Structural Analysis of Bacterial Glycoconjugates

A thesis submitted to
the **Department of Life Sciences**, **Imperial College London**
for the **Degree of Doctor of Philosophy**

Tiandi Yang

Hypotheses non fingo

by Isaac Newton

"I have not as yet been able to discover the reason for these properties of gravity from phenomena, and I do not feign hypotheses. For whatever is not deduced from the phenomena must be called a hypothesis; and hypotheses, whether metaphysical or physical, or based on occult qualities, or mechanical, have no place in experimental philosophy. In this philosophy particular propositions are inferred from the phenomena, and afterwards rendered general by induction."

Declaration of Originality

I declare hereby that all the works included in this thesis have not been preciously or concurrently submitted for any other qualifications, such as degree and diploma, to any other university or institute. All the contents in this thesis are the result of my own independent research unless otherwise indicated.

Tiandi Yang

Declaration of Copyright

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence.

Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work

Tiandi Yang

Acknowledgement

First of all, I want to deeply thank my parents for funding my PhD research. More importantly, all of my academic achievement, if there is any, should be ultimately attributed to my mothers' educational effect upon my young heart when I was a little child.

I would like to express my tremendous gratitude to my supervisors, Prof. Anne Dell and Dr. Stuart M. Haslam for their great helps in all respects of my research, and also Prof. Barbara Mulloy for her helps in NMR research and many inspiring discussions.

I would also like to thank all my previous and current colleagues, Dr. Aristotelis Antonopolous, Laura Bouche, Dr. Qiushi Chen, Dr. Paola Grassi, Dr. Paul Hitchen, Dr. Nan Jia, Dongli Lu, Dr. Simon North, Dr. Poh-choo Pang, Dr. Maria Panico, Dinah Rahman, Dr. Grigorji Sutov, Dr. Valeria V. Ventura and Dr. Gang Wu. You have provided great helps and suggestions for my research.

It was a great honour that I have got opportunites to make many friends in London, and I would hereby like to appeciate Jin Wang, Li Li, Min Kuang, Qian Du, Wenqi Shi and Yifan Yu.

We should also thank our collaborators: Prof. Barry Marshall, Dr. Hong Li and Dr. Mohammed Benghezal from the University of Western Australia; Dr. Alan R Brown and Dr. Matthew Robinson from the University of Exeter; Fan Zhu, Dr. Hua Zhang and Prof. Hui Wu from the University of Alabama; Prof. Brendan W. Wren, Dr. Carmen Denman and Dr. Esmeralda Valiente from the London School of Hygiene and Tropical Medicine.

In addition, appreciations should be given to Dr. Jan Marchant and Dr. Yingqi Xu from the Cross-Faculty NMR center, Imperial College London, for help setting up NMR experiments.

Abstract

Bacterial glycoconjugates are a group of highly complex macromolecules that are deeply implicated in various biological processes. They can be crudely divided into several subgroups mainly including bacterial glycolipids and glycoproteins. Their structural complexity once hindered the elucidation of the biological roles they are playing. Based on the development of modern chromatography, mass spectrometry and nuclear magnetic resonance, research described in this thesis analysed the structures of multiple glycoconjugates such as glycolipids and glycoproteins.

Lipopolysaccharides are characteristic bacterial glycolipids. The *Pseudomonas aeruginosa* lipid A and *Helicobacter pylori* lipopolysaccharide was investigated by a combination of chemical, MS and/or NMR strategies. For the lipid A, it was proved to be modified by 4-amino-arabinose, which is associated with bacterial antibiotic resistance. For the lipopolysaccharide, the specificities of mutiple putative glycosyltransferases, including HP1284, HP1283, HP1578, HP0102 and WaaL ligase, involved in its biosynthesis were characterised. The structural data suggest *Helicobacter pylori* WaaL ligase has a relaxed specificity *in vivo* that transfers an un-expected long O-antigen to a short core-oligosaccharide. In addition, our data shed light on the regulation of bacterial O-antigen biosynthesis.

The heavily glycosylated fimbrial-associated protein 1 of *Streptococcus parasanguinis* was subjected to in-depth glycomics and glycoproteomics, which helps determine the specificities of multiple putative glycosyltransferases such as GALT1/2 and GLY. More specifically, GALT1 was proved to be a bi-functional glycosyltransferase that transfers a glucose and an N-acetyl-glucosamine. GALT2 and GLY was deteremined to transfer a glucose and a rhamnose, respectively. Based on these conclusions, our study helps elucidate the glycosylation pathway of fimbrial-associated protein 1.

The recent characterisation of bacterial oligosaccharyltransferases has the potential to revolutionise the production of glycoconjugate vaccines. The putative O-linked oligosaccharyltransferases system in *Burkholderia thailandensis* was investigated based on a zwitterionic hydrophilic interaction liquid chromatography mass spectrometry method, which helps expand the toolbox of glyco-engineering. By using a simpilied liquid chromatography-mass spectrometry method, multiple glycoproteins in *Coxiella burnetii* were semi-identified.

During these studies, many classical chemistries were resurrected to inspire novel chemical strategies and facilitate the instrumental analysis. We hope the methods included in this thesis can contribute to further research where comprehensive structural characterisations of complex bacterial glycoconjugates are required.

Publications

Zhang, H., Zhu, F., **Yang, T.**, Ding, L., Zhou, M., Li J., Haslam, S. M., Dell, A., Erlandsen, H., and Wu, H., The highly conserved domain of unknown function 1792 has a distinct glycosyltransferase fold. ***Nat. Commun.* 5**, 4339 (2014)

Zhang, H., Zhou, M., **Yang, T.**, Haslam, S. M., Dell, A., and Wu, H., A new helical binding domain mediates a glycosyltransferase activity of a bifunctional protein. ***J. Biol. Chem.*** Published on-line.

Zhu, F., Zhang, H., **Yang, T.**, Haslam, S. M., Dell, A., Wu, H., Engineering and dissecting the glycosylation pathway of a *Streptococcal* serine-rich repeat adhesin, under revision after submitted to ***J. Biol. Chem.***

Li, H.†, **Yang, T.**†, Liao, T.†, Nilsson, H., Fulurija, A., Debowski, A., Haslam, S. M., Molloy, B., Dell, A., Stubbs, K. A., Marshall B. J., and Benghezal, M., The Core Domain of *Helicobacter pylori* Lipopolysaccharide is Unexpectedly Short but Critical for Colonisation, under revision after submitted to ***PLoS pathog.*** (†Authors contributed equally)

Abbreviation List

A. pleuropneumoniae	*Actinobacillus pleuropneumoniae*
ABC	adenosine triphosphate binding cassette
ADP	adenosine diphosphate
AMBIC	ammonium bicarbonate
AQ	2-alkyl-4-quinolone
Ara4N	4-amino-L-arabinose
ATP	adenosine triphosphate
B. cenocepacia	*Burkholderia cenocepacia*
B. thailandensis	*Burkholderia thailandensis*
BacdiNAc	diacetamido bacillosamine
C. burnetii	*Coxiella burnetii*
C. jejuni	*Campylobacter jejuni*
CAD	collisionally activated dissociation
CAMP	cationic antimicrobial peptide
CI	chemical ionisation
CID	collision-induced dissociation
CMP	cytidine monophosphate
CPS	capsular polysaccharides
CRM	charged residue model
Ct	C-terminal
D,D-Hep	D-glycero-D-manno-Hep
DC	direct current
DC-SIGN	dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin
DNA	deoxyribonucleic acid
DPC	dodecyl phosphocholine
DTT	dithiothreitol
DUF	domain of unknown function
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic
EI	electron ionisation

ESI	electrospray ionization
ETD	electron transfer dissociation
FAB	fast atom bombardment
FD	field desorption
FI	field ionisation
FID	free induction decay
Fru	fructose
Fuc	fucose
FWHM	full width at half maximum
Gal	galactose
GalT	galactosyltransferase
GC	gas chromatography
GDP	guanosine diphosphate
Glc	glucose
GlcN	glucosamine
GlcNAc	N-acetylglucosamine
GlcNAc	N-acetyl glucosamine
GlcT	glucosyltransferase
GT	glycosyltransferase
H. influenzae	*Haemophilus influenzae*
H. pylori	*Helicobacter pylori*
Hep	heptoses
HepT	heptosyltransferase
Hex	hexose
HexN	hexosamine
HexNAc	N-acetyl hexosamine
HSQC	heteronuclear single quantum coherence
IAA	iodoacetic acid
IEM	ion evaporation model
IL-6	interleukin-6
KDO	2-keto-3-deoxyoctanoic acid

KO	D-glycero-a-D-talo-oct-2-ulopyranosylonate
L,D-Hep	L-glycero-D-manno-Hep
LC	liquid chromatography
LD	laser desorption
LOS	lipooligosaccharide
LPS	lipopolysaccharides
MALDI	matrix-assisted laser desorption/ionization
Man	mannose
ManT	mannosyltransferases
MS	mass spectrometry
MD-2	lymphocyte antigen 96
NF-κB	nuclear factor kappa-light-chain-enhancer of activated B cells
NGT	N-glycosyltransferase
NMD	nucleotide-binding domain
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser effect spectroscopy
Nt	N-terminal
Otase	oligosaccharyltransferase
P. aeruginosa	*Pseudomonas aeruginosa*
PAO1	*P. aeruginosa* serotype O5
PCR	polymerase chain reaction
PEtN	phosphoethanolamine
PHB	poly-3-hydroxybutyrate
PICI	positive ion chemical ionisation
PMT	protein mannosyltransferases
ppm	part per million
Q-TOF	quadrupole time-of-flight
QIT	quadrupole ion trap
QMS	quadrupole mass filter
RF	radio frequency
Rha	rhamnose
Rib	ribose

RNA	ribonucleic acid
RRKM	Rice-Ramsperger-Kassel-Marcus
RT	retention time
S. enterica	*Salmonella enterica*
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SS1	Strain Sydney 1
TCS	two-component systems
TIR	Toll/interleukin-1 receptor
TLR4	Toll-like receptor 4
TMD	transmembrane domain
TMS	trimethylsilyl
TOCSY	totally correlation spectroscopy
TOF	Time-of-flight
UDP	uridine diphosphate
Und-P	undecaprenyl phosphate
V. cholerae	*Vibrio cholerae*
ZIC-HILIC	zwitterion–ion hydrophilic interaction chromatography

Contents

Index of Figures

Index of Table

Chapter 1 Introduction

This chapter provides physical, chemical, biochemical and biological backgrounds for this thesis. It is divided into four independent sections, including fundamental knowledge for mono- and poly-saccharides (**Section 1.1**), bacterial glycoconjugates (**Section 1.2**), Mass Spectrometry (MS, **Section 1.3**) and Nuclear Magnetic Resonance (NMR, **Section 1.4**). Due to the highly collaborative nature of research described in this thesis, there are biological implications, such as host immunity elicited by bacterial glycoconjugates, which cannot be fully covered in this chapter. Relevant biological background material is explained and discussed in a case-specific manner in the following chapters.

1.1 Carbohydrates

Carbohydrates are the most abundant naturally occurring organic chemicals. They are ketones (ketose) or aldehydes (aldose) that carry multiple hydroxyl groups. Though ketoses are involved in the metabolism of carbohydrates, only the aldoses are discussed throughout this thesis (except for KDO that is a ketose, when terms such as monosaccharides are used, they are referring to aldoses). They often share a $C_m(H_2O)_n$ empirical formula, which is the origin of the term "carbohydrate", hydrates of carbon. The term "glycan" has been widely adopted to denote carbohydrate oligomers and is used throughout this thesis.

The most biologically relevant monosaccharide is arguably glucose (Glc) that is widely known as an energy source for the majority of creatures including human beings. The initial effort to interpret the pathway of glycolysis was made in the 1850s, when the field of biochemistry just began to emerge. However, the wide acceptance of biological functions other than energy sources, mediated by oligo/polysaccharides and glycoconjugates, occurred much later, which was partially due to their structural complexity.

Research described in this thesis focuses on structural glycobiology with an emphasis on MS- and NMR- based analyses of bacterial glycoconjugates. Prior to going deeper into the biological implications of bacterial glycoconjugates and biophysical methods used for their structural investigations, the first section of this thesis introduces monosaccharides and oligo/polysaccharides. The biogenesis of these carbohydrates, especially the role of glycosyltransferases is introduced and their structural complexity is explained in order to validate the necessity of structural glycobiology from chemical and biochemical points of view.

1.1.1 Monosaccharides

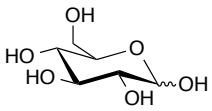
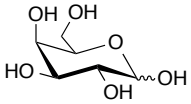
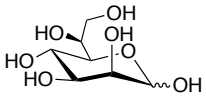
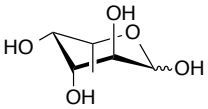
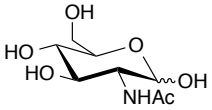
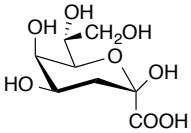
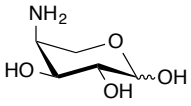
Monosaccharides are the simplest carbohydrates that cannot be further hydrolysed into smaller sugar units. Their classification is normally based on the numbers of carbon atoms they possess, for instance, monosaccharides having five and six carbons are called pentose and hexose, respectively. The sugar units found in deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are pentoses; and our most familiar sugar, Glc, is a hexose. There are many different kinds of monosaccharides (see **Section 1.1.1.1**), and a large number of them can be found in living organisms. A complete summary of all biology-related monosaccharides is not realistic for the sake of this thesis, as bacterial glycans can be potentially composed of any monosaccharide. The names, abbreviations and chemical structures of a few selected monosaccharides that are relevant to this thesis are listed in **Table 1-1**. Because the chemical formulas and structures of monosaccharides are often too complicated and inconvenient to be used for their illustration, a symbolic system has been established to denote monosaccharides (see last column of **Table 1-1**).

1.1.1.1 Stereochemistry of Monosaccharides

The complexity of monosaccharides is due to their stereochemistry. Pentoses are taken as examples to introduce the stereochemistry of monosaccharides. A monosaccharide normally carries N minus two chiral centres, where N equals to its carbon number. Hence, pentose has three chiral centres and can exist in eight (2^3) different stereoisomers that are four pairs of enantiomers.

These stereoisomers can be presented by the Fischer projections, which are shown in **FIG. 1-1.** The Fischer projection is advantageous mainly on account of its straightforward showing D- and L-configurations that are defined relative to glyceraldehyde. Briefly, the monosaccharide has the D-configuration when the hydroxyl group on the carbon next to the CH_2OH group is present on the right in its Fischer projection (dashed squares on D- and L- ribose in **FIG. 1-1**). This carbon is also called the configurational carbon. In addition, the carbon atoms of pentoses are numbered from 1, the carbon of the aldehyde group, to 5, the carbon of the CH_2OH group, which can also be easily shown by the Fischer projections (blue numbers on D-arabinose in **FIG. 1-1**). The configurational carbon in pentoses is carbon 4.

Table 1-1 Table of Selected Monosaccharides

Monosaccharide	Abbreviation	Chemical Structure	Symbol
D-Glucose	Glc		
D-Galactose	Gal		
L,D-Heptose	Hep		
L-Fucose	Fuc		
N-acetyl-Glucosamine	GlcNAc		
3-deoxy-D-manno-oct-2-Ulosonic acid	KDO		
4-amino-4-deoxy-L-Arabinose	Ara4N		N/A

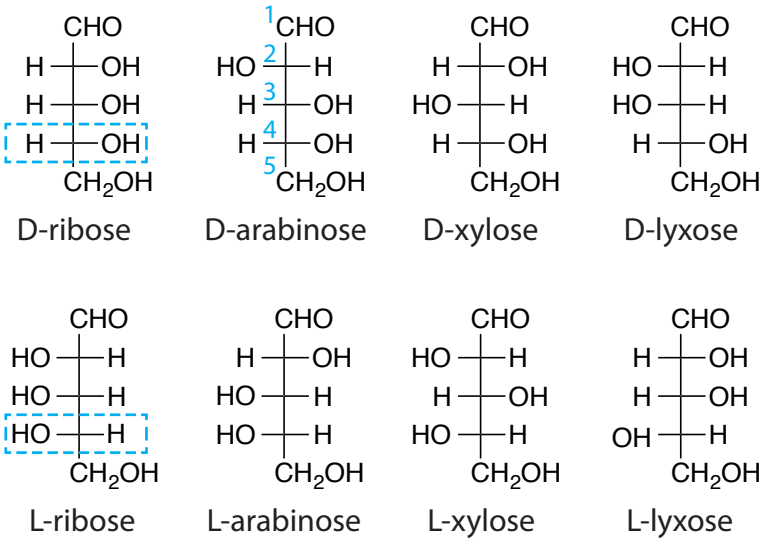


Figure 1-1 Fischer Projections of pentoses. The configurational carbon of pentoses is exemplified by D- and L-ribose (dashed squares), and the numbering of carbon in pentoses is represented on D-arabinose.

Monosaccharides can exist in linear and more commonly ring structures. During the formation of the ring, more than one hydroxyl groups can serve as the nucleophile, forming pyranoses (six-membered ring) or furanoses (five-membered ring). The ring-formed monosaccharides can be conveniently presented as Haworth projections. The Haworth projections of D-arabinose are shown in **FIG. 1-2**, from which, pyranoses and furanoses can be readily identified. Moreover, a new anomeric centre is created during the ring formation, which gives rise to α- and β-configurations. These two configurations are determined by comparing the anomeric and configurational carbons, *e.g.* if they have the same D- or L- configuration, the monosaccharide is an α-anomer.

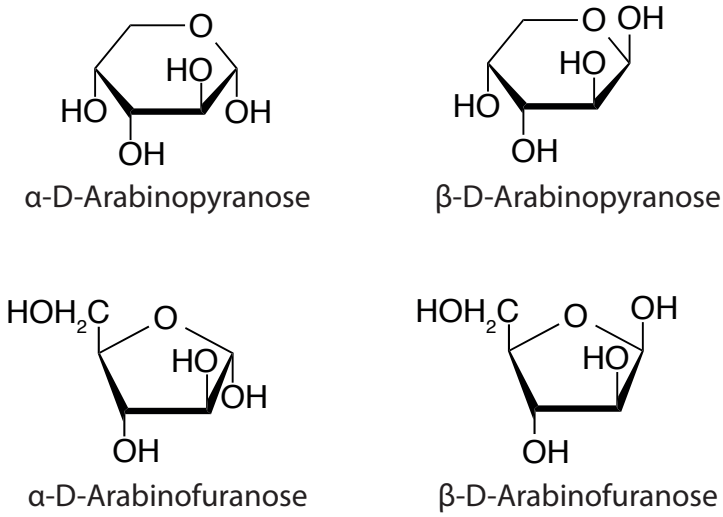


Figure 1-2 Haworth Projections of D-arabinose.

The complex stereochemistry of monosaccharides is one of the major challenges for the structural investigation of glycans, especially MS-based glycomics that is widely used in this thesis. Most information provided by MS is directly related to the molecular weights of ions (see **Section 1.3**), which means very little, if any, stereochemical information can be easily extracted from the mass spectra (there are exceptions, see **Section 4.2.3**). Chromatography-MS has been a powerful tool to unambiguously identify monosaccharides when monosaccharide standards are available, as epimers and even derivatised enantiomers can be separated by chromatography and detected by MS. However, some monosaccharide standards can be extremely difficult to synthesise. As a result, NMR (see **Section 1.4**) has to be used in some instances to complement MS (see **Section 3.0**).

1.1.1.2 Biosynthesis of Monosaccharides

Though chemical synthesis of monosaccharides can be highly tedious and complicated, they are "routinely synthesised" in cells. Glc, for example, can be biosynthesised by prokaryotes via photosynthesis and chemosynthesis, and by eukaryotes via gluconeogenesis. Glc can be further synthesised into literally all kinds of biologically occurring monosaccharides including hexose (Hex), hexosamine (HexN) and N-acetyl hexosamine (HexNAc) and so forth.

Biosynthesis of monosaccharides is closely related to their metabolism, and therefore very complicated. This section aims to briefly review important biochemical processes involved in the biosynthesis of monosaccharides that are relevant to my thesis research. More detailed information on the metabolism of monosaccharides can be found in the KEGG database (http://www.genome.jp/kegg/).

Though two main metabolic pathways introduced in this section are shown to be initiated with Glc (**pathways 0** in **FIG. 1-3** and **1-4**), this sugar, along with all the other monosaccharides, is not a good substrate for most enzymes involved in monosaccharide metabolism. Instead each is normally first phosphorylated to a monosaccharide-phosphate (Pi) by kinases (such as glk for Glc) and converted to a sugar nucleotide by a nucleotidyltransferase (an uridyltransferases UGP2 for Glc) before further reactions (**FIG. 1-3**, **reaction 1**). Many D-monosaccharides are assembled into glycans by living organisms, therefore most monosaccharide abbreviations mean D-monosaccharide in this thesis, except for Fuc and Rha stand for L-fucose and L-rhamnose.

Monosaccharides can be converted into one another under the catalysis of isomerases or epimerases. For instance, fructose-6-Pi (Fru-6-Pi) can be directly isomerised from Glc-6-Pi by an isomerase GPI. Note that the term isomerisation is used here because Fru is a ketose. The conversion of some hexoses can be slightly more complicated, *i.e.* they need to be synthesised into sugar-nucleotides first and then epimerised into other sugar-nucleotides by corresponding epimerases. For instance, galE converts uridine diphosphate Glc (UDP-Glc) to UDP-galactose (UDP-Gal, **FIG. 1-3**, **reaction 2**).

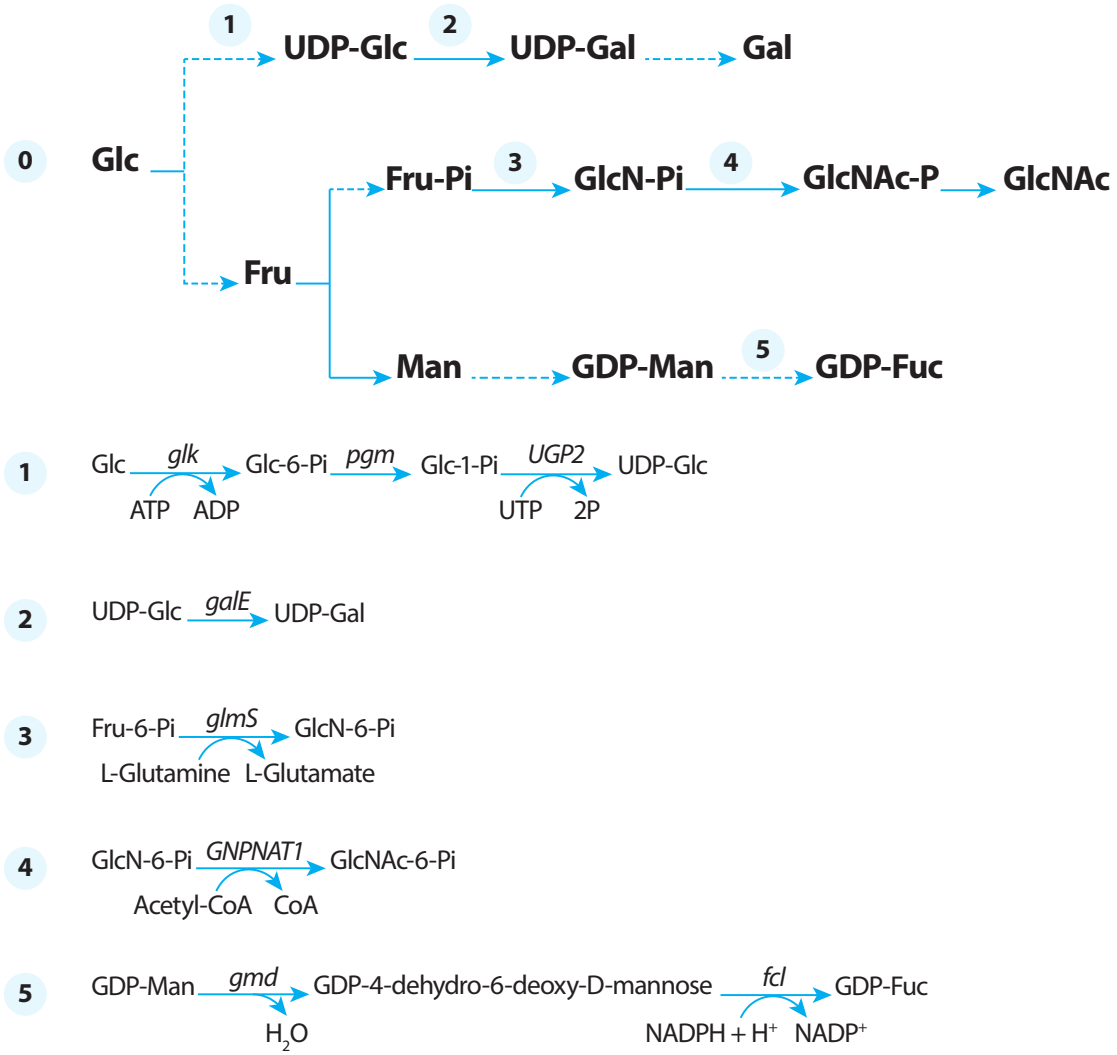


Figure 1-3 Metabolism of hexoses, HexNAcs and deoxy-hexoses. General biosynthesis of hexose (Hex, exemplified by Gal), HexN (exemplified by GlcN), HexNAc (exemplified by GlcNAc) and deoxy-Hex (exemplified by Fuc) is shown in Pathway 0. Filled lines indicate reactions are finished within one step, and dashed lines indicate the reactions require multiple steps. Five characteristic types of reactions are summarised in reactions 1-5, which are examples of the assembly of monosaccharide-nucleoside (1), epimerisation (2), transamination (3), acetylation (4) and elimination-reduction (5) respectively. The corresponding genes/enzymes and other compounds involved in the reactions are annotated above and below the arrows respectively. Note that the genes/enzymes shown here are not complete, and it is possible that other genes/enzymes have the same activities. The shown pathways are extracted from the KEGG database.

Fru is a very important intermediate along various metabolic pathways of monosaccharides. An amino group can be transferred by an aminotransferase (glmS) from L-glutamine on to Fru-6-Pi, forming glucosamine-6-Pi (GlcN-6-Pi, **FIG. 1-3**, **reaction 3**), which can be further acetylated by an acetyltransferase (GNPANT1) to generate GlcNAc-6-Pi (**FIG. 1-3**, **reaction 4**). Fru can also be converted into mannose (Man), and guanosine diphosphate Man (GDP-Man) can be synthesised into GDP-fucose (GDP-Fuc), which bridges the hexose and deoxy-hexose (**FIG. 1-3**, **reaction 5**). In order to do this, GDP-Man is first dehydrated by a dehydratase (gmd) and reduced by an oxidoreductase (fcl).

Pentoses and heptoses (Heps, carbohydrates with seven carbons) can be synthesised from Glc via the pentose phosphate pathway (partially shown in **FIG. 1-4**, **pathway 0**). In addition, intermediates in this pathway are significant for the biosynthesis of monosaccharides that compose bacterial lipopolysaccharides (LPS, see **Section 1.2.1**). Glc-6-Pi is oxidised by a dehydrogenase G6PD to form a lactone, and subsequently hydrolysed and ring-opened by a lactonase (PGLS). The produced carboxylic group is further oxidised by another dehydrogenase (PGD), resulting in gluconate-6-Pi losing a carbon by releasing a carbon dioxide. The final product of this series of oxidations is Rib-5-Pi, which is a key intermediate linking pentoses and heptoses to hexoses.

For pentose synthesis, ribulose-5-Pi is epimerised into xylulose-5-Pi by an epimerase rpe and isomerised into ribose-5-Pi (Rib-5-Pi) by an isomerase (rpiA). The latter product can be directly hydrolysed into Rib.

For heptose biosynthesis, xylulose-5-Pi and Rib-5-Pi react with each other under the catalysis of a transketolase (tktA). During the reaction, one carbon is translocated from Rib-5-Pi to Xylulose-5-Pi, forming seduheptulose-7-P. Seduheptulose-7-Pi is the precursor for L-glycero-D-manno-Hep (L,D-Hep) and D-glycero-D-manno-Hep (D,D-Hep) that are found in inner cores of bacterial LPSs (see **Section 1.2.1.2)**. Their biosynthesis in Gram-negative bacteria involves a remarkable process that isomerises D,D-Hep-7-Pi to D,D-Hep-1-Pi by the addition and elimination of Pi rather than isomerisation (**FIG. 1-4**, **pathway 1**, Kneidinger *et al.* 2002). Many enzymes are involved in this process, and interestingly the genes encoding these enzymes can be clustered on the chromosome in some bacteria, such as *Helicobacter pylori* (*H. pylori*, Raetz & Whitfield 2002). Fundamentally, sedoheptulose-7-Pi is isomerised by GmhA and phosphorylated by HldE, forming a di-phosphorylated intermediate that is hydrolysed at the 7-position phosphate, generating adenosine diphosphate (ADP) -D,D-Hep that can be converted into ADP-L,D-Hep by an epimerase HldD.

Furthermore, Rib-5-Pi can be synthesised into 2-keto-3-deoxyoctanoic acid (KDO, **FIG. 1-4**, **pathway 2**). KDO is an octo-carbon acidic monosaccharide found in bacterial and plant glycoconjugates, and its biosynthesis is almost constantly conserved through prokaryotes to eukaryotes. Rib-5-Pi is first isomerised by KdsD, forming Ara-5-Pi that reacts with phosphoenol pyruvate under the catalysis of a KDO-8-Pi synthase (KdsA) to generate KDO-8-Pi. KDO-8-Pi is hydrolysed by a phosphatase (KdsC) to produce KDO (Smyth & Marchant 2013). Cytidine monophosphate-KDO (CMP-KDO) assembled from KDO can be directly used in the Raetz Pathway (see **Section 1.2.1.1**).

In addition to the different types of monosaccharides introduced above, special monosaccharide derivatives can be incorporated into bacterial glycoconjugates. Glycoconjugates are biological macromolecules that are formed by covalently linking glycans to other biochemical

compounds such as lipids and proteins. Some bacterial glycoconjugates such as LPS can be modified by monosaccharide derivatives. An excellent example for monosaccharide derivatives is 4-amino-L-arabinose (Ara4N, note the arabinose here is in L-configuration), which modifies lipid A molecules, increasing antibiotic resistance of bacteria (see **Section 1.2.1.1** and **3.1.1**). Biosynthesis of Ara4N is not really derived from the pentose phosphate pathway. However, as both of these two pathways involve an extensive use of oxidases, they are summarised in a same figure. It is apparent from the **FIG. 1-4**, Pathway 3 that shows the biosynthesis of Ara4N, several oxidative steps catalysed by oxidases UgD and ArnA convert UDP-Glc to UDP-L-threo-pentapyranos-4-ulose, onto which an amino group can be transferred by an aminotransferase (ArnB), forming Ara4N (Breazeale *et al.* 2003).

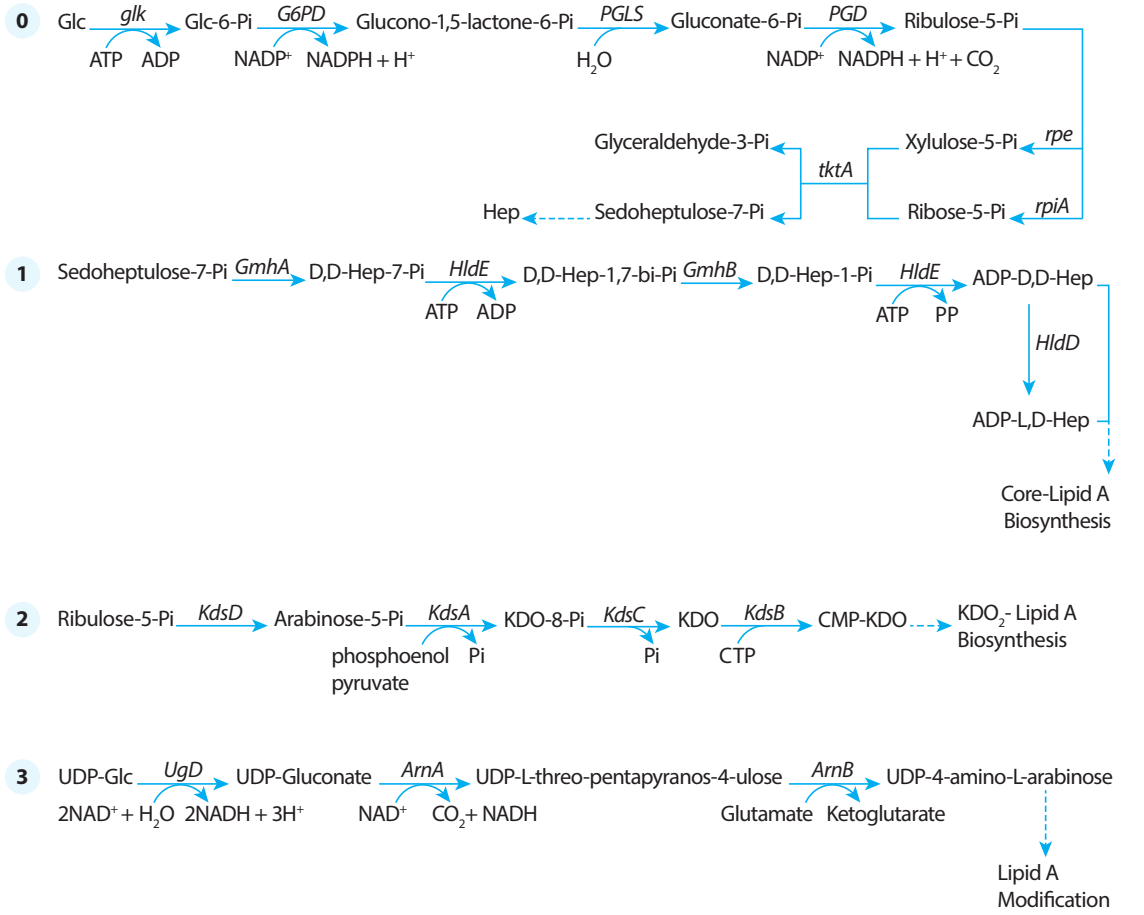


Figure 1-4 Biosynthesis of pentose, heptose, acidic and special sugars. The pentose phosphate pathway is partially shown in the Pathway 0. Biosynthesis of L,D- and D,D-Hep is shown in the Pathway 1. Biosynthesis of acidic sugars (exemplified by KDO) and special sugars (exemplified by Ara4N) are shown in the Pathway 2 and 3 respectively. More detailed descriptions of this figure can be found in the legend of **FIG. 1-3**.

1.1.2 Oligo/Polysaccharides

Monosaccharides can be successively assembled into oligo/polysaccharides by glycosyltransferases (GTs). Oligosaccharides are oligomers of monosaccharides and polysaccharides are the polymers. The classification of oligomer and polymer is vague and subjective, though polysaccharides normally have more than ten monosaccharide units. During the assembly process, a glycosidic bond formed between the hemiacetal carbon of one monosaccharide and a hydroxyl group of another monosaccharide. Normally, a hemiacetal is conserved on one end of the oligo/polysaccharides, which is called the reducing end. Based on this reducing end, the sequences of monosaccharides can be defined. In addition, another structural feature, linkage, is introduced for oligo/polysaccharides with the formation of the glycosidic bond. This section introduces the activities of GTs and their structural consequences, sequences and linkages of monosaccharides.

1.1.2.1 Glycosyltransferases

It is obvious from their names that GTs are enzymes transferring sugars. Indeed, GTs transfer monosaccharides from sugar nucleotides, sugar phosphates and their lipid carriers to monosaccharide acceptors ranging from small molecules to marcomolecules such as glycans, proteins and lipids, thereby generating new glycosidic bonds.

There are currently 98 families of GTs listed in the CAZy database (Feb. of 2016), whereas only two GT folds have been firmly established, GT-A and GT-B (Lairson *et al*. 2008). A characteristic GT-A folding includes the Asp-X-Asp (DXD) motif coordinating a divalent metal ion. In addition, the GT-A architecture is composed of two adjacent Rossmann-like folds that can be found in nucleotide-binding proteins. With regards to the GT-B fold, it also possesses two Rossmann-like folds facing each other, but they are more flexibly linked in the GT-B fold compared with the GT-A fold.

A GT can be classified as a retaining or inverting GT based on its catalytic outcome, as the α- or β-configurations of monosaccharides can either be retained and inverted during their transfer. The different catalytic outcomes are in fact determined by two different mechanisms (**FIG. 1-5**) that the GTs adopt to transfer monosaccharides.

A characteristic S_N2-like mechanism was proposed for the inverting GTs (Kozmon & Tvaroška 2006). Briefly, the hydroxyl group of the acceptor is semi-deprotonated by the carboxyl group on the side chain, which facilitates its nucleophilic attack to the anomeric carbon of the donor (**FIG. 1-5 a**). An oxocarbenium ion-like transition state is formed before the elimination of the nucleoside leaving group, generating a configurationally inverted glycosidic bond.

A hypothetical double-displacement mechanism (**FIG. 1-5 b**) was proposed for retaining GT activities (Koshland 1953), yet little evidence has been found to support it (Lairson *et al.* 2008; Persson *et al.* 2001). This mechanism requires a covalently linked glycosyl-enzyme intermediate formed by a nucleophilic addition of the carboxyl group side chain on the catalytic centre to the sugar donor, and an elimination of the nucleoside leaving group of the donor. The eliminated nucleoside group helps activate the hydroxyl group of the sugar acceptor, which nucleophilically attacks the anomeric carbon in the glycosyl-enzyme intermediate, forming a new glycosidic bond. During the whole processes, two nucleophilic attacks happen, inverting the anomeric configuration twice. Therefore, the overall reaction retains the configuration.

Generally speaking, each GT has a very strong specificity, *i.e.* it recognises both sugar donor and acceptor. For this reason, a sequence of an oligo/polysaccharide can be determined during its template-free biosynthesis. For instance, during the biosynthesis of a trisaccharide with a sequence (see **Section 1.1.2.2**) of Glc-Gal-Man, a galactosyltransferase (GalT) recognises the Man and transfers a Gal to the Man, generating a Gal-Man disaccharide before the activity of a glucosyltransferase (GlcT) that recognises Gal and transfers a Glc. Thus, sequential activities of multiple GTs essentially guarantee the sequences of oligo/polysaccharides in principle.

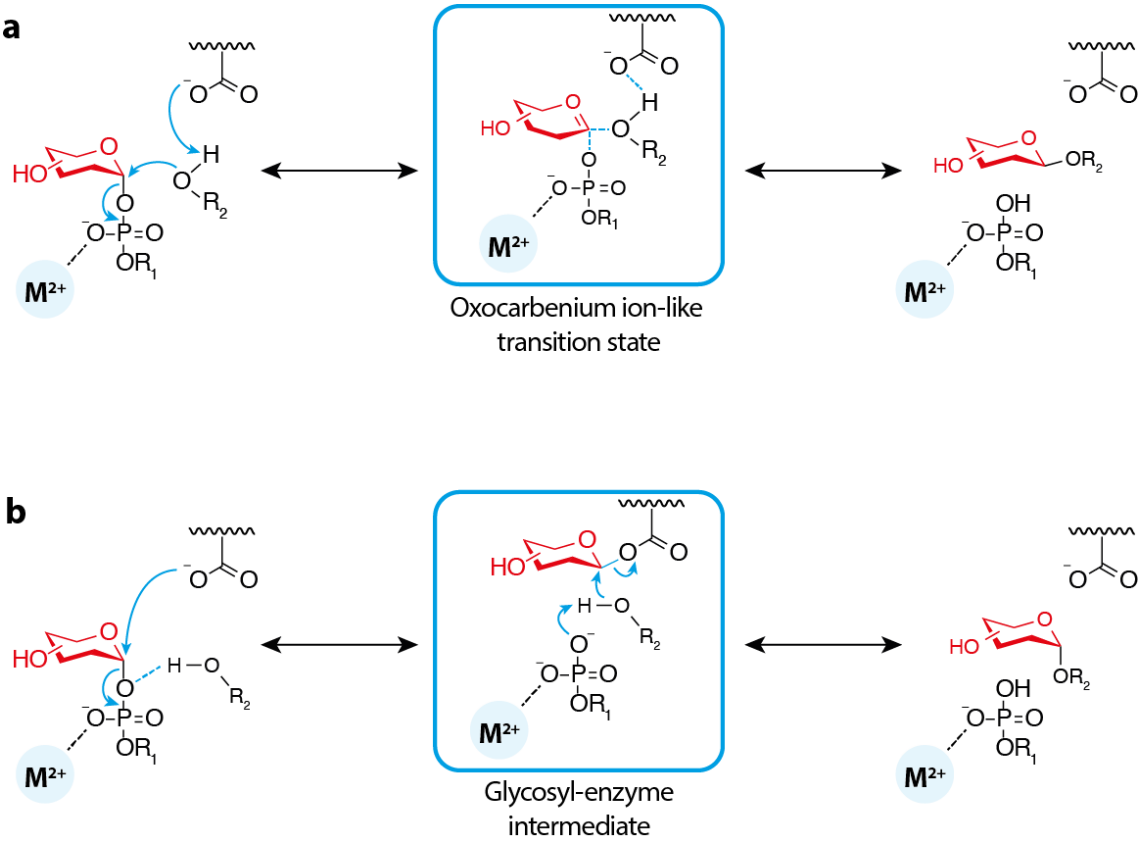


Figure 1-5 Proposed mechanisms for inverting (a) and retaining (b) GTs. The two intermediates are shown in the blue squares. The red boat shaped molecules represent the monosaccharide residues, and the blue arrows represent electron transfer. The M^{2+} stands for a bivalent metal ion. The carboxylic groups linking to the waved lines indicates they are on the GTs. Adapted from (Lairson *et al.* 2008).

1.1.2.2 Sequences and Linkages

When monosaccharides are assembled into oligosaccharides and polysaccharides, new structural features, sequences and linkages, emerge. This section takes several oligosaccharides and polysaccharides (**FIG. 1-6**) as examples to introduce these features.

Each glycan has a sequence. It was mentioned that oligo/polysaccharides normally have a hemiacetal group at one end that is called the reducing end. The other end (or ends) of the glycan are called the non-reducing end(s). The sequences of glycans are determined based on the reducing and non-reducing ends and written in such a way that the non-reducing end(s) are on the left and the reducing end is on the right. For instance, lactose (β-Gal-β-1,4-Glc, **FIG. 1-6 a**) has a sequence of Gal-Glc, indicating Glc is at the reducing end and Gal is at the non-reducing end. A sugar (β-Glc-β-1,4-Gal, **FIG. 1-6 b**) with a sequence of Glc-Gal cannot be called lactose because the sequence is different, though both of them are made up with Glc and Gal.

Glycosidic bonds in a glycan are defined based on their linkages. Each monosaccharide has more than one hydroxyl group, and all of them, including the one within the hemiacetal group, can be potential sites to form new glycosidic bonds. As a result, the linkage of each glycosidic bond needs to be determined in order to uniquely describe a glycan. For instance, β-maltose (β-Glc-α-1,4-Glc **FIG. 1-6 c**) and the β-Glc-α-1,6-Glc disaccharide (**FIG. 1-6 d**) share the same Glc reducing end. However, the non-reducing Glc is linked at different positions, *i.e.* position 4 in β-maltose and position 6 in the disaccharide. Therefore, they are structurally different. In addition, α and β configurations of monosaccharides also need to be considered. The β-maltose and cellobiose (β-Glc-β-1,4-Glc, **FIG. 1-6 e**) are different as the anomeric configurations of their reducing end Glc are different from each other.

All the rules introduced for the disaccharides above also work for more complicated oligosaccharides and even polysaccharides. One of the most familiar polysaccharides is starch that is basically poly-Glc molecules, and we are taking it as an example here. Fundamentally, starch is composed of two kinds of poly-Glc that are called amylose and amylopectin. Amylose (**FIG. 1-6 f**) is a simple linear polymer that comprises α-1,4-linked glucose, whereas amylopectin (**FIG. 1-6 g**) is the most structurally complicated glycan that is introduced by far in this thesis. Being similar to amylose, amylopectin also consists of α-1,4-linked glucose chains. However, these chains are arranged in a highly branched manner. The branches connect to the α-1,4-linked glucose chains with α-1,6 glycosidic bonds, which occur every 24 to 30 glucose units. Amylopectin serves as an excellent example of branched polysaccharides where monosaccharides simultaneously linking to more than two other monosaccharides can be found.

As a summary, this section has introduced the structural features and biogenesis of mono-, oligo- and polysaccharides. It is clear that glycans are structurally complex molecules. In biological

systems, glycans are often coupled with other biological macromolecules such as proteins and lipids, forming complex glycoconjugates such as those described in the following sections.

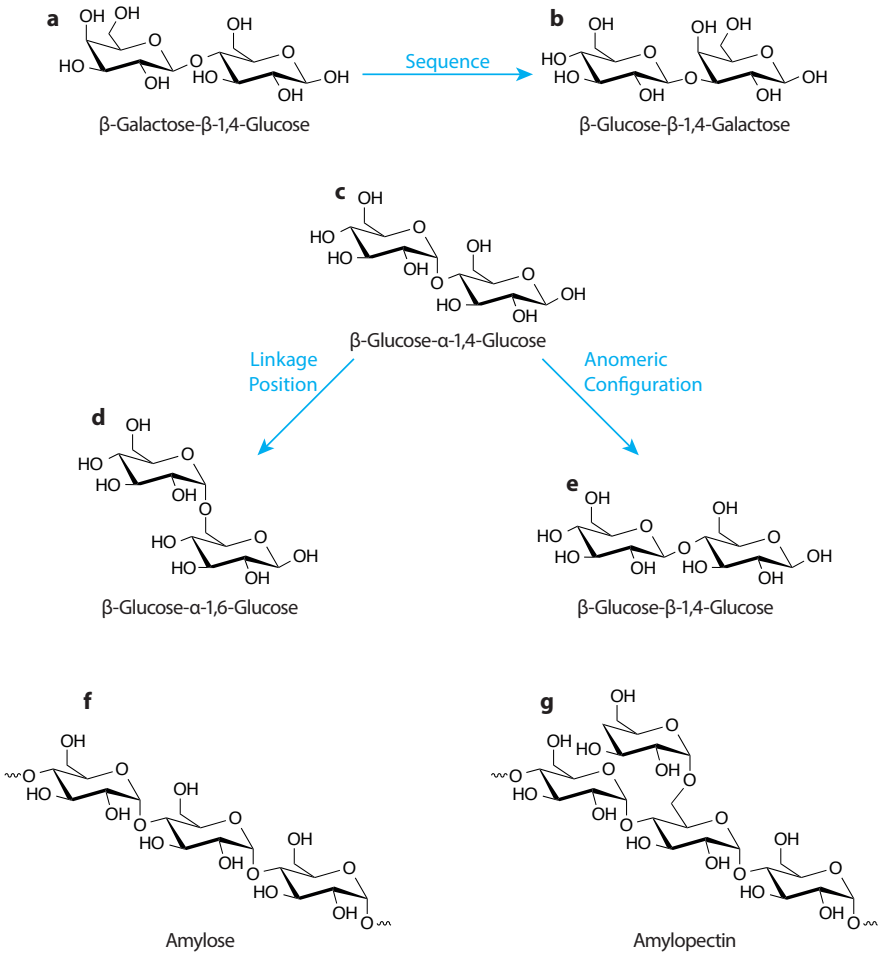


Figure 1-6 Examples of oligosaccharides and polysaccharides. The blue arrows and texts indicate the structural differences between relevant disaccharides.

1.2 Bacterial Glycoconjugates

It is long known that bacteria make use of glycans to assemble variable glycoconjugates, which can be exemplified by peptidoglycans, capsular polysaccharides (CPS) and LPS. Peptidoglycan is the main component of the bacterial cell wall. It is made by attaching β-1,4-linked N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid to short peptide chains. CPS are homo- or hetero- polymers composed of various monosaccharides. One of its best examples is the CPS of Gram-positive *Streptococcus pneumonia*, which protects the organism from host immune responses, allowing the organism to colonise its hosts (Hyams *et al*. 2010). LPS are glycolipids formed by covalently linking polysaccharide chains to lipids. It acts as the outermost layer of the characteristic two-membrane systems of Gram-negative bacteria. The first part of this section will focus on LPS, which provides a general background for **Chapter 3**.

In addition to the classical bacterial glycoconjugates above, bacterial proteins can also be modified by mono- and polysaccharides, generating glycoproteins. Glycosylation of bacterial proteins is a relatively more recent research topic compared with other bacterial glycoconjugates such as LPS. This is because it was long believed that only eukaryotes glycosylate their proteins. However, it is now accepted that both N-linked and O-linked bacterial glycoproteins are widespread (Iwashkiw *et al*. 2013; Nothaft & Szymanski 2010). This knowledge has largely arisen from the usage of genomic sequencing and analysis (Szymanski *et al*. 1999; Linton *et al*. 2002). Bacterial glycoproteins are beginning to be implicated in biomedical research and human health (De Gregorio & Rappuoli 2014). For instance, the characterisation of bacterial glycosylation systems is facilitating the development of glycoprotein vaccines against multiple human pathogens (Wacker *et al*. 2002; Madhi *et al*. 2013; Lin *et al*. 2001). As **Chapter 4** and **5** in this thesis involve glycoproteomic investigations of bacterial glycoproteins, the second part of this section introduces protein glycosylation systems in bacteria.

1.2.1 Lipopolysaccharides

Research on LPS has been entangled with bacteriology since its nascent age. In 1892, Richard Pfeiffer, a student of Robert Koch, who is well known as the founder of modern bacteriology, discovered that an injection of heat-killed *Vibrio cholerae* (*V. cholerae*) into animals still leads to fever and shock (Beutler 2003). It was realised that there must be toxins enclosed in bacterial cell walls, which are released after the lysis of bacteria and cause the symptoms in the animal models. He, therefore, named the causative agent as "endotoxins". However, it was not possible for him to chemically define "endotoxins" during his career. It took other several decades to reveal that endotoxins are glycolipids, and a more chemical term "lipopolysaccharide" is given to the endotoxins.

Subsequent research has diverged into two categories: first, to further characterise structures and biosynthesis of LPS in multiple bacterial models; and secondly, to seek for immune mechanisms underlying the symptoms caused by LPS in animal models.

For the structural and biosynthetic aspect, it is now understood that LPS has three main components: the lipid A molecule that anchors the whole LPS to bacterial outer membranes; the core-oligosaccharide extending on the lipid A and the O-antigen stretching out of the core-lipid A. The general biosynthesis of lipid A and O-antigens has been very well studied and documented for model bacteria such as *Escherichia coli* (*E. coli*) and *Salmonella* (Raetz & Whitfield 2002; Whitfield & Trent 2014). For the immunity aspect, Toll-Like Receptors (TLRs; Ip *et al.* 1993, Lemaitre *et al.* 1996) have been identified as the main receptor for LPS and the elicited cellular signalling has been in-depth characterised, if not completely elucidated (Takeda *et al.* 2003; Beutler 2003).

It is reasonable to conclude that tremendous success has been made in both categories, especially for the second endeavour, which has directly given rise to the 2011 Nobel prize in Physiology and Medicine (Poltorak, He, *et al.* 1998; Poltorak, Smirnova, *et al.* 1998; Takeda *et al.* 2003; Beutler 2003; Lemaitre *et al.* 1996). This section introduces all three components of LPS mostly from a structural and biosynthetic point of view, though immune responses against it will be briefly included.

1.2.1.1 Lipid A

The structure and biosynthesis of lipid A are highly conserved throughout different Gram-negative bacteria (Raetz & Whitfield 2002). The general architecture (**FIG. 1-7**) of Lipid A is an acylated and amidated 1-6 linked di-glucosamine (di-GlcN) that is further phosphorylated.

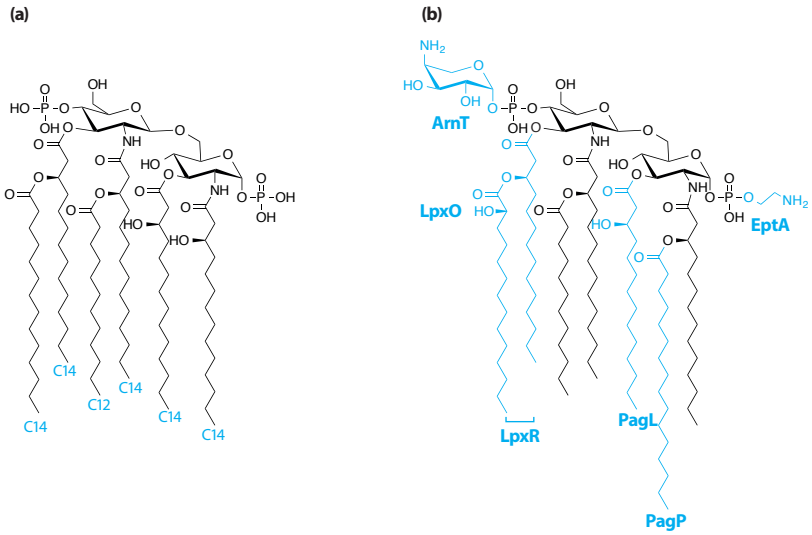


Figure 1-7 Chemical structure of Lipid A. (a) The classical structure of lipid A. (b) Potential modifications on lipid A. ArnT transfer 4-amino-arabinos, EptA transfers phosphoethanolamine (PEtH), PagP transfers palmitate to the 2 position of lipid A, LpxR and PagL hydrolyse lipid chains on the 3 and 3' position respectively and LpxO oxidises the secondary acyl chain at 3' position. The latter three genes cannot be found in *E. coli* strain K12.

The conserved biosynthesis of KDO_2-lipid A is shown in **FIG. 1-8** (Raetz & Whitfield 2002). This pathway is best understood in *E. coli* and is largely based on the work of Christian R. H. Raetz, and therefore it is often referred to as the Raetz pathway. It begins with the acylation of UDP N-acetyl glucosamine (UDP-GlcNAc) at its 3-position (Anderson & Raetz 1987). The N-acetyl group of the produced 3-acylated UDP-GlcNAc is irreversibly hydrolysed by LpxC, forming 3-acylated UDP-glucosamine (GlcN; Anderson *et al.* 1988, Young *et al.* 1995). It is N-acylated by LpxD (T. M. Kelly *et al.* 1993), generating 2,3-diacyl UDP-GlcN whose UDP group is subsequently hydrolysed by LpxH (Babinski *et al.* 2002). The hydrolysis product (2,3-diacyl GlcN-Pi) reacts with its precursor (3-O-acylated UDP-GlcNAc) under the catalysis of LpxB to produce an acylated diGlcN-Pi (Crowell *et al.* 1986; Radika & Raetz 1988). This product is phosphorylated by a kinase, LpxK, forming an important intermediate, lipid IV_A (Garrett *et al.* 1997). Two KDOs are transferred to the lipid IV_A by WaaA from cytidine monophosphate KDO (CMP-KDO) donors (Clementz & Raetz 1991), and more lipids are transferred by LpxL and LpxM, eventually synthesising the KDO_2-lipid A (Brozek & Raetz 1990).

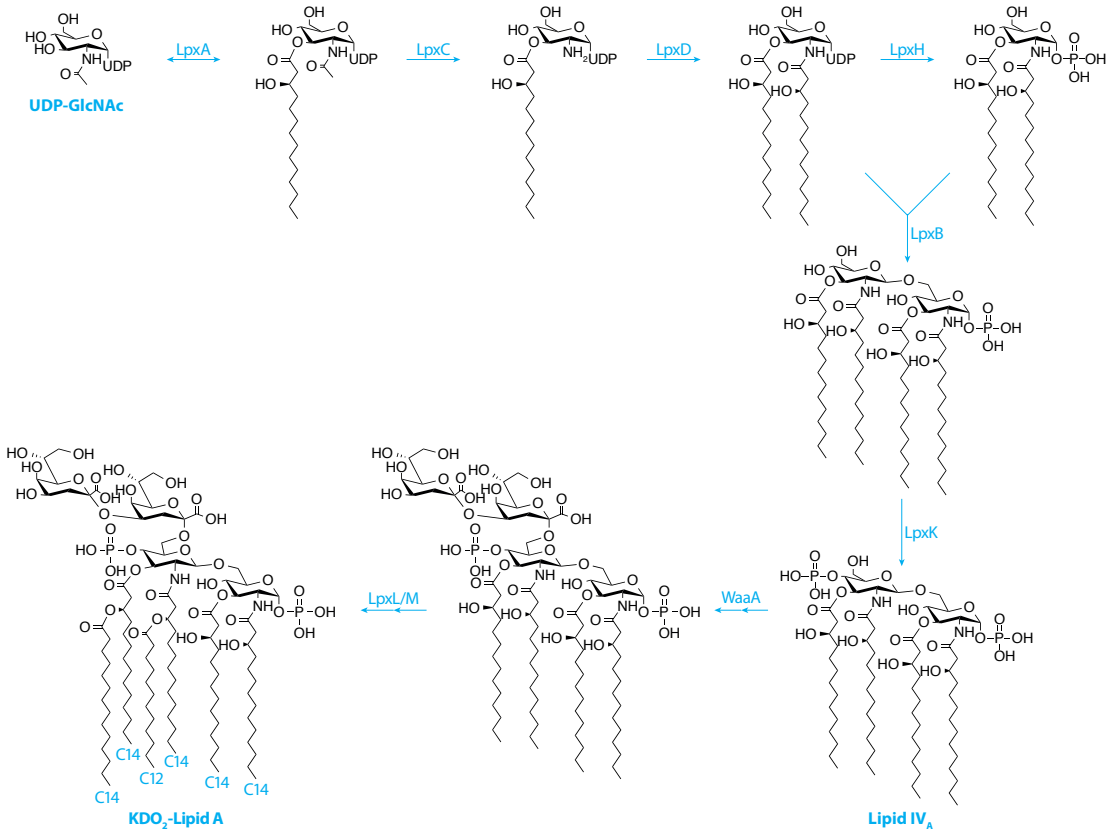


Figure 1-8 The Raetz pathway, conserved biosynthesis of Kdo_2-lipid A. Adapted from (Whitfield & Trent 2014; Raetz & Whitfield 2002).

The length of the lipid chains in lipid A is delicately regulated (Wyckoff *et al.* 1998). For instance, the catalytic centre of LpxA that catalyses the first acetylation process locates in the middle of two adjacent domains that incorporate different lipids at different rates. As a result, only favoured lipids can be transferred to UDP-GlcNAc efficiently, *e.g.* a 14-carbon acyl chain is preferred in *E. coli*, whereas a 10-carbon acyl chain is preferred in *Pseudomonas aeruginosa* (*P. aeruginosa*). In addition, more lipid transferases are encoded in bacterial genomes and their expressions are also regulated. For example, a gene *lpxP* similar to *lpxL* (see the last step in **FIG. 1-8**) can be found in *E. coli*., which encodes an enzyme that transfers an un-saturated 16-carbon acyl chain (Carty *et al.* 1999). Its expression is sensitive to temperature, thus adjusting the lipid length in lipid A in a temperature-dependent manner (Carty *et al.* 1999).

Lipid A stimulates severe inflammatory responses, which is the main reason why LPS is called endotoxin. The Toll-like receptor 4 (TLR4) serves as the main receptor of lipid A in mammals (Lemaitre *et al.* 1996). It is the first and best characterised among all mammalian TLRs. TLR4 has two domains, an intra-cellular Toll/interleukin-1 receptor (TIR) domain (Rock *et al.* 1998) and an extracellular domain (Park *et al.* 2009). The latter can interact with the lymphocyte antigen 96 (MD-2) molecule forming a TLR4-MD-2 interface. When the lipid A binds to the TLR4-MD-2 complex, most of its lipid chains are buried in the hydrophobic pocket in MD-2, leaving a single lipid chain exposed on the surface. This exposed lipid interacts with a second TLR4, contributing to its dimerization (Park *et al.* 2009; Shimazu *et al.* 1999). In addition, the interaction between the two phosphate groups and cationic amino acids also plays an important role in TLR dimerization (Rietschel *et al.* 1994). The dimerization initiates the TLR4 signalling that eventually leads to the activation of nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) that transcribes pro-inflammatory cytokines (Ip *et al.* 1993; Lemaitre *et al.* 1996). As a summary, the lipid chains and negatively charged phosphate groups are the main molecular pattern recognized by TLR4 that leads to inflammation active signalling.

Bacteria have evolved pathways to modify their lipid A. For some bacteria, lipid A modifications are a response to their growth conditions that can be sensed by their two-component systems (TCSs, see **Section 3.1.1**; Guo *et al.* 1998, Gunn, Lim, *et al.* 1998). As a result, despite the conservation of the overall biosynthesis and structure of lipid A, different modification pathways yield minor lipid A diversity (Raetz *et al.* 2007).

Potential modifications on lipid A are illustrated in **FIG. 1-7 b**. Several enzymes have been characterised to be involved in modifying the lipid chains. PagP transfers palmitate to a different position (Bishop *et al.* 2000). In *Salmonella,* PagL (Trent, Pabich, *et al.* 2001) can hydrolyse lipid on the 3 position, and LpxO (Gibbons *et al.* 2000) can hydroxylate lipid on 3' positions. In addition to the modifications on lipid chains, the phosphate groups can be capped by other functional groups including Ara4N and phosphoethanolamine (PEtN). They are transferred by ArnT (Breazeale *et al.* 2003) and EptA (Stead *et al.* 2010; Kim *et al.* 2006), respectively.

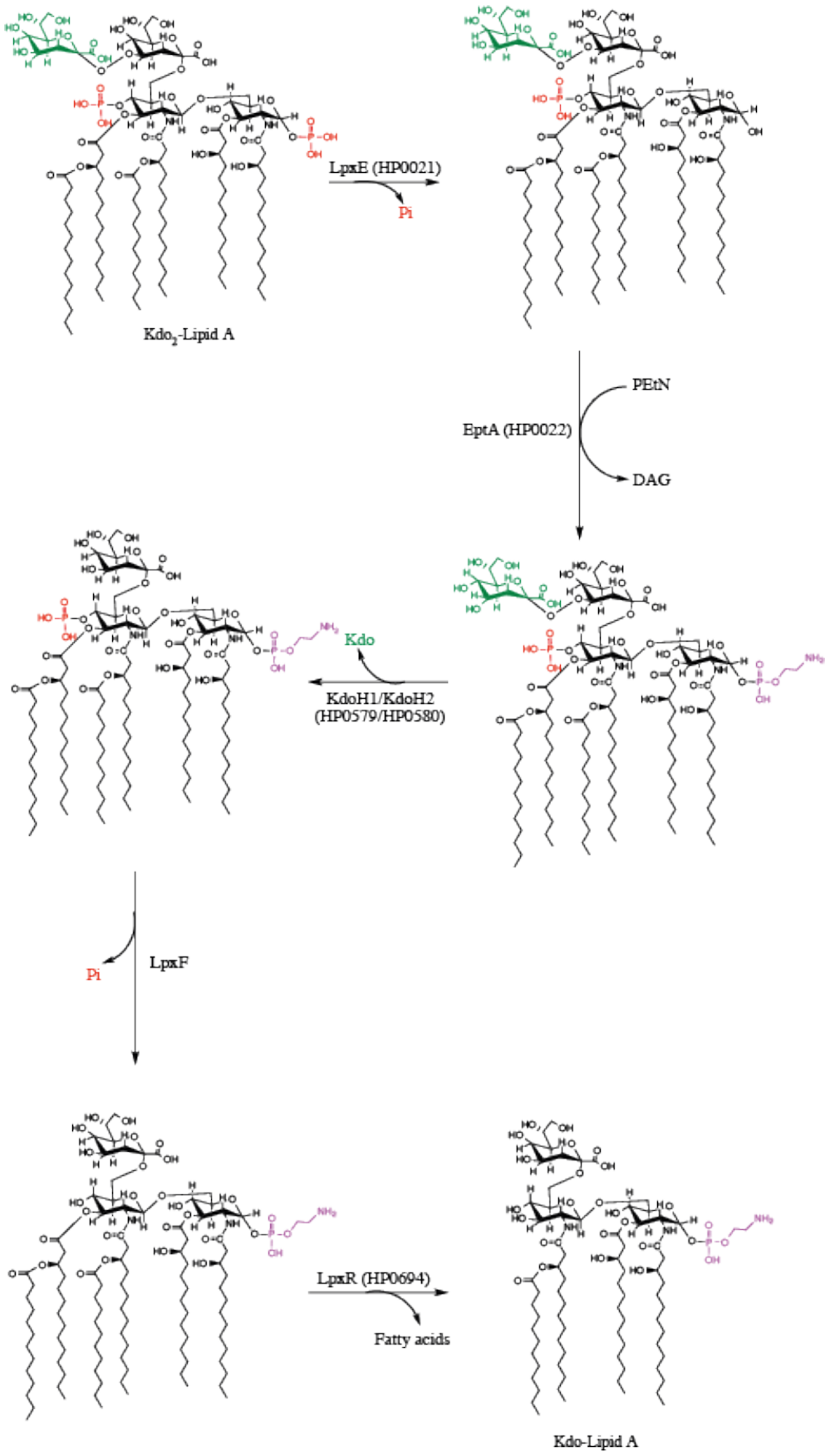


Figure 1-9 Modification of KDO_2-lipid A in *H. pylori* strain 26695. Adapted from (Stead *et al.* 2010).

The modifications of lipid A can benefit bacteria by the evasion of host immunity, which can be exemplified by *H. pylori* lipid A (see **Section 3.2.1**). The fundamental architecture of *H. pylori* lipid A is a di-GlcN acylated by four (two fewer than *E. coli* and *Salmonella*) lipid chains, and it is further modified by PEtN instead of negatively charged phosphate groups (Stead *et al.* 2010). Its modification pathway is shown in **FIG. 1-9**. As previously mentioned, the final product of the Raetz pathway is KDO_2-lipid A modified by two phosphate groups (see above). One of them (red, on position 1) is first removed and replaced by PEtN under the activities of LpxE (HP0021) and EtpA (HP0022), respectively. HP0579/HP0580 hydrolyses the outer KDO (green) from KDO_2-lipid A, and the remaining phosphate group (red, on the 4'-position) is removed by LpxR (HP0694). Finally, two fatty acid chains attaching on the 3' position get eliminated, giving rise to PEtN modified KDO-lipid A with two fewer lipid chains.

The modified *H. pylori* lipid A significantly reduces the dimerization of TLR4, thus suppressing the inflammatory activity in the host. Fewer (four) lipid chains and the deletion of a phosphate group has been proved to be capable of reducing the inflammatory activity 100-fold, respectively (Teghanemt *et al.* 2005; Rietschel *et al.* 1994). In addition, the replacement of negatively charged phosphates by PEtN can further benefits the organism by blocking the binding of cationic antimicrobial peptides (CAMPs) in the host (Cullen *et al.* 2011). This kind of modification has also been implicated in the resistance of human CAMPs attack in other pathogens such as *V. cholerae* (Matson *et al.* 2010; Clements *et al.* 2007).

1.2.1.2 Core-oligosaccharides

The core-oligosaccharides bridge the lipid As and O-antigens. LPSs can terminate with core-oligosaccharides due to mutations of certain enzymes involved in LPS synthesis, and they are often called rough LPSs (or lipooligosaccharide, LOS) as opposite to smooth LPSs that are terminated with O-antigens. The core-oligosaccharide can be further divided into two parts: the inner and outer cores, in spite of their classification being somewhat arbitrary. Notably, the KDO transferred to the lipid IV_A during lipid A biosynthesis (see **Section 1.2.1.1**) is considered as a part of the inner-core, which means that the assembly of the core-oligosaccharide initiates before the completion of lipid A biosynthesis.

Like the lipid A, the structure and biosynthesis of the core-oligosaccharide have been best characterised in *E. coli* and *Salmonella* (E. V. Vinogradov *et al.* 1999; Kaniuk *et al.* 2004; Muller-Loennies *et al.* 2003; Heinrichs *et al.* 1998; K. Amor *et al.* 2000; Olsthoorn *et al.* 1998), although more structural data are being gradually collected for more organisms (Bystrova *et al.* 2004; Lukasiewicz *et al.* 2006; Lukasiewicz *et al.* 2009; Altman *et al.* 2013). Bacteria normally make use of characteristic non-human monosaccharides including Hep and KDO to assemble the main chains of core-oligosaccharides. The term "non-human" is used as Hep and KDO cannot be synthesised by human beings, which means that an inhibition of their metabolism can potentially produce novel antibiotics. Notably, KDO and its derivatives such as

D-glycero-D-talo-octulosonic acid (KO) are the only monosaccharides found in all known LPS structures, reflecting the conserved nature of the Raetz pathway (Raetz & Whitfield 2002). The biosynthesis of the building blocks of the core-oligosaccharides is discussed in **Section 1.1.1.2**, and therefore this section only focuses on the assembly and modification of the core-oligosaccharide.

The core-oligosaccharide is less conserved than lipid A. At least two different though very similar core structures were found in *Salmonella* (Olsthoorn *et al.* 1998), and no less than five were discovered in *E. coli* (E. V. Vinogradov *et al.* 1999; Kaniuk *et al.* 2004; Heinrichs *et al.* 1998; Muller-Loennies *et al.* 2003). Generally speaking, the outer-core is slightly more diverse than the inner-core. The five documented core-oligosaccharides found in *E. coli* are shown in **FIG. 1-10**. It is fairly obvious that the inner-cores are very similar to each other and mainly composed of heptoses and KDOs with phosphate and PEtN modifications, whereas the structures of the outer-cores tend to be random.

The core oligosaccharide is directly assembled on KDO_2-lipid A (Whitfield & Trent 2014), which is mediated by sequential activities of multiple glycosyltransferases (GTs). Taking *E. coli* R1 for example, the involved enzymes are shown in **FIG. 1-11** (Raetz & Whitfield 2002). In addition to the GTs assembling the oligosaccharide backbone (Heinrichs *et al.* 1998), WaaP and WaaY modifies it with PEtN and phosphate groups (Yethon *et al.* 1998). The mutation of the heptosyltransferase (HepT) transferring Hep in the inner-core region leads to the deep-rough phenotype (Parker *et al.* 1992). These mutants normally show a very hydrophobic surface property and they become very sensitive to hydrophobic compounds such as fatty acids. This may indicate the importance of LPS as a major component of the bacterial outer membrane.

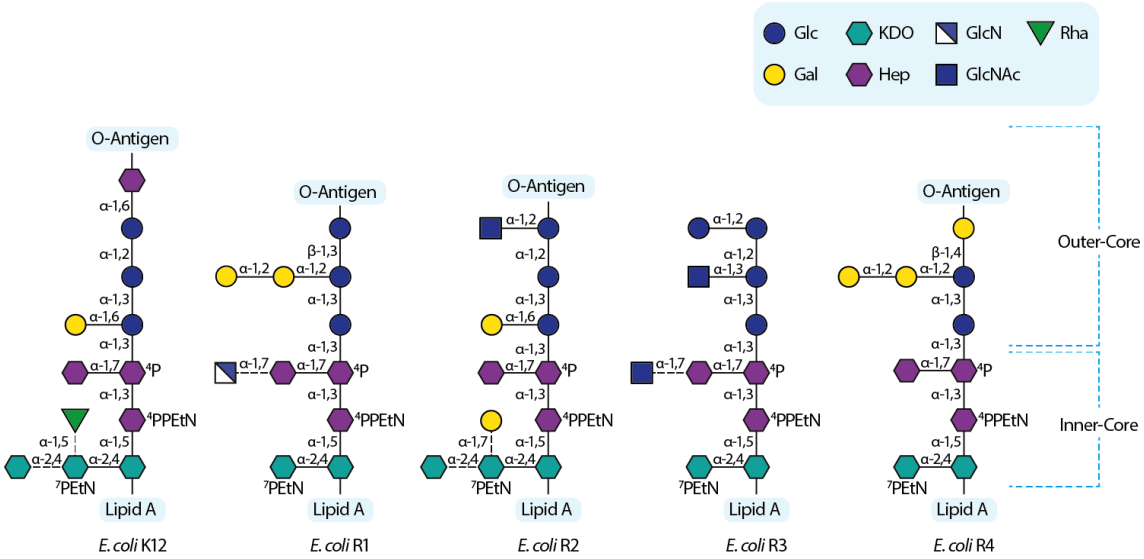


Figure 1-10 Core-oligosaccharide region of *E. coli*. Five documented different core structures are shown. Modifications such as phosphate (Pi) and phosphoethanolamine (PEtN) are directly written near the involved monosaccharides. The nonstoichiometric modifications are drawn with dash lines. The inner- and outer- core regions are illustrated at the right of the structures. Adapted from (Raetz & Whitfield 2002).

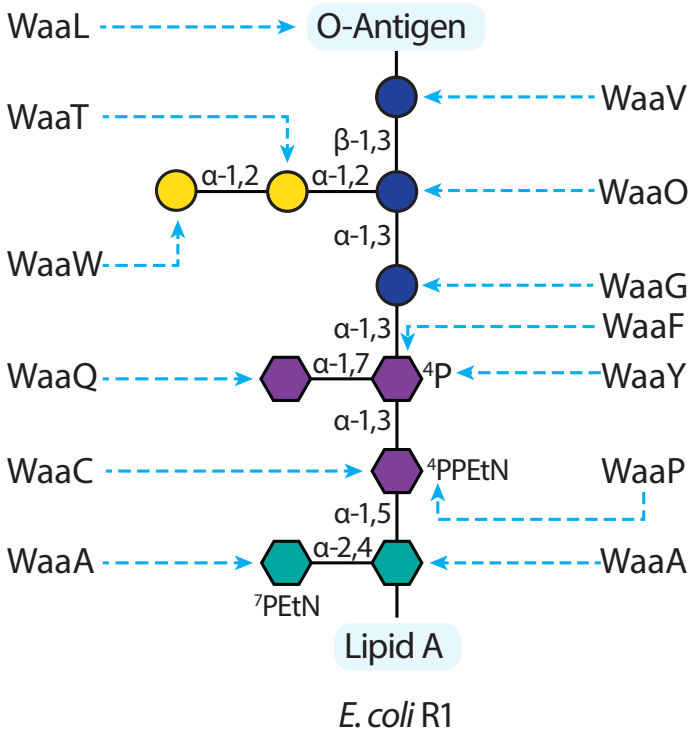


Figure 1-11 Structure and biosynthesis of Core-oligosaccharide region of *E. coli* R1. The involved transferases are annotated near the monosaccharides and functional groups. Adapted from (Raetz & Whitfield 2002).

It is normally a good strategy to initiate genomic investigations before attempting structural elucidation of the core-oligosaccharides, because genomic analysis has revealed that the GTs shown in **FIG. 1-11** can be conserved among different bacteria (Raetz & Whitfield 2002), and structural analysis of core-oligosaccharides can be sometimes highly challenging. However, it is worth emphasising that predicted homology does not necessarily mean the same function. Therefore, structural analysis of the core-oligosaccharide is still essential for scrutinising the proposed gene functions and ultimately determining the proposed structures. As this thesis resolves the structures of LPS based on MS and NMR methods, genomic aspects of core-oligosaccharides are not further discussed here and more detailed information can be found in review articles (Raetz & Whitfield 2002) and textbooks such as ***Endotoxin in Health and Disease*** (edited by Helmut Brade *et al.*).

The biosynthesis of KDO_2-lipid A and assembly of core-oligosaccharide occurs at the cytoplasmic surface of inner-membrane (Whitfield & Trent 2014). Following core-lipid A biosynthesis, it is transported to the periplasmic surface. The detailed mechanism is described in **Section 1.2.1.4** but first, the O-antigens are introduced.

1.2.1.3 O-antigens

O-antigens are the most diverse part of LPS. Different organisms and even different strains or serotypes can carry distinct O-antigens. Sometimes structures of O-antigens are used as diagnostic markers specific to bacterial serotypes. Therefore, a simple summarisation of different structures of O-antigens is not meaningful and this section mainly focuses on biosynthesis of O-antigens.

Biosynthesis of O-antigens is completely parallel to core-lipid A (Whitfield & Trent 2014), and all known O-antigens are synthesised via one of three pathways which are the Wzy-dependent, the adenosine triphosphate (ATP) binding cassette (ABC)-transporter-dependent and the synthase-dependent pathways (Raetz & Whitfield 2002). The latter one is very rare, in fact only one known LPS O-antigen is synthesised via this pathway (Keenleyside & Whitfield 1995; Keenleyside & Whitfield 1996). This section introduces the Wzy- and ABC-transporter-dependent pathways for LPS O-antigens.

Both of these pathways share the same initial reaction. This reaction involves transferring a monosaccharide phosphate from a nucleotide monosaccharide to an undecaprenyl phosphate (Und-P) carrier, which is catalysed by WecA in *E. coli* (Alexander & Valvano 1994) and WbaP in *Salmonella* (L. Wang *et al.* 1996). WecA is a GlcNAc-1-Pi transferase first characterised in the Wzy-dependent pathway of enterobacterial common antigen biosynthesis (Meier-Dieter *et al.* 1992). It is a conserved enzyme that can be found in different organisms in addition to *E. coli*, such as *Shigella* (Yao & Valvano 1994) and *H. pylori* (Hug *et al.* 2010). WecA shows a relaxed specificity *e.g.* it transfers GalNAc-1-Pi in some *E. coli* strains (P. A. Amor & Whitfield 1997). WbaP reversibly transfers a Gal-1-Pi to Und-P and was first characterised in the Wzy-dependent pathway for biosynthesising O-antigens in *Salmonella* (L. Wang *et al.* 1996).

The Wzy-dependent pathway has been best studied in *Salmonella enterica* (*S. enterica*) type B (**FIG. 1-12 a**) and branched LPSs are normally synthesised via this pathway. The O-antigen of *S. enterica* group E1 has a Man-rhamnose-Gal (Man-Rha-Gal) tri-saccharide repeating unit that is assembled onto the Und-P carrier by sequential activities of GTs (D. Liu *et al.* 1993). Each of the Und-PP linked tri-saccharides is subsequently transported from the cytoplasmic side of the inner membrane to the periplasmic polymerisation site by a so-called "flippase" Wzx (D. Liu *et al.* 1996). The Und-PP linked tri-saccharides then are polymerised by a polymerase Wzy that transfers the previously assembled tri-saccharide polymer from its Und-PP carrier to the non-reducing end of the newly transported tri-saccharide unit, thus elongating the O-antigen (**FIG. 1-12 a**, Robbins *et al.* 1967). The released Und-PP carrier is probably hydrolysed to Und-P for recycling. A Wzz enzyme is the last component of the Wzy-dependent pathway (Batchelor *et al.* 1991). It acts as a "regulator" of the chain length of O-antigens. It is proposed that this enzyme interacts with Wzy and WaaL ligase (see **Section 1.2.1.4**),

deciding whether certain O-antigens should be further elongated or terminated and transferred to the core-lipid A. Some organisms such as *P. aeruginosa* have two chromosomal *wzz* genes (Rocchetta *et al.* 1999), which is indicative of its importance.

The Wzy-dependent pathway is readily identifiable based on simple genomic analysis. Multiple components of this pathway are highly conserved through different organisms. Wzx and Wzz and their homologs can be found in all gene clusters encoding this pathway. Moreover, the fact that the Wzy mutant carries LPS with only one O-antigen (Man-Rha-Gal) unit is another defining feature of this pathway.

The Wzy-dependent pathway for synthesizing LPS is remarkably similar to the oligosaccharyl transferase pathway (OTase) for glycosylating bacterial proteins (Iwashkiw *et al.* 2013). The OTases are normally annotated as an O-antigen polymerase in databases, and some bacterial glycoproteins carry their O-antigen units.

The second popular pathway for synthesising O-antigens is the ABC-transporter-dependent pathway (**FIG. 1-12 b**). O-antigens synthesised via this pathway tend to be linear as opposed to the Wzy-dependent pathway. All known ABC-transporter-dependent pathways initiate with the WecA activity that transfers a GlcNAc-Pi to an Und-P carrier (Raetz & Whitfield 2002; Rick *et al.* 1994). The O-antigen is assembled onto the Und-PP linked GlcNAc primer at the cytoplasmic surface of the inner-membrane, and the fully synthesised Und-PP linked O-antigen is transported to the periplasmic surface of the inner-membrane by an ABC-transporter instead of the Wzx "flippase" (Reizer *et al.* 1992). Therefore, no ubiquitous polymerase is required for the ABC-transporter-dependent pathway, and the transporter functions only once for each LPS molecule.

FIG. 1-12 b shows the ABC-transporter dependent pathway in *E. coli* O8 that expresses mannan as its O-antigen. As mentioned, the mannan O-antigen is directly assembled on the Und-PP linked GlcNAc primer by multiple mannosyltransferases (ManT) that are involved in this process. Though most of the mannan is synthesised by WbdA and WbdB, the first Man is transferred to the primer by a distinct ManT, WbdC (Kido *et al.* 1995; Greenfield *et al.* 2012). Therefore, the first Man is considered as an adaptor linking the initial GlcNAc and repeating O-antigen unit. Similar monosaccharide or oligosaccharide adaptors have also been observed in many other O-antigens synthesised via the ABC-transporter dependent pathway, which indicates that the transfer of the adaptor is constitutional to this pathway. The mannan O-antigen is subsequently *en bloc* transported by the ABC-transporter to the periplasmic space where it is ligated to core-lipid A (see **Section 1.2.1.4**).

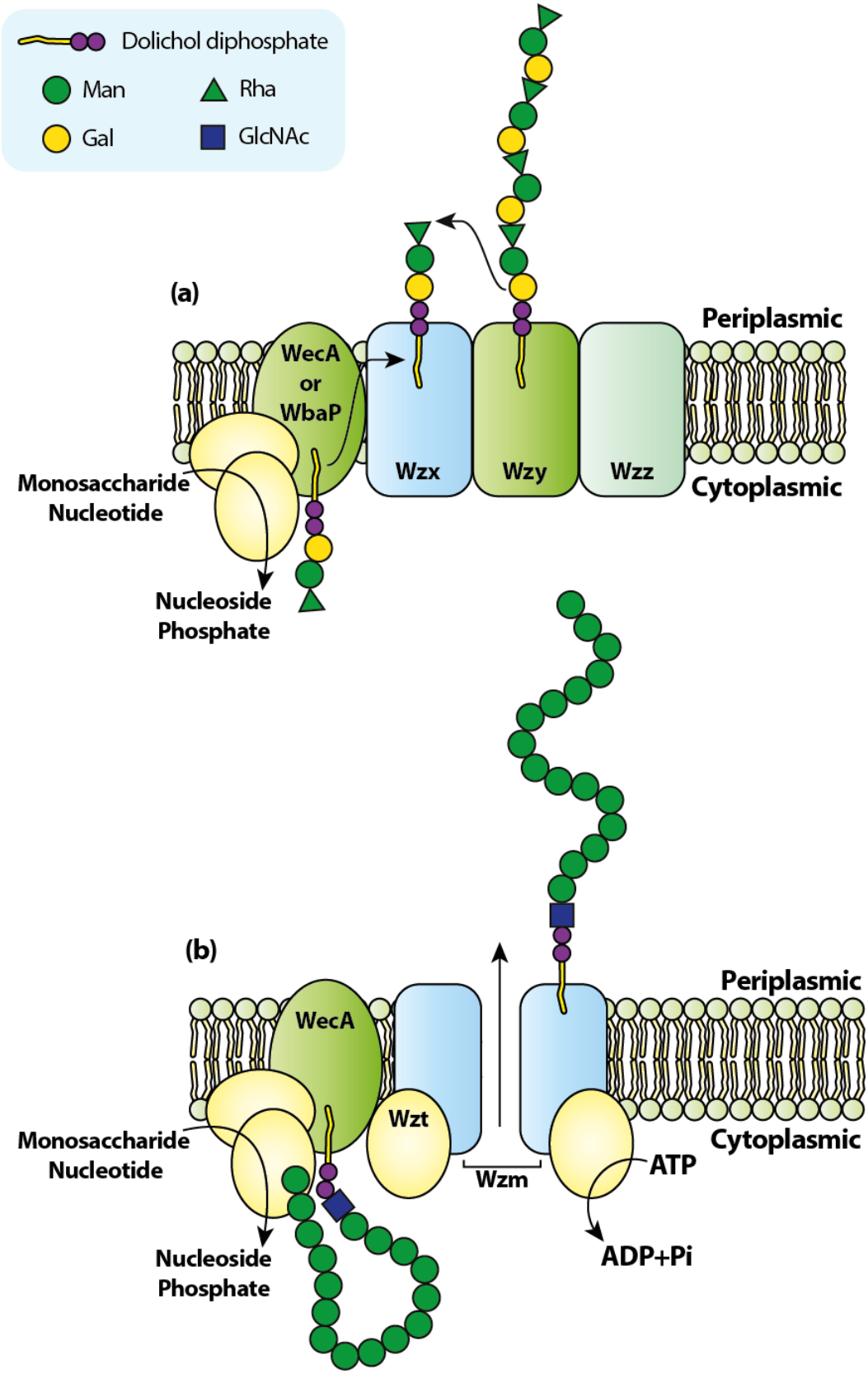


Figure 1-12 Diagram for Wzy-dependent (a) and ABC transporter-dependent (b) pathway. Adapted from (Raetz & Whitfield 2002)

1.2.1.4 Ligation and Transportation

In O-antigen biosynthesis, the fully synthesised O-antigens are eventually transported to the periplasmic side of the inner membrane, and the transportation itself is considered as a part of the pathway. However, core-lipid A is synthesised at the cytoplasmic side of the inner-membrane, and therefore needs to be transported to the periplasmic side where the ligation occurs (**FIG. 1-13**).

Core-lipid A is transported by MsbA (Doerrler *et al.* 2001). Its crystal structure indicates that each MsbA polypeptide contains one transmembrane domain (TMD) and one nucleotide-binding domain (NBD; Ward *et al.* 2007). Therefore, it is in fact a half-ABC-transporter and functions when forming multiplexes. The transportation function of MsbA is proposed to be based upon the nucleotide binding status of its NBD, which decides the conformation of its transmembrane helices and alters the direction of accessibilities of its lumen. More specifically, when the NBD of MsbA binds to nucleotides, its lumen is open towards the cytoplasmic space, and when the nucleotides are absent, its lumen is no longer accessible from the cytoplasmic space (Ward *et al.* 2007). The transportation process also plays a quality-control step of core-lipid A biosynthesis, *i.e.* mis-synthesised structures cannot be efficiently transported (Doerrler & Raetz 2002; Meredith *et al.* 2006). Therefore, it is not surprising that most modifications on core-lipid A can only happen after transportation (Whitfield & Trent 2014).

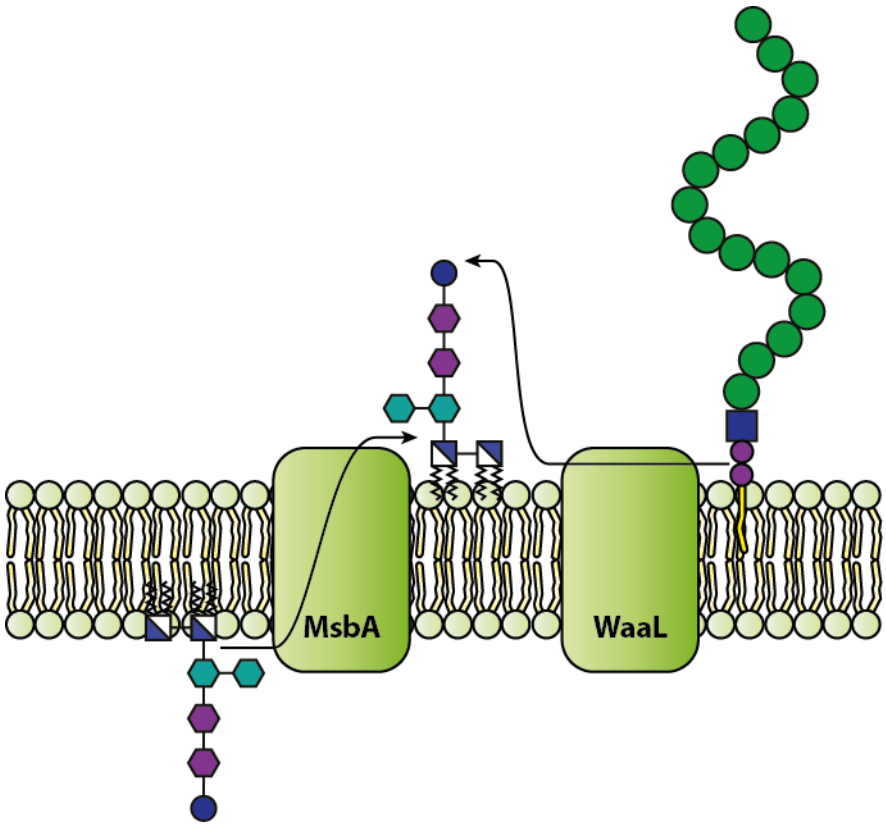


Figure 1-13 Transportation of core-lipid A by an ABC-transporter MsbA and ligation of O-antigens to core-lipid A by a ligase WaaL.

As mentioned, fully synthesised O-antigens are ligated onto core-lipid A at the periplasmic surface of inner-membranes by the WaaL ligase. Its crystal structure is not yet solved and therefore the detailed mechanism for ligation is still elusive. But it was proposed that WaaL behaves like an inverting GT (Ruan *et al.* 2012).

A fully synthesised LPS molecule is transported and inserted into the outer membrane by lipopolysaccharide transporter machinery (LpxT) that consists of multiple proteins (**FIG. 1-14**). The LPS is first extracted by an ABC transporter complex (Sperandeo *et al.* 2008; Ruiz *et al.* 2008), in which two LptB proteins comprise the NBDs, and LptF/LptG proteins form the TMDs. The extracted LPS is subsequently transferred towards the outer membrane through LptC and LptA proteins. The dimer of the LptA protein dimer act as a bridge connecting LptC

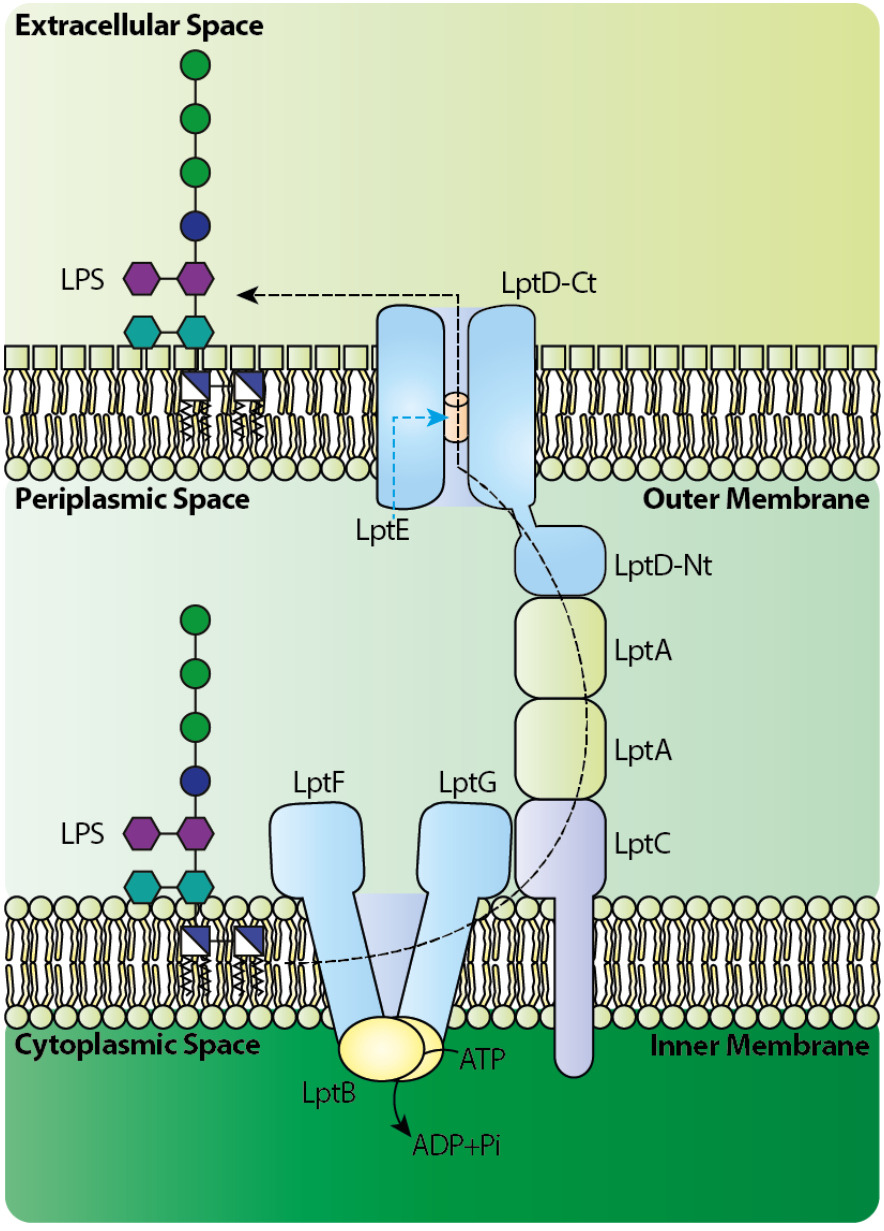


Figure 1-14 Transportation system of LPS in *E. coli*. The system is composed of LptA, B, C, D and E proteins that exist in a ration of 2:2:1:1:1. The whole system is driven by ATP hydrolysis. The dashed black arrow line shows the route of transportation. Ct and Nt stand for C- and N- terminal respectively. Adapted from (Whitfield & Trent 2014).

and the N-terminal domain of LptD (Bowyer *et al.* 2011). The C-terminal domain of LptD forms a giant transmembrane domain composed of 26 anti-parallel-stranded β-barrels, which hold the LptE that stabilises LptD C-terminal domain (Dong *et al.* 2014; Qiao *et al.* 2014). This protein complex flips the LPS from the periplasmic side of the outer membrane to its extracellular side and inserts it there. When considering LPS as a detergent-like molecule, *i.e.* it has a highly hydrophilic polysaccharide domain and a highly hydrophobic lipid domain, the question arises as to how this system binds to the LPS molecules. The answer to this question can be partially given by a characteristic β-jellyroll folding found in the crystal structures of Lpt A, C, F and G (Villa *et al.* 2013; Sperandeo *et al.* 2011; Suits *et al.* 2008; Tran *et al.* 2010). It was proposed that the acyl chains of lipid A are accommodated in these proteins in a way similar to its binding to the TLR4-MD2 complex. For the remaining components of this system, *i.e.* the LptD/E complex; it was proposed that LptE helps bind to the lipids and LptD forms a hydrophilic channel for the whole LPS.

The LPS transportation system has been almost exclusively researched in *E. coli* and it is not necessarily always conserved among other organisms (Sutcliffe 2010). Therefore, further investigations on other systems are essential for a comprehensive understanding of LPS transportation.

1.2.2 Bacterial Glycoproteins

The first bacterial glycoprotein was discovered in 1975 (Sleytr 1975), which is far later than LPS and other classical bacterial glycoconjugates, though they were not necessarily known to be glycoconjugates at first. On the contrary, glycosylation of eukaryotic proteins was first characterised in 1938 (Neuberger 1938), which is the reason why people believed for decades that glycosylation of proteins only existed in eukaryotes. To date, both N- and O- linked glycoproteins have been found in bacteria, and the corresponding biosynthetic pathways have been very well characterised (Iwashkiw *et al*. 2013; Nothaft & Szymanski 2010). Bacterial proteins are glycosylated in two distinct ways, the oligosaccharyltransferase (OTase)–dependent pathway and the OTase-independent sequential addition of monosaccharides. Both pathways can post-translationally transfer N- and O-glycans to bacterial proteins. The biological implications of the glycosylation of bacterial proteins are normally proposed to be associated with evasion of host immunity, bacterial adhesion and mobility, but additional as yet undiscovered roles are likely. This section introduces the biosynthesis of bacterial proteins, and covers in brief the biological implications.

1.2.2.1 Oligosaccharyltransferase-dependent Pathway

The OTase is a glycosyltransferase that can *en bloc* transfer an oligosaccharide from its lipid carrier to a protein substrate. Eukaryotic OTase enzymes have been characterised for some time, and it is established that they act to transfer the N-glycan precursor to a consensus sequence (N-X-S/T, where X is not proline) in eukaryotic proteins (R. Kornfeld & S. Kornfeld 1985). Bacterial N-linked oligosaccharyltransferase (N-OTase) was first found in *Campylobacter jejuni* (*C. jejuni*; Szymanski *et al*. 1999, Linton *et al*. 2002). Its pglB gene encodes the N-OTase that transfers a hexa-saccharide with a sequence of GalNAc-GalNAc-(Glc)-GalNAc-GalNAc-2,4-diacetamido bacillosamine (BacdiNAc) to an extended consensus sequences (see **FIG. 1-15 a**) of D/E-X_1-N-X_2-S/T (X_1 and X_2 are not proline; Wacker *et al*. 2002). *C. jejuni* PglB is encoded in its *pgl* gene locus, which encodes epimerases such as PglE, glycosyltransferases such as PglC and an ABC-transporter PglK (J. Kelly *et al*. 2006; Linton *et al*. 2002).

Since then, pglB and its homologs have been found in a growing number of organisms including *Campylobacter*, *Helicobacter*, *Sulfurovum* and *Nitratiruptor* (Nothaft & Szymanski 2010). According to the X-ray crystallography of PglB (Perez *et al*. 2015), it is composed of two domains. Its transmembrane domain contains an acidic catalytic centre and binding sites for protein acceptors and lipid carriers, and its periplasmic domain is involved in the recognition of the consensus sequence. An inverting GT function of PglB relies on the acidic catalytic centre, DXD motif and a bivalent metal (M^{2+}) ion, which is characteristic for GT-A folding. However, it seems the DXD motif coordinating M^{2+} is bi-functional, simultaneously

directing the asparagine acceptor to the acidic catalytic centre and stabilising the lipid carrier, which differs from GT-A family enzymes whose M^{2+} ions merely share the latter function.

Though no O-linked oligosaccharyltransferase (O-OTase) has been found in eukaryotes yet, it has been reported in multiple bacterial species such as *Burkholderia cenocepacia* (*B. cenocepacia*) and *V. cholerae* (Gebhart *et al.* 2012; Iwashkiw *et al.* 2013). The first characterised O-OTase was PilO in some *P. aeruginosa* strains (Castric 1995), which is probably horizontally acquired as it is absent in many other strains. Other O-OTase PglO or PglL can be found in the genomes of all *Neisseria meningitides* (*N. meningitides*) and *N. gonorrhoeae* species (Power *et al.* 2006). As these two enzymes are nearly the same, they are often both referred to as PglL directly.

All characterised O-OTases have a Wzy_C domain (Power *et al.* 2006) that is also found in the WaaL ligase (see **Section 1.2.1.4**). It is therefore proposed that the OTase mediated O-glycosylation has an evolutionary relationship to bacterial LPS biosynthesis (Hug *et al.* 2010). Indeed, the OTase-dependent pathway (**FIG. 1-15**) is remarkably similar to the biosynthesis of LPS O-antigens (see **Section 1.2.1.3**).

The N-OTase system in *Campylobacter* and the O-OTase system in *Burkholderia* are shown (**FIG. 1-15 a** and **b**) to exemplify the OTase-dependent pathway (Wacker *et al.* 2002; Lithgow *et al.* 2014). An initial GT transfers the monosaccharide-phosphate from the corresponding monosaccharide nucleotide to the Und-P carrier at the cytoplasmic surface of the inner-membrane, forming Und-PP linked monosaccharide as a primer. More monosaccharides are transferred to the primer, forming an UndPP-linked oligosaccharide that is flipped to the periplasmic side of the inner membrane, and the oligosaccharide is transferred to a protein acceptor by OTase.

Both PglB and PglL have a relaxed specificity, *i.e.* they can transfer a wide range of glycans onto multiple protein acceptors (Iwashkiw *et al.* 2012; Ciocchini *et al.* 2013). The specificity of PglL is broader than PglB, as the latter one can only transfer oligosaccharides with a reducing end monosaccharide containing N-acetyl groups whereas no structural requirement has been found for PglL. In addition, bacterial OTase loci have been proved to be functionally transferrable into *E. coli* (Wacker *et al.* 2002), which makes the OTase-dependent pathway a promising tool for producing novel glycoconjugate vaccines. Indeed, such investigations have been carried out for multiple human pathogens, and glyco-vaccines proved to provide protection on mice models (Cuccui *et al.* 2013; Rondini *et al.* 2015) and in clinical trials (Shao *et al.* 2006; Grijalva *et al.* 2007). In addition, the immunological mechanisms underlying the protection provided by glycoconjugate vaccines are being solved (Avci *et al.* 2011). Note that each of these two enzymes has their own advantages and disadvantages to produce glyco-vaccines. The specificity of PglL is broader than PglB, whereas the glycosylation mechanism of PglB is far better

understood than PglL. More research efforts on the three-dimensional structure of PglL and the mechanism of its GT function may render PglL with a wider potential for application.

As the OTase dependent glycosylation system can affect many different proteins in a single bacterium, it is not surprising it plays multiple biological roles. OTase-dependent glycosylation of bacterial proteins directly affects the virulence of various pathogens such as *C. jejuni*, *Neisseria* and *B. cenocepacia*. First of all, the OTase-dependent glycosylation machinery is associated with immune evasion. For instance, the GalNAc containing N-glycan in *C. jejuni* is recognised by galactose-type lectins expressed on human macrophages, which suppresses the expression of an immune stimulating cytokine, interleukin-6 (IL-6, Sorge *et al.* 2009). It was also proposed that O-glycans in *Neisseria* shield certain proteins from immune recognitions (Ku *et al.* 2009). Secondly, the glycosylation is related to mobility, *e.g.* a disruption of the glycosylation system of *B. cenocepacia* results in reduced mobility, and reduced its virulence compared to the wild-type (Lithgow *et al.* 2014).

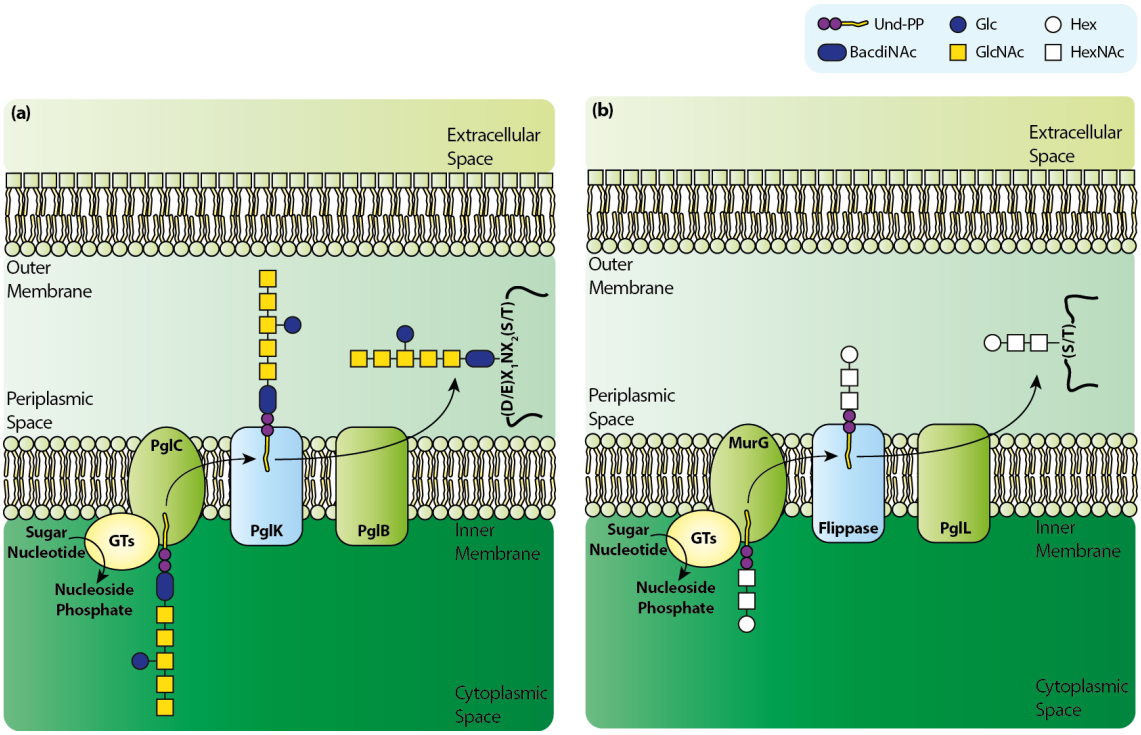


Figure 1-15 Model of the OTase-dependent pathway. The N-OTase-dependent pathway mediated by PglB in *Campylobacter jejuni* (a) and The O-OTase-dependent pathway mediated by PglL in *Burkholderia cenocepacia* (b) were chosen as examples (Wacker *et al.* 2002; Lithgow *et al.* 2014). In both cases, an amino sugar transferase initiates the pathway by transferring an N-acetyl containing monosaccharide-phosphate (BacdiNAc-P in a, hexNAc-P in b) to the UndP carrier forming a primer. The oligosaccharides are then assembled onto the primer and flipped from the cytoplasmic side of the inner-membrane to the periplasmic side. The PglB transfers the shown hepta-saccharide carried by Und-PP to the asparagine within the D/E-X_1-N-X_2-S/T (X_1 and X_2 are not proline) consensus sequences. PglL transfers the tri-saccharide in the figure to serine or threonine.

1.2.2.2 Oligosaccharyltransferase-independent Pathways

Bacterial proteins can also be glycosylated in an OTase-independent manner, and this mechanism is a lot simpler compared with the OTase-dependent pathway. The bacterial OTase-independent pathway is similar to the mammalian O-glycosylation system. Fundamentally, an initial GT directly modifies an acceptor protein with monosaccharides to generate a glycoprotein modified with monosaccharides. More GTs can assemble oligosaccharides onto the monosaccharides, forming glycoproteins carrying oligosaccharides. Both N- and O-linked glycoproteins can be synthesised via this pathway, and it glycosylates a broad-spectrum of important bacterial proteins such as flagellins and adhesins.

N-linked glycoproteins glycosylated by bacterial N-glycosyltransferase (NGT) are one of the newest characterised families of bacterial glycoproteins synthesised via the OTase-independent pathway, though bacterial N-linked glycoproteins were initially found in the early 1990s (Erickson & Herzberg 1993). HMW1C was the first discovered NGT that is responsible for N-glycosylation of HMW1A in *Haemophilus influenzae* (*H. influenzae*; Gross *et al*. 2008; Grass *et al*. 2010). HMW1C recognises the classical N-X-S/T (X is not proline) consensus sequence and transfers a single Glc onto the asparagine (**FIG. 1-16 a**; Gross *et al*. 2008, Grass *et al*. 2010). A homologue was later found in *Actinobacillus pleuropneumoniae* (*A. pleuropneumoniae*) (Schwarz *et al*. 2011) and this homolog is functional when transferred into *H. influenza* (Choi *et al*. 2010). *A. pleuropneumoniae* NGT (apNGT) has a relaxed monosaccharide-nucleotide and acceptor specificity *in vitro*. It can transfer Glc, Gal and xylose (Xyl) from both UDP and GDP; and when it is reconstituted in *E. coli*, apNGT tolerates N-X-S/T/A/G/V/N sequences as acceptors (Naegeli *et al*. 2014). In addition, it can even sometimes glycosylate serine and glutamine (Naegeli *et al*. 2014).

The majority of the glycoproteins synthesised via the OTase-independent pathway carry O-linked glycans (Iwashkiw *et al*. 2013; Nothaft & Szymanski 2010). The most characteristic examples are bacterial flagella. The flagellum is a whip-like system that allows cells to move, which can be found in all three kingdoms of life. Bacterial flagella are composed of a bacterial filament protein, flagellin, which has been known to be O-glycosylated for decades. One of the most prevalent sugars attached on flagellin is pseudaminic acid, a nine-carbon acidic sugar that mimics mammalian sialic acids (**FIG. 1-16 b**) (Thibault *et al*. 2001). It can be found in various human pathogens such as *H. pylori*, *C. jejuni* and *Aeromonas caviae* (Tabei *et al*. 2009; Schirm *et al*. 2003). Another example is legionaminic acid, which can be found in *Clostridium botulinum* and is very similar to pseudaminic acid (Twine *et al*. 2008). There is not yet a defined consensus sequence for O-glycosylation of flagellin, and it is interesting that glycosylation of flagellin differs drastically from one organism to another. For instance, there is only one O-glycosylation site in *Burkholderia* flagellin (Scott *et al*. 2011), whereas there are up to 19 sites in *Campylobacter* flagellin (Schirm *et al*. 2005).

In addition, *C. jejuni* flagellin is modified by a single monosaccharide at one site (Thibault *et al.* 2001), whereas some *Pseudomonas* flagellins contain several oligosaccharides (Schirm *et al.* 2004).

O-glycosylation of bacterial flagellin seems to be essential for its biogenesis and in turn affecting bacterial mobility and virulence. A deficiency in glycosylation leads to a failure in producing flagellar filament in most studied organisms, though a minor portion of the deficient strains can still express inactivated filaments. This can be explained by the hypothesis that flagella glycans contribute to stabilising and assembling flagellin. Other proposed functions for flagellar glycans are mainly concentrated on immune evasion through host mimicry and blocking host recognitions (Logan 2006; Nothaft & Szymanski 2010; Iwashkiw *et al.* 2013). However, these proposed functions should not be automatically generalised to all bacteria. In some radical examples such as *P. aeruginosa*, glycosylation of flagellin does not affect flagellar formation and bacterial motility, although it stimulates immune responses and significantly increases virulence (Arora *et al.* 2005; Verma *et al.* 2005).

Bacterial adhesins are by definition surface proteins that are mediating bacterial adhesion to host cells. Their glycosylation can increase bacterial adhesion and biofilm formation, facilitating colonisation. Moreover, experiments *in vitro* indicate a relationship between the OTase-independent glycosylation of adhesin and invasion of epithelial cells. O-glycosylation of adhesins is similar to flagellin, *i.e.* no consensus sequences and common glycan structures are characteristic. Numerous bacterial adhesins are proved to be O-glycosylated. For example, three adhesins AIDA-I (Moormann 2002; Benz & Schmidt 1989), TibA (Lindenthal & Elsinghorst 1999) and Ag43 (Sherlock *et al.* 2006; Knudsen *et al.* 2008) of *E. coli* were suggested to be heptosylated. *Streptococci* and *Staphylococci* produce serine-rich adhesins that are heavily O-glycosylated (more detailed information can be found in **Section 4.1**).

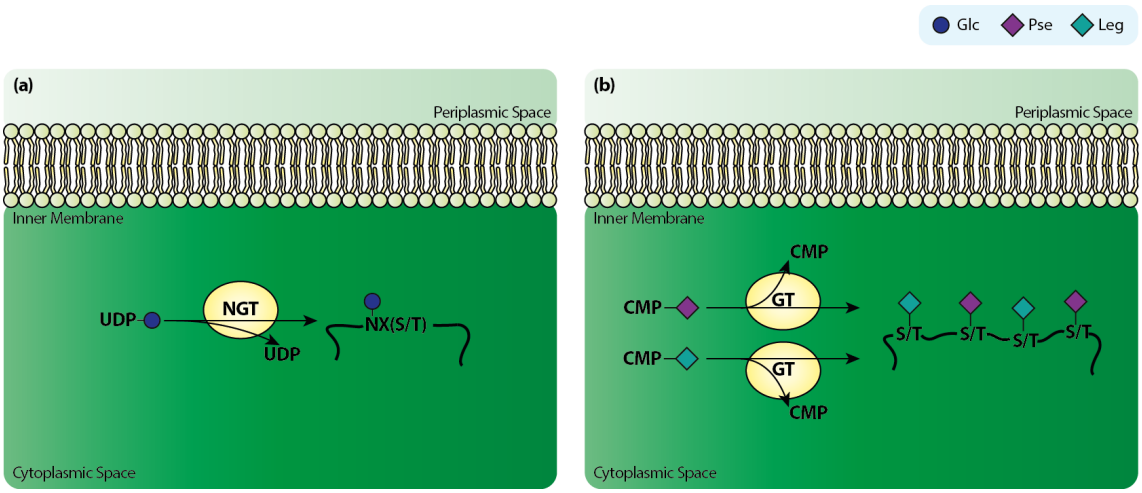


Figure 1-16 Model of the OTase-independent pathway. N-glucosylation mediated by NGT in H. influenza (a) and O-glycosylation of bacterial flagellin (b) are taken as examples. In both cases, the bacterial proteins are directly glycosylated by corresponding GTs.

O-mannosylated proteins are found in various bacteria. The O-mannosylation pathway is conserved through all three kingdoms of life (Lommel & Strahl 2009). The key process of this pathway is to transfer a mannose from its lipid carrier onto acceptor proteins under the catalysis of protein mannosyltransferases (PMTs; Lommel & Strahl 2009). The first bacterial O-mannosylated protein (Espitia & Mancilla 1989) and the bacterial enzyme (VanderVen *et al.* 2005) described in the literature as an O-mannosyltransferase is found in *Mycobacteria tuberculosis* (*M. tuberculosis*), which shows more than 20% homology to eukaryotic PMT. In bacteria, the initially transferred mannose can be elongated into oligo-mannoses (Liu *et al.* 2013), *e.g.* proteins can carry 10 mannoses at a single glycosylation site in *M. tuberculosis* (Iwashkiw *et al.* 2013). More O-mannosylated proteins have been characterised in other bacteria such as *Corynebacterium glutamicum* (Mahne *et al.* 2006) and *Streptomyces coelicolor* (Wehmeier *et al.* 2009).

1.3 Mass Spectrometry

The work described in this thesis would not have been possible without the more than one century development of MS as an analytical tool (McLafferty 2011; Hoffmann 2007; Cole 2011). MS is used to determine structures and quantities of substances based on measuring their mass-to-charge ratios (*m/z*). Sir Joseph John Thomson built the first mass spectrometer, or more accurately its prototype, in Cambridge during his exploration towards the composition of cathode rays. The device deflected charged particles by using electromagnetic fields, and quantitatively identified two isotopes of Ne. F. W. Aston continued Thompson's work to build the first fully functional MS. The application of electromagnetic focusing helped him identify more than two hundred naturally occurring isotopes at the Cavendish Laboratory. This discovery directly inspired the invention of the "calutron" that was used for enriching U-235 as a part of the Manhattan Project during World War-II. He received the 1921 Nobel Prize in Chemistry, two years after the first report of his MS. In his MS book, Isotopes, published in 1922, he mentioned the application of an "electron impact" method first used by A. J. Dempster for ionising several small organic compounds (Dempster 1918). Since then, MS began its service to chemical investigations.

In 1946, the time-of-flight (TOF) as a method for mass analysis was proposed by William Stephens (Wolff & Stephens 1953; Stephens 1946), which was realised by A. E. Cameron and D.F. Eggers two years later (Cameron & Eggers 1948). This mass analyser much later turned out to be a perfect partner of Matrix-assisted Laser Desorption/Ionisation (MALDI) established by K. Tanaka in the 1980s (Tanaka *et al.* 1988).

The subsequent 1950s was a flourishing period for MS. In 1952, the quasi-equilibrium theory (Rosenstock *et al.* 1952) and RRKM (Rice-Ramsperger-Kassel-Marcus; Marcus 1952) were proposed to explain the unimolecular fragmentation of ions, which led to R. A. Marcus winning the 1992 Nobel Prize in Chemistry. In 1953, W. Paul and co-workers introduced the quadrupole analyser (Hoffmann 2007) and the ion trap, and he received the 1989 Nobel in Physics. In 1956, MS was first used as a detector for gas chromatography (GC) by R. Gohlke and F. McLafferty (McLafferty 1957), following up the invention and application of GC. Liquid chromatography (LC) was later coupled to MS in the 1970s (McLafferty 2011; Fenn *et al.* 1989). These opened a new era where the combination of chromatography and MS has been widely applied.

In the 1960s, M. Dole initially proposed the idea of electrospray ionisation (ESI; Dole *et al.* 1968). His promising proposal inspired J. B. Fenn to investigate ESI-MS operated under both positive (Yamashita & Fenn 1984a) and negative modes (Yamashita & Fenn 1984b). In 1985, Fenn and his co-workers first made use of liquid chromatography (LC)-ESI-MS to characterise peptides and proteins (Whitehouse *et al.* 1985). Three years later, K. Tanaka first used a "prototype" of MALDI to ionise intact proteins (Tanaka *et al.* 1988).

Their work accelerated the biological application of MS, and they won the 2002 Nobel Prize in Chemistry.

MS is now being used as a routine method in many labs around the world to analyse all four types of biological macromolecules, nucleic acids, proteins, glycans and lipids. The establishment of tandem MS and dissociation techniques such as collision-induced dissociation (CID; McLafferty & Bryce 1967) and electron transfer dissociation (ETD; Syka *et al*. 2004) has greatly improved MS as a tool for structural determination of polymers especially biological macromolecules.

During a typical MS experiment, chemical compounds first get ionised in the ion source, producing ions that are separated in the mass analyser based on their *m/z*. Separated batches of ions with same *m/z* are detected in the detector, producing a mass spectrum. The following sections give an overview of modern MS, focusing on ionisation and mass analysis.

1.3.1 Ionisation

In all MS experiments, chemical or biological compounds need to be ionised before further separation and analysis. The device where ionisation happens is called the ion source. Variable ionisation techniques were developed, and this section focuses on electron ionisation (EI), ESI and MALDI, which were used throughout research described in this thesis. In addition, desorption methods such as field, laser and plasma desorption are very briefly mentioned, as they can be considered as prototypes of ESI and MALDI. Chemical ionisation (CI) and fast atom bombardment (FAB) are introduced, as they are associated with the ETD and CID that are used for MS/MS analysis in this thesis.

1.3.1.1 Electron Ionisation

Electron ionisation is the oldest ionisation technique used in MS and was first introduced in 1918 by A.J. Dempster (Dempster 1918), and it can be considered as the foundation of all ionisation methods. This "antique" is still widely applied around the world today.

A diagram illustrating the setup of an electron ionisation source is shown in **FIG. 1-17**. A potential heats the cathodic filament, which gives off electrons. These electrons are subsequently accelerated by an electron accelerating potential. During their travel through the ion source towards an anode, they orthogonally collide with the analytes in the gaseous phase. These molecules get ionised and become ions that are guided and filtered by several layers of electric fields into a mass analyser.

Electron ionisation was also called electron impact ionisation by tradition, as a classical but not accurate interpretation of this technique involves electrons produced in the ion source

impacting and expelling electrons from the analyte molecules.

As a matter of fact, EI should be considered under a context of quantum physics. Each electron produced by the emitter has a certain wavelength determined by the following **Eq. 1-1**,

$$\lambda=h/mv \qquad (Eq.1\text{-}1)$$

where λ, m and v are the wavelength, mass and velocity of the electron respectively, and h is the Plank constant. The energy E carried by this electron can be calculated by the following **Eq. 1-2**.

$$E=hv/\lambda \qquad (Eq.1\text{-}2)$$

When this energy matches the transition energy of an electron in a molecule, the analyte electron is activated to a higher energy level. Ionisation happens when the energy of the produced electron is high enough to excite and expel an electron out of the analyte.

The accelerated electrons produced in the ion source normally carry around 70 eV energy, which is sufficiently high to break the majority of the analyte into fragment ions. Therefore, EI is referred to as a "hard" ionisation technique. Because the EI requires gaseous samples and is only capable of ionising relatively small molecules, it is often coupled with GC and the quadrupole mass analyser. For example, GC-EI-MS is widely used for determining identities and linkages of monosaccharides (see **Section 3.0**; Sweeley *et al.* 1963; Albersheim *et al.* 1967).

1.3.1.2 Chemical Ionisation and Electron Transfer Dissociation

As oppose to "hard" EI that normally fragments the analyte molecules, CI is a "soft" ionisation technique that produces molecular ions, which can act as a complement to EI. As CI was introduced slightly later than another "soft" field ionisation (FI) method, it is in fact the

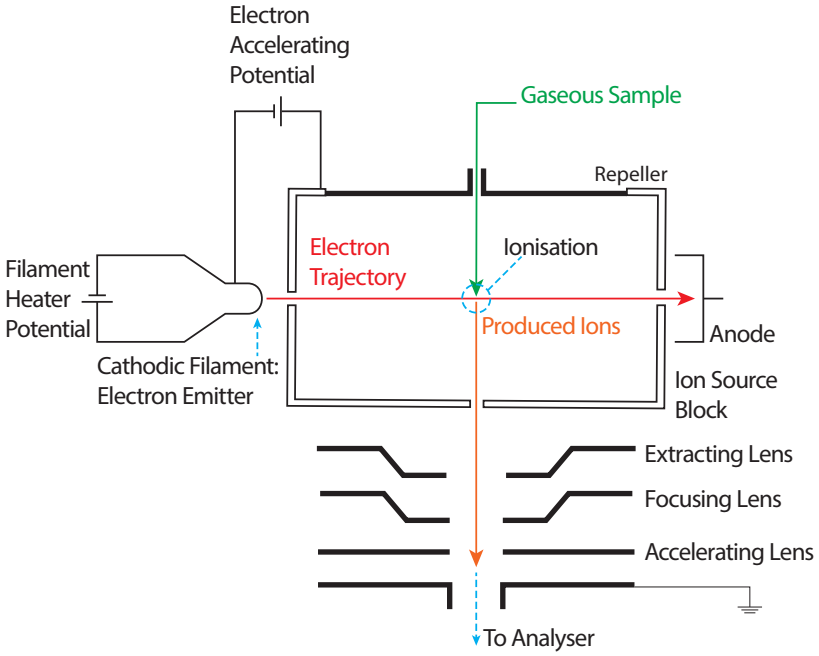


Figure 1-17 Diagram of the electron ionisation source. Adapted from (Hoffmann 2007).

second "soft" ionisation technique.

The setup of the CI source is very similar to the EI source (comparing **FIG. 1-17** and **1-18**). Gaseous samples and an excessive amount of a CI reagent such as methane, ammonia and isobutane are injected into an EI source. Because the amount of reagent is far larger than the sample, electron ionisation of the reagent is preferred and produces reagent ions. These ions chemically react with the sample, forming mostly the molecular ions of the sample.

CI normally produces positive ions. Under most circumstances, CI means positive ion chemical ionisation (PICI). PICI is not directly used in this thesis as softer ionisation techniques such as ESI and MALDI have been routinely used in research. However, fragmentation pathways for derivatised glycans under positive soft ionisation conditions were first defined by PICI. Also, negative ion chemical ionisation, although it was less commonly exploited, is the foundation of ETD in the thesis.

ETD is an MS/MS dissociation method for characterising post-translational modifications such as phosphorylation and glycosylation of proteins. For glycoproteomic investigations, it cleaves peptide bonds instead of glycosidic bonds, therefore the glycosylation sites can be mapped. The ETD experiments often need to be carried out in an ion trap type instrument (see **Section 1.3.2.2**) equipped with a CI source. In order to carry out ETD experiments, an ETD reagent such as anthracene is first negatively ionised in the CI source and injected into the ion trap, where electrons carried by the negatively charged ETD reagent ions are transferred to the peptides, cleaving their peptide bonds. A proposed ETD mechanism is shown in **FIG. 1-19**.

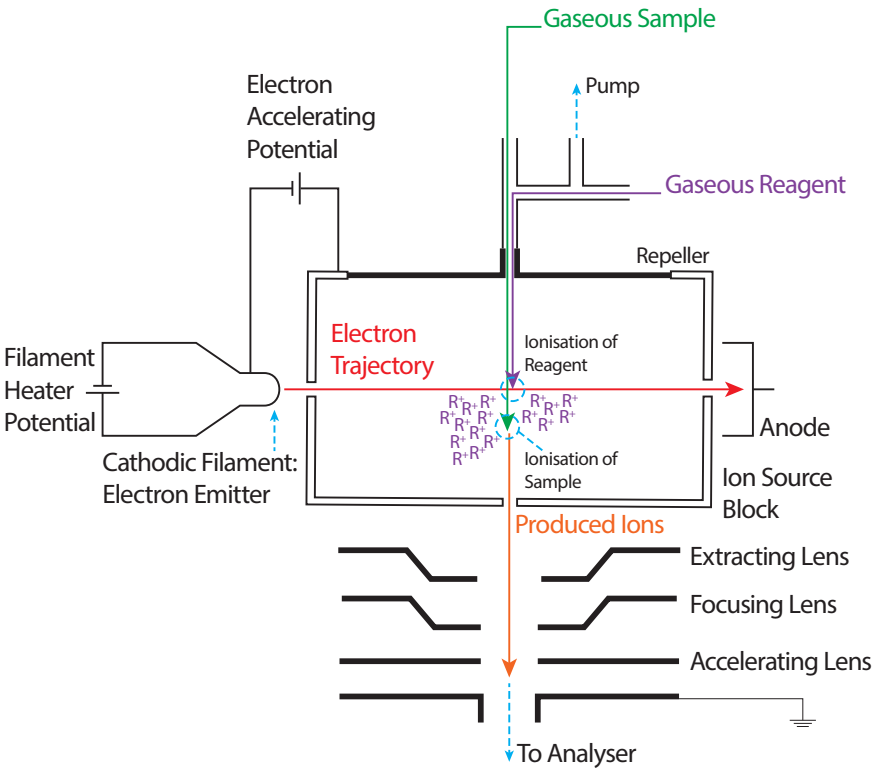


Figure 1-18 Diagram of chemical ionisation source. Adapted from (Hoffmann 2007)

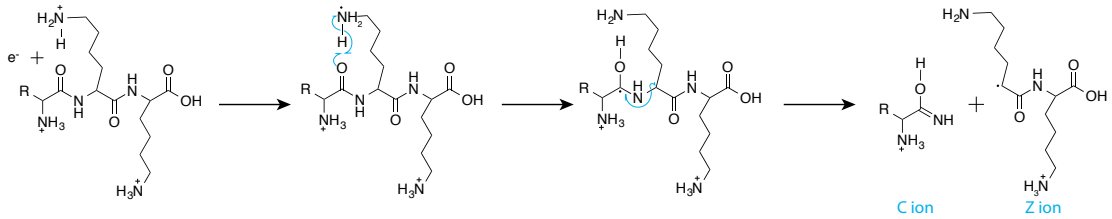


Figure 1-19 A proposed mechanism for ETD fragmentation. The blue arrows stand for transferring of an electron. Adaoed from (Syka *et al.* 2004).

It is obvious from the mechanism that only doubly or multiply charged peptides yield detectable products after the ETD fragmentation, as one of the charges is quenched during the transfer of the electron. The transfer of a proton from the C-terminal side of a peptide bond to the N-terminal side is coupled with an electron transfer process, forming a radical at the carbonyl group of the peptide bond. The fragmentation happens when the nitrogen-carbon bond next to the peptide bond cleaves, forming a double bond at the N-terminal side and a new radical at the C-terminal side. Overall, the ETD experiments normally produce C- and Z-type ions as shown in **FIG. 1-19** (Syka *et al.* 2004).

1.3.1.3 Desorption Ionisation

Ionisation sometimes conjugates with desorption, and field desorption (FD) is the first method based on their combination. For FD experiments, a solution of analytes is dropped on a tungsten or rhenium filament covered with carbon micro-needles. The samples gradually precipitate with the evaporation of the solution. Meanwhile, the filament is heated by a potential difference, melting the precipitated sample. The samples eventually accumulate and desorb at the tip of the micro-needles. When exerting a sufficiently high electric field on the needles, electrons of the samples are transferred to the filament, thus achieving the ionisation of the samples. The invention of FD was based FI that is a high electric field induced ionisation process, and FD is the basis of two most commonly used ionisation methods, MALDI and ESI.

Many other desorption methods such as plasma desorption, laser desorption (LD) have been inspired by the introduction of FD. For LD, high energy laser pulses are focused on sample surfaces, vaporising and ionizing the samples. A major disadvantage of LD is that its application heavily relies on the physical properties such as volatility of the samples, and it is actually a rather "hard" technique that normally fragments molecules larger than 500 Da. These disadvantages have been overcome by MALDI (see **Section 1.3.1.5**).

1.3.1.4 Fast Atom Bombardment and Collision-induced Dissociation

The ionisation technique bridges LD and MALDI is FAB. FAB along with liquid secondary ion MS are secondary ion MS techniques that emit ions by irradiating a surface with a beam of atoms or primary ions. Their samples need to be dissolved in a non-volatile liquid matrix such as glycerol. The usage of the matrix later became characteristic for MALDI (see **Section 1.3.1.5**).

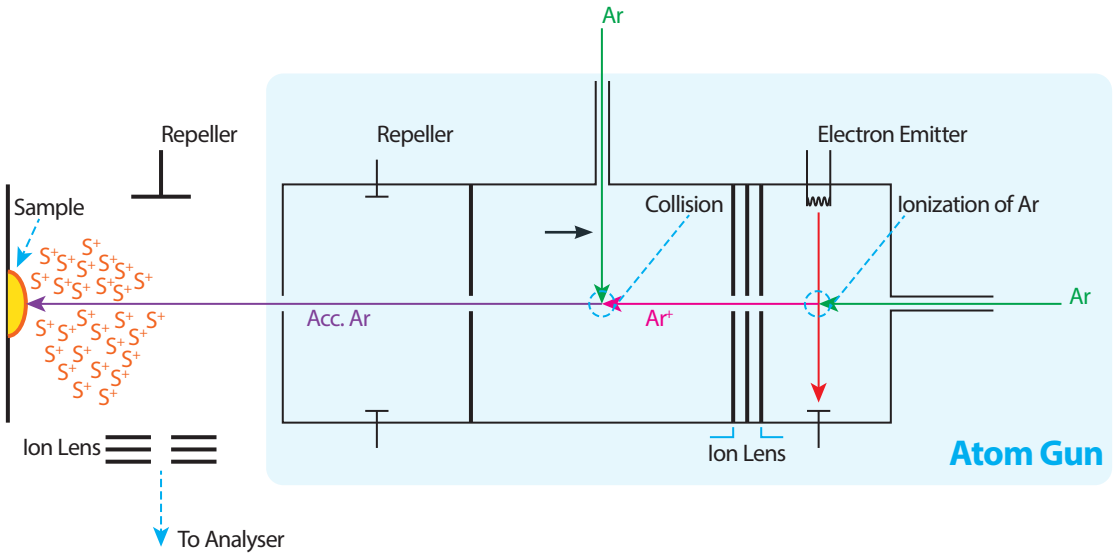


Figure 1-20 Diagram of an Atom Gun and FAB ion source. Adaped from (Hoffmann 2007).

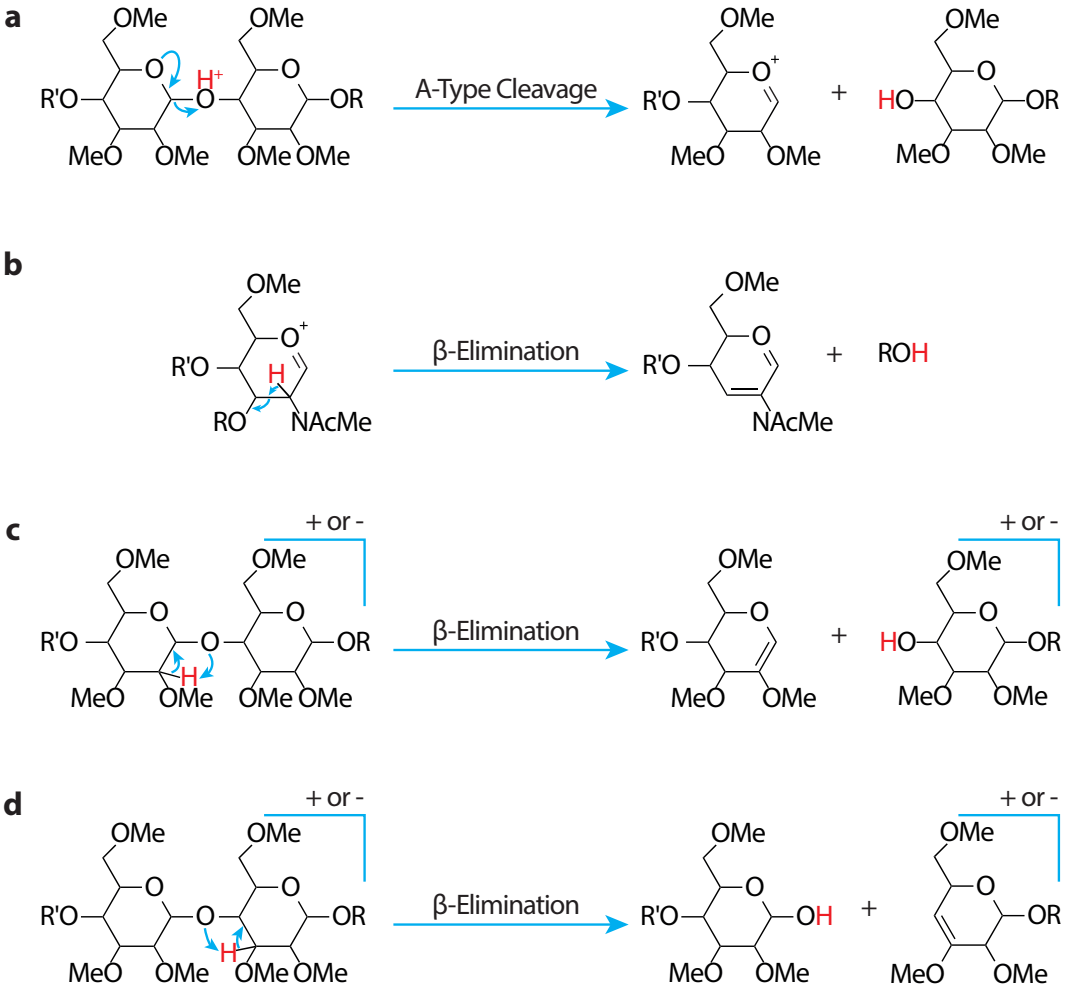


Figure 1-21 Mechanisms for fragmentations of permethylated glycans. Glycans are normally permethylated before MS characterization by MS. Therefore, the fragment pathways are shown for permethylated glycans. (a) An oxonium type ion is formed at the non-reducing side of a glycosidic bond. (b) If the oxonium ion contains an N-acetyl group at its 2-position, it can eliminate its 3-position group. (c) and (d) show an elimination mechanism based on proton migration. This reaction can occur at the non-reducing end side (c) or reducing end side (d) of a glycosidic bond.

For FAB, argon or xenon atoms are usually accelerated and focused on the mixture of matrix and sample. The accelerated atoms are generated by a device called the atom gun (**FIG. 1-20**). Taking argon as an example, the argon atoms are ionised and accelerated to collide with argon atoms. The collisional process can be described by the following equation.

$$Ar^{\cdot+}_{(rapid)} + Ar_{(slow)} \qquad Ar_{(rapid)} + Ar^{\cdot+}_{(slow)} \qquad \textit{(Eq.1-3)}$$

A repeller is set after the collision in the atom gun to remove argon ions so that only energetic atoms can leave the atom gun. These atoms hit the surface of the sample/matrix solution and eject sample ions out of the solution. The ions are then guided by an accelerating potential into the mass analyser.

There are two major pathways for the fragmentation of permethylated glycans (Dell *et al.* 1983). A-type cleavage (**FIG. 1-21 a**) occurs at the non-reducing side of a glycosidic bond, forming an oxonium ion. This reaction is normally observed for amino sugars that can further eliminate their 3-position group (**FIG. 1-21 b)**, which produces diagnostic ions for 3-linked HexNAc. Another major fragmentation pathway is β-elimination, which couples with a proton migration process. This reaction can occur at the non-reducing end (**FIG. 1-21 c**) or reducing end (**FIG. 1-21 d**) side.

1.3.1.5 Matrix-assisted Laser Desorption/Ionisation

MALDI is a powerful soft ionisation technique that is capable of producing molecular ions of various high-molecular-weight compounds even if they are non-volatile and thermally labile. It has been widely adopted for analysing biological macromolecules such as proteins, oligonucleotides and glycans. For a typical MALDI experiment, the sample is pre-mixed with a matrix and the mixture is dropped on a MALDI plate. The matrix is normally an organic chemical with a large conjugation system that strongly absorbs laser light with a certain wavelength. The mixture is allowed to be sufficiently dried before MALDI analysis, during which the samples are co-crystallised and embedded into the matrix molecules. When firing laser pulses to the co-crystals, the matrix molecules are excited and rapidly heated because of their accumulating a large amount of energy. This sublimates and ablates a portion of the matrix crystals into the gaseous phase, which simultaneously desorbs intact sample molecules.

The rest of the MALDI processes remains unclear and the ionisation of samples can possibly happen at any point. One of the most widely accepted proposals for the mechanism of MALDI involves a proton transfer from the matrix to the sample molecules right before they enter into the gas phase. At this point, the co-crystal system has the largest possible energy, because the time interval is insufficient for its relaxation.

MALDI is mainly coupled with the TOF mass analyser, as it minimises the spatial distribution of the produced ions. A diagram of MALDI-TOF mass spectrometry is shown in **FIG. 1-26**, **Section 1.3.2.3**.

1.3.1.6 Electrospray Ionisation

ESI is a soft ionisation technique that can produce multiply charged ions from the liquid phase, which is perfect for interfacing the LC and essential for certain MS/MS analysis such as ETD experiments. In addition, the *m/z* of an ion is significantly reduced when it carries more than one charge, thus greatly expanding the mass ranges of mass analysers. For instance, ions of intact proteins produced by an ESI source can carry more than 70 charges (Tipton *et al*. 2012), which means as long as the molecular weights of the proteins are less than 140 KDa, they are readily detectable by mass analysers with a mass range up to 2 KDa in principle.

Because of its superior features, ESI has become one of the most widely used ion sources and its mechanism has been extensively researched (Cole 2011). A diagram of an ESI source is shown in **FIG. 1-22 a**. In the positive ion mode of ESI, a positive electric potential is imposed to a capillary tip that can be the outlet of an LC or a syringe. Under this field, the solution of the sample is injected into the ion source, forming a spray of positively charged droplets emerging from the capillary tip. The solvent contained in these droplets is gradually evaporated,

generating gaseous ions. A mixture of ions, charged droplets and solvent vapour enters into a countercurrent chamber through an orifice, and the latter two species are removed by a stream of nitrogen. The ions are guided by the electric potential and a pressure difference into a heated capillary, and eventually enter into a low-pressure chamber, where another electric field between the capillary and skimmer accelerates the ions and propels them into high vacuum mass analysers.

ESI produces ions via a three-step mechanism: first, producing charged droplets at the tip of the ES capillary; secondly, reducing the volumes of the drops by the evaporation-induced shrinkage and electric repulsion-induced fission; thirdly, charges transferring from the droplets to the analytes.

The charged droplets are produced via the electrophoretic mechanism. Normally, a high voltage of 2-3 kV is imposed to the capillary (**FIG. 1-22 b**). The strength of this field between the capillary tip opposite a planar counter-electrode can be empirically evaluated with the following **Eq. 1-4** (Pfeifer & Hendricks 1968):

$$E_C=2V_C/[r_C \ln(4d/r_C)] \qquad (Eq.1\text{-}4)$$

where E_C is the electric field, V_C is the applied potential, r_C is the capillary outer radius, and d is the distance from capillary tip to the counter-electrode.

As the spray capillary tip is very thin, the electric field at the tip is high enough (E_C around 106 V/m) to penetrate the solution near the capillary tip, causing a polarization of the solution near the meniscus of the liquid. As a result, the solution with even a trace amount of electrolytes is sufficiently conducting, and the positive ions in the solution move towards the counter-electrode. This directly leads to an enrichment of positive ions near the meniscus surface that is distorted to a cone-shape, which is called the Taylor cone. When a sufficiently high field is imposed, a fine jet carrying an excessive amount of positive charges emerges from the tip of the Taylor cone, which produces the first batch of charged droplets. The whole process can be interpreted from an electrochemical point of view, *i.e.* the ESI source can be considered as a special electrolytic cell, where the ion transportation occurs in the gaseous phase rather than the liquid phase.

The second step of the ion production during the ESI process is the evaporation-induced shrinkage and electric repulsion-induced fission (**FIG. 1-22 c**). The produced charged droplets continuously shrink due to the evaporation of the solvent in an ambient gas, which is normally nitrogen. As the charge carried by the droplets remains nearly constant during the shrinkage, the already high charge density of the droplets is getting even higher with the evaporation of the solvent, which causes an increase in the repulsion of the charges within the droplets. When this repulsion overcomes the surface tension of the solvent, it eventually results in their Coulomb fissions. The fissions normally produce smaller charged progeny droplets.

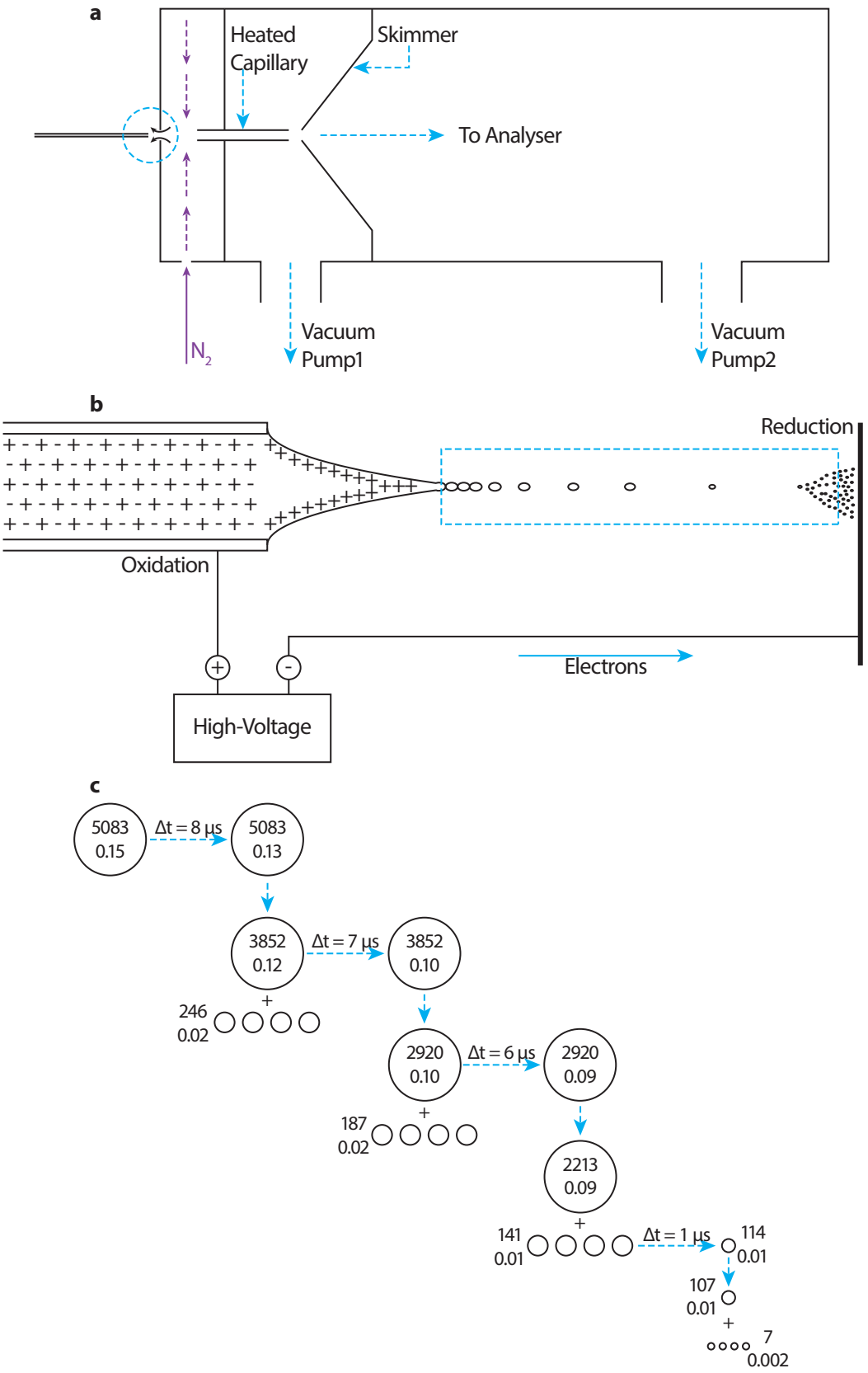


Figure 1-22 The mechanism of ESI. (a) a diagram of an ESI source. The blue circle is zoomed into (b) that shows the polarisation of the sample solution and emergence of the charged droplets. The blue square is zoomed into (c) that shows the shrinkage of the droplets. Each circle is a charged droplet. The numbers in or around the circles show their charge numbers (up) and radii (μm, down). Horizontal arrows annotated with time intervals indicate evaporation of solvents and vertical arrows indicate the Coulomb fissions. Adapted from (Cole 2011).

The condition for the Coulomb fissions is described by the Rayleigh equation **(Eq. 1-5)** (Rayleigh 1882),

$$Q_{Ry}=8\pi(\varepsilon_0\gamma R^3)^{1/2} \qquad (Eq.1\text{-}5)$$

where Q_{RY} is the charge carried by the droplet, γ is the surface tension of the solvent, R is the radius of the droplet and ε_0 is the electrical permittivity. For water, the Coulomb fission happens when the droplets carry the charge at approximately 90% of the Rayleigh limit and produces droplets carry the charge at around 65% of the limit. The droplets keep undergoing the Coulomb fission until they turn into minute droplets that give rise to ions.

Two mechanisms for ultimately producing ions from the minute droplets have been proposed, the ion evaporation model (IEM) and charged residue model (CRM; Cole 2011). The former model is normally used for describing the ionisation of small chemicals and peptides, and the latter model is better suited for the ion production processes of large analytes such as proteins.

Iribarne and Thomson proposed small ions can evaporate from the droplets directly, and the rate constant K_I for emission of ions is given by **Eq. 1-6** (Iribarne & Thomson 1976),

$$K_I=(K_BT/h)e^{-\Delta G^{\neq}/KT} \qquad (Eq.1\text{-}6)$$

where K_B is the Boltzmann's constant, T is the thermodynamic temperature of the droplet, h is the Plank constant. $\Delta G^{\neq}$ is the free energy of activation that needs to be interpreted from a transition state model of the ion evaporation processes (**FIG. 1-23**).

The energy barrier in this model is mainly due to the difference of two opposing forces; the attraction between the escaping ion and droplet, and the repulsion between the escaping ion and the remaining ions. The former force is larger than the latter at a short distance and smaller at a large distance. These two forces become equal to each other at a certain distance, where the transition state occurs. In addition to the energy barrier caused by the two forces, anther portion of the free energy required to transfer the ion from a neutral droplet to the gas phase (the droplet and the gas phase have a same volume) needs to be considered. The IEM can predict the boundary between the ion evaporation and the Coulomb fission. For positive ions, at a radius lower than around 8.4 nm, ion evaporation replaces the Coulomb fission.

Large molecules such as proteins are probably produced in the way described by the RCM, which is a relatively simple model proposing that when the charged droplets evaporated completely, the charges will be transferred to the analytes (Winger *et al.* 1993). The importance of this model replies on the fact that it can predict the number of charges carried by certain proteins. Smith and co-workers found an empirical correlation between the molecular mass, M, and the average charge, Z_{av},

$$Z_{av}=aM^b \qquad (Eq.1\text{-}7)$$

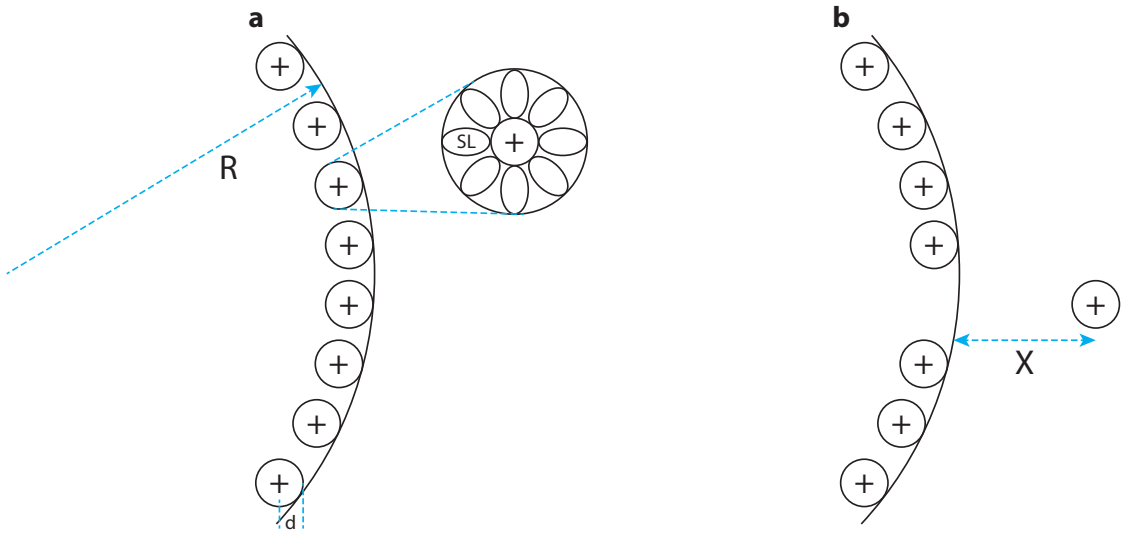


Figure 1-23 Schematics of the IEM. Under an equilibrium state (a), the ions are surrounded by the solvent molecules (SL), and they are located just underneath the surface of the droplet. In the transition state (b), one of the ions is ejected out of the surface due to a thermal activation. The X represents the distance where the forces of attraction and repulsion are equal to each other. Adapted from (Cole 2011).

where a and b are constant. This equation was initially established for dendrimers, and proved to fit the data for globular proteins. A good agreement between the curve predicted by this equation and experimental data can be observed when a=0.078 and b=0.5 with M has a unit of mega Da (La Mora 2000; Cole 2011). A much better agreement can be observed when the molecular mass in the equation is replaced by the surface area S (Kaltashov & Mohimen 2005; Cole 2011).

$$Z_{av}=aS^{b} \qquad (Eq.1\text{-}8)$$

This equation can explain the phenomenon that denatured proteins normally carry more charges than their native states.

The CRM theory leads to a problem why only monomers are normally observed, *i.e.* if CRM is correct, one should also expect to observe oligomers of proteins. The answer to this question is simply that oligomers of proteins indeed exist. However, as their intensities are very low and rapidly decreasing from dimer, trimer to oligomers, it is highly challenging to observe them. The observation of high mutlimers of proteins by using a mass spectrometer with a very high mass range has been reported (Winger *et al.* 1993).

1.3.2 Mass Analysis

In MS experiments, the ion separation and mass analysis are performed by mass analysers. More specifically, the produced ions are normally accelerated in and after leaving the ion sources, and guided into the mass analysers, in which they are separated based on their *m/z* so that they can be eventually detected. Three parameters are used to evaluate the quality of a mass analyser, the resolution, mass range and transmission ratio.

The resolution (*R*) measures how well a mass analyser can resolve two ions with very similar *m/z*. It can be defined by **Eq. 1-9**:

$$R=m/dm \qquad (Eq.1\text{-}9)$$

where *dm* is the smallest distinguishable mass difference between two MS peaks at *m/z m* and (*m-dm*). In other words, it is the resolving power. Its value can be taken as the peak width at certain percentages such as 0.5%, 5% and 50% of the peak height. When the percentage is set to be 50%, and the term *dm* represents the full width at half maximum (FWHM).

The mass range measures how wide *m/z* a mass analyser can deal with. As it is always defined based on single charged ions (*z*=1), it can be directly reported in the unit of Da.

The transmission ratio decides sensitivity, and a higher transmission ratio means a higher sensitivity. It is the ratio between the number of ions surviving to arrive at the detector and ions produced in the ion source.

The next sections introduce the mechanisms of the sector mass analysers, quadrupole mass filters (QMS), quadrupole ion traps (QIT) and TOF mass analysers, which are used in research included in this thesis. Different mass analysers can be linked in tandem to each other, forming the so-called hybrid mass analysers. They combine the advantages (and also limitations) of these mass analysers. The hybrid quadrupole time-of-flight (Q-TOF) mass analyser is emphasised at the end of this section.

1.3.2.1 Magnetic/Electrostatic Sectors

It is long known that the trajectories of ions with different *m/z* are separated within a magnetic field, and this knowledge was applied for making the earliest mass analyser, the magnetic sector mass analyser.

When an ion leaves the ion source, it carries a certain amount of kinetic energy, E_K provided by a potential difference, *V*:

$$E_K=mv^2/2=qV \qquad (Eq.1\text{-}10)$$

where m, v and q are the mass, velocity and charge of the ion, respectively. If this ion perpendicularly moves into a magnetic field, it is submitted to the Lorentz force, F_L:

$$F_L=qvB \qquad (Eq.1\text{-}11)$$

where q stands for the charge carried by the ion and B is the strength of the magnetic field. This force only changes the direction of the ions' movement without affecting their velocity. Therefore, it acts as the centrifugal force (F_C), curving the trajectory of the ion,

$$F_C=(mv^2)/r=qvB \qquad (Eq.1\text{-}12)$$

where r is the radius of its curved trajectory. From **Eq. 1-12**, it is easy to obtain the following relationship:

$$r=mv/qB \qquad (Eq.1\text{-}13)$$

It is obvious that the radius of the ion, and therefore the trajectory of the ion, are merely dependent on the strength of magnetic field when all other terms are determined, *i.e.* its kinetic energy is fixed before entering into the mass analyser.

By combining **Eq. 1-10** and **1-13**, it is easy to obtain the following important equation:

$$m/q=(r^2B^2)/2V \qquad (Eq.1\text{-}14)$$

It reveals that when the trajectory of the ions is fixed, for each given strength of magnetic field, only ions with a certain m/z can move through the analyser and arrive at the detector. Also, when the magnetic field is fixed, only the ions with a certain m/z can travel through the analyser with a determined radius. Hence, by scanning the strength of the magnetic field or gradually altering the direction of the mass analyser, the separation of ions with different m/z can be achieved.

The ions with the same mass and charge disperse in the magnetic sector analyser when they have slightly different kinetic energies. The dispersion can be described by the following equation obtained by rearranging **Eq. 1-14**,

$$r=(2mE_K)^{1/2}/qB \qquad (Eq.1\text{-}15)$$

The dispersed ions can be re-focused in an electrostatic field. Being similar to the case introduced for the magnetic sector mass analysers, when an ion moves into an electrostatic field, it is submitted to an electrostatic force (F_E):

$$F_E=qE \qquad (Eq.1\text{-}16)$$

where E is the strength of the electrostatic field. For an electrostatic sector mass analyser, the field is produced by two circular electrodes, making the force constantly perpendicular to the velocity of ions. Thus, the force will also only change the direction of the ion's velocity and acts as the centrifugal force:

$$F_C=(mv^2)/r=qE \qquad (Eq.1\text{-}17)$$

and,

$$r=(mv^2)/qE \qquad (Eq.1\text{-}18)$$

Therefore, the radius can be adjusted by changing the strength of the electrostatic field in order to re-focus the dispersed ions.

A mass analyser based on all the theories above is the so-called "double-focusing" mass analyser (**FIG. 1-24**). The magnetic sector can be set prior or after the electrostatic sector along the ions' moving direction. The introduction of high field magnets to this type of instrument has rendered it a high mass limit (up to around 15 KDa) and a decent resolution (50K - 100K).

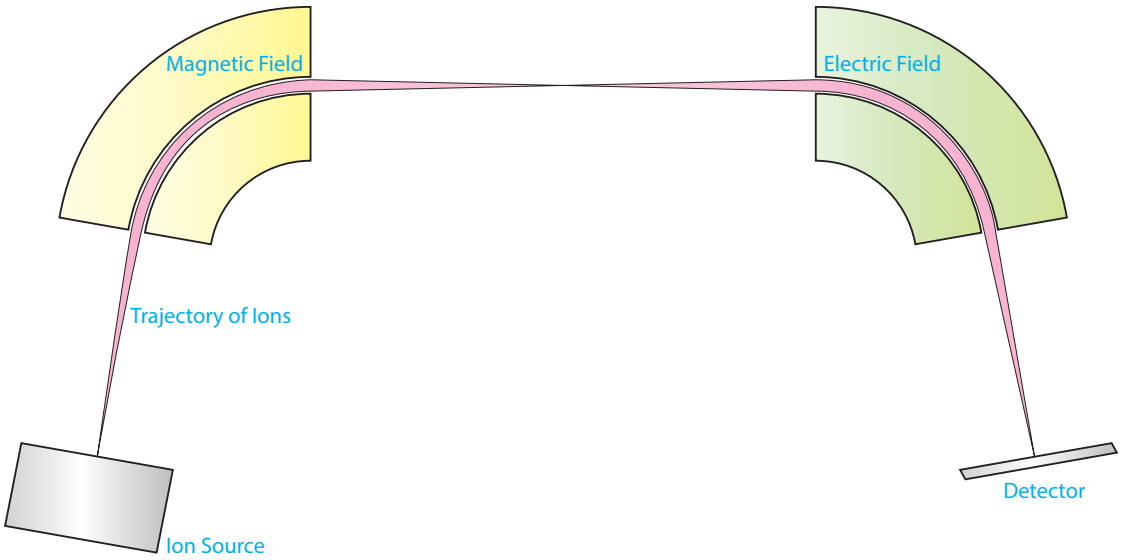


Figure 1-24 A double-focusing mass analyser employing the reverse Nier-Johnson geometry (Matsuda 1981).

1.3.2.2 The Quadrupole, Triple-Quadrupole and Ion Trap

Radio frequency (RF) and direct current (DC) fields can manipulate ions, based on which, multiple mass analysers have been invented (Hoffmann 2007; March 1997; Schwartz *et al.* 2002). This section introduces four types of such instruments: QMS and QIT are emphasised, and the linear quadrupole ion trap (LQIT) and Orbitrap are briefly introduced.

QMS is built by symmetrically positioning four rod-like electrodes in parallel (**FIG. 1-25 a**), which ideally creates a hyperbolic-shaped inner surface. The electrodes are held in position by insulating materials such as ceramics. A high-quality 2D quadrupole field is formed near the axis of ions by applying the RF plus DC waveforms to the four electrodes. More specifically, a potential of $U-V\cos(\omega t)$ is applied to one pair of rods and its opposite potential of $-U+V\cos(\omega t)$ is applied to another pair (see the zoomed in region of **FIG. 1-25 a**). Here U and $V\cos(\omega t)$ are the magnitudes of the DC and RF fields, respectively. The movement of ions in the electric field produced by the QMS is defined by the second order Mathieu differential equation. Though integration of the equation is beyond the scope of the thesis, the spatial positions of ions are determined by the two Mathieu parameters below:

$$a_u = a_x = -a_y = 2eU/(mr_0^2\pi^2 f_{RF}^2) \qquad \text{(Eq.1-19)}$$

$$q_u = q_x = -q_y = eV/(mr_0^2\pi^2 f_{RF}^2) \qquad \text{(Eq.1-20)}$$

where r_0 is the nearest distance between the electrodes and the central axis of the QMF, f_{RF} is the frequency of the RF voltage, and e and m are the charge and mass of the ions, respectively. For a given instrument, r_0 and f_{RF} are fixed values, and e and m are determined for ions. As a result, the Mathieu parameters and the positions of the ions are solely dependent on the U and V values.

The Mathieu equation reveals that for given U and V values, only ions with the correct m/z can go through the mass analyser, *i.e.* as long as the X and Y coordinates of an ion are less than r_0, it can successfully travel through the QMF.

Other ions are filtered by the QMF, which is the origin of the name "quadrupole mass filter". When scanning the magnitudes of RF and DC fields from low to high, ions with low to high m/z are transmitted. When the magnitude of DC is set to zero, QMF is constantly operated under a RF field and it acts as a mass guider instead of a mass filter, which allows the transmission of ions with a wide range of m/z. Ions can be guided to collide with each other and/or inert gases within this device, and therefore it is also called a collisional cell.

For practical considerations, the typical mass range of a QMF is 10-4000 Da, throughout which isotopic resolution can be maintained. These parameters are actually poorer relative to other mass analysers. However, multiple advantages of QMF, including the ease of operation,

low expense, and adaptability to multiple ion sources guarantee it a wide range of applications, especially the ones that do not require superb mass range and resolution such as monosaccharide analysis. All GC-EI-MS analyses described in this thesis are carried out with a QMF mass analyser.

In addition, the popularity of the QMF largely relies on the significant advantage that it can be coupled to various mass analysers including the QMF itself to carry out the MS/MS experiments. The widely applied QqQ or triple-quadrupole mass spectrometer is set up by tandemly linking three QMFs and parking the middle quadrupole mass analyser at the collision cell mode (**FIG. 1-25 a**). This configuration renders the mass spectrometer the ability to select and fragment certain ions, which not only increases the signal-to-noise ratio (S/N), but also makes structural elucidation based on MS far more reliable. The Q-TOF mass analyser is another example of a combination of QMF with other mass analysers, which makes a great contribution to the field of proteomics and is introduced in **Section 1.3.2.3** in detail.

The QIT is built upon a theoretical basis that is similar to QMF, but it is a three-electrode device. The basic architecture of QIT is shown in **FIG. 1-25 b**. Two of the electrodes sharing a same conical shape are called end caps, and the third sand-clock shaped electrode is called the ring electrode. The two end caps sandwich the ring electrode, and they are held by insulating materials such as ceramic. The geometry of the electrodes generates a hyperbolic shaped inner face, which is reminiscent of QMF. There is a hole at the centre of each end cap, which is used for either injection or ejection of the ions. A RF voltage with a magnitude of $V\cos(2\pi f_{RF}t)$ is applied to the ring electrode. In contract to the 2D quadrupole field generated by QMF, a 3D quadrupole field is produced by this RF voltage. Again, the trajectories of ions are given by the Matthieu differential equations, and the Matthieu parameters under a QIT context are shown below:

$$a_z=-2a_r=(-4eU_{DC})/[m(r_0^2+2z_0^2)\pi^2 f_{RF}^2] \quad \text{(Eq.1-21)}$$

$$q_z=-2q_r=(-eV_{RP})/[m(r_0^2+2z_0^2)\pi^2 f_{RF}^2] \quad \text{(Eq.1-22)}$$

where U_{DC} is the DC amplitude, V_{RF} is the RF amplitude, f_{RF} is the frequency of the RF voltage, e and m stand for the charge and mass of an ion, respectively, r_0 represent the inner radius of the ring electrode and z_0 is the minimum axial distance between the centre of the QIT and the end-caps.

Under the combinations of certain Matthieu parameters *i.e.* a_z and q_z, the ions are trapped inside the QIT. The space enclosed by these parameters in a 2D coordinate system is called the stable area. At the boundaries of the stable area, ions start being "unstable" and leaving the QIT. When the QIT is operated under the Matthieu parameters outside the stable area, it carries out an unstable scan and ejects ions. For instance, ions with lower *m/z* can be ejected first, followed by ions with higher *m/z*.

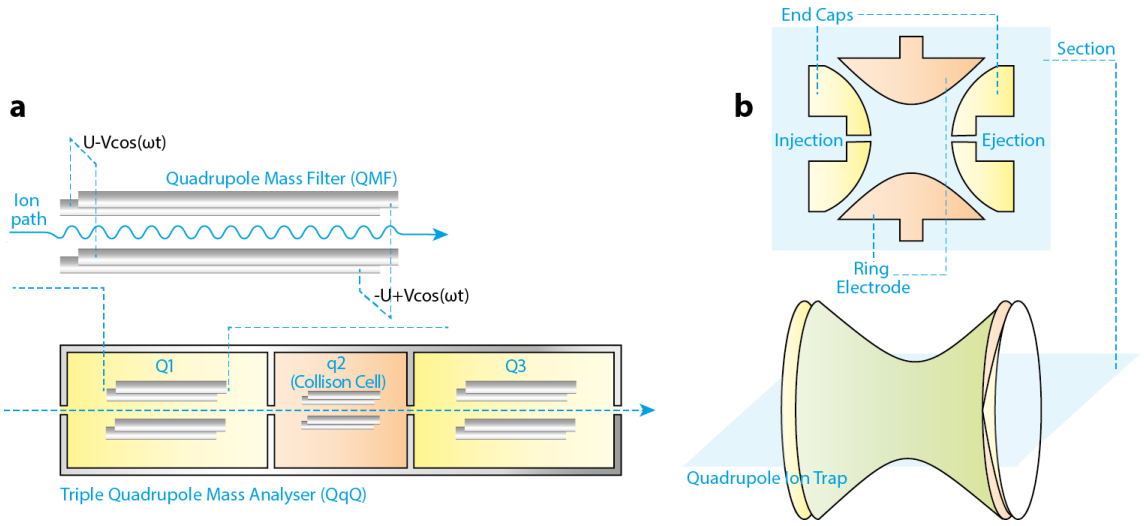


Figure 1-25 Schematic representation of the triple-quadrupole (a) and the quadrupole ion trap (QIT, b) mass analysers. The zoomed-in region in (a) shows a quadrupole mass filter (QMF) composing four parallel electrodes. A section view for the QIT is shown in the upper panel of (b).

The ejection of ions can be achieved by the resonance ejection. Briefly, an ion with a certain *m/z* resonates in the QIT with a characteristic frequency. A dipole signal with the same frequency needs to be applied to the end caps in order to eject the ions. The ejected ions can be detected by the detectors, generating final mass spectra. Notably, the ejection of ions can be selective, *i.e.* target ions can be trapped in the QIT with unwanted ions being ejected. Based on the selective ejection, certain ions can be isolated and fragmented, so that MS/MS or even MS^n experiments are possible in QIT instruments.

The QIT can be utterly sensitive, *i.e.* with a detection limit at attomolar level. Its mass range is similar to QMF, and the resolution is up to around 5000. Some of the glycoproteomic research in **Section 5.2** was carried out with the amaZon Speed ETD mass spectrometer that is a QIT instrument.

The LQIT and Orbitrap are the other two widely used trap type instruments. LQIT is very similar to QMF, although there are two major differences between them. First, LQIT needs to generate a field along the z-axis used for trapping the ions, which is absent in QMF. Secondly, LQIT needs to adopt either a radial or an axial ejection method to eject ions, whereas ions with suitable *m/z* values can move through the QMF themselves. The Orbitrap is a spindle-like instrument operated with electrostatic fields. It is an outstanding mass analyser in terms of resolution, and therefore often coupled with LQIT and QIT for the final stage mass analysis.

1.3.2.3 The Time-of-Flight and Quadrupole-Time-of-Flight

The TOF mass analyser is built upon an obvious fact that ions with different *m/z* values are accelerated to different velocities in a same electric field. It takes different time intervals for the ions to travel with different velocities through a field-free drift tube towards the detector. Thus, ions with different *m/z* can be separated and recorded into a mass spectrum.

Fundamentally, the overall energy (E_K) carried by ions when they are leaving the accelerating field is defined by:

$$E_K=zeU=(1/2)mv^2 \qquad \textit{(Eq.1-23)}$$

where *m*, *z* and *v* are the mass, number of charge and speed of ions respectively, and *U* is the potential difference of the accelerating field. If we assume the length of the drift tube is *D*, it is easy to obtain the relationship between *m/z* and the time of flight.

$$t=D/v=D[(m/z)\times 1/(2eU)]^{1/2} \qquad \textit{(Eq.1-24)}$$

The **Eq. 1-24** shows the square root of *m/z* is positively correlated with the time of flight in the drift tube, which means it takes longer time for the bigger ions to arrive at the detector.

There are several major advantages of the TOF mass analyser. Most importantly, its mass range is very wide. Theoretically, the upper mass limit of the TOF mass analyser can be infinite. In reality, singly charged ions with a mass of up to around 300 KDa have been reported to be characterised by TOF. As no scanning is required for TOF analysers, the transmission ratio of ions is very high, which leads to a high sensitivity. An attomolar level sensitivity is achievable.

Though the concept of the TOF is rather simple, the practical situation is very complicated in order to get a high resolution, especially when it is coupled with a gaseous phase ionisation source.

The major contributors of the low resolution of the MALDI-TOF mass spectrometer are the spatial and kinetic distributions. Specifically, the ions with the same *m/z* are not necessarily generated at the same position in the ion source, and they carry different initial energies. Consequently, it takes the ions with the same *m/z* slightly different time to arrive at the detector, and the difference in the time of flight broaden the MS peaks, leading to a decrease in the resolution.

The modern TOF mass analyser is normally coupled with a MALDI source where ionisation happens after short laser pulses and at the surface of a metal plate to avoid the (significant) spatial distribution of the ions. Even so, the basic MALDI-TOF mass spectrometers still struggle to obtain a high resolution.

The simplest way to increase the resolution of the MALDI-TOF instrument is to elongate the travelling time of ions in the drift tube, which can be achieved by two ways, decrease accelerating potentials and increase the length of drift tubes. However, it simultaneously decreases the sensitivity of the TOF analyser when decreasing the potential, and it is of practical difficulty to increase the length of drift tubes outside a certain range. Therefore, these simple solutions are not ideal. Two other methods, the delayed extraction and reflectron TOF, were invented within this context, which partially solves the issue of resolution.

For delayed extraction, the generated ions are allowed to move in a field free region in order to focus their energies (Vestal *et al.* 1995). After a short delay, a pulsed voltage is applied to the ions. The ions with smaller or negative (towards to back of ion plate) velocities are exposed to the accelerating voltage longer and obtain larger kinetic energies that compensate for the smaller initial energies. Careful adjustments of the duration allow for field-free moving, and the amplitude of the pulsed voltages effectively focuses the energies of the ions with different *m/z* and restores the decreased resolution. The delayed extraction method is mass-dependent. Generally speaking, a shorter duration and weaker pulse are used for focusing the energies of ions with a smaller *m/z*.

The Reflectron TOF is a mass-independent method for focusing the energies of ions, which is developed by Boris Mamyrin in 1975 (Mamyrin *et al.* 1973). Very briefly, the reflectron is an electrostatic reflector device or "ion mirror" imposed at the opposite end of the drift tube to the ion source, and it is often composed of a series of ring electrodes and grids, generating a decelerating field. Ions with higher kinetic energies travel into the decelerating field deeper than ions with lower kinetic energies, and thus taking longer time in the field. As a result, ions with different velocities take a very similar time to arrive at the detector after the reflectron, which sharpens MS peaks and increases the resolution. More recently, the multiple stage reflectron TOF has been implemented to further increase the resolution.

MS/MS experiments can be carried out by TOF/TOF instruments. In such instruments, a mass selection gate that is used for selecting ions and a collisional chamber where fragmentations occur are set in the drift tube (**FIG. 1-26**). When the instruments are operated under the MS/MS mode, only certain parental ions are allowed to go through the selection gate and enter into the collisional chamber where they collide with each other or inert gas to generate their fragment ions. These fragment ions can be re-accelerated after fragmentation and travel through the rest space of the drift tube, being reflected and detected, which generates MS/MS spectra. Such instruments are very important for MS-based glycomics, and an Applied Biosystems 4800 plus TOF/TOF mass spectrometer was used for glycomic research described in this thesis.

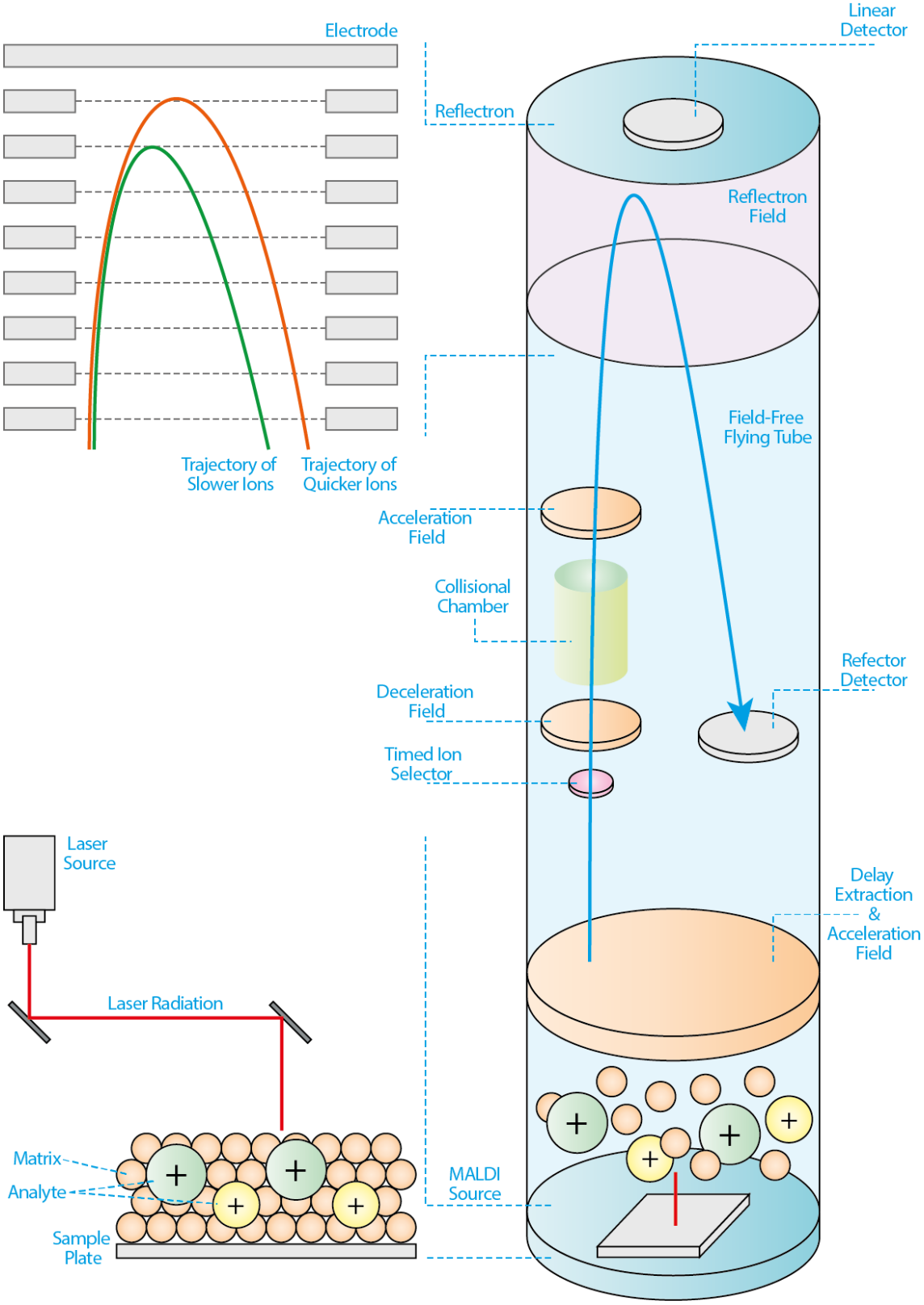


Figure 1-26 Schematic representation of an Applied Biosystem 4800 MALDI-TOF/TOF mass spectrometer. The MALDI ion source (bottom panel) and the reflectron (upper panel) are zoomed in and shown at the right side of the figure.

Though the TOF can be relatively easily coupled to MALDI sources, it was initially highly challenging to be coupled to ESI sources because of the spatial distribution of ions in the ion sources. An orthogonal acceleration TOF mass spectrometer was introduced to overcome this problem (Dawson & Guilhaus 1989). In this type of instrument, a flow of ions generated by an ESl source is focused into a thin beam that is subsequently injected and stored in a space. A voltage pulse generated by an electric field with a direction orthogonal to the moving direction of the ion beam is applied to the ions, pushing them into a TOF mass analyser to be analysed. As the ions are focused into a beam in this configuration, there is seldom any component of the initial velocities towards the TOF direction, which minimised the spatial and kinetic distribution.

The introduction of the orthogonal acceleration has inspired the invention of the hybrid quadrupole-TOF (Q-TOF) instruments (Morris et al. 1996), which is made by connecting a number of QMFs (or other multipole and LQIT) devices to an orthogonal TOF mass analyser (see the blue background region at the bottom-left of **FIG. 1-27**). The principle of this high-performance mass spectrometry is simple. A first QMF (Q_0) constantly parked at the RF mode is directly interfaced to ESI sources (or other atmospheric pressure sources), being used for focusing and guiding the ion beams. A subsequent QMF (Q_1) is used as a first mass analyser. It can either "open a wide band" to guide ions moving towards the orthogonal TOF in the MS mode or select parental ions for the MS/MS experiments. The third QMF (Q_2) allows the ions going through in the MS mode, and is essentially used as a collisional cell under the MS/MS mode. An orthogonal accelerating reflectron TOF acts as a final mass analyser, providing a wide mass range, a high sensitivity and a relatively high resolution. An Applied BioSystems QSTAR mass spectrometer and a Waters Synapt G2-Si High definition MS (HDMS) are the Q-TOF type instruments used in research described in the thesis. The former one was used for obtaining data for glycoproteomic research in **Section 4.3**, and the latter one was used for obtaining high-resolution MS and MS/MS data for proteomic and glycoproteomic research described in **Section 5.2.2**.

The Synapt G2-Si HDMS has a highly sophisticated geometry (**FIG. 1-27**), where the QMF, LQIT, mobility cell, orthogonal-TOF and dual-stage reflectron are incorporated into a single instrument. Therefore, it is used as an example to further discuss the geometry of the Q-TOF type instruments. The ions are produced in an ESI source and guided into a QMF by a two-stage step-wave ion guide, and the QMF is used as a mass filter to select ions for MS/MS mode. In this configuration, the ion guide and the QMF are essentially playing roles as the Q_0 and Q_1, respectively. The ions are subsequently guided into three sets of ring electrodes that composing an ion trap area, an ion mobility cell, and a transfer line. Both the ion trap area and transfer line can be used as collisional cells, while the mobility cell can separate ions based on their conformations. All of these ring electrodes can be considered as the Q_3. The ions are eventually separated by a TOF mass analyser with a dual-stage reflectron and detected.

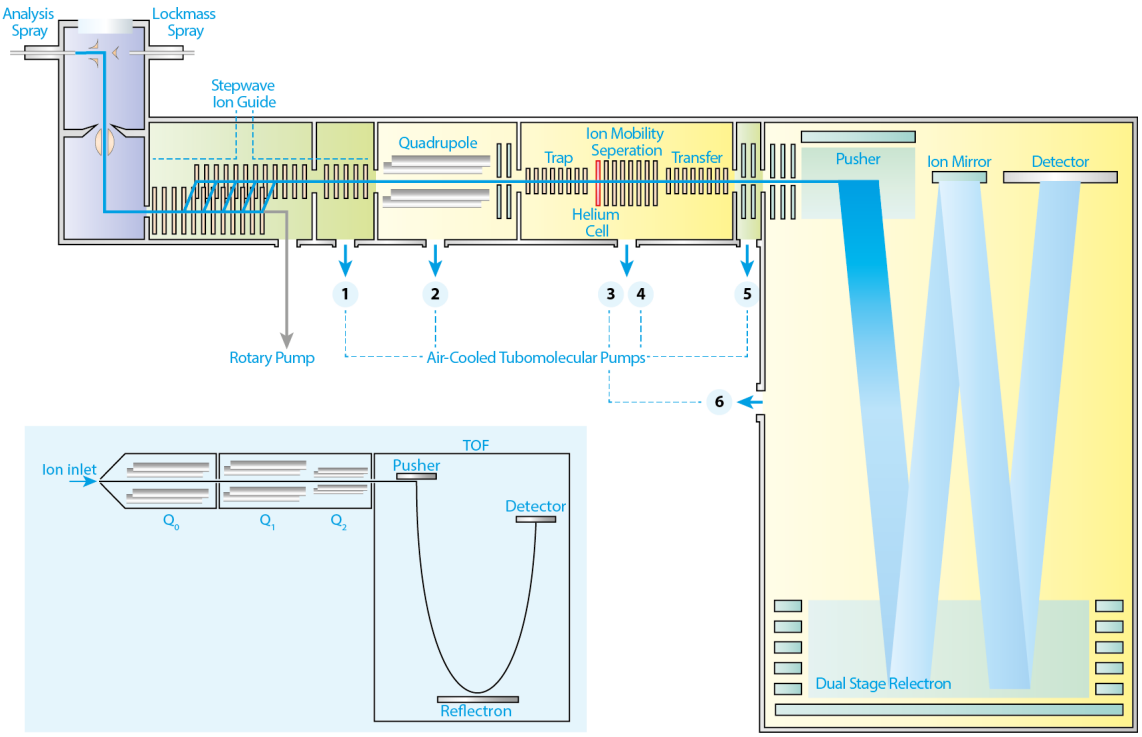


Figure 1-27 Schematic representation of the principle components of a Waters Synapt G2-Si HDMS. The fundamental geometry of a Q-TOF type instrument is shown at the left bottom with a blue background. Adapted from (Pringle *et al.* 2007).

This instrument allows a nominal resolution at the level of 10 K.

The fragmentation of the ions in a collisional cell is often referred to as collisionally activated dissociation (CAD) or CID (Paizs & Suhai 2005). For peptides, a low-energy collision induced fragmentation of a parental ion mostly produces acylium (B) or ammonium (Y) daughter ions (**FIG. 1-28 a** and **c**). The B ions can further lose a carbon monoxide to form A ions (**FIG. 1-28 b**). The sequencing of a peptide is performed by extrapolating a series of ions that potentially matches the sequence of the peptide towards the high or low end of an MS/MS spectrum. Note that the mass of amino acid residues need to be added to 1 Da (a proton) and 19 Da (a hydroxyl group plus two protons) to obtain correct masses for the N- and C- terminal ions of a peptide, respectively.

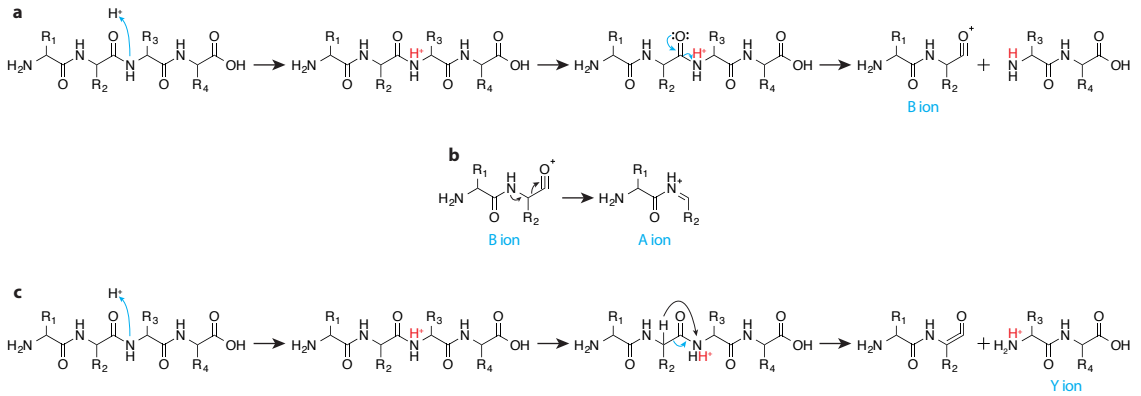


Figure 1-28 Proposed mechanisms of peptide fragmentations during the CID processes. The production of the B, A and Y ions are shown in (a), (b) and (c), respectively. The blue arrows stand for transferring electron pairs and the curved black arrow in (c) stands for proton migration.

1.4 Nuclear Magnetic Resonance

Nuclear Magnetic Resonance (NMR) is one of the most widely used analytical techniques for the structural determination of organic compounds. In the field of glycobiology, NMR can provide information on the sequences, and more importantly, the anomeric configurations of the glycans. Therefore, it can be used as a complementary technique to MS. This section briefly introduces the most fundamental theories of NMR and explains the basic ideas of micelle NMR that are used in **Section 3.2.3**. Note that this section only aims to provide a brief introduction of NMR; a comprehensive review on different NMR experiments and their mathematical and physical basis is beyond the scope of this thesis.

1.4.1 Fundamental Physics

The physics of the NMR phenomenon is briefly summarised in this section.

1.4.1.1 Nuclear Angular and Magnetic Momentum

Most nuclei have a nuclear angular momentum *P* and it is quantized,

$$P=(h/2\pi)[I(I+1)]^{1/2} \qquad (Eq.1\text{-}25)$$

where I is the angular momentum quantum number and *h* is the Plank's constant. The angular momentum $\boldsymbol{P}$ of a nucleus is associated with its magnetic momentum $\boldsymbol{\mu}$ in a proportional way.

$$\boldsymbol{\mu}=\gamma\boldsymbol{P} \qquad (Eq.1\text{-}26)$$

In this equation, both the angular momentum $\boldsymbol{P}$ and the magnetic momentum $\boldsymbol{\mu}$ are vectors and γ is a constant for a certain nucleus, called the magnetogyric ratio. It determines the sensitivity of a nucleus to NMR experiments, *i.e.* nuclei with larger magnetogyric ratios are more sensitive. For instance, the gyromagnetic ratios of the proton and ^{13}C have values of around 267.5×10^6 and 67.23×10^6 ($radT^{-1}s^{-1}$) respectively, which means protons are approximately four times more sensitive than ^{13}C.

By combining the **Eq. 1-25** and **1-26**, we can easily obtain the relationship between μ and *I*.

$$\mu=(\gamma h/2\pi)[I(I+1)]^{1/2} \qquad (Eq.1\text{-}27)$$

It is obvious that when nuclei have $I=0$, both of their *P* and μ are zero, which means they do not have a nuclear magnetic moment and cannot be observed by NMR spectroscopy. ^{12}C and ^{16}O are among the best examples for this small category of nuclei. Many other nuclei have $I\neq0$, and can be observed in NMR experiments.

1.4.1.2 The spinning Nuclei in a Static Magnetic Field

When a spinning nucleus is placed in a static magnetic field B_0, its angular momentum P takes up orientations such that its components along the direction of the B_0 are half-integral and integral numbers of $h/2\pi$,

$$P=mh/2\pi \qquad (Eq.1\text{-}28)$$

where m is the magnetic quantum number of the spinning nucleus, and can take up $(2I+1)$ values raging from I, I-1, … to -I. This phenomenon is called directional quantization. For nuclei such as protons and ^{13}C, which have I=1/2, the values of m can be +1/2 and -1/2; for nuclei such as ^{2}H and ^{14}N, which have I=1, the values of m can be +1, 0 and -1 (**FIG. 1-29**).

A classical interpretation of the nuclei spinning in external magnetic fields arises from the behaviour of a spinning top. The precession frequency or the Larmor frequency ν_L of a spinning nucleus is proportional to the magnetic flux density B_0 of the external magnetic field.

$$\nu_L=|\gamma/2\pi|B_0 \qquad (Eq.1\text{-}29)$$

Because of the directional quantization, the precessing nuclei can only take up certain orientations, and this angle for protons is 54˚55′.

The energy E of a spinning nucleus is anti-proportional to the flux density of the external magnetic field, in which this nucleus is placed,

$$E=-\mu_Z B_0=(m\gamma hB_0)/2\pi \qquad (Eq.1\text{-}30)$$

where the ratio μ_Z is the component of the magnetic momentum μ along the direction of the magnetic field, and it equals $m\gamma h/2\pi$. As the spinning nuclei take up $(2I+1)$ orientations, their energy E has $(2I+1)$ values. Nuclei with an m=±1/2, such as the proton, possess two different energy levels in an external magnetic field. When m=1/2 (α state), their μ_Z has the same direction with the magnetic field; and when m=-1/2 (β state), their μ_Z is anti-parallel to the magnetic field (**FIG. 1-30**).

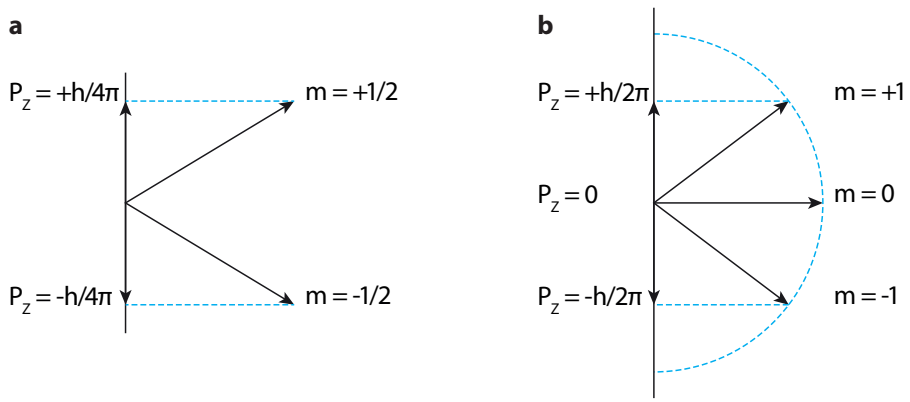


Figure 1-29 A diagram of directional quantization. Nuclei have I=1/2 and I=1 are shown in (a) and (b) respectively.

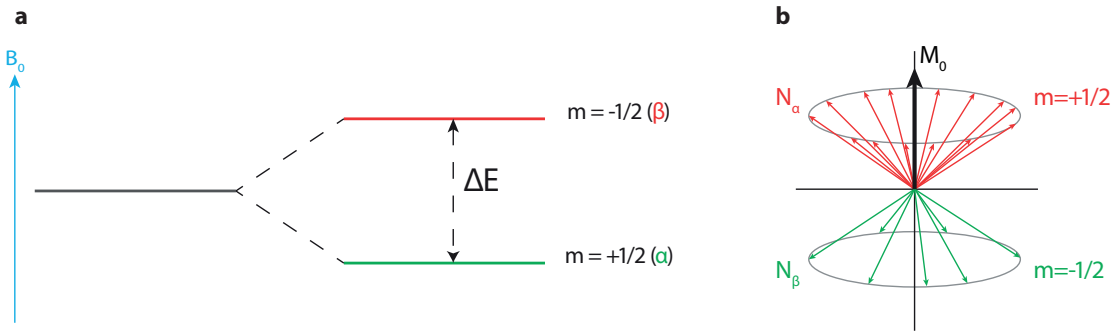


Figure 1-30 Spinning nuclei in an external magnetic field. (a) For nuclei with m=±1/2, their energy splits into two levels, an energetically lower α state (green) and a higher β state (red). (b) Nuclei in the α state have a same direction with the external magnetic field and in the β state have an opposite direction. The fact that more nuclei are in the α state results in a macroscopic magnetization M_0.

The energy difference ΔE between these two energy states can be calculated by the following equation,

$$\Delta E=-m_\beta\gamma B_0 h/2\pi-(-m_\alpha\gamma B_0)h/2\pi=\gamma B_0 h(m_\alpha-m_\beta)/2\pi \qquad (Eq.1\text{-}31)$$

Because,

$$m_\alpha-m_\beta=+1/2-(-1/2)=1 \qquad (Eq.1\text{-}32)$$

we have

$$\Delta E=(\gamma B_0 h)/2\pi \qquad (Eq.1\text{-}33)$$

It is obvious from **FIG. 1-30 a** that the α state is energetically lower than the β state, which suggests under an equilibrium condition, there are more nuclei at the α state than the β state. The distribution of the nuclei population is described by the Boltzmann statistics,

$$N_\beta/N_\alpha\approx 1-(\gamma h B_0)/(2\pi k_B T) \qquad (Eq.1\text{-}34)$$

where k_B is the Boltzmann constant and T is the thermodynamic temperature. It can be readily found that $N_\alpha>N_\beta$, further proving more nuclei are in the α state. The distribution of the nuclei on the different energy levels results in a macroscopic magnetization M_0 (**FIG. 1-30 b**).

It is worth noting that the energy difference between these two states is actually very small. As all pulsed NMR experiments involve the excitation of the nuclei at the lower energy state to the higher energy state, this small difference is important for explaining two features of NMR experiments. First, even for the most sensitive nucleus, the proton, NMR is very insensitive relative to techniques such as MS. Secondly, the sensitivity of NMR can be enhanced by increasing the strength of the external magnetic field.

1.4.1.3 The Resonance Condition and Pulse Angle

In NMR experiments, certain nuclei are transited from a lower energy level to a higher energy level by absorbing energy from an electromagnetic wave, and the transition only happens under the condition that the wave has an appropriate frequency ν_1.

$$\nu_1=\Delta E/h \qquad (Eq.1\text{-}35)$$

Nuclei like protons only have two different levels, thus only one transition needs to be considered. For nuclei (such those with I=1) that split their energy into more than two levels in an external magnetic field, the transition can only happen between two adjacent energy levels. NMR happens when the transition frequency matches the Larmor frequency. By combination of **Eq. 1-33** and **1-35**, we obtain the resonance condition.

$$\nu_L=\nu_1=|\gamma/2\pi|B_0 \qquad (Eq.1\text{-}36)$$

The term "resonance" is borrowed from classical physics, according to which the resonance of two independent waves happens when both of them are at the same frequency. For NMR, the resonance frequency falls into the range of RF.

Modern Fourier Transform (FT, see below) NMR experiments are achieved by irradiating the nuclei of interest with a short RF pulse. More specifically, an RF generator operates for a very short time τ_P (also called pulse duration, usually at μs level), which produces a continuous band of frequencies symmetrical about the central frequency ν_{RF}. The ν_{RF} is determined by the external magnetic field B_0 provided by NMR spectrometers and the nuclei of interest. When the pulse is applied to a sample, only a part of its frequency band that is proportional to τ_P^{-1} can induce the transitions. Therefore, ν_{RF} and τ_P should be correctly chosen in order to cover all the frequencies needed for the NMR experiments.

Before the pulse is applied, we can consider macroscopic magnetization M_0 spinning about the Z-axis at the Larmor frequency in a normal 3D coordinate system (**FIG. 1-31 a**); for simplicity however a rotating coordinate frame is normally used. The rotating coordinate frame is a system where the XOY plane is rotating around the Z-axis at the Larmor frequency,

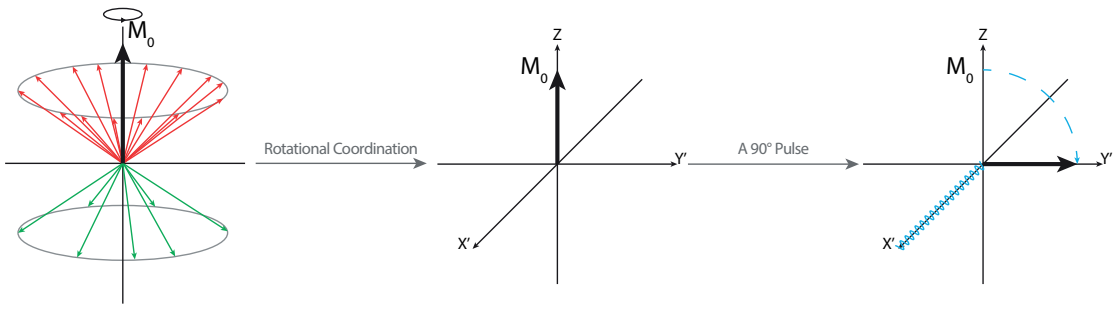


Figure 1-31 Diagrams of the rotating frame and pulse angle. In normal lab frame, the macroscopic magnetization M0 is spinning around the Z-axis. In the rotating frame, it is static and points towards the Z+ direction. After a 90° towards X'- direction, it is rotated to the Y'+ direction.

resulting in a X′OY′ plane. In the system, M0 is static and pointing towards the direction of the B_0, *i.e.* the +Z direction (**FIG. 1-31 b**). The RF pulse results in an applied field B_1 that manipulates M_0 by tipping the magnetization vector away from its original alignment with the Z-axis towards the Y-axis, along which the receiver coil is aligned. The rotational direction of M_0 is decided by the "right-hand rule". For instance, a RF towards the -X′, direction rotates M_0 to the +Y′ direction (**FIG. 1-31 c**).

The pulse angle Θ is defined by the following equation,

$$\Theta=\gamma B_{1i}\tau_P \qquad (Eq.1\text{-}37)$$

where B_{1i} is the amplitude of the component of B_1 at the frequency of NMR transition. It is obvious that the pulse angle is proportional to the duration of the pulse (τ_P) for a certain nucleus.

The most commonly used pulse angles are 180° and 90°. A 180° pulse is obtained by irradiating an RF field to the M_0 for a sufficiently long time to completely reverse the direction of M0. The microscopic result of a 180° pulse is the reversed number of the nuclei at the α and β states. After the irradiation of a 90° pulse, M_0 is rotated into the X′OY′ and the magnitude of its component in the Y′OZ plane is turned to zero. The microscopic result of a 90° pulse is a phenomenon called phase coherence. Its physical nature is beyond the scope of the thesis. A simplified explanation is that there are a small fraction of protons precessing in phase, batching together, though equal numbers of the nuclei are at the α and β states. The phase coherence is different from the saturation, when all the protons are equally distributed at both energy levels and therefore no absorption is possible.

The simplest pulsed NMR experiments involve the irradiation of a sufficiently relaxed spin system by a 90° pulse followed by recording the relaxation process that is often called free induction decay (FID). The pulse sequence is shown in **FIG. 1-32**, which is the basis of all pulsed NMR experiments, especially ^{1}H-NMR experiments in this thesis. The recorded NMR signals are in the time domain, which needs to be transferred by a mathematical treatment called the FT into the frequency domain, in order to obtain the interpretable NMR spectra.

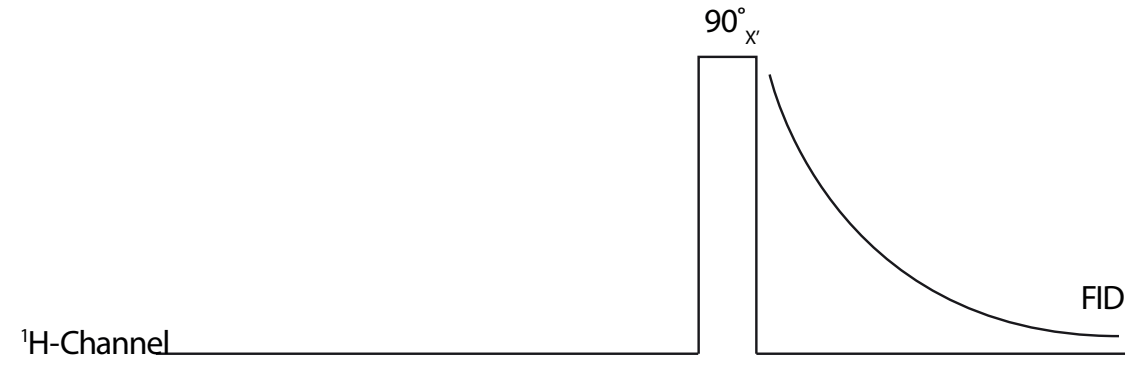


Figure 1-32 Pulse sequence of ^{1}H NMR experiments. The FID (the decay curve) is immediately recorded after a 90° pulse (the square annotated with 90°x).

1.4.2 NMR Parameters

This section introduces several important NMR parameters including the relaxation times, chemical shift and coupling constant. All conclusions provided by NMR in this thesis are drawn from these parameters. Another potentially important NMR parameter is the intensities of peaks, which can be difficult to measure for the NMR spectra of complex glycans due to their relatively low resolutions. Therefore, it is not covered in this section.

1.4.2.1 The Relaxation Times and Micelle NMR

Relaxation is a physical process, in the course of which a perturbed system returns into its equilibrium state. In NMR, relaxation is the process that restores a spin system to its equilibrium state after the irradiation of the RF pulse. The time taken for a system to relax is called the relaxation time. As there are two types of relaxation in NMR, spin-lattice and spin-spin relaxation, there are two relaxation times for the NMR experiments, *i.e.* the spin-lattice relaxation time that is often referred to as T_1 and the spin-spin relaxation time that is usually denoted as T_2. The relaxation times are important parameters for NMR spectroscopy, as they (especially T_2) determine the line width of NMR peaks.

The spin-lattice relaxation time T_1 is defined by the following equation,

$$dM_Z/dt=-(M_Z-M_0)/T_1 \qquad (Eq.1\text{-}38)$$

where M_Z is the Z-axis component of M_0. From a kinetic point of view, T_1^{-1} is the rate constant of a first-order process, during which, M_0 restores its equilibrium state after perturbations caused by pulses, *i.e.* the component of M_0 on the Z axis is re-established. The spin-lattice relaxation is directly associated with a decay of energy. In other words, the absorbed energy is given off by the system to the surrounding environments that is referred to as the "lattice".

The T_1 is affected by multiple factors including atoms covalently linked to the nuclei of interest as well as molecular size and motility. Taking ^{13}C for example, the T_1 is significantly reduced from CH_3 to CH_2 to CH groups. In addition, T_1 is shorter for larger and less mobile molecules. For the NMR spectra of LPS shown in **Section 3.2.3**, their resolutions are relatively low partially due to the large sizes of LPS samples.

The spin-spin relaxation time T2 is defined by the following equations:

$$dM_{X'}/dt=-M_{X'}/T_2 \quad ; \quad dM_{Y'}/dt=-dM_{Y'}/T_2 \qquad (Eq.1\text{-}39)$$

where $M_{X'}$ and $M_{Y'}$ are the components of M_0 at X′and Y′ axis, respectively. The spin-spin relaxation time T_2 is determined by how fast the $M_{X'}$ and $M_{Y'}$ decay. During spin-spin relaxation, the overall energy of the spin system is constant and only the phase coherence is lost. Spin-spin relaxation is a result of an inhomogeneity of the external magnetic field throughout volumes of sample molecules, *i.e.* the effective magnetic fields are

different from one nucleus to another in the same sample molecule. Therefore, these nuclei have slightly different Larmor frequencies, some of which "move" faster than the transverse magnetisation $M_{X'}$ or $M_{Y'}$, and some of which "move" slower, leading to a fanning-out of the transverse magnetisation.

The relationship between molecular mobility and the relaxation time provides the opportunity to record NMR spectra of the intact glycolipids incorporated into micelles, the technique used in **Section 3.2.3**. When LPS samples are mixed with detergents in an aqueous solution, dodecyl phosphocholine (DPC), the hydrophobic lipid A part of the LPS buries into the DPC molecules, with the hydrophilic polysaccharide part stretching out in the aqueous phase (**FIG. 1-33**). Internal flexibility of the polysaccharide in the aqueous phase gives it greater mobility than lipid A, leading to the NMR signals of the polysaccharide part being far sharper than the signals of lipid A. The sharp polysaccharide signals stand out of the broadened lipid A signals, facilitating the NMR experiments for the intact LPS, especially its O-antigen part.

1.4.2.2 The Chemical Shift and the Coupling Constant

In addition to the relaxation times, the chemical shift and coupling constant are other important factors to be considered in NMR experiments.

The chemical shift is similar to spin-spin relaxation to some extent: the effective magnetic fields imposed on different nuclei are slightly different from one another. Being different from spin-spin relaxation where the field difference is due to the instruments, chemical shift is caused by the electron cloud surrounding the nuclei.

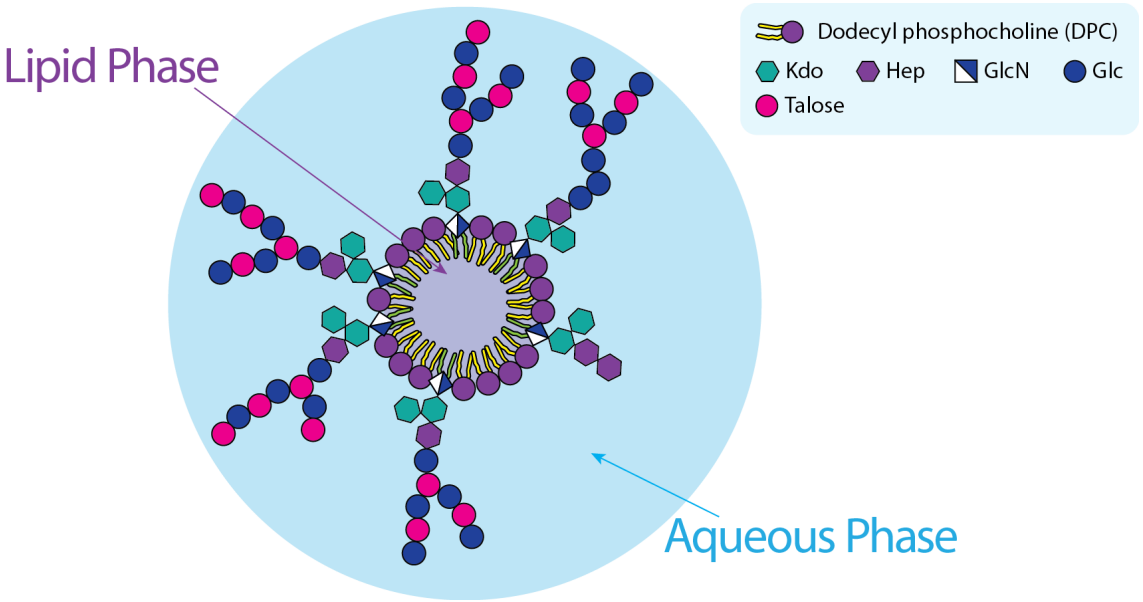


Figure 1-33 Diagram of micelle NMR. Lipid A parts of LPS molecules are embedded into DPC micelles, and the polysaccharide parts are stretching in the aqueous phase. This DPC micelle embedded LPS complex renders more mobility to the polysaccharide parts, facilitating the recording of their signals.

The electron cloud shields the nucleus, resulting in the effective magnetic field (B_{eff}) imposed on the nucleus being different from the constant external field (B_0).

$$B_{eff}=B_0-\sigma B_0=(1-\sigma)B_0 \qquad (Eq.1\text{-}40)$$

where σ is the shielding constant that is independent from the external field and solely determined by the chemical environment of the nucleus. For the proton, this constant is of the order of 10^{-5}, and is larger for bigger nuclei. The resonance condition for the shielded nuclei needs to be adjusted, and is shown in the following equation.

$$\nu_1=\gamma/2\pi(1-\sigma)B_0 \qquad (Eq.1\text{-}41)$$

As the nuclei have different shielding constants when they are in the different chemical environments, it is obviously from **Eq. 1-41** that chemically non-equivalent nuclei have different resonance frequencies.

The absolute resonance frequency is not a good parameter for NMR experiments especially when comparing NMR spectra obtained by different NMR spectrometers, as the resonance frequency is dependent on the external magnetic fields and different instruments produce different fields. For practical considerations, reference chemicals such as tetramethylsilane are often added to the samples. Values of the chemical shift are normally reported in the unit of "part per million" ("ppm"). The chemical shift of the protons in tetramethylsilane molecules are defined to be zero. For glycans, the anomeric protons often have larger chemical shifts than the non-anomeric protons, and the protons with the α-configuration often have larger chemical shifts than those with the β-configuration.

Another important phenomenon of NMR is the coupling that can be divided into two categories, the indirect coupling or the scalar coupling mediated by chemical bonds, and the direct coupling through space. The latter one is normally averaged to zero when measured in a low-viscosity liquid, therefore this section will only introduce the indirect spin-spin coupling.

The detailed physical nature of the nuclear coupling is beyond the scope of this thesis. For the indirect spin-spin coupling, a simplified explanation can be given in such a way that neighbouring magnetic dipoles within a molecule interact with each other, slightly altering their resonance frequencies. For example, when recording a proton NMR spectrum of an AX type molecule, the NMR signal of proton A and proton X each split into two signals due to the coupling between A and X. The interval between the two signals is the indirect coupling constant between A and X, which is normally denoted as $J_{A,X}$. The magnitude of the coupling constant can be affected by multiple factors such as the type of involved nuclei and the number and geometry of bonds between the nuclei.

It is possible to get rid of a certain indirect coupling by doing double resonance experiments. In order to do such experiments, RF in a second frequency in addition to the resonance frequency ν_L is applied. This new frequency equals the resonance frequency of the target nucleus that needs to be decoupled, which removes the splitting without affecting other peaks. It is possible to remove all the splittings in a spectrum by irradiating a continuous sequence of "composite" pulses, which is called the broad-band decoupling. The NMR spectra obtained by such experiments are greatly simplified, but lose all information on the coupling. This technique can be used, for example, to remove proton-carbon couplings from a ^{13}C spectrum.

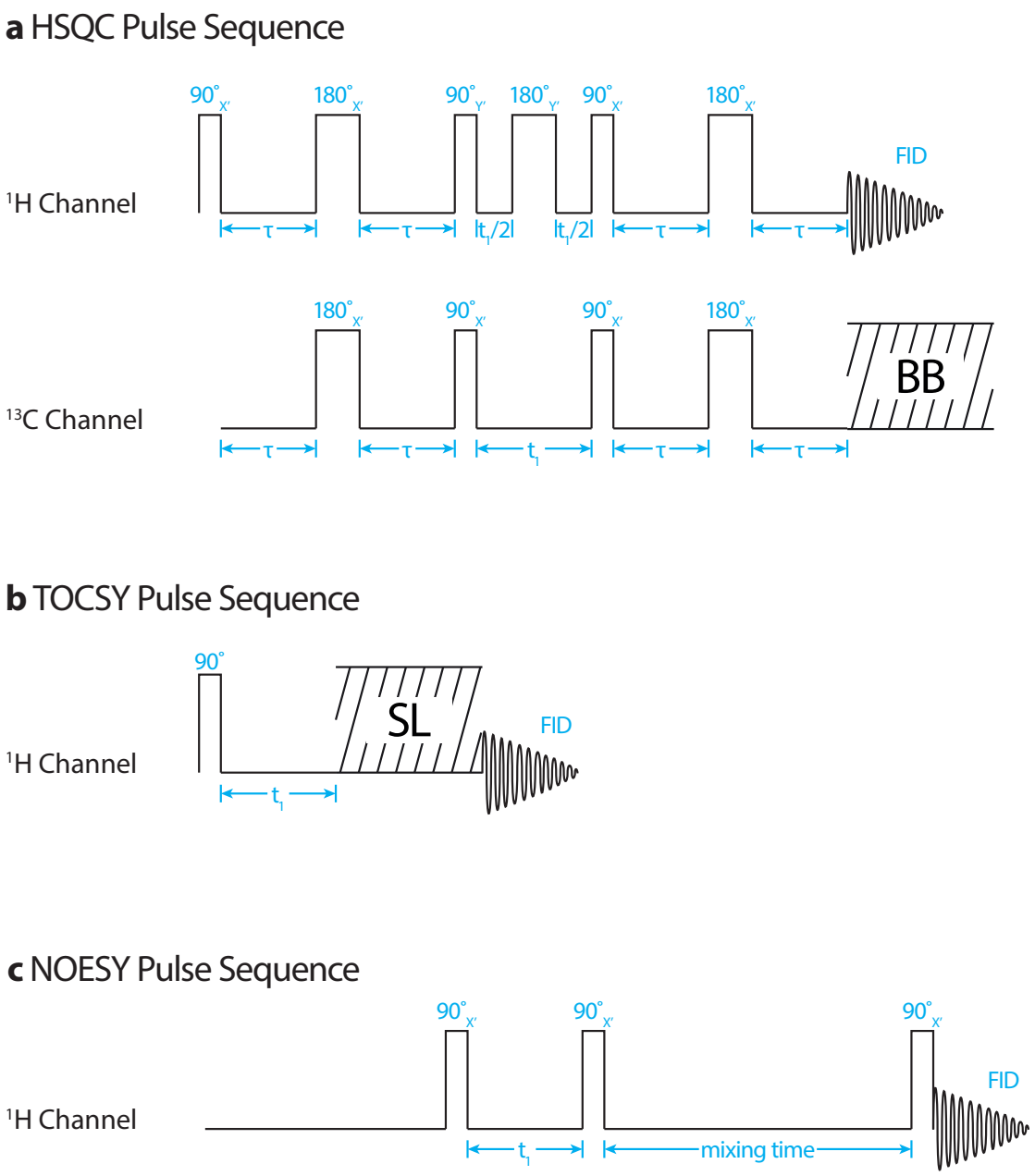


Figure 1-34 The pulse sequences of HSQC (a), TOCSY (b) and NOESY (c). The squares annotated with the pulse angles represent the RF pulses. BB stands for broad band decoupling and SL stands for spin lock.

1.4.3 Two-dimensional NMR

Though two dimensional NMR experiments have become routine in many labs today, their mathematical and physical bases are quite complex. This section provides a brief introduction on three 2D NMR experiments, Heteronuclear Single Quantum Coherence (HSQC), Totally Correlation Spectroscopy (TOCSY) and Nuclear Overhouser Effect Spectroscopy (NOESY), used in **Chapter 3** from three aspects, the pulse sequence, strategy, and information that can be drawn out of the experiments.

Fundamentally, 2D NMR experiments share a common pulse procedure, *i.e.* in addition to the preparation and detection phases in 1D NMR experiments, an additional evolution and mixing phase is inserted between them. Multiple 1D NMR experiments are accumulated in order to carry out a 2D experiment, and the evolution time (t_1) is gradually increased stepwise through the series of 1D experiments. Two FTs are needed to completely transfer the NMR signals from the time domain into the frequency domain.

HSQC is a reverse 2D heteronuclear H,C-correlated NMR technique. The pulse sequence of HSQC experiments is shown in **FIG. 1-34 a**. It adopts a two-step strategy and the pulse sequence on the proton channel is actually modified from another complex 1D pulsed experiment that enhances the NMR signals based on the polarisation transfer. The first step of HSQC is to transfer the polarisation from the protons to the ^{13}C nuclei. The system is allowed to evolve for a time t_1 and then the polarisation is transferred back to the protons in the second step. BB-decoupling of the ^{13}C channel is turned on during the collection of the proton FID. The ^{1}H magnetisations modulated by the ^{13}C resonances are eventually recorded, therefore HSQC experiments can provide valuable information on H-C correlations, *i.e.* which proton is covalently linked to which carbon.

TOCSY is an NMR experiment that provides information on the homonuclear coupling within a spin system. Taking TOCSY of an oligosaccharide as an example, it shows couplings of protons within a monosaccharide unit, which is very helpful for determining the identifications and the stereochemistries of the monosaccharides composing the oligosaccharide. The pulse sequence of TOCSY is shown in **FIG. 1-34 b**, which is directly adopted from a 1D selective TOCSY experiment. After a 90˚ degree pulse of the selective protons, the spin system is allowed to evolve for a gradually incremented time t1 followed by a spin lock as a mixing phase. During this phase, the scalar coupling is dominating the chemical shift and the polarisation is transferred from the selective protons to others. The FID is immediately recorded after the mixing phase. In this way, ^{1}H magnetisations modulated by the scalar couplings are recorded in TOCSY, and therefore the cross peaks in 2D TOCSY spectra indicate two protons are coupling with each other through chemical bonds.

NOESY is a powerful NMR technique that characterises the 3D structures of molecules. Its pulse sequence is shown in **FIG. 1-34 c**. When the sequence is applied to a proton-proton system where they are near in space but no scalar coupling exists between them, the first 90° pulse rotates both magnetisations to the transverse direction, and they are allowed to evolve around the vertical direction with their Larmor frequencies for a time t1 that is gradually increased. As the Larmor frequencies of these two protons are not identical, when the second 90° pulse turns both magnetisations back to the vertical direction, their magnitudes are different. During the mixing phase between the second and third 90° pulses, the polarisation transfers between these two protons. The resulting magnetisations are turned to the transverse direction again by the final 90° pulse, so that the FID can be recorded. Two FTs then generate a 2D NOESY spectrum showing cross-signals that are indicative of a proximity in space between the two protons.

Chapter 2 Material and Methods

This chapter describes material and methods used in research included in this thesis. For material, sources of chemical agents, enzymes and consumables are listed (**Section 2.1.1**), and types of chromatography, MS and NMR instruments are introduced (**Section 2.1.2**). The softwares used for analysing the data are also included (**Section 2.1.3**). The biological samples obtained from our collaborators are detailedly discussed in **Section 2.2** (note that the preparations of samples were not carried out by the author, relevant information was provided by my collaborators). All the experimental procedures are listed in **Section 2.3**.

2.1 Material

2.1.1 Chemical Reagents, Enzymes and Consumables

Alfa Aesar (Lanchashire, UK): Acetic anhydride, methyl iodide and deuterated methyl iodide;

BOC (Guildford, UK): Argon, helium and nitrogen gases;

Cambridge Isotope laboratories, Inc. (Andover, USA): dodecylphosphocholine (D38, 99%);

Thermal Scientific Inc. (USA): Tri-Sil TP derivatising agent

ROMIL (Waterbeach, UK): acetonitrile (MeCN), ammonia, acetic acid, methanol (MeOH), chloroform, propan-1-ol (PrOH), dimethylsulfoxide, sodium hydroxide, trifluoroacetic acid and formic acid.

Sigma-Aldrich Corporation (Poole, UK): 2,5-dihydrosylbenzoic acid, iodoacetic acid (IAA), potassium borohydride, sodium acetate, hexanes, sodium borodeuteride, pyridine, acetic anhydride, ammonium bicarbonate (AMBIC), deuterium oxide (99%), deuterium oxide ("pure", 99.990%), chromium (VI) oxide, glycerol, dodecylphosphocholine (D38-DPC, 99%), dithiothreitol (DTT), monosaccharide standards (glucose, galactose, GlcNAc, GalNAc, heptose, rhamnose, fucose, arabitol) and Pur-A-Lyzer™ Mini Dialysis Kit (1 KDa, MWCO)

Promega Corporation (Hampshire, UK): trypsin (porcine)

Applied Biosystems (Warrington, UK): 4700 Mass Standards kit

Fisher Scientific (Loughborough, UK): Slide-A-Lyzer® dialysis cassettes (3.5 KDa, MWCO)

Invitrogen (Paisley, UK): NuPAGE® SDS-PAGE Midi Gel System (Tri-acetate, pre-cast), NuPAGE® LDS sample buffer, NuPAGE® reducing agent, NuPAGE® Tri-acetate SDS running buffer, Novex Colloidal blue staining kit and Coomassie R-250

2.1.2 Instruments

API Q-STAR Hybrid LC-MS/MS system on-line linked to a Dionex UltiMate 3000 liquid charomatography (LC) fitted with a Pepmap® C18 nanocapillary column (75 μm × 150 mm);

Applied Biosystems 4800 plus MALDI-TOF/TOF mass spectrometer;

Bruker 456-GC/SCION™ SQ instrument fitted with a BR-5ms column (15 m, 0.25 mmID, 0.25 μm; composition: 5% phenyl 95% dimethyl arylene siloxane);

Bruker amaZon Speed ETD mass spectrometer on-line linked to a Dionex UltiMate 3000 LC fitted with a Thermo Acclaim PepMap® RSLC column (75 μm × 25 cm, nanoViper, C18, 2 μm, 100 Å);

Bruker Avance III 600 MHz, fitter with a TXI/TCI cryoprobe;

Dionex UltiMate 3000 LC systems fitted with a SeQuant® ZIC®-HILIC column (5 μm, 200 Å; PEEK 250 mm × 4.6 mm);

Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX®-5MS column (30 m 0.25 mmID, 0.25 μm; 5% diphenyl 95% dimethylpolysiloxanephase);

Perspetive Biosystems Voyager-DE STR system MALDI mass spectrometer;

Waters SYNAPT G2-S mass spectrometer equipped with a nano-flow ESI source on-line linked to a Waters ACQUITY UPLC® M-Class system fitted with an ACQUITY UPLC® column (HSS T3 1.8 μm 75 μm × 150 mm).

2.1.3 Softwares

The MALDI-TOF and TOF/TOF data were interpreted by using a Data Explorer software (version 4.9);

The amaZon data were interpreted by using a Compass DataAnalysis software (version 4.2 SR2);

The Q-Star data were interpreted by using an Analyst TF software (version 1.6);

The SYNAPT data were interpreted by using a MassLynx software (version 4.1);

The GC-MS data were interpreted by using either an MS data review (version 8.0.1) or a TurboMass (version 6.1) software;

The Mascot search was performed by either a Bruker Daltonics BioTool (version 3.2 SR2) or ProteinScape (version 3.1) software;

A GlycoWorkBench software (version 2.1) was used to facilitate understanding glycan structures.

2.2 Samples

2.2.1 *Pseudomonas aeruginosa* LPS Samples

P. aeruginosa serotype O5 (PAO1) LPS samples were obtained from Dr. Alan Brown's lab at the University of Exeter, UK. A $MgCl_2$-ethylenediaminetetraacetic acid (EDTA) method was used to extract LPS samples from two WTs and three mutants.

Two WT PAO1 were cultured in media containing 2 mM and 20 μM Mg^{2+}, respectively. The mutant include a PhoQ knock-out, and PAO1 with a site mutation on PmrB[M292T] and PhoQ[V260G]. Overlap extension polymerase chain reaction (PCR) was used to construct either a fragment mapping to the flanking regions of PhoQ (for knock-out), or PhoQ-/PmrB-specific fragments that harboured altered coding sequence within the overlap region (for single-site mutation). The fragments were cloned into pEX19Gm, and the resulting plasmids were mobilised into PAO1 by triparental mating with the pRK2013 helper plasmid as previously described (Hoang *et al.* 1998). pEX19Gm integrants were selected on Lysogeny broth agar supplemented with gentamicin and trimethoprim, and the integrants were assessed to confirm the expected gentamicinr/sucroses phenotype and the presence of the construct (by PCR), prior to being resolved by plating on Lysogeny broth agar supplemented with 5 % sucrose. Resulting colonies were screened for the presence of the desired mutation by PCR. Genome sequencing of the resulting mutants was performed to confirm that the desired gene replacement had occurred without lesions.

2.2.2 *Helicobacter pylori* LPS Samples

H. pylori G27 LPS samples were obtained from the Ondek Corporation, Australia. Either a $MgCl_2$-EDTA or a hot-phenol/water method was used to extract LPS samples from WT G27 and seven mutants including ΔWaaL, ΔHP1284, ΔHP1578, ΔHP0102, ΔHP0156, ΔHP1283 and ΔHPG27_1230.The amount of samples used for the micelle NMR experiments are listed in **Table 2-1**.

Clean deletion of HP1283, HP1284 and HP0102 in the G27 genome was made using counter-selection, based on G27ΔHP1283::RC, G27ΔHP1284::RC and G27ΔHP0102::RC, respectively. G27HP1283::RC, G27HP1284::RC, G27HP0102::RC and G27HP0805::RC were made by inserting *rpsL-cat* into the corresponding gene using counter-selection method. Brief, HP1283, HP1284, HP0102 and HP0805 genes were amplified from 26695 genomic DNA, and the PCR products were then individually TA-cloned into pCR®2.1-TOPO® vector (Invitrogen Life Technologies) to give pCR2.1-1283, pCR2.1-1284 and pCR2.1-0102. The *rpsL-cat* cassette was liberated from pENT-RC by *Bgl*II (New England BioLabs) digest and inserted into above-constructed pCR2.1 derivatives using MuA transposase (Thermo Fisher Scientific) *in vitro* transposition to give pCR2.1-1283-RC, pCR2.1-1284-RC and pCR2.1-0102-RC, which were used to transform strain G27 to $G27^{Cm}$. Colonies growing on chloramphenicol selection plates were subcultured and diagnostic PCR using chromosomal DNA was performed to confirm the insertion of *rpsL-cat* into the targeted genes. The genomic DNA from G27HP1283::RC, G27HP1284::RC and G27HP0102::RC were used to transform strains 26695 to generate 26695HP1283::RC, 26695HP1284::RC and 26695HP0102::RC, respectively.

For the construction of the clean deletion strains, two DNA fragments flanking to the target gene were amplified from 26695 genomic DNA using corresponding primers. The two DNA fragments were joined by splicing by overlap extension PCR. The PCR product was used to transform G27 RC strains. Colonies growing on streptomycin selection plates were subcultured and diagnostic PCR using chromosomal DNA was performed to confirm the deletion of target genes.

Table 2-1 The H. pylori LPS samples

Sample	Mass (mg)
G27 WT	8.7
ΔWaaL	7.5
ΔHP1284	12
ΔHP1578	7
ΔHP0102	10
ΔHP0156	8.4
ΔHP1283	7.8
ΔHPG27_1230	9.1

The ΔWaaL, ΔHP1578, ΔHP0156 and G27ΔHPG27_1230 were constructed by difH/XerH recombination. Briefly, two DNA fragments flanking to the target gene were amplified from G27 genomic DNA using corresponding primers. The two DNA fragments were joined by splicing by overlap extension PCR to give a fragment containing the DNA regions flanking the target gene separated by a unique *Bam*HI site. This PCR product was treated with My*Taq*™ DNA polymerase (Bioline Meridian Life Science) and ligated into pGEM®-T Easy Vector (Promega). The difH flanked *rpsL-cat* cassette was liberated from pDifWT-RC by *Bam*HI (New England BioLabs) digest and cloned into the unique *Bam*HI site, which was used to transform strain G27 to $G27^{Cm}$. Colonies growing on chloramphenicol selection plates were directly subcultured onto plates supplemented with streptomycin. Single colonies were subcultured onto nonselective plates and diagnostic PCR using chromosomal DNA was performed to confirm the deletion of target gene.

Note that the above three paragraphs were adapted from my collaborator, Hong Li's PhD Thesis, **Lipopolysaccharide biosynthesis and protein glycosylation in *Helicobacter pylori***.

2.2.3 Fap1 Protein Samples

Fap1 proteins were obtained from Prof. Hui Wu's lab at the University of Alabama, USA. The native Fap1 protein purified from *S. parasanguinis* FW213 and recombinant Fap1 proteins (**Table 2-2**) co-expressed with different putative GTs in the *E. coli* BL21 strains were analysed.

To obtain differentially modified Fap1 ("mutant" Fap1), plasmid pVPT-Gtf12, pVPT-Gtf123, pVPT-Gtf123-dT1, pVPT-Gtf123-dGT1-Gly or pVPT-Gtf123-dGT1-GalT2 was co-transformed with pET-28a-rFap1-R1 and pHSG576, respectively, into BL21. To produce fully modified Fap1 ("mature" Fap1), plasmids pET-28a-rFap1-R1, pVPT-Gtf123-dGT1-GalT2 and pHSG576-Gly were co-transformed into BL21. All plasmids were confirmed by DNA sequencing (above information is from the submitted manuscript ***Engineering and dissecting the glycosylation pathway of a Streptococcal serine-rich repeat adhesin***). The Fap1 were expressed and purified using the same method described previously (Zhang *et al*. 2014). The recombinant Fap1 proteins only contains a short RI domain, and its sequence is listed below:

MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRDPANEVKLSSENFDSEKAEEK
ISLSQSESASESVSESISESVSESVSTSESVSESVSESVSESISESVSESISESISESVSESTST
SIVLSESGAASGNKAT

The native Fap1 proteins were purified as previously descrbied (Wu *et al*. 1998).

Table 2-2 Recombinant Fap1 proteins

No. of Samples	Putative GTs
1	GTF1/2
2	GTF1/2/3
3	GTF1/2/3 GALT1
4	GTF1/2/3 GALT1/2
5	GTF1/2/3 GALT1/2 GLY
6	GTF1/2/3 DUF1792

2.2.4 Bacterial Whole Cell Lysates

Various bacterial whole cell lysates were obtained from Prof. Brendan Wren's lab at LSHTM. The samples were extracted from *B. cenocepacia* (*B. cenocepacia*) K56-2, *Burkholderia thailandensis* E264 WT, *pglL* mutant and *Coxiella burnetii* (*C. burnetii*).

For *pglL* mutation, downstream and upstream homologous DNA fragments flanking to the target gene (~1200 bp) were amplified using Phusion™ High-Fidelity 2x PCR master mix with GC buffer with 6% DMSO from purified genomic DNA (PureLink Genomic DNA Kit, Invitrogen). Amplification conditions were as follows: denaturation at 98 °C for 30 s, 25 cycles of amplification at 98 °C for 10 s, annealing at 62 °C, and extension at 72 °C at 1 min. 30 s, a final extension of 72 °C for 10 min. These fragments were assembled in a splicing-by-overlapping-extension PCR and purified by using a gel extraction kit (MinElute Gel Extraction Kit Qiagen). The insert was ligated into the pJET1.2/blunt cloning vector (Thermo Fisher Scientific) and was transformed into *E. coli* Top10 cells (Invitrogen Life Technologies). Clones were screened by PCR and Sanger sequencing. The deletion construct was sub-cloned via ligation at a 1:3 vector insert ratio (NEB T4 ligase) into the XbaI-digested dephosphorylated pDM4 vector via the restriction sites and transformed into *E. coli* MFD cells. Overnight cultures of the donor strain (*E. coli*) and the recipient strain (*B. thailandensis* E264) were mixed in a 1:3 (v/v) ratio. 100 µl of the mixture was applied to a pre-warmed Lysogeny broth agar and incubated for no longer than 16 hours at 37 °C. Cells were resuspended and serially diluted in PBS and cultured overnight at 37 °C on selective Lysogeny broth agar containing chloramphenicol. *B. thailandensis* colonies were screened for presence of the deletion fragment.(above information is from a paper manuscript under preparation)

The *Burkholderia* whole cell lysates were harvested according to a published protocol (Velapatiño *et al*. 2013). Around 1 mg dried bacterial whole cell lysates were resuspended in the aqueous solution of 6 M urea, 2 M thiourea, 50 mM AMBIC. The de-natured proteins were incubated in a 500 µL AMBIC buffer containing 20 mg/mL DTT for 1 h at room temperature. The solution was subsequently added to 500 µL AMBIC containing 120 mg/mL IAA, and the mixture were stood at room temperature in the dark for 1 h. The samples were then dialyzed at 4 ˚C against a 4.5 L AMBIC buffer overnight using a mini dialysis kit (Sigma-Aldrich, Pur-A-Lyzer™ Mini Dialysis Kit, 1 KDa cut-off).

The *B. thailandensis* periplasmic proteins were prepared based on previous research (Feldman *et al*. 2005). The periplasmic extracts were performed by osmotic shock lysis. The cells were incubated at 0°C for 2–4 h in an acqueous solution containing 20% sucrose, 30 mM Tris Cl (pH 8.0) and 1 mM EDTA, and then in 5 mM $MgSO_4$ (20 OD 600 units/ml). Centrifugation was used to enrich the periplasmic proteins in the supernatant.

The *C. burnetii* whole bacterial lysates were prepared based on previous methods (Deringer *et al.* 2011). The bacterial cells were suspended in phosphate-buffered saline (PBS) with 0.25 M sucrose and were centrifuged at 6000 g for 10 min at room temperature. The supernatant was discarded and the precipitated cells were washed in cold Tris-buffered saline (TBS; 50 mM Tris/HCl, 150 mM NaCl, pH 7.4) and subjected to centrifugation again. The pelleted cells were re-suspended and sonicated in 250 µL 40 mM Tris buffer (pH 8.8) containing 60 U benzonase, phenylmethane sulfonyl fluoride, bug buster, 1 mg/mL lysozyme and 2 mM $MgCl_2$. Four rounds of 15 min pulses were applied to the cells. The sonicated samples were ultra-centrifuged at 21952 g for 15 min, and the supernatant was collected and transferred to a fresh protein low-bind microfuge tube. The obtained proteins were then de-natured by a 40 mM ammonia bicarbonate solution containing 2M thiurea, 12M urea and of 10 mM DTT.

2.3 Methods

Methanolysis. The LPS sample was dissolved in a 0.5 M methanolic-HCl solution and incubated at 50 °C for 40 min. The supernatant was collected and then dried under a stream of nitrogen.

Mild Acid Hydrolysis. The LPS sample was dissolved in 200 µL of a 0.1 M aqueous solution of sodium acetate (pH=4.4, adjusted by acetic acid) and incubated at 100 °C for 4 hrs. The produced lipid A molecules were extracted with a mixture of chloroform and MeOH (4:1, V/V). Chloroform and MeOH were dried under a stream of nitrogen.

Mild HF Hydrolysis. The LPS sample was hydrolysed by HF as previously described (Haslam *et al.* 1997) with minor modifications. The LPS samples were dissolved in 50 µL of 48% HF and incubated at 4°C for 24 hrs. The reagent was removed under a stream of nitrogen.

Mild Periodate Oxidation. The LPS sample was dissolved in 100 µL of a 20 mM ammonium acetate (100 mM, pH=6.5) solution of sodium periodate. The mixture was incubated at 4 °C for 20 hrs in the dark, and the reaction was terminated by adding a drop of ethylene glycol and keeping at room temperature for 60 min. The solution was lyophilised, and 400 µL of a 10 mg/mL ammonium hydroxide (2 M) solution of sodium borohydride was added to the dried sample and incubated at room temperature for 2 hrs. The reaction was quenched by adding 3-5 drops of acetic acid and purified by a 50E-8C Dowex® ion-exchange resin.

Reductive Elimination. The O-glycans were reductively eliminated from glycoprotein samples according to published methods (Haslam *et al.* 2006). Briefly, the lyophilised protein sample was dissolved in 400 µL of a 55 mg/mL 0.1 M potassium hydroxide solution of potassium borohydride, and was incubated at 45 °C for 18 hrs. The reaction was quenched by adding 5-6 drops of acetic acid, and the eliminated O-glycans were purified by the 50E-8C Dowex® ion-exchange resin.

50E-8C Dowex® Ion-exchange Purification of Glycans. The O-glycans were purified by a Dowex® ion-exchange column according to published methods (Haslam *et al.* 2006). The glycan was dissolved in 5% acetic acid and loaded onto the Dowex® column. The same solution was used for elution. The collected solution was concentrated and lyophilised. Excessive borates were co-evaporated with 10% methanolic acetic acid under a stream of nitrogen.

Permethylation and Deuteropermethylation. The glycan (glycoconjugate) samples were permethylated according to published methods (Haslam *et al.* 2006). Sodium hydroxide (three to five pellets per sample) was crushed in 3 mL dry dimethyl sulfoxide. The resulting slurry (0.75 mL) and 0.85 mL methyl iodide (deuterated methyl iodide) were added to the sample, and the solution was agitated at room temperature for 60 min. The reaction was quenched by

adding 2 ml ultrapure water with shaking, and samples were extracted with chloroform (2 ml). The chloroform solution of the samples was washed with ultrapure water for two times before chloroform was removed under a stream of nitrogen.

Sep-pak® C18 Purification of Glycans. The Sep-pak® C18 column was conditioned by successively washing it with MeOH (5 ml), ultrapure water (5 ml), MeCN (5 ml) and ultra-pure water (3×5 ml). The dried glycan sample was dissolve in 100 µL MeOH and added in 100 µL ultrapure water. The solution was loaded onto the column and the glycans were washed with 5 mL ultrapure water and successively eluted with 3 mL of 15%, 35%, 50%, 75% or 15%, 50%, 75% aqueous solutions of MeCN. All the MeCN eluents were collected and lyophilised, and the samples are ready for spotting on a MALDI plate.

Peracetylation. A previously described method was used for peracetylation (Haslam *et al.* 1997). Glycans were incubated in a 400 µL pyridine/acetic anhydride (v/v=1:1) mixture at 80 °C for 3 hrs, and the reagents were removed under a stream of nitrogen. The acetylated glycans were subsequently dissolved in chloroform and washed three times with ultrapure water. The chloroform was removed under a stream of nitrogen.

Chromium (VI) Oxidation. These experimental procedures were adapted from a previous study (Khoo & Dell 1990). 10 mg CrO_3 was added to 100 µL of acetic acid, and the resulting slurry was used for oxidising peracetylated glycans at 50 °C for 3 hrs. After quenching the reaction with water, the product of oxidation was extracted with chloroform, washed with ultrapure water twice. The chloroform was removed under a stream of nitrogen.

In-Solution Digestion. The protein pellets were incubated in 500 µL of a 2 mg/mL 50 mM AMBIC solution of DTT at 50 C˚ for 60 min. 500 µL of a 12 mg/mL 50 mM AMBIC solution of IAA was added to the solution before an incubation in the dark at room temperature for 90 min. The solution was then dialyzed for 36- 48 hrs in a mini dialysis kit with a molecular mass cut-off of 1 KDa at 4 ˚C against a 50 mM AMBIC buffer that is regularly replaced every 12 hrs. The dialyzed sample was collected and lyophilised, and then added in trypsin (1/50, w/w) and a sufficient amount of ABMIC buffer (50 mM, pH=8.4 adjusted with ammonia) to make sure all samples are covered by the solution. The solution was incubated at 37 °C overnight, and the digestion was quenched by heating the solution to 100 °C.

Sep-pak® HLB plus C18 Purification of Peptides. The Sep-pak® HLB plus C18 column was conditioned by successively eluting MeOH (5 ml), 5% acetic acid (5 ml), PrOH (5 ml) and 5% acetic acid (5 ml). The AMBIC solution of peptide (see **In-Solution Digestion**) was loaded on the column and successively eluted with 10 mL 5% acetic acid and 4 mL of 20%, 40% and 100% aqueous solution of PrOH. The eluents were collected and lyophilised.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE Purification of Glycoproteins). The gel cassette was washed with ultrapure water and the wells were washed with the running buffer before usage. The protein sample and benchmark standard was dissolved in the sample buffer. The inner chamber of the electrophoresis instrumentation was filled with the running buffer. After making sure there was no leakage, the solutions of the protein and the standard were loaded into the gel wells and the outer chamber was filled with running buffer. The electrophoresis was kept running under a constant 200 V, until the front line reached the bottom of the gel. The collected gel was added to a fixing solution containing 50 ml ultrapure water, 40 ml MeOH and 10 ml acetic acid and incubated for 10 min at room temperature with shaking. 20 mL Stainer A (Colloidal Blue Staining Kit, see **Section 2.1.1**), 55 ml ultrapure water, 20 ml MeOH were premixed and added to the fixing solution. The gel was incubated for 10 min at room temperature with shaking. The gel was further incubated overnight after 5 mL Stainer B (Colloidal Blue Staining Kit, see **Section 2.1.1**) was added to the solution.

Destaining of the Gel. The gel was put on a piece of clean glass and each of the stained parts were excised and collected separated. Each of the excised gels was further diced into small cubes and moved to a clean Eppendorf tube. For each tube, 200 µL of 50 mM AMBIC (pH 8.4. adjusted with ammonia) and 200 µL of MeCN were added to the gel pieces. The gel pieces were incubated at room temperature for 5 min and centrifuged. The supernatant was then removed and the gel pieces were dried in a SpeedVAC instrument.

In-Gel Digestion. The dried gel pieces were incubated with 200 µL of a 50 mM AMBIC (pH 8.4. adjusted with ammonia) solution of DTT (10 mM) at 56 °C for 30 min. The DTT solution was removed and the gel pieces were washed with 200 µL MeCN. MeCN was dried in a SpeedVAC instrument. The dried gel pieces were then incubated with 200 µL of the 50 mM AMBIC solution of IAA (55 mM) in the dark for 30 min. The IAA solution was removed and the gel pieces were washed with 200 µL MeCN. MeCN was dried in a SpeedVAC instrument. The dried gel pieces were incubated at 37 °C overnight in 20 µL of the AMBIC solution containing 0.5 µg trypsin. The digestion was quenched by heating the solution to 100 °C. The AMBIC solution was collected. And the gel pieces were successively washed with 200 µL of the AMBIC solution, the AMBIC solution/MeCN (1:1, V/V) and MeCN. All the solutions were collected and dried in a SpeedVAC instrument.

GC-EI-MS linkage analysis. Partially methylated alditol acetates were prepared as previously described (Haslam *et al.* 2006). Either a Bruker 456-GC/SCION^{TM} SQ instrument fitted with a BR-5ms column or a PerkinElmer Clarus 500 instrument fitted with a RTX®-5MS column was used to carry out the experiments. For the former instrument, the sample was injected into the column at 60 °C, and the temperature increases to 300 °C over 30 min at a rate of 8 °C/min; for the latter instrument, the sample was injected into the column at 65 °C,

and the temperature increases to 140 °C at a rate of 25 °C/min, to 200 °C at a rate of 5 °C/min, to 300 °C at a rate of 10 °C/min and held at 300 °C for 5 min.

GC-EI-MS TMS analysis. Monosaccharide TMS derivatives were prepared as previously described (Zhang *et al*, 2014). A Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX®-5MS column was used to carry out the experiments. The following temperature programme was used: the sample was injected into the column at 65 °C. The temperature was increased to 140 °C at the rate of 25 °C/min, and then to 200 °C at the rate of 5 °C/min. The temperature was finally raised to 300°C at a rate of 10°C/min and was held for 5min.

LC-ESI-MS. LC-ESI-MS experiments were performed by three MS instruments (the Q-STAR amaZon and SYNAPT).

The following LC gradient was used for Q-STAR: the column was eluted from 95.5 % (v/v) water/acetonitrile solution to 95.5 % (v/v) acetonitrile/water solution. Both solutions contained 0.05% formic acid. Note that this gradient was not directly operated by the author.

For amaZon, the LC was constantly running at 0.3 μL/min flow rate and 35 °C. It was initially conditioned with 97/3 (v/v) solution A (0.1% formic acid in ultra-pure water) / solution B (0.1% formic acid in 80% aqueous solution of MeCN) for 4 min. The concentration of solution B was increased to 40% over 30 min, kept for 1 min, further increased to 99% over 4 min, decreased to 3% over 1 min and kept for 20 min.

A longer LC programme was used for SYNAPT. Briefly, the LC was also constantly running at 50 °C with a flow rate of 0.3 μL/min. The LC was initially eluting with 97/3 (v/v) solution A (0.1% formic acid in ultra-pure water) / solution C (0.1% formic acid in MeCN), and the concentration of solution C was gradually increased to 50% over 90 min and kept for 5 min. The concentration of solution A was increased back to 97% within 0.5 min and kept for 14.5 min.

Zwitterion–ion hydrophilic interaction chromatography (ZIC-HILIC) enrichment of glycopeptides. Bacterial protein extracts were dissolved in 80% aqueous solution of MeCN containing 5% formic acid. The ZIC-HILIC column was conditioned with the same solution at flow rates decreased from 2.5 mL/min to 500 μL/min over 8 min before sample injection. The flow rate was set at 500 μL/min for the rest of the program. The column was eluted with 80% aqueous solution of MeCN containing 5% formic acid for 20 min, and this solution was replaced by 5% formic acid over 6 min. The column was then eluted with 5% formic acid for 10 min, and 5% formic acid was replaced by 80% aqueous solution of MeCN containing 5% formic acid over 5 min. 0-10, 10-20, 20-26, 26-33 min eluents were collected, and the glycopeptides were contained within the 26-33 min eluents. This eluent was lyophilised and ready for LC-ESI-MS analysis.

MALDI-TOF and TOF/TOF. Either a Voyager DE-STRTM MALDI–TOF or a 4800 plus MALDI–TOF/TOF mass spectrometer was used for recording MALDI-TOF spectra. MALDI-TOF/TOF spectra were acquired with the latter instrument. The 4700 Calibration standard kit and fibrinopeptide B was used for calibrating TOF and TOF/TOF mode, respectively. The collision energy for TOF/TOF was set to 1 kV, and the collision gas was argon. 2, 5-dihydroxybenzoic acid and 3,4-diaminobenzophenone were used as matrix. Permethylated samples were dissolved in 10 µL MeOH. The solution was premixed with 20 mg/mL matrix with a ratio of 1:1 (v/v), before 1 µL of the mixture was spotted on the plate.

Micelle NMR. The LPS samples were incorporated into micelles as previously described (Acquotti & Sonnino 2000). Briefly, the LPS sample was subjected to the H-D exchange, pre-mixed with deuterated DPC in an estimated molar ratio of 1: 40. The mixture was then dissolved in a 50 mM deuterated potassium phosphate buffer (pD=6), and was transferred to a 5 mm NMR tube for spectroscopy. ^{1}H NMR and 2D TOCSY, HSQC and NOESY NMR spectra were recorded at 30 °C using a Bruker Avance III 600MHz NMR spectrometer equipped with a TXI/TCI cryoprobe.

Chapter 3 Structural Analysis of the Lipopolysaccharides from *Pseudomonas aeruginosa* and *Helicobacter pylori*

This chapter describes structural analyses aimed at characterising bacterial LPS, based on MS and NMR strategies. The chapter contains two sections, showing the structural investigations of *P. aeruginosa* LPS lipid A and the polysaccharide of *H. pylori* LPS, respectively. A combination of all the analytical strategies contained in these two sections provides facile and systematic approaches to the structural elucidation of LPS with high sensitivity.

3.0 Rationale for Methods Employed in LPS Characterisation

Though it was long believed that glycan structures are extremely difficult to elucidate, tremendous progress in the field of glycomics have been made during the past three decades, partially due to the advancement in MS.

Many eukaryotes, especially human beings, can only make use of a limited repertoire of monosaccharides, and the biosynthesis of their glycans has been very well characterised. This knowledge has great facilitated eukaryotic glycomics, *i.e.* their glycans can not only be qualitatively analysed with high sensitivity but also semi-quantitatively compared.

However, the case is quite different for prokaryotes. These "simple" organisms are capable of producing both highly "complex" and "complicated" polysaccharides and glycoconjugates, in terms of the monosaccharide compositions and sequences. The LPS can be considered as an exemplar of bacterial glycoconjugates. These large glycolipids contain many monosaccharides that are not observed in eukaryotes, such as KDO and heptose, and these monosaccharides can be assembled into all kinds of sequences. Furthermore, there is a lack of firm structural correlation between the LPSs extracted from different organisms. As mentioned in **Section 1.2.1.3**, the O-antigens are highly diverse and sometimes can be different from one strain to another. This means a structural and biosynthetic elucidation on one strain will normally not make much contribution to others. In addition, the LPS can be modified by multiple functional groups such as PEtN and Ara4N, which further complicates its structural analysis.

As a consequence, the in-depth structural analysis of bacterial glycoconjugates such as LPS, often requires a larger amount (relative to eukaryote glycans) of material as well as tedious experiments for the separation and enrichment of the samples, and the chemical degradation that facilitates the following instrumental analysis. The analytical strategies included in this chapter provide a standardised and facile starting point for LPS structural elucidation. As these methods can be very elusive to people who are not very familiar with this topic, it is worth summarising the strategies (**FIG. 3-1**) before showing the results.

First of all, the monosaccharide composition of the LPS needs to be elucidated (or partially determined) before the attempts of sequencing (**FIG. 3-1 Stage 1**). The question "what is the monosaccharide composition of an LPS sample" can be answered from three aspects, *i.e.* the identities, stereochemistries (α and β configuration) and linkages of the monosaccharides comprising the LPS.

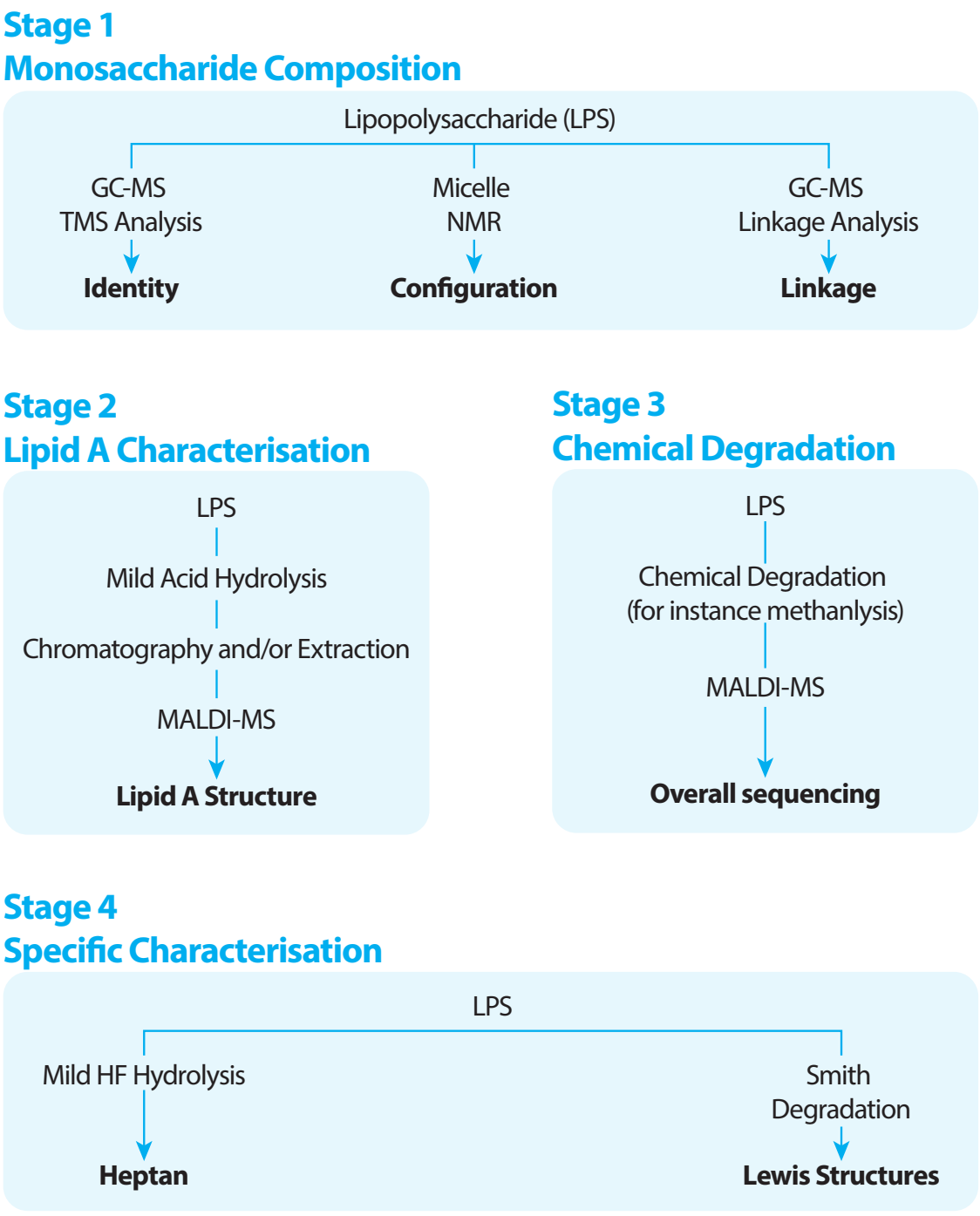


Figure 3-1 A diagram for the analytical strategies used in this chapter for the LPS structural elucidation. Specifically, the whole analytical processes are divided into four stages. Mild acid hydrolysis (Stage 2) is used for analysing *P. aeruginosa* lipid A, while all methods are used for elucidating *H. pylori* LPS glycan (mild acid hydrolysis did not work very well for *H. pylori* LPS, therefore the data are not shown in this thesis).

The identities and linkages can be determined by GC-EI-MS of TMS derivatives (TMS analysis, **Section 3.2.1**) and by linkage analysis (**Section 3.2.3**), respectively. Both of the GC-MS methods were well established decades ago. Very briefly, LPS samples are routinely hydrolysed and derivatised in order to produce highly volatile glycosides that are compatible with GC.

The stereochemistries can be characterised by NMR spectroscopy (**Section 3.2.2**). A novel micelle NMR technique was adopted in this thesis. As the LPS is a glycolipid whose lipid part is highly hydrophobic and the glycan part is very hydrophilic, it is difficult to be equally distributed in either aqueous or organic phase, which hinders NNR experiments. As this solubility issue can be partially solved by dissolving the LPS samples with detergents, the intact LPS molecules were incorporated into micelles for recording NMR spectra with a decent quality (see **Section 1.4.2.1**).

Each of the experiments above, especially the NMR analysis, requires a relatively large amount of material. In addition, it is preferable to begin with the intact samples for linkage and NMR analysis, so it is recommended to start the structural analysis of the LPS with GC-ESI-MS and NMR. In addition, there are two more benefits to run the NMR experiments before further MS-based sequencing. First, certain modifications on the LPS can be readily detected by NMR (see **Section 3.2.3**). Secondly, the sizes of the LPS can be characterised by NMR. If the NMR spectra indicate the LPS carries a short glycan chain, the MS spectra of the intact LPS should be recorded in order to determine its molecular weight before the chemical degradations.

Once the monosaccharide composition is confidently established, subsequent experiments should focus on exploring their sequences. Because the large size and hydrophilicity of the LPS greatly hinder the ESI and MALDI processes, the LPS is normally chemically degraded to produce its small fragments that are normally derivatised (such as permethylation and per-acetylation) for further MS investigations. In this chapter, MALDI-TOF and MALDI-TOF/TOF were used for characterising permethylated glycans degraded from the LPS (see **Section 3.1.2** and **3.2.5-7**). The chemically degraded LPS can also be directly analysed by LC-NMR.

Among all the chemistries that can be used for degrading LPS, mild acid hydrolysis is usually the first reaction chosen, especially for lipid A analysis (**FIG. 3-1 Stage 2**). The mild acid hydrolysis can be carried out by incubating the LPS with various acids such as diluted acetic acid (Manzo *et al*. 2001), HCl (Isshiki *et al*. 2003) and buffered sodium acetate (Kooistra *et al*. 2003) at elevated temperatures. For the non-dissolvable LPS samples, detergent-assisted mild acid hydrolysis was investigated (Caroff *et al*. 1988). The detergent such as SDS was suggested to significantly facilitate the hydrolysis by increasing the solubility of LPS (Caroff *et al*. 1988). These reactions preferably cleave the glycosidic bond between the KDO and the lipid A, producing the lipid A together with the remaining part of the LPS, which can be relatively simply separated due to their differences in hydrophilicity for further analysis. Note that these reactions can non-specifically cleave the weaker glycosidic bonds, generating small glycans. Mild acid hydrolysis was used to cleave the lipid A from the *P. aeruginosa* LPS (see **Section 3.1.2**) and this approach was also investigated for the characterization of the *H. pylori* LPS polysaccharides. However, the latter experiments did not yield useful information, so the data are not shown in the main text (see **FIG. A1**).

Some LPS samples are very large and/or stable, and cannot be characterised even after mild acid hydrolysis and the *H. pylori* LPS falls very well into this category. Such LPS needs to be degraded in a harsher chemical environment, for example that provided by the methanolysis strategy (**FIG. 3-1 Stage 3**). Methanolysis is capable of cleaving relatively stable glycosidic bonds and hydrolysing the LPS into small pieces, which played a key role in sequencing the *H. pylori* LPS (see **Section 3.2.5**). Interestingly, the LPS retained its phosphorylation after the methanolysis, which contributed to the characterisation of the functional groups, such as phosphate and PEtN, modifying the LPS,

Though the methanolysis can potentially provide valuable information on all three parts of the LPS, it lacks the selectivity required to define all structural motifs. For instance, it hydrolysed the long poly-lacNAcs expressed by *H. pylori* into shorter units, demolishing information on their overall lengths. As a consequence, more selective chemical reactions (**FIG. 3-1 Stage 4**) including mild HF hydrolysis and the Smith degradation (mild periodate oxidation) were adopted for complementing the methanolysis. The former reaction was used for characterising the heptan (**Section 3.2.6**), and the latter reaction was used for characterising the Lewis antigens (**Section 3.2.7**). *H. pylori* expresses a conserved yet special tri-oligosaccharide (called the "Trio", see **Section 3.2.1**) containing a Fuc1-3GlcNAc glycosidic bond that was proved to be labile to the mild HF hydrolysis (Yan *et al*. 2012; Haslam *et al*. 1997). The mild HF hydrolysis cleaved the Trio bridging the O-antigen and the core-oligosaccharide lipid A of the *H. pylori* LPS, forming a large glycan containing a full-length heptan that can potentially be detected by MALDI-TOF. The mild periodate oxidation was a very effective way to characterise the Lewis structures, as under optimised conditions, the Lewis structures were retained while most other LPS structures were completely oxidised and destroyed.

3.1 Structural Analysis of *Pseudomonas aeruginosa* Lipid A

3.1.1 Background

P. aeruginosa is a Gram-negative bacterium that normally inhabits the soil and aqueous surfaces. It is an opportunistic human pathogen that causes a wide range of diseases, including pneumonia, the urinary tract, surgical sites and bloodstream infections (Hauser 2009). It is a highly adaptable organism that can survive with the supplementation of a low nutrient in multiple physical conditions. For instance, it survives on the surfaces of medical devices, which renders it opportunities to infect hospitalised patients (Gellatly & Hancock 2013). In fact, the severe *P. aeruginosa* infections often do happen in hospitals, and patients with immune-deficiencies, such as neutropenia and burn victims, comprise the most vulnerable population.

The treatment of *P. aeruginosa* is highly challenging, partially due to its intrinsic, acquired and adaptive abilities to resist multiple antibiotics, such as β-lactam (Lister *et al.* 2009). The intrinsic resistance is a phenotype that is directly encoded in the genome. The intrinsic resistant methods of *P. aeruginosa* include the low outer-membrane permeability that can block antibiotic uptake, efflux pumps that can eject numerous antibiotics, and the expression of β-lactamases that can hydrolyse β-lactam antibiotics. *P. aeruginosa* can also horizontally acquire antibiotic-resistance related genes to obtain extra resistant ability. For instance, *P. aeruginosa* can acquire a second β-lactamase gene that further improves its β-lactam resistance. In addition to the ability of intrinsic resistance, *P. aeruginosa* can adapt itself to environmental stresses by different ways, such as forming biofilms, which also contributes to its antibiotic resistance.

The modifications of *P. aeruginosa* lipid A have also been linked to antibiotic resistance, on which extensive research efforts have focused. Previous studies showed that *P. aeruginosa* lipid A could be modified with palmitate and Ara4N, as shown in **FIG. 3-2** (Vallée-Réhel & Zähringer 2004; Gunn, Lim, *et al.* 1998). These modifications were suggested to be associated with resistance to cationic antimicrobial peptides and antibiotics, such as polymyxin B.

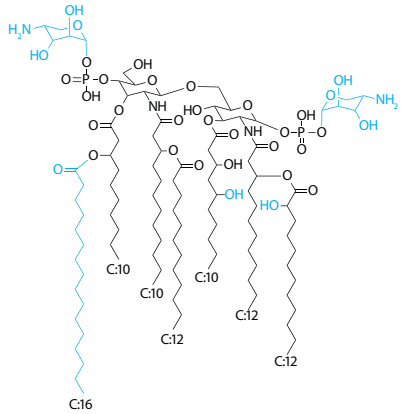


Figure 3-2 Structural Model for *P. aeruginosa* lipid A. The blue coloured parts are potential modifications of its lipid A.

The addition of palmitate to the lipid A of *P. aeruginosa* is reminiscent of the lipid A modification in *E. coli* and *S. typhimurium*, which are very well understood (see **Section 1.2.1.1**). In the latter two bacteria, an enzyme encoded by *papP* mediates the acylation of their lipid As on position 2 (Bishop *et al.* 2000). In *P. aeruginosa*, the palmitate is added to position 3′instead of position 2 (compare **FIG. 3-2** with **FIG. 1-7**). This was believed to be due to a lack of an ortholog of the *pagP* gene in *P. aeruginosa* and the existence of another enzyme mediating this modification. However, recent research (Thaipisuttikul *et al.* 2014) identified a gene *PA1343* as the *pagP* ortholog encoding the lipid transferase adding the palmitate to the different site (position 3′) of the lipid A.

The attachment of Ara4N to the two phosphate groups is another lipid A modification in *P. aeruginosa* (**FIG. 3-2**). This process is mediated by an amino-arabinose transferase (ArnT) and the modification was first discovered and characterised in *S. typhimurium* (Vaara *et al.* 1981; Trent, Ribeiro, *et al.* 2001). Ara4N has been reported to be present on the *P. aeruginosa* lipid A by multiple researchers (Gunn, Lim, *et al.* 1998; Vallée-Réhel & Zähringer 2004; Bhat *et al.* 1990).

The expressions of the PagP and ArnT enzymes are regulated by the TCS (see **FIG. 3-3**; Groisman 2001). More specifically, the expression of pagP is dependent on the PhoP/Q TCS that senses the extracellular concentration of Mg^{2+} (Vescovi *et al.* 1996; Gunn, Belden, *et al.* 1998). The PhoQ protein is a transmembrane sensor whose periplasmic domain is proposed to sense divalent cations (Vescovi *et al.* 1997), and its cytoplasmic domain

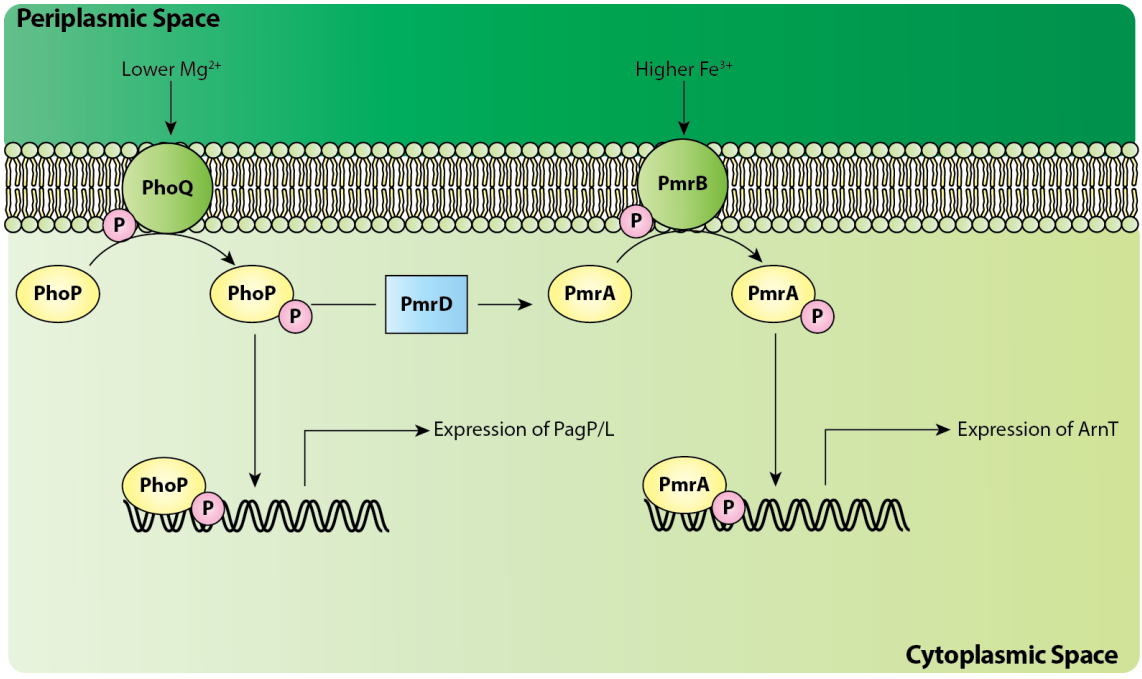


Figure 3-3 The model of PhoP/Q, PmrA/B TCSs in *P. aeruginosa*. PhoP is phosphorylated under low Mg^{2+} concentration that can be sensed by PhoQ. Phospho-PhoP can activate multiple genes including PagP/L and PmrD. PmrD can activate the PmrA/B TCS. PmrA can also be activated by PmrB at high Fe^{3+} concentration, which will regulate expression of ArnT.

can mediate auto-phosphorylation. The PhoP protein is a cytoplasmic regulator that can activate multiple genes including PagP. When the bacteria live in a Mg^{2+} rich environment (for instance, mM level), their PhoQ proteins will de-phosphorylate the PhoP protein, thus abolishing PhoP activating genes (Castelli *et al.* 2000). In this way, transcriptions of the PhoP-activated genes are promoted in a medium with a low Mg^{2+} concentration (for instance, μM level) and repressed under the Mg^{2+} rich condition.

Another important gene regulated by the PhoP/Q TCS is pmrD. It is responsible for linking the PhoP/Q TCS to the PmrAB TCS, *i.e.* the expression of pmrD can post-transcriptionally activate the PmrA/B TCS via an unknown mechanism (Kox *et al.* 2000). The PmrA/B TCS is very similar to the PhoP/Q TCS, where PmrB parallels to PhoQ as the transmembrane sensor and PmrA serves a role as an intracellular regulator like PhoP (Groisman 2001). In contrast to PhoQ sensing divalent cations, PmrB senses Fe^{3+} (Wösten *et al.* 2000). The Fe^{3+} ions can bind to the cytoplasmic domain of PmrB, facilitating its auto-phosphorylation, and in turn phosphorylating the PmrA protein (Wösten *et al.* 2000). The phospho-PmrA regulator subsequently promotes the expressions of multiple genes including ArnT (**FIG. 3-3**; Wösten & Groisman 1999). Either low level Mg^{2+} or high level Fe^{3+} is sufficient to promote the PmrA-dependent genes and this ability to sense multiple extracellular signals might increase bacterial adaptation to variable environments (Wösten *et al.* 2000). Interestingly, the PhoPQ/PmrAB TCS mutations also lead to an activation of the ArnT in *in vitro*-required and clinically isolated *P. aeruginosa* that are resistant to polymyxin (Moskowitz *et al.* 2004; Barrow & Kwon 2009).

P. aeruginosa can also sense 2-alkyl-4-quinolone (AQ) molecules in addition to the environmental cationic concentrations, which is associated with its pathogenic factors such as biofilm formation (de Kievit 2009; Smith & Iglewski 2003; Keller & Surette 2006). Our collaborators from the University of Exeter discovered an elevated AQ activity in the *P. aeruginosa* PhoPQ/PmrAB TCS site-specific mutants (see below) that show an increased polymyxin resistance. As the antibiotic resistance phenotype is a sign of lipid A modification, it was proposed that the quorum sensing system could be affected by the TCSs, *i.e.* the TCS-regulated modification of lipid A might influence the secretion of membrane vesicles that can transport AQ molecules.

In order to confirm whether the TCS mutants obtain the ability of the polymyxin resistance through the lipid A modification, LPS samples were extracted from two WT *P. aeruginosa* serotype O5 (PAO1) and its three mutant strains by our collaborators. The WT PAO1 was cultured in 2 mM and 20 μM Mg^{2+} media respectively, and three mutant *P. aeruginosa* include two site mutations on the PmrB[M292T] and PhoQ[V260G] respectively, and a phoQ knock-out (see **Section 2.2.1**). This section describes the MS-based structural characterisation of the lipid A extracted from PAO1.

3.1.2 Mild Acid Hydrolysis

Mild acid hydrolysis is a classical way to selectively cleave the glycosidic bond between the KDO and the acylated GlcN (see the blue arrow in **FIG. 3-4** and **Section 3.0**). In this study, the LPS samples were hydrolysed in a sodium acetate buffer (pH=4.6, adjusted by acetic acid). The hydrolysed LPS samples were extracted by a chloroform/methanol (v/v 4:1) solution, and were further washed with ultra-pure water for three times before drying down under a stream of nitrogen. The dried samples were then analysed by a MALDI-TOF spectrometer operated in the negative ion mode.

The MALDI-TOF spectra of the lipid A extracted from the WT and mutant PAO1 are shown in **FIG. 3-5 A** and **B**. The former figure shows the MS spectra with a mass range from *m/z* 1350 to 1650, and the latter figure shows the spectra with a mass range from *m/z* 1650 to 1950. In both figures, **a** and **b** show the spectra of the LPS from WT PAO1 cultured in 2 mM and 20 μM Mg^{2+} respectively; **c**, **d** and **e** show the spectra of the LPS from PmrB[M292T], PhoQ[V260G] and ΔphoQ PAO1 respectively. All of the MS peaks corresponding to lipid As or its fragments are coloured red. Because of the limited space, the lipid A structures were not annotated on the MS spectra. Instead a complete peak list abstracted from the spectra and the proposed lipid A structures is shown in **Table 3-1**. Only the WT spectra were partially annotated with mass shifts due to different modifications, as all the MS peaks have been included in those two spectra except for an MS peak at *m/z* 1769.1 (-16 Da shift from 1785.0).

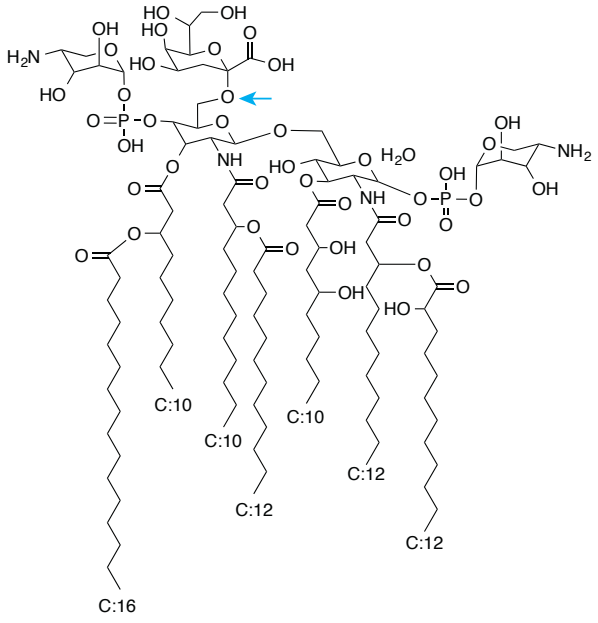


Figure. 3-4 Mild Acid Hydrolysis of PAO1 LPS. The blue arrow highlights the glycosidic bond between KDO and actylated GlcN, which is more labile than others and therefore can be selectively cleaved under an optimised mild acid environment.

The lipid A samples from PmrB[M292T] and PhoQ knockout are similar to each other, but more complex because they have all of the structures found in WT samples in varying abundances. The obtained MS spectrum for PhoQ[V260G] is of lower quality than the others: only minor peaks, for instance, the peaks at *m/z* 1365.7 and 1496.7, were observed. Nevertheless, these peaks indicate this sample is similar to the other two mutant samples, especially the lipid A is modified by the Ara4N.

The peaks at *m/z* 1365.7 were present in all the spectra (see **Table 3-1**), and correspond to a di-GlcN acylated with 4 lauric acids and a capric acid (C10). This structure was proposed based on previous research, as no MS/MS experiment was successfully carried out with the MALDI-TOF/TOF spectrometer (Vallée-Réhel & Zähringer 2004). The higher MS peaks were attributed to different modifications on the acylated di-GlcN structure. The mass shift of 16 Da, such as *m/z* 1365.8 to 1381.8, corresponded to an addition of a hydroxyl group; and the mass shift of 102 Da, such as *m/z* 1365.8 to 1467.7, corresponded to an addition of a sodium phosphate group. Phosphorylation and C10 acylation increased the *m/z* by 80 Da and 170 Da respectively. In addition, an addition of Ara4N characteristically shifted the *m/z* by 131 Da. All these modifications were successively added onto the acylated di-GlcN structure, giving rise to all the observed MS peaks.

The MS data showed that all of the samples were mixtures containing lipid As with different numbers of phosphates and fatty acids. In addition, the Ara4N was detected in all the samples except for the lipid A extracted from WT *P. aeruginosa* cultured in high level Mg^{2+} (**FIG. 3-5 A a**). The lipid A from the WT *P. aeruginosa* cultured in the concentrated Mg^{2+} medium contained the simplest WT structures, but the major components were modified by the Ara4N. In addition, the abundance of the hydroxylated species in this sample is relatively higher (compare the MS peaks at *m/z* 1496.8 and 1512.8).

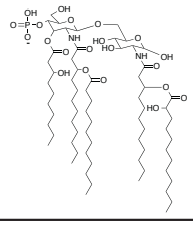
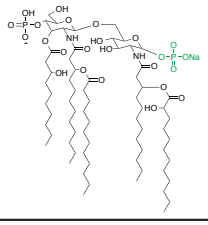
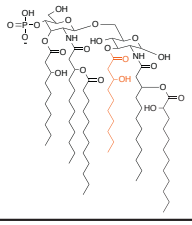
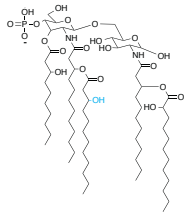
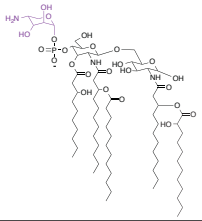
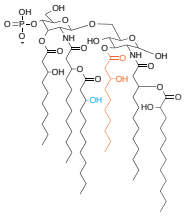
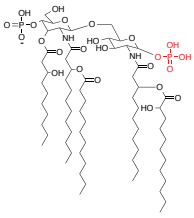
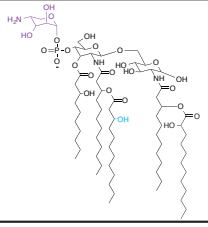
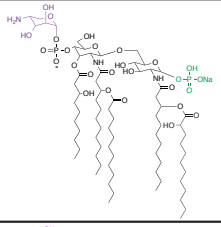
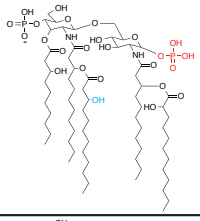
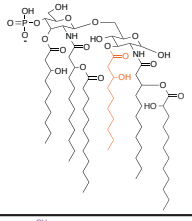
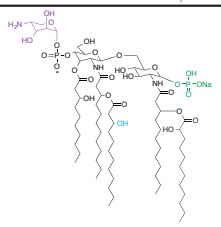
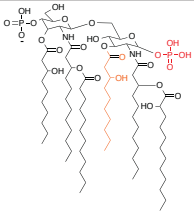
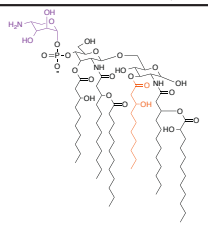
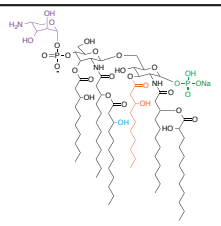
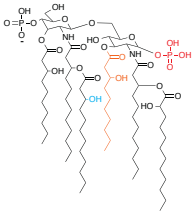
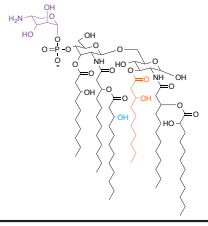
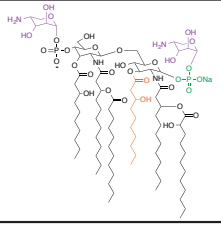
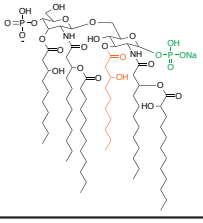
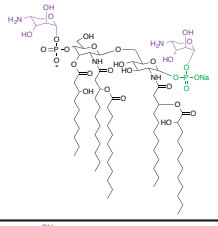
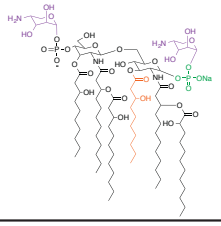
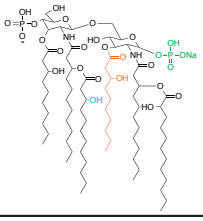
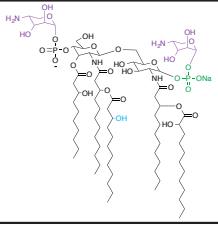
m/z	Proposed Structures	*m/z*	Proposed Structures	*m/z*	Proposed Structures
1365.8		1467.7		1535.9	
1381.8		1483.7		1551.9	
1445.7		1496.8		1598.7	
1461.7		1512.8		1614.7	
1615.9		1666.9		1784.8	
1631.9		1682.9		1899.8	
1637.8		1729.7		1915.8	
1653.8		1745.8			

Table 3-1 List of the Lipid A structures detected in the MALDI-TOF spectra of the PAO1 LPS samples. The structures were proposed based on previous research. The modifications of the lipid A are shown in different colours, *e.g.* blue for the hydroxyl groups, red for the phosphates, green for the sodium phosphates, orange for the capric acids, and purple for the Ara4N. Note that the same colours are used for antotating the MALDI-TOF spectra.

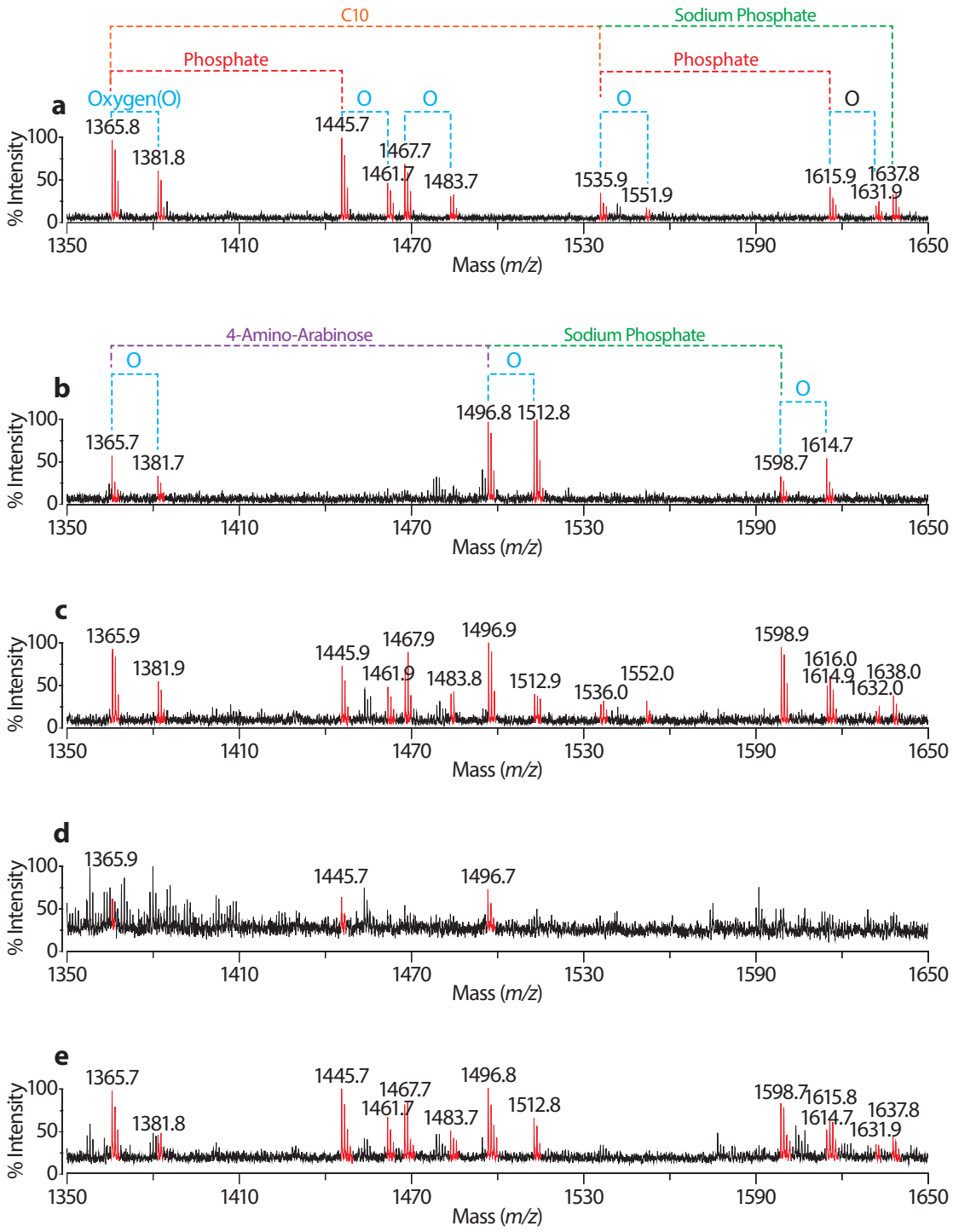


Figure 3-5 A MALDI-TOF spectra (mass range from *m/z* 1350 to1650) of PAO1 WT and mutants LPS. (a) and (b) show spectra of LPS from WT PAO1 cultured in 2 mM and 20 μM Mg^{2+} respectively. (c), (d) and (e) show spectra of LPS from PmrB[M292T], PhoQ[V260G] and ΔphoQ PAO1 respectively. The peaks at *m/z* 1365.7 are corresponding to a di-GlcN acylated with four lauric acids and a capric acid. The higher MS peaks correspond to the same structure carrying different modifications (blue for the hydroxyl groups, red for the phosphates, green for the sodium phosphates, orange for the capric acids, and purple for the Ara4N).

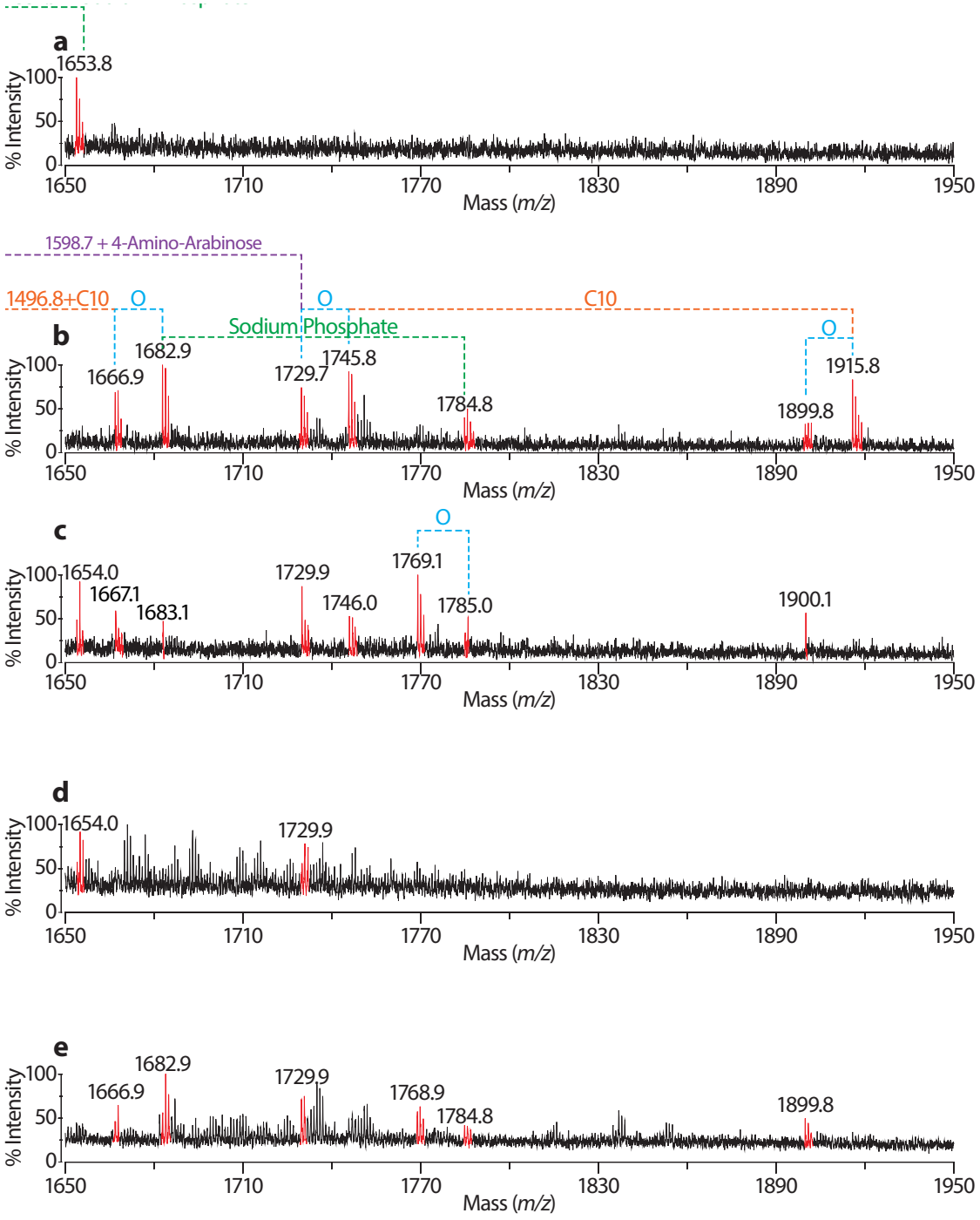


Figure 3-5 B MALDI-TOF spectra (mass range from *m/z* 1650 to 1950) of PAO1 WT and mutants LPS. More detailed information can be found in the legend of **FIG.3-5 A**.

3.1.3 Discussion and Conclusions

The MS analysis in this section re-confirmed that the WT PAO1 lipid A is modified by Ara4N when cultured with a low Mg^{2+} concentration, or one of the PhoP/Q and PmrA/B TCSs is mutated. This modification covers the negatively charged phosphate groups, thus blocking the ionic interaction between positively charged polymyxin and the phosphate groups on the lipid A. This contributes to the increased antibiotic resistance ability of the mutant *P. aeruginosa* strains.

As mentioned, our collaborator initially observed an increased AQ activity in the PAO1 PmrB[M292T] and PhoQ[V260G] strains. The MS data helped confirm that there exists the Ara4N modification in both of them, and the PhoQ knock-out strain. However, subsequent experiments carried out by our collaborator showed an unparalleled activity of AQ molecules. The PhoQ knock-out strain essentially showed a wildtype AQ activity, indicating that their hypothesis that the TCS-regulated modifications of lipid A are responsible for the increased AQ activity is not fully correct.

In addition to the Ara4N, the hydroxyl group and the capric acids were observed to modify the lipid A. These modifications are not mediated by the TCSs, and as expected, they were observed in all the PAO1 LPS samples. The hydroxylation of lipid A has been characterised in *Salmonella*, which expresses an LpxO enzyme capable of hydroxylating the acyl chain on the 3′ position (Raetz 2001). Two putative LpxO homologs (LpxO1 and LpxO2) were found in *P. aeruginosa*, hydroxylating the acyl chain on the 2′ position (Vallée-Réhel & Zähringer 2004). The incorporation of the capric acids into the KDO_2-lipid A was suggested to be mediated by LpxA. As its carboxyl ruler mechanism was introduced in **Section 1.2.1.1**, it is not further discussed here.

When the WT PAO1 was cultured in the Mg^{2+} rich medium, its LPS is hardly modified by any Ara4N. However, when it is cultured in a low Mg^{2+} environment, the Ara4N modified lipid A dominates the MS spectrum. The mutant PAO1 LPS samples are fundamentally the mixtures of the former two samples, indicating that the PAO1 mutant strains are not able to efficiently modify their lipid As. The biological implication underlying these observations still remains elusive.

3.2 Structural Analysis of *H. pylori* LPS Glycans

3.2.1 Background

Helicobacter pylori (*H. pylori*) is a Gram-negative bacterium colonizing human stomachs (Marshall *et al*. 1985). It affects at least half of all people around the world; and is a leading cause of human gastritis, stomach ulcer and cancer (Marshall *et al*. 1985; Suerbaum & Michetti 2002). Notably, its infection can last for life without proper antibiotic treatments (Peek & Blaser 2002; Salama *et al*. 2013).

H. pylori infection is typically acquired during childhood (Suerbaum & Michetti 2002). The risk of the infection highly depends on the social-economical development (Malaty & Graham 1994): the infection is more prevalent in developing countries. Previous research estimated that the infection rate of *H. pylori* can be as high as 70-80% in developing countries, while this number is around 20-50% in developed countries (Suerbaum & Michetti 2002).

The infection of *H. pylori* significantly increases the risk of diffuse-type gastric adenocarcinoma (Suerbaum & Michetti 2002; Peek & Blaser 2002). The persistent colonization of pathogenic *H. pylori* causes chronic superficial gastritis that can lead to more serious inflammation such as atrophic gastritis. Long-term inflammation can ultimately result in dysplasia and adenocarcinoma.

The pathogenicity of *H. pylori* relies on multiple factors (Salama *et al*. 2013; Suerbaum & Michetti 2002) including the survival in the acidic niche (Eaton *et al*. 1991; Weeks *et al*. 2000; Rolig *et al*. 2011; Bonis *et al*. 2010; Sycuro *et al*. 2010; Sachs *et al*. 2003), the establishment of persistent infection (Mahdavi 2002; Salama *et al*. 2013), and the secreted cellular toxins (such as CagA and VacA; Bridge & Merrell 2013). Notably, the suppression of the human immune system greatly facilitates the chronic infection. Through co-evolution with human beings for over 60,000 years (Falush *et al*. 2003), *H. pylori* has evolved multiple methods to suppress the human immune responses, *e.g.* it is able to evade the recognition of multiple pattern recognition receptors (PRRs) and modulate T cell responses (Salama *et al*. 2013).

H. pylori LPS is closely related to its pathogenicity. Its LPS mimics human Lewis-type antigens, which is proposed to improve cell adhesion at the initial stage of infection. More importantly, the LPS plays a vital role in suppressing TLRs and Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin (DC-SIGN) induced inflammation, thus helping establish a persistent *H. pylori* infection. Therefore, the elucidation of LPS structure is a pre-requisite to completely understanding its pathogenicity.

Because of its biological importance, *H. pylori* LPS has been in-depth researched. The detailed structures of LPS from several *H. pylori* strains including 26695, strain Sydney 1 (SS1) and serogroup O:3 have been reported based on NMR and MS analysis. Multiple glycosyltransferases involved in its LPS biosynthesis were characterized, as shown in **FIG. 3-6 c**. Moreover, recent research has revisited, and more importantly has amended the previously published structures of the LPS from *H. pylori* strain 26695 (Altman, Chandan, Li & E. Vinogradov 2011b) and SS1 (see **FIG. 3-6 b** and **a**; Altman, Chandan, Li & E. Vinogradov 2011a). In particular, NMR has provided important information on the structure and biosynthesis of the core-oligosaccharide.

The NMR spectra showed that the inner core of the LPS from both strains is a branched hexa-saccharide whose main chain is a tri-heptose (Hep) -KDO substituted on the distal Hep with a glucose-galactose (Glc-Gal) disaccharide. This inner core is extended with a novel tri-saccharide (the novel Trio) with a sequence of Hep-fucose–N-acetyl-glucosamine (Hep-Fuc-GlcNAc). The shared inner-core and novel Trio structures in these two strains may suggest they constitute a wider conserved property through different *H. pylori* strains. The biosynthesis of the *H. pylori* inner-core was partially characterised. Two heptosyltransferases (HepT; HP0279 and HP1191) were demonstrated to catalyse the addition of Heps to the KDO-lipid A, generating the inner-core-KDO-lipid A, onto which the outer-core can be assembled.

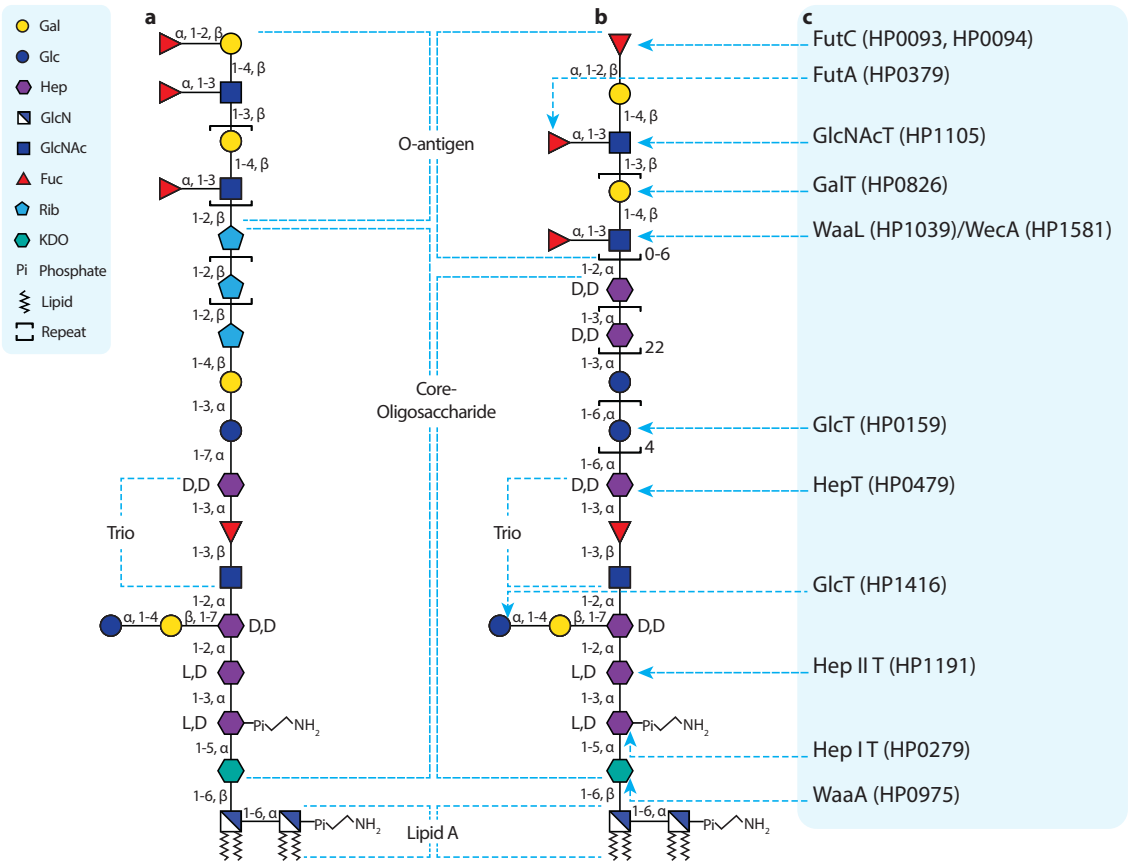


Figure 3-6 Revisited structures of the LPS from *H. pylori* strain Sydney 1 (a) and 26695 (b). The proposed nominations of different parts of LPS are shown. Previously characterised *H. pylori* GTs involved in the biosynthesis of LPS are shown in (c).

The difference between the two LPSs arises from the outer-core: in strain 26695, the outer-core is a glucan linking to a heptan, whereas in strain SS1, it is a Gal-Glc disaccharide linking to a riban. Several GTs involved in the outer-core assembly have been identified, including a HepT (HP0479; Altman, Chandan, Li & E. Vinogradov 2011b, Hiratsuka *et al.* 2005) and a GlcT, (HP0159; Altman *et al.* 2008, Langdon 2005, Logan *et al.* 2005, Moran *et al.* 2004).

Recent research efforts aiming to create glycoconjugate vaccines based on the LPS extracted from certain HepT mutants concluded that when the core glucan cannot be capped with the heptan, it will continue elongating to around eleven glucose units. Proteins modified by the glucan are immunogenic (Altman *et al.* 2012).

H. pylori expresses mammalian-type poly-N-acetyl-lactosamine (lacNAc) and Lewis antigens (including Lewis X, Y, A and B) as its O-antigens. In addition, a phenomenon called phase variation was observed (Appelmelk *et al.* 2000): by adopting a strand-slipping mechanism, different epitopes (such as Lewis epitopes) can be expressed. It has been proposed that *H. pylori* synthesises its O-antigen via a pathway similar to the ABC transporter dependent pathway (compare **FIG. 3-7** and **FIG. 1-12** in **Section 1.2.1.3**) based on the discovery that a functional flippase in *H. pylori* is a homolog of the one in *C. jejuni* (Hug *et al*, 2010). The model of the ABC-transporter dependent pathway in *H. pylori* is as follows.

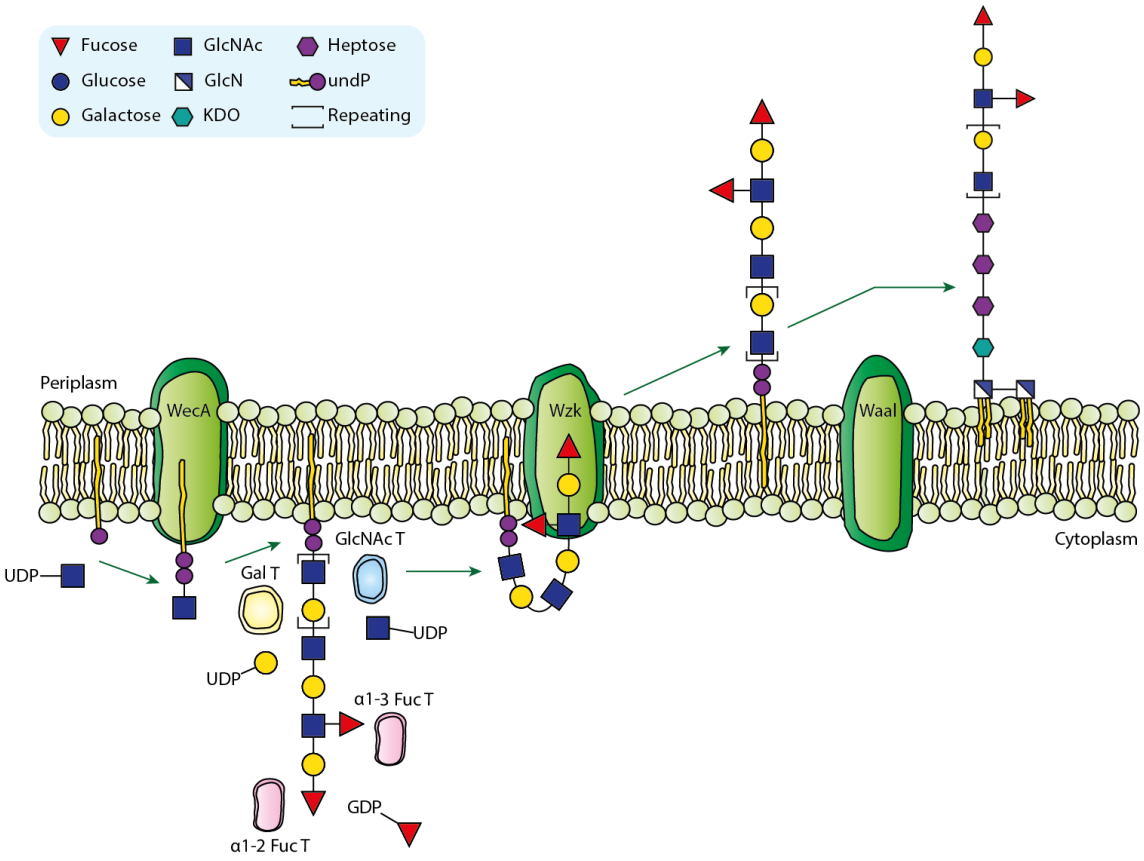


Figure 3-7 Diagram for the proposed biosynthesis of the O-antigen in *H. pylori*. Adapted from (Hug *et al*, 2010).

The biosynthesis of *H. pylori* O-antigens initiates with transferring a GlcNAc on a lipid carrier called undecaprenol pyrophosphate (UndP) by WecA (HP1581) at the inner layer of the cytoplasmic membrane. Afterwards, repeating lacNAc units are assembled onto the GlcNAc-UndP carrier, until fucosyltransferases (FucT) cap the antigens by producing the Lewis epitopes. The lacNAc residues are synthesized by a GlcNAcyltransferase (GlcNAcT, HP1105; Logan *et al.* 2005) and a galactosyltransferase (GalT, HP0826; Altman, Chandan, Li & E. Vinogradov 2011a).

One Fuc can be transferred to the GlcNAc residue in the terminal lacNAc by a FucT (FutA, HP0379), forming Lewis X (Ge *et al.* 1997). A second Fuc can be added to the terminal Gal by another FucT (FutC, HP0093/HP0094), producing Lewis Y (G. Wang, Rasko, *et al.* 1999; G. Wang, Boulton, *et al.* 1999). Enzymes for synthesizing Lewis A and B have also been reported (Rasko *et al.* 2000). Interestingly, the order of activities of FucTs in *H. pylori* is opposite to that in humans in synthesizing Lewis B. The fully synthesized O-antigens are transferred to the periplasmic face of the inner-membrane by the "flippase" Wzk (the previously mentioned *C. jejuni* flippase homolog).

It is obvious from **FIG. 3-6** that not all glycosyltransferases involved in the *H. pylori* LPS biosynthesis have been clarified. Based on genomic analysis, our collaborator characterized several putative *H. pylori* GTs, such as HP1284 and HP1283, involved in the LPS biosynthesis. Interestingly, the mutation in their homologous genes in *H. pylori* strain X47 that is able to colonize mouse seems to result in a complete loss of the colonization ability (**FIG. 3-8**). Therefore, in order to characterise the functions of the putative GTSs and more comprehensively understand the pathogenicity of *H. pylori*, current research aims to elucidate the structures of the LPSs from the WT and mutant *H. pylori* model systems.

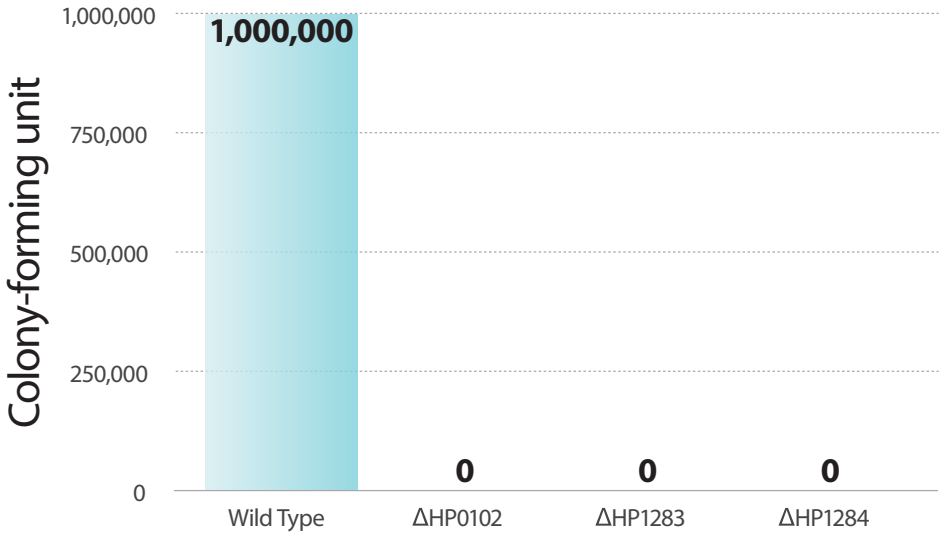


Figure 3-8 The mouse colonization experiment. *H. pylori* X47 wild-type (WT) and isogenic ΔHP1283, ΔHP1284 and ΔHP0102 mutants were used for challenging eight-week old C57BL/6J mice. After two weeks, mice were sacrificed so that colony-forming unit (CFU) can be measured. The detection limit was less than 200 CFU per stomach. [collaborators' data]

The LPS samples were extracted from WT and mutant (ΔHP0156, G27ΔHPG27_1230, ΔHP1284, ΔHP1283, ΔHP1578, ΔHP0102 and ΔWaaL) *H. pylori* strain G27. The detailed analytical strategies were discussed in **Section 3.0**. Briefly, the intact LPS molecules were too big to be directly analysed by MS, and they therefore were subjected to chemical degradation (methanolysis, HF hydrolysis and Smith degradation) before Matrix-assisted Laser desorption/ionisation (MALDI)-time of flight (TOF) characterisation. Methanolysis gives valuable insights into the structure of core oligosaccharide-lipid A, while mild HF hydrolysis and the Smith degradation characterise the heptan and poly-lacNAc, respectively. Important MS peaks were chosen for MALDI-TOF/TOF analysis. The samples were also characterised by NMR and GC-EI-MS in order to determine the identities, configurations and linkages of contained monosaccharides.

The WT LPS is very similar, if not exactly the same, as the previously reported *H. pylori* strain 26695 LPS, and our data strongly suggest that the *H. pylori* G27 LPS O-antigen is a lot longer than previously proposed. Therefore, the definitions of O-antigen and core-oligosaccharide are different. For the convenience of showing and discussing results, the nominations of different parts of the G27 LPS is shown in **FIG. 3-9**.

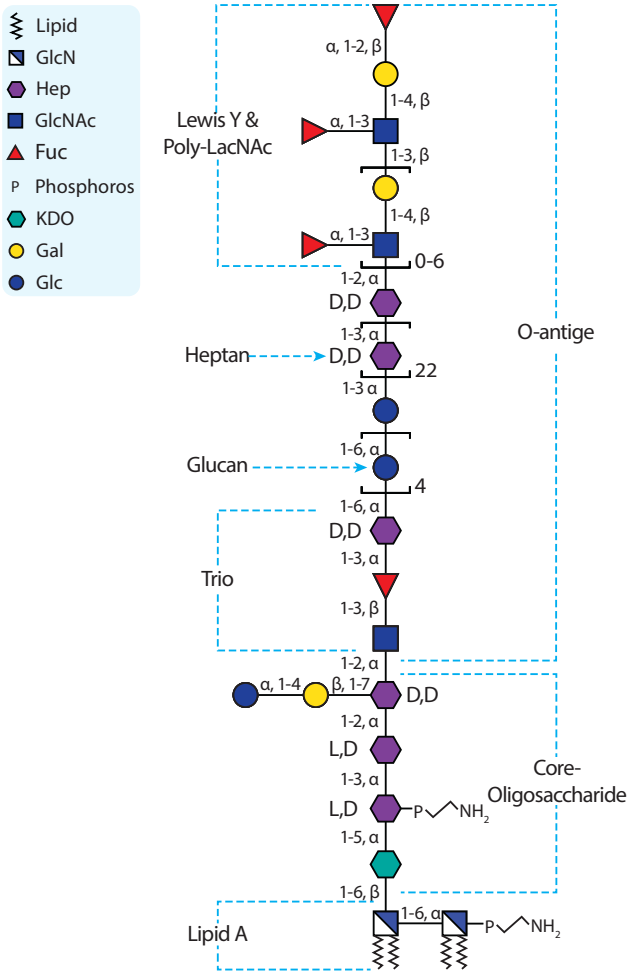
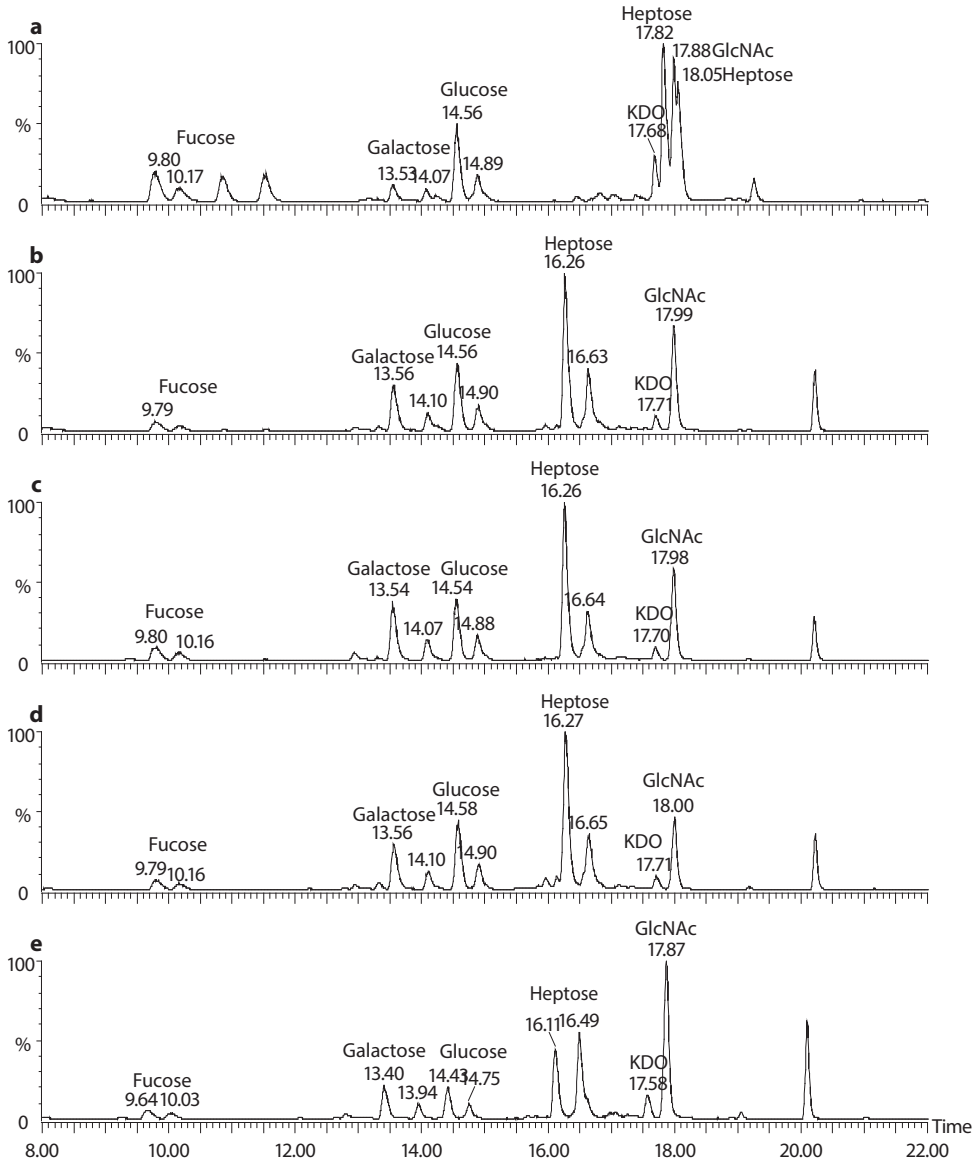


Figure 3-9 Full Structure of *H. pylori* G27 WT LPS. The nominations of different parts in the LPS are shown.

3.2.2 GC-EI-MS Trimethylsilyl Analysis

LPS samples were completely methanolysed into monosaccharides that were further TMS derivatised and analysed by GC-EI-MS. The monosaccharide standard containing Gal, Glc, GlcNAc, KDO and Hep was prepared and analysed with LPS samples sequentially so that the identities of monosaccharide in LPS samples could be determined by comparing the retention times (RTs) and MS fingerprints.

The TMS analysis chromatograms are shown in separate figures based on the relative amount of Fuc. LPS samples from WT, ΔHP0156, G27ΔHPG27_1230, ΔHP1284 contained relatively more Fuc, which indicates they probably carry Lewis structures. A low amount of Fuc was detected in ΔHP1283, ΔHP1578 LPS, and only a trace amount of Fuc was found in ΔHP0102 and ΔWaaL LPS.



Fligure 3-10 A The TMS analysis of full standards (a) and WT (b), ΔHP0156 (c), G27ΔHPG27_1230 (d), ΔHP1284 LPS (e). The corresponding chromatograms annotated with retention times (RT) and monosaccharides are shown.

Total ion chromatograms from GC-EI-MS analysis of standards and WT, ΔHP0156, G27ΔHPG27_1230, ΔHP1284 LPS are shown in **FIG. 3-10 A**. The results showed that all of these LPS samples contain Fuc, Gal, Glc, Hep, GlcNAc and KDO. Notably, the Hep found in LPS samples was different from the commercially available L-heptose standard. As previously mentioned, they all contain a relatively high amount of fucose, indicating they carry Lewis structures.

Analogous results for ΔHP1283, ΔHP1578, ΔHP0102 and ΔWaaL LPS are shown in **FIG. 3-10 B**. The data indicated that the kind of monosaccharides contained by them are exactly the same as previously shown samples, *i.e.* Fuc, Gal, Glc, GlcNAc, KDO and Hep was observed

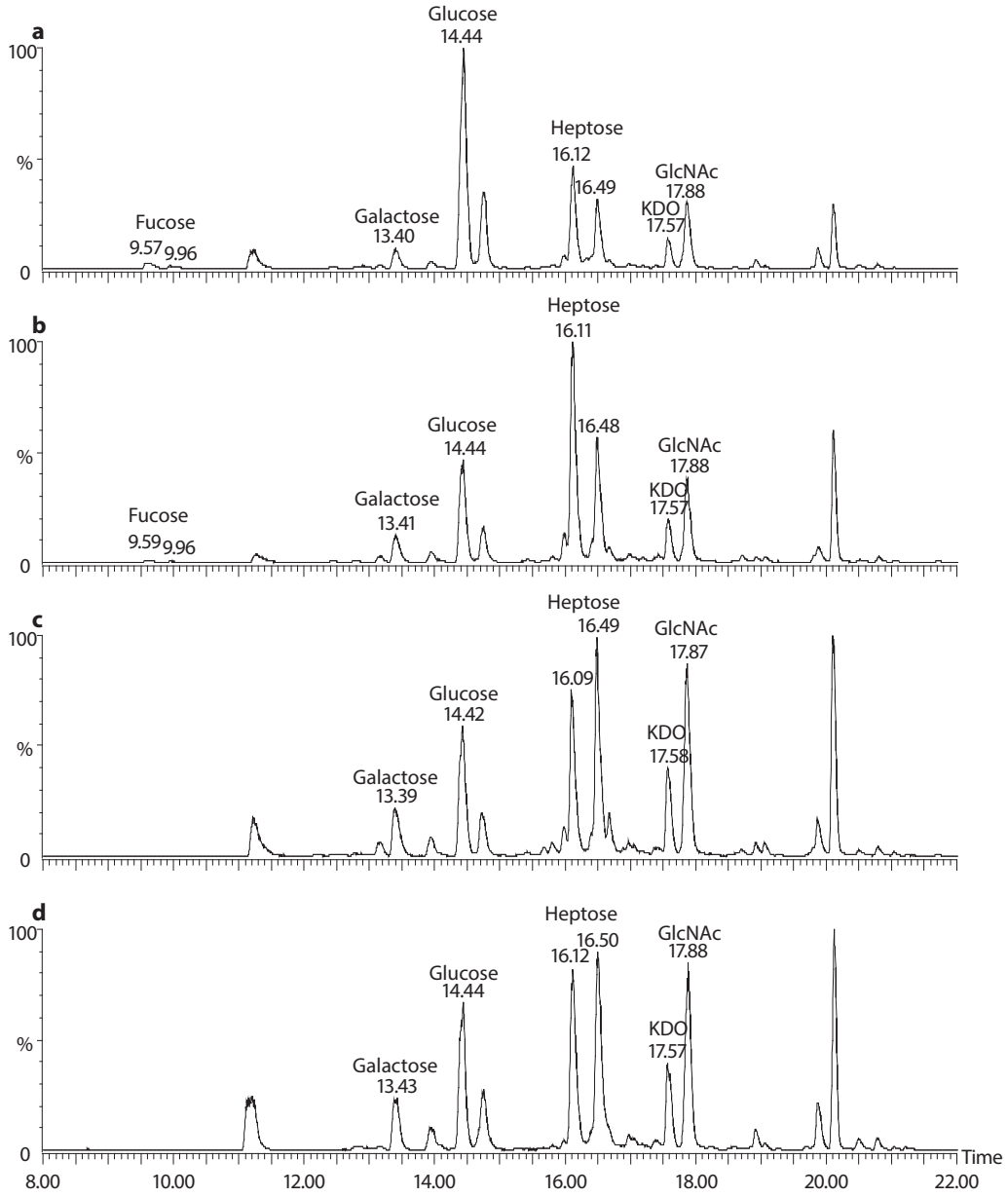


Figure. 3-10 B The TMS analysis of ΔHP1283 (a), ΔHP1578 (b), ΔHP0102 (c) and ΔWaaL LPS (d). The corresponding chromatograms annotated with retention times (RT) and monosaccharides are shown.

in the chromatograms. However, it was apparent that the ΔHP1283, ΔHP1578 LPS contain a lower amount of Fuc and the ΔHP0102 and ΔWaaL LPS only contain a trace amount of Fuc, *i.e.* the peaks for fucose were not visible unless greatly zooming these two chromatograms in for observation. In addition, it was interesting to find that a significant portion of the monosaccharides contained in ΔHP1283 LPS is glucose, indicating it has a longer glucan chain. This is consistent with the later MS analysis (see **Section 3.2.5.3**).

All the chromatograms were further investigated quantitatively, and the results are shown in **Table 3-2**. The peaks found in chromatograms were integrated and the area of each monosaccharide was divided by the area of all peaks. According to the table, it was further confirmed that LPS samples from WT, ΔHP0156, G27ΔHPG27_1230, ΔHP1284 contain the highest levels of Fuc; and that ΔHP1283 LPS contains extra Glc.

Table 3-2 Quantitative analysis of the GC-MS data. The name of the samples are shown in the first column, and for each sample, the ratios of monosaccharides are shown in the same row. The ratios were calculated by dividing the areas of each monosaccharide peak with all peaks in the samples. ΔHP1230 stands for G27ΔHPG27_1230 in order to save space in the table.

	Fucose	Galactose	Glucose	Heptose	KDO	GlcNAc
WT	0.06	0.15	0.19	0.38	0.03	0.19
ΔHP0156	0.07	0.16	0.18	0.39	0.02	0.17
ΔHP1230	0.05	0.15	0.21	0.43	0.02	0.14
ΔHP1284	0.05	0.11	0.11	0.35	0.05	0.33
ΔHP1283	0.02	0.05	0.51	0.26	0.04	0.12
ΔHP1578	0.01	0.22	0.19	0.42	0.05	0.12
ΔHP0102	trace	0.10	0.21	0.38	0.09	0.22
ΔWaaL	trace	0.09	0.23	0.41	0.08	0.19

3.2.3 Micelle NMR

All *H. pylori* LPS samples were analysed by micelle NMR. LPS samples were mixed with DPC, and all acid protons in the system were exchanged with deuterium. The micelle incorporated LPSs were then analysed by various NMR methods directly, *e.g.* by ^{1}H NMR, TOCSY, and NOESY and HSQC.

The resolution of NMR spectra is greatly affected by the size and solubility of the analytes, *i.e.* the smaller and more soluble the samples are, the better their NMR spectra are resolved. Although solubility of the samples was improved by incorporating the LPS into micelles, the micelle incorporated LPSs tend to be clustered together, which decreases the overall resolution of the NMR spectra. As a result, the resolution of the micelle NMR spectra was still inherently lower than the NMR spectra of isolated oligosaccharides. Despite the defect, most of the obtained NMR spectra were of acceptable quality, and the resolution of some spectra, such as those of the ΔHP1283 and ΔHP1578 LPS, was high enough to identify multiple anomeric proton signals. In addition, the spectra NMR of the ΔWaaL and ΔHP0102 were relative simpler though they were not excellently resolved, and extensive structural information was extracted from their NMR spectra.

This section first discusses the NMR signals corresponding to the PEtN modification of LPS. Though this modification is relatively more difficult to detect by MS-based methods, it can be readily identified because of the diagnostic NMR signals of PEtN. TOCSYs for all the *H. pylori* G27 LPS samples indicate the existence of PEtN, and the TOCSY of the WT G27 LPS assigned with the putative signals of PEtN is shown in **FIG. 3-11**.

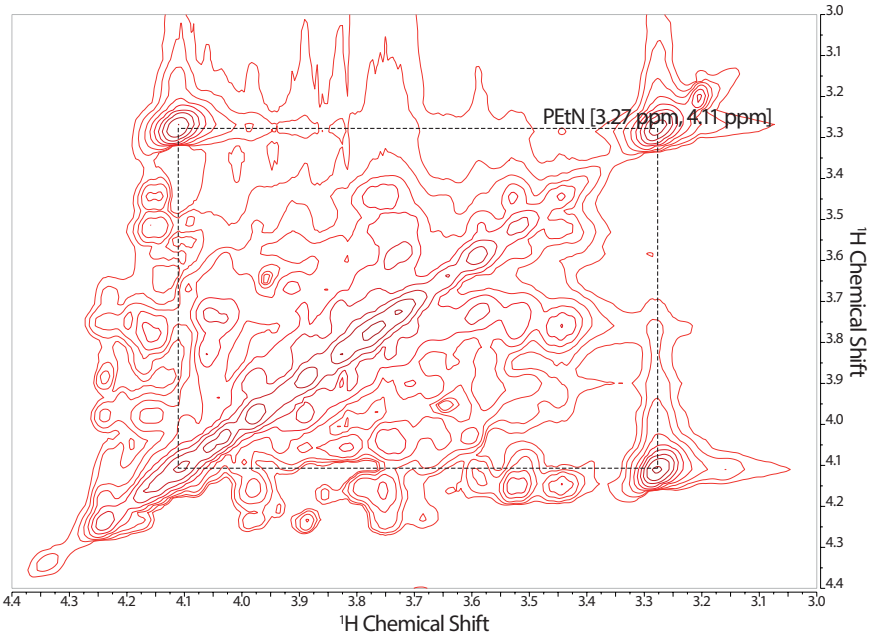


Figure 3-11 Expansion of the TOCSY spectrum of the LPS extracted from the WT *H. pylori* strain G27. The LPS was incorporated into the DPC micelles prior to the NMR experiments. The NMR spectrum was recorded by using a Bruker Avance III 600MHz NMR spectrometer equipped with a TXI/TCI cryoprobe. Cross-peaks in the spectrum arising from PEtN are indicated.

In the spectra, a signal at 3.27 ppm was observed and linked to a signal at 4.11 ppm by a strong cross-peak. The chemical shift of 3.27 ppm is characteristic of amine compounds, encouraging us to assign it to the protons of PEtN that was reported to modify *H. pylori* LPS previously (see **Section 1.2.1.1**).

The remaining content focuses on the stereochemistry of monosaccharides (α and β configurations) contained in different LPS samples. The spectra of the ΔWaaL and ΔHP0102 LPS are the simplest in terms of the monosaccharide composition, indicating they are most seriously truncated. These spectra were best interpreted and discussed in detail. Other mutant spectra were analysed based on the comparison between the spectra of each mutant and ΔWaaL/ΔHP0102.

Note that the interpretations of all the NMR spectra were tentative rather than definitive, though all of the proposed assignments were highly consistent with the previous reported chemical shifts, and information extracted from the NMR analysis fully matched the conclusions made based on other analytical techniques such as MS. In-depth NMR research on the LPS from several *H. pylori* LPS was carried out by other groups (see **Section 3.2.1**). Based on the previous studies, this NMR study only aims to characterise the stereochemistries of the monosaccharides contained in the LPS samples. More detailed NMR research is beyond the scope of this thesis.

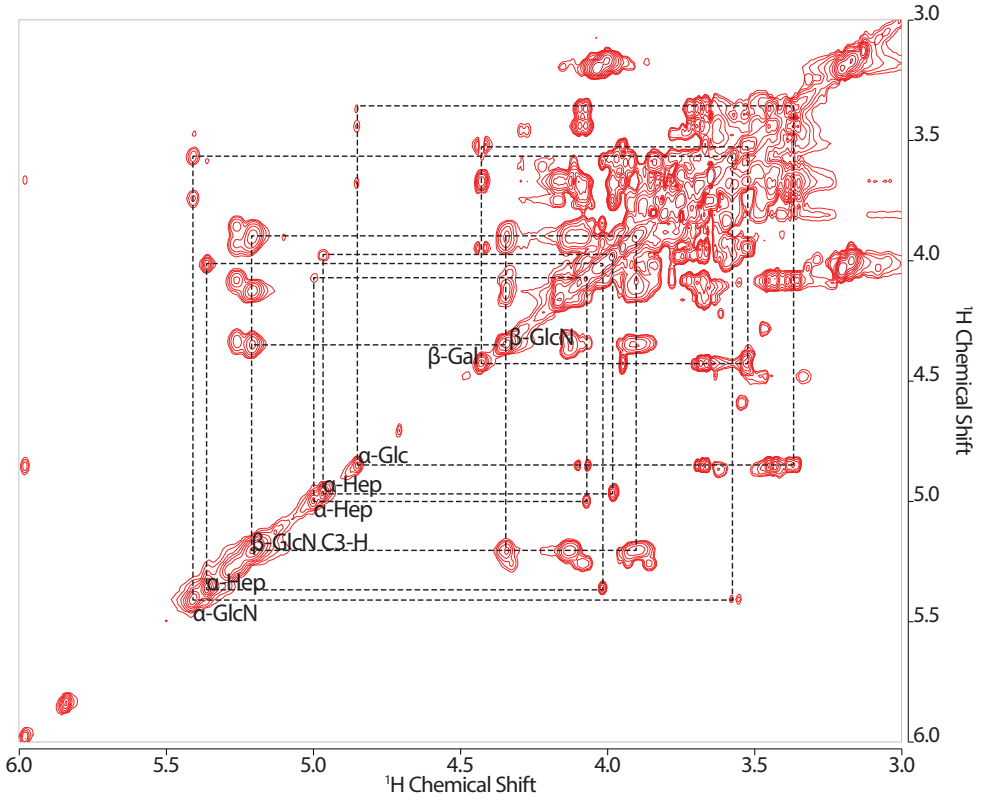


Figure 3-12 Expansion of the TOCSY spectrum of the LPS extracted from *H. pylori* strain G27 ΔWaaL. The LPS was incorporated into DPC micelles prior to the NMR experiments. The NMR spectrum was recorded by using a Bruker Avance III 600MHz NMR spectrometer equipped with a TXI/TCI cryoprobe. The spectrum was partially assigned with the interpretable monosaccharide spin systems.

3.2.3.1 ΔWaaL and ΔHP0102

TOCSY of ΔWaaL LPS is shown in **FIG. 3-12**. Although this spectrum was much simpler relative to those of other samples, some cross peaks with chemical shifts between 3.0 and 4.5 ppm still overlap, which hindered its complete interpretation. Nevertheless, peaks corresponding to seven anomeric protons were identified (**FIG. 3-12**), and their chemical shifts are listed in **Table 3-3**.

In principle, the assignment of TOCSY can be further confirmed by HSQC as the latter spectra provide extra information on the ^{13}C chemical shifts. Unfortunately, most of the anomeric carbon-proton cross peaks in the HSQC spectra were not visible, which was another main reason why the NMR data could not be fully interpreted.

It was intriguing that the anomeric protons from only seven monosaccharides were identified (note that KDO is a ketose and does not have anomeric protons), which indicates that mutation of the ligase much more seriously truncates the LPS than previously expected. Based on this observation, we proposed that the O-antigen of *H. pylori* LPS is much longer, containing not only the poly-lacNAc and Lewis epitopes but also the heptan, glucan and the novel Trio. This proposal was further supported by the MS analysis through the rest of this chapter.

TOCSY of ΔHP0102 LPS is shown in **FIG. 3-13**. By comparing this spectrum with the ΔWaaL spectrum, it was easy to observe an additional signal at 5.02 ppm, which was assigned to the anomeric proton of a GlcNAc. In addition, the characteristic signal for N-acetyl protons at 2.05 ppm is significantly higher in the spectrum of ΔHP0102 LPS, indicating it has more GlcNAc than the ΔWaaL LPS (compare **FIG. 3-14 a** and **b**).

Table 3-3 Tentative ^{1}H and ^{13}C Chemical shifts of the ΔWaaL LPS. All chemical shift values are expressed in ppm. ND stands for non-detectable, and the bar means non-existence. When only one chemical shift value is shown, it is the ^{1}H chemical shift.

Unit	H/C 1	H/C 2	H/C 3	H/C 4
6-linked β-GlcN (major)	4.43	3.98	5.28/71.66	4.20/64.18
6-linked β-GlcN (minor)	4.42	4.01	5.32	4.16
6-linked α-GlcN 1-Pi	5.48	3.81	2.63*	3.54*/60.48
5-linked KDO	-	-	2.10(e)/1.86	4.16
3-linked α-L,D-Hep	5.08	4.16/70.99	ND	ND
2-linked α-L,D-Hep	5.43	4.09	3.99	3.82
2,7-linked α-D,D-Hep	5.04	4.09/72.90	ND	ND
4-linked β-Gal	4.50	3.61/71.98	3.74	4.02/77.97
Terminal α-Glc	4.93/100.96	3.52/73.06	3.75/73.82	3.45/70.44

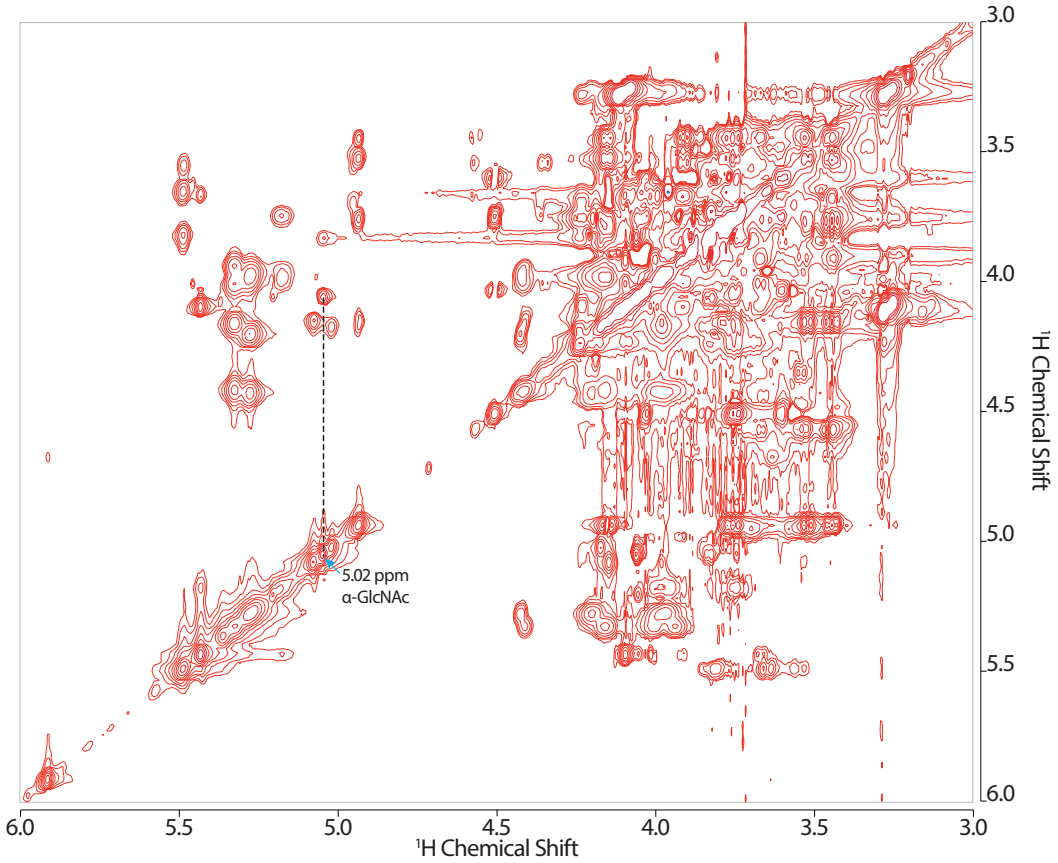


Figure 3-13 Expansion of the TOCSY spectrum of the LPS extracted from the *H. pylori* strain G27 ΔHP0102. More detailed information can be found in the legend of **FIG. 3-12**.

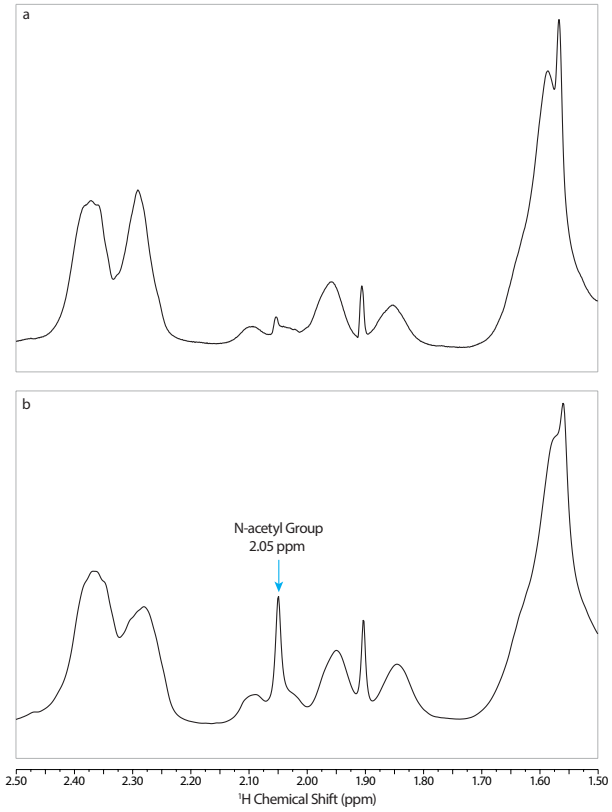


Figure 3-14 Expansion of the ¹H NMR spectrum of the LPS extracted from the *H. pylori* strain G27 ΔWaaL (a) and ΔHP0102 (b). The spectra are assigned with the putative chemical shift of N-acetyl groups. More detailed experimental information can be found in the legend of **FIG.3-13**.

Table 3-4 Tentative ^{1}H and ^{13}C assignments for an α–GlcNAc that was observed in ΔHP0102 but not in ΔWaaL. More detailed information can be found in the legend of **Table 3-3**.

Unit	H/C 1	H/C 4	N-acetyl H/C
3-linked α-GlcNAc	5.02	4.18/77.99	2.10/

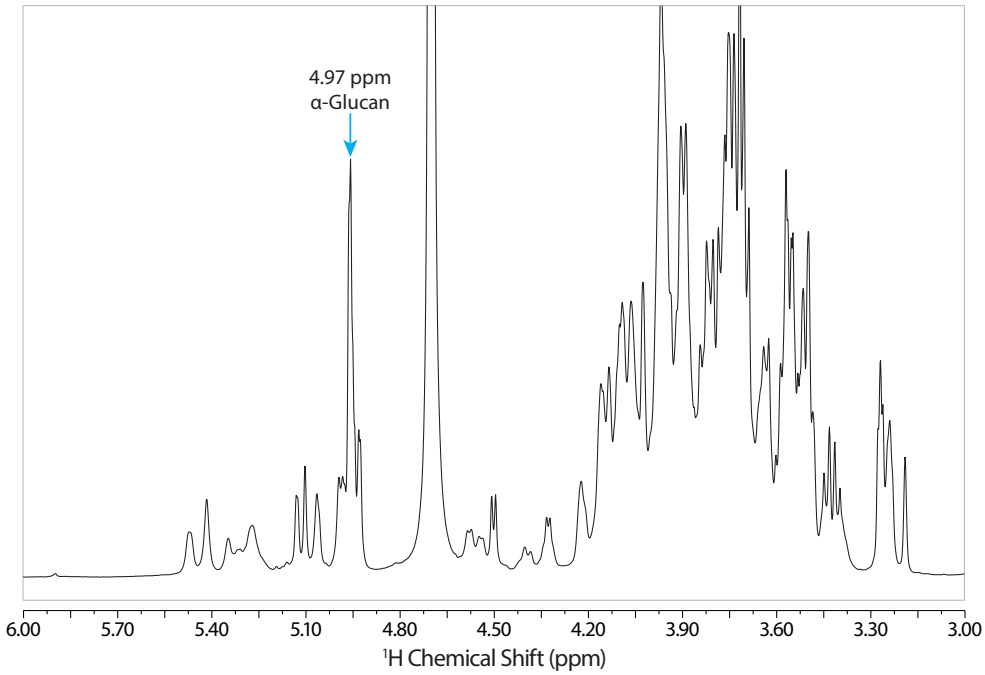


Figure 3-15 Expansion of the 1H NMR spectrum of the *H. pylori* ΔHP1283 LPS. A signal tentatively assigned to α-glucan is marked. More detailed information can be found in the legend of **FIG. 3-12**.

Previous research assigned the signal at 4.91 ppm to the anomeric proton of a β-GlcNAc, however, this chemical shifts fall in the range of an anomeric proton with an α-configuration (normally 4.8 to 6 ppm). The signal observed here shifted even more down field, therefore, we tentatively decided to assign it to an α–GlcNAc (**Table 3.4**).

3.2.3.2 ΔHP1283 and ΔHP1578

The ^{1}H NMR spectrum of the ΔHP1283 LPS is shown in **FIG. 3-15**. It was interesting to observe a very strong signal at chemical shift 4.97 ppm. We decided to assign this signal to the anomeric protons of a 6-linked α-glucan, on the basis of its characteristic chemical shift and strong intensity. The 6-linked α-glucan was found in the LPS from *H. pylori* serogroup O:3 (Altman *et al.* 2013), and our data match well with the previous data.

The existence of the 6-linked α-glucan was further suggested by the TOCSY spectrum of the ΔHP1283 LPS (**FIG. 3-15**). Several very strong and well resolved cross correlations (at chemical shift 3.91, 3.71, 3.58 and 3.51 ppm) of the signal at chemical shift 4.97 ppm were found. These ^{1}H chemical shifts were further confirmed by checking heteronuclear correlation cross peaks on the HSQC spectrum. Though the HSQC data are not shown here, the observed ^{13}C chemical shifts were shown in **Table 3-5**. Another peak at 5.14 ppm was ascribed to the anomeric proton of the glucose unit in the glucan linking to the Trio.

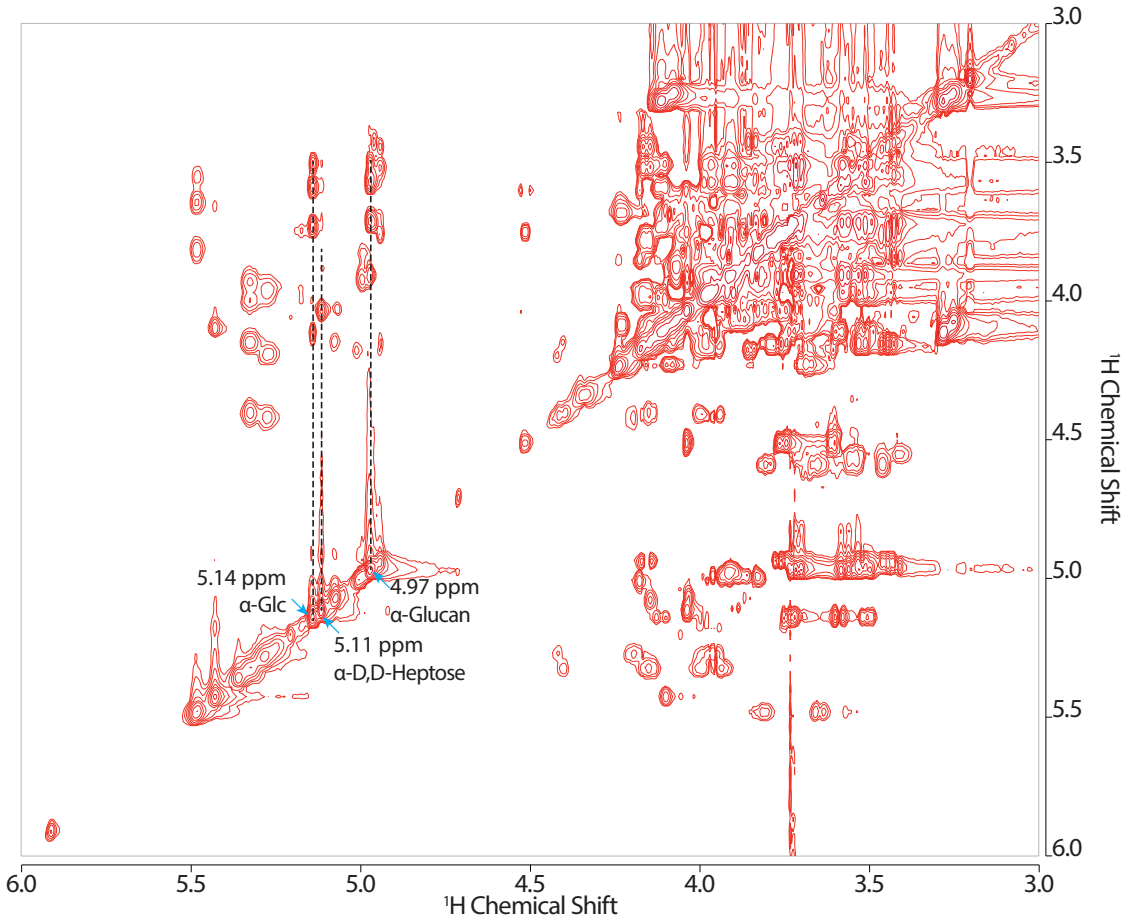


Figure 3-16 A Expansion of the TOCSY spectrum of the LPS extracted from the *H. pylori* strain ΔHP1283. The spectrum is annotated with the tentative assignments of resonances from monosaccharides that were not observed in ΔWaaL/ΔHP0102 LPS. More detailed information can be found in the legend of **FIG. 3-12**.

When the TOCSY is zoomed into the region of high field (chemical shift 0.7-1.7 ppm), a strong cross peak (1.14, 4.34 ppm, **FIG.3-16 B**) that was not observed for ΔWaal/ΔHP0102 was found. A signal at 1.15 ppm can normally be due to a deoxy monosaccharide in an NMR spectrum of sugars. Considering previous research revealed a fucose in the Trio motif (Altman *et al.* 2013; Altman, Chandan, Li & E. Vinogradov 2011b; Altman, Chandan, Li & E. Vinogradov 2011a), we decided to assign this signal to the methyl protons on the fucose C6. This was also supported by a piece of HSQC data where a cross peak at 1.15, 16.27 ppm was found. Unfortunately, the associated anomeric proton resonance hides under other signals, so could not be identified with confidence.

In addition, the signal at 5.11 ppm was tentatively assigned to the anomeric proton of the α-D,D-heptose in the Trio. Though this assignment matches previously published data (Altman *et al.* 2013), not much supporting evidence for this assignment could be found; for instance, only one cross peak at 5.11, 4.04 ppm was observed in TOCSY and no cross peak was found in the HSQC spectrum.

The TOCSY spectrum of ΔHP1578 LPS is shown in **FIG. 3-17**. This sample gave NMR spectra similar to that of the ΔHP1283 LPS, except for several more signals that could correspond to a 3-linked α-heptan (**Table 3-6**). Notably like glucan, the chemical shifts of the heptose linked to glucan is different from those of the heptan. Being similar to the 6-linked α-D,D-heptose signals observed in TOCSY of the ΔHP1283 LPS, limited supporting evidence was found for this assignment.

Table 3-5 Tentative ^{1}H and ^{13}C assignments for monosaccharides that were observed in ΔHP1283 but not in ΔHP0102. More detailed information can be found in the legend of **Table 3-3**.

Unit	H/C 1	H/C 2	H/C 3	H/C 4	H/C 5	H/C 6
3-linked α-Fuc	4.96	ND	ND	ND	4.34/68.09	1.15/16.27
6-linked α-D,D-Hep	5.11	4.04	ND	ND	ND	ND
6 linked α-Glc	5.14	3.59/72.64	3.74/74.57	3.51/70.75	4.12/70.95	ND
6-linked α-Glucan	4.97/98.85	3.58/72.50	3.71/74.52	3.51/70.70	3.91/71.43	ND

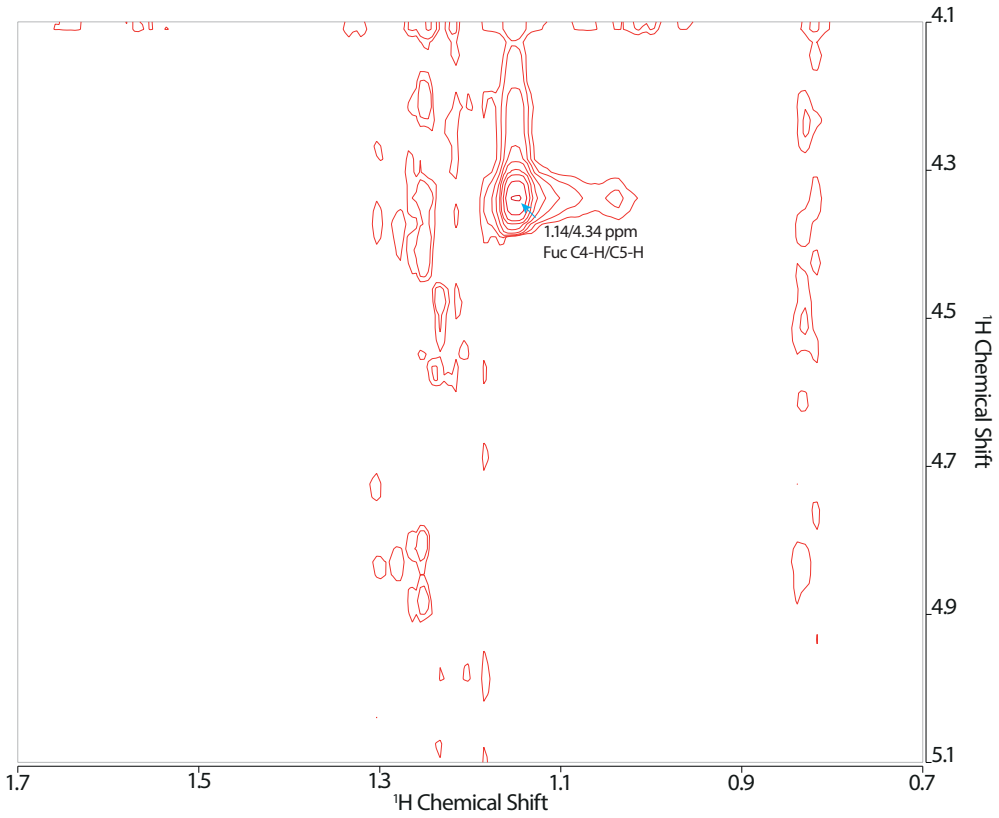


Figure 3-16 B Expansion of the TOCSY spectrum of the LPS extracted from *H. pylori* strain ΔHP1283. The spectrum is zoomed into the region of the fucose C6-H chemical shift and tentatively assigned. More detailed information can be found in the legend of **FIG. 3-12**.

Table 3-6 Tentative ^{1}H and ^{13}C assignments for the α–heptan that was observed in ΔHP1578 but not in ΔHP283. More detailed information can be found in the legend of **Table 3-3**.

Unit	H/C 1	H/C
3-linked α-Hep	5.10/ND	4.23/70.50
3-linked α-Heptan	5.21/ND	4.21

Table 3-7 Tentative ^{1}H and ^{13}C assignments for fucoses observed in WT LPS. The NMR signals corresponding to two α-Fuc residues 3-linked to Gal ware observed, which suggests the existence of both terminal Lewis Y and repeating Lewis X. More detailed information can be found in the legend of **Table 3-3**.

Unit	H/C 1	H/C 6	H/C
3-linked α-Fuc (Trio)	ND	1.15/16.29	4.34/67.99
Terminal α-Fuc (1-3 Gal)	ND	1.23/16.49	4.87
Terminal α-Fuc (1-3 Gal)	ND	1.15/16.29	4.80/67.80
Terminal α-Fuc (1-2 Gal)	5.27	1.26/16.52	3.97/69.19, 4.24

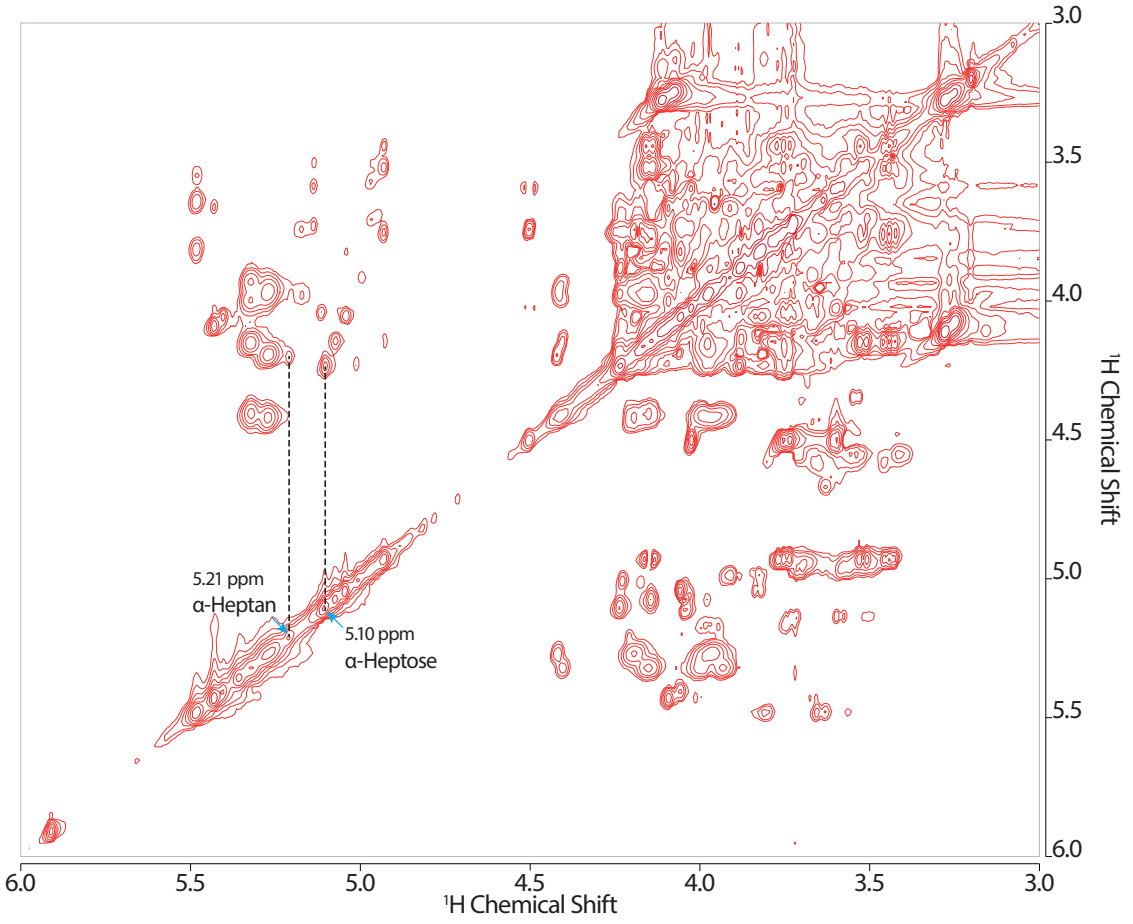


Figure 3-17 Expansion of the TOCSY spectrum of the LPS extracted from *H. pylori* strain G27 ΔHP1578. The spectrum is notated with tentative assignments for monosaccharides that are not observed in ΔWaaL/ΔHP0102/ΔHP1283 LPS. More detailed information can be found in the legend of **FIG. 3-12**.

3.2.3.3 WT and ΔHP1284

The TOCSY spectra of WT *H. pylori* strain G27 LPS are shown in **FIG. 3-18 A** and **B**. It was extremely difficult to identify any new signals corresponding to the monosaccharides in this region of chemical shift (3-6 ppm). This was mainly because the peaks were very broad and their intensities were irregular, *i.e.* some weak signals hid under strong ones becoming unidentifiable. This indicates well the complexity of the WT G27 LPS.

In order to obtain more information, this spectrum was zoomed to the region where the signals for the 3-linked α-Fuc were observed (**FIG. 3-17**), as the signals were present in a highly separated region of the spectrum, and were representative of the Lewis antigens that were not observed in the described samples. As expected, more signals that could be assigned to the protons on the Fuc C-6 were found (**Table 3-7**), which suggests the WT *H. pylori* strain G27 carries Lewis antigens.

In addition to the NMR signal at 1.15 ppm that was assigned to the C-6 proton of the 3-linked α-Fuc in the Trio, three more ^{1}H signals potentially corresponding to the C-6 proton of Fucs were observed. The signal at 1.26 ppm was assigned to the C-6 proton of the terminal α-Fuc 1-2 linked to a Gal, which is characteristic of the Lewis Y epitope. Two cross peaks at 1.15, 4.34 ppm and 1.15, 4.80 ppm were resolved in the TOCSY of the WT LPS, and they were tentatively assigned to the terminal α-Fucs 1-3 linked to Gals. They are probably representative of terminal Lewis Y and the poly-Lewis X. These signals were further supported by cross peaks in HSQC.

The spectra of the ΔHP1284 LPS were very informative, as they indicated the ΔHP1284 LPS has all the Lewis antigens, heptan, glucan and Trio structures; but it misses the α-glucose (the anomeric proton at 4.93 ppm) found in the spectra of all the other samples. The TOCSY spectrum of the ΔHP1284 LPS is shown in **FIG. 3-19**, and it clearly showed that the α-glucose signals are missing. This is indicative of an incorrectly synthesized core-oligosaccharide.

Indeed, the ΔHP1284 LPS is very unusual as it has a normal O-antigen but an incompletely synthesized core-oligosaccharide. Further evidence is provided by the rest of the chapter. Notably, though the previously described signals at chemical shift 5.05 and 4.45 ppm corresponding to the α-D,D-heptose and β-Gal in the core-oligosaccharide are also missing in this spectrum, they are not the best evidence for an incorrectly synthesised core because the signal at 5.05 ppm is also missing in TOCSY of the ΔHP1283 LPS, and the signal at 4.45 ppm is missing in several other TOCSYs including that of WT LPS. These inconsistencies could be attributed to the irregularity of the peak intensities, *i.e.* the NMR signals at 4.45 and 5.05 ppm could be too weak to be observed in the corresponding spectra.

The NMR data for ΔHP0156 and G27ΔHPG27_1230 LPS were very similar to WT LPS, therefore, they are not shown and further discussed in this section.

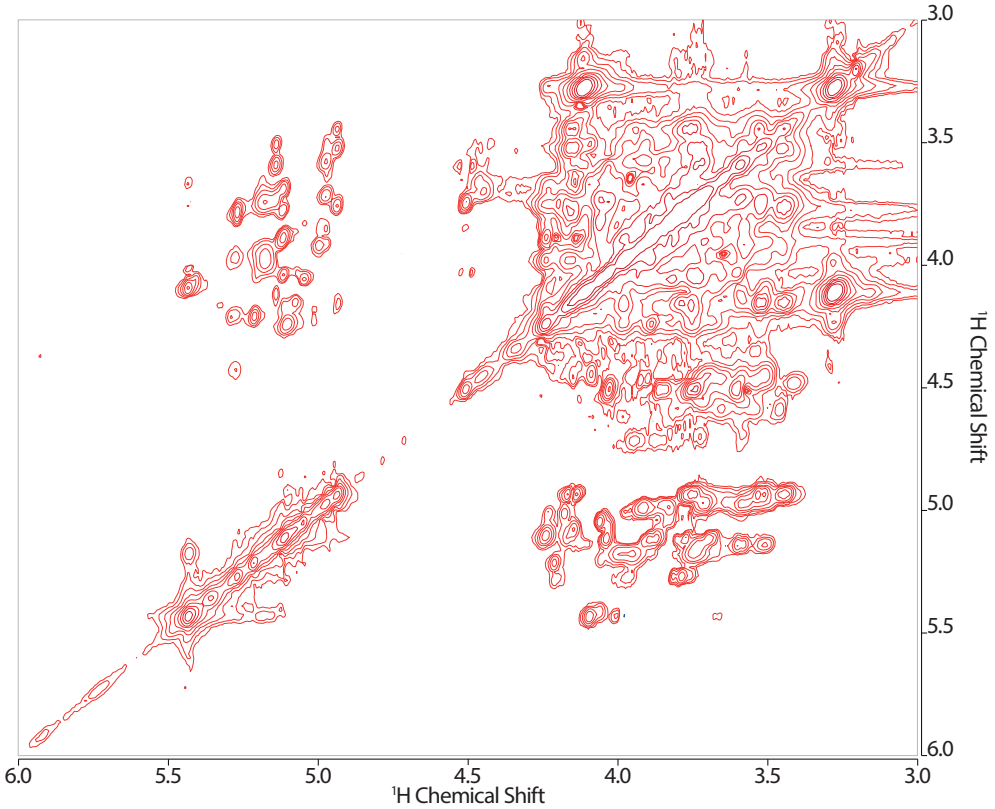


Figure 3-18 A Expansion of the TOCSY spectrum of the LPS extracted from WT *H. pylori* strain G27. More detailed information can be found in the legend of **FIG. 3-12**.

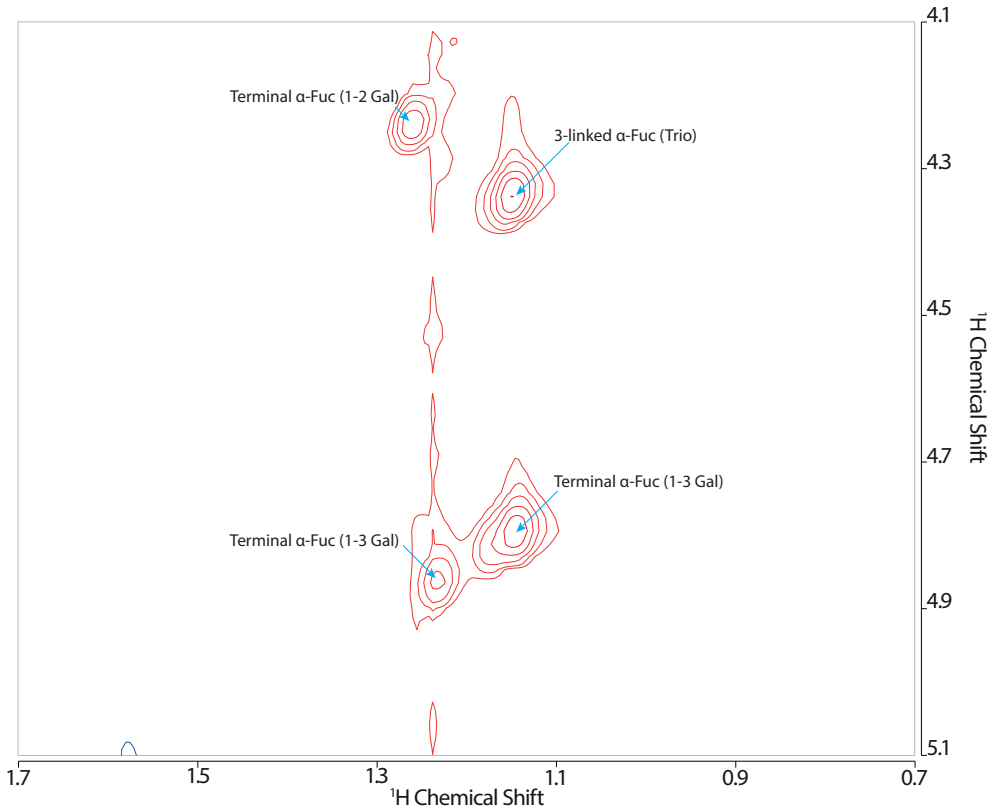


Figure 3-18 B Expansion of the TOCSY spectrum of the LPS extracted from WT *H. pylori* strain G27. The spectrum is zoomed into the region of fucoses C6-H signals, and tentative assignments are shown. More detailed information can be found in the legend of **FIG. 3-12**.

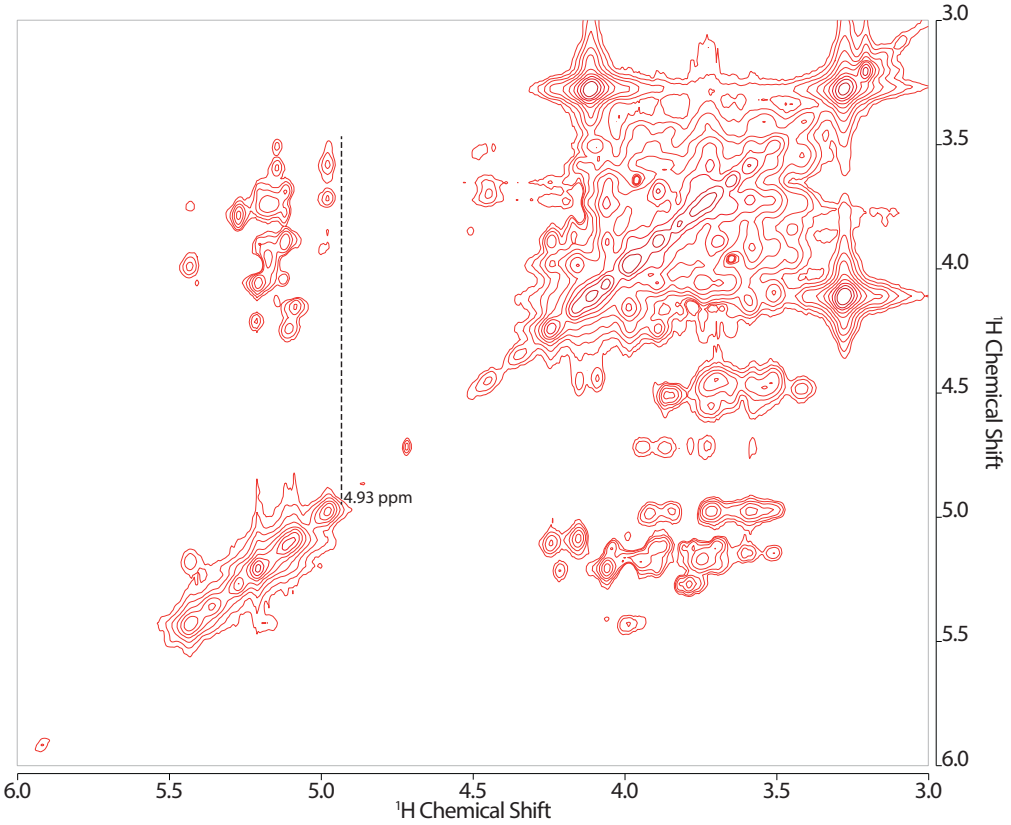


Figure 3-19 Expansion of the TOCSY spectrum of the LPS extracted from the WT *H. pylori* strain G27 ΔHP1284. It does not contain the signals corresponding to the α-glucose, indicating that this monosaccharide is missing in the motif of the core-oligosaccharide. More detailed information can be found in the legend of **FIG. 3-12**.

3.2.4 GC-EI-MS Linkage Analysis

Linkage analyses were done on six samples including the WT, ΔHP1284, ΔHP1578, ΔHP1283, ΔHP0102 and ΔWaaL LPSs. As mentioned in **Section 3.0**, the linkage analysis has been well established for decades. Briefly, the samples were permethylated and completely hydrolysed into partially methylated monosaccharides that were further reduced by $NaBD_4$ and acetylated. The partially methylated alditol acetates (PMAAs) were analysed by GC-EI-MS in order to determine the linkages of the monosaccharide. The obtained Total Ion Count (TIC) chromatograms are shown in **FIG. 3-20** and the retention time and characteristic ions used for determining monosaccharide linkage analysis are summarized in **Table 3-8**. Note that the D/L configurations cannot be identified by a simple linkage analysis, the annotations of the D,D-Hep and L,D-Hep were based on previous research (Altman, Chandan, Li & E. Vinogradov 2011b).

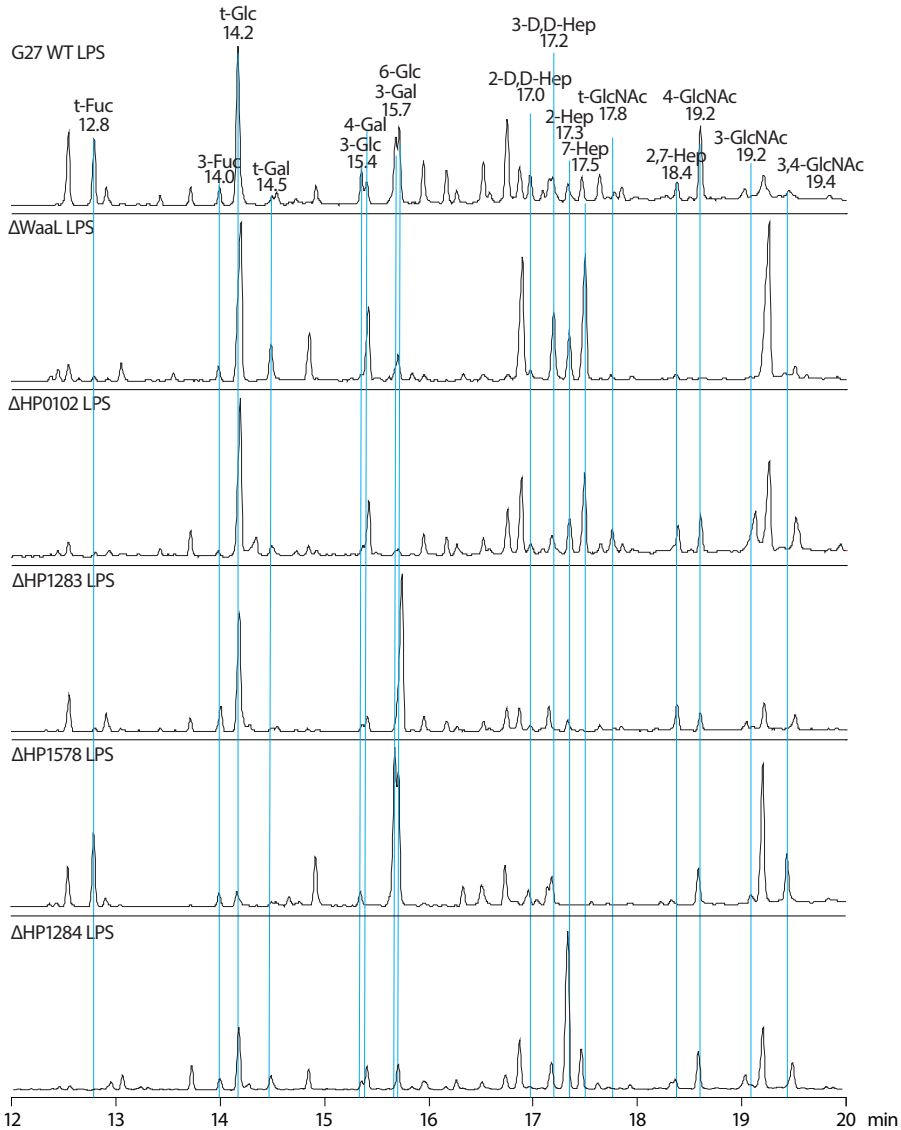


Figure 3-20 Linkage Analysis of the LPS samples from G27 WT, ΔWaaL, ΔHP0102, ΔHP1283 and ΔHP1284. The Total ion count (TIC) chromatograms are shown. The retention times (RT) and monosaccharides are annotated above the chromatogram of the WT LPS. Note that not all of the peaks are annotated mainly because of contamination.

Table 3-8 GC-EI–MS linkage analysis of partially methylated alditol acetates from LPS samples

	RT (min)	Characteristic Ions
t-Fuc	12.8	89+102+115+118+131+175
3-Fuc	14.0	87+118+131+174+190+234+247
t-Glc	14.2	87+102+118+129+145+161+162+205
t-Gal	14.5	87+102+118+129+145+161+162+205
3-Glc	15.4	101+118+129+161+202+217+234
4-Gal	15.4	87+99+102+113+118+129+131+162+173+233
2-Gal	15.6	87+88+101+129+130+161+190+205+234
3-Gal	15.7	101+118+129+161+202+217+234
6-Glc	15.7	87+102+118+129+162+189+233
2-DD-Hep	17.0	89+101+130+190+205+234+249+306+350
3-DD-Hep	17.2	101+118+202+205+234+276+290+318+350
2-Hep	17.3	89+101+130+190+205+234+249+306+350
7-Hep	17.5	102+117+118+162+189+233+277+321+322
3-Hep	17.6	101+118+202+205+234+276+290+318+350
t-GlcNAc	17.8	117+129+143+145+159+203+205+247+258
2,7-D,D-Hep	18.4	117+130+190+201+233+277+350
4-GlcNAc	18.6	117+129+143+159+203+233
3-GlcNAc	19.1	117+129+159+231+258+275+318
6-GlcNAc	19.2	117+129+159+189+346
3,4-GlcNAc	19.4	117+129+159+189+346

These chromatograms were quite complicated due to some under-permethylation, contamination and complexity of the samples. Under-permethylation altered the elution time and mass fingerprints of the PMAAs, giving rise to multiple peaks that do not match any glycan standards. There are three main sources of contamination. Firstly, non-sugar material such as plastics caused contamination. These materials co-eluted with the PMAAs, leading to a drastic increase in their peak intensities. Secondly, naturally occurring glycans contaminated the samples. For instance, cellulose could contaminate the samples if tissues were used during the sample preparations. And finally, the LPS samples could cross-contaminate each other. For example, all the monosaccharides contained in the WT LPS were detected in the ΔWaaL LPS in a trace amount. This might be because a tiny amout of the WT LPS was mixed into the ΔWaaL LPS sample.

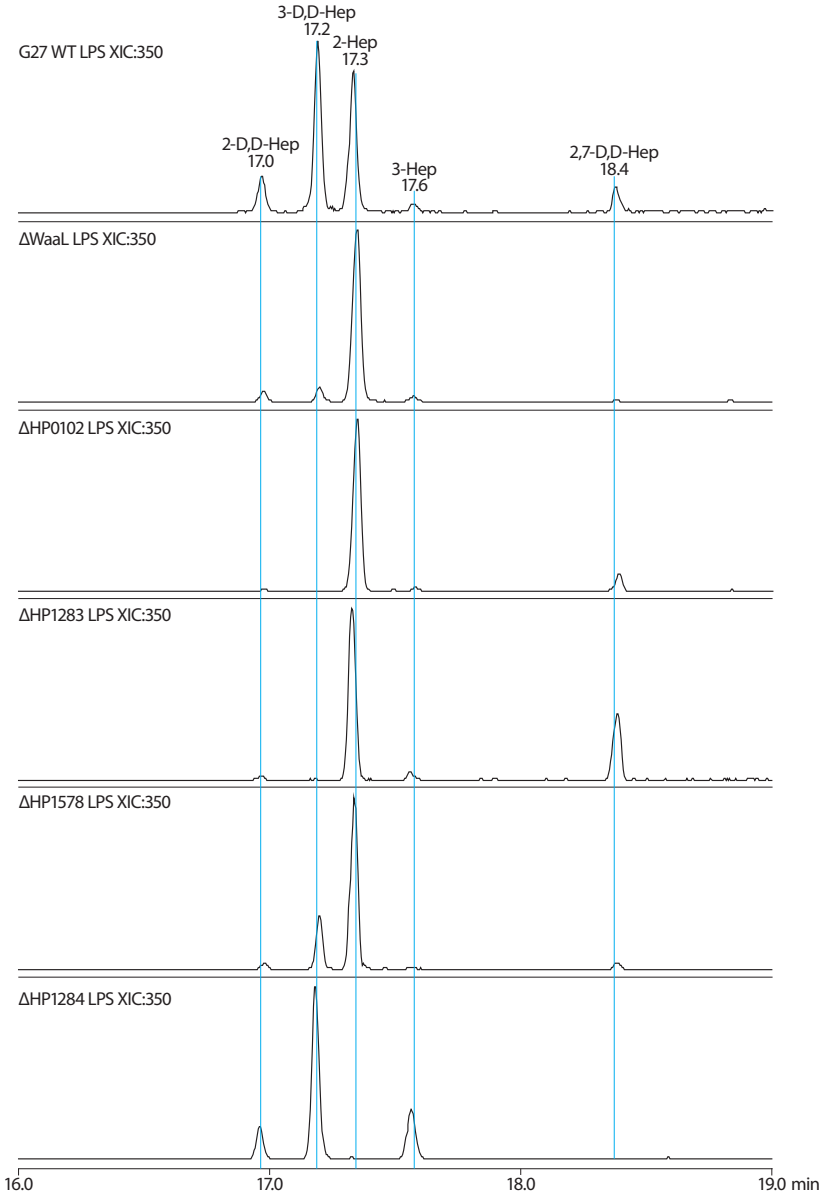


Figure 3-21 Linkage Analysis of the LPS samples from G27 WT, ΔWaaL, ΔHP0102, ΔHP1283 and ΔHP1284. The chromatograms are extracted with *m/z* 350 and shown. The retention times (RT) and monosaccharides are annotated above the chromatogram of the WT LPS.

Some characteristic ions were extracted from the TIC chromatograms to produce the Extract Ion Count (XIC) chromatograms. In this way, certain types of PMAAs were highlighted, facilitating the data interpretation. The ion at *m/z* 350 was chosen for representing the PMAAs of different Heps, as it is a relatively large fragment ion that normally cannot be observed in the MS fingerprints of the PMAAs of hexoses and hexNAcs. The *m/z* 350 XIC chromatograms are shown in **FIG. 3-21**. The 6-Hep in the Trio structure was not detected in any sample, which is a reminder that certain low-abundant PMAAs could be missed. The 3-D,D-Hep was not detected in the LPS from ΔHP1283, ΔHP0102 and ΔWaaL, indicating they do not carry the heptan. For the ΔWaaL LPS, instead of the 2,7-D,D-Hep, a 7-D,D-Hep was detected, suggesting a linear core-oligosaccharide structure (see **Section 3.2.5.3**).

The peak corresponding to the 7-D,D-Hep is not present in **FIG. 3-22** as the largest fragment ion observed in its MS spectrum is at *m/z* 322. For the ΔHP1284 LPS, neither the 2,7-D,D-Hep nor the 7-D,D-Hep was observed, which is fully consistent with the conclusion that this sample misses the outer-most heptose in the core-region (see **Section 3.2.5.4**).

The XIC chromatograms are advantageous due to their facilitating quantitative comparisons of different PMAAs. For instance, the peak corresponding to the 3-D,D-Hep was the strongest in the *m/z* 350 XIC of the G27 WT LPS, indicating the heptan is 3-linked. However, it is necessary to be cautious to evaluate the quantitative data, because types of linkages greatly affect the peak intensity, *i.e.* the peaks corresponding to the terminal sugars are normally very strong. In addition, contaminations can also alter the abundance of the peak. For instance, the peak corresponding to 2-Hep was very strange in the *m/z* 350 XIC of the G27 WT LPS, *i.e.* its intensity was nearly the same as the 3-D,D-Hep peak. This is because some contaminations co-eluted with 2-Hep, which drastically increased the intensity of the peak. The same phenomenon was observed in the *m/z* 350 XIC of ΔHP1578.

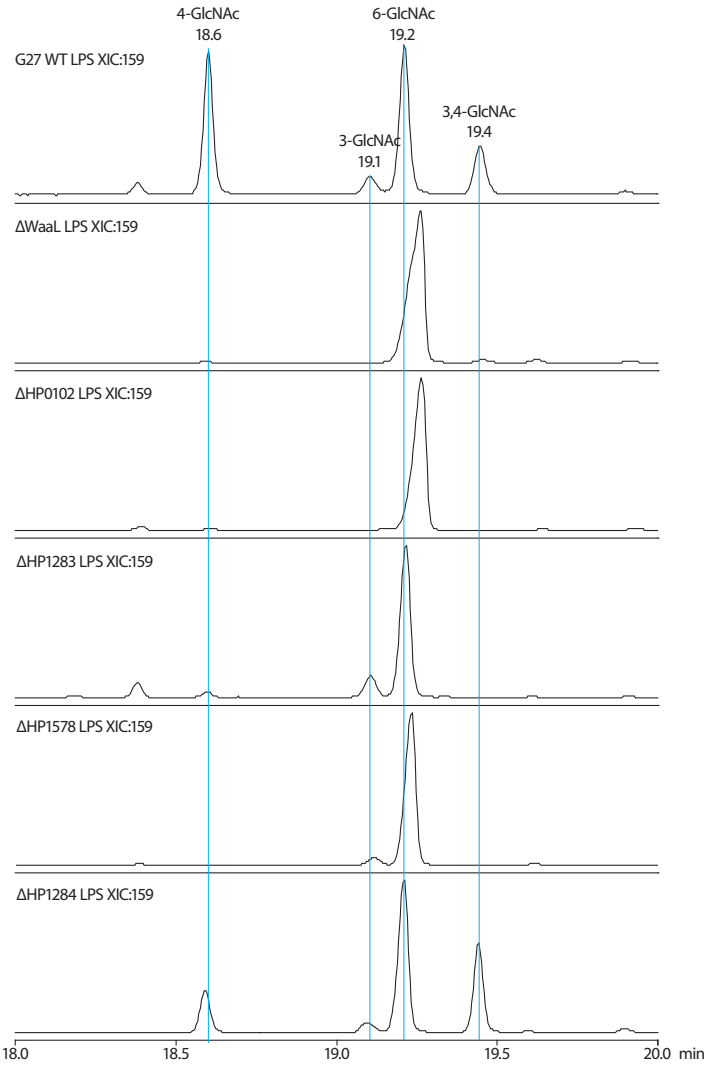


Figure 3-22 Linkage Analysis of the LPS samples from G27 WT, ΔWaaL, ΔHP0102, ΔHP1283 and ΔHP1284. The chromatograms are extracted with *m/z* 159 and shown. The retention times (RT) and monosaccharides are annotated above the chromatogram of the WT LPS.

The fragment ion at *m/z* 159 containing a nitrogen atom was chosen for representing the PMAAs of different GlcNAcs (**FIG. 3-22**). It was obvious that all samples contained a very strong peak putatively corresponding to 6-GlcNAc that is probably from lipid A. This was difficult to be confirmed due to a lack of the 6-GlcNAc standard. The elution of this peak is slightly delayed in the XICs of the ΔWaaL and ΔHP0102 LPS because of a minor over-loading. As expected, the PMAAs of the 4-GlcNAc and 3,4-GlcNAc that are from Poly-lacNAc/Lewis structures were observed in the XICs of the WT and ΔHP1284 LPS, indicating that they contain the Poly-lacNAc/Lewis structures. The peak corresponding to the 3-GlcNAc shifted to the fairly strong peak corresponding to the t-GlcNAc in the *m/z* 159 XIC of the HP0102 LPS, which strongly indicates that it is truncated from the fucose in the Trio. Neither of these two peaks was observed in the XIC of ΔWaaL, suggesting that it is more seriously truncated.

More XICs were generated in order to completely understand the linkage data. For instance, a strong peak corresponding to 6-Glc was observed in the ΔHP1283 LPS, which is consistent with the NMR data, indicating this LPS carries a long glucan. The ΔHP1284 LPS contained no t-Glc, further suggesting that it carries an incompletely synthesised core-oligosaccharide.

3.2.5 Methanolysis

The LPS samples were subjected to methanolysis under a mild acidic condition that can catalyse the reversible acetal reactions of sugars. The overall result of methanolysis is reminiscent of hydrolysis: a methoxyl group (CH_3O) instead of a hydroxyl group (OH) is added to the reducing end of a sugar, thus cleaving the glycosidic bond (**FIG. 3-23**). The methanolysed LPSs were then permethylated (or deutero-permethylated) before being analysed by MALTI-TOF and MALDI-TOF/TOF.

3.2.5.1 G27 WT

Methanolysed products of the WT LPS after permethylation and deutero-permethylation were subjected to the MALDI-TOF analysis (**FIG. 3-24 a** and **b**). The red peaks in the spectra correspond to sodiated permethylated glycans, and they are annotated with the mass-to-charge ratio (*m/z*) and structures.

The monosaccharide compositions of each peak were proposed based on *m/z*. For example, *m/z* 518 is the sum of a hexose residue (*m/z* 204), a hexNAc residue (*m/z* 245), reducing and non-reducing ends (*m/z* 46) and a sodium ion (*m/z* 23). Because the deduced glycan structures were very similar to the published results of LPS for the strain 26695, and GC-EI-MS analysis (see **Section 3.2.2** and **3.2.4**) did not show any new monosaccharide compared with the strain 26695 LPS, the exact identities of the monosaccharides were proposed based on the published structure of strain 26695 LPS (Altman, Chandan, Li & E. Vinogradov 2011b). Take the peak at *m/z* 518 for example, we postulated the hexose to be Gal and the hexNAc to be GlcNac. In this way, we deduced that the peak at *m/z* 518 corresponds to a lacNAc structure.

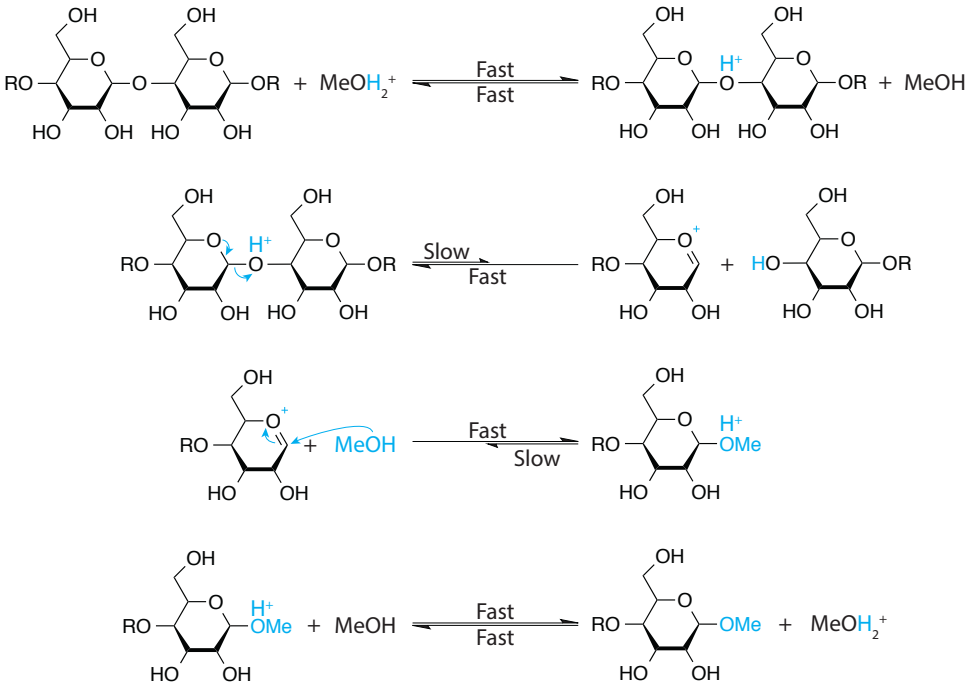


Figure 3-23 Mechanism of Methanolysis (Juvet & Chiu 1961). The blue arrows stand for the transferring of the electron pairs.

As no solid phase extraction method was exploited to purify the products after methanolysis, both spectra showed rather complicated glycan profiles, *i.e.* all three components of the LPS (lipid A, the core-oligosaccharide and O-antigen) were observed. Both permethylation and deutero-permethylation were used for derivatizing the same sample in order to confidently confirm the annotation of the MS spectra. Indeed, most methanolysed products were found in both MS spectra, which strengthens the proposed overall LPS structure.

The peaks corresponding to lacNAc increments were observed in both spectra. In order to simplify the annotation, only the lacNAc peaks at *m/z* 518.2 and 539.5 were annotated, while in fact peaks corresponding to two and three lacNAc units were also found, which is consistent with the Smith degradation data (see **Section 3.2.7**). The clusters of the peaks at *m/z* 695.3, 899.4, 1103.5, 1307.5 and 1151.6 in the permethylation spectrum (**FIG. 3-24 a**) and 725.5, 928.7, 1151.8, 1364.9 and 1578.1 in the deutero-permethylation spectrum (**FIG. 3-24 b**) were detected, which corresponds to the linear glucan-Hep-Fuc structure with a glucose number ranging from one to five.

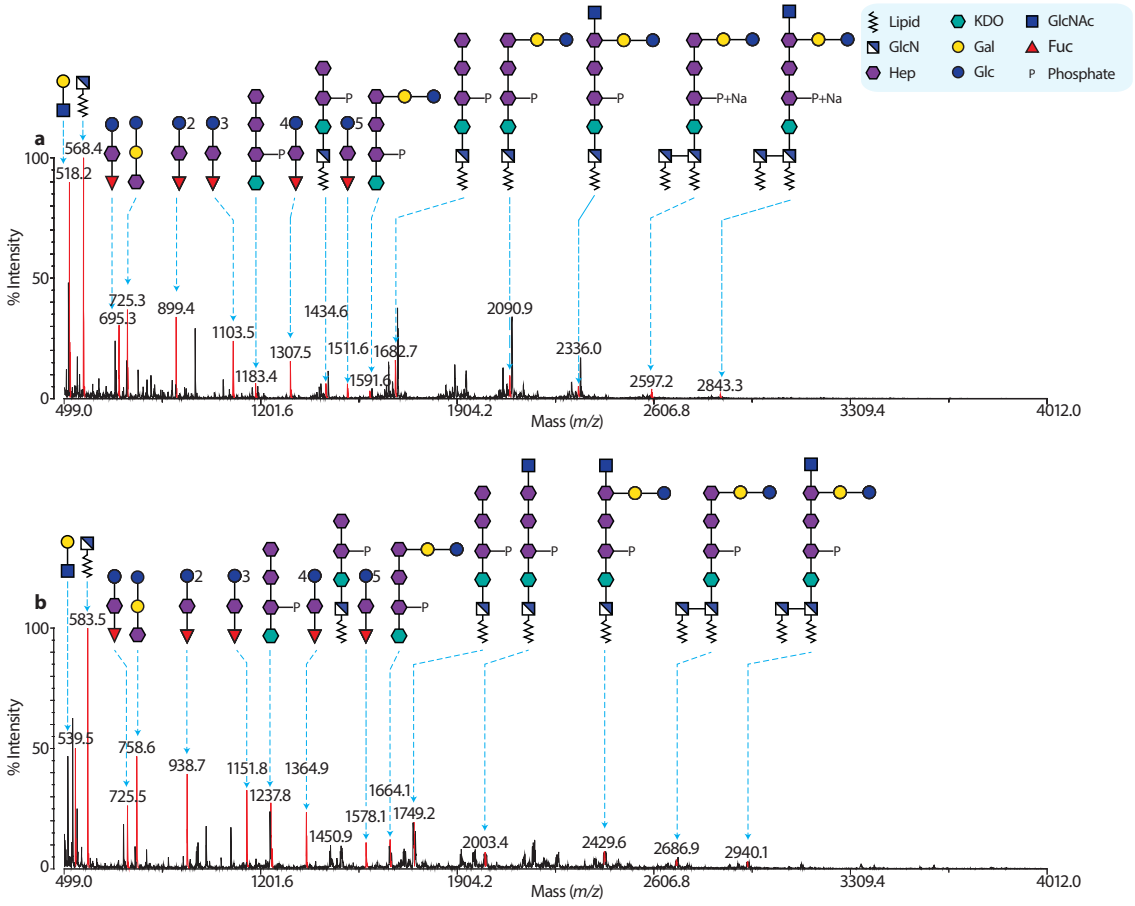


Figure 3-24 The MALDI-TOF spectra of the permethylated (a) and deutero-permethylated (b) WT LPS after methanolysis. The major MS peaks corresponding to fully permethylated glycans are coloured red, and they are annotated with *m/z* and glycan structures. When these peaks are too weak to annotate, under-permethylated peaks are annotated.

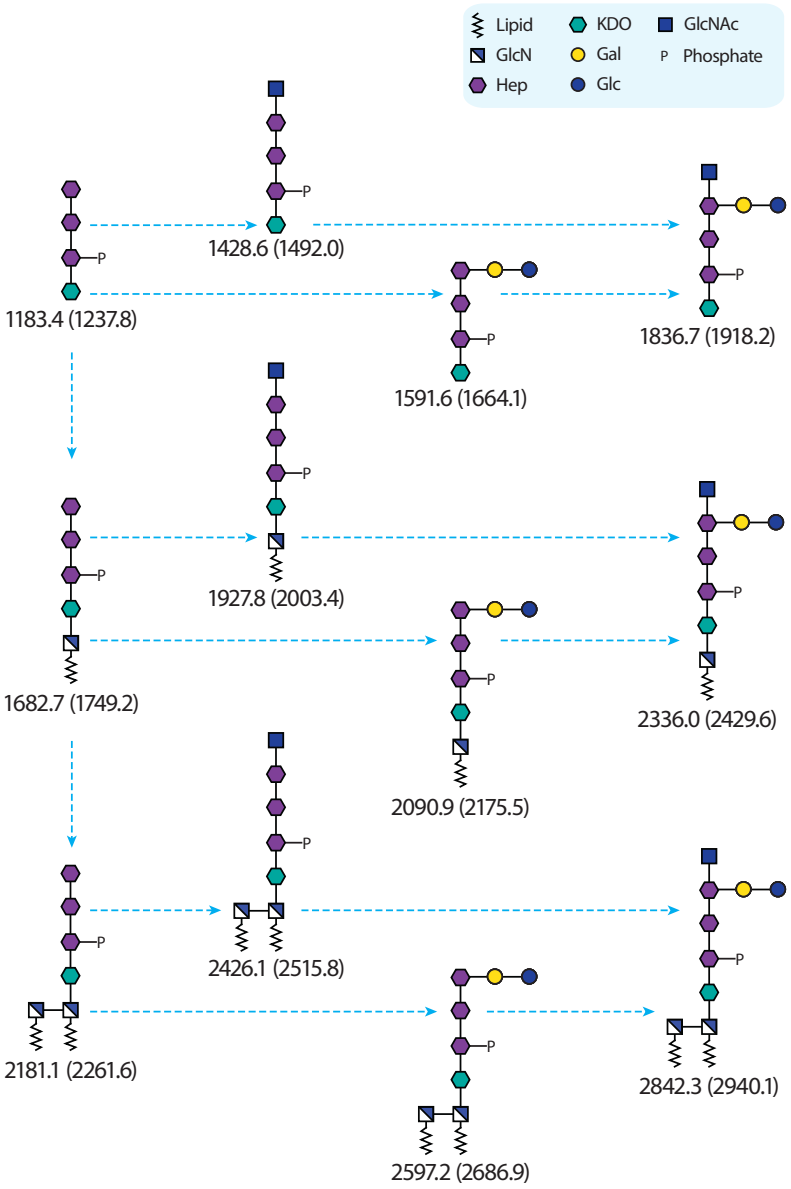


Figure 3-25 Diagram of the methanolysis products giving rise to the MS pattern of the core-oligosaccharide lipid A parts of the *H. pylori* G27 LPS samples. The methanolysis products are annotated with both permethylated and deutero-permethylated *m/z* values. The latter ones are shown in the brackets.

The core-lipid A part was not completely cleaved; consequently, their very strong signals contributed to a substantial content of the MS spectra. Fundamentally, the spectra showed a pattern consistent with two GlcN-lipids, a Glc-Gal and a GlcNAc being successively attached to a tri-Hep-KDO chain. This pattern arises from incomplete methanolysis of the glycosidic linkages in the LPS core domain. The glycan sequences giving rise to this pattern is shown clearly in **FIG. 3-25**.

All of the *m/z* values shown in **FIG. 3-25** were found in the MS spectra of permethylated or deutero-permethylated samples, but some of the MS peaks showed extremely low intensities, hiding under the strong peaks, and therefore cannot be annotated in the figure. For instance, the peak at *m/z* 1428.6 was too weak to be properly annotated in **FIG. 3-24 a**. Because of the tri-Hep-KDO chain being phosphorylated, prominent under-permethylation was observed.

For under-permethylation, one of the hydroxyl groups on the phosphate was sodiated, leading to the overall mass 8 Da higher than fully permethylated mass (23-15=8). As a result, some of the methanolysis products mainly existed in underpermethylated forms. For example, in **FIG.3-24 a**, the peak at *m/z* 2843.5 corresponded to the underpermethylated core-oligosaccharide lipid A capped with a GlcNAc.

The MS peaks attributed to the poly-lacNAc, glucan and core-oligosaccharide domains were further characterized by MALDI-TOF/TOF. As the poly-lacNAc structures were better characterized by the Smith degradation (see **Section 3.2.7**), only the MS/MS spectra of glucan-Hep-Fuc and GlcNAc-core-oligosaccharide are shown in **FIG. 3-26** and **3-27**, respectively. The MS/MS spectra in **FIG. 3-26** promisingly proved a linear structure formed by attaching a glucan on a Hep-Fuc disaccharide. All the MS/MS spectra in **FIG. 3-27** were dominated by the minus 126 Da peaks that were attributed to a neutral loss of a permethylated phosphate group. **FIG. 3-27 a**, **b** and **c** strongly suggested the core-oligosaccharide is a linear hexa-saccharide with a sequence of Glc-Gal-tri-Hep-KDO. A very weak peak at *m/z* 795.3 in **FIG. 3-27 d** is diagnostic to the di-Hep-KDO chain, indicating the Glc-Gal disaccharide attaching on the outer-most heptose. Though a branched structure was strongly supported by the 2.7-Hep detected by the GC-EI-MS Linkage analysis (see **Section 3.2.4**), no direct evidence shown in the MS/MS spectra confirmed the attachment site of the O-antigen.

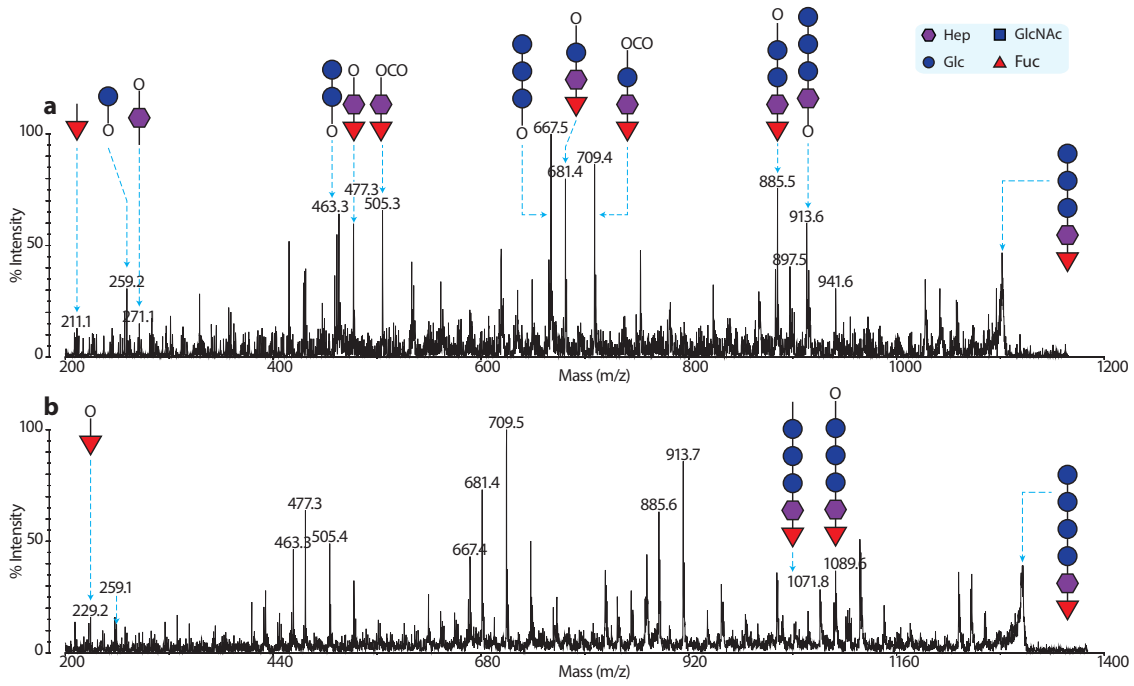


Figure 3-26 The MALDI-TOF/TOF spectra of two methanolysis products whose *m/z* values suggested that they contain parts of the glucan domain. MS/MS spectra of peaks at *m/z* 1103 and 1307 are shown in (a) and (b) respectively. Spectrum (a) is fully annotated with *m/z* and glycan structures, and only the peaks not annotated in (a) are annotated with glycan structures in (b).

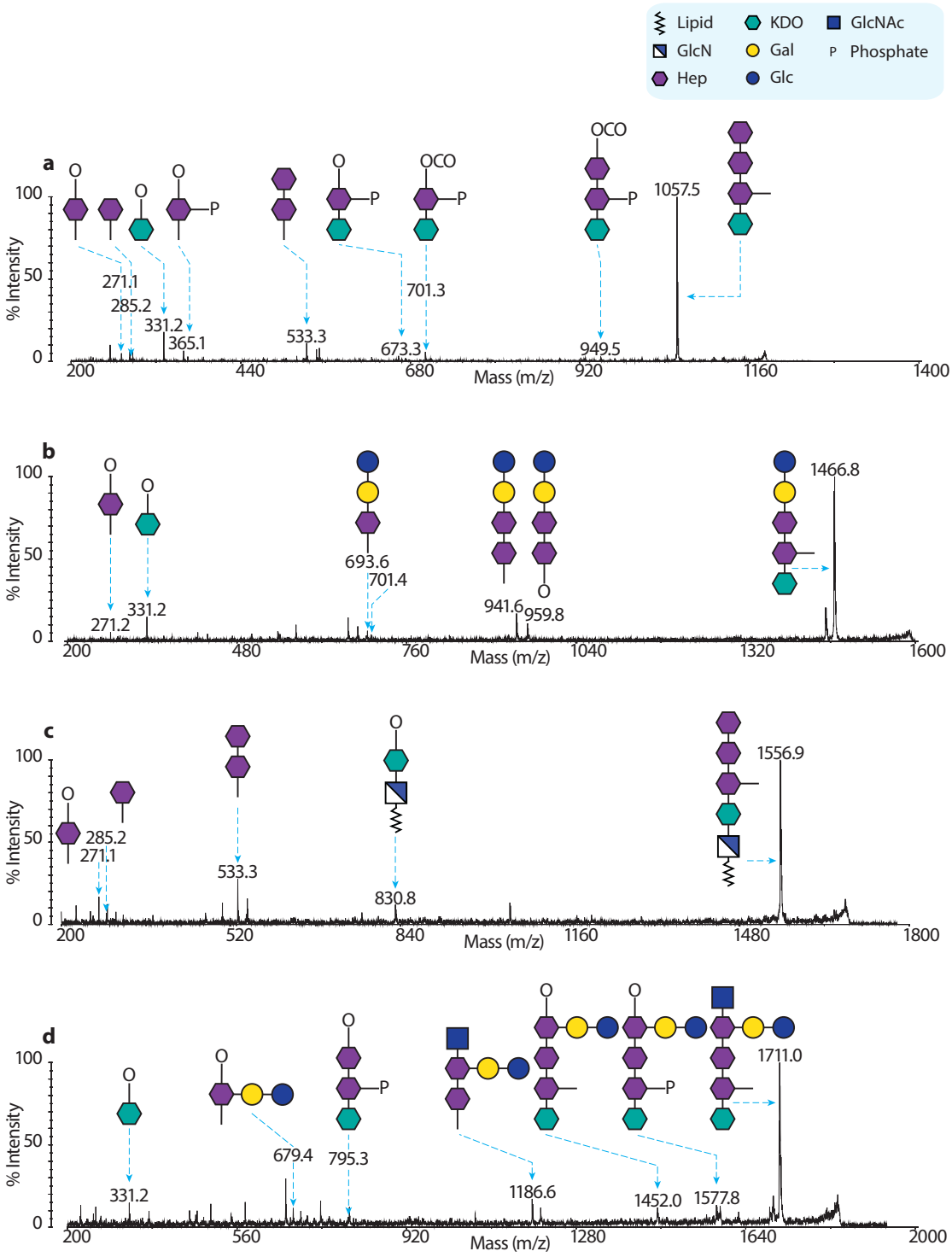


Figure 3 27 The MALDI-TOF/TOF spectra of the core oligosaccharide. MS/MS spectra of peaks at *m/z* 1183, 1591, 1682 and 1836 are shown in (a), (b), (c) and (d) respectively.

3.2.5.2 G27ΔHPG27_1230 and ΔHP0156 LPS

By adopting the same method, other mutant LPS samples were analysed. The MALDI-TOF spectra of the G27ΔHPG27_1230 and ΔHP0156 LPS are very similar to that of WT LPS (**FIG. 3-28**).

Because of the similarity between the WT and G27ΔHPG27_1230 and ΔHP0156 LPS samples, the MS spectra shown in **FIG. 3-28** were zoomed into a shorter mass range (1000-3500) so that the peaks could be better shown and to avoid repeatedly showing very similar data. The peaks attributed to the glycans are coloured red and annotated with *m/z* and structures. Briefly, the methanolysed products from all three components of LPS were found, strongly suggesting they are not truncated and there is no significant difference with the WT LPS. Interestingly, the peak at *m/z* 1416 corresponding to the tri-LacNAc structure is much stronger in the spectrum of G27ΔHPG27_1230 than the WT and ΔHP0156 LPS. The same phenomenon is also observed in the Smith degradation data (see **Section 3.2.7**), indicating it has a longer poly-lacNAc profile.

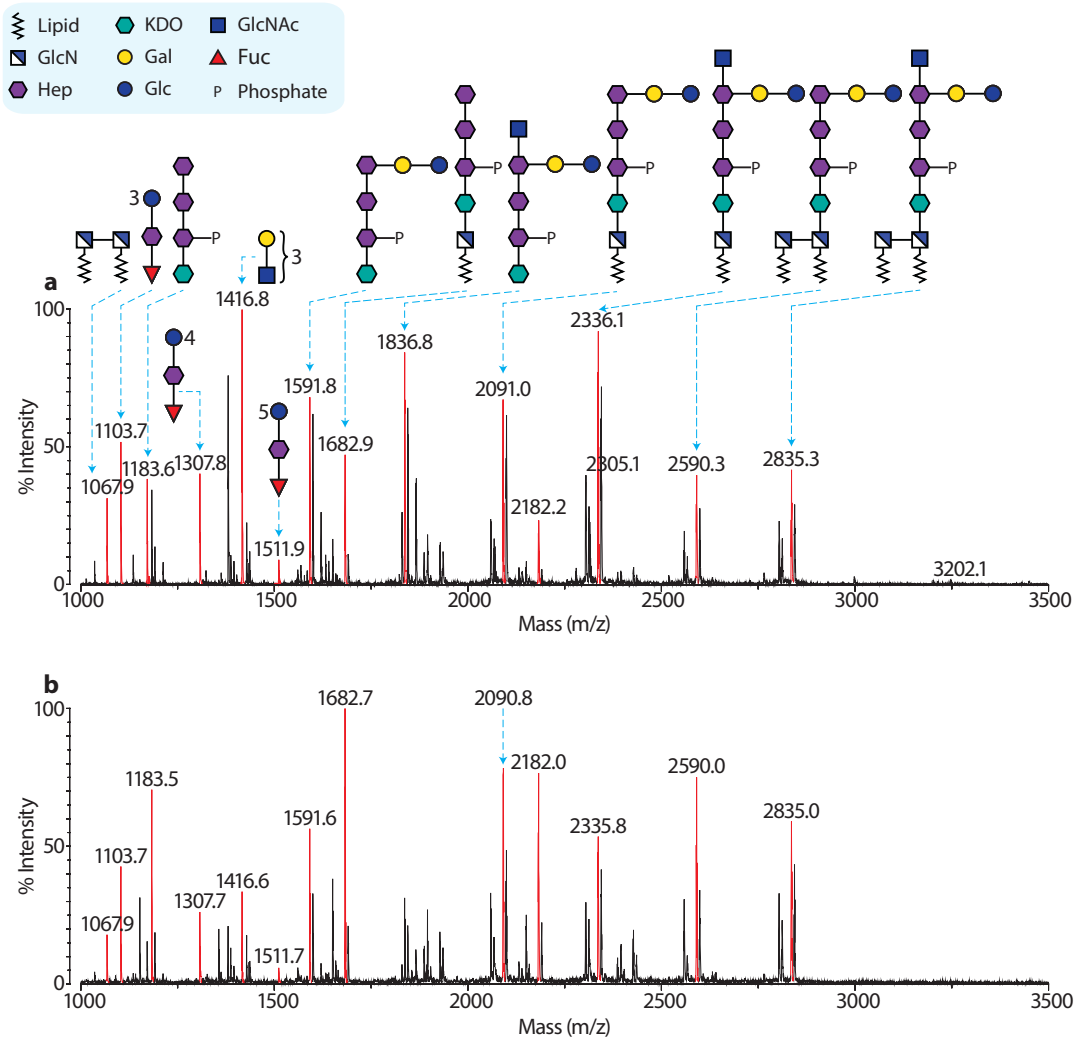


Figure 3-28 The MALDI-TOF spectra of the G27ΔHPG27_1230 (a) and ΔHP0156 (b) LPS. The LPS were subjected to methanolysis and permethylation before analysis. The spectrum (a) is fully annotated, all the peaks found in the spectrum (b) are included in (a), and they are therefore not annotated. More detailed information can be found in **FIG. 3-24**.

3.2.5.3 ΔWaaL, ΔHP0102 and ΔHP1578

The spectra of the ΔHP0102, ΔHP1578 and ΔWaaL LPSs showed no peaks corresponding to the poly-LacNAc and glucan, indicating they are seriously truncated, whereas the peaks corresponding to the core-oligosaccharide are clearly shown in these MALDI-TOF (**FIG. 3-29**). The spectrum of the ΔHP0102 LPS gives the best signals, therefore it was fully annotated with *m/z* and glycan structures. For the other two spectra (the ΔHP1578 and ΔWaaL LPS), only the peaks not present in the spectrum of the ΔHP0102 LPS were annotated.

In the MS spectrum of the ΔHP0102 LPS (**FIG. 3-29 a**), the successive addition pattern (see **FIG. 3-25**) was observed again, *e.g.* the peak at *m/z* 1183.2 shifted to the peaks at *m/z* 1690.3 and 2189.4 by adding one or two GlcN-lipids. Note the latter two glycans were under-permethylated, and therefore the masses were 8 Da higher than the permethylated structures (see **FIG. 3-25**, **Section 3.2.5.1**). The additions of a Glc-Gal, a GlcNAc and both of them to the peak at *m/z* 1183.2 resulted in the peaks at *m/z* 1591.2, 1436.2 and 1844.2. The same pattern existed for the peaks at *m/z* 1690.3 and 2189.4.

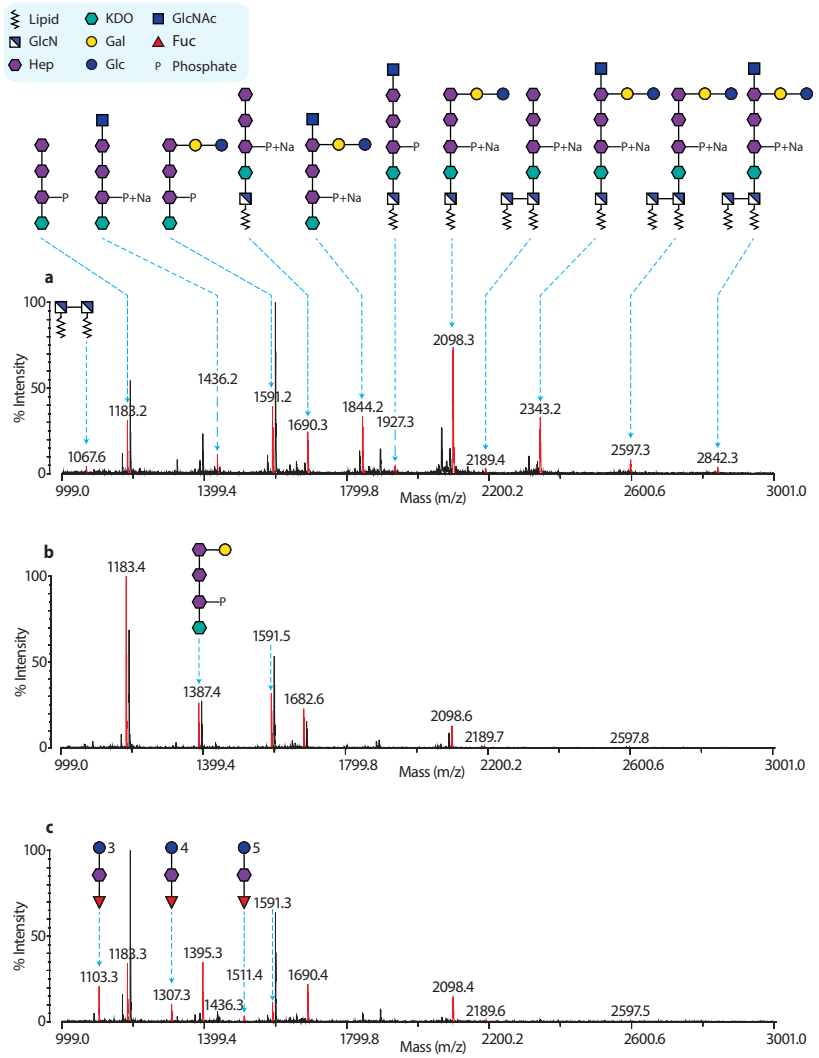


Figure 3-29 The MALDI-TOF spectra of the truncated LPS samples. The MS spectra of ΔHP0102 (a), ΔWaaL (b) and ΔHP1578 (c) LPS are shown. Spectrum (a) is fully annotated with *m/z* and glycan structures, and only the peaks not annotated in (a) are annotated with glycan structures in the remaining two spectra.

For the ΔWaaL LPS (**FIG. 3-29 b**), only the peak at *m/z* 2098.6 due to the addition of the Glc-Gal disaccharide was observed (no peak corresponding to the addition of a GlcNAc was found), which strongly indicates it is truncated from the Trio, *i.e.* the ΔWaaL strain only carries a hexa-saccharide with a sequence of Glc-Gal-tri-Hep-KDO.

In addition, the cluster of the peaks at *m/z* 1103, 1307 and 1511 from the glucan was absent in the spectra of ΔHP0102 and ΔWaaL LPS, also indicating that both of them carry seriously truncated LPSs. Notwithstanding the spectra of above two mutants, the cluster was found in the spectrum of the ΔHP1578 LPS (**FIG. 3-29 c**), implying it is probably longer than the other two LPSs.

The observation that the ΔWaaL LPS was truncated from the Trio is very interesting. WaaL is a conserved enzyme used for linking the fully synthesized O-antigen (see **Section 1.2.1.4**) to the core-oligosaccharide. Therefore, it is reasonable to propose that the O-antigen of *H. pylori* in fact starts with the Trio instead of the LacNAc units. This proposal is relevant to the structural analysis of ΔHP1284 LPS (see **Section 3.2.5.4**).

3.2.5.4 ΔHP1283 and ΔHP1284

The ΔHP1283 and ΔHP1284 LPS gave interesting MS spectra, *i.e.* the former sample carries a much longer glucan chain instead of glucan-heptan-lacNAc-Lewis structures as the O-antigen, and the latter one expresses an incompletely synthesised core-oliogosaccharide without affecting the O-antigen.

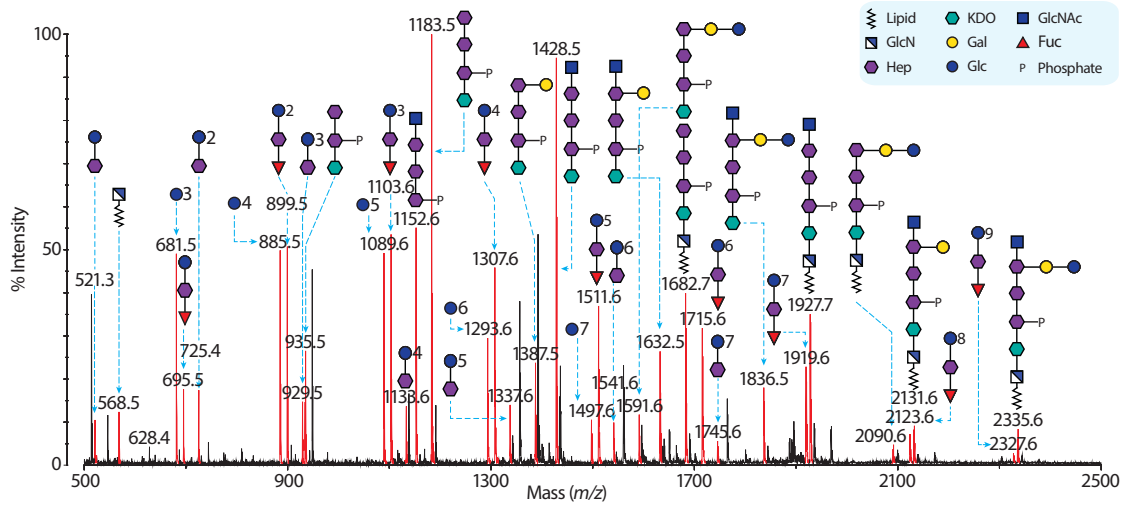


Figure 3-30 The MALDI-TOF spectrum of the ΔHP1283 LPS. The LPS was methanolysed and permethylated before analysis. The red peaks are corresponding to glycans and are annotated with glycan structures and *m/z*.

FIG. 3-30 shows the MALDI-TOF spectrum of the methanolysed products of the ΔHP1283 LPS. The signals from the core-oligosaccharide were exactly the same as the previously described samples, and no peak corresponding to the lacNAc was observed. However, this spectrum was characteristically dominated by clusters of peaks with an increment of 204 in *m/z*, which indicates it carries a longer glucan. More specifically, the cluster of peaks at *m/z* 521.3, 725.4, 929.5, 1133.6, 1337.6, 1541.6, 1745.6, 1949.6 and 2153.6 (1949.6 and 2153.6 were too small to be annotated) corresponded to the glucan-Hep; and the cluster of peaks at *m/z* 681.5, 885.5, 1089.6, 1293.6, 1497.6 and 1701.6 (1701.6 was too small to be annotated) corresponded to the glucan; and the cluster of peaks at *m/z* 695.5, 899.5, 1103.5, 1307.6, 1511.6, 1715.6, 1919.6, 2123.6 and 2327.6 corresponded to the glucan-Hep-Fuc. The longest observed glucan chain contained nine glucose units. However, as the sample was chemically degraded under a relatively harsh condition, it might carry a longer glucan chain. This observation is consistent with the NMR and GC-EI-MS analyses (see **Section 3.2.2**, **3.2.3.2** and **3.2.4**).

The most interesting sample is the ΔHP1284 LPS, as its spectrum showed a completely different pattern (**FIG. 3-31**) from all the other samples. It turned out that one of the Heps with the branched Gal-Glc in the core-oligosaccharide was absent, suggesting HP1284 is involved in the heptose transferring. According to previous research, only the GT of the outer-most Hep in the inner-core was not characterized yet, and it is therefore assumed that HP1284 is related to this Hep, to which the Gal-Glc disaccharide links.

The peaks at *m/z* 943.3 and 1188.3 (the under-permethylated versions of 935.2 and 1180.2, respectively) in the MS spectrum of ΔHP1284 LPS were further subjected to MALDI-TOF/TOF analysis (**FIG. 3-32**). They were chosen in order to avoid the dominant minus 126 Da peak, and the peak at *m/z* 1188.3 was a lot stronger than the one at *m/z* 1180.2. The peak at *m/z* 929 in **FIG. 3-32 b** convincingly proved a GlcNAc extending out of the diHep-KDO core.

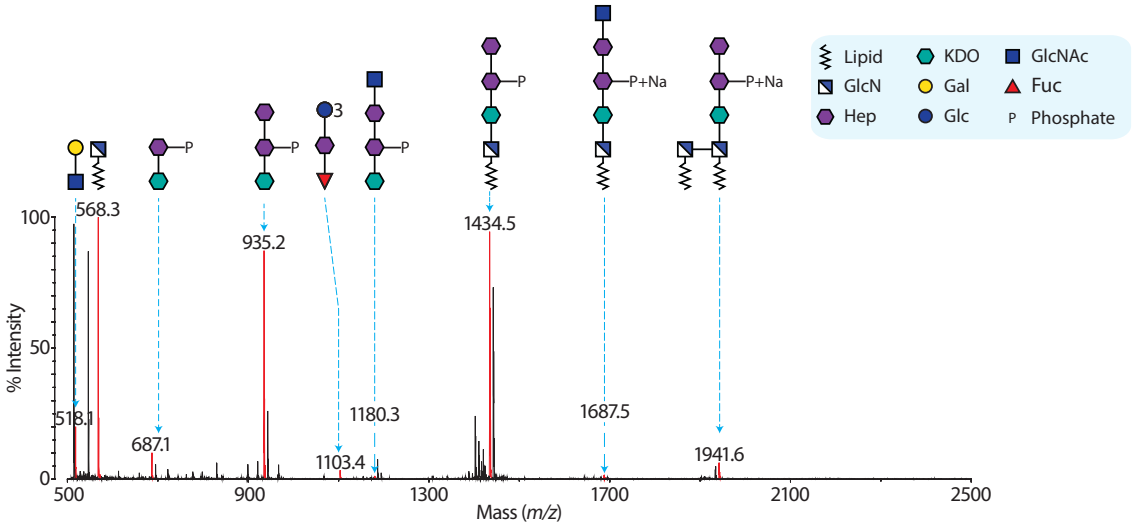


Figure 3 31 The MALDI-TOF spectrum of the ΔHP1284 LPS. The LPS was methanolysed and permethylated before analysis. The red peaks are corresponding to glycans and are annotated with glycan structures and *m/z*.

It is widely known that each GT has a specificity that transfers a glycan from a determined donor to a certain acceptor. When one of the GTs along the biosynthetic pathway of a glycan malfunctions, the next one will be no longer able to recognize the incorrectly synthesized glycan, and therefore cannot transfer the subsequent monosaccharide onto it. As a consequence, a mutation leading to an incompletely synthesised core-oligosaccharide normally completely prohibits further synthesis, which truncates the whole LPS. At this point, it is rather abnormal to find the ΔHP1284 LPS carrying all the remaining polysaccharide.

However, the ligase could have a relaxed specificity that recognizes the lipid carrier instead of the glycans (see **Section 1.2.1.4** and **1.2.2.1**). If we propose that the O-antigen begins with the Trio, and it is transferred intact to the incorrectly synthesized core-oligosaccharide by the ligase that has a relaxed specificity, we could explain why an aberrant inner-core does not affect the remaining LPS synthesis. Note that the ΔWaaL LPS data also support this proposal (see **Section 3.2.5.3**).

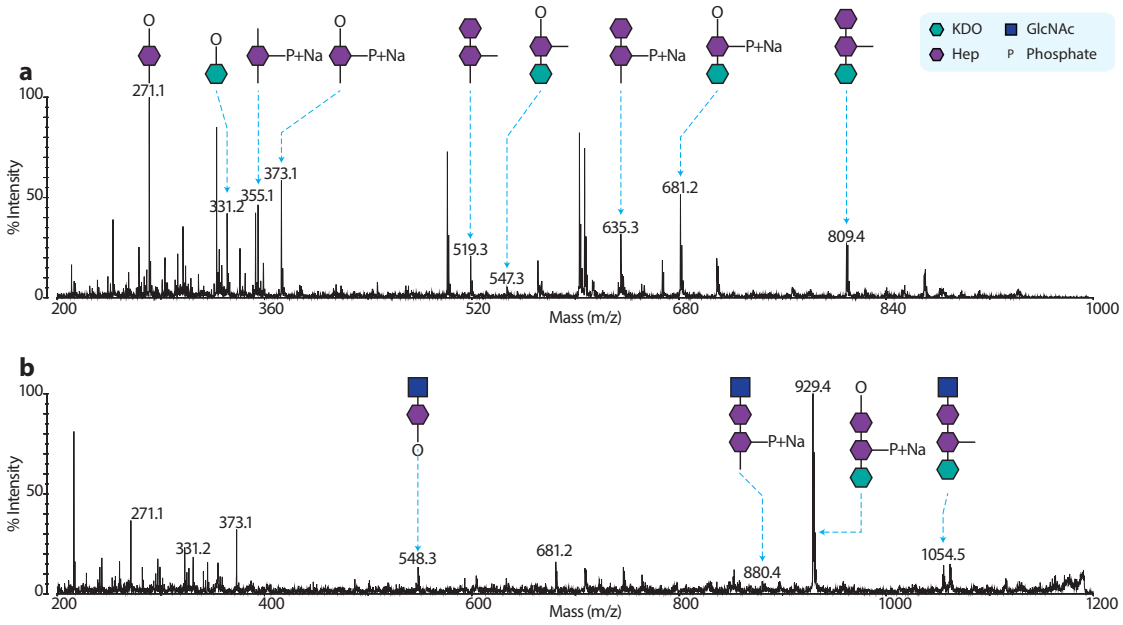


Figure 3-32 The MALDI-TOF/TOF analysis of the ΔHP1284 LPS. The MS/MS spectra of peaks at *m/z* 943.3 and 1188.3 are shown in (a) and (b) respectively. Spectrum (a) is fully annotated with *m/z* and glycan structures, and only the peaks not annotated in (a) are annotated with glycan structures in (b).

3.2.6 Mild HF Hydrolysis

The experiments described so far characterised the poly-lacNAc, glucan, core-oligosaccharide and lipid A parts of the *H. pylori* G27 LPS. Based on these data, it seems the structure of the G27 LPS is very similar, if not identical, to the previously published structure of the strain 26695 LPS. However, methanolysis analysis did not provide any information on the heptan that was found in the strain 26695 LPS, and previous research did not specify the number of its repeating units. In order to characterize the heptan, the WT, ΔHP1284 and ΔHP1578 were further analysed after mild HF hydrolysis that was proved to preferentially cleave the Fuc1-3GlcNAc glycosidic bond (Haslam *et al.* 1997; Yan *et al.* 2012).

The MALDI-TOF analysis of the HF hydrolysed WT G27 LPS (**FIG. 3-33 a**) revealed that it contained a rather large heptan (its *m/z* is more than 7 KDa). The observed highest peak at *m/z* 7221.7 corresponded to a Hep23-Glc5-Hep-Fuc structure, indicating the heptose in the WT LPS repeats at least 23 times. This data further confirmed that the glucan contained five glucose units. The MS spectrum of the HF hydrolysed ΔHP1284 LPS showed a very similar pattern to the MS spectrum of the WT LPS (**FIG. 3-33 b**), supporting that the G27 ΔHP1284 strain carries the WT O-antigen. The presumably intact heptan-glucan-Hep-Fuc structure was observed in both spectra, indicating the heptan and glycan are rather resistant to the HF hydrolysis.

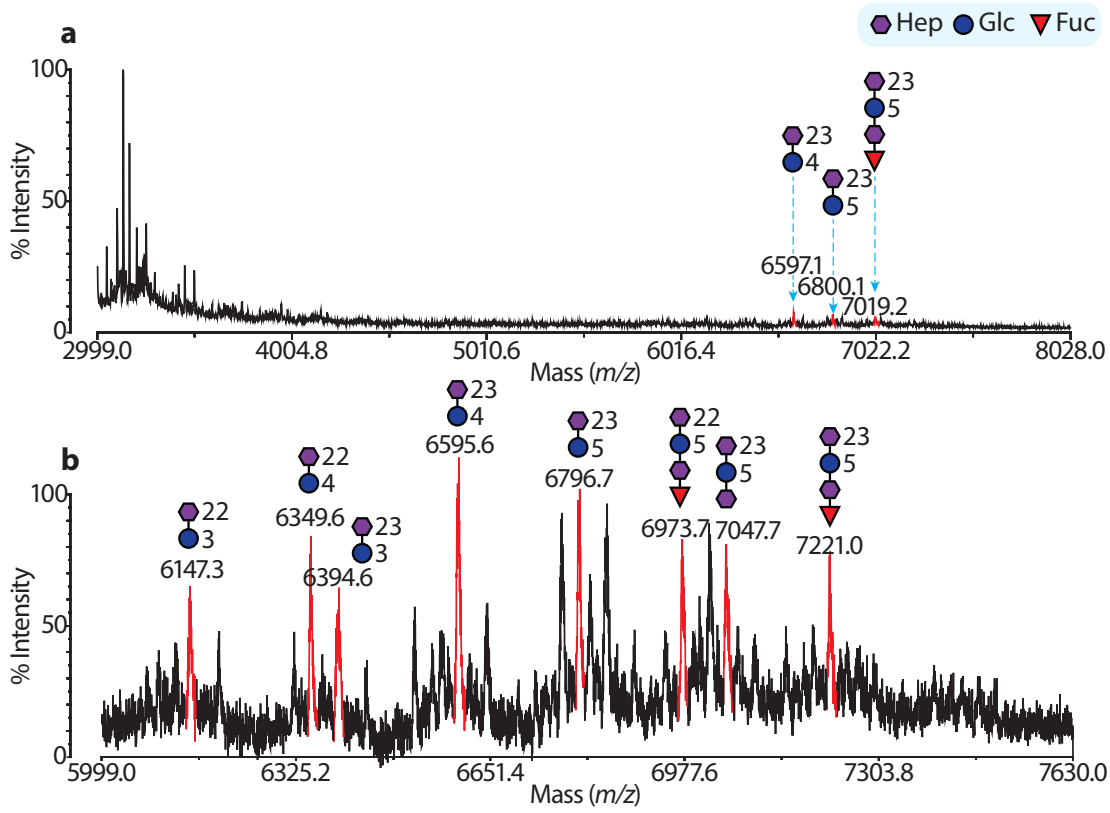


Figure 3-33 The MALDI-TOF spectra of the permethylated G27 WT (a) and ΔHP1284 (b) LPS after mild HF hydrolysis. The red peaks are corresponding to glycans and are annotated with glycan structures and averaged *m/z*.

The ΔHP1578 LPS was a lot more complicated than the WT (**FIG. 3-34**). The cluster of peaks dominating the spectrum shown in **FIG. 3-34 a** corresponded to the heptan-Glc5-Hep-Fuc with the heptose numbers ranging from 6 to 24. As the data of the WT and ΔHP1284 LPS suggested the resistance of the heptan to HF hydrolysis, the ΔHP1578 LPS sample actually contained a mixture of the LPS with different lengths of heptan. Notably, the biggest structure observed contained a heptan with 24 repeating units, which was slightly larger than the heptan in G27 WT LPS. This indicates that the addition of the lacNAcs functions as a regulator terminating the elongation of the heptan. A similar situation was observed for the ΔHP1283 LPS, where the missing heptan leads to a much longer glucan. However, it is highly possible that other regulation mechanisms are involved in determining the length of the heptan as it is only slightly longer when the lacNAcs are mutated.

In order to further confirm the proposed annotation for the MS spectra, the peaks at *m/z* 3000.7 and 3248.8 (**FIG. 3-34 a**) were chosen for MS/MS analysis (**FIG. 3-35**), as they were easier to get fragmented, and were representative of the observed peak cluster (from *m/z* 3000.7, 3248.8 to 7466.3). In **FIG. 3-35**, a peak at *m/z* 1498 attributed to a single-cleaved glycan with a sequence of Glc5-Hep-Fuc was found in both MS/MS spectra. The peak cluster with *m/z* lower than this peak showed an interval of 204 Da, which corresponds to a glucose; while the cluster with *m/z* higher than it showed an interval of 248 Da, which corresponds to a heptose. This strongly indicated that the peak cluster observed in **FIG. 3-34 a** contains a 5-repeating-unit glucan and different lengths of heptose.

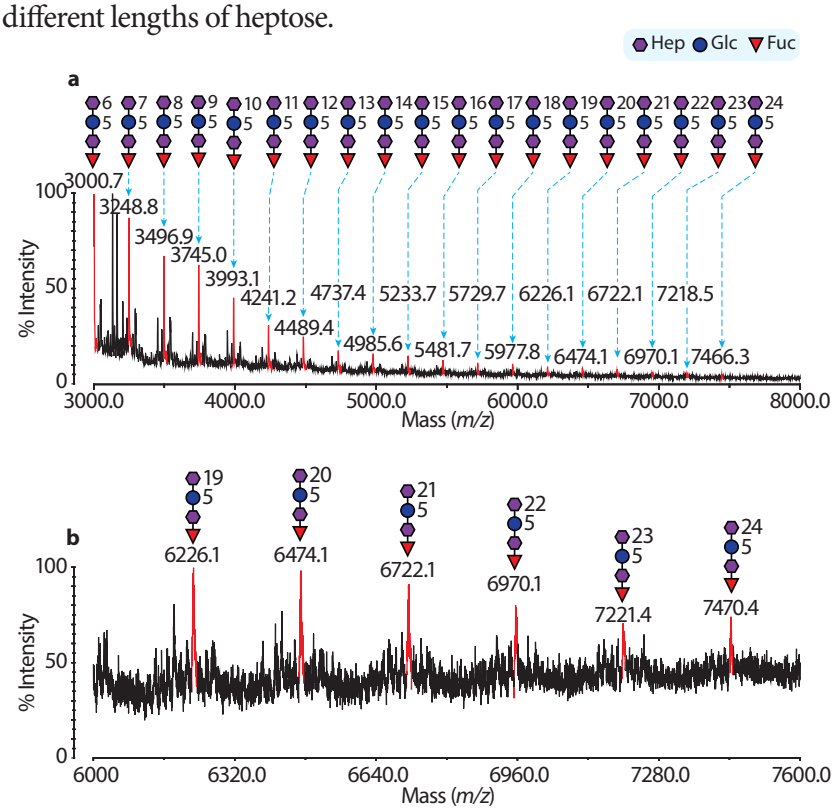


Figure 3-34 The MALDI-TOF spectra of the permethylated ΔHP1578 LPS after mild HF hydrolysis. The spectrum is shown in two different mass ranges, 3000.0-8000.0 (a) and 6000.0-7600.0 (b). The red peaks are corresponding to glycans and are annotated with glycan structures. In addition, for (a), the peaks are annotated with theoretical *m/z*, and for (b), they are annotated with observed averaged *m/z*.

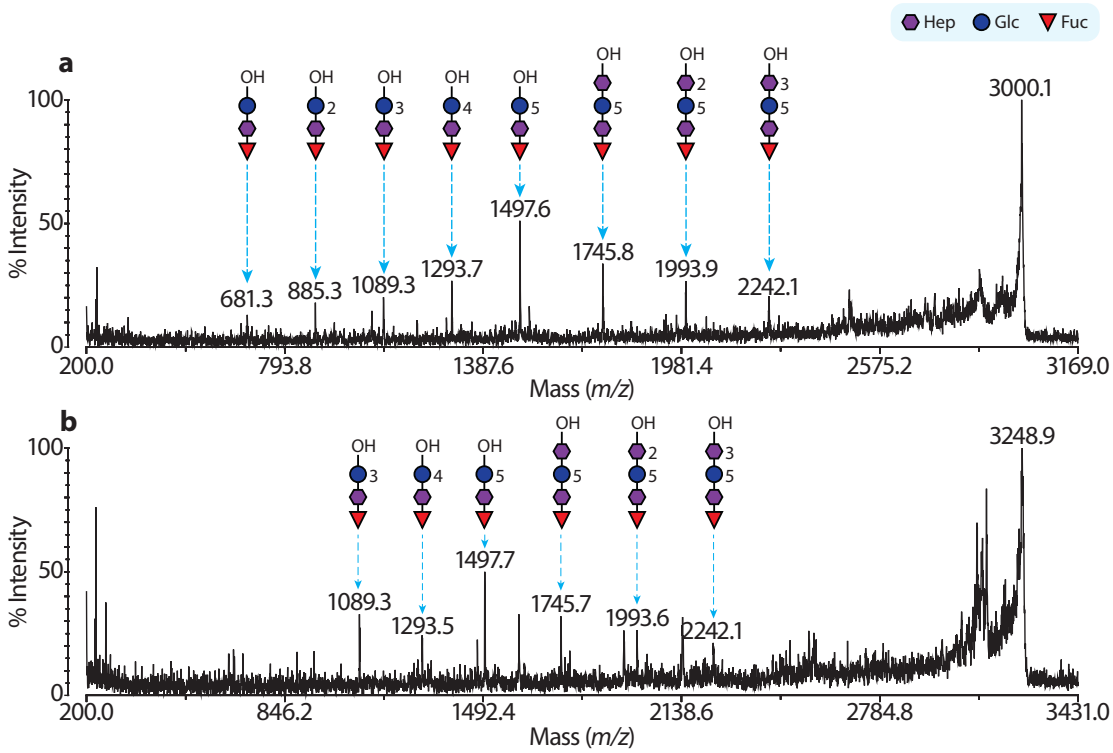


Figure 3-35 MALDI-TOF/TOF analysis of the permethylated ΔHP1578 LPS after mild HF hydrolysis. The MS/MS spectra of the peak at *m/z* 3000.7 and 3248.8 are shown in (a) and (b) respectively.

3.2.7 Smith Degradation

In order to further characterize the poly-lacNAc structures of the G27 LPSs, they were oxidized by sodium periodate ($NaIO_4$) under mild conditions and reduced by sodium borohydride ($NaBH_4$). During the process, glycans carrying neighbouring hydroxyl groups were oxidized and degraded, whereas the ones without them were protected from degradation. The mechanism of the Smith degradation is shown in **FIG. 3-36**.

Specifically, when exploiting the reaction to analyse the *H. pylori* strain G27 LPSs; the poly-lacNAc structures were kept, while most other glycans were degraded. For instance, if the O-antigen was capped by a Lewis Y structure, both terminal Fucs were degraded, leaving a LacNAc which gave rise to a signal at *m/z* 534; if the O-antigen was capped by a Lewis X structure (or merely LacNAc), the terminal Gal and Fuc were degraded, giving rise to a peak at *m/z* 779 corresponding to a GlcNAc-lacNAc structure. Therefore, this is an effective method to characterize the length of the lacNAc structures as they are kept nearly intact during the reaction, which provides highly predictable MS spectra.

The ΔHP1283, ΔHP1578, ΔHP0102 and ΔWaaL LPSs showed no peaks corresponding to Lewis O-antigens (data not shown in the main text, see **FIG. A2**), along with previous Western blotting data suggesting they do not carry Lewis O-antigens. Only the MS spectra of mild oxidized LPS from G27 WT and ΔHPG27_HP1230, ΔHP0156 and ΔHP1284 are shown (**FIG. 3-37 b**, **a**, **c** and **d**). The peaks corresponding to the sodiated lacNAc and poly-LacNAc structure are coloured red and annotated with their *m/z* values. Because the spectrum of ΔHPG27_HP1230 LPS showed the strongest peaks and all spectra showed similar poly-LacNAc profiles, only this spectrum was annotated with glycan structures. The data showed that its LPS O-antigens are poly-lacNAc mixtures containing one to eight repeating units. As expected, the lacNAcs were extended from both *m/z* 534.3 that is indicative of the Lewis Y epitope and *m/z* 779.5 that is indicative of the Lewis X epitope, confirming the existence of both Lewis structures. The maximum length of the G27ΔHPG27_1230 LPS is similar to that of the ΔHP0156 LPS, whereas the longest poly-lacNAcs only contained seven or six repeating units in both WT and ΔHP1284 LPSs. As the

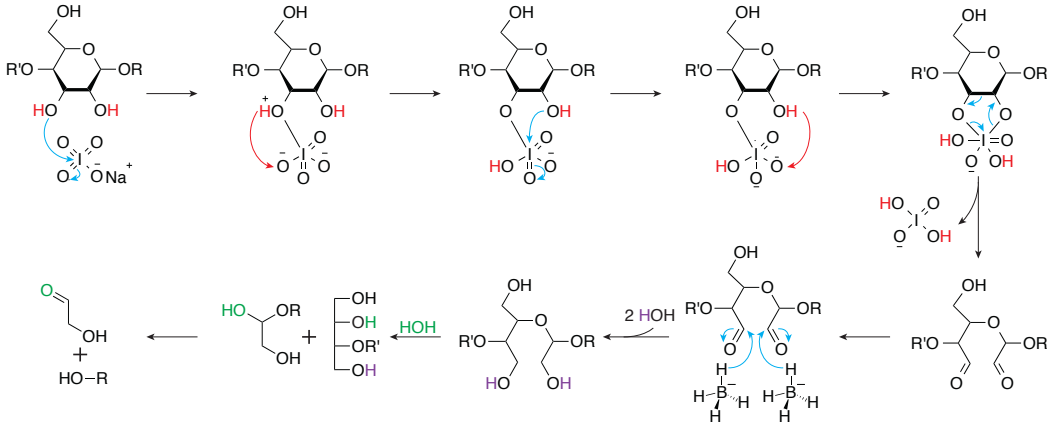


Figure 3-36 Mechanism of the smith degradation. Red arrows stand for proton transfer, and blue arrows for electron pair transfer.

peaks corresponding to the longer O-antigens had very small intensities, most of them were not annotated in the spectra.

Intriguingly, the G27ΔHPG27_1230 LPS showed a profile of longer O-antigens, *i.e.* the peak at *m/z* 1667.9 corresponding to four lacNAc units dominated its spectrum, while in all other samples, the intensity of that peak was similar to a peak at *m/z* 1228.9 which is from three LacNAc units.

The proposed poly-LacNAc structure was further confirmed by MALDI-TOF/TOF. The MS/MS spectra of the peak at *m/z* 779.5 and *m/z* 1228.7 are shown (**FIG. 3-38**). Both spectra were dominated by a peak at *m/z* 520.3, which corresponds to a single-cleavage lacNAc structure. By comparing the two spectra, it was concluded that the peak at *m/z* 1228.9 corresponds to a LacNAc unit linearly extended on the structure at *m/z* 779.3, indicating that all the poly-LacNAc structures are linear.

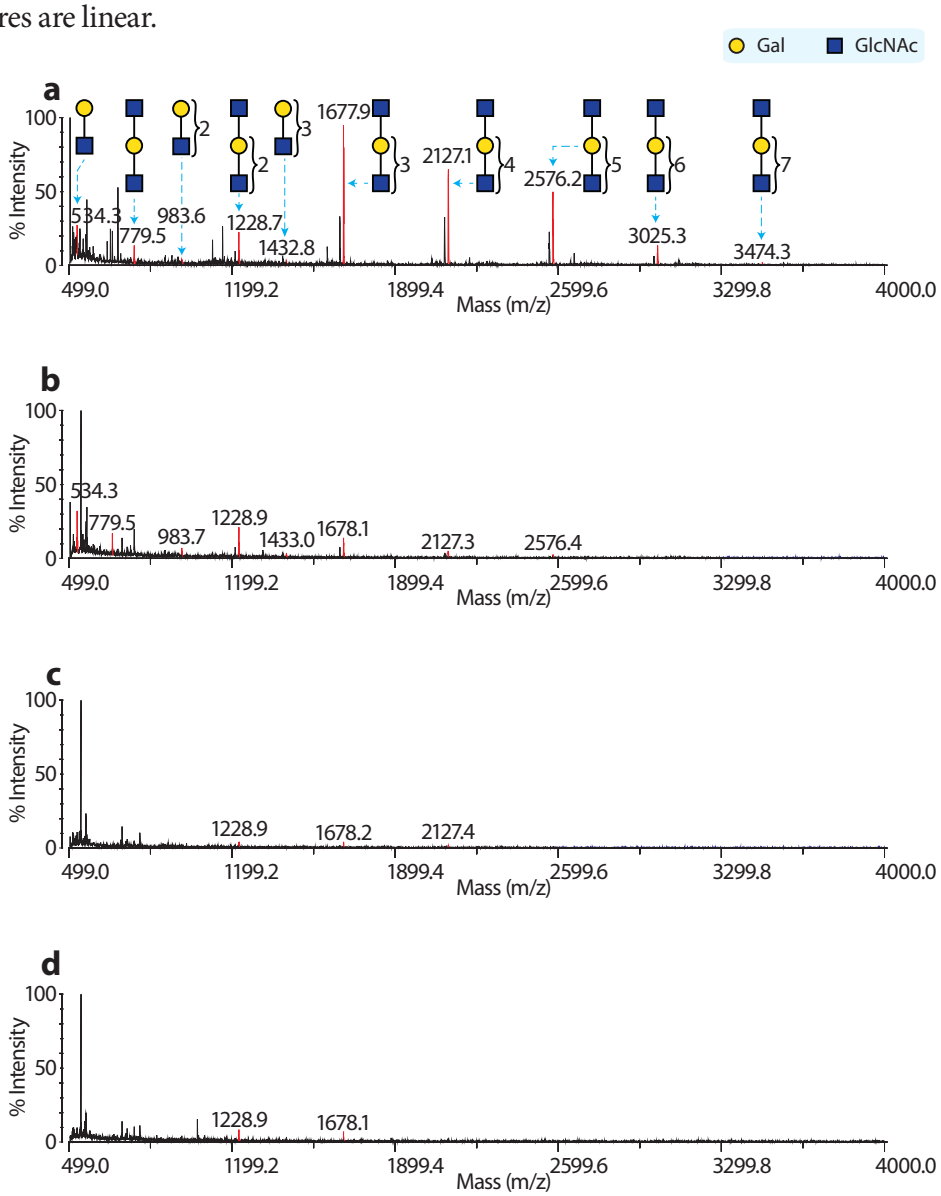


Figure 3-37 The MALDI-TOF of ΔHPG27_1230 (a), WT (b), ΔHP0156 (c) and ΔHP1284 (d) LPS after Smith Degradation. As the spectrum (a) gives best signals, only this spectrum is annotated with *m/z* and structures. The peaks in all other spectra are only annotated with *m/z*.

3.2.8 Discussion and Conclusions

The differences in the chemical stability of different glycosidic bonds provide a philosophical perspective that any single glycan structure can be elucidated however complex it is. The LPS is no exception, though its structural elucidation is a highly skilled work. The analysis of the LPS samples extracted from WT *H. pylori* G27 and its mutants (ΔHP0156, G27ΔHPG27_1230, ΔHP1284, ΔHP1283, ΔHP1578, ΔHP0102 and ΔWaaL) relied on various chemical, MS and NMR strategies (see **Section 3.0**). Each of these methods has its own advantages and disadvantages, and therefore their combination is essential to completely solve a LPS structure.

Mild acid hydrolysis is a classical reaction mostly chosen for the lipid A investigations. Its popularity relies on it preferentially cleaving the lipid A from the LPS (**FIG. 3-4**). This specificity is based on the instability of the KDO-GlcNAc glycosidic bond, which means it is not a universal reaction functioning for all LPSs. For instances, some LPS contains D-glycero-a-D-talo-oct-2-ulopyranosylonate (KO) may resist the reaction (Isshiki *et al*. 2003). In addition, the yield of the reaction is rather low, and normally, a 1-10 mg level sample is required (Manzo *et al*. 2001).

Mild HF hydrolysis and Smith degradation are another two well established chemistries used for "fragmenting" glycans. The mechanism of both reactions were fully explained (see **Section 3.2.6** and **3.2.7**), and therefore their products are highly predictable when some structural information is available. However, these reactions especially the Smith degradation alone are difficult to be used in *de novo* structural elucidation. In addition, these two reactions require poisonous reagents and relatively complicated experimental operations.

Methanolysis is a chemically simple reaction, and has been a step in the protocol of the GC-EI-MS TMS analysis. To my knowledge, this is the first time it is used to in-completely hydrolysis the LPS. The overall result of the methanolysis is rather similar to the mild HF hydrolysis (only high *m/z* ends of the mild HF hydrolysis spectra are shown in **Section 3.2.6**). It seems that the former reaction is slightly stronger, as the latter reaction kept the heptan structures, which means some important structure can be destroyed during the reaction. The great advantage of the methanolysis is its simplicity and sensitivity. After optimisation, the experiments described in **Section 3.2.3** can be finished within 10 hrs and they only require 10-100 μg level starting materials.

Both GC methods (see **Section 3.2.2** and **3.2.4**) were intensively researched decades ago, and are routinely performed ever since. As MS is used as a detector, they are both very sensitive. Generally speaking, the TMS analysis is more sensitive than the linkage analysis, probably due to the simplicity of the former method. The linkage analysis is more diagnostic, because the MS fragmentation patterns of the PMAAs are highly characteristic. The major issues of both methods is the strict dependency on the monosaccharide standards that are almost for sure not commercially available for most LPSs. Therefore, the chemical and biological synthetic

methods for monosaccharides are in a high demand. Presumably, their advancements can greatly facilitate the structural elucidation of LPSs.

The micelle NMR strategies (see **Section 3.2.3**) were initially used for characterising the gangliosides (Poppe *et al.* 1994) and rough-mode LPS (Zähringer *et al.* 2014). We adopted the same method to analyse the intact LPS molecules, which provides valuable information on the anomeric configurations of the monosaccharides. As mentioned, at least 5-10 mg samples are needed for the micelle NMR experiments. In principle, these materials can be re-used for the chemical degradation, such as methanolysis, and the MS characterisations. The resolution of the NMR spectra is significantly affected by the size and homogeneity of the LPS sample. Though decent NMR spectra were successfully obtained, a purification and further optimisation are likely to significantly increase the quality of the NMR spectra.

MS has emerged as a not only powerful but also ultimate tool to sequence the glycans, and it plays a central role in analysing the LPSs. Two MALDI ion sources were used to ionise the permethylated glycans cleaved from the LPS samples. Attempts were made to record the MALDI-TOF spectra of the HP0102 mutant LPS, and they turned out to be successful eventually (data not shown). As the HP0102 mutant LPS is in a rough-mode, an investigation of the smooth mode LPSs could be very interesting. In addition, LC-ESI-MS could be better than MALDI-TOF in terms of characterising intact LPSs, on which further research efforts should focus. As almost all LPSs contain repeating units, a method measuring the molecular weight of the intact LPSs should be highlighted.

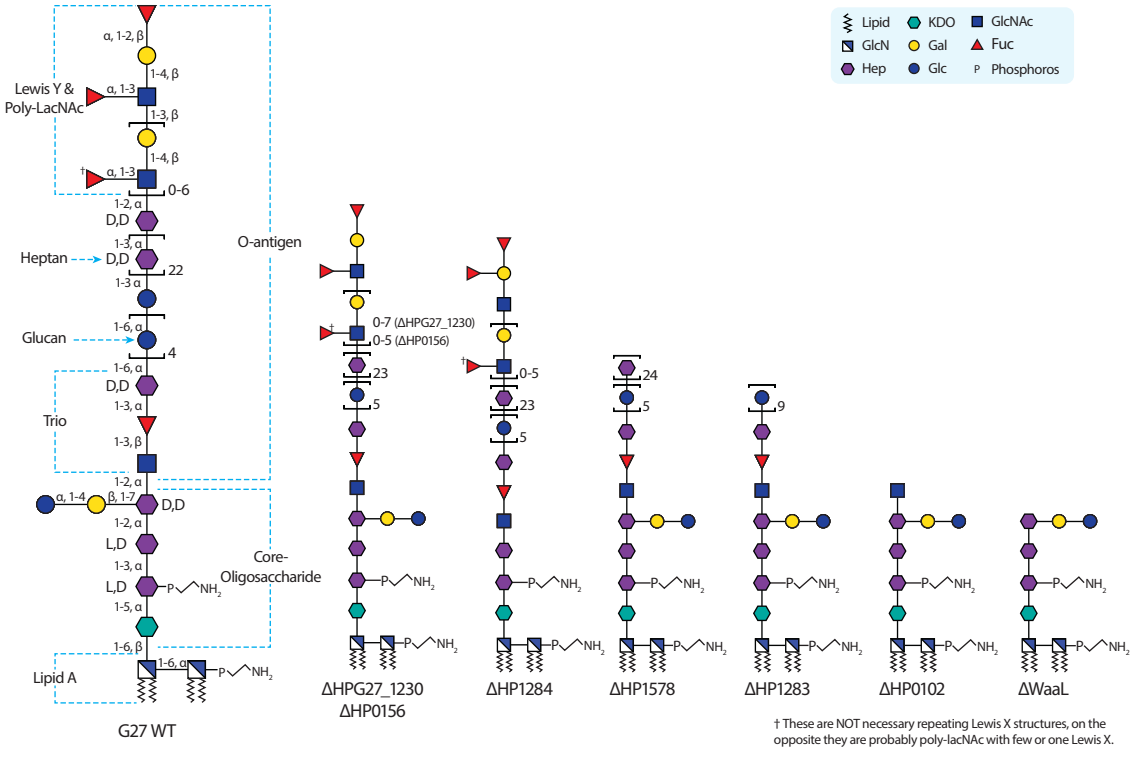


Figure 3-38 Diagram of the proposed LPS structures

Overall, our data consistently supported several structural conclusions (**FIG. 3-38**). The structure of strain G27 WT LPS is very similar to that of strain 26695 LPS. The latter structure was very well characterised by in-depth NMR analysis (Altman, Chandan, Li & E. Vinogradov 2011b). The *H. pylori* strain 26695 and HP0826::Kan LPS were completely deacylated and deaminated, and the N-deamination reaction transformed GlcNAc into anhydromannose, simultaneously cleaving the glycan chains. The glycan fragments were charecterised by NMR and capillary electrophoresis-MS (CE-MS). The CE-MS experiments on the HP0826 mutant LPS successfully evaluated the length of its glucan to be 11-12 glucose units, though very little information was collected for the heptan and the glycans were not fully sequenced. The researchers claimed a similar result was obtained for the WT strain 26695 LPS, though the length of neither the heptan nor the glucan was specified. However, it seems that the HP0826 mutant LPS lacks the Lewis O-antigens and carries an extended glucan.

Therefore, our data on the WT strain G27 LPS supplement previous research in terms of characterizing the lengths of the poly-lacNAc/Lewis structures, heptan and glucan. The MS spectra recorded after mild HF hydrolysis revealed that the glucan and heptan contains no less than 5 and 23 repeating units, respectively. The length of the glucan is similar to the previously reported *H. pylori* serogroup O:3 LPS (Altman *et al.* 2013). Smith degradation provided information on the length of the poly-lacNAc backbones, though their fucosylations were removed. Further analysis could focus on identifying the number of the Lewix X epitopes contained in the O-antigen.

The ΔHP0156 LPS is very similar to the G27 WT LPS. It seems that the mutation of HP0156 does not have any significant effect on the LPS structure. The G27ΔHPG27_1230 LPS carries a longer poly-LacNAc profile. All of the samples including the WT, ΔHP0156, G27ΔHPG27_1230, ΔHP1284 LPS are mixtures carrying poly-lacNAc/Lewis antigens with different numbers of the repeating unit. However, a majority of the G27ΔHPG27_1230 LPS normally carries poly-lacNAc with 4 to 6 units, whereas others normally carry poly-lacNAc with 3 to 4 units. This indicates the function of G27ΔHPG27_1230 may be related to the regulation of the α-1,2-fucosyltransferase that caps poly-lacNAc elongation.

The ΔHP0102 LPS lacks all the structures from the Trio except for the GlcNAc in the Trio region, which could indicate HP0102 is the FucT that transfers the fucose in the Trio. The ΔWaaL LPS are most seriously truncated, *i.e.* ΔWaaL lacks all the structures from the Trio. The data suggest *H. pylori* O-antigen is considerably longer than previously proposed and that the ligase is able to transfer a very large O-antigen to a short core-oligosaccharide. The ΔHP1284 LPS lacks the outer-most Hep and the Glc-Gal disaccharide attaching on it without affecting the remaining polysaccharide structures. This indicates that HP1284 could be the HepT (HepT III) that transfers the outer-most heptose in the core-oligosaccharide. In addition, this observation also supports the function of the ligase.

The ΔHP1578 LPS lacks the poly-LacNAc/Lewis structures; instead it carries heptans with different length. This indicates that HP1578 could be the GlcNAcT that transfers the GlcNAc to the heptan, preventing the heptan from further elongation. The ΔHP1283 LPS carries a longer glucan and lacks the poly-lacNAc/Lewis structures and heptan. This indicates that the HP1283 could be the HepT that transfers the heptose to the glucan, preventing its further synthesis. The elongated glucan has been observed before (Altman *et al*. 2012). The *H. pylori* strain 26695 HP0826 mutant LPS carrying a longer glucan but lacking the Lewis O-antigens was chemically conjugated to the tetanus toxoid or bovine serum albumin carrier proteins. The produced glycoproteins were proved to be promising candidates for the novel glycoconjugates vaccines against *H. pylori*. Because the O-antigen was redefined based on the WaaL mutant LPS, the conjugation of glucan to carrier proteins can be completed in genetically modified *E. coli* systems (see **Section 5.1**) in principle, which could develop into new projects.

The proposed GT functions above can be further supported by two considerations. First, the genomic analysis performed by our collaborators suggested that the genes encoded GTs. Secondly, it seems the mutations do not prohibit the metabolism of monosaccharides. For example, the mutation of HP1578 did not affect the GlcNAc in the Trio, which means it is not essential for the GlcNAc biosynthesis. Therefore, it is safe to conclude that HP1578 is essential for the GlcNAcT activity, based on the structural alteration caused by its mutation. Nevertheless, the GT functions above have not been confirmed by biochemical assays, and the structural analysis of the mutant LPS solely cannot determine the functions of mutated genes. It is worth mentioning that it can be extremely difficult to obtain donor and especially acceptor glycans structures for the biochemical assay used for confirming bacterial GT activities, which again helps emphasize the importance of the synthetic strategies for glycan and glycoconjugate standards.

The structural analysis helps better define the LPS especially the O-antigen biosynthesis in *H. pylori*. The *H. pylori* O-antigen is probably synthesised via the ABC transporter-dependent pathway. An UndPP-GlcNAc primer is first assembled at the cytoplasmic space of the inner membrane. A Fuc and a Hep are subsequently transferred to the primer, generating the Trio that acts as the adaptor between the core-oligosaccharide and O-antigen. A glucan, a heptan and poly-lacNAc/lewis structures are successively synthesised on the Trio, forming the full O-antigen. The regulation mechanism of the O-antigen length remains elusive in the ABC transporter-dependent pathway, on which our data might shed light. It seems that the homopolymeric O-antigen keeps elongating until a GT caps it with a different monosaccharide, which leads to either a varied homopolymeric elongation or the termination of O-antigen synthesis. It is still unclear though how the incorporation of the "capping GT" is modulated. The completely synthesised O-antigen is transferred to the short core-oligosaccharide by the WaaL ligase. As the structural alteration caused by the mutation of HP1284 in the core-region does not affect the function of the ligase, it probably has a relaxed specificity.

LPSs without the fully synthesised core-oligosaccharides but carrying O-antigens were have been observed in multiple organisms (Loutet *et al.* 2006; Skurnik *et al.* 1999; Ramjeet *et al.* 2005). Our data further confirmed the relaxed specificity of the WaaL ligase. Because the ligase shares the same Wzy_C domain with the bacterial OTase that can recognize multiple donors and acceptors, it is not unexpected that the WaaL ligase can share the relaxed specificity. The detailed mechanisms of the reactions catalyzed by both enzymes are unclear, and a 3D structure of the Wzy_C domain is essential to explain their functions.

It is interesting that our collaborators found mutation of HP0102, HP1283 or HP1284 leads to a complete loss of the colonization ability. This could be further characterised by immunohistochemical experiments to identify whether the inability to colonize is because of a reduced adhesion and/or the stimulation of immune responses. The HP0102 and HP1283 mutant LPSs do not carry the Lewis epitopes that were proved to be important for both of the adhesion and immune suppression of *H. pylori*, which could explain the loss of colonization ability. Though the main adhesions of *H. pylori* are babA and sialic acid–binding adhesin (SabA) (Boren *et al.* 1993; Ilver *et al.* 1998; Mahdavi 2002; D. J. Evans & D. G. Evans 2000), its Lewis-type O-antigen was demonstrated to play a limited role as an "adhesin" (Lozniewski *et al.* 2003; Mahdavi *et al.* 2003). In addition, the fucosylated Lewis-type epitopes can promote the expression of the immune-suppressing cytokines via the DC-SIGN signaling pathway (Gringhuis *et al.* 2009). Therefore, it is highly possible that the interruption of the Lewis epitopes in the HP0102 and HP1283 mutant LPSs leads to the drastic reduction in the colonization due to a combination of the adhesion and immunological reasons and subsequent works should concentrate on the biological assays.

The HP1284 mutant LPS only misses a tri-saccharide motif in the core-oligosaccharide, and it is difficult to fully explain the phenomenon based on current data. The observations that the LPS core-oligosaccharides are required for the bacterial virulence are not unprecedented at all (Skurnik *et al.* 1999; Ramjeet *et al.* 2005; Loutet *et al.* 2006), which could partially answer the question. In all of these studies, gel data were used to evaluate the overall length of the LPS and the antimicrobial peptides/serum assays were done to investigate the biological implications of the core-oligosaccharides. They consistently showed that the core-oligosaccharide is involved in the host immunity and/or antimicrobial peptides/serum resistance, and is therefore associated with the bacterial virulence. In addition, several *E. coli* adhesins were proved to be heptosylated (Moormann 2002; Lindenthal & Elsinghorst 1999; Sherlock *et al.* 2006), and the glycosylation of the Ag43 protein was suggested to mediate the bacterial adhesion processes. Therefore, it is worth investigating the possibility whether HP1284 modify *H. pylori* proteins as a HepT.

As a conclusion beyond structural glycobiology, the remaining questions of current *H. pylori* LPS research address the idea that the elucidation of biological problems can only rely on a combination of physical, chemical, biochemical, and biological approaches, and the structural analysis alone is far less sufficient to answer any biological questions.

Chapter 4 Glycomics and Glycoproteomics of Fimbriae-Associated Protein

This chapter describes MS-based glycomic and glycoproteomic analysis of serine rich fimbriae-associated protein 1 (Fap1). The structures of O-glycans carried by Fap1 were in-depth analysed, and the glycosylation sites of recombinant Fap1 were partially mapped. Analytical strategies included in this chapter can *de novo* determine the structures of O-glycans from bacterial glycoproteins with high sensitivity.

4.1Background

Fimbriae-Associated Protein 1 (Fap1) is the main adhesin of Gram-positive *Streptococcus parasanguinis* (*S. parasanguinis*, (Becker *et al.* 2002). Its main niche is saliva-covered surfaces of human teeth. The adhesion of *S. parasanguinis* along with other oral *Streptococci* acts as the foundation to which other oral bacteria can attach, eventually forming dental plaque (**FIG. 4-1**) (Froeliger & Fives-Taylor 2001; Garnett *et al.* 2012). In addition, it can also infect human heart valves, which leads to infective endocarditis (Que & Moreillon 2011).

The adhesive ability is the key to the "pathogenesis" of the organism and is mediated by its surface adhesins such as Fap1. Fap1 was proved to be essential for fimbriae biogenesis, bacterial adhesion and dental plaque formation (Wu *et al.* 1998). For instance, previous experiments demonstrated that a Fap1 mutant shows no fimbriae and reduced ability in adhering (Wu *et al.* 1998; Fives-Taylor & Thompson 1985).

Fap1 was the earliest characterized serine-rich protein. According to genomic analysis, Fap1 contains 2570 amino acid residues in total (**FIG. 4-2**). These amino acids form a long protein with a signal sequence (SS, yellow) at the N-terminus and a classic cell wall anchor (CW, yellow) at the C-terminus. Between these two terminal domains, there are four alternative non-repeat (NRI and NRII, green) and repeat (RI and RII, red) domains. The NRI domain is next to the signal sequence and the RII domain is adjacent to the cell wall anchor. Notably, the RII is a very long domain that has more than 2000 amino acids, and around 50% of them are serine. Serine-rich proteins sharing a similar structure (alternative non-repeating and repeating domains) can also be found in other *Streptococci* as well as *Staphylococci*.

Within the serine-rich protein family, Fap1 is the first one that is proven to be glycosylated (Stephenson *et al.* 2002). Previous structural research showed that the glycans attaching on Fap1 contain GlcNAc, Glc and Rha (its D/L-configurations were not determined). Meanwhile, it was proposed that the glycans are O-linked to the serine residues because of the prevalent existence of serine and a failure in removing the glycan by using N-glycanase (Stephenson *et al.* 2002). It should be noted, though, that the latter observation dose not completely rule out N-glycosylation, as it is highly possible that N-glycanase is not capable of hydrolysing bacterial N-glycans. In addition, the sequence and detailed structures of the glycans were not elucidated prior to the work in this thesis.

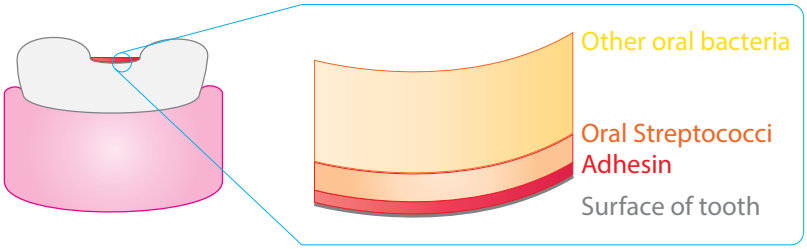


Figure 4-1 Role of oral bacterial adhesion in dental plaque formation. Oral *Streptococci* first adhere to human teeth and other oral bacteria subsequently attach onto colonized *Streptococci*.

Figure 4-2 Schematics of six domains in Fap1. A signal sequence (SS, yellow) and cell wall (CW, yellow) anchor are at its N- and C-terminus, respectively. Between them are four alternative repeat (RI and RII, red) and non-repeat (NRI and NRII, green) domains. The long line and the numbers at the bottom present the number of the amino acid (A.A.).

The heavy glycosylation of Fap1 makes a significant contribution to the biological function of *S. parasanguinis* fimbriae, *e.g.* Fap1 is believed to be improperly synthesized without correct glycosylation. Moreover, a monoclonal antibody capable of blocking its adhesive ability can only recognize glycosylated Fap1 (Bu *et al.* 2008; Zhou & Wu 2009). Therefore, it is of biological importance to understand the glycosylation of Fap1.

Preliminary research has previously been carried out to elucidate the glycosylation of Fap1. The genomic context of the fap1 gene in the genome of *S. parasanguinis* FW213 is shown in **FIG. 4-3**. Within this gene cluster, six genes were predicted to code glycosyltransferases (GTs), *gly*, *gtf3*, *galT1*, *galT2*, *gtf1* and *gtf2*. Other genes were attributed to the section of Fap1, *e.g. secY2* and *secA2* encoding accessory Sec proteins are essential for Fap1 biogenesis, and *gap1*, *2* and *3* were proposed to regulate Fap1 secretion.

Glycosylation of Fap1 is an OTase-independent (see **Section 1.2.2.2**) multiple-step process. GTF1/2/3 encoded by *gtf1/2/3* respectively were identified as GTs and proved to be involved in the biosynthesis of the Fap1 glycans (Wu *et al.* 1998; Zhu *et al.* 2011). The GTF1/2 complex was determined to be responsible for adding a GlcNAc on the protein (Wu *et al.* 1998). This has been proven by the fact that a deletion in either of them leads to an un-glycosylated phenotype. The activity of this complex was believed to be a conserved first step in the glycosylation of not only Fap1 in *S. parasanguinis* but also many other serine-rich proteins in other Gram-positive organisms such as *Staphylococcus*. Afterwards, a GTF3 enzyme transfers a Glc onto the GlcNAc residue (Zhu *et al.* 2011). This was investigated by comparing the GC-MS monosaccharide compositions of recombinant Fap1 expressed with GTF1/2 and GTF 1/2/3, respectively. All of these GTs were very well studied, *e.g.* the crystal structures of their product proteins were resolved and their functions were characterized by biochemical assays.

Figure 4-3 Diagram for the genomic context of the fap1 gene in the genome of *S. parasanguinis* FW213. The red arrow stands for genes encoding Fap1. The green genes encode Sec proteins that are essential for Fap1 secretion and the blue genes encode GTs. Putative genes encode more GTs are shown as grey arrows with blue strokes, and functions of pure grey genes are not fully understood yet.

Further steps in the glycosylation pathway remain elusive, *i.e.* functions of other three additional putative GTs are yet to be confirmed. For example, GALT1 was predicted to be a GT with two active sites, one of which, the domain of unknown function 1792 (DUF1792), was believed to function as a GlcT, adding another glucose residue on the previously synthesized Glc-GlcNAc, and another was believed to subsequently transfer a GlcNAc. Similar functional speculations were made for other putative GTs encoding by *galT2* and gly found in the genome of the organisms, *e.g.* GALT2 and GLY were proposed to be a rhamnosyltransferase (RhaT) and a GlcT, respectively. However, these proposals were only supported by *in vitro* biochemical assays carried out by our collaborators, instead of more concrete *in vivo* investigations.

This study aimed to elucidate the glycosylation of Fap1 based on rigorous glycomic and glycoproteomic data. The structures of the glycans carried by Fap1 were determined, and the specificities of each putative GT were clarified.

As it is very difficult to obtain a sufficient amount of *S. parasanguinis* Fap1 by direct extraction to carry out in-depth glycomic analysis, recombinant Fap1 that carries a much shorter serine-rich domain expressed in *E. coli* was used as a model for glycomic and glycoproteomic analysis. The amino acid sequence of the recombinant Fap1 is shown below:

MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRDPANEVKLSSENFDSEKAEEK
ISLSQSESASESVSESISESVSESVSTSESVSESVSESVSESISESVSESISESISESVSESTST
SIVLSESGAASGNKAT

The first 34 amino acids mainly comprise the Polyhistidine-tag and the remaining part mimics the serine-rich domain of Fap1. When Fap1 is co-expressed with a full set of putative GTs, it is called mature Fap1. If only a portion of the proposed GTs are co-expressed, they are mutant Fap1 (see **Section 2.2.3**). Briefly, Fap1 with GTF123, GTF123/GALT1, GTF123/GALT1/2 and GTF123/GALT1/2/GLY are the "mutants" that were studied.

The glycomic strategies used in this chapter are summarized in **FIG. 4-4**. The identities of monosaccharides contained by Fap1 glycans were determined by GC-EI-MS (**a**) before further investigations. For glycomic analysis, the Fap1 glycans were reductively eliminated, purified, chemically derivatised (such as permethylation and per-acetylation) and subjected to MALDI-TOF and MALDI-TOF/TOF analysis (**b** and **c**). The acetylated Fap1 glycans were oxidised by CrO_3 for characterising anomeric configurations (**d**). A small portion of the permethylated Fap1 glycans was hydrolysed, acetylated and further analysed by GC-EI-MS in order to determine the linkages of monosaccharides (**e**). For glycoproteomic analysis, mature Fap1 samples were in-gel digested with trypsin and analysed by LC-ESI-Q-TOF.

After the glycosylation of the mature Fap1 proteins expressed in *E. coli* had been fully elucidated, Fap1 extracted from *S. parasanguinis* (native Fap1) was collected and glycomically investigated by permethylation and MS analysis in order to establish whether the same glycans were found in the native material.

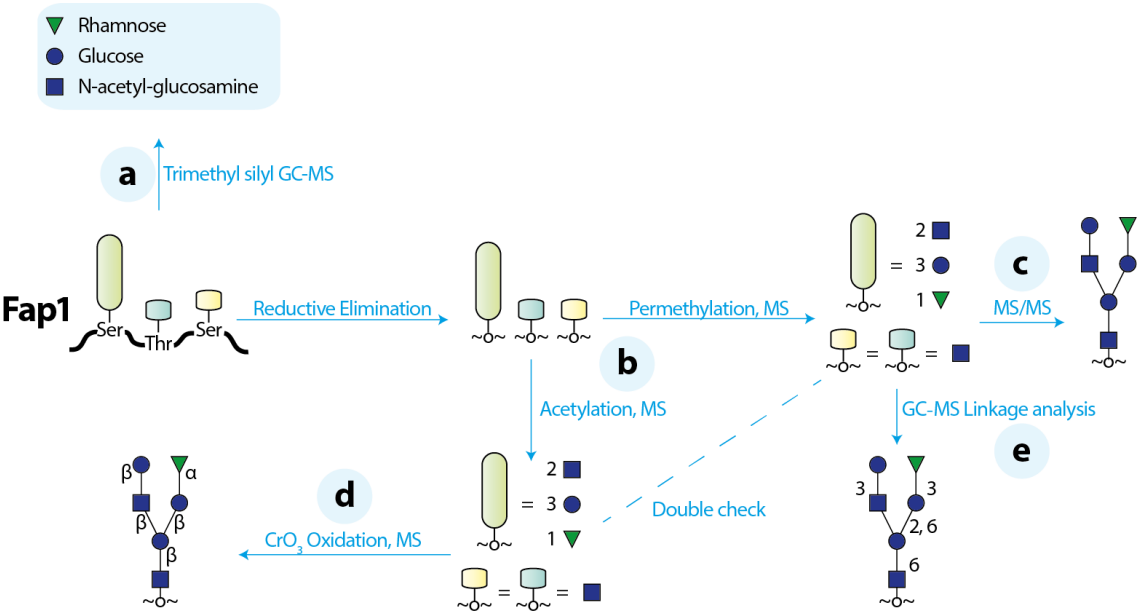


Figure 4-4 Diagram for the analytical glycomic strategies used in this chapter. The sample was initially subjected to trimethyl silyl GC-MS analysis (a). The Fap1 glycans were released, derivatised and analysed by MS (b). The permethylated glycans were sequenced by MS/MS (c), and their linkages were analysed by GC-MS linkage analysis (e). The peracetylated glycans were oxidised by CrO_3 and characterised by MS (d).

4.2 Glycomic Analysis of Mature Fap1 Glycan

4.2.1 GC-EI-MS TMS Analysis

Mature Fap1 was initially subjected to TMS derivatization and GC-EI-MS analysis in order to clarify the identities of constituent monosaccharides. As previous research had indicated that Fap1 may carry Glc, Gal, GlcNAc and GalNAc, two sets of monosaccharide standards were analysed together with mature Fap1. A full standard contained all these expected monosaccharides, and a partial standard contained only Glc and GlcNAc. The resultant chromatograms of the samples (purple for the mature Fap1, green for the full standard and red for the partial standard) are shown in **FIG. 4-5**. Based on the retention times, it is very easy to identify that the mature Fap1 glycan is composed of Rha, Glc and GlcNAc, which acts as the basis of further annotations on the MS spectra.

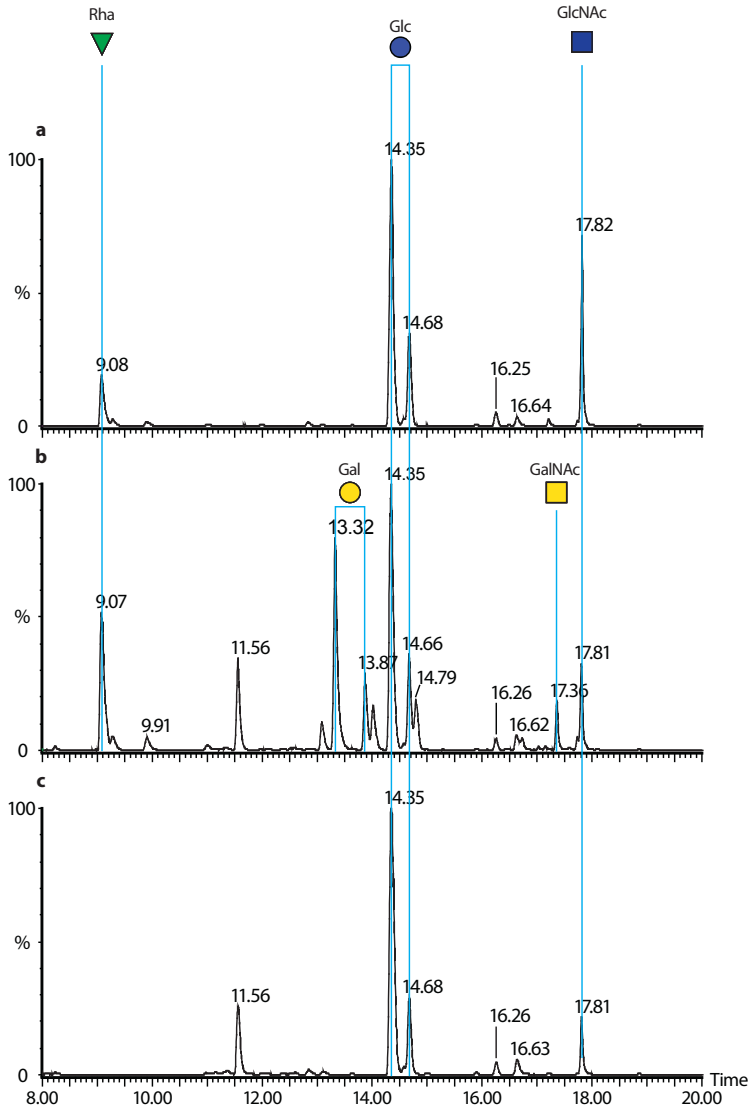


Figure 4-5 GC-EI-MS TMS sugar analysis. The mature Fap1 Samples and monosaccharide standards were methanolysed and TMS derivatised. GC-MS chromatograms of mature fap1 (a), Full standards (b) and partial standards (c) are shown. The comparison of retention times indicates the Fap1 glycan is composed of rhamnose, glucose and GlcNAc.

4.2.2 Glycan Profile of Mature Fap1

Glycans attaching on mature Fap1 were reductively eliminated, permethylated (or deutero-permethylated) and subjected to MALDI-TOF analysis (**FIG. 4-6**). Spectra **a** and **b** show permethylated glycans eluted with the 35% and 50% aqueous solution of acetonitrile (MeCN), respectively; **c** and **d** show deutero-permethylated glycans eluted with the 35% and 50% aqueous solution of MeCN, respectively. Peaks corresponding to sodiated glycans in the mass spectra are coloured red and annotated with *m/z* and glycan structures. Other peaks in the low *m/z* areas are attributed to matrix, and those near the annotated ones are due to under-permethylation (*m/z* lower 14), under-deutero-permethylation (*m/z* under 17) and potassiated glycans (*m/z* higher 16).

Annotation of the spectra was based upon the assumption that all these glycans are synthesized within the same pathway. MALDI-TOF/TOF analysis also made a great contribution to sequence these glycans (see below).

Five different structures were unambiguously identified in the mass spectra of both permethylated and deutero-permethylated glycans, which strengthens the fact that these signals were due to the glycans rather than matrix and contamination. Peaks at *m/z* 330.0 and 1362.1, corresponding to a single GlcNAc and a presumably fully synthesized branched glycan (see the glycan structure annotated at *m/z* 1362.1 in **FIG. 4-6 b**), respectively, are more intense than others. When taking the fact that many potential glycosylation sites can be found in the protein into consideration, it is reasonable to make a conclusion that many sites are glycosylated with only GlcNAc, and merely some of them are further synthesized to the branched hexa-saccharide.

The intensities of other glycan peaks were inconsistent in the spectra of the permethylated and detero-permethylated glycans. For example, the relative intensity between the peak at *m/z* 738 and 912 in the **FIG. 4-6 a** obviously differed from the that between the peak at *m/z* 774 and 954 in the **FIG. 4-6 c**, despite peaks at *m/z* 738 and 774, and at *m/z* 912 and 954 corresponding to a same tri-saccharide and tetra-saccharide, respectively. A difference in the ionization efficiency between the normal and deuterated material may cause this phenomenon.

All the glycan peaks found in the MS spectra were further subjected to MALDI-TOF/TOF analysis to confirm the proposed structures. MS/MS analyses on samples based on different derivatization methods showed a consistent fragmentation pattern for each individual glycan peak. Selected MS/MS spectra of permethylated glycans that were proposed to be linear and branched structures are shown in **FIG. 4-7** and **4-9**, respectively.

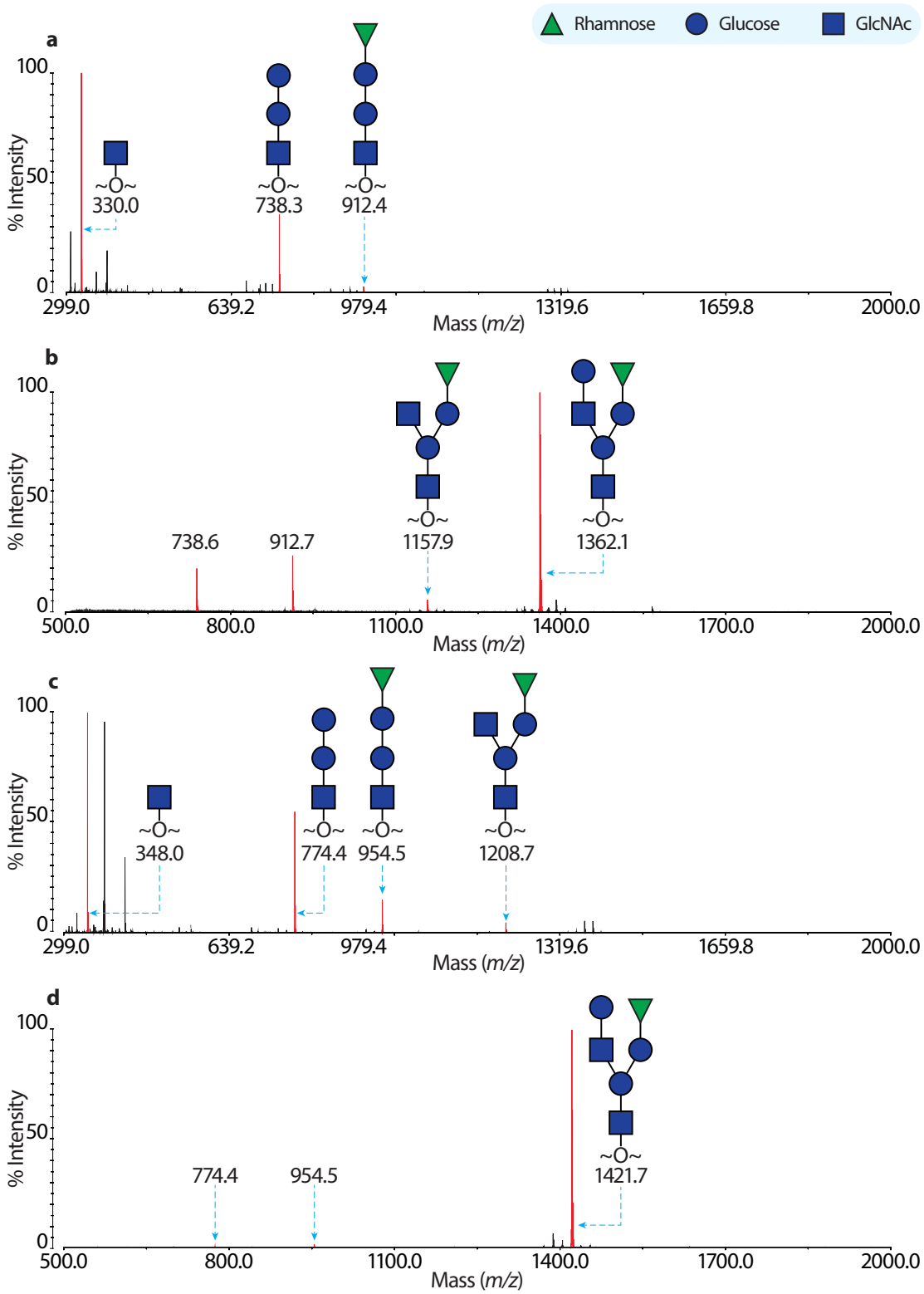


Figure 4-6 The MALDI-TOF profile of permethylated (deutero-permethylated) mature Fap1 glycan. Spectra (a) and (b) show 35% and 50% MeCN proportions of permethylated mature Fap1 glycan, (c) and (d) show 35% and 50% MeCN proportions of deutero-permethylated mature Fap1 glycan. The MS peaks corresponding to glycans are coloured red and annotated with *m/z* values and glycan structures.

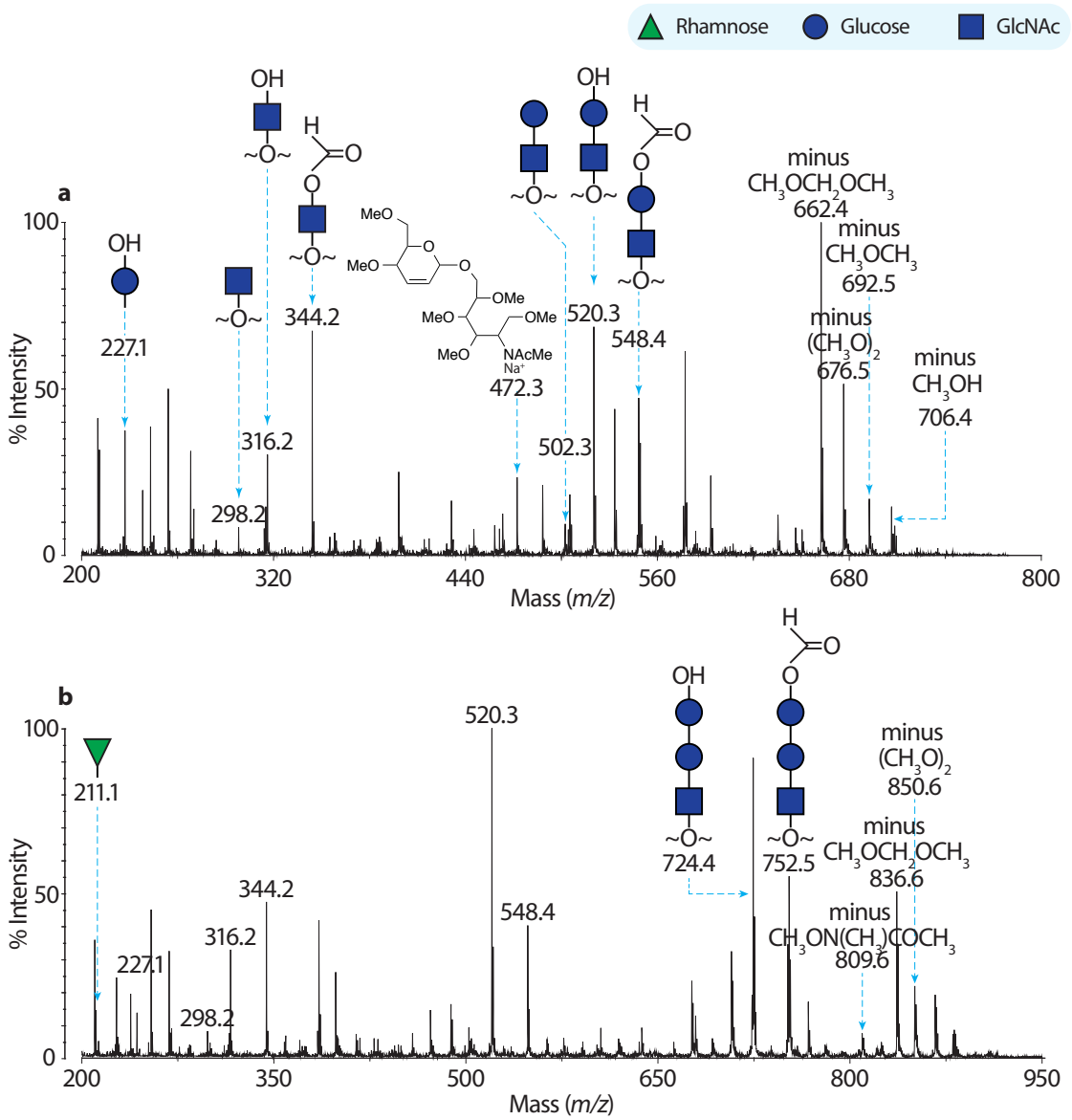


Figure 4-7 MS/MS analysis of putative linear Fap1 glycan. MALDI-TOF/TOF spectra of MS peak at *m/z* 738.3 and 912.7 are shown in (a) and (b) respectively. The spectrum (a) is fully annotated. For the spectrum (b), only the peaks present in (b) but not (a) are annotated.

The proposed linear glycan structures were confirmed by the MALDI-TOF/TOF analysis (**FIG. 4-7**), and the MS/MS spectra of the MS peaks at *m/z* 738.3 and 912.7 are shown in a and b, respectively. The fragmentation patterns strongly indicate both glycans share a linear structure. A strong peak at *m/z* 520.3 corresponding to a single cleavage Glc-GlcNAc structure were found in both spectra. Satellite peaks at *m/z* 548.4 due to a 1,5 position cross-ring cleavage were also observed, which further validated the annotation of the peak at *m/z* 520.3. Moreover, the peaks at *m/z* 316.2 in **FIG. 4-7 a** and peaks at *m/z* 316.2 and 724.4 in **FIG. 4-7 b** also supported linear structures. Notably, a peak at *m/z* 472.3 indicates an either 1,2 or 1,3 linked Glc-Glc glycosidic bond.

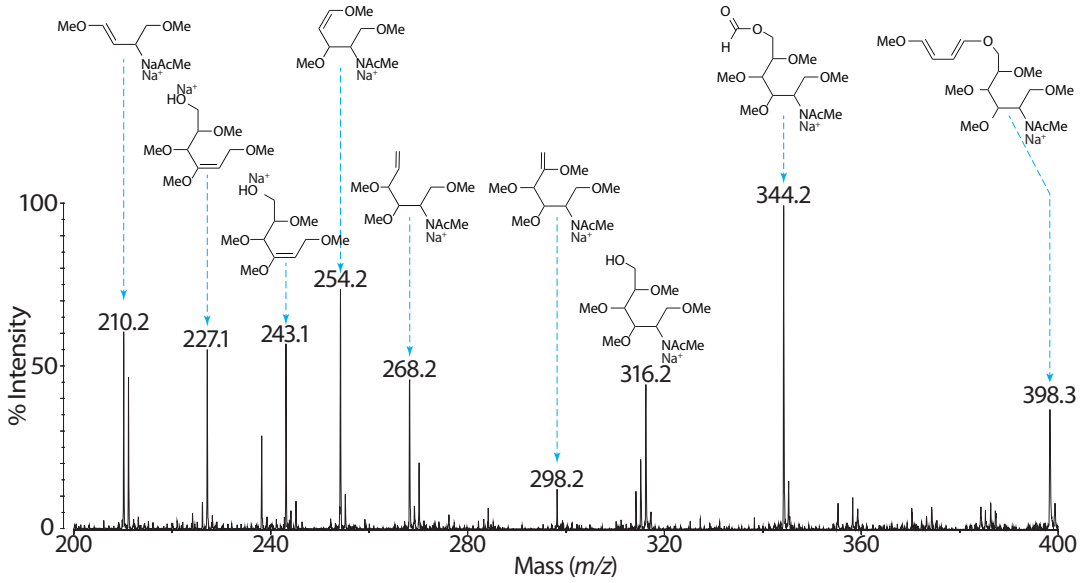


Figure 4-8 The MALDI-TOF/TOF spectrum of the MS peak at *m/z* 738.3 zoomed into a mass range of *m/z* 200-400 Da.

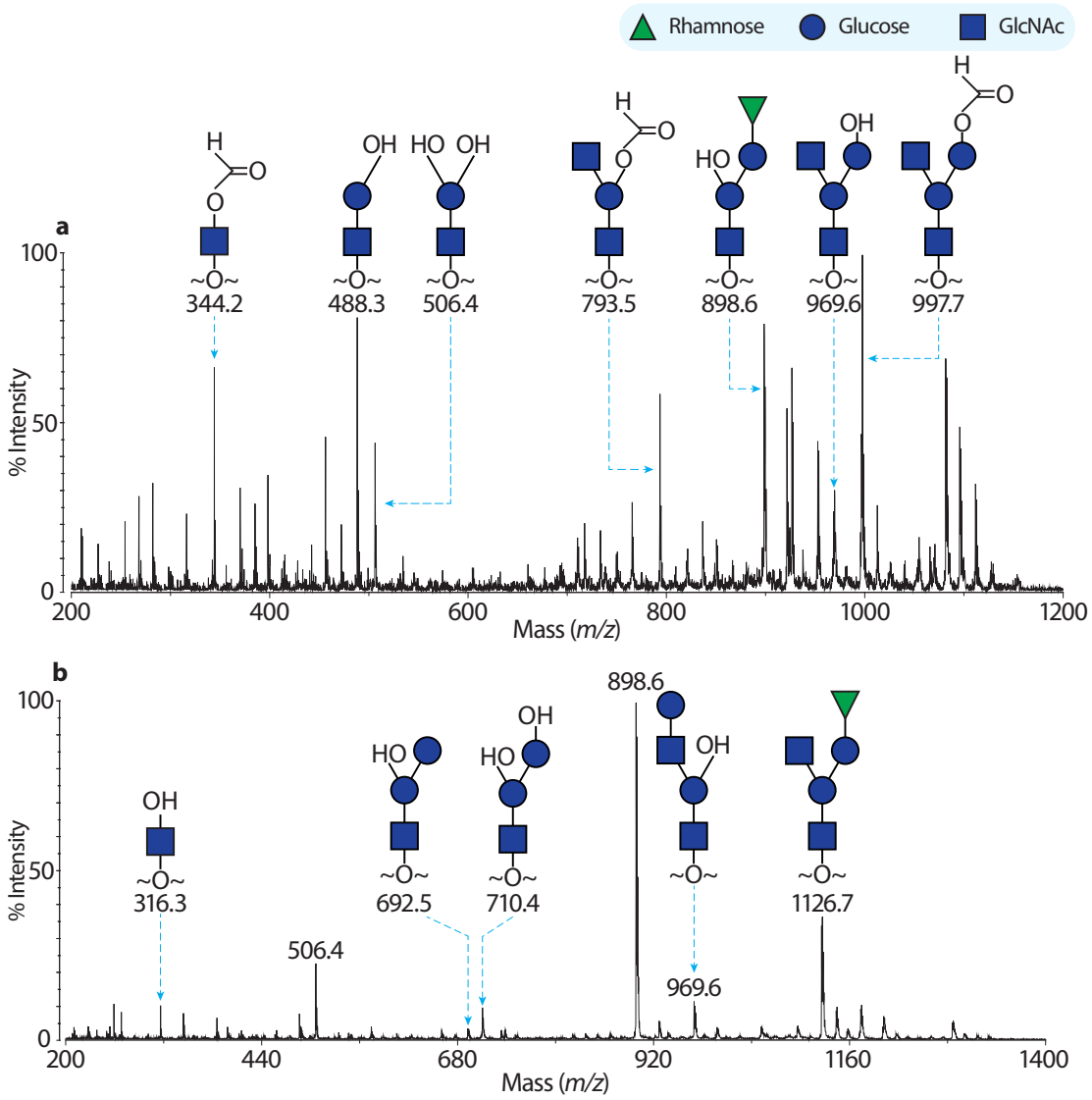


Figure 4-9 MS/MS analysis of putative branched Fap1 glycans. MALDI-TOF/TOF spectra of MS peak at *m/z* 1157.9 and 1362.1 are shown in (a) and (b), respectively. The spectrum (a) is fully annotated. For the spectrum (b), only the peaks present in (b) but not (a) are annotated.

There exist a lot of cross-ring fragments, which is probably because the small sizes of these glycans significantly facilitated fragmentations. Low *m/z* peaks corresponding to cross-ring fragments were annotated with chemical structures in **FIG. 4-8** that shows the MS/MS spectrum of the MS peak at *m/z* 738.3 zoomed into a mass range of 200-400 Da. The peaks at *m/z* 268.2 and 298.2 indicate a 6-linked Glc-GlcNAc glycosidic bond.

The proposed branched Fap1 glycans were also characterised by the MALDI-TOF/TOF analysis of (**FIG. 4-9**). The MS/MS spectra of the MS peak at *m/z* 1157.9 and 1362.1 are shown in **a** and **b**, respectively. In both spectra, the previously mentioned peak at *m/z* 520.3 in **FIG. 4-7** shifted to *m/z* 506.4 that corresponds to a doubly cleaved structure. In addition, the peaks at *m/z* 898.6 and 969.6 (**FIG. 4-9 b**) correspond to a loss of one of the two branches which further indicates these two glycans share a branched structure. These structural investigations match the previously predicted function of GALT1. As soon as a second GlcNAc is transferred to the glycans, their structures become branched.

As a summary, the glycomic studies of the mature Fap1 strongly indicate that it carries multiple glycans ranging from a single GlcNAc to the branched hexa-saccharide with a sequence of Rha-Glc-(Glc-GlcNAc)-Glc-GlcNAc.

4.2.3 Peracetylation and Chromium (VI) Oxidation

For further characterising the anomeric configurations of each monosaccharide, the Fap1 glycans were peracetylated and oxidized by CrO_3. As the anomeric carbon with an α-configuration is more stable than a β-configuration during the oxidation process, β-anomeric carbons are oxidized while α-carbons remain largely unaffected under an optimized condition. Each oxidation reaction adds 14 Da to the molecular weight of the glycan. Thus, by comparing the *m/z* before and after oxidation, anomeric configurations can be assigned.

All Fap1 samples were analysed by using the above method, and the MALDI-TOF spectra of peracetylated Fap1 glycan before and after CrO_3 oxidation are shown in **FIG. 4-10 a** and **b**, respectively. Again, red peaks corresponding to sodiated glycans are annotated with *m/z* and structures. Only four structures were detected, which is less than those found in the spectra of permethylated glycans. This was attributed to inherent lower sensitivity of the peracetylation method.

The peak at *m/z* 1032.2 corresponding to the tri-saccharide with a sequence of Glc-Glc-GlcNAc in **FIG. 4-10 a** shifted to a new peak at *m/z* 1060.3 after oxidation (b). The shift of 28 Da in *m/z* indicates both anomeric carbons in this structure have β-configuration. Meanwhile, the peak at *m/z* 1262.3 **FIG. 4-10 a** corresponding to a tetra-saccharide with a sequence of Rha-Glc-Glc-GlcNAc also shifted 28 Da in *m/z*, resulting in a peak at *m/z* 1290.4 in **FIG. 4-10 b**. This strongly suggests the Rha has an α-configuration. In addition, the molecular weight of the fully synthesized hexa-saccharide incremented 56 Da after oxidation, implying that there is only one α-anomeric carbon in this structure. Overall, except for the terminal Rha is α-linked, all other monosaccharides (except for GlcNAc at the reducing end) are β-linked.

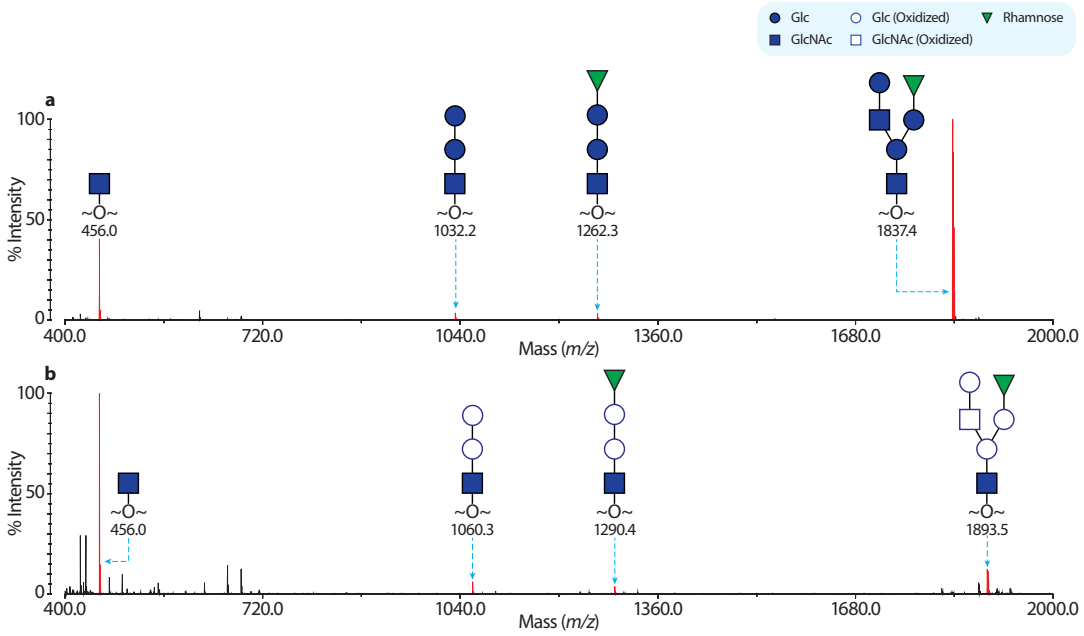


Figure 4-10 The MALDI-TOF spectra of the peracetylated mature Fap1 glycan before (a) and after (b) CrO_3 oxidation. The MS peaks corresponding to glycans are coloured red and annotated with *m/z* values and glycan structures.

4.2.4 GC-EI-MS Linkage Analysis

In order to elucidate linkages of the monosaccharides in Fap1, the glycans released from mature Fap1 were permethylated, hydrolysed, reduced and acetylated. The resultant partially permethylated alditol acetates (PMAAs) were characterized by GC-EI-MS. The results are summarised in **Table 4-1**.

Linkage analysis data confirmed the t-Rha and t-Glc to be two non-reducing ends, and 6-linked GlcNAc to be the reducing end sugar. In addition, 3-linked Glc, 3-linked GlcNAc and 2, 6-linked Glc were also observed. These observations are fully consistent with the proposed branched structure based on MS and MS/MS analysis. For instance, the cross-ring fragments indicate the existence of 6-linked GlcNAc and 2 or 3-linked Glc.

Based on all the MS-based glycomic studies, it is proposed that the fully synthesized Fap1 glycan is a branched hexa-saccharide with a sequence of α-Rha-1,3-β-Glc-1-(β-Glc-1,3-GlcNAc1-)2,6-β-Glc-1,3-β-GlcNAc.

Table 4-1 Linkage analysis of Fap1 glycan

Type of Monosaccharide	Characteristic Ions	Retention Time (min)
Terminal-Rha	175+162+131+102	16.74
Terminal-Glc	205+162+161+118	18.62
3-Glucose	277+234+161+118	19.83
6-GlcNAcitol	218+174+161+130+117	21.24
2,6-Glucose	190+189+130+129	21.36
3-GlcNAc	275+203+159+117	23.81

4.3 Glycoproteomics of Mature Fap1

The previous section describes the MS-based glycomic studies of Fap1 glycans. In order to better understand the glycosylation of Fap1, the mature Fap1 sample was in-gel digested and analysed by LC-ESI-Q-TOF. Results are shown in **FIG. 4-11**.

The MS spectra of three observed glycopeptides are shown in spectra **FIG. 4-11 a**, **b** and **c**, respectively. The MS peaks in the latter two spectra corresponded to two glycoforms of a same peptide with a sequence of LSSENFDSEKAEEK. Note that the glycoform carrying a GlcNAc plus the fully synthesised hexa-saccharide was observed and annotated in **FIG. 4-11 c**. It is obvious to find that the sequence of the glycopeptide contains one missed-cleavage lysine site. A completely cleaved glycopeptide carrying two GlcNAcs with a sequence of LSSENFDSEK was found, whose MS peak was shown in **FIG. 4-11 a**.

The CID MS/MS spectra of the MS peak in **FIG. 4-11 a** and **c** are shown in **FIG. 4-11 d** and **e**, respectively. These MS/MS spectra played a significant role in annotating the MS spectra of glycopeptides. At the low *m/z* end of both MS/MS spectra, a cluster of strong peaks at *m/z* 204.0, 186.0, 168.0, 144.0, 138.0 and 126.0 that corresponds to a GlcNAc and its fragments were observed, which are strong indicators of a glycopeptide. Meanwhile, at the high *m/z* end, two peaks with a *m/z* interval of 203 were observed (*m/z* 1155.2 and 1138.3 in d and *m/z* 1612.3 and 1815.5 in e), which are characteristic of the fragmented peptide backbones and glycopeptides carrying a single GlcNAc. Because the molecular weight of the glycopeptide can be calculated from its molecular ion, the molecular weight difference between the glycopeptide and peptide is due to glycosylation. For instance, the calculated molecular weight of the triply charged ion at *m/z* 951.9 is 2852.7 Da, and the detected molecular weight of the peptide backbone in its MS/MS spectrum is 1611.3 Da. Their mass difference, 1241.4 Da, corresponds to a GlcNAc residue (203.1 Da) plus the fully synthesised hexa-saccharide residue (1038.4 Da). In this way, glycosylation of Fap1 was further characterised. Note that these glycans are not permethylated, therefore the masses of monosaccharides referred to here are different from their previously mentioned masses.

The peptide backbones were simultaneously sequenced by CID MS/MS. The peptide found in **FIG. 4-11 a** was very well fragmented in its MS/MS spectrum (d), *i.e.* most Y ions were observed and the sequence was very confidently assigned. However, for the glycopeptide carrying the fully synthesised hexa-saccharide (**FIG. 4-11 c**), only a few B and Y ions from the C terminals were observed, which can be attributed to the relatively larger hydrophilic glycan moiety.

The glycoproteomic data suggest that the mature Fap1 mainly carries two glycans, a single GlcNAc and the fully synthesised hexa-saccharide, which is fully consistent with the glycomic data. All of the other glycans found in MALDI glycan profiles are probably precursors

of the hexa-saccharide. In addition, as the detected glycopeptides were found to carry more GlcNAc residues than the fully synthesised glycan, it is reasonable to presume that only a small portion of GlcNAc residues are further elongated to the hexa-saccharide.

It is noteworthy that only one glycopeptide carrying the fully synthesised Fap1 glycan was observed. This can be due to two reasons. First of all, although mature Fap1 is a recombinant protein and much shorter than native Fap1, it still has a relatively long serine-rich motif, which greatly hinders trypsin digestion. In-silica digestion revealed a single serine-rich peptide with a molecular weight of over seven KDa after tryptic treatment, and this weight turns a lot bigger when taking a larger number of potential glycan moieties into consideration. It is very difficult to detect such a big peptide in a normal bottom-up proteomic analysis. In addition, as heavily glycosylated peptides are highly hydrophilic, they are a lot more difficult to be ionized than normal peptides. Consequently, their MS signals are significantly suppressed and hard to be detected.

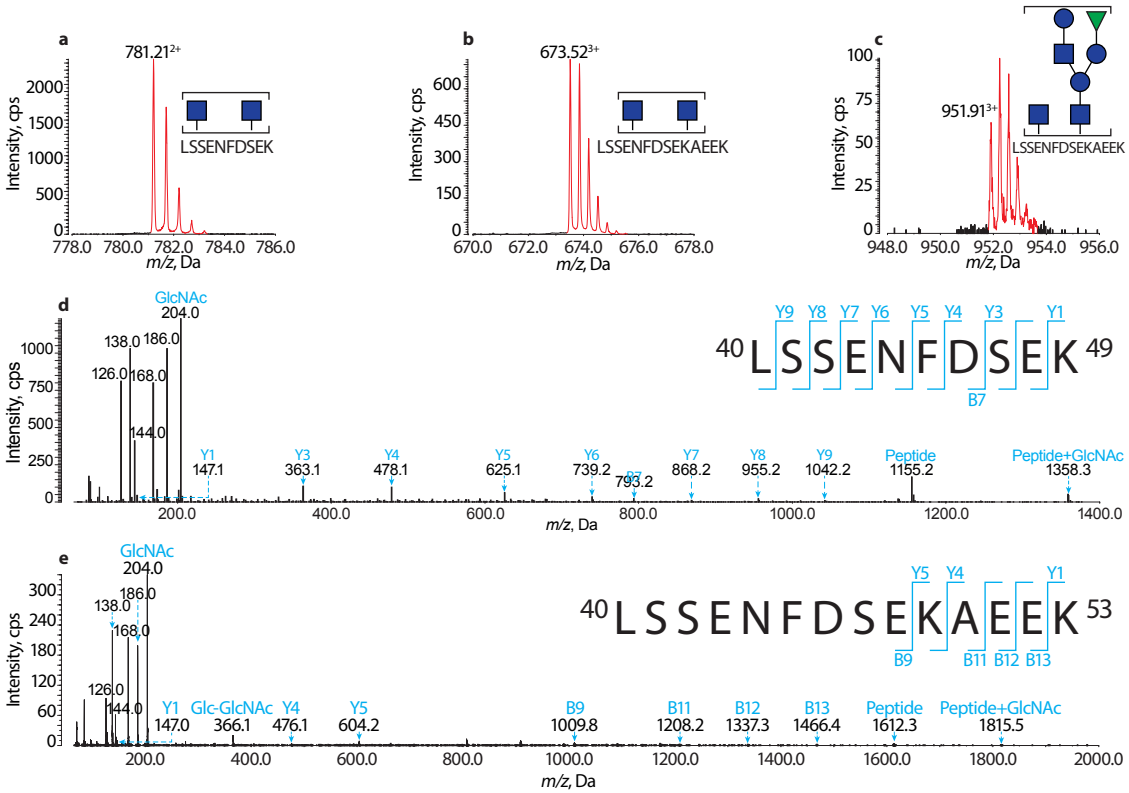


Figure 4-11 MS-based Proteomics of mature Fap1. The MS spectrum of a glycopeptide with a peptide sequence of LSSENFDSEK carrying two GlcNAcs, and glycopeptides with a peptide sequence of LSSENFDSEKAEEK carrying two GlcNAcs and one GlcNAc plus the fully synthesised Fap1 hexa-saccharide are shown in (a), (b) and (c) respectively. The MS peaks are coloured red and annotated with the peptide sequences and glycans modifying the peptides. The MS/MS spectra of the MS peaks in (a) and (c) are shown in (d) and (e), respectively. The corresponding Y and B ions are annotated on the MS/MS spectra and peptide sequences (the numbers shown on the two terminals show the number of amino acids in the mature Fap1 sequence). The clusters of strong peaks at *m/z* 204.0, 186.0, 168.0, 144.0, 138.0 and 126.0 correspond to a GlcNAc and its fragments.

4.4 Glycan Profile of Mutant Fap1

In order to further confirm the proposed structures are correct and elucidate the biosynthetic pathway, MS profiles of permethylated glycans from mutant Fap1 samples were recorded (**FIG. 4-12**).

The glycan synthesized by GTF1/2/3, and GTF1/2/3 plus GALT1 gave rise to MS spectra a and b, respectively. When one more GT, GALT2, was co-expressed with Fap1 in *E. coli*, the protein was modified with the glycan shown in spectrum c. As mentioned earlier, there are two presumably GT domains in GALT1s. In order to confirm this, Fap1 co-expressed with GTF1/2/3 and the DUF1792 of GALT1 was prepared and analysed. The result is shown in spectrum d.

A single peak at *m/z* 534.1 corresponding to Glc-GlcNAc was detected in **FIG. 4-12 a**. This structure had been hard to be identified in the profile of mature Fap1 glycan (**FIG. 4-6**), as its signal was very weak and suppressed by signals from the matrix. However, when Fap1 was only co-expressed with GTF1/2/3, the strong peak at *m/z* 534.1 was clearly observed. The function of GTF3 was therefore concluded to transfer a Glc to GlcNAc, which is consistent with previous research (Zhu *et al*. 2011).

When GALT1 were added to the co-expression system, two strong peaks were detected. One of them was at *m/z* 738.5 and attributed to the previously observed linear Glc-Glc-GlcNAc structure (**FIG. 4-6**). Another new peak was at *m/z* 983.7, corresponding to a branched GlcNAc(Glc)-Glc-GlcNAc structure. This observation supports the speculated two-domain feature of the GALT1 enzyme, and suggests one of them functioning as a GlcT and another acting as a GlcNAcT. Moreover, when its GlcNAcT domain is mutated, GalT1 only transfers a Glc, confirming that the remaining DUF1792 domain transfers Glc (**FIG. 4-12 d**). It was interesting to observe the MS peak at *m/z* 534.1 completely disappeared in this spectrum, indicating all precursors are transformed into products and GALT1 is a highly efficient enzyme.

Four previously discussed MS peaks were found in **FIG. 4-12 c**. One peak at *m/z* 738.5 and another weak peak at *m/z* 983.8 corresponding to two precursors respectively had been observed in **FIG. 4-12 b**. Meanwhile, peaks at *m/z* 912.7 and 1157.9 corresponding to two products of GALT2 catalyst had been observed in **FIG. 4-6**. The mass difference of 174 between peaks at *m/z* 912.7 and 738.5, and at *m/z* 1157.9 and 983.8 indicates GALT2 transferring a Rha. By comparing the relative ratios of peaks at *m/z* 738.5 and 983.8 in **FIG. 4-12 b** and **c**, it was obvious that the intensity of the latter peak was significantly reduced, indicating more branched structures were used by GALT2. Consistently, the branched product dominated the spectrum suggested more branched structures were generated by GALT2. Collectively, GALT2 is more efficient in transferring Rha to the branched structure, which could explain why the peak at *m/z* 983.8 was extremely weak and hard to be identified in **FIG. 4-6**.

To sum up, all glycans found in mature Fap1 were found in mutant Fap1, which strongly indicates that all of them are synthesized within the same pathway. In addition, the specificity of each enzyme was clarified, thus establishing a biosynthetic pathway (see **FIG. 4-14**, **Section 4.6**).

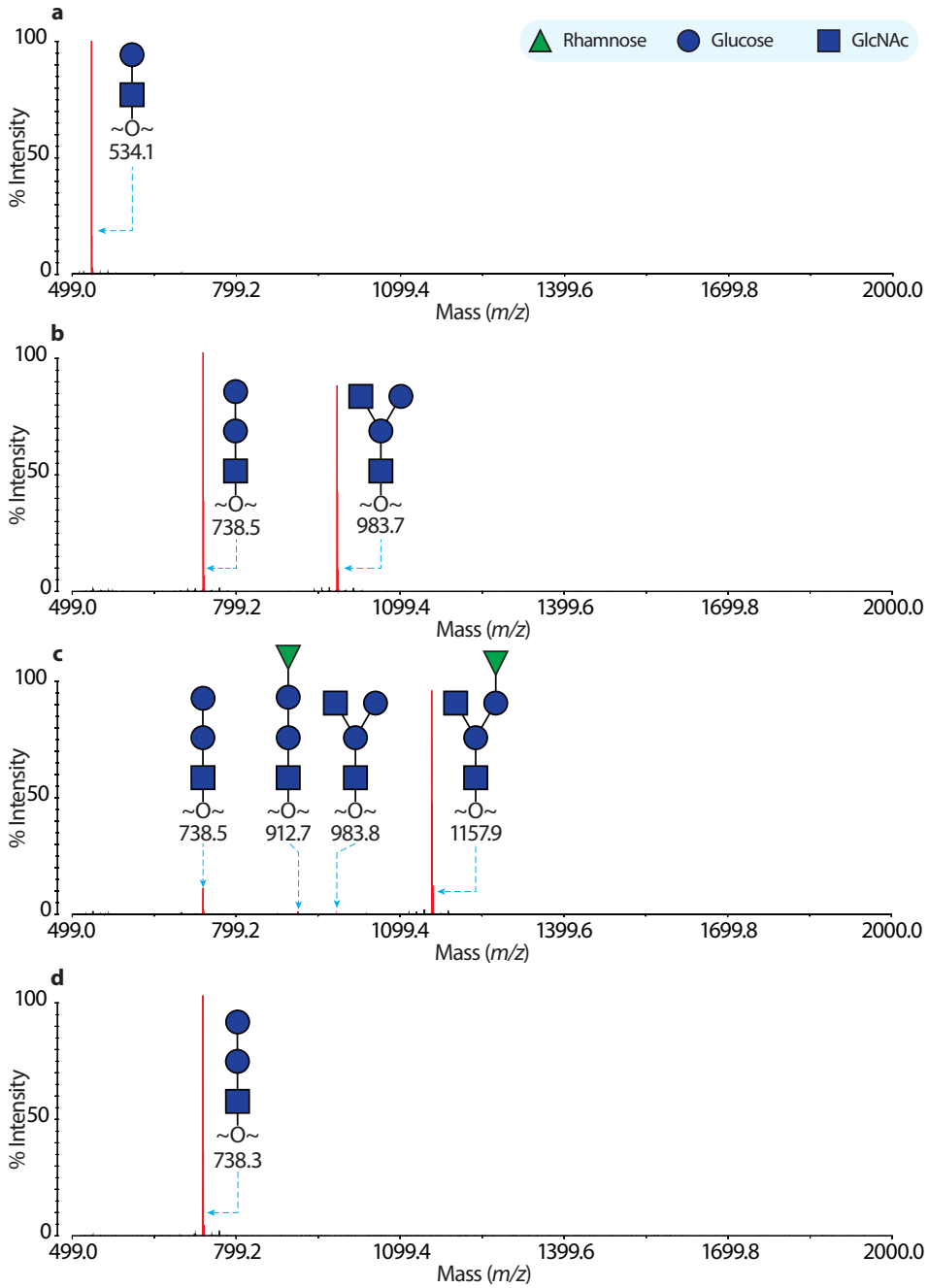


Figure 4-12 The MALDI-TOF profile of permethylated glycan from mutant Fap1. MS spectra of recombinant Fap1 co-expressed with GTF1/2/3 (a), GTF1/2/3 and GALT1 (b), GTF1/2/3 and GALT1/2 (c), and GTF1/2/3 and GALT1-DUF1792 domain are shown. The MS peaks corresponding to glycans are coloured red and annotated with *m/z* values and glycan structures.

4.5 Glycan Profile of Native Fap1 Glycans

As mature Fap1 is a recombinant protein expressed in *E. coli*, although its glycosylation had been in-depth characterised, it is still important to analyse native Fap1 protein directly extracted from *S. parasanguinis*. However, as native Fap1 could not be harvested with a high amount, only MS-based permethylation analysis was done to compare the native and recombinant Fap1 glycans.

The MALDI-TOF spectra of permethylated glycans released from native Fap1 are shown in **FIG. 4-13**, and 35% and 50% MeCN eluents are shown in a and b, respectively. Red MS peaks were attributed to glycans and annotated with *m/z* and structures. Being similar to recombinant Fap1, five different glycan structures ranging from a GlcNAc to the hexa-saccharide were detected, and therefore the annotation of these peaks was based upon the glycan structures of mature Fap1. The consistency between the recombinant and native Fap1 glycans further enhanced the reliability of recombinant Fap1 as a model of native Fap1. Again, the peaks at *m/z* 329.9 and 1361.5 were the most intense, and all other peaks at *m/z* 738.2, 912.1 and 1157.3 were rather minor. This is distinct from the spectra of mature Fap1 where the latter three peaks were of fair intensity, which indicates the involved GTs are more efficient in *S. parasanguinis*.

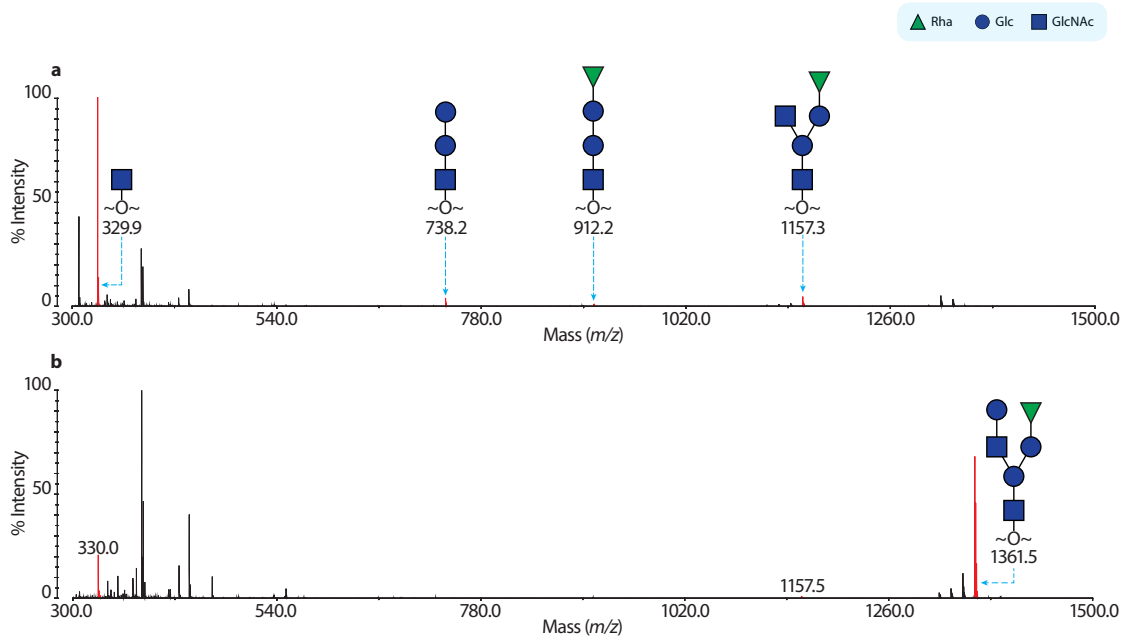


Figure 4-13 MALDI-TOF spectra of permethylated glycans released from native Fap1. The 35% and 50% MeCN eluents are shown in (a) and (b) respectively. The MS peaks corresponding to glycans are coloured red and annotated with *m/z* values and glycan structures.

4.6 Discussion and Conclusions

Combined with the glycomic and glycoproteomic strategies, this study confidently defined glycosylation of Fap1. Fap1 is a heavily O-glycosylated serine-rich protein mainly carrying GlcNAcs and the branched hexa-saccharides with a sequence of α-Rha-1,3-β-Glc-1-(β-Glc-1,3-GlcNAc1-)2,6-β-Glc-1,3-β-GlcNAc. The MS data indicate that many of its potential O-glycosylation sites are first GlcNAcylated and only a small portion of GlcNAc residues is further elongated to the hexa-saccharides. A proposed biosynthetic pathway of the hexa-saccharide is elucidated and shown in **FIG. 4-14**.

The fully synthesised hexa-saccharide carried by Fap1 is the most complicated post-translational modification (PTM) of serine-rich proteins characterised by far, *e.g.* another serine-rich (Srr1) protein found in *S. agalactiae* only carries GlcNAc. Nevertheless, it seems glycosylation of all serine-rich proteins initiates with GlcNAcylation mediated by a GTF1/2 (GTFA/B) complex (Chaze *et al.* 2014). In *S. parasanguinis*, a GTF3 enzyme transfers a Glc to the 6-position of the GlcNAc residue, whereas the GlcNAc residue is 6-acylated in S. agalactiae, potentially blocking further glycosylation. It is still unknown whether the acylation is actively regulated by the organism.

In addition to the previous studied GTs, GTF1/2 and 3, the specificities of three predicted GTs, GALT1/2 and GLY are determined in this study. GALT1 is an intriguing two-domain enzyme. Its N-terminal domain (DUF1792) possesses a novel folding that cannot be straightforwardly assigned to GT-A, B or C families, *i.e.* it has a DXE motif that does not directly interact with bivalent metal ion and a special topology relative to other GT folds. This domain transfers a Glc to the Glc-GlcNAc structure before the GT-A folded C-terminal domain of GALT1 transfers a branching GlcNAc.

It is unprecedented that a two-domain GT is related to glycosylation of serine-rich proteins, despite this kind of enzyme having been characterised before in *Saccharomyces cerevisiae* for instance (O'Reilly *et al.* 2006). Incorporating two GT domains into a single enzyme may dramatically increase its efficiency. Indeed, GALT1 is a highly active enzyme that converts almost all the precursors to products (compare **FIG.4-12 a** and **b**).

GALT1 essentially creates two antennae for the Fap1 glycan. Subsequently, a Rha is transferred onto the Glc by GALT2 and a Glc is transferred on the the GlcNAc by GLY. GALT2 has a semi-relaxed specificity, *i.e.* it can recognise both linear and branched glycan acceptors, though the latter one is greatly favoured (see **FIG.4-12 c**). GLY is a widely spreading GT that is found in many *Streptococci* as well as *Lactobacillus*. In certain organisms such as *S. pneumonia*, its multiple homologs can be found, indicating it might be playing an important role in glycosylation of serine-rich proteins.

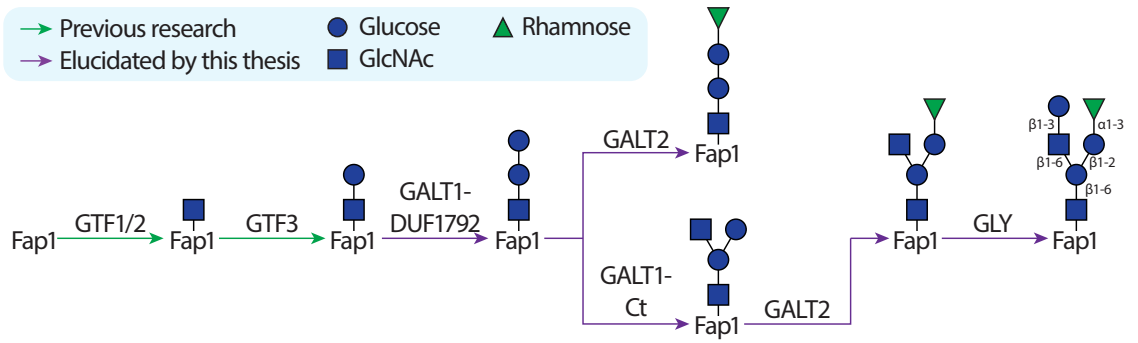


Figure 4-14 Biosynthetic pathway of the Fap1 hexa-saccharide. The green and purple arrows present biosynthesis elucidated by previous and current research, respectively. The GTs involved in each biosynthetic step are annotated above the arrows.

Albeit glycosylation of Fap1 has been in-depth analysed in the current work, functions of several other proteins including GAP1/2/3 encoded in the Fap1 gene cluster are not fully understood. Moreover, glycosylation of most other serine-rich proteins is yet to be researched. All the proposed GT functions and especially their characteristics are still waiting to be estimated under a broader context provided by further research in this area.

Chapter 5 MS-based Characterisation of Bacterial O-Glycosylation Systems

This chapter describes research efforts aimed at improving current methods for investigating putative bacterial glycosylation systems. First, the previously reported ZIC-HILIC LC-ESI-MS method was applied to whole cell lysates and periplasmic proteins of *B. thailandensis* in order to search for glycopeptides, and in turn characterise the type of glycosylation. When optimising this analytical method for *B. thailandensis*, it was noticed that glycopeptides can be readily detected by a simple LC-ESI-MS analysis. This approach was successfully used to identify glycosylation in another organism, *C. burnetii*. The analytical strategies described in this chapter could make contributions towards characterising targets for novel glycoconjugate vaccines as well as furthering the understanding of bacterial glycosylation systems.

5.1 Background

Vaccines have become one of the most effective "weapons" of human beings to fight against infectious diseases, since they were initially invented. The history of immunology essentially started with Edward Jenner's introducing the smallpox vaccine. Though extremely complicated immunological signalling networks underlie the protection provided by the vaccines, its fundamental logic is rather simple, *i.e.* a vaccine of a pathogen stimulates the host immune activities against the pathogen, protecting the host from the future infections. Therefore, vaccines protect people from suffering infectious diseases in the first place, which is far more advantageous than passive antibiotic-based treatment. This is especially true in our era when a growing number of the multidrug-resistant bacteria are being identified.

Traditionally, the main sources of vaccines are pathogens themselves in a "poison-reduced" form, *e.g.* live attenuated or killed. The live attenuated vaccines normally lead to both humoral and cellular immunity as the vaccinated hosts are exposed to multiple immunogenic epitopes. However, these live attenuated pathogens could revert to their full-virulent states, causing potential risks for the vaccinated people, which can be reflected by the fact that these vaccines cannot be used for vaccinating immunosuppressed people. The relatively safer killed vaccines, on the contrary, only stimulate humoral immunity, which means multiple vaccinations are essential to maintaining the protection.

Because of the apparent limitations of both kinds of vaccines, research efforts were made to create safer vaccines that stimulate memory immune responses. One of the most successful strategies is to chemically bond bacterial polysaccharides to carrier proteins, generating the so-called glycoconjugate vaccines. The polysaccharide parts efficiently induce the humoral immunity, and the protein carriers induce immunological memory. In addition, these vaccines are safe to be administered to immunosuppressed people. Glycoconjugate vaccines have been licenced that protect people from various pathogens such as *Haemophilus influenza* (Theodoratou *et al.* 2010) and *Neisseria meningitidis* (Poland 2010). Despite the supreme advantages, the preparation of glycoconjugate vaccines can be very difficult. For instance, the LPS O-antigens are normally good candidates for producing glycoconjugate vaccines, but they are extremely difficult to purify from the pathogens without contamination with lipid A moieties that could be lethal when directly injected. Moreover, the chemical reactions used for coupling the polysaccharides to the carrier proteins are hard to control, leading to variations between different batches. Therefore, there is in a need for a facile and economic method to produce glycoconjugate vaccines. The solution for this problem could be buried in the field of bacterial glycobiology.

Bacterial glycobiology is attracting more and more research interest mainly because of the recently characterised bacterial protein glycosylation systems, namely the OTase-dependent pathway (see **Section 1.2.2.1**). The OTases are the glycosyltransferases with relaxed specificities that modify proteins with oligosaccharides. OTase-dependent glycosylation systems have been identified in multiple bacterial pathogens such as Campylobacter, Neisseria, Vibrio and *Burkholderia*. The pathogens with a mutated OTase normally showed reduced virulence, and it was therefore proposed that the OTase-dependent glycosylation of bacterial proteins is associated with the evasion of host immunity, bacterial adhesion and mobility. Wider biological implications that are yet to be determined are also highly likely.

Though the detailed mechanisms of bacterial O-OTase-dependent glycosylation still remain elusive, N-OTase has emerged as a promising tool for creating glycoconjugate vaccines since the discovery of the first OTase gene, pglB. The pglB gene was initially discovered in *Campylobacter jejuni*, and it encodes an N-OTase that has a relaxed specificity transferring any

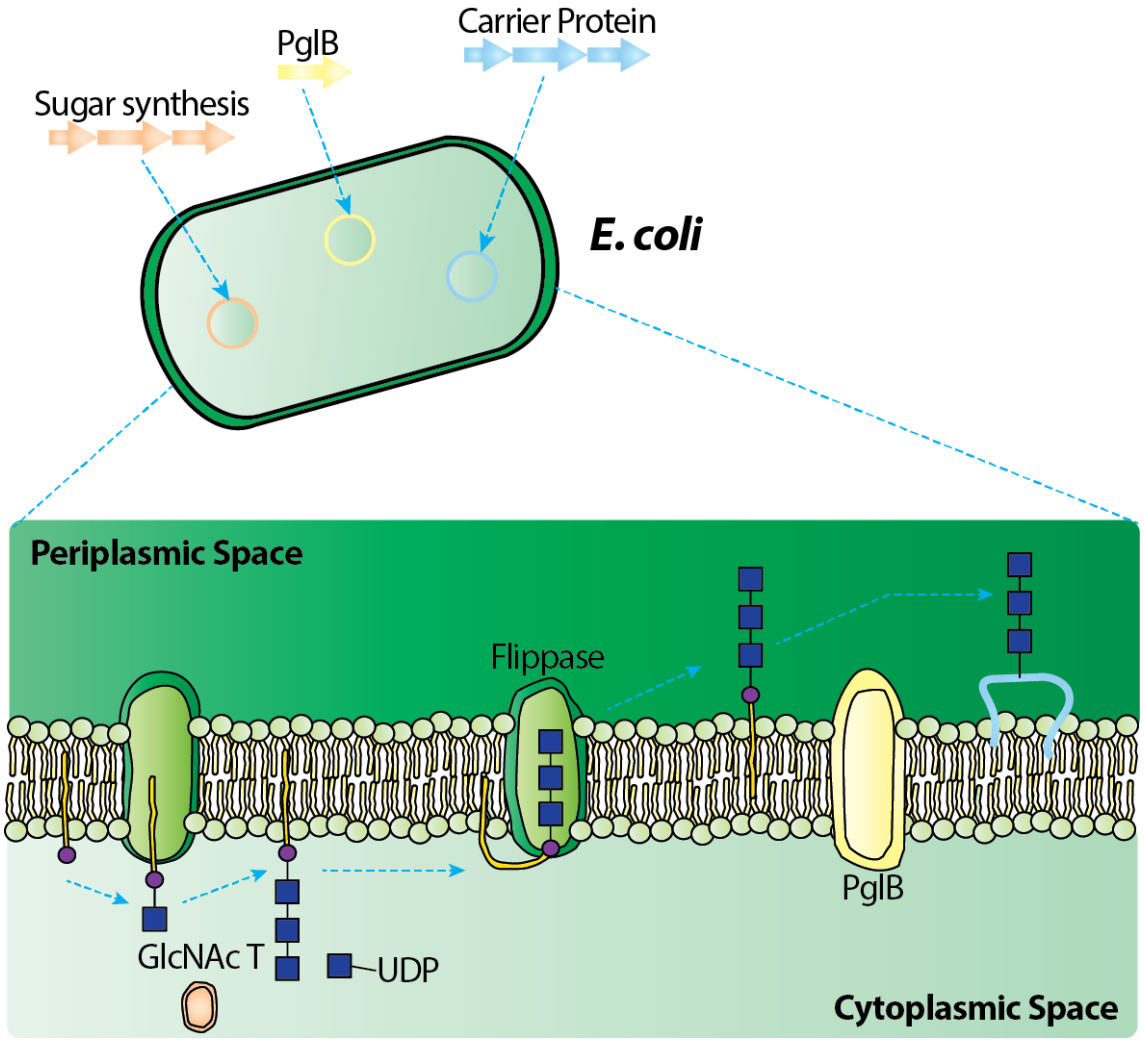


Figure 5-1 Diagram illustrating the principles of protein glycosylation mediated by the OTase in the genetically modified *E. coli*. Briefly, plasmids encoding the OTase (PglB in this case), carrier proteins and enzymes used for synthesising the polysaccharides are transferred into the *E. coli*. The polysaccharides are assembled on an Und-PP carrier and are flipped from the cytoplasmic to the periplasmic space, where they are transferred to the carrier proteins by PglB.

oligosaccharide containing an amino-sugar as its reducing end to the D/E-X-N-X-S/T consensus sequence. The PglB can be functionally expressed in *E. coli*, *i.e.* when co-expressing it with an acceptor protein together with polysaccharide synthetases results in glycosylation with certain glycans (**FIG. 5-1**), thus producing glycoconjugate vaccines.

As mentioned in **Section 1.2.2.1**, bacterial O-OTases such as PglL have also been characterised in addition to the N-OTase. PglL was suggested to have an even wider specificity than PglB, which can transfer almost any oligosaccharide to serine/threonine residues. This makes it a potentially better OTase for use in producing glycoconjugate vaccines. However, PglL is not as well understood as PglB, *e.g.* no consensus sequence was determined for PglL. More PglL homologs therefore need to be investigated in order to more comprehensively understand their biological implication as a widespread enzyme family utilised by various pathogens and to explore their potential for producing glycoconjugate vaccines.

Previous work has revealed putative *pglL* genes in *B. thailandensis* and *V. cholerae*, and confirmed their functions in a genetically modified *E. coli* system. *B. thailandensis* is a Gram-negative bacterium that has been occasionally reported as a pathogen (Glass *et al*. 2006). The organism attracts numerous research interests due to it being used as model to investigate the pathogenicity of the human pathogen, *B. pseudomallei* that is the etiological agent of melioidosis (Haraga *et al*. 2008). It was also used as a live vaccine against *B. pseudomallei*, and it seems the vaccine efficiency is dependent on whether or not the candidate *B. thailandensis* strains express the heptose-containing CPS (Scott *et al*. 2013). Therefore, a heptosylated *B. thailandensis* protein could be used as a glycoconjugate vaccine against *B. pseudomallei*.

Our collaborators at the London School of Hygiene and Tropical Medicine (LSHTM) established the *pglL* mutant of *B. thailandensis* to confirm the native function of the PglL OTase in *B. thailandensis*, and also to investigate the potentiality of *B. thailandensis* PglL to produce glycoconjugate vaccines. A proteomic screening of the glycopeptides in *B. thailandensis* is a prerequisite to these research aims. As a result, a ZIC-HILIC method was modified from previous research to enrich the glycopeptides from the bacterial whole cell lysates. ZIC-HILIC is an LC method that separates substances based on their hydrophilicity, and its overall effect is similar to a normal phase LC column. Its detailed mechanisms are beyond the scope of this thesis. Fundamentally, the column is made by chemically coupling zwitterions to a stationary phase. When the inner surface of the column is exposed to a water-containing mobile phase, the water molecules are trapped by the zwitterions, forming a water layer next to the inner surface. As the glycopeptides are more hydrophilic than the un-modified peptides, when the whole cell lysate is injected into the ZIC-HILIC system that is running with an organic mobile phase, the glycopeptides are "grasped" by the water layer and the peptides are directly washed away with the mobile phase, potentially achieving the enrichment of the glycopeptides that can be further analysed by LC-ESI-MS.

5.2 Results and Discussion

This section contains MS-based proteomic/glycoproteomic analysis of three sets of samples. Prior to analysing the *B. thailandensis* whole cell lysate, the previously characterised *B. cenocepacia* whole cell lysates were investigated by our ZIC-HILIC LC-ESI-MS method to confirm the system was working properly (**Section 5.2.1**). The method was subsequently optimised to analyse the *B. thailandensis* whole cell lysate and periplasmic proteins (**Section 5.2.2**). A preliminary study of the *C. Burnetii* whole cell lysate based solely on LC-ESI-MS is shown in the last section (**Section 5.2.3**), which explores the possibility of screening for glycopeptides without a complicated multiple-dimensional LC separation. *C. Burnetii* is a Gram-negative bacterial pathogen causing Q fever. Genomic analysis indicated that it carries glycosylation systems, and therefore it was used as a model to explore the MS methods here. Its detailed biology is beyond the scope of the thesis, and was therefore not introduced in **Section 5.1**.

5.2.1 ZIC-HILIC LC-ESI-MS Analysis of *Burkholderia cenocepacia*

ZIC-HILIC is a relatively novel method used in enriching for bacterial glycopeptides. Previous work mainly relied on commercially available ZIC-HILIC resins to pack mini-columns. However, the resins were no longer supplied before research described in the thesis started. Therefore, we decided to use an LC system equipped with a pre-packed ZIC-HILIC column to carry out the enrichment of the glycopeptides.

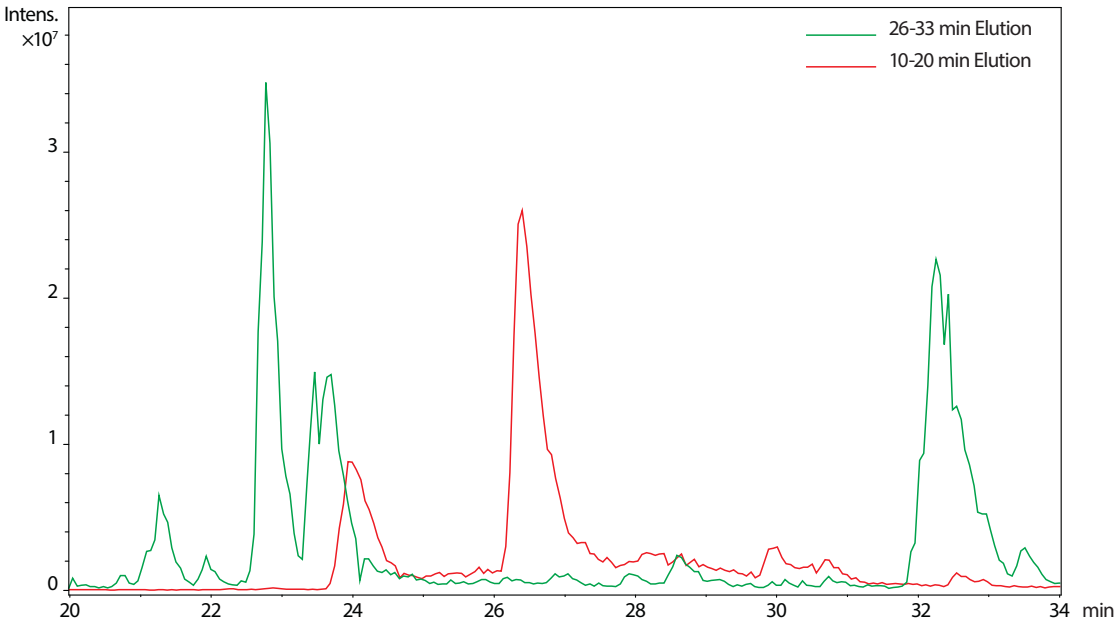


Figure 5-2 Liquid Chromatograms of the hydrophobic (10-20 min eluent from ZIC-HILIC, red line) and hydrophilic (26-33 min eluent from ZIC-HILIC, green line) eluents. An ESI-QIT MS instrument (amaZon Speed ETD) was used as a detector. Briefly, the LC was eluted with a gradient from 97/3 (v/v) solution A (0.1% formic acid in ultra-pure water) / solution B (0.1% formic acid in 80% aqueous solution of MeCN) to 1/99 (v/v) solutionA/B. More detailed LC programme can be found in the **Chapter 2**.

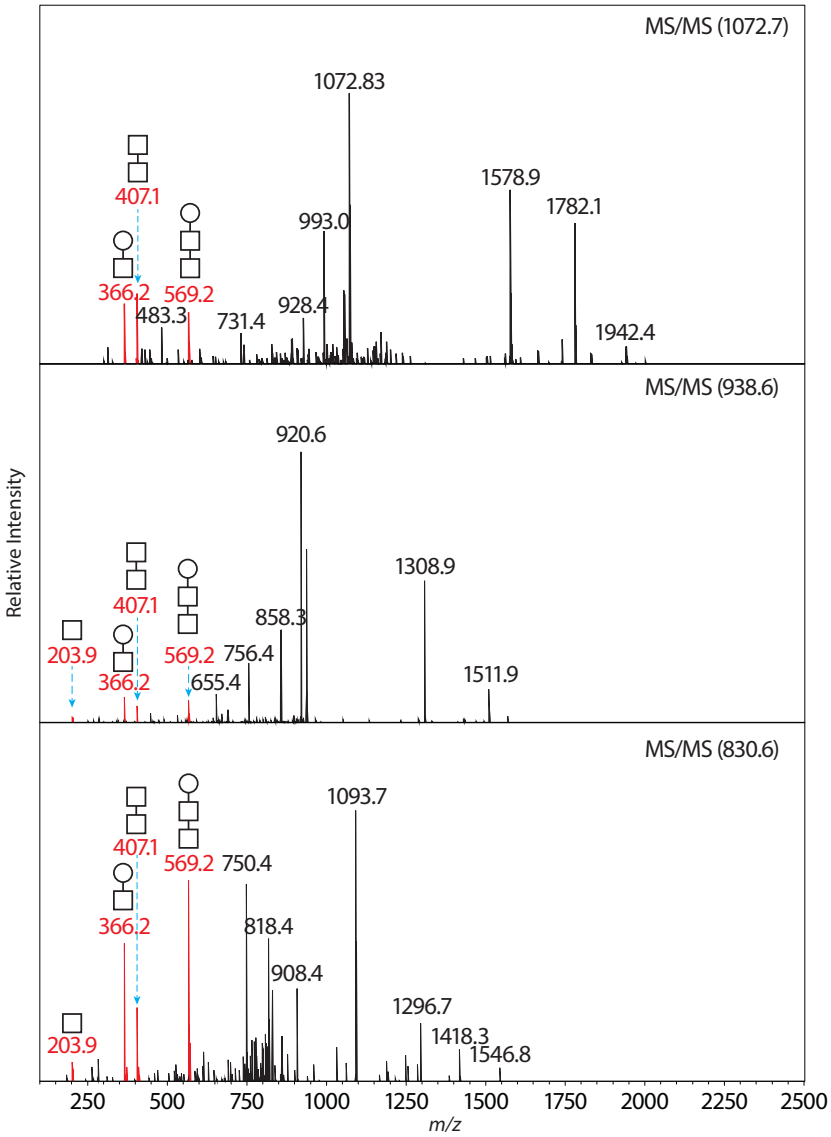


Figure 5-3 CID MS/MS analysis of the enriched glycopeptides. The MS peaks selected for MS/MS are shown in the up-right corner of each MS/MS spectra. The glycan peaks are coloured red and annotated with putative glycan structures. The squares stand for HexNAc and the circles stand for Hex. The peptide peaks are not annotated as the sequences of the glycopeptides could not be determined by low-energy CID MS/MS.

In order to test whether the ZIC-HILIC system was working properly, our collaborators prepared the previously characterised *B. cenocepacia* whole cell lysate, and we used the cell lysate as a standard sample. The glycopeptides in the samples were enriched by ZIC-HILIC first and then analysed by LC-ESI-QIT. More specifically, the whole cell lysate sample was dissolved in an 80% aqueous solution of MeCN containing 5% formic acid and injected into the ZIC-HILIC column. The same solution was used for washing away the unmodified peptides, while a 20% aqueous solution of MeCN containing 5% formic acid was used to elute the glycopeptides. The eluents were collected for the further LC-ESI-MS analysis (**FIG. 5-1**). **FIG. 5-1** shows the liquid chromatograms of the un-modified peptide (red line) and the glycopeptide (green line) eluents. It is clear from the chromatograms that the hydrophobic and hydrophilic compounds are contained in different eluents, indicating the glycopeptides were likely to be separated from the un-modified peptides.

In the hydrophilic portion of the whole cell lysate standard, three glycopeptides were readily identified based on the CID MS/MS fragmentation (**FIG. 5-3**). As the low-energy CID experiments carried out by a QIT instrument mainly cleave the glycosidic bonds instead of the peptide bonds, though the glycan structures were determined, the glycopeptides were not successfully sequenced. The peaks at *m/z* 204, 366, 407 and 569 in the spectra correspond to the residues of a HexNAc, Hex-HexNAc, di-HexNAc and Hex-Di-HexNAc respectively, which is fully consistent with the previously identified glycans carried by the *B. cenocepacia* glycopeptides.

As a trial experiment, the data were sufficient to prove that our ZIC-HILIC set-up is capable of enriching glycopeptides, and therefore we did not make further efforts to sequence these glycopeptides. However, it might be interesting to revisit these data in the future, as the glycopeptides we found are probably different from the previously identified ones.

5.2.2 ZIC-HILIC LC-ESI-MS Analysis of *Burkholderia thailandensis*

The same analytical strategies were subsequently used for analysing the whole cell lysates extracted from the WT and pglL mutant of *B. thailandensis*. To our surprise, though multiple *B. thailandensis* protein hits were indicated by the Mascot database (data not shown), no glycopeptide was found in both whole cell lysate samples. This could be due to a couple of reasons. First, the organism might not glycosylate its proteins under the culturing condition. Secondly, the whole cell lysate sample was too complex, and the ionisation of the hydrophilic glycopeptides was suppressed by the un-modified peptides.

In order to address whether glycopeptides might be suppressed in the complex protein mixtures, our collaborators extracted the periplasmic proteins for further MS analysis. The periplasmic proteins were chosen mainly because the OTase dependent glycosylation of the bacterial proteins occurs at the periplasmic space (see **FIG. 5-1**), where the concentration of the glycopeptides is the highest in principle.

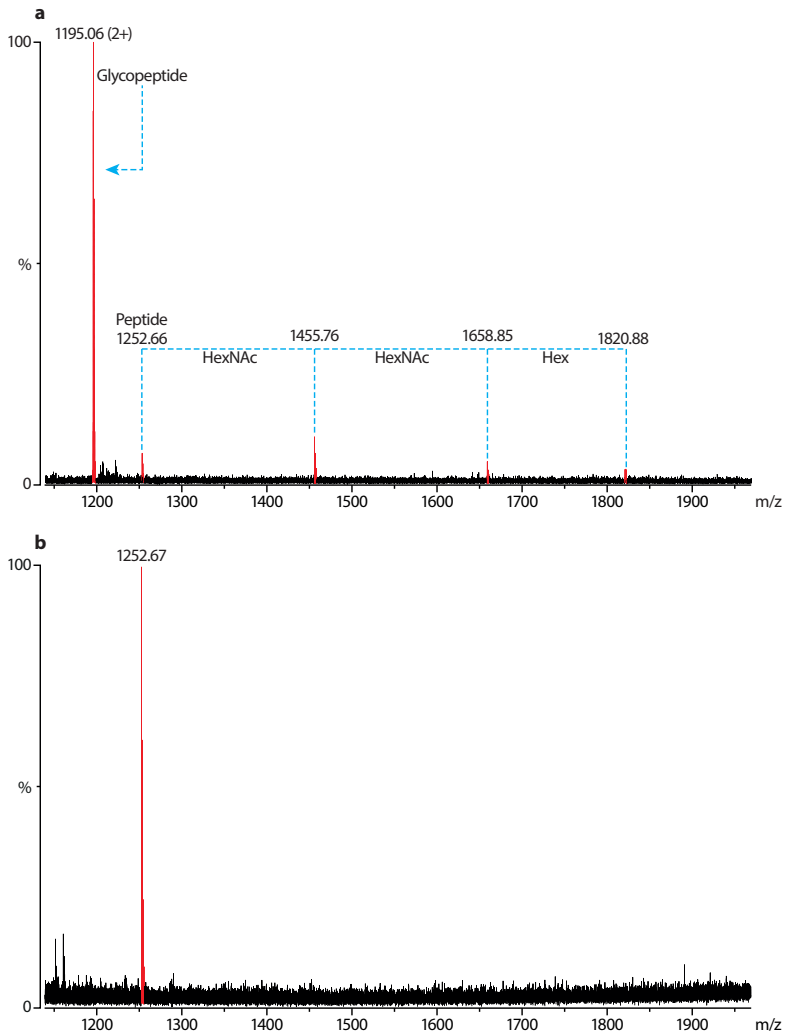


Figure 5-4 MS analysis of the WT (a) and pglL mutant (b) *B. thailandensis* periplasmic protein samples. The MS peaks corresponding to the glycopeptides and peptide are coloured red and annotated with *m/z*. Only the peak at *m/z* 1195.6 is doubly charged, and therefore a 2+ was shown in a bracket after the *m/z*. An in-source fragmentation was observed for the WT glycopeptide, and therefore the mass shifts corresponding to the HexNAc and Hex (annotated under the blue dash line) can be found in the spectrum (a).

One glycopeptide present in the WT *B. thailandensis* periplasmic protein sample but not the pglL mutant sample was identified based on LC-ESI-Q-TOF (**FIG. 5-4**). A strong MS peak at *m/z* 1195.06 corresponding to the doubly charged molecular ion of the glycopeptide was found in the MS spectrum of the WT *B. thailandensis* periplasmic protein sample (**FIG. 5-3 a**). An in source fragmentation of *m/z* 1195.06 [2+] was observed in the spectrum, *i.e.* the singly charged MS peak at *m/z* 1262.66 was assigned to the peptide backbone, and a cluster of MS peaks at *m/z* 1455.67, 1658.85 and 1820.88 corresponded to sequential sugar increments. The observed mass shifts between the peptide backbone and the glycopeptide fragments indicate the glycan carried by this glycopeptide is a Hex-HexNAc-HexNAc tri-saccharide, which is exactly the same as the reported *B. cenocepacia* glycan structure (Lithgow *et al.* 2014). In addition, the mass difference between the calculated mass of the glycopeptide (2388.12 Da) and the mass of the peptide backbone (1251.66 Da) is 1136.46 Da, corresponding to two tri-saccharides (two times 568 Da), indicating this peptide has two glycosylation sites.

The peak at *m/z* 1252.67 was observed in the MS spectra of the pglL mutant sample, while all the peaks corresponding to the glycopeptide and its fragments were absent. This strongly indicated that the pglL gene is involved in the glycosylation of the observed peptide.

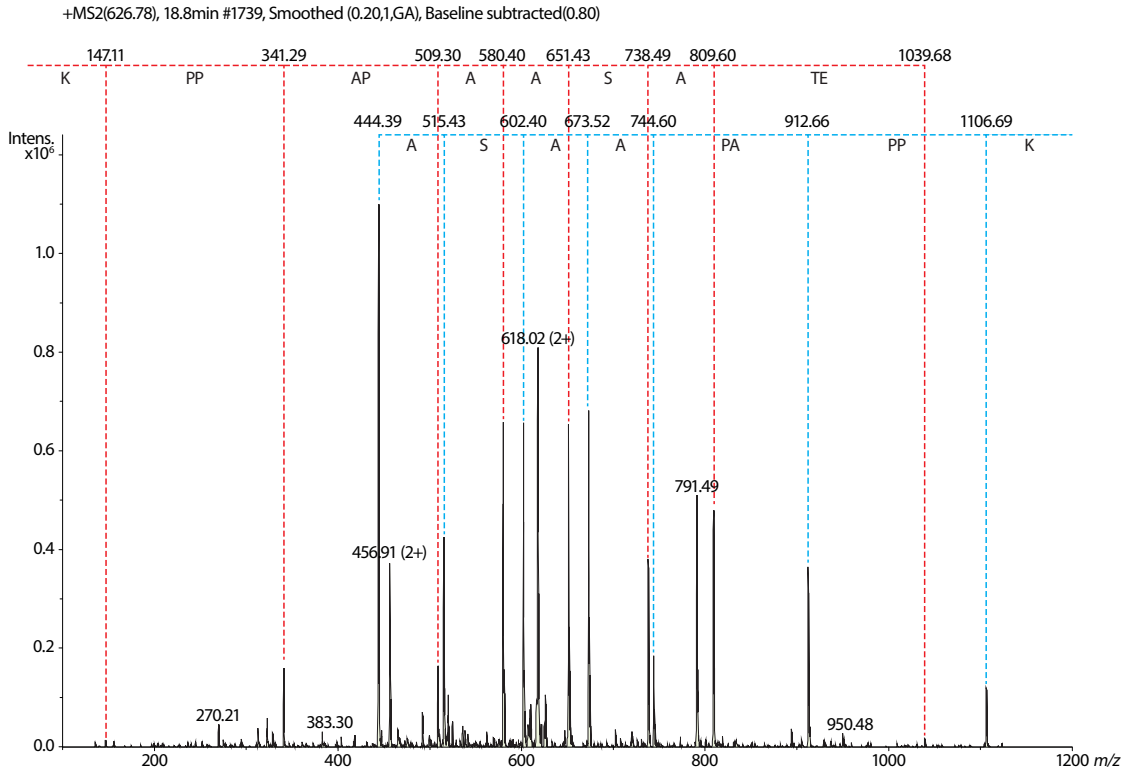


Figure 5-5 A CID MS/MS sequencing of the peptide backbone (the MS peak at *m/z* 1252.66). The red dash lines annotate the sequence from the C- to N- termini, and the blue dash lines annotated it from the N- to C-termini. The *m/z* values and the amino acids corresponding to the mass shifts are annotated above and under the dash lines, respectively. Both annotations strongly indicate the peptide contains an ASAAPA sequence and its C-terminus is K. The complete de-novo sequencing of this peptide was not possible mainly because of the low peak intensities at the low mass range area and the low collisional energy.

In order to confirm the identity of the glycopeptide, the peptide backbone was subjected to CID MS/MS analysis (**FIG. 5-5 A**) performed by an ESI-QIT instrument. The fragmentation pattern did not hit any protein in the Mascot database, and therefore the spectrum was manually interpreted. The partial sequencing based on CID MS/MS strongly indicated that the peptide contains an ASAAPA sequence and has a K at its C-terminus. Based on all the observed mass shifts, we proposed that the sequence of the peptide backbone is GRETASAAPAPPK. Note that a confirmative *de novo* sequencing of the peptide was not possible, which is mainly due to the limitations of the QIT instruments, *i.e.* the relatively low resolution, 1/3 cut–off rule and the low collisional energy.

The peptide backbone was further subjected to ETD MS/MS analysis (**FIG. 5-5 B**). The proposed sequence contains multiple prolines, which greatly prohibits the ETD processes (Mikesh *et al.* 2006). Consequently, the spectrum could not be fully understood. Nevertheless, the spectrum indicated the peptide contains a K at the C-terminus, and also contains S, which is consistent with the CID MS/MS analysis.

It is interesting to find that a manually interpretable MS/MS spectrum does not hit any protein in the Mascot database. This phenomenon is very common for several types of proteins, including the recombinant, not well-digested and modified proteins. Most recombinant proteins are not recorded by the database, and obviously cannot be matched. Some proteins are

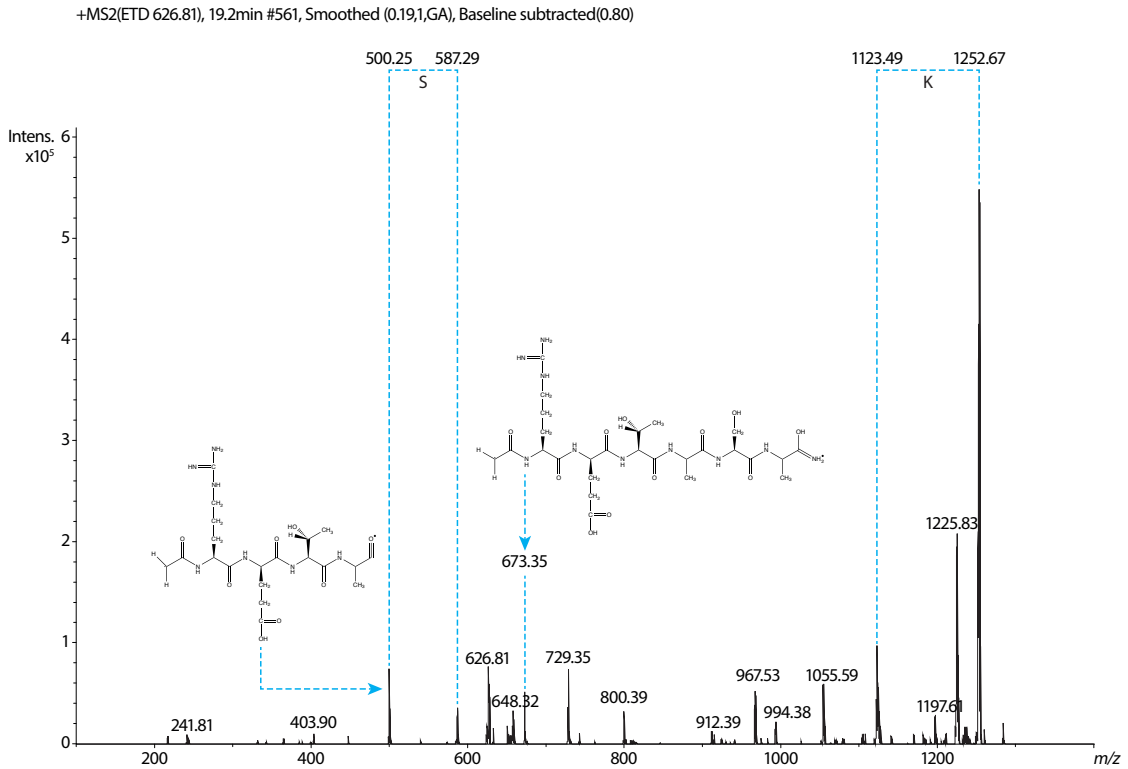


Figure 5-5 B ETD MS/MS sequencing of the peptide backbone (the MS peak at *m/z* 626.81[2+]). The mass shift at the high mass end indicates the peptide contain a K at the C-terminus. The spectrum is not fully understood, and partially annotated with some proposed peptide structures. Note that these peptide structures are 1 Da higher or lower than the observed peak, indicating there exist some proton transferring processes.

not well digested by certain proteases, and routine MS characterisations cannot discover their signals. The latter group of proteins generate abnormal fragmentation patterns when subjected to MS/MS analysis, which does not match information in the database. The recombinant Fap1 protein discussed in the previous chapter belongs to all of these three categories, and Mascot search was not able to identify it, as expected. We speculate that the peptide we observed here belongs to the third category. When manually checking the proteome of *B. thailandensis*, we noticed that an intracellular poly-3-hydroxybutyrate (PHB) depolymerase produces a peptide with a sequence of ETASAAPAK after trypsin digestion, which matches very well with our annotation. Note that the CID MS/MS data strongly support the peptide contains an ASAAPA…K sequence. The only difference between the depolymerase and the observed peptide is a 194 Da mass shift that could be attributed to a novel modification.

Previous research has shown that an opportunistic Gram-negative bacterial pathogen, *Acinetobacter baumannii*, predominantly O-glycosylates serine residues that are located after an alanine residue (Scott *et al.* 2014). This is consistent with our data indicating that an ASAA sequon is glycosylated in *B. thailandensis*, which could mean that there exist consensus sequences for the bacterial OTase dependent O-glycosylation. The consensus sequence can provide clues for the future characterisation of bacterial glycosylation systems, and it is vital to the logical design and mass production of the glycoconjugates vaccines. A mechanistic study of the O-OTase is needed to further support this proposal, and more importantly, a piece of determinant evidence could be offered by a 3D structure of the Bacterial O-OTases or even the Wzy_C domain.

Though extensive research efforts were made to find more glycopeptides in *B. thailandensis*, they were unsuccessful. It has been reported that some bacteria only glycosylate their proteins under challenging culturing condition, *e.g.* a low nutrition supply. Therefore, it is possible that though an OTase system is encoded in the genome of *B. thailandensis*, it was not activated to a "fully functional state" under the culturing condition. A similar phenomenon was observed for *V. Cholerae* whose glycosylation system is functionally transferable in *E. coli*, but MS-based glycoproteomics did not identify any glycopeptide (personal communication with N. Scott). In addition, the biological experiments that have now been performed by our collaborators suggest that the glycosylation system seems to be not essential or even not important for *B. thailandensis*. The pglL knock-out did not alter their growth rate and motility, though slightly reducing the ability to form biofilms. In fact, the mutation even increased its serum resistance and the macrophage survival rate. The importance of pglL seems to rely on increasing the adhesion ability, *e.g.* a dramatic reduction in adhering to the human alveolar epithelial cells was observed by our collaborators. The obvious candidate mediating the adhesion is pilin. However, the SDS-PAGE analysis of the pilin proteins indicated they are not glycosylated. All these facts could mean the O-OTase system in *B. thailandensis* is not very effectively functional.

It should also be noted that the analytical methods could be better optimised for screening glycopeptides based on the fact that the identified glycopeptide was not found in the whole cell lysate sample. Therefore, it is yet to be determined whether or not the organism glycosylates other proteins under the culturing conditions used.

5.2.3 Glycoproteomic Screening of the *Coxiella burnetii* Glycopeptides

When optimising the ZIC-HILIC method for *B. thailandensis*, we noticed that LC-ESI-MS was capable of detecting the glycopeptide before the enrichment (data not shown), which led us to propose that a simple LC-ESI-MS analysis could be sometimes sufficient for screening the glycopeptides even in a highly complex mixture.

We therefore used the *C. Burnetii* whole cell lysate as a standard to test our hypothesis. To our knowledge, no glycoproteomic work has been done before, and no glycosylation system was reported in this organism, though the genetic analysis carried out by our collaborators indicated putative glycan-related genes. Multiple glycopeptides were readily detected after LC-ESI-MS, and all of them were shown to carry a single HexNAc (CID MS/MS spectra of some of the glycopeptides are shown in **FIG. 5-6**).

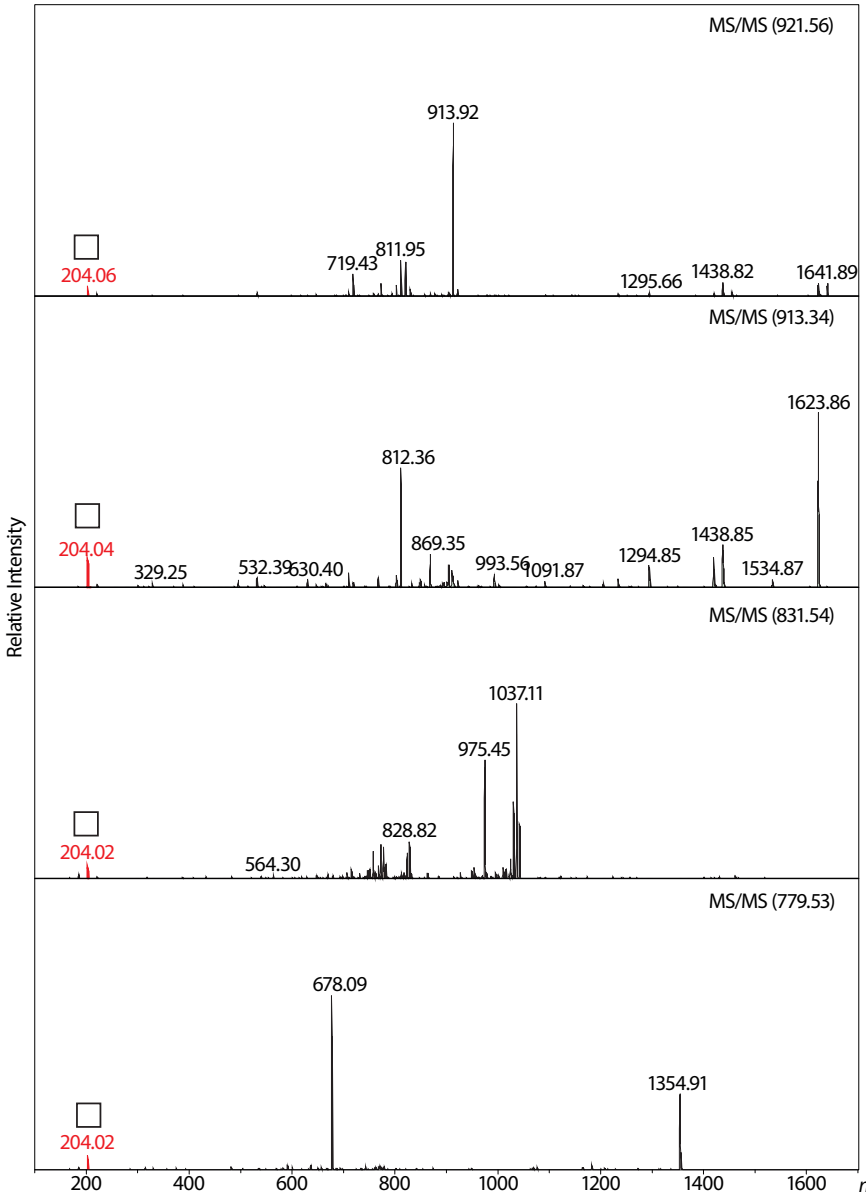


Figure 5-6 CID MS/MS analysis of the glycopeptides found in the *C. Burnetii* whole cell lysate. More detailed information can be found in the figure legend of **FIG.5-3**.

The limitations of the QIT instrument again constrained our abilities to sequence the glycopeptide, *e.g.* it is obvious from **FIG. 5-6** that most peptides were not fully fragmented. As a result, *de novo* sequencing was not possible based on these MS/MS spectra, and the fragmentation pattern did not match any protein in the Mascot database. Nevertheless, we managed to extract some information from these spectra. Most importantly, the MS/MS analysis is sufficient to confirm these glycopeptides carry HexNAc, based on the cluster of characteristic peaks at *m/z* 138, 144, 168 and 186 corresponding to the fragments of HexNAc (the MS/MS spectra here were not annotated with these peaks as they have low intensities; see **Section 4.3** for an explanation of how they are produced). In addition, glycopeptides carrying more than one HexNAc were observed. For example, the peptide at *m/z* 921 probably carries three HexNAc residues. Moreover, an apparent water loss was observed, *e.g.* the MS/MS analysis strongly indicated that the MS peak at *m/z* 913 is generated when the MS peak at *m/z* 922 loses a water molecule, and they probably share the same sequence.

The MS3 analysis was carried out in order to further characterise these glycopeptides. Most of the MS3 experiments were successfully performed, and good quality MS3 spectra were obtained. An automatic sequencing of the peptides based on the MS3 analysis is not supported by the Mascot database, so attempts were made to *de novo* sequence the glycopeptides.

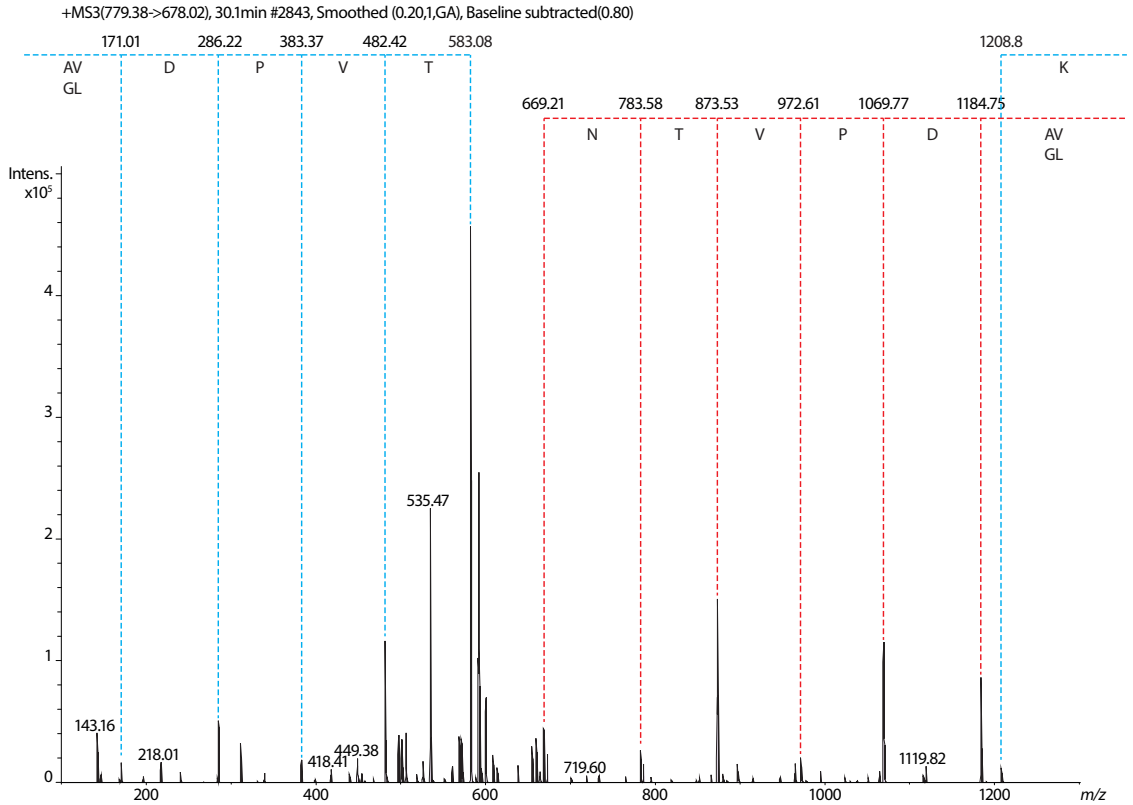


Figure 5-7 CID MS3 sequencing of the peptide backbone (the MS/MS peak at *m/z* 678.02 [2+]). The fragmentation pattern indicates an AV(GL)DPVT sequence at the N-terminus of the glycopeptide, and the C-terminal amino acid is K. more detailed information can be found in the figure legend of **FIG. 5-4**.

Unfortunately, most of the observed MS3 fragments possibly corresponded to the double-cleavage peptide structures, which makes the sequencing work utterly difficult. We managed to partially sequence some of the glycopeptides, and an example is shown in **FIG. 5-7**. The spectrum indicated an AV(GL)DPVT sequence at the N-terminus of the glycopeptide that contains a K at its C-terminus. This proposed sequence contains a T residue where the HexNAcylation probably occurs.

Because of the failure to completely sequence the glycopeptides, our data cannot provide concrete evidence to confirm a *C. Burnetii* glycosylation system. As the Mascot quires returned more than seven hundred protein hits, the majority of which are from Coxiella, it is highly possible that the discovered glycopeptides are derived from these Coxiella proteins. However, contaminating proteins were also suggested to exist in the *C. Burnetii* sample according to the Mascot search, among which was the hen egg white lysozyme protein. This enzyme was proved to be a glycosylated glycosidase (Nakamura *et al*. 1993; Asperger *et al*. 2015; Toshima *et al*. 2003). In addition to being contaminations, the existence of such proteins could hydrolyse the glycans carried by the glycoproteins in the samples. This may greatly affect the glycan structures, and in turn our MS characterisations, *i.e.* we were unable to exclude the possibility that longer glycan structures stretching out of HexNAc had been cleaved. It will be very interesting to repeat the MS analysis on a freshly made *C. Burnetii* sample that is more strictly purified.

5.3 Further Work

Because of the limited time and resources such as instrumentations, though the LC-MS-based analytical strategies were established to enrich and characterise the bacterial glycopeptides, the diagnostic data that unambiguously confirm the *in vivo* function of the *B. thailandensis* O-OTase system were not successfully obtained. Several novel experiments could provide the definitive data, proving or disproving the function of the *B. thailandensis* O-OTase, in addition to the obvious option of tedious 2D-Gel experiments.

First of all, a glyco-fishing experiment can be done. The glyco-fishing experiment is similar to the processes for producing the glycoconjugate vaccines. Fundamentally, a histidine-tagged protein that contains the ASAA sequon can be expressed in *B. thailandensis*. This protein can be easily purified due to the histidine-tag, and a simple glycoproteomic analysis can in principle confirm whether or not the *B. thailandensis* O-OTase is functional.

Secondly, a metabolic labelling method can be used for identifying the glycopeptides. More specifically, analogues to HexNAc can be used for culturing *B. thailandensis*. These compounds carry the functional groups on which the orthogonal chemistry can be done. If these compounds are incorporated into the glycopeptides by *B. thailandensis*, they can be used as "tags" to enrich the glycopeptides. In addition, they can also be used for creating chromophores in order to observe the *in vivo* spatial distributions and functions of the glycopeptides.

Both of these two methods above in fact make efforts to "tag" the glycoproteins so that they can be easily purified. They are designed to overcome one of the main challenges met in the current research that the signals of the glycopeptides tend to be suppressed in a complex mixture. As long as the glycoproteins can be effectively enriched from the mixture, the following MS characterisations and data interpretations will in principle get significantly simpler.

Thirdly, the same LC-MS strategies can be used for analysing the *B. pseudomallei* whole cell lysate. Because it is a human pathogen causing severe diseases, the characterisation of its O-OTase system can be more valuable. In addition, these experiments will also help test whether or not the enrichment strategies are effective.

Final Remarks

Bacterial glycoconjugates are of vital biological and medical importance; as they mediate bacteria and host interactions, they are the targets of many antibiotics and they themselves can be used as vaccines.

All bacterial cells are encapsulated by glycoconjugates: monosaccharides are the major component of both peptidoglycans and LPSs that are constituent element of bacterial surfaces. They are deeply implicated in improving colonisation, resisting antimicrobial peptides and modulating host immune signallings. In addition, because of the essentiality of bacterial glycoconjugates to bacteria, since the discovery of penicillin, β-lactam antibiotics have been designed and synthesised in order to inhibit their biosynthesis. Moreover, due to the recent characterisation of bacterial OTases and the advancement in glycoengineering, recombinantly synthesised glycoconjugates are being emerged as a reliable source of vaccines.

Extensive research efforts were made to elucidate the structures of glycoconjugates from all three domains of life during the past four to five decades, and research described in this thesis is essentially the consequence of summarising, following and slightly improving these academic heritages. This piece of PhD research is directly devoted to structural analysis of various bacterial glycoconjugates, which undoubtedly contributes to both comprehensively explaining bacterial pathogenicity and biochemically investigating bacterial glycosylation machinery. However, its ultimately purpose is to establish systematic and universally applicable strategies to structurally determining any glycoconjugates.

A number of success was achieved. For instance, highly sensitive and facile chemical-MS methods were developed to analyse lipid As and LPSs, a novel micelle NMR technique was investigated to characterise intact glycolipid molecules, systematic glycomic and glycoproteomic strategies were assembled to de novo elucidate the glycosylation pathway of serine-rich protein, and an LC-MS glycoproteomic procedure was recruited to screen bacterial glycopeptides with a high-throughput.

Even so, the ultimate goal of this thesis is still far from being achieved. For instance, it lacks methods to efficiently dissociate proteins during top-down analysis; quantification of proteins and glycans under a biological context largely relies on isotopic labelling, which can be expensive when processing a large amount of samples; the stereochemistry of glycans is difficult to be determined with a high sensitivity; and it is still highly challenging to enrich the modified proteins from unmodified proteins. It seems ground-breaking advancements in fundamental biophysics and biochemistry are still being waited for to solve these issues.

Final Remarks

It is widely accepted today that a biological problem needs to be solved from both structural and functional aspects. Unfortunately, the biological roles that glycoconjugates are playing in not only prokaryotes but also eukaryotes are even worse established than the analytical techniques. It is not abnormal that the structure of a glycoconjugate has been almost fully determined, whereas its function remains elusive.

As a matter of fact, an unambiguous function is often difficult to be assigned to a single glycan at molecular level, even though biological implications are indicated by phenotypes caused by genomic mutations. Taking glycoproteins for example, as proteins tend to be glycosylated at multiple sites, and it is nearly impossible to only mutate one glycan without affecting the protein backbone and other glycans; any observed phenotype cannot be exclusively attributed to a single glycan. In addition, the biological function of glycans is largely based on lectin-glycan recognition. However, the lectin-glycan bindings tend to be weak interactions *e.g.* it is drastically lower than antibody-peptide binding. As a result, it is usually the multivalent interactions that support the lectin-glycan binding, which means multiple glycans and lectins are mediating a single biological function.

Because of all the difficulties mentioned above, glycobiology especially bacterial glycobiology is still highly challenging relative to other sophisticated biological fields. Nevertheless, it is the challenges that guarantee the interest of glycobiology, just like a high risk often suggests a high profit.

References

Acquotti, D. & Sonnino, S., 2000. Use of nuclear magnetic resonance spectroscopy in evaluation of ganglioside structure, conformation, and dynamics. *Methods in Enzymology*, 312, pp.247–272.

Albersheim, P. *et al*., 1967. A method for the analysis of sugars in plant cell-wall polysaccharides by gas-liquid chromatography. *Carbohydrate Research,* 5, pp.340–345.

Alexander, D.C. & Valvano, M.A., 1994. Role of the *rfe* gene in the biosynthesis of the *Escherichia coli* O7-specific lipopolysaccharide and other O-specific polysaccharides containing N-acetylglucosamine. *Journal of Bacteriology*, 176(22), pp.7079–7084.

Altman, E. *et al*., 2012. Design and immunological properties of *Helicobacter pylori* glycoconjugates based on a truncated lipopolysaccharide lacking Lewis antigen and comprising an α-1,6-glucan chain. *Vaccine*, 30(50), pp.7332–7341.

Altman, E. *et al*., 2008. Effect of the HP0159 ORF mutation on the lipopolysaccharide structure and colonizing ability of *Helicobacter pylori*. *FEMS Immunology & Medical Microbiology*, 53(2), pp.204–213.

Altman, E. *et al*., 2013. Lipopolysaccharide structure of *Helicobacter pylori* serogroup O:3. *Carbohydrate Research*, 378, pp.139–143.

Altman, E., Chandan, V., Li, J. & Vinogradov, E., 2011a. A reinvestigation of the lipopolysaccharide structure of *Helicobacter pylori* strain Sydney (SS1). *FEBS Journal*, 278(18), pp.3484–3493.

Altman, E., Chandan, V., Li, J. & Vinogradov, E., 2011b. Lipopolysaccharide structures of *Helicobacter pylori* wild-type strain 26695 and 26695 HP0826::Kan mutant devoid of the O-chain polysaccharide component. *Carbohydrate Research*, 346(15), pp.2437–2444.

Amor, K. *et al*., 2000. Distribution of core oligosaccharide types in lipopolysaccharides from *Escherichia coli*. *Infection and Immunity*, 68(3), pp.1116–1124.

Amor, P.A. & Whitfield, C., 1997. Molecular and functional analysis of genes required for expression of group IB K antigens in *Escherichia coli*: Characterization of the his-region containing gene clusters for multiple cell-surface polysaccharides. *Molecular Microbiology*, 26(1), pp.145–161.

Anderson, M.S. & Raetz, C.R.H., 1987. Biosynthesis of Lipid A Precursors in *Escherichia Coli*. A Cytoplasmic Acyltransferase That Converts UDP-N-Acetylglucosamine to UDP-3-

O-(*R*-3-Hydroxymyristoyl)-N-Acetylglucosamine. *Journal of Biological Chemistry*, 262(11), pp.5159–5169.

Anderson, M.S., Robertson, A.D. & Macher, I., 1988. Biosynthesis of Lipid A in *Escherichia coli*: Identification of UDP-3-*O*-[(*R*)-3-hydroxymyristoyl]-α-D-glucosamine as a Precursor of UDP-N^2,O^3-Bis[(*R*)-3-Hydroxymyristoyl]-α-D-glucosamine. *Biochemistry*, 27(6), pp.1908–1917.

Appelmelk, B.J. *et al*., 2000. Phase Variation in H Type I and Lewis a Epitopes of *Helicobacter pylori* Lipopolysaccharide. *Infection and Immunity*, 68(10), pp.5928–5932.

Arora, S.K. *et al*., 2005. Role of motility and flagellin glycosylation in the pathogenesis of *Pseudomonas aeruginosa* burn wound infections. *Infection and Immunity*, 73(7), pp.4395–4398.

Asperger, A. *et al*., 2015. Low Abundant N-linked Glycosylation in Hen Egg White Lysozyme Is Localized at Nonconsensus Sites. *Journal of Proteome Research*, 14(6), pp.2633–2641.

Avci, F.Y. *et al*., 2011. A mechanism for glycoconjugate vaccine activation of the adaptive immune system and its implications for vaccine design. *Nature Medicine*, 17(12), pp.1602–1609.

Babinski, K.J., Ribeiro, A.A. & Raetz, C.R.H., 2002. The *Escherichia coli* gene encoding the UDP-2,3-diacylglucosamine pyrophosphatase of lipid A biosynthesis. *Journal of Biological Chemistry*, 277(29), pp.25937–25946.

Barrow, K. & Kwon, D.H., 2009. Alterations in two-component regulatory systems of phoPQ and pmrAB are associated with polymyxin B resistance in clinical isolates of *Pseudomonas aeruginosa*. *Antimicrobial agents and chemotherapy*, 53(12), pp.5150–5154.

Batchelor, R.A. *et al*., 1991. Regulation by a novel protein of the bimodal distribution of lipopolysaccharide in the outer membrane of *Escherichia coli*. *Journal of Bacteriology*, 173(18), pp.5699–5704.

Becker, M.R. *et al*., 2002. Molecular analysis of bacterial species associated with childhood caries. *Journal of Clinical Microbiology*, 40(3), pp.1001–1009.

Benz, I. & Schmidt, M.A., 1989. Cloning and expression of an adhesin (AIDA-I) involved in diffuse adherence of enteropathogenic *Escherichia coli*. *Infection and Immunity*, 57(5), pp.1506–1511.

Beutler, B., 2003. Innate immune responses to microbial poisons: discovery and function of the Toll-like receptors. *Annual Review of Pharmacology and Toxicology*, 43, pp.609–628.

Bhat, R. *et al*., 1990. Structural studies of lipid A from *Pseudomonas aeruginosa* PAO1: occurrence of 4-amino-4-deoxyarabinose. *Journal of Bacteriology*, 172(12), pp.6631–6636.

Bishop, R.E. *et al*., 2000. Transfer of palmitate from phospholipids to lipid A in outer membranes of Gram-negative bacteria. *The EMBO journal*, 19(19), pp.5071–5080.

Bonis, M. *et al*., 2010. A M23B family metallopeptidase of *Helicobacter pylori* required for cell shape, pole formation and virulence. *Molecular Microbiology*, 78(4), pp.809–819.

Boren, T. *et al*., 1993. Attachment of *Helicobacter pylori* to Human Gastric Epithelium Mediated by Blood-Group Antigens. *Science*, 262(5141), pp.1892–1895.

Bowyer, A. *et al*., 2011. Characterization of interactions between LPS transport proteins of the Lpt system. *Biochemical and biophysical research communications*, 404(4), pp.1093–1098.

Breazeale, S.D., Ribeiro, A.A. & Raetz, C.R.H., 2003. Origin of lipid A species modified with 4-amino-4-deoxy-L-arabinose in polymyxin-resistant mutants of *Escherichia coli*. An aminotransferase (ArnB) that generates UDP-4-deoxyl-L-arabinose. *Journal of Biological Chemistry*, 278(27), pp.24731–24739.

Bridge, D.R. & Merrell, D.S., 2013. Polymorphism in the *Helicobacter pylori* CagA and VacA toxins and disease. *Gut Microbes*, 4(2), pp.101–117.

Brozek, K.A. & Raetz, C.R.H., 1990. Biosynthesis of Lipid A in *Escherichia coli*. Acyl Carrier Protein-Dependent Incorporation of Laurate and Myristate. *Journal of Biological Chemistry*, 265(26), pp.15410–15417.

Bu, S. *et al*., 2008. Interaction between Two Putative Glycosyltransferases Is Required for Glycosylation of a Serine-Rich Streptococcal Adhesin. *Journal of Bacteriology*, 190(4), pp.1256–1266.

Bystrova, O.V. *et al*., 2004. Full structure of the lipopolysaccharide of *Pseudomonas aeruginosa* immunotype 5. *Biochemistry. Biokhimiya*, 69(2), pp.170–175.

Cameron, A. E. & Eggers, D. F., Jr. An Ion "Velocitron". *Review of Scientific Instruments*, 19, 605–607 (1948).

Caroff, M., Tacken, A. & Szabo, L., 1988. Detergent-Accelerated Hydrolysis of Bacterial Endotoxins and Determination of the Anomeric Configuration of the Glycosyl Phosphate Present in the "Isolated Lipid A" Fragment of the *Bordetella Pertussis* Endotoxin. *Carbohydrate Research*, 175(2), pp.273–282.

Carty, S.M., Sreekumar, K.R. & Raetz, C.R.H., 1999. Effect of cold shock on lipid A biosynthesis in *Escherichia coli*. Induction At 12 degrees C of an acyltransferase specific for palmitoleoyl-acyl carrier protein. *Journal of Biological Chemistry*, 274(14), pp.9677–9685.

References

Castelli, M.E., García Véscovi, E. & Soncini, F.C., 2000. The phosphatase activity is the target for Mg^{2+} regulation of the sensor protein PhoQ in *Salmonella*. *Journal of Biological Chemistry*, 275(30), pp.22948–22954.

Castric, P., 1995. pilO, a gene required for glycosylation of *Pseudomonas aeruginosa* 1244 pilin. *Microbiology (Reading, England)*, 141 (Pt 5)(5), pp.1247–1254.

Chaze, T. *et al*., 2014. O-Glycosylation of the N-terminal region of the serine-rich adhesin Srr1 of *Streptococcus* agalactiae explored by mass spectrometry. *Molecular & Cellular Proteomics*, 13(9), pp.2168–2182.

Choi, K.-J. *et al*., 2010. The Actinobacillus pleuropneumoniae HMW1C-like glycosyltransferase mediates N-linked glycosylation of the *Haemophilus influenzae* HMW1 adhesin. *PLoS ONE*, 5(12), p.e15888.

Ciocchini, A.E. *et al*., 2013. Development and validation of a novel diagnostic test for human brucellosis using a glyco-engineered antigen coupled to magnetic beads. *PLoS neglected tropical diseases*, 7(2), p.e2048.

Clements, A. *et al*., 2007. Secondary acylation of *Klebsiella pneumoniae* lipopolysaccharide contributes to sensitivity to antibacterial peptides. *Journal of Biological Chemistry*, 282(21), pp.15569–15577.

Clementz, T. & Raetz, C.R.H., 1991. A gene coding for 3-deoxy-D-manno-octulosonic-acid transferase in *Escherichia coli*. Identification, mapping, cloning, and sequencing. *Journal of Biological Chemistry*, 266(15), pp.9687–9696.

Cole, R.B., 2011. Electrospray and MALDI Mass Spectrometry 2nd ed. R. B. Cole, John Wiley & Sons, Inc.

Crowell, D.N., Anderson, M.S. & Raetz, C.R.H., 1986. Molecular cloning of the genes for lipid A disaccharide synthase and UDP-N-acetylglucosamine acyltransferase in *Escherichia coli*. *Journal of Bacteriology*, 168(1), pp.152–159.

Cuccui, J. *et al*., 2013. Exploitation of bacterial N-linked glycosylation to develop a novel recombinant glycoconjugate vaccine against Francisella tularensis. *Open biology*, 3(5), p.130002.

Cullen, T.W. *et al*., 2011. *Helicobacter pylori* versus the host: remodeling of the bacterial outer membrane is required for survival in the gastric mucosa. N. Salama, ed. *PLoS Pathogens*, 7(12), p.e1002454.

Dawson, J.H.J. & Guilhaus, M., 1989. Orthogonal-acceleration Time-of-flight Mass Spectrometer. *Rapid Communications in Mass Spectrometry*, 3(5), pp.155–159.

De Gregorio, E. & Rappuoli, R., 2014. From empiricism to rational design: a personal perspective of the evolution of vaccine development. *Nature Reviews Immunology*, 14(7), pp.505–514.

de Kievit, T.R., 2009. Quorum sensing in *Pseudomonas aeruginosa* biofilms. *Environmental microbiology*, 11(2), pp.279–288.

Dell, A. *et al.*, 1983. Fast-Atom-Bombardment Mass-Spectrometry for Carbohydrate-Structure Determination. *Carbohydrate Research*, 115, pp.41–52.

Dempster, A.J., 1918. A new Method of Positive Ray Analysis. *Physical Review*, 11(4), pp.316–325.

Deringer, J.R. *et al.*, 2011. Immunoreactive Coxiella burnetii Nine Mile proteins separated by 2D electrophoresis and identified by tandem mass spectrometry. *Microbiology*, 157(Pt 2), pp.526–542.

Doerrler, W.T. & Raetz, C.R.H., 2002. ATPase activity of the MsbA lipid flippase of *Escherichia coli*. *Journal of Biological Chemistry*, 277(39), pp.36697–36705.

Doerrler, W.T., Reedy, M.C. & Raetz, C.R.H., 2001. An *Escherichia coli* mutant defective in lipid export. *Journal of Biological Chemistry*, 276(15), pp.11461–11464.

Dole, M. et al., 1968. Molecular Beams of Macroions. *The Journal of Chemical Physics*, 49, pp.2240–2249.

Dong, H. *et al.*, 2014. Structural basis for outer membrane lipopolysaccharide insertion. *Nature*, 511(7507), pp.52–56.

Eaton, K.A. *et al.*, 1991. Essential Role of Urease in Pathogenesis of Gastritis Induced by *Helicobacter pylori* in Gnotobiotic Piglets. *Infection and Immunity*, 59(7), pp.2470–2475.

Erickson, P.R. & Herzberg, M.C., 1993. Evidence for the Covalent Linkage of Carbohydrate Polymers to a Glycoprotein From *Streptococcus sanguis*. *Journal of Biological Chemistry*, 268(32), pp.23780–23783.

Espitia, C. & Mancilla, R., 1989. Identification, Isolation and Partial Characterization of Mycobacterium-Tuberculosis Glycoprotein Antigens. *Clinical and Experimental Immunology*, 77(3), pp.378–383.

Evans, D.J. & Evans, D.G., 2000. *Helicobacter pylori* adhesins: review and perspectives. *Helicobacter*, 5(4), pp.183–195.

Falush, D. *et al.*, 2003. Traces of human migrations in *Helicobacter pylori* populations. *Science*,

299(5612), pp.1582–1585.

Feldman, M.F. *et al*., 2005. Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 102(8), pp.3016–3021.

Fenn, J.B. *et al*., 1989. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science*, 246(4926), pp.64–71.

Fives-Taylor, P.M. & Thompson, D.W., 1985. Surface properties of *Streptococcus sanguis* FW213 mutants nonadherent to saliva-coated hydroxyapatite. *Infection and Immunity*, 47(3), pp.752–759.

Froeliger, E.H. & Fives-Taylor, P., 2001. *Streptococcus parasanguis* fimbria-associated adhesin fap1 is required for biofilm formation. *Infection and Immunity*, 69(4), pp.2512–2519.

Garnett, J.A. *et al*., 2012. Structural insight into the role of *Streptococcus parasanguinis* Fap1 within oral biofilm formation. *Biochemical and biophysical research communications*, 417(1), pp.421–426.

Garrett, T.A., Kadrmas, J.L. & Raetz, C.R.H., 1997. Identification of the gene encoding the *Escherichia coli* lipid a 4′-kinase. Facile phosphorylation of endotoxin analogs with recombinant LpxK. *Journal of Biological Chemistry*, 272(35), pp.21855–21864.

Ge, Z. *et al*., 1997. Cloning and Heterologous Expression of an α1,3-Fucosyltransferase Gene from the Gastric Pathogen *Helicobacter pylori*. *Journal of Biological Chemistry*, 272(34), pp.21357–21363.

Gebhart, C. *et al*., 2012. Characterization of exogenous bacterial oligosaccharyltransferases in *Escherichia coli* reveals the potential for O-linked protein glycosylation in *Vibrio cholerae* and *Burkholderia thailandensis*. *Glycobiology*, 22(7), pp.962–974.

Gellatly, S.L. & Hancock, R.E.W., 2013. *Pseudomonas aeruginosa*: new insights into pathogenesis and host defenses. *Pathogens and disease*, 67(3), pp.159–173.

Gibbons, H.S. *et al*., 2000. Oxygen Requirement for the Biosynthesis of the *S*-2-Hydroxymyristate Moiety in *Salmonella typhimurium* Lipid A. Function of LpxO, A new Fe^{2+}/α-Ketoglutarate-dependent Dioxygenase Homologue. *Journal of Biological Chemistry*, 275(42), pp.32940–32949.

Glass, M.B. *et al*., 2006. Pneumonia and septicemia caused by *Burkholderia thailandensis* in the United States. *Journal of Clinical Microbiology*, 44(12), pp.4601–4604.

Grass, S. *et al*., 2010. The *Haemophilus influenzae* HMW1C protein is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin. *PLoS Pathogens*, 6(5), p.e1000919.

Greenfield, L.K. *et al*., 2012. Biosynthesis of the polymannose lipopolysaccharide O-antigens from *Escherichia coli* serotypes O8 and O9a requires a unique combination of single- and multiple-active site mannosyltransferases. *Journal of Biological Chemistry*, 287(42), pp.35078–35091.

Grijalva, C.G. *et al*., 2007. Decline in pneumonia admissions after routine childhood immunisation with *pneumococcal* conjugate vaccine in the USA: a time-series analysis. *The Lancet*.

Gringhuis, S.I. *et al*., 2009. Carbohydrate-specific signaling through the DC-SIGN signalosome tailors immunity to *Mycobacterium tuberculosis*, HIV-1 and *Helicobacter pylori*. *Nature Immunology*, 10(10), pp.1081–1088.

Groisman, E.A., 2001. The pleiotropic two-component regulatory system PhoP-PhoQ. *Journal of Bacteriology*, 183(6), pp.1835–1842.

Gross, J. *et al*., 2008. The *Haemophilus influenzae* HMW1 adhesin is a glycoprotein with an unusual N-linked carbohydrate modification. *Journal of Biological Chemistry*, 283(38), pp.26010–26015.

Gunn, J.S., Belden, W.J. & Miller, S.I., 1998. Identification of PhoP-PhoQ activated genes within a duplicated region of the *Salmonella typhimurium* chromosome. *Microbial Pathogenesis*, 25(2), pp.77–90.

Gunn, J.S., Lim, K.B., *et al*., 1998. PmrA-PmrB-regulated genes necessary for 4-aminoarabinose lipid A modification and polymyxin resistance. *Molecular Microbiology*, 27(6), pp.1171–1182.

Guo, L. *et al*., 1998. Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. *Cell*, 95(2), pp.189–198.

Haraga, A. *et al*., 2008. *Burkholderia thailandensis* as a model system for the study of the virulence-associated type III secretion system of *Burkholderia pseudomallei*. *Infection and Immunity*, 76(11), pp.5402–5411.

Haslam, S., North, S. & Dell, A., 2006. Mass spectrometric analysis of N- and O-glycosylation of tissues and cells. *Current Opinion in Structural Biology*, 16(5), pp.584–591.

Haslam, S.M. *et al*., 1997. Characterisation of the phosphorylcholine-containing N-linked oligosaccharides in the excretory-secretory 62 kDa glycoprotein of *Acanthocheilonema viteae*.

Molecular and biochemical parasitology, 85(1), pp.53–66.

Hauser, A.R., 2009. The type III secretion system of *Pseudomonas aeruginosa*: infection by injection. *Nature Publishing Group*, 7(9), pp.654–665.

Heinrichs, D.E. *et al*., 1998. The assembly system for the outer core portion of R1- and R4-type lipopolysaccharides of *Escherichia coli*. The R1 core-specific beta-glucosyltransferase provides a novel attachment site for O-polysaccharides. *Journal of Biological Chemistry*, 273(45), pp.29497–29505.

Hiratsuka, K. *et al*., 2005. Identification of a D-glycero-D-manno-Heptosyltransferase Gene from *Helicobacter pylori*. *Journal of Bacteriology*, 187(15), pp.5156–5165.

Hoang, T.T. *et al*., 1998. A broad-host-range Flp-FRT recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked Pseudomonas aeruginosa mutants. *Gene* 212, pp.77-86.

Hoffmann, E. de, 2007. *Mass Spectrometry* Third Edition, John Wiley & Sons Ltd.

Hug, I. *et al*., 2010. *Helicobacter pylori* Lipopolysaccharide Is Synthesized via a Novel Pathway with an Evolutionary Connection to Protein N-Glycosylation J. N. Weiser, ed. *PLoS Pathogens*, 6(3), p.e1000819.

Hyams, C. *et al*., 2010. The *Streptococcus pneumoniae* capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. *Infection and Immunity*, 78(2), pp.704–715.

Ilver, D. *et al*., 1998. *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science*, 279(5349), pp.373–377.

Ip, Y.T. *et al*., 1993. Dif, a dorsal-related gene that mediates an immune response in *Drosophila*. *Cell*, 75(4), pp.753–763.

Iribarne, J.V. & Thomson, B.A., 1976. On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics*, 64(6), pp.2287–9.

Isshiki, Y., Zähringer, U. & Kawahara, K., 2003. Structure of the core-oligosaccharide with a characteristic D-glycero-α-D-talo-oct-2-ulosylonate-(2→4)-3-deoxy-D-manno-oct-2-ulosonate [α-Ko-(2→4)-Kdo] disaccharide in the lipopolysaccharide from Burkholderia cepacia. *Carbohydrate Research*, 338(23), pp.2659–2666.

Iwashkiw, J.A. *et al*., 2012. Exploiting the *Campylobacter jejuni* protein glycosylation system for glycoengineering vaccines and diagnostic tools directed against brucellosis. *Microbial cell*

factories, 11(1), p.13.

Iwashkiw, J.A. *et al*., 2013. Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation. *Molecular Microbiology*, 89(1), pp.14–28.

Juvet, R.S., Jr & Chiu, J., 1961. Gas Chromatography. IV. The Thermodynamics and Kinetics of the Alcoholysis of Acetals. *Journal of American Chemical Society*, 83(7), pp.1560–1563.

Kaltashov, I.A. & Mohimen, A., 2005. Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. *Analytical Chemistry*, 77(16), pp.5370–5379.

Kaniuk, N.A. *et al*., 2004. Chromosomal and plasmid-encoded enzymes are required for assembly of the R3-type core oligosaccharide in the lipopolysaccharide of *Escherichia coli* O157:H7. *Journal of Biological Chemistry*, 279(30), pp.31237–31250.

Keenleyside, W.J. & Whitfield, C., 1996. A novel pathway for O-polysaccharide biosynthesis in *Salmonella enterica* serovar Borreze. *Journal of Biological Chemistry*, 271(45), pp.28581–28592.

Keenleyside, W.J. & Whitfield, C., 1995. Lateral Transfer of Rfb Genes: a Mobilizable ColE1-Type Plasmid Carries the $Rfb_{O:54}$ (O-54 Antigen Biosynthesis) Gene Cluster From *Salmonella enterica* Serovar Borreze. *Journal of Bacteriology*, 177(18), pp.5247–5253.

Keller, L. & Surette, M.G., 2006. Communication in bacteria: an ecological and evolutionary perspective. *Nature Reviews Microbiology*, 4(4), pp.249–258.

Kelly, J. *et al*., 2006. Biosynthesis of the N-linked glycan in *Campylobacter jejuni* and addition onto protein through block transfer. *Journal of Bacteriology*, 188(7), pp.2427–2434.

Kelly, T.M. *et al*., 1993. The firA Gene of *Escherichia coli* Encodes UDP-3-*O*-(*R*-3-Hydroxymyristoyl)-Glucosamine N-Acyltransferase. The Third Step of Endotoxin Biosynthesis. *Journal of Biological Chemistry*, 268(26), pp.19866–19874.

Khoo, K.-H. & Dell, A., 1990. Assignment of anomeric configurations of pyranose sugars in oligosaccharides using a sensitive FAB-MS strategy. *Glycobiology*, 1(1), pp.83–91.

Kido, N. *et al*., 1995. Expression of the O9 Polysaccharide of *Escherichia coli*: Sequencing of the *Escherichia coli* O9 *rfb* Gene Cluster, Characterization of Mannosyl Transferases, and Evidence for an ATP-Binding Cassette Transport System. *Journal of Bacteriology*, 177(8), pp.2178–2187.

Kim, S.-H. *et al*., 2006. Phosphoethanolamine substitution in the lipid A of *Escherichia coli* O157 : H7 and its association with PmrC. *Microbiology*, 152(Pt 3), pp.657–666.

Kneidinger, B. *et al*., 2002. Biosynthesis Pathway of ADP-L-glycero-β-D-manno-Heptose in *Escherichia coli*. *Journal of Bacteriology*, 184(2), pp.363–369.

References

Knudsen, S.K. *et al*., 2008. Effect of glycosylation on the extracellular domain of the Ag43 bacterial autotransporter: enhanced stability and reduced cellular aggregation. *The Biochemical journal*, 412(3), pp.563–577.

Kooistra, O. *et al*., 2003. Structure of a highly phosphorylated lipopolysaccharide core in the ΔalgC mutants derived from *Pseudomonas aeruginosa* wild-type strains PAO1 (serogroup O5) and PARC1 (serogroup O3). *Carbohydrate Research*, 338, pp.2667–2677.

Kornfeld, R. & Kornfeld, S., 1985. Assembly of asparagine-linked glycoproteins. 54, pp.631–664.

Koshland, D. E., 1953. Stereochemistry and the mechanism of enzymatic reactions. *Biological Reviews* 28, pp.416-436.

Kox, L., Wösten, M. & Groisman, E.A., 2000. A small protein that mediates the activation of a two-component system by another two-component system. *The EMBO journal*, 19(8), pp.1861–1872.

Kozmon, S. & Tvaroška, I., 2006. Catalytic mechanism of glycosyltransferases: Hybrid quantum mechanical/molecular mechanical study of the inverting N-acetylglucosaminyltransferase I. *Journal of the American Chemical Society*, 128, pp.16921–16927.

Ku, S.C. *et al*., 2009. The pilin O-glycosylation pathway of pathogenic *Neisseria* is a general system that glycosylates AniA, an outer membrane nitrite reductase. *Biochemical and biophysical research communications*, 378(1), pp.84–89.

La Mora, De, J.F., 2000. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Analytica chimica acta*, 406(1), pp.93–104.

Lairson, L.L. *et al*., 2008. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annual Review of Biochemistry*, 77(1), pp.521–555.

Langdon, R., 2005. Analysis of the role of HP0208, a phase-variable open reading frame, and its homologues HP1416 and HP0159 in the biosynthesis of *Helicobacter pylori* lipopolysaccharide. *Journal of Medical Microbiology*, 54(8), pp.697–706.

Lemaitre, B. *et al*., 1996. The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in Drosophila adults. *Cell*, 86(6), pp.973–983.

Lin, F. *et al*., 2001. The efficacy of a *Salmonella typhi* Vi conjugate vaccine in two-to-five-year-old children. *Journal of Medicine*, 344(17), pp.1263–1269.

Lindenthal, C. & Elsinghorst, E.A., 1999. Identification of a glycoprotein produced by

enterotoxigenic *Escherichia coli*. *Infection and Immunity*, 67(8), pp.4084–4091.

Linton, D. *et al*., 2002. Identification of N-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in *Campylobacter jejuni*. *Molecular Microbiology*, 43(2), pp.497–508.

Lister, P.D., Wolter, D.J. & Hanson, N.D., 2009. Antibacterial-resistant *Pseudomonas aeruginosa*: clinical impact and complex regulation of chromosomally encoded resistance mechanisms. *Clinical Microbiology Reviews*, 22(4), pp.582–610.

Lithgow, K.V. *et al*., 2014. A general protein O-glycosylation system within the *Burkholderia cepacia* complex is involved in motility and virulence. *Molecular Microbiology*, 92(1), pp.116–137.

Liu, C.-F. *et al*., 2013. Bacterial protein-O-mannosylating enzyme is crucial for virulence of *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, 110(16), pp.6560–6565.

Liu, D. *et al*., 1993. Glycosyl transferases of O-antigen biosynthesis in *Salmonella enterica*: identification and characterization of transferase genes of groups B, C2, and E1. *Journal of Bacteriology*, 175(11), pp.3408–3413.

Liu, D., Cole, R.A. & Reeves, P.R., 1996. An O-antigen processing function for Wzx (RfbX): A promising candidate for O-unit flippase. *Journal of Bacteriology*, 178(7), pp.2102–2107.

Logan, S.M., 2006. Flagellar glycosylation - a new component of the motility repertoire? *Microbiology*, 152, pp.1249–1262.

Logan, S.M. *et al*., 2005. Novel biosynthetic functions of lipopolysaccharide *rfaJ* homologs from Helicobacter pylori. *Glycobiology*, 15(7), pp.721–733.

Lommel, M. & Strahl, S., 2009. Protein O-mannosylation: conserved from bacteria to humans. *Glycobiology*, 19(8), pp.816–828.

Loutet, S.A. *et al*., 2006. A complete lipopolysaccharide inner core oligosaccharide is required for resistance of *Burkholderia cenocepacia* to antimicrobial peptides and bacterial survival in vivo. *Journal of Bacteriology*, 188(6), pp.2073–2080.

Lozniewski, A. *et al*., 2003. Influence of Lewis Antigen Expression by *Helicobacter pylori* on Bacterial Internalization by Gastric Epithelial Cells. *Infection and Immunity*, 71(5), pp.2902–2906.

Lukasiewicz, J. *et al*., 2006. Structure of the lipid A-inner core region and biological activity of *Plesiomonas shigelloides* O54 (strain CNCTC 113/92) lipopolysaccharide. *Glycobiology*, 16(6), pp.538–550.

Lukasiewicz, J. *et al*., 2009. Two Kdo-heptose regions identified in *Hafnia alvei* 32 lipopolysaccharide: the complete core structure and serological screening of different *Hafnia* O serotypes. *Journal of Bacteriology*, 191(2), pp.533–544.

Madhi, S.A. *et al*., 2013. Considerations for a phase-III trial to evaluate a group B *Streptococcus* polysaccharide-protein conjugate vaccine in pregnant women for the prevention of early- and late-onset invasive disease in young-infants. *Vaccine*, 31, pp.D52–D57.

Mahdavi, J., 2002. *Helicobacter pylori* SabA Adhesin in Persistent Infection and Chronic Inflammation. *Science*, 297(5581), pp.573–578.

Mahdavi, J. *et al*., 2003. Limited Role of Lipopolysaccharide Lewis Antigens in Adherence of *Helicobacter pylori* to the Human Gastric Epithelium. *Infection and Immunity*, 71(5), pp.2876–2880.

Mahne, M. *et al*., 2006. The *Corynebacterium glutamicum* gene pmt encoding a glycosyltransferase related to eukaryotic protein-O-mannosyltransferases is essential for glycosylation of the resuscitation promoting factor (Rpf2) and other secreted proteins. *FEMS microbiology letters*, 259(2), pp.226–233.

Malaty, H.M. & Graham, D.Y., 1994. Importance of Childhood Socioeconomic-Status on the Current Prevalence of *Helicobacter pylori* Infection. *Gut*, 35(6), pp.742–745.

Mamyrin, B. A., Karataev, V. I. & Shmikk, D. V., 1973. The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution. *Journal of Experimental and Theoretical Physics*, 37, pp. 45–48.

Manzo, E. *et al*., 2001. A very efficient method to cleave Lipid A and saccharide components in bacterial lipopolysaccharides. *Carbohydrate Research*, 333(4), pp.339–342.

March, R.E., 1997. An introduction to quadrupole ion trap mass spectrometry. *Journal of Mass Spectrometry*, 32(4), pp.351–369.

Marcus, R.A., 1952. Unimolecular dissociations and free radical recombination reactions. *The Journal of Chemical Physics*, 20(3), pp.359–364.

Marshall, B.J. *et al*., 1985. Attemp to fulfill Koch's postulates for pylori campylobacter. *The Medical Journal of Australia*, 142, pp.436–449.

Matson, J.S. *et al*., 2010. Polymyxin B resistance in El Tor *Vibrio cholerae* requires lipid acylation catalyzed by MsbB. *Journal of Bacteriology*, 192(8), pp.2044–2052.

Matsuda, H., 1981. High-Resolution High-Sensitivity Mass-Spectrometer for Heavy Molecule

Analysis. *Japanese Journal of Applied Physics*, 20(11), pp.L825–L828.

McLafferty, F.W., 2011. A century of progress in molecular mass spectrometry. *Annual Review of Analytical Chemistry*, 4, pp.1–22.

McLafferty, F.W., 1957. Mass Spectrometry in Chemical Research and Production. *Applied Spectroscopy*, 11(4), pp.148–156.

McLafferty, F.W. & Bryce, T.A., 1967. Metastable-ion characteristics: characterization of isomeric molecules. *Chemical Communications (London)*, (23), pp.1215–1217.

Meier-Dieter, U. *et al*., 1992. Nucleotide sequence of the *Escherichia coli rfe* gene involved in the synthesis of enterobacterial common antigen. Molecular cloning of the rfe-rff gene cluster. *Journal of Biological Chemistry*, 267(2), pp.746–753.

Meredith, T.C. *et al*., 2006. Redefining the requisite lipopolysaccharide structure in *Escherichia coli*. *ACS Chemical Biology*, 1(1), pp.33–42.

Mikesh, L.M. *et al*., 2006. The utility of ETD mass spectrometry in proteomic analysis. *Biochimica et biophysica acta*, 1764(12), pp.1811–1822.

Moormann, C., 2002. Functional Substitution of the TibC Protein of Enterotoxigenic *Escherichia coli* Strains for the Autotransporter Adhesin Heptosyltransferase of the AIDA System. *Infection and Immunity*, 70(5), pp.2264–2270.

Moran, A.P. *et al*., 2004. Role of *Helicobacter pylori rfaJ* genes (HP0159 and HP1416) in lipopolysaccharide synthesis. *FEMS microbiology letters*, 241(1), pp.57–65.

Morris, H.R. *et al*., 1996. High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Communications in Mass Spectrometry*, 10(8), pp.889–896.

Moskowitz, S.M., Ernst, R.K. & Miller, S.I., 2004. PmrAB, a two-component regulatory system of *Pseudomonas aeruginosa* that modulates resistance to cationic antimicrobial peptides and addition of aminoarabinose to lipid A. *Journal of Bacteriology*, 186(2), pp.575–579.

Muller-Loennies, S., Lindner, B. & Brade, H., 2003. Structural Analysis of Oligosaccharides from Lipopolysaccharide (LPS) of *Escherichia coli* K12 Strain W3100 Reveals a Link between Inner and Outer Core LPS Biosynthesis. *Journal of Biological Chemistry*, 278(36), pp.34090–34101.

Naegeli, A. *et al*., 2014. Substrate specificity of cytoplasmic N-glycosyltransferase. *Journal of Biological Chemistry*, 289(35), pp.24521–24532.

References

Nakamura, S. *et al*., 1993. Hyperglycosylation of Hen Egg-White Lysozyme in Yeast. *Journal of Biological Chemistry*, 268(17), pp.12706–12712.

Neuberger, A., 1938. Carbohydrates in protein: The carbohydrate component of crystalline egg albumin. *The Biochemical journal*, 32(9), pp.1435–1451.

Nothaft, H. & Szymanski, C.M., 2010. Protein glycosylation in bacteria:sweeter than ever. *Nature Reviews Microbiology*, 8(11), pp.1–14.

Olsthoorn, M. *et al*., 1998. Identification of a Novel Core Type in *Salmonella* Lipopolysaccharide. Complete Structural Analysis of the Core Region of the Lipopolysaccharide from *Salmonella enterica* sv. Arizonae 062. *Journal of Biological Chemistry*, 273(7), pp.3817–3829.

O'Reilly, M.K., Zhang, G. & Imperiali, B., 2006. In vitro evidence for the dual function of Alg2 and Alg11: essential mannosyltransferases in N-linked glycoprotein biosynthesis. *Biochemistry*, 45(31), pp.9593–9603.

Paizs, B. & Suhai, S., 2005. Fragmentation pathways of protonated peptides. *Mass Spectrometry Reviews*, 24(4), pp.508–548.

Park, B.S. *et al*., 2009. The structural basis of lipopolysaccharide recognition by the TLR4. *Nature*, 458(7242), pp.1191–1195.

Parker, C.T. *et al*., 1992. Role of the *rfaG* and *rfaP* Genes in Determining the Lipopolysaccharide Core Structure and Cell Surface Properties of *Escherichia coli* K-12. *Journal of Bacteriology*, 174(8), pp.2525–2538.

Peek, R.M. & Blaser, M.J., 2002. *Helicbacter pylori* and gastrointestinal tract adenocarcinomas. *Nature Reviews Cancer*, 2(1), pp.28–37.

Perez, C. *et al*., 2015. Structure and mechanism of an active lipid-linked oligosaccharide flippase. *Nature*, 524(7566), pp.433–438.

Persson, K. *et al*., 2001. Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. *Nature Structural Biology*, 8(2), pp.166–175.

Pfeifer, R.J. & Hendricks, C.D., 1968. Parametric studies of electrohydrodynamic spraying. *AIAA Journal*, 6(3), pp.496–502.

Poland, G.A., 2010. Prevention of meningococcal disease: current use of polysaccharide and conjugate vaccines. *Clinical Infectious Diseases*, 50 Suppl 2(s2), pp.S45–53.

Poltorak, A., He, X., *et al*., 1998. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr

mice: mutations in *Tlr4* gene. *Science*, 282(5396), pp.2085–2088.

Poltorak, A., Smirnova, I., *et al*., 1998. Genetic and Physical Mapping of the *Lps* Locus: Identification of the Toll-4 Receptor as a Candidate Gene in the Critical Region. *Blood cells, molecules & diseases*, 24(3), pp.340–355.

Poppe, L. *et al*., 1994. Carbohydrate Dynamics at a Micellar Surface: GD1a Headgroup Transformations Revealed by NMR spectroscopy. *Biophysical Journal*, 66(5), pp.1642–1652.

Power, P.M., Seib, K.L. & Jennings, M.P., 2006. Pilin glycosylation in *Neisseria meningitidis* occurs by a similar pathway to wzy-dependent O-antigen biosynthesis in *Escherichia coli*. *Biochemical and biophysical research communications*, 347(4), pp.904–908.

Pringle, S.D. *et al*., 2007. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry*, 261(1), pp.1–12.

Qiao, S. *et al*., 2014. Structural basis for lipopolysaccharide insertion in the bacterial outer membrane. *Nature*, 511(7507), pp.108–111.

Que, Y.-A. & Moreillon, P., 2011. Infective endocarditis. *Nature Reviews Microbiology*, 8(6), pp.322–336.

Radika, K. & Raetz, C.R.H., 1988. Purification and properties of lipid A disaccharide synthase of *Escherichia coli*. *Journal of Biological Chemistry*, 263(29), pp.14859–14867.

Raetz, C.R.H., 2001. Regulated covalent modifications of lipid A. *Journal of Endotoxin Research*, 7(1), pp.73–78.

Raetz, C.R.H. & Whitfield, C., 2002. Lipopolysaccharide endotoxins. *Annual Review of Biochemistry*, 71, pp.635–700.

Raetz, C.R.H. *et al*., 2007. Lipid A modification systems in gram-negative bacteria. *Annual Review of Biochemistry*, 76, pp.295–329.

Ramjeet, M. *et al*., 2005. Truncation of the lipopolysaccharide outer core affects susceptibility to antimicrobial peptides and virulence of *Actinobacillus pleuropneumoniae* serotype 1. *Journal of Biological Chemistry*, 280(47), pp.39104–39114.

Rasko, D.A. *et al*., 2000. Synthesis of mono- and di-fucosylated type I Lewis blood group antigens by *Helicobacter pylori*. *European Journal of Biochemistry*, 267(19), pp.6059–6066.

Rayleigh, L., 2009. XX. On the equilibrium of liquid conducting masses charged with electricity. *Philosophical Magazine Series 5*, 14(87), pp.184–186.

Reizer, J., Reizer, A. & Saier, M.H., 1992. A New Subfamily of Bacterial Abc-Type Transport-Systems Catalyzing Export of Drugs and Carbohydrates. *Protein Science*, 1(10), pp.1326–1332.

Rick, P.D., Hubbard, G.L. & Barr, K., 1994. Role of the *rfe* Gene in the Synthesis of the O8 Antigen in *Escherichia coli* K-12. *Journal of Bacteriology*, 176(10), pp.2877–2884.

Rietschel, E.T. *et al*., 1994. Bacterial endotoxin: molecular relationships of structure to activity and function. *FASEB journal*, 8(2), pp.217–225.

Robbins, P.W. *et al*., 1967. Direction of chain growth in polysaccharide synthesis. *Science*, 158(3808), pp.1536–1542.

Rocchetta, H.L., Burrows, L.L. & Lam, J.S., 1999. Genetics of O-antigen biosynthesis in *Pseudomonas aeruginosa*. *Microbiology and Molecular Biology Reviews*, 63(3), pp.523-553.

Rock, F.L. *et al*., 1998. A family of human receptors structurally related to *Drosophila* Toll. *Proceedings of the National Academy of Sciences*, 95(2), pp.588–593.

Rolig, A.S., Carter, J.E. & Ottemann, K.M., 2011. Bacterial chemotaxis modulates host cell apoptosis to establish a T-helper cell, type 17 (Th17)-dominant immune response in *Helicobacter pylori* infection. *Proceedings of the National Academy of Sciences*, 108(49), pp.19749–19754.

Rondini, S. *et al*., 2015. Design of glycoconjugate vaccines against invasive African *Salmonella enterica* serovar *Typhimurium*. *Infection and Immunity*, 83(3), pp.996–1007.

Rosenstock, H.M. *et al*., 1952. Absolute Rate Theory for Isolated Systems and the Mass Spectra of Polyatomic Molecules. *Proceedings of the National Academy of Sciences*, 38(8), pp.667–678.

Ruan, X. *et al*., 2012. The WaaL O-antigen lipopolysaccharide ligase has features in common with metal ion-independent inverting glycosyltransferases. *Glycobiology*, 22(2), pp.288–299.

Ruiz, N. *et al*., 2008. Identification of two inner-membrane proteins required for the transport of lipopolysaccharide to the outer membrane of *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 105(14), pp.5537–5542.

Sachs, G. *et al*., 2003. The gastric biology of *Helicobacter pylori*. *Annual Review of Physiology*, 65, pp.349–369.

Salama, N.R., Hartung, M.L. & Müller, A., 2013. Life in the human stomach: persistence strategies of the bacterial pathogen *Helicobacter pylori*. *Nature Review Microbiology*, 11(6), pp.385–399.

Schirm, M. *et al*., 2005. Identification of unusual bacterial glycosylation by tandem mass spectrometry analyses of intact proteins. *Analytical Chemistry*, 77(23), pp.7774–7782.

Schirm, M. *et al*., 2004. Structural and genetic characterization of glycosylation of type a flagellin in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 186(9), pp.2523–2531.

Schirm, M. *et al*., 2003. Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Molecular Microbiology*, 48(6), pp.1579–1592.

Schwartz, J.C., Senko, M.W. & Syka, J.E.P., 2002. A two-dimensional quadrupole ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry*, 13(6), pp.659–669.

Schwarz, F. *et al*., 2011. Cytoplasmic N-glycosyltransferase of *Actinobacillus pleuropneumoniae* is an inverting enzyme and recognizes the NX(S/T) consensus sequence. *Journal of Biological Chemistry*, 286(40), pp.35267–35274.

Scott, A.E. *et al*., 2011. Flagellar glycosylation in *Burkholderia pseudomallei* and *Burkholderia thailandensis*. *Journal of Bacteriology*, 193(14), pp.3577–3587.

Scott, A.E. *et al*., 2013. Protection against experimental melioidosis following immunization with live *Burkholderia thailandensis* expressing a manno-heptose capsule. *Clinical and vaccine immunology : CVI*, 20(7), pp.1041–1047.

Scott, N.E. et al., 2014. Diversity within the O-linked protein glycosylation systems of acinetobacter species. Molecular & Cellular Proteomics, 13(9), pp.2354–2370.

Shao, P.L. *et al*., 2006. Safety and immunogenicity of heptavalent pneumococcal conjugate vaccine booster in Taiwanese toddlers. *Journal of the Formosan Medical Association*, 105(7), pp. 542-549

Sherlock, O. *et al*., 2006. Glycosylation of the self-recognizing *Escherichia coli* Ag43 autotransporter protein. *Journal of Bacteriology*, 188(5), pp.1798–1807.

Shimazu, R. *et al*., 1999. MD-2, a molecule that confers lipopolysaccharide responsiveness on Toll-like receptor 4. *The Journal of Experimental Medicine*, 189(11), pp.1777–1782.

Skurnik, M. *et al*., 1999. The lipopolysaccharide outer core of *Yersinia enterocolitica* serotype O:3 is required for virulence and plays a role in outer membrane integrity. *Molecular Microbiology*, 31(5), pp.1443–1462.

Sleytr, U.B., 1975. Heterologous reattachment of regular arrays of glycoproteins on bacterial surfaces. *Nature*, 257(5525), pp.400–402.

Smith, R.S. & Iglewski, B.H., 2003. *Pseudomonas aeruginosa* quorum sensing as a potential antimicrobial target. *Journal of Clinical Investigation*, 112(10), pp.1460–1465.

Smyth, K.M. & Marchant, A., 2013. Conservation of the 2-keto-3-deoxymanno-octulosonic

acid (Kdo) biosynthesis pathway between plants and bacteria. *Carbohydrate Research*, 380, pp.70–75.

Sorge, N.M.V. *et al*., 2009. N-glycosylated proteins and distinct lipooligosaccharide glycoforms of *Campylobacter jejuni* target the human C-type lectin receptor MGL. *Cellular microbiology*, 11(12), pp.1768–1781.

Sperandeo, P. *et al*., 2008. Functional Analysis of the Protein Machinery Required for Transport of Lipopolysaccharide to the Outer Membrane of *Escherichia coli*. *Journal of Bacteriology*, 190(13), pp.4460–4469.

Sperandeo, P. *et al*., 2011. New insights into the Lpt machinery for lipopolysaccharide transport to the cell surface: LptA-LptC interaction and LptA stability as sensors of a properly assembled transenvelope complex. *Journal of Bacteriology*, 193(5), pp.1042–1053.

Stead, C.M. *et al*., 2010. Removal of the outer Kdo from *Helicobacter pylori* lipopolysaccharide and its impact on the bacterial surface. *Molecular Microbiology*, 78(4), pp.837–852.

Stephens, W.E., 1946. A pulsed mass spectrometer with time dispersion. In Bulletin of the American Physical Society. Physical Review, pp. 22.

Stephenson, A.E. *et al*., 2002. The Fap1 fimbrial adhesin is a glycoprotein: antibodies specific for the glycan moiety block the adhesion of *Streptococcus parasanguis* in an in vitro tooth model. *Molecular Microbiology*, 43(1), pp.147–157.

Suerbaum, S. & Michetti, P., 2002. *Helicobacter pylori* infection. *New England Journal of Medicine*, 347(15), pp.1175–1186.

Suits, M.D.L. *et al*., 2008. Novel structure of the conserved gram-negative lipopolysaccharide transport protein A and mutagenesis analysis. *Journal of Molecular Biology*, 380(3), pp.476–488.

Sutcliffe, I.C., 2010. A phylum level perspective on bacterial cell envelope architecture. *Trends in Microbiology*, 18(10), pp.464–470.

Sweeley, C.C. *et al*., 1963. Gas-liquid chromatography of trimethylsilyl derivatives of sugars and related substances. *Journal of American Chemical Society*, 85(16), pp.2497–2507.

Sycuro, L.K. *et al*., 2010. Peptidoglycan crosslinking relaxation promotes Helicobacter pylori's helical shape and stomach colonization. *Cell*, 141(5), pp.822–833.

Syka, J. *et al*., 2004. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences*, 101(26), pp.9528–9533.

Szymanski, C.M. *et al*., 1999. Evidence for a system of general protein glycosylation in

Campylobacter jejuni. *Molecular Microbiology*, 32(5), pp.1022–1030.

Tabei, S.M.B. *et al*., 2009. An Aeromonas caviae genomic island is required for both O-antigen lipopolysaccharide biosynthesis and flagellin glycosylation. *Journal of Bacteriology*, 191(8), pp.2851–2863.

Takeda, K., Kaisho, T. & Akira, S., 2003. Toll-like Receptors. *Annual Review of Immunology*, 21(1), pp.335–376.

Tanaka, K. *et al*., 1988. Protein and polymer analyses up to m/z 100000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 2(20), pp.151–153.

Teghanemt, A. *et al*., 2005. Molecular basis of reduced potency of underacylated endotoxins. *Journal of Immunology*, 175(7), pp.4669–4676.

Thaipisuttikul, I. *et al*., 2014. A divergent *Pseudomonas aeruginosa* palmitoyltransferase essential for cystic fibrosis-specific lipid A. *Molecular Microbiology*, 91(1), pp.158–174.

Theodoratou, E. *et al*., 2010. The effect of *Haemophilus influenzae* type b and *pneumococcal* conjugate vaccines on childhood pneumonia incidence, severe morbidity and mortality. *International journal of epidemiology*, 39 Suppl 1(Supplement 1), pp.i172–85.

Thibault, P. *et al*., 2001. Identification of the carbohydrate moieties and glycosylation motifs in *Campylobacter jejuni* flagellin. *Journal of Biological Chemistry*, 276(37), pp.34862–34870.

Tipton, J.D. *et al*., 2012. Nano-LC FTICR tandem mass spectrometry for top-down proteomics: routine baseline unit mass resolution of whole cell lysate proteins up to 72 kDa. *Analytical Chemistry*, 84(5), pp.2111–2117.

Toshima, G. *et al*., 2003. Histidine-114 at subsites E and F can explain the characteristic enzymatic activity of guinea hen egg-white lysozyme. *Bioscience, biotechnology, and biochemistry*, 67(3), pp.540–546.

Tran, A.X., Dong, C. & Whitfield, C., 2010. Structure and Functional Analysis of LptC, a Conserved Membrane Protein Involved in the Lipopolysaccharide Export Pathway in *Escherichia coli*. *Journal of Biological Chemistry*, 285(43), pp.33529–33539.

Trent, M.S., Pabich, W., *et al*., 2001. A PhoP/PhoQ-induced Lipase (PagL) That Catalyzes 3-O-Deacylation of Lipid A Precursors in Membranes of *Salmonella typhimurium*. *Journal of Biological Chemistry*, 276(12), pp.9083–9092.

Trent, M.S., Ribeiro, A.A., *et al*., 2001. An inner membrane enzyme in *Salmonella* and

Escherichia coli that transfers 4-amino-4-deoxy-L-arabinose to lipid A: induction on polymyxin-resistant mutants and role of a novel lipid-linked donor. *Journal of Biological Chemistry*, 276(46), pp.43122–43131.

Twine, S.M. *et al*., 2008. Flagellar glycosylation in *Clostridium botulinum*. *FEBS Journal*, 275(17), pp.4428–4444.

Vaara, M. *et al*., 1981. Characterization of the Lipopolysaccharide from the Polymyxin-Resistant *pmrA* Mutants of *Salmonella Typhimurium*. *FEBS Letters*, 129(1), pp.145–149.

Vallée-Réhel, K. & Zähringer, U., 2004. Lipid A components from *Pseudomonas aeruginosa* PAO1 (serotype O5) and mutant strains investigated by electrospray ionization ion-trap mass spectrometry. *Journal of Mass Spectrometry*, 39(5), pp.505–513.

VanderVen, B.C. *et al*., 2005. Export-mediated assembly of mycobacterial glycoproteins parallels eukaryotic pathways. *Science*, 309(5736), pp.941–943.

Velapatiño, B. *et al*., 2013. Total protein extraction and 2-D gel electrophoresis methods for Burkholderia species. *Journal of visualized experiments : JoVE*, (80), p.e50730.

Verma, A. *et al*., 2005. Roles of specific amino acids in the N terminus of *Pseudomonas aeruginosa* flagellin and of flagellin glycosylation in the innate immune response. *Infection and Immunity*, 73(12), pp.8237–8246.

Vescovi, E.G. *et al*., 1997. Characterization of the bacterial sensor protein PhoQ. Evidence for distinct binding sites for Mg^{2+} and Ca^{2+}. *Journal of Biological Chemistry*, 272(3), pp.1440–1443.

Vescovi, E.G., Soncini, F.C. & Groisman, E.A., 1996. Mg^{2+} as an extracellular signal: Environmental regulation of *Salmonella* virulence. *Cell*, 84(1), pp.165–174.

Vestal, M.L., Juhasz, P. & Martin, S.A., 1995. Delayed Extraction Matrix-Assisted Laser-Desorption Time-of-Flight Mass-Spectrometry. *Rapid Communications in Mass Spectrometry*, 9(11), pp.1044–1050.

Villa, R. *et al*., 2013. The *Escherichia coli* Lpt transenvelope protein complex for lipopolysaccharide export is assembled via conserved structurally homologous domains. *Journal of Bacteriology*, 195(5), pp.1100–1108.

Vinogradov, E.V. *et al*., 1999. The structures of the carbohydrate backbones of the lipopolysaccharides from *Escherichia coli* rough mutants F470 (R1 core type) and F576 (R2 core type). *European Journal of Biochemistry*, 261(3), pp.629–639.

Wacker, M. *et al*., 2002. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into E. coli. *Science*, 298(5599), pp.1790–1793.

Wang, G., Boulton, P.G., *et al*., 1999. Novel *Helicobacter pylori* α1, 2-fucosyltransferase, a key enzyme in the synthesis of Lewis antigens. *Microbiology*, 145(11), pp.3245–3253.

Wang, G., Rasko, D.A., *et al*., 1999. Molecular genetic basis for the variable expression of Lewis Y antigen in *Helicobacter pylori*: analysis of the α (1, 2) fucosyltransferase gene. *Molecular Microbiology*, 31(4), pp.1265–1274.

Wang, L., Liu, D. & Reeves, P.R., 1996. C-terminal half of *Salmonella enterica* WbaP (RfbP) is the galactosyl-1-phosphate transferase domain catalyzing the first step of O-antigen synthesis. *Journal of Bacteriology*, 178(9), pp.2598–2604.

Ward, A. *et al*., 2007. Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proceedings of the National Academy of Sciences*, 104(48), pp.19005–19010.

Weeks, D.L. *et al*., 2000. A H^+-Gated Urea Channel: The Link Between *Helicobacter pylori* Urease and Gastric Colonization. *Science*, 287(5452), pp.482–485.

Wehmeier, S. *et al*., 2009. Glycosylation of the phosphate binding protein, PstS, in *Streptomyces coelicolor* by a pathway that resembles protein O-mannosylation in eukaryotes. *Molecular Microbiology*, 71(2), pp.421–433.

Whitehouse, C.M. *et al*., 1985. Electrospray Interface for Liquid Chromatographs and Mass Spectrometers. *Analytical Chemistry*, 57(3), pp.675–679.

Whitfield, C. & Trent, M.S., 2014. Biosynthesis and export of bacterial lipopolysaccharides. *Annual Review of Biochemistry*, 83, pp.99–128.

Winger, B.E. *et al*., 1993. Observation and Implications of High Mass-to-Charge Ratio Ions From Electrospray-Ionization Mass-Spectrometry. *Journal of the American Society for Mass Spectrometry*, 4(7), pp.536–545.

Wolff, M.M. & Stephens, W.E., 1953. A pulsed mass spectrometer with time dispersion. *Review of Scientific Instruments*, 24, pp.616–617.

Wösten, M. & Groisman, E.A., 1999. Molecular characterization of the PmrA regulon. *Journal of Biological Chemistry*, 274(38), pp.27185–27190.

Wösten, M. *et al*., 2000. A signal transduction system that responds to extracellular iron. *Cell*, 103(1), pp.113–125.

References

Wu, H. et al., 1998. Isolation and characterization of Fap1, a fimbriae-associated adhesin of *Streptococcus parasanguis* FW213. *Molecular Microbiology*, 28(3), pp.487–500.

Wyckoff, T. et al., 1998. Hydrocarbon rulers in UDP-N-acetylglucosamine acyltransferases. *Journal of Biological Chemistry*, 273(49), pp.32369–32372.

Yamashita, M. & Fenn, J.B., 1984a. Electrospray Ion Source. Another Variation on the Free-Jet Theme. *Journal of Physical Chemistry*, 88(20), pp.4451–4459.

Yamashita, M. & Fenn, J.B., 1984b. Negative Ion Production with the Electrospray Ion Source. *Journal of Physical Chemistry*, 88(20), pp.4671–4675.

Yan, S. *et al*., 2012. Galactosylated fucose epitopes in nematodes: increased expression in a Caenorhabditis mutant associated with altered lectin sensitivity and occurrence in parasitic species. *Journal of Biological Chemistry*, 287(34), pp.28276–28290.

Yao, Z. & Valvano, M.A., 1994. Genetic analysis of the O-specific lipopolysaccharide biosynthesis region (*rfb*) of *Escherichia coli* K-12 W3110: identification of genes that confer group 6 specificity to *Shigella flexneri* serotypes Y and 4a. *Journal of Bacteriology*, 176(13), pp.4133–4143.

Yethon, J.A. *et al*., 1998. Involvement of waaY, waaQ, and waaP in the modification of *Escherichia coli* lipopolysaccharide and their role in the formation of a stable outer membrane. *Journal of Biological Chemistry*, 273(41), pp.26310–26316.

Young, K. *et al*., 1995. The envA Permeability/cell Division gene of *Escherichia coli* Encodes the Second enzyme of Lipid A biosynthesis. UDP-3-*O*-(*R*-3-hydroxymyristoyl)-*N*-acetylglucosamine deacetylase. *Journal of Biological Chemistry*, 270(51), pp.30384–30391.

Zähringer, U. *et al*., 2014. NMR-based structural analysis of the complete rough-type lipopolysaccharide isolated from Capnocytophaga canimorsus. *Journal of Biological Chemistry*, 289(34), pp.23963–23976.

Zhang, H. *et al*., 2014. The highly conserved domain of unknown function 1792 has a distinct glycosyltransferase fold. *Nature Communications,* 5:4339.

Zhou, M. & Wu, H., 2009. Glycosylation and biogenesis of a family of serine-rich bacterial adhesins. *Microbiology*, 155(2), pp.317–327.

Zhu, F. *et al*., 2011. Structural and Functional Analysis of a New Subfamily of Glycosyltransferases Required for Glycosylation of Serine-rich *Streptococcal* Adhesins. *Journal of Biological Chemistry*, 286(30), pp.27048–27057.

Appendices

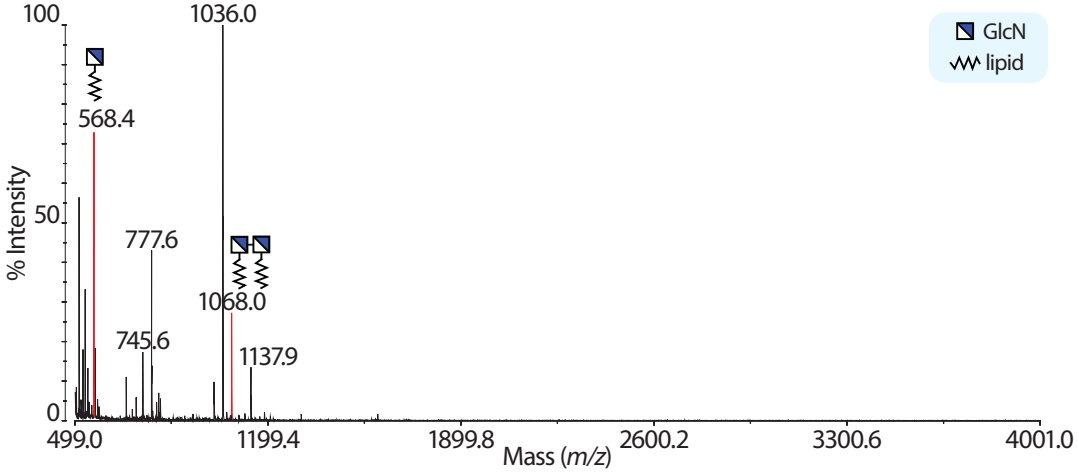


Figure A1. WT G27 LPS after mild acid hydrolysis. Briefly, the LPS sample was hydrolysed in 5% acetic acid at 100 °C for 120 min and subjected to centrifuge. The supernatant was collected, lyophilised, permethylated and then analysed by MALDI-TOF. The red peaks correspond to hydrolysed lipid A structures.

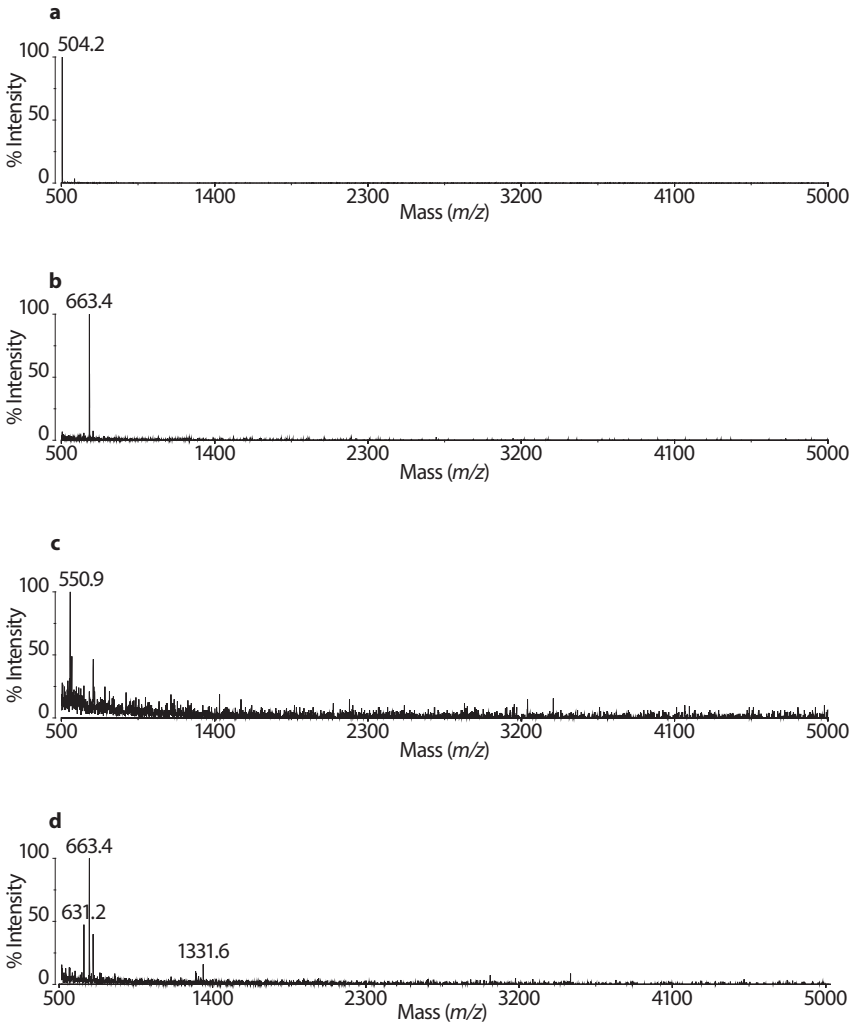


Figure A2. Smith degradation of LPS samples extracted from ΔHP1283 (a), ΔHP1578 (b), ΔHP0102 (c) and ΔWaaL G27 (d). No peaks corresponding to LacNAc-type were observed.

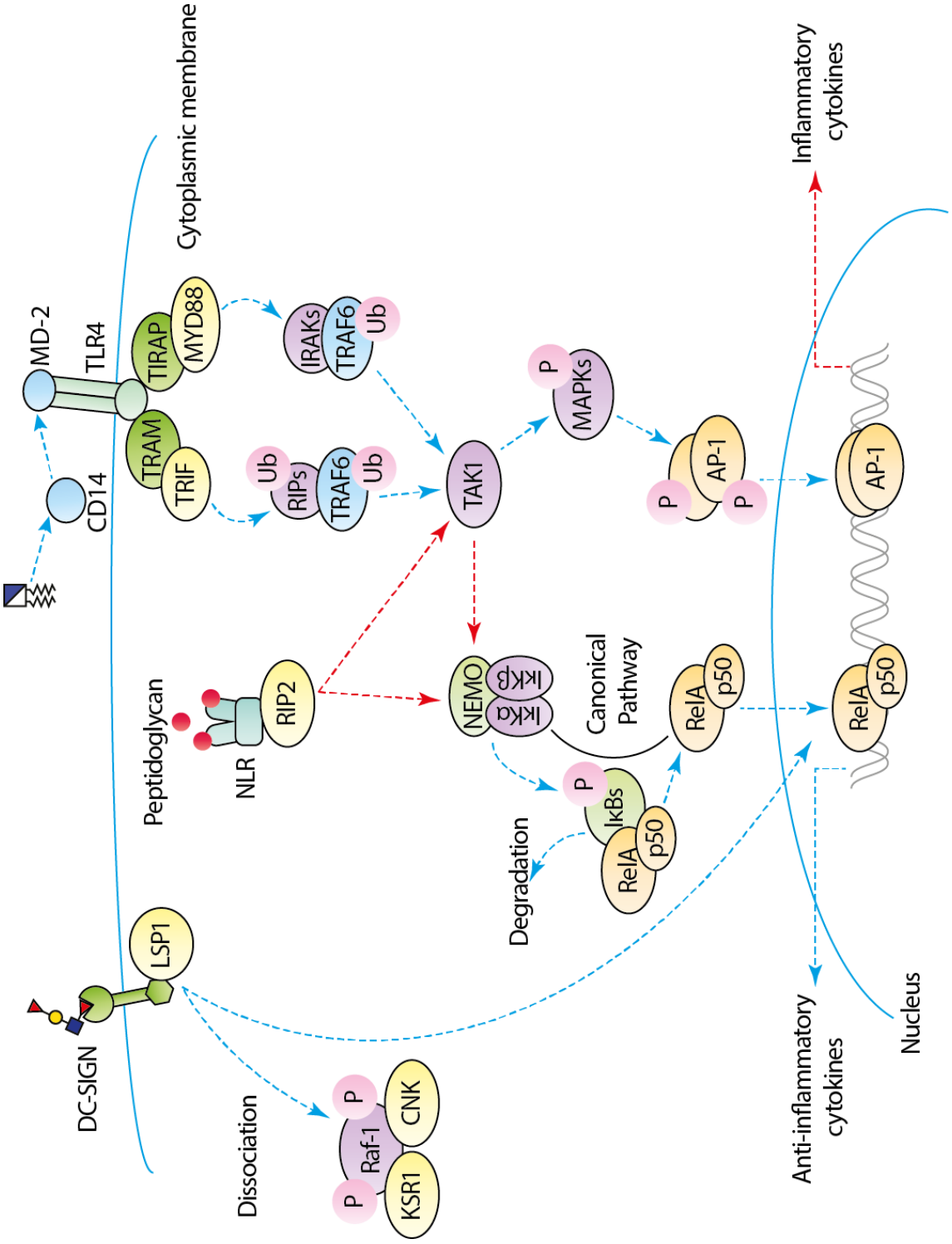


Figure A3. An example of innate immunity against glycoconjugates

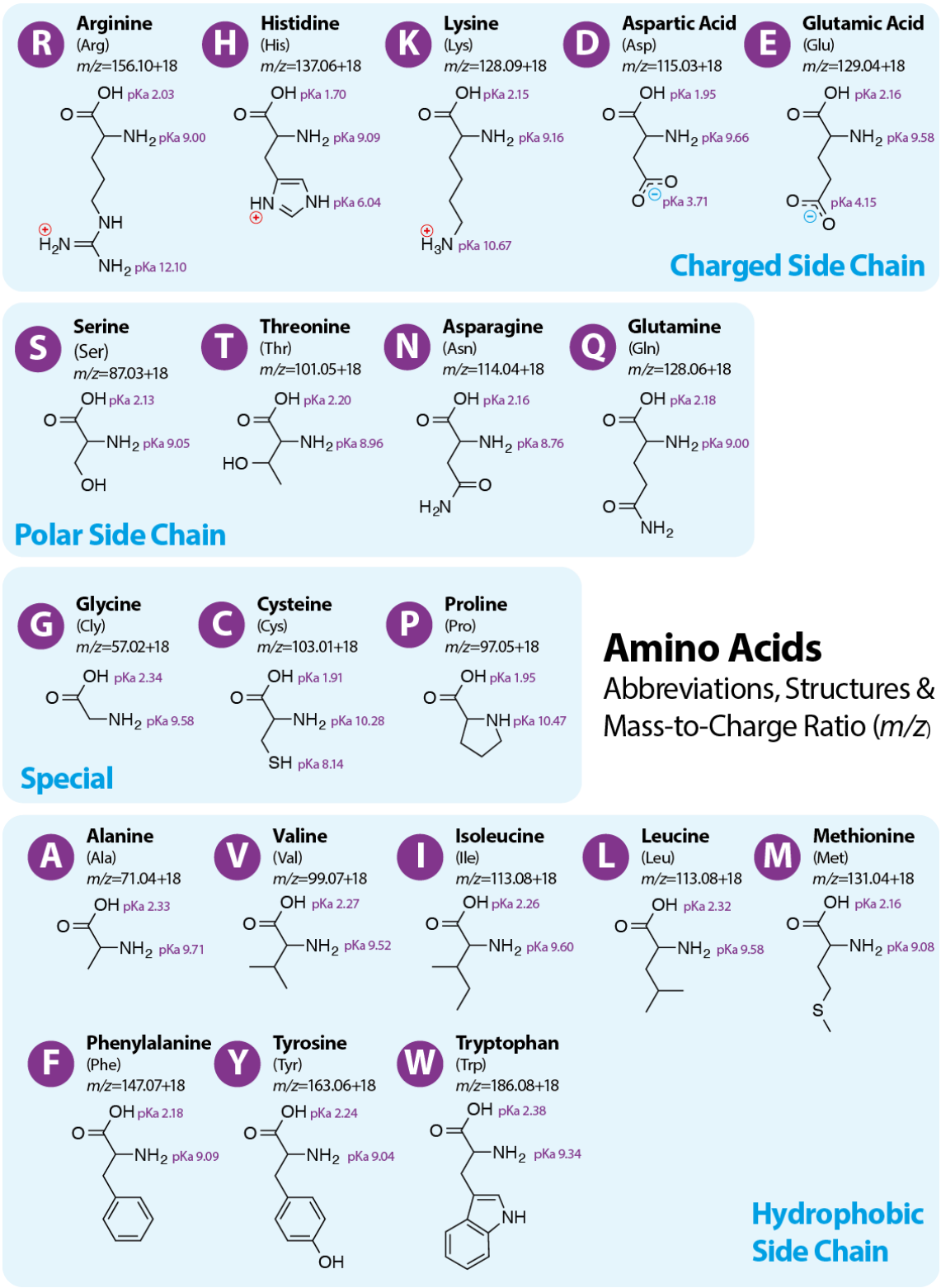


Figure A4. List of Amino Acids