Imperial College London

[dstl]

STRUCTURE ANALYSIS OF BIOLOGICALLY IMPORTANT PROKARYOTIC GLYCOPOLYMERS

A thesis submitted to Imperial College of Science, Technology and Medicine for the Degree of Doctor of Philosophy

Submitted by

Valeria Victoria Ventura

IMPERIAL COLLEGE LONDON

Department of Life Sciences

London SW7 2AZ

UNITED KINGDOM

July 2013

Für meine Eltern und Grosseltern,
die unermüdlich gearbeitet haben, um ihren Kindern eine bessere Zukunft zu bieten.

To my parents and grandparents,
who worked so relentlessly to offer their children a better future.

Ai miei genitori ed ai miei nonni,
che hanno lavorato duramente per poter offrire ai loro figli un futuro migliore.

wo chiemte mer hi
wenn alli seite
wo chiemte mer hi
und niemer giengti
für einisch z'luege
wohi dass me chiem
we me gieng

Kurt Marti

where would we get to
if everyone said
where would we get to
and nobody would go
to look
where we would get to
if we would go

Kurt Marti

ABSTRACT

Of the many post-translational modifications organisms can undertake, glycosylation is the most prevalent and the most diverse. The research in this thesis focuses on the structural characterisation of glycosylation in two classes of glycopolymer (lipopolysaccharide (LPS) and glycoprotein) in two domains of life (bacteria and archaea). The common theme linking these subprojects is the development and application of high sensitivity analytical techniques, primarily mass spectrometry (MS), for studying prokaryotic glycosylation. Many prokaryotes produce glycan arrangements with extraordinary variety in composition and structure. A further challenge is posed by additional functionalities such as lipids whose characterisation is not always straightforward. Glycosylation in prokaryotes has a variety of different biological functions, including their important roles in the mediation of interactions between pathogens and hosts. Thus enhanced knowledge of bacterial glycosylation may be of therapeutic value, whilst a better understanding of archaeal protein glycosylation will provide further targets for industrial applications, as well as insight into this post-translational modification across evolution and protein processing under extreme conditions.

The first sub-project focused on the S-layer glycoprotein of the halophilic archeaon *Haloferax volcanii*, which has been reported to be modified by both glycans and lipids. Glycoproteomic and associated MS technologies were employed to characterise the N- and O-linked glycosylation and to explore putative lipid modifications. Approximately 90% of the S-layer was mapped and N-glycans were identified at all the mapped consensus sites, decorated with a pentasaccharide consisting of two hexoses, two hexuronic acids and a methylated hexuronic acid. The O-glycans are homogeneously identified as a disaccharide consisting of galactose and glucose. Unexpectedly it was found that membrane-derived lipids were present in the S-layer samples despite extensive purification, calling into question the predicted presence of covalently linked lipid. The *H. volcanii* N-glycosylation is mediated by the products of the *agl* gene cluster and the functional characterisation of members of the *agl* gene cluster was investigated by MS analysis of *agl*-mutant strains of the S-layer.

Burkholderia pseudomallei is the causative agent of melioidosis, a serious and often fatal disease in humans which is endemic in South-East Asia and other equatorial regions. Its LPS is vital for serum resistance and the O-antigen repeat structures are of interest as vaccine targets. *B. pseudomallei* is reported to produce several polysaccharides, amongst which the already characterised ‘typical’ O-antigen of K96243 represents 97% of the strains. The serologically distinct ‘atypical’ strain 576 produces a different LPS, whose characterisation is the subject of this research project. MS strategies coupled with various hydrolytic and chemical derivatisation methodologies were employed to define the composition and potential sequences of the O-antigen repeat unit. These MS strategies were complemented by a novel NMR technique involving embedding of the LPS into micelles. Taken together the MS and NMR data have revealed a highly unusual O-antigen structure for atypical LPS which is remarkably different from the typical O-antigen.

The development of structural analysis tools in MS and NMR applicable to the illustrated types of glycosylation in these prokaryotes will give a more consistent approach to sugar characterisation and their modifications thus providing more informative results for pathogenicity and immunological studies as well as pathway comparisons.

THESIS PUBLICATIONS

Ngugi SA, **Ventura VV**, Qazi O, Harding SV, Kitto GB, Estes DM, Dell A, Titball RW, Atkins TP, Brown KA, Hitchen PG, Prior JL (2010) *Lipopolysaccharide from Burkholderia thailandensis E264 provides protection in a murine model of melioidosis*. ***Vaccine*** 3;28(47):7551-5. (doi: 10.1016) (Epub 2010 Sep 15)

Kaminski L, Abu-Qarn M, Guan Z, Naparstek S, **Ventura VV**, Raetz CR, Hitchen PG, Dell A, Eichler J (2010) *AglJ adds the first sugar of the N-linked pentasaccharide decorating the Haloferax volcanii S-layer glycoprotein.* ***J Bacteriol.*** 192(21):5572-9. (doi: 10.1128) (Epub 2010 Aug 27)

Yurist-Doutsch S, Magidovich H, **Ventura VV**, Hitchen PG, Dell A, Eichler J (2010) *N-glycosylation in Archaea: on the coordinated actions of Haloferax volcanii AglF and AglM.* ***Mol Microbiol.***;75(4):1047-58. (doi: 10.1111)

Magidovich H, Yurist-Doutsch S, Konrad Z, **Ventura VV**, Dell A, Hitchen PG, Eichler J (2010) *AglP is a S-adenosyl-L-methionine-dependent methyltransferase that participates in the N-glycosylation pathway of Haloferax volcanii.* ***Mol Microbiol.*** 76(1):190-9. (doi: 10.1111) (Epub 2010 Feb 10)

ORIGINALITY DECLARATION

The work presented in this thesis has not been previously or concurrently submitted for any other degree, diploma or other qualification and is the result of the author's own independent investigation unless otherwise stated.

Valeria V. Ventura
May 2013

COPYRIGHT DECLARATION

ACKNOWLEDGEMENTS

There are many people who have contributed to the realisation of this thesis and who I am immeasurably grateful for.

I would like to express my profoundest gratitude to my wonderful mentor Professor Anne Dell for her unwavering support and motivation throughout my PhD. Your brilliant approach to leadership and scientific research have been enormously inspirational and I would like to thank you for reinforcing my confidence as a scientist and will always fondly remember these last few years of teaching as the perfect foundation for my scientific career. I could not have wished for a better role model. My sincere thanks extended to Dr. Stuart Haslam, as well as Professor Howard Morris and Dr. Maria Panico, whose extremely welcoming and encouraging training will stay with me in all my future endeavours.

To our collaborators and supervisors outside of the laboratory, I would like to thank Professor Jerry Eichler and his team, especially the supportive team at the Defence Science and Technology Laboratory (dstl) for all your help and stimulating brainstorming. Dr. Joann Prior, Sarah Ngugi, Dr. Andrew Scott, Dr. Claire Lonsdale and Dr. Laura Marshall, it was been such a pleasure to learn from you and explore these projects with you all. My sincere thanks go to Dr. Barbara Mulloy for her positive and kind coaching and support and I am very grateful for the help of Dr. Jan Marchant at the Cross Faculty NMR centre as well as the appreciated assistance of Michael-John O'Connor and Daniel Binet at the Mass Spectrometry Research and Training Centre in Jersey.

To the past and present members of the laboratory whose interesting discussions, friendliness and welcome assistance in more taxing times have made this experience enormously enjoyable and memorable. I would like to especially thank Dr. Paul Hitchen, whose guidance, help and answers to all my incessant questions over the last five years has been tremendously appreciated. Thank you to Dr. Poh-Choo Pang, Dr. Aristotelis Antonopolous, Dr. Paola Grassi, Dr. Kévin Canis, Dr. Rebecca Harrison, Dr. Bérangère Tissot, Dr. Simon North, Dr. Elham Peyfoon, Dr. Nazri Ismael, Sang-Jun Lee, Alana Trollope, Dr. David Damerell, Dr. Alessio Ceroni, Lesley Harris, Tony Etienne, Dr. Rositsa Karamanska, Dr. Jihye Jang-Lee, Dinah Rahman, Grigorji Sutov, Gang Wu, Qiushi Chen, Nan Jia, Matthew Choo, Tiandi Yang and Laura Bouché - I will never forget you.

A big thank you to all my friends who have offered welcome breaks and kept me thoroughly entertained, throughout my PhD, I appreciate all of you so much. A big shout out goes to Maudy, Kika, Kiki, Kiwi, Alla, Marina, Maria, Rita, Jason, Esther, Kati, Räbi and all the other IC survivors from the HC and the Union; life would not be the same without you all.

Thank you to my amazing family; my hardworking and courageous grandparents, who moved their lives to give their children more possibilities, my aunts Ariane and Monika for all your advice and everyone else that makes up my wonderful family. I am so grateful for all the crazy, loud and happy moments over the years and those to come.

Finally and most importantly, the inner circle. I am unbelievably grateful to share my life's journey with you all. Each of you inspires me every day. Mammuri and Pebbl, thank you so much for all your support at every stepping stone of my life, your endless love and affection, never doubting me and always making me believe that I could do anything in the world – you encourage me to be the best version of myself. My brothers Manuel and Luis, your determination and courage motivate me in all my aspirations and I could not wish for better companions on this adventure.

Above all, my deepest love and gratitude goes to my best friend and partner in crime, Virya. You have not only been there for me throughout all the ups and downs of my scientific career, but also supported me at every hurdle that comes with life over the last few years. Thank you for all the laughter, fun and experiences; you bring sunshine to my life. You have taught me so much, most importantly finding my own strength and for that I will be forever grateful.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

2-AB	2-aminobenzamide
α-L-Ara4N	4-amino-4deoxy-L-arabinose
ABC	ATP-binding cassette
ACDP	Advisory Committee on Dangerous Pathogens
AgI	Archaeal glycosyl transferase
AMAC	2-aminoacridone
API	Atmospheric pressure ionisation
ArtA	Archaeosortase A
BM	*Burkholderia mallei*
BP	*Burkholderia. pseudomallei*
BT	*Burkholderia thailandensis*
CAD	Collisionally activated dissociation
CAMPs	Cationic antimicrobial peptides
CDC	US Centres for Disease Control and Prevention
CI	Chemical ionisation
CID	Collision-induced dissociation
CMP	Cytidine monophosphate
COSY	Correlated spectroscopy
CPS	Capsular polysaccharides
CTL	Cytotoxic T cells
CZE	Capillary zone electrophoresis
DABP	Diaminobenzophenone
DATDH	2,4-diacetamido-2,4,6-trideoxyhexose
DHB	2,5-dihydroxybenzoic acid
Dol-P	Dolichol phosphate
Dol-PP	Dolichol pyrophosphate
Dol-P-P-GlcNAc	Dolichol-pyrophosphate N-acetylglucosamine
DPC	Perdeuterated dodecylphosphocholine
dstl	Defence Science and Technology Laboratory
DTT	Dithioreitol
ECD	Electron capture dissociation
EI-MS	Electron impact mass spectrometry
EPS	Expopolysaccharides
ER	Endoplasmic reticulum
ESI-MS	Electrospray ionisation mass spectrometry
FAB-MS	Fast atom bombardment mass spectrometry
FD	Field desorption
FID	Free induction decay
FT-ICR	Fourier transform ion cyclotron resonance
GC	Guanine-cytosine
GC-MS	Gas chromatography mass spectrometry
HABA	2-(4'-hydroxylbenzeneazo) benzoic acid
HCCA	α-cyano-4-hydroxy cinnamic acid
Hep	L-glycero-D-manno-heptose
HMBC	Heteronuclear multiple-bond spectroscopy
HPLC	High performance liquid chromatography
HSQC	Heteronuclear single-quantum coherence
IAA	Iodoacetic acid
IDA	Information-dependent-acquisition
IFN-γ	Interferon-γ
IL	Interleukin
ISD	In-source decay
Kdo	3-deoxy-D-manno-oct-2-ulosonic acid
Ko	D-glycero-D-talo-octu-2-ulsonic acid
Leg	Legionaminic acid
LINAC	Linear accelerating high pressure collision
LIT	Linear ion trap
LLO	Lipid-linked oligosaccharide

LolC	Lipoprotein-releasing system transmembrane protein
LPS	Lipopolysaccharide
MAC	membrane attack complex
MALDI-MS	Matrix assisted desorption ionisation mass spectrometry
MDR	Multidrug resistance
MOD	Ministry of Defence
M_r	Relative molecular weight
MS/MS	Tandem mass spectrometry
MTTD	Mean time to death
MWCO	Molecular weight cut off
MyD88	Myeloid differentiation primary response gene 88
NHS	N-hydroxysuccinimide
NIH	US National Institutes of Health
NK	Natural killer
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
OmpA	Outer membrane protein A
ORF	Open reading frames
OST	Oligosaccharyltransferase
PAMP	Pathogen-associated-molecular-pattern
PD-MS	Plasma-desorption mass spectrometry
Pgl	Protein glycosyl transferase
ppGalNAcT	Polypeptide-N-acetylgalactosaminyltransferase
PRR	Pattern recognition receptors
PSD	Post-source decay
Pse	Pseudaminic acid
QIT	Quadrupole ion trap
QS	Quorum-sensing
Q-TOF	Quadrupole with time-of-flight
RF	Radiofrequency
SA	Sinapinic acid
SDS-PAGE	Sodium dodecyl suflate-polyacrylamide gel electrophoresis
S-layer	Surface-layer
SP	Signal peptidase
T3SS	Type III secretion system
T6SS	Type VI secretion system
Th	T helper
TIS	Time-ion selector
TLR	Toll-like receptors
TMS	Trimethylsilyl
TNF-α	Tumour necrosis factor α
TOCSY	Total correlation spectroscopy
TOF	Time-of-flight
TRIF	TIR-domain-containing adaptor-inducing interferon-β
Und-P	Undecaprenyl phosphate

Chapter 1
Introduction

1 INTRODUCTION

1.1 GLYCOSYLATION AND ITS SIGNIFICANCE ACROSS THE THREE DOMAINS OF LIFE

1.1.1 Overview

Glycosylation is the ubiquitous and covalent attachment of carbohydrates to proteins and lipids. All domains of life express an abundance of glycopolymers on their cell surfaces where they play important roles in cell-cell interactions. Of the many post-translational modifications proteins can undergo, glycosylation is the most prevalent and the most diverse. Today, it is clear that both N-glycosylation and O-glycosylation, once believed to be restricted to eukaryotes also occur in bacteria and archaea (Eichler and Adams, 2005, Messner, 2004, Abu-Qarn et al., 2008a). In fact, prokaryotic glycoproteins and glycolipids rely on a wider variety of monosaccharide constituents than those of eukaryotes. In recent years, substantial progress in describing the enzymes involved in bacterial and archaeal glycosylation pathways has been made. It is becoming clear that enhanced knowledge of bacterial glycosylation enzymes may be of therapeutic value as this post-translational modification is seen with increasing frequency in pathogenic species (Szymanski and Wren, 2005), while the demonstrated ability to introduce bacterial glycosylation genes into *Escherichia coli* represents a major step forward in glycoengineering.

The three-domain system as a biological classification was introduced by Carl Woese in 1977 (Woese and Fox, 1977). It divides cellular life forms into archaea, bacteria and eukaryote domains (**Figure 1.1**). In particular, it emphasises the separation of prokaryotes into two groups, originally called eubacteria (now bacteria) and archaebacteria (now archaea). Woese argued that since there was an increasing emphasis on molecular level comparisons of genes, initially on the basis of differences in their 16S rRNA genes as the primary factor in their classification, genetic similarity is more important than outward appearance and behaviour. Furthermore the eukaryotic genetic make-up at the molecular level related it more closely to the archaebacteria, rather than the eubacteria.

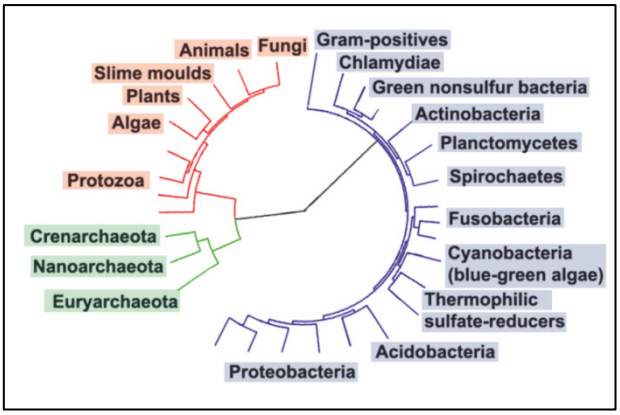


Figure 1.1 – Phylogenetic tree of life.
This image of a highly resolved tree of life, was first published by Ciccarelli and colleagues who generated the image based on completely sequenced genomes (Ciccarelli et al., 2006). Archaea are highlighted in green and eukaryotes are highlighted in red, whilst bacteria are highlighted in blue.

With advances in molecular biology and improved analytical techniques a significant change of perception has taken place regarding prokaryotic glycoproteins. Glycosylation of proteins from prokaryotes is no longer considered an uncommon feature of a few organisms but has been demonstrated for many archaea and bacteria. So far, several differences in structure and biosynthesis have been observed in comparison to eukaryotic glycoproteins. Further, many prokaryotes produce polysaccharide arrangements (bacterial capsules, cell wall peptidoglycan or lipopolysaccharides) with extraordinary diversity in composition and structure. These glycopolymers are typically much more complex than those produced by plants and animals and are often partially composed of unusual monosaccharides rarely seen elsewhere in nature. Over 100 different monosaccharides have been found in bacteria and their polysaccharides are often immunogenic to humans and are responsible for the specific immunological properties of a bacterial strain or type.

Currently, heightened attention is focused on protein glycosylation in bacteria, primarily because of the increasing frequency with which this post-translational modification is seen in pathogenic species. Accordingly, research conducted in several laboratories over the past decade has served to decipher the N-glycosylation pathway of the intestinal pathogen *Campylobacter jejuni*, making it the first bacterium for which such a complete pathway has been described (Szymanski et al., 2003, Szymanski et al., 1999). In addition, O-glycosylation has been shown to occur in several other bacterial pathogens, such as *Neisseria gonorrhoea* and *Helicobacter pylori*, with glycosylation-defective mutants displaying attenuated virulence-associated properties (Szymanski and Wren, 2005). Such pathways of protein glycosylation in these pathogens, may offer novel therapeutic targets, and the achievement of expressing the *C. jejuni pgl* gene cluster in *E. coli*, raises the hope that protein N-glycosylation will be added to the menu of biological and biotechnological processes (Wacker et al., 2006, Wacker et al., 2002). Much less is known of the steps involved in this post-translational modification in archaea, despite the fact that the first non-eukaryotic N-glycosylated protein was discovered over three decades ago in the haloarchaeon *Halobacterium salinarum* (Mescher and Strominger, 1976) and that N-linked glycoproteins are more prevalent in archaea than in bacteria (Eichler and Adams, 2005, Messner, 2004). These issues are addressed in more detail in the following sections (**Sections 1.1.2.1, 1.1.2.2 and 1.1.2.3**). First, though, key features of eukaryotic glycosylation are summarised so that similarities and differences in prokaryotic glycosylation can be more easily discussed.

1.1.1.1 N-linked glycosylation in eukaryotes

N-linked glycans in eukaryotic glycoproteins are attached to the amide nitrogens of asparagine side chains in the consensus sequon Asn-X-Ser/Thr, where X can be any amino acid except proline (Marshall, 1974). Nearly all N-glycans share a common core sugar sequence (Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ1-Asn-X-Ser/Thr. N-glycans of higher eukaryotes can be classified into oligomannose, in which only mannose residues are attached to the core, complex N-glycans, where antennae are attached to the core, and finally hybrid N-glycans in which only mannose residues are attached to the Manα1-6 arm of the core and one or two antennae are on the Manα1-3 arm (Varki et al., 2009).

The general biosynthesis of eukaryotic N-glycans begins on the cytoplasmic face of the endoplasmic reticulum (ER) membrane with the transfer of GlcNAc-P from UDP-GlcNAc, catalysed by GlcNAc-1-

phosphotransferase, to the lipid-like precursor dolichol-phosphate (Dol-P) to generate dolichol pyrophosphate N-acetylglucosamine (Dol-P-P-GlcNAc). Dolichol is a polyisoprenol lipid comprised of five-carbon isoprene units linked linearly in a head-to-tail fashion. The number of isoprene units in dolichol varies within cells, cell types and organisms. The presence of 19 isoprene units means that the hydrophobic portion of this lipid is far longer than the fatty acid tails on membrane phospholipids, and this extended hydrophobic region is inserted into the lipid bilayer possibly in a helical or folded conformation. A second GlcNAc and five mannose residues are subsequently transferred in a stepwise manner from UDP-GlcNAc and GDP-Man, to generate $Man_5GlcNAc_2$-P-P-Dol on the cytoplasmic side of the ER. The $Man_5GlcNAc_2$-P-P-Dol precursor is transferred across the ER membrane bilayer so that the glycan becomes exposed to the lumen of the ER, by a "flippase" (Helenius et al., 2002). $Man_5GlcNAc_2$-P-P-Dol is then extended in the ER lumen by the addition of four mannose residues transferred from Dol-P-Man. Assembly of the Dol-P-P-glycan precursor is completed with the addition of three glucose residues donated by Dol-P-Glc. Dol-P-Man and Dol-P-Glc donors are formed on the cytoplasmic side of the ER from GDP-Man and UDP-Glc by transfer of the respective sugar to Dol-P and must therefore also be flipped across the ER bilayer for use by the mannosyltransferases and glucosylatransferases that catalyse the synthesis of the mature N-glycan precursor $Glc_3Man_9GlcNAc_2$-P-P-Dol (Kornfeld and Kornfeld, 1985, Burda and Aebi, 1999). Once assembled the entire 14 sugar glycan is transferred *en bloc* from the lipid to a consensus sequon in a protein that is being synthesized and translocated through the ER membrane. A multisubunit protein complex in the ER membrane is responsible for catalysing this transfer. It is called the oligosaccharyltransferase (OST) and binds to the membrane-anchored Dol-P-P-oligosaccharide and transfers the glycan to the nascent protein by cleavage of the high-energy GlcNAc-P bond, releasing Dol-P-P in the process (Silberstein and Gilmore, 1996). Yeast OST complexes have been purified and are comprised of nine different membrane-bound subunits, where Stt3p seems to be the catalytic subunit, which is accompanied by a total of seven additional proteins whose roles remain poorly understood (Lennarz, 2007, Kelleher and Gilmore, 2006, Alaimo et al., 2006, Schulz and Aebi, 2009). This process is depicted in **Figure 1.2a**.

The protein-bound N-glycan is subsequently remodelled in the ER and Golgi by a complex series of reactions catalysed by membrane-bound glycosidaseS and glycosyltransferases. The initial steps appear to be conserved among all eukaryotes and are now known to have key roles in regulating glycoprotein folding via interactions with ER chaperones that recognize specific features of the trimmed glycan. Processing or trimming of $Glc_3Man_9GlcNAc_2$-Asn begins with the sequential removal of glucose residues by α-glucosidases I and II. Before exiting the ER many glycoproteins are acted on by an ER α-mannosidase I, which specifically removes a terminal α1-2Man to yield a $Man_8GlcNAc_2$ isomer. A further α-mannosidase I-like protein has an important role in the recognition of misfolded glycoproteins, targeting them for ER degradation. Glycans, containing between five and nine mannose residues are called high mannose oligosaccharides (Varki et al., 2009).

Biosynthesis of hybrid and complex N-glycans in multicellular eukaryotes is initiated in the medial-Golgi by the GlcNAcT-I catalysed addition of GlcNAc to the 2-position of the mannose α1-3 in the core of $Man_5GlcNAc_2$. The majority of N-glycans are then trimmed by α-mannosidase II, to form $GlcNAcMan_3GlcNAc_2$. Once the last two mannose residues outside the core are removed, a second GlcNAc is added to the 2-position of the mannose α1-6 in the core by the action of a GlcNAcT-II to yield the

precursor for all biantennary complex N-glycans. Tri- and tetra-antennary glycans result from the initiation of additional antennae by GlcNAcT-IV and V. Hybrid glycans are formed if the $GlcNAcMan_5GlcNAc_2$ glycan is not acted on by α-mannosidase II, leaving the peripheral residues intact and unmodified in the mature glycoprotein. Complex and hybrid N-glycans may carry a "bisecting" GlcNAc that is attached to the β-mannose of the core by GlcNAcT-III. Further sugar additions, mostly occurring in the trans-Golgi, convert the limited repertoire of hybrid and branched N-glycans into an extensive array of mature, complex N-glycans. This includes sugar additions to the core, which in vertebrates is the addition of fucose as an α1-6 linkage to the GlcNAc adjacent to the Asn in the core. Invertebrates modify their cores more extensively with fucose, xylose and sometimes other sugars (Schiller et al., 2012). Elongation of the branching GlcNAc residues by sugar additions to produce the ubiquitous building block Galβ1-4GlcNAc is referred to as the "LacNAc" sequence. Capping or decoration of elongated branches involve the addition of sialic acid, fucose, GalNAc and sulphate to the branches (Varki et al., 2009).

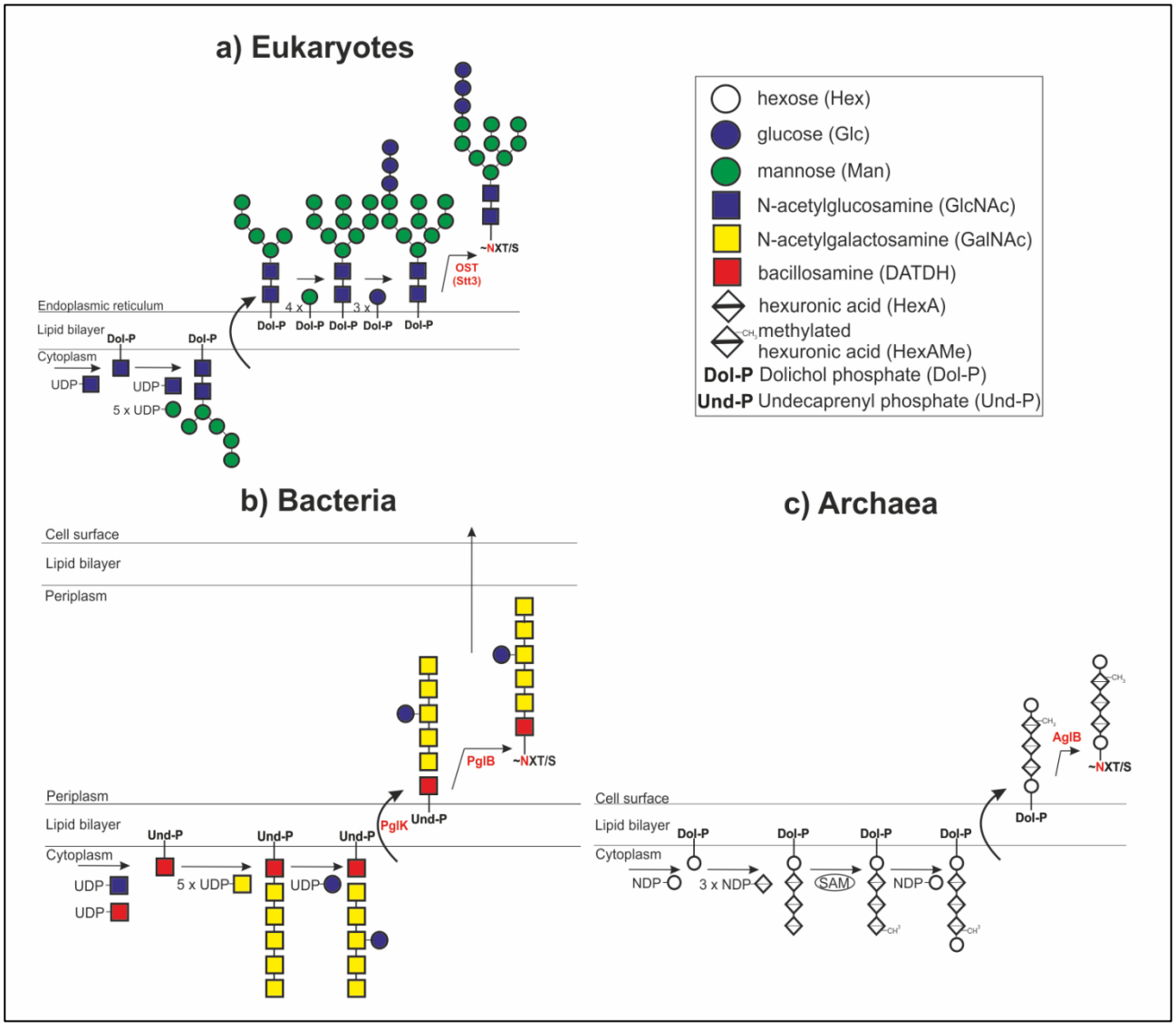


Figure 1.2 – General N-glycan biosynthesis pathways in the three domains of life.
The above figure depicts the similarities and differences in N-glycosylation pathways in eukaryotes, bacteria and archaea. The figure collates prokaryotic N-glycosylation pathways from studies in *H. marismortui* (archaea) and *C. jejuni* (bacteria) and was adapted from a review by Dell and colleagues (Dell et al., 2010)

The N-linked glycans of an individual glycoprotein are usually heterogeneous and glycoprotein molecules with a common polypeptide chain but bearing different glycans are called glycoforms. The extent of this heterogeneity varies from protein to protein. High mannose structures are ubiquitous in all eukaryotes but are particularly abundant in lower eukaryotes. A large number of mammalian glycoproteins bear complex N-linked glycans that exhibit heterogeneity in the degree of branching, the presence or absence of core fucose and bisecting GlcNAc residues, the presence or absence of polylactosamine extensions and a great diversity of terminal epitopes many of which are sialylated and/or fucosylated (Taylor and Drickamer, 2011). The complex type glycans of invertebrates are characterised by short antennae, highly modified cores, lack of sialylation and the presence of functionalities such as O-methyl and phosphorylcholine (Wilson, 2002, Haslam et al., 1997). In yeast and fungi, the high mannose glycans are extended rather than trimmed, and they serve a structural role, forming the outer wall of these single-celled eukaryotes. High mannose structures of N-glycans therefore represent an evolutionary precursor to the more complex glycans and a pre-existing pathway seems to have been adapted to generate glycans with more complex terminal elaborations needed for recognition purposes in multicellular organisms (Taylor and Drickamer, 2011).

1.1.1.2 O-glycosylation in eukaryotes

O-linked glycans are extremely diverse in both structure and function and the full extent of this diversity has not yet been established. In contrast to N-linked glycans, the groups of O-linked glycans are built on different protein-glycan linkages, in which sugars such as GalNAc, Fuc, GlcNAc, Man, Xyl, Gal and Glc can be attached to serine, threonine, tyrosine, hydroxyproline and hydroxylysine as is shown in **Figure 1.3** (Spiro, 2002, Elhammer et al., 1993).

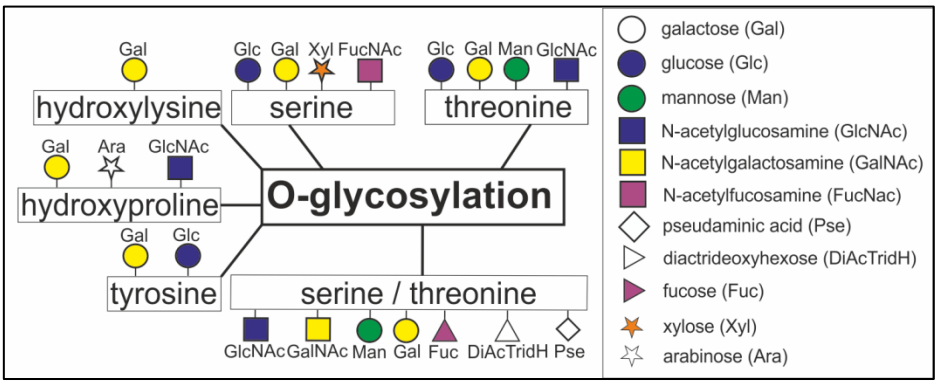


Figure 1.3 – O-linked glycosylation in eukaryotes.
The schematic depicts the variety of protein-glycan linkages used by eukaryotes in O-linked glycosylation as adapted from Spiro and colleagues (Spiro, 2002).

The best understood O-glycosylation in eukaryotes is mucin-type glycosylation in which O-glycans are covalently α-linked via a GalNAc moiety to serine or threonine. The first step of mucin-type O-glycosylation is the transfer of GalNAc from UDP-GalNAc to serine or threonine residues, which is catalysed by a polypeptide-N-acetylgalactosaminetransferase (ppGalNAcT), of which there are at least 21. Most of these enzymes are found in the Golgi and the subcellular localization of ppGalNAcTs and other glycosyltransferases involved in O-glycosylation has a critical role in determining the range of O-glycans synthesised by a cell. In contrast to N-glycosylation, no lipid-linked intermediates or glycosidases are involved in mucin biosynthesis and processing. The first sugar added to the protein in mucin biosynthesis

creates the Tn antigen (GalNAc-Ser/Thr). There are eight GalNAc-linked core structures (cores 1-8) in eukaryotes, most of which may be further substituted by other sugars. The ubiquitous mucin cores are 1 and 2 (Wopereis et al., 2006). GalNAc is converted to core 1 (Galβ1-3GalNAc, also called the T antigen) by a core 1 β1-3 galactosyltransferase. In many serum glycoproteins and mucins, the T antigen is substituted by sialic acids at C-3 of galactose and at C-6 of GalNAc, which add a negative charge to the glycan and prevent further modifications of core 1. Core 2 mucin O-glycans are branched core 1 structures that are produced in many tissues, including the intestinal mucosa. The synthesis of cores 3 and 4 (GlcNAcβ1-3GalNAc and its branched analogue, respectively) appears to be restricted mostly to mucous epithelia from the gastrointestinal and respiratory tracts and the salivary glands.

Although mucin-type O-glycosylation predominates in vertebrates, this is not the case in other eukaryotes where the alternative protein-glycan linkages mentioned earlier often predominate. Indeed, even in mammals, O-glycans attached via mannose, fucose or glucose, although rare, are important for certain functions (Varki et al., 2009).

1.1.2 Glycosylation in prokaryotes

Prokaryotes produce a variety of glycoconjugates and polysaccharides of enormous structural diversity and complexity. These glycans include many unusual sugars not found in vertebrates, including 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo), mannoheptose (Hep) and variously modified hexoses, which have important roles in the biology and pathogenicity of bacterial cells. Protein glycosylation was first demonstrated in the late 1930's and was long thought to exist only in eukaryotes, but in 1970 the S-layer glycoproteins of *Halobacterium salinarum* (Mescher and Strominger, 1976) and some *Clostridium* species were found to be glycosylated (Messner, 2004). Diverse N- and O-linked glycans have now been identified and characterised in archaeal species (predominantly N-glycans) and in many bacteria (more O-glycans). The proteins modified in this way are found in the surface layer, the outer and inner membranes, the periplasm and pili and flagella or secreted from the cell. The glycans also undergo other modifications such as sulphation, acetylation and methylation. These glycoproteins have important roles including maintenance of shape, adhesion and antigenicity.

1.1.2.1 Protein glycosylation in bacteria

1.1.2.1.1 Overview

Although bacterial glycans are more diverse than their eukaryotic counterparts, there are many similarities in the structures and biosynthetic pathways in mucosal pathogens, indicating that these modifications might have similar roles in multiple organisms (Szymanski and Wren, 2005). Studies of the biosynthetic pathways of prokaryotic glycoproteins are just beginning, but genetic information suggests that these systems range in complexity. For example, several species of *Campylobacter* contain gene clusters encoding a "general" glycosylation system that modify various proteins, whereas in other cases, the genes whose products modify a specific protein are located adjacent to the gene encoding the protein itself. *C. jejuni* has N- and O-linked glycosylation systems and is the cause of bacterial gastroenteritis worldwide and is thus used as a model

organism in bacterial glycosylation (Szymanski and Wren, 2005). The first protein N-glycosylation pathway was elucidated in *C. jejuni* and a cluster of 12 genes called the *pgl* locus (*p*rotein *gl*ycosylation) was found to be responsible for the synthesis of a heptasaccharide on bactoprenol and the transfer of the chain *en bloc* to asparagine residues. Discovery of the O-linked flagellin glycosylation system was based on the results of studies demonstrating that *Campylobacter* flagellins are sensitive to periodate oxidation and that they bind to a sialic-acid-specific lectin. The latter was shown to be due to the presence of pseudaminic acid and legionamic acid (analogs of sialic acid) decorating the flagella (Doig et al., 1996). Since then a wealth of information has been gathered about this system through studies in the native host, heterologous expression in *E. Coli* and *in vitro* analyses (Linton et al., 2005, Szymanski et al., 1999, Wacker et al., 2002, Young et al., 2002).

1.1.2.1.2 N-glycosylation in bacteria

The N-linked protein glycosylation system in bacteria was originally thought to be involved in lipopolysaccharide (LPS) biosynthesis, but was subsequently shown to have a role in the glycosylation of multiple *C. jejuni* proteins (Szymanski et al., 1999). The N-glycans on *C. jejuni* glycoproteins were found to comprise a heptasaccharide and its biosynthetic precursors. The heptasaccharide is linked to Asn via 2,4-diacetamido-2,4,6-trideoxyhexose (DATDH, also called bacillosamine) and has the sequence GalNAc-α1,4-GalNAc-α1,4-[Glcβ1,3]-GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-DATDH-β1). In *C. jejuni*, the pgl gene cluster encodes, among others, five putative glycosyltransferases (PglA, PglC, PglH, PglI and PglJ) involved in the assembly of the heptasaccharide on a lipid carrier and PglB, responsible for the *en bloc* transfer of the glycan from the carrier to protein, in a pathway similar to that employed in eukaryotes (**Figure 1.2**) (Kelly et al., 2006, Linton et al., 2005). The central enzyme of the Pgl system is the oligosaccharyltransferase (OST) PglB, an integral membrane protein of 82kDa, which shares significant primary sequence similarity with Stt3 of eukaryotic cells (Szymanski et al., 1999, Wacker et al., 2006, Wacker et al., 2002, Yan et al., 2003). PglB contains a conserved carboxy-terminal catalytic motif Trp-Trp-Asp-Tyr/Trp-Gly, of which the second and third amino acids are essential (Yan and Lennarz, 2002). In the *C. jejuni* OST, the eukaryotic sequon, N-X-T/S is N-terminally extended to Asp-Glu-X_1-Asn-X_2-Ser/Thr, where X_1 and X_2 can be any amino acid except proline (Kowarik et al., 2006), and although D-Q-N-A-T is the optimal bacterial acceptor sequence, not all sequons are glycosylated (Chen et al., 2007).

In *C. jejuni* the heptasaccharide is built on the cytoplasmic side of the inner membrane on a lipid-linked precursor undecaprenyl phosphate (Und-P) (Linton et al., 2005, Reid et al., 2009, Alaimo et al., 2006, Kelly et al., 2006). The undecaprenyl pyrophosphate-linked heptasaccharide is assembled in the cytosol by the sequential addition of sugars from nucleotide-activated donors. The resulting lipid-linked oligosaccharide (LLO) is then translocated across the inner membrane into the periplasmic space by protein glycosylation K (PglK), an ATP-binding cassette (ABC)-type transporter “flippase” (Alaimo et al., 2006). The oligosaccharide is transferred to the amino group of Asn in the protein consensus sequence, or released into the periplasm as free oligosaccharides by the OST PglB (**Figure 1.2b**) (Wacker et al., 2002, Nothaft and Szymanski, 2010).

Using synthetic glycan donors and a short acceptor peptide it was shown that PglB not only accepts saccharides of various lengths (2-7 sugars) *in vitro* but also accepts different sugars at the reducing end (Glover et al., 2005). X-ray crystallography and NMR of a *C. jejuni* glycoprotein showed glycosylation of a surface-exposed flexible loop that would be expected to be accessible to PglB (Nothaft and Szymanski, 2010). Many diverse proteins are modified by the N-glycosylation pathway in *C. jejuni*, so it is not surprising that disruption of this pathway has pleiotropic effects for the bacterium, including reduced protein immunoreactivity with both human and animal sera (Szymanski and Wren, 2005, Szymanski et al., 1999), a reduced ability to adhere to and invade human epithelial cells *in vitro* and a decrease in mouse and chicken colonisation *in vivo* (Nothaft and Szymanski, 2010, van Sorge et al., 2009, Szymanski et al., 2002, Hendrixson and DiRita, 2004, Jones et al., 2004).

The availability of an increasing number of complete bacterial genome sequences has revealed that organisms encoding orthologs of the *C. jejuni* N-glycan pathway constituents are much more widespread than originally predicted. Although the glycosyltransferase content and *pgl* gene distributions vary, the characteristic enzyme of the pathway, PglB is present in at least 49 species, with five species possessing two copies of *pglB* and all homologs containing the conserved Trp-Trp-Asp-Tyr/Trp-Gly motif that has been shown to be indispensable for PglB function in *C. jejuni* (Wacker et al., 2002, Szymanski and Wren, 2005). The apparent diversity in *pgl* gene distribution contrasts with the partial conservation of the neighbouring chromosomal DNA sequences and suggests that the *pgl* cluster has evolved under selective pressure in its host environment, a typical feature of pathogenicity-related gene clusters (Nothaft and Szymanski, 2010). The N-linked glycosylation locus is highly conserved among the proteobacteria and lacks putative phase-variable genes and N-linked glycosylation takes place in the periplasm so that only proteins secreted across the inner membrane will become N-glycosylated suggesting that N-linked glycosylation allows for protection against proteolytic cleavage, enhancement of protein stability or signals for cellular sorting similar to eukaryotic N-linked glycans (Helenius and Aebi, 2001, Marceau and Nassif, 1999).

1.1.2.1.3 O-glycosylation in bacteria

O-Glycosylation is more widespread in bacteria than N-glycosylation and a detailed review of this type of glycosylation is outside the scope of this thesis. Some common themes are exemplified by the examples given below.

C. jejuni modifies its flagellar proteins with unusual O-linked monosaccharides, which can constitute up to 10% of the protein mass. The modifications are necessary for flagellum assembly and affect secretion of virulence-modulating proteins, bacterial colonisation of the gastrointestinal tract, auto-agglutination and biofilm formation (Guerry, 2007). The predominant sugars attached to the C*ampylobacter* flagellum are derivatives of pseudaminic acid (Pse) or legionaminic acid (Leg), which are analogues of sialic acid. The flagellar apparatus of *C. jejuni* spans both the inner and outer membranes and O-linked glycoyslation of flagellin monomers is thought to occur at the cytoplasm interface. Cytidine monophosphate (CMP)-activated sugars including CMP-pseudaminic acid, CMP-legionaminic acid and their derivatives are individually added to surface exposed serine or threonine residues in the flagellin subunit, by glycosyltransferase FlaA (Nothaft and Szymanski, 2010). *P. aeruginosa* O-glycosylate their proteins in a similar manner to *C. jejuni* and their

pili contain O-linked pseudaminic acid derivatives, as do proteins of other Gram-negative pathogens, such as *H. pylori* and *C. coli* (Abu-Qarn et al., 2008a).

In contrast to the above organisms *Neisserial* pathogens have an O-glycosylation machinery that has some of the hallmarks of protein N-glycosylation in *C. jejuni*. The pili of *N. meningitidis* (Stimson et al., 1995) and *N. gonorrhoea* (Hegge et al., 2004) contain galactosylated O-glycans linked to serine via DATDH, the same sugar that attaches the *C. jejuni* N-glycans to Asn (**Section 1.1.2.1.2**). The precursor O-glycan is assembled on the cytoplasmic side of the inner membrane on a lipid anchor, flipped into the periplasm where an O-oligosaccharyltransferase, PglL then transfers the glycan *en bloc* to the protein acceptor (Nothaft and Szymanski, 2010, Faridmoayer et al., 2007). This is extremely unusual, since all other O-glycans studied to date are assembled via sequential addition of their monosaccharide constituents directly onto a protein acceptor (Aas et al., 2007, Dell et al., 2010).

Another class of O-glycoproteins in bacteria are the surface-layer (S-layer) proteins. For glycobiologists S-layer glycoproteins are of historical importance because they were the first prokaryotic glycoproteins to be identified over 40 years ago (Mescher and Strominger, 1976, Sleytr and Thorne, 1976), albeit the “bacteria” then studied were later classified as archaea (**Section 1.1.1**). Bacterial S-layer glycoproteins have been mainly found on Gram-positive organisms although recently they have been identified on a couple of Gram-negative species (Ristl et al., 2011). Bacterial S-layer O-glycans have been found attached to tyrosine as well as serine and threonine. They contain both common and rare sugars as well as functionalities such as sulphate and phosphate. They are often comprised of repeating units reminiscent of bacterial polysaccharides (**Section 1.1.2.2**) and are therefore usually much larger than the O-glycans described above.

1.1.2.2 Polysaccharides in bacteria

1.1.2.2.1 Overview

Bacterial polysaccharides are typically much more complex than those produced by plants and animals and they often contain unusual monosaccharides seldom seen elsewhere in nature. Bacterial polysaccharides have a variety of functions including roles as adhesion receptors, protection from desiccation, and resistance to infection by phage and resistance to destruction by their vertebrate hosts. Microbial polysaccharides are used in the food industry as thickeners, emulsifiers, gelling or viscosifying agents. Pathogenic bacteria commonly produce a thick, mucous-like layer of polysaccharide, which is like a capsule that cloaks antigenic proteins on the surface of the bacteria, which would otherwise produce an immune response. Capsular polysaccharides are usually water soluble, acidic and consist of regularly repeating subunits of 1-6 monosaccharides. The assembly of capsules involves genes that are clustered in the bacterial chromosome in three continuous regions. This arrangement allows a simple mechanism for changing capsule types by merely swapping different serotype cassettes (Varki et al., 2009). LPS are the major constituent of the outer membrane of most Gram-negative bacteria. This class of bacterial polysaccharide is discussed in more detail below (**Section 1.1.2.2.2**) because one of the projects (**Chapter** 4) in this thesis is devoted to LPS characterisation.

Lipid A constitutes the hydrophobic component of LPS (**Figure 1.4**) which is located in the outer leaflet of the outer membrane. Attached to lipid A moieties, and displayed on the surface of bacterial cells, are polysaccharides composed of core oligosaccharides and variable numbers of O-antigen repeats (Raetz et al., 2007, Raetz and Whitfield, 2002). Lipid A comprises fatty acid residues linked both as esters and amides to C-3, C-3' and C-2, C-2' respectively of the β-1-6-linked-2-amino-2-deoxy-D-glucopyranose disaccharide. The C-1 and C-4' of the backbone disaccharide are phosphorylated whereas the C-6' is the site of attachment of the carbohydrate chain. The fatty acid residues that are directly linked to the disaccharide backbone (primary acyl groups) are 3-hydroxylated and hence can be esterified with non-hydroxylated or sometimes 2-hydroxylated fatty acids (secondary acyl groups) (Wang and Quinn, 2010). Lipid A usually constitutes a family of molecular species that differ in the type and/or the number of fatty acid residues. More frequently occurring fatty acids are 12:0 (3-OH), 14:0 (3-OH), 16:0 (3-OH), 10:0 (3-OH), 12:0 and 14:0. Longer-chain fatty acids as well as fatty acids with an odd carbon number or branched acyl groups have also been reported. Other possible sources of lipid A heterogeneity are oxidation of a 3-hydroxyl group of a fatty acid to a keto group, lack of a phosphate group (usually at C-4') or the presence of further substituents on the phosphate groups, such as a monosaccharide, most commonly 4-amino-4deoxy-L-arabinose (α-L-Ara4N). Lipid A is responsible for the major bioactivity of endotoxin and is recognised by the immune system as a pathogen-associated molecule, but the detailed structure of LPS varies from one bacterium to another and this variation could affect the virulence of the bacterium (Galanos et al., 1985).

The carbohydrate moiety of LPS consists of two structurally and biosynthetically different parts, the core region, found in all LPS (Ogawa et al., 2007) and the O-antigen. LPS is designated as S or R type depending on whether O-antigen is present or not (Raetz and Whitfield, 2002). These are so-called because the presence or absence of the O-antigen determines the shape of the bacterial colony, which appears as smooth (S) or rough (R), respectively. The core oligosaccharide consists of up to 15 monosaccharide residues, and inner and outer moieties may be identified. The inner part generally begins with a 3-deoxy-D-manno-oct-2-ulsonic acid (Kdo) residue, which is linked at C-6' of the non-reducing glucosamine residue of lipid A. This first Kdo residue is either 4-phosphorylated or 4-glycosylated with the second Kdo residue or as in *B. cepacia*, a D-glycero-D-talo-octu-2-ulsonic acid (Ko) residue (Toman et al., 2004). In many bacteria the inner part contains several L-glycero-D-manno-heptose (Hep) residues (Ogawa et al., 2007). The outer core contains more traditional monosaccharides and is more structurally and compositionally diverse (Muller-Loennies et al., 2003). The core oligosaccharide often shows heterogeneity owing to the presence of phosphate substituents and/or some constituent monosaccharides in non-stoichiometric amounts.

The O-antigen is the most variable part of the LPS and usually represents a regular polysaccharide built up of repeating units consisting of one to six monosaccharide residues, whilst some O-antigens are homopolymers. The heterogeneity of the O-antigens with respect to the number of repeating units is responsible for the ladder-like profile of the smooth LPS in SDS-PAGE gels. Further decorations of the O-antigens with non-sugar components, such as O-acetyl groups, various N-acyl groups, ether-linked methyl and carboxymethyl groups, pyruvic acid acetals, amino acids linked to carboxyl groups of hexuronic acids, amino alcohols linked via phosphate groups and other modifications are common. Sugar compositions of the O-antigens seem to have a relationship with the bacterial environment, for instance, 6-deoxy sugars appear to be prevalent in phytopathogenic bacteria (Toman et al., 2004), whereas acidic monosaccharides are

frequent in marine bacteria (Suits et al., 2008). **Figure 1.4** shows a cartoon schematic of a lipopolysaccharide and modifications.

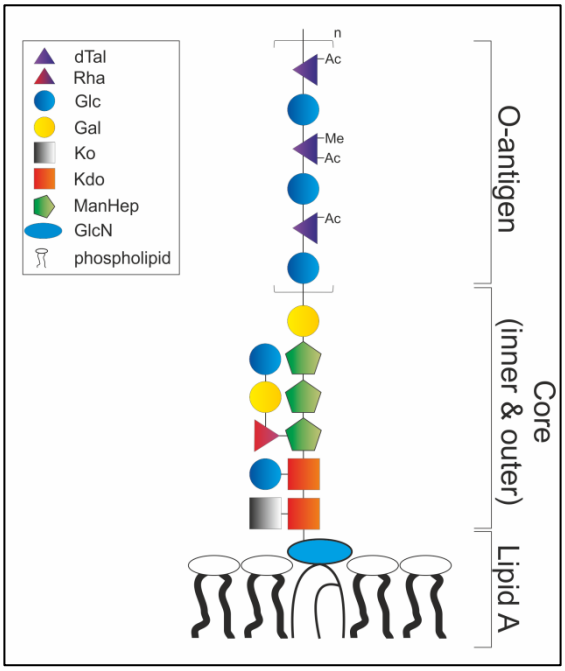


Figure 1.4 – Cartoon representation of a lipopolysaccharide and modifications.
The schematic shows a cartoon version of a lipopolysaccharide and its modifications, such as is found in *B. pseudomallei* K96243. The O-antigen is only found in smooth LPS and is lacking in rough LPS, whilst the core oligosaccharide is composed of an inner and outer core and the lipid A is embedded in the cell membrane.

Most enzymes and genes coding for proteins responsible for the biosynthesis and export of LPS in *E. coli* have been identified and, based on genetic information, many are shared by most Gram-negative bacteria (Steeghs et al., 1997). Nevertheless, the detailed structure of LPS differs from one bacterium to another, consistent with the recent discovery of additional enzymes and gene products that can modify the basic structure of LPS in some bacteria, especially pathogens. These modifications are not required for survival but are tightly regulated in the cell and closely related to the virulence of bacteria.

1.1.2.2.2 Biosynthesis of LPS

The first stage of the LPS biosynthetic pathway in *E. coli* is the synthesis of Kdo_2-lipid A (Doerrler, 2006, Raetz and Whitfield, 2002). This pathway is mediated by nine enzymes and takes place in the cytoplasm and on the inner surface of the inner membrane. Three soluble enzymes (LpxA, LpxC and LpxD) catalyse the addition of two 3:OH fatty acid chains to the C-2 and C-3 of the UDP-GlcNAc to form UDP-diacyl-GlcN (Buetow et al., 2007, Williams and Raetz, 2007). LpxH then hydrolyses UDP-diacyl-GlcN to form lipid X and LpxB condenses lipid X and its precursor UDP-diacyl-GlcN to form disaccharide-1-P (Crowell et al., 1987), whilst LpxK is a kinase that phosphorylates the C-4' of the disaccharide-1-P to form lipid IV_A (Garrett et al., 1998). KdtA is a bifunctional enzyme that incorporates two Kdo residues at the C-6' of the lipid IV_A using the sugar nucleotide CMP-Kdo as a donor (Brozek et al., 1989). Kdo_2-lipid IV_A undergoes further reactions catalysed by LpxL and LpxM to form Kdo_2-lipid A (Brozek and Raetz, 1990). The nine enzymes involved in the biosynthesis of Kdo_2-lipid A all have relatively high specificity for their respective substrates. LpxA, LpxD, LpxL and LpxM are all acyltransferases, but they selectively catalyse different substrates and employ

different acyl donors. The minimal LPS structure needed for the viability of *E. coli* is lipid IV_A although such *E. coli* mutants exhibit highly attenuated growth. LpxL adds a secondary lauroyl residue and LpxM adds a myristoyl residue to the distal glucosamine unit (Wang and Quinn, 2010).

The core oligosaccharides are sequentially assembled on lipid A at the cytoplasmic surface of the inner membrane in a process that involves a number of membrane-associated glycosyltransferases, using nucleotide sugars as donors. The biosynthesis of core oligosaccharides is rapid and efficient, suggesting that the glycosyltransferases function in a coordinated complex. In *E. coli* and *Salmonella*, genes required for the biosynthesis of core oligosaccharides exist in three operons: *gmhD*, *waaQ*, *kdtA* operons (Roncero and Casadaban, 1992). In *E. coli* K-12, the *gmhD* operon contains four genes *gmhD-waaF-waaC-waaL* that are required for the biosynthesis of the inner core oligosaccharides (Schnaitman and Klena, 1993). The *gmhD*, *waaF* and *waaC* genes encode proteins involved in the biosynthesis and transfer of Hep (Whitfield et al., 1997, Brozek et al., 1989).

The O-antigen is synthesised on a lipid carrier on the cytoplasmic surface of the inner membrane, using sugar nucleotides as donors, After synthesis, both core-lipid A and O–antigen are transported to the periplasmic face of the inner membrane, where the O-antigen is polymerised by Wzy and Wzz and ligated to the core-lipid A by an enzyme encoded by the w*aaL* gene, resulting in a nascent LPS (Abeyrathne et al., 2005, Raetz et al., 2007).

The transport of core-lipid A is carried out by a membrane protein MsbA (Doerrler et al., 2004). In *E. coli* MsbA is a homodimer and each monomer contains six transmembrane helices and a cytosolic ATP-binding domain, analogous to the ABC-type transporter-dependent pathway found in bacterial N-glycosylation (Ward et al., 2007). MsbA is highly conserved in Gram-negative bacteria and shares homology with the multidrug resistance (MDR) proteins of eukaryotes as such members of the large ABC-transporter superfamily are widely distributed in all domains of life. O-antigen transport across the inner membrane is mediated by Wzx protein (Alaimo et al., 2006, Raetz et al., 2007). Wzx proteins from different bacteria have similar hydropathy profiles and can complement each other in the translocation of different O-antigen sugar precursors but no sequence homology or conserved residues are found amongst Wzx proteins (Feldman et al., 1999). There is evidence that Wzx proteins might function by recognising the first sugar phosphate bound to Und-P (Marolda et al., 2004).

LptA, LptB, LptC, LptF and LptG shuttle the nascent LPS from the periplasmic face of the inner membrane to the inner surface of the outer membrane, where LptD and LptE assemble LPS into the outer surface of the outer membrane (Ruiz et al., 2008, Sperandeo et al., 2007, Sperandeo et al., 2008). It appears that some of these proteins may function as complexes (Bos et al., 2004). The ABC-transporter LptBFG functioning with LptC and LptA translocates LPS to the inner leaflet of the outer membrane (Sperandeo et al., 2007, Sperandeo et al., 2008). When LptA or LptB, or both were depleted, LPS was found to accumulate in the periplasm (Sperandeo et al., 2007). The outer membrane, nascent LPS is exported to the outer leaflet by LptD and LptE (Bos et al., 2004, Ma et al., 2008, Sperandeo et al., 2007, Sperandeo et al., 2008, Wu et al., 2006). The transport of the LPS into the periplasm and to the outer surface of the bacteria is described in **Figure 1.5**.

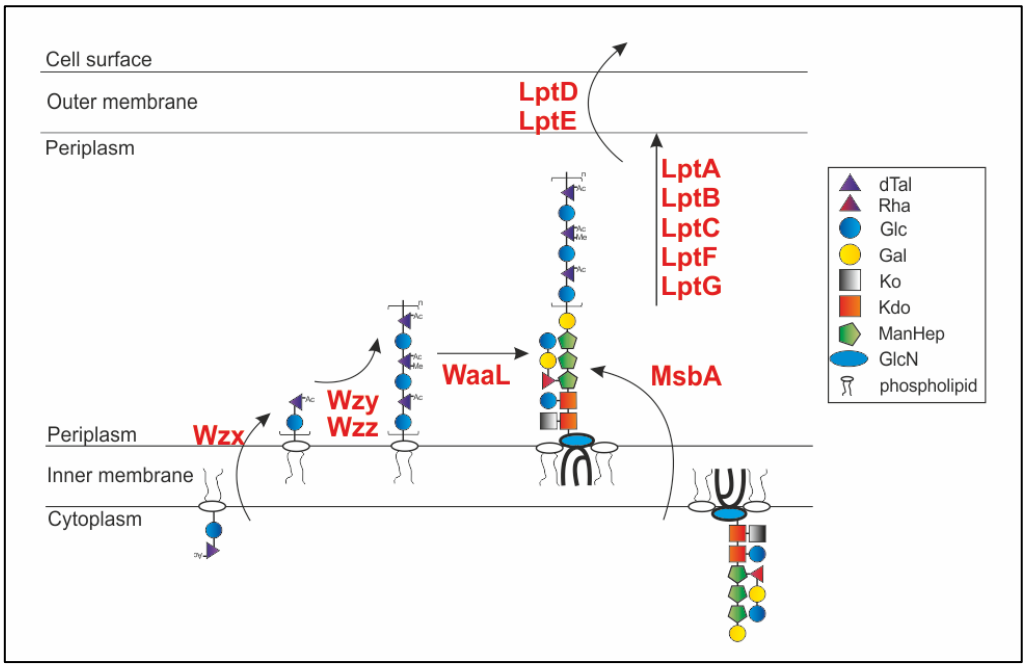


Figure 1.5 – LPS biosynthesis and transport across the bacterial membranes.
The schematic depicts the enzymes involved in O-antigen transport, polymerisation and Lipid A-core transport across the inner membrane as well as O-antigen and Lipid A-core ligation to produce the LPS. Further, the enzymes responsible for shuttling the LPS across the periplasm and delivery to the outer membrane are shown.

Bacteria have evolved mechanisms to modify their LPS structure, even the most conserved part, lipid A. It has been shown that modifications can involve both the hydrophilic disaccharide region as well as the hydrophobic acyl chain domain. Modifications to the lipid A usually occur at the periplasmic face of the inner membrane or in the outer membrane. The structure modification of lipid A might help the bacteria to resist the cationic antimicrobial peptides (CAMPs) released by the host immune system, or to evade recognition by the innate immune receptor TLR4 (Wang and Quinn, 2010). Some modifications of the lipid A are under the control of the PhoP-PhoQ system and/or PmrA-PmrB system (Guo et al., 1997). The PhoP-PhoQ is a two-component system that governs virulence, mediates the adaption to Mg^{2+}-limiting environment and regulates numerous cellular activities in Gram-negative bacteria (Gibbons et al., 2005, Guo et al., 1998). A PmrA-PmrB two-component system is also required for *S. enterica* virulence in mice. Several enzymes have been reported to modify the fatty acyl chain region of lipid A, such as the membrane proteins PgaP, PagL, LpxR and LpxO. The length of the fatty acid chains in the lipid A differs in different Gram-negative bacteria. For example, the fatty acyl chains of *E. coli* lipid A are 12 or 14 carbons long, while that of *F. novicida* lipid A are 16-18 carbons long (Bainbridge et al., 2008, Shaffer et al., 2007).

Changes to the hydrophilic region of lipid A can also occur. Lipid A usually contains two phosphate groups which impart a net negative charge to the molecule. The negative charges of lipid A allow the binding of positively charged CAMPs. To evade the attack by the immune system some bacterial pathogens have evolved less negatively-charged variation of lipid A. Modification to the hydrophilic region of lipid A focuses on the removal or decoration of the phosphate groups at the C-1 and C-4' and can include the addition of the amine-containing residues such as α-L-Ara4N and phosphoethalonamine. These modifications result in resistance to CAMPs and polymyxin B and are controlled by the PmrA-PmrB two-component system (Wang and Quinn, 2010). To lower overall phosphate, these groups can also be removed by *lpxE* and *lpxF* encoding proteins (Wang et al., 2004, Wang et al., 2006).

1.1.2.3 Protein glycosylation in archaea

1.1.2.3.1 Overview

Archaea are single cell prokaryotes that have rigid cell walls with diverse structures. They contain many unusual lipids with repeating isoprenyl groups linked to glycerol and an S-layer of glycoproteins in a lattice-like arrangement attached to the membrane. They lack the peptidoglycan found in almost all bacteria and instead contain a pseudomurein layer, which is similar to the peptidoglycan structure. Archaea have no nuclear membrane, possess unique ancient evolutionary history for which they are considered some of the oldest species of organisms on Earth and often characterised by living in extreme environments. Generally, archaea and bacteria are quite similar in size and shape; however despite these visual similarities to bacteria, archaea possess genes and several metabolic pathways that are more closely related to those in eukaryotes, notably the enzymes involved in transcription and translation. They exploit a much greater variety of energy sources than eukaryotes, which range from familiar organic compounds such as sugars, to using ammonia, metal ions or even hydrogen gas as nutrients. Salt-tolerant archaea, the halobacteria, use sunlight as a source of energy and other species of archaea fix carbon. They reproduce asexually and divide by binary fission, fragmentation or budding. Archaea are extremophiles that live in a broad range of habitats, such as soils, oceans and marshlands. Typical archaea include methanogens, which metabolise hydrogen and carbon dioxide into methane, halophiles that thrive in salt and thermoacidophiles, which grow in acid and high temperatures of up to 110ºC. Archaea express proteins that are able to remain properly folded and functional in the face of extremes of salinity, temperature and other adverse physical conditions that would normally lead to protein denaturation, loss of solubility and aggregation. No clear examples of archaeal pathogens or parasites are known, but they are often mutualists or commensals. They also have some importance in technology, with methanogens used to produce biogas and as part of sewage treatment, whilst enzymes from extremophiles are exploited in biotechnology for their ability to resist high temperature and organic solvents (Eichler and Adams, 2005).

Glycosylation in archaea, similarly to eukaryotes and bacteria, has structural roles. This fits with the role for the sulphated S-layer glycoprotein oligosaccharide chains in maintaining the rod shape of *Halobacterium salinarum* cells and it was noted that similarities exist in the overall structures of the S-layer glycoprotein and proteoglycans (Wieland, 1988, Bernfield et al., 1999). Furthermore, the glycosylation of archaeal proteins has been implicated in protein assembly and function. In archaeal flagellins, glycosylation is associated with proper flagellar assembly, since upon bacitracin-mediated interference with flagellin glycosylation, a loss of *Methanococcus deltae* flagellation was observed microscopically (Kalmokoff et al., 1992). Coping with the often harsh environmental conditions encountered by archaea serves as the basis for yet another hypothesised role of archaeal protein glycosylation. In comparison of the glycosylation profiles of S-layer glycoproteins from the moderate halophile *Haloferax volcanii* and the extreme halophile *H. salinarum*, it was noted that the latter experiences a higher degree of glycosylation than the former (Mengele and Sumper, 1992). Moreover, the glycan moieties of the extreme halophile were enriched in sulphated glucuronic acid subunits as opposed to the neutral sugars found in the moderate halophile, which provides the *H. salinarum* S-layer glycoprotein with a drastically increased surface charge density relative to its *H. volcanii* counterpart. The enhanced negative surface charge associated with protein glycosylation and the resulting protection

that this would afford in the face of acidic conditions have been offered as the role of *Sulfolobus acidocaldarius* cytochrome b558/566 glycosylation (Hettmann et al., 1998, Zahringer et al., 2000). It has also been suggested that a significant amount of the protein surface is shielded from the pH 2 environment by the high degree of glycosylation (Zahringer et al., 2000, Voorhorst et al., 1996). Glycosylation has also been implicated in the stabilisation of thermophilic archaeal glycoproteins.

1.1.2.3.2 N-glycosylation in archaea

Like eukaryotic glycosylation, archaeal N-glycosylation occurs on Asn-X-Ser/Thr consensus sequences and it has been proposed that the elaborate process responsible for N-glycosylation in eukaryotes originated from a simpler archaeal system (**Figure 1.2c**). Indeed, many of the fundamental steps and central components involved in eukaryal protein N-glycosylation are also present in archaea (Eichler and Adams, 2005). However there are important differences between the N-glycosylation machinery in the two kingdoms (**Figure 1.2**). For example although dolichyl pyrophosphate (Dol-PP) is the carrier in the cytoplasmic assembly of yeast LLO's, both dolichol phosphate (Dol-P) and Dol-PP have been found as carriers in archaea (Abu-Qarn et al., 2008a, Chaban et al., 2006, Wieland, 1988). Also, while OSTs are found across the range of archaeal phenotypes, the N-glycosylation machinery of hyperthermophilic archaea may well rely on fewer components than do the parallel systems.

Glycoproteins from about 25 species of archaea have been reported of which about ten have had their N-glycans partially or fully characterized (Abu-Qarn et al., 2008a, Chaban et al., 2006). The best understood are the S-layers of the halophiles *Halobacterium salinarum* and *Haloferax volcanii*, together with the S-layers and flagellins of several methanogens. An overview of halophilic S-layer glycosylation is given later (**Section 1.2.2**) because this is the subject of one of the projects in this thesis (**Chapter** 3). Here some general points about archael N-glycan structure are summarized. Most of the characterised archaeal N-glycans are unbranched and have fewer than six sugar residues. An exception to the latter is *H. salinarum* which, in addition to bearing a trisaccharide composed of monosulphated glucuronic acid linked via glucose at about ten consensus sites, has one N-linked site occupied by a GalNAc-linked polysaccharide comprised of multiple repeats of a sulphated pentasaccharide composed of GlcNAc, GalNAc, Gal, GalA, and 3-O-methyl-GalA (Paul and Wieland, 1987). The N-glycans attached to the S-layer of *Methanothermus fervidus* are hexasaccharides containing 3-O-methylmannose, Man, and GalNAc (Karcher et al., 1993). In *M. voltae* the flagellins and S-layer proteins are glycosylated with a tri- or tetrasaccharide composed of GlcNAc, GlcdiNAcA, a threonine-substituted ManNAcA and an uncharacterized sugar (Chaban et al., 2009, Voisin et al., 2005). An even more unusual tetrasaccharide has been found on the pilins and flagellins of *M. maripaludis*. This has the sequence Sug-4-ManNAc3NAmA6Thr-4-GlcNAc3NAcA-3-GalNAc where Sug is a highly unusual aldolase diglycoside (Kelly et al., 2009). In contrast to these linear glycans which contain a great diversity of monosaccharides, a small number of extreme thermophiles have been reported to have branched N-glycans comprised of sugars such as mannose and GlcNAc (Peyfoon et al., 2010, Yang and Haug, 1979). Interestingly *Sulfolobus* species have glycosylation bearing some of the hallmarks of high mannose eukaryotic glycans, including branched mannose and a chitobiose core (Peyfoon et al., 2010, Zahringer et al., 2000).

1.1.2.3.3 O-glycosylation in archaea

Extremely little is known of the O-glycosylation process in archaea, although S-layers containing N- and O-linked glycans have been found to occur, such as was found to be the case in both *H. salinarum* and *H. volcanii*. Thr-rich regions adjacent to the predicted membrane spanning domain of the S-layer glycoprotein, were found to be modified with galactose-glucose disaccharides (Mescher and Strominger, 1976, Sumper et al., 1990, Calo et al., 2010b). There has not been much investigation since into the nature of O-linked glycosylation in archaea.

1.2 THE HALOPHILIC ARCHAEON *HALOFERAX VOLCANII*

Halobacteria are located throughout the world in salt ponds and lakes and may exist in the form of dormant or living cells, biopolymers in rocks, salt crystals or as evaporates in desert regions. *Haloferax volcanii* was first identified by Mullakhanbhai and Larsen in 1975 as an organism that resides largely in the bottom sediment of the Dead Sea, being the predominant organism there. They found that the organism possessed the properties of the halobacteria, but differed from known species in the fact that its cells were disc shaped and often cupped when grown under optimum conditions, and that the optimum requirements for sodium chloride for the organism was in the range 1.7-2.5M, close to the salinity found in the Dead Sea, which is about half of that generally reported for the halobacteria. They further discovered that the tolerance for magnesium chloride was very high, allowing the organism to grow well in media containing magnesium chloride in the concentrations found in the Dead Sea. The organism was assigned to the genus *Halobacterium* and described as *Halobacterium volcanii* (Mullakhanbhai and Larsen, 1975) and was later re-named *Haloferax volcanii*.

1.2.1 S-layer glycoproteins in haloarchaea

S-layer proteins are among the most abundant proteins synthesised by the archaeal cell and have the unique feature of self-assembling into 2D lattices and can be isolated with relative ease in substantial amounts for biochemical and structural studies. Molecular masses of S-layer proteins typically range from 40 to 170kDa (Sleytr and Beveridge, 1999, Sleytr et al., 2007) and N-glycosylation is the most frequent posttranslational modification these proteins undergo although O-linked glycans are sometimes present (Jarrell et al., 2010). Other covalent modifications include lipid attachment, phosphorylation and methylation (Schaffer and Messner, 2004, Claus et al., 2002, Eichler and Adams, 2005). Although found in numerous archaeal species, the S-layers of the haloarchaea remain the best studied (Eichler, 2003, Schaffer and Messner, 2001, Schaffer and Messner, 2004). Investigations of S-layer protein glycosylation of haloarchaea were initiated by Strominger and colleagues (Mescher and Strominger, 1976), extended by Sumper and Wieland and colleagues (Sumper et al., 1990) and is currently extensively being performed by Eichler in collaboration with our laboratory (Abu-Qarn et al., 2008a).

The first description of an S-layer was reported in 1956 when electron microscopic examination of *Halobacterium halobium* (*Halobacterium salinarum*) cells revealed a surface presenting morphological units organised in a hexagonal pattern (Houwink, 1956). Later examination of thin-sectioned haloarchaeal cells

revealed the presence of a 17nm thick cell wall beyond the plasma membrane (Kirk and Ginzburg, 1972, Steensland and Larsen, 1969, Stoeckenius and Rowen, 1967). Blaurock and colleagues, later relied on X-ray diffraction to demonstrate a protein layer lying beyond the haloarchaeal plasma membrane at a distance of 8nm, with a periplasmic-like space being formed from morphological subunits assuming an “inverted-parabola shape” (Blaurock et al., 1976). Enzymatic iodination of the surface proteins, together with proteolytic treatment, revealed this surface (S)-layer to contain the S-layer glycoprotein. This was the first prokaryotic glycoprotein to be described in detail (Lechner and Sumper, 1987, Mescher and Strominger, 1976). Kessel and colleagues next proposed a three-dimensional reconstruction of the *Halobacterium* (*Haloferax*) *volcanii* S-layer glycoprotein and cell envelope after considering the primary sequence of the *H. salinarum* S-layer glycoprotein, as well as the earlier X-ray diffraction data and electron microscopic images of negatively stained cell envelopes (**Figure 1.6**). Kessel and colleagues revealed the structure of the surface glycoprotein to a resolution of 2nm. The glycoprotein was found to be arranged on a p6 lattice with a lattice constant of 16.8nm and formed 4.5nm high, dome-shaped, morphological complexes with a narrow pore at the apex opening into a “funnel” towards the cell membrane. Six radial protrusions were found to emanate from each morphological complex and to join around the 3-fold axis to provide lateral connectivity (Kessel et al., 1988).

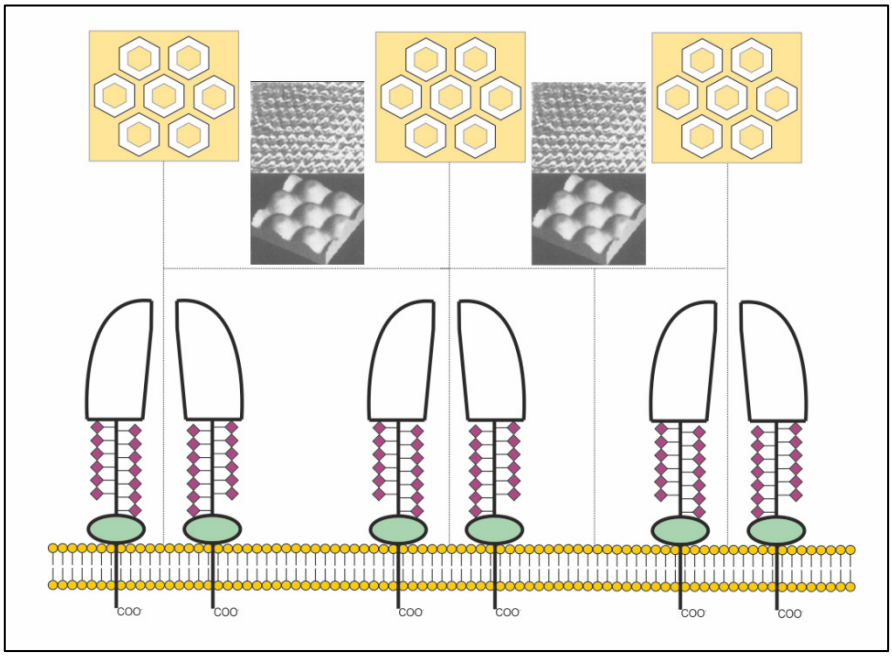


Figure 1.6 – Schematic of the proposed structural presentation of the S-layer glycoprtoein of *H. volcanii*.
The figure describes the *H. volcanii* S-layer glycorptoein organisation and view of the surface at different axes, as adapted from Kessel and colleagues (Kessel et al., 1988), which was proposed from all available structural information discovered at the time. The diagram combined information from X-ray studies of the envelope by Blaurock and colleagues (Blaurock et al., 1976), the primary structure published by Sumper and colleagues (Lechner and Sumper, 1987) as well as the three dimensional information from electron microscopy by Kessel and colleagues.

The gene coding for the S-layer protein of *H. volcanii* was first cloned and sequenced by Sumper and co-workers (Sumper et al., 1990). The mature polypeptide is composed of 794 amino acids and is preceded by a typical signal sequence of 34 amino acid residues. A highly hydrophobic stretch of 20 amino acids at the C-terminal end is proposed to be the transmembrane domain. Clusters of threonine residues are located adjacent to this membrane anchor. The S-layer protein is a glycoprotein containing both N- and O-glycosylation sites, as is shown in **Figure 1.7**. The primary structure of the S-layer glycoprotein from *H.*

volcanii and *H. halobium* were compared and found to exhibit distinct differences, despite the fact that three dimensional reconstruction from electron micrographs revealed no structural differences at least to the 2.5nm level attained (Sumper et al., 1990).

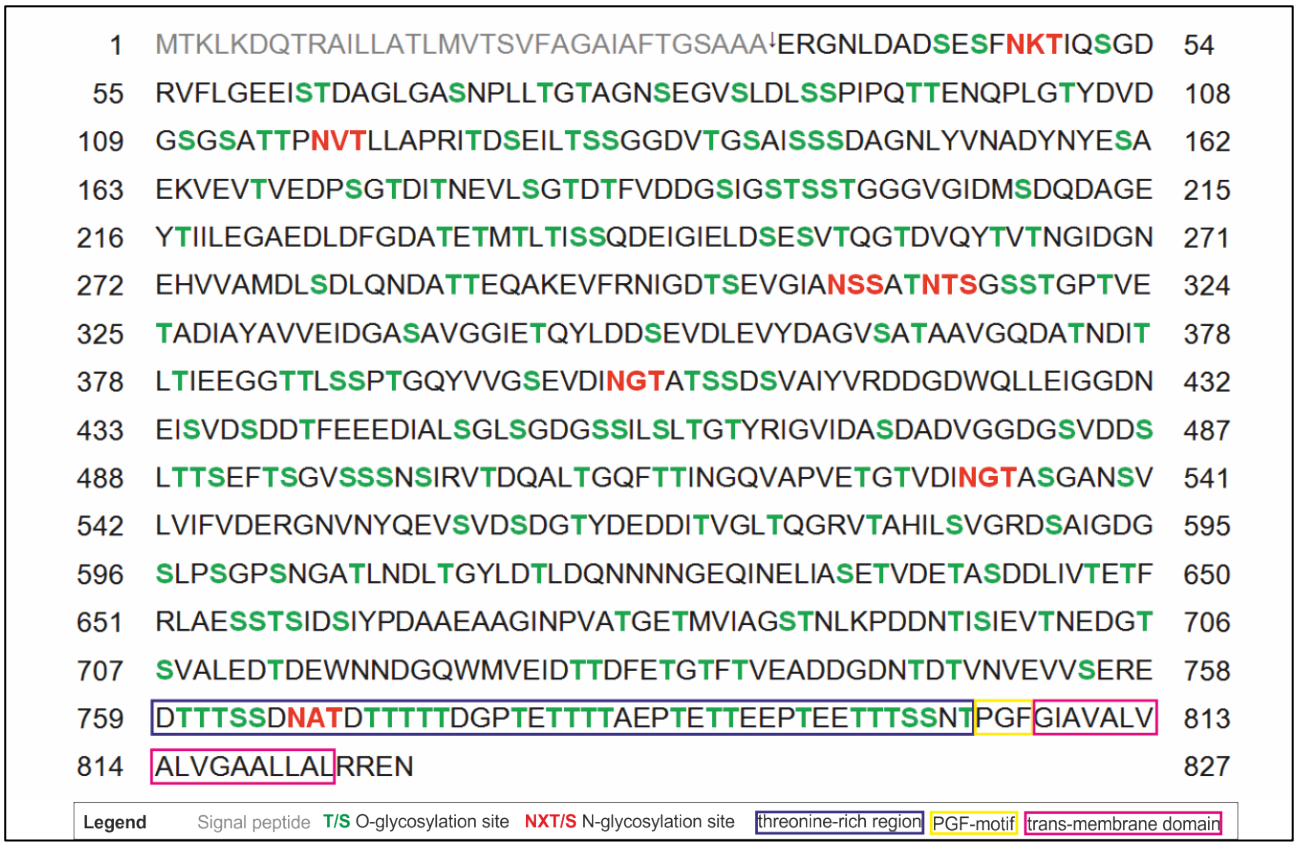


Figure 1.7 – Amino acid sequence of the *H. volcanii* S-layer glycoprotein.
The sequence is derived from UniProtKB/SwissProt (Entry: P25062), as published by Sumper and colleagues (Sumper et al., 1990). The signal peptide is coloured grey, whilst putative N- and O-glycosylation sites are shown in red and green respectively. The C-terminal region contains a threonine-rich region (purple), a PGF motif (yellow) as well as a suggested trans-membrane domain (pink).

1.2.2 Glycosylation of the S-layer glycoprotein of *H. volcanii*

Examination of *H. volcanii* proteins by SDS-PAGE and PAS or lectin staining revealed numerous putative glycosylated bands. A 190kDa protein was most intensely stained, although other protein bands were also clearly seen (Eichler, 2000). There are seven Asn-X-Ser/Thr sequons found in the *H. volcanii* S-layer glycoprotein that could potentially be glycosylated by N-glycans (**Figure 1.7**). Outputs of structural studies carried out more than twenty years ago by Mengele and colleagues suggested that at least four of these sites were occupied and that more than one type of N-glycan was present. They proposed that Asn-47 and Asn-532 were glycosylated with linear chains of β-1-4-linked glucose residues, that Asn-308 was modified by a glucose and idose-containing oligosaccharide, and that Asn-313 carried a polysaccharide composed of glucose, idose and galactose (Mengele and Sumper, 1992).

When *H. volcanii* S-layer N-glycosylation was re-visited over a decade later using more powerful structural methods it became clear that these early compositional assignments were incorrect. Thus, at the time the work described in this thesis (**Chapter** 3) was initiated, it had been shown by glycoproteomic analyses at Imperial College London that the glycan at Asn-47 was actually a pentasaccharide comprising two hexoses,

two hexuronic acids and a 190Da subunit, possibly corresponding to a methylated hexuronic acid or a di-methylated hexose (Abu-Qarn et al., 2007). The same glycan was found at the Asn-117 consensus site.

In parallel to the mass spectrometric work, our collaborators in Israel (Eichler and colleagues) were undertaking characterisation of the N-glycosylation process. Some of the *a*rchaeal *gl*ycosylation (*agl*) genes implicated in this post-translational modification had been identified on the basis of their homologies to known N-glycosylation components in eukaryotes or bacteria (Szymanski and Wren, 2005) and their roles in N-glycan biosynthesis were being explored by the Eichler group. Thus, putative *H. volcanii* homologs of genes involved in the eukaryal or bacterial N-glycosylation were identified by bioinformatics and reverse transcription polymerase chain reaction confirmed that the proposed N-glycosylation genes were being transcribed. Deletion of some genes showed slower growth and interfered with S-layer glycosylation, represented by modified migration on SDS-PAGE and glycostaining and the structural consequences of gene deletion were investigated by glycoproteomics (Abu-Qarn and Eichler, 2006). Subsequently additional *agl* genes were identified either based upon their proximity to previously identified *agl* sequences or upon re-annotation of the region of the genome where all but one of the previously identified *agl* sequences clustered (Abu-Qarn et al., 2008b, Yurist-Doutsch et al., 2008, Abu-Qarn et al., 2007, Yurist-Doutsch and Eichler, 2009). Further characterisation of the N-glycosylation pathway has been achieved as part of this thesis (**Section 3.4**).

In addition to N-glycosylation, the *H. salinarum* and *H. volcanii* S-layer glycoproteins also undergo O-glycosylation. In each case, a threonine-rich region upstream of the predicted membrane-spanning domain of the protein (**Figure 1.7**) has been proposed to be a mucin-like domain that is decorated at numerous positions by glucosyl(1-2)galactose disaccharides (Mescher and Strominger, 1976, Sumper et al., 1990). Essentially nothing further is presently known of the archaeal O-glycosylation process. The O-glycan structures were re-investigated as part of this thesis (**Section 3.3**) in order to rule out the possibility that the limitations of the analytical procedures employed by these early workers had resulted in errors in O-glycan assignments, as was the case with the N-glycans (see above).

1.2.3 Lipid as a potential posttranslational modification of the *H. volcanii* S-layer

The covalent attachment of lipids is a common modification experienced by both eukaryal and bacterial proteins and can involve myristoyl or palmitoyl acyl groups and isoprenyl polymers of various lengths, or aminoglycan-linked phospholipids. These can be added at the amino terminus, the carboxy terminus or at internal residues via ester, thioester, thioether or amide bonds, or through mediating functionalities such as phosphodiesters. Lipid modification of proteins is largely a posttranslational event and serves a variety of roles, such as enhancing membrane affinity of the modified protein, signal transduction, embryogenesis and pattern formation, protein trafficking through the secretory pathway and evasion of the immune response by infectious parasites (Eichler and Adams, 2005).

One of the defining traits of archaea distinguishing them from eukaryotes and bacteria is the chemical composition of their membrane phospholipids (Kates, 1993). Unlike eukaryal and bacterial phospholipids, which are built on a glycerol-3-phosphate backbone, archaeal phospholipids are based on the opposite steroisomer, glycerol-1-phosphate. The archaeal phospholipids contain polyisoprenyl side-chains rather than

the acyl groups employed by eukaryotes and bacteria. The archaeal phospholipids further rely on ether bonds to link the isoprenyl side chains to the glycerol-1-phosphate backbone (Eichler and Adams, 2005). Archaeal phospholipids are generally organised into the bilayer structure that is also present in eukaryal and bacterial cells, although tetra-ether lipid-based monolayers can also be found in thermophilic and hyperthermophilic archaea (Koga et al., 1993). Whereas phospholipids and other polar lipids (phosphoglycolipids, glycolipids and sulfolipids) account for the vast majority of archaeal membrane lipids, archaeal membranes also contain acetone-soluble nonpolar lipid species, primarily neutral squalenes and other isoprenoid-based polymers (Kates, 1978, Kates, 1992).

Lipid-modified proteins have been reported from a wide range of species of archaea. In the haloalkaliphile *Natronobacterium pharaonis*, a small blue copper protein was proposed to undergo amino-terminal lipid modification based on the presence of the so-called lipopbox sequence motif near the start of predicted amino acid sequence (Mattar et al., 1994). Although widespread in eukaryotes, GPI-anchored proteins have not been observed in bacteria, but have been seen in archaea, as *Sulfolobus acidocaldarius* was shown to possess three proteins that incorporate radiolabeled precursors of the GPI anchor moiety (Kobayashi et al., 1997).

Growth of *Halobacterium cutirubrum* and *Halobacterium salinarum* in the presence of radiolabeled [^{3}H] mevalonate, a precursor for the isoprene building block used to synthesise archaeal lipids (Boucher et al., 2004, Smit and Mushegian, 2000), led to the appearance of several proteins radiolabeled through the covalent attachment of a lipid entity. Sagami and colleagues showed that a derivative of [^{3}H] mevalonic acid is incorporated into a number of specific proteins in *H. halobium* and *H. cutirubrum* and that the major radioactive material released by treatment with methyl iodide was an unknown compound, which was analysed by reverse and normal phase high performance liquid chromatographies followed by mass spectrometry (Sagami et al., 1994). The identified parent ion showed the compound to be a diphytanylglyceryl methyl thioether, suggesting that *Halobacteria* contained specific proteins with a novel type of modification of a cysteine residue of proteins with a diphytanylglyceryl group in thioether linkage (Sagami et al., 1995).

During further analysis of the isoprenoid-modified proteins in *H. salinarum* using other radiolabeled isoprenyl derivatives to obtain insight into the nature of the isoprenoid binding manner, radioactive PAS-stained peptides recovered from the gel were treated with methyl iodide. No radioactive materials were released from the peptides in the sulfonium salt cleavage reaction or under the acidic conditions at 37°C, however under the acidic conditions at 95°C, radioactive materials were released from the peptides with 90% recovery and revealed that the S-layer glycoprotein is modified by a second novel group, diphytanylglyceryl phosphate, which is attached through an as yet uncharacterised linkage. Amino acid sequencing placed the modification near an O-glycosylated Thr-rich stretch found in the C-terminal region of the protein, upstream of the single transmembrane domain (Kikuchi et al., 1999). Kessel and colleagues had speculated that the unusual structural threonine-rich cluster element in the S-layer of haloarchaea serves as a spacer between the membrane-binding domain and the extracellular domain of the cell-surface glycoprotein, thus creating an interspace that may be regarded as analogous to the periplasmic space of Gram-negative bacteria. The

hydroxyl group of the side chain of serine or threonine next to each membrane-binding domain was therefore suggested as a possible modification site (Kessel et al., 1988).

In *H. volcanii*, the S-layer glycoprotein was found to be first synthesised as an immature precursor, possessing a lower apparent molecular mass than the final version of the protein. Pulse-chase radio-labelling and cell-fractionation studies were employed by Eichler and colleagues to reveal that newly synthesised S-layer glycoprotein seemed to undergo a maturation step following the translocation of the protein across the plasma membrane. Post-translational conversion into the mature form of the protein is largely completed within the first 10 min following the appearance of the full-length polypeptide. The S-layer glycoprotein modification was suggested to require Mg^{2+}-dependent membrane association of the protein, as no maturation was detected in the absence of Mg^{2+}, which is known to be involved in attachment of the S-layer glycoprotein to the membrane (Sumper et al., 1990). This processing step was detected as an increase in the apparent molecular mass of the S-layer glycoprotein and was suggested to be unaffected by inhibition of protein synthesis and unrelated to glycosylation of the protein. Maturation resulted in an observed increase in hydrophobicity of the protein as revealed by enhanced detergent binding and modification of the *H. volcanii* S-layer glycoprotein was suggested to offer an explanation for these observations (Eichler, 2001).

Eichler and Konrad suggest that the *H. volcanii* S-layer glycoprotein was also modified by a mevalonic acid-based lipid moiety and that such modification was responsible for the maturation of the *H. volcanii* S-layer glycoprotein. *H. volcanii* cells were therefore incubated with a [^{3}H]-radiolabelled version of the isoprene precursor, mevalonic acid. The [^{3}H] mevalonic acid label appeared to remain associated with the S-layer glycoprotein peptide fragment after de-lipidation, proteolytic digestion and SDS-PAGE. Konrad and Eichler proposed that neither phosphate ester nor pyrophosphate ester lipids were released under the mild acidic conditions, which led to the suggestion that the mevalonic acid-derived moiety could be covalently linked to to the protein, rather than simply being loosely associated with it (Konrad and Eichler, 2002).

It was hypothesised from the studies described above that the *H.volcanii* S-layer was likely to be substituted with a lipid moiety in the mucin-like domain. However, the S-layer has a C-terminal hydrophobic sequence which has all the properties required for it to be a robust membrane-spanning domain (Lechner and Sumper, 1987, Sumper et al., 1990, Wakai et al., 1997). It was therefore puzzling that an additional membrane anchor in the form of a covalent lipid would be required. To cast more light on this issue we felt that it was important to rigorously establish whether the S-layer is, indeed, covalently modified with lipid. The chemical nature of the lipid association with the S-layer glycoprotein is therefore also explored in this thesis (**Section 3.6**).

1.2.3.1 Characterisation of polar lipids of the halophilic archaeon *H. volcanii*

As part of a study to identify novel lipids with immune adjuvant activity, a structural comparison was made by Sprott and colleagues between the polar lipids from two halophiles, *Haloferax volcanii* and *Planococcus H8* (Sprott et al., 2003). The total polar lipids of *H. volcanii* accounted for 83.5% by weight and fast atom bombardment mass spectrometry (FAB-MS) and thin layer chromatography were used to identify and quantify the major signals corresponding to the phospholipids. *H. volcanii* polar lipid extracts consisted of

44% archaetidylglycerol methylphosphate, 35% archaetidylglycerol, 4.7% of archaeal cardiolipin, 2.5% archaetidic acid, and 14% sulphated glycolipids 1 and 2 as are summarised in **Table 1.1**.

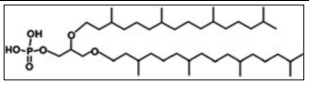
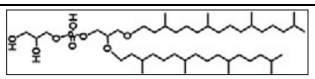
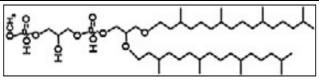
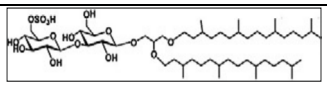
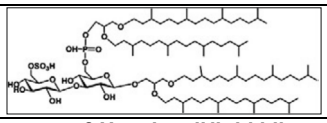
Lipid moiety	Name	Structure	% Abundance	m/z	+Na
PA	archaetidic acid		2.5	731.5	
PG	archaetidylglycerol		35.2	805.6	
PGP-CH_3	archaetidylglycerol methylphosphate		43.8	899.5	921.5
S-GL-1	sulphate-diglycosyl archaeol-1		5.2	1055.6	
S-GL-2	sulphate-diglycosyl archaeol-2		8.7	1770.2	1792.2

Table 1.1 – Summary of *H. volcanii* lipid bilayer population.
The table describes the total lipid population of *H. volcanii*, as characterised described by Sprott and colleagues (Sprott et al., 2003).

1.3 *BURKHOLDERIA PSEUDOMALLEI*

1.3.1 *Burkholderia pseudomallei* the causative agent of melioidosis

The genus *Burkholderia* comprises more than 40 different species and these occupy a wide array of ecological niches. Traditionally, the *Burkholderia* species are known as plant pathogens and soil bacteria with a couple of important exceptions, including *Burkholderia mallei* and *Burkholderia pseudomallei*, which are primary pathogens for humans and animals (Coenye and Vandamme, 2007).

The *Burkholderia* genus, was formerly part of the genus *Pseudomonas*, which was first described more than 100 years ago and its classification and description have been revised on several occasions. In 1973 its taxonomic heterogeneity was revealed by the work of Palleroni and co-workers who delineated five major species clusters, also referred to as rRNA homology groups (Palleroni et al., 1973). The name *Pseuodomonas* was confined to rRNA homology group I organisms because they comprised the type species, *Pseudomonas aeruginosa*, whilst *Pseudomonas cepacia* and six other species belonging to *Pseudomonas* rRNA group II (*P. solanacaerum*, *P. pickettii*, *P. gladioli*, *P. mallei*, *P. pseudomallei* and *P. caryophylli*) were transferred to the new genus *Burkholderia* (Yabuuchi et al., 1992). The name *Burkholderia* was chosen to honour Walter Burkholder's seminal work in the 1940s and 1950s on a disease of onion bulbs that he termed "sour skin" or "slippery skin", basing his description of the causative agent on this bacterial rot disease on seven isolates from decaying onion bulbs which he named *Pseudomonas cepacia* (Coenye and Vandamme, 2007).

The saprophytic bacterium *B. pseudomallei* is the causative agent of melioidosis, a serious and often fatal infectious disease endemic in Southeast Asian countries and Northern Australia (Dance, 1991, Dance, 2000). Melioidosis is also frequently reported in areas of China and the Far East and there have been

sporadic cases in areas of Asia and Central America (**Section 1.3.2**) (Cheng and Currie, 2005). The pathologist Alfred Whitmore and his assistant C. S. Krishnaswami first described melioidosis as a "glanders-like" disease among morphia addicts in Rangoon, Burma in 1911 (Whitmore, 1913). They distinguished this disease causing agent from the organism causing glanders by its relatively rapid growth, motility and the lack of the Strauss reaction when it was injected to guinea pigs, used at the time to diagnose glanders (Whitmore, 1913). The disease was later termed melioidosis by Stanton and Fletcher in 1932 (Cheng and Currie, 2005). This Gram-negative environmental bacterium has been variously known as *Bacillus pseudomallei*, *Bacillus whitmorii* (Bacille de Whitmore), *Malleomyces pseudomallei*, *Pseudomonas pseudomallei* and since 1992, *Burkholderia pseudomallei* (Yabuuchi et al., 1992). In the latter half of the 20th century, melioidosis emerged as an infectious disease of major public health importance in South East Asia and Northern Australia (Currie et al., 2000c, Chaowagul et al., 1989).

B. pseudomallei induced melioidosis can present itself in various forms, including fever, septicaemia, pulmonary and other chronic diseases and melioidosis has a broad clinical spectrum, ranging from asymptomatic condition to acute fatal septicaemia (**Section 1.3.3.1**) (Sarkar-Tyson et al., 2007, Cheng and Currie, 2005). The organism is innately resistant to a large number of antibiotics and inadequate antibiotic treatment is associated with a high rate of relapse (**Section 1.3.3.4**) (Anuntagool et al., 2006). *B. pseudomallei* is also well adapted to its many hosts, with virulence factors such as the production of proteases, lipases and catalase (**Section 1.3.7**). Furthermore it produces a glycocalyx polysaccharide capsule suggested to be an important virulence determinant, alongside cell-associated antigens, such as pili and capsular polysaccharides (CPS), as well as lipopolysaccharides (LPS) (**Section 1.3.5**).

1.3.2 Epidemiology

B. pseudomallei is visualised as a Gram-negative bacillus with bipolar staining. It is vacuolated, slender, has rounded ends and has previously been described as having a "safety-pin" appearance. The bacterium is 2-5μm length and 0.4-0.8μm in diameter and motile due to its capability of self-propulsion using flagella. In culture, the organism demonstrates differing colonial morphology with mostly smooth colonies initially and dry or wrinkled colonies on further incubation.

B. pseudomallei is oxidase and gelatinase positive and can be distinguished from the closely related but less pathogenic *Burkholderia thailandensis* by its ability to assimilate arabinose (Smith et al., 1997). *B. pseudomallei* is a very resilient organism that is capable of surviving in hostile environmental conditions, including prolonged nutrient deficiency of up to 10 years (Wuthiekanun et al., 1995b) and antiseptic and detergent solutions (Gal et al., 2004). Furthermore it is able to survive in acidic environments, such as pH 4.5 for up to 70 days (Dejsirilert et al., 1991) and in a wide temperature range from 24ºC to 32ºC. Cultures typically become positive in 24 to 48 hrs, a rapid growth rate, which differentiates the organism from *B. mallei*, which typically takes a minimum of 72 hours to grow. It is likely that harsh environmental conditions may confer a selective advantage for the growth of *B. pseudomallei*.

1.3.2.1 Geographical epidemiology

Although the area where melioidosis is endemic has generally been regarded as restricted to the latitudes of 20ºN and 20ºS in Southeast Asia and Northern Australia, large outbreaks have occurred outside this area (Dance, 1991, Dance, 2000). In Thailand, *B. pseudomallei* is widely distributed in soil and more particularly in pooled surface water such as in rice paddies as well as the roots of various plants (Wuthiekanun et al., 1995a, Holden et al., 2004). Humans in prolonged contact with contaminated water and soil, such as rice farmers, are therefore more likely to contract the disease and most infections occur during the rainy season when bacteria are leached from the soil (Lazar Adler et al., 2009). *B. pseudomallei* is especially problematic in North-Eastern Thailand where approximately 20% of community-acquired septicaemias can be attributed to this bacterial pathogen (Chaowagul et al., 1989). In Ubon Ratchathani province in north-eastern Thailand, 4.4 cases per 100'000 have been noted (Suputtamongkol et al., 1994). In this region up to 80% of children are positive for antibodies against *B. pseudomallei* by the age of four years.

The global map of melioidosis continues to expand with examples of case clusters in Brazil, recognition that melioidosis is endemic in the pacific nation of New Caledonia and cases following the Asian tsunami. It was found that outbreaks of melioidosis occurred following typhoon season and flooding in several endemic areas (Chaowagul et al., 1989, Currie et al., 2000c). In endemic regions up to 85% of human infection occurs during the monsoon wet season (Currie and Jacups, 2003).

Molecular tools used to infer genetic relatedness between isolates of *B. pseudomallei* have demonstrated that environmental isolates can be identical to epidemiologically related human or animal strains (Currie et al., 1994, Currie et al., 2001). Furthermore, it was found that recurrent infection is usually due to relapse with the same strain rather than re-infection with a different strain (Currie et al., 2000a, Desmarchelier et al., 1993). These studies have demonstrated considerable diversity in isolates, suggesting that the introduction into the regions where *B. pseudomallei* is found is not a recent event (Pitt et al., 2000). This contrasts with clonal outbreaks in regions where the organism is not endemic (Currie et al., 1994).

The finding that the antigenically similar *B. thailandensis* was relatively avirulent led to a suggestion that different strains of *B. pseudomallei* may have differing virulence. Two other studies have suggested that clinical presentation or outcome may depend on the strain type (Pitt et al., 2000). A sequence-based successor to multilocus enzyme electrophoresis, multilocus sequence typing has been developed for *B. pseudomallei*. This technique is based on allelic genetic analysis beyond the outbreak situation due to its reproducibility and the low rate of genetic change in the allelic sites, which allows for comparisons between strains typed at different laboratories through an internet-based database. This study confirmed the diversity seen in isolates worldwide and placed *B. mallei* within its wider *B. pseudomallei* group, both being distinct form *B. thailandensis* (Godoy et al., 2003).

1.3.3 Clinical features of melioidosis

B. pseudomallei is capable of causing a broad spectrum of diseases with protean clinical manifestations, but acute pulmonary and septicaemic forms are of particular concern due to the high incidence of mortality

associated with these cases even when vigorous chemotherapeutic intervention is implemented (Leelarasamee and Bovornkitti, 1989).

1.3.3.1 Disease presentation and diagnosis

Melioidosis may present as either a chronic or acute disease and there are no pathonomonic signs of infection, hindering prompt diagnosis. In an acute infection the symptoms include fever, malaise, abscess formation, pneumonia and sepsis. Even with aggressive antibiotic therapy, septicaemia caused by *B. pseudomallei* has a mortality rate of about 40%. The chronic version of melioidosis is often described as a milder version of the symptoms, but may last for months or years. It has been shown that reactivation of chronic *B. pseudomallei* infections have occurred in Vietnam veterans up to 18 years after their last exposure to the bacteria, which nicknamed the condition “the Vietnamese time bomb” (Koponen et al., 1991, Howe et al., 1971). Reactivation is correlated with the onset of other illnesses including influenza, Type 2 diabetes or cancer (Mays and Ricketts, 1975). The majority of infected persons remain asymptomatic, however latent infection with subsequent reactivation is well recognised (Ngauy et al., 2005).

Variation of the clinical presentation and severity of melioidosis was noted to be due to one or more of three factors. These include a variation in bacterial strains, including the presence or absence of virulence factors, a variation in the host immune response and a variation in acquisition. There is a large body of evidence that suggests that host factors, particularly age and comorbidites are of prime importance in determining the pattern of disease (Cheng and Currie, 2005).

The mean incubation period of *B. pseudomallei* is about 9 days, but ranges from 1 - 21 days (Currie et al., 2004). Patients with latent melioidosis may be symptom free for decades. Diagnosis is made by culturing the organism from a clinical sample and a complete screen inclusive of blood culture, sputum culture, urine culture, throat swab and culture of any aspirated pus is be performed on patients with suspected melioidosis. Mortality rates from melioidosis vary from 15% in centres where intensive care is available to greater than 50% in regions with poor health care (Cheng and Currie, 2005, White, 2003).

1.3.3.2 Infection by *B. pseudomallei*

Infection by *B. pseudomallei* in humans usually occurs through percutaneous inoculations, abrasions of the skin or inhalation, while ingestion is more common in animals (Currie, 2008). Due to the fact that farmers in endemic regions rarely wear protective footwear when harvesting crops, the feet of the individuals often show signs of repeated trauma and injury, presenting the route of infection (Chaowagul et al., 1989), whilst inhalation of aerosols created by heavy rains presents a further route of infection.

Inhalation was initially thought to be the primary mode of acquisition based on studies of US soldiers in Vietnam where it was noted that helicopter crews seemed to have high incidence of the disease and although sporadic cases have continued to surface in the US from remote exposure, the feared epidemic failed to materialise. The finding that periods of heavy rainfall are associated not only with higher numbers of cases but also pneumonic presentations and cases of increased severity may suggest a shift to inhalation

during extreme weather events (Currie and Jacups, 2003). It is now believed that inoculation is the major mode of acquisition. Ingestion has been suggested as a mode of infection and this route has been implicated in animals by findings of infected gastroheptaic nodes in pigs, which have also been noted in humans. The contribution of this route of infection, however, is undefined, although contamination of potable water has been implicated as the point source in two outbreaks (Currie et al., 2001). Acquisition of infection from animals has been described but is believed to be very uncommon as is person to person transmission and laboratory-acquired infection (Cheng and Currie, 2005, Dance, 2000).

1.3.3.3 Melioidosis risk factors

Approximately 20% of persons with disease have no apparent risk factors. However, epidemiological studies suggested patients with worsened physical wellbeing were more susceptible to infection, where individuals with diabetes, compromised liver or renal function showed to have increased risk of infection (Currie et al., 2004). The use of steroids is associated with an increased risk of melioidosis and chronic obstructive pulmonary disease and the consumption of kava and alcohol have also been implicated (Currie et al., 2000b, Suputtamongkol et al., 1994, Suputtamongkol et al., 1999, White, 2003). A number of case studies have noted an intriguing association with chronic granulomatous disease that might highlight the under-recognised role of neutrophil defects in pathogenesis. The notable lack of an association between HIV infection and melioidosis severity and the very strong link to diabetes suggests that after initial infection with *B. pseudomallei,* phagocytic function and not cell-mediated immunity may be critical to prevent the dramatic progressive dissemination of the bacteria seen in acute septicaemic melioidosis (Currie, 2008).

1.3.3.4 Antibiotic resistance and treatment

B. pseudomallei exhibits resistance to diverse groups of antibiotics including third-generation cephalosporins, penicillins, rifamycins and aminoglycosides. In addition its relative resistance to quinolones and macrolides limits therapeutic options for the treatment of melioidosis (Cheng and Currie, 2005). *B. pseudomallei* is, however, susceptible to kanamycin although this antibiotic is no longer used in the treatment of clinical melioidosis as transposon mutation analysis has revealed that the efflux system AmrAB-OprA confers resistance (Cheng and Currie, 2005). The choice of antibiotic regimen has not been shown to have an impact on mortality within the first 48 hours of admission and severe melioidosis in Thailand is still associated with a case fatality rate of approximately 50%, whilst in Australia mortality rate is still significant and approaches 20% of all patients with melioidosis (White, 2003, Currie et al., 2000b). The most common antibiotic resistance mechanisms that have been described in *Burkholderia* are those associated with limited permeability, thus inactivating enzymes and altered drug targets are rarely described. The role of limited permeability as an antibiotic resistance mechanism in *Burkholderia* may be related to its polysaccharide structures, limited porin size or multidrug efflux pumps, whereas biofilm mediated antibiotic resistance has been reported rarely (Coenye and Vandamme, 2007).

1.3.3.5 Relapse associated melioidosis

Relapse after apparently successful treatment is well described and is associated with mortality similar to that for the initial episodes (Currie et al., 2000a). Factors associated with a higher risk of relapse included poor adherence to therapy, the use of doxycycline monotherapy or amoxicillin-clavuanate in the eradication phase and severe disease. In the majority of cases, relapse was due to reactivation of the original infecting strain. Relapses be due to re-infection with a different strain was demonstrated in 4-7% of cases in Thailand and Australia, or the initial infection being caused by multiple strains (Currie et al., 2000a, Maharjan et al., 2005). This means that an effective vaccine against the pathogen would have to encompass all important strains.

1.3.4 *Burkholderia* genomes and species variation

Examination of the completed *Burkholderia* genome sequences has shown that for most species capture and retention of foreign DNA is essential to their genome evolution. The conservation of the guanine-cytosine (GC) content is a defining feature of bacterial species, where essential genes that are core to the genome have a consistent GC composition that is reflective of the average of the species as a whole. Foreign regions of DNA can be identified by several signatures within the DNA sequence, where looking for regions of atypical GC content is frequently predictive of DNA that has been recently acquired in evolutionary terms (Coenye and Vandamme, 2007). The genomes of both *B. mallei* ATCC 23344 and the clinical strain K96243 of *B. pseudomallei* have been sequenced and annotated (Holden et al., 2004, Nierman et al., 2004). The latter is composed of two circular chromosomes of 4.07Mb and 3.17Mb, showing significant functional partitioning of the genes between them (**Figure 1.8**).

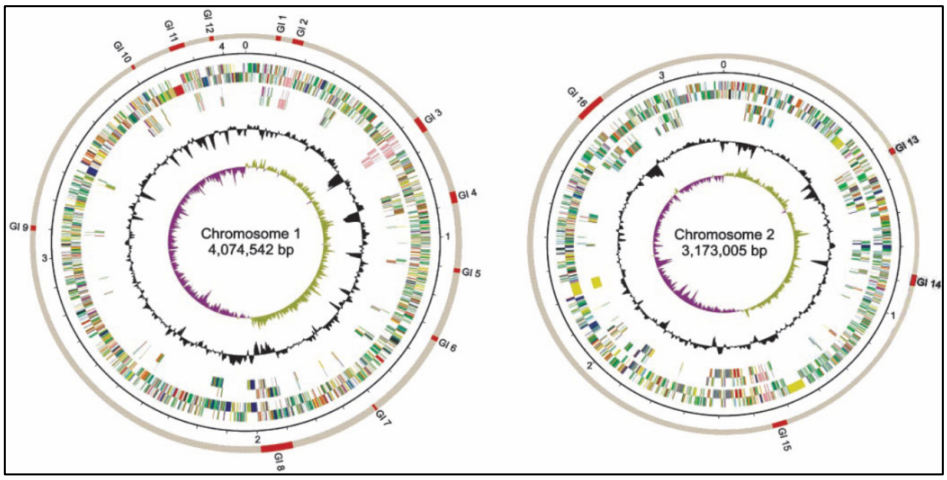


Figure 1.8 – Circular diagrams of the *B. pseudomallei* chromosomes.
The figure shows chromosome 1(4.07mbp) and chromosome 2 (3.17mbp) of *B. pseudomallei* K96243 as described by Holden and colleagues (Holden et al., 2004)

Chromosome 1 contains a higher proportion of coding sequences involved in core functions, such as macromolecule biosynthesis, amino acid metabolism, cofactor and carrier synthesis, nucleotide and protein biosynthesis, chemotaxis and mobility. Chromosome 2, by contrast, contains a greater proportion of genes encoding accessory functions associated with adaption to atypical conditions, osmotic protection and iron acquisition, secondary metabolism, regulation and laterally acquired DNA. In addition chromosome 2

contains a greater proportion of coding sequences with matches to hypothetical proteins or proteins that have no database matches at all (Holden et al., 2004). At 7.2 Mb, the *B. pseudomallei* genome is large in comparison with the typical prokaryotic genome and horizontal acquisition of DNA appears to have been intrinsic to the evolution of this organism. Many regions within both chromosomes showed some of the characteristics of genomic islands acquired through very recent lateral transfer, such as anomalies in the percentage of GC content or dinucleotide frequency signature of the DNA in these regions.

1.3.4.1 *B. pseudomallei* and *B. mallei* genomes in comparison

B. mallei is the causative agent of the highly infectious equine disease known as glanders. It has no known reservoir other than horses, mules and donkeys and is a host-adapted pathogen that does not persist in the environment outside its equine host. The contagious and potentially lethal nature of human infection with *B. mallei* also makes this species well suited for use as an agent of bioterrorism and *B. mallei* was one of the first biological weapons of the 20th century, being used by Germany during World War I. Glanders has since been eradicated from most countries due to stringent infection control measures, including the immediate slaughter of affected animals.

Genomic comparisons of *B. pseudomallei* with closely and more distantly related bacteria revealed a greater level of gene order conservation and a greater number of orthologous genes on the large chromosome, suggesting that the two replicons have distinct evolutionary origins. The investigation revealed 16 genomic islands that together make up 6.1% of the genome, which were shown to be variably present in a collection of invasive and soil isolates but entirely absent from the clonally related organism *B. mallei*. These comparative analyses of *B. pseudomallei* and *B. mallei* have identified many coding regions that may contribute to the phenotypic differences between the two species. These phenotypes include known virulence determinants, such as flagella and type III secretion systems (T3SS) as well as potential virulence determinants such as surface polysaccharides, exoproteins, fimbriae, pili and putative adhesisn. Further determinants include drug resistance and potential environmental survival functions including various secondary metabolite pathways, numerous catabolic pathways, transport systems and stress response proteins (Holden et al., 2004).

Although *B. mallei* strains appear to have recently evolved from a *B. pseudomallei* ancestor, there are large differences in the sizes of the genomes, the *B. pseudomallei* genome is 1.31 Mb larger than that of *B. mallei*, reflecting the adaption of the pathogen to grow in a restricted infectious environment (Coenye and Vandamme, 2007). The absence of genomic islands in the *B. mallei* genome suggests that gene acquisition does not play as significant a role in the genetic variation of this species as it does in *B. pseudomallei*. Holden and colleagues propose that gene loss appears to have been an important source of genetic variation in the recent evolution of *B. mallei* (Holden et al., 2004), whilst Nierman and colleagues, suggest that numerous insertion sequences and simple sequence repeats point to an alternate mechanism for genetic variability in *B. mallei* (Nierman et al., 2004).

1.3.4.2 *B. pseudomallei* and *B. thailandensis* genomes in comparison

Environmental isolates from central and north-eastern Thailand were collected and tentatively identified as *B. pseudomallei* strains, based on their growth on selective media, biochemical profiling and latex agglutination assays (Smith et al., 1995, Wuthiekanun et al., 1995a). However, Brett and colleagues, whilst attempting to develop a transposon based mutagenesis system for *B. pseudomallei* observed a number of phenotypic and genotypic dissimilarities between clinical isolates and some of the environmental isolates, specifically the differences associated with the decreased virulence for Syrian golden hamsters in comparison to the true *B. pseudomallei* strains (Brett et al., 1997, DeShazer et al., 1997). These environmental isolates had a more than 10^5 fold decrease in virulence relative to *B. pseudomallei* in this animal model of acute melioidosis (Brett et al., 1997). Wuthiekanun and co-workers demonstrated differences in the biochemical characteristics of *B. pseudomallei* isolates and true isolates, suggesting that these environmental isolates may represent a new species of *B. pseudomallei* (Brett et al., 1998). Brett and colleagues cloned and sequenced more than 95% of the 16S rDNA from *B. pseudomallei* 1026b, which is a virulent clinical isolate similar to K96243 and the newly isolated *B. pseudomallei*-like E264, to assess the relatedness of these organisms to each other, finding 15 nucleotide dissimilarities in all 1488 base pairs of the 16S rRNA upon alignment of the two 16S rRNA alleles, thus confirming the presence of a new *Burkholderia* species to which they proposed the name *B. thailandensis*.

B. thailandensis has colonies which are smooth and glossy with pink pigmentation, in comparison to the rough, wrinkled and dark purple pigmentation found in *B. pseudomallei* colonies and is capable of assimilating L-arabinose (Brett et al., 1997). It is sensitive to tetracycline, ceftayidime and trimethoprim, but resistant to aminoglycosidases. *B. thailandensis* and *B. pseudomallei* strains are structurally and immunologically similar to one another and the differences in virulence were further investigated to identify the genetic factors responsible for the enhanced virulence of *B. pseudomallei* (Brett et al., 1998).

The genomes of *B. pseudomallei* and *B. thailandensis* both comprise two highly syntenic chromosomes with comparable numbers of coding regions, protein family distributions and horizontally acquired genomic islands. They share an extensive repertoire of genes involved in core metabolism, accessory pathways, structure-based superfamilies and bacterial virulence factors, however virulence related genes appear to have undergone accelerated change, to adapt to the challenge of infecting and surviving in human or animal hosts in *B. pseudomallei*. The *B. thailandensis* chromosomes are 3.8Mb and 2.9Mb (Yu et al., 2006).

The *B. thailandensis* genome contains at least 15 regions exhibiting either atypical GC content or stretches of bacteriophage related genes and phage-like integrases, which collectively encompass 4-5% of the entire genome. These genomic islands are not found in *B. pseudomallei* and *B. mallei*, consistent with their acquisition by *B. thailandensis* subsequent to the *B. pseudomallei*-*B. thailandensis* divergence. The genomic islands of *B. thailandensis* seem to occur in the same relative genomic location as those in *B. pseudomallei*, suggesting that these locations may represent genomic hot spots or landmarks for the acquisition of horizontally acquired sequences. Thus the acquisition and loss of large-scale genomic material seems to represent a major driving force in bacterial evolution and often plays a critical role in the development of novel microbial phenotypes.

Although *B. thailandensis* and *B. pseudomallei* occupy similar ecological niches, biochemical analysis has identified several phenotypic differences between the species including the ability of *B. thailandensis* but not *B. pseudomallei* to assimilate the carbon sources arabinose and xylose (Moore et al., 2004, Smith et al., 1997). *B. thailandensis* was shown to contain an eight-gene arabinose assimilation operon on chromosome 2 that is absent in *B. pseudomallei*, where this region has been replaced by protein clusters. Introduction of this operon into *B. pseudomallei* resulted in the downregulation of a number of type III secretion genes and the strain displayed reduced virulence in Syrian hamsters (Brett et al., 1998, Reckseidler et al., 2001, Wuthiekanun et al., 1996, Wiersinga and van der Poll, 2009). Similarly *B. thailandensis* contains a 64kb region on chromosome 1 encoding several genes involved in xylose metabolism, which is absent and replaced in *B. pseudomallei*. This suggests an evolutionary model where the horizontal acquisition by *B. pseudomallei* may have also resulted in the deletion of the xylose gene cluster, demonstrating how horizontal transfer events can often result in simultaneous gene acquisition and loss (Whitfield and Roberts, 1999).

Variations in surface component proteins have been shown to contribute to virulence in several pathogenic species, such as is seen with fimbriae, short pilus structures allowing bacteria to adhere to environmental surfaces and host cells (Kespichayawattana et al., 2004). *B. pseudomallei* contains twice as many fimbrial gene clusters as *B. thailandensis* and has been shown to be more efficient than *B. thailandensis* in adhering to and invading host cells. Furthermore it has been shown that *B. pseudomallei* contains a large gene cluster involved in the synthesis and export of capsular polysaccharides, a major determinant of virulence and this cluster is absent in *B. thailandensis*. The specific location of the *B. pseudomallei* capsular gene cluster within its genome is likely to be non-random, as it replaces a pre-existing 10-gene cluster in *B. thailandensis* already dedicated toward the metabolism and processing of polysaccharide structures (**Figure 1.9**). The original cluster in *B. thailandensis* contains several genes involved in polysaccharide assembly, raising the possibility that a key event in the pathogenic evolution of *B. pseudomallei* was to replace a pre-existing or ancestral polysaccharide coat with an alternative pathogenic variant capable of resisting challenges by the immune system of infected hosts (Reckseidler et al., 2001).

Strong conservation between the *B. pseudomallei* and *B. thailandensis* proteomes were found in the core metabolic pathways such as amino acid metabolism, cofactor and carrier synthesis, nucleotide and protein biosynthesis, consistent with the ability of *B. pseudomallei* and *B. thailandensis* to occupy similar environmental niches (Smith et al., 1997). Unexpectedly, the *B. pseudomallei* and *B. thailandensis* proteomes also appear to share significant similarities in their virulence components and 71% of potential virulence genes found in *B. pseudomallei* are also present in *B. thailandensis* at an average similarity of greater than 80%, including two type III secretion systems (T3SS2 and T3SS3), antibiotic resistance genes, type IV pili-generating proteins, hemolysin-related genes ad several adhesion factors and proteases (Holden et al., 2004). However, even though both *B. pseudomallei* and *B. thailandensis* share two T3SS, required for the full virulence of *B. pseudomallei* in a hamster model of infection (Warawa and Woods, 2002), it has recently been shown that arabinose exposure may downregulate T3SS expression and activity (Moore et al., 2004). The absence of an arabinose assimilation operon in *B. pseudomallei* might thus have contributed to the increased virulence of this species.

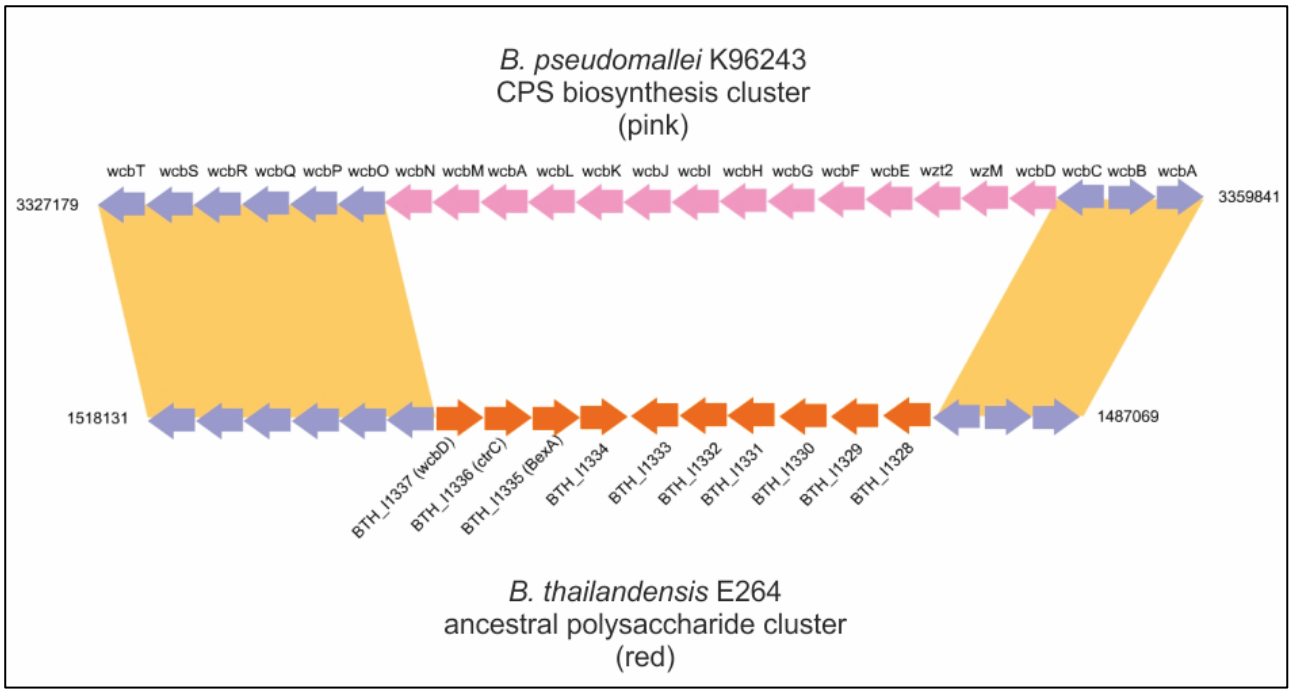


Figure 1.9 – *B. pseudomallei* and *B. thailandensis* polysaccharide biosynthesis clusters.
The schematic was adapted from the genomic comparison study by Yu assessing the genome of *B. pseudomallei* K96243 and *B. thailandensis* E264 (Yu et al., 2006). The diagram shows the capsular polysaccharide biosynthesis cluster of *B. pseudomallei* K96243 (top – pink) with an ancestral polysaccharide cluster of *B. thailandensis* E264 (bottom – red).

Genomic comparisons between pathogens and non-pathogenic relatives have played an important role in identifying the mechanisms responsible for acquisition of virulence in the natural environment. Yu and colleagues discuss that gene mutation, gene deletion and gene acquisition on the part of *B. pseudomallei* are likely to represent the major evolutionary drivers of virulence and that other proposed mechanisms of pathogen evolution including chromosomal rearrangement and bacteriophage-mediated recombination may thus play a less relevant role in the pathogenic evolution (Yu et al., 2006).

B. pseudomallei can only be experimentally manipulated under biosafety level 3 (BSL 3) conditions, but *B. thailandensis* is non-pathogenic for humans and animals although it displays several phenotypic characteristics similar to *B. pseudomallei*. As suggested by Nierman and colleagues research on the 16S rRNA phylogeny studies, *B. thailandensis* is closely related to *B. pseudomallei*, whilst the phylogenetic similarity between *B. pseudomallei* and *B. mallei* suggested it to be a derivative or clone of *B. pseudomallei* (Nierman et al., 2004). *B. thailandensis* on the other hand, was thought to have diverged from *B. pseudomallei* and *B. mallei* approximately 47 Million years ago, suggesting that although *B. thailandensis* is avirulent, it is likely to be highly evolutionary related to virulent *B. pseudomallei* and thus a good candidate for comparative genomic analysis (Yu et al., 2006).

1.3.4.3 *B. pseudomallei* as an accidental pathogen

The *B. pseudomallei* genome shares a core set of 2590 genes with other members of the *Burkholderia* family and was recently shown to be highly dynamic (Lin et al., 2008, Sim et al., 2008, Tuanyok et al., 2007, Tumapa et al., 2008). It was found that 86% of the *B. pseudomallei* K96243 genome is common to all strains across the 94 Southeast Asian strains isolated from a variety of clinical, environmental and animal sources. These represent the core genome including genes common to all strains in a population and those involved

in essential functions. The remaining 14% of the genome were found to be variably present across the strain panel investigated by Sim and co-workers. Core genes were significantly over-represented in several functions necessary for basic bacterial growth and survival, including amino acid metabolism, inorganic ion transport, nucleotide metabolism and protein translation. Besides these basic housekeeping functions the *B. pseudomallei* core genes were also significantly enriched in commonly encountered virulence related genes such as those involved in the production of secretion proteins, capsular polysaccharides, exoproteins, adhesions, fimbriae and pili. This would suggest that most if not all *B. pseudomallei* isolates are likely to possess a common "virulence" machinery and notably many of these conventional virulence genes are also found in other related species such as *B. thailandensis* that although non-infectious in mammals, are able to infect other species such as nematodes. This is consistent with the possibility that *B. pseudomallei* might have descended from a pathogenic ancestor with a non-mammalian host (Sim et al., 2008).

It is very likely that the "virulent" combination of genes found in *B. pseudomallei* has emerged for reasons other than to cause human disease, particularly since cases of human or animal infection are relatively rare compared to the density of *B. pseudomallei* in the soil. In contrast to bacteria which are obligately associated with eukaryotic hosts, soil bacteria such as *B. pseudomallei* commonly face extreme and unpredictable biotic and abiotic challenges including extreme temperature shifts, solar radiation, variable humidity, competition for nutrients and the requirement to survive ingestion by predatory protozoa, nematodes, the production of bacteriocides from other bacteria and phage infection. It thus seems entirely plausible that genes facilitating survival against these environmental challenges might have also indirectly enhanced the microbes' ability to colonize and "accidently" infect human host, particularly when the host is immunocompromised (Sim et al., 2008).

At 7.2Mb, the genome of *B. pseudomallei* represents one of the most complex bacterial genomes sequenced to date. Certain environmental microorganisms can cause severe human infections, even in the absence of an obvious requirement for transition through an animal host for replication, which is termed accidental virulence. Nandi and co-workers compared eleven isolate genomes of *B. pseudomallei* and the existence of several new genes in the *B. pseudomallei* reference genome were found by mRNA transcripts, database homologs and presence of ribosomal binding sites. These exhibited significant levels of positive selection distributed across many cellular pathways including carbohydrate and secondary metabolism (Nandi et al., 2010). Functional experiments revealed that certain positively selected genes might enhance mammalian virulence by interacting with host cellular pathways or utilising host nutrients. Evolutionary modifications improving *B. pseudomallei* environmental fitness may thus have indirectly facilitated the ability to colonise and survive in mammalian hosts.

1.3.4.4 *B. pseudomallei* typical and atypical serotypes

Diagnosis of melioidosis remains a problem and definitive diagnosis still relies on the isolation and identification of bacteria from clinical specimens. Attempts to develop reliable immunologic and molecular assays to replace the more time-consuming bacteriologic diagnosis have been made, but these new diagnostic methods remain non-validated (Anuntagool et al., 2000b, Wuthiekanun et al., 2002). *B. pseudomallei* is generally considered a rather homogenous species, but colony variation occurs, and

variation in antibiotic susceptibility profiles and biochemical profiles using a number of commercial bacteriological identification systems have been reported (White, 2003, Cheng and Currie, 2005). In addition to these phenotypic variations, physicochemical and antigenic heterogeneity of the LPS prepared from *B. pseudomallei* was reported, which would be responsible for false seronegativity in the patients with *B. pseudomallei* infection (Anuntagool et al., 2000a, Anuntagool et al., 1998).

In a highly pathogenic bacterial species such as *B. pseudomallei*, LPS has a major role in stimulating host innate immune response during infection (Nelson et al., 2004). Previous studies have shown that *B. pseudomallei* LPS is required for serum resistance and virulence (DeShazer et al., 1998). It has been well established in many bacterial diseases that overstimulation of the host cells by LPS can lead to the features of septic shock and septicaemia is a major cause of death. Cellular recognition of LPS by the innate immune system triggers the proinflammatory cytokines by host cells, which aids in the clearance of the pathogen. With high concentrations of antibodies to LPS associated with survival in severe melioidosis demonstrated a potential role for *B. pseudomallei* LPS in protective immunity (Ho et al., 1997). The LPS mutant *B. pseudomallei* strain SRM117, which lacked the O-antigenic polysaccharide moiety, was more susceptible to macrophage killing during the early phase of infection than its parental strain 1026b (Arjcharoen et al., 2007).

Anuntagool and collaborators reported 96% of the LPS from clinical isolates possessed a sodium dodecyl suflate-polyacrylamide gel electrophoresis (SDS-PAGE) silver staining ladder profile referred to as a typical ladder and approximately 3% that exhibited different ladder characteristics referred to as an atypical ladder. The remaining 1% had no ladder detectable in the high molecular weight region but possessed a low molecular weight silver-staining band below the 29kDa marker that was designated as a no ladder LPS. All three LPS types were antigenically distinct as evident from immunoblot reactivity against pooled sera from patients from whom the typical or atypical LPSs were isolated (Anuntagool et al., 2000a). To investigate the possible biological significance of the different LPS phenotypes, Anuntagool and colleagues then expanded their study to include environmental isolates as well as clinical isolates form patients with different clinical manifestations. They extracted LPS from 1327 isolates and tested with pooled patient sera (Anuntagool et al., 1998). In this larger series of isolates, the proportion of isolates possessing different LPS patterns was similar to that in the initial report, showing a ratio of 97:2:1, reflecting typical, atypical and no-ladder LPS ratios. Overall 99% of the isolates possessed smooth type LPS exhibiting two different ladder profiles as well as being serologically distinct, most likely reflecting two different repeating polysaccharide side chains, termed O-antigens.

The most common serotype is referred to as serotype A (97%), whilst the less abundant one (2%) is known as serotype B, referring to the previous terms of typical and atypical LPS respectively (Anuntagool et al., 2000a). The remaining 1% without a ladder profile and exhibiting negative seroreactivity, against sera from patients infected with serotype A and B isolates, was suggested to represent only a lipid A-core oligosaccharide, known as rough LPS (Anuntagool et al., 2006). **Figure 1.10** shows the SDS-PAGE gel and silver stain of *B. pseudomallei* 576 (serotype B), *B. pseudomallei* K96243 (serotype A) and *B. thailandensis* with their respective ladder profiles.

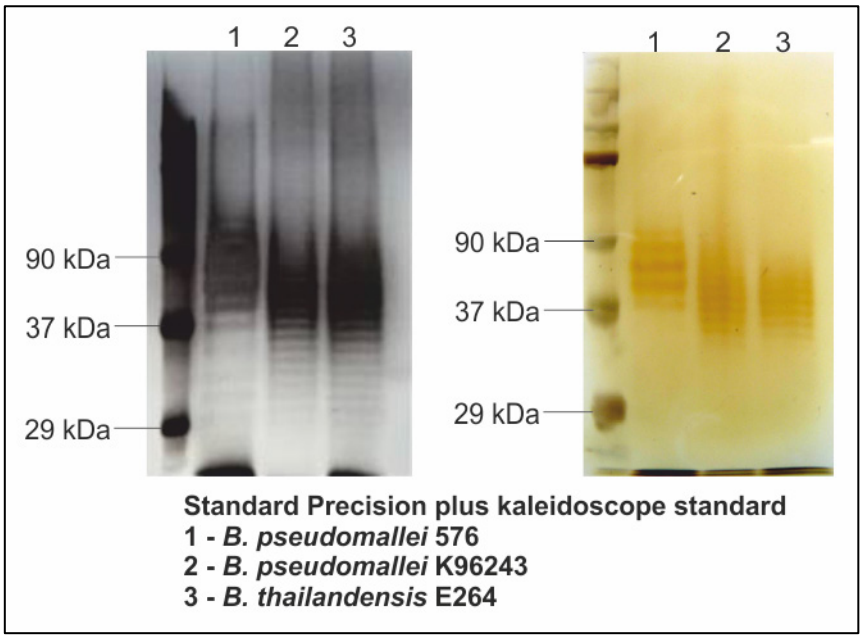


Figure 1.10 –Silver stain and SDS-PAGE of *B. pseudomallei* and *B. thailandensis*.
The figure shows the silver stain and SDS-PAGE gel of *B. pseudomallei* 576 (serotype B), *B. pseudomallei* K96243 (serotype A) as well as *B. thailandensis* E264 (serotype A) as run on a 10% novex SDS 1.5mm, 10 well gel alongside a precision plus kaleidoscope standard. The gels were run by our collaborators at dstl.

A much higher proportion of the Australian isolates possessed the two less common serotype B and rough LPS types (19.3%) in comparison with the Thai isolates (2.7%). The LPS of clinical isolates were more heterogeneous than the LPS of environmental isolates, whilst all environmental isolates had the smooth type LPS and all except one belonged to serotype A. When quantitated for the capacity to produce biofilm the three LPS types differed from one another. Those with a rough LPS exhibited the highest biofilm-producing capacity and those with a smooth type serotype A profile showed the lowest mean biofilm producing capacity. The increased capacity to produce biofilm by rough type LPS serotypes, may allow the organism to evade host defences and therefore survive better inside the host. Whether biofilm production is important in the pathogenesis of melioidosis is still unclear (Taweechaisupapong et al., 2005, Anuntagool et al., 2006).

Since LPS has been implicated as one of the virulence factors for a number of Gram-negative bacteria, including *B. pseudomallei* (DeShazer et al., 1998), possible associations between the LPS types and various clinical parameters of the patients from whom these isolates were obtained were investigated. Anuntagool and co-workers found no association between the LPS type and disease severity, clinical manifestations or underlying risk factors. The two less abundant LPS patterns were found more in isolates obtained from patients with relapse melioidosis than from those with primary infection, suggesting that the bacteria with the less common LPS (serotype B) are able to survive better in their host in comparison to those with the more common LPS type (serotype A). Since each LPS type was found to be antigenically distinct, antigenic polymorphism among the *B. pseudomallei* LPS may allow the bacteria to evade host immunity. These uncommon LPS types were not associated with a survival advantage *ex-vivo* because bacteria with different LPS phenotypes were equally resistant to killing and equally able to replicate in 30% normal human serum (Anuntagool et al., 2006).

Due to the importance of LPS phenotypic diversity for serology and diagnostics Tuanyok and co-workers investigated the genetic and molecular basis of differential LPS phenotypes in a large *B. pseudomallei*

population using bioinformatics, phenotypic characterisation as well as population genetics approaches (Tuanyok et al., 2012). Comparative analysis was used to assess the diversity of genes responsible for the biosynthesis of the O-antigen moiety of LPS in *B. pseudomallei* using publicly available genomes to identify differences within LPS biosynthetic genes. A total of 27 *B. pseudomallei*, 10 *B. mallei* and 3 *B. thailandensis* and two *B. oklahomensis* isolates were compared to identify the LPS O-antigen biosynthesis genes. Three different O-antigen biosynthesis gene categories, or genotypes, were identified. The *B. pseudomallei* O-antigen biosynthesis genes were assigned to two major groups; group A referring to LPS genotype A, was identical or very similar to the O-antigen biosynthesis operon observed in *B. pseudomallei* 1026b and K96243, and group B, referring to LPS genotype B, which was found in an atypical LPS strain 576 (**Figure 1.11**).

DeShazer and colleagues had previously identified seven genes in the O-antigen biosynthesis operon in *B. pseudomallei* 1026b by Tn5-OT182 mutagenesis, which were responsible for O-antigen biosynthesis and serum resistance and included *mlB*, *mlD*, *wbiA*, *wbiC*, *wbiE*, *wbiG* and *wbiI*. The gene *wbiE* of *B. pseudomallei* strain K96243 and gene BUC_3396 of strain 576 were used to represent the presence of LPS genotypes A and B, respectively and a total of 999 *B. pseudomallei* strains from different geographic locations and epidemiological origins, including clinical, animal and environmental strains, were tested for their LPS genotypes. Tuanyok and collaborators further observed a frame-shift mutation in the O-antigen biosynthesis *wbiI* gene of *B. pseudomallei* strains MSHR1655, which was correlated with its rough phenotype. The mutation was found to encode an oligosaccharide epimerase/dehydratase and was conserved in all O-antigen biosynthesis gene clusters of *B. pseudomallei*, suggesting that a mutation in this gene probably impacts on the synthesis of the O-antigen (Cheng et al., 2004).

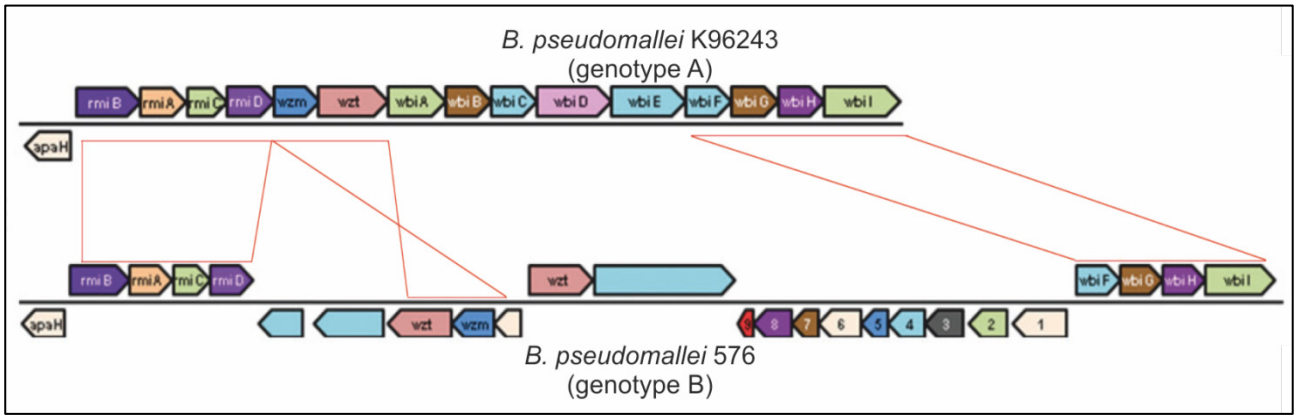


Figure 1.11 – Biosynthesis gene cluster for the O-antigen moiety in *B. pseudomallei* K96243 and 576.
The schematic is adapted from Tuanyok and colleagues (Tuanyok et al., 2012) who compared the LPS O-antigen biosynthesis gene cluster of *B. pseudomallei* K96243 of serotype A (top) with that of *B. pseudomallei* 576 of serotype B (bottom).

The LPS genotype A was found in most *B. pseudomallei* and all *B. mallei* and *B. thailandensis* genomes examined. It was also the most common genotype found in both Australian (85.3%) and Southeast Asian (97.7%) strain populations, whilst the LPS genotype B was relatively rare in Southeast Asian strains (2.3%), but was found in 13.8% of Australian strains. Additionally, five of the Australian strains were not able to be typed using PCR and they were further analysed for O-antigen biosynthesis gene identification using whole genome sequencing. Comparative genomics demonstrated that many genes in this new cluster were similar to those of the LPS genotype B genes of *B. pseudomallei* 576 and were distinct from K96243 LPS genotype

A genes. They were therefore newly identified as a variant of LPS genotype B and were designated as genotype B2. Because the LPS genotypes B and B2 were frequently found in Australia but not in Southeast Asia, it could be argued that this finding may be due to different therapies used for clinical cases in these two endemic locations, however, most of the samples were collected before antibiotic treatment and some of the B strains were collected form soil and animals, thus confirming that the occurrence of LPS types B and B2 in Australia is not associated with the exposure to antibiotics or treatment therapy. B2 strains were also killed in growth media containing 30% N-hydroxysuccinimide (NHS), whereas the LPS genotype B strains were resistant (Tuanyok et al., 2012).

The LPS of *B. pseudomallei* was not assumed to readily switch its LPS phenotype from A to B as the gene composition of the different genotypes are very different and a simple switching mechanism is difficult to envision, whilst some rough LPS strains were even found to have mutations in their O-antigen biosynthesis genes. Further biochemical characterisation and structural elucidation of the different O-antigens and lipid A of *B. pseudomallei* LPS should provide additional insights into the possible role of LPS in pathogenesis, diagnosis and vaccine development and these investigations are a focus of this thesis (**Chapter** 4).

1.3.5 *B. pseudomallei* polysaccharide antigens and their structures

B. pseudomallei produces several surface-associated molecules that have been proposed for use as vaccine components and serological diagnostic agents. These include lipopolysaccharides (LPS), capsular polysaccharides (CPS), expopolysaccharides (EPS) and flagella (Brett et al., 1994, Brett and Woods, 1996, Bryan et al., 1994). The immunological importance of the polysaccharides has been demonstrated by the reduction in the lethality of infection in mice and diabetic rats when passively immunised with antibodies to the LPS or CPS. Immunisation of mice with LPS or CPS provides protection against subsequent challenge with *B. pseudomallei*, suggesting that these polysaccharides have potential as vaccine candidates (Sarkar-Tyson et al., 2007).

The LPS and CPS of *B. pseudomallei* have been intensively studied and their gene clusters have been identified (Sarkar-Tyson et al., 2007, Cheng and Currie, 2005). The LPS O-antigen seems to have properties typical for bacterial O-antigens, such as providing resistance to serum killing, whilst the function of the CPS is less clear but it may provide resistance to phagocytosis by reducing levels of complement C3 deposition. Further putative gene clusters of polysaccharides termed type III O-PS and type IV O-PS were also reported (Cheng and Currie, 2005, Sarkar-Tyson et al., 2007). These clusters are not found in *B. mallei*, but are conserved in *B. thailandensis* and their putative functions are polysaccharide biosynthesis and transport, indicating that these putative polysaccharide clusters have a possible role in environmental survival. Furthermore there is an already previously reported exopolysaccharide (DeShazer et al., 1998) and studies have shown glycosylation of the *B. pseudomallei* flagella (Scott et al., 2011).

The O-antigen is the most structurally diverse LPS component within a species and over 170 known structures have been found in *E. coli*. The complication with LPS as a potential vaccine target is the fact that there are four different *B. pseudomallei* O-antigen genotypes, including LPS genotypes A, B, B2 and a rough LPS type. Furthermore, sero-crossreactivity is common among O-antigens of *Burkholderia* species (Stone et

al., 2012). As previously discussed, most *B. pseudomallei* strains express the genotype A O-antigen, making it the most abundant structure, whilst the atypical genotypes B and B2 are related serologically, but have distinct ladder banding profiles on SDS-PAGE (Tuanyok et al., 2012). The only LPS O-antigen that has been characterised is serotype A, whilst LPS type B has not been found in any other species, other than *B. pseudomallei* and type B2 was recently described in a *B. thailandensis* species, found in soil and water in northern Australia.

1.3.5.1 LPS structures

Knirel and collaborators were the first group to structurally characterise the polysaccharides of *B. pseudomallei* in 1992. They demonstrated the presence of two partially O-acetylated O-antigenic polysaccharides, which they named PS-I and PS-II. Their methylation analysis and ^{1}H and ^{13}C NMR spectroscopy including NOE experiments showed PS-I to have the structure [-3)-β-$_{D}$-glucopyranose-(1-3)-6-deoxy-α-$_{L}$-talopyranose-(1-] and PS-II to have the structure [-3)-2-O-acetyl-6-deoxy-β-$_{D}$-manno-heptopyranose-(1-] (Knirel et al., 1992). In 1995, Perry and co-workers further characterised the structures of the serologically typical strain of *B. pseudomallei* strain 304b. They reported two smooth type LPS, and termed them O-PS I and O-PS II reflective of Knirel's annotation. They found that 33% of the L-6dTal*p* residues in O-PS II bear 2-O-methyl and 4-O-acetyl substituents while the other L-6dTal*p* residues carry only 2-O-acetyl substituents (**Table 1.2**) (**Figure 1.4**) (Perry et al., 1995, Knirel et al., 1992). Rabbit polyclonal antibody and a monoclonal antibody raised against a tetanus toxoid conjugate of *B. pseudomallei* O-PS II were found to be protective against challenge in an animal model of infection. The same antibodies recognised more than 40 strains of *B. pseudomallei* isolated from infected patients in North-Eastern Thailand, indicating the presence of a common antigen.

A cluster of 15 genes was located that is required for biosynthesis of the type II O-PS moiety of LPS in *B. pseudomallei* and seems to provide resistance to complement-mediated bacteriolysis. Furthermore, it was found that *B. pseudomallei* has the ability to invade and survive within the bloodstream, which is important in the pathogenesis of septicaemic melioidosis (DeShazer et al., 1998). DeShazer and colleagues were the first to suggest that type I and II O-PS were not both LPS moieties and that type I O-PS was in fact a CPS (**Section 1.3.5.2**), whilst the type II O-PS was the LPS. Serum sensitive mutants SRM108, SRM109 and SRM117 were deficient in the LPS moiety but produced the CPS moiety and were less virulent than the parental strain 1026b in three animal models of melioidosis. The C5b-C9 membrane attack complex (MAC) was also found to be deposited on the bacterial surface and on serum sensitive mutants implying a correlation between LPS presence and serum resistance. Immunogold electron microscopy studies further demonstrated that the lack of LPS altered the distribution and appearance of the CPS. DeShazer and colleagues explained this by suggesting that the CPS was responsible for preventing the integration of the MAC complex into the bacterial outer membrane and that the LPS moiety plays an indirect, supportive and maintaining, role by influencing the organisation and or arrangement of the CPS to effect MAC complex formation far away from the bacterial surface or preventing its insertion into the outer membrane. This was assumed as the closely related *B. thailandensis* wasn't shown to produce CPS, yet was resistant to 30% NHS (DeShazer et al., 1998).

The O-antigen of *B. mallei*, is virtually identical to the LPS of *B. pseudomallei* except that it lacks acetyl modifications at the O-4 position of the L-6d-Tal*p* residues. The genes required for the synthesis of the *B. mallei* LPS O-antigen have been identified (Burtnick et al., 2002). These genes are clustered on an approximately 19kb region of chromosome 1 and encode 16 ORFs (Nierman et al., 2004). These ORFs are identical to those encoding the LPS O-antigen of *B. pseudomallei* (DeShazer et al., 1998). Sequence alignment of the *B. pseudomallei* and *B. mallei* LPS biosynthetic regions revealed 99% identity at the nucleotide level and the genes comprising the *B. mallei* O-antigen biosynthetic operon were named as per the identical genes found in *B. pseudomallei*. As there are no differences between the O-antigen loci, it has been suggested that the different pattern of O-acetylation is due to the presence of an unlinked locus that is either not present or non-functional in the genome of *B. mallei* (Burtnick et al., 2002, Coenye and Vandamme, 2007). The presence of O-acetyl groups on the LPS moiety was concluded to have consequences when the O-antigen is considered as a component of a vaccine that protects against both *B. mallei* and *B. pseudomallei* (Burtnick et al., 2002, Brett et al., 2003).

B. thailandensis and *B. pseudomallei* also have similar O-antigens, and *B. thailandensis* has similar modification patterns as *B. pseudomallei* (Cheng and Currie, 2005, Sarkar-Tyson et al., 2007). The *wbiA* allele in *B. thailandensis* and *B. pseudomallei* showed conserved amino acid motifs that define a family of inner membrane trans-acylases, suggesting that the function of the *wbiA* gene would be similar in both species. Brett and colleagues investigated whether the *wbiA* gene was required for the expression of the LPS moiety (Brett et al., 2003). A protective monoclonal antibody reacted with the LPS of the wild-type *B. pseudomallei*, but not with a *B. pseudomallei wbiA* null mutant. It was shown that WbiA activity was required for the acetylation of the L-6dTal*p* residues at the O-2 position and that such structural modification of the LPS was critical for the recognition with the antibody (Brett et al., 2003). It was further reported that the O-antigen acetylase (OacA) gene is known to be involved in the acetylation at the O-4 position of *B. thailandensis*, with its homolog in *B. pseudomallei* K96243 being identified as BPSL1936 (Brett et al., 2011). Point mutations were inserted in the *oacA* genes of both species, to determine if the gene played only a single role in the side group modifications of the talose residues or a dual role in combination with the synthesis of the O-antigen. Both strains were tested for O-antigen production and immunogenic specificity and OacA was found not to be involved in the synthesis of the O-antigen, as the mutations still showed bands and strong interaction with *B. mallei* LPS-specific monoclonal antibody. However the OacA was speculated to possess both acetyletransferase and methyltransferase activites, as both modifications of O-antigens were affected (Brett et al., 2011).

Anuntagool and colleagues found variable serotypes in the heterogeneity among the *B. pseudomallei* LPS, where bacteria with the less common LPS survive better in hosts than those with the more common LPS type (Anuntagool et al., 2006, Sarkar-Tyson et al., 2007). Since each LPS type is antigenically distinct, antigenic polymorphism among the *B. pseudomallei* LPS may allow the bacteria to evade host immunity. Therefore, structural elucidation of the different O-antigens and the lipid A of *B. pseudomallei* LPS should provide additional insights into the possible role of LPS in pathogenesis, diagnosis and vaccine development. The O-antigen repeating unit of serotype A LPS has been characterised (Anuntagool et al., 2000a, DeShazer et al., 1998, Knirel et al., 1992, Perry et al., 1995, Brett et al., 2003), as well as its lipid A structure (Novem et al., 2009). The O-antigens of the strains of serotypes B and B2 have not been

structurally characterised, and the O-antigen repeat unit structure of serotype B, strain *B. pseudomallei* 576, is presented in this thesis (**Chapter** 4).

1.3.5.2 Capsular polysaccharides

The capsule forming polysaccharide of *B. pseudomallei* K96243 was reported to be an unbranched homopolymer with the structure [-3)-2-O-acetyl-6-deoxy-β-$_D$-manno-heptopyranose-(1-] by Knirel and colleagues (**Table 1.2**) (Knirel et al., 1992). However, as described in the previous section, Perry and colleagues subsequently suggested that this sequence as the O-antigen component of a second LPS found on *B. pseudomallei* which they called O-PS I. They suggested that simultaneous production of two or more LPS O-antigens by a Gram-negative bacterium was not unusual, such as is seen with *Salmonella boecker* (group H), which produced two distinct O-antigens, whilst *Salmonella madelia* has been demonstrated to produce three structurally different LPS O-antigens (Perry et al., 1995). Nevertheless, Reckseidler and co-worker's later investigations suggested that this polysaccharide is a capsule rather than an O-antigen moiety. An important piece of evidence was the discovery that the genes involved in the production of this polysaccharide demonstrated strong homology to the genes involved in the production of CPS in many organisms including *E. coli*, *N. meningitides and H. influenza* (Perry et al., 1995, Reckseidler et al., 2001).

1.3.5.3 Other oligo- and polysaccharides

Further types of glycopolymers, in addition to LPS and CPS, are expressed on the surface of *B. pseudomallei.* These are exopolysaccharides (EPS) and glycoproteins.

It remains to be determined whether more than one EPS is expressed by this organism. EPS has been reported to be comprised of an unbranched polymer of repeating tetrasaccharide units with the structure [-3)-2-O-acetyl-β-$_D$-Gal*p*-(1-4)-α-$_D$-Gal*p*-(1-3)-β-$_D$-Gal*p*-(1-5)- β-$_D$-KDO*p*-(2-] (**Table 1.2**) (DeShazer et al., 1998, Masoud et al., 2009). However, upon genomic analysis of *B. pseudomallei*, further to the already characterised LPS (type II O-PS) and CPS (type I O-PS), two additional putative gene clusters implicated in polysaccharide biosynthesis termed type III O-PS and type IV O-PS were reported (Cheng and Currie, 2005, Sarkar-Tyson et al., 2007). The type III O-PS is encoded by a cluster of 13 genes to which have been assigned putative functions related to polysaccharide biosynthesis and transport (Holden et al., 2004), whilst type IV O-PS genes show putative functions of CPS biosynthesis and polysaccharide transport proteins. The EPS structure described above is present in both *B. pseudomallei* and *B. mallei* yet the genome of the latter did not contain the putative type III and type IV O-PS clusters, suggesting that these clusters are not responsible for the EPS biosynthesis. (Sarkar-Tyson et al., 2007). Further work is required to establish the function of the type III and IV gene clusters.

Like EPS, the extent of protein glycosylation in *B. pseudomallei* is poorly understood. However it is now firmly established that the flagellins in this organism are O-glycosylated. Scott and colleagues used a gel-based glycoproteomics approach to identify glycoproteins of *B. pseudomallei* and the closely related but non-pathogenic *B. thailandensis*, and found that the flagellin proteins of both species were modified by novel glycans (Scott et al., 2011). The identification of the major flagellar filament protein FliC as a putative

glycoprotein is not without precedent as this protein has been shown to be glycosylated in a range of other bacteria such as *H. pylori*, *C. botulinum*, *C. difficile*, *C. jejuni* and *P. aeruginosa*. Previous work had indicated that the FliC proteins of four *B. pseudomallei* strains have experimentally determined masses some 3.3kDa larger than expected and it was suggested that posttranslational modifications might explain this discrepancy. Inactivation of *rmlB* resulted in non-glycosylated FliC in *B. pseudomallei* and also to abrogate motility. Interestingly *rmlB* mutants still have flagellar filaments on their surfaces visible by transmission electron microscopy, indicating that the loss of motility is not due to defects in export and assembly but rather in the function of the flagellar filament itself. Furthermore it was shown that the *rmlB* mutation also resulted in loss of O-antigen synthesis of the LPS, as the *rmlB* gene is located within the gene cluster responsible for LPS O-antigen. This, however, would also suggest that the defects in motility could be due to defects in membrane structure.

Species	Serotype	Antigen	Structural details
BP	A	CPS	-3)-2-O-acetyl-6-deoxy-β-$_D$-manno-heptopyranose-(1-
BP	A	O-PS III	Genetic locus only
BP	A	O-PS IV	Genetic locus only
BP	A	EPS	-3)-2-O-acetyl-β-$_D$-Gal*p*-(1-4)-α-$_D$-Gal*p*-(1-3)-β-$_D$-Gal*p*-(1-5)-β-$_D$-KDO*p*-(2-
BP	A	LPS	-3)-β-$_D$-glucopyranose-(1-3)-6-deoxy-α-$_L$-talopyranose-(1- 33% variable L-6dTal*p*: 2-O-methyl / 4-O-acetyl
BP	**B**	**LPS**	**uncharacterised**
BP	**B2**	**LPS**	**uncharacterised**
BT	A	LPS	-3)-β-$_D$-glucopyranose-(1-3)-6-deoxy-α-$_L$-talopyranose-(1- variable L-6dTal*p* 4-O-acetyl
BM	A	LPS	-3)-β-$_D$-glucopyranose-(1-3)-6-deoxy-α-$_L$-talopyranose-(1- variable L-6dTal*p* 2-O-acetyl

Table 1.2 – Summary of polysaccharide structures of *B. pseudomallei* and *B. thailandensis* known to date.

1.3.6 Infection models employed in the study of melioidosis

Melioidosis has several disease outcomes ranging from asymptomatic, acute, chronic or latent. These are determined by the host immune response. The use of infection models is central to the study of microbial pathogenesis; in combination with genetic, immunologic and antigen purification techniques, much can be learned regarding the pathogenesis of disease due to microbes. A number of animal models have been developed in order to understand the pathogenesis of melioidosis to function as a systems for assessing the efficacy of antibiotics and vaccine candidates *in vivo* (Coenye and Vandamme, 2007).

One of the most frequently used animal models for the study of *B. pseudomallei* pathogenesis is the Syrian golden hamster model of acute *B. pseudomallei* infection. It was demonstrated many years ago that a number of animals could be infected with *B. pseudomallei* and the hamster proved to be the most susceptible and was used to assess the lethality of a number of strains. Rats have been shown to be relatively resistant to *B. pseudomallei* infection, however recently a number of mouse models have been developed in order to represent the various manifestations of melioidosis and the major routes of infection by *B. pseudomallei* (Coenye and Vandamme, 2007).

Murine models of acute (BALB/c) and chronic (C57BL/6) infections mimic these immune states in humans. The acute melioidosis of BALB/c mice is characterised by a significantly stronger innate immune response (Hoppe et al., 1999, Ulett et al., 2000) where hyperproduction of proinflammatory cytokines result in an

inappropriate cellular response that fails to control the infection and contributes to tissue destruction and multiple organ failure. Chronic infection (C57BL/6 mice) demonstrates moderate cytokine increase that allows mice to temporarily confine *B. pseudomallei* within phagocytes, allowing time for an adaptive immune response to occur (Hoppe et al., 1999, Ulett et al., 2000).

Most studies conducted to investigate treatments or therapies for melioidosis have involved murine models using BALB/c or C57BL/6 mice. BALB/c mice are highly susceptible to infection and are less resistant to infection, as compared to C57BL/6 mice. The Th1 response of C576BL/6 mice is important for macrophage activation, whereas BALB/c mice provoke a Th2 biased response inhibiting macrophage activation and stimulation of antibody production. These variations in disease presentation, suggest macrophage activation to be of importance in combating melioidosis (Titball et al., 2008).

Murine model studies have also shown protective immunity against *B. pseudomallei* to be induced by a range of living and nonliving immunogens, however the strongest response was induced by live attenuated immunogens. The problem with such vaccines lies with concerns about latency, thus making it unlikely that such vaccines will be appropriate for use in humans (Sarkar-Tyson and Titball, 2010).

1.3.7 *B. pseudomallei* virulence factors

A number of virulence factors have been proposed to be involved in the pathogenesis of *B. pseudomallei*. These include: CPS, LPS, type III secretion systems (T3SS), flagella and quorum-sensing (QS) molecules, type IV pili-mediated adherence and secreted proteins, including haemolysin, olipases and proteases, the type VI secretion system (T6SS), secreted factors and regulatory genes, which play a more moderate role in virulence (Essex-Lopresti et al., 2005, Wiersinga et al., 2006, Lazar Adler et al., 2009). LPS as a potential virulence factor is discussed below because of the relevance to the work described in this thesis.

As described earlier (**Section 1.3.4.4**), four distinct antigenic types of LPS of *B. pseudomallei* have been reported. These showed no immunological cross reactivity, but they shared similar endotoxic levels in the limulus amoebocyte lysate assay as well as the levels of macrophage activation (Anuntagool et al., 2000a, Anuntagool et al., 2006). Furthermore, LPS was shown to play a role in the resistance of *B. pseudomallei* to host cationic antimicrobial peptides and complement-mediated killing by DeShazer and colleagues, who showed that serum sensitive mutants of *B. pseudomallei* lack the LPS moiety. The LPS-deficient mutants were attenuated in hamsters, guinea pigs and diabetic rats, due to being susceptible to complement mediated killing. The proteins lacking in the serum-sensitive mutants were involved in bacterial polysaccharide biosynthesis, suggesting the LPS to be essential for *B. pseudomallei* serum resistance and virulence (DeShazer et al., 1998). In addition, *B. pseudomallei* mutants with transposon insertions within genes responsible for LPS core biosynthesis (*waaF*) and the regulation of the peptidoglycan and phospholipid synthesis (*lytB*) were identified as polymyxin-B sensitive (Burtnick and Woods, 1999).

Patients with localised and septicaemic melioidosis produce antibodies against both the CPS and LPS, however, the most potent immunogen in both types of infection was shown to be the LPS moiety (Ho et al., 1997), which would indicate that the LPS may be a critical constituent of any vaccine designed for the

prevention of melioidosis (Brett and Woods, 1996). In-vitro studies in mouse macrophage cell lines using LPS mutants suggested a further possible role for the O-antigenic polysaccharide moiety of LPS in internalisation and intracellular survival of *B. pseudomallei*. In comparison to *E. coli* LPS, however, the LPS of *B. pseudomallei* is a less potent activator of macrophages (West et al., 2008, Utaisincharoen et al., 2000, Arjcharoen et al., 2007).

1.3.8 Host defence and response during melioidosis

The compartmentalised gene-expression profile of important inflammatory molecules has been described in murine and human melioidosis (Wiersinga et al., 2008c, Wiersinga et al., 2007a). Interferon-γ (IFN-γ) is essential for resistance against *B. pseudomallei* infection with additionally vital roles played by tumour necrosis factor α (TNF-α), interleukin (IL)-12 and IL-18, whilst murine studies on CD4+ and CD8+ T cells, highlighted the importance of CD4 (Barnes et al., 2008, Easton et al., 2007, Haque et al., 2006b, Wiersinga et al., 2007c, Haque et al., 2006a).

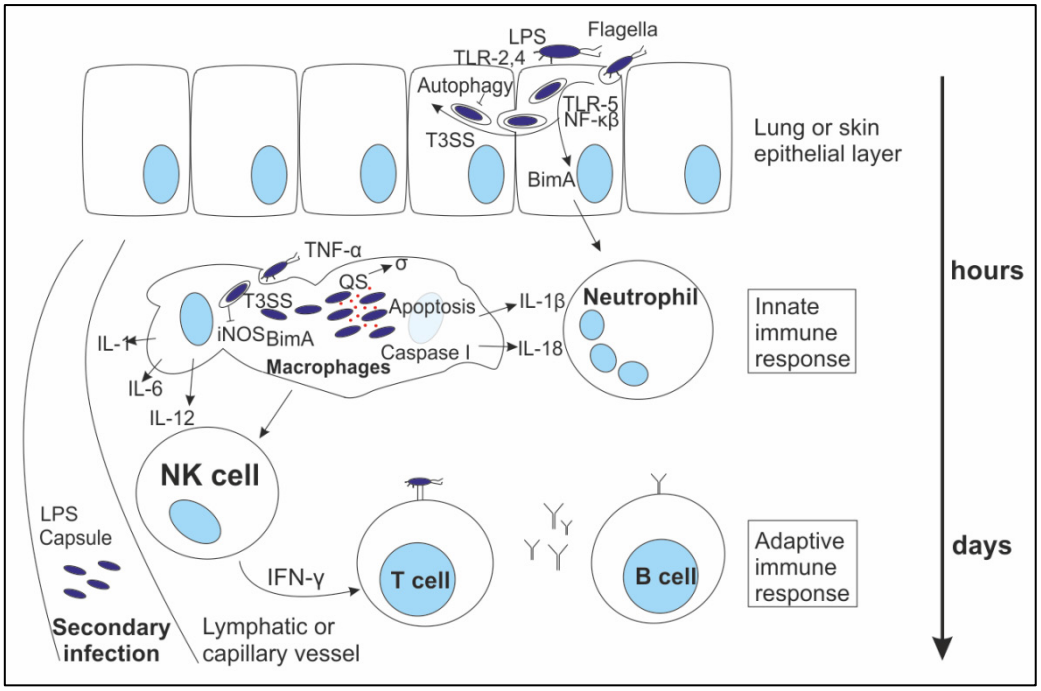


Figure 1.12 – Molecular and cellular pathogenesis of melioidosis.
The schematic as adapted from Lazar-Adler and colleagues (Lazar Adler et al., 2009) represents the current understanding of *B. pseudomallei* pathogenesis from the environment, to epithelial attachment by LPS, CPS and flagella followed by the invasion of epithelial cells. Within these cells, the T3SS effectors assist in vacuolar escape, thus evading host autophagy and BimA-mediated actin polymerisation allows for intracellular motility. Reruitment of innate immune cells is stimulated by the activation of TLR2 and TLR4 by *B. pseudomallei* LPS and flagella, which result in proinflammatory cytokine release and provide an intracellular niche for replication of the bacteria. Secondary spread occurs via the lymphatic or capillary vessels. With the progression of the infection, the adaptive immune response with T cells and B cells occurs.

Antibody formation is important for protection against infection by *B. pseudomallei*, but does not seem to be essential for primary resistance (Haque et al., 2006b). Neutrophils on the other hand are rapidly recruited to the primary site of infection after infection and depletion of neutrophils is associated with a 1000-fold increase in pulmonary bacterial loads within 4 days (Barnes et al., 2001, Easton et al., 2007, Wiersinga et al., 2008a, Wiersinga et al., 2008d). The myeloid differentiation primary response gene 88 (MyD88) is the key signalling adaptor for most toll-like receptors (TLRs) and the IL-1 and IL-18 receptors are crucially

involved in protective neutrophil recruitment (Wiersinga et al., 2008d). Macrophage-depleted mice further showed an accelerated mortality during experimentally induced melioidosis compared with untreated mice, suggesting an important role for macrophages in early host defence (Breitbach et al., 2006). **Figure 1.12** shows a summary of the molecular and cellular basis of pathogenesis of melioidosis as caused by *B. pseudomallei* as described in the following sections (**Section 1.3.8.1** and **Section 1.3.8.2**)

1.3.8.1 *B. pseudomallei* interaction with the innate immune system

B. pseudomallei activates the alternative complement pathway, but as previously established the membrane attack complex is deposited on the capsular polysaccharide and is therefore not bactericidal. Furthermore opsonisation with a complement enhances, but is not essential for uptake by phagocytes and does not necessarily result in intracellular killing of the bacteria (Kespichayawattana et al., 2000). The CPS and LPS further explain *B. pseudomallei* resistance to lysosomal defensins and cationic peptides, which also allow the bacteria to survive within phagocytes and human serum (Gan, 2005).

Following infection with *B. pseudomallei*, mouse tissue shows rapid influx and activation of neutrophils and acute infection is established upon depletion of neutrophils in C57BL/6 mice showing their importance in immunity (Easton et al., 2007). Macrophages were shown to be essential for the control of *B. pseudomallei* as the depletion of macrophages in BALB/c or C57BL/6 mice significantly increased mortality rates (Breitbach et al., 2006). However *B. pseudomallei* infection in BALB/c mice failed to attract macrophages to the same extent as in C57BL/6 mice and these cells were proposed to temporarily contain *B. pseudomallei*, resulting in chronic melioidosis (Barnes et al., 2001). In comparison to healthy individuals, the macrophages from patients showed reduced levels of lysosomal fusion, which resulted in a higher number of bacterial numbers, suggesting acute melioidosis to stem from an ineffective cellular innate immune response (Puthucheary and Nathan, 2006).

1.3.8.1.1 Toll-like receptors and *B. pseudomallei* recognition

Toll-like receptors (TLRs) are innate immune receptors that together with other pattern recognition receptors (PRRs) are the first to detect host invasion by pathogens and form the crucial link between the innate and adaptive immunity. TLRs and PRRs recognise conserved motifs on pathogens termed “pathogen-associated-molecular-patterns” (PAMPs), which in *B. pseudomallei* include LPS, lipid A, the peptidoglycan of the bacteria, flagella, T3SS as well as DNA and mediate an inflammatory immune response (West et al., 2008, Wiersinga et al., 2007b). Activation of TLRs occurs via various signalling adaptor proteins, including MyD88 and TRIF. MyDD88 knock-out mice showed increased susceptibility to *B. pseudomallei* infection due to reduced neutrophil recruitment and activation (Wiersinga et al., 2008d).

B. pseudomallei LPS is suggested to be the prototypical bacterial ligand for TLR4. The lipid A usually interacts with TLR4 in addition to proteins CD14 and MD-2 to induce a pro-inflammatory signal. MD-2 is a soluble molecule that is associated with the extracellular domain of TLR4 and is generally regarded as essential for LPS recognition by TLR4. Both TLR2 and TLR4 signal via the adaptor molecule MyD88 although TLR4 may also signal in a MyD88-independent fashion via the adaptor molecule TIR-domain-

containing adaptor-inducing interferon-β (TRIF) (West et al., 2008). It is still not clear whether purified *B. pseudomallei* LPS signals via TLR2, TLR4 or both; however, during experimental melioidosis CD14 and TLR2 deficient mice displayed a markedly improved host defence. This was reflected by strong survival advantages together with decreased bacterial loads, reduced proinflammatory cytokine levels and reduced organ injury. TLR4 deficient mice, however, were indistinguishable from wild type mice with respect to bacterial outgrowth and survival in experimentally induced melioidosis, which undermines TLR4 as main receptor (West et al., 2008, Wiersinga et al., 2008b). However other MyD88-dependent signalling pathways may also be important in the host response to *B. pseudomallei* infection.

1.3.8.2 *B. pseudomallei* and the adaptive immune response

Human memory T cell responses to *B. pseudomallei* was investigated by Tippayawat and co-workers, who tested blood from 133 healthy individuals who lived in endemic areas with no history of melioidosis. In animal models of melioidosis, the production of IFN-γ is critical for resistance, but in humans the characteristics of IFN-γ production and the bacterial antigens that are recognised by the cell-mediated immune response have not been defined. *B. pseudomallei* is a potent activator of human peripheral blood natural killer (NK) cells for innate production of IFN-γ. Healthy individuals in this study, with serological evidence of exposure to *B. pseudomallei* and patients recovered from active melioidosis developed CD4+ and CD8+ T cells that recognised whole bacteria and purified proteins LolC, OppA and PotF, which are members of the *B. pseudomallei* ABC transporter family. The response was primarily mediated by terminally differentiated T cells of the effector-memory phenotype and correlated with the titre of anti-*B. pseudomallei* antibodies present in the serum. Individuals living in a melioidosis-endemic region therefore showed clear evidence of T cell priming for the ability to make IFN-γ, correlating with their serological status. The ability to detect T cell responses to defined *B. pseudomallei* proteins in large numbers of individuals now provides the opportunity to screen candidate antigens for inclusion in protein or polysaccharide-conjugate subunit vaccines against this important disease (Tippayawat et al., 2009).

1.3.9 *B. pseudomallei* as a bioweapon

The lack of a vaccine against *B. pseudomallei* and melioidosis makes this a potential candidate for bioweaponisation. With the high rate of infectivity via aerosols and resistance to many common antibiotics, both *B. pseudomallei* and *B. mallei* have been classified as category B priority pathogens by the US National Institutes of Health (NIH) and the US Centres for Disease Control and Prevention (CDC), which spurred a dramatic interest in the microorganism for research into virulence factors and vaccine development (Jeddeloh et al., 2003, Warawa and Woods, 2002, Coenye and Vandamme, 2007, Rotz et al., 2002). *B. mallei* was used as a bioweapon in both World War I and II, specifically in the first World War, where *B. mallei* was used by the Central Powers to infect Russian horses with glanders on the eastern front, with significant effects (Cheng and Currie, 2005, Bondi and Goldberg, 2008). It is also believed that biological weapons research using *B. pseudomallei* occurred in the former USSR, although the extent of this effort and the possibility of engineered antibiotic resistant strains remain unknown. Other countries with military interest in *B. pseudomallei* included the United States and possibly Egypt (Cheng and Currie, 2005).

The potential risk posed by *B. pseudomallei* as a bioweapon is uncertain. Melioidosis carries a potentially high mortality rate and its causative agent has intrinsic antibiotic resistance and a wide host range. However weaponisation has not been known to have been performed, the disease does not spread from person to person and the susceptibility of immunocompetent individuals after inhalation is not clear. Attempts to create a vaccine for this infection have been made, which would benefit military personnel, the group targeted most likely by intentional release, as well as individuals that live in areas where melioidosis is endemic (Bondi and Goldberg, 2008).

1.3.10 Glycoconjugate vaccines

Reports exist dating back to 200BC in China and India of people purposely inoculating themselves with infections in order to protect themselves from other diseases, but the modern era of human vaccination is ascribed to Edward Jenner's discovery in 1796 using cowpox to protect against smallpox (Guo and Boons, 2009). Louis Pasteur acknowledged Jenner's pioneering research by retaining the word vaccination to describe his own efforts in the prevention of rabies and anthrax in his immunological research. Vaccines have since become a vastly important and successful countermeasure to the threat of infectious disease, providing protection by inducing humoral and/or cellular immunity to disease-causing pathogens (Astronomo and Burton, 2010).

The field of carbohydrate vaccines can be traced back to the 1920's and 1930's. Heidelberger and Avery made the connection between *Pneumococcal* serotypes and a "soluble specific substance", the capsular polysaccharides (Heidelberger and Avery, 1923). Francis and Tillet further noted that patients immunised intradermally with serotype-specific polysaccharides, would elicit antibodies against heterologous types of pneumococcal species (Tillett and Francis, 1929). In 1931 Avery and Goebel demonstrated that non-immunogenic carbohydrates could become antigenic when covalently attached to proteins. They created the first conjugate of the *Streptococcus pneumoniae* type 3 polysaccharide by linking it to horse serum globulin, which induced specific antibodies in rabbits that were previously unresponsive to the pure polysaccharide (Guo and Boons, 2009, Avery and Goebel, 1931). Heidelberger further established that vaccination with *Pneumococcal* capsular polysaccharide could be used to elicit persistent antibody-mediated immunity (Heidelberger et al., 1950). Due to the advent in chemotherapeutics and antibiotics, research into carbohydrate vaccines remained idle, but renewed interest was shown upon the steady increase in antibiotic resistance by many pathogenic bacteria.

Half a century after initial discoveries, Jennings used bacterial capsular polysaccharides coupled to a carrier protein to render it immunogenic and produced a commercial vaccine (Jennings, 1983). PneumoVax® was launched by Merck as the first polysaccharide vaccine, composed of unconjugated capsular polysaccharides isolated from 14 pneumonia serotypes (now upgraded to 23 out of approximately 90 serotypes in PneumoVax 23®). This vaccine induces protection against approximately 90% of infections caused by these pathogens, but high risk groups, including neonates, children under 2 years of age, elderly and immunocompromised population, generally elicit a poor antibody response and do not induce adequate protection (Merck & Co., 2012).

Glycoconjugate vaccine creation requires the full structural characterisation of one or multiple antigens of a specific pathogen, its synthesis or isolation as well as its conjugation via a linker to an appropriate immunogenic carrier protein or other immunostimulant, all the while maintaining the immunological integrity of the antigen. The conjugate vaccine has to clear immunological studies in animal models, where host response is further studied and assessed on antibodies elicited, target interaction and efficacy, before entering clinical trial evaluation, prior to widespread use (Guo and Boons, 2009).

1.3.10.1 Carbohydrate-based antigens

Glycans contribute to the immune system, through their recognition as carbohydrate antigens, as is exemplified by the ABO(H) blood group system. The common epitope is the H antigen, which is found in individuals with the O blood type. This can be elaborated with a GalNAc or a Galactose residue to produce the structures that specify the A and B blood types respectively. Individuals will develop antibodies to the antigen epitope (A or B) not synthesised by their own body, because similar glycans are found in various food substances (Taylor and Drickamer, 2011, Guo and Boons, 2009). Differences in carbohydrate structures found on the cells of humans and other animal species can also cause an antibody response. For example, where most vertebrates contain a gene for a α-galactosyltransferase, capping various glycan structures by forming a Galα1-3Gal linkage, this gene, albeit present in humans, is not expressed. Again exposure to similar epitopes on food substances will induce antibody production against the α-galactosyl residue, also posing as an obstacle for xenotransplantation of organs, as α-gal specific natural antibodies bind to the endothelium of vascularised xenografts and the complement system is activated, which leads to the activation of the coagulation cascade and rapid graft rejection (Guo and Boons, 2009, Taylor and Drickamer, 2011).

In nature animals and humans mount massive humoral responses against the polysaccharides of Gram-negative bacteria, in part due to the abundant range of possibly varying glycan epitopes decorating the lipid A and the presence of unusual monosaccharide components (Guo and Boons, 2009). Bacterial glycosylation therefore presents a relatively easy target for vaccine development, due to the characteristic differences between these microbial and human glycans. However, the difficulties in creating effective vaccines composed of non-toxic, immunogenic polysaccharides lie with just this chemical diversity found among the serologic variability displayed by O-antigens, where a broad acting O-antigen targeted vaccine must effectively consist of a highly complex mixture of well-defined characterised glycans (Guo and Boons, 2009), presenting the main aim of the work described in **Chapter** 4.

1.4 MASS SPECTROMETRY IN GLYCOBIOLOGY

1.4.1 Overview

The analytical technique of mass spectrometry dates back to the early 20th century, when Sir Joseph John Thomson, was the first person to demonstrate the mass to charge ratio of the electron in 1906 using his “positive ray parabola” and becoming the first person to obtain a mass spectrum. The concept was first put into practice by his student Francis W. Aston, working in the Cavendish Laboratory in Cambridge in 1919,

who designed an instrument to measure the mass of elements, especially isotopes. In the past four decades tremendous advances and developments in mass spectrometry have led to its status as one of the most powerful analytical tools in structural glycobiology. Mass spectrometry is the current method of choice for the structural analysis of glycoconjugates, due to its ultra-high sensitivity and ability to characterise individual components within complex, heterogeneous mixtures (Morris et al., 1978) and its accurate molecular weight determination of minute quantities, sometimes requiring less than femtomole (10^{-15} mole) amounts of material.

Mass spectrometry involves the production of gas-phase ions from analytes, where the resulting ions are accelerated out of the ionisation source into a mass analyser, where they are separated according to their mass to charge ratio (m/z) and detected. The three main components of a mass spectrometer are the ionisation source, one or more mass analysers and the ion detector. The separated ions reach the detector where signals are recorded generating a mass spectrum consisting of the m/z rations and their relative abundances.

Mass spectrometry has been used for many years to define carbohydrate structures found on prokaryotic glycoconjugates. More recently, knowledge of genome sequences has led to the development of new strategies for the analysis of glycoproteins that exploit genomic information (Hitchen and Dell, 2006) and proteomics has become an invaluable high-throughput technique for functional genomics. The techniques of electrospray ionisation (ESI) and matrix-assisted laser desorption/ionisation (MALDI) mass spectrometry have become the routine methods of choice for prokaryote glycoconjugate analysis. The key to both techniques is the ability to analyse large biological molecules without degradation.

1.4.2 Ionisation techniques

There are various types of ionisation modes that are available and used depending on the type of compound analysed and the specific information required. In mass spectrometry, different mass analysers can be used in combination with a range of different ionisation techniques.

Structural analysis of glycans using mass spectrometry dates back to the 1960's, when analysis of low molecular weight volatile molecules was carried out using electron impact (EI) and chemical ionisation (CI) techniques. EI is a "hard" ionisation technique also known as electron bombardment, used to identify small molecules up to 1000Da by extensive energy input and creation of numerous fragment ions, making this technique unsuitable for analysis of molecular ions (Ando et al., 1977). CI on the other hand is a "soft" ionisation technique, with a similar ion source to EI. However, the ionisation is less energetic upon interaction of the sample with the reagent gas and the ions produced are molecular ions with little fragmentation. Both ionisation techniques are limited by the requirement for the sample to be in the gaseous phase before ionisation. On the other hand, if samples are sufficiently volatile, for example if polar groups are derivatised, then these methods are very powerful for structure analysis.

Field desorption (FD) ionisation was developed in the late 1960's (Beckey, 1969), which allowed the study of non-volatile compounds of masses up to 5000Da, yet the technique proved difficult to use and was only employed in a handful of laboratories (Linscheid et al., 1981). In 1976 plasma-desorption mass spectrometry

(PD-MS) was introduced by MacFarlane and Torgerson (Macfarlane and Torgerson, 1976), which allowed the detection of biomolecules of up to 20'000Da and was used for the mass determination of proteins. However, PD-MS relied on fission products of radioactive Californium for ionisation/desorption and was accompanied by poor resolution and peak broadening, due to the large kinetic energy distribution carried by the molecular ions and was therefore not used widely. It was not until the introduction of the high field magnet by Howard Morris and colleagues in 1977 (Morris et al., 1981a) and fast atom bombardment mass spectrometry (FAB-MS) by Barber and co-workers in the 1980's (Barber et al., 1981), that routine structural characterisation of large involatile glycoconjugates was feasible (Dell and Ballou, 1983, Egge et al., 1983, Morris et al., 1981a, Morris et al., 1981b). FAB-MS was widely used for glycan analysis for over twenty years, but has since been superseded by the more versatile technologies of MALDI (Karas and Hillenkamp, 1988) and ESI (Fenn et al., 1989).

In the following sections the three ionisation techniques (EI, ESI, MALDI) employed in the two projects reported this thesis are described (**Section 1.4.2.1**, **Section 1.4.2.3** and **Section 1.4.2.4**, respectively). FAB-MS is also briefly introduced, because glycan fragmentation pathways seen in current MALDI and ESI experiments were originally defined by FAB-MS (**Section 1.4.2.2.**) Also many sample handling methodologies optimised for FAB-MS are still employed in conjunction with other ionisation techniques today.

1.4.2.1 Electron impact mass spectrometry (EI-MS)

Electron impact has enjoyed considerable success in the analysis of small carbohydrates, particularly monosaccharides and is still valuable as the ionisation method in combined gas chromatography/mass spectrometry systems (GC-MS) for the determination of composition and linkage. Linkage analysis is a technique whereby polysaccharides are permethylated, hydrolysed and then further acylated for structural investigation (Hellerqvist, 1990, Lindberg and Lonngren, 1978). In GC-EI-MS the gaseous sample is introduced into the source from an online gas chromatograph where it is bombarded with electrons from Rhenium or Tungsten filaments with energy of 70eV, well above the covalent bond strength of molecules, to form highly energetic radical molecular cations and imparting enough internal energy to fragment the molecular ion into stable fragments by the ejection of an unpaired electron.

Quantitative monosaccharide analysis provides estimated molar ratios of individual sugars and may suggest the presence of specific oligosaccharide classes. This involves cleavage of all glycosidic linkages, fractionation of the resulting monosaccharides, detection and quantification. The monosaccharides are first chemically modified at their hydroxyl and aldehyde groups. Reduction of the aldehyde of a free monosaccharide followed by acetylation of its hydroxyl groups provides a "peracylated alditol acetate". These monosaccharides may be analysed and compared with authentic standards. The hydroxyl groups of free monosaccharides generated by glycan hydrolysis can also be converted to trimethylsilyl ethers (TMS). These per-O-trimethylsilyl derivatives are widely used for monosaccharide compositional analysis by GC-MS.

Linkage analysis was previously also known as “methylation analysis” and is a well-established approach for determining linkage positions. The principle of this method is to introduce a stable substituent onto each free hydroxyl group of the native glycan, which in this case is an ether-linked methyl group (**Figure 1.13**). The glycosidic linkages which are much more labile than the ether-linked methyl groups are then cleaved by acid hydrolysis, producing individual methylated monosaccharides with free hydroxyl groups at the positions that were previously involved in a linkage. The partially methylated monosaccharides are further derivatised to produce volatile molecules amenable to GC-MS analysis (Dell and Morris, 2001). The most common strategy involves reduction of the monosaccharides to produce alcohols at C-1, by eliminating the formation of ring structures, followed by derivatisation of the free hydroxyl groups, usually by acetylation. The individual components of the mixture of partially methylated alditol acetates can then be identified by a combination of GC retention time and EI-MS fragmentation pattern. The limitation of this technique is that it does not provide information on the oligosaccharide sequence or anomericity of the monosaccharide linkages (Varki et al., 2009).

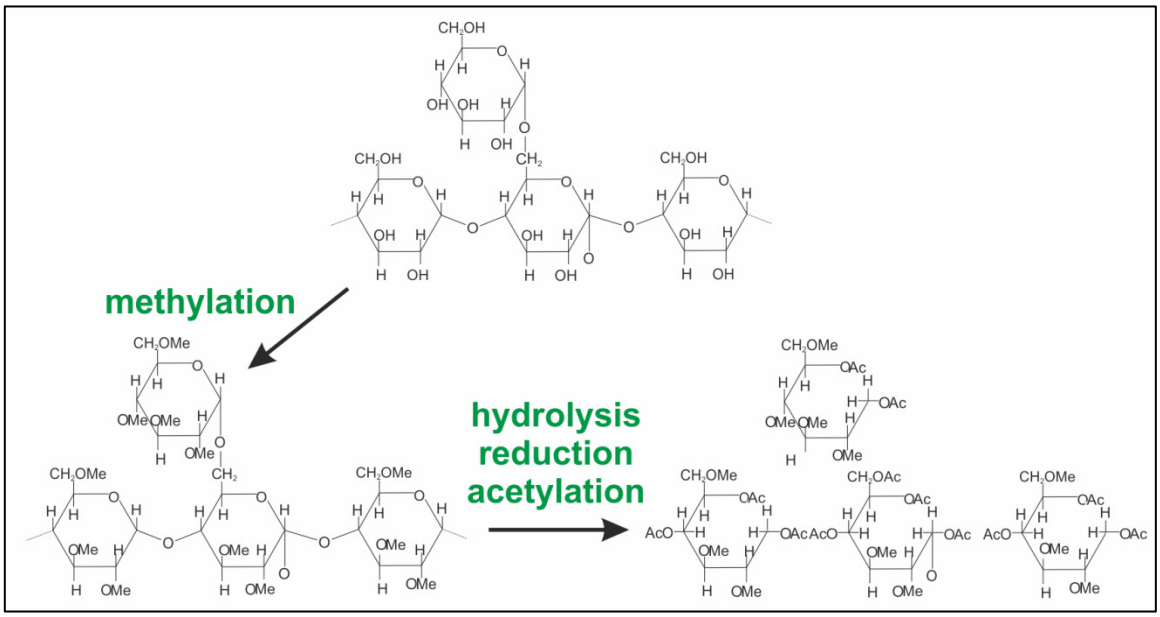


Figure 1.13 – Example of linkage analysis.
The schematic shows the procedure of linkage analysis of a α1,4-linked glucose chain with an α1,6-linked glucose branch. The terminal α1,6 glucose gives rise to glucitol, which is acetylated at C-1 and C-5, whilst the two α1,4 glucose units on either side of the branching monosaccharide are acetylated at C-1, C-4 and C-5. The branching glucose is acetylated at C-1, C-4, C-5 and C-6. The schematic was adapted from Varki and colleagues (Varki et al., 2009)

1.4.2.2 Fast atom bombardment mass spectrometry (FAB-MS)

Fast atom bombardment (FAB) was introduced in 1981 by Barber and colleagues (Barber et al., 1981) and was the major ionisation method for oligosaccharides for the next decade opening the field to most areas of biomedical research. Although it is still used to monitor products of chemical synthesis, it is now regarded as mainly of historical interest with the advent of ESI and MALDI. This method was one of the first techniques that allowed large intact carbohydrates to be ionised efficiently (Dell, 1987, Dell and Morris, 2001, Morris et al., 1981b).

With FAB the sample is dissolved in a suitable solvent and an aliquot is added to a drop of a relatively non-volatile liquid matrix on the target. The most common FAB matrices are glycerol, thioglycerol or 3-nitrobenzyl

alcohol. The probe is then inserted into the mass spectrometer. The matrix is then bombarded with a beam of high kinetic energy atoms, such as inert gases xenon or argon or ions, such as Caesium (Cs^+). The bombarding atoms or ions destroy the surface monolayer and energy is transferred to underlying layers resulting in ionisation of the sample molecules, which are "sputtered" from the surface. This allows desorption and ionisation either through protonation/deprotonation through the coupling with cations or anions in a positive or negative mode and entrance into the gas phase. Gas-phase ions are therefore generated without prior volatisation of the sample, allowing the analysis of polar, involatile and thermally labile compounds (**Figure 1.14**). Both positive and negative ions are produced during the sputtering process, and either can be recorded by the appropriate choice of instrumental parameters. Molecules are ionised by the addition of a proton or cation such as sodium, potassium, ammonium (positive ion formation) or by the loss of a proton or addition of an anion such as chloride (negative ion formation). During ionisation some internal energy is imparted to the molecule resulting in fragmentation of labile bonds and these molecular and fragment ions are accelerated and detected. FAB-MS can yield information on the degree of heterogeneity and the type and sites of glycosylation and branching patterns of N- and O-linked glycans and was primarily used to identify polar molecules such as peptides and oligosaccharides up to 6000Da using a high field magnet double focusing ZAB instrument developed in this laboratory (Dell et al., 1993).

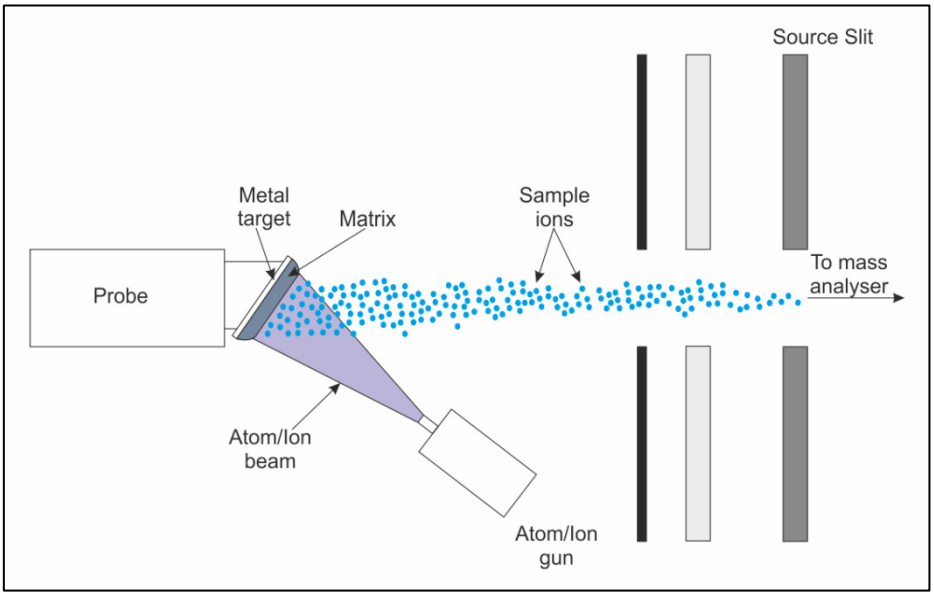


Figure 1.14 – Fast atom bombardment ionisation mechanism.
The schematic representation shows the fast atom bombardment mechanism. The diagram was adapted from Dell and colleagues (Dell et al., 1993)

1.4.2.3 Electrospray ionisation (ESI-MS)

Electrospray ionisation is one of the so-called atmospheric pressure ionisation (API) techniques and was first described by Malcolm Dole and collaborators in the late 1960's, but came into prominence in the late 1980's when John Fenn and colleagues demonstrated the use of electrospray for the ionisation of high mass biopolymers, for which he was honoured with a share in the 2002 Nobel Prize in Chemistry (Fenn et al., 1989). ESI has now become one of the most widely used ionisation techniques for proteomic and glycoproteomic analyses.

ESI-MS is a method by which a stream of liquid containing the sample of interest is directly injected into the atmospheric pressure ion source of a mass spectrometer. For pure samples dissolved in the mobile phase, a direct injection “loop” can be used, especially when interfaced to chromatographic procedures such as capillary zone electrophoresis (CZE), high performance liquid chromatography (HPLC) and in particular “nanospray” ESI. The sample is dissolved in a mobile phase, pumped through a fine stainless steel capillary and is introduced into the source in solution through a narrow golden-tipped glass capillary. The potential difference, of typically 3-6kV, is applied between the capillary and counter electrode situated at the entrance cone. A spray of microdroplets is therefore generated with the assistance of nitrogen. These then traverse a series of “skimmers”, imposing a charge accumulation on the liquid surface and form highly charged droplets. Upon evaporation, the surface charge density is increased whilst droplet size decreases, resulting in the Raleigh stability limit being reached, overcoming the surface tension to cause breakdown into even smaller droplets and the creation of a charged molecular species, devoid of solvent, ready for analysis either by magnetic sector or more commonly quadrupole analysers (**Figure 1.15**). Large molecules such as proteins and glycoproteins tend to carry multiple charges and since the mass spectrometer is really an instrument which measures mass to charge ratio (m/z) rather than mass, even very large molecules exceeding 100kDa can be amenable to ESI-MS analysis. ESI-MS is particularly valuable as a method to review whether a protein is post-translationally modified by examination of any mass difference between the observed signal and that calculated as the sum of the amino acids present in the sequence. There is a distribution of signals carrying varying numbers of net positive (or negative if working in the negative ion mode) charges on basic (or acidic) sites in the molecule (Dell et al., 1993).

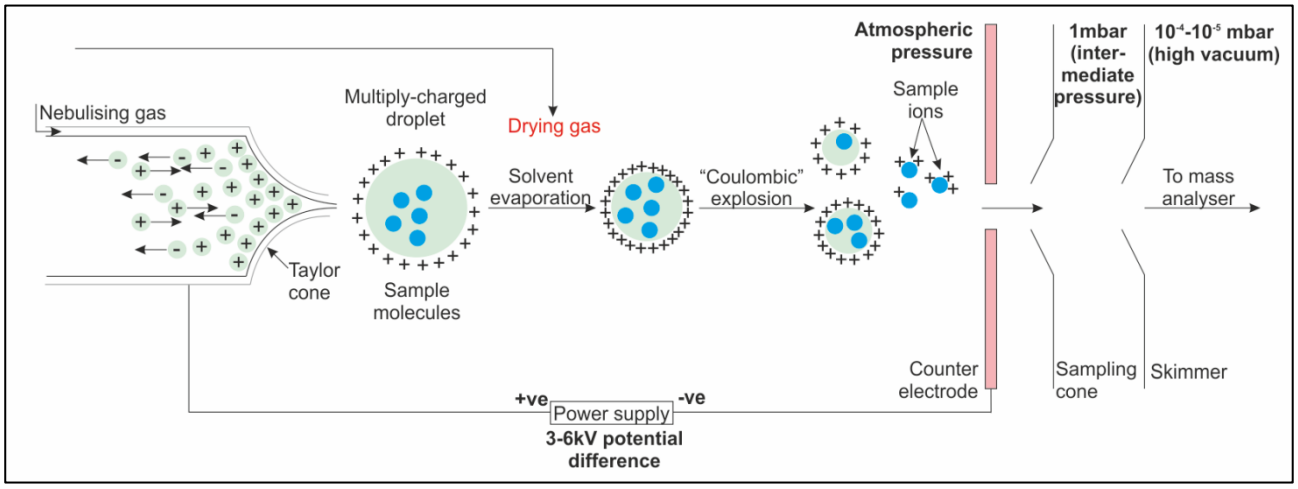


Figure 1.15 – Electrospray ionisation mechanism.
The schematic representation shows the electrospray ionisation mechniams in the positive ion mode. The diagram was adapted from Dell and colleagues (Dell et al., 1993)

The compatibility of electrospray with liquid separation techniques has resulted in LC-MS becoming extensively used for protein and carbohydrate analysis and micro-separation methods suitable for glycan analysis have been developed in many laboratories (Wuhrer et al., 2005, Mechref and Novotny, 2006, Novotny and Mechref, 2005, Greer and Morris, 1997, Morris et al., 1996, Dell et al., 2000). Normal-phase separations provide good correlation between structure, molecular weight and retention time (Royle et al., 2002), allowing some structural information to be obtained directly from the elution profile, although the buffers often result in doubly charged ions in the mass spectra. Wuhrer and colleagues have described a nanoscale LC-MS technique using normal-phase (NP) LC-MS that is capable of producing fragmentation

data from small N-linked glycans at the low-femtomole level (Wuhrer et al., 2004). Early ESI sources ran at flow rates of a few µl/min, but newer nanospray sources operate at flow rates of 10-30nl/min increasing sensitivity and the possibility for online nanoLC for complex mixtures. Furthermore the Z-spray nanospray source of some commercial instruments moves the ions in a Z-formation, allowing the removal of neutral elements prior to sample focusing and resulting in the enhancement of sensitivity.

1.4.2.4 Matrix assisted desorption ionisation mass spectrometry (MALDI-MS)

Ionisation based on laser desorption was introduced in the early 1960's and later developed and adapted for the analysis of biomolecules, however its limitations in poor ionisation and high molecular ion fragmentation were only overcome in the 1980's with supplementary use of a matrix, leading to the matrix-assisted laser desorption ionisation (MALDI) technique. Koichi Tanaka was later honoured with part of the 2002 Nobel Prize in Chemistry for showing that proteins can be ionised by MALDI. The use of MALDI to ionise carbohydrates was first reported by Karas and Hillenkamp (Karas and Hillenkamp, 1988) in one of their early papers on the technique and applied to N-linked glycans by Mock and colleagues in 1991 (Mock et al., 1991). It has since become one of the most popular techniques for glycan analysis.

Unlike FAB-MS this process uses a crystalline, rather than liquid matrix and a pulsed beam of photons, rather than a continuous beam of atoms or ions. The sample is embedded in a low molecular weight UV absorbing "crystalline" matrix. The matrix is chosen to have an absorption maximum near the wavelength of the pulsed laser that is used to ionise the sample. Nitrogen lasers with a wavelength of 337nm at 10Hz are widely used, as also are solid state lasers which can give similar wavelengths upon frequency tripling. The matrix absorbs the pulsed laser energy and enough energy is transferred to the sample to ionise it at room temperature, resulting in a dramatic increase in both sensitivity and mass range of analysable compounds. Irradiation of the crystals by the laser causes the accumulation of a large amount of excitation energy in the matrix molecules, resulting in the desorption of the sample and matrix ion formation from the surface of the crystal, by proton or cation transfer between the photo-excited matrix and analyte (**Figure 1.16**) (Knochenmuss, 2006). MALDI is a "softer" ionisation technique than FAB and produces mainly singly charged molecular ions of the form $[M+H]^+$ or $[M+Na]^+$, with little fragmentation, making this an ideal technique for "fingerprinting" or mass profiling of complex glycan mixtures. However, fragmentation can occur or be induced with higher laser intensity or acceleration voltage, either by in-source decay (ISD) or post-source decay (PSD) (Dell and Morris, 2001, Harvey, 2006).

Early MALDI experiments were characterised by poorly resolved spectra where ions of similar m/z ratio could not be separated, due to the kinetic energy distribution of ions. This meant that ions with the same m/z value but different kinetic energies reached the detector at slightly different times, resulting in peak broadening. This was improved by "delayed extraction", where the pulse of ions is kept in the source for a short time after the laser pulse in order to normalise their energies by collision processes (Hoffmann et al., 2007). After the delay, an extraction pulse applied to the now monoenergetic ions leaving the source with the same m/z value thus improves resolution.

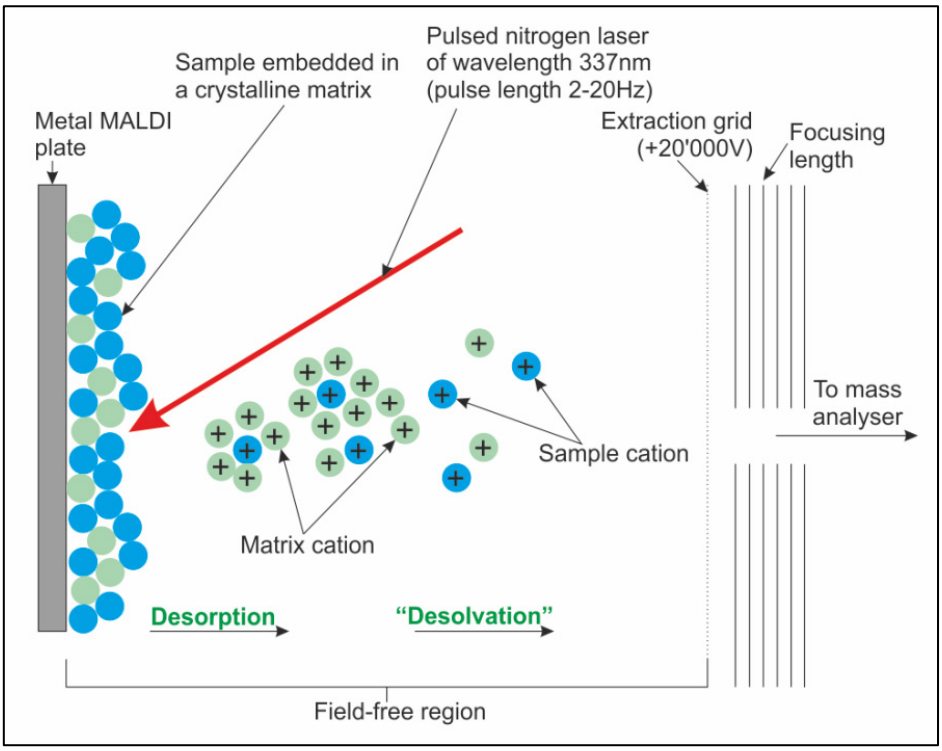


Figure 1.16 – Matrix assisted laser desorption ionisation mechanism.
The schematic representation shows the matrix assisted laser desorption ionisation mechanism in the positive ion mode.

The choice of matrix is crucial in imparting efficient ionisation of a given sample and many matrices have been developed, but one of the earliest, 2,5-dihydroxybenzoic acid (DHB), is still the most popular for carbohydrate and glycoprotein analysis (Harvey, 2006, Mock et al., 1991, Harvey, 1999, Mohr et al., 1995). Furthermore α-cyano-4-hydroxy cinnamic acid (CHCA) is used for peptides and glycopeptides, whereas sinapinic acid (SA) is the matrix of choice for higher molecular weight proteins or glycoproteins. Analysis with diaminobenzophenone (DABP) and 2-(4'-hydroxylbenzeneazo) benzoic acid (HABA) are important for analysis of glycolipids.

1.4.3 Mass Analysers

A number of mass analysers are commonly used in biopolymer mass spectrometry including double focusing instruments, quadrupole mass analysers, time-of-flight (TOF), quadrupole ion traps and Fourier transform ion cyclotron resonance (FT-ICR). Field instruments utilise the behaviour of charged particles moving through field regions, whilst sector instruments incorporate an electromagnetic field and an electric field for energy focusing. Quadrupole and ion trap instruments incorporate a combination of radiofrequency and direct-current fields. Ions entering a field experience a deflecting force, depending on the strength of the field and the m/z ratio of the molecular ion. By scanning the field strength all the ions produced in the same ion source are sequentially focused at the detector, which is usually a photomultiplier, an electron multiplier or a multichannel plate array, creating a mass spectrum. In time-of-flight instruments the time taken for an ion to travel a set distance is measured very accurately, as it is directly related to the m/z value of the ion.

Double focusing analysers consist of magnetic sectors, that separate the ions of different m/z values and electrostatic analysers that focus ions according to their kinetic energy, allowing high resolution mass

spectrometric analysis (Dell et al., 1993). The introduction of MALDI and ESI ionisation techniques stimulated the development of new analysers and double-focusing analysers have now been largely superseded by more versatile and sensitive instrumentation. EI sources are most commonly found on quadrupole instruments or on magnetic sector instruments, whilst FAB sources are most commonly found with magnetic sector instruments. MALDI and ESI sources are most commonly found with TOF and triple quadrupole, ion trap or Q-TOF instruments respectively.

1.4.3.1 Ion traps and ion mobility

Ion traps allow the “trapping” of a small mass range of m/z values using oscillating electric fields generated by radiofrequency (RF) voltages applied to electrodes of opposite polarity. Scanning the RF or magnetic field allows ions to be detected sequentially. For ion detection, the voltages are altered to destabilise ion motions, resulting in the ejection of the ions from the storage cell to the detector. There are two types of ion trap analysers, the 3D quadrupole ion trap (QIT) or the 2D linear ion trap (LIT).

Fourier transform ion cyclotron resonance (FT-ICR) allows for the determination of multiple ions at once with the highest resolution mass analyser and most powerful ion trap (Marshall et al., 1998). FT-ICR ions are subjected to a simultaneous RF electric field perpendicular to the magnetic field of a static, superconducting high field magnet, causing the ions to follow spiral paths. The m/z values are measured based on their cyclotron frequencies, which is the frequency of an ion moving perpendicular to the direction of a uniform magnetic field, where the frequency of ion rotation is inversely proportional to the m/z ratio. The instruments are however, hugely expensive and the maintenance of the high magnetic field is difficult, which resulted in their disuse. The high resolution and mass accuracy is currently being exploited with the recently developed electron capture dissociation (ECD) soft fragmentation technique, which mainly induces fragmentation of the peptide backbone, retaining the more labile modifications such as glycosylation and thus defining site attachment (Hakansson et al., 2001). The electrostatic trap, “Orbitrap”, is a high-resolution mass analyser, which was initially proposed by Makarov and later commercially produced, which electrostatically maintains the ions in orbit around a central electrode similar to the FT-ICR mass spectrometer, but does not require an expensive superconducting magnet, still providing high resolution and mass accuracy of data. In combination with a linear ion trap, it has since proved a powerful tool for glycomics and proteomics.

1.4.3.2 Quadrupole Analysers

The quadrupole mass filter was invented in the mid 1950’s but not used until the 1970’s. This analyser uses a quadrupolar electrical field comprising RF and DC components to separate the ions. It consists of four parallel rods with circular or hyperbolic cross-sections. An electrical field is created by the application of a fixed direct current component and an oscillating radiofrequency perpendicular to each other, producing opposite current polarity in adjacent rods. Changing the voltage on the field, allows only certain ions with the correct m/z value pass through the filters without obstruction towards the detector, unlike those with different m/z values which will collide with the rods and are lost (**Figure 1.17**). Thus in the quadrupole mass analyser, ion separation is based on the stability of the ion trajectory through the four parallel rods.

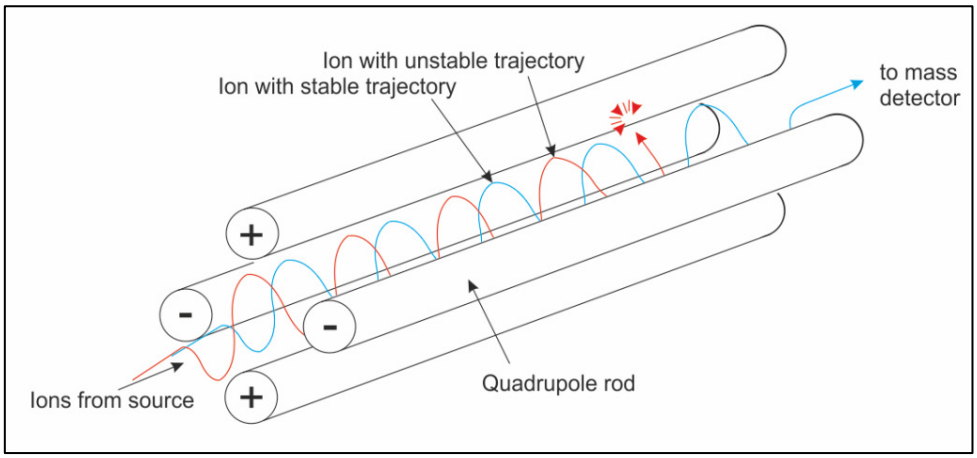


Figure 1.17 – Ion path through a quadrupole analyser.
The schematic shows a quadrupole analyser where the red ion possesses the selected m/z value, thus passing out of the quadrupole into the mass detector, whereas the path of the blue ion is intercepted by a quadrupole rod, which results in the ion being discharged before reaching the mass detector and thus not being detected.

Quadrupole mass filters can detect m/z values of up to about 4000, are very cost effective and robust. They have the advantage of rapid scanning and operate at relatively high pressures and can therefore be readily interfaced with a wide variety of inlet systems, which is useful in combination with gas chromatography and thus ideal for GC-MS and ESI-MS. However, its drawbacks include the relatively low resolution and limited mass range, yet due to its efficient selectivity of molecules at specific m/z values; quadrupole mass analysers are considered mass filters and often used as the first analyser in tandem mass spectrometry.

1.4.3.3 Time-of-flight

In time-of-flight analysers, the ions are separated by differences in velocities as they move in a field-free, "flight tube" towards the detector, which allows for analysis of an "unlimited" mass range. The separation principle of TOF analysers is based on the fact that ions of different m/z values will traverse the field-free region at different velocities and hence reach the detector at different times. To determine the flight time, all ions, regardless of their m/z values must enter the flight tube simultaneously, which is achieved in a pulsed process. However, due to the spread of kinetic energies originally imparted to the ions during the ionisation process and time taken for ion formation, ions of the same m/z value will have a spread of velocities that contribute to a spread in the signal detected, resulting in reduced resolution. To improve the resolution two developments have been introduced to compensate for such variation, such as the previously discussed delayed extraction and a reflectron.

A reflectron, or ion mirror, is a reflecting electric field consisting of a series of rings/grids with voltages that increase up to a value slightly greater than the voltage of the ion source, so that the more energetic ions penetrate further into the field to such an extent that their increased path length just compensates for their increased velocity, resulting in a reduction of the spread of initial kinetic energies of the ions produced and an improvement in resolution (**Figure 1.18**). High mass, slow moving ions cannot penetrate the reflectron and so high resolution reflectron mode comes at the expense of sensitivity, due to the physical constraints applied by the reflectron and limits the mass range to about 5'000Da as well as causing fragmentation of molecules accompanied by a decrease in detection levels. This means that reflectrons are used for the analysis of low to medium molecular weight molecules, but deactivated in favour of the sensitivity of linear

analysis to detect larger ions. The pulsed nature and delayed extraction mechanisms employed in MALDI-MS make this technique ideally suited to being coupled to a TOF mass analyser. MALDI-TOF mass spectrometers with delayed extraction and reflectron as well as linear operation modes were used for much of the data collection in the projects discussed in this thesis.

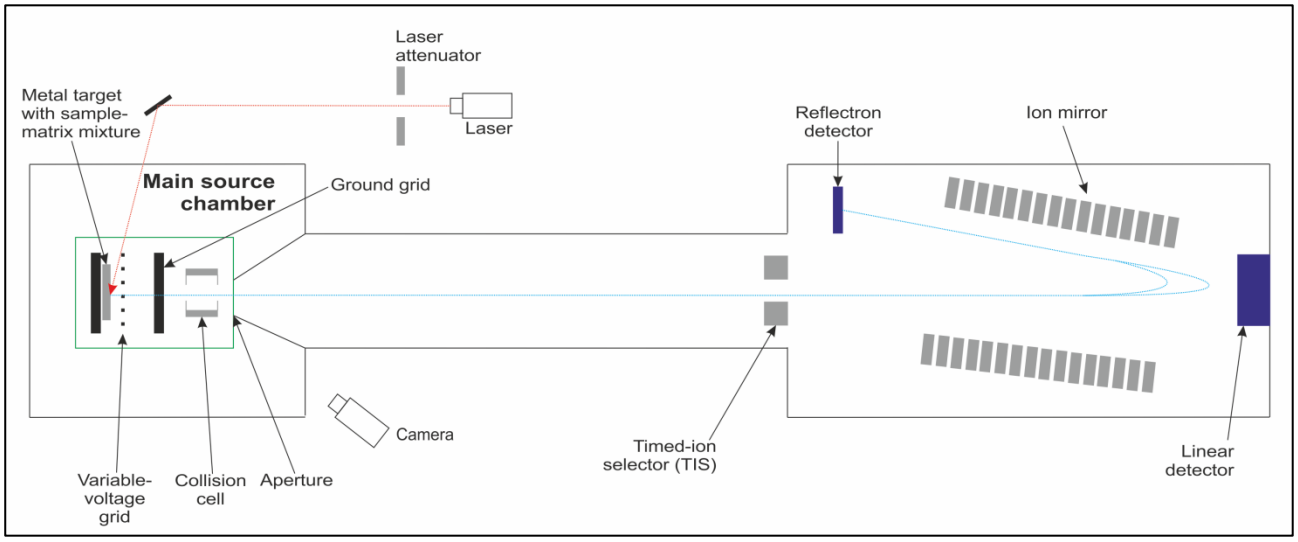


Figure 1.18 – Base components of a TOF analyser.
The above schematic depicts a Voyager-DE™ STR System with a time-of-flight analyser.

1.4.4 Tandem mass spectrometry

In mass spectrometry the molecular ion annotations can yield information on the theoretical composition in monosaccharides, amino acids or lipids and confirm the presence or absence of modifying groups such as sulphate or phosphate. However fragmentation of the molecular ion allows for rigorous sequencing, including branch determination and site occupation in glycoproteomics. Both MALDI-MS and ESI-MS are soft ionisation techniques, yielding very few fragment ions and are thus ideal for mass profiling of glycan and glycopeptide mixtures, however for further sequence analysis, molecular ion fragmentation is central and is induced by employing two mass analysers in tandem.

Many modern mass spectrometers contain two or more mass analysers in tandem, the term MS/MS referring to such coupling when used for induced fragmentation studies. These can consist of the same or different types of analysers. Most MS/MS experiments are performed on instruments equipped with triple quadrupole, hybrid quadrupole with time-of-flight (Q-TOF) (Morris et al., 1996) and more recently time-of-flight with time-of-flight (TOF-TOF) analysers (Vestal and Campbell, 2005).

In MS/MS experiments, the first mass analyser performs the mass selection of a precursor ion of interest from the mixture of ions produced at the source. The selected ion is then passed into a pressurised collision cell containing an inert gas such as argon or xenon and undergoes collision-induced dissociation (CID), also known as collisionally activated dissociation (CAD). This collision promotes fragmentation of the precursor ion and these fragments are subsequently re-accelerated towards the second mass analyser, which is used to separate the resulting fragment ions and finally detected. These fragment ions present a unique fingerprint pattern of the selected precursor ion. Tandem mass spectrometers have now become an indispensable tool

affording detailed structural characterisation of individual components as well as protein or glycan sequencing.

The data acquired in this thesis was obtained using an API ESI quadrupole orthogonal acceleration time of flight (Q-STAR Pulsar) mass spectrometer, a Waters Xevo G2S Q-TOF mass spectrometer, an Applied Biosystems MALDI-TOF-TOF mass spectrometer as well as a Voyager-DE STR System MALDI-TOF™ mass spectrometer.

1.4.4.1 Quadrupole orthogonal acceleration time of flight (Q-TOF)

One of the instruments used in the work for this thesis is the API Q-STAR™ Hybrid LC-MS/MS system, a high performance hybrid quadrupole orthogonal acceleration time of flight (Q-TOF) mass spectrometer coupled to an ESI source, which is linked to a nanoLC column, enabling high sensitivity MS/MS experiments in an online fashion. This instrument consists of a first quadrupole used as a mass filter, coupled with a hexapole collision cell enabling low energy fragmentation, together with an orthogonal time-of-flight analyser. In MS mode the quadrupole analyser and collision cell act only as ion guides whilst the TOF acts as the only mass analyser, however with MS/MS the ion filter capability of the first quadrupole analyser is used to select the precursor ions of interest, which are then accelerated into the hexapole collision cell, where they undergo fragmentation via collision with nitrogen or argon gas, with the TOF used to mass analyse the resulting fragment ions before they reach the detector (**Figure 1.19**).

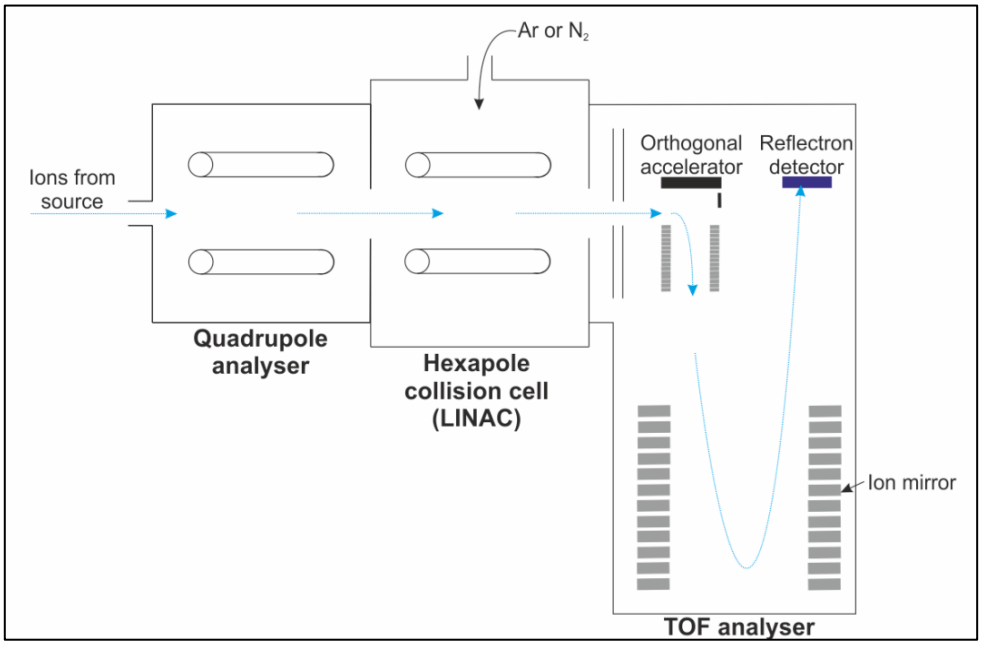


Figure 1.19 – Schematic of a Q-TOF analyser.
The diagram depicts a quadrupole analyser, hexapole collision cell and orthogonal acceleration time-of-flight (TOF), as are found in the API QSTAR™ Pulsar Hybrid LC-MS/MS system.

The Q-STAR is a “second generation” orthogonal acceleration instrument with improved scanning rate, ion optics, transmission and sensitivity. Furthermore it possesses a novel linear accelerating high pressure collision cell (LINAC) that enables the pressure of the collision cell to be raised to an extent where ion fragmentation and collection can be obtained with almost 100% efficiency.

1.4.4.2 Tandem time-of-flight

Most MS/MS and ion trap fragmentations have been limited by multiple low energy CID collisions. High energy CID on the other hand is sufficient to cause fragmentation as the result of a single collision, thus providing additional information-rich cleavage products. MALDI-TOF/TOF instruments are capable of producing CID fragments at both low and high energies, whilst preserving the resolution and sensitivity of the single TOF instrumentation (Vestal and Campbell, 2005), their only disadvantage being the relatively low parent ion resolution compared to a Q-TOF geometry. The MALDI-TOF/TOF instrument used in the production of data described in this thesis was an Applied Biosystems 4800 MALDI-TOF/TOF™ (**Figure 1.20**).

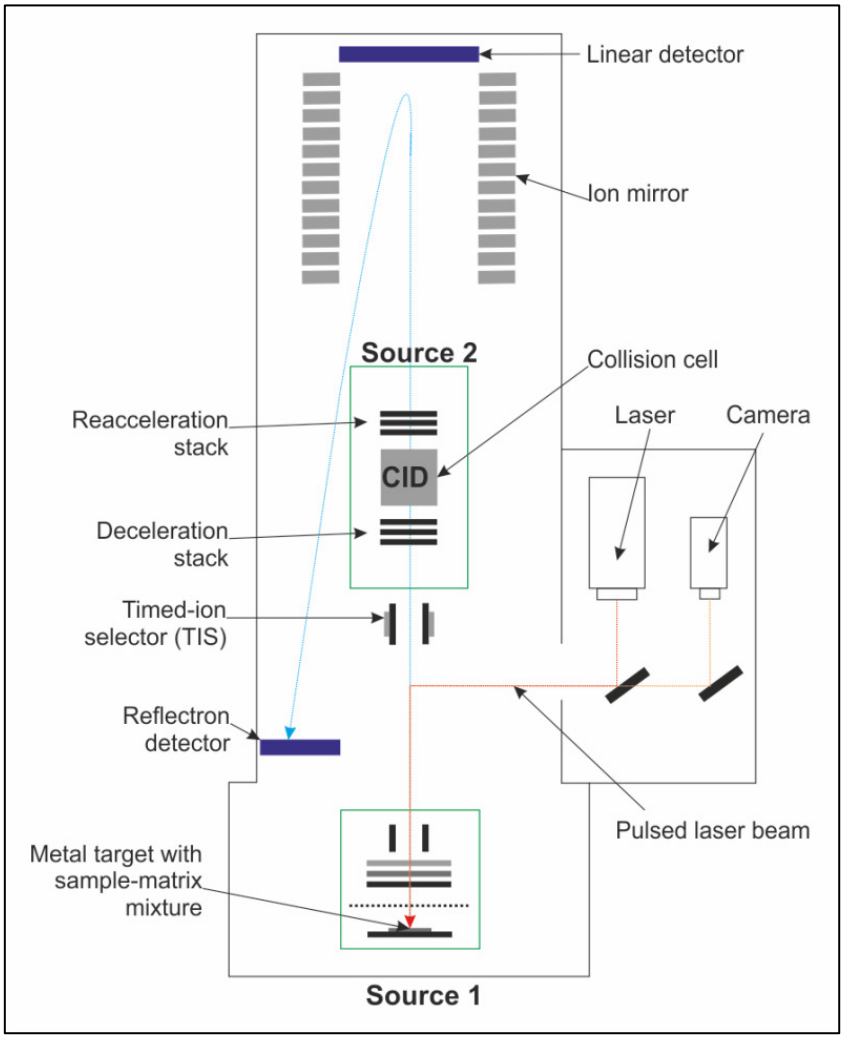


Figure 1.20 – Schematic of a MALDI-TOF/TOF analyser.
The 4800™ Analyzer from Applied biosystems is a double focusing instrument with linear and reflectron flight tubes, allowing for the selection of individual precursor ions for further fragmentation in the collision cell.

MALDI-TOF/TOF instruments are made up of a linear delayed extraction MALDI-TOF, a collision cell and a second TOF analyser. The ion source optics and electronics of the initially developed MALDI-TOF remain essentially unchanged but additional elements incorporated for the MS/MS mode include a time-ion selector (TIS). In the MS mode, ions are extracted and guided directly to the detector; however, in the MS/MS mode the molecular ions produced in the source pass through the first TOF and are then focused at the timed-ion-selector (TIS). This enables the selection of a precursor ion according to its velocity through the first TOF

analyser, termed the Bradbury-Neilson gate, whilst other ions are deflected away. A time delay generator is programmed to open the TIS at the lightest mass of interest reaches the gate, and closed when the highest mass of interest has passed through, resulting in the retardation of selected ions by the deceleration lens and the entrance into the collision cell at a designated collision energy defined by the potential energy between the source and the collision cell (i.e. 1kV, 2kV). Variation in the collision energy is achieved by the adjusting of the ion source potential to that of the collision cell, where the pressure of the inert gas, such as argon, is controlled to provide satisfactory fragmentation. Fragment ions produced in the collision cell therefore travel with the same velocity as the decelerated precursor ions, until they are re-accelerated by a high voltage pulse into the second TOF analyser, reaching the detector via the reflectron and providing high resolution MS/MS data of the selected m/z parent ion.

1.4.5 Characterisation of peptides, glycopeptides and carbohydrates

Fragment ions can be formed within the ion source of the mass spectrometer [in-source decay (ISD)], within the flight tube of the time-of-flight instruments [post-source decay (PSD)], or by collisional activation (CID) in a collision cell. ISD, which is common with FAB spectra, can be a problem because Y-type ions are isobaric with native glycans, thus distorting the glycan profile. PSD ions are sometimes difficult to observe, are often poorly resolved and generally require relatively large amounts of sample. CID-produced ions, on the other hand, suffer from none of these disadvantages.

1.4.5.1 Fragmentation of peptides

If an ion is created by cationisation (protonation or sodiation) the inherent stability of an even electron oxonium (O-protonated) or ammonium (N-protonated) ion allows for a “quasimolecular ion” to be formed that normally presents minimal natural fragmentation. Such creation of charge is the basis for further analysis and detection. In MALDI and ESI sources positively charged peptides are formed as so called $[M+H]^+$ or quasimolecular ions. Initially this is thought to occur at amine groups in amino acid side chains (Lys or Arg) or at the peptide amino-terminus itself ($R\text{-}NH_2 + H^+ \rightarrow R\text{-}NH_3^+$). The charge may then be passed around the molecule due to its flexibility, including the peptide bond where fragments of interest are generated. In negative mode, the functional groups that readily lose protons include carboxylic acids ($R\text{-}CO_2\text{-}H \rightarrow R\text{-}CO_2^- + H^+$) and alcohols ($R\text{-}OH \rightarrow R\text{-}O^- + H^+$) and these more susceptibly form negatively charged $[M-H]^-$ molecular ions.

As discussed previously, ESI and MALDI allow the examination of intact molecules to provide molecular weight information, yet this is of limited use in detailed structural studies where it was found that digestion strategies were needed to break the problem down into a series of solvable structural units, the peptides and glycopeptides. The digestion of proteins by various enzymes, will thus form peptide fragments, such as in the case of trypsin digestion, which cleaves after lysine or arginine amino acids and will create peptide fragments that possess a C-terminal NH_2, which are often observed as $[M+H]^+$ molecular ions in the positive ion mode. Further derivatisation methods for protein and peptide samples are discussed in **Chapter** 2.

Peptides fragment in a reasonably well-documented manner, and fragments may be classified into two categories of peptides that are derived from cleavage of bonds within the peptide backbone chain, to give amino-terminal (N-) and carboxy-terminal (C-) fragment ions (Morris et al., 1981b, Marino et al., 1988) and are annotated using the nomenclature described by Roepstorff, Fohlman and Biemann (Biemann, 1988, Roepstorff and Fohlman, 1984, Morris et al., 1981a).

Using appropriate low collision energy, three types of cleavages can occur within the peptide chain at three different types of bonds, such as NH-CH, CH-CO and CO-NH. In singly charged species, any bond cleavage gives rise to two species, one charged and one neutral, with only the charged species being able to be accelerated and detected by the mass spectrometer. The position of the charged moiety is dependent on the chemistry and related proton affinity of the two species and thus six types of fragments can be observed, that are respectively labelled a_n, b_n, c_n, when the charge is retained on the N-terminus, and x_n, y_n and z_n ions when the charge is maintained on the C-terminus (**Figure 1.21**). The CO-NH peptide bond is the most susceptible cleavage site, thus b and y" ions are the most prominent fragments observed during MS/MS analysis on peptides. The mass difference between two adjacent b or y" ions provides information on the mass and thus identity of the amino acid residue, allowing sequencing of the peptide in question. A summary of the most common amino acids including their composition, structure and monoisotopic masses is shown in **Table 1.3**. The two other types of fragments found in most spectra of doubly or higher charge states result from cleavage of at least two internal bonds in the peptide chain. An internal fragment will have lost the initial N- and C-terminal of the peptide, whilst a single amino acid internal fragment with just one side chain is called an immonium ion.

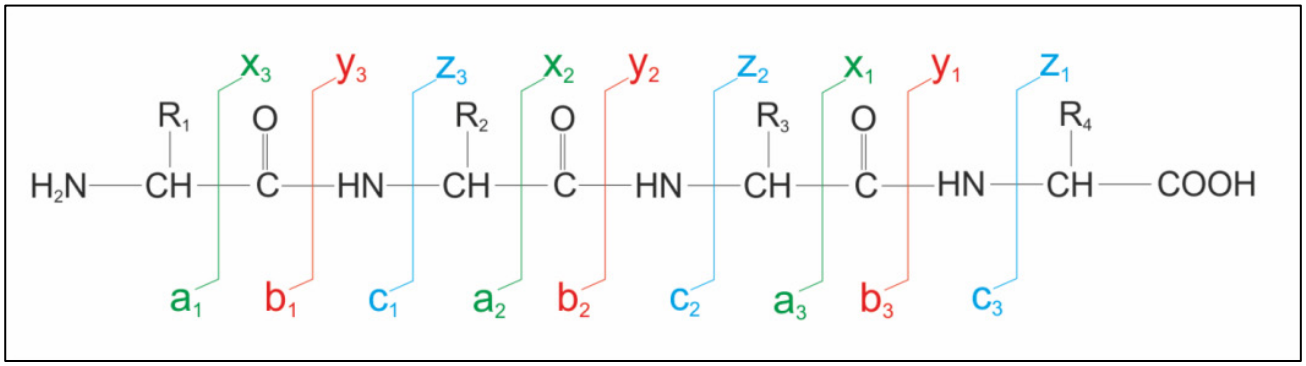


Figure 1.21 – Fragmentation of the peptide backbone.
The nomenclature of the peptide backbone was first proposed by Roepstorff and Fohlman (Roepstorff and Fohlman, 1984).

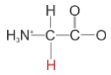
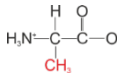
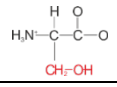
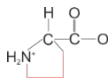
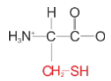
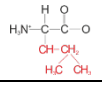
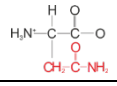
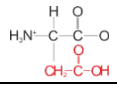
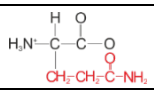
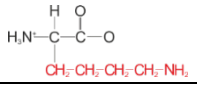
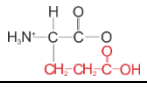
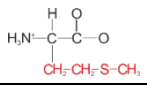
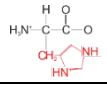
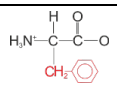
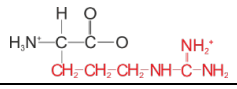
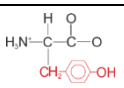
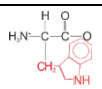
1 letter code	3 letter code	Name and composition	Residue structure	Average Mass (Da)
G	Gly	Glycine C_2H_3NO		57.05
A	Ala	Alanine C_3H_5NO		71.08
S	Ser	Serine $C_3H_5NO_2$		87.08
P	Pro	Proline C_5H_7NO		97.12
V	Val	Valine C_5H_9NO		99.13
T	Thr	Threonine $C_4H_7NO_2$		101.12
C	Cys	Cysteine C_3H_5NOS		103.14
I	Ile	Isoleucine $C_6H_{11}NO$		113.16
L	Leu	Leucine $C_6H_{11}NO$		113.16
N	Asn	Asparagine $C_4H_6N_2O_2$		114.10
D	Asp	Aspartic acid $C_4H_5NO_3$		115.09
Q	Gln	Glutamine $C_5H_8N_2O_2$		128.13
K	Lys	Lysine $C_6H_{12}N_2O$		128.17
E	Glu	Glutamic acid $C_5H_7NO_3$		129.17
M	Met	Methionine C_5H_9NOS		131.20
H	His	Histidine $C_6H_7N_3O$		137.14
F	Phe	Phenylalanine C_9H_9NO		147.18
R	Arg	Arginine $C_6H_{12}N_4O$		156.19
Y	Tyr	Tyrosine $C_9H_9NO_2$		163.17
W	Trp	Tryptophan $C_{11}H_{10}N_2O$		186.21

Table 1.3 – Amino acid residues, compositions, structures and monoisotopic masses.

1.4.5.2 Glycan derivatisation

Purified native carbohydrates can directly be analysed by mass spectrometry, but due to their extensive inter-molecular hydrogen bonding and the lack of primary amino functions, they do not transfer into the gas phase as efficiently as other biomolecules (Dell et al., 1993, North et al., 2010). It is therefore advantageous to derivatise glycans prior to analysis in order to remove hydrogen bonding and allow high sensitivity structural analysis. Furthermore, the increased stability of labile bonds imparted on glycans by derivatisation allows for characteristic and predictable fragmentation patterns, important in *de novo* sequencing of glycans. **Table 1.4** summarises some of the monosaccharide families reviewed in this thesis along with their composition, ring structure of an example moiety as well as their underivatised or derivatised monoisotopic masses.

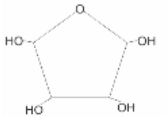
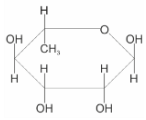
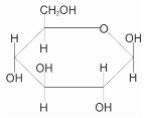
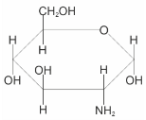
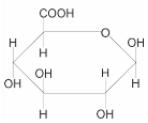
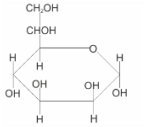
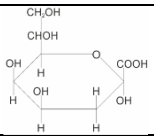
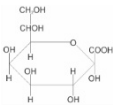
Symbol / Code	Name / Composition / Example	Exemplar ring Structure	Underivatised monoisotopic mass (Average)	Permethylated monoisotopic mass (Average)	Perdeutero-methylated monoisotopic mass (Average)
Pent	Pentose $C_5H_{10}O_5$ (Xyl)		132.12	160.17	166.13
dHex	DeoxyHexose $C_6H_{12}O_5$ (Rha)		146.06	174.20	180.13
Hex	Hexose $C_6H_{12}O_6$ (Glc)		162.14	204.22	213.16
HexN	Hexosamine $C_6H_{13}NO_5$ (GlcN)		161.16	217.27	229.07
HexA	Hexuronic Acid $C_6H_{10}O_7$ (GlcA)		176.13	218.21	227.03
Hep	D-glycero-D-manno-heptose $C_7H_{17}O_9$ (ManHep)		192.12	248.14	260.15
Kdo	D-glycero-D-talo-octu-2-ulsonic acid $C_8H_{14}O_8$ (Kdo)		220.06	276.08	288.09
Ko	D-glycero-D-talo-octu-2-ulsonic acid $C_8H_{14}O_9$ (Ko)		236.09	306.07	321.11

Table 1.4 – Monosaccharide residues, composition, structure and masses.

Derivatisation of the reducing terminus, usually by reductive amination is necessary for detection of carbohydrates by optical methods and is frequently used for mass spectral analysis for purposes such as enhancement of signals or modification of fragmentation patterns. Small aromatic amines such as 2-

aminobenzamide (2-AB), 2-aminoacridone (AMAC), benzylamine, or 2-AP are used extensively for fluorescent detection and increase the proton affinity of the molecules such that $[M+H]^+$ ions are formed sometimes in preference to the more usual $[M+Na]^+$, particularly with ESI where they have been reported to enhance signal strength (Harvey, 2010). Protection of functional hydroxyl and amide groups in carbohydrate characterisation is carried out by permethylation or peracetylation (**Section 2.4.8.1**). Permethylation involves the exchange of protons in such functional groups with hydrophobic methyl groups, as was first introduced by Hakomori and colleagues (Hakomori, 1964).

1.4.5.3 Fragmentation of carbohydrates

Carbohydrate fragmentation pathways were established using EI-MS and FAB-MS instrumentation (Dell et al., 1993) and are preserved, regardless of the ionisation methods used. The accepted nomenclature for describing the fragmentation of carbohydrates was introduced by Domon and Costello in 1988 (Domon and Costello, 1988, Spina et al., 2004, Stephens et al., 2004). The simplest fragmentation at low collision energy, results from cleavage of glycosidic bonds to yield reducing and non-reducing end fragment ions. Ions that retain the charge at the reducing terminus are designated X (cross-ring), Y and Z (glycosidic linkage), whereas ions with the charge at the non-reducing terminus are A (cross-ring), B and C (glycosidic linkage). Sugar rings are numbered from the non-reducing end for A, B and C ions and from the reducing end for the others (**Figure 1.22**) (Harvey, 2010).

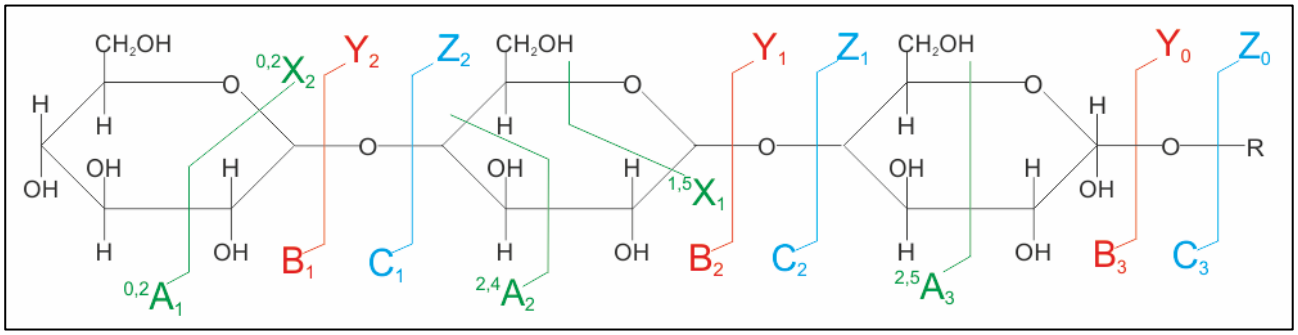


Figure 1.22 – Systematic nomenclature of carbohydrate fragmentation.
Cleavage of the oligosaccharide occurs by breaking of the glycosidic bonds (B, C, Y, Z), and across rings (A and X). Fragment ions containing the non-reducing terminal are A, B and C types, whilst the reducing end fragments are X, Y and Z types. The subscripts of A and X denote the bonds broken in order to form the respective fragments. This nomenclature was first published by Domon and Costello (Domon and Costello, 1988).

In the positive ion mode some native samples and most permethylated or peracetylated derivatives undergo A-type cleavages on the non-reducing side of glycosidic bonds to produce an oxonium ion, which occurs predominantly at amino sugar residues, such as HexNAc residues are present in the sequence (see A-type cleavage in **Figure 1.23**). In the case of cleavage at the reducing side of a residue, a rearrangement is often observed via β-elimination in which the loss of a water molecule and the formation of a double bond between C-1 and C-2 can be observed (Dell et al., 1993). The second most common type of glycosidic cleavage, β-cleavage, occurs when the charge on the fragment ion is not located at the point of cleavage. Both positive and negative mode mass spectrometry can produce β-cleavages and the resulting fragment ions can be reducing or non-reducing, depending on which of the two bonds to the glycosidic oxygen was cleaved, thus providing sequence and branching information (see β-cleavage in **Figure 1.23**) (Dell, 1987). At higher collision energy, cross-ring cleavages are observed. Ring cleavages arise from the sequential movement of

electron pairs around the ring resulting in the breakage of single bonds and the formation of double bonds (see Ring cleavages in **Figure 1.23**) (Dell et al., 1993). Cross-ring cleavages are helpful for assigning linkages and fragment ions can occur from two or more cleavage events in different parts of the molecule. This is seen when fragment ions arise from ring cleavages when they are formed together with β-cleavage ions, are sodium cationised or occurring some distance from a site carrying a permanent positive or negative charge (sulphate or phosphate moiety), in which case these dominate the mass spectrum observed.

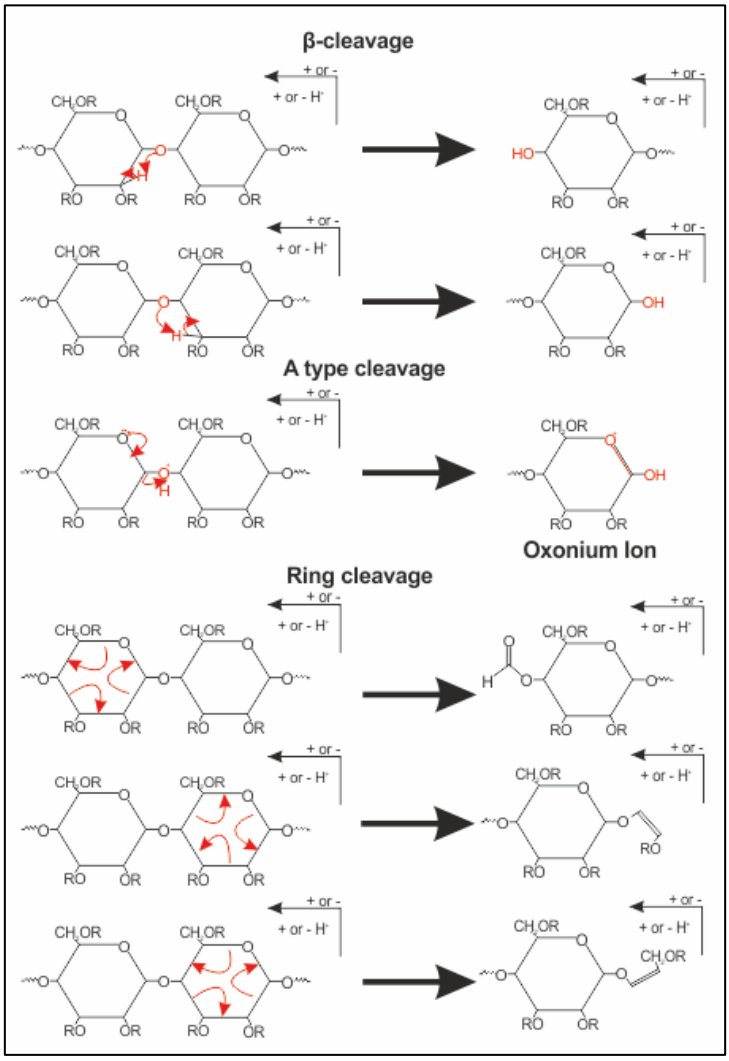


Figure 1.23 – Fragmentation and cleavages of oligosaccharides.
Glycosidic cleavages with hydrogen transfer are known as β-cleavages, where the fragment ions can be either reducing (upper panel) or non-reducing (lower panel). A type cleavages arise from β-elimination from the 3-position and form oxonium ions. Examples of ring cleavages are shown in the bottom three panels. These schematics are adapted from Dell and colleagues (Dell et al., 1993)

1.5 NUCLEAR MAGNETIC RESONANCE IN GLYCOBIOLOGY

1.5.1 Overview

In the study of prokaryotic carbohydrates, whilst mass spectrometric glycoproteomic techniques can provide details on the nature of the glycan and the site of attachment, there is often insufficient information to define carbohydrate structures rigorously, without ancilliary studies involving for example glycosidases to define anomeric configuration or sugar identity. Detailed structural characterisation of a novel prokaryotic glycan

requires further techniques and proves difficult due to the fact that few biosynthetic pathways have been determined in detail for prokaryotes, as opposed to eukaryotes. However, complementary techniques can be employed to overcome these limitations and supplement mass spectrometric information, such as using chemical degradations to characterise structural features such as sugar type and branching. Furthermore, nuclear magnetic resonance (NMR) spectroscopy is often essential in the determination of novel or unusual sugars that are a common feature of prokaryotic glycoconjugates (Castric et al., 2001). The greatly increased sample amounts required for NMR analysis can prove problematic but the introduction of cryoprobe technology for increased sensitivity in NMR analysis has shown potential for detailed structural information from limited amounts of material (Voisin et al., 2005).

NMR is a powerful tool for *de novo* structural characterisation of a glycan. Because this method is non-destructive, the same sample can later be used for other destructive approaches, such as mass spectrometry and linkage analysis. NMR spectroscopy is based on the observation that magnetic nuclei such as ^{1}H, ^{13}C, ^{31}P and ^{15}N can absorb energy at characteristic radiofrequencies when placed in a strong magnetic field. The resonance frequency of a particular nucleus, expressed as the chemical-shift (δ), is sensitive to its chemical environment, making NMR a valuable technique for structural studies. Furthermore, for isolated glycans, such as hydrolysed repeat units, the resonance lines show fine structure, referred to as spin-spin (or scalar) coupling, originating from interactions with the spin states of nearby nuclei sharing bonding electrons. This, however, may be lost in the analysis of larger macromolecular compounds, even with the advantages of modern nuclear Overhauser and 2D methods. A molecule typically has a characteristic NMR spectrum that can be used as a fingerprint (Hard and Vliegenthart, 1993).

When enough sample is available, the anomericity of a particular monosaccharide residue in a glycan can usually be determined by ^{1}H-NMR spectroscopy. The anomeric resonances (H-1 signals) appear in a well-resolved region in a spectrum, where the number of signals is an indicator of the number of residues present in a sample. These signals show characteristic doublets with a splitting (J constant) that is significantly larger for β anomers than for α-anomers. A simple ^{1}H-NMR spectrum can provide the entire primary structure of a glycan if ^{1}H-NMR spectra of well-characterised glycans of related structures are available for comparison. Complete structural elucidation requires full assignment of both the ^{1}H and ^{13}C NMR spectra of a glycan. This is accomplished by a combination of two-dimensional NMR techniques such as correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY) for ^{1}H. These techniques allow the assignment of the ^{1}H signals of individual monosaccharide residues, after which the heteronuclear single-quantum coherence (HSQC) experiment can be used to extend the assignment to the ^{13}C spectrum. A further NMR approach for glycan sequencing relies exclusively on two dimensional ^{1}H-NMR spectroscopy using through space coupling effects, such as nuclear Overhauser effects (NOEs), as the sole source of evidence for linking, position and sequence (Varki et al., 2009).

The range of detailed information provided by NMR experiments is invaluable in the studies of both covalent structures and the conformational properties accompanying their dynamics. Unlike proteins, polysaccharides are not folded into stable globular structure in which inter-residue linkages have fixed or highly restricted conformations, but are to some extent flexible as a result of rotation around glycosidic bonds. Polysaccharides from bacteria give remarkably good NMR spectra despite their high molecular weight, due

to their internal mobility and it is often possible to determine the structure of the repeat unit by NMR without need for depolymerisation. Of all the structurally determined classes of polysaccharides bacterial polysaccharides form the most numerous, particularly LPS O-antigens, CPS and other exopolysaccharides of Gram-negative bacteria (Schiller et al., 2012).

1.5.2 NMR theory

When a sample is placed in a static external magnetic field (B_0) all nuclei will be aligned in the applied field, in which spins aligned with the external field will possess a slightly lower energy than those oriented in the opposite direction, leading to a small net population difference between the two states. In a 1D experiment the application of a short, high-power, radiofrequency pulse perpendicular to the z-axis gives rise to an additional transient magnetic field (B_1), which exerts a torque on the net magnetisation of the sample and forces it into the x-y plane to give transverse magnetisation, exciting a broad range of ^{1}H frequencies (**Figure 1.24a**). The net magnetisation can now be visualised as a vector representing the spin of a nucleus A that is moving (precessing) in the x-y plane (**Figure 1.24b**), which induces the oscillating radiofrequency signal in the x-y plane measured in NMR. Owing to relaxation processes, the signal will decay and return to equilibrium and is therefore called the free induction decay (FID) (**Figure 1.24c**). The total FID is the sum of all frequencies from spins in the sample and can be converted by Fourier transformation into the frequency domain giving rise to signals in the spectrum (**Figure 1.24d**). The resonance positions of these peaks are called their chemical-shifts (δ) and are expressed as parts per million (ppm) of the spectrometer frequency (Hard and Vliegenthart, 1993).

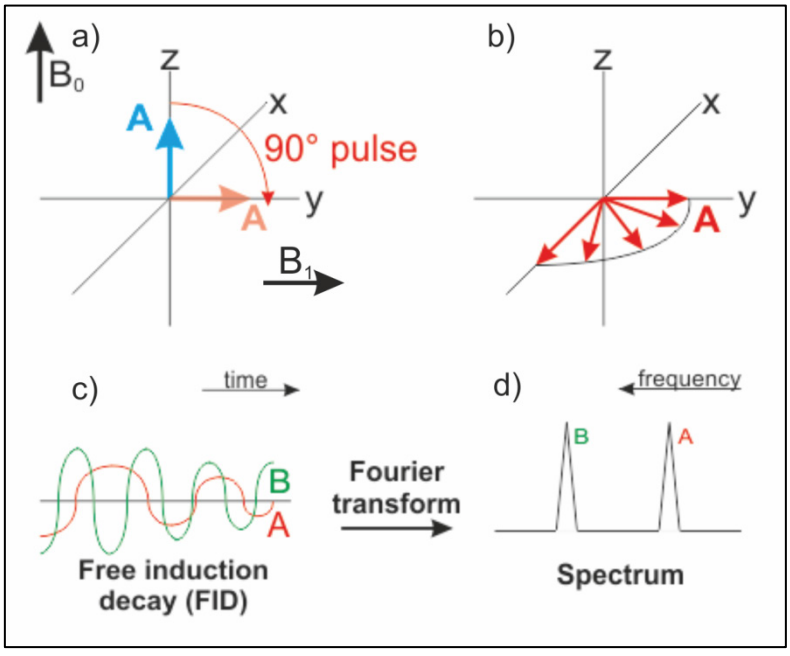


Figure 1.24 – Theory of 1D NMR.
In this schematic vector A describes the net magnetisation of a sample in a magnetic field, which after a 90° pulse precesses in the x-y plane producing a decaying signal that can be converted into a frequency-domain signal by Fourier transform. The schematic was adapted from Hård and colleagues (Hard and Vliegenthart, 1993)

The basic difference between 1D and 2D NMR spectroscopy is the addition of radiofrequency pulses and delay (t1), which are incremented after every acquisition, thus most 2D NMR experiments can schematically be divided into four time periods: preparation, evolution (t1), mixing and acquisition (t2). The various 2D

pulse sequences differ mainly in the type of mixing period, which is the period during which spins are induced to interact with each other. The two types of interactions which can be selected during the mixing period are through-space interactions due to cross-relaxation (NOEs) and through-bond interactions, termed spin-spin (J)-coupling. During the preparation period the spins are allowed to return to thermal equilibrium and during the acquisition period the transverse magnetisation created by the last pulse is detected. The systematic augmentation of t1 gives rise to a modulation in the amplitude of the peaks in the spectra obtained after Fourier transformation in the t2 direction. If the peaks are modulated only by their own frequency a second Fourier transformation in the t1 direction gives rise to so-called diagonal peaks. However, if during the mixing period the spins have interacted with each other the second Fourier transformation will also reveal off-diagonal peaks, the so-called cross-peaks (Hard and Vliegenthart, 1993).

1.5.3 NMR Spectroscopy

All NMR spectrometers consist of a magnet that can produce an intense and homogeneous field, a radiofrequency transmitter for perturbation of the nuclei in the sample and receiver coils for the observation of their responses, which are placed close to the sample in the magnet. Furthermore, a radio-frequency receiver, a computer for manipulation of the radiofrequency pulses, as well as storage and handling of the acquired data are required.

The correlated spectroscopy (COSY) experiment was the first 2D NMR experiment to be proposed. COSY spectra provide information on directly coupled protons. From the fine structure of the cross-peaks in COSY-type experiments information can be extracted about coupling constants, which in turn can be used to estimate dihedral angles in small carbohydrates in a pure sample. With larger polysaccharides or macromolecular complexes investigated, such fine structure is more difficult to elucidate, due to slower tumbling in solution, requiring longer 90º pulses, which is accompanied by the loss of resolution. Total correlation spectroscopy (TOCSY) is a spin-locked version of the COSY experiment. This experiment extends the information on all correlated spins, that are directly or indirectly coupled within a spin system, such as a monosaccharide. Nuclear Overhauser effect spectroscopy (NOESY) methods have been widely applied in structural work on oligosaccharides because cross-peaks are obtained between protons close in space and this information can be used in the form of structural constraints when calculating three-dimensional structures of oligosaccharides. NOEs are often used to determine linkage positions between monosaccharides, but the results should be examined critically since the strongest NOE between adjacent monosaccharide residues need not always be the proton at the linkage carbon. However, the observed NOEs are ensemble averages originating from all different conformations present of the oligosaccharide in solution. The heteronuclear single-quantum coherence spectroscopy (HSQC) experiment belongs to a family of heteronuclear inverse (proton detected) experiments (heteronuclear decoupling), where it is possible to observe carbon resonances split by their directly bonded protons, thus allowing assignment of carbon type. Heteronuclear multiple-bond spectroscopy (HMBC), on the other hand generates cross-peaks between ^{13}C and ^{1}H atoms separated by several chemical bonds. The way of sequencing oligosaccharides is to use long-range ^{1}H-^{13}C couplings over the glycosidic linkages, allowing the unambiguous determination of linkage positions (Hard and Vliegenthart, 1993).

1.5.4 Carbohydrate NMR

Due to the complexity of their structures, no single technique will provide enough information on the characteristics of specific glycoconjugates and thus a multi-technique approach of a combination of analytical techniques is often applied to the study of bacterial glycoconjugates. Oligosaccharides are usually investigated using NMR spectroscopy as deuterium-exchanged samples, so that only non-exchangeable C-linked protons are observed in ^{1}H NMR spectra. The majority of monosaccharide protons of the constituent residues resonate between 3.2 and 4.1ppm, as is visible as the bulk region in a 1D spectrum. Protons that do not resonate in this region include anomeric protons, acetyl and fatty acyl groups or protons shifted out of the envelope of unresolved signals (bulk region) under the influence of substituents such as phosphate, methyl or acetyl groups.

The chemical shift patterns and couplings can be translated into structural information, based on comparison to patterns in a library of relevant reference compounds, but the successful application of this approach requires careful calibration of the experimental technique. In COSY experiments, spin systems corresponding to monosaccharide constituents can be traced and interpretation starts from an anomeric signal or from any other well resolved signal. COSY and TOCSY spectra do not provide monosaccharide sequence information, due to the absence of coupling over the glycosidic linkage, and often NOESY and ROESY spectra are used for this purpose. ^{13}C spectroscopy is at least three orders of magnitude less sensitive than ^{1}H spectroscopy, partly because of the low natural abundance of ^{13}C (1.1%) and partly because the ^{13}C nucleus is inherently less sensitive than ^{1}H, however ^{13}C spectra have a wider chemical shift range than ^{1}H spectra, mitigating the problem of overlapping signals (Schiller et al., 2012). HSQC spectra provide important correlations and owing to the unusually large dispersion of the ^{13}C signals, these spectra have great value for interpretation. In the determination of the structure of the repeating unit of polysaccharides the HSQC experiment is very helpful; however it is currently not feasible to elucidate the distribution of the acetyl groups over the polysaccharide chain. The HMBC experiment is tuned at small (1-10Hz) heteronuclear couplings, such as is useful for sequence information as it yields through bond correlations, providing cross-peaks between each anomeric carbon atom and the proton of the adjacent sugar residue as well as between the anomeric proton and the ring carbon atom of the adjacent sugar residue. This is extremely useful in the assignment of the repeating structure of a typical LPS O-antigen (Vliegenthart and Woods, 2006).

With advances in genomics alongside rapid technological improvements in MS and NMR analysis, prokaryotic glycomics has taken on a new life, with many novel glycan structures being identified and opening up possibilities for glycoengineering. The assignment of genetic information to glycan biosynthetic pathways is complex, combining bioinformatic analysis, linked to systematic mutagenesis and functional analysis of individual genes and structural analysis from MS or NMR, all of which ultimately creates a challenging task to weave the different strands of information together to provide a holistic understanding of the roles that glycans play in prokaryotes (Kay et al., 2010).

1.6 PROJECT AIMS

The main objectives of this PhD project are to apply and improve high sensitivity methodologies for the characterisation of bacterial and archaeal glycopolymers by focusing on the the S-layer glycoprotein of the halophilic archaeon *H. volcanii* and the LPS O-antigen of *B. pseudomallei* serotype 576. The specific aims of each project are:

1.6.1 *Haloferax volcanii*

- to functionally characterise several *agl* genes in the *agl* gene cluster of the S-layer glycoprotein of the halophilic archaeon *H. volcanii* in order to provide a better understanding of the N-glycan biosynthetic pathway
- to definitively establish N- and O-glycan heterogeneity
- to map the S-layer glycoprotein using state of the art mass spectrometric techniques in order to define N-glycosylation occupancy and putative lipid modification

1.6.2 *Burkholderia pseudomallei*

- to characterise the structure of the *B. pseudomallei* 576 LPS O-antigen repeat unit using various mass spectrometric and NMR techniques to provide information for glycoconjugate vaccine creation

Chapter 2
Materials and Methods

2 MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Chemical reagents

- **Alfa Aesar (Lancashire, UK):** acetic anhydride ($(CH_3CO)_2O$) and methyl iodide (CH_3I), deuterated methyl iodide (CD_3I)
- **BDH (Gilingham, UK):** hydrochloric acid (HCl)
- **BOC (Guildford, UK):** nitrogen and argon gases
- **Cambridge isotope Laboratories, Inc.** (Andover, USA): Dodecylphosphocholine (D38, 99%)
- **Fisher Scientific (Loughborough, UK):** Slide-A-Lyzer® dialysis cassettes (3.5kDa MWCO)
- **Fluka (Poole, UK):** ammonium hydrogen carbonate (ambic; NH_4HCO_3), dithioreitol (DTT), sodium hydroxide pellets, 'Dowex' 50W-X8 (H^+) 50-100 mesh, 'Dowex' 1X8 chloride form (Cl^-) 20-50 mesh, sodium periodate, pyridine (C_5H_5N), ethylene glycol ($C_2H_6O_2$), potassium hydroxide (KOH)
- **Invitrogen (Paisley, UK):** NuPAGE® SDS-PAGE Midi Gel System (Tris-acetate, pre-cast), NuPAGE® LDS sample buffer, NuPAGE® reducing agent, NuPAGE® Tris-Acetate SDS Running Buffer Novex Colloidal blue staining Kit and Coomassie R-250.
- **NBS Biologicals (Cambridgeshire, UK):** Spectra/Por® Float-A-Lyzer® G2 (with biotech cellulose ester membranes; 0.1-0.5 MWCO)
- **NORRELL Inc. (Landisville, UK):** NMR tubes
- **Pierce (Rockford, USA):** tri-Sil 'Z' Derivatising agent (TMS), Snakeskin® dialysis tubing (7kDa molecular weight cut off (MWCO)), Float-A-Lyzer G2, 8M guanidine hydrochloride (GuHCl)
- **Purite Ltd. (Oxfordshire, UK):** ultrapure 18 $M\Omega \cdot cm^3$ distilled/deionised (MilliQ®) water from Purite Neptune ultrapure water purification system was used for the preparation of all aqueous solutions
- **ROMIL (Waterbeach, UK):** acetonitrile (ACN), ammonia (NH_4), acetic acid (CH_3COOH), methanol (CH_3OH), chloroform ($CHCl_3$), propan-1-ol (C_3H_7OH), 1-butanol (C_4H_9OH), dimethylsulfoxide (DMSO), sodium hydroxide (NaOH) and trifluoroacetic acid (TFA), formic acid (HCO_2H)
- **Sigma-Aldrich Corporation (Poole, UK):** α-cyano-4-hydroxycinnaminic acid (HCCA), 2,5-dihydroxybenzoic acid (DHB), sinapinic acid (SA), 2-(4-hydroxyphenylazo)benzoic acid (HABA), iodoacetic acid (IAA), potassium borohydride (KBH_4), sodium acetate (CH_3COONa), hexanes (C_6H_{14}), sodium borodeuteride ($NaBD_4$), 2-amino-2-hydroxymethyl-propane-1, 2-diol (Tris; $(HOCH_2)_3CNH_2$), sodium chloride (NaCl), sodium dodecyl sulphate (SDS), pyridine (C_5H_5N), acetic anhydride ($(CH_3CO)_2O$), ammonium acetate (CH_3COONH_4), deuterium oxide (D_2O, 99%), deuterium oxide (D_2O, "100%", 99.990%)

2.1.2 Enzymes

- **Promega Corporation (Hampshire, UK):** trypsin (porcine), chymotrypsin (bovine pancreas), Glu-C (*S. aureus* V8), endoproteinase Lys-C (*L. enzymogenes*), Asp-N (*P. fragi*), proteinase K (*T. album* limber), pepsin (porcine stomach), cyanogen bromide (CNBr)

- **Roche Applied Science (Sussex, UK):** pronase (*S. griseus*)
- **Sigma-Aldrich Corporation (Poole, UK):** trypsin (porcine EC 3.4.21.4), chymotrypsin (EC 3.4.21.1), Benzonase® nuclease (*E. coli* EC 3.1.30.2), prolidase (porcine kidney EC 3.4.13.9), elastase (porcine pancreas EC3.4.21.36)

2.1.3 Lipopolysaccharide, peptide and saccharide standards

- **Applied Biosystems (Warrington, UK):** 4700 Mass Standards kit, including the peptide standards: bradykinin, fragment 1-8 (FW 904.0), angiotensin I (FW 1296.5), adrenocorticotropic hormone fragment (ACTH) 1-17 (FW 2093.4, ACTH fragment 18-39 (FW 2465.7), ACTH fragment 7-38 (FW 3659.2), linear MALDI-TOF BSA Calibration Standard Kit containing lyophilised bovine serium albumin and [Glu^1]-Fibrinopeptide B human (FW 1570.7)
- **Dextra Laboratories (Reading, UK):** oligosaccharide standards (Manα1-3(Manα1-6)Man, Galα1-3Glcα1-3Glc, Galα1-3Gal, Galβ1-4Gal)
- **Sigma-Aldrich Corporation (Poole, UK):** monosaccharide standards (inositol, arabitol, mannoheptose, neuraminic acid, glucuronic acid, galacturonic acid, N-acetylglucosamine, N-acetylgalactosamine, mannose, glucose, galactose, rhamnose, fucose, xylose, arabinose), LPS from *E. coli* O26:B6 (purified by phenol extraction), LPS from *E. coli* O128:B12 (purified by phenol extraction)

2.2 EQUIPMENT AND CONSUMABLES

2.2.1 Equipment

- Perkin Elmer Clarus 500 GC-MS instrument fitted with either Rtx®-5MS (crossbond® 5% dimethylpolysiloxane, 30m, 0.25mmID, 0.25µm df) or a Stabilwax® (crossbond® Carbowax® polyethylene glycol, 30m, 0.32mmID, 0.25µm df)
- PerSeptive Biosystems Voyager-DE STR System MALDI^{TM} mass spectrometer
- Applied Biosystems 4800 MALDI-TOF/TOF^{TM} mass spectrometer
- Dionex, Sunnyvale, USA UltiMate 3000 with a probot spotter and fitted with a Pepmap analytical C_{18} nanocapillary (75µm ID x 15cm length) for offline-nanoLC separation prior to MALDI-TOF/TOF analysis
- API Q-STAR^{TM} Hybrid LC-MS/MS system linked to a nanoLC column
- Xevo G2S Q-TOF LC-MS/MS mass spectrometer on-line to a Waters Acquity UPLC with microbore reverse phase column
- Bruker Avance III 600 MHz, fitted with a TXI/TCI cryoprobe

2.2.2 Consumables

- **Bennett Scientific (Devon, UK):** culture tube caps
- **Bruker – Michrom Bioresources, Inc. (Auburn, CA, USA):** "protein concentration and desalting" and "titanium dioxide phosphopeptide" MicroTraps
- **Fisher Scientific (Loughborough, UK):** 10ml glass syringes
- **Owens Polyscience Ltd. (Cheshire, UK):** Teflon inserts
- **Presearch (Basingstoke, UK):** C_{18} microtrap peptide cartridges
- **Sigma-Aldrich (Poole, UK):** Lo-bind® eppendorf tubes
- **Thermo Scientific (Basingstoke, UK):** Thermo Savant SPD121P Speed Vac connected to a RVT4104 refrigerated vapour trap and Edwards XDS10 pump for volume reduction and concentration, Thermo Savant ModuloD freeze dryer for lyophilisation, IEC Centra CL3 centrifuge for centrifugation
- **VWR International Ltd (Leicestershire, UK):** automatic multi-tube vortexer used for shaking, screw cap glass culture tubes (13x100mm corning, 7.5ml), eppendorfs, glass Pasteur pipettes (150mm and 230mm)
- **Waters Ltd. (Hertfordshire, UK):** reverse phase Classic C_{18} (360mg, 55-150µm), Oasis® HLB, C_{18} (Plus) Sep-Pak cartridges

2.3 SAMPLE PREPARATRION

All biological samples used in the projects discussed in this thesis were prepared by research collaborators. Methods for their preparation are briefly given in **Section 2.3.1** and **Section 2.3.2**.

2.3.1 *H. volcanii* samples

All *H. volcanii* S-layer glycoproteins were kindly supplied by Prof. Jerry Eichler and purified by the Eichler laboratory at Ben-Gurion University of the Negev, Israel. These are summarised in **Table 2.1**.

Species	Strain
H. volcanii S-layer glycoprotein	WR536 (WT)
H. volcanii S-layer glycoprotein	ΔAglM (HVO_1531)
H. volcanii S-layer glycoprotein	ΔAglP (HVO_1522)
H. volcanii S-layer glycoprotein	ΔAglJ (HVO_1517)

Table 2.1 – Table summarising all the *H. volcanii* glycoprotein samples provided by the Eichler group

2.3.1.1 *H. volcanii* cell growth conditions

The *H. volcanii* strain WR536 (H53) wild type/parent strain was obtained from M. Mevarech (Tel Aviv University). All wild type S-layer glycoprotein samples and gene deleted samples were grown in complete medium containing 3.4M NaCl, 0.15M $MgSO_4.7H_2O$, 1mM $MnCl_2$, 4mM KCl, 3mM $CaCl_2$, 0.3% (w/v) yeast extract, 0.5% (w/v) tryptone, 50mM Tris-HCl at pH 7.2 at 40°C (Mevarech and Werczberger, 1985).

2.3.1.2 *H. volcanii* S-layer glycoprotein release from cells

The S-layer glycoprotein was released using EDTA. The cell pellet was washed with 2M NaCl, 50mM Tris–HCl at pH 7.2 and resuspended in minimal medium to which 0.5M EDTA was added. The cells were further incubated for 3-4hrs at 37°C and harvested by centrifugation. The supernatant was dialyzed against 50mM Tris-HCl at pH 7.2, precipitated and lyophilised. In the case of gene deletion strains, these were run on a 10% SDS-PAGE gel and the bands cut out and provided for analysis.

2.3.1.3 S-layer glycoprotein gene deletions

For the deletion of the *H. volcanii agl* genes in question, regions flanking the coding sequences were amplified and restriction sites introduced upstream and downstream. To confirm the deletion of the *agl* gene at DNA level, PCR amplification was performed using primers against an internal region of the respective *agl* gene (Kaminski et al., Magidovich and Eichler, 2009, Yurist-Doutsch et al., 2008).

2.3.2 *Burkholderia* samples

All *B. pseudomallei*, *B. thailandensis* and *B. mallei* samples were kindly supplied by Dr. Joann Prior and her team at the Defence Science and Technology Laboratory (dstl), part of the Ministry of Defence (MOD) at Porton Down, Wiltshire, United Kingdom. These are summarised in **Table 2.2**.

Species	Strain
B. pseudomallei	K96243 purified LPS
B. pseudomallei	K96243 *ΔCPS*
B. pseudomallei	576 purified LPS
B. pseudomallei	576 ΔLPS
B. pseudomallei	576 *ΔCPS*
B. thailandensis	E264 purified LPS
B. mallei	WT

Table 2.2 – Table to summarise all *Burkholderia* samples provided by dstl

2.3.2.1 *Burkholderia* bacterial growth

Bacteria were cultured on nutrient agar (bioMérieux) or statically in nutrient broth at 37°C for 18hrs. For enumeration of bacteria, samples were serially diluted in nutrient broth before plating. The Advisory Committee on Dangerous Pathogens (ACDP) lists *B. pseudomallei* and *B. mallei* as category III pathogens. All manipulations were carried out in class III microbiological safety cabinets located in designated ACDP containment level III laboratories (Nelson et al., 2004).

2.3.2.2 *Burkholderia* LPS extraction

LPS was extracted from lyophilised *Burkholderia* strains or mutants using a combination of previously published methods (Westphal and Luderitz, 1953, Paul et al., 1987). Bacterial cell pellets were digested for 16hrs at 4°C with 15’000Units of lysozyme (Sigma, UK) per mg of bacteria, prior to digestion with 20µg/ml of

DNase I and RNase A (Sigma, UK) for a further 16hrs at room temperature. A modified hot phenol method followed (Nelson et al., 2004).

2.3.2.3 Gel electrophoresis and silver staining of *Burkholderia* LPS

The LPS was separated by SDS-PAGE, where 10µl of each sample were electrophoresed for approximately 2hrs at 100mV in a 12.5% separating gel with 4.5% stacking gel and visualised by silver staining (Chart, 1994).

2.4 METHODS

2.4.1 Gel electrophoresis of *H. volcanii* S-layer glycoprotein

For gel electrophoresis of *H. volcanii* S-layer glycoprotein samples, the NuPAGE® SDS-PAGE Midi Gel System, Tris-acetate pre-cast polyacrylamide gels were used (Invitrogen, Paisly, UK), allowing for separation of a wide range of molecular weight proteins. The samples were denatured at 70°C for 10 mintues prior to electrophoresis and run on 3-8% Tris-acetate gels (pH 8.1) with NuPAGE® LDS sample buffer and reducing agent, as well as NuPAGE® Tris-Acetate SDS Running Buffer, at 150V (constant) for approximately 1hr. The gels, samples and molecular weight markers were stained with the Novex Colloidal blue staining Kit and Coomassie R-250.

2.4.2 Preparation of *H. volcanii* S-layer glycoproteins for MS analysis

For in-gel digestion of S-layer glycoproteins, the bands of interest were excised (if not provided) and destained in 400µl of 50% (v/v) acetonitrile in 0.1M ammonium bicarbonate at pH 8.4. If cysteine residues in the amino acid sequence of the protein in question are likely to be present, disulphide bridges will generally need to be reduced (10mM dithiothreitol at 56ºC for 30mins) and carboxymethylated (55mM iodoacetic acid at room temperature for 30mins). However, since there were no cysteines reported in the predicted amino acid sequence of the *H. volcanii* S-layer glycoprotein and therefore this step was not normally performed.

2.4.3 Proteolytic digests and glycopeptide purification of the *H. volcanii* S-layer

For both in-gel and in-solution digestion of the purified S-layer glycoprotein of *H. volcanii* the sample was incubated with the appropriate enzyme at a suitable concentration ratio, in the respective optimal buffer pH and temperature overnight as is described below.

2.4.3.1 Trypsin

For digestion, the S-layer glycoprotein was incubated with trypsin at 37°C overnight in 50mM ammonium bicarbonate buffer at pH 8.5 at a protease:protein ratio of 1:20 (w/w). Trypsin hydrolyses the peptide bonds at the carboxyl side of lysine and arginine residues, unless they are followed by a proline residue.

2.4.3.2 Chymotrypsin

For digestion, the S-layer glycoprotein was incubated with chymotrypsin at 37°C overnight in 50mM ammonium bicarbonate buffer at pH 7.8 at a protease:protein ratio of 1:20 (w/w). Chymotrypsin selectively hydrolyses the peptide bonds on the C-terminal side of tyrosine, phenylalanine, tryptophan and leucine and secondary hydrolysis may occur on the C-terminal side of methionine, histidine and asparagine.

2.4.3.3 Glu-C

For digestion, the S-layer glycoprotein was incubated with Glu-C at 37°C overnight in 50mM ammonium bicarbonate buffer at pH 7.8 at a protease:protein ratio of 1:20 (w/w). Glu-C is a serine protease that specifically cleaves at the C-terminus of glutamic acid and at a 300-fold lower rate also aspartic acid.

2.4.3.4 Pronase

Pronase allows for the total degradation of proteins and the production of small glycopeptides from purified glycoproteins. The peptides are digested to the size of only a few amino acids. The S-layer glycoprotein was incubated with pronase at 37°C overnight in 50mM ammonium bicarbonate buffer at pH 7.8 at a 70000 u/g working concentration (0.5-2mg/ml solution).

2.4.3.5 AspN

For digestion, the S-layer glycoprotein was incubated with AspN at 37°C overnight in 50mM ammonium bicarbonate buffer at pH 7.8 at a protease:protein ratio of 1:20 (w/w). AspN is a zinc metalloendopeptidase, which selectively cleaves peptide bonds on the N-terminal side of aspartic acid residues.

2.4.3.6 Prolidase

For sequencing of the C-terminal region of the S-layer glycoprotein, it was digested with prolidase. Prolidase is a cytosolic exopeptidase that cleaves on the N-terminal side of proline of small peptides. The requirement of smaller peptides warranted for the sample to first be digested with Glu-C (**Section 2.4.3.3**) and then further digested with prolidase at 37°C overnight in 50mM ammonium bicarbonate buffer at pH 8.4 at a protease:protein ratio of 1:100 (w/w).

2.4.3.7 Elastase

For sequencing of the C-terminal region of the S-layer glycoprotein, it was digested with elastase, a protease that cleaves specific peptide bonds on the carboxyl side of small hydrophobic amino acids such as glycine, alanine and valine. The sample was incubated with elastase at 37°C for 3hrs in 50mM ammonium bicarbonate buffer at pH 8.4 at a protease:protein ratio of 1:100 (w/w).

2.4.3.8 Cyanogen bromide

For sequencing of the C-terminal region of the S-layer glycoprotein, it was digested with cyanogen bromide, which hydrolyses peptide bonds at the C-terminus of methionine residues, except when followed by a threonine residue, or if methionine is oxidised. This reaction is used to reduce the size of polypeptide segments for identification and sequencing. The sample was incubated with a suitable concentration of CNBr in H_2O for 4hrs. The incubation was stopped by drying under nitrogen.

2.4.3.9 Double digests

In the case of double digests being performed on the S-layer glycoprotein of *H. volcanii*, the first enzyme incubation was stopped by heat deactivation of the enzyme used and the sample dried down in a Thermo Savant SpeedVac concentrator (Thermo Fisher Scientific Inc., Waltham, USA). This was followed by re-suspension of the sample in the appropriate buffer for the subsequent enzymatic digest.

2.4.4 Purification and concentration of digested *H. volcanii* S-layer glycopeptides

For the removal of contaminant, such as reagents, enzymes as well as salts, the digested glycopeptides were purified and concentrated for mass spectrometric analysis. Various techniques were employed for the purification of *H. volcanii* S-layer glycopeptides.

2.4.4.1 Sep-Pak® purification propan-1-ol / 5% acetic acid system

Sep-Pak® Classic C_{18} cartridges were attached to 10ml glass syringes and conditioned by eluting sequentially with 5ml methanol, 5ml 5% (v/v) acetic acid, 5ml propan-1-ol and 15ml 5% (v/v) acetic acid. The sample was loaded, washed with 20ml of 5% (v/v) acetic acid for the removal of hydrophilic contaminants and eluted sequentially with 4ml of each 20%, 40%, and 100% (v/v) propan-1-ol in 5% (v/v) acetic acid. The glycopeptides were eluted in the 20% and 40% (v/v) propan-1-ol fractions, which were combined, dried down on a SpeedVac® and lyophilised.

2.4.4.2 Michrom MicroTrap™ cartridge purification

Michrom trap cartridges are packed with application specific HPLC column packing material for the concentration, desalting and detergent removal from samples prior to analysis by HPLC, LC-MS and MALDI-MS. The two traps used were the “peptide concentration and desalting” trap as well as the “titanium dioxide phosphopeptide” trap.

The “protein concentration and desalting” column contains a medium pore, large particle, polymeric reversed-phase packing material with retention similar to a C_8 phase and is designed to bind protein digests, peptides and other small molecules. It was washed with 5-10 trap volumes (25µl) of Solvent B (90% ACN, 10% H_2O, 0.1% TFA, v/v/v), equilibrated with 5-10 trap volumes (25µl) of Solvent A (2% ACN, 98% H_2O,

0.1% TFA, v/v/v). The sample was dissolved in 20µl of Solvent A and loaded, whilst the salt was removed by flushing the cartridge with 5 trap volumes of Solvent A and the concentrated sample eluted with Solvent B.

The “titanium dioxide phosphopeptide” trap contains a medium pore, large particle titanium dioxide material and is designed to bind negatively charged phosphopeptides, while most other peptides pass through due to their net positive charge. It was used to concentrate the N-glycosylated S-layer glycopeptides, due to the HexA presence, which resemble such phosphopeptides. The trap was washed with 500µl of the Elutant Buffer (1% NH_4OH, 2% ACN, 97% H_2O, v/v/v, pH 12) and equilibrated with 500µl of the Load Solvent (5% TFA, 2% ACN, 93% H_2O, v/v/v). The sample was dissolved in 20µl of Load Solvent and the cartridge flushed with 500µl of Wash Solvent (5% TFA, 80% ACN, 15% H_2O, v/v/v). The non-phosphorylated peptides were removed with 100µl Load Solvent, whilst the phosphorylated peptides/N-glycopeptides were eluted with 100µl of Elutant Buffer.

2.4.5 Release of glycans from *H. volcanii* S-layer glycopeptides

Eukaryal N-glycans are all built on the chitobiose core, which links the first GlcNAc residue to the Asn residue of the peptide backbone. This constant and predictable linkage can be exploited using various enzymes, including peptide-N-glycanase F (PNGase F) (Kelly et al., 2009). Furthermore, O-glycosidase is available and capable of releasing some O-glycans, specific to certain core types. These conserved linkages to the peptide backbone, however, are not found in prokaryotic glycoproteins and are thus not applicable for the study of the N- and O-glycosylation of the S-layer glycoprotein of *H. volcanii*. Due to the specificity of the O-glycosidase enzyme, most O-glycans in eukaryotes are chemically released by way of alkaline β-elimination (Dell et al., 1993). The following methods were employed in this thesis in order to release the glycans from the protein for further analysis.

2.4.5.1 Reductive elimination

The β-elimination reaction is carried out under reducing condition to prevent glycan degradation by the “peeling” effect of reducing the terminal/linking residues to their alditol forms and hence is commonly referred to as reductive elimination. This can be carried out directly after enzymatic digests to release both N- and O-glycans, albeit N-glycans are more difficult to visualise. The samples, post enzymatic digests and propanol Sep-Pak® purification, were dissolved in 800µl of 1M potassium borohydride (KBH_4) in a 0.1M solution of potassium hydroxide (KOH) and incubated at 45°C for 20-22hrs. The reaction was terminated by dropwise addition of glacial acetic acid.

2.4.5.2 Ion exchange clean-up

Cationic salts, amino acids and peptides were removed from the sample using a Dowex® (50W-X8 (H^+) 50-100 mesh) column. The Dowex cation exchange column was constructed using a Pasteur pipette plugged at the tapered end with a small amount of glass wool, where a piece of silicon tubing was placed with a flow-blocking adjustable clip. The Dowex was washed with 15ml of 5% (v/v) acetic acid and the sample was loaded and eluted with 2x2ml of 5% (v/v) acetic acid. The collected eluate was lyophilised and the excess

borates were removed as volatile methyl borate compounds by co-evaporation with acetic acid in methanol 10% (v/v) under a stream of nitrogen (4x500µl).

2.4.6 *H. volcanii* S-layer glycoprotein lipid purification and derivatisation

The speculated lipid association of unidentified lipid moieties with the S-layer glycoprotein was investigated using the S-layer glycoprotein sample grown and purified as previously discussed (**Section 2.3.1**). The nature of association of the lipid with the S-layer glycoprotein was investigated using a variety of hydrolysis and separation techniques, discussed below.

2.4.6.1 Purification of lipids by chloroform extraction

Due to the hydrophobic nature of glyco- and phospholipids, simple separation of chloroform and water fractions would remove lipid moieties from the more hydrophilic peptides. The S-layer glycoprotein sample was dissolved in 50% chloroform in water (v/v) and the two fractions separated (2x), isolating non-covalently attached lipids from the S-layer glycoprotein in the chloroform fraction.

2.4.6.2 Hydrofluoric acid (HF) hydrolysis

Hydrofluoric acid treatment was employed to hydrolyse any phosphodiester linkages that might be present between the lipid moieties and the S-layer glycoprotein. The sample was transferred to Lo-bind® eppendorf tubes and the dried sample was incubated with 200µl of hydrofluoric acid (HF) on ice for 24hrs at 4°C. The reaction was terminated by drying under a gentle stream of nitrogen and the protein and lipid moieties separated by chloroform extraction prior to MALDI-MS and MS/MS analysis.

2.4.6.3 Base hydrolysis

For base hydrolysis, the S-layer glycoprotein was combined with 0.8ml PBS, 1ml $CHCl_3$ and 2ml methanol (MeOH) in a glass tube with a Teflon-lined cap to yield a single phase solution ($CHCl_3$:MeOH:PBS = 1:2:0.8, v/v/v). Following addition of 200µl (15N) KOH and intermittent mixing by vortex for 5mins, the mixture was sonicated in a water bath for 15mins at room temperature. The sample was further centrifuged (3000rpm) for 5mins at room temperature and the supernatant was transferred to a fresh glass tube with a Teflon-lined cap. 1ml $CHCl_3$ and 1ml PBS were added to yield a two-phase Bligh-Dyer mixture ($CHCl_3$:MeOH:PBS = 2:2:1.8, v/v/v). After mixing, the sample was again centrifuged (3000rpm) for 5mins at room temperature to separate the phases (Kandiba et al., 2013). The upper phase was removed and the lower $CHCl_3$ phase was dried under a stream of nitrogen. Both lipid and protein fractions were analysed by MALDI-MS and MALDI-MS/MS.

2.4.7 *B. pseudomallei* LPS hydrolysis procedures and de-lipification

Generally lipopolysaccharides are too large to be analysed intact by mass spectrometry. Furthermore, the amphiphilic nature of LPS renders them difficult to dissolve in aqueous solutions or solvents, which

complicates their analysis. In addition, the heterogeneity of LPS, arising from the number and length of fatty acyl chains and phosphate substitutions favours the removal of lipid A prior to the study of O-antigens. The most prominent hydrolysis techniques employed is mild acid hydrolysis, but various other suitable hydrolysis techniques were investigated, due to the unusual LPS characteristics of *B. pseudomallei* 576.

2.4.7.1 Mild acetic acid hydrolysis

The purified LPS sample was hydrolysed in 1%, 2% and 3% (v/v) acetic acid and heated at 100ºC for 1 or 2hrs and centrifuged at 3500rpm for 30mins to remove the polysaccharides in the supernatant from the lipid A in the pellet. Mild acid hydrolysis acts on the acid-labile ketosidic bond of Kdo to the GlcN backbone, linking the core oligosaccharide to lipid A in most LPS. The mild nature of this procedure prevents cleavage of other glycosidic bonds in the rest of the LPS.

2.4.7.2 Methanolysis

For methanolic acid hydrolysis, samples were incubated in various concentrations of methanolic-hydrogen chloride prepared by adding a suitable volume of acetylchloride to dry methanol. Samples were incubated at a range of different temperatures and for different lengths of time and the reaction terminated by removing the reagent under a gentle stream of nitrogen prior to MALDI-MS and MS/MS analysis to monitor hydrolysis (Grobe et al., 2013).

2.4.7.3 Ammonolysis

Ammonolysis refers to any reaction with ammonia that is analogous to hydrolysis, in which a bond is broken and an NH_2 group is appended to one fragment. LPS was hydrolysed in 200µl of 10M ammonium hydroxide at 150ºC for 18hrs. The sample was further cooled, dried and concentrated under a constant stream of nitrogen, dissolved in water and the insoluble components removed by centrifugation (3000g x 30mins at 4ºC). The supernatant was recovered and dried down (Mattos et al., 2005).

2.4.7.4 SDS combined hydrolysis

This technique was devised from a combination of published and previously used hydrolysis techniques applied to bacterial LPS de-lipification (Novem et al., 2009). The LPS was hydrolysed in 1% SDS (sodium dodecyl sulphate) in 10mM sodium acetate (pH 4.5) at 100°C for 1hr. The sample was freeze dried and the sodium dodecyl sulphate (SDS) removed by re-suspension in distilled water followed by sonication for dispersal of the sample. The sample was further centrifuged at 2000g for 30mins in acidified ethanol (0.5% HCl in ethanol) to separate the lipid A (pellet) from the polysaccharide (supernatant). The polysaccharide sample was further dialysed in 50mM ammonium bicarbonate buffer at pH 8.4 to remove any further traces of SDS using a Float-A-Lyzer® G2 for 24hrs, lyophilised and hydrolysed by mild acid hydrolysis (1% acetic acid (v/v), 100°C, 2hrs). The sample could then be further derivatised prior to MALDI-MS and MS/MS analysis.

2.4.8 Chemical derivatisation of carbohydrates for MS analyses

As discussed in **Chapter 1**, native glycans can be analysed directly by mass spectrometry, but do not ionise as efficiently due to their hydrophilic nature and are thus derivatised prior to analysis by mass spectrometry. Derivatisation methods can either employ a reducing end tag or protect most or all of the functional groups of the carbohydrates.

2.4.8.1 Permethylation

Permethylation involves the exchange of protons in hydroxyl and amide groups for hydrophobic methyl groups and involves the successive base-catalysed ionisation of these functional groups followed by methylation. The technique was introduced by Hakomori and colleagues (Chou and Yang, 1979) who described the use of methylsulphinyl carbanion from dimethyl sulphoxide (DMSO) as the base and methyl iodide as the methyl donor and the chemistry was optimised for high sensitivity application to peptides and carbohydrates by the discovery in isotope dilution MS experiments that the kinetics is on the second timescale rather than the hours and days originally used (Morris, 1972). This led to all subsequent methods utilising permethylation times of minutes in order to minimise byproducts and maximise sensitivity. Permethylation is now commonly catalysed by weaker bases such as sodium hydroxide and has become the staple derivatisation step in many glycomic strategies (Dell, 1990).

For specific oligosaccharide analysis, samples were methylated using the sodium hydroxide procedure. A slurry of NaOH-anhydrous DMSO (1ml) was added to the sample, followed by 0.5ml iodomethane. The reaction mixture was agitated on an automatic multi-tube vortexer for 30mins at room temperature and the reaction quenched by dropwise addition of water. The permethylated glycans were extracted in 1ml of chloroform, washing with 4ml of water (4x) and dried under a gentle stream of nitrogen gas. The resulting glycans were then purified by Sep-Pak® C_{18} using the aqueous ACN system.

2.4.8.2 Deutero-permethylation

The sample was deuteropermethylated using the sodium hydroxide procedure. A slurry of NaOH-anhydrous DMSO (1ml) was added to the sample, followed by 0.5ml deuterated iodomethane-d3. The reaction mixture was agitated on an automatic multi-tube vortexer for 30mins at room temperature and quenched by dropwise addition of water. The deuteropermethylated glycans were extracted in 1ml of chloroform, washing with 4ml of water (4x) and dried under a gentle stream of nitrogen gas. The resulting glycans were then purified by Sep-Pak® C_{18} using the aqueous ACN system as described below.

2.4.8.3 Reduction

To reduce the oligosaccharides prior to methylation, they were incubated with 10mg/ml of $NaBH_4$ (in 2M aqueous NH_3) at room temperature for 2hrs. After borate removal with acidified methanol co-evaporation, they were permethylated as described previously (**Section 2.4.8.1**).

2.4.8.4 Deutero-reduction

To deuteroreduce the oligosaccharides prior to methylation, they were incubated with 10mg/ml of $NaBD_4$ (in 2M aqueous NH_3) at room temperature for 2hrs. After borate removal with acidified methanol co-evaporation, they were permethylated as described previously (**Section 2.4.8.1**).

2.4.8.5 Sep-Pak® clean up: Aqueous acetonitrile system

Purification of the permethylated sample was achieved by reverse-phase chromatography using C_{18} Sep-Pak® cartridges. Sep-Pak® Classic C_{18} cartridges were attached to 10ml glass syringes and conditioned successively with 5ml of methanol, water, ACN and 15ml of water. The sample was dissolved in 200µl of 50% (v/v) H_2O/MeOH and loaded onto the cartridge. The sample was washed with 5ml of water and then eluted stepwise with 3ml of each 15%, 35%, 50%, 75% and 100% (v/v) aqueous ACN. The fraction volumes were reduced on the SpeedVac® and lyophilised.

2.4.8.6 TMS derivatisation

For sugar composition, trimethylsiylated (TMS) derivatives of the polysaccharides were prepared by methanolysis (1M methanolic-acetylchloride, 80ºC, 16hrs) and dried under nitrogen. The samples were re-N-acetylated (methanol/pyridine/acetic acid (50/1/5, v/v/v), room temperature, 15mins) with the reagent removed by drying under nitrogen. Trimethylsiylation was carried out with tri-Sil 'Z' Derivatising agent (room temperature, 30mins). The samples were purified by centrifugation and the supernatant was re-dissolved in hexanes prior to GC-MS analysis.

2.4.8.7 Alditol acetates

For compositional analysis by alditol acetate, the non-permethylated samples were hydrolysed (2M TFA, 121ºC, 2hrs), dried under nitrogen and the monosaccharides reduced (10mg/ml $NaBD_4$ in 2M aqueous NH_3, room temperature, 2hrs). After borate removal with acidified methanol co-evaporation, they were re-acetylated (acetic anhydride, 100ºC, 1hr), purified by washing with chloroform and water (4x) and prepared for GC-MS by suspension in hexanes.

2.4.8.8 Linkage analysis

Linkage information was achieved by the analysis of partially methylated partially acetylated alditol acetate derivatives. Samples were permethylated as described followed by hydrolysis (2M TFA, 121ºC, 2hrs) and reduction (10mg/ml $NaBD_4$ in 2M aqueous NH_3, room temperature, 2hrs). After borate removal with acidified methanol co-evaporation, the monosaccharide samples were re-acetylated (acetic anhydride, 100ºC, 1hr), purified by washing with chloroform and water (4x) and prepared for GC-MS by suspension in hexanes (Albersheim, 1967).

2.4.8.9 Fatty acid methyl ester analysis

Fatty acid methyl esters were produced for lipid analysis and samples were hydrolysed in 200µl methanolic-acetylchloride (1M) overnight in 80ºC and the reaction terminated by drying under a stream of nitrogen. The samples were dissolved in chloroform and washed several times with water (4x), then dried under nitrogen and ultimately prepared for GC-MS by suspension in hexanes.

2.4.8.10 Esterification of carboxylic groups using methanolic HCl

This procedure was employed for the identification of the 190Da moiety contained in the N-glycan of the *H. volcanii* S-layer glycoprotein. The digested glycopeptides were incubated with 2M HCl/MeOH for 15mins, dried under nitrogen, then purified by nanoLC and analysed by MALDI-TOF.

2.4.9 *B. pseudomallei* polysaccharide preparation for NMR analysis

Similarly to mass spectrometry, contaminations in the sample preparation for NMR can be caused by reagents and salts and will affect data collection. Purification of the polysaccharides prior to NMR analysis is discussed below (**Section 2.4.9.1**). Initially the hydrolysed polysaccharide fractions (O-antigen and core) assessed, were found to produce very heterogeneous spectra, due to hydrolysis at different monosaccharide linkages, where too much information was presented in a crowded spectrum. To avoid these problems of heterogeneity and solubility of the LPS sample, a different approach was devised for obtaining suitable NMR data in the form of mixed micelles.

2.4.9.1 Ion-exchange purification

Polysaccharide purification and separation of the hydrolysed O-antigen and core fractions of the polysaccharide was achieved by ion-exchange chromatography. The separation was based on the carboxylic acid group of the Kdo and Ko components of the core, where Kdo has a pK_a of 2 - 3. A column was assembled by packing a Pasteur pipette fitted with a piece of silicone tubing at its tapered end with Cl^- Dowex® beads. The column was eluted with 20ml of water. The sample was loaded onto the column and eluted with 10ml of H_2O, 5 ml of 5% acetic acid (v/v) and 5ml of 0.1M HCl into separate glass culture tubes. The O-antigen was found in the H_2O fraction, whilst the core was found in the 5% acetic acid (v/v) and the 0.1M HCl fraction. The fractions were lyophilised and deuterium exchanged for NMR analysis.

2.4.10 Creation of mixed micelles

The method employed for creating mixed LPS-micelles was derived from and originally performed by Sandro Sonnino and Domenico Acquotti (Acquotti and Sonnino, 2000). Their ganglioside-micelles were created by mixing the ganglioside sample (dried) with perdeuterated dodecylphosphocholine (DPC), where synthetic DPC mixed with a small amount of gangliosides formed small spherical micelles. The gangliosides were mixed with DPC at a molar ratio of 1:40. As these mixed-micelles have only previously been used in relation

to gangliosides, and not been seen with LPS, a pilot study was created to assess the feasibility of this technique.

2.4.10.1 Pilot study to create mixed micelles using *E. coli* LPS O26 and O128

2mg of the acquired LPS standards of *E.coli* LPS (serotypes O26 and O128) were mixed with 3.2mg of DPC in the presence of deuterated potassium phosphate buffer (50mM, 0.4ml, p^2H 6). The mixed-micelle-LPS samples were further lyophilised and deuterium exchanged for NMR analysis.

2.4.10.2 Creating mixed micelles for *B. pseudomallei* 576 LPS

10mg of *B. pseudomallei* LPS was mixed with 16mg of DPC in the presence of deuterated potassium phosphate buffer (50mM, 1ml, p^2H 6). The mixed-micelle-LPS samples were further lyophilised and deuterium exchanged for NMR analysis. Upon identification of RNA and DNA nucleotide signals in the NMR spectra, the micelles were further treated with Benzonase® endonuclease (250 Units, in 50nM Tris-HCl, 1mM MgCl2, 0.1mg/ml BSA) for 6hrs at 37°C. The Benzonase was then digested with pronase (250µl of 20mg/ml) overnight at 37°C and dialysed in potassium phosphate buffer (50mM, 0.4ml, pH 6) for 48hrs. The mixed-micelle-LPS samples were lyophilised and deuterium exchanged again for NMR analysis.

2.4.10.3 Deuterium exchange

In NMR, oligosaccharides are usually visualised as deuterium-exchanged samples, so that only non-exchangeable C-linked protons are observed in NMR spectra. The purified polysaccharides and mixed-micelle-LPS, core and O-antigen fractions, as separated by ion-exchange chromatography, were lyophilised and re-dissolved in 100% D_2O for removal of exchangeable protons. This was repeated 3 times to ensure minimal ^1H proton exchange throughout NMR analysis.

2.4.11 Mass spectrometric analysis

The mass spectrometric techniques employed in this thesis were introduced in **Section 1.4**, and consisted of an approach utilising the following instruments: GC-MS, MALDI-TOF, MALDI-TOF/TOF and ESI Q-TOF mass spectrometers. Details of the operating parameters are described as follows.

2.4.11.1 GC-MS

GC-MS analysis was carried out on a Clarus 500 GC-MS instrument (Perkin Elmer, Cambridgeshire, UK). The sugar samples were dissolved in hexanes and injected on a RTX-5MS fused silica capillary column (30mx0.25mm internal diameter). For analysis of the TMS derivatives, the oven was held at 65ºC for 1min before being increased to 140ºC at a rate of 25ºC min^{-1}, then to 200ºC at a rate of 5ºC min^{-1} and finally to a temperature of 300ºC at a rate of 10ºC min^{-1} and held for 5mins. For analysis of the alditol acetate derivatives, the oven was held at 60ºC for 1min and increased to 190ºC at 20ºC min^{-1}, from where the temperature is increased to 230ºC at 1ºC min^{-1}. The final temperature increment is to 300ºC raised at 25ºC

min^{-1} and held for a total of 5mins. The programme for the linkage analysis is started at 60ºC for 1 minute, and increased to 300ºC in a single ramp at 8ºCmin^{-1} and held for 5 minutes. For FAME GC-MS analysis, the capillary column of the GC-MS instrument was changed to a Stabilwax Crossbond® Carbowax® polyethylene glycol column (30m x 0.32mm internal diameter). The oven programme used starts at 90°C, which is held for 4mins. This is increased to 130°C at a rate of 20°C min^{-1} and held for 1 minute and finally increased to 250°C at a rate of 5°C min^{-1} and held for 5mins.

2.4.11.2 MALDI-TOF MS and TOF/TOF MS/MS of carbohydrates

Derivatised samples were dissolved in 10-20µl of methanol and 1µl aliquots were combined with an equal volume of the sample specific matrix (10mg/ml) in their appropriate solvents (typically DHB for glycans and lipids, and HCCA for peptides) and spotted onto a MALDI target plate, before drying under vacuum. MALDI-TOF was performed using a Voyager-DE STR Elite mass spectrometer workstation equipped with delayed extraction technology (Perseptive Biosystems, Hertfordshire, UK). MALDI-TOF MS and TOF/TOF MS/MS were also performed using a 4800 MALDI-TOF/TOF™ Analyser (Applied Biosystems, Darmstadt, Germany) mass spectrometer. The 4700 Calibration Standard Kit, calmix was used as the external calibrant for both instruments. In the linear mode where sinapinic acid was used as a matrix for analysis of carbohydrates in the 60-70kDa range, the BSA Calibration Standard Kit was used. For MALDI-TOF/TOF the molecular ions were selected and subjected to collision induced dissociation (CID), where the collision energy was set to 1kV and pressurised argon was used as the collision gas ($3.5x10^{-6}$ Torr).

2.4.11.3 Offline LC-MALDI-TOF MS and TOF/TOF MS/MS of glycopeptide

For offline liquid chromatography matrix-assisted laser desorption ionization-time of flight (MALDI-TOF/TOF-MS Analysis), the protease-generated peptides were separated by using an UltiMate 3000 LC system (Dionex, Sunnyvale, CA), fitted with a Pepmap analytical C_{18} nanocapillary (15cm length x 75µm internal diameter; Dionex). The digests were loaded onto the column and eluted using Solvent A (0.1% (v/v) trifluoroacetic acid in 2% ACN and Solvent B (0.1% TFA in 90% ACN), in the following gradient: 0–60% Solvent B (0– 36min), 60–90% Solvent B (36–37min), 90% Solvent B (37–40min) and 100% Solvent A (40–41min). Eluting fractions were mixed with α-cyanohydroxy-cinnamic acid matrix at a concentration of 3.3mg/ml and spotted onto a metal MALDI target plate. MALDI TOF/TOF MS profiling was performed using the 4800 MALDI-TOF/TOF™ Analyser (Applied Biosystems, Darmstadt, Germany) in the positive reflectron mode and set for delayed extraction, with major peaks selected using a data dependent acquisition method for collision induced dissociation (CID) and sequencing by tandem mass spectrometry.

2.4.11.4 Online LC-ESI MS and MS/MS with Q-STAR and Q-TOF mass spectrometers

Enzymatic digests of glycopeptides were also analysed by online liquid chromatography and were resuspended in 0.1% (v/v) formic acid and were analysed by nano-LC-ESI-MS/MS using a reverse-phase nano-HPLC system (LC Packings, Dionex Ltd., Camberley, UK) connected to a quadrupole TOF mass spectrometer (API Q-STAR® Pulsar, Applied Biosystems/MDS Sciex, Toronto, Canada). Online peptide separation was achieved by a binary nano-HPLC gradient generated by an UltiMate pump fitted with a

Famos autosampler and a Switchos microcolumn switching module (Dionex Ltd., Camberley, UK), utilising a pre-microcolumn C_{18} cartridge and an analytical C_{18} nanocapillary (15cm x 75mm ID, PepMap). The digests were first loaded onto the precolumns and eluted with 0.1% formic acid prior to transferral onto the analytical nanocapillary HPLC column and eluted using a gradient of Solvent A (0.05% (v/v) formic acid in a 99.5% (v/v) ACN mixture). The instrument was pre-calibrated using 10-100fmol/µl of [Glu1]-fibrinopeptide B human / 5% (v/v) acetic acid (1:3, v/v). In the MS/MS mode, the collision gas utilised was nitrogen and the pressure was maintained at $3.5x10^{-6}$ Torr). In Q-TOF experiments on the Xevo G2S similar procedures were used.

2.4.12 NMR analysis

One and two dimensional NMR spectra were obtained on a Bruker Avance III 600 MHz Ultra-Shield DRX50 NMR spectrometer fitted with a TXI/TCI cryoprobe, run on Linux computers with Bruker Topspin software. All NMR spectra were recorded at room temperature in 100% D_2O. The internal chemical shift reference used for calibration of individual experiments was acetone at 2.04 ppm.

2.4.13 Data processing and interpretation

The GC-MS data acquired were processed and analysed using TurboMass 5.4 GC-MS Software (Perkin Elmer, Cambridgeshire, UK), whilst the obtained MALDI-TOF MS and TOF/TOF MS/MS data were processed using Data Explorer® 4.9 Software (Applied Biosystems, Darmstadt, Germany) and 4800 Series Explorer™ Software (Matrix Science, London, UK). The MS data were generally noise filtered (correction factor of 0.7). For data handling of ESI Q-STAR MS and MS/MS data Analyst QS Software (Applied Biosystems, Darmstadt, Germany) with an automatic information-dependent-acquisition (IDA) function was used.

To aid in the assignment of peptide molecular ion peaks MassLynx Mass Spectrometry Software (Waters, Hertfordshire, UK) was used for in-silico digestion of the protein in question, allowing virtual digestion with various enzymes and predicting peptide molecular ions, missed cleavages, methionine oxidation and cysteine carboxymethylation. For correct identification of the protein being analysed Mascot searches were performed. Data from all MS/MS spectra were used to search the Swiss Prot database for peptide sequences consistent with the fragment ions using the Mascot Search Engine (Matrix Science, London, UK). The “gold standard” of manual/visual inspection was used to oversee all interpretations.

NMR 1D and 2D data was processed and presented using MestreNova® (Mestrelab Research, Chemistry Software Solutions). All spectral and structural assignments were carried out manually with CorelDraw X3 Graphics Suite Software.

Chapter 3

Haloferax volcanii S-layer glycoprotein analysis

3 *HALOFERAX VOLCANII* S-LAYER GLYCOPROTEIN ANALYSIS

3.1 EXPERIMENTAL STRATEGY

This section describes efforts made in profiling the S-layer glycoprotein of the halophilic archaeon *H. volcanii*, including investigations made into defining the N-glycosylation site occupancy and heterogeneity as well as a putative lipid modification. Furthermore, several *agl* genes of the agl gene cluster of the S-layer N-glycan biosynthesis pathway were functionally characterised.

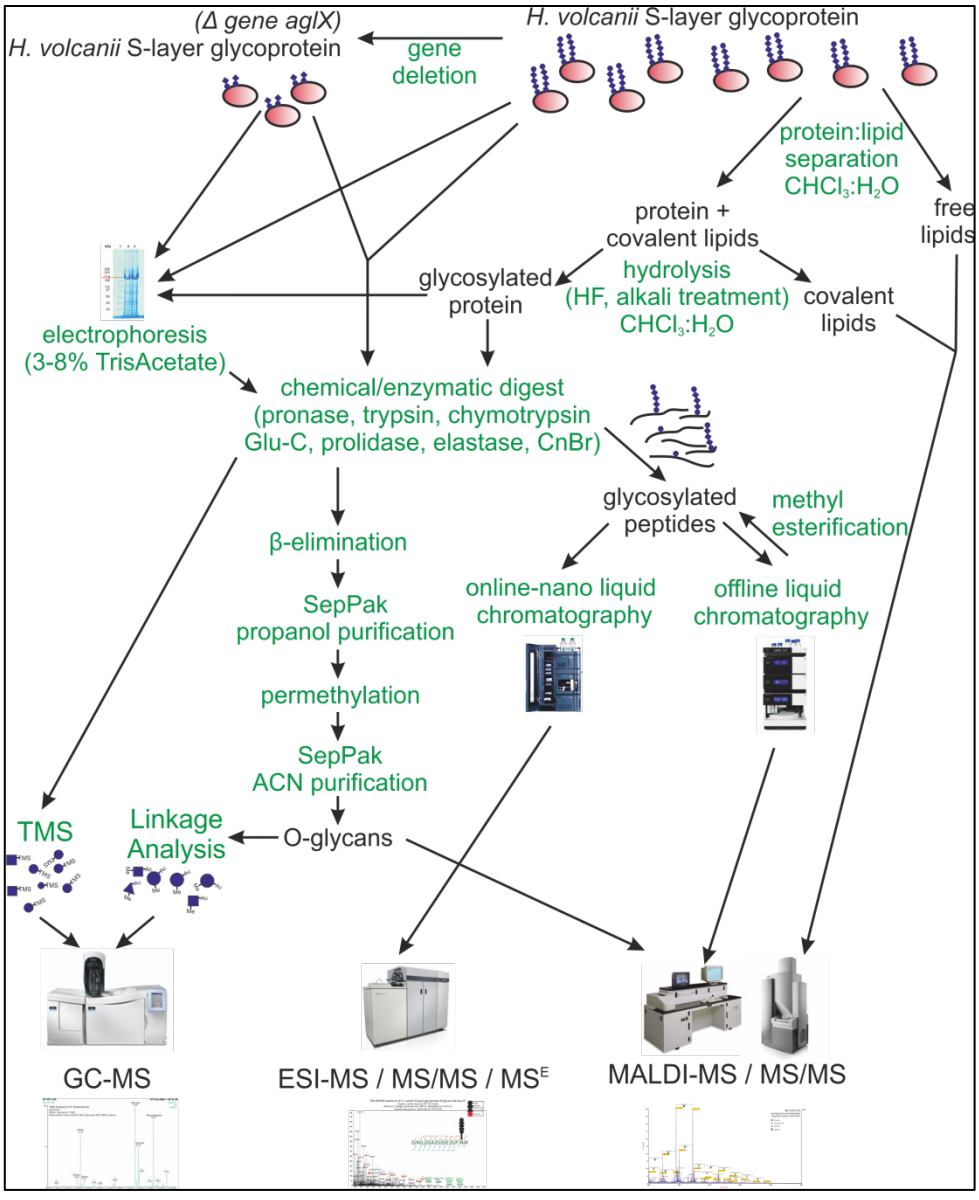


Figure 3.1 – Simplified representation of the experimental strategy employed in this section. Methods employed in the study of the S-layer glycoprotein of *H. volcanii*, including the glycan monosaccharide composition, O-glycan linkage and structure, protein and glycosylation site mapping, potential lipid modification and functional studies.

This work has been carried out in collaboration with Prof. Jerry Eichler from the Eichler laboratory at Ben-Gurion University of the Negev, Israel. All *H. volcanii* S-layer glycoprotein samples were grown, purified and provided by our collaborators (**Section 2.3.1.1**). The wild type strain WR536 was provided as released by EDTA and either purified by dialysis and delivered as a lyophilised sample or as bands run on SDS-PAGE

gels (**Section 2.3.1.2**). For the study of the functional roles of certain *agl* genes, the deletion of the *H. volcanii agl* genes in question was achieved by our collaborators and provided as gel bands cut out of SDS-PAGE gels run (**Section 2.3.1.3**). The gene-deleted strains analysed included *ΔAglM* (*ΔHVO_1531*), *ΔAglP* (*ΔHVO_1522*) and *ΔAglJ* (*ΔHVO_1517*).

The approaches employed in the experimental strategy in this project are two-fold (summarised in **Figure 3.1**). The functional characterisation of the genes responsible for the N-glycosylation of the S-layer glycoprotein was performed by comparing gene deleted strains with the wild type *H. volcanii* S-layer glycoprotein as purified by gel electrophoresis, enzymatically digested, separated by liquid chromatography and analysed by MALDI-MS and MALDI-MS/MS. In the venture to gain further understanding by fully profiling the S-layer glycoprotein, various techniques were employed shedding light in a site-specific analysis strategy on the composition of the glycan moieties N- and O-linked to the S-layer glycoprotein, as well as the confirmation of the amino acid sequence and investigation into the possible role of lipids as a posttranslational modification. This required a range of analysis techniques including GC-MS, MALDI-MS and MS/MS and ESI-MS, ESI-MS/MS, as well as MS^E, post enzymatic and or chemical derivatisation or hydrolysis.

3.2 COMPOSITIONAL ANALYSIS OF ALL S-LAYER GLYCANS

For insight into the composition of the N- and O-linked glycoprotein of the S-layer of *H. volcanii*, a series of compositional analysis techniques were employed, where wild type samples for the WR536 S-layer N- and O-glycans decorating the S-layer protein were assessed for composition by GC-MS. The wild type WR536 sample as extracted and purified by the Eichler group (**Section 2.3.1**), was digested with pronase enzyme, an enzyme that hydrolyses the peptide backbone into short peptides of only a few amino acids (**Section 2.4.3.4**). This was done to improve solubility and allow for further steps of hydrolysis of the glycopeptides into monosaccharides for subsequent derivatisation by TMS or alditol acetate methods (**Section 2.4.8.6** and **Section 2.4.8.7**).

TMS and alditol acetate derivatised samples were analysed by GC-MS and run alongside commercial standards using the retention times of elution on the chromatogram as well as the specific mass spectrometric fingerprints of each monosaccharide in order to determine the monosaccharide composition of the glycans present on the S-layer glycoprotein. Notably, hexuronic acid monosaccharides are less easily assessed using the alditol acetate procedure, because acidic sugars are not easily amenable to analysis by GC-MS without additional manipulation, as adequate chromatography cannot be easily achieved for compounds containing a carboxyl group. Considering that the N-glycan of the S-layer glycoprotein of *H. volcanii* contains hexuronic acids, the digested S-layer glycoprotein was therefore derivatised with a trimethylsilyl reagent prior to TMS analysis by GC-MS (**Section 2.4.8.6**).

The GC-MS chromatograms for TMS derivatives of the S-layer derived glycans plus standards are shown in **Figure 3.2**. Comparison of the two chromatograms as well as examination of the mass spectrometric fingerprints of each moiety eluted from the gas chromatography column allowed for full annotation of the chromatogram showing the monosaccharide composition of the S-layer glycans of *H. volcanii*.

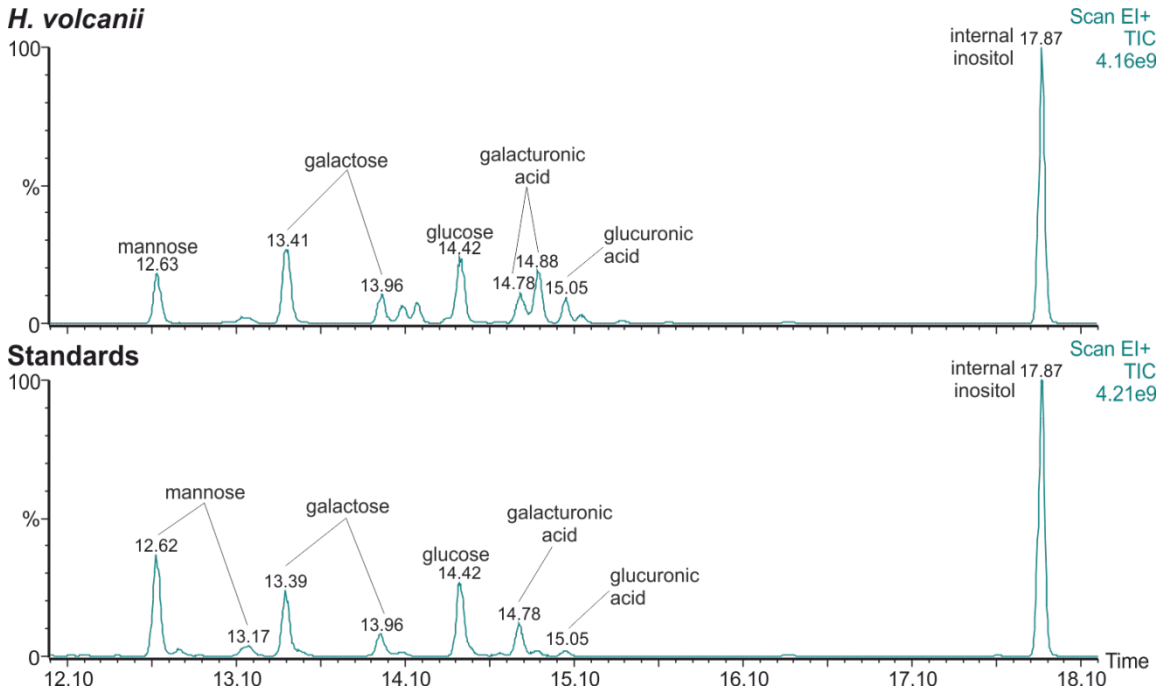


Figure 3.2 – GC-MS chromatogram of TMS derivatised glycans of the *H. volcanii* S-layer glycoprotein.
The glycan composition of the *H. volcanii* S-layer glycoprotein (top panel) was run alongside commercially obtained standards, at 50nM each (bottom panel), as run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column. Both preparations were supplemented with 50nM of the internal standard inositol.

The N- and O-glycans of the S-layer glycoprotein, were found to contain mannose (elution time: 12.63, **Figure 3.3**), galactose (elution time: 13.41 and 13.96), glucose (elution time: 14.42), galacturonic acid (elution time: 14.78 and 14.88, **Figure 3.4**) and glucuronic acid (elution time: 15.05). The TMS chromatogram peaks and fingerprints were checked for the possible presence of naturally present sugar modifications, such as O-methylation that would be expected to be preserved during sugar release from the S-layer. No naturally occurring modification was observed in the TMS chromatogram.

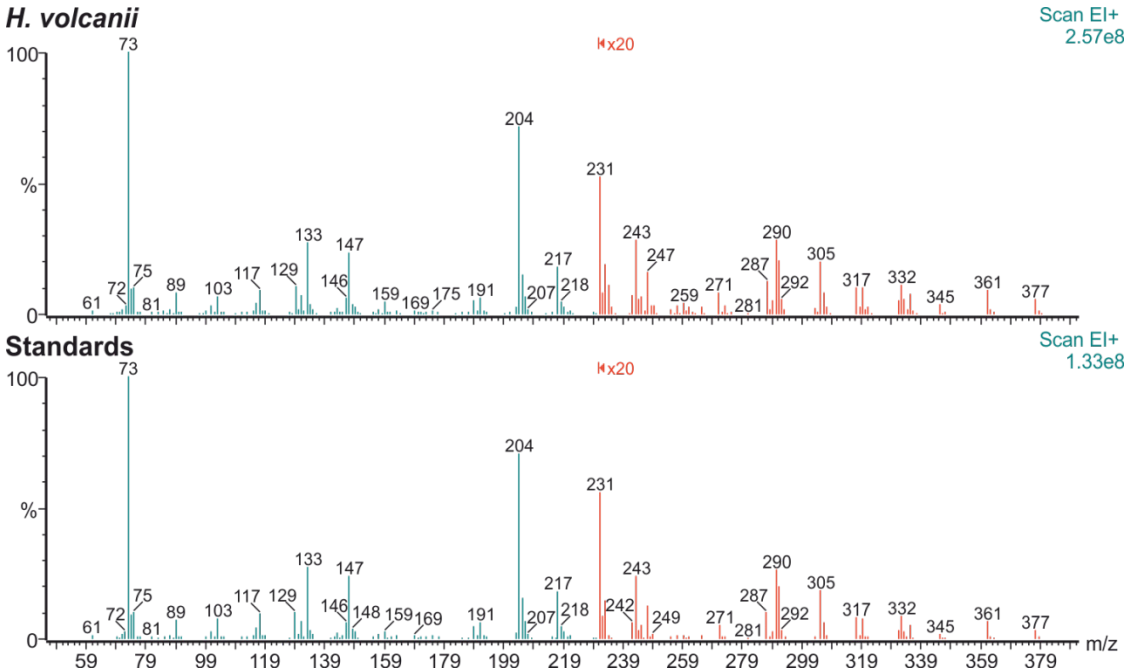


Figure 3.3 – Mass spectrometric fingerprint of TMS derivatised mannose.
The fingerprint is from the GC-MS chromatogram of the TMS derivatised mannose moiety (top panel) and that of the commercially obtained standard for mannose (bottom panel), as run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column. The fingerprint is magnified from m/z 231 onwards and is highlighted in red.

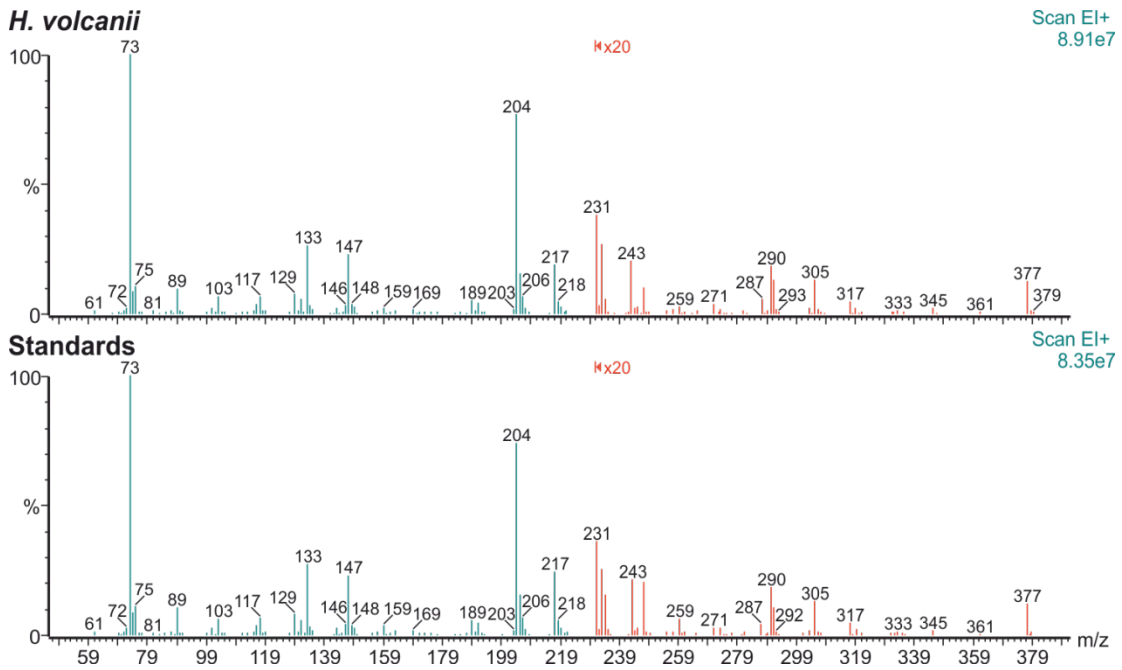


Figure 3.4 – Mass spectrometric fingerprint of TMS derivatised galacturonic acid.
The fingerprint is from the GC-MS chromatogram of the TMS derivatised galacturonic acid moiety (top panel) and that of the commercially obtained standard for galacturonic acid (bottom panel), as run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column. The fingerprint is magnified from m/z 231 onwards and is highlighted in red.

3.2.1 Summary discussion of TMS compositional analysis

The O-glycan disaccharide of the S-layer glycoprotein was previously suggested to be composed of a glucose and galactose disaccharide (Sumper et al., 1990) and this is supported by their presence in the GC-MS chromatogram. However, the *H. volcanii* S-layer N-glycan also contains two hexose moieties, of which only the final hexose has been identified to be a mannose monosaccharide (Calo et al., 2011, Girrbach et al., 2000, Cohen-Rosenzweig et al., 2012) (**Figure 3.2** and **Figure 3.3**) (**Section 3.7.2**). The hexose moiety linked to the asparagine residues in the S-layer glycoprotein has not been identified yet, and could therefore be considered to be any of the hexose moieties present in the TMS chromatogram. Furthermore the hexuronic acid moieties present in the N-glycan pentasaccharide of the *H. volcanii* S-layer glycoprotein were previously suggested to be consistent with glucuronic acid moieties, however, the presence of galacturonic acid moieties in the TMS chromatogram indicates that the N-glycan pentasaccharide contains both glucuronic and galacturonic acid moieties (**Figure 3.2** and **Figure 3.4**) (Magidovich et al., 2010, Yurist-Doutsch et al., 2010).

Unfortunately, no suitable enzymatic or chemical hydrolysis method is available for the selective release of the N- and O-glycans from the S-layer. Therefore, the lack of such separation did not aid in the discrimination of the monosaccharide identifications in the N- and O-glycan moieties. Interestingly, however, the amount of glucuronic acid and galacturonic acid is not substantially less than the mannose, galactose and glucose. The uronic acids are confined to the N-glycans whilst the hexoses are in both the N- and O-glycans. Their comparable abundance suggests that the stoichiometry of the O-glycans is unlikely to be very different from the N-glycans. The proposal that the O-glycan disaccharides occupy multiple sites in the threonine-rich region towards the C-terminal of the *H. volcanii* S-layer (Mescher and Strominger, 1976, Sumper et al., 1990), therefore does not seem to be supported by the TMS data in **Figure 3.2**.

3.3 ANALYSIS OF THE O-GLYCAN OF THE S-LAYER GLYCOPROTEIN

In the analysis to investigate the O-glycan moiety initially identified by Mescher and Sumper and colleagues, the extracted and purified S-layer glycoprotein was digested by trypsin (**Section 2.4.3.1**). The peptide/glycopeptide sample mixture was further purified and contaminants removed by the Sep-Pak® purification propan-1-ol / 5% acetic acid system (**Section 2.4.4.1**). The glycopeptides were further treated to release the glycan moieties by the reductive β-elimination reaction (**Section 2.4.5**). The glycan moieties were permethylated, purified and separated using the aqueous acetonitrile Sep-Pak® system (**Section 2.4.8**). Analysis by MALDI-MS and MALDI-MS/MS allowed for confirmation of the O-glycan structure, whilst further derivatisation of the O-glycan bearing ACN fraction into partially methylated alditol acetates analysed by GC-MS allowed for verification of the published linkage information of the O-glycan moiety of the *H. volcanii* S-layer glycoprotein.

3.3.1 MALDI mass spectrometry of *H. volcanii* S-layer O-glycans

The purified and derivatised glycans were analysed by a 4800 MALDI-TOF/TOF™ Analyser mass spectrometer. The released O-glycan moiety of the S-layer glycoprotein of *H. volcanii* was found in the 35% ACN fraction of the Sep-Pak® clean up and the MALDI-MS spectrum of the 35% ACN fraction is shown in **Figure 3.5**. A single O-glycan molecular ion of composition Hex-Hexitol was identified at m/z 493.24 and confirmed by MALDI-MS/MS (**Figure 3.6**). Furthermore, upon investigation of the MALDI-MS spectrum, no extended version of the hexose disaccharide was identified.

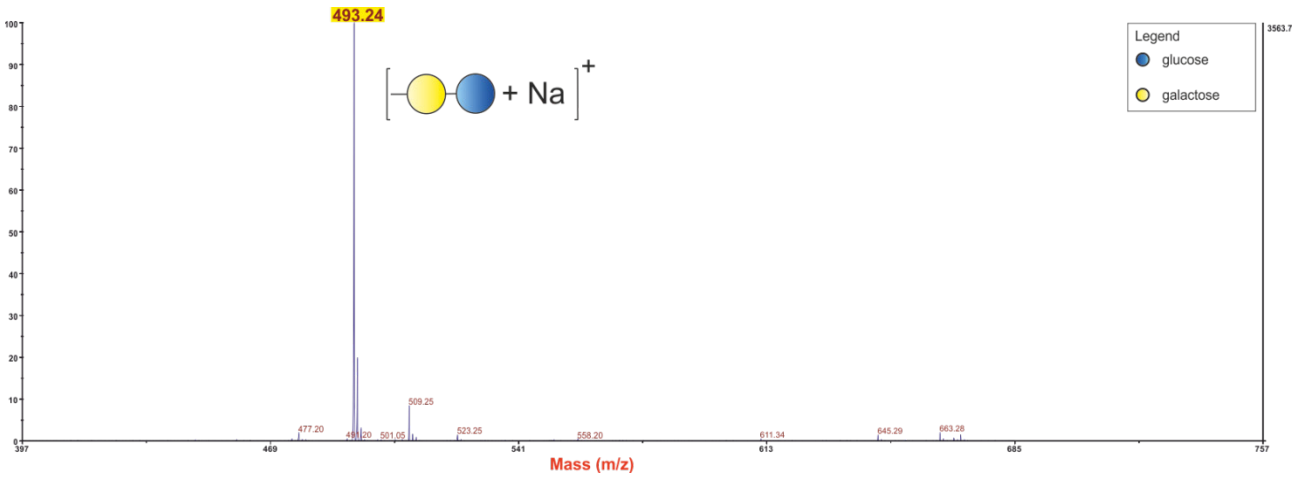


Figure 3.5 – MALDI-MS spectrum of *H. volcanii* S-layer O-glycans.
The spectrum shows the O-glycan moieties of the *H. volcanii* S-layer glycoprotein as digested by trypsin and released from their respective serine/threonine amino acid residues by β-elimination and further derivatised by permethylation (Sep-Pak 35% ACN). The sample was analysed on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer. Non-annotated peaks correspond to partial degradation products which did not relate to the O-glycan moiety investigated.

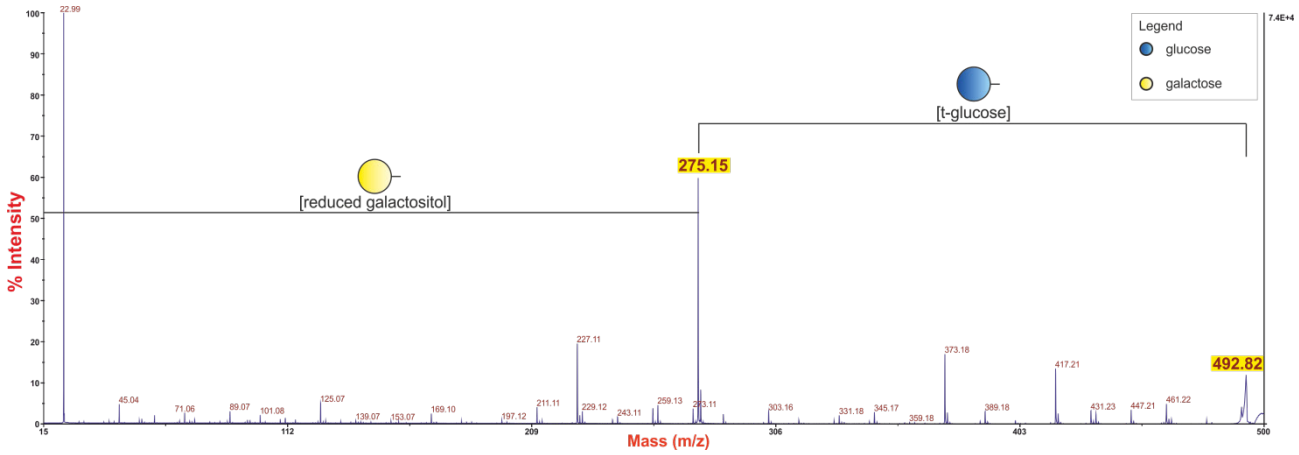


Figure 3.6 – MALDI-MS/MS spectrum of molecular ion at m/z 493 of *H. volcanii* S-layer O-glycans.
The *H. volcanii* S-layer glycoprotein as digested by trypsin and released from their respective serine/threonine amino acid residues by β-elimination and further derivatised by permethylation (Sep-Pak 35% ACN). The sample was analysed on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer. Non-annotated peaks correspond to partial degradation and cross-ring cleavages of the disaccharide.

3.3.2 Linkage analysis of the O-glycan moieties of the S-layer glycoprotein

To confirm the linkage profile published previously, the O-glycan containing sample fraction recovered by β-elimination and Sep-Pak purification was further derivatised to create partially methylated alditol acetates that were analysed by GC-MS (**Section 2.4.8.8**). Linkage analysis was performed on the O-glycan and the linkage possibilities identified *ab initio* using the mass spectrometric fingerprints and known fragmentation pathways as well as comparisons with the mass spectrometric fingerprints of standards. **Figure 3.7** shows the linkage chromatogram obtained from the O-glycan containing fraction of the sample, clearly showing a 2-linked galactitol (**Figure 3.8**) and terminal glucose (**Figure 3.9**). It is to be noted that upon release from the peptide by β-elimination, the linking sugar is converted into a galactocitol moiety, thus rendering a different mass spectrometric fingerprint to one of a 2-linked hexose.

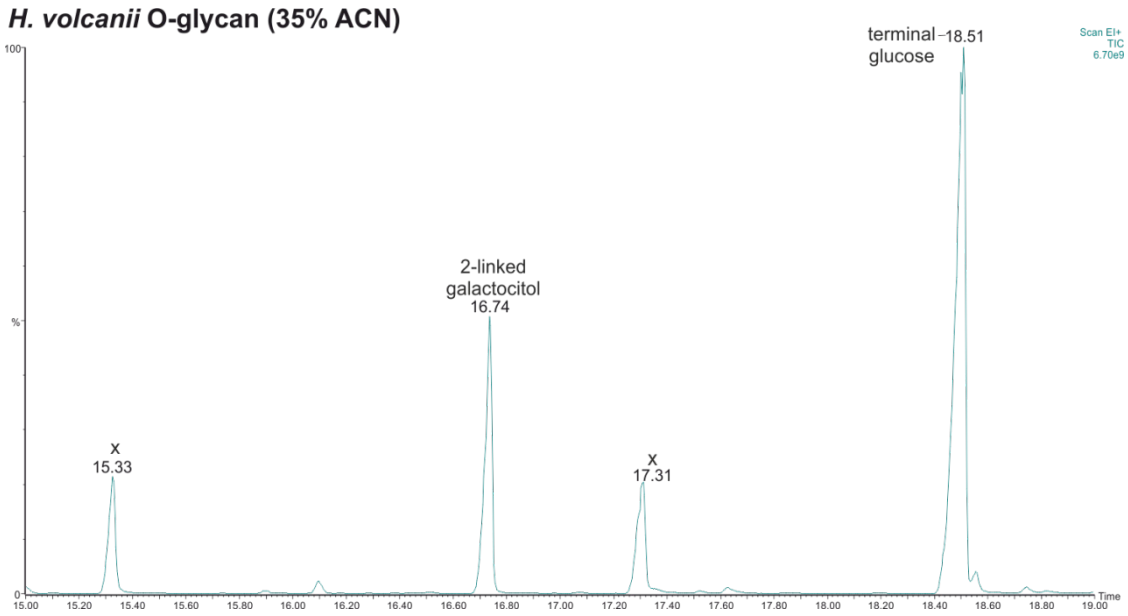


Figure 3.7 – GC-MS chromatogram of the linkage anlysis of the *H. volcanii* O-glycan.
The linkage analysis chromatogram shows the presence of the 2-linked galactositol and terminally linked glucose moiety, as run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column.The annotation was completed using the mass spectrometric fingerprints and fragmentation pathways of the monosaccharide structures and in comparison to standards. Non-sugar impurities present in the chromatogram are labelled x.

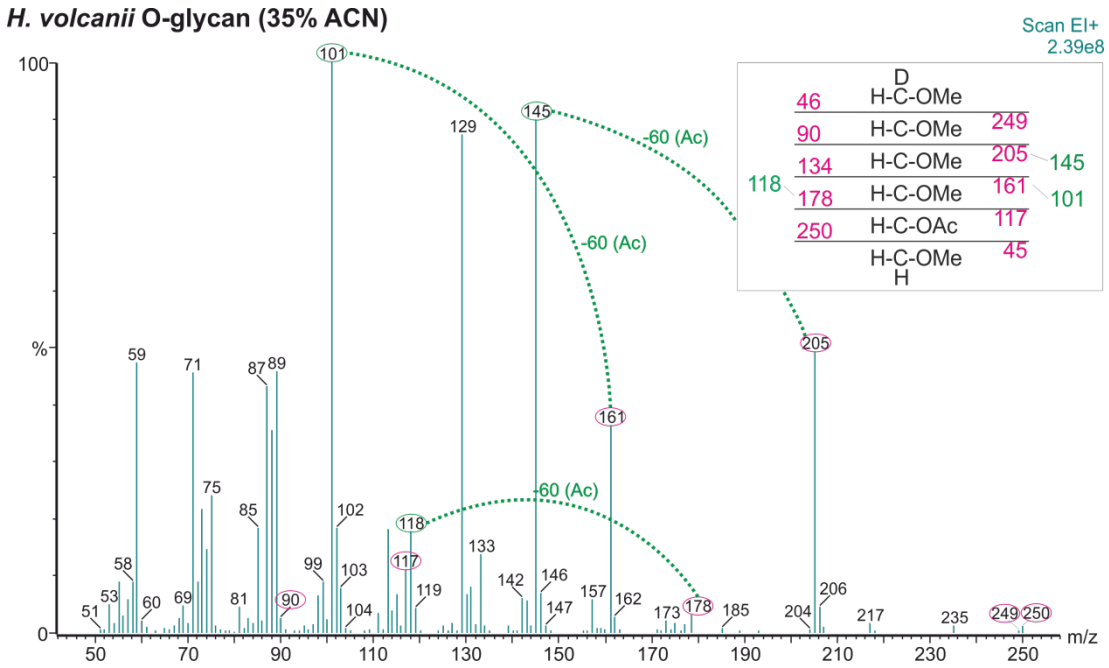


Figure 3.8 – Mass spectrometric fingerprint of the 2-linked galactositol in the *H. volcanii* O-glycan.
The mass spectrometric fingerprint of the 2-linked galactositol was obtained from the linkage analysis chromatogram run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column.The fragmentation pattern of the monosaccharide is annotated, showing the main fragments (pink) of the molecule as well as sub-fragments arising from the loss of functional groups (green). The structure of the derivatised monosaccharide is shown in the schematic.

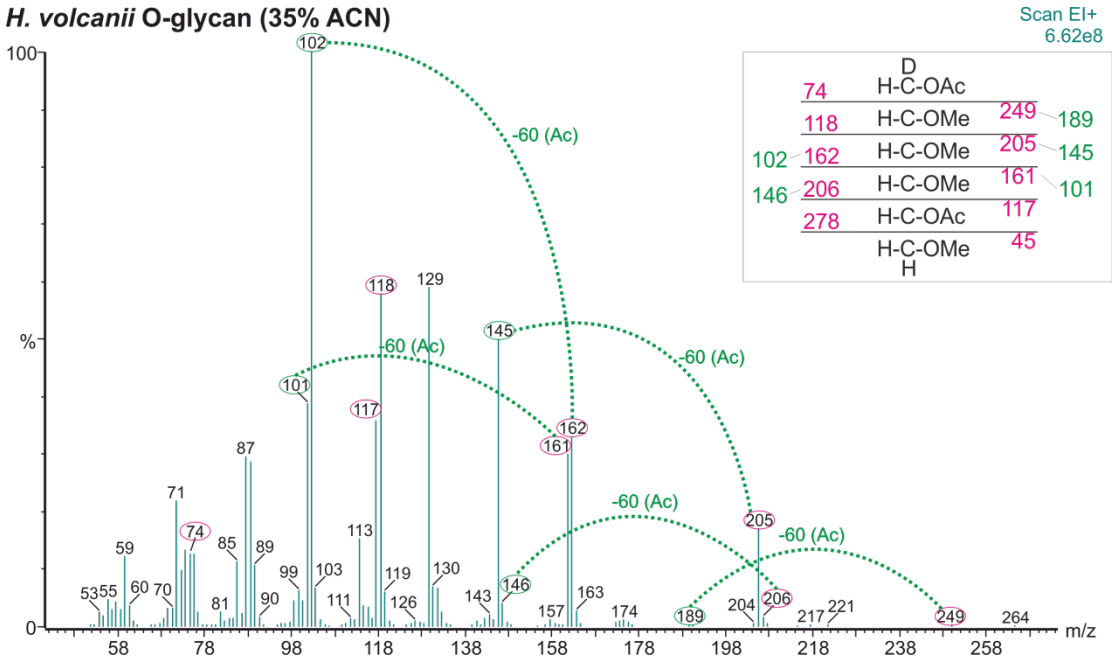


Figure 3.9 – Mass spectrometric fingerprint of the terminal glucose in the *H. volcanii* O-glycan.
The mass spectrometric fingerprint of the terminal glucose was obtained from the linkage analysis chromatogram run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column.The fragmentation pattern of the monosaccharide is annotated, showing the main fragments (pink) of the molecule as well as sub-fragments arising from the loss of functional groups (green). The structure of the derivatised monosaccharide is shown in the schematic.

3.3.3 Summary discussion of S-layer glycoprotein O-glycan analysis

The O-glycan moiety of the S-layer glycoprotein was initially characterised by Sumper and colleagues, who degraded the S-layer by alkaline β-elimination, liberating the disaccharide O-glycan (Sumper et al., 1990). The saccharide fraction identified in their study contained galactose and glucose in a 1:1 stoichiometry. They identified the galactose moiety, as the monosaccharide directly linked to threonine or serine in an O-

glycosidic linkage. This information was confirmed by the purification of the O-glycans of the *H. volcanii* S-layer glycoprotein and their derivatisation for analysis by MALDI-MS and MALDI-MS/MS as well as further analysis by linkage analysis by GC-MS. The results presented here confirm that the O-glycan is composed of a glucose(1-2)galactose disaccharide and that no other O-glycans are present.

3.4 FUNCTIONAL CHARACTERISATION OF *AGL* GENES BY MS ANALYSIS

In *H. volcanii*, N-glycosylation is mediated by the products of the *agl* gene cluster. The members of the *agl* gene cluster are responsible for the assembly and attachment of the pentasaccharide. The characterisation of the N-glycosylation process in *H. volcanii* was undertaken by Abu-Qarn and co-workers (**Section 1.2.2**). The putative *H. volcanii* homologs of genes involved in the eukaryal or bacterial N-glycosylation were identified by bioinformatics and reverse transcription polymerase chain reaction confirmed that the proposed N-glycosylation genes were being transcribed. Deletion of some genes showed slower growth and interfered with surface layer glycosylation, represented by modified migration on SDS-PAGE and glycostaining (Abu-Qarn and Eichler, 2006). Subsequently additional *agl* genes were identified either based upon their proximity to previously identified *agl* sequences or upon re-annotation of the region of the genome where all but one of the previously identified *agl* sequences clustered (**Table 3.1**) (Abu-Qarn et al., 2008b, Yurist-Doutsch et al., 2008, Abu-Qarn et al., 2007, Yurist-Doutsch and Eichler, 2009).

Encoded Protein	Suggested function
AglB	oligosaccharyltransferase
AglD, AglE, AglG, AglI	glycosyltransferases
AglF	glucose-1-phosphate uridyltransferase
AglM	**uncharacterised**
AglP	**uncharacterised**
AglJ	**uncharacterised**

Table 3.1 – Previous Agl proteins found to be involved in the assembly of the pentasaccharide

To deduce the function of the members of the *agl* gene cluster responsible for the assembly and attachment of the pentasaccharide, samples of parent (wild type) and mutant (gene-deleted) S-layer glycoproteins were analysed and compared. SDS-PAGE gel pieces received from the Eichler group, containing the S-layer glycoprotein from the *H. volcanii* parent and mutant strains, were subjected to in-gel tryptic digestion (**Section 2.4.2**). The obtained peptides were then separated by liquid chromatography and examined by MALDI mass spectrometry (**Figure 3.10**).

The products generated in this manner included the N-terminal ‘ERGNLDADSESFNK’ peptide (m/z 1581) with a missed cleavage (Sumper et al., 1990), encompassing the glycosylated Asn-47 residue. The MALDI-MS analysis of the wild type WR536 glycoprotein shows the S-layer glycoprotein-derived peptide containing Asn-47 decorated with a pentasaccharide and its precursors, including a hexose (glycopeptide at m/z 1743), two hexuronic acids (glycopeptides at m/z 1919 and 2095, respectively), a 190Da species (glycopeptides at m/z 2285) and a final hexose subunit (glycopeptides at m/z 2448). The N-glycopeptide shown in **Figure 3.11** was used as a “reporter” molecule for investigating changes in glycosylation resulting from targeted mutations of the *agl* genes described in the next three sections.

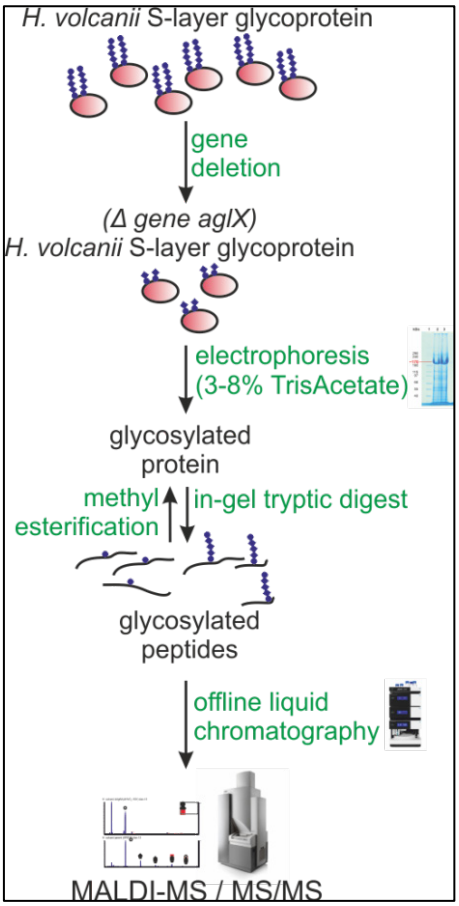


Figure 3.10 – Experimental strategy employed for functional characterisation of *H. volcanii agl* genes.
Methods employed in the study of the *agl* genes responsible for the biosynthesis of the *H. volcanii* S-layer glycoprotein N-glycan.

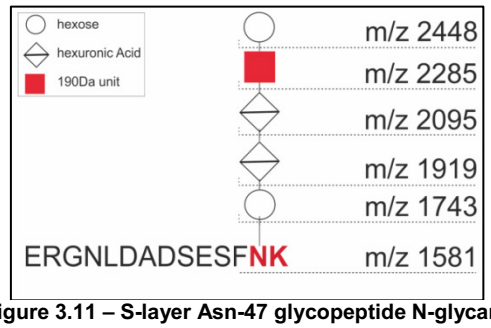


Figure 3.11 – S-layer Asn-47 glycopeptide N-glycan.
The annotated m/z values in the schematic correspond to the molecular ions of the peptide alone and together with all intermediates of the mature pentasaccharide

3.4.1 Function of AglM

AglM (HVO_1531) was thought to participate in the addition of the hexuronic acid located at position two of the pentasaccharide decorating the S-layer glycoprotein. The involvement of AglM (HVO_1531) in this post-translational modification was tested by comparison of *H. volcanii* parent strain with the AglM-deleted cells. Comparative data for the wildtype and mutant are presented in **Figure 3.12**. Data from the wildtype are in the lower panel. As explained above, molecular ions are observed for the pentasaccharide and its biosynthetic precursors. The relative abundances of the five glycoforms seen in this spectrum were characteristic of all wildtype samples examined. In the case of the *aglM* (*HVO_1531*) gene deleted strain (**Figure 3.12** - *ΔHVO_1531* – *ΔaglM* – upper panel), the Asn-47-containing peptide is only decorated with

one monosaccharide moiety (m/z 1743), which allows for the conclusion that AglM is involved in the biosynthesis of the hexuronic acid found at position two of the native pentasaccharide decorating the *H. volcanii* S-layer glycoprotein. With the involvement of HVO_1531 in protein N-glycosylation thus verified, the gene was renamed *aglM*, according to the nomenclature first proposed by Chaban and colleagues (Chaban et al., 2006) for genes involved in archaeal N-glycosylation.

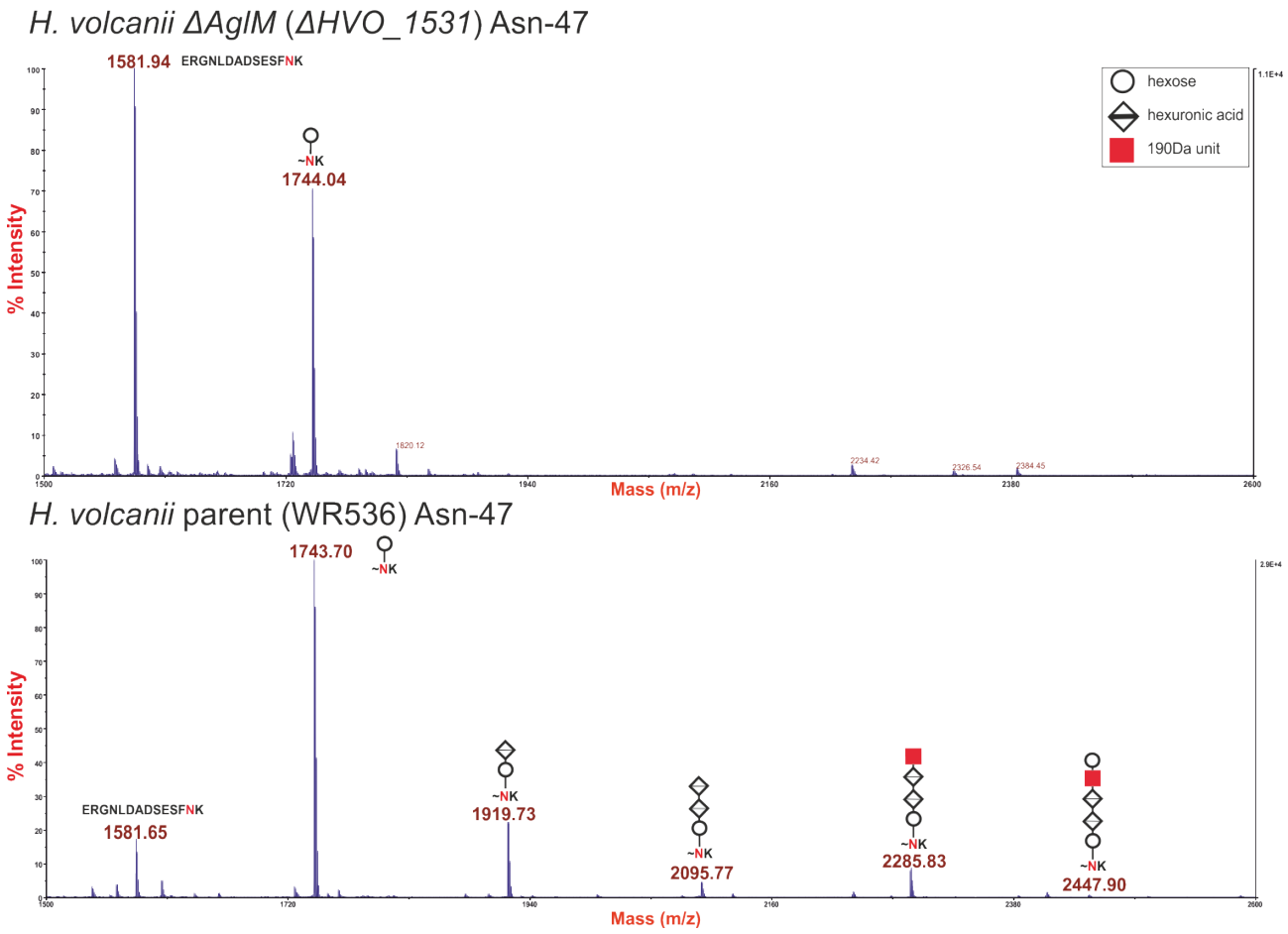


Figure 3.12 – MALDI-MS of the Asn-47-containing tryptic peptide (AglM).
The samples were derived from gel pieces containing the *H. volcanii* S-layer glycoprotein from the parent (WR536) and *aglM*-- deleted strains (*ΔaglM* – *ΔHVO_1531*). The sample was analysed on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer after in-gel trypsin digestion and separation by an offline UltiMate 3000 LC system. Non-annotated signals are not glycopeptides.

3.4.2 Function of AglP

Previous analysis had assigned the 190Da species at position four of the glycoprotein-linked pentasaccharide to either a dimethylated hexose or a methylated hexuronic acid (Abu-Qarn et al., 2007). To determine how the absence of AglP affected N-glycosylation in *H. volcanii*, mass spectrometry was employed to compare the N-linked glycan profile of the S-layer glycoprotein in cells of the parent (**Figure 3.13** – parent - WR536 – lower panel) and those lacking the AglP producing gene (HVO_1522). The same peptide derived from the AglP deleted S-layer glycoprotein, shows a smaller glycan (**Figure 3.13** – *ΔaglP* – *ΔHVO_1522* – upper panel). The tetrasaccharide found on this S-layer glycoprotein, contains the Asn-linked hexose (m/z 1743) and the two hexuronic acids (m/z 1919 and m/z 2095, respectively). The 190Da species observed in the parent strain is replaced by a 176Da species (m/z 2271). The *aglP* deletion therefore results

in a glycan moiety of the mutant strain that does not only lack the final hexose, but also presents a subunit at position four that is 14Da smaller than the 190Da species in the parent strain at this position.

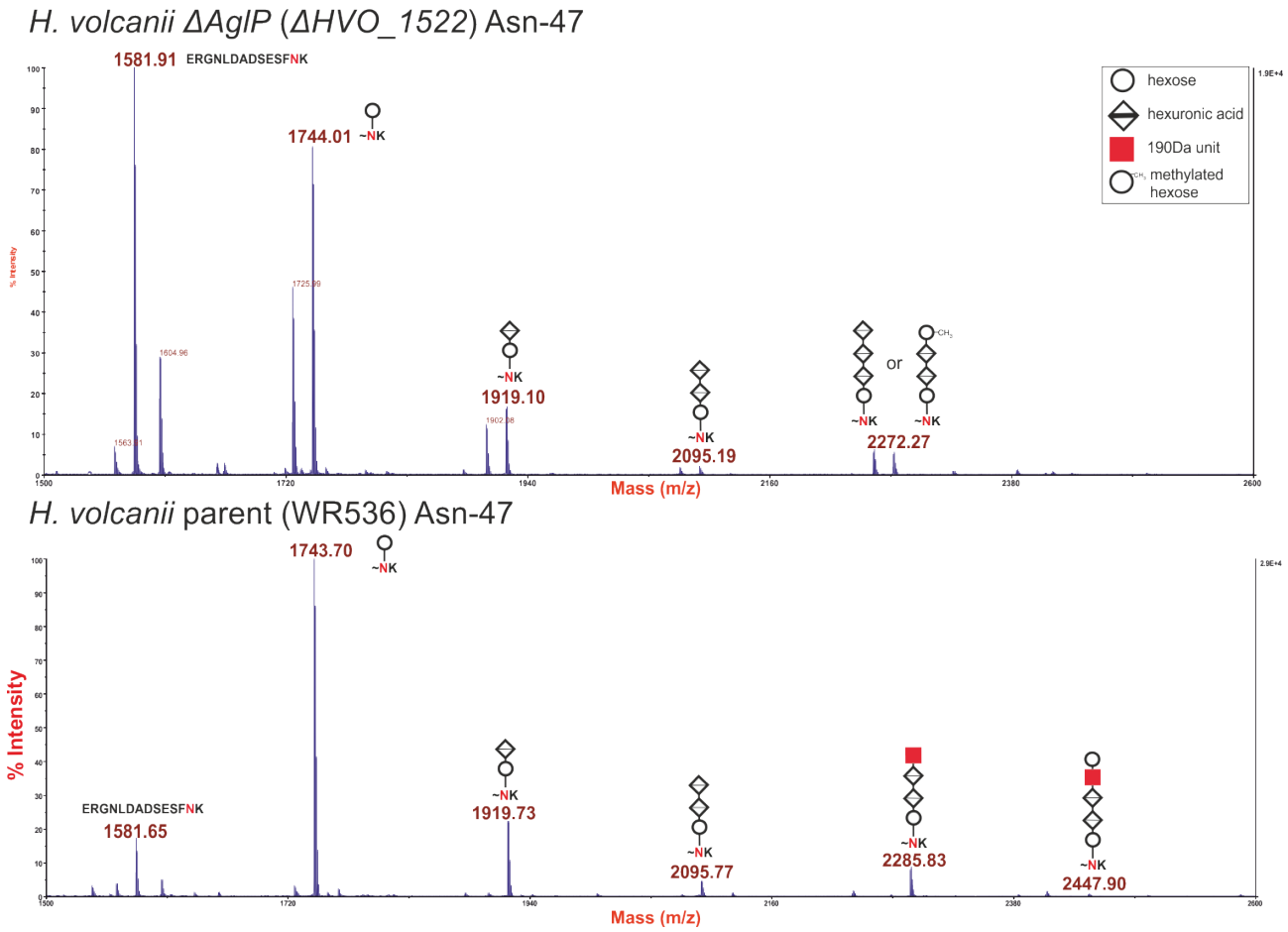


Figure 3.13 – MALDI-MS analysis of the Asn-47-containing tryptic glycopeptides (AglP).
The samples were derived from gel pieces containing the *H. volcanii* S-layer glycoprotein from the parent (WR536) and *aglP*-deleted strains (*ΔaglP – ΔHVO_1522*). The sample was analysed on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer after in-gel trypsin digestion and separation by an offline UltiMate 3000 LC system. Non-annotated signals are not glycopeptides.

As such, it would appear that in cells lacking AglP, the 176Da saccharide found at position four of the pentasaccharide is lacking a methyl group, as observed by the decrease in mass of this reporter glycopeptides by 14 mass units, yielding either methylated hexose or hexuronic acid. To distinguish between these possibilities, methyl esterification of the S-layer glycopepetides derived from the *aglP*-deleted cells was performed, as described in **Section 2.4.8.10** (Abu-Qarn et al., 2007). As a result of such treatment, carboxylic acid groups are converted to their methyl esters, reflected as a 14Da mass shift for each carboxylic acid moiety. The trisaccharide-bearing Asn-47-containing S-layer glycopeptide derived from the parent strain, showed a shift of m/z 98 for this peptide following esterification, explained by the transformation of the seven carboxylic acid groups found on the two Glu, the two Asp and the C-terminal residue of the peptide, as well as on the two hexuronic acids found at positions two and three of the N-linked saccharide (data not shown).

When methyl-esterification of the Asn-13-containing S-layer glycoprotein peptide derived from aglP-deleted cells was performed (**Figure 3.14**), the mono-, di- and trisaccharide-bearing peptides were shifted 70, 84 and 98Da, respectively, consistent with the same glycopeptides derived from the background strain, reflecting the

presence of five, six and seven target carboxylic groups, respectively. However, unlike the tetrasaccharide-bearing glycopeptide from the background strain, the m/z 2271 peak from the AglP-lacking cells was shifted by 112 mass units to m/z 2383, reflecting the presence of eight carboxylic moieties. This corresponds to an additional methyl group on the tetrasaccharide glycan derived from the AglP-lacking cells. It can thus be concluded that the 176 Da species found at position four of the tetrasaccharide linked to the S-layer glycoprotein derived from the aglP-deleted cells corresponds to a hexuronic acid. By extension, the 190Da saccharide found at position four of the N-linked pentasaccharide in the parent strain corresponds, therefore, to the methyl ester of a hexuronic acid (**Figure 3.14**).

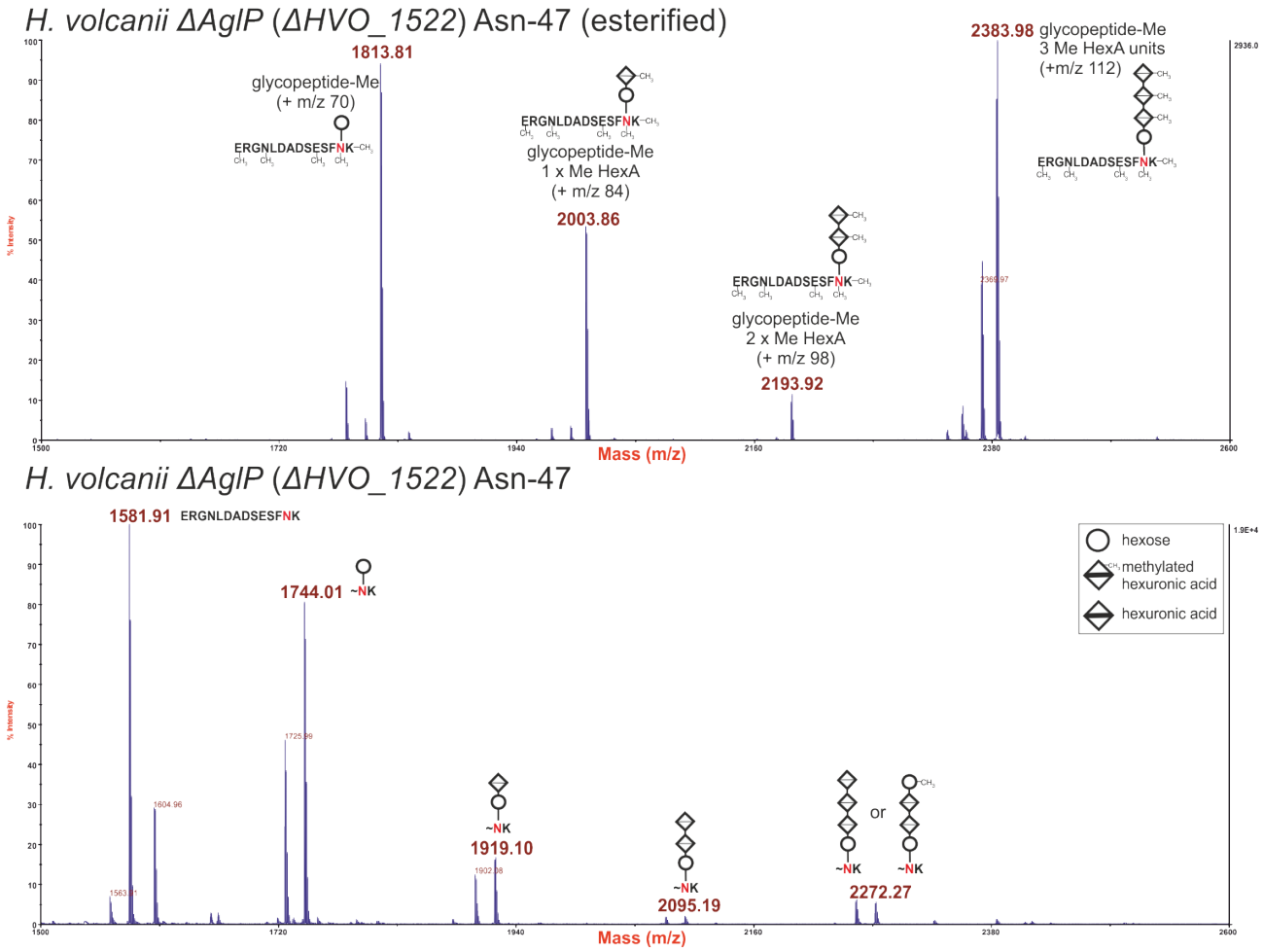


Figure 3.14 – MALDI-MS analysis of the Asn-47-containing tryptic glycopeptides (AglP - esterified). The samples were derived from gel pieces containing the *H. volcanii* S-layer glycoprotein from the *aglP*-deleted strains (*ΔaglP – ΔHVO_1522*) as well as the *aglP*-deleted strains (*ΔaglP – ΔHVO_1522*) post esterification by 1M methanolic acetylchloride. The sample was analysed on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer after in-gel trypsin digestion and separation by an offline UltiMate 3000 LC system and esterification. The clusters of signals below each of the annotated molecular ions are due to incomplete esterification of a portion of the sample under the conditions used and/or lactonization (intervals of 14 and 18Da, respectively).

3.4.3 Function of AglJ

To examine the effect of the absence of AglJ on N-glycosylation of the S-layer glycoprotein in *H. volcanii*, the glycan profile of the parent (**Figure 3.15** – parent - WR536 – lower panel) was compared with that of the S-layer glycoprotein in cells lacking the AglJ protein. Examination of the glycopeptide derived from cells lacking *aglJ* (**Figure 3.15** – *ΔaglJ – ΔHVO_1517* – upper panel) shows only a minor peak for the glycopeptides decorated with one attached hexose (m/z 1743), a peak, which is much smaller than that observed of the

Asn-linked hexose in the parent strain. Furthermore, there is a major peak corresponding to the nonmodified peptide (m/z 1581), whilst peaks corresponding to di-, tri-, tetra-, and pentasaccharide-modified peptides were completely absent in this sample.

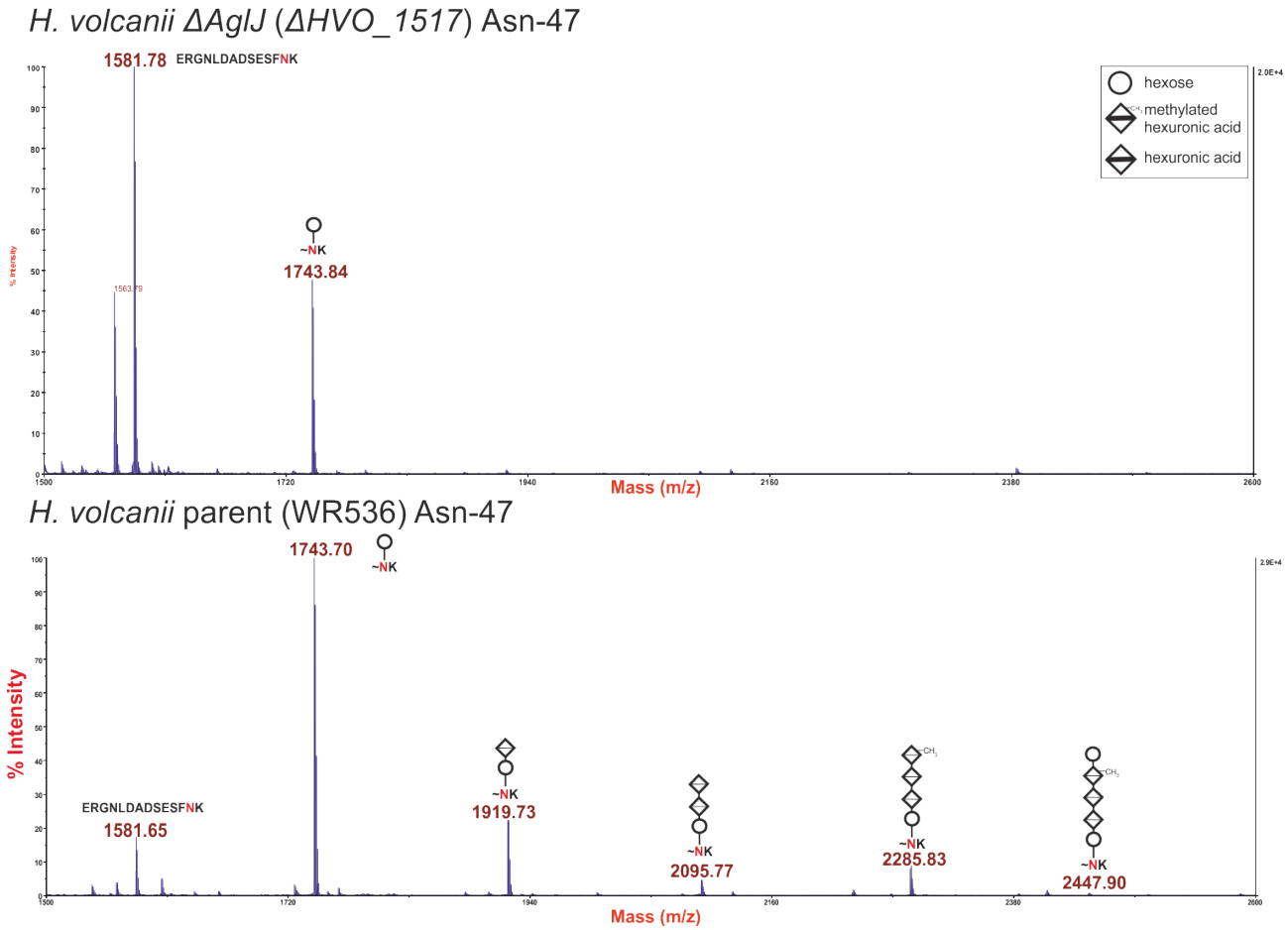


Figure 3.15 – MALDI-MS analysis of the Asn-47-containing tryptic glycopeptides (AglJ).
The samples were derived from gel pieces containing the *H. volcanii* S-layer glycoprotein from the parent (WR536) and *aglJ*-deleted strains (*ΔaglJ – ΔHVO_1517*). The sample was analysed on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer after in-gel trypsin digestion and separation by an offline UltiMate 3000 LC system. Non-annotated signals are not glycopeptides.

3.4.4 Summary discussion of functional characterisation of *agl* genes

For *H. volcanii*, AglB, AglD, AglE, AglF, AglG and AglI were shown to participate in the assembly and attachment of a pentasaccharide to select Asn residues of the S-layer protein, a reporter of N-glycosylation in this species. Specifically, AglG, AglI, AglE and AglD are thought to be glycosyltransferases involved in adding the second, third, fourth and fifth pentasaccharide subunits, respectively (Abu-Qarn et al., 2008b, Abu-Qarn et al., 2007, Yurist-Doutsch et al., 2008), whilst AglF was shown to be a glucose-1-phosphate uridyltransferase (Yurist-Doutsch et al., 2008). The functional characterisation of the three investigated *agl* genes, AglM, AglJ and AglP, expand the limints of the *agl* cluster to include AglM (HVO_1531), AglJ (HVO_1531) and AglP (HVO_1522). The gene encoding AglM was found in the open reading frame lying directly downstream to *aglB* and is currently the last gene of the cluster.

When assessing the product AglM from the open reading frame downstream from the *agl* gene cluster, it was found that AglM participates in the biosynthesis of hexuronic acids contained within the S-layer

pentasaccharide. In complementation to mass spectrometric studies, bioinformatics-based approaches and activity assays were carried out to suggest that AglM is a sugar nucleotide dehydrogenase, responsible for formation of the hexuronic acid located at position two of the pentasaccharide that decorates the *H. volcanii* S-layer glycoprotein. It was further shown that *aglB*, *aglF*, *aglG* and *aglI* show a coordinated transcription response in the face of various growth conditions, despite the fact that they do not share a common promoter (Yurist-Doutsch et al., 2010). This suggests that the *agl* gene cluster acts as a functional genomic unit, and like many other archaeal biochemical pathways, could imply that the N-glycosylation process of *H. volcanii* is regulated at the gene level.

Studies by our collaborators showed that AglM acts as a UDP-glucose dehydrogenase, able to transform UDP-glucose into UDP-glucuronic acid in a NAD^+-dependent manner across a wide range of salt concentrations. AglM was also able to utilize UDP-galactose or GDP-mannose as substrates, albeit to a much lesser extent than UDP-glucose. AglF, on the other hand, was shown to function as a glucose-1-phosphate uridyltransferase, catalyzing the biosynthesis of UDP-glucose from glucose-1-phosphate and UTP. Through the use of a coupled assay developed for this study, it was shown that the two enzymes are able to function in a sequential manner, forming UDP-glucuronic acid from glucose-1-phosphate and UTP, in the presence of NAD^+. The ability of AglM and AglF to function in unison implies that AglM is involved in the biosynthesis of the hexuronic acids found at both positions two and three of the pentasaccharide decorating the *H. volcanii* S-layer glycoprotein. By contrast, AglF only participates in the biogenesis of the hexuronic acid found at position three (Yurist-Doutsch et al., 2010)

Deletion of the *aglP*-encoding gene resulted in an S-layer glycoprotein modified by a tetrasaccharide rather than the pentasaccharide detected on the native protein. The fact that in the tetrasaccharide observed in the *ΔaglP* cells that contained a 176Da species, rather than the 190Da species found at position four of the pentasaccharide in the parent strain corresponded to the contribution of a methyl group (14Da difference). Previous studies had identified the 190Da species to correspond to either a dimethylated hexose or a methyl ester of hexuronic acid (Abu-Qarn et al., 2007). Methanolic acetylchloride based methyl esterification prior to MS analysis of the carboxylic groups was undertaken, confirming the presence of a methyl-ester of hexuronic acid at position four of the N-linked oligosaccharide. This indicates that AglP functions as a methyltransferase. With *in vitro* assays for AglP by our collaborators, the SAM-dependent methyltransferase activity of the protein was shown (Magidovich et al., 2010).

Moreover, it was found by our collaborators that AglP-mediated methylation occurs on the lipid-linked tetrasaccharide precursor of the pentasaccharide, which is eventually transferred to the S-layer glycoprotein. With the cytoplasmic location of AglP, it is further also unlikely that the methyltransferase modifies the oligosaccharide after its transfer to the S-layer glycoprotein, a process thought to occur on the external surface of the cell (Lechner and Wieland, 1989). On the other hand, the active site of AglD, responsible for adding the final hexose to the pentasaccharide transferred to the S-layer glycoprotein, was found to face the cytoplasm (Plavner and Eichler, 2008).

The methylation of a monosaccharide in the *H. volcanii* S-layer glycoprotein, is not the first example of such modification of an oligosaccharide and in *H. salinarum*, SAM-dependent methylation of a glucose subunit

was previously reported (Lechner et al., 1985). It was further shown that inhibition of such methylation prevented transfer of the oligosaccharide to its target protein. The work presented in this thesis shows that the *H. volcanii* S-layer glycoprotein is modified by a tetrasaccharide in cells deleted of *aglP* and hence incapable of oligosaccharide methylation, arguing against a role for methylation as a delivery mechanism (Magidovich et al., 2010).

The combination of the deletion of *aglJ* gene (HVO_1517) with mass spectrometric analysis, collected evidence showing the participation of AglJ in adding the first hexose to the pentasaccharide, as is shown by the smaller peak of the Asn-linked hexose. The deletion of *aglJ*, however did not fully prevent the attachment of this hexose residue to the S-layer glycoprotein. Furthermore, in cells lacking AglJ, no further sugar subunits were added to the monosaccharide-modified S-layer glycoprotein, pointing to the importance of the sugar added through the actions of AglJ for proper N-glycosylation (Kaminski et al., 2010).

The following Agl proteins have been added to the repertoire of proteins found to be involved in the assembly of the pentasaccharide (**Table 3.2**) and the current working model of the *H. volcanii* S-layer glycoprotein N-glycan biosynthesis was extended with information of the work described in this thesis and is shown in **Figure 3.16**.

Encoded Protein	Suggested function
AglB	oligosaccharyltransferase
AglD, AglE, AglG, AglI	glycosyltransferases
AglF	glucose-1-phosphate uridyltransferase
AglM	**addition of hexuronic acid (position 2, 3) - dehydrogenase**
AglP	**modification of 190Da species - methyltransferase**
AglJ	**addition of hexose (position 1) - glycosyltransferase**

Table 3.2 – Current understanding of Agl proteins involved in the assembly of the pentasaccharide.

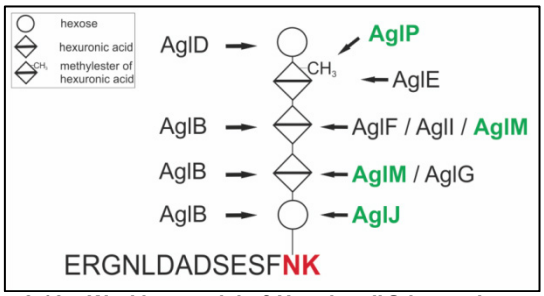


Figure 3.16 – Working model of *H. volcanii* S-layer glycoprotein

3.5 S-LAYER SEQUENCE PROFILING AND N-GLYCOSYLATION SITE OCCUPANCY

The *H. volcanii* S-layer glycoprotein amino acid sequence was initially investigated and published by Sumper and colleagues who analysed the purified and enzymatically digested S-layer glycopeptides by an automated gas phase sequencer (Sumper et al., 1990). In this work, in order to confirm the protein sequence published and identify the site specific occupation of N-glycosylation sites, a comprehensive map of the *H. volcanii* S-layer glycoprotein was generated using a series of biochemical techniques coupled with high sensitivity MALDI-MS, MALDI-MS/MS, ESI-MS, ESI-MS/MS as well as MS^E methods. The S-layer glycoprotein sequence is 827 amino acids in size, prior to signal peptide cleavage of the first 34 amino acids,

creating a mature polypeptide of 793 amino acids, with an expected calculated molecular weight of circa 81'732Da.

A range of enzymatic digests were performed in the aim to produce suitably sized peptide fragments accessible for characterisation by both MALDI and ESI ionisation for manual annotation of the peptide sequences. The enzymes used in the production of all datasets included: trypsin, chymoptrypsin, Glu-C (V8), AspN, prolidase and elastase (**Section 2.4.3**). For chemical digestion cyanogen bromide was also used (**Section 2.4.3.8**). These enzymatic and chemical degradation procedures were used separately or in tandem. The resultant mixture of peptide fragments from each of the single or double digests investigated was submitted to purification using offline or online liquid chromatography separation methods prior to analysis by MALDI-MS and MS/MS as well as ESI-MS, ESI-MS/MS and MS^E either at Imperial College or at the MS-RTC facility in St. Saviour, Jersey (**Figure 3.17**).

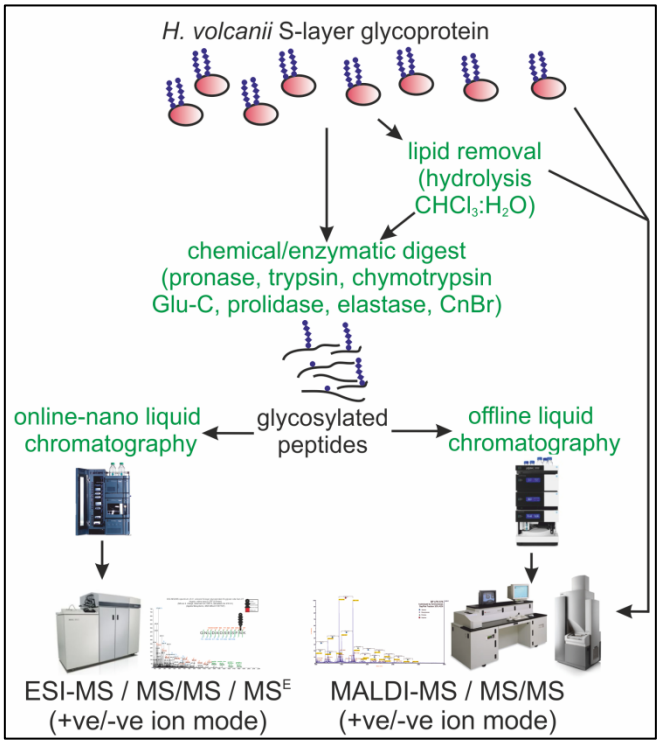


Figure 3.17 – Methods employed in mapping the S-layer glycoprotein sequence and glycosylation.

3.5.1 S-layer glycoprotein coverage by MALDI-MS and ESI-MS using various proteolytic digests

To allow for sequencing of the whole S-layer glycoprotein the lyophilised sample sent by the Eichler group was dissolved in H_2O and slowly adjusted to the required pH for enzymatic and treatment, due to the ease with which the protein precipitated and its absolute insolubility at acidic pH. The *H. volcanii* S-layer sample was further digested in solution, with a series of enzymes in the appropriate pH and protein:protease concentration as required for optimal enzyme activity. The enzymatic digests were commonly incubated at 37°C overnight, unless otherwise stated (**Section 2.4.3**). Due to the absence of cysteines in the protein sequence, the samples were not reduced or carboxymethylated. The published protein sequence was *in silico* digested using MassLynx to predict the molecular ions created by the specific enzymes employed in

the elucidation of the peptide map, giving insight into the expected molecular ions assessed by mass spectrometry and aiding with manual annotation of spectra.

The *H. volcanii* S-layer glycoprotein was digested with trypsin, hydrolysing the peptide bonds at the carboxyl side of lysine and arginine residues (except if followed by a proline residue). Bearing in mind the published S-layer protein sequence, digestion with trypsin produced rather large peptides and was therefore also coupled with Glu-C enzyme, cleaving at the C-terminus of glutamic acid and at a 300-fold lower rate also aspartic acid, to create a second dataset of smaller, more amenable peptides for analysis in a second series of experiments. Further digestion by chymotryspin was achieved for MALDI and ESI analysis. Chymotrypsin selectively hydrolyses the peptide bonds on the C-terminal side of tyrosine, phenylalanine, tryptophan and leucine and secondary hydrolysis occurs on the C-terminal side of methionine, histidine and asparagine, thus creating a series of smaller peptides that are amenable to mass spectrometric ionisation. Finally the S-layer glycoprotein was also digested using AspN enzyme, a zinc metalloendopeptidase, which selectively cleaves peptide bonds on the N-terminal side of aspartic acid residues, producing small peptides allowing for information on N-glycosylation.

Upon encountering difficulties in sequencing the C-terminal region of the S-layer glycoprotein, it was digested with prolidase and elastase due to the recurring presence of their substrate residues in the C-terminal region. Prolidase is a cytosolic exopeptidase that cleaves on the N-terminal side of proline of small peptides. The requirement of smaller peptides warranted for the sample to first be digested with Glu-C and then further digested with prolidase. Furthermore, elastase was used as a protease that cleaves specific peptide bonds on the carboxyl side of small hydrophobic amino acids such as glycine, alanine and valine. In addition, the C-terminal region of the S-layer glycoprotein was investigated using CNBr, which hydrolyses peptide bonds at the C-terminus of methionine residues, except when followed by a threonine residue, or if methionine is oxidised. CNBr digested peptides are generally quite large, with the exception of the N-terminal, as suggested by MassLynx information from *in silico* digestion and the digest was both performed by CNBr only as well as on a previously trypsin digested sample.

The digests were initially concentrated and desalted using Michrom MicroTraps™ and run in a solution of 40µl 0.1% TFA, where a 1µl injection was used for offline LC-MALDI-MS and MS/MS analysis. However, upon comparison between samples desalted and purified using the michrom microtraps and those separated by only LC-MS, this technique was abandoned in favour of offline and online LC-MS. The samples were generally assessed in the positive ion mode using both MALDI and ESI techniques, except in the case of AspN, where the negative ion mode was used additionally for data collection (data not shown). The sugar residues identified in such glycoproteomic experiments are analysed underivatised and therefore retain their underivatised mass, which differ from the masses of the same permethylated monosaccharides (**Table 1.4**).

Upon separation by offline liquid chromatography and analysis by MALDI-MS and MALDI-MS/MS, the spectra generated were evaluated by Mascot, a program allowing for identification of constituent proteins by database comparisons of raw mass spectrometric data against molecular sequence databases. This was done to verify the protein analysis and exclude contaminations throughout manual processing. All enzymatic digests were correctly identified as S-layer glycoproteins of *H. volcanii*, thus ruling out the presence of

contaminating glycoproteins, although manual interpretations did later find some evidence of minor amounts of other *H. volcanii* protein sequences. The rest of the mass spectra were manually assessed for further sequence information and compiled. **Figure 3.18** shows all amino acid residues and sequences identified from the chymotryptic digest. An example of MALDI-MS/MS annotation for the peptide sequence ^{283}QNDATTEQAKEVF^{295}, from the molecular ion at 1480.7 is shown in **Figure 3.19**, which allowed for the identification of the sequence by annotation of the mass differences between two consecutive ions within a series (eg. b-ions and y" ions).

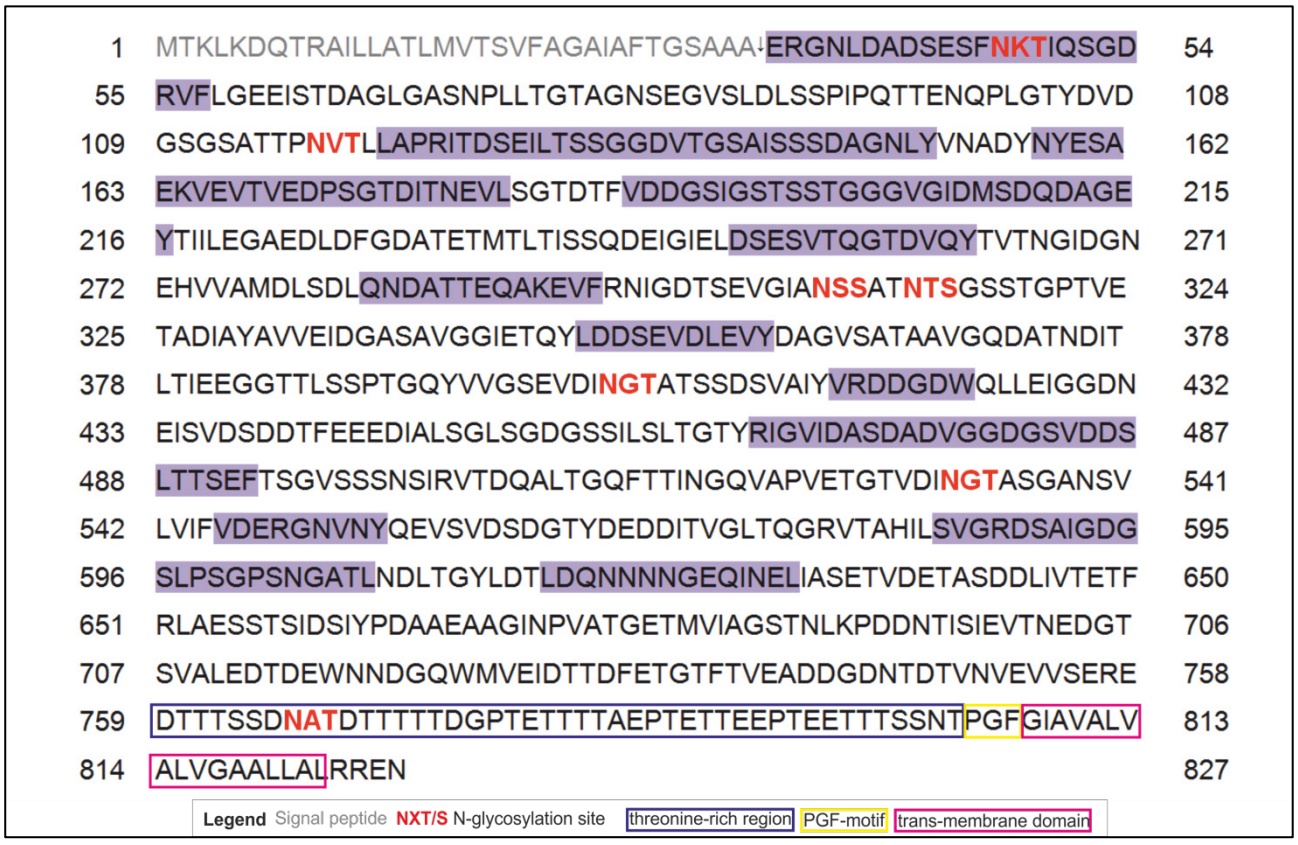


Figure 3.18 – S-layer mapping coverage by MALDI-MS/MS post chymotryptic digest.
Partial mapping coverage of *H. volcanii* S-layer glycoprotein by chymotryptic digest, where the identified peptide sequences are highlighted in purple. Separation was achieved by offline liquid chromatography and analysis by MALDI-MS and MALDI-MS/MS.

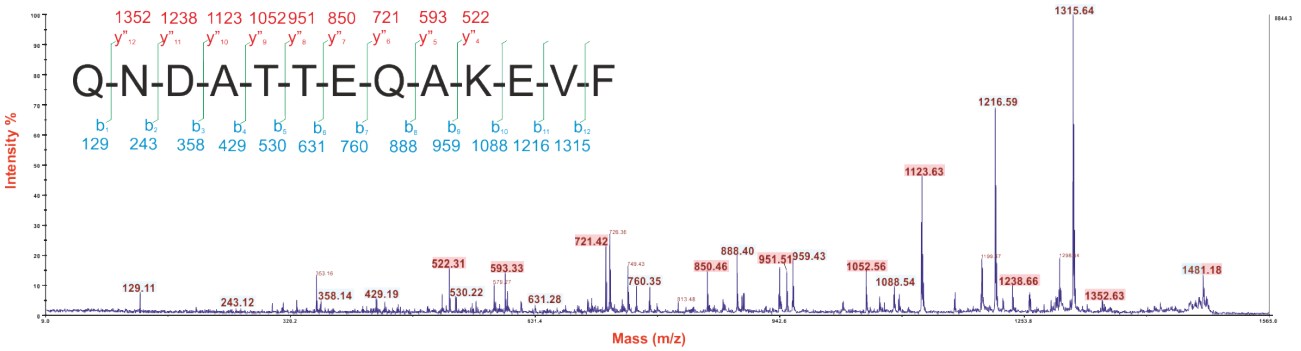


Figure 3.19 – MALDI-MS/MS spectrum of the molecular ion at m/z 1480.7.
The S-layer glycoprotein was digested with chymotrypsin and separated by an offline UltiMate 3000 LC system and assessed by MALDI-MS and MALDI-MS/MS on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer. The peptide sequence assignment of ^{283}QNDATTEQAKEVF^{295} is shown in the schematic. Fragment b and y" ions are labelled in blue and red respectively.

MALDI-MS and MALDI-MS/MS analysis for obtaining peptide sequence information was limited to the fact that generally only singly charged molecular ions are generated, thus restricting the number of molecular ions generated from which information was to be gained by MALDI-MS/MS. With this in mind, the enzymatic digests were further analysed by online nano-LC ESI-MS and data-dependent MS/MS. The MS and MS/MS data across the elution profiles were then manually examined for information on peptide sequences corresponding to *in silico* expected fragments either glycosylated or unglycosylated, where all spectra exhibiting putative carbohydrate fragment ions were scrutinised for the presence of molecular ions corresponding to glycopeptides.

As a feature of the ESI ionisation process, the formation of a series of multiply charged ions is common, creating molecular ions of the form $[M+nH]^{n+}$ or $[M-nH]^{n-}$. The charge state of each ion can be used to determine the actual mass of the peptides or glycopeptides instead of their m/z ratio. Whereas the isotopic peak distribution will be identical to the one of singly-charged peaks in MALD-MS, multiply-charged ions on the other hand, will exhibit different m/z increments according to the charge state, such that doubly, triply or quadruply charged isotopic peak distribution will show m/z increments between isotopes of m/z 0.5, 0.33 or 0.25, respectively. The relative molecular weight (M_r) of the identified peptides from individual peaks with a range of proton distributions allows for identification of different peaks with the same composition, yet differently charged. The various enzymatic digest preparations of the S-layer glycoprotein were further manually assessed for sequence information by online nano-LC-ESI-MS/MS data.

Figure 3.20 – Mapping coverage of all enzymatic digests.
Complete comprehensive mapping coverage of *H. volcanii* S-layer glycoprotein by enzymatic digests, including trypsin, chymotrypsin, trypsin/Glu-C, prolidase, elastase, CNBr and trypsin/CNBr, separation by offline/online liquid chromatography and analysis by MALDI-MS and MALDI-MS/MS or online nano-ESI-MS and ESI-MS/MS.

The data and information from all enzymatic digests as analysed by offline LC-MALDI-MS/MS and online nano-ESI-MS/MS techniques were collated and **Figure 3.20** shows the S-layer glycoprotein amino acid sequence highlighting the combined information map as derived from enzymatic tryptic, chymotryptic, double tryptic and Glu-C as well as elastase and prolidase digests and chemical digestion by CNBr and trypsin, followed by CNBr methods. The N-terminal signal sequence was not mapped, as it is cleaved prior to the mature status of the sample extracted by the Eichler group. The map in **Figure 3.20** shows roughly 90% of the protein mapped, including six of the seven putative N-glycosylation sites (**Section 3.5.2**). The C-terminal region of the sequence including the threonine rich region that is thought to be O-glycosylated as well as the final N-glycosylation site were not able to be described at the time of thesis submission. This was, despite extensive investigations directed at detecting C-terminal peptides taking into account possible glycan and lipid modifications, which were perhaps obscuring access to the C-terminal region. Some of these experiments have called into question the proposed lipid modification of the S-layer (described in **Section 3.6**).

3.5.2 Comprehensive map of N-glycosylation sites of *H. volcanii* S-layer

Work carried out in **Section 3.4** for the functional characterisation of the *agl* genes in the *agl* gene cluster of *H. volcanii* only focused on the first of the seven putative N-glycosylation sites and throughout previous investigations into the S-layer glycoprotein and its N-glycan, only two of the seven putative N-glycosylation sites were found to be occupied. This was possibly due to the ease with which the peptides containing these sites could be isolated and investigated and the focus of these studies, which lay with the identification of the N-glycan decorating these glycosylation sites, rather than the sites themselves. The N-linked glycosylation sites were shown to be occupied by a pentasaccharide N-glycan containing two hexoses, two hexuronic acids and a methylated hexuronic acid, which was found at Asn-47 and Asn-117, whilst Asn-404 was only identified as an unglycosylated peptide (Abu-Qarn et al., 2007). Prior to the work in this thesis, no studies had been undertaken to create a comprehensive map of these seven putative N-glycosylation sites. This was undertaken as described in this section with the aim to get a broad understanding on the N-glycosylation site occupancy and the heterogeneity of glycan occupation in the *H. volcanii* S-layer glycoprotein.

With the S-layer glycoprotein mapping data achieved using a series of enzymatic digests and assembled using MALDI-MS and ESI-MS techniques, information on the N-glycosylation sites on their respective peptide fragments could be gathered similarly (**Section 3.5.1**). Within the mapped protein sequence of the S-layer glycoprotein, which includes the first six putative N-glycosylation sites, all have been found to be occupied by the full pentasaccharide N-glycan. A selection of tables are listed here to represent the findings from enzymatic digests, including chymotrypsin and trypsin followed by Glu-C as analysed by ESI-MS and ESI-MS/MS and AspN data accumulated by MS^E analysis, to summarise the relevant glycan occupation of significant N-glycosylation sites (**Table 3.3**, **Table 3.4** and **Table 3.5**)

A selection of relevant ESI-MS/MS spectra are further presented here to show N-glycosylation site occupation and annotation of the peptide sequences and the respective glycan moieties (**Figure 3.21** and **Figure 3.22**). ES-MS/MS and high energy MS^E spectra provide definitive evidence for glycosylation and also offer additional information about the nature of the peptide backbone, such as is seen from the b and y" ion

signals annotated in the spectra in blue and red respectively. Further indications of glycosylation come from the presence of major low mass singly charged signals corresponding to carbohydrate fragments, however such signals are limited to well stabilised fragments from carbohydrate monosaccharides, with ionisable functional groups, such as HexNAc.

***H. volcanii* S-layer glycoprotein digested by chymotrypsin (ESI-MS and ESI-MS/MS)**					
RT (min)	**Observed (m/z)**	**Charge**	**Calculated MW**	**Assignment**	**Glycosylation**
26.4	1066.0	2	2130.0	47-57	$Hex_2.HexA_2.MeHexA$
24.8	984.9	2	1967.8	47-57	$Hex.HexA_2.MeHexA$
23.9	889.9	2	1777.8	47-57	$Hex.HexA_2$
22.1	801.9	2	1601.8	47-57	Hex.HexA
65.4	1170.2	3	3507.6	106-131	*$Hex_2.HexA_2.MeHexA$*
63.4	1116.2	3	3345.6	106-131	$Hex.HexA_2.MeHexA$
52.5	1726.0	3	5175.0	296-330	*$(Hex_2.HexA_2.MeHexA)$ x 2*
51.4	1672.0	3	5013.0	296-330	*$Hex_2.HexA_2.MeHexA$, $Hex.HexA_2.MeHexA$*
49.9	1618.0	3	4851.0	296-330	$(Hex.HexA_2.MeHexA)$ x 2
77.3	930.1	3	2787.3	396-416	$Hex.HexA_2.MeHexA$
53.9	1776.8	2	3551.6	515-542	*$Hex_2.HexA_2.MeHexA$*
52.5	1695.8	2	3389.6	515-542	*$Hex.HexA_2.MeHexA$*

Table 3.3 – List of molecular ions observed in S-layer glycopeptides digested by chymotrypsin.
Summary of the observed m/z values used to calculate the relative molecular weight of the peptides and glycan moieties as corroborated from ESI-MS/MS data. The glycosylation annotations in italics were derived from MS data only.

***H. volcanii* S-layer glycoprotein digested by trypsin and Glu-C (ESI-MS and ESI-MS/MS)**					
RT (min)	**Observed (m/z)**	**Charge**	**Calculated MW**	**Assignment**	**Glycosylation**
15.1	681.2	2	1360.4	45-48	$Hex_2.HexA_2.MeHexA$
84.5	1515.7	5	7573.5	56-124	$Hex.HexA_2.MeHexA$
47.6	994.8	3	2981.4	99-124	$Hex.HexA_2$
50.0	1053.5	3	3157.5	99-124	Hex.HexA
50.8	1116.8	3	3347.4	99-124	$Hex.HexA_2.MeHexA$
52.6	1170.9	3	3509.7	99-124	$Hex_2.HexA_2.MeHexA$
33.0	1835.2	2	3668.4	304-324	*$(Hex_2.HexA_2.MeHexA)$ x 2*
32.5	1754.1	2	3506.2	304-324	*$Hex_2.HexA_2.MeHexA$, $Hex.HexA_2.MeHexA$*
31.3	1673.1	2	3344.3	304-324	*$(Hex.HexA_2.MeHexA)$ x 2*
82.9	1328.0	4	5308.0	505-549	$Hex.HexA_2.MeHexA$
82.9	1368.5	4	5470.0	505-549	$Hex_2.HexA_2.MeHexA$

Table 3.4 – List of molecular ions observed in S-layer glycopeptides, digested by trypsin and Glu-C.
Summary of the observed m/z values used to calculate the relative molecular weight of the peptides and glycan moieties as corroborated from ESI-MS/MS data. The glycosylation annotations in italics were derived from MS data only.

***H. volcanii* S-layer glycoprotein digested by AspN (MS^E)**					
RT (min)	**Observed (m/z)**	**Charge**	**Calculated MW**	**Assignment**	**Glycosylation**
2.9	1008.9	2	2015.8	42-53	$Hex.HexA_2.MeHexA$
2.9	1089.9	2	2177.8	42-53	$Hex_2.HexA_2.MeHexA$
6.8	1288.1	2	2574.1	108-126	$Hex.HexA_2.MeHexA$
6.8	1369.1	2	2736.2	108-126	$Hex_2.HexA_2.MeHexA$
5.8	1975.2	2	3948.4	300-326	$(Hex.HexA_2.MeHexA)$ x 2
0.9	1569.85	1	1568.5	402-410	$Hex_2.HexA_2.MeHexA$
1.2	1731.6	1	1730.6	402-410	$Hex_2.HexA_2.MeHexA$
12.0	1191.0	2	2380.1	530-546	$Hex.HexA_2.MeHexA$
12.3	1272.1	2	2542.1	530-546	$Hex_2.HexA_2.MeHexA$

Table 3.5 – List of molecular ions observed in S-layer glycopeptides, digested by AspN.
Summary of the observed m/z values used to calculate the relative molecular weight of the peptides and glycan moieties on the specific peptides as corroborated from MS^E data.

Figure 3.21 shows the ESI-MS/MS spectrum of the molecular ion found at m/z 1081.9^{2+}, a doubly charged signal selected from the MS data, showing the tryptic peptide ^{37}GNLDADSESFNK^{48}, encompassing the first putative N-glycosylation site at Asn-47. The calculated relative molecular weight of this doubly charged molecular ion is 2161.8 mass units and with the peptide in question only having a theoretical mass of 1295.6 mass units, this peptide is glycosylated with the pentasaccharide (Hex_2.$HexA_2$.HexAMe at a mass of m/z 866). Using the b and y" ions ions generated, the signals can be mapped to the N- and C-terminal of this tryptic glycopeptide. Furthermore, the spectrum in **Figure 3.21** allows for annotation of the N-glycan monosaccharide units as labelled in green.

Figure 3.22 shows the ESI-MS/MS spectrum of the molecular ion with an m/z observed at 1170.9^{3+}, a triply charged signal selected from the MS data, showing the doubly digested peptide ^{99}NQPLGTYDVDGSGSATTPNVTLLAPR^{124} after enzymatic treatment with trypsin and Glu-C enzyme, displaying two missed cleavages. This peptide contains the second putative N-glycosylation site at Asn-117. The calculated relative molecular weight of this triply charged MS ion is 3509.7 mass units and with the peptide in question only having a theoretical mass of 2643.9 mass units, this peptide is glycosylated with the pentasaccharide (Hex_2.$HexA_2$.HexAMe at a mass of m/z 866). Using the b and y" ions generated, the signals can be mapped to the N- and C-terminal of this S-layer glycopeptide, whilst the N-glycan monosaccharide units are seen attached to the stable proline y" ion at m/z 980.6 and are labelled in green.

The N-glycosylation sites Asn-308 and Asn-313 are not shown here, as MS/MS data for these sites could only be obtained as per their occupation by the first four subunits of the N-glycan pentasaccharide, and thus only show tetrasaccharide occupation. However, the occupation of these sites by the pentasaccharide was clear from the MS data, as highlighted in italics in the tables above (**Table 3.3** and **Table 3.4**). This is often the case for glycopeptides terminating in simple sugars such as hexose, where the hexose is partially lost due to the internal energy uptake on ionisation and the automated MS/MS software then choses the more abundant des-Hex signal for MS/MS analysis.

MS^E technology allows for the generation of fragment ions on every sample moiety or molecular ion detected simultaneously, thus recording exact mass precursor and fragment ion information, whilst at the same time obtaining accurate quantitative profiles from every detectable component in any sample. This is achieved by rapid alternating between two functions and recording two channels of data continuously. The first channel at low energy acquires low energy precursor ion spectra, such as is normally seen in MS spectra, whilst the second high energy channel complemented with collision energy, records elevated energy fragment ion spectra, similar to MS/MS spectra. MS^E records data without discrimination and every mass in the sample is measured. The spectra for each component can be aligned in time, providing comprehensive data and quantitative information on the sample.

Further information was derived from MS^E analysis of an AspN digest. This was done in both the positive and negative ion modes allowing for the assignment of N-glycosylation on all first six N-glycosylation sites in the protein (**Table 3.5**). Presented here are two spectra of the N-glycosylation sites Asn-404 (**Figure 3.23**), acquired at high energy, and Asn-532 for which low energy MS^E data is shown (**Figure 3.24**).

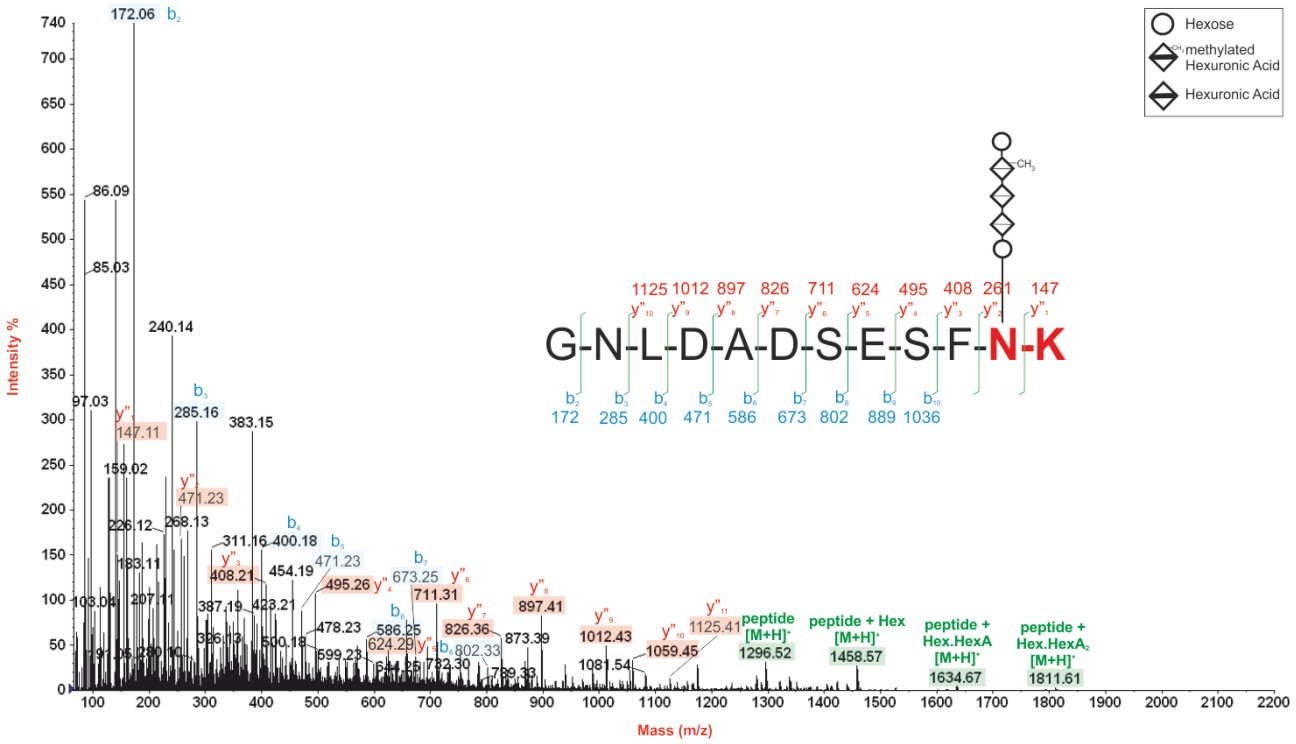


Figure 3.21 – ESI-MS/MS spectrum of m/z 1081.9^{2+} showing the S-layer N-glycan site Asn-47.
The ESI-MS/MS spectrum allows for peptide sequence assignment of ^{37}GNLDADSESFNK^{48}, such as is represented in the schematic, whilst b and y" ions are respectively labelled and highlighted in blue and red. Signals for the N-glycan units decorating this peptide are labelled in green. The data was accumulated post tryptic digest followed by separation on a reverse-phase nano-HPLC system and analysis on an Applied Biosystems API Q-STAR® Pulsar mass spectrometer. The data was accumulated in the positive ion mode.

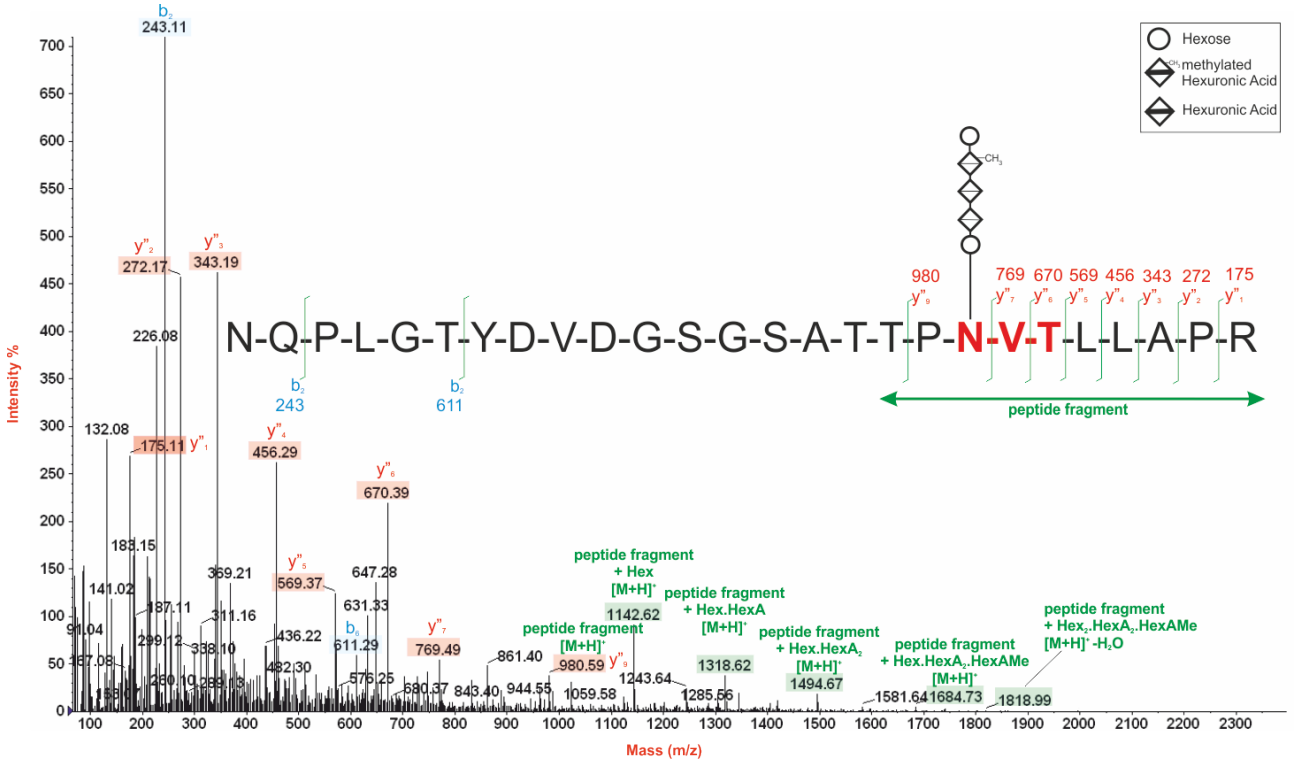


Figure 3.22 – ESI-MS/MS spectrum of m/z 1170.9^{3+} showing the S-layer N-glycan site Asn-117.
The ESI-MS/MS spectrum allows for peptide sequence assignment of ^{99}NQPLGTYDVDGSGSATTPNVTLLAPR^{124}, such as is represented in the schematic, whilst b and y" ions are respectively labelled and highlighted in blue and red. Signals for the N-glycan units decorating this peptide are labelled in green. The data was accumulated post tryptic and Glu-C digest followed by separation on a reverse-phase nano-HPLC system and analysis on an Applied Biosystems API Q-STAR® Pulsar mass spectrometer.The data was accumulated in the positive ion mode.

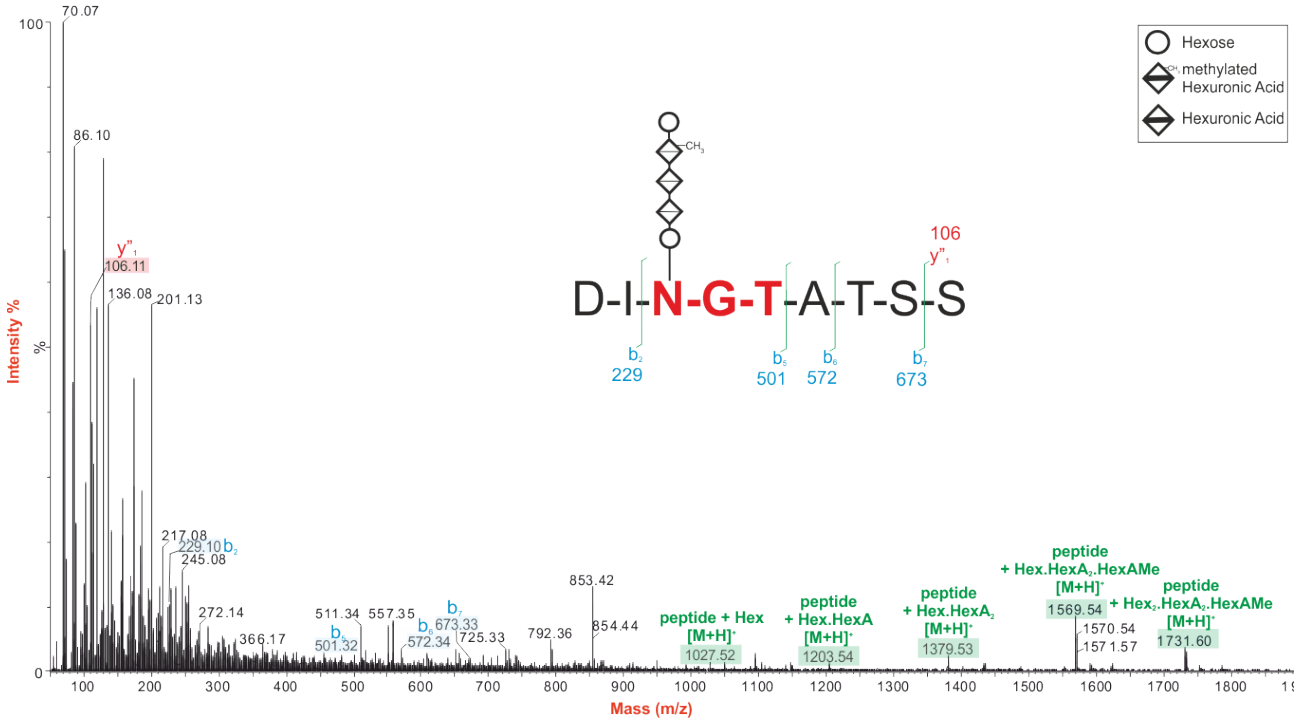


Figure 3.23 – High energy MS^E spectrum of m/z 1731.6$^+$ showing the S-layer N-glycan site Asn-404.
The MS^E high energy spectrum allows for peptide sequence assignment of ^{402}DINGTATSS^{410}, such as is represented in the schematic, whilst b and y" ions are respectively labelled and highlighted in blue and red. Signals for the N-glycan units decorating this peptide are labelled in green. The data was accumulated post AspN digest followed by separation on a Waters Acquity UPLC with microbore reverse phase column and analysis on a Waters Xevo G2S Q-TOF mass spectrometer. The data was accumulated in the positive ion mode.

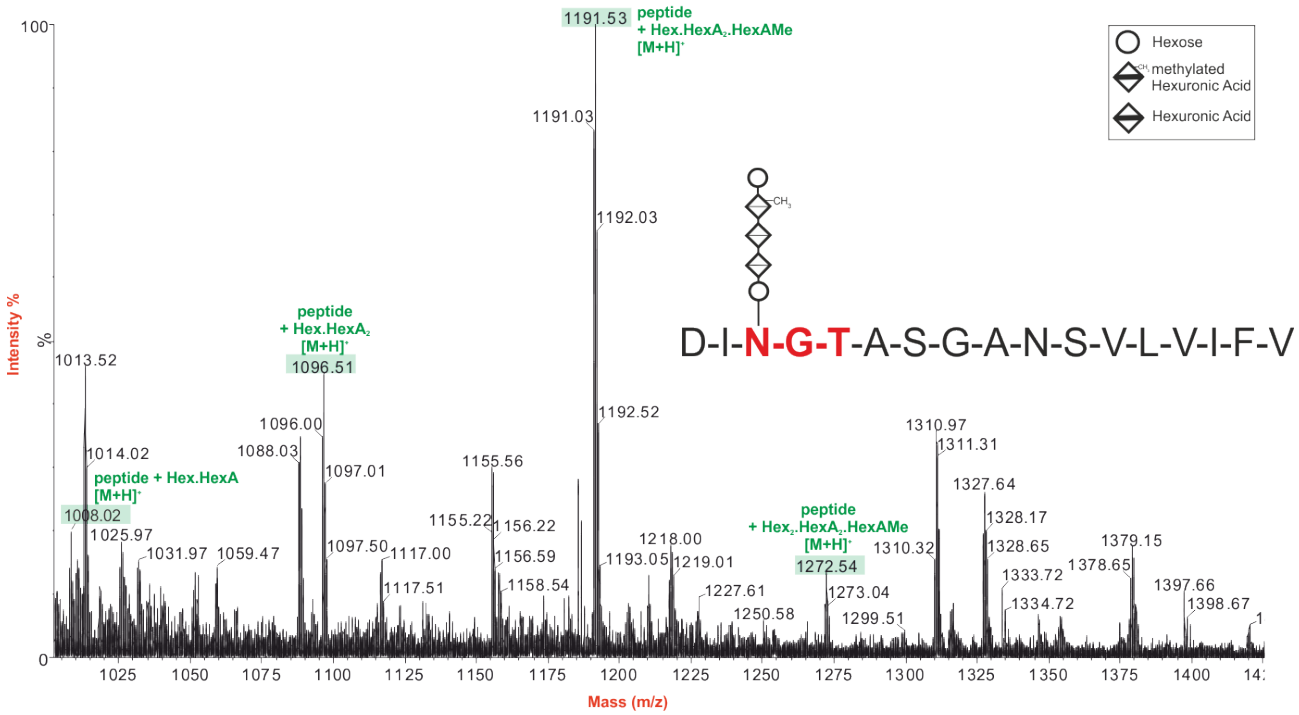


Figure 3.24 – Low energy MS^E spectrum showing the S-layer N-glycan site Asn-532.
The MS^E low energy spectrum is expanded around the signals representing the N-glycosylation site Asn-532 with m/z 1272.1^{2+} showing the fully occupied peptide sequence ^{530}DINGTASGANSVLVIF^{546} with the pentasaccharide N-glycan. This is represented in the schematic, whilst signals for the N-glycan units decorating this peptide are labelled in green. The data was accumulated post AspN digest followed by separation on a Waters Acquity UPLC with microbore reverse phase column and analysis on a Waters Xevo G2S Q-TOF mass spectrometer. The data was accumulated in the positive ion mode.

Figure 3.23 shows the high energy MS^E spectrum of the molecular ion with an m/z observed at 1731.6, a singly charged molecular ion, showing the peptide ^{402}DINGTATSS^{410} after enzymatic treatment with AspN in the positive ion mode. This peptide contains the fifth putative N-glycosylation site at Asn-404. The calculated relative molecular weight of this molecular ion is 1730.6 mass units and with the peptide in question only having a theoretical mass of 864.4 mass units, this peptide is shown to be glycosylated with the pentasaccharide (Hex_2.$HexA_2$.HexAMe at a mass of m/z 866). Annotation of the b and y" ions generated, as labelled in blue and red respectively, allow for these signals to be mapped to the N- and C-terminal of this glycopeptide, whilst annotation of the N-glycan monosaccharide units are labelled in green.

A low energy MS^E spectrum is shown in **Figure 3.24**, exhibiting assignment of the S-layer glycopeptide ^{530}DINGTATSGANSVLVIFV^{546}, as digested by AspN in the positive ion mode. The spectrum presented is expanded around the doubly charged molecular ions, labelled and highlighted in green, representing molecular ions of this peptide glycosylated with various intermediates of the N-glycan pentasaccharide (Hex_2.$HexA_2$.HexAMe). This spectrum demonstrates the sixth putative N-glycosylation site at Asn-532 to be glycosylated with the full pentasaccharide moiety.

3.5.3 Summary discussion of the comprehensive mapping of the S-layer glycoprotein

The results presented in this section confirm the amino acid sequence published previously by Sumper and colleagues, as far as could be identified by mass spectrometric techniques of approximately 90% of the S-layer glycoprotein. Previously only two of the seven putative N-glycosylation sites were proposed to be occupied by the N-glycan pentasaccharide, Asn-47 and Asn-117, whilst Asn-404 was suggested to be unoccupied (Abu-Qarn et al., 2007). However, the research presented here shows the occupation of the first six N-glycosylation sites, comprising Asn-47, Asn-117, Asn-308, Asn-313, Asn-404 and Asn-532. The C-terminal region was not characterised, despite various approaches and thus, the final N-glycosylation site cannot be commented on. All occupied N-glycosylation sites, were found to be glycosylated with the full N-glycan pentasaccharide, although there was also some evidence of shorter structures occurring naturally in the sample that were not generated by MS decomposition.

All N-glycosylation sites were found to be occupied by the pentasaccharide as well as its biosynthetic precursors. From overall assessment and comparison of the MS data accumulated throughout this series of experiments, the various intermediates of the N-glycan pentasaccharide (Hex_2.$HexA_2$.HexAMe) were found to be amongst the glycoforms of the S-layer. However it seems that occupation with the full N-glycan pentasaccharide is more difficult to achieve than its tetrasaccharide precursor (Hex.$HexA_2$.HexAMe). The last hexose of the pentasaccharide was recently identified as mannose and was found to be added onto the growing N-glycan from a separate dolichol carrier, whereas the previous four monosaccharide moieties making up the N-glycan pentasacharide are assembled, modified and transferred from a distinct dolichol carrier *en bloc* (Cohen-Rosenzweig et al., 2012). The matter of this observation of heterogeneity of the N-glycan of the *H. volcanii* S-layer glycoprotein is further discussed at the end of this chapter (**Section 3.7.2**).

Despite many efforts, the C-terminal domain was not observed and the S-layer glycoprotein sample was therefore investigated for potential modifications such as lipids.

3.6 STUDIES ON S-LAYER GLYCOPROTEIN LIPID ASSOCIATION

With the mapping of the S-layer glycoprotein resulting in a comprehensive map only short of the C-terminal region, further investigation was made into the reasons for this absence. This section describes the experimentation undertaken to seek information to confirm or discount the suggested lipid modification of the S-layer glycoprotein.

The sample provided as extracted by the Eichler group was investigated for possible lipid moieties and if discovered, any information on a possible covalent linkage. In Eichler's laboratory the S-layer glycoprotein was released using EDTA, to inhibit protease activity, after washing of the cell pellets with 2M NaCl, 50mM Tris–HCl at pH 7.2 (**Section 2.3.1.2**).

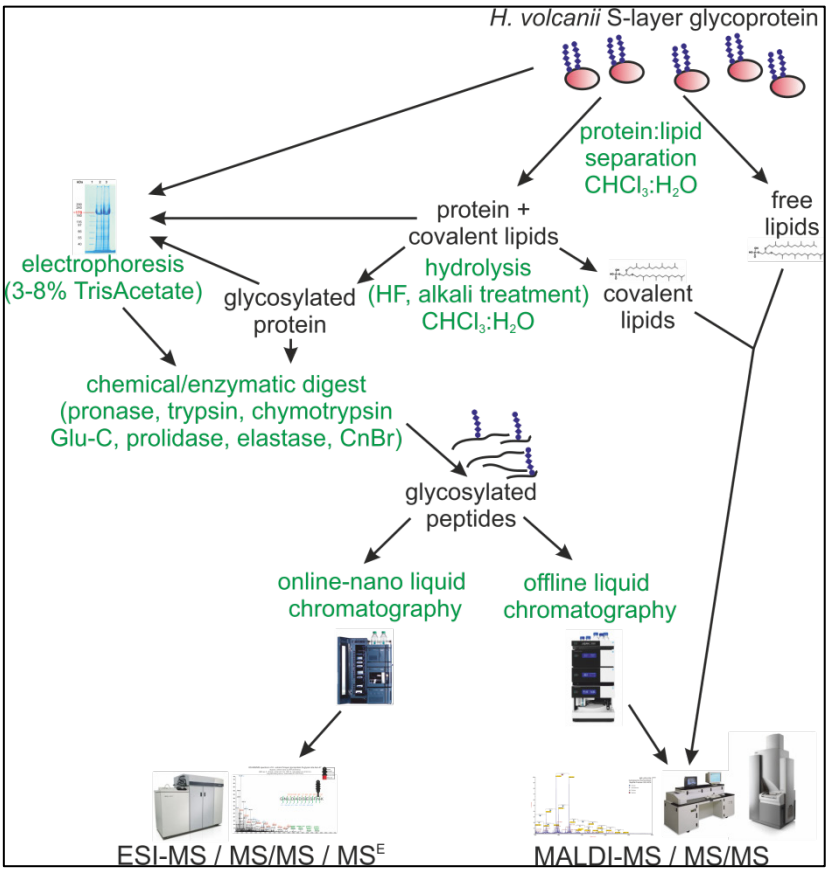


Figure 3.25 – Methods investigating the potential lipid modification of the S-layer glycoprotein.
This schematic describes the various approaches employed in the investigation of the S-layer glycoprotein C-terminal region, as proposed to be modified by lipid moieties.

Figure 3.25 summarises the methods employed in this section. The sample was washed in chloroform and water to separate any unbound lipid moieties from the S-layer glycoprotein (**Section 2.4.6.1**) and whilst the chloroform fraction was assessed by MALDI-MS and MALDI-MS/MS, the protein-containing fraction was further hydrolysed to separate the protein from any covalently linked lipid moieties present. The methods of choice were HF treatment to hydrolyse any phosphodiester linked moieties and an alkaline treatment described by the Eichler group (**Section 2.4.6.2** and **Section 2.4.6.3**) (Kandiba et al., 2013), upon which a further chloroform:water separation aimed to give rise to a pure protein fraction and a lipid fraction of the released lipids. Both hydrolysis techniques were optimised in a series of repeat attempts to ensure full

release of any lipid moieties covalently linked to any portion of the S-layer glycoprotein. These methods were assessed alongside controls of unhydrolysed samples. The lipid fractions were further concentrated and spotted in conjunction with their water fraction counterparts containing the S-layer glycoprotein using a series of matrices for analysis by MALDI-MS and MALDI-MS/MS in both the negative and positive ion mode. The matrices employed included DHB, HABA and DABP for the lipid moieties used on their own and in suitable combinations, as well as HCCA and SA matrices for assessment of the water fraction containing any peptides.

3.6.1 Lipid population of *H. volcanii* found in the S-layer glycoprotein extract

To investigate the lipid presence acting as a modification of the S-layer glycoprotein, the sample was processed as described in **Figure 3.25**. All fractions were examined for the potential identification of any lipid moieties, if present, expected to be covalently linked directly to the peptide or via the O-glycans. These experiments were repeated multiple times and the hydrolysis and separation procedures optimised to assure maximum confidence in the results presented here. FAME analysis was initially performed on the S-layer glycoprotein samples (data not shown), confirming the presence of lipid moieties in the sample.

The free, non-covalently bound lipid fraction as separated from the S-layer glycoprotein sample by chloroform:water extraction was shown to contain archaetidylglycerol (PG, m/z 805.68), archaetidylglycerol methylphosphate (PGP-Me, m/z 899.66) as visible separately or in adducts with sodium (PGP-Me +Na, m/z 921.65), sulphate-diglycosyl archaeol-1 (S-GL-1, m/z 1055.74) and sulphate-diglycosyl archaeol-2 in aggregation with sodium (S-GL-2, m/z 1792.24), as is shown in **Figure 3.26**. These lipid moieties were identified as freely associated with the S-layer glycoprotein sample as extracted by the Eichler group and are equivalent to the lipid moieties classified as belonging to the lipid bilayer of *H. volcanii* cells published by Sprott and colleagues (Sprott et al., 2003).

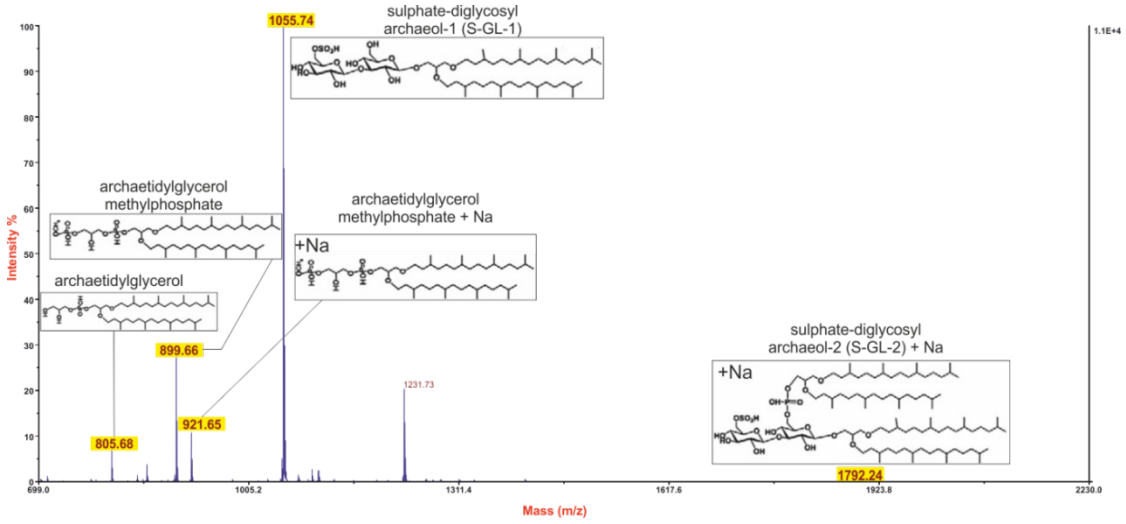


Figure 3.26 – MALDI-MS spectrum of the *H. volcanii* S-layer glycoprotein free lipid fraction.
The MALDI-MS spectrum shows the free lipids present in the purified S-layer glycoprotein sample as separated from the protein by chloroform:water extraction, analysed in the negative ion mode as was accumulated on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer. The highlighted and labelled peaks show the lipid moieties present in the fraction.

The molecular ion at m/z 1231.73 was assessed by MS/MS, but did not yield the expected S-GL-1 covalently linked to a hexose. The assignment of the lipid moieties presented was verified by MALDI-MS/MS of each of the molecular ions, such as is shown for the MALDI-MS/MS spectrum of the molecular ion at m/z 1055.74, of the S-GL-1 moiety (**Figure 3.27**).

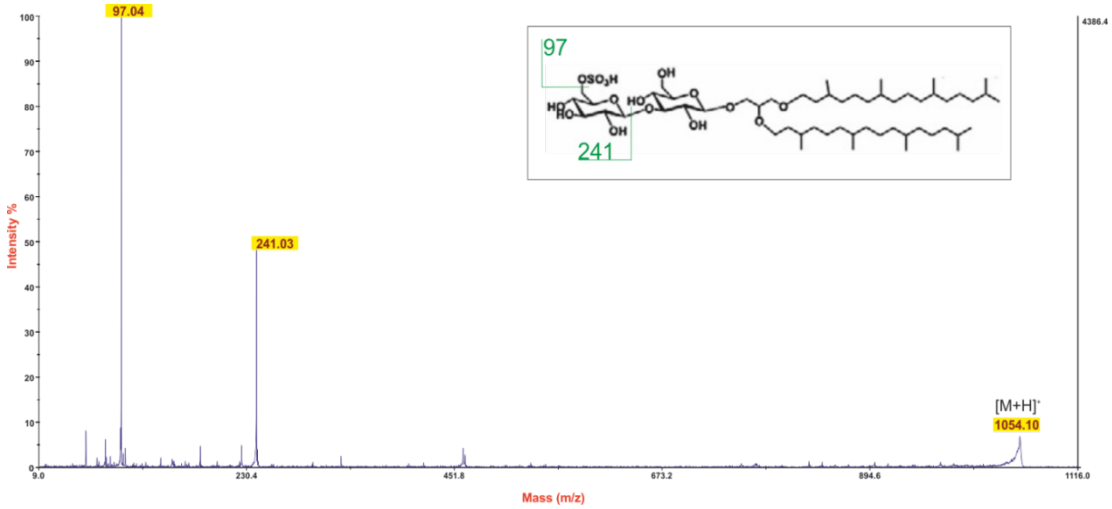


Figure 3.27 – MALDI-MS/MS spectrum of the molecular ion at m/z 1055.74, S-GL-1 (free lipid fraction).
The MALDI-MS/MS spectrum shows the fragmentation of molecular ion at 1055.74, S-GL-1 from the free lipid fraction, present in the purified S-layer glycoprotein sample as separated from the protein by chloroform:water extraction, analysed in the negative ion mode as was accumulated on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer. The structure of the glycolipid is shown in a separate box with indications to fragmentations observed.

The sample fraction obtained post HF treatment presented MALDI-MS and MS/MS data similar to the lipid assessment of the non-covalently associated lipid population, in as much that no newly identified lipid moieties related to modification of the S-layer glycoprotein could be observed in the spectrum. **Figure 3.28** shows the MALDI-MS spectrum of the lipid moieties extracted post HF hydrolysis. HF treatment allows for the hydrolysis of any phosphodiester linkage, such as are observed in archaetidylglycerol methylphosphase (PGP-Me, m/z 899.66) and sulphate-diglycosyl archaeol-2 (S-GL-2, m/z 1770.24, +Na m/z 1792.24), explaining the absence of these signals in the spectrum.

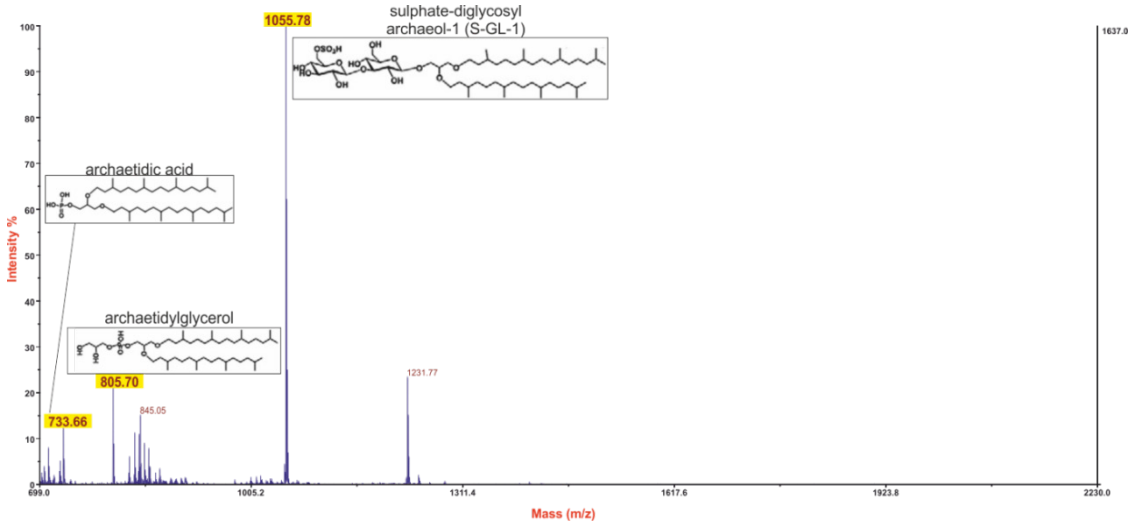


Figure 3.28 – MALDI-MS spectrum of the S-layer glycoprotein released lipid moieties.
The MALDI-MS spectrum shows lipid moieites released or generated by HF treatment of the purified S-layer glycoprotein sample as separated from the protein by chloroform:water extraction post hydrolysis, analysed in the negative ion mode, as was accumulated on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer. The highlighted and labelled peaks show the lipid moieties present in the fraction.

The presence of archaetidic acid (PA, m/z 733.66), could be suggested to represent a new lipid moiety released from the S-layer glycoprotein by HF treatment, however, such hydrolysis of the S-GL-2 glycolipid will yield the archaetidic acid and S-GL-1 moieties, and thus cannot be definitively proposed as such and further investigations will be required for elucidation of the archaetidic acid moiety from different sample extractions and purifications. Comparison with control samples of unhydrolysed S-layer glycoprotein and analysis of the peptide containing counterpart fractions in water, which contained the same group of lipid moieties, suggested these lipids though tightly associated with the S-layer glycoprotein, were not phosphodiester-linked to the protein. The reproduction of the alkaline hydrolysis procedure described by Kandiba and colleagues did not yield any informative data, even though this procedure was repeated multiple times (Kandiba et al., 2013).

To exclude the possibility that the lipid moieties found potentially covalently linked to any glycan or peptide region, were too large to be identified, due to the sizes of these modified sections of the S-layer glycoprotein, the sample was digested by pronase enzyme, an enzyme that hydrolyses the peptide backbone into short peptides of only a few amino acids. This should allow the identification of any lipid moieties, even if linked via the disaccharide O-glycan or directly to any peptide, as the presence of such monosaccharides or amino acids would be observable in the MALDI-MS and MS/MS spectra in the positive or negative ion mode. This however, also did not yield any evidence for covalent attachment of the lipid moieties present in the sample, as the resultant mass spectra showed the same peaks as found in the other fractions (data not shown).

From the series of experiments described above, no further evidence of the C-terminal region of the S-layer glycoprotein was found and the sample purified and extracted by the Eichler group was found to contain the lipid moieties, previously identified by Sprott and colleagues as consistent with the major lipid populations of the *H. volcanii* membrane bilayer (**Table 1.1**), when the samples were assessed by MALDI-MS and MALDI-MS/MS in both the positive and negative ion mode. The only differences observable were that in both ion modes, the relative abundances of each of the lipid moieties identified were not consistent with the published works by Naparstek and Sprott (Sprott et al., 2003, Naparstek et al., 2012). The high abundance of the m/z 1055 S-GL-1 glycolipid compared with the barely visible m/z 1770 S-GL-2 glycolipid, suggests that the presence of the additional lipid moiety on the latter prevents tight association with the S-layer.

3.6.2 S-layer glycoprotein electrophoresis studies

The mature S-layer glycoprotein, as published by Sumper and colleagues, was shown be composed of 793 amino acids with an expected calculated molecular weight of approximately 81,732Da. However, upon analysis of the extracted and deglycosylated S-layer glycoprotein on an SDS-PAGE gel, the apparent molecular mass of the polypeptide running at approximately 170kDa, is nearly double the expected size of the protein (Sumper et al., 1990). Even with the herein presented information on N- and O-glycosylation of the *H. volcanii* S-layer glycoprotein contributing to additional molecular mass, such an extreme irregularity in electrophoretic behaviour seems unlikely to be explained by glycosylation. The abnormal electrophoretic behaviour was also seen in the S-layer glycoprotein of *H. halobium*, where Kikuchi and colleagues had suggested covalent lipid attachment to *H. halobium* in (Kikuchi et al., 1999), which brought about suggestions on the electrophoretic retardation being due to a proposed lipid moiety. Eichler and colleagues

therefore also suggested lipid modification of *H. volcanii* as the reason for the aberrant electrophoretic mobility of the S-layer glycoprotein (Konrad and Eichler, 2002).

The S-layer glycoprotein was run on Tris-Acetate 3-8% gradient gels due to the large mass of the glycoprotein and run alongside an unstained HiMark™ ladder, for mass annotation and stained with Coomassie blue stainer (**Section 2.4.1**) (**Figure 3.29**, well number 2 and 3 and **Figure 3.30**, well number 5). To further elucidate the unusual electrophoretic migration of the S-layer glycoprotein, the S-layer glycoprotein was treated with alkaline and HF treatment described for the release of any covalently linked lipids and separated using chloroform and water (**Section 2.4.6**) and run alongside the untreated S-layer glycoprotein separated from non-covalently linked lipid moieties (**Figure 3.29**). These experimental fractions were run in wells 7-10, however, no protein moieties could be visualised. Furthermore, the aspect of a possible dimerisation was investigated by chemical denaturing of the protein prior to electrophoresis, which however, did not present any different data and the protein ran identically to the control sample and the S-layer glycoprotein could be visualised at roughly 170kDa again (data not shown), dismissing the concept of possible dimerization. In addition, the sample was reduced and carboxymethylated in case mis-annotation of the amino acid sequence had occurred and cysteine residues were to be present in the S-layer glycoprotein, which could allow dimersiation via cystine formation (**Figure 3.30**, well number 3). However, this experiment also proved negative.

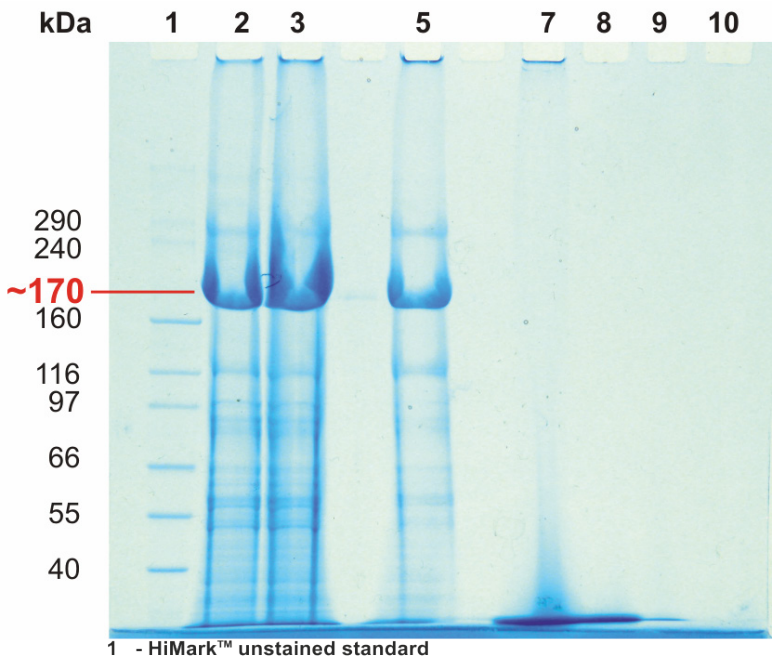


Figure 3.29 – Tris-Acetate (3-8%) gel of *H. volcanii* S-layer glycoprotein (hydrolysed lipid moieties). The samples run on the gel included a HiMark™ unstained standard, the control sample of purified S-layer glycoprotein, as separated by chloroform:water and post alkaline and HF treatment (both fractions). The gel was run at 150V for 60mins and stained with Coomassie blue stain. Annotated in red is the *H. volcanii* S-layer glycoprotein size as visualised on the gel.

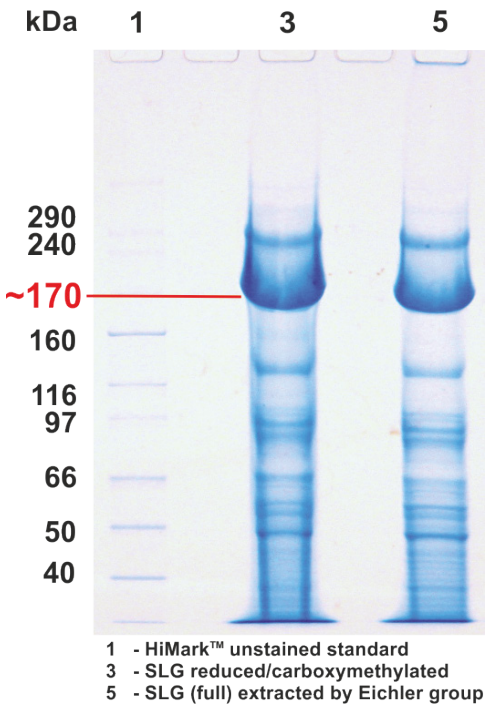


Figure 3.30 – Tris-Acetate (3-8%) gel of *H. volcanii* S-layer glycoprotein (reduced/carboxymethylated). The samples run on the gel included a HiMark™ unstained standard, the control sample of purified S-layer glycoprotein, as well as reduced and carboxymethylated prior to electrophoresis. The gel was run at 150V for 60mins and stained with Coomassie blue stain. Annotated in red is the *H. volcanii* S-layer glycoprotein size as visualised on the gel.

3.6.3 Further investigation into the C-terminal region of the S-layer glycoprotein

The lack of information obtained on the C-terminal region, as well as the irregular and inconclusive behaviour of the S-layer glycoprotein during electrophoresis, highlighted the importance of knowing the true molecular weight of the S-layer. Determining this by MALDI-MS was therefore attempted. The S-layer glycoprotein was therefore spotted on MALDI plates with a series of appropriate matrices, for linear-TOF analysis in both the negative and positive ion modes. To supplement this, the sample was further investigated by MALDI-MS and MS/MS, post hydrolysis by trypsin and CNBr, which were aimed to have created peptide fragments of the C-terminal domain. This C-terminal region was assessed by MALDI-MS, keeping in mind its glycosylation with one N-glycan and a few O-linked disaccharides, potentially resulting in a high molecular weight peptidoglycolipid. MALDI-MS was used, as there is less prospect of obtaining ESI data with the charges predicted for the C-terminal tryptic peptide. Unfortunately, despite considerable experimentation none of these molecular weight investigations produced meaningful MALDI spectra (data not shown). Therefore, no new in formation of any covalently linked glycolipid or phospholipid moiety on the elusive C-terminal region of the S-layer glycoprotein was obtained.

3.6.4 Summary discussion on lipid association studies of the S-layer glycoprotein

Based on the data collected by the various hydrolysis and enzymatic procedures employed in the elucidation of the C-terminal region suggested to be modified by a lipid moiety, the MALDI-MS data does not provide any evidence that the lipids present in the S-layer glycoprotein sample are covalently linked to the S-layer glycoprotein, but would rather suggest that lipid moieties of the lipid bilayer are in strong non-covalent association with the protein.

The lipid moieties observed coincide with the published lipid population identified by Sprott and colleagues from the *H. volcanii* lipid bilayer (Sprott et al., 2003). *H. volcanii* polar lipids were found to be composed of phospholipids and sulphated glycolipids S-GL-1 and S-GL-2. The phospholipids are archaetidylglycerol (PG 805.6Da) and archaetidylglycerol-methyl-phosphate (PGP-CH_3 899.5Da and with sodium 921.5Da) were the major signals identified. Further, a major signal of 1055.6 m/z was detected as expected for a sulphate-diglycosyl archaeol, characteristically found in the polar lipids extracted from archaeal extreme halophiles (Kamekura and Kates, 1999). The sulphated diglycosyl lipid S-GL-1 from *H. volcanii* was characterized by detailed NMR analysis and found to be 6-HSO_3-D-Manp-a1-2-D-Glcp-a1,1-[sn-2,3-di-O-phytanylglycerol. The structural identity of S-GL-1 to a portion of the S-GL-2 molecule suggests a precursor relationship during biosynthesis of S-GL-2 (Sprott et al., 2003). To date, however the enzymes responsible for adding the glucose and mannose residues to the diphytanylglycerol moiety have not been identified, although a putative biosynthesis pathway has been proposed. Naparstek and colleagues tested whether AglB or any of the glycosyltransferases involved in *H. volcanii* S-layer glycoprotein N-glycosylation were also involved in the *H. volcanii* glycolipid biogenesis by deletion of the known *agl* genes using NMR. The results showed different glcyosyltransferases were involved in the biosynthesis of N-linked glycoproteins and glycolipids in archaea (Naparstek et al., 2010).

The lipids found to be present in the herein examined sample are more likely to be co-extracted with the EDTA procedure used by the Eichler group, where 0.5M EDTA was used to release the S-layer from the *H. volcanii* cells. Such methods could lead to the release of the S-layer glycoprotein in close association with the lipid bilayer, through its transmembrane domain at the C-terminal, and thus presented in strong association with such moieties, rather than by covalent attachment. This is further explained by the independently published works of Kandida and colleagues of the Eichler group, suggesting differences in the S-layer presentation upon the use of two separate extraction procedures, EDTA and Triton X-100 (Kandiba et al., 2013). The investigations carried out by Kandida and colleagues suggested that two distinct populations of the S-layer glycoprotein could be observed upon solubilisation with the different methods, presenting mass spectrometric data proposing that only the EDTA-solubilised version of the protein was modified by an archaetidic acid. This lipid moiety, however, will also be produced upon alkaline hydrolysis of the S-GL-2 sulphate-diglycosyl archaeol lipid found in the lipid bilayer of *H. volcanii* (Sprott et al., 2003, Naparstek et al., 2012).

For further assessment of such differences in extraction techniques and to confirm the data presented here, additional sample preparations are required for analysis. Such S-layer preparations were requested from our collaborators but were not received by the time of thesis submission.

On another note, our detection of non-covalent lipids in the S-layer preparations whilst finding no evidence for their covalent attachment, has prompted us to question the hypothesis that an archaeosortase mechanism could be involved in the putative lipid modification of the S-layer (Haft et al., 2012, Kandiba et al., 2013). As mentioned earlier, Eichler's laboratory has recently reported that the S-layer glycoprotein solubilised by EDTA, has an archaetidic acid moiety that could be detected after base hydrolysis. Eichler has proposed that the S-layer glycoprotein could be initially synthesized as an integral membrane protein that subsequently undergoes a processing event in which the extracellular portion of the protein is separated

from the membrane-spanning domain and transferred to a waiting lipid moiety (Kandiba et al., 2013). This was explained by the proposed identification of an enzyme suspected to catalyse a cleavage or transfer reaction as is found parallel to the protein processing system found in Gram-positive bacteria using sortases (Haft et al., 2012). Sortases in bacteria recognize a carboxyl terminal sequence motif, the PGF-motif, immediately upstream of a C-terminal membrane-spanning domain in the target protein, cleave the protein at this site and then covalently attach the released proteins to the cell wall and facilitate their transfer to a waiting lipid moiety. In the *H. volcanii* S-layer glycoprotein, a PGF motif is present just upstream of the predicted C-terminal trans-membrane domain and is thought to be recognized by a bioinformatically identified archaeosortase, ArtA. (Kandiba et al., 2013, Haft et al., 2012). It is thought that the transpeptidation catalysed by the archaeosortase would parallel a similar protein processing system found in Gram-positive bacteria, where the sortases recognise this PGF motif for cleavage of the protein at such a site followed by covalent attachment of the released proteins to a waiting lipid moiety in the cell wall. This process requires high activation energy and the process by which such transpeptidation would occur has not been studied in archaea.

The presence of a membrane-spanning domain in the amino acid sequence of the S-layer glycoprotein of *H. volcanii*, however, would make the activity of archaeosortases redundant. Furthermore, if this theory was in fact feasible, it has to be noted that archaeosortases might be non-specific and could cleave anywhere, with the remaining C-terminal region of the S-layer glycoprotein being too small to visualise by the mass spectrometric techniques employed. A potential lipid modification of membrane inserted proteins in archaea may reflect a primitive version of a model of protein membrane association widely used in eukaryotes, such as is seen in C-terminal GPI-anchors, however, this still maintains the role of the membrane spanning domain present. Furthermore, in view of the solubility problems encountered throughout the various experiments, the C-terminal region of the S-layer glycoprotein and its associated modifications by lipid moieties could preferentially precipitate and thus would not be observable in the experiments conducted. A final possibility for the lack of C-terminal mapping data is that the gene sequence may be incorrect. Further work will be required to determine whether this might be the case.

3.7 GENERAL DISCUSSION

3.7.1 Roles of post-translational modifications of the S-layer glycoprotein in *H. volcanii*

In the S-layer glycoprotein of *H. volcanii*, the use of deletion strains has provided considerable insight into the importance of N-glycosylation to the cell. Strains lacking the ability to perform N-glycosylation, due to the absence of the oligosaccharyltransferase, AglB, or only able to partially recruit the N-glycosylation pathway, due to an absence of other Agl proteins, presented various phenotypes, including an S-layer of modified architecture, showing increased susceptibility to proteolytic digestion, enhanced S-layer glycoprotein release into the growth medium and slower growth in medium of increasing salt (Abu-Qarn et al., 2008b, Abu-Qarn et al., 2007, Yurist-Doutsch et al., 2008). Indeed, differential transcription of various Agl proteins in response to differing growth conditions, reflected by reverse transcription or real time PCR, points to N-glycosylation as being an adaptive process in *H. volcanii* (Jarrell et al., 2010).

As discussed by Mengele and Sumper, halophilic archaeal proteins are enriched in acidic amino acids to cope with life in hypersaline environments, however, this strategy wouldn't offer a response to transient changes in salinity, as would post-translational modifications. The S-layer glycoprotein was suggested to undergo differential glycosylation in terms of the N-glycan decorating the N-glycosylation sites, upon growth in distinct environments of salinity (Guan et al., 2012). Guan and colleagues investigated the Dol-P carriers involved in the assembly of the N-glycan, as well as the S-layer glycoprotein by LC-ESI-MS in cells grown in high and low salt concentrations of 3.4M NaCl and 1.75M NaCl respectively. Guan and colleagues found that in high salt Asn-47 and Asn-117 were modified by a pentasaccharide, while dolichol phosphate was modified by a tetrasaccharide comprising the first four pentasaccharide residues. When the same targets were considered from cells grown in low salt, substantially less pentasaccharide was detected and cells grown at low salinity contained dolichol phosphate modified by a distinct tetrasaccharide absent in cells grown at high salinity. This tetrasaccharide composed of a deoxyhexose, two hexoses and a sulphated or phosphorylated hexose was found to modify Asn-532 in cells grown in low salt, whereas no glycan was found to decorate this residue in cells grown in the high salt-medium. Guan and Eichler therefore suggested that *H. volcanii* modulated the N-linked glycans in response to changes in environmental salinity, which however, is surprising as sulphated or phosphorylated monosaccharide moieties would be expected in high salt environment rather than lower concentrations of salinity. The presently investigated samples were grown at 3.4M NaCl, corresponding to the high salt environment discussed by Guan and colleages of the Eichler group. However, the data presented by Guan and colleagues suggested Asn-404 to be unoccupied, which is rectified in the data presented here.

3.7.2 Functional characterisation of N-glycan biosynthesis of the *H. volcanii* S-layer

In the effort to elucidate the biosynthesis pathway and the gene products involved in the N-glycosylation of the S-layer glycoprotein of *H. volcanii*, Eichler and colleagues in collaboration with our laboratory previously attempted to functionally characterise the putative homologs of genes involved in the eukaryal or bacterial N-glycosylation, with the studies described below (Abu-Qarn et al., 2008b, Abu-Qarn et al., 2007, Yurist-Doutsch et al., 2008).

The combination of gene deletion and mass spectrometry of a reporter glycopeptide carrying the pentasaccharide, demonstrated the involvement of AglD and AglB in the N-glycosylation of the S-layer glycoprotein. AglB was identified as the sole component of the oligosaccharide transferase complex. AglD was shown to be involved in the loading of the final hexose residue onto the tetrasaccharide. AglD includes a soluble, cytoplasmic-facing N-terminal half, associated with the membrane via the integral C-terminal portion of the molecule, spanning the membrane six times. This represents a topology reminiscent of those eukaryal glycosyltransferases containing active sites facing the ER lumen, responsible for the transfer of mannose or glucose residues from charged dolichol carriers to the heptasaccharide following its reorientation from the cytoplasm to the lumen of the ER membrane (Abu-Qarn et al., 2007). Kaminski and Eichler further attempted to identify the residues of importance for the activity of AglD who designed an *in vivo* assay. Site-directed mutagenesis at positions encoding residues conserved in archaeal homologs showed the restoration of AglD function in *H. volcanii aglD* deletion strains that were transformed to express the plasmid-encoded versions of AglD. This showed that Asp-110 and Asp-112 were elements of the Asp-X-Asp motif of AglD, which

interacts with metal cations associated with nucleotide-activated sugar donors, whilst Asp-201 was predicted to be the catalytic base of the enzyme (Kaminski and Eichler, 2010).

In addition to these genes, *H. volcanii* was found to encode at least nine other putative glycosyltransferases, such as *aglE*, which is located only eight ORFs upstream from *aglB*. Upon gene deletion experiments in combination with mass spectrometric techniques, it was suggested that AglE was involved in the loading the 190 Da sugar subunit of the *H. volcanii* S-layer glycoprotein onto the growing trisaccharide. This process was presumed to be achieved by loading of the 190Da sugar subunit onto a lipid carrier followed by transferral onto the 176Da sugar species of the three-membered pentasaccharide precursor via the actions of a distinct transferase, or that AglE directly adds the 190Da sugar subunit to the growing oligosaccharide. Topological analysis assigned AglE as an integral membrane protein, with its N-terminus and putative active site facing the cytoplasm, suggesting the latter option (Abu-Qarn et al., 2008b).

Examination of those ORFs found upstream of *aglB* revealed further sequences and the gene products of *aglF*, *aglI* and *aglG* were found to be involved in the biosynthesis of the S-layer pentasaccharide. These three genes are found immediately upstream from the gene encoding the archaeal oligosaccharide transferase, AglB. It was shown that AglF and AglI are involved in the addition of the hexuronic acid found at position three of the pentasaccharide, whilst AglG was shown to contribute to the addition of the hexuronic acid found at position two (Yurist-Doutsch et al., 2008). Given their proximities in the *H. volcanii* genome, the transcription profiles of *aglF*, *aglI*, *aglG* and *aglB* were considered and whilst *aglF* and *aglI* share a common promoter, transcription of the four genes is co-ordinated, as was demonstrated by variation in growth conditions and the consequent analysis of transcript levels in *H. volcanii* cells. Such changes in N-glycosylation gene transcription levels offer additional support for the adaptive role of this post-translational modification in *H. volcanii*. Furthermore, the deletion of these three gene products did not compromise *H. volcanii* survival, but presented a modified S-layer (Yurist-Doutsch et al., 2008).

Yurist-Doutsch and colleagues re-addressed the annotation of the gene clusters that includes *aglB*, *aglE*, *aglF*, *aglG*, *aglI* and *aglJ* and found further genes encoding AglS, AglQ and AglR (Yurist-Doutsch and Eichler, 2009), whilst the actions of AglM, AglP and AglJ are discussed in this thesis.

Whilst investigations in this project focussed on the full profiling of the S-layer glycoprotein and the possibility of a lipid modification, the Eichler group continued research into further genes identified by Yurist-Doutsch. Their research led to further published findings on AglR and AglS. The function of the AglR protein in the biosynthesis of the N-glycan was suggested to be similar to that of its homologue Wzx, which acts as a flippase in LPS bioysnthesis. AglR was found to be responsible for delivering the dolichol phosphate linked mannose as the final subunit of the N-linked glycan to be transferred to the nascent pentasaccharide. It was found that in cells lacking AglR, glycan-charged dolichol phosphate, including mannose-charged dolichol phosphate accumulated, resulting in the S-layer not incorporating mannose into its N-glycan (Kaminski et al., 2012).

Furthermore, the role of AglS was also found to be required for the final mannose addition to the N-glycan pentasaccharide, by acting as a dolichol phosphate-mannose mannosyltransferase and thus mediating the

transfer of mannose from its dolichol carrier to the nascent tetrasaccharide (Cohen-Rosenzweig et al., 2012). This further underlines the difference in biosynthesis pathways of *H. volcanii* to the related *H. marismortui* S-layer glycoprotein **(Section 3.7.2.1)**.

The current understanding of the N-glycosylation biosynthesis pathway in *H. volcanii* S-layer glycoprotein is represented in **Figure 3.31**.

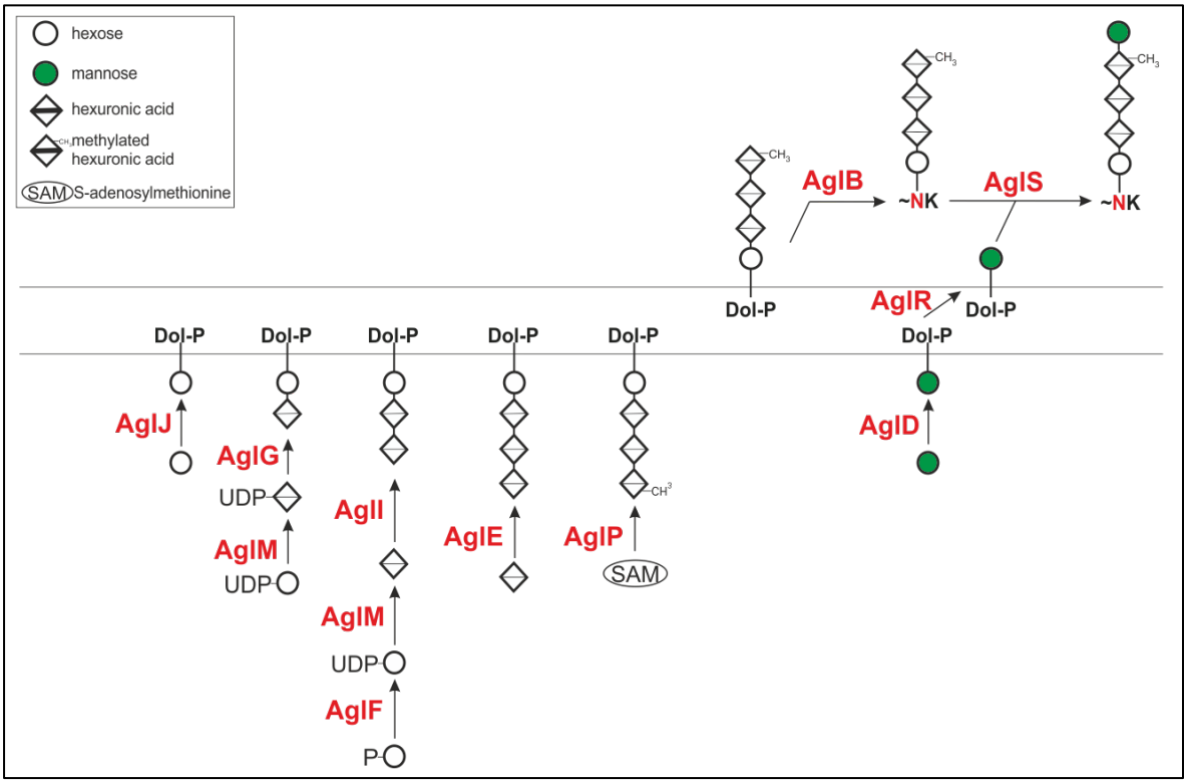


Figure 3.31 – Current understanding of N-glycosylation biosynthesis in the S-layer glycoprotein.
Schematic of up to date knowledge of various gene functions of the *agl* gene cluster in *H. volcanii* as adapted and summarised from Jarrell and colleagues (Guerry, 2011, Jarrell et al., 2010).

3.7.2.1 The biosynthesis of the N-glycan of the S-layer glycoprotein and glycoengineering

Kunz and colleagues first reported the presence of C_{55} and C_{60} dolichol phosphates in *H. volcanii* including glycan-modified species (Kuntz et al., 1997). Guan and colleagues then used high sensitive LC-MS to detect and characterise glycan-charged phosphodolichols in *H. volcanii* (Guan et al., 2010, Guan and Eichler, 2011). Apart from the complete pentasaccharide, *H. volcanii* was found to contain dolichol phosphates charged with the same glycan series as found N-linked to the S-layer glycoprotein and cells lacking components of the N-glycosylation machinery presented dolichol phosphates void of, or bearing truncated glycans. In contrast to the sequential assembly of the first four pentasaccharide subunits onto a common dolichol phosphate, the fifth subunit of the pentasaccharide, mannose, was detected on its own distinct lipid carrier, suggested by the finding that AglD, involved in the addition of the fifth pentasaccharide subunit, acts in a manner seemingly independent of the other Agl proteins. This is not unexpected, as *aglD* is the only gene not found in the *agl* gene island present in the *H. volcanii* genome (Yurist-Doutsch and Eichler, 2009). It was found to be unlikely for there to be a short lived pentasaccharide-modified phosphodolichol species, as such entities were readily observed in another halophilic archaea originating from the dead Sea, such as

Haloarcula marismortui (Calo et al., 2011). *H. marismortui* is a second haloarchaeon also originating from the Dead Sea, which similar to *H. volcanii* decorates the S-layer glycoprotein with the same N-linked pentasaccharide and employs dolichol phosphate as lipid glycan carriers. In *H. marismortui* the complete pentasaccharide is assembled on dolichol phosphate and only then transferred the N-glycan to the target protein *en bloc*, unlike the biosynthesis process explored by *H. volcanii*, reminiscent of the process observed in eukaryal N-glycosylation (Calo et al., 2011, Guerry, 2011).

The *H. volcanii* complete genome sequence has been made available and shows homologs to many other *agl* gene sequences, which are often clustered into gene islands. Calo and colleagues detected OST-encoding aglB genes in numerous species and suggested that the N-glycosylation seen in the N-linked glycans of archaeal glycoproteins identified to date implies that an enormous array of enzymes are involved in this posttranslational modification. The relatively well-defined N-glycosylation pathway and molecular tools for gene manipulation of the *H. volcanii* S-layer glycoprotein, proposed this archaeon as a promising candidate for glycoengineering. The ability of three haloarchaeal homologs of *H. volcanii* AglD to replace the native enzyme was further considered. AglD is involved in adding the final hexose of the pentasaccharide N-linked to the S-layer glycoprotein. *H. volcanii ΔaglD* cells were transformed with the AglD homologs from *H. marismortui* rrnAC1873, *H. walsbyi* 1489A and *H. salinarum* OE1782 to restore the S-layer glycosylation as was assessed by SDS-PAGE migration, glycostaining and mass spectrometry profiles. All three homologs of AglD were found to introduce the final hexose to the N-linked S-layer glycoprotein pentasaccharides even though not all were found to be functional and transcribed in their native host under the tested conditions. *H. volcanii* was therefore shown to express foreign enzymes and incorporate them into its N-glycosylation pathway, even though such enzymes were not central to this posttranslational modification in the native host. These results represent the first steps in assessing the suitability of *H. volcanii* as a platform for the synthesis of variable N-linked glycans relying on introduced N-glycosylation pathway components originating from other archaeal species (Calo et al., 2010a).

3.8 FURTHER DIRECTIONS

At the time of the submission of this thesis, unfortunately the C-terminal of the S-layer glycoprotein of *H. volcanii* was still undetermined, along with its potential modifications associated with the C-terminal region. Further investigations into the S-layer glycoprotein of *H. volcanii* will be focused on such characterisation of the C-terminal region along with the final putative N-glycosylation site. In the search for the C-terminal region of the S-layer glycoprotein matters of potential lipid modification, cleavage of the region and transfer onto lipid moieties by archaeosortases as well as assessment of different S-layer glycoprotein preparations in terms of extraction from the *H. volcanii* cells will be investigated.

With regards to glycosylation sites, the last putative N-glycosylation site has not been able to be described in work presented in this thesis, and with further investigations into the C-terminal domain, this last N-glycosylation site will be explored for occupation by the N-glycan. With the continued aim of creating a full comprehensive map of the S-layer glycoprotein, the O-glycosylation sites present a further goal yet to be investigated.

On the subject of potential lipid modification as discussed in various published papers, the exact sample preparations described in the published works will need to be investigated for the potential covalent association of any lipid moieties linked to the S-layer glycoprotein, which could also be responsible for obscuring the C-terminal region.

A recent study on the archaeosortase ArtA protein in *H. volcanii* by Halim and colleagues suggests the S-layer glycoprotein as one of its substrates, where *artA* deletion strains show poor growth, differences in cell shape and S-layer architecture, as well as impaired motility of the cells (Abdul Halim et al., 2013). Furthermore, their identification of two peptides in the C-terminal region of the S-layer glycoprotein by LC-MS/MS in the artA deleted strains of *H. volcanii* lead, and the absence of said peptides in the wild type parent strains, led them to suggest ArtA to proteolytically process the S-layer glycoprotein and anchor it in the lipid bilayer. No further identification of any other C-terminal peptide sequences, or glyco-lipo-protein could be identified. It is therefore clear that similar problems are encountered by other groups in the study of the *H. volcanii* S-layer glycoprotein.

This means that thorough mass spectrometric analysis of all S-layer glycoprotein preparations, including lipid extractions and archaeosortase deleted strains should be conducted. Improvement of mass spectrometric data collection could be geared towards enhancing the ionisation capacity of peptides post enzymatic or chemical digestion by esterification, whilst hydrolysis procedures aimed at the release of chemical bonds responsible for the covalent attachment of such moieites, without the degradation of the rest of the protein will also need to be developed for characterisation of both potential modifications and the S-layer glycoprotein. Further sample preparations with consideration given to extraction procedures and growth conditions (high and low salinity), will be investigated upon their receipt from the Eichler group and assessed for glycosylation patterns and protein sequence to allow for full understanding of the processing and assembly of the S-layer glycoprotein of *H. volcanii*.

Chapter 4

Burkholderia pseudomallei 576

LPS analysis

4 *BURKHOLDERIA PSEUDOMALLEI* 576 LPS ANALYSIS

4.1 EXPERIMENTAL STRATEGY

This chapter focuses on the structural elucidation of the *B. pseudomallei* 576 LPS O-antigen to provide information for the creation of a comprehensive glycoconjugate vaccine against all melioidosis causing bacterial agents. Due to the ACDP listing of *B. pseudomallei* as a category III pathogen, the work on this project was was carried out in collaboration with Dr. Joann Prior and her team at the Defence Science and Technology Laboratory (dstl), part of the Ministry of Defence (MOD) at Porton Down, Wiltshire, United Kingdom. The LPS samples investigated derived from strains including *B. thailandensis* E264, *B. pseudomallei* K96243 (serotype A) and 576 (serotype B). Whilst the capsule mutants (*ΔCPS*) of both *B. pseudomallei* K96243 and 576 were analysed in comparison, the *B. pseudomallei* 576 (*ΔLPS*) strain was examined for assessment of background carbohydrate contamination (**Section 2.3.2**). All strains were investigated for compositional and linkage information and used for data comparison with *B. pseudomallei* 576 LPS, which remained the focus for structural characterisation by MALDI mass spectrometry and NMR techniques.

As previously discussed the structural characterisation of glycosylation in prokaryotes cannot rely on predictable biosynthesis pathways as is often the case for eukaryotes, which makes the elucidation of unknown structures, such as presented by this project on the structural characterisation of the *B. pseudomallei* LPS O-antigen, a much more difficult task (**Figure 4.1**).

Although most of the methodologies employed in the structural description of the O-antigen repeat unit are widely used in LPS characterisation, their implementation was not necessarily straightforward. Hydrolytic methods in particular had to be developed as the project and understanding of the LPS evolved. In the elucidation of the *B. pseudomallei* 576 LPS O-antigen repeat unit a variety of structural analysis techniques were combined. Our approach included determination of the monosaccharide composition by GC-MS analysis of TMS and alditol acetate derivatives (**Section 2.4.8.6** and **Section 2.4.8.7**), prepared from hydrolysed material, complemented by linkage analysis which provided insight into the arrangement of the monosaccharaides within the O-antigen repeat unit (**Section 2.4.8.8**). For MALDI-MS and MALDI-MS/MS sequencing of the O-antigen repeat unit, the polysaccharide portions of the LPS of *B. pseudomallei* 576 had to be released from the lipid A moiety to allow for solubilisation and hydrolysis of the polysaccharide for analysis. A multitude of hydrolysis techniques were employed, adapted and created in order to achieve lipid A release and polysaccharide hydrolysis into manageable and observable repeat units (**Section 2.4.7**). Mass spectrometric techniques allowed for the postulation of a proposed repeat unit structure of the O-antigen, yet with an assortment of choices. NMR spectroscopy techniques aimed to support these structural proposals by providing further information on the linkage and anomeric stereochemistry of the *B. pseudomallei* 576 LPS O-antigen (**Section 2.4.9** and **Section 2.4.10**).

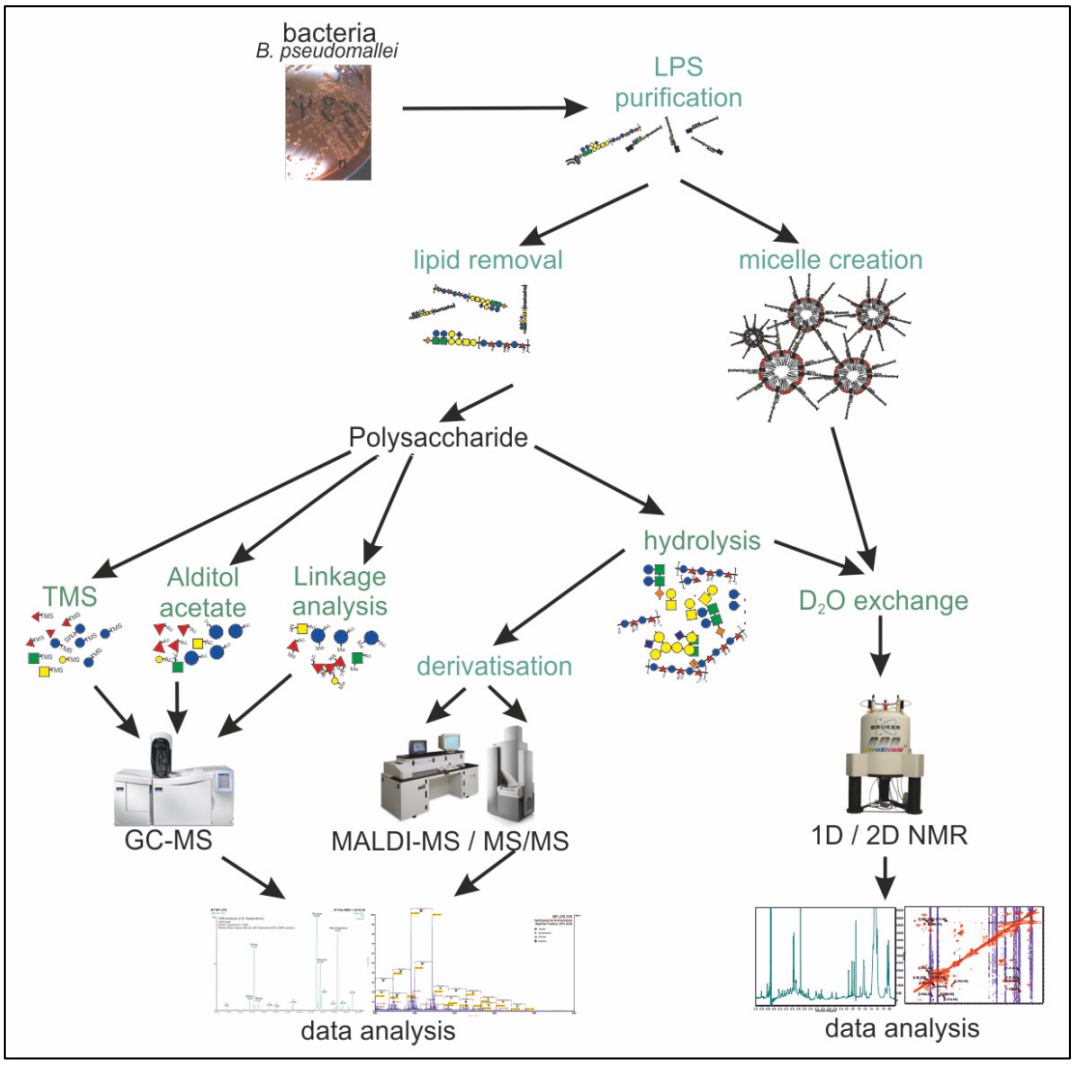


Figure 4.1 - Experimental strategy employed in O-antigen structural characterisation.
The schematic describes the methods employed in the elucidation of the O-antigen repeat unit of the *B. pseudomallei* LPS serotype 576

4.2 *B. PSEUDOMALLEI* LPS COMPOSITIONAL AND LINKAGE ANALYSIS

The *B. pseudomallei* LPS samples were analysed by TMS and alditol acetate analysis and run alongside commercial standards using the retention times of elution on the chromatogram as well as the specific mass spectrometric fingerprints of each monosaccharide in order to determine the monosaccharide composition of the O-antigen (**Figure 4.2** and **Figure 4.6** respectively). Linkage analysis was performed on the LPS and various linkage possibilities were identified *ab initio* using the mass spectrometric fingerprints and known fragmentation pathways as well as comparisons with the mass spectrometric fingerprints of standards (**Figure 4.10**).

The *B. pseudomallei* 576 LPS was also compared to the LPS of *B. pseudomallei* K96243, K96243 ΔCPS, 576 and *B. thailandensis*, all prepared by TMS, alditol acetate and linkage techniques and analysed by GC-MS alongside each other. This way the samples for *B. pseudomallei* strains K96243, 576, K96243 ΔCPS (K96423 without capsule) and *B. thailandensis* wild type strain were assessed for quality and composition and linkage (**Figure 4.12**, **Figure 4.13**, **Figure 4.14)**.

4.2.1 Compositional analysis of *B. pseudomallei* 576 LPS by TMS

The LPS of *B. pseudomallei* 576 was derivatised by the TMS technique described previously (**Section 2.4.8.6**). The samples were analysed by GC-MS fitted with an RTX-5MS column and run alongside commercially obtained standards encompassing a range of monosaccharides (**Figure 4.2**). Comparison of the two chromatograms as well as examination of the mass spectrometric fingerprints of each moiety eluted from the gas chromatography column allowed for full annotation of the chromatogram showing the monosaccharide composition of the *B. pseudomallei* 576 LPS. The monosaccharides present in the sample of *B. pseudomallei* 576 were shown to be the pentose xylose (elution time: 10.38min, **Figure 4.3**), deoxyhexose rhamnose (elution time: 9.15min, **Figure 4.4**), hexoses galactose and glucose (elution time: 13.68/14.43mins, **Figure 4.5**).

The deoxymannoheptose peak was assigned in comparison with other LPS chromatograms of another *B. pseudomallei* serotype K96243 (**Figure 4.12**), which were known to contain such a monosaccharide moiety as part of their capsule as described in literature (Knirel et al., 1992, Perry et al., 1995). The galactose moiety present in the chromatogram was investigated using an LPS free sample of *B. pseudomallei* 576 (*B. pseudomallei ΔLPS*) run alongside the wild type *B. pseudomallei* 576 LPS. This experiment illustrated that the galactose moiety arose from the background environment during the growth of the bacteria, as it was also found to be present in the TMS chromatograms of *B. pseudomallei ΔLPS* as well as L-broth medium used in the preparation of the sample (results not shown).

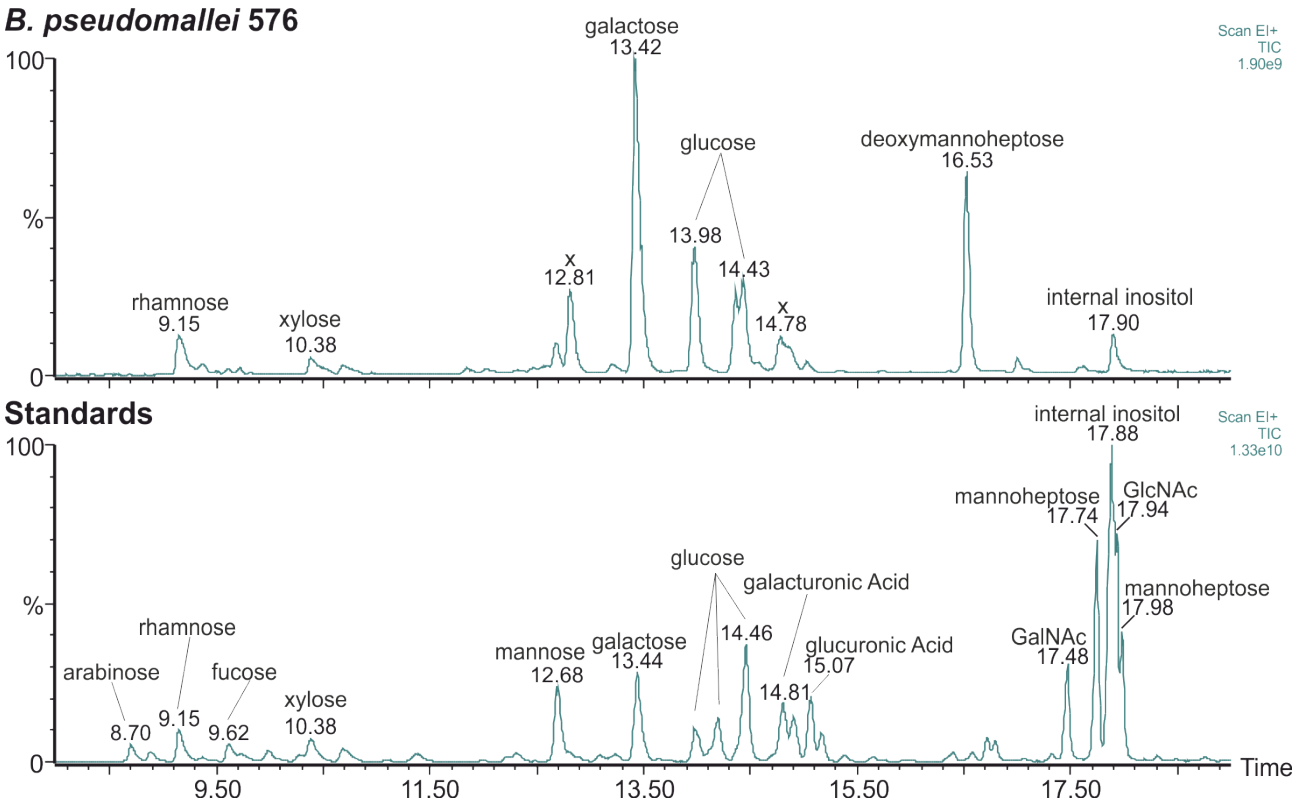


Figure 4.2 – GC-MS analysis of TMS derivatised LPS of *B. pseudomallei* 576.
The sample (top panel) was run alongside commercially obtained standards, at 50nM each (bottom panel) on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column. Both preparations were supplemented with 50nM of the internal standard inositol and non-sugar impurities present in the chromatogram are labelled x.

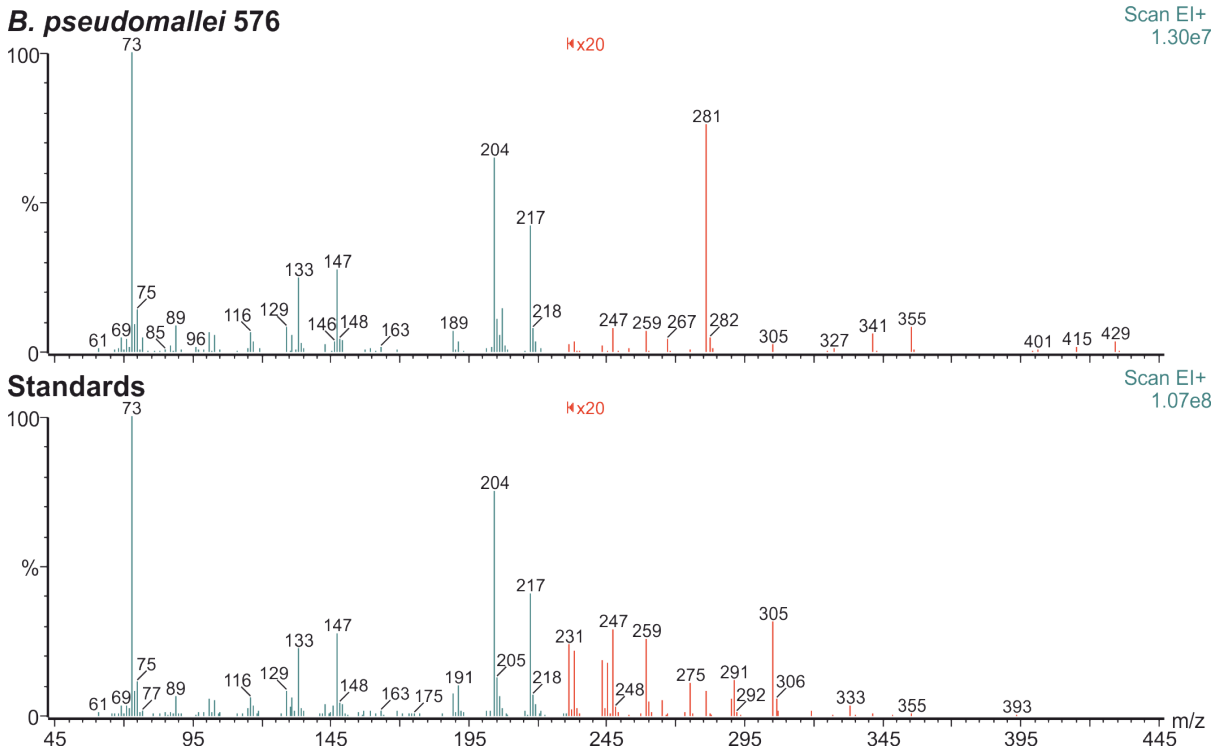


Figure 4.3 – Mass spectrometric fingerprint of xylose from the GC-MS chromatogram.
The mass spectrometric fingerprint of xylose was obtained from TMS derivatised LPS of *B. pseudomallei* 576 (top panel) and presented with that of the commercially obtained standard for xylose (bottom panel), as run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column. The fingerprint is magnified from m/z 231 onwards and is highlighted in red.

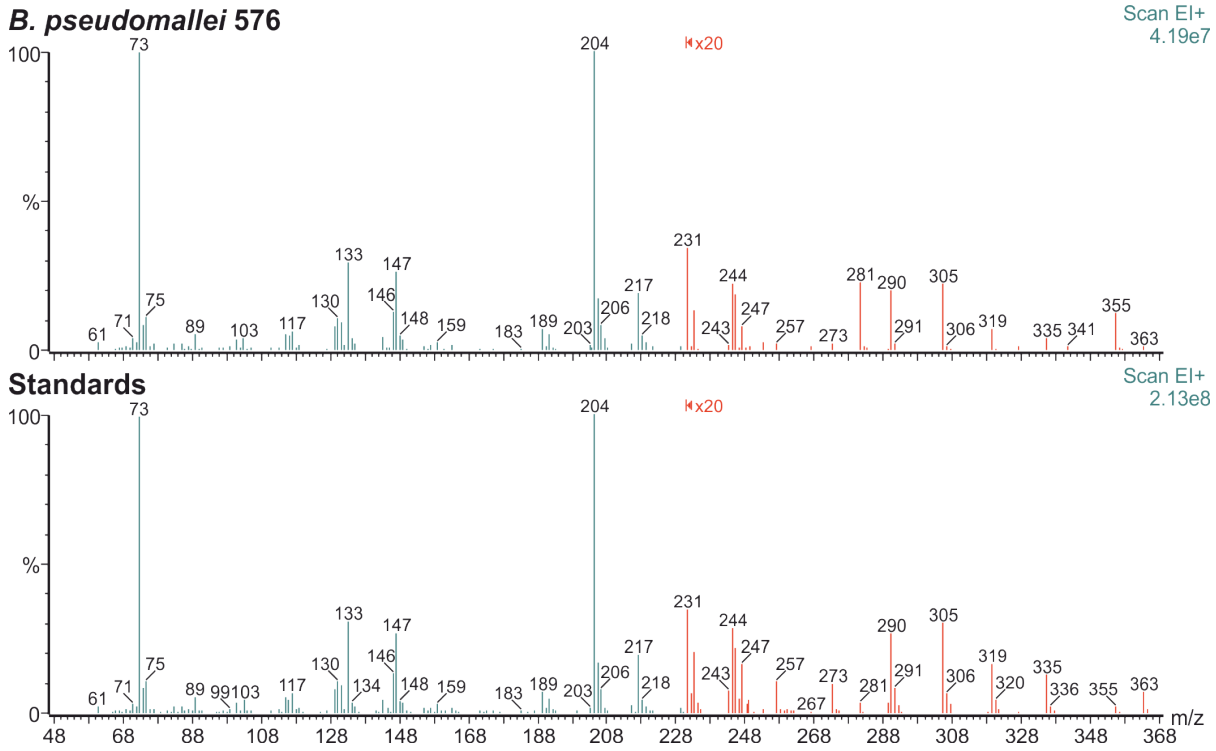


Figure 4.4 – Mass spectrometric fingerprint of rhamnose from the GC-MS chromatogram.
The mass spectrometric fingerprint of rhamnose was obtained from TMS derivatised LPS of *B. pseudomallei* 576 (top panel) and presented with that of the commercially obtained standard for rhamnose (bottom panel), as run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column. The fingerprint is magnified from m/z 231 onwards and is highlighted in red.

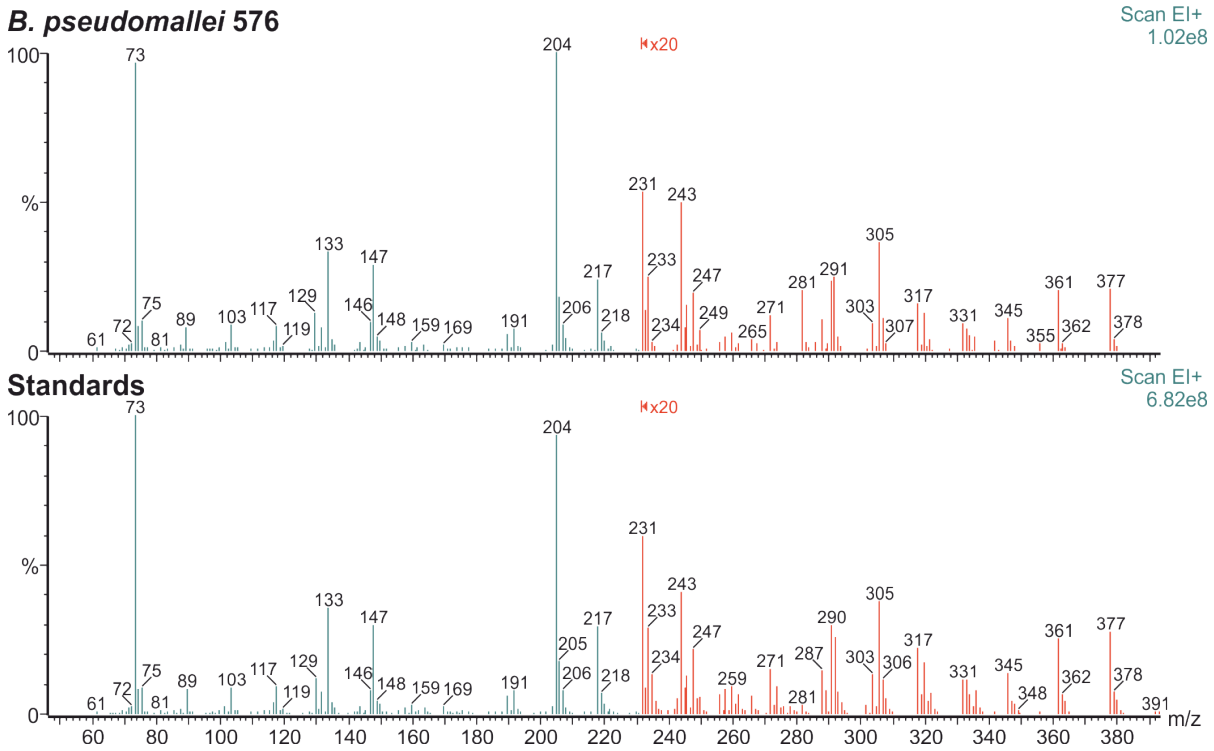


Figure 4.5 – Mass spectrometric fingerprint of glucose from the GC-MS chromatogram.
The mass spectrometric fingerprint of glucose was obtained from TMS derivatised LPS of *B. pseudomallei* 576 (top panel) and presented with that of the commercially obtained standard for glucose (bottom panel), as run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column. The fingerprint is magnified from m/z 231 onwards and is highlighted in red.

4.2.2 Compositional analysis of *B. pseudomallei* 576 LPS by the alditol acetate method

The LPS of *B. pseudomallei* 576 was hydrolysed and derivatised to form alditol acetates and analysed by GC-MS (**Section 2.4.8.7**) (**Figure 4.6**). Comparison of the two chromatograms as well as examination of the mass spectrometric fingerprints of each moiety eluted from the gas chromatography column allowed for full annotation of the chromatogram showing the monosaccharide composition of the *B. pseudomallei* 576 LPS. The mass spectrometric fingerprints (**Figure 4.7**, **Figure 4.8** and **Figure 4.9**) were assessed to determine the structures of the monosaccharide moieties as per the fragmentation and further losses of functional groups from the fragments, such as is observed by the loss of ketenes and acetate groups. The monosaccharides present in the sample of *B. pseudomallei* 576 were shown to be the pentose xylose (elution time: 19.94min, **Figure 4.7**), deoxyhexose rhamnose (elution time: 19.41min, **Figure 4.8**), and hexoses galactose and glucose (elution time: 23.71mins, **Figure 4.9**).

Similarly to the TMS chromatograms the deoxymannoheptose peak was assigned in comparison with *B. pseudomallei* K96243 chromatogram (**Figure 4.13**) and information from literature, whilst the galactose moiety seemed to arise from the sample preparation employed by our collaborators using L-broth media containing galactose, as was explained by an alditol acetate experiment using an LPS free sample of *B. pseudomallei* 576 (*B. pseudomallei ΔLPS*) run alongside the wild type *B. pseudomallei* 576 LPS (data not shown). The mannoheptose peak identified in the chromatogram is understood to potentially arise from the core portion of the LPS, which is known to contain mannoheptose moieties. The structure of the core polysaccharide of the *B. pseudomallei* 576 LPS has not been defined or published yet.

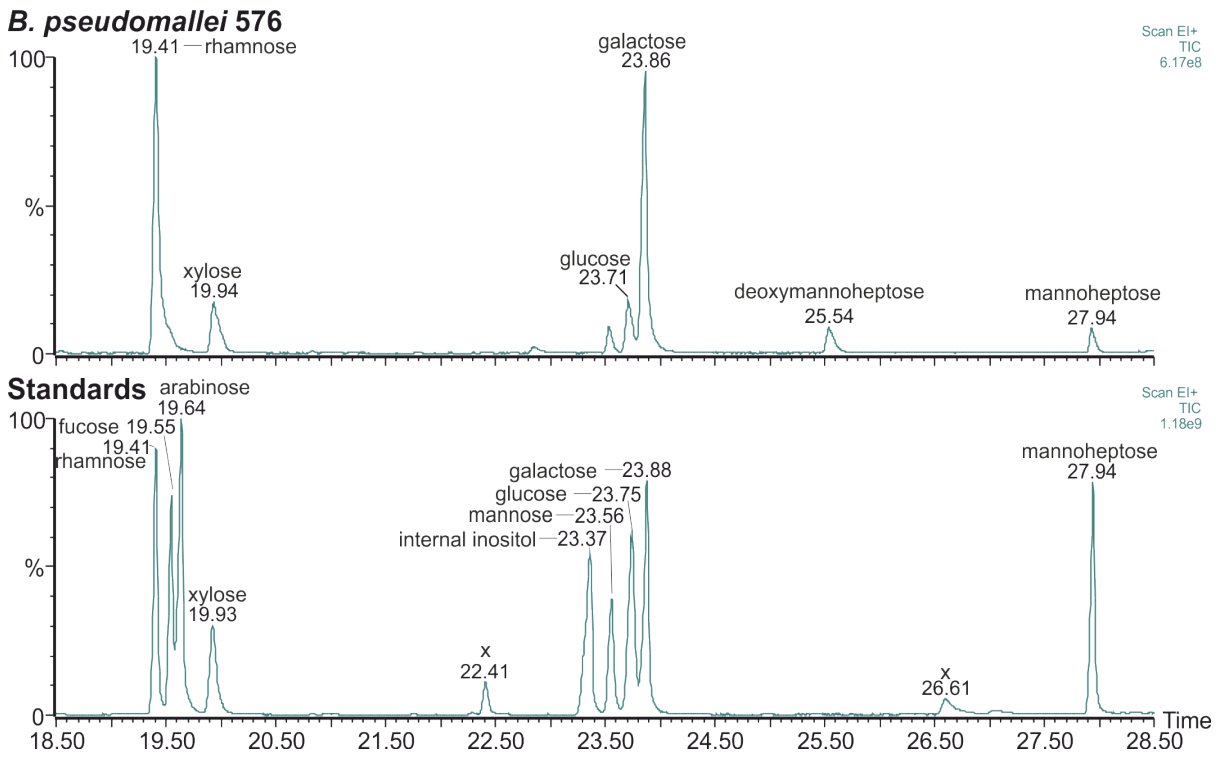


Figure 4.6 – GC-MS alditol acetate analysis of *B. pseudomallei* 576 LPS.
The sample (top panel) was run alongside commercially obtained standards, at 50nM each (bottom panel). The samples were run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column. The standards preparation was supplemented with 50nM of the internal standard Inositol. Non-sugar impurities present in the chromatogram are labelled x.

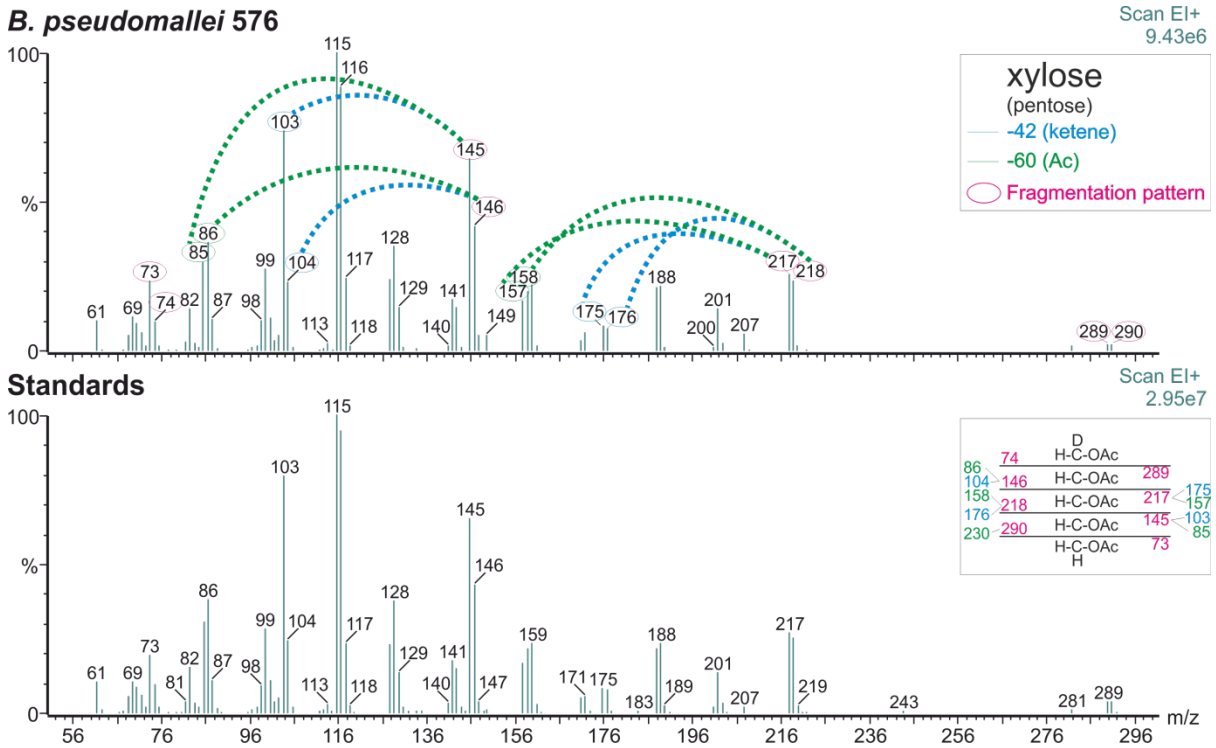


Figure 4.7 – Mass spectrometric fingerprint of xylose from the GC-MS chromatogram.
The mass spectrometric fingerprint of xylose was obtained from alditol acetate derivatives of the *B. pseudomallei* 576 LPS (top panel) and presented with that of the commercially obtained standard for xylose (bottom panel). The samples were run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column. The fragmentation pattern of the monosaccharide is annotated in the top panel, showing the main fragments (pink) of the molecule as well as sub-fragments arising from the loss of functional groups (blue and green). The structure of the derivatised monosaccharide is shown in a separate schematic (bottom panel).

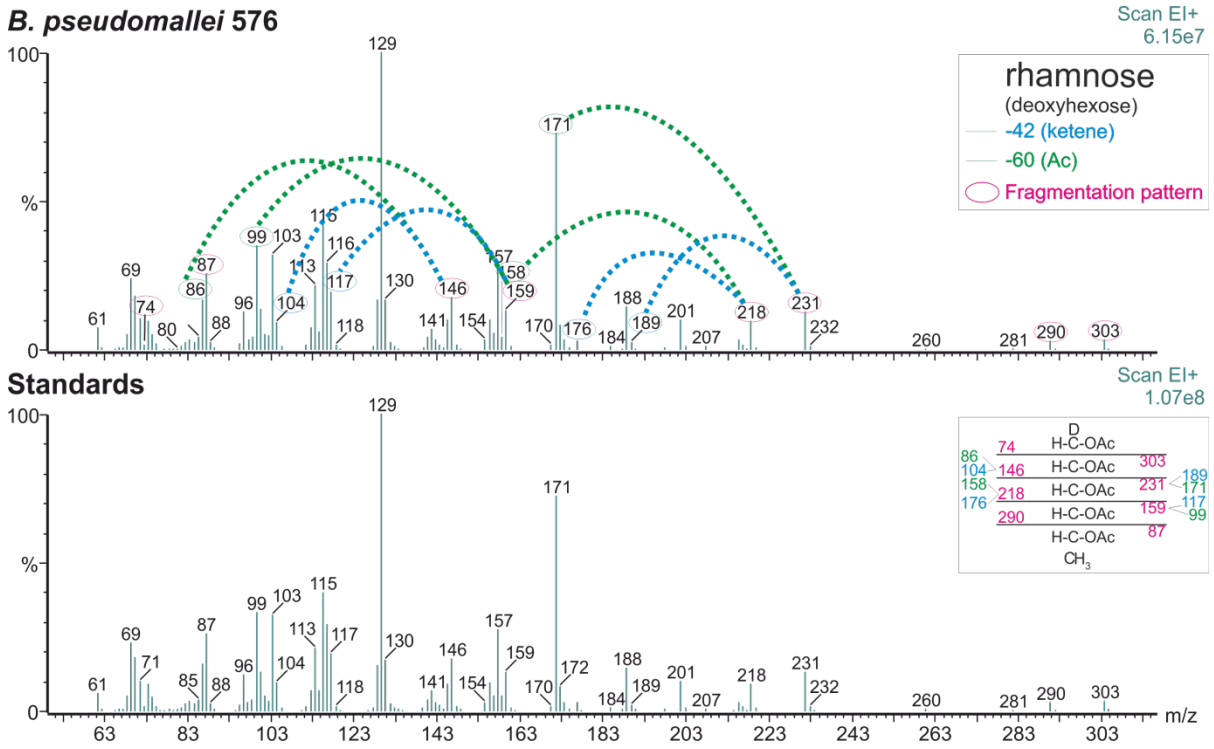


Figure 4.8 – Mass spectrometric fingerprint of rhamnose from the GC-MS chromatogram.
The mass spectrometric fingerprint of rhamnose was obtained from alditol acetate derivatives of the *B. pseudomallei* 576 LPS (top panel) and presented with that of the commercially obtained standard for rhamnose (bottom panel). The samples were run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column.The fragmentation pattern of the monosaccharide is annotated in the top panel, showing the main fragments (pink) of the molecule as well as sub-fragments arising from the loss of functional groups (blue and green). The structure of the derivatised monosaccharide is shown in a separate box (bottom panel).

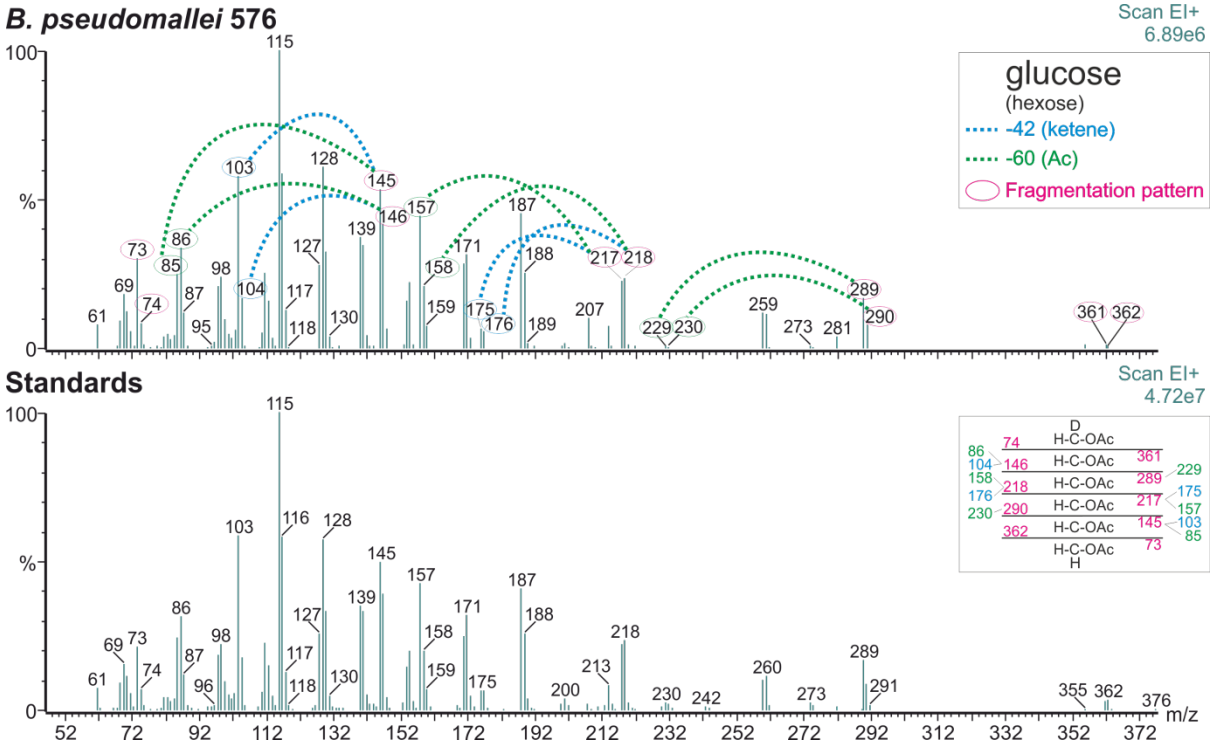


Figure 4.9 – Mass spectrometric fingerprint of glucose from the GC-MS chromatogram.
The mass spectrometric fingerprint of glucose was obtained from alditol acetate derivatives of the *B. pseudomallei* 576 LPS (top panel) and presented with that of the commercially obtained standard for glucose (bottom panel). The samples were run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column. The fragmentation pattern of the monosaccharide is annotated in the top panel, showing the main fragments (pink) of the molecule as well as sub-fragments arising from the loss of functional groups (blue and green). The structure of the derivatised monosaccharide is shown in a separate box (bottom panel).

4.2.3 Linkage analysis of *B. pseudomallei* 576 LPS

Linkage analysis allows for the determination of the carbon atoms involved in a linkage, where free hydroxyl groups in the glycan are substituted by an ether-linked methyl group. The glycosidic linkages are then hydrolysed by acid and the implicated hydroxyl groups can be further acetylated to produce partially methylated alditol acetates (Dell and Morris, 2001, McNeil et al., 1982). The *B. pseudomallei* LPS was derivatised into partially methylated alditol acetates to provide insight into the potential linkages present in the O-antigen repeat unit (**Section 2.4.8.8**). The samples were analysed by GC-MS fitted with an RTX-5MS column (**Figure 4.10**). Examination of the mass spectrometric fingerprints of each moiety eluted from the gas chromatography column allowed for full annotation of the chromatogram showing the monosaccharide linkages of the *B. pseudomallei* 576 LPS. The mass spectrometric fingerprints, such as is demonstrated for the terminal pentose, 2-linked deoxyhexose, 3,4-linked deoxyhexose and 3-linked hexose (**Figure 4.11**), were assessed to determine the linkages of the monosaccharide moieties as per the fragmentation and further losses of functional groups from the fragments, such as is observed by the loss of ketenes and acetate groups. The linkage data revealed a terminal pentose and a terminal hexose moiety. The deoxyhexose linkages include 2-, 3-, 4-, 3,4- and 2,3-linked moieties. Further hexose moieties show 3-, 6- and 2,6-linkages. Compared with what would be expected for the typical strain K96243 (serotype A) LPS (**Figure 4.14a**), the serotype B 576 LPS chromatogram is remarkably complex, with double linkages indicating a potentially branched structure. It is to be noted, though, that some of the annotated chromatogram peaks will have arisen from monosaccharide linkages found in the LPS core, so further information is required to fully annotate the linkage arrangement of the O-antigen of the LPS.

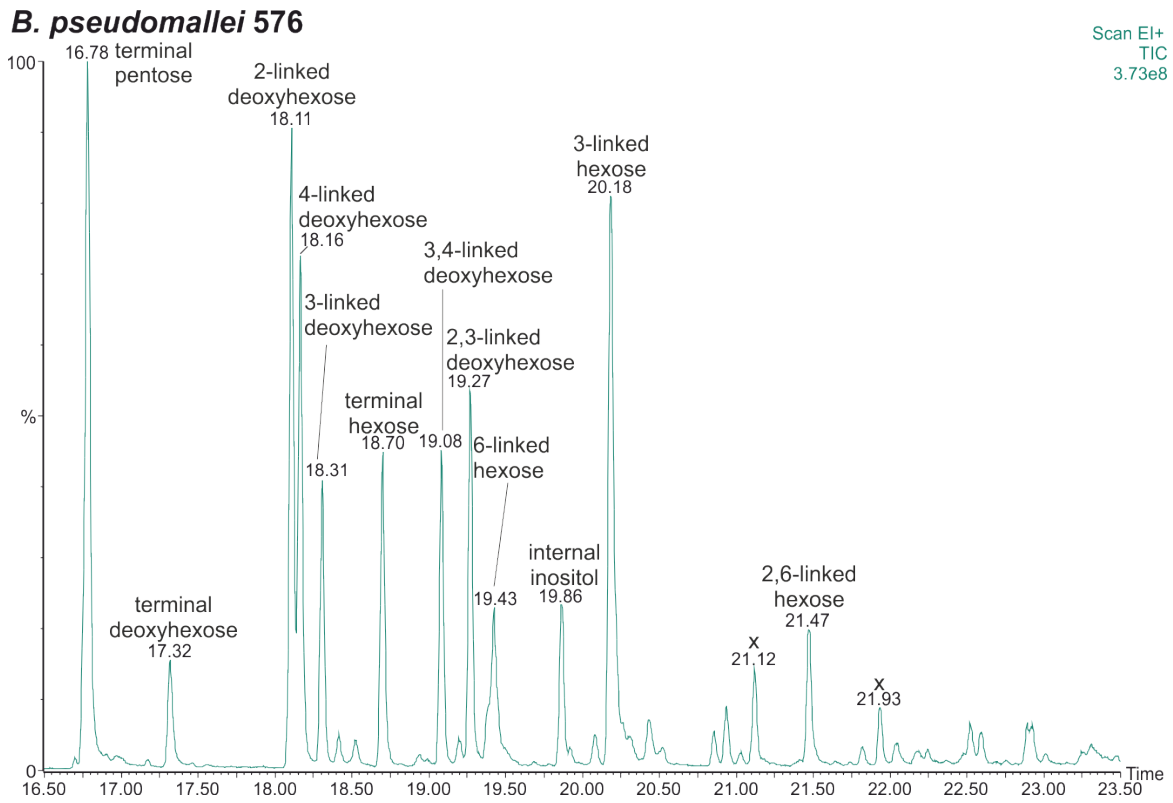


Figure 4.10 – GC-MS linkage analysis of *B. pseudomallei* 576 LPS with 50nM of inositol. The samples were run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column. The annotation was completed using the mass spectrometric fingerprints and fragmentation pathways of the monosaccharide structures. Non-sugar impurities present in the chromatogram are labelled x.

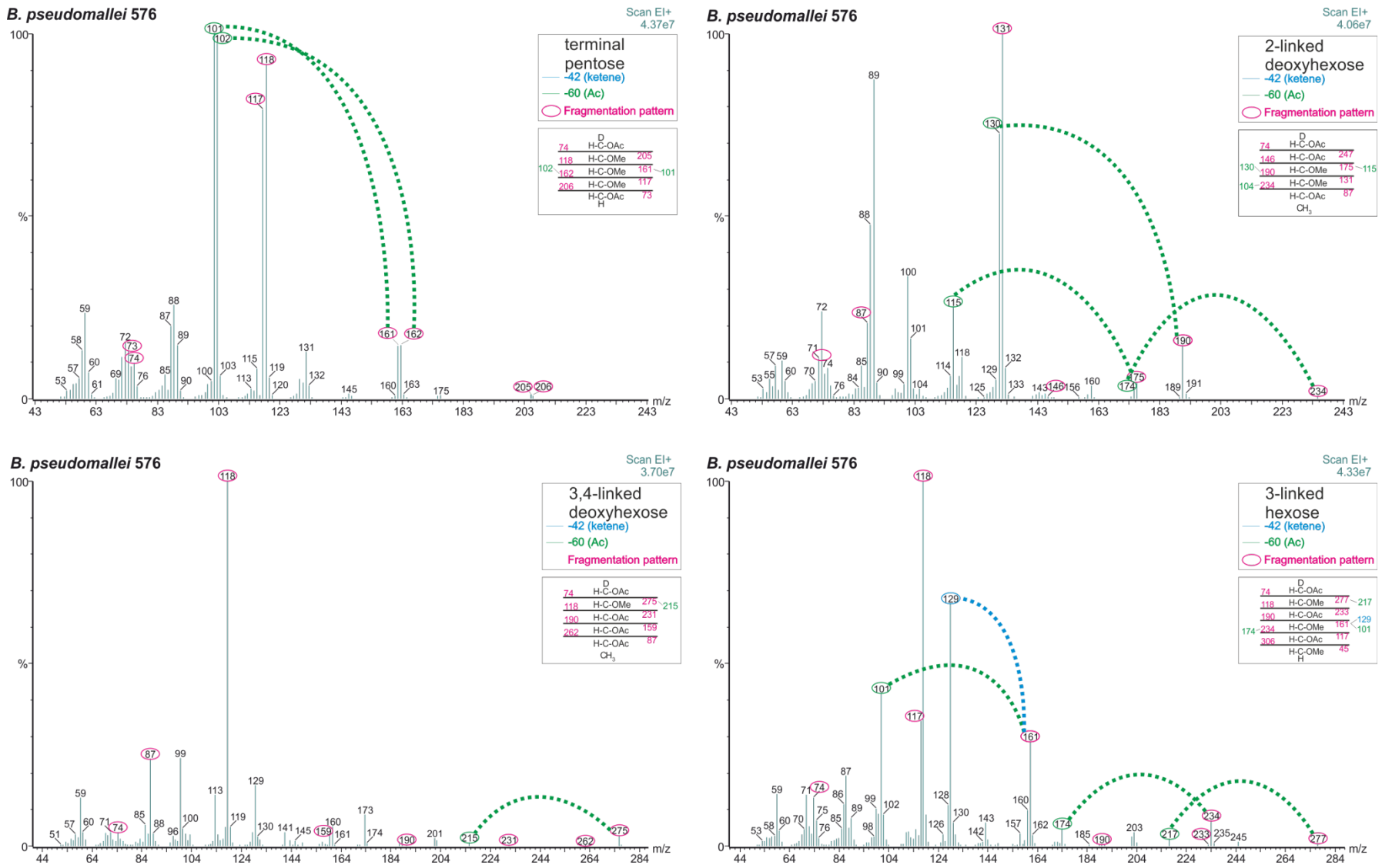


Figure 4.11 – Mass spectrometric fingerprints of linkage signals.
The mass spectrometric fingerprints of a terminal pentose, 2-linked deoxyhexose, 3,4-linked deoxyhexose and 3-linked hexose were obtained from the GC-MS chromatogram of the linkage analysis of the *B. pseudomallei* 576 LPS, run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column. The fragmentation patterns of the monosaccharides are annotated, showing the main fragments (pink) of the molecule as well as sub fragments arising from the loss of functional groups (blue and green). The structures of the derivatised monosaccharides are shown in a separate box.

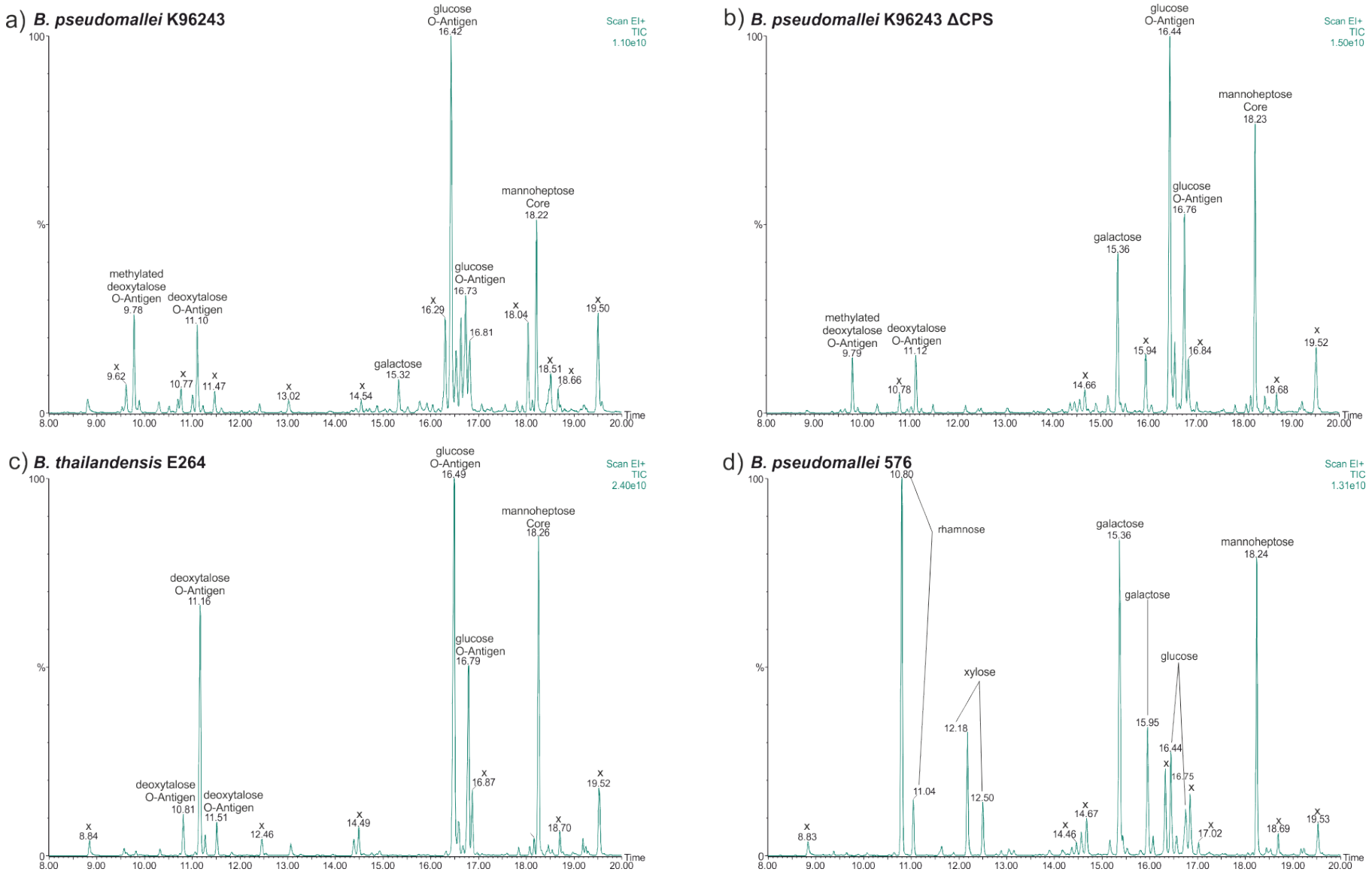


Figure 4.12 – GC-MS analysis of TMS derivatised LPS from *B. pseudomallei* and *B. thailandensis*.
The chromatograms for the wild type and capsule mutant of strain K96423 are shown in (a) and (b) respectively. The chromatogram for the *B. thailandensis* wild type is shown in (c) and that of *B. pseudomallei* strain 576 in (d). All samples were run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column. Non-sugar impurities present in the chromatogram are labelled x.

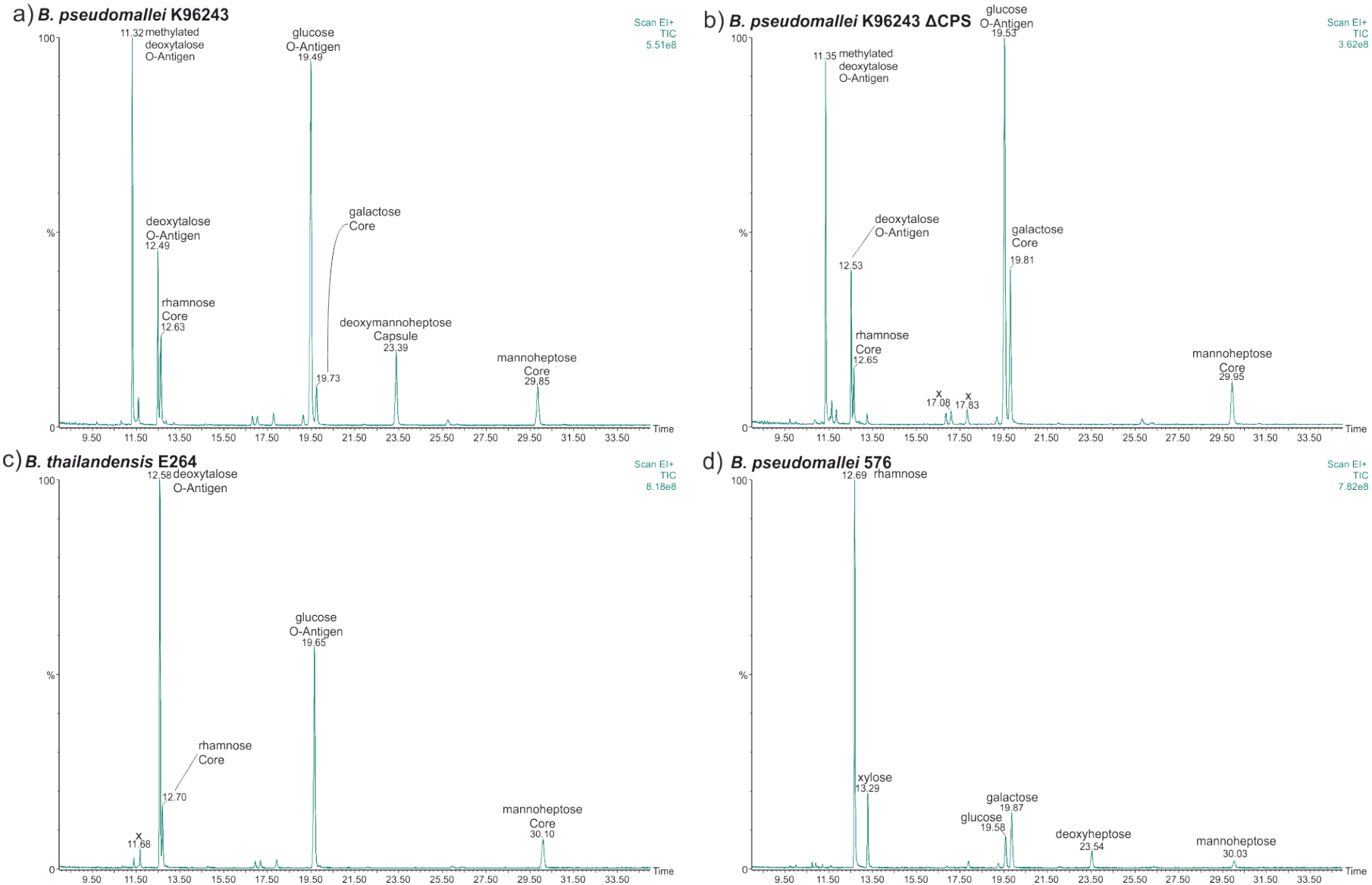


Figure 4.13 – GC-MS alditol acetate analysis comparisons.
The figure presents alditol acetate chromatograms for *B. pseudomallei* strains K96423 (a), capsule mutant (b) and 576 (d) as well as *B. thailandensis* (c). All samples were run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column. Non-sugar impurities present in the chromatogram are labelled x.

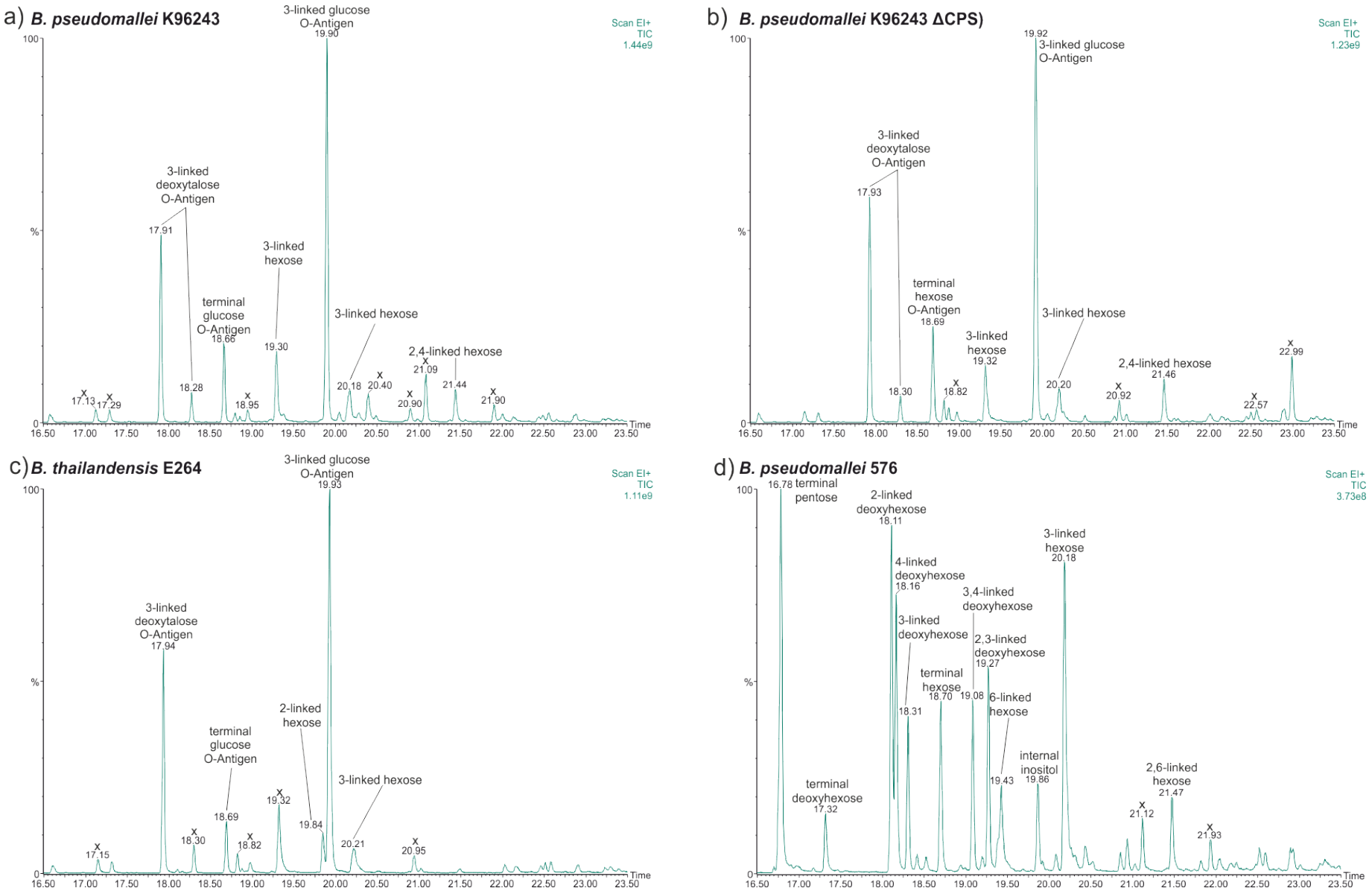


Figure 4.14 – GC-MS of linkage analysis comparisons.
The figure presents alditol acetate chromatograms for *B. pseudomallei* strains K96423 (a), capsule mutant (b) and 576 (d) as well as *B. thailandensis* (c).). All samples were run on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a RTX-5MS column. Non-sugar impurities present in the chromatogram are labelled x.

4.2.4 Comparison of 576 LPS composition and linkage with related strains

The derivatised *B. pseudomallei* 576 LPS was further compared to the derivatised LPS moieties of related *B. pseudomallei* and *B. thailandensis*. All LPS were analysed by GC-MS as TMS, alditol acetate and partially methylated alditol acetate derivatives in order to assess the composition and linkage of the serotype B (576) as it differs from related serotype A LPS (*B. pseudomallei* K96243 and *B thailandensis* E264). The samples analysed included the LPS of *B. pseudomallei* K96243, K96243 *ΔCPS* (K96243 without capsule), 576 and *B. thailandensis* E264.

4.2.4.1 Comparisons of TMS findings

The evaluated chromatograms indicate the composition of the *B. pseudomallei* samples consisting of deoxytalose and glucose in both the K96423 wild type (**Figure 4.12a**) and capsule mutant (**Figure 4.12b**) of that same strain. Peaks for both deoxytalose and methylated deoxytalose are observable in the *B. pseudomallei* samples. All chromatograms for *B. pseudomallei* K96243, K96243 *ΔCPS* as well as *B. thailandensis* (**Figure 4.12c**) show compositions consistent with published data for serotype A (Brett et al., 1998, Knirel et al., 1992, Perry et al., 1995). Galactose was identified and is thought to arise from either background carbohydrate contamination from L-broth medium, however it is to be noted that the core of the *B. pseudomallei* LPS has not yet been identified and thus galactose could potentially be present in the LPS core. The composition of the *B. pseudomallei* 576 (**Figure 4.12d**), was previously identified to contain rhamnose and glucose. The mannoheptose moiety is most likely from the core unit of the LPS. The data described earlier for strain 576 confirmed the presence of the pentose moiety, xylose, which was not observed in the *B. pseudomallei* strain K96423 or *B. thailandensis* LPS. Furthermore the glucose moiety is present at equal relative ratios to talose in strain K96243, in strain 576 the rhamnose moiety is relatively more abundant as compared to the rest of the monosaccharide moieties identified.

4.2.4.2 Comparisons of alditol acetate analysis

The chromatograms for *B. pseudomallei* and *B. thailandensis* serotype A LPS support the compositional understanding from the TMS results and confirm the published structure for *B. pseudomallei* K96423 (**Figure 4.13a**). Again, two separate peaks for the deoxytalose moiety are identified, with slight differences in their mass spectrometric fingerprints, due to partial 2-O-methylation, as suggested in literature (Reckseidler et al., 2001). Peaks for rhamnose, galactose and mannoheptose, presumably from the LPS core are also present, or in the case of galactose also from the growth media. As suggested by the TMS results, glucose is present at a similarly high intensity as the deoxytalose moiety to make up the repeating heteropolymer unit of the O-antigen. A deoxymannoheptose peak can be seen around 23.39mins on the chromatogram as the main component of the capsule. The *B. pseudomallei* capsule mutant (**Figure 4.13b**) shows many similarities to that of its wild type, but is missing the deoxymannoheptose peak representing the capsule, similar to the *B. thailandensis* (**Figure 4.13c**) chromatogram where the peak for the capsule sugar moiety is also absent. The *B. pseudomallei* strain 576 (**Figure 4.13d**) chromatogram shows a rhamnose peak as well as a xylose peak. There is also a capsule sugar component as shown by the peak around 23.54mins, which has a deoxymannoheptose mass spectrometric fingerprint. The rhamnose peak is much larger when compared to

the deoxyhexose moieties in compositional analysis of the other LPS samples, suggesting increased presence in the O-antigen repeat unit.

4.2.4.3 LPS linkage analysis and comparisons

Linkage analysis on the *Burkholderia* samples gave insight into the organisation of the repeating oligosaccharide units making up the O-antigen chain on each strain and serotype. The *B. pseudomallei* wild type O-antigen forms a disaccharide unit of 3-linked glucose with 3-linked deoxytalose linked to the core structure of the LPS via the deoxytalose unit. The non-reducing terminal glucose is also observed in the chromatogram (**Figure 4.14a**). The linkage data of the capsule mutant *B. pseudomallei* LPS (**Figure 4.14b**) is similar to its wild type and the repeating heteropolymer unit is made up of 3-linked glucose and 3-linked deoxytalose components, with the latter linked to the core element of the LPS. The *B. thailandensis* wild type strain (**Figure 4.14c**) supports a similar composition, where 3-linked glucose and 3-linked deoxytalose are prominent in the chromatogram, accompanied by the terminal glucose peak that suggests the deoxytalose linkage to the core unit. The core structure of these LPS strains has not yet been characterised and thus the 2-, 2,4-linked and 3-linked hexose moieties in excess of these assignments cannot be assigned precisely. It has been established that galactose is present in these samples as per the sample preparation procedures used by our collaborators however, due to the possibility of as of yet unidentified moieties arising from core signals, these cannot be definitively assigned. The complex linkage chromatogram of *B. pseudomallei* 576 strain (**Figure 4.14d**) supports initial findings from TMS and alditol acetates about the complexity of the O-antigen heteropolymer, with a terminally linked pentose, deoxyhexose as well as a terminal hexose evident in the chromatogram. Furthermore, the deoxyhexoses seem to be 2-, 3-, 4-, 2,3- and 3,4-linked, of which some of the signals will most probably be determined as arising from the rhamnose moiety in the O-antigen repeat unit, but care has to be taken as some signals will possibly have arisen from core linkages. The main hexose peaks show a 3-linked and 6-linked hexose mass spectrometric fingerprint, while the 2,6-linked hexose peak is proposed to arise from the core unit linkages or the galactose polymer arising from the growth medium.

4.2.5 Compositional and linkage analysis summary and discussion

The purified samples of *B. pseudomallei* and *B. thailandensis* were analysed for sugar composition as TMS derivatised monosaccharides in order to assess the quality of the sample and get an idea of the composition of the O-antigen sugars in the unknown strain *B. pseudomallei* 576 O-antigen. The two peaks observed for deoxytalose shown in the *B. pseudomallei* chromatograms are consistent with the discussed literature, where it was found that 33% of the deoxytalose residues have been shown to contain 2-O-methyl and 4-O-acetyl substituents, while the other deoxytalose residues contain only 2-O-acetyl substituents (Brett et al., 2003, Reckseidler et al., 2001). While the assignment of the known structures proved the sample's suitable quality, the unknown strain 576 LPS of *B. pseudomallei*, showed relatively higher variation in monosaccharide peaks, as compared to K96423 and *B. thailandensis*. Xylose was observed in strain 576 but not in any of the other samples, indicating a more complex repeat unit for this strain.

The alditol acetate information further substantiates the published data on the *B. pseudomallei* K96243 LPS composition and monosaccharide modifications, as seen by the distinct deoxytalose peaks for the 2-O-methylated and non-methylated units, while the prepared minimal capsule sample seems identical, bar the capsule deoxymannoheptose moiety. Furthermore, the experiments reinforce the notion that the *B. thailandensis* LPS has a similar enough structure to be potentially used for protection studies, as suggested by complementary immunological studies (Ngugi et al., 2010). The differences seem to lie in the missing capsule moiety, most likely contributing to its lower virulence nature. A summary of the proposed composition of the O-antigens investigated is shown in **Table 4.1**.

The *B. pseudomallei* strain K96243 wild type and minimal capsule strain LPS both show matching linkage chromatograms, consistent with the published data on the heteropolymer illustrating it to consist of 3-linked glucose and 3-linked deoxytalose to make up the O-antigen part of the LPS (Knirel et al., 1992, Perry et al., 1995). The linkage information on the *B. thailandensis* LPS, shows a similarly simple chromatogram, differing in the presence of a 2-linked hexose peak, which most likely derives from the core polysaccharide. On the other hand, the *B. pseudomallei* strain 576, shows linkage data of higher complexity compared to the other samples. The xylose unit revealed from the TMS and alditol acetate analysis presents itself as a terminal sugar in the O-antigen polymer, in addition to the terminal deoxyhexose and hexose moieties. This leads to the conclusion that the repetitive oligosaccharide unit, unlike the typical LPS of the K96243 strain, is branched; as is further supported by the possible 2,3- or 3,4-linked deoxyhexose peak, providing the basis for such branching on the repeat unit of the O-antigen.

***Burkholderia* sample**	**Suggested sugar Composition**
B. pseudomallei K96243 LPS & CPS	O-antigen: glucose and deoxytalose Core: rhamnose, galactose, mannoheptose Capsule: deoxymannoheptose
B. pseudomallei K96243 LPS only (*ΔCPS*)	O-antigen: glucose and deoxytalose Core: rhamnose, galactose, mannoheptose Capsule: N.A.
B. pseudomallei 576 LPS & CPS	O-antigen: xylose , rhamnose, glucose Core: mannoheptose Capsule: deoxymannoheptose
B. thailandensis WT LPS	O-antigen: glucose and deoxytalose Core: rhamnose, mannoheptose Capsule: N. A.

Table 4.1 – Suggested sugar compositions of *Burkholderia* samples.

The suggested difference in deoxyhexose monosaccharide subunits employed in the O-antigen repeat unit by the two distinct serotypes A (*B. pseudomallei* K96243) and B (*B. pseudomallei* 576), calls into question why and how such a change would evolve in two such closely related serotypes of the same bacterium. Generally, the biosynthesis of 6-deoxyhexoses occurs from a nucleotidylated hexose, such as dTDP-glucose by dTDP-glucose 4,6-dehydratase (RmlB) from which diverse deoxyhexoses can be generated, including dTDP-6-deoxytalose and dTDP-rhamnose (Thibodeaux et al., 2008). The biosynthesis of rhamnose requires the actions of glucose-1-phosphate thymidyltransferase (RmlA), dTDP-glucose 4,6-dehydratase (RmlB), dTDP-4-hexulose 3,5 epimerase (RmlC) and finally dTDP-6-deoxy-lyxo-4-hexulose reductase / 4-KR (RmlD). The biosynthetic gene cluster in *B. pseudomallei* K96243 and 576 (serotypes A and B) contain the dTDP-rhamnose biosynthesis genes of rmlABCD (**Figure 1.11**), however not the 4-KR enzymes appropriate for the creation of dTDP-6-deoxytalose from it's the dTDP-6-deoxy-xylo-4-hexulose. 6-deoxytalose, however is the

C4 epimer of rhamnose and it was therefore proposed that an epimerase (WbiB) was responsible for the synthesis of dTDP-6-deoxytalose from dTDP-rhamnose (DeShazer et al., 1998). Furthermore, recent studies on *B. thailandensis* demonstrated epimerase activity that interconverted dTDP-rhamnose and dTDP-6-deoxytalose (Yoo et al., 2011). This suggests an explanation for such a switch in monosaccharide incorporation with a possible lack of the *wbiB* gene present in the *B. pseudomallei* 576 serotype B biosynthesis cassette. This would further be support the *B. pseudomallei* evolutionary model, whereby both gene acquisition and loss occurs in horizontal transfer events, as was previously discussed in **Section 1.3.4**.

4.3 MS SEQUENCING OF THE *B. PSEUDOMALLEI* LPS O-ANTIGEN REPEAT

As previously discussed, the size and amphiphilic nature of LPS makes them difficult to analyse intact by mass spectrometry. Structure determination is usually attained by independent analyses of the lipid A and carbohydrate moieties after their separation by hydrolysis at the Kdo linkage. In favourable cases, MALDI data from intact LOS has been possible to be acquired (Silipo et al., 2006). Unfortunately we found that *B. pseudomallei* 576 LPS was refractory to direct MS analysis. Thus when samples were analysed using linear MALDI-MS in a sinapinic acid matrix in both the positive and negative ion mode, no molecular ions were observed (results not shown). Therefore it was necessary to hydrolyse the polysaccharide into smaller pieces prior to MALDI-MS. In this section the various hydrolytic conditions that were tested for optimum hydrolysis of the polysaccharides are described together with the resulting MALDI-MS and MS/MS data (**Section 2.4.7**). Furthermore, to understand the efficiency of lipid A removal by each of the methods, we analysed the samples with their respective non-hydrolysed samples as standards by FAME analysis and compared them with standards for the acyl chains (**Section 2.4.8.9**).

4.3.1 Mild acid hydrolysis

Mild acid hydrolysis is widely used to liberate the core and O-antigen polysaccharide from the Lipid A by selective cleavage of the acid-labile glycosidic linkage between Kdo in the core oligosaccharide and the lipid A (**Section 2.4.7.1**). To examine the suitability of mild acid hydrolysis to release the lipid A from the rest of the polysaccharide, the mild acid hydrolysed LPS moieties from *B. pseudomallei* 576 and *B. thailandensis* were evaluated by fatty acid methyl ester analysis. This allowed for comparison of lipid moieties remaining after mild acid hydrolysis and centrifugation to separate the polysaccharide moieties from the lipid A. Whilst the O-antigen of various strains of the same species of bacteria can show extreme differences in structures, the lipid A and core unit of the LPS are generally quite conserved, suggesting *B. thailandensis* to be a suitable comparison for such analyses with *B. pseudomallei* 576.

B. pseudomallei 576 and *B. thailandensis* wild type samples were compared to the same sample post mild 1% acetic acid hydrolysis. The FAME analysis for *B. thailandensis* mild acid hydrolysis (**Figure 4.15**), showed the absence of 3-hydroxyl fatty acids present at 20.86mins (C14 3:OH) and 24.11mins (C16 3:OH) post 1% acetic acid hydrolysis. The selected ion chromatograms for the ion m/z 103, which is specific for the 3-hydroxy fatty acids, confirmed the absence of the methyl esters post hydrolysis. In contrast, the FAME analysis for *B. pseudomallei* 576 mild acid hydrolysis (**Figure 4.16**) showed methyl esters of C14 3:OH and C16 3:OH at 20.87mins and 24.10mins respectively, even after hydrolysis with 1% acetic acid. Experiments

with *B. pseudomallei* 576 in comparison with *B. pseudomallei* K96243 as well as increased concentrations of acetic acid used for incubation showed similar results (data not shown). This suggests that mild acid hydrolysis and centrifugation does not ensure the loss of the lipid A moiety from the 576 LPS, despite the conditions used being effective for cleavage of the other LPS samples examined.

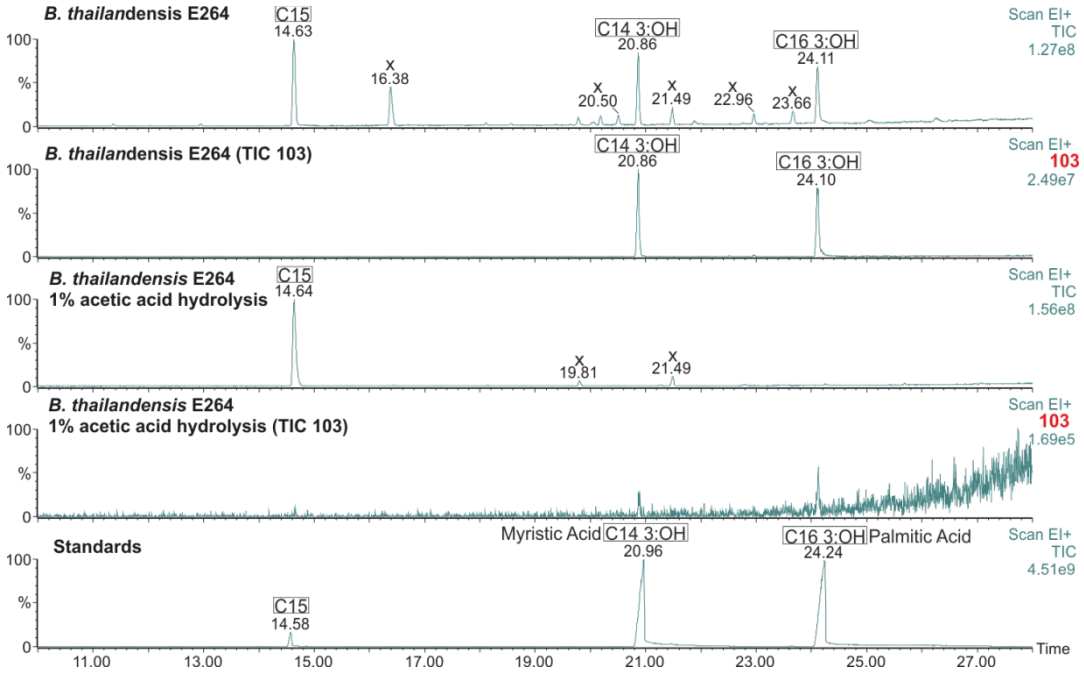


Figure 4.15 – FAME analysis chromatograms of *B. thailandensis* E264.
The chromatograms presented are of *B. thailandensis* E264 and post 1% acetic acid hydrolysis. The sample was run alongside commercially obtained standards (C15, C14 3:OH and C16 3:OH) on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a Stabilwax column. The total ion chromatograms for m/z 103 are specific to the 3-hydroxy fatty acids present in the lipid A. Non-FAME impurities present in the chromatogram are labelled x.

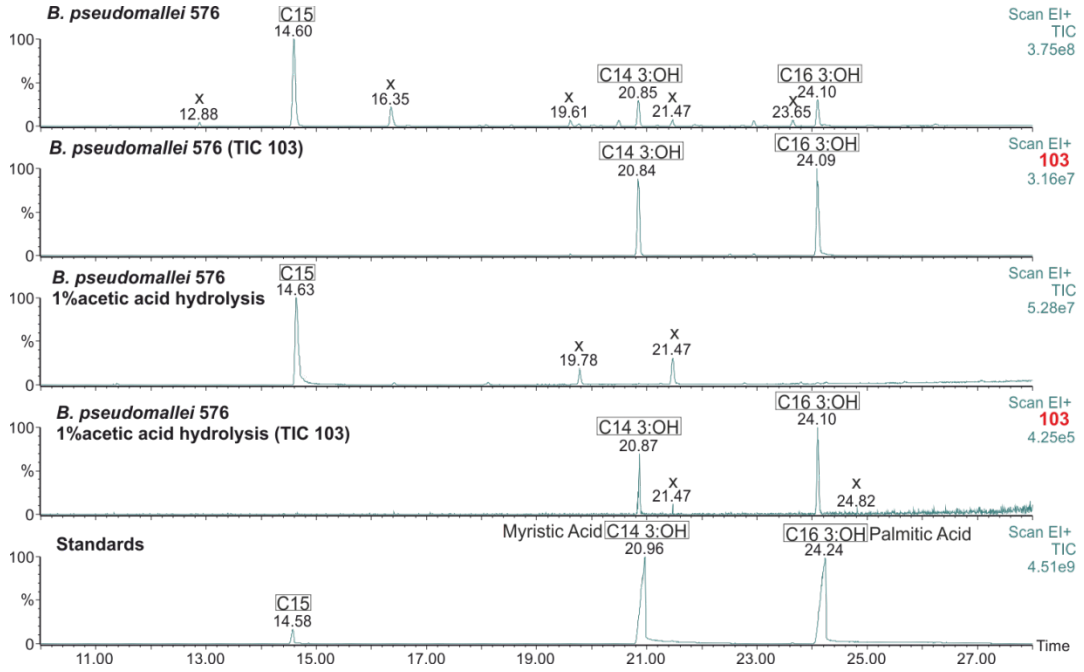


Figure 4.16 – FAME analysis chromatograms of *B. pseudomallei* 576.
The chromatograms presented are of *B. pseudomallei* 576 and post 1% acetic acid hydrolysis. The sample was run alongside commercially obtained standards (C15, C14 3:OH and C16 3:OH) on a Perkin Elmer Clarus 500 GC-MS instrument fitted with a Stabilwax column. The total ion chromatograms for m/z 103 are specific to the 3-hydroxy fatty acids present in the lipid A. Non-FAME impurities present in the chromatogram are labelled x.

4.3.2 Methanolysis

Mild acid hydrolysis is generally understood to be too mild a hydrolysis procedure to fully fragment O-antigen polysaccharides into their respective repeating units. However, the procedure can be very useful in cases where acid labile linkages are present. In this instance, MALDI-MS showed no data indicative of hydrolysis fragments (results not shown). These results suggest that the O-antigen polysaccharides lack acid-sensitive glycosidic linkages. Previous investigation into hydrolysis of the O-antigen polysaccharides of the *B. pseudomallei* K96243 (serotype A) by methanolysis showed the heteropolymer repeating units. The same hydrolysis conditions, however, achieved no fragmentation for *B. pseudomallei* 576 (serotype B) by mass spectrometry (**Section 2.4.7.2**). Hydrolysis of the samples investigated by methanolysis at 1M and 0.1M MeOH-HCl analysed at 15, 30, 45, 60, 75 and 90min intervals at room temperature provided no additional information (results not shown).

Figure 4.17 shows the MALDI spectrum for *B. pseudomallei* K96243 hydrolysed at 50ºC in 0.25M MeOH-HCl for 15mins, showing the heteropolymer units made up of glucose and deoxytalose. While hydrolysis conditions to achieve effective hydrolysis on *B. pseudomallei* K96243 and *B. thailandensis* E264 strains were established, it proved far more difficult to break the *B. pseudomallei* 576 into fragments observable by MS. The other samples for *B. pseudomallei* K96243, the capsule mutant and the *B. thailandensis* wild type, showed similar hydrolysis fragments using the above mentioned conditions (results not shown). The *B. pseudomallei* 576 sample, however, did not show any hydrolysis pattern, despite variations in temperature, optimal MeOH-HCl concentrations and reaction times investigated (results not shown).

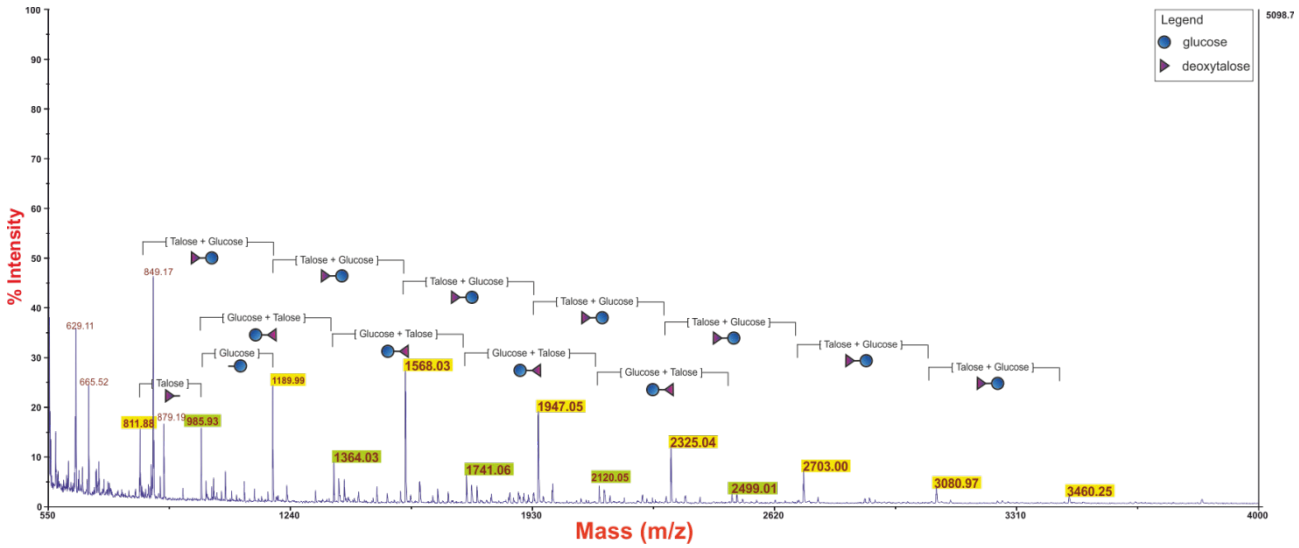


Figure 4.17 – MALDI-MS of *B. pseudomallei* LPS K96243 post methanolysis.
The MALDI-MS spectrum above shows the *B. pseudomallei* K96243 LPS, post hydrolysis with methanolysis in 0.25M MeOH-HCl for 15mins at 50ºC and permethylation. The data was accumulated on a PerSeptive Biosystems Voyager-DE STR System MALDI mass spectrometer. The O-antigen repeat unit of the repeating glucose-deoxytalose polymer is annotated.

4.3.3 Ammonolysis

As previous hydrolysis methods proved ineffective in harvesting suitably sized polysaccharide repeating units for analysis by MALDI-MS of the atypical LPS *B. pseudomallei* 576, ammonolysis was employed (**Section 2.4.7.3**). This is a process used to isolate the polysaccharide while maintaining the fidelity of the

repeating structure, in the case of polysaccharides containing certain more susceptible glycosidic linkages. The *B. pseudomallei* 576 strain LPS was hydrolysed using ammonolysis and the MALDI-MS data shown in **Figure 4.18** and **Figure 4.19**, which confirmed compositional information and given a greater insight into the complexity of the O-antigen.

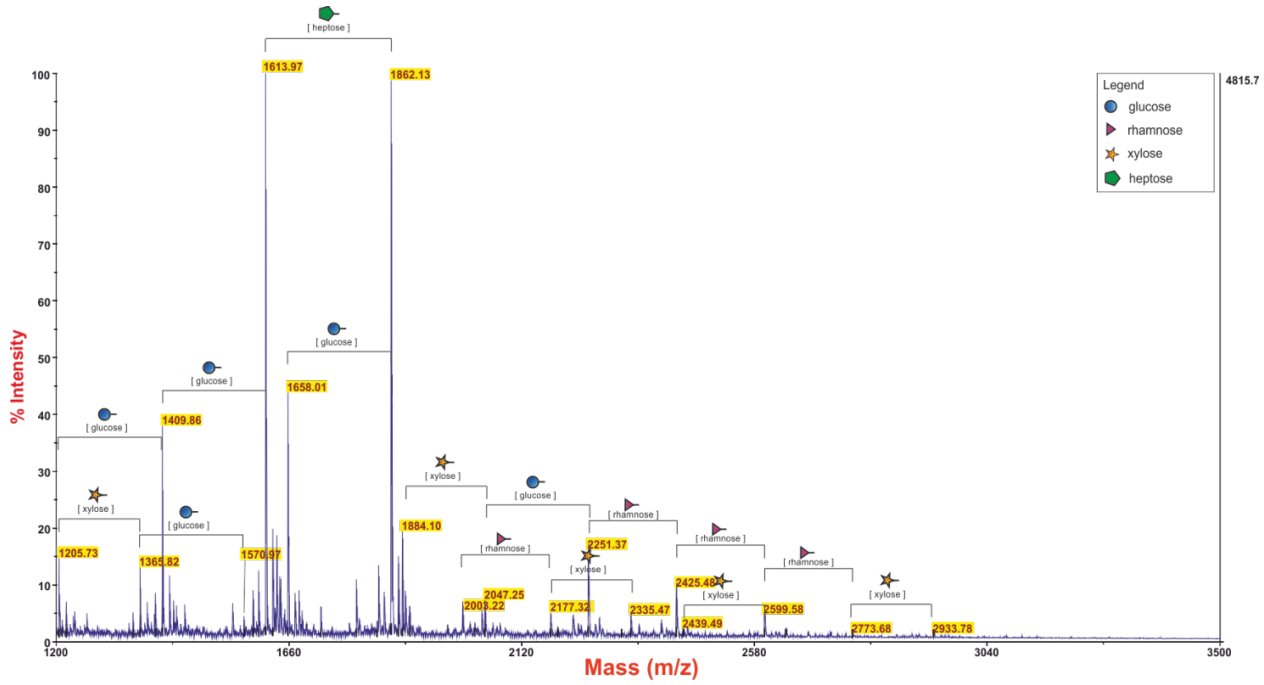


Figure 4.18 – MALDI-MS of *B. pseudomallei* LPS 576 hydrolysed by ammonolysis (35% ACN).
The MALDI-MS spectrum above shows the *B. pseudomallei* 576 LPS, as hydrolysed by ammonolysis in 10M ammonium hydroxide for 18hrs at 150°C and permethylated. The spectrum shows the 35% ACN fraction from the Sep-Pak® purification process and was analysed on a PerSeptive Biosystems Voyager-DE STR System MALDI mass spectrometer. The monosaccharide moieties identified are annotated in the spectrum.

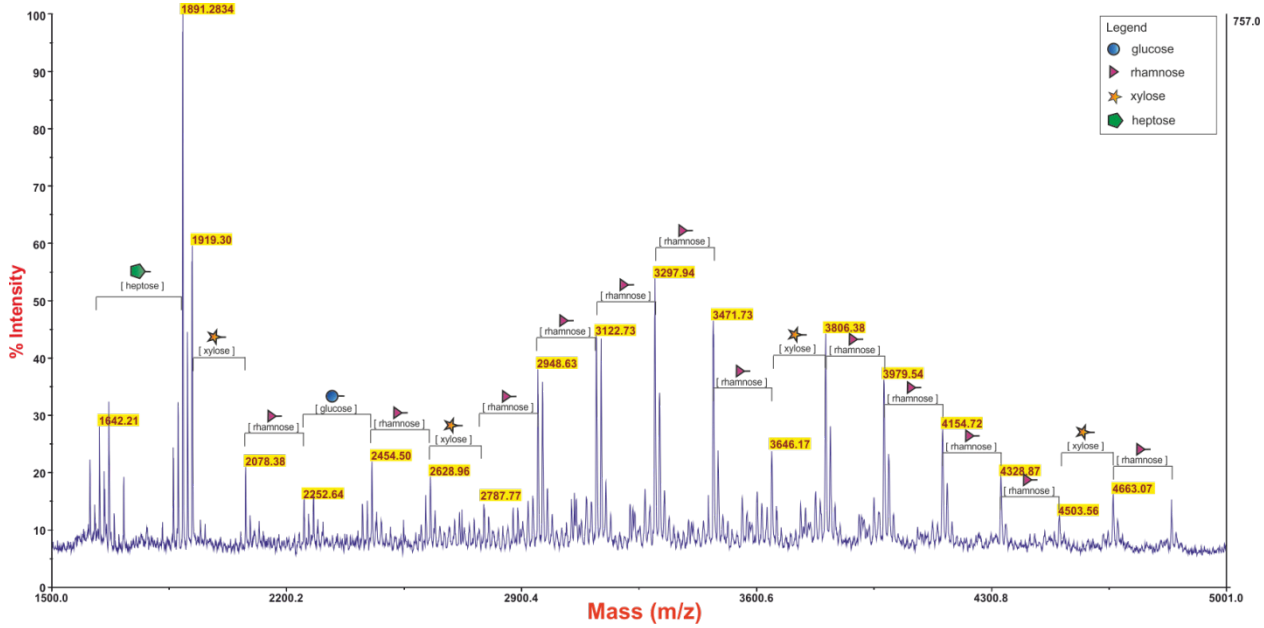


Figure 4.19 – MALDI-MS of *B. pseudomallei* LPS 576 hydrolysed by ammonolysis (100% ACN).
The MALDI-MS spectrum above shows the *B. pseudomallei* 576 LPS, as hydrolysed by ammonolysis in 10M ammonium hydroxide for 18hrs at 150°C and permethylated. The spectrum shows the 100% ACN fraction from the Sep-Pak® purification process and was analysed on a PerSeptive Biosystems Voyager-DE STR System MALDI mass spectrometer. The monosaccharide moieties identified are annotated in the spectrum.

The figures show a repeated sequence of monosaccharides of the heteropolymer. As suggested by compositional analysis these recurring units are made up of xylose, rhamnose and glucose monosaccharides. Both figures additionally show the presence of mannoheptose moieties, which are assumed to derive from the core (**Section 1.1.2.2.1**, **Section 4.3.4.1** and **Figure 4.37**). The ammonolysis procedure used on *B. pseudomallei* 576 as presented in these figures, produced fragmentation patterns by MALDI-MS, however repetition of the experiments on the samples investigated produced no further information on fragmentation of the polysaccharides (results not shown).

4.3.4 SDS hydrolysis

Further investigation into a suitable hydrolysis procedure for the *B. pseudomallei* 576 LPS led us to follow up on the work of Novem and colleagues (Novem et al., 2009), who recently published their findings on the structure of the lipid A. The technique used to separate the lipid A from the rest of the polysaccharide using sodium dodecyl sulphate (SDS) in sodium acetate was employed and combined with mild acid hydrolysis, allowing for MALDI-MS and MS/MS data collection on the core and O-antigen of *B. pseudomallei* 576 (**Section 2.4.7.4**). The samples were further derivatised by deuteropermethylation and reduced or deuteroreduced prior to permethylation. MALDI-MS and MALDI-MS/MS data analysis of these samples, provided further information on the O-antigen repeat unit. The use of the adjusted SDS hydrolysis procedure provided an additional benefit such that different Sep-Pak® purification fractions, contained separate and clearly unrelated series of molecular ions. The higher concentrations of ACN purification fractions seemed to contain the core domain of the LPS, which was analysed similarly to the O-antigen preparations and described in **Section 4.3.4.1**.

Hydrolysis of *B. pseudomallei* 576 LPS with this adjusted method of hydrolysis produced MALDI-MS data as annotated in the spectrum in **Figure 4.20**. MALDI-MS data showing a “ladder” of molecular ions (flagged in yellow) separated by mass increments consistent with predicted monosaccharides, as visible in the annotated spectrum. The spectrum shows mass differences of peaks consistent with xylose, rhamnose and hexose monosaccharides, as supported by GC-MS composition data. Each of these molecular ions was selected for MS/MS analysis.

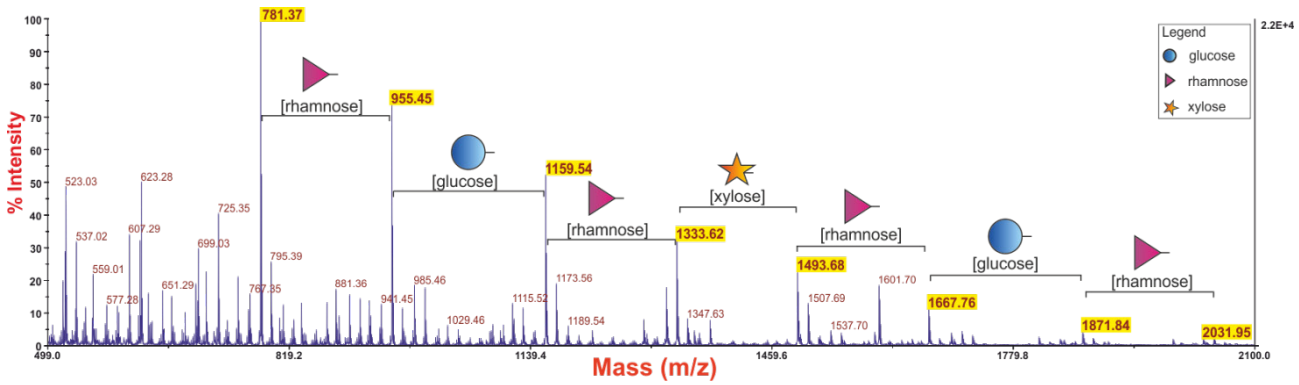


Figure 4.20 – MALDI-MS spectrum of *B. pseudomallei* 576 LPS hydrolysed by the adapted SDS method. The annotated signal pattern shows the presence of xylose, rhamnose and glucose monosaccharides in the O-antigen repeat. The sample was analysed on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer, after hydrolysis with SDS, acidified ethanol and acetic acid and derivatisation by permethylation. This spectrum was accumulated from the 50% ACN Sep-Pak® purification fraction.

In the MALDI-MS spectrum in **Figure 4.20**, the highlighted molecular ions were further isolated for MS/MS analysis. The MALDI-MS/MS spectra of the molecular ions at m/z 781.37, 955.45, 1159.54, 1333.62 and 1493.68 are shown in **Figure 4.21**, **Figure 4.22**, **Figure 4.23**, **Figure 4.24** and **Figure 4.25** respectively.

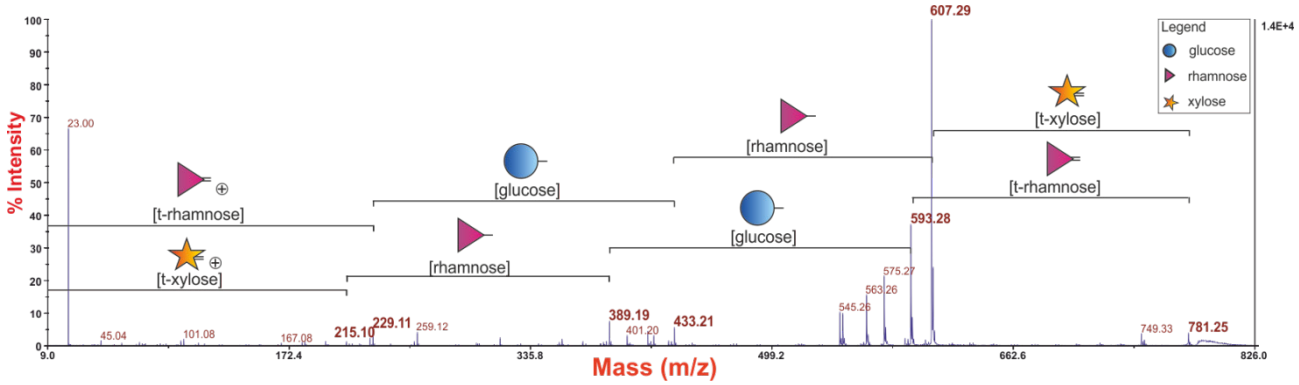


Figure 4.21 – MALDI-MS/MS spectrum of the molecular ion at m/z 781.37.
The molecular ion was obtained from *B. pseudomallei* 576 LPS preparation analysed on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer, after hydrolysis with SDS, acidified ethanol and acetic acid and derivatisation by permethylation. This spectrum was accumulated from the molecular ion at m/z 781.37 found in the 35% ACN Sep-Pak® purification fraction.

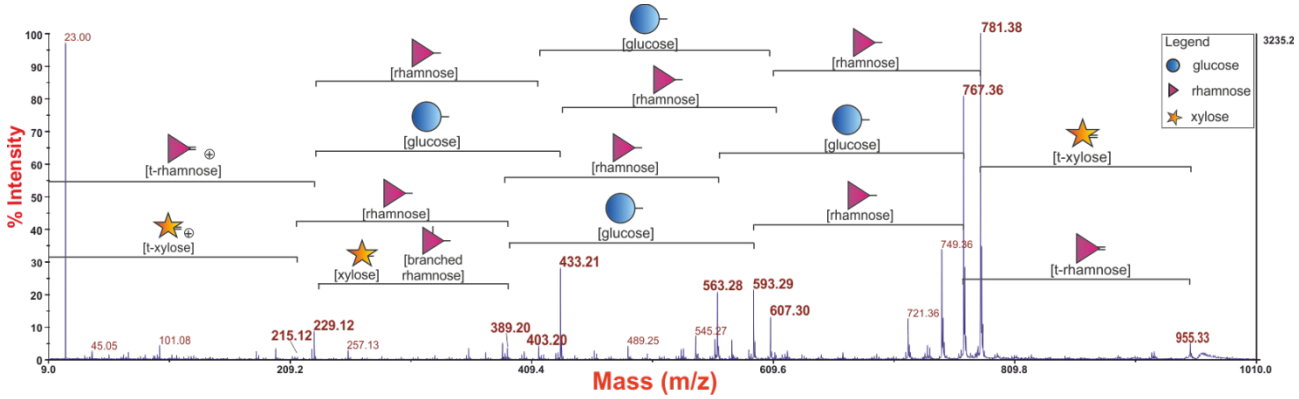


Figure 4.22 – MALDI-MS/MS spectrum of the molecular ion at m/z 955.45.
The molecular ion was obtained from *B. pseudomallei* 576 preparation analysed on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer, after hydrolysis with SDS, acidified ethanol and acetic acid and derivatisation by permethylation. This spectrum was accumulated from the molecular ion at m/z 955.45 found in the 35% ACN Sep-Pak® purification fraction.

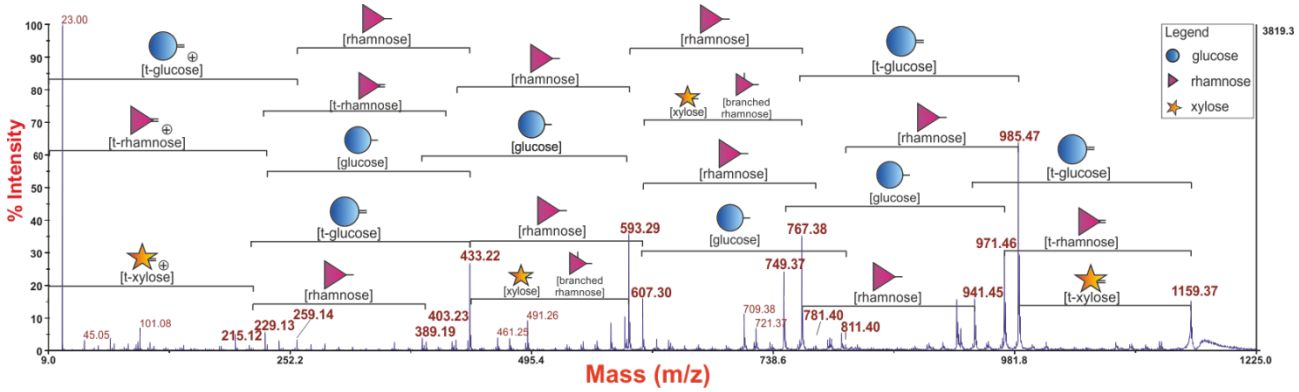


Figure 4.23 – MALDI-MS/MS spectrum of the molecular ion at m/z 1159.54.
The molecular ion was obtained from *B. pseudomallei* 576 LPS preparation analysed on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer, after hydrolysis with SDS, acidified ethanol and acetic acid and derivatisation by permethylation. This spectrum was accumulated from the molecular ion at m/z 1159.54 found in the 50% ACN Sep-Pak® purification fraction.

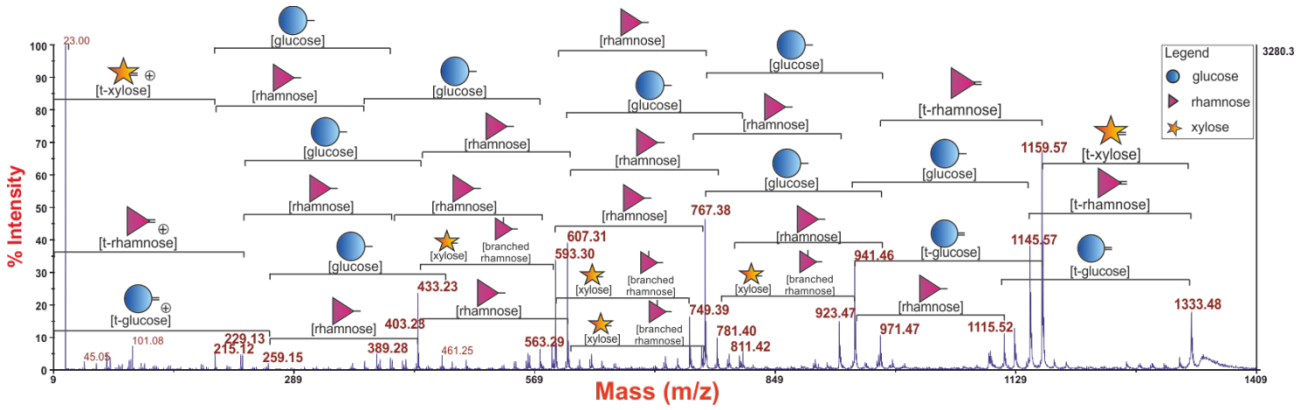


Figure 4.24 – MALDI-MS/MS spectrum of the molecular ion at m/z 1333.62.
The molecular ion was obtained from *B. pseudomallei* 576 LPS preparation analysed on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer, after hydrolysis with SDS, acidified ethanol and acetic acid and derivatisation by permethylation. This spectrum was accumulated from the molecular ion at m/z 1333.62 found in the 50% ACN Sep-Pak® purification fraction.

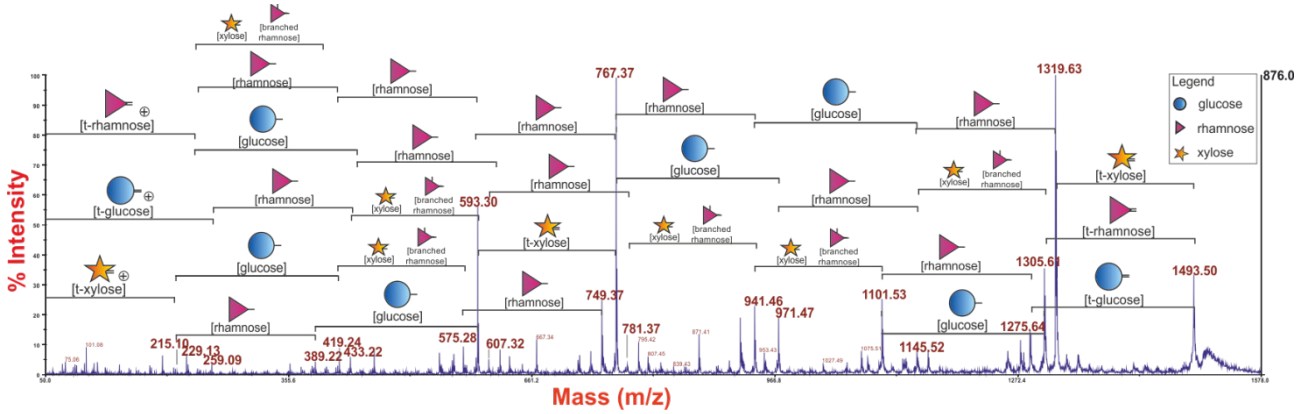


Figure 4.25 – MALDI-MS/MS spectrum of the molecular ion at m/z 1493.68.
The molecular ion was obtained from *B. pseudomallei* 576 LPS preparation analysed on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer, after hydrolysis with SDS, acidified ethanol and acetic acid and derivatisation by permethylation. This spectrum was accumulated from the molecular ion at m/z 1493.68 found in the 50% ACN Sep-Pak® purification fraction.

The observed fragmentation patterns in these figures, as well as other MS/MS spectra of further molecular ions (results not shown) and understanding of linkage and composition gained from GC-MS analysis, allowed the postulation of various structures of the O-antigen repeat unit.

The molecular ions presented in these figures, specifically the molecular ion at m/z 781.37 and 1493.68 would suggest a composition of a pentose, a hexose and two deoxyhexoses per O-antigen repeat unit, as is presented in **Figure 4.26** and **Figure 4.27** respectively. It should be noted, however, that this sequence is not unambiguous due to the multiple options of fragmentation pattern assignment. However, the proposed structures are not the sole fragmentation patterns observed in the MALDI-MS/MS spectra and thus a series of possible structures and fragmentation patterns is shown in **Figure 4.28**.

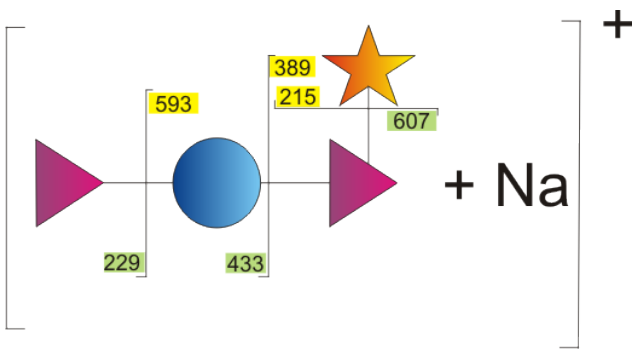


Figure 4.26 – Cartoon representation of a proposed structure of the molecular ion at m/z 781.37. This representation of the proposed smallest repeat unit, as hydrolysed and permethylated, suggests a composition of a xylose, a glucose and two rhamnose monosaccharides per O-antigen repeat unit.

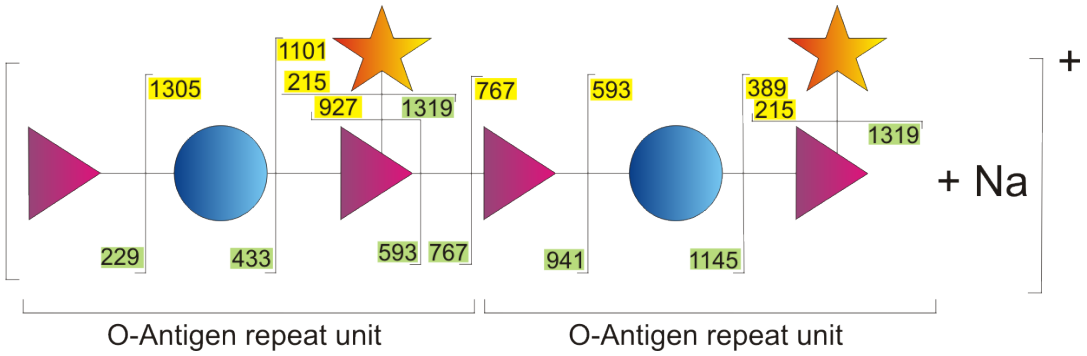


Figure 4.27 – Cartoon representation of a proposed structure of the molecular ion at m/z 1493.68. This representation of proposed two repeat units, as hydrolysed and permethylated, is suggested to be composed of two xyloses, two glucoses and four rhamnoses.

To help resolve ambiguities, the sample was further prepared with different derivatisation procedures post hydrolysis. Once the samples were hydrolysed they were either permethylated, deuteropermethylated, reduced or deuteroreduced for comparison (**Section 2.4.8.2**, **Section 2.4.8.3** and **Section 2.4.8.4**). The reduction step was important for differentiating reducing and non-reducing fragment ions. D-labels were important for confirming tentative assignements. This, in combination with repetitions to improve and add data, allowed for further evidence to support the proposed structure of the O-antigen repeat unit.

All major molecular ions observed in the MS spectra were subjected to MS/MS and data for the molecular ion at m/z 1493.68 were chosen for detailed assessment because its composition is likely to correspond to two repeat units of two pentoses, two hexoses and four deoxyhexoses, which based on results from GC-MS analysis can be inferred as two xyloses, two glucoses and four rhamnoses (**Figure 4.29** and **Figure 4.30**). The fragment ions from the permethylated samples were very useful for partial assignment of the structure of the O-antigen repeat unit. Unfortunately, however, the ions could not be unequivocally assigned to a unique sequence of the monosaccharides, due to the multiple options of fragmentation pattern assignment. As found in **Figure 4.25**, these include, for example, the rhamnose fragment between m/z 1145.52 and m/z 971.47, the glucose fragment between m/z 971.47 to m/z 767.37 or the rhamnose fragment between m/z 433.22 and m/z 259.09. These fragments and many more add uncertainty to the assignment of the sequence. Furthermore, any fragment assignment of m/z 174 could suggest the presence of a rhamnose or a terminal xylose.

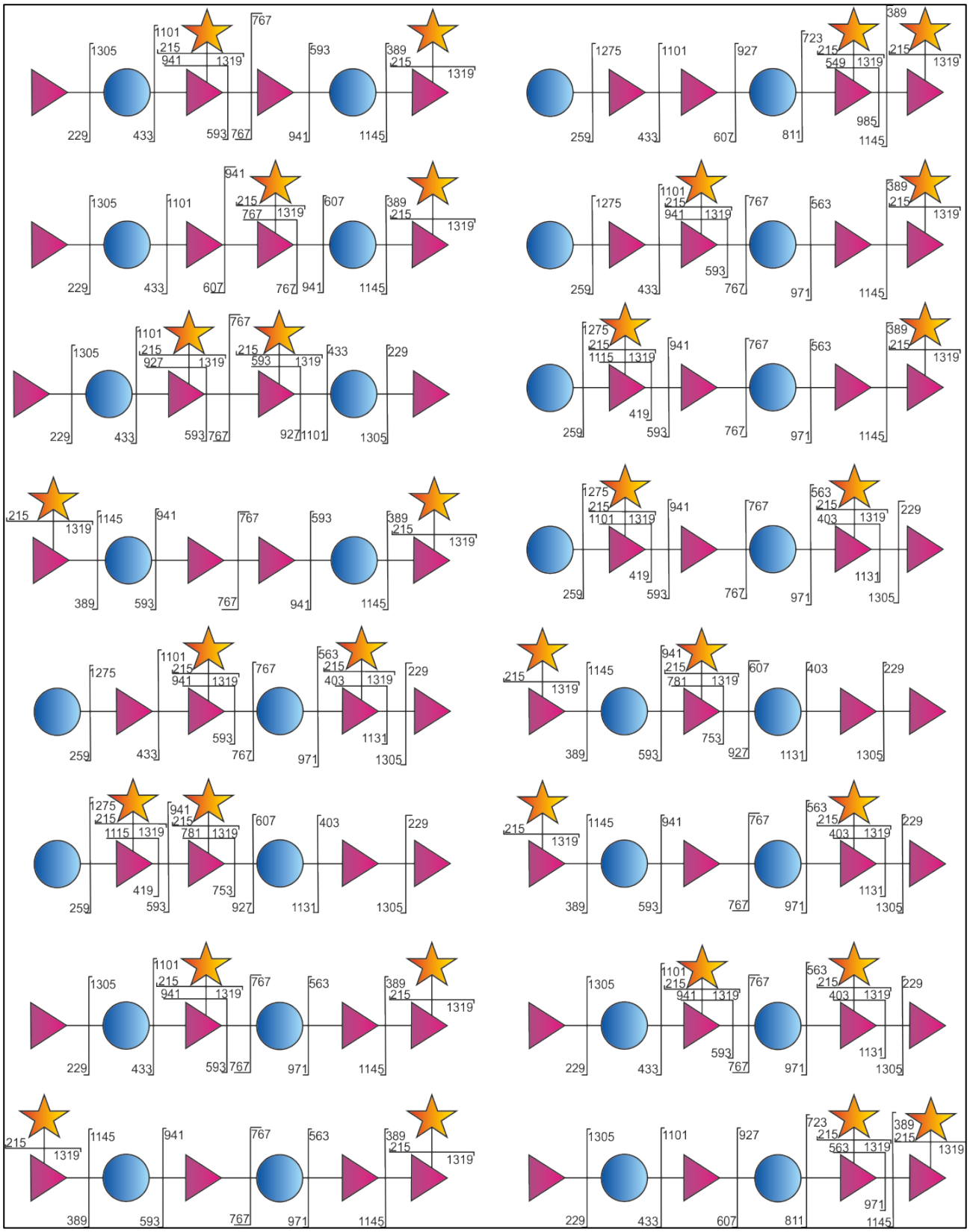


Figure 4.28 – Series of cartoon structures possible for the molecular ion at m/z 1493.68.
The figure shows a series of possibilities accumulated from the MALDI-MS/MS analyses of every molecular ion found in the series of MALDI-MS molecular ions corresponding to the O-antigen.

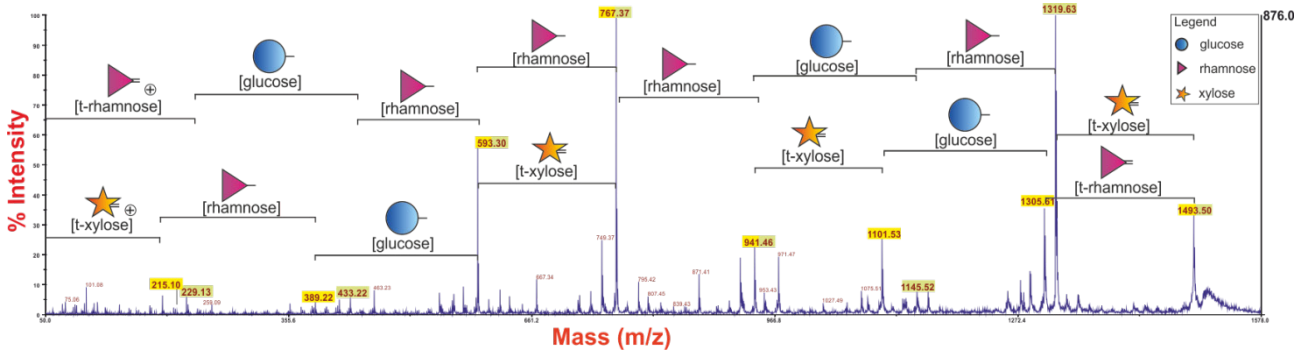


Figure 4.29 – MALDI-MS/MS spectrum of *B. pseudomallei* 576 LPS molecular ion at m/z 1493.50.
This sample was hydrolysed in SDS, acidified ethanol and acetic acid, permethylated and the molecular ion was present in the purified by Sep-Pak® fraction of 50% ACN. The spectrum was accumulated on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer. The fragmentation annotated suggests a composition of two xyloses, two glucoses and four rhamnoses. This is likely to correspond to two O-antigen repeat units.

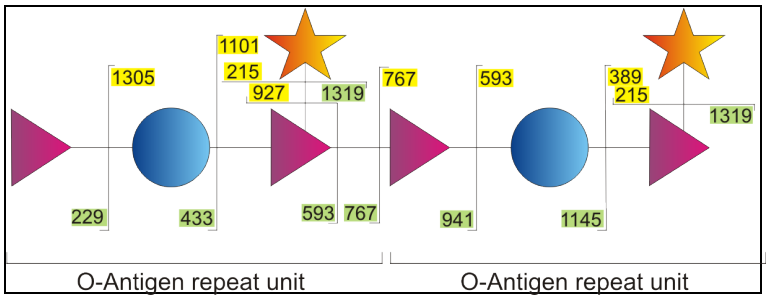


Figure 4.30 – Proposed structure of two O-antigen repeat units, from MS/MS of m/z 1493.50.
The schematic also takes into account other MS/MS spectra and compositional and linkage data from GC-MS. This sample was prepared by derivatisation by permethylation prior to MS and MS/MS analysis.

The MALDI-MS data for the different derivatisation procedures, produced the same molecular ion distribution for fragments as is shown in **Figure 4.20**, with the exception that masses shifted according to the derivatisation procedure performed. In each MALDI-MS spectrum all molecular ions related to the O-antigen repeat series were selected for MS/MS analysis. Presented here are the MALDI-MS/MS spectra of the molecular ions corresponding to that of m/z 1493.68, in **Figure 4.20**, which was suggested to represent two O-antigen repeat units. These molecular ions were m/z 1554.01, m/z 1509.67 and m/z 1510.63 for the deuteropermethylated, reduced and deuteroreduced samples respectively.

The deuteropermethylation distinguished peaks related to the molecular ion in question from those arising from non-methylated contaminants, as these were the only ones showing a shift in mass, as shown in **Figure 4.31**, with the respective proposed structural assignment in **Figure 4.34**. Furthermore, reduction or deuteroreduction of the oligosaccharides aided in identifying the direction of cleavage thus allowing identification of the reducing end of the O-antigen repeat unit. The reduced and deuteroreduced sample fragmentations are shown in **Figure 4.33** and **Figure 4.35** respectively, which allowed for the identification of specific reducing end ions, including reducing-end rhamnose, reducing-end xylose-rhamnose and reducing-end glucose. The most significant peak for the reducing-end ions are m/z 405.19 and m/z 406.21 for the reduced and deuteroreduced sample, respectively. This corresponds to a reducing end ion of reducing-end xylose-rhamnose. The inferred structural proposals for these reduced and deuteroreduced MALDI-MS/MS spectra are shown in **Figure 4.34** and **Figure 4.36**, respectively.

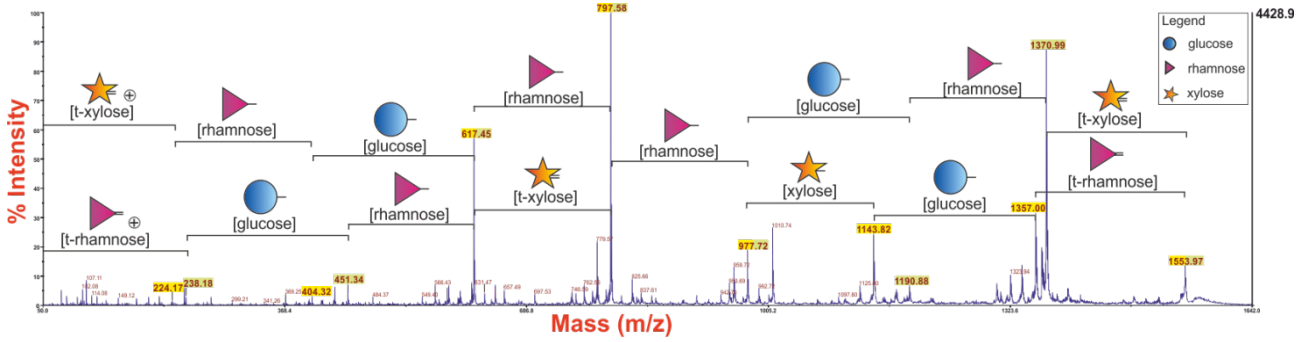


Figure 4.31 – MALDI-MS/MS spectrum of *B. pseudomallei* 576 LPS molecular ion at m/z 1554.01.
This sample was hydrolysed in SDS, acidified ethanol and acetic acid, deuteropermethylated and the molecular ion was present in the purified by Sep-Pak® fraction of 50% ACN. The spectrum was accumulated on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer. The fragmentation annotated suggests a composition of two xyloses, two glucoses and four rhamnoses. The m/z differences corresponding to the monosaccharides are larger, due to the use of deuterium instead of hydrogen in the permethylation procedure (see Table 1.4). This spectrum confirms the proposed structure from the permethylated sample shown in Figure 4.32.

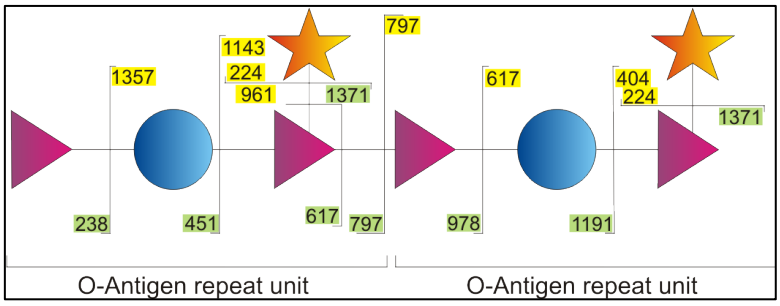


Figure 4.32 – Proposed structure of two O-antigen repeat units from MS/MS of m/z 1554.01.
The schematic also takes into account other MS/MS spectra and compositional and linkage data from GC-MS. This sample was prepared by derivatisation by deuteropermethylation prior to MS and MS/MS analysis.

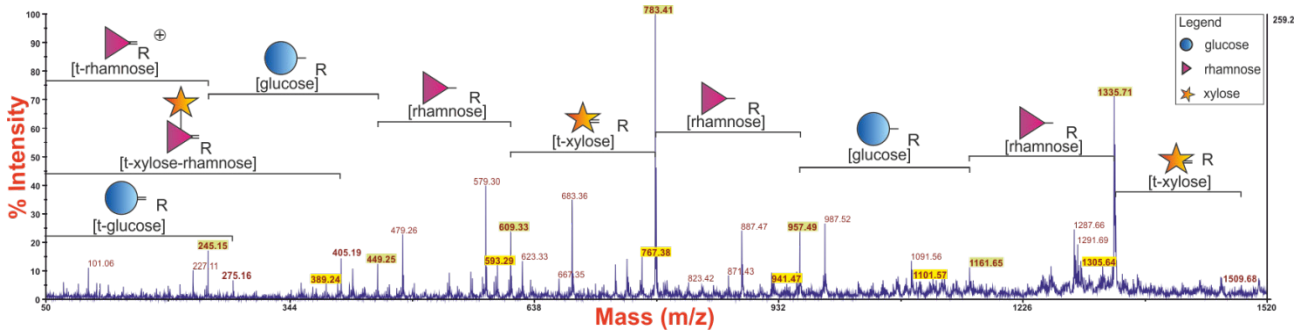


Figure 4.33 – MALDI-MS/MS spectrum of *B. pseudomallei* 576 LPS molecular ion at m/z 1509.67.
This sample was hydrolysed in SDS, acidified ethanol and acetic acid and reduced prior to permethylation and the molecular ion was present in the purified by Sep-Pak® fraction of 50% ACN. The spectrum was accumulated on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer. The fragmentation annotated suggests a composition of two xyloses, two glucoses and four rhamnoses. The reducing end of the fragments found in the MALDI-MS/MS spectrum will have a slightly larger mass (+ m/z 16) in comparison with the non-reducing end monosaccharide signal differences. This spectrum confirms the proposed structure from the permethylated sample shown in Figure 4.32.

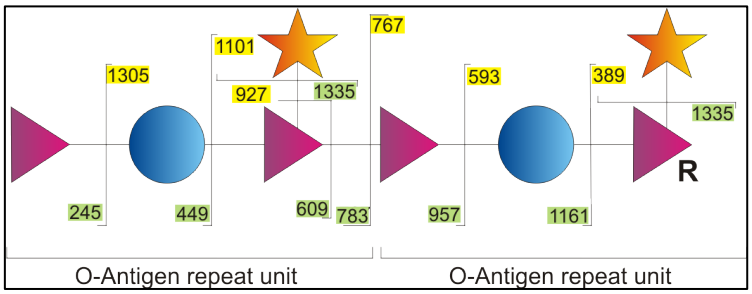


Figure 4.34 – Proposed structure of two O-antigen repeat units, from MS/MS spectrum of m/z 1509.67.
The schematic also takes into account other MS/MS spectra and compositional and linkage data from GC-MS. This sample was reduced before derivatisation by permethylation prior to MS and MS/MS analysis.

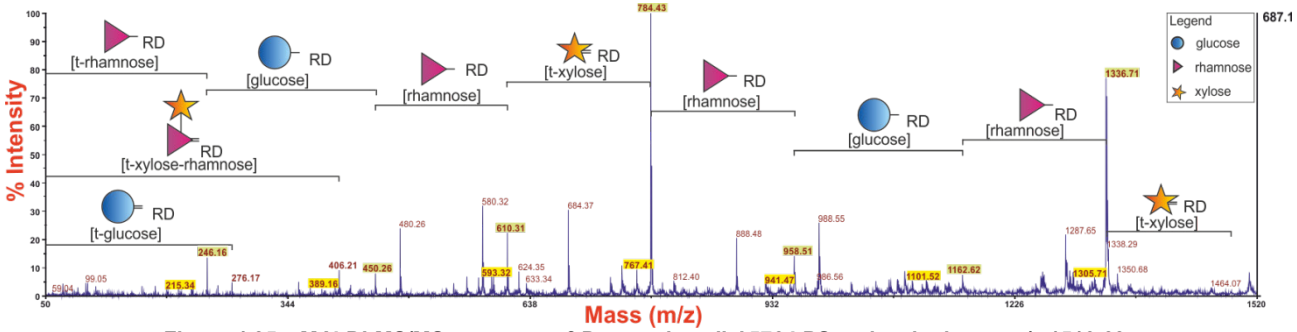


Figure 4.35 – MALDI-MS/MS spectrum of *B. pseudomallei* 576 LPS molecular ion at m/z 1510.63.
This sample was hydrolysed in SDS, acidified ethanol and acetic acid and deuteroreduced prior to permethylation and the molecular ion was present in the purified by Sep-Pak® fraction of 50% ACN. The spectrum was accumulated on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer. The fragmentation annotated suggests a composition of two xyloses, two glucoses and four rhamnoses. The reducing end of the fragments found in the MALDI-MS/MS spectrum will have a slightly larger mass (+ m/z 17) in comparison with the non-reducing end monosaccharide signal differences. This spectrum confirms the proposed structure from the permethylated sample shown in Figure 4.32.

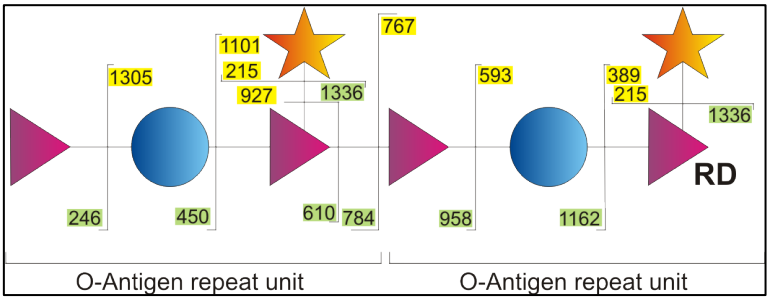


Figure 4.36 – Proposed structure of two O-antigen repeat units, from MS/MS of m/z 1510.63.
The schematic also takes into account other MS/MS spectra and compositional and linkage data from GC-MS. This sample was reduced before derivatisation by permethylation prior to MS and MS/MS analysis.

The proposed structure of the O-Antigen repeat unit in **Figure 4.30** is primarily based on information deduced from the fragment ions that arose from MS/MS of the molecular ions at m/z 1493.68, m/z 1554.01, m/z 1509.67 and m/z 1510.63 for the four derivatisation procedures employed. This was combined with information from other MS/MS spectra from different molecular ions as well as, compositional and linkage data from GC-MS. These experiments provided confidence that the proposed structures of the hydrolysis products are representative of most of the elements of the repeat unit of *B. pseudomallei* 576 LPS. However it is important to bear in mind that labile portions, such as acetyl or methyl group modifications of the repeat unit may have been lost or masked during derivatisation or hydrolysis.

4.3.4.1 SDS hydrolysis of the LPS core domain

Upon investigation of the LPS O-antigen through the use of the adjusted SDS hydrolysis procedure described in **Section 2.4.7.4**, some of the Sep-Pak® purification fractions recovered, clearly contained molecular ion sequences unrelated to that of the O-antigen. These fractions, mainly higher concentrations of ACN, were likely to comprise the core domain of the LPS of *B. pseudomallei* 576. The MALDI-MS spectra of these fractions were therefore assessed and individual molecular ions subjected to MALDI-MS/MS analysis as was done for the LPS O-antigen containing fractions. Furthermore, the samples were also deuteropermethylated and reduced as well as deuteroreduced prior to permethylation for further data accumulation on the core domain of the LPS.

The MALDI-MS spectrum presented in **Figure 4.37** shows the assumed core domain cluster of molecular ions found in fractions of the higher concentrations of ACN post Sep-Pak® purification. The spectrum identified in the 75% ACN fraction, was reproduced when deuteropermethylating, as well as reducing and deuteroreducing the sample prior to further derivatisation by permethylation (data not shown). The monosaccharide moieties presented in the spectrum include 3-deoxy-D-manno-oct-2-ulsonic acid (Kdo) and D-glycero-D-talo-octu-2-ulsonic acid (Ko) as well as mannoheptose, hexose and deoxyhexose. The Kdo, mannoheptose and hexose moieties are generally expected in the core domains of bacterial LPS, however, the finding of Ko led to further understanding of reasons for the inadequate hydrolysis of the LPS with standard procedures and the need for further development of such techniques.

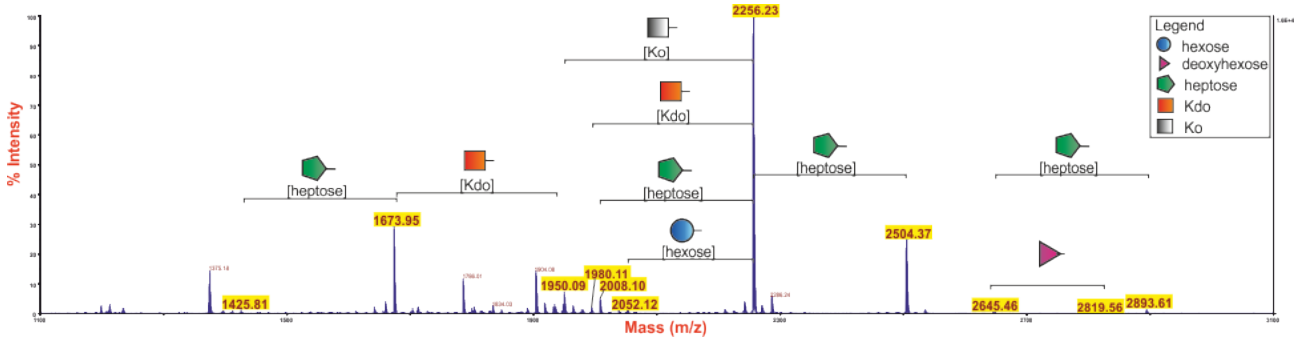


Figure 4.37 – MALDI-MS spectrum of *B. pseudomallei* 576 core domain cluster.
The sample was hydrolysed in SDS, acidified ethanol and acetic acid and permethylated. The molecular ions corresponding to the core domain cluster were present in the 75% ACN fraction of the Sep-Pak® purified sample. The spectrum was accumulated on an Applied Biosystems 4800 MALDI-TOF/TOF mass spectrometer. The monosaccharide constitution of the core domain cluster includes Ko, Kdo, Hep, Hex and dHex moieties.

Albeit accumulating a series of data sets, including MALDI-MS and MALDI-MS/MS data of differentially derivatised samples of the core domain of the *B. pseudomallei* LPS, the focus of this project remained with the structural characterisation of the O-antigen of the *B. pseudomallei* 576 LPS, rather than its core domain. Therefore no further work was undertaken at this stage to identify the structure of the core domain.

4.3.5 Summary conclusions and discussion of LPS O-antigen repeat unit elucidation

4.3.5.1 Mild acid hydrolysis for lipid removal

The lipid A in LPS is connected to the core polysaccharide that normally is easily separated by exploiting the labile linkage between the Kdo of the core and the lipid A, which can be hydrolysed to separate the two. The core structure of the LPS of *B. pseudomallei*, however, has not been determined yet. It was clear from the mild acid hydrolysis techniques employed with *B. pseudomallei* and *B. thailandensis* samples, that this technique, whilst suitable for *B. thailandensis* and *B. pseudomallei* K96243 (serotype A) samples, may not be useful for the *B. pseudomallei* 576 LPS (serotype B).

Preliminary results on *B. cepacia* showed that the LPS possess the following unusual chemical feature; the sugar that interlinks the core and lipid A, which usually is 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) in enterobacterial LPS, was not the only moiety involved in the linkage to the lipid A in *B. cepacia* LPS. In *B. cepacia* 2-keto-octulsonic acid (Ko) has been shown to most commonly replace the second Kdo in linking the core directly to lipid A forming a more acid stable ketosidic linkage, where the additional hydroxyl group is

thought to play a role in the stability of this linkage (Isshiki et al., 1998, Kawahara et al., 1987, Masoud et al., 2009). With the SDS hydrolysis procedure adapted for *B. pseudomallei* 576 degradation, information could be gathered on the core domain of the LPS (**Section 4.3.4.1**), showing that this also seems true for *B. pseudomallei* lipid A-core linkage in *B. pseudomallei* 576 LPS. These studies on the separated and hydrolysed core polysaccharide show the presence of a mass increment consistent with the presence of Ko (m/z 306 for permethylated monosaccharides). Furthermore, the linkage between the core and lipid A was resistant against mild acid hydrolysis in *B. cepacia* (Isshiki et al., 1998). In the case of *B. pseudomallei*, it was found that the linkage between the core and the lipid A is susceptible to mild acid hydrolysis only for *B. pseudomallei* K96243 and *B. thailandensis* LPS (serotype A). In *B. pseudomallei* 576 (serotype B), this linkage seems to be resistant against mild acid hydrolysis and could be due to similar reasons as in *B. cepacia*. Therefore, with the possible presence of such a distinctive core of *B. pseudomallei* 576, the linkage between the lipid A and the core polysaccharide may not be as easily broken, depending on the monosaccharide the lipid A is connected to.

Furthermore, it has been shown that the major lipid A species in *B. pseudomallei* consists of a biphosphorylated disaccharide backbone, which is modified with 4-amino-4-deoxy-arabinose (Ara4N) at both phosphates and penta-acylated with fatty acids C14:3OH, C16:3OH and either C14:O or C14:2OH (Novem et al., 2009). It was abundantly clear that mild acid hydrolysis was not producing easily separable lipid A and polysaccharide moieties for further studies, and completely different approaches were then pursued to remove the lipid A from the *B. pseudomallei* 576 LPS, for further structural elucidation. Investigation into a different release by acid hydrolysis through sodium dodecyl sulphate in sodium acetate and more aggressive methanolysis as employed by Novem and colleagues proved suitable for collection of O-antigen repeat unit structural analysis (Novem et al., 2009).

4.3.5.2 O-antigen repeat unit structure and core domains

It was established by MALDI-MS analysis that the LPS, even released from the lipid A, was too large to be studied as a whole and therefore a series of hydrolysis methods were carried out to determine the best parameters to achieve regular breaking of the polymers to an optimal size for structural determination by MALDI-MS. Hydrolysis techniques including various concentrations of methanolic acetylchloride, ammonolysis as well as mild acid hydrolysis were employed to find the best hydrolysis parameters to observe the polysaccharide units making up the O-antigen. The polysaccharides of different strains seemed to behave differently to various hydrolysis techniques, which could be explained by their variations in structure, as some bonds might be more unstable than others in distinct arrangements. The preliminary set of experiments of the listed hydrolysis procedures employed, generated inconclusive results even though the ammonolysis method allowed for some insight by MALDI-MS, but was not reproducible.

Further work into the structural characterisation of the O-antigen repeat unit of *B. pseudomallei* LPS using sodium dodecyl sulphate (SDS) hydrolysis in sodium acetate in combination with different derivatisation techniques, provided enough information to propose a theoretical structure and sequence of the O-antigen repeat unit by MALDI-MS and MALDI-MS/MS. The proposed structure in **Figure 4.30** is based on information deduced from the fragment ions that arose from tandem mass spectrometry of various molecular

ions as derivatised by different techniques as well as compositional and linkage data from GC-MS (**Section 4.2**). These experiments provided confidence that the proposed structures of the hydrolysis products are representative of most of the elements of the repeat unit of *B. pseudomallei* 576 LPS. However it is important to bear in mind that labile portions of the repeat unit may have been lost during hydrolysis or derivatisation. It should also be noted that earlier linkage analysis experiments showed many peaks with different linkage profiles, which could not be fully explained by the tentative sequence, which is further discussed in **Section 4.4.4**.

Core domain elucidation of bacterial LPS moieties has been undertaken by various laboratories, however, this does not seem a trivial task, as is seen in the structural characterisation of the core domain of the *H. pylori* LPS as described by Altman and colleagues (Altman et al., 2011a, Altman et al., 2011b). The LPS of *H. pylori* has been investigated by various groups for a long time and only the last few years have shown the possibility to define the outer core and establish major structural differences from previous conclusions made from earlier mass spectrometric data. Furthermore, such characterisation required approximately 60-80mg of LPS sample roughly 40x the total sample amounts used for the various experiments described in this thesis. This highlights the difficulty of bacterial glycopolymer characterisation, however, data presented in this thesis and advances made in this field prove progress in structural characterisation of prokaryotic glycopolymers.

4.4 *B. PSEUDOMALLEI* LPS ANALYSIS BY NMR

With all the mass spectrometric data culminating in a proposed structure, further work was undertaken to confirm the validity of the conclusions that had been arrived at. NMR methods were devised to provide information on the linkage and anomeric stereochemistry of the hydrolysates and narrow down the possible O-antigen structure. Preliminary NMR analysis was performed on a Bruker Avance III 600 MHz spectrometer equipped with a TCI cryoprobe on a sample that had been solubilised by the same SDS hydrolysis technique that provided the MALDI-MS and MALDI-MS/MS data. With the understanding that hydrolysed sample preparations provided excessive amounts of information due to the heterogeneity of the hydrolysed O-antigen fragments, further investigations were made into a different method for structural analysis by NMR. A method embedding native LPS moieties into mixed-DPC-LPS micelles as adapted from Acquotti and Sonnino (Acquotti and Sonnino, 2000) was piloted with a standard LPS and further employed for analysis of the intact LPS of *B. pseudomallei* 576 LPS.

4.4.1 NMR of the hydrolysed *B. pseudomallei* O-antigen

The sample, previously hydrolysed by SDS hydrolysis in sodium acetate was prepared in bulk for analysis by NMR spectroscopy. Multi-dimensional NMR techniques were used to gain further insight into the linkage and composition of the O-antigen repeat unit.

The LPS was hydrolysed using the SDS combined method (**Section 2.4.7.4**) and the separation of the O-antigen and core fractions of the polysaccharide was achieved by ion-exchange chromatography (**Section 2.4.9.1**). The separation was based on the carboxylic acid group of the Kdo and Ko components of the core,

where Kdo has a pK_a of 2-3. The O-antigen was found in the H_2O fraction, whilst the core was found in the 5% acetic acid and the 0.1M HCl fraction. The fractions were dried down and lyophilised.

In proton NMR spectroscopy, most carbohydrate signals are crowded between 3ppm and 4ppm. The signals from the anomeric protons (H-1 bound to C-1 of any monosaccharide residues) are found downfield, allowing for interpretation of the anomeric stereochemistry of the monosaccharide residues. In the hydrolysed *B. pseudomallei* sample, it is shown that there are resonances relating to monosaccharides in both α- and β-form (**Figure 4.38**). This means that there are multiple α- and β-anomeric residues, as is expected for a complex sample. From the 1D proton NMR, we can further tentatively confirm the deoxyhexose moiety discovered by GC-MS, MALDI-MS and MALDI-MS/MS, as seen from the signals between 1.1 – 1.3ppm ^{1}H chemical shift, which is consistent with the methyl residues found on deoxyhexoses, such as rhamnose. The intense signal found at 3.41ppm, could derive from a methoxy substituent of a monosaccharide moiety, however this could also arise from residual ethanol moieties from the hydrolysis procedure.

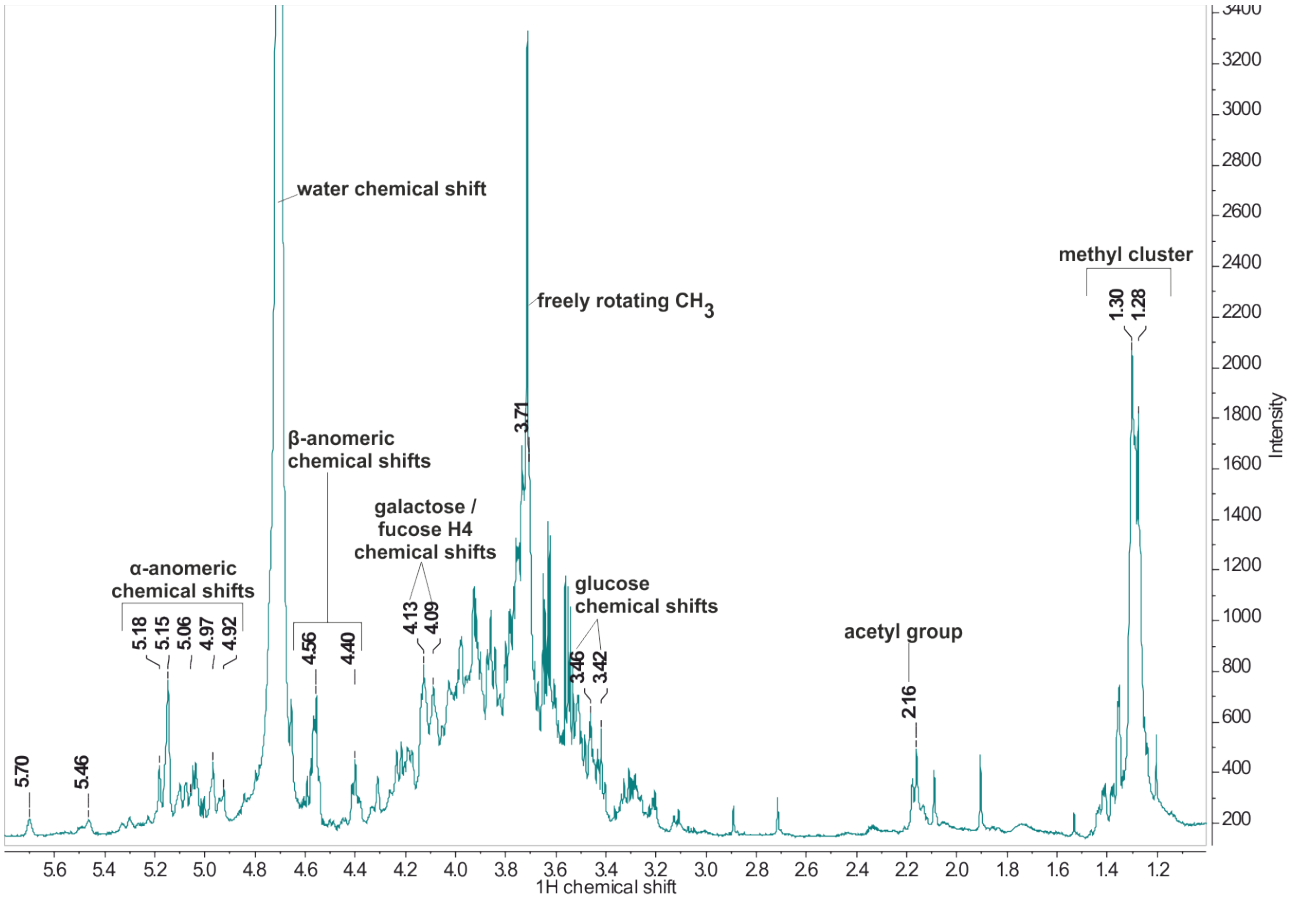


Figure 4.38 – 1D ^{1}H NMR spectrum of hydrolysed *B. pseudomallei* 576 O-antigen.
The sample was hydrolysed by SDS, acidified ethanol and acetic acid and purified by ion exchange chromatography. The spectrum was acquired on a Bruker Avance III 600 MHz spectrometer equipped with a TXI/TCI cryoprobe in 100% D_2O. Anomeric protons are found in the 4.2-6.0ppm range, whilst non-anomeric protons are found in the 3.0-4.2ppm range.

Homonuclear correlation spectra (COSY, TOCSY) are used to assign the ^{1}H spectrum. TOCSY experiments are similar to COSY experiments, where cross peaks of coupled protons are observed. In COSY spectra the cross peaks are observed only for nuclei directly coupled, whereas TOCSY spectra show signals for nuclei that are connected by a chain of couplings. This allows for signal allocation for a whole spin system, or monosaccharide in carbohydrate NMR. This is achieved by a repetitive series of pulses, which cause

isotropic mixing; the longer the isotropic mixing time, the further the polarisation can spread throughout the spin system, generally allowing cross-peaks for all protons in a monosaccharide.

The 2D homonuclear ^{1}H-^{1}H correlation NMR spectra (TOCSY) of the hydrolysed *B. pseudomallei* 576 O-antigen were accumulated at different mixing times, which are used in carbohydrate NMR to build up assignments for individual spin systems associated with monosaccharide residues within the O-antigen structure (data not shown). These however, did not produce enough data for assignment of the spin systems involved in the O-antigen structure, as only fragmentary spin systems could be identified and full spin system assignment could therefore not be deduced from the spectra. The weak and fragmentary nature of the spectra was attributed to the minimal amounts of sample available and the complexity of the mixture of O-antigen fragments in the hydrolysate. Sample amounts are a crucial determinant for NMR data accumulation, where an increase in sample amount exponentially decreases the time taken to accrue NMR spectra. This means that an increase in sample concentration has a greater effect on data quality than extending the experiment times.

A ROESY experiment was further conducted, which explores the correlations between nuclei in space and could be used to confirm assignments and provide sequence and linkage position information. The sample quantity, however, proved too small to accumulate enough data and results are not shown.

A couple of correlation signals were identified in the spectra and used for assignment of the cross-peak lattices on the heteronuclear ^{1}H-^{13}C single quantum correlation spectra (HSCQ), which is used to detect through-bond correlations between nuclei of two different types, in this case ^{1}H and ^{13}C nuclei. **Figure 4.39** shows the 2D heteronuclear total ^{1}H-^{13}C correlation NMR spectrum of the hydrolysed *B. pseudomallei* 576 O-antigen (HSQC-TOCSY), where the ^{13}C chemical shift scale was “folded” and thus minimised to allow for more data collection on a small sample amount. Spectral folding, or spectral aliasing is a frequently used phenomenon in NMR spectroscopy, if a large spectral width is required for data accumulation, which increases the experimental recording times. Folding of the spectral width, allows for signal presentation at frequencies, different from their real frequency. Tuning of acquisition parameters are crucial, when applying such folding or aliasing, to ensure that folded or aliased peaks do not overlap with other peaks present in the spectrum, that cannot easily be distinguished from these folded peaks. Spectral folding allows for an increased number of scans, accumulating more data and only slightly reduces the resolution. The coloured lattice arrangements assigned on the spectrum show spin systems, where the ^{1}H and ^{13}C correlations within a monosaccharide residue are connected. This allows comparison with information on chemical shift patterns reported in literature in order to identify monosaccharides. The spectrum shows very poorly resolved signals that seem to overlap. This again, could be due to the fact that the O-antigen of the LPS was hydrolysed, which allows for solubilisation for NMR, but produces a very heterogeneous sample.

From the 1D proton NMR spectrum it can be concluded that the proposed elements of the O-antigen repeat unit show concurring signals in the NMR spectra, confirming the presence of a deoxyhexose as well as hexoses. Further to this information, a large methyl signal was observed, which could suggest a methylated monosaccharide present in the O-antigen of *B. pseudomallei* 576, as was previously identified in the wild type serotype A of *B. pseudomallei* K96243. However, due to the hydrolysis method containing solvents with

such methyl groups, this signal could arise from residual solvent presence. Additionally, various anomeric signals were shown, suggesting multiple α- and β-anomeric residues. A tentative assignment of the deoxyhexose as coinciding with literature values for α-rhamnose could be made. Furthermore, partial information from both the homo- and heteronuclear spectra are consistent with a β-hexose moiety.

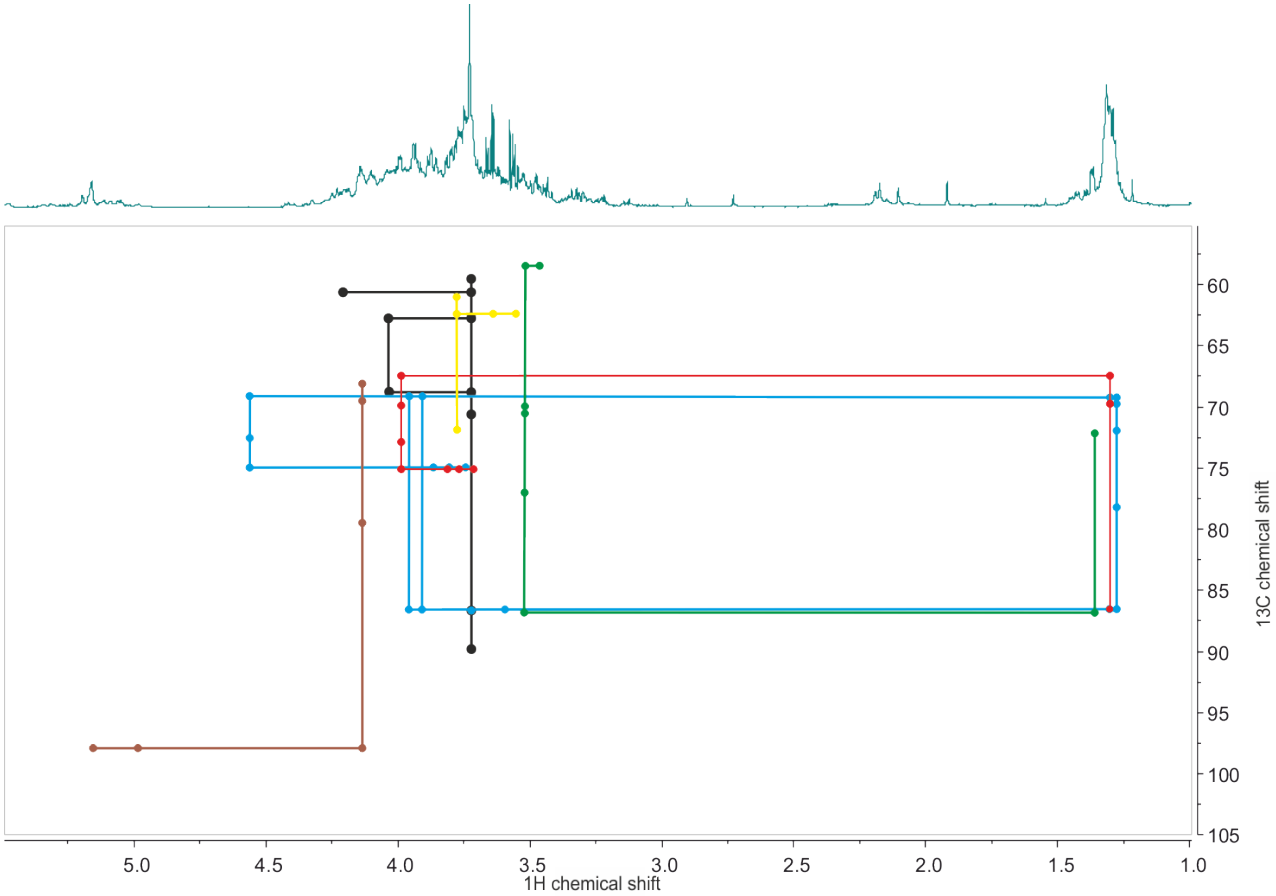


Figure 4.39 – 2D HSQC-TOCSY of hydrolysed *B. pseudomallei* 576 O-antigen.
The sample was hydrolysed by SDS, acidified ethanol and acetic acid and purified by ion exchange chromatography. The 2D heteronuclear ^{1}H-^{13}C total correlation NMR spectrum was acquired on a Bruker Avance III 600 MHz spectrometer equipped with a TXI/TCI cryoprobe in 100% D_2O. The coloured lattice assignments show partial correlations of some unique spin system.

The O-antigen sample was hydrolysed at different monosaccharide linkages, resulting in a mixture of different, yet similar compounds. These fragments will give very similar but not identical signals for each monosaccharide in the repeat unit, resulting from the slight change in environment in that specific compound. As a result, too much information is presented as a crowded spectrum containing many weak signals rather than a minimal number of strong signals, which makes complete assignment impossible.

4.4.2 Pilot study of micelle-LPS

To avoid the above mentioned problems of heterogeneity and solubility of the LPS sample, which are exacerbated by the lipid A, a different approach was proposed for obtaining suitable NMR data. In this experiment, instead of trying to solubilise the sample into organic solvents, an unhydrolysed sample of *B. pseudomallei* 576 was investigated as mixed micelles with a perdeuterated detergent, such as dodecylphosphocholine in aqueous solutions (**Section 2.4.10.1**) as is discussed by Acquotti and Sonnino who describe this method with gangliosides (Acquotti and Sonnino, 2000). This protocol would simplify the

spectrum by avoiding hydrolysis of the sample and allowing for NMR analysis on the full sample amount as extracted by our collaborators at dstl. Mock experiments using model compounds were therefore developed using standardised LPS from *E. coli* (serotypes O26 and 128) to create micelles with dodecylphosphocholine.

The previously characterised LPS O-antigen of the LPS of *E. coli* O26 was prepared in micelles and analysed by NMR to assess this as a technique to elucidate the O-antigen structure of *B. pseudomallei* LPS. With the determination of this technique as viable for the interpretation of non-hydrolysed LPS structures it is possible to deduce the O-antigen repeat unit and its modifications as a whole, where the mixed DPC-micelles of LPS and dodecylphosphocholine mimic the membrane. Furthermore this technique produces micelles of almost spherical shape that show rotational mobility, carrying LPS chains that show internal mobility, thus allowing for high-resolution NMR spectra and avoiding unnecessary sample loss through sample processing.

NMR spectroscopic techniques employed in this series of experiments were aimed at re-producing initial result, published on the characterisation of the O-antigen of *E.coli* LPS O26. The structure of the O-antigen of *E. coli* O26 LPS was published by Manca and colleagues in 1996 at Stockholm University in Sweden (Manca et al., 1996). Sugar and methylation analyses as well as 2D NMR spectroscopy were used to characterise the O-antigen.

The following structure was proposed by Manca and colleagues (Manca et al., 1996):

$$\rightarrow 3)\text{-}\alpha\text{-}_{L}\text{-Rhap-}(1\rightarrow 4)\text{-}\alpha\text{-}_{L}\text{-FucpNAc-}(1\rightarrow 3)\text{-}\beta\text{-}_{D}\text{-GlcpNAc-}(1\rightarrow$$

The chemical shifts (ppm) of the signals produced by the experiments performed by Manca and colleagues are shown in **Table 4.2**.

Sugar residue	**H/C**							
	1	**2**	**3**	**4**	**5**	**6**	**NAc**	**CO**
→4)-α-$_{L}$-FucNAc-(1→	5.01	4.17	3.96	3.87	4.43	1.19	1.97	
	98.6 [174]	50.6	68.0	82.0	68.0	16.3	23.0	175.1
→3)-α-$_{L}$-Rha-(1→	4.83 (1.9)	4.28	3.90	3.49	4.03	1.22		
	102.6 [169]	70.9	81.1	71.7	70.4	17.4		
→3)-β-$_{D}$-GlcNAc-(1→	4.68 (8.4)	3.89	3.70	3.55	3.47	3.77;3.89	2.04	
	103.5 [163]	56.5	79.6	69.4	76.4	61.5	23.0	175.0

Table 4.2 – *E. coli* O26 O-antigen chemical shifts published by Manca and colleagues.
The table above lists the chemical signals (ppm) observed by Manca and colleagues, including the signals in the ^{1}H and ^{13}C NMR spectra of the O-deacetylated *E. coli* O26 O-antigen polysaccharide (Manca et al., 1996).

The mixed-DPC-LPS micelles of *E. coli* LPS O26 were run at the Cross Faculty NMR Centre at Imperial College London, and the 1D proton NMR spectrum shown in **Figure 4.40** was obtained. The anomeric residues were annotated with lettered labels A, B and C to denote the anomeric hydrogen residues and the respective spin systems of the three monosaccharides. The obtained chemical shifts (ppm) are shown in **Table 4.3**.

Sugar residue	H						
	1	2	3	4	5	6	NAc
A →4)-α-$_L$-FucNAc-(1→	5.02	4.12	3.97	~3.86	4.44	1.20	1.97
	97.6	51.0	67.0	80.9	67.1		
B →3)-α-$_L$-Rha-(1→	4.82	4.33	~3.89	3.47	4.06	1.21	
	101.5	69.9	80.2	70.8	69.4		
C →3)-β-$_D$-GlcNAc-(1→	~4.66	~3.90	3.69	3.55	3.47	3.77; 3.87	2.03
		55.6	78.2	68.4	75.5	60.5	

Table 4.3 – Chemical shift (ppm) signals of the mixed-DPC-LPS micelles of *E. coli* LPS O26.
The values noted here were identified from the 1D proton NMR spectrum shown in Figure 4.40 and Figure 4.41. The values are to two decimal places but can vary by a decimal place within the spectra due to the difficulty of calibration at low resolution.

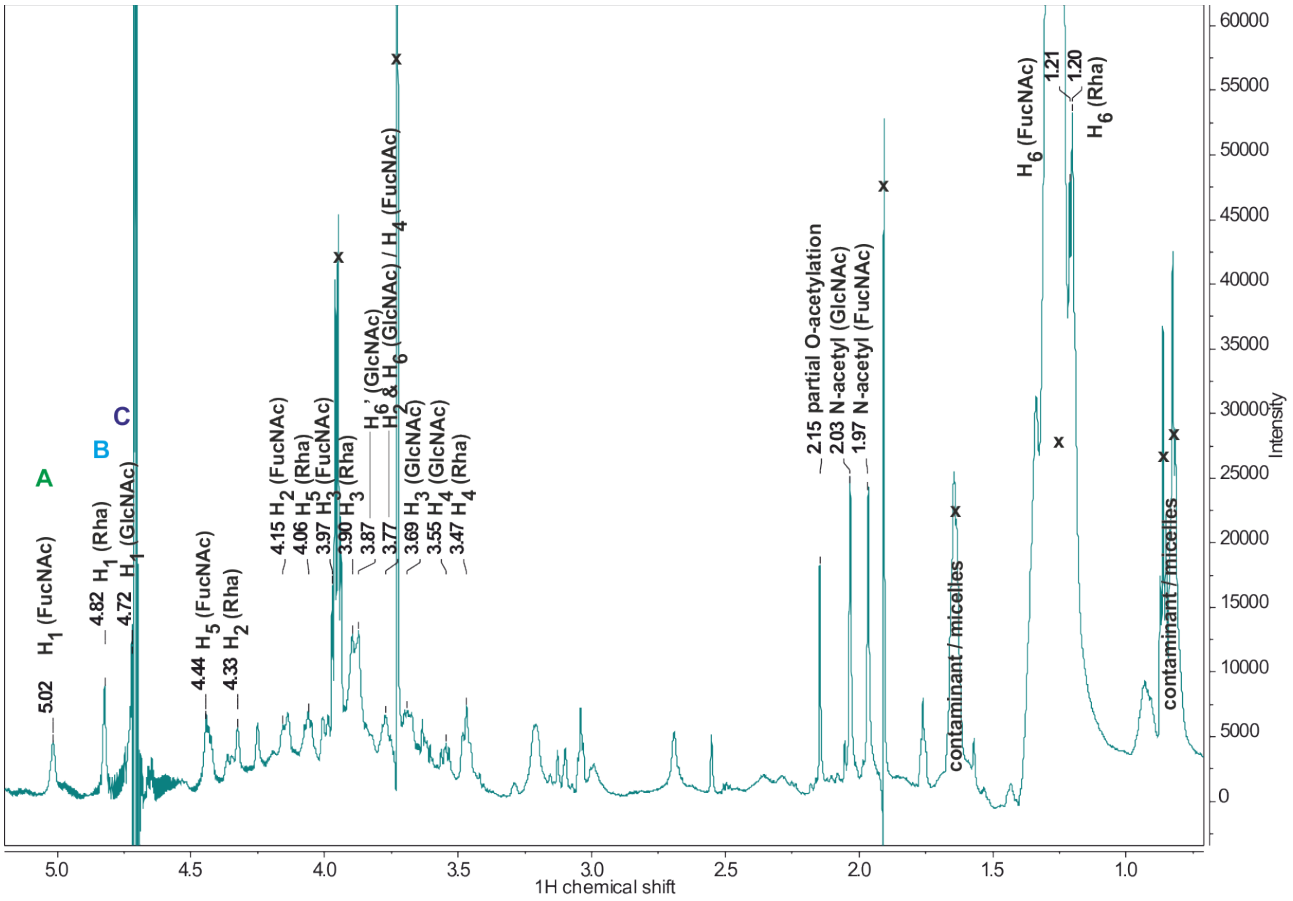


Figure 4.40 – 1D ^{1}H NMR spectrum of the *E. coli* O26 LPS embedded in mixed DPC-LPS-micelles.
The spectrum was acquired on a Bruker Avance III 600 MHz spectrometer equipped with a TXI/TCI cryoprobe in 100% D_2O. Anomeric residues of each monosaccharide spin system are labelled with letters in colour. Contaminants and micelle signals are labelled with x.

The signal at 2.15 ppm in **Figure 4.40** is discussed to disappear on alkaline treatment by Manca and colleagues. The O-acetylation of the O-antigen is partial, as was deducible from the peak intensity, which is relatively less intense than other signals in the spectrum. This would also suggest that there are two different versions of the O-antigen, which could also be the case for the *B. pseudomallei* 576 LPS, where it is common for bacterial polysaccharides to be modified further by acetylation, methylation or phosphorylation. The H-1 of the GlcNAc monosaccharide is buried beneath the D_2O peak and therefore obscured from precise assignment.

The 2D TOCSY NMR spectrum shown in **Figure 4.41** allowed for partial spin system assignments, which could be interpreted for each monosaccharide in the O-antigen repeat unit. The coloured spin system assignments for each monosaccharide present is carried forward from the assignemnts of the anomeric

hydrogen signals found in the 1D spectrum. A heteronuclear HSQC was run for the mixed DPC-LPS-micelles (results not shown) and cross peaks from each monosaccharide could be annotated.

Although a lot of information could be deduced from the spectra, the signals were still faint and the spectral accumulation needed to be run for longer times and particularly with more sample.

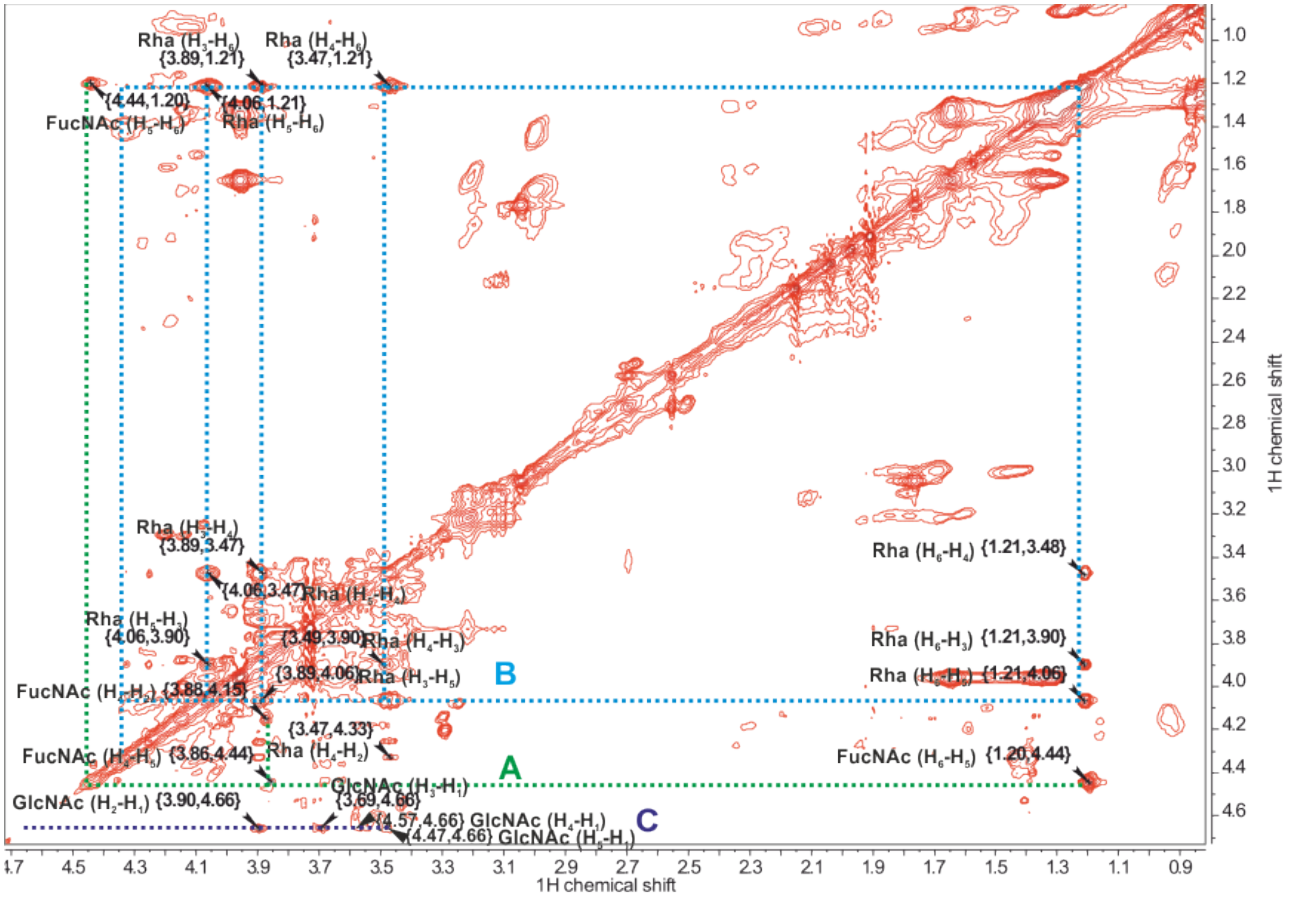


Figure 4.41 – 2D TOCSY of the *E. coli* O26 LPS embedded in mixed-DPC-micelles.
The 2D ^{1}H-^{1}H total correlation NMR spectrum shows the LPS of *E. coli* O26 LPS embedded in mixed DPC-LPS-micelles. The spectrum was acquired on a Bruker Avance III 600 MHz spectrometer equipped with a TXI/TCI cryoprobe in 100% D_2O. Spin systems of respective monosaccharide residues are shown in colour.

4.4.3 NMR analysis of *B. pseudomallei* LPS embedded in micelles

Investigation into the method using micelles to embed LPS moieties for assessment of the repeat unit structure of their O-antigens proved effective, where heterogeneity and solubility issues of the samples were overcome. The O-antigen repeat unit of *E. coli* serotype O26 was successfully assigned using both 1D and 2D NMR spectroscopy. Therefore further *B. pseudomallei* 576 LPS sample was prepared in the same way for analysis of the O-antigen repeat unit (**Section 2.4.10.2**). Since the literature on structural elucidation by NMR of LPS embedded in micelles is rather short and to overcome the problem of signals belonging to the micelle creating moieties such as DPC, NMR spectra of micelle-DPC only samples were accumulated to allow for exclusion of the respective signals. All spectra were calibrated to acetone supplemented in one of the samples and collectively aligned.

The 1D proton NMR spectrum of *B. pseudomallei* 576 LPS embedded in DPC-mixed micelles is shown in **Figure 4.42**. Anomeric hydrogen resonances were annotated with letters to denote specific spin systems that could be present in the O-antigen repeat unit. These are labelled with different colours to carry over assignments into spin lattice arrangements on 2D NMR spectra. Eight anomeric residues were denoted, evidently far too many to be consistent with the mass spectrometry derived proposed O-antigen repeat unit of four monosaccharides per repeat unit and these anomeric signals were further investigated in 2D TOCSY experiments. The 1D NMR spectrum of the micelle-embedded moieties of LPS from *B. pseudomallei* in comparison to that of the hydrolysed sample shows similarities in both the presence of the glucose moiety as well as the methyl cluster and partial O-acetylation signals previously identified.

Furthermore there is an overall pattern especially visible in the α and β anomeric region. With the objective of full characterisation of any O-antigen repeat unit or polysaccharide by NMR, the anomeric signals of the H-1 carbons (α and β anomers) were initially assessed to accumulate information on the global organisation of the O-antigen repeat unit. The number of signals found in the anomeric region provides an indication of the number of monosaccharide spin systems and hence the number of monosaccharide residues present in the O-antigen repeat unit, due to the repetitive environment encountered by each monosaccharide in a repeat unit. The proton signals corresponding to the monosaccharides spin systems could be traced from each anomeric proton signal. With the information collected by compositional and linkage analysis, as well as MALDI-MS and MALDI-MS/MS data in combination with literature values, the structure of the O-antigen in question could be projected.

A 2D TOCSY NMR spectrum was accumulated as shown in **Figure 4.43**. This allowed for partial spin system assignemnt, which could be interpreted to suggest specific monosaccharides in the O-antigen repeat unit of *B. pseudomallei* 576 LPS, by corroboration with literature values. The coloured spin system assignment for each monosaccharide present is carried forward from the assignment of the anomeric proton signals found in the 1D spectrum.

Analysis of the 2D TOCSY NMR spectrum of the *B. pseudomallei* 576 LPS embedded in micelles as prepared from a newly provided sample for NMR analysis, showed that the NMR spectra contained signals far outside the predicted chemical shifts for normal sugars (**Figure 4.44**). Upon investigation of the nature of such signals as are found between 5.5ppm and 8.0ppm, it became clear that signals representative of RNA and DNA nucleotides were present in the spectrum. Furthermore, these signals showed cross-peaks to the region investigated for the O-antigen repeat unit monosaccharide hydrogen signals, and thus created an overcrowded envelope region, that was too complex to investigate. To overcome such signal crowding and for purification of the LPS embedded in the micelles, the sample had to be further treated with Benzonase® endonuclease , which was then digested with pronase enzyme (**Section 2.4.10.2**). Analysis of the dialysed and re-exchanged sample allowed for a much cleaner spectrum **Figure 4.45**. This allowed for a much purer sample preparation for analysis by NMR and a 2D TOCSY NMR spectrum of the *B. pseudomallei* 576 LPS is shown in **Figure 4.46**, post sample purification and clean up.

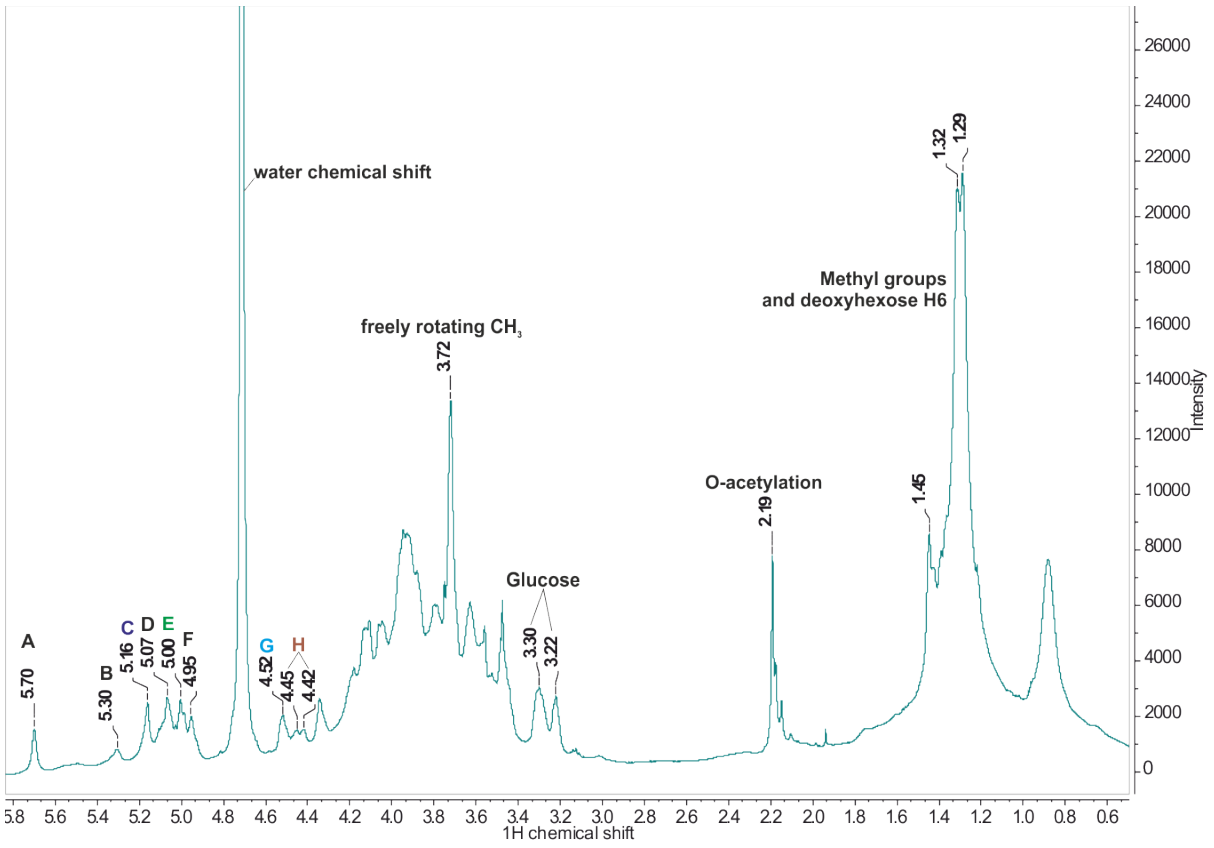


Figure 4.42 – 1D ^{1}H NMR spectrum of the *B. pseudomallei* 576 LPS in micelles.
The spectrum was acquired on a Bruker Avance III 600 MHz spectrometer equipped with a TXI/TCI cryoprobe in 100% D_2O. Proposed anomeric residues of possible monosaccharide spin system are labelled with letters in colour. Contaminants and micelle signals are labelled with x.

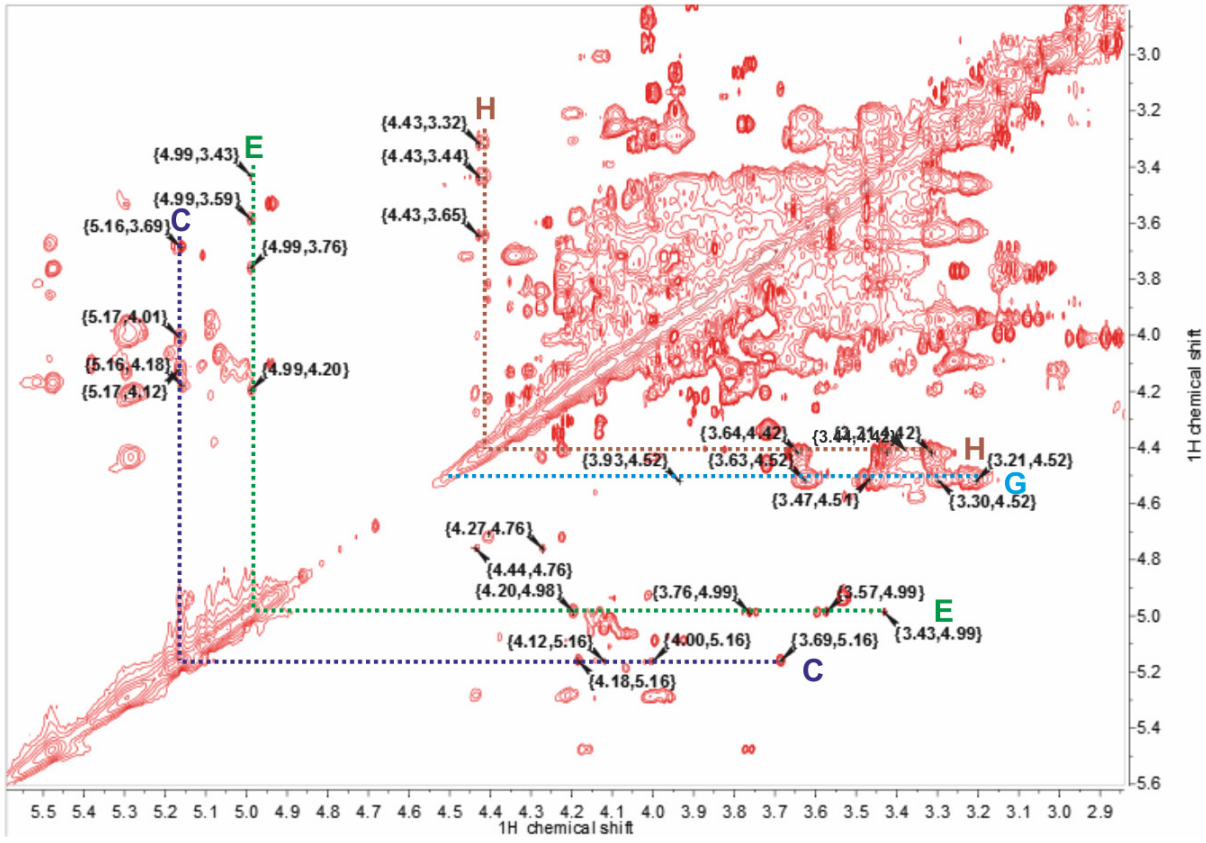


Figure 4.43 – 2D TOCSY spectrum of the *B. pseudomallei* 576 LPS in micelles.
The spectrum was acquired on a Bruker Avance III 600 MHz spectrometer equipped with a TXI/TCI cryoprobe in 100% D_2O. Proposed spin system assignments of respective monosaccharide residues are labelled with letters and highlighted in colour.

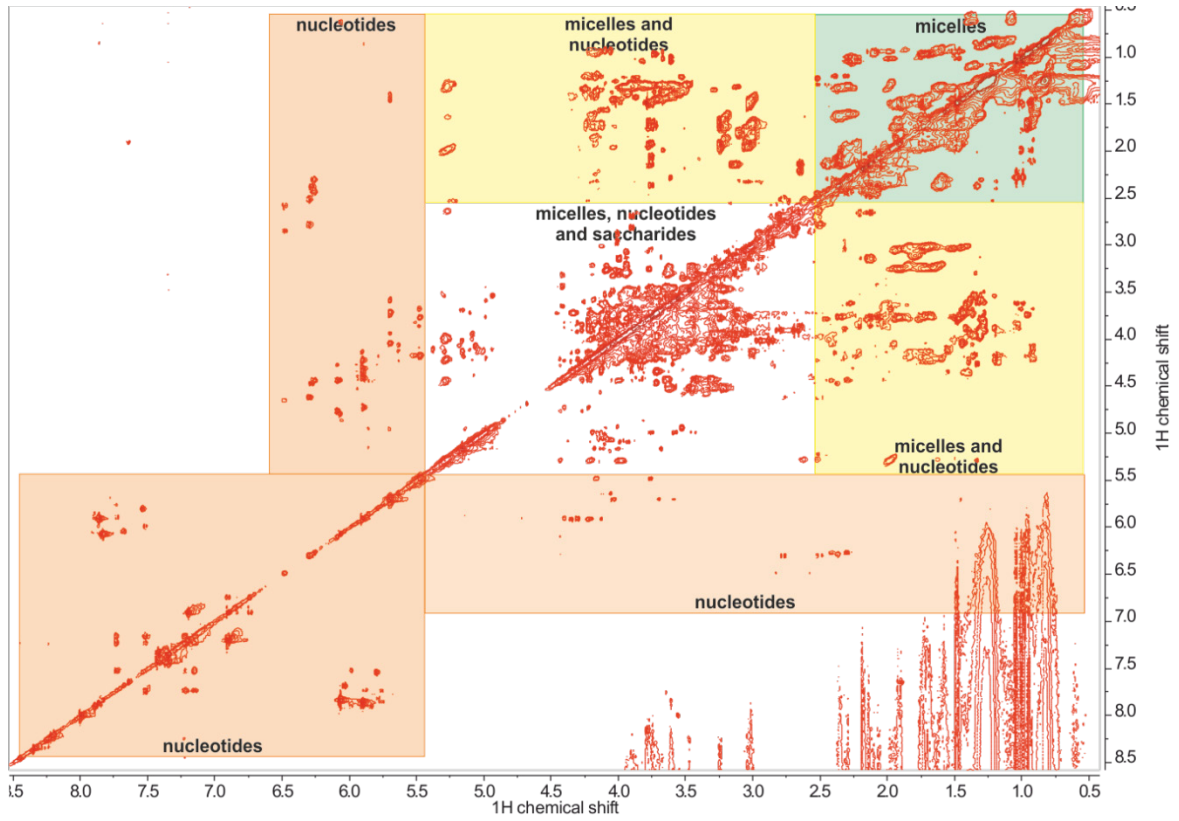


Figure 4.44 – 2D TOCSY spectrum of the *B. pseudomallei* 576 LPS (contamination).
The sample was embedded in mixed DPC-LPS-micelles. The ^{1}H-^{1}H total correlation NMR spectrum was acquired on a Bruker Avance III 600 MHz spectrometer equipped with a TXI/TCI cryoprobe in 100% D_2O. The coloured areas show contamination signals arising from the sample preparation, which contained nucleotide (RNA and DNA) signals.

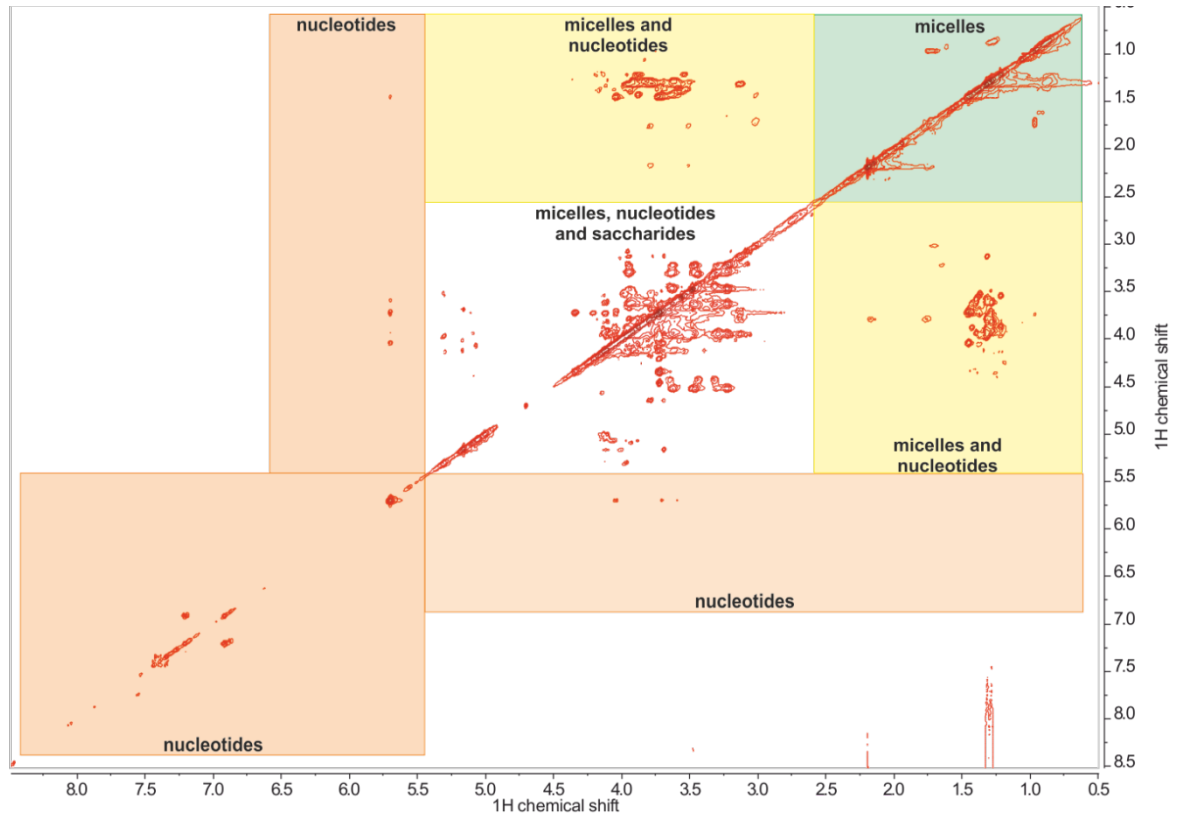


Figure 4.45 – 2D TOCSY spectrum of the *B. pseudomallei* 576 LPS (clean).
The sample was embedded in mixed DPC-LPS-micelles. The ^{1}H-^{1}H total correlation NMR spectrum was acquired on a Bruker Avance III 600 MHz spectrometer equipped with a TXI/TCI cryoprobe in 100% D_2O. The coloured areas previously showed contamination signals arising from the sample preparation which contained nucleotide (RNA and DNA) signals, which were removed by treatment with benzonase, pronase and dialysis.

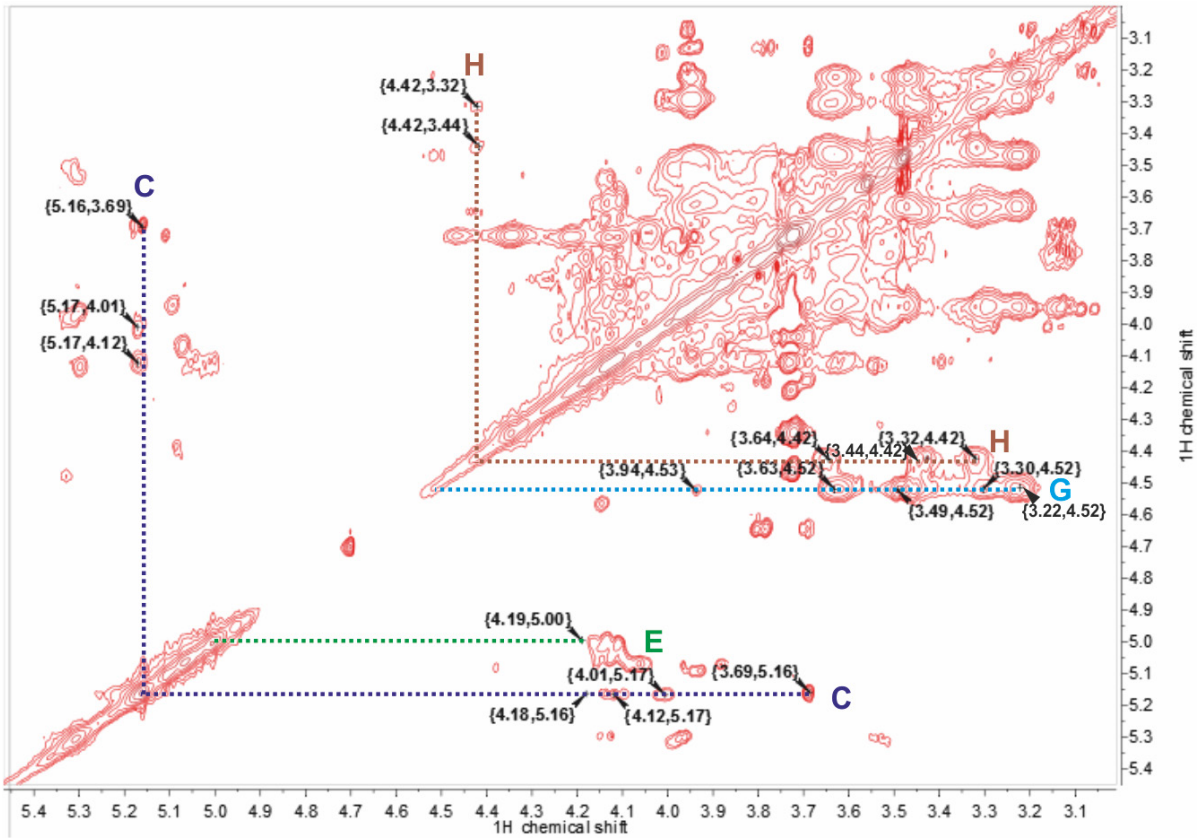


Figure 4.46 – 2D TOCSY spectrum of the *B. pseudomallei* 576 LPS in micelles (clean). The spectrum was acquired on a Bruker Avance III 600 MHz spectrometer equipped with a TXI/TCI cryoprobe in 100% D_2O. Proposed spin system assignments of respective monosaccharide residues are labelled with letters and highlighted in colour. Contamination signals arising from the sample preparation which contained nucleotide (RNA and DNA) signals were removed by treatment with benzonase, pronase and dialysis.

The assessment and annotation of the two 2D TOCSY NMR spectra of *B. pseudomallei* 576 LPS embedded in micelles allowed for the identification of spin systems associated with the anomeric signals found in the 1D NMR spectrum. The anomeric signals labelled A and B in **Figure 4.42** were initially understood to be monosaccharide spin systems from interpretation of the TOCSY spectrum. Anomeric carbon chemical shifts are generally found between 90-110ppm in pyranose or slightly above 110ppm for furanose sugars. However upon assessment of the HSQC-TOCSY carbon chemical shifts respective to the H-1 anomeric chemical shifts, the signals associated with the anomeric carbons, were not in this region. In bacterial polysacharides it is not unprecedented that non-monosaccharide components can be associated with the O-antigen moiety of an LPS, however no mass spectrometric data suggested the presence of such non-carbohydrate moieties to be included in the O-antigen repeat unit and thus the signals corresponding to spin systems A and B were determined not to be part of the O-antigen and thus disregarded at the present time. The rest of the spin systems annotated in the 1D and the 2D TOCSY NMR spectra of the *B. pseudomallei* 576 LPS O-antigen were analysed and combined as is shown in **Table 4.4**.

It is clear from the 1D proton NMR spectrum of the *B. pseudomallei* 576 LPS embedded micelles, that the creation of micelles allowed for suitable solubilisation of the LPS, however due to the now much larger nature of this new macromolecular complex, the multiplicity (fine structure) of the LPS O-antigen repeat unit is lost from the NMR spectrum. This was additionally worsened by the necessity to further purify the sample provided, where sample loss could have occurred during dialysis for removal of the contaminants, whilst

repeated D_2O exchanges might have allowed the micelles to coalesce resulting in larger micelles being formed, thus decreasing resolution of signals due to the size of the macromolecular complex.

Residue assignment	^{1}H chemical shift (ppm)
C	5.16, 4.18, 4.12, 4.00, 3.69, 1.2-1.4
D	5.09, 4.22, 3.99, 3.94
E	5.00, 4.20, 3.76, 3.57, 3.43, 1.2-1.4
F	4.94, 4.10, 4.01, 3.53
G	4.52, 3.93, 3.63, 3.47, 3.30, 3.21
H	4.43, 3.64, 3.44, 3.31

Table 4.4 – Chemical shift signals of proton spin systems of *B. pseudomallei* 576 LPS.
The proton spin system signals of the *B. pseudomallei* 576 LPS sample, as embedded in mixed DPC-LPS-micelles, were identified from the 1D proton NMR spectrum (Figure 4.42) as well as 2D TOCSY NMR spectra (Figure 4.43 and Figure 4.46). The values are to two decimal places but can vary by a decimal place within the spectra due to the difficulty of calibration at low resolution.

In **Table 4.4**, six spin systems are identified, however, mass spectrometric analysis allowed for the proposition of an O-antigen repeat unit for *B. pseudomallei* 576 LPS of a repeat unit containing four monosaccharides (one pentose, two deoxyhexoses and one hexose, denoted as a xylose, two rhamnoses and a glucose moiety). Due to the novelty of the method employed in creating mixed micelles with LPS for analys by NMR and with the time constraints on thesis submission, the method could not be optimised further. From the data obtained by NMR in the limited time permitted, the proposed O-antigen repeat unit structure cannot be verified and further analysis will be necessary to corroborate the postulated structure. There is unfortunately not enough data from the TOCSY data to allow for full spin system assignment of each monosaccharide. However, some of the anomeric signals could be ruled out as signals deriving from monosaccharides and possible assignments can be made with the support of literature values for ^{1}H chemical shifts. This will by no means validate the structure of the O-antigen unit investigated, but is presented here as a possibility that is in line with the data and structural proposition derived from mass spectrometric data.

The spin systems derived from analysis of the 1D and 2D TOCSY NMR spectra in comparison with literature values, allowed for a tentative proposition of monosaccharide identities that support previous findings by MALDI-MS, MALDI-MS/MS and GC-MS suggesting a proposed O-antigen repeat unit of a four monosaccharide repeat unit branched at a deoxyhexose (rhamnose) moiety, containing a xylose, two rhamnoses and a glucose monosaccharide moiety. Using the data available, the provisional assignments of the monosaccharide identities shown in **Table 4.5** offer the best possible interpretations that can be made at the time of thesis submission.

The unannotated spin systems shown in **Table 4.5** did not allow for identification of any monosaccharides and further information is required for complete description, however variations in environments can be expected in different parts of the O-antigen repeat unit, allowing for the occurrence of further spin systems. Furthermore, it is widely understood that bacterial LPS moieties can be extensively modified, such as is seen by acetylation and methylation, which will undoubtedly allow for such differences. The 1D proton NMR spectrum describes such modification, if only partial, to be present in this O-antigen repeat unit, providing a possible explanation for the chemical shift systems denoted by D and F.

Residue	^{1}H chemical shift (ppm)	Tentative assignment	Reference
C	5.16, 4.18, 4.12, 4.00, 3.69, 1.2-1.4	α-rhamnose (2,3-linked)	(Czerwicka et al., 2011)
	5.18, 4.19, 3.96, 3.71, 3.51, 1.27		
D	5.09, 4.22, 3.99, 3.94		
E	5.00, 4.20, 3.76, 3.57, 3.43, 1.2-1.4	α-rhamnose (2-linked)	(Mattos et al., 2005)
	4.97, 4.19, 3.89, 3.84, 3.57, 1.31		
F	4.94, 4.10, 4.01, 3.53		
G	4.52, 3.93, 3.63, 3.47, 3.30, 3.21	β-glucose (3-linked)	(Perry et al., 1995)
	4.62, 3.62, 3.49, 3.47, 3.45, 3.75/3.86		
H	4.43, 3.64, 3.44, 3.31	β-xylose (terminal)	(Molinaro et al., 2003)
	4.55, 3.69, 3.45, 3.33, 3.37/3.91		

Table 4.5 – Tentative assignments of systems in *B. pseudomallei* 576 LPS O-antigen.
The table shows the proposed monosaccharide residue assignments for the O-antigen of the *B. pseudomallei* 576 LPS and the associated reference signals from published data (italics). The signals were identified from the 1D proton NMR spectrum as well as the 2D TOCSY NMR spectra shown in Figure 4.42, Figure 4.43 and Figure 4.46, respectively. The values are to two decimal places but can vary by a decimal place within the spectra due to the difficulty of calibration at low resolution.

Inspection of the 2D homonuclear NOESY NMR spectrum of *B. pseudomallei* 576 LPS as embedded in micelles also showed the previously annotated spin system arrangements as derived from the 2D TOCSY experiments. This is shown in **Figure 4.47**. The spin systems of the previously annotated monosaccharide residues are shown in colour, whilst new chemical shifts, not previously seen in the 2D TOCSY experiments are highlighted in green. The highlighted crosspeaks generally suggest possible correlation to a different spin system/monosaccharide. Due to the partial nature of the assignment owing to the inadequate data collection, such sequence information could not definitively be deduced at the present time.

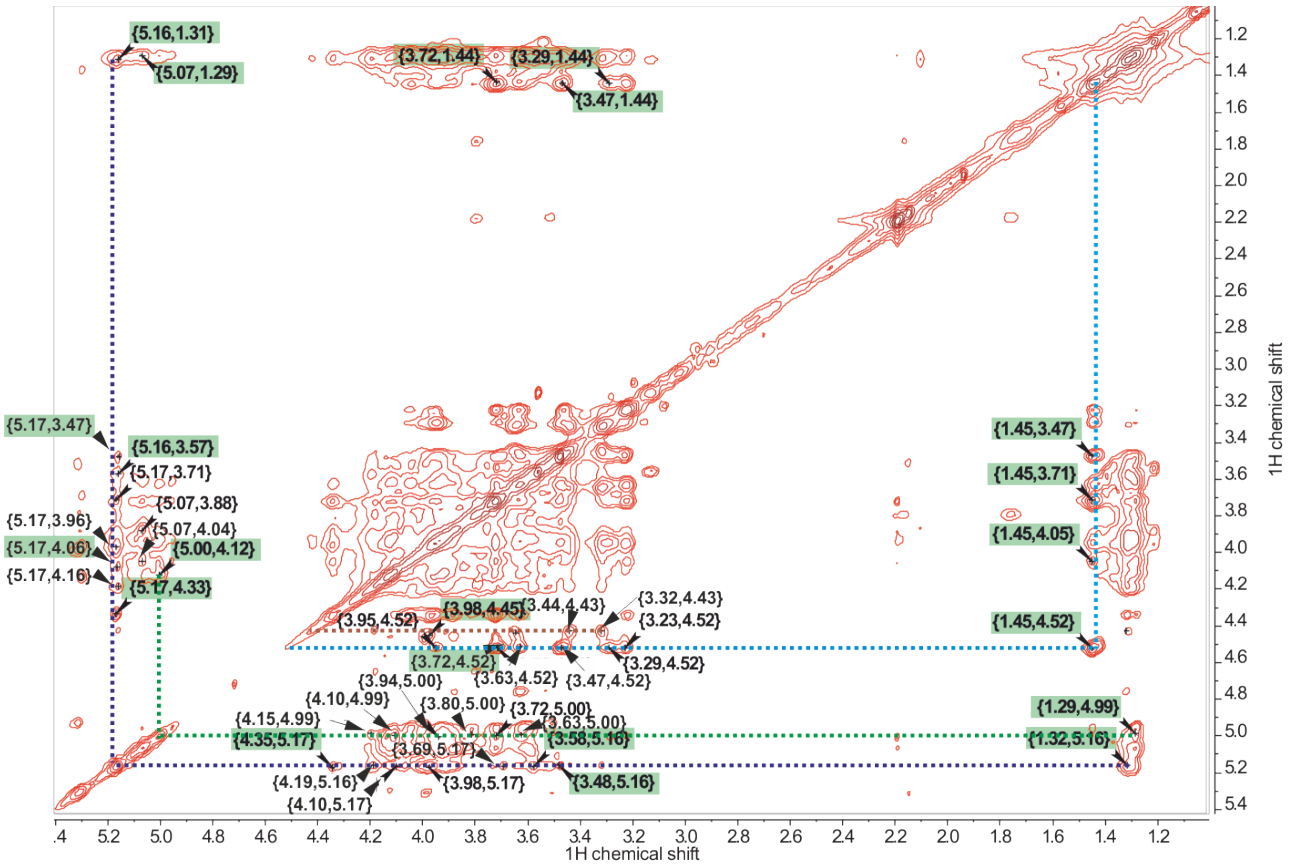


Figure 4.47 – 2D NOESY spectrum of *B. pseudomallei* 576 LPS in micelles.
The 2D homonuclear ^{1}H-^{1}H correlation NMR spectrum was acquired on a Bruker Avance III 600 MHz spectrometer equipped with a TXI/TCI cryoprobe in 100% D_2O. Spin systems of respective monosaccharide residues are shown in colour, whilst new chemical shifts, not previously seen in the 2D TOCSY experiments are highlighted in green, suggesting possible correlation to a different spin system/monosaccharide.

A 2D heteronuclear HSQC-TOCSY was run for the mixed DPC-LPS-micelles of *B. pseudomallei* 576. Again, it is important to note that the data collected up to the time of submission of this thesis, does not allow for full assignment of spin systems or characterisation of the O-antigen repeat unit, however, an attempt was made to propose a possibility that sustains the proposed O-antigen repeat structure from mass spectrometric analyses. The heteronuclear HSQC-TOCSY run for the mixed DPC-LPS-micelles of *B. pseudomallei* 576 is shown in **Figure 4.48.** The coloured lattice assignment is presented to show the partially annotated total correlation lattice of the proposed tentative β-glucose moiety.

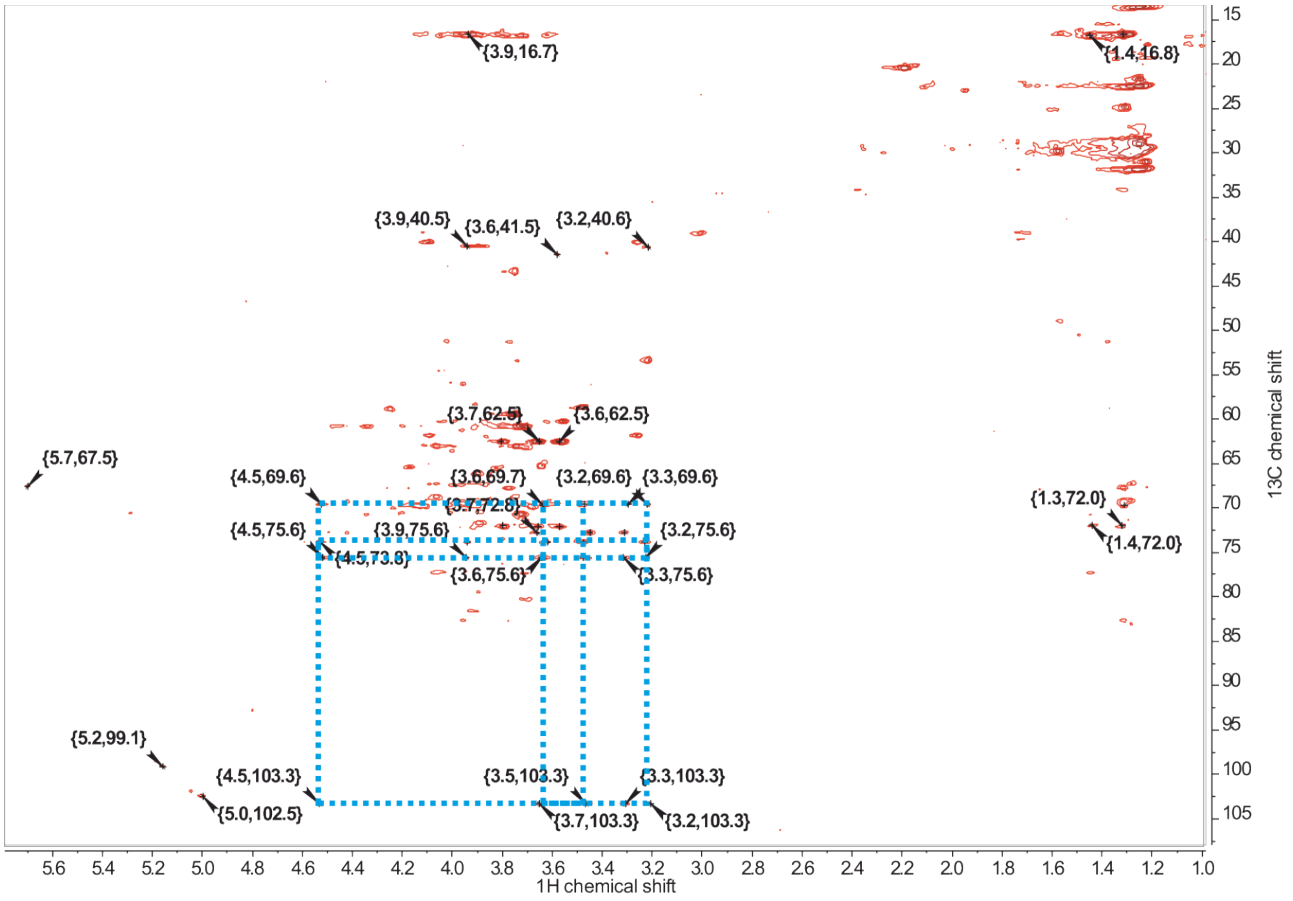


Figure 4.48 – 2D HSQC-TOCSY spectrum of *B. pseudomallei* 576 LPS in micelles.
The 2D heteronuclear ^{1}H-^{13}C total correlation NMR spectrum was acquired on a Bruker Avance III 600 MHz spectrometer equipped with a TXI/TCI cryoprobe in 100% D_2O. The coloured lattice assignment (blue) shows the partial correlations of the proposed β-glucose spin system.

Assessment of the 2D heteronuclear HSQC-TOCSY NMR spectrum shown in **Figure 4.48**, presented the opportunity to allocate carbon chemical shifts to the proposed monosaccharides that were identified in the TOCSY spectra, to further assign the proposed possible O-antigen repeat unit structure, as is presented in **Figure 4.36**. The information collected in **Table 4.6** is only one of a few possibilities of structural assignments that can be deduced from the present data, and should not be construed as definite. Further data collection and experiments are required to support any claims to this proposed structure.

Sugar residue	H/C						Reference
C →2,3)-α-Rha-(1→	5.16	4.18	4.12	4.00	3.69	1.2-1.4	(Czerwicka et al., 2011)
	5.18	*4.19*	*3.96*	*3.71*	*3.51*	*1.27*	
	99.1		69.8	67.5	72.1		
	101.4	*78.0*	*75.9*	*70.2*	*72.8*	*17.1*	
E →2)-α-Rha-(1→	5.00	4.20	3.76	3.57	3.43	1.2-1.4	(Gaur et al., 1998)
	4.97	*4.19*	*3.89*	*3.84*	*3.57*	*1.31*	
	102.5	70.0	76.3	72.1	69.6		
	102.4	*70.3*		*78.3*			
G →3)-β-Glc-(1→	4.52	3.93	3.63	3.47	3.30	3.21	(Perry et al., 1995)
	4.62	*3.62*	*3.49*	*3.47*	*3.45*	*3.75/3.86*	
	103.3	73.8	75.6	75.5	72.7	73.8	
	102.4	*82.7*	*68.7*	*74.8*	*76.6*	*61.5*	
B →t)-β-Xyl-(1→	4.43	3.64	3.44	3.31			(Molinaro et al., 2003)
	4.55	*3.69*	*3.45*	*3.33*	*3.37/3.91*		
	103.3	75.6	72.7	75.6			
	105.5	*71.5*	*77.8*	*75.6*	*65.5*		

Table 4.6 – Full tentative assignment of the chemical shifts of the *B. pseudomallei* 576 LPS O-antigen. The table shows the proposed H/C and monosaccharide residue assignments for the O-antigen of the *B. pseudomallei* 576 LPS and the associated reference signals from published data (italics). The signals were identified from the 1D proton NMR spectrum, the 2D TOCSY NMR spectra as well as the 2D HSQC-TOCSY NMR spectrum shown in Figure 4.42, Figure 4.43, Figure 4.46 and Figure 4.48, respectively. The values are to two decimal places but can vary by a decimal place within the spectra due to the difficulty of calibration at low resolution.

4.4.4 Summary and discussion of NMR analysis of LPS

The NMR data presented in this section, albeit not conclusive, provided further steps in the full structural elucidation of the O-antigen repeat unit of the *B. pseudomallei* 576 LPS. The NMR spectra obtained from hydrolysed and purified O-antigen repeat units as prepared for mass spectrometric analysis, proved difficult for *de novo* interpretation of the data as the signals were too complex owing to the heterogeneous nature of the hydrolysed sample. This brought about the search for further methods to analyse the LPS in its natural environment and research conducted, culminated in the development of a method embedding the LPS moieties into mixed-DPC-LPS micelles for analysis on soluble LPS. This technique was initially tested in a pilot study revealing concurrent results for an industrially acquired LPS of *E. coli* serotype O26, allowing for the structural determination of the LPS O-antigen.

The *B. pseudomallei* LPS embedded in micelles was further analysed by NMR using various 1D and 2D experiments. The NMR study fell short of definitive structural characterisation, as only partial information could be deduced from the sample. Problems underlying complete data collection were reflected by micelle amalgamation during repeat purification and deuterium exchange of samples. Furthermore, the resultant size of these macromolecular complexes diminished resolution and multiplicity as required for identification of the fine structure of any polysaccharide. No sequence information could be deduced from the interpreted data. The data accumulated in this series of experiments can be returned to for interpretation, once the micelle method has been further optimised with appropriate planning and more information is gathered about the structure of the O-antigen repeat unit.

However, certain proposals on the O-antigen repeat unit structure as suggested by mass spectrometric data (**Section 4.2** and **Section 4.3**) could be emphasized and supported, such as the presence of glucose in the repeat unit structure, which is reminiscent of the hexose present in the *B. pseudomallei* K96243 LPS, whereas the replacement of deoxytalose with rhamnose is explained in **Section 4.2.5**.

Anomeric resonances identified by 1D proton NMR along with assignment of their respective spin systems using TOCSY experiments and further investigation of their NOESY and HSQC-TOCSY spectra proposed partial insight into the monosaccharides comprised in the O-antigen repeat unit of *B. pseudomallei* 576. In an attempt to extract as much information of current data as possible a plausible presentation of the O-antigen repeat unit is proposed in **Figure 4.49**. This provides a likely organisation of the O-antigen repeat unit in *B. pseudomallei* 576, as was previously suggested to be a possibility from mass spectrometric MALDI-MS/MS data.

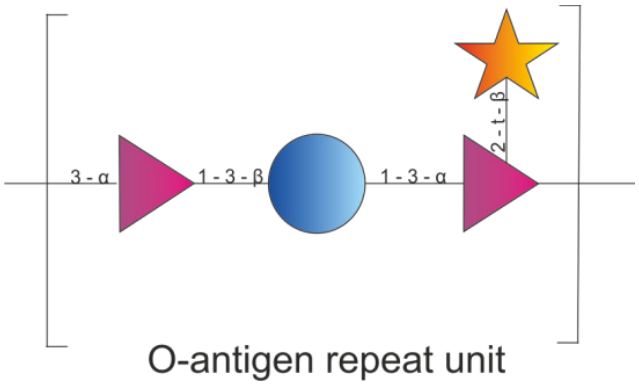


Figure 4.49 – Proposed structural repeat unit of the O-antigen of *B. pseudomallei* 576 LPS.

When considering the suggested linkages inferred from the NMR data presented in this section, the complex chromatogram of the initial linkage analysis needs to be considered (**Section 4.2.3**). The proposed structure reflects selected linkage data consistent with the GC-MS data; however some linkage chromatogram peaks were discounted for a variety of reasons, including the presence of signals arising from core domain linkages as well as the possibility of partial O-acetylation as is commonly observed in O-antigen and core polysaccharides of bacteria. The core polysaccharide was found to contain a series of monosaccharides including rhamnose, hexose, mannoheptose as well as Kdo and Ko. The presence of rhamnose and hexose in the core domain allows for some allocation of less prominent linkage peaks to such core domain linkages, including 4-linked, 3-linked and terminal deoxyhexose peaks, along with 6-linked, terminal and 2,6-linked hexose peaks in **Figure 4.10**. Furthermore, the unassigned anomeric signals and their spin systems corresponding to spin systems D and F in **Table 4.4** and **Table 4.5**, contain H-2 signals with relatively high values above 4ppm, characteristic of rhamnose residues. The identification of these further signals as arising from rhamnose systems, presents additional explanation for the complexity of the linkage chromatogram and the presence of uninterpreted signals. This further highlights the tentative nature of the O-antigen repeat unit proposal, for which the most convincing data was chosen. Moreover, it is clear from NMR data that the O-antigen repeat unit is further modified by partial O-acetylation, which presents another reason for the complexity of the linkage data in **Section 4.2.3**. Natural acetylation of any monosaccharide residue would affect its linkage mass fingerprint, such that more linkages than are present could be induced.

The four-monosaccharide repeat unit structure of *B. pseudomallei* 576 would be double the two-monosaccharide (glucose-deoxytalose) repeat unit structure of the O-antigen found in *B. pseudomallei* K96243. This seems consistent with the silver stain ladder profiles of the two serotypes, where *B. pseudomallei* 576 (serotype B) shows larger distance between the rungs in comparison with the *B. pseudomallei* K96243 (serotype A) profile as shown on in **Figure 1.10**.

The technique devised for NMR analysis of intact soluble LPS molecules as embedded in micelles, is not deemed to be at fault for the incomplete data analysis if starting samples of utmost purity are investigated. As far as is known, the method developed here has not previously been used for structural analysis of LPS by NMR and thus needs further optimisation, which was not possible to the desired extent for the work presented in this thesis. Fine-tuning of the method, however, would assess the mixed micelle sizes as related to tumbling in solution and resolution, as well as DPC-LPS-micelle ratios. LPS molecules will form micelles *in vivo* in respective solvents, which would be another avenue to be explored in the complete elucidation of the LPS O-antigen of *B. pseudomallei* 576.

4.5 CURRENT APPROACHES TO GLYCOCONJUGATE VACCINE CREATION

Genome sequencing has provided the springboard for new insights into the relatedness of *B. pseudomallei*, *B. thailandensis* and *B. mallei*, as well as their potential virulence factors and vaccine candidates, but vaccine creation requires a clearer understanding of the complexity of the lifestyle of the bacteria, their survival in the host and environment as well as the history of infection and disease (Bondi and Goldberg, 2008).

The nature of the bacteria, capable of residing in the vacuoles of eukaryotic cells, complicates antibiotic treatment and highlights the need to generate cell-mediated immunity to combat the infection (Stevens et al., 2005). The eradication of *B. pseudomallei* was proposed to be dependent on the induction of a Th1 type immune response at the onset of melioidosis, because even in the presence of protective humoral antibodies, *B. pseudomallei* was found to still be able to invade epithelial or phagocytic cells, which resulted in a long term intracellular latent infection before relapse occurred many years later (Healey et al., 2005, Jones et al., 1996, Bondi and Goldberg, 2008). Sterile immunity has rarely been reported other than at very low challenge doses and bacteria were isolated from mice that were alive at the end of these studies, suggesting that antibodies, though able to delay the onset of disease, were not able to completely eradicate the infection (Sarkar-Tyson et al., 2009).

Various vaccination strategies are being investigated with regards to *B. pseudomallei* induced melioidosis. Significant protection could be demonstrated in animal models, following vaccination with attenuated strains of *B. pseudomallei* (Haque et al., 2006a). The use of dendritic cells as a delivery vector was investigated by Healey and colleagues, whilst Chen and co-workers explored the use of a DNA vaccine against the *fliC* flagellin structural gene and the use of conjugate vaccines with CPS and LPS has been under investigation by Brett and Nelson and their teams (Nelson et al., 2004, Brett and Woods, 1996).

4.5.1.1 Whole-cell vaccines for *B. pseudomallei*

The creation of non-viable, whole-cell bacterial preparations have traditionally represented a starting point in the development of vaccines and have been shown to be effective in the prevention of cholera, typhoid fever, whooping cough, anthrax and plague (Bondi and Goldberg, 2008). This technique was employed against *B. mallei* in BALB/c mice and showed immune responses with cytokines (IFN-γ, IL-4 and IL-10), however the mice seemed to be unable to generate a directed Th1 or Th2-like response (Amemiya et al., 2002). Heat-

inactivated *B. pseudomallei* immunogens have also been found to provide some protection in the mouse model of infection (Barnes and Ketheesan, 2007, Sarkar-Tyson et al., 2009). Immunisation with killed whole cell vaccines provided protection against subsequent challenge with *B. pseudomallei* in mice, however, similar to the *B. mallei* model, sterile immunity was difficult to achieve. One approach to overcome this problem and to enhance the protective response induced by inactivated preparations, was to combine them with immune adjuvants that promote the production of Th1 and proinflammatory cytokines as well as the maturation and activation of professional antigen-presenting cells (Elvin et al., 2006).

4.5.1.2 Live-attenuated vaccines for *B. pseudomallei*

Various vaccination strategies that have been investigated have demonstrated the necessity of a cell-mediated immune response in conjunction with a humoral response to successfully clear infection by *B. pseudomallei*. This mechanism of immunity was explored by Haque and colleagues where IFN-γ was found to be important for protection against acute sepsis in mice. Immunisation with live immunogens was reported to activate antigen-specific CD4+ T cells which produced high levels of IFN-γ, placing high importance on IFN-γ (Haque et al., 2006a, Haque et al., 2006b, Santanirand et al., 1999). To determine the roles of specific T cells in protection, immunised mice were depleted of CD4+ and CD8+ T cells before and after *B. pseudomallei* challenge. CD4+ T cell depleted mice were substantially more susceptible to infection compared to controls, whereas CD8+ T cell depletion had no effect on the vaccine mediated protection, suggesting that protection generated by this vaccine is mediated by CD4+ T cells only (Atkins et al., 2002, Easton et al., 2007, Haque et al., 2006a). The vaccinated mice, however, were unable to completely clear challenge infection, even in the presence of both CD4+ and CD8+ cells, further suggesting the necessity of cytotoxic T lymphocytes to clear the host cells (Haque et al., 2006b).

Bacterial polysaccharides have been found to be potent stimulators of the host immune system and response, thus have previously represented a critical component of subunit and conjugate vaccines for clinically relevant diseases including *Haemophilus influenza* type b, *Pneumococcal* and *Meningococcal* infections (Bondi and Goldberg, 2008).

LPS have been shown to be an immunodominant antigen, as recognised in patients infected with *B. pseudomallei*, where anti-LPS antibodies were found at higher levels in surviving patients and those with nonsepticaemic melioidosis as compared to those with septicaemic melioidosis and those that died (Wiersinga et al., 2006). This suggests that antibodies raised against this antigen, may protect the host from death, prompting a more efficient host immune response including cell-mediated killing, thus making this a noteworthy potential approach to vaccine development. The capsular polysaccharide of *B. pseudomallei* was also shown to be an important virulence determinant in Syrian golden hamsters and BALB/c mice (Perry et al., 1995, Reckseidler et al., 2001). Both LPS and CPS were evaluated as subunit vaccine candidates against *B. pseudomallei* by Nelson and colleagues; however protection by this procedure was not complete (Nelson et al., 2004).

It was suggested that the antibodies were most likely directed against these cell-surface polysaccharide of *B. pseudomallei*, as the passive transfer of polyclonal sera against the antigens into naive mice was found to double their survival time after subsequent challenge. Nelson and co-workers used LPS and CPS to

immunise BALB/c mice, where the polysaccharide antigens induced varying antibody responses. LPS vaccinated mice mostly developed IgM and IgG3 responses, whilst CPS vaccinated mice developed mainly IgG2b responses. When challenged by intraperitoneal injection, mice exhibited the highest level of protection upon immunisation with LPS, where 50% survival was seen at the last day of monitoring (day 35), with a mean time to death (MTTD) of 17.6 days. Mice immunised with CPS also showed increased MTTD, however exhibited 100% mortality by day 28.

The use of polysaccharides to achieve active immunity was shown to provide some, but not comprehensive protection against experimental murine melioidosis via the intraperitoneal route and the polysaccharide moiety would have to be complemented with a protein subunit to stimulate the protective cell-mediated immune response, thus increasing the efficiency of the subunit vaccine (Nelson et al., 2004). It was further suggested that the natural exposure to *B. thailandensis* in melioidosis endemic areas and the subsequent development of an antibody response without overt infection, may also provide protection against melioidosis (Nelson et al., 2004).

Studies presented in this thesis with our collaborators at the Defence Science and Technology Laboratory (dstl), showed that LPS from *B. thailandensis* provided protection against a lethal challenge of *B. pseudomallei* in a mouse model of melioidosis (**Section 4.2**). Sugar analysis of the LPS from *B. thailandensis* E264 confirmed similar composition and structures of the O-antigen (**Figure 4.12**, **Figure 4.13** and **Figure 4.14**). This was also suggested by the ladder profile on a silver stained LPS (**Figure 1.10**), showing the O-antigen polysaccharide had a similar structure to that of the LPS of *B. pseudomallei* K96243, bar the capsule and O-methylation discussed previously (**Section 1.3.4.2**). This addresses the importance of understanding the structure of the various O-antigens present in the *B. pseudomallei* strains for a comprehensive subunit based vaccine creation. Immunisation of BALB/c mice with both *B. thailandensis* and *B. pseudomallei* LPS provided similar levels of protection upon a challenge with a lethal dose of *B. pseudomallei* K96243, suggesting the use of *B. thailandensis* LPS as a potential candidate to make up a subunit vaccine against pathogenic *B. pseudomallei*. Mice immunised with the LPS from *B. thailandensis* showed 50% survival, compared to 66% survival of the mice immunised with *B. pseudomallei* LPS. The mean time to death (MTTD) of naive mice was 2.2 days, whilst the MTTD of mice immunised with *B. thailandensis* and *B. pseudomallei* LPS was 32.8 and 31.5 days respectively. The production of *B. thailandensis* is associated with reduced costs and hazards and thus was suggested to present a very suitable molecule to be studied regarding the development of a vaccine against melioidosis infection (Ngugi et al., 2010).

This phenomenon is reflected in geographical epidemiology studies, that found that the rate of the closely related but less virulent *B. thailandensis*, may account for the variation in disease throughout Southeast Asia and may also account for the much higher rates of seropositivity seen in Thailand compared to the areas of endemicity of northern Australia (Trakulsomboon et al., 1999). This is further supported by the high prevalence of melioidosis in the Issan region of north-eastern Thailand, which contrasts with low prevalence in Laos to the east of the Mekong River and in Cambodia further south (Phetsouvanh et al., 2001). Despite the recovery of *B. pseudomallei* from soil isolates in the Vientiane region, only a handful of cases are registered in Laos (Wuthiekanun et al., 2005). This paradox could be explained by the uncertainty associated

with the seropositivity rates in Southeast Asia, which may represent exposure to the less pathogenic *B. thailandensis* (White, 2003).

The effectiveness of live attenuated vaccines depends both on the number of different antigens presented and on the ability to stimulate both the antibody and cell-mediated arms of the immune system. The likelihood of an effective live attenuated vaccine against *B. pseudomallei* being developed may be limited by the bacterium's potential to establish a persistent infection. Successful development of such vaccines may depend on gaining a fuller understanding of the genetic basis underlying latency and how this can be effectively disabled (Sarkar-Tyson and Titball, 2010). LPS and CPS of *B. pseudomallei* have been suggested to be protective antigens and important determinants of virulence, however, even though CPS and LPS show promise as vaccine candidates, immunisation with these antigens alone does not appear to provide complete protection against disease or sterile immunity (Sarkar-Tyson et al., 2007).

4.5.1.3 Vaccines aimed at other *B. pseudomallei* antigens

Patients recovering from melioidosis presented antibodies that reacted with *B. pseudomallei* proteins, including the lipoprotein-releasing system transmembrane protein (LolC), an adenosine triphosphate-binding cassette transporter protein and two outer membrane proteins (OmpAs) of *B. pseudomallei* that reacted with melioidosis (Hara et al., 2009, Harland et al., 2007). The potential of inducing protective immunity after immunisation with individual *B. pseudomallei* antigens is there, but it has become clear that single component vaccines, at least delivered as a purified polysaccharide or protein, will not provide complete protection against disease and polysaccharide conjugate vaccines as well as DNA vaccines are being investigated.

Flagella have been implicated in cell invasion of phagocytic and nonphagocytic cells (Chuaygud et al., 2008) and electron microscopy studies have demonstrated the presence of flagella and the variable expression of pili on *B. pseudomallei* (Vorachit et al., 1995). Mutagenesis screens for genes involved in motility of the bacterium identified 19 unique genetic loci (DeShazer et al., 1997), whilst bioinformatic analysis of the K96243 genome of *B. pseudomallei* identified 13 gene clusters predicted to be involved in the synthesis of type I fimbria, type IV pili and Tad-like pili (Holden et al., 2004). The flagella of clinical and environmental *B. pseudomallei* isolates are antigenically conserved and polyclonal rabbit antisera raised against flagellin or a flagellin polysaccharide antigen conjugate passively protected infant diabetic rats from challenge with *B. pseudomallei* (DeShazer et al., 1997). The flagella of a range of different bacterial species have been shown to have a role in virulence, through either motility directly or adherence. DNA vaccines are potent stimulators of both humoral and cell-mediated immunity. Candidate DNA vaccines against melioidosis have been investigated based on the ability to promote cell-mediated immunity. Studies on the *B. pseudomallei* flagellar subunit gene *fliC* have shown that the FliC naked DNA resulted in the induction of both antibody and cell-mediated responses, and reduced the number of bacteria found in the liver and spleen of challenged mice as compared to the control mice (Chen et al., 2006). The potency of this naked DNA construct was further improved by incorporating CpG oligodeoxynucleotide motifs into the DNA plasmid/*fliC* construct.

On the basis of investigating the genes encoding the known immunogenic outer membrane proteins of *B. pseudomallei*, outer membrane protein A (OmpA) was assessed as a potential candidate for vaccine targets using a bioinformatics based approach. OmpA's are known to be virulence factors in bacterial pathogenesis and are capable of inducing both humoral and cytotoxic responses (Kim et al., 2000). Out of the 12 ORFs assessed, two OmpAs (Omp3 and Omp7) were successfully cloned, expressed as soluble proteins and purified. They were found to induce 50% protection upon immunisation with the respective OmpA via the intraperitoneal route, examined over 21 days post-challenge (Hara et al., 2009).

The immunogenicity of bacterial polysaccharides can be improved by covalently linking them to protein carriers to form a conjugate. Polysaccharide-protein glycoconjugates work by promoting T-cell-dependent immune responses, thus enhancing immunologic memory (Jones, 2005, Wuorimaa et al., 2001). Glycomic research is now opening doors for carbohydrate vaccine researchers to better tackle the challenges inherent to carbohydrate vaccine development and to expand the field to encompass a broader spectrum of diseases (Astronomo and Burton, 2010).

4.6 FUTURE DIRECTIONS

Efforts made into the structural elucidation of the *B. pseudomallei* 576 LPS, although incomplete, have nevertheless provided a plausible identification of the structural repeat unit of the O-antigen. Mass spectrometric data allowed for determination of the monosaccharides comprised in the O-antigen repeat unit, as well as provided structural information on the constitution of the O-antigen repeat unit resulting in the proposal of a series of possibilities. The complementing NMR analyses conducted have not yet produced evidence of the explicit configuration of said O-antigen repeat unit, but have provided valuable insight into the methods employable and the analytical strategies to finalise such a goal.

Further experimentation on the side of the mass spectrometric approach in such complete structural characterisation will focus on the further modifications decorating most bacterial O-antigen and that have been evident throughout the strategies employed in this research. Bacterial core and O-antigen polysaccharides are commonly further modified by acetyl, methyl or phosphate groups. The presence of phosphates on the core or O-antigen polysaccharides can be investigated by dephosphorylation with hydrofluoric acid (HF), which has been shown to selectively remove phosphate and phosphate linked moieties that are then lost during purification upon derivatisation of the samples (Prehm et al., 1975). Comparison of samples treated or untreated with such dephosphorylation techniques will provide insight into this modification. Comparisons of de-O-acylated and untreated samples can also further give insight into the presence of acetylation on specific monosaccharides.

Although mass spectrometry is an invaluable tool for structural characterisation, limitations to the complete elucidation of LPS include the anomeric configuration of individual sugars and the location within a molecule of sugar epimers. Generally such determination is investigated using NMR, but the possibility of the determination of the anomeric configuration exist in the form of commercially available enzymes, which are, on the other hand, largely unavailable for the vast array of bacterial sugars. The determination of absolute configuration of the monosaccharides (D- and L-) can be achieved in an adjusted method, where hydrolysis

occurs in the presence of butanolic-HCl prior to TMS derivatisation and GC-MS analysis in the presence of standards.

The method devised for LPS analysis by NMR in creating mixed-DPC-LPS micelles is still in its infancy and will further need to be fine-tuned with regards to avoid tumbling and resolution problems faced in the series of experiments described in this thesis, due to micelle coagulation and the creation of macromolecular complexes that prevent the accumulation of information on the multiplicity of the polysaccharide investigated. Further experiments not attempted at the present time will include heteronuclear multiple-bond spectroscopy (HBMC), which allows for correlation of ^{1}H and ^{13}C atoms separated by several chemical bonds, thus allowing for sequence information over the glycosidic linkages and allowing for unequivocal determination of linkage positions of the *B. pseudomallei* 576 LPS.

Chapter 5
Overview and Concluding remarks

5 OVERVIEW AND CONCLUDING REMARKS

This thesis documents two related projects, where high sensitivity methodologies including mass spectrometry and nuclear magnetic resonance were employed and developed with the aim of characterising bacterial and archaeal glycopolymers. Appropriate structural analysis methods were exploited for the evaluation of prokaryotic problems with the focus on the LPS O-antigen of *B. pseudomallei* 576 and the S-layer glycoprotein of the halophilic archaeon *H. volcanii*. The projects documented in this thesis demonstrate sample and goal specific strategies employed in order to gain further understanding of glycosylation in prokaryotic organisms.

The first project, discussed in **Chapter** 3, used a series of complementary high sensitivity mass spectrometric methods to unambiguously demonstrate the sequence fidelity as well as site-specific N-glycosylation pattern of the S-layer glycoprotein of *H. volcanii*. Glycosylation composition, location as well as insight into a further putative post-translational modification of the protein were also investigated. Such methods coupled with biological experiments in genomics and proteomics, further allowed for functional characterisation of several *agl* genes in the *agl* gene cluster of *H. volcanii* providing more insight into the N-glycan biosynthetic pathway in archaea. These investigations demonstrated the diversity of mass spectrometry as a technology that can be exploited to achieve a variety of different goals addressing biological and structural problems.

Investigations using GC-MS and MALDI-MS/MS led to evidence for the composition of the glycan moieties present on the S-layer glycoprotein as well as the confirmation of the O-glycan present, highlighting the necessity for enzymatic and chemical methods appropriate for each glycosylation type in prokaryotes. Functional characterisation by MALDI-MS of the pentasaccharide N-glycan decorating the first glycosylation site, provided insight into three further *agl* genes from the *agl* gene cluster, responsible for such N-glycosylation, demonstrated the importance of the rapport between genomic experimentation and mass spectrometric mapping, in order to get an overview of the N-glycosylation biosynthesis process in archea. This work sparked further interest in the overall site-specific occupation of this N-glycan throughout the S-layer glycoprotein. A host of different protein degradation, separation and mass spectrometric ionisation techniques, including MALDI-MS/MS, ESI-MS/MS and MS^E, were employed to create a comprehensive map of the S-layer glycoprotein, with the exception of the C-terminal domain. Six out of seven predicted N-glycosylation consensus sites were mapped and found to be occupied with the same pentasaccharide N-glycan together with its biosynthetic precursors. The use of various analysis techniques specifically matched to the differences in biologically and chemically prepared samples, exhibits the multifaceted applications of mass spectrometry.

The problems encountered in the profiling of the C-terminal region, which includes the final N-glycosylation site, a threonine-rich region with expected O-glycosylation as well as a putative membrane spanning domain, pushed investigations into the direction of characterising a potential lipid modification of the S-layer glycoprotein. With the objective geared towards providing definitive proof of previously proposed covalent lipid modification, mass spectrometric techniques were fine-tuned and a structural, high sensitivity approach was employed, different from the biological experiments that gave rise to such suggestions. No such

validations could be made on the current sample preparations, opening the debate on possibilities of such lipid association or the mechanisms by which these could occur. Other than accumulating information on protein glycosylation in the *H. volcanii* S-layer and thus providing further understanding of this process in archaea, these studies raised further intriguing questions specific to archaeal mechanisms of glycoprotein modification, lipid anchoring of the protein in the lipid bilayer, as well as creating a basis for investigation of glycosylation in halophiles as a mechanism for adaption to variations in environments.

The second project sought to characterise the structure of the O-antigen repeat unit of the serotype B LPS of an atypical strain of *B. pseudomallei*, a pathogen that is the causative agent of melioidosis. **Chapter** 4 describes the O-antigen structural information that emerged from several chemical and derivatisation techniques teamed with high sensitivity mass spectrometric techniques as well as a novel approach to nuclear magnetic resonance for LPS analysis. Various sample handling methods were employed and devised, including new hydrolysis techniques for investigation of the O-antigen repeat unit for analysis by mass spectrometry, as well as creating a new approach for LPS analysis by NMR, which involved embedding the LPS into micelles. Mass spectrometry and nuclear magnetic resonance techniques are complementary approaches allowing for the accumulation of different information on the investigated sample for structural characterisation.

The results of this work are aiding the development of a comprehensive glycoconjugate vaccine. In the structural characterisation of LPS, no specific approach is commonly employed and methods have to be improved, modified and adapted to the unknown LPS under investigation, as only little is known on the structural biosynthesis of O-antigen units of the LPS in different bacteria. This is further complicated by the many variations in further modifications that monosaccharide repeat units can undergo. However, the importance of correct structural assignment of O-antigens, inclusive of further modifications by methylation, phosphorylation or acetylation is not to be underestimated when the final aim of such structural characterisation lies with the creation of an all-encompassing glycoconjugate vaccine. This warranted a multi-faceted approach exploring several avenues for investigation, resulting in a range of techniques explored throughout this project with the purpose of producing suitable sample preparations analysable by mass spectrometry.

The development of a procedure circumventing such difficulties in hydrolysis techniques tailored to the structure of the LPS, was proposed in **Section 4.4**, when the LPS was embedded in micelles for analysis by multi-dimensional NMR. This technique permits LPS analysis without degradation allowing assessment of all further glycan modifications and avoiding the limitations encountered with other techniques, where degradation could lead to the loss of glycan modifications during sample preparation. This technique, albeit in its infancy, when further fine-tuned can be applied to any phenol extracted LPS moiety, due to the relative similarity of lipid A properties across the board and thus presents enormous potential.

Substantial progress has been made in each of the above projects. Nevertheless the findings presented in this thesis are not complete, but further directions in each project are discussed in **Section 3.8** and **Section 4.6**, which describe current issues to be addressed in future work in order to round off these investigations.

Researchers are only now beginning to appreciate the variety of prokaryotes capable of protein glycosylation and the wider range of monosaccharides found in the glycan moieties of prokaryotic glycoproteins. Studying glycosylation in relatively less-complicated prokaryotic systems provides the opportunity to elucidate and exploit glycoprotein and lipopolysaccharide biosynthetic pathways. A review of the genetic organisation, glycan structures and function of glycosylation systems in prokaryotes may help us to understand glycosylation processes in more complex eukaryotic systems and how these can be used for glycoengineering. Continued investigation into the bacterial N-glycosylation and O-glycosylation process will advance glycoengineering efforts as well as have therapeutic value in the development of new antibacterial agents. A more comprehensive understanding of N-glycosylation and O-glycosylation in archaea also carries enormous applied potential, given the possible links between glycosylation of archaeal proteins and their ability to withstand diverse physical challenges. Furthermore, such studies give insight into this post-translational modification across evolution as well as protein processing under extreme conditions.

The field of glycomics in the prokaryotic domain has slightly lagged behind its eukaryotic counterpart, in part, due to the unexplored processes underlying glycosylation in archaea and bacteria as compared to the established pathways described in eukaryotic cells. The simpler glycosylation systems encountered in prokaryotes, however, produce glycopolymers and glycoconjugates of an extremely high complexity in composition and structure in relation to those found in eukaryotes. With the advances of molecular biology and with improved analytical techniques to some extent described in this thesis, a significant change of perception has taken place regarding prokaryotic polysaccharides and glycoproteins. Glycosylation in prokaryotes is no longer considered a specific feature of certain organisms but has been demonstrated for many archaea and bacteria.

The work described in this thesis has brought together many established techniques and described the development of new methods in the field of protein, glycan and lipid characterisation. It is hoped that this research, not only allows for the creation of a fully comprehensive glycoconjugate vaccine against *B. pseudomallei* and related melioidosis-causing bacteria, but also provides further insight into glycosylation of important archaeal proteins.

Chapter 6

References

6 REFERENCES

AAS, F. E., VIK, A., VEDDE, J., KOOMEY, M. & EGGE-JACOBSEN, W. 2007. Neisseria gonorrhoeae O-linked pilin glycosylation: functional analyses define both the biosynthetic pathway and glycan structure. *Mol Microbiol,* 65**,** 607-24.

ABDUL HALIM, M. F., PFEIFFER, F., ZOU, J., FRISCH, A., HAFT, D., WU, S., TOLIC, N., BREWER, H., PAYNE, S. H., PASA-TOLIC, L. & POHLSCHRODER, M. 2013. Haloferax volcanii archaeosortase is required for motility, mating, and C-terminal processing of the S-layer glycoprotein. *Mol Microbiol,* 88**,** 1164-75.

ABEYRATHNE, P. D., DANIELS, C., POON, K. K., MATEWISH, M. J. & LAM, J. S. 2005. Functional characterization of WaaL, a ligase associated with linking O-antigen polysaccharide to the core of Pseudomonas aeruginosa lipopolysaccharide. *J Bacteriol,* 187**,** 3002-12.

ABU-QARN, M. & EICHLER, J. 2006. Protein N-glycosylation in Archaea: defining Haloferax volcanii genes involved in S-layer glycoprotein glycosylation. *Mol Microbiol,* 61**,** 511-25.

ABU-QARN, M., EICHLER, J. & SHARON, N. 2008a. Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Curr Opin Struct Biol,* 18**,** 544-50.

ABU-QARN, M., GIORDANO, A., BATTAGLIA, F., TRAUNER, A., HITCHEN, P. G., MORRIS, H. R., DELL, A. & EICHLER, J. 2008b. Identification of AglE, a second glycosyltransferase involved in N glycosylation of the Haloferax volcanii S-layer glycoprotein. *J Bacteriol,* 190**,** 3140-6.

ABU-QARN, M., YURIST-DOUTSCH, S., GIORDANO, A., TRAUNER, A., MORRIS, H. R., HITCHEN, P., MEDALIA, O., DELL, A. & EICHLER, J. 2007. Haloferax volcanii AglB and AglD are involved in N-glycosylation of the S-layer glycoprotein and proper assembly of the surface layer. *J Mol Biol,* 374**,** 1224-36.

ACQUOTTI, D. & SONNINO, S. 2000. Use of Nuclear Magnetic Resonance Spectroscopy in Evaluation of Gangliosid Structure, COnformation and Dynamics. *In:* MERRILL, A. H., JR. & HANNUN, Y. A. (eds.) *Sphingolipid Metabolism and Cell Signaling Part B.* London, UK: Academic Press Limited.

ALAIMO, C., CATREIN, I., MORF, L., MAROLDA, C. L., CALLEWAERT, N., VALVANO, M. A., FELDMAN, M. F. & AEBI, M. 2006. Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. *EMBO J,* 25**,** 967-76.

ALBERSHEIM, P. 1967. A method for the analysis of sugars in plant cell-wall polysaccharides by gas-liquid chromatography. *Carbohydr Res,* 5**,** 340-345.

ALTMAN, E., CHANDAN, V., LI, J. & VINOGRADOV, E. 2011a. Lipopolysaccharide structures of Helicobacter pylori wild-type strain 26695 and 26695 HP0826::Kan mutant devoid of the O-chain polysaccharide component. *Carbohydr Res,* 346**,** 2437-44.

ALTMAN, E., CHANDAN, V., LI, J. & VINOGRADOV, E. 2011b. A reinvestigation of the lipopolysaccharide structure of Helicobacter pylori strain Sydney (SS1). *FEBS J,* 278**,** 3484-93.

AMEMIYA, K., BUSH, G. V., DESHAZER, D. & WAAG, D. M. 2002. Nonviable Burkholderia mallei induces a mixed Th1- and Th2-like cytokine response in BALB/c mice. *Infect Immun,* 70**,** 2319-25.

ANDO, S., KON, K., NAGAI, Y. & MURATA, T. 1977. Chemical ionization and electron impact mass spectra of oligosaccharides derived from sphingoglycolipids. *J Biochem,* 82**,** 1623-31.

ANUNTAGOOL, N., ARAMSRI, P., PANICHAKUL, T., WUTHIEKANUN, V. R., KINOSHITA, R., WHITE, N. J. & SIRISINHA, S. 2000a. Antigenic heterogeneity of lipopolysaccharide among Burkholderia pseudomallei clinical isolates. *Southeast Asian J Trop Med Public Health,* 31 Suppl 1**,** 146-52.

ANUNTAGOOL, N., INTACHOTE, P., WUTHIEKANUN, V., WHITE, N. J. & SIRISINHA, S. 1998. Lipopolysaccharide from nonvirulent Ara+ Burkholderia pseudomallei isolates is immunologically indistinguishable from lipopolysaccharide from virulent Ara- clinical isolates. *Clin Diagn Lab Immunol,* 5**,** 225-9.

ANUNTAGOOL, N., NAIGOWIT, P., PETKANCHANAPONG, V., ARAMSRI, P., PANICHAKUL, T. & SIRISINHA, S. 2000b. Monoclonal antibody-based rapid identification of Burkholderia pseudomallei in blood culture fluid from patients with community-acquired septicaemia. *J Med Microbiol,* 49**,** 1075-8.

ANUNTAGOOL, N., WUTHIEKANUN, V., WHITE, N. J., CURRIE, B. J., SERMSWAN, R. W., WONGRATANACHEEWIN, S., TAWEECHAISUPAPONG, S., CHAIYAROJ, S. C. & SIRISINHA, S. 2006. Lipopolysaccharide heterogeneity among Burkholderia pseudomallei from different geographic and clinical origins. *Am J Trop Med Hyg,* 74**,** 348-52.

ARJCHAROEN, S., WIKRAIPHAT, C., PUDLA, M., LIMPOSUWAN, K., WOODS, D. E., SIRISINHA, S. & UTAISINCHAROEN, P. 2007. Fate of a Burkholderia pseudomallei lipopolysaccharide mutant in the mouse macrophage cell line RAW 264.7: possible role for the O-antigenic polysaccharide moiety of lipopolysaccharide in internalization and intracellular survival. *Infect Immun,* 75**,** 4298-304.

ASTRONOMO, R. D. & BURTON, D. R. 2010. Carbohydrate vaccines: developing sweet solutions to sticky situations? *Nat Rev Drug Discov,* 9**,** 308-24.

ATKINS, T., PRIOR, R. G., MACK, K., RUSSELL, P., NELSON, M., OYSTON, P. C., DOUGAN, G. & TITBALL, R. W. 2002. A mutant of Burkholderia pseudomallei, auxotrophic in the branched chain amino acid biosynthetic pathway, is attenuated and protective in a murine model of melioidosis. *Infect Immun,* 70**,** 5290-4.

AVERY, O. T. & GOEBEL, W. F. 1931. Chemo-Immunological Studies on Conjugated Carbohydrate-Proteins : V. The Immunological Specifity of an Antigen Prepared by Combining the Capsular Polysaccharide of Type Iii Pneumococcus with Foreign Protein. *J Exp Med,* 54**,** 437-47.

BAINBRIDGE, B. W., KARIMI-NASER, L., REIFE, R., BLETHEN, F., ERNST, R. K. & DARVEAU, R. P. 2008. Acyl chain specificity of the acyltransferases LpxA and LpxD and substrate availability contribute to lipid A fatty acid heterogeneity in Porphyromonas gingivalis. *J Bacteriol,* 190**,** 4549-58.

BARBER, M., BORDOLI, R. S., GARNER, G. V., GORDON, D. B., SEDGWICK, R. D., TETLER, L. W. & TYLER, A. N. 1981. Fast-atom-bombardment mass spectra of enkephalins. *Biochem J,* 197**,** 401-4.

BARNES, J. L. & KETHEESAN, N. 2007. Development of protective immunity in a murine model of melioidosis is influenced by the source of Burkholderia pseudomallei antigens. *Immunol Cell Biol,* 85**,** 551-7.

BARNES, J. L., ULETT, G. C., KETHEESAN, N., CLAIR, T., SUMMERS, P. M. & HIRST, R. G. 2001. Induction of multiple chemokine and colony-stimulating factor genes in experimental Burkholderia pseudomallei infection. *Immunol Cell Biol,* 79**,** 490-501.

BARNES, J. L., WILLIAMS, N. L. & KETHEESAN, N. 2008. Susceptibility to Burkholderia pseudomallei is associated with host immune responses involving tumor necrosis factor receptor-1 (TNFR1) and TNF receptor-2 (TNFR2). *FEMS Immunol Med Microbiol,* 52**,** 379-88.

BECKEY, H. D. 1969. Field desorption mass spectrometry: A technique for the study of thermally unstable substances of low volatility. *International Journal of Mass Spectrometry and Ion Physics,* 2**,** 500 - 502.

BERNFIELD, M., GOTTE, M., PARK, P. W., REIZES, O., FITZGERALD, M. L., LINCECUM, J. & ZAKO, M. 1999. Functions of cell surface heparan sulfate proteoglycans. *Annu Rev Biochem,* 68**,** 729-77.

BIEMANN, K. 1988. Contributions of mass spectrometry to peptide and protein structure. *Biomed Environ Mass Spectrom,* 16**,** 99-111.

BLAUROCK, A. E., STOECKENIUS, W., OESTERHELT, D. & SCHERFHOF, G. L. 1976. Structure of the cell envelope of Halobacterium halobium. *J Cell Biol,* 71**,** 1-22.

BONDI, S. K. & GOLDBERG, J. B. 2008. Strategies toward vaccines against Burkholderia mallei and Burkholderia pseudomallei. *Expert Rev Vaccines,* 7**,** 1357-65.

BOS, M. P., TEFSEN, B., GEURTSEN, J. & TOMMASSEN, J. 2004. Identification of an outer membrane protein required for the transport of lipopolysaccharide to the bacterial cell surface. *Proc Natl Acad Sci U S A,* 101**,** 9417-22.

BOUCHER, Y., KAMEKURA, M. & DOOLITTLE, W. F. 2004. Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Mol Microbiol,* 52**,** 515-27.

BREITBACH, K., KLOCKE, S., TSCHERNIG, T., VAN ROOIJEN, N., BAUMANN, U. & STEINMETZ, I. 2006. Role of inducible nitric oxide synthase and NADPH oxidase in early control of Burkholderia pseudomallei infection in mice. *Infect Immun,* 74**,** 6300-9.

BRETT, P. J., BURTNICK, M. N., HEISS, C., AZADI, P., DESHAZER, D., WOODS, D. E. & GHERARDINI, F. C. 2011. Burkholderia thailandensis oacA mutants facilitate the expression of Burkholderia mallei-like O polysaccharides. *Infect Immun,* 79**,** 961-9.

BRETT, P. J., BURTNICK, M. N. & WOODS, D. E. 2003. The wbiA locus is required for the 2-O-acetylation of lipopolysaccharides expressed by Burkholderia pseudomallei and Burkholderia thailandensis. *FEMS Microbiol Lett,* 218**,** 323-8.

BRETT, P. J., DESHAZER, D. & WOODS, D. E. 1997. Characterization of Burkholderia pseudomallei and Burkholderia pseudomallei-like strains. *Epidemiol Infect,* 118**,** 137-48.

BRETT, P. J., DESHAZER, D. & WOODS, D. E. 1998. Burkholderia thailandensis sp. nov., a Burkholderia pseudomallei-like species. *Int J Syst Bacteriol,* 48 Pt 1**,** 317-20.

BRETT, P. J., MAH, D. C. & WOODS, D. E. 1994. Isolation and characterization of Pseudomonas pseudomallei flagellin proteins. *Infect Immun,* 62**,** 1914-9.

BRETT, P. J. & WOODS, D. E. 1996. Structural and immunological characterization of Burkholderia pseudomallei O-polysaccharide-flagellin protein conjugates. *Infect Immun,* 64**,** 2824-8.

BROZEK, K. A., HOSAKA, K., ROBERTSON, A. D. & RAETZ, C. R. 1989. Biosynthesis of lipopolysaccharide in Escherichia coli. Cytoplasmic enzymes that attach 3-deoxy-D-manno-octulosonic acid to lipid A. *J Biol Chem,* 264**,** 6956-66.

BROZEK, K. A. & RAETZ, C. R. 1990. Biosynthesis of lipid A in Escherichia coli. Acyl carrier protein-dependent incorporation of laurate and myristate. *J Biol Chem,* 265**,** 15410-7.

BRYAN, L. E., WONG, S., WOODS, D. E., DANCE, D. A. & CHAOWAGUL, W. 1994. Passive protection of diabetic rats with antisera specific for the polysaccharide portion of the lipopolysaccharide isolated from Pseudomonas pseudomallei. *Can J Infect Dis,* 5**,** 170-8.

BUETOW, L., SMITH, T. K., DAWSON, A., FYFFE, S. & HUNTER, W. N. 2007. Structure and reactivity of LpxD, the N-acyltransferase of lipid A biosynthesis. *Proc Natl Acad Sci U S A,* 104**,** 4321-6.

BURDA, P. & AEBI, M. 1999. The dolichol pathway of N-linked glycosylation. *Biochim Biophys Acta,* 1426**,** 239-57.

BURTNICK, M. N., BRETT, P. J. & WOODS, D. E. 2002. Molecular and physical characterization of Burkholderia mallei O antigens. *J Bacteriol,* 184**,** 849-52.

BURTNICK, M. N. & WOODS, D. E. 1999. Isolation of polymyxin B-susceptible mutants of Burkholderia pseudomallei and molecular characterization of genetic loci involved in polymyxin B resistance. *Antimicrob Agents Chemother,* 43**,** 2648-56.

CALO, D., EILAM, Y., LICHTENSTEIN, R. G. & EICHLER, J. 2010a. Towards glycoengineering in archaea: replacement of Haloferax volcanii AglD with homologous glycosyltransferases from other halophilic archaea. *Appl Environ Microbiol,* 76**,** 5684-92.

CALO, D., GUAN, Z., NAPARSTEK, S. & EICHLER, J. 2011. Different routes to the same ending: comparing the N-glycosylation processes of Haloferax volcanii and Haloarcula marismortui, two halophilic archaea from the Dead Sea. *Mol Microbiol,* 81**,** 1166-77.

CALO, D., KAMINSKI, L. & EICHLER, J. 2010b. Protein glycosylation in Archaea: sweet and extreme. *Glycobiology,* 20**,** 1065-76.

CASTRIC, P., CASSELS, F. J. & CARLSON, R. W. 2001. Structural characterization of the Pseudomonas aeruginosa 1244 pilin glycan. *J Biol Chem,* 276**,** 26479-85.

CHABAN, B., LOGAN, S. M., KELLY, J. F. & JARRELL, K. F. 2009. AglC and AglK are involved in biosynthesis and attachment of diacetylated glucuronic acid to the N-glycan in Methanococcus voltae. *J Bacteriol,* 191**,** 187-95.

CHABAN, B., VOISIN, S., KELLY, J., LOGAN, S. M. & JARRELL, K. F. 2006. Identification of genes involved in the biosynthesis and attachment of Methanococcus voltae N-linked glycans: insight into N-linked glycosylation pathways in Archaea. *Mol Microbiol,* 61**,** 259-68.

CHAOWAGUL, W., WHITE, N. J., DANCE, D. A., WATTANAGOON, Y., NAIGOWIT, P., DAVIS, T. M., LOOAREESUWAN, S. & PITAKWATCHARA, N. 1989. Melioidosis: a major cause of community-acquired septicemia in northeastern Thailand. *J Infect Dis,* 159**,** 890-9.

CHART, H. 1994. *Methods in Practical Laboratory Bacteriology*, CRC Press, London, Tokyo.

CHEN, M. M., GLOVER, K. J. & IMPERIALI, B. 2007. From peptide to protein: comparative analysis of the substrate specificity of N-linked glycosylation in C. jejuni. *Biochemistry,* 46**,** 5579-85.

CHEN, Y. S., HSIAO, Y. S., LIN, H. H., YEN, C. M., CHEN, S. C. & CHEN, Y. L. 2006. Immunogenicity and anti-Burkholderia pseudomallei activity in Balb/c mice immunized with plasmid DNA encoding flagellin. *Vaccine,* 24**,** 750-8.

CHENG, A. C. & CURRIE, B. J. 2005. Melioidosis: epidemiology, pathophysiology, and management. *Clin Microbiol Rev,* 18**,** 383-416.

CHENG, A. C., GODOY, D., MAYO, M., GAL, D., SPRATT, B. G. & CURRIE, B. J. 2004. Isolates of Burkholderia pseudomallei from Northern Australia are distinct by multilocus sequence typing, but strain types do not correlate with clinical presentation. *J Clin Microbiol,* 42**,** 5477-83.

CHOU, M. W. & YANG, S. K. 1979. Combined reversed-phase and normal-phase high-performance liquid chromatography in the purification and identification of 7,12-dimethylbenz[a]anthracene metabolites. *J Chromatogr,* 185**,** 635-54.

CHUAYGUD, T., TUNGPRADABKUL, S., SIRISINHA, S., CHUA, K. L. & UTAISINCHAROEN, P. 2008. A role of Burkholderia pseudomallei flagella as a virulent factor. *Trans R Soc Trop Med Hyg,* 102 Suppl 1**,** S140-4.

CICCARELLI, F. D., DOERKS, T., VON MERING, C., CREEVEY, C. J., SNEL, B. & BORK, P. 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science,* 311**,** 1283-7.

CLAUS, H., AKCA, E., DEBAERDEMAEKER, T., EVRARD, C., DECLERCQ, J. P. & KONIG, H. 2002. Primary structure of selected archaeal mesophilic and extremely thermophilic outer surface layer proteins. *Syst Appl Microbiol,* 25**,** 3-12.

COENYE, T. & VANDAMME, P. 2007. *Burkholderia - Molecular Microbiology and Genomics,* Norfolk, Horizon Bioscience.

COHEN-ROSENZWEIG, C., YURIST-DOUTSCH, S. & EICHLER, J. 2012. AglS, a novel component of the Haloferax volcanii N-glycosylation pathway, is a dolichol phosphate-mannose mannosyltransferase. *J Bacteriol,* 194**,** 6909-16.

CROWELL, D. N., REZNIKOFF, W. S. & RAETZ, C. R. 1987. Nucleotide sequence of the Escherichia coli gene for lipid A disaccharide synthase. *J Bacteriol,* 169**,** 5727-34.

CURRIE, B., SMITH-VAUGHAN, H., GOLLEDGE, C., BULLER, N., SRIPRAKASH, K. S. & KEMP, D. J. 1994. Pseudomonas pseudomallei isolates collected over 25 years from a non-tropical endemic focus show clonality on the basis of ribotyping. *Epidemiol Infect,* 113**,** 307-12.

CURRIE, B. J. 2008. Advances and remaining uncertainties in the epidemiology of Burkholderia pseudomallei and melioidosis. *Trans R Soc Trop Med Hyg,* 102**,** 225-7.

CURRIE, B. J., FISHER, D. A., ANSTEY, N. M. & JACUPS, S. P. 2000a. Melioidosis: acute and chronic disease, relapse and re-activation. *Trans R Soc Trop Med Hyg,* 94**,** 301-4.

CURRIE, B. J., FISHER, D. A., HOWARD, D. M., BURROW, J. N., LO, D., SELVA-NAYAGAM, S., ANSTEY, N. M., HUFFAM, S. E., SNELLING, P. L., MARKS, P. J., STEPHENS, D. P., LUM, G. D., JACUPS, S. P. & KRAUSE, V. L. 2000b. Endemic melioidosis in tropical northern Australia: a 10-year prospective study and review of the literature. *Clin Infect Dis,* 31**,** 981-6.

CURRIE, B. J., FISHER, D. A., HOWARD, D. M., BURROW, J. N., SELVANAYAGAM, S., SNELLING, P. L., ANSTEY, N. M. & MAYO, M. J. 2000c. The epidemiology of melioidosis in Australia and Papua New Guinea. *Acta Trop,* 74**,** 121-7.

CURRIE, B. J. & JACUPS, S. P. 2003. Intensity of rainfall and severity of melioidosis, Australia. *Emerg Infect Dis,* 9**,** 1538-42.

CURRIE, B. J., JACUPS, S. P., CHENG, A. C., FISHER, D. A., ANSTEY, N. M., HUFFAM, S. E. & KRAUSE, V. L. 2004. Melioidosis epidemiology and risk factors from a prospective whole-population study in northern Australia. *Trop Med Int Health,* 9**,** 1167-74.

CURRIE, B. J., MAYO, M., ANSTEY, N. M., DONOHOE, P., HAASE, A. & KEMP, D. J. 2001. A cluster of melioidosis cases from an endemic region is clonal and is linked to the water supply using molecular typing of Burkholderia pseudomallei isolates. *Am J Trop Med Hyg,* 65**,** 177-9.

CZERWICKA, M., MARSZEWSKA, K., BYCHOWSKA, A., DZIADZIUSZKO, H., BRZOZOWSKI, K., LOJKOWSKA, E., STEPNOWSKI, P. & KACZYNSKI, Z. 2011. Chemical structure of the O-polysaccharide isolated from Pectobacterium atrosepticum SCRI 1039. *Carbohydr Res,* 346**,** 2978-81.

DANCE, D. A. 1991. Melioidosis: the tip of the iceberg? *Clin Microbiol Rev,* 4**,** 52-60.

DANCE, D. A. 2000. Melioidosis as an emerging global problem. *Acta Trop,* 74**,** 115-9.

DEJSIRILERT, S., KONDO, E., CHIEWSILP, D. & KANAI, K. 1991. Growth and survival of Pseudomonas pseudomallei in acidic environments. *Jpn J Med Sci Biol,* 44**,** 63-74.

DELL, A. 1987. F.A.B.-mass spectrometry of carbohydrates. *Adv Carbohydr Chem Biochem,* 45**,** 19-72.

DELL, A. 1990. Preparation and desorption mass spectrometry of permethyl and peracetyl derivatives of oligosaccharides. *Methods Enzymol,* 193**,** 647-60.

DELL, A. & BALLOU, C. E. 1983. Fast-atom-bombardment, negative-ion mass spectrometry of the mycobacterial O-methyl-D-glucose polysaccharide and lipopolysaccharides. *Carbohydr Res,* 120**,** 95-111.

DELL, A., GALADARI, A., SASTRE, F. & HITCHEN, P. 2010. Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *Int J Microbiol,* 2010**,** 148178.

DELL, A., KHOO, K., PANICO, M., MCDOWELL, R. A., ETIENNE, A. T., REASON, A. J. & MORRIS, H. R. 1993. FAB-MS and ES-MS of glycoproteins. *In:* FUKUDA, M. & KOBATA, A. (eds.) *Glycobiology - A practical Approach.* Oxford: Oxford University Press.

DELL, A. & MORRIS, H. R. 2001. Glycoprotein structure determination by mass spectrometry. *Science,* 291**,** 2351-6.

DELL, A., MORRIS, H. R., EASTON, R. L., HASLAM, S. M., PANICO, M., SUTTON-SMITH, M., REASON, A. J. & KHOO, K. 2000. Structural analysis of oligosaccharides: FAB-MS, ES-MS, MALDI-MS. *In:* ERNST, B., HART, G. & SINAY, P. (eds.) *Carbohydrates in Chemistry and Biology.* Weinheim: Wiley-VCH Verlag GmbH, Weinheim, Germany.

DESHAZER, D., BRETT, P. J., CARLYON, R. & WOODS, D. E. 1997. Mutagenesis of Burkholderia pseudomallei with Tn5-OT182: isolation of motility mutants and molecular characterization of the flagellin structural gene. *J Bacteriol,* 179**,** 2116-25.

DESHAZER, D., BRETT, P. J. & WOODS, D. E. 1998. The type II O-antigenic polysaccharide moiety of Burkholderia pseudomallei lipopolysaccharide is required for serum resistance and virulence. *Mol Microbiol,* 30**,** 1081-100.

DESMARCHELIER, P. M., DANCE, D. A., CHAOWAGUL, W., SUPUTTAMONGKOL, Y., WHITE, N. J. & PITT, T. L. 1993. Relationships among Pseudomonas pseudomallei isolates from patients with recurrent melioidosis. *J Clin Microbiol,* 31**,** 1592-6.

DOERRLER, W. T. 2006. Lipid trafficking to the outer membrane of Gram-negative bacteria. *Mol Microbiol,* 60**,** 542-52.

DOERRLER, W. T., GIBBONS, H. S. & RAETZ, C. R. 2004. MsbA-dependent translocation of lipids across the inner membrane of Escherichia coli. *J Biol Chem,* 279**,** 45102-9.

DOIG, P., KINSELLA, N., GUERRY, P. & TRUST, T. J. 1996. Characterization of a post-translational modification of Campylobacter flagellin: identification of a sero-specific glycosyl moiety. *Mol Microbiol,* 19**,** 379-87.

DOMON, B. & COSTELLO, C. E. 1988. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra fo glycoconjugates. *Glycoconj J,* 5**,** 397-409.

EASTON, A., HAQUE, A., CHU, K., LUKASZEWSKI, R. & BANCROFT, G. J. 2007. A critical role for neutrophils in resistance to experimental infection with Burkholderia pseudomallei. *J Infect Dis,* 195**,** 99-107.

EGGE, H., PETER-KATALINIC, J., PAZ-PARENTE, J., STRECKER, G., MONTREUIL, J. & FOURNET, B. 1983. Carbohydrate structures of hen ovomucoid. A mass spectrometric analysis. *FEBS Lett,* 156**,** 357-62.

EICHLER, J. 2000. Novel glycoproteins of the halophilic archaeon Haloferax volcanii. *Arch Microbiol,* 173**,** 445-8.

EICHLER, J. 2001. Post-translational modification of the S-layer glycoprotein occurs following translocation across the plasma membrane of the haloarchaeon Haloferax volcanii. *Eur J Biochem,* 268**,** 4366-73.

EICHLER, J. 2003. Facing extremes: archaeal surface-layer (glyco)proteins. *Microbiology,* 149**,** 3347-51.

EICHLER, J. & ADAMS, M. W. 2005. Posttranslational protein modification in Archaea. *Microbiol Mol Biol Rev,* 69**,** 393-425.

ELHAMMER, A. P., POORMAN, R. A., BROWN, E., MAGGIORA, L. L., HOOGERHEIDE, J. G. & KEZDY, F. J. 1993. The specificity of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase as inferred from a database of in vivo substrates and from the in vitro glycosylation of proteins and peptides. *J Biol Chem,* 268**,** 10029-38.

ELVIN, S. J., HEALEY, G. D., WESTWOOD, A., KNIGHT, S. C., EYLES, J. E. & WILLIAMSON, E. D. 2006. Protection against heterologous Burkholderia pseudomallei strains by dendritic cell immunization. *Infect Immun,* 74**,** 1706-11.

ESSEX-LOPRESTI, A. E., BODDEY, J. A., THOMAS, R., SMITH, M. P., HARTLEY, M. G., ATKINS, T., BROWN, N. F., TSANG, C. H., PEAK, I. R., HILL, J., BEACHAM, I. R. & TITBALL, R. W. 2005. A type IV pilin, PilA, Contributes To Adherence of Burkholderia pseudomallei and virulence in vivo. *Infect Immun,* 73**,** 1260-4.

FARIDMOAYER, A., FENTABIL, M. A., MILLS, D. C., KLASSEN, J. S. & FELDMAN, M. F. 2007. Functional characterization of bacterial oligosaccharyltransferases involved in O-linked protein glycosylation. *J Bacteriol,* 189**,** 8088-98.

FELDMAN, M. F., MAROLDA, C. L., MONTEIRO, M. A., PERRY, M. B., PARODI, A. J. & VALVANO, M. A. 1999. The activity of a putative polyisoprenol-linked sugar translocase (Wzx) involved in Escherichia coli O antigen assembly is independent of the chemical structure of the O repeat. *J Biol Chem,* 274**,** 35129-38.

FENN, J. B., MANN, M., MENG, C. K., WONG, S. F. & WHITEHOUSE, C. M. 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science,* 246**,** 64-71.

GAL, D., MAYO, M., SMITH-VAUGHAN, H., DASARI, P., MCKINNON, M., JACUPS, S. P., URQUHART, A. I., HASSELL, M. & CURRIE, B. J. 2004. Contamination of hand wash detergent linked to occupationally acquired melioidosis. *Am J Trop Med Hyg,* 71**,** 360-2.

GALANOS, C., LUDERITZ, O., RIETSCHEL, E. T., WESTPHAL, O., BRADE, H., BRADE, L., FREUDENBERG, M., SCHADE, U., IMOTO, M., YOSHIMURA, H. & ET AL. 1985. Synthetic and natural Escherichia coli free lipid A express identical endotoxic activities. *Eur J Biochem,* 148**,** 1-5.

GAN, Y. H. 2005. Interaction between Burkholderia pseudomallei and the host immune response: sleeping with the enemy? *J Infect Dis,* 192**,** 1845-50.

GARRETT, T. A., QUE, N. L. & RAETZ, C. R. 1998. Accumulation of a lipid A precursor lacking the 4'-phosphate following inactivation of the Escherichia coli lpxK gene. *J Biol Chem,* 273**,** 12457-65.

GAUR, D., GALBRAITH, L. & WILKINSON, S. G. 1998. Structural characterisation of a rhamnan and a fucorhamnan, both present in the lipopolysaccharide of Burkholderia vietnamiensis strain LMG 10926. *Eur J Biochem,* 258**,** 696-701.

GIBBONS, H. S., KALB, S. R., COTTER, R. J. & RAETZ, C. R. 2005. Role of Mg2+ and pH in the modification of Salmonella lipid A after endocytosis by macrophage tumour cells. *Mol Microbiol,* 55**,** 425-40.

GIRRBACH, V., ZELLER, T., PRIESMEIER, M. & STRAHL-BOLSINGER, S. 2000. Structure-function analysis of the dolichyl phosphate-mannose: protein O-mannosyltransferase ScPmt1p. *J Biol Chem,* 275**,** 19288-96.

GLOVER, K. J., WEERAPANA, E., NUMAO, S. & IMPERIALI, B. 2005. Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from Campylobacter jejuni. *Chem Biol,* 12**,** 1311-5.

GODOY, D., RANDLE, G., SIMPSON, A. J., AANENSEN, D. M., PITT, T. L., KINOSHITA, R. & SPRATT, B. G. 2003. Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, Burkholderia pseudomallei and Burkholderia mallei. *J Clin Microbiol,* 41**,** 2068-79.

GREER, F. M. & MORRIS, H. R. 1997. Fast-atom bombardment and electrospray mass spectrometry of peptides, proteins, and glycoproteins. *Methods Mol Biol,* 64**,** 147-63.

GROBE, N., WEIR, N. M., LEIVA, O., ONG, F. S., BERNSTEIN, K. E., SCHMAIER, A. H., MORRIS, M. & ELASED, K. M. 2013. Identification of prolyl carboxypeptidase as an alternative enzyme for processing of renal angiotensin II using mass spectrometry. *Am J Physiol Cell Physiol*.

GUAN, Z. & EICHLER, J. 2011. Liquid chromatography/tandem mass spectrometry of dolichols and polyprenols, lipid sugar carriers across evolution. *Biochim Biophys Acta,* 1811**,** 800-6.

GUAN, Z., NAPARSTEK, S., CALO, D. & EICHLER, J. 2012. Protein glycosylation as an adaptive response in Archaea: growth at different salt concentrations leads to alterations in Haloferax volcanii S-layer glycoprotein N-glycosylation. *Environ Microbiol,* 14**,** 743-53.

GUAN, Z., NAPARSTEK, S., KAMINSKI, L., KONRAD, Z. & EICHLER, J. 2010. Distinct glycan-charged phosphodolichol carriers are required for the assembly of the pentasaccharide N-linked to the Haloferax volcanii S-layer glycoprotein. *Mol Microbiol,* 78**,** 1294-303.

GUERRY, P. 2007. Campylobacter flagella: not just for motility. *Trends Microbiol,* 15**,** 456-61.

GUERRY, P. 2011. N-linked glycosylation in Archaea: two paths to the same glycan. *Mol Microbiol,* 81**,** 1133-5.

GUO, L., LIM, K. B., GUNN, J. S., BAINBRIDGE, B., DARVEAU, R. P., HACKETT, M. & MILLER, S. I. 1997. Regulation of lipid A modifications by Salmonella typhimurium virulence genes phoP-phoQ. *Science,* 276**,** 250-3.

GUO, L., LIM, K. B., PODUJE, C. M., DANIEL, M., GUNN, J. S., HACKETT, M. & MILLER, S. I. 1998. Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. *Cell,* 95**,** 189-98.

GUO, Z. & BOONS, G.-J. 2009. Carbohydrate-Based Vaccines and Immunotherapies. *In:* WANG, B. (ed.) *Wiley series in drug discovery and development.* Hoboken, N.J.: Wiley.

HAFT, D. H., PAYNE, S. H. & SELENGUT, J. D. 2012. Archaeosortases and exosortases are widely distributed systems linking membrane transit with posttranslational modification. *J Bacteriol,* 194**,** 36-48.

HAKANSSON, K., COOPER, H. J., EMMETT, M. R., COSTELLO, C. E., MARSHALL, A. G. & NILSSON, C. L. 2001. Electron capture dissociation and infrared multiphoton dissociation MS/MS of an N-glycosylated tryptic peptic to yield complementary sequence information. *Anal Chem,* 73**,** 4530-6.

HAKOMORI, S. 1964. A Rapid Permethylation of Glycolipid, and Polysaccharide Catalyzed by Methylsulfinyl Carbanion in Dimethyl Sulfoxide. *J Biochem,* 55**,** 205-8.

HAQUE, A., CHU, K., EASTON, A., STEVENS, M. P., GALYOV, E. E., ATKINS, T., TITBALL, R. & BANCROFT, G. J. 2006a. A live experimental vaccine against Burkholderia pseudomallei elicits CD4+ T cell-mediated immunity, priming T cells specific for 2 type III secretion system proteins. *J Infect Dis,* 194**,** 1241-8.

HAQUE, A., EASTON, A., SMITH, D., O'GARRA, A., VAN ROOIJEN, N., LERTMEMONGKOLCHAI, G., TITBALL, R. W. & BANCROFT, G. J. 2006b. Role of T cells in innate and adaptive immunity against murine Burkholderia pseudomallei infection. *J Infect Dis,* 193**,** 370-9.

HARA, Y., MOHAMED, R. & NATHAN, S. 2009. Immunogenic Burkholderia pseudomallei outer membrane proteins as potential candidate vaccine targets. *PLoS One,* 4**,** e6496.

HARD, K. & VLIEGENTHART, J. F. G. 1993. Nuclear magnetic reasonance spectroscopy of glycoprotein-derived carbohydrate chains. *In:* FUKUDA, M. & KOBATA, A. (eds.) *Glycobiology - A practical Approach.* Oxford: Oxford University Press.

HARLAND, D. N., CHU, K., HAQUE, A., NELSON, M., WALKER, N. J., SARKAR-TYSON, M., ATKINS, T. P., MOORE, B., BROWN, K. A., BANCROFT, G., TITBALL, R. W. & ATKINS, H. S. 2007. Identification of a LolC homologue in Burkholderia pseudomallei, a novel protective antigen for melioidosis. *Infect Immun,* 75**,** 4173-80.

HARVEY, D. J. 1999. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass Spectrom Rev,* 18**,** 349-450.

HARVEY, D. J. 2006. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update covering the period 1999-2000. *Mass Spectrom Rev,* 25**,** 595-662.

HARVEY, D. J. 2010. Carbohydrate Analysis by ESI and MALDI. *In:* COLE, R. B. (ed.) *Electrospray and MALDI Mass Spectrometry.* Second Edition ed. Hoboken, New Jersey, USA: John Wiley & Sons.

HASLAM, S. M., KHOO, K. H., HOUSTON, K. M., HARNETT, W., MORRIS, H. R. & DELL, A. 1997. Characterisation of the phosphorylcholine-containing N-linked oligosaccharides in the excretory-secretory 62 kDa glycoprotein of Acanthocheilonema viteae. *Mol Biochem Parasitol,* 85**,** 53-66.

HEALEY, G. D., ELVIN, S. J., MORTON, M. & WILLIAMSON, E. D. 2005. Humoral and cell-mediated adaptive immune responses are required for protection against Burkholderia pseudomallei challenge and bacterial clearance postinfection. *Infect Immun,* 73**,** 5945-51.

HEGGE, F. T., HITCHEN, P. G., AAS, F. E., KRISTIANSEN, H., LOVOLD, C., EGGE-JACOBSEN, W., PANICO, M., LEONG, W. Y., BULL, V., VIRJI, M., MORRIS, H. R., DELL, A. & KOOMEY, M. 2004. Unique modifications with phosphocholine and phosphoethanolamine define alternate antigenic forms of Neisseria gonorrhoeae type IV pili. *Proc Natl Acad Sci U S A,* 101**,** 10798-803.

HEIDELBERGER, M. & AVERY, O. T. 1923. The Soluble Specific Substance of Pneumococcus. *J Exp Med,* 38**,** 73-9.

HEIDELBERGER, M., DILAPI, M. M., SIEGEL, M. & WALTER, A. W. 1950. Persistence of antibodies in human subjects injected with pneumococcal polysaccharides. *J Immunol,* 65**,** 535-41.

HELENIUS, A. & AEBI, M. 2001. Intracellular functions of N-linked glycans. *Science,* 291**,** 2364-9.

HELENIUS, J., NG, D. T., MAROLDA, C. L., WALTER, P., VALVANO, M. A. & AEBI, M. 2002. Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature,* 415**,** 447-50.

HELLERQVIST, C. G. 1990. Linkage analysis using Lindberg method. *Methods Enzymol,* 193**,** 554-73.

HENDRIXSON, D. R. & DIRITA, V. J. 2004. Identification of Campylobacter jejuni genes involved in commensal colonization of the chick gastrointestinal tract. *Mol Microbiol,* 52**,** 471-84.

HETTMANN, T., SCHMIDT, C. L., ANEMULLER, S., ZAHRINGER, U., MOLL, H., PETERSEN, A. & SCHAFER, G. 1998. Cytochrome b558/566 from the archaeon Sulfolobus acidocaldarius. A novel highly glycosylated, membrane-bound b-type hemoprotein. *J Biol Chem,* 273**,** 12032-40.

HITCHEN, P. G. & DELL, A. 2006. Bacterial glycoproteomics. *Microbiology,* 152**,** 1575-80.

HO, M., SCHOLLAARDT, T., SMITH, M. D., PERRY, M. B., BRETT, P. J., CHAOWAGUL, W. & BRYAN, L. E. 1997. Specificity and functional activity of anti-Burkholderia pseudomallei polysaccharide antibodies. *Infect Immun,* 65**,** 3648-53.

HOFFMANN, D., IINUMA, Y. & HERRMANN, H. 2007. Development of a method for fast analysis of phenolic molecular markers in biomass burning particles using high performance liquid chromatography/atmospheric pressure chemical ionisation mass spectrometry. *J Chromatogr A,* 1143**,** 168-75.

HOLDEN, M. T., TITBALL, R. W., PEACOCK, S. J., CERDENO-TARRAGA, A. M., ATKINS, T., CROSSMAN, L. C., PITT, T., CHURCHER, C., MUNGALL, K., BENTLEY, S. D., SEBAIHIA, M., THOMSON, N. R., BASON, N., BEACHAM, I. R., BROOKS, K., BROWN, K. A., BROWN, N. F., CHALLIS, G. L., CHEREVACH, I., CHILLINGWORTH, T., CRONIN, A., CROSSETT, B., DAVIS, P., DESHAZER, D., FELTWELL, T., FRASER, A., HANCE, Z., HAUSER, H., HOLROYD, S., JAGELS, K., KEITH, K. E., MADDISON, M., MOULE, S., PRICE, C., QUAIL, M. A., RABBINOWITSCH, E., RUTHERFORD, K., SANDERS, M., SIMMONDS, M., SONGSIVILAI, S., STEVENS, K., TUMAPA, S., VESARATCHAVEST, M., WHITEHEAD, S., YEATS, C., BARRELL, B. G., OYSTON, P. C. & PARKHILL, J. 2004. Genomic plasticity of the causative agent of melioidosis, Burkholderia pseudomallei. *Proc Natl Acad Sci U S A,* 101**,** 14240-5.

HOPPE, I., BRENNEKE, B., ROHDE, M., KREFT, A., HAUSSLER, S., REGANZEROWSKI, A. & STEINMETZ, I. 1999. Characterization of a murine model of melioidosis: comparison of different strains of mice. *Infect Immun,* 67**,** 2891-900.

HOUWINK, A. L. 1956. Flagella, gas vacuoles and cell-wall structure in Halobacterium halobium; an electron microscope study. *J Gen Microbiol,* 15**,** 146-50.

HOWE, C., SAMPATH, A. & SPOTNITZ, M. 1971. The pseudomallei group: a review. *J Infect Dis,* 124**,** 598-606.

ISSHIKI, Y., KAWAHARA, K. & ZAHRINGER, U. 1998. Isolation and characterisation of disodium (4-amino-4-deoxy-beta-L- arabinopyranosyl)-(1-->8)-(D-glycero-alpha-D-talo-oct-2-ulopyranosylona te)- (2-->4)-(methyl 3-deoxy-D-manno-oct-2-ulopyranosid)onate from the lipopolysaccharide of Burkholderia cepacia. *Carbohydr Res,* 313**,** 21-7.

JARRELL, K. F., JONES, G. M., KANDIBA, L., NAIR, D. B. & EICHLER, J. 2010. S-layer glycoproteins and flagellins: reporters of archaeal posttranslational modifications. *Archaea,* 2010.

JEDDELOH, J. A., FRITZ, D. L., WAAG, D. M., HARTINGS, J. M. & ANDREWS, G. P. 2003. Biodefense-driven murine model of pneumonic melioidosis. *Infect Immun,* 71**,** 584-7.

JENNINGS, H. J. 1983. Capsular polysaccharides as human vaccines. *Adv Carbohydr Chem Biochem,* 41**,** 155-208.

JONES, A. L., BEVERIDGE, T. J. & WOODS, D. E. 1996. Intracellular survival of Burkholderia pseudomallei. *Infect Immun,* 64**,** 782-90.

JONES, C. 2005. Vaccines based on the cell surface carbohydrates of pathogenic bacteria. *An Acad Bras Cienc,* 77**,** 293-324.

JONES, M. A., MARSTON, K. L., WOODALL, C. A., MASKELL, D. J., LINTON, D., KARLYSHEV, A. V., DORRELL, N., WREN, B. W. & BARROW, P. A. 2004. Adaptation of Campylobacter jejuni NCTC11168 to high-level colonization of the avian gastrointestinal tract. *Infect Immun,* 72**,** 3769-76.

KALMOKOFF, M. L., KOVAL, S. F. & JARRELL, K. F. 1992. Relatedness of the flagellins from methanogens. *Arch Microbiol,* 157**,** 481-7.

KAMEKURA, M. & KATES, M. 1999. Structural diversity of membrane lipids in members of Halobacteriaceae. *Biosci Biotechnol Biochem,* 63**,** 969-72.

KAMINSKI, L., ABU-QARN, M., GUAN, Z., NAPARSTEK, S., VENTURA, V. V., RAETZ, C. R., HITCHEN, P. G., DELL, A. & EICHLER, J. AglJ adds the first sugar of the N-linked pentasaccharide decorating the Haloferax volcanii S-layer glycoprotein. *J Bacteriol,* 192**,** 5572-9.

KAMINSKI, L., ABU-QARN, M., GUAN, Z., NAPARSTEK, S., VENTURA, V. V., RAETZ, C. R., HITCHEN, P. G., DELL, A. & EICHLER, J. 2010. AglJ adds the first sugar of the N-linked pentasaccharide decorating the Haloferax volcanii S-layer glycoprotein. *J Bacteriol,* 192**,** 5572-9.

KAMINSKI, L. & EICHLER, J. 2010. Identification of residues important for the activity of Haloferax volcanii AglD, a component of the archaeal N-glycosylation pathway. *Archaea,* 2010**,** 315108.

KAMINSKI, L., GUAN, Z., ABU-QARN, M., KONRAD, Z. & EICHLER, J. 2012. AglR is required for addition of the final mannose residue of the N-linked glycan decorating the Haloferax volcanii S-layer glycoprotein. *Biochim Biophys Acta,* 1820**,** 1664-70.

KANDIBA, L., GUAN, Z. & EICHLER, J. 2013. Lipid modification gives rise to two distinct Haloferax volcanii S-layer glycoprotein populations. *Biochim Biophys Acta,* 1828**,** 938-43.

KARAS, M. & HILLENKAMP, F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem,* 60**,** 2299-301.

KARCHER, U., SCHRODER, H., HASLINGER, E., ALLMAIER, G., SCHREINER, R., WIELAND, F., HASELBECK, A. & KONIG, H. 1993. Primary structure of the heterosaccharide of the surface glycoprotein of Methanothermus fervidus. *J Biol Chem,* 268**,** 26821-6.

KATES, M. 1978. The phytanyl ether-linked polar lipids and isoprenoid neutral lipids of extremely halophilic bacteria. *Prog Chem Fats Other Lipids,* 15**,** 301-42.

KATES, M. 1992. Archaebacterial lipids: structure, biosynthesis and function. *Biochem Soc Symp,* 58**,** 51-72.

KATES, M. 1993. Biology of halophilic bacteria, Part II. Membrane lipids of extreme halophiles: biosynthesis, function and evolutionary significance. *Experientia,* 49**,** 1027-36.

KAWAHARA, K., BRADE, H., RIETSCHEL, E. T. & ZAHRINGER, U. 1987. Studies on the chemical structure of the core-lipid A region of the lipopolysaccharide of Acinetobacter calcoaceticus NCTC 10305. Detection of a new 2-octulosonic acid interlinking the core oligosaccharide and lipid A component. *Eur J Biochem,* 163**,** 489-95.

KAY, E., LESK, V. I., TAMADDONI-NEZHAD, A., HITCHEN, P. G., DELL, A., STERNBERG, M. J., MUGGLETON, S. & WREN, B. W. 2010. Systems analysis of bacterial glycomes. *Biochem Soc Trans,* 38**,** 1290-3.

KELLEHER, D. J. & GILMORE, R. 2006. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology,* 16**,** 47R-62R.

KELLY, J., JARRELL, H., MILLAR, L., TESSIER, L., FIORI, L. M., LAU, P. C., ALLAN, B. & SZYMANSKI, C. M. 2006. Biosynthesis of the N-linked glycan in Campylobacter jejuni and addition onto protein through block transfer. *J Bacteriol,* 188**,** 2427-34.

KELLY, J., LOGAN, S. M., JARRELL, K. F., VANDYKE, D. J. & VINOGRADOV, E. 2009. A novel N-linked flagellar glycan from Methanococcus maripaludis. *Carbohydr Res,* 344**,** 648-53.

KESPICHAYAWATTANA, W., INTACHOTE, P., UTAISINCHAROEN, P. & SIRISINHA, S. 2004. Virulent Burkholderia pseudomallei is more efficient than avirulent Burkholderia thailandensis in invasion of and adherence to cultured human epithelial cells. *Microb Pathog,* 36**,** 287-92.

KESPICHAYAWATTANA, W., RATTANACHETKUL, S., WANUN, T., UTAISINCHAROEN, P. & SIRISINHA, S. 2000. Burkholderia pseudomallei induces cell fusion and actin-associated membrane protrusion: a possible mechanism for cell-to-cell spreading. *Infect Immun,* 68**,** 5377-84.

KESSEL, M., WILDHABER, I., COHEN, S. & BAUMEISTER, W. 1988. Three-dimensional structure of the regular surface glycoprotein layer of Halobacterium volcanii from the Dead Sea. *EMBO J,* 7**,** 1549-54.

KIKUCHI, A., SAGAMI, H. & OGURA, K. 1999. Evidence for covalent attachment of diphytanylglyceryl phosphate to the cell-surface glycoprotein of Halobacterium halobium. *J Biol Chem,* 274**,** 18011-6.

KIM, S. K., DEVINE, L., ANGEVINE, M., DEMARS, R. & KAVATHAS, P. B. 2000. Direct detection and magnetic isolation of Chlamydia trachomatis major outer membrane protein-specific CD8+ CTLs with HLA class I tetramers. *J Immunol,* 165**,** 7285-92.

KIRK, R. G. & GINZBURG, M. 1972. Ultrastructure of two species of halobacterium. *J Ultrastruct Res,* 41**,** 80-94.

KNIREL, Y. A., PARAMONOV, N. A., SHASHKOV, A. S., KOCHETKOV, N. K., YARULLIN, R. G., FARBER, S. M. & EFREMENKO, V. I. 1992. Structure of the polysaccharide chains of Pseudomonas pseudomallei lipopolysaccharides. *Carbohydr Res,* 233**,** 185-93.

KNOCHENMUSS, R. 2006. Ion formation mechanisms in UV-MALDI. *Analyst,* 131**,** 966-86.

KOBAYASHI, T., NISHIZAKI, R. & IKEZAWA, H. 1997. The presence of GPI-linked protein(s) in an archaeobacterium, Sulfolobus acidocaldarius, closely related to eukaryotes. *Biochim Biophys Acta,* 1334**,** 1-4.

KOGA, Y., NISHIHARA, M., MORII, H. & AKAGAWA-MATSUSHITA, M. 1993. Ether polar lipids of methanogenic bacteria: structures, comparative aspects, and biosyntheses. *Microbiol Rev,* 57**,** 164-82.

KONRAD, Z. & EICHLER, J. 2002. Lipid modification of proteins in Archaea: attachment of a mevalonic acid-based lipid moiety to the surface-layer glycoprotein of Haloferax volcanii follows protein translocation. *Biochem J,* 366**,** 959-64.

KOPONEN, M. A., ZLOCK, D., PALMER, D. L. & MERLIN, T. L. 1991. Melioidosis. Forgotten, but not gone! *Arch Intern Med,* 151**,** 605-8.

KORNFELD, R. & KORNFELD, S. 1985. Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem,* 54**,** 631-64.

KOWARIK, M., YOUNG, N. M., NUMAO, S., SCHULZ, B. L., HUG, I., CALLEWAERT, N., MILLS, D. C., WATSON, D. C., HERNANDEZ, M., KELLY, J. F., WACKER, M. & AEBI, M. 2006. Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J,* 25**,** 1957-66.

KUNTZ, C., SONNENBICHLER, J., SONNENBICHLER, I., SUMPER, M. & ZEITLER, R. 1997. Isolation and characterization of dolichol-linked oligosaccharides from Haloferax volcanii. *Glycobiology,* 7**,** 897-904.

LAZAR ADLER, N. R., GOVAN, B., CULLINANE, M., HARPER, M., ADLER, B. & BOYCE, J. D. 2009. The molecular and cellular basis of pathogenesis in melioidosis: how does Burkholderia pseudomallei cause disease? *FEMS Microbiol Rev,* 33**,** 1079-99.

LECHNER, J. & SUMPER, M. 1987. The primary structure of a procaryotic glycoprotein. Cloning and sequencing of the cell surface glycoprotein gene of halobacteria. *J Biol Chem,* 262**,** 9724-9.

LECHNER, J. & WIELAND, F. 1989. Structure and biosynthesis of prokaryotic glycoproteins. *Annu Rev Biochem,* 58**,** 173-94.

LECHNER, J., WIELAND, F. & SUMPER, M. 1985. Transient methylation of dolichyl oligosaccharides is an obligatory step in halobacterial sulfated glycoprotein biosynthesis. *J Biol Chem,* 260**,** 8984-9.

LEELARASAMEE, A. & BOVORNKITTI, S. 1989. Melioidosis: review and update. *Rev Infect Dis,* 11**,** 413-25.

LENNARZ, W. J. 2007. Studies on oligosaccharyl transferase in yeast. *Acta Biochim Pol,* 54**,** 673-7.

LIN, C. H., BOURQUE, G. & TAN, P. 2008. A comparative synteny map of Burkholderia species links large-scale genome rearrangements to fine-scale nucleotide variation in prokaryotes. *Mol Biol Evol,* 25**,** 549-58.

LINDBERG, B. & LONNGREN, J. 1978. Methylation analysis of complex carbohydrates: general procedure and application for sequence analysis. *Methods Enzymol,* 50**,** 3-33.

LINSCHEID, M., D'ANGONA, J., BURLINGAME, A. L., DELL, A. & BALLOU, C. E. 1981. Field desorption mass spectrometry of oligosaccharides. *Proc Natl Acad Sci U S A,* 78**,** 1471-5.

LINTON, D., DORRELL, N., HITCHEN, P. G., AMBER, S., KARLYSHEV, A. V., MORRIS, H. R., DELL, A., VALVANO, M. A., AEBI, M. & WREN, B. W. 2005. Functional analysis of the Campylobacter jejuni N-linked protein glycosylation pathway. *Mol Microbiol,* 55**,** 1695-703.

MA, B., REYNOLDS, C. M. & RAETZ, C. R. 2008. Periplasmic orientation of nascent lipid A in the inner membrane of an Escherichia coli LptA mutant. *Proc Natl Acad Sci U S A,* 105**,** 13823-8.

MACFARLANE, R. D. & TORGERSON, D. F. 1976. Californium-252 plasma desorption mass spectroscopy. *Science,* 191**,** 920-5.

MAGIDOVICH, H. & EICHLER, J. 2009. Glycosyltransferases and oligosaccharyltransferases in Archaea: putative components of the N-glycosylation pathway in the third domain of life. *FEMS Microbiol Lett,* 300**,** 122-30.

MAGIDOVICH, H., YURIST-DOUTSCH, S., KONRAD, Z., VENTURA, V. V., DELL, A., HITCHEN, P. G. & EICHLER, J. 2010. AglP is a S-adenosyl-L-methionine-dependent methyltransferase that participates in the N-glycosylation pathway of Haloferax volcanii. *Mol Microbiol,* 76**,** 190-9.

MAHARJAN, B., CHANTRATITA, N., VESARATCHAVEST, M., CHENG, A., WUTHIEKANUN, V., CHIERAKUL, W., CHAOWAGUL, W., DAY, N. P. & PEACOCK, S. J. 2005. Recurrent melioidosis in patients in northeast Thailand is frequently due to reinfection rather than relapse. *J Clin Microbiol,* 43**,** 6032-4.

MANCA, M. C., WEINTRAUB, A. & WIDMALM, G. 1996. Structural studies of the Escherichia coli O26 O-antigen polysaccharide. *Carbohydr Res,* 281**,** 155-60.

MARCEAU, M. & NASSIF, X. 1999. Role of glycosylation at Ser63 in production of soluble pilin in pathogenic Neisseria. *J Bacteriol,* 181**,** 656-61.

MARINO, G., PUCCI, P., MALORNI, A. & MORRIS, H. R. 1988. Analysis of post-translational modifications by FAB mass spectrometry. *Adv Exp Med Biol,* 231**,** 651-7.

MAROLDA, C. L., VICARIOLI, J. & VALVANO, M. A. 2004. Wzx proteins involved in biosynthesis of O antigen function in association with the first sugar of the O-specific lipopolysaccharide subunit. *Microbiology,* 150**,** 4095-105.

MARSHALL, A. G., HENDRICKSON, C. L. & JACKSON, G. S. 1998. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev,* 17**,** 1-35.

MARSHALL, R. D. 1974. The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins. *Biochem Soc Symp***,** 17-26.

MASOUD, H., PERRY, M. B., BRISSON, J. R., UHRIN, D., LI, J. & RICHARDS, J. C. 2009. Structural elucidation of the novel core oligosaccharide from LPS of Burkholderia cepacia serogroup O4. *Glycobiology,* 19**,** 462-71.

MATTAR, S., SCHARF, B., KENT, S. B., RODEWALD, K., OESTERHELT, D. & ENGELHARD, M. 1994. The primary structure of halocyanin, an archaeal blue copper protein, predicts a lipid anchor for membrane fixation. *J Biol Chem,* 269**,** 14939-45.

MATTOS, K. A., TODESCHINI, A. R., HEISE, N., JONES, C., PREVIATO, J. O. & MENDONCA-PREVIATO, L. 2005. Nitrogen-fixing bacterium Burkholderia brasiliensis produces a novel yersiniose A-containing O-polysaccharide. *Glycobiology,* 15**,** 313-21.

MAYS, E. E. & RICKETTS, E. A. 1975. Melioidosis: recrudescence associated with bronchogenic carcinoma twenty-six years following initial geographic exposure. *Chest,* 68**,** 261-3.

MCNEIL, M., DARVILL, A. G., AMAN, P., FRANZEN, L. E. & ALBERSHEIM, P. 1982. Structural analysis of complex carbohydrates using high-performance liquid chromatography, gas chromatography, and mass spectrometry. *Methods Enzymol,* 83**,** 3-45.

MECHREF, Y. & NOVOTNY, M. V. 2006. Miniaturized separation techniques in glycomic investigations. *J Chromatogr B Analyt Technol Biomed Life Sci,* 841**,** 65-78.

MENGELE, R. & SUMPER, M. 1992. Drastic differences in glycosylation of related S-layer glycoproteins from moderate and extreme halophiles. *J Biol Chem,* 267**,** 8182-5.

MERCK & CO., I. 2012. *Pneumovax23 (Pneumococcal Vaccine Polyvalent)* [Online]. Merck Sharp & Dohme Corporation. [Accessed 13.03.2013.

MESCHER, M. F. & STROMINGER, J. L. 1976. Purification and characterization of a prokaryotic glucoprotein from the cell envelope of Halobacterium salinarium. *J Biol Chem,* 251**,** 2005-14.

MESSNER, P. 2004. Prokaryotic glycoproteins: unexplored but important. *J Bacteriol,* 186**,** 2517-9.

MEVARECH, M. & WERCZBERGER, R. 1985. Genetic transfer in Halobacterium volcanii. *J Bacteriol,* 162**,** 461-2.

MOCK, K. K., DAVEY, M. & COTTRELL, J. S. 1991. The analysis of underivatized oligosaccharides by matrix-assisted laser desorption mass spectrometry. *Biochem Biophys Res Commun,* 177**,** 644-51.

MOHR, M. D., BORNSEN, K. O. & WIDMER, H. M. 1995. Matrix-assisted laser desorption/ionization mass spectrometry: improved matrix for oligosaccharides. *Rapid Commun Mass Spectrom,* 9**,** 809-14.

MOLINARO, A., BEDINI, E., FERRARA, R., LANZETTA, R., PARRILLI, M., EVIDENTE, A., LO CANTORE, P. & IACOBELLIS, N. S. 2003. Structural determination of the O-specific chain of the lipopolysaccharide from the mushrooms pathogenic bacterium Pseudomonas tolaasii. *Carbohydr Res,* 338**,** 1251-7.

MOORE, R. A., RECKSEIDLER-ZENTENO, S., KIM, H., NIERMAN, W., YU, Y., TUANYOK, A., WARAWA, J., DESHAZER, D. & WOODS, D. E. 2004. Contribution of gene loss to the pathogenic evolution of Burkholderia pseudomallei and Burkholderia mallei. *Infect Immun,* 72**,** 4172-87.

MORRIS, H. R. 1972. Studies towards the complete sequence determination of proteins by mass spectrometry; a rapid procedure for the successful permethylation of histidine containing peptides. *FEBS Lett,* 22**,** 257-260.

MORRIS, H. R., DELL, A. & MCDOWELL, R. A. 1981a. Extended performance using a high field magnet mass spectrometer. *Biomed Mass Spectrom,* 8**,** 463-73.

MORRIS, H. R., PANICO, M., BARBER, M., BORDOLI, R. S., SEDGWICK, R. D. & TYLER, A. 1981b. Fast atom bombardment: a new mass spectrometric method for peptide sequence analysis. *Biochem Biophys Res Commun,* 101**,** 623-31.

MORRIS, H. R., PAXTON, T., DELL, A., LANGHORNE, J., BERG, M., BORDOLI, R. S., HOYES, J. & BATEMAN, R. H. 1996. High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Commun Mass Spectrom,* 10**,** 889-96.

MORRIS, H. R., THOMPSON, M. R., OSUGA, D. T., AHMED, A. I., CHAN, S. M., VANDENHEEDE, J. R. & FEENEY, R. E. 1978. Antifreeze glycoproteins from the blood of an antarctic fish. The structure of the proline-containing glycopeptides. *J Biol Chem,* 253**,** 5155-62.

MULLAKHANBHAI, M. F. & LARSEN, H. 1975. Halobacterium volcanii spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Arch Microbiol,* 104**,** 207-14.

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. *J Biol Chem,* 278**,** 34090-101.

NANDI, T., ONG, C., SINGH, A. P., BODDEY, J., ATKINS, T., SARKAR-TYSON, M., ESSEX-LOPRESTI, A. E., CHUA, H. H., PEARSON, T., KREISBERG, J. F., NILSSON, C., ARIYARATNE, P., RONNING, C., LOSADA, L., RUAN, Y., SUNG, W. K., WOODS, D., TITBALL, R. W., BEACHAM, I., PEAK, I., KEIM, P., NIERMAN, W. C. & TAN, P. 2010. A genomic survey of positive selection in Burkholderia pseudomallei provides insights into the evolution of accidental virulence. *PLoS Pathog,* 6**,** e1000845.

NAPARSTEK, S., GUAN, Z. & EICHLER, J. 2012. A predicted geranylgeranyl reductase reduces the omega-position isoprene of dolichol phosphate in the halophilic archaeon, Haloferax volcanii. *Biochim Biophys Acta,* 1821**,** 923-33.

NAPARSTEK, S., VINAGRADOV, E. & EICHLER, J. 2010. Different glycosyltransferases are involved in lipid glycosylation and protein N-glycosylation in the halophilic archaeon Haloferax volcanii. *Arch Microbiol,* 192**,** 581-4.

NELSON, M., PRIOR, J. L., LEVER, M. S., JONES, H. E., ATKINS, T. P. & TITBALL, R. W. 2004. Evaluation of lipopolysaccharide and capsular polysaccharide as subunit vaccines against experimental melioidosis. *J Med Microbiol,* 53**,** 1177-82.

NGAUY, V., LEMESHEV, Y., SADKOWSKI, L. & CRAWFORD, G. 2005. Cutaneous melioidosis in a man who was taken as a prisoner of war by the Japanese during World War II. *J Clin Microbiol,* 43**,** 970-2.

NGUGI, S. A., VENTURA, V. V., QAZI, O., HARDING, S. V., KITTO, G. B., ESTES, D. M., DELL, A., TITBALL, R. W., ATKINS, T. P., BROWN, K. A., HITCHEN, P. G. & PRIOR, J. L. 2010. Lipopolysaccharide from Burkholderia thailandensis E264 provides protection in a murine model of melioidosis. *Vaccine,* 28**,** 7551-5.

NIERMAN, W. C., DESHAZER, D., KIM, H. S., TETTELIN, H., NELSON, K. E., FELDBLYUM, T., ULRICH, R. L., RONNING, C. M., BRINKAC, L. M., DAUGHERTY, S. C., DAVIDSEN, T. D., DEBOY, R. T., DIMITROV, G., DODSON, R. J., DURKIN, A. S., GWINN, M. L., HAFT, D. H., KHOURI, H., KOLONAY, J. F., MADUPU, R., MOHAMMOUD, Y., NELSON, W. C., RADUNE, D., ROMERO, C. M., SARRIA, S., SELENGUT, J., SHAMBLIN, C., SULLIVAN, S. A., WHITE, O., YU, Y., ZAFAR, N., ZHOU, L. & FRASER, C. M. 2004. Structural flexibility in the Burkholderia mallei genome. *Proc Natl Acad Sci U S A,* 101**,** 14246-51.

NORTH, S. J., JANG-LEE, J., HARRISON, R., CANIS, K., ISMAIL, M. N., TROLLOPE, A., ANTONOPOULOS, A., PANG, P. C., GRASSI, P., AL-CHALABI, S., ETIENNE, A. T., DELL, A. & HASLAM, S. M. 2010. Mass spectrometric analysis of mutant mice. *Methods Enzymol,* 478**,** 27-77.

NOTHAFT, H. & SZYMANSKI, C. M. 2010. Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol,* 8**,** 765-78.

NOVEM, V., SHUI, G., WANG, D., BENDT, A. K., SIM, S. H., LIU, Y., THONG, T. W., SIVALINGAM, S. P., OOI, E. E., WENK, M. R. & TAN, G. 2009. Structural and biological diversity of lipopolysaccharides from Burkholderia pseudomallei and Burkholderia thailandensis. *Clin Vaccine Immunol,* 16**,** 1420-8.

NOVOTNY, M. V. & MECHREF, Y. 2005. New hyphenated methodologies in high-sensitivity glycoprotein analysis. *J Sep Sci,* 28**,** 1956-68.

OGAWA, T., ASAI, Y., MAKIMURA, Y. & TAMAI, R. 2007. Chemical structure and immunobiological activity of Porphyromonas gingivalis lipid A. *Front Biosci,* 12**,** 3795-812.

PALLERONI, N. J., KUNISAWA, R., CONTOPOULOU, R. & DOUDOROFF, M. 1973. Nucleic acid homologies in the genus *Pseudomonas*. *Int J Syst Bacteriol,* 23**,** 333-339.

PAUL, A. V., TADA, H., VON DER HELM, K., WISSEL, T., KIEHN, R., WIMMER, E. & DEINHARDT, F. 1987. The entire nucleotide sequence of the genome of human hepatitis A virus (isolate MBB). *Virus Res,* 8**,** 153-71.

PAUL, G. & WIELAND, F. 1987. Sequence of the halobacterial glycosaminoglycan. *J Biol Chem,* 262**,** 9587-93.

PERRY, M. B., MACLEAN, L. L., SCHOLLAARDT, T., BRYAN, L. E. & HO, M. 1995. Structural characterization of the lipopolysaccharide O antigens of Burkholderia pseudomallei. *Infect Immun,* 63**,** 3348-52.

PEYFOON, E., MEYER, B., HITCHEN, P. G., PANICO, M., MORRIS, H. R., HASLAM, S. M., ALBERS, S. V. & DELL, A. 2010. The S-layer glycoprotein of the crenarchaeote Sulfolobus acidocaldarius is glycosylated at multiple sites with chitobiose-linked N-glycans. *Archaea,* 2010.

PHETSOUVANH, R., PHONGMANY, S., NEWTON, P., MAYXAY, M., RAMSAY, A., WUTHIEKANUN, V. & WHITE, N. J. 2001. Melioidosis and Pandora's box in the Lao People's Democratic Republic. *Clin Infect Dis,* 32**,** 653-4.

PITT, T. L., TRAKULSOMBOON, S. & DANCE, D. A. 2000. Molecular phylogeny of Burkholderia pseudomallei. *Acta Trop,* 74**,** 181-5.

PLAVNER, N. & EICHLER, J. 2008. Defining the topology of the N-glycosylation pathway in the halophilic archaeon Haloferax volcanii. *J Bacteriol,* 190**,** 8045-52.

PREHM, P., STIRM, S., JANN, B. & JANN, K. 1975. Cell-wall lipopolysaccharide from Escherichia coli B. *Eur J Biochem,* 56**,** 41-55.

PUTHUCHEARY, S. D. & NATHAN, S. A. 2006. Comparison by electron microscopy of intracellular events and survival of Burkholderia pseudomallei in monocytes from normal subjects and patients with melioidosis. *Singapore Med J,* 47**,** 697-703.

RAETZ, C. R., REYNOLDS, C. M., TRENT, M. S. & BISHOP, R. E. 2007. Lipid A modification systems in gram-negative bacteria. *Annu Rev Biochem,* 76**,** 295-329.

RAETZ, C. R. & WHITFIELD, C. 2002. Lipopolysaccharide endotoxins. *Annu Rev Biochem,* 71**,** 635-700.

RECKSEIDLER, S. L., DESHAZER, D., SOKOL, P. A. & WOODS, D. E. 2001. Detection of bacterial virulence genes by subtractive hybridization: identification of capsular polysaccharide of Burkholderia pseudomallei as a major virulence determinant. *Infect Immun,* 69**,** 34-44.

REID, C. W., STUPAK, J., SZYMANSKI, C. M. & LI, J. 2009. Analysis of bacterial lipid-linked oligosaccharide intermediates using porous graphitic carbon liquid chromatography-electrospray ionization mass spectrometry: heterogeneity in the polyisoprenyl carrier revealed. *Anal Chem,* 81**,** 8472-8.

RISTL, R., STEINER, K., ZARSCHLER, K., ZAYNI, S., MESSNER, P. & SCHAFFER, C. 2011. The s-layer glycome-adding to the sugar coat of bacteria. *Int J Microbiol,* 2011.

ROEPSTORFF, P. & FOHLMAN, J. 1984. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom,* 11**,** 601.

RONCERO, C. & CASADABAN, M. J. 1992. Genetic analysis of the genes involved in synthesis of the lipopolysaccharide core in Escherichia coli K-12: three operons in the rfa locus. *J Bacteriol,* 174**,** 3250-60.

ROTZ, L. D., KHAN, A. S., LILLIBRIDGE, S. R., OSTROFF, S. M. & HUGHES, J. M. 2002. Public health assessment of potential biological terrorism agents. *Emerg Infect Dis,* 8**,** 225-30.

ROYLE, L., MATTU, T. S., HART, E., LANGRIDGE, J. I., MERRY, A. H., MURPHY, N., HARVEY, D. J., DWEK, R. A. & RUDD, P. M. 2002. An analytical and structural database provides a strategy for sequencing O-glycans from microgram quantities of glycoproteins. *Anal Biochem,* 304**,** 70-90.

RUIZ, N., GRONENBERG, L. S., KAHNE, D. & SILHAVY, T. J. 2008. Identification of two inner-membrane proteins required for the transport of lipopolysaccharide to the outer membrane of Escherichia coli. *Proc Natl Acad Sci U S A,* 105**,** 5537-42.

SAGAMI, H., KIKUCHI, A. & OGURA, K. 1995. A novel type of protein modification by isoprenoid-derived materials. Diphytanylglycerylated proteins in Halobacteria. *J Biol Chem,* 270**,** 14851-4.

SAGAMI, H., KIKUCHI, A., OGURA, K., FUSHIHARA, K. & NISHINO, T. 1994. Novel isoprenoid modified proteins in Halobacteria. *Biochem Biophys Res Commun,* 203**,** 972-8.

SANTANIRAND, P., HARLEY, V. S., DANCE, D. A., DRASAR, B. S. & BANCROFT, G. J. 1999. Obligatory role of gamma interferon for host survival in a murine model of infection with Burkholderia pseudomallei. *Infect Immun,* 67**,** 3593-600.

SARKAR-TYSON, M., SMITHER, S. J., HARDING, S. V., ATKINS, T. P. & TITBALL, R. W. 2009. Protective efficacy of heat-inactivated B. thailandensis, B. mallei or B. pseudomallei against experimental melioidosis and glanders. *Vaccine,* 27**,** 4447-51.

SARKAR-TYSON, M., THWAITE, J. E., HARDING, S. V., SMITHER, S. J., OYSTON, P. C., ATKINS, T. P. & TITBALL, R. W. 2007. Polysaccharides and virulence of Burkholderia pseudomallei. *J Med Microbiol,* 56**,** 1005-10.

SARKAR-TYSON, M. & TITBALL, R. W. 2010. Progress toward development of vaccines against melioidosis: A review. *Clin Ther,* 32**,** 1437-45.

SCHAFFER, C. & MESSNER, P. 2001. Glycobiology of surface layer proteins. *Biochimie,* 83**,** 591-9.

SCHAFFER, C. & MESSNER, P. 2004. Surface-layer glycoproteins: an example for the diversity of bacterial glycosylation with promising impacts on nanobiotechnology. *Glycobiology,* 14**,** 31R-42R.

SCHILLER, B., HYKOLLARI, A., YAN, S., PASCHINGER, K. & WILSON, I. B. 2012. Complicated N-linked glycans in simple organisms. *Biol Chem,* 393**,** 661-73.

SCHNAITMAN, C. A. & KLENA, J. D. 1993. Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol Rev,* 57**,** 655-82.

SCHULZ, B. L. & AEBI, M. 2009. Analysis of glycosylation site occupancy reveals a role for Ost3p and Ost6p in site-specific N-glycosylation efficiency. *Mol Cell Proteomics,* 8**,** 357-64.

SCOTT, A. E., TWINE, S. M., FULTON, K. M., TITBALL, R. W., ESSEX-LOPRESTI, A. E., ATKINS, T. P. & PRIOR, J. L. 2011. Flagellar glycosylation in Burkholderia pseudomallei and Burkholderia thailandensis. *J Bacteriol,* 193**,** 3577-87.

SHAFFER, S. A., HARVEY, M. D., GOODLETT, D. R. & ERNST, R. K. 2007. Structural heterogeneity and environmentally regulated remodeling of Francisella tularensis subspecies novicida lipid A characterized by tandem mass spectrometry. *J Am Soc Mass Spectrom,* 18**,** 1080-92.

SILBERSTEIN, S. & GILMORE, R. 1996. Biochemistry, molecular biology, and genetics of the oligosaccharyltransferase. *FASEB J,* 10**,** 849-58.

SILIPO, A., MOLINARO, A., COMEGNA, D., STURIALE, L., CESCUTTI, P., GAROZZO, D., LANZETTA, R. & PARRILLI, M. 2006. Full Structural Characterisation of the Lipooligosaccharide of a Burkholderia pyrrocinia Clinical Isolate. *Eur J Org Chem,* 2006**,** 4874-4883.

SIM, S. H., YU, Y., LIN, C. H., KARUTURI, R. K., WUTHIEKANUN, V., TUANYOK, A., CHUA, H. H., ONG, C., PARAMALINGAM, S. S., TAN, G., TANG, L., LAU, G., OOI, E. E., WOODS, D., FEIL, E., PEACOCK, S. J. & TAN, P. 2008. The core and accessory genomes of Burkholderia pseudomallei: implications for human melioidosis. *PLoS Pathog,* 4**,** e1000178.

SLEYTR, U. B. & BEVERIDGE, T. J. 1999. Bacterial S-layers. *Trends Microbiol,* 7**,** 253-60.

SLEYTR, U. B., EGELSEER, E. M., ILK, N., PUM, D. & SCHUSTER, B. 2007. S-Layers as a basic building block in a molecular construction kit. *FEBS J,* 274**,** 323-34.

SLEYTR, U. B. & THORNE, K. J. 1976. Chemical characterization of the regularly arranged surface layers of Clostridium thermosaccharolyticum and Clostridium thermohydrosulfuricum. *J Bacteriol,* 126**,** 377-83.

SMIT, A. & MUSHEGIAN, A. 2000. Biosynthesis of isoprenoids via mevalonate in Archaea: the lost pathway. *Genome Res,* 10**,** 1468-84.

SMITH, M. D., ANGUS, B. J., WUTHIEKANUN, V. & WHITE, N. J. 1997. Arabinose assimilation defines a nonvirulent biotype of Burkholderia pseudomallei. *Infect Immun,* 65**,** 4319-21.

SMITH, M. D., WUTHIEKANUN, V., WALSH, A. L. & WHITE, N. J. 1995. Quantitative recovery of Burkholderia pseudomallei from soil in Thailand. *Trans R Soc Trop Med Hyg,* 89**,** 488-90.

SPERANDEO, P., CESCUTTI, R., VILLA, R., DI BENEDETTO, C., CANDIA, D., DEHO, G. & POLISSI, A. 2007. Characterization of lptA and lptB, two essential genes implicated in lipopolysaccharide transport to the outer membrane of Escherichia coli. *J Bacteriol,* 189**,** 244-53.

SPERANDEO, P., LAU, F. K., CARPENTIERI, A., DE CASTRO, C., MOLINARO, A., DEHO, G., SILHAVY, T. J. & POLISSI, A. 2008. Functional analysis of the protein machinery required for transport of lipopolysaccharide to the outer membrane of Escherichia coli. *J Bacteriol,* 190**,** 4460-9.

SPINA, E., STURIALE, L., ROMEO, D., IMPALLOMENI, G., GAROZZO, D., WAIDELICH, D. & GLUECKMANN, M. 2004. New fragmentation mechanisms in matrix-assisted laser desorption/ionization time-of-flight/time-of-flight tandem mass spectrometry of carbohydrates. *Rapid Commun Mass Spectrom,* 18**,** 392-8.

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

SPROTT, G. D., LAROCQUE, S., CADOTTE, N., DICAIRE, C. J., MCGEE, M. & BRISSON, J. R. 2003. Novel polar lipids of halophilic eubacterium Planococcus H8 and archaeon Haloferax volcanii. *Biochim Biophys Acta,* 1633**,** 179-88.

STEEGHS, L., JENNINGS, M. P., POOLMAN, J. T. & VAN DER LEY, P. 1997. Isolation and characterization of the Neisseria meningitidis lpxD-fabZ-lpxA gene cluster involved in lipid A biosynthesis. *Gene,* 190**,** 263-70.

STEENSLAND, H. & LARSEN, H. 1969. A study of the cell envelope of the halobacteria. *J Gen Microbiol,* 55**,** 325-36.

STEPHENS, E., MASLEN, S. L., GREEN, L. G. & WILLIAMS, D. H. 2004. Fragmentation characteristics of neutral N-linked glycans using a MALDI-TOF/TOF tandem mass spectrometer. *Anal Chem,* 76**,** 2343-54.

STEVENS, M. P., STEVENS, J. M., JENG, R. L., TAYLOR, L. A., WOOD, M. W., HAWES, P., MONAGHAN, P., WELCH, M. D. & GALYOV, E. E. 2005. Identification of a bacterial factor required for actin-based motility of Burkholderia pseudomallei. *Mol Microbiol,* 56**,** 40-53.

STIMSON, E., VIRJI, M., MAKEPEACE, K., DELL, A., MORRIS, H. R., PAYNE, G., SAUNDERS, J. R., JENNINGS, M. P., BARKER, S., PANICO, M. & ET AL. 1995. Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose. *Mol Microbiol,* 17**,** 1201-14.

STOECKENIUS, W. & ROWEN, R. 1967. A morphological study of Halobacterium halobium and its lysis in media of low salt concentration. *J Cell Biol,* 34**,** 365-93.

STONE, J. K., MAYO, M., GRASSO, S. A., GINTHER, J. L., WARRINGTON, S. D., ALLENDER, C. J., DOYLE, A., GEORGIA, S., KAESTLI, M., BROOMALL, S. M., KARAVIS, M. A., INSALACO, J. M., HUBBARD, K. S., MCNEW, L. A., GIBBONS, H. S., CURRIE, B. J., KEIM, P. & TUANYOK, A. 2012. Detection of Burkholderia pseudomallei O-antigen serotypes in near-neighbor species. *BMC Microbiol,* 12**,** 250.

SUITS, M. D., SPERANDEO, P., DEHO, G., POLISSI, A. & JIA, Z. 2008. Novel structure of the conserved gram-negative lipopolysaccharide transport protein A and mutagenesis analysis. *J Mol Biol,* 380**,** 476-88.

SUMPER, M., BERG, E., MENGELE, R. & STROBEL, I. 1990. Primary structure and glycosylation of the S-layer protein of Haloferax volcanii. *J Bacteriol,* 172**,** 7111-8.

SUPUTTAMONGKOL, Y., CHAOWAGUL, W., CHETCHOTISAKD, P., LERTPATANASUWUN, N., INTARANONGPAI, S., RUCHUTRAKOOL, T., BUDHSARAWONG, D., MOOTSIKAPUN, P., WUTHIEKANUN, V., TEERAWATASOOK, N. & LULITANOND, A. 1999. Risk factors for melioidosis and bacteremic melioidosis. *Clin Infect Dis,* 29**,** 408-13.

SUPUTTAMONGKOL, Y., HALL, A. J., DANCE, D. A., CHAOWAGUL, W., RAJCHANUVONG, A., SMITH, M. D. & WHITE, N. J. 1994. The epidemiology of melioidosis in Ubon Ratchatani, northeast Thailand. *Int J Epidemiol,* 23**,** 1082-90.

SZYMANSKI, C. M., BURR, D. H. & GUERRY, P. 2002. Campylobacter protein glycosylation affects host cell interactions. *Infect Immun,* 70**,** 2242-4.

SZYMANSKI, C. M., LOGAN, S. M., LINTON, D. & WREN, B. W. 2003. Campylobacter--a tale of two protein glycosylation systems. *Trends Microbiol,* 11**,** 233-8.

SZYMANSKI, C. M. & WREN, B. W. 2005. Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Microbiol,* 3**,** 225-37.

SZYMANSKI, C. M., YAO, R., EWING, C. P., TRUST, T. J. & GUERRY, P. 1999. Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Mol Microbiol,* 32**,** 1022-30.

TAWEECHAISUPAPONG, S., KAEWPA, C., ARUNYANART, C., KANLA, P., HOMCHAMPA, P., SIRISINHA, S., PROUNGVITAYA, T. & WONGRATANACHEEWIN, S. 2005. Virulence of Burkholderia pseudomallei does not correlate with biofilm formation. *Microb Pathog,* 39**,** 77-85.

TAYLOR, M. E. & DRICKAMER, K. 2011. *Introduction to Glycobiology,* Great Clarendon, Oxford University Press.

THIBODEAUX, C. J., MELANCON, C. E., 3RD & LIU, H. W. 2008. Natural-product sugar biosynthesis and enzymatic glycodiversification. *Angew Chem Int Ed Engl,* 47**,** 9814-59.

TILLETT, W. S. & FRANCIS, T. 1929. Cutaneous Reactions to the Polysaccharides and Proteins of Pneumococcus in Lobar Pneumonia. *J Exp Med,* 50**,** 687-701.

TIPPAYAWAT, P., SAENWONGSA, W., MAHAWANTUNG, J., SUWANNASAEN, D., CHETCHOTISAKD, P., LIMMATHUROTSAKUL, D., PEACOCK, S. J., FELGNER, P. L., ATKINS, H. S., TITBALL, R. W., BANCROFT, G. J. & LERTMEMONGKOLCHAI, G. 2009. Phenotypic and functional characterization of human memory T cell responses to Burkholderia pseudomallei. *PLoS Negl Trop Dis,* 3**,** e407.

TITBALL, R. W., RUSSELL, P., CUCCUI, J., EASTON, A., HAQUE, A., ATKINS, T., SARKAR-TYSON, M., HARLEY, V., WREN, B. & BANCROFT, G. J. 2008. Burkholderia pseudomallei: animal models of infection. *Trans R Soc Trop Med Hyg,* 102 Suppl 1**,** S111-6.

TOMAN, R., GARIDEL, P., ANDRA, J., SLABA, K., HUSSEIN, A., KOCH, M. H. & BRANDENBURG, K. 2004. Physicochemical characterization of the endotoxins from Coxiella burnetii strain Priscilla in relation to their bioactivities. *BMC Biochem,* 5**,** 1.

TRAKULSOMBOON, S., VUDDHAKUL, V., THARAVICHITKUL, P., NA-GNAM, N., SUPUTTAMONGKOL, Y. & THAMLIKITKUL, V. 1999. Epidemiology of arabinose assimilation in burkholderia pseudomallei isolated from patients and soil in Thailand. *Southeast Asian J Trop Med Public Health,* 30**,** 756-9.

TUANYOK, A., AUERBACH, R. K., BRETTIN, T. S., BRUCE, D. C., MUNK, A. C., DETTER, J. C., PEARSON, T., HORNSTRA, H., SERMSWAN, R. W., WUTHIEKANUN, V., PEACOCK, S. J., CURRIE, B. J., KEIM, P. & WAGNER, D. M. 2007. A horizontal gene transfer event defines two distinct groups within Burkholderia pseudomallei that have dissimilar geographic distributions. *J Bacteriol,* 189**,** 9044-9.

TUANYOK, A., STONE, J. K., MAYO, M., KAESTLI, M., GRUENDIKE, J., GEORGIA, S., WARRINGTON, S., MULLINS, T., ALLENDER, C. J., WAGNER, D. M., CHANTRATITA, N., PEACOCK, S. J., CURRIE, B. J. & KEIM, P. 2012. The genetic and molecular basis of O-antigenic diversity in Burkholderia pseudomallei lipopolysaccharide. *PLoS Negl Trop Dis,* 6**,** e1453.

TUMAPA, S., HOLDEN, M. T., VESARATCHAVEST, M., WUTHIEKANUN, V., LIMMATHUROTSAKUL, D., CHIERAKUL, W., FEIL, E. J., CURRIE, B. J., DAY, N. P., NIERMAN, W. C. & PEACOCK, S. J. 2008. Burkholderia pseudomallei genome plasticity associated with genomic island variation. *BMC Genomics,* 9**,** 190.

ULETT, G. C., KETHEESAN, N. & HIRST, R. G. 2000. Cytokine gene expression in innately susceptible BALB/c mice and relatively resistant C57BL/6 mice during infection with virulent Burkholderia pseudomallei. *Infect Immun,* 68**,** 2034-42.

UTAISINCHAROEN, P., TANGTHAWORNCHAIKUL, N., KESPICHAYAWATTANA, W., ANUNTAGOOL, N., CHAISURIYA, P. & SIRISINHA, S. 2000. Kinetic studies of the production of nitric oxide (NO) and tumour necrosis factor-alpha (TNF-alpha) in macrophages stimulated with Burkholderia pseudomallei endotoxin. *Clin Exp Immunol,* 122**,** 324-9.

VAN SORGE, N. M., BLEUMINK, N. M., VAN VLIET, S. J., SAELAND, E., VAN DER POL, W. L., VAN KOOYK, Y. & VAN PUTTEN, J. P. 2009. N-glycosylated proteins and distinct lipooligosaccharide glycoforms of Campylobacter jejuni target the human C-type lectin receptor MGL. *Cell Microbiol,* 11**,** 1768-81.

VARKI, A., CUMMINGS, R. D., ESKO, J. D., FREEZE, H. H., STANLEY, P., BERTOZZI, C. R., HART, G. W. & ETZLER, M. E. 2009. *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press.

VESTAL, M. L. & CAMPBELL, J. M. 2005. Tandem time-of-flight mass spectrometry. *Methods Enzymol,* 402**,** 79-108.

VLIEGENTHART, F. G. & WOODS, R. J. 2006. *NMR spectroscopy and computer modeling of carbohydrates: recent advances,* Washington, DC, Oxford University Press, New York.

VOISIN, S., HOULISTON, R. S., KELLY, J., BRISSON, J. R., WATSON, D., BARDY, S. L., JARRELL, K. F. & LOGAN, S. M. 2005. Identification and characterization of the unique N-linked glycan common to the flagellins and S-layer glycoprotein of Methanococcus voltae. *J Biol Chem,* 280**,** 16586-93.

VOORHORST, W. G., EGGEN, R. I., GEERLING, A. C., PLATTEEUW, C., SIEZEN, R. J. & VOS, W. M. 1996. Isolation and characterization of the hyperthermostable serine protease, pyrolysin, and its gene from the hyperthermophilic archaeon Pyrococcus furiosus. *J Biol Chem,* 271**,** 20426-31.

VORACHIT, M., LAM, K., JAYANETRA, P. & COSTERTON, J. W. 1995. Electron microscopy study of the mode of growth of Pseudomonas pseudomallei in vitro and in vivo. *J Trop Med Hyg,* 98**,** 379-91.

WACKER, M., FELDMAN, M. F., CALLEWAERT, N., KOWARIK, M., CLARKE, B. R., POHL, N. L., HERNANDEZ, M., VINES, E. D., VALVANO, M. A., WHITFIELD, C. & AEBI, M. 2006. Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proc Natl Acad Sci U S A,* 103**,** 7088-93.

WACKER, M., LINTON, D., HITCHEN, P. G., NITA-LAZAR, M., HASLAM, S. M., NORTH, S. J., PANICO, M., MORRIS, H. R., DELL, A., WREN, B. W. & AEBI, M. 2002. N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science,* 298**,** 1790-3.

WAKAI, H., NAKAMURA, S., KAWASAKI, H., TAKADA, K., MIZUTANI, S., AONO, R. & HORIKOSHI, K. 1997. Cloning and sequencing of the gene encoding the cell surface glycoprotein of Haloarcula japonica strain TR-1. *Extremophiles,* 1**,** 29-35.

WANG, X., KARBARZ, M. J., MCGRATH, S. C., COTTER, R. J. & RAETZ, C. R. 2004. MsbA transporter-dependent lipid A 1-dephosphorylation on the periplasmic surface of the inner membrane: topography of francisella novicida LpxE expressed in Escherichia coli. *J Biol Chem,* 279**,** 49470-8.

WANG, X., MCGRATH, S. C., COTTER, R. J. & RAETZ, C. R. 2006. Expression cloning and periplasmic orientation of the Francisella novicida lipid A 4'-phosphatase LpxF. *J Biol Chem,* 281**,** 9321-30.

WANG, X. & QUINN, P. J. 2010. Lipopolysaccharide: Biosynthetic pathway and structure modification. *Prog Lipid Res,* 49**,** 97-107.

WARAWA, J. & WOODS, D. E. 2002. Melioidosis vaccines. *Expert Rev Vaccines,* 1**,** 477-82.

WARD, A., REYES, C. L., YU, J., ROTH, C. B. & CHANG, G. 2007. Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc Natl Acad Sci U S A,* 104**,** 19005-10.

WEST, T. E., ERNST, R. K., JANSSON-HUTSON, M. J. & SKERRETT, S. J. 2008. Activation of Toll-like receptors by Burkholderia pseudomallei. *BMC Immunol,* 9**,** 46.

WESTPHAL, O. & LUDERITZ, O. 1953. [Chemical and biological analysis of highly purified bacterial polysaccharides]. *Dtsch Med Wochenschr,* 78**,** 17-9.

WHITE, N. J. 2003. Melioidosis. *Lancet,* 361**,** 1715-22.

WHITFIELD, C., AMOR, P. A. & KOPLIN, R. 1997. Modulation of the surface architecture of gram-negative bacteria by the action of surface polymer:lipid A-core ligase and by determinants of polymer chain length. *Mol Microbiol,* 23**,** 629-38.

WHITFIELD, C. & ROBERTS, I. S. 1999. Structure, assembly and regulation of expression of capsules in Escherichia coli. *Mol Microbiol,* 31**,** 1307-19.

WHITMORE, A. 1913. An Account of a Glanders-like Disease occurring in Rangoon. *J Hyg (Lond),* 13**,** 1-34 1.

WIELAND, F. 1988. Structure and biosynthesis of prokaryotic glycoproteins. *Biochimie,* 70**,** 1493-504.

WIERSINGA, W. J., DE VOS, A. F., DE BEER, R., WIELAND, C. W., ROELOFS, J. J., WOODS, D. E. & VAN DER POLL, T. 2008a. Inflammation patterns induced by different Burkholderia species in mice. *Cell Microbiol,* 10**,** 81-7.

WIERSINGA, W. J., DE VOS, A. F., WIELAND, C. W., LEENDERTSE, M., ROELOFS, J. J. & VAN DER POLL, T. 2008b. CD14 impairs host defense against gram-negative sepsis caused by Burkholderia pseudomallei in mice. *J Infect Dis,* 198**,** 1388-97.

WIERSINGA, W. J., DESSING, M. C., KAGER, P. A., CHENG, A. C., LIMMATHUROTSAKUL, D., DAY, N. P., DONDORP, A. M., VAN DER POLL, T. & PEACOCK, S. J. 2007a. High-throughput mRNA profiling characterizes the expression of inflammatory molecules in sepsis caused by Burkholderia pseudomallei. *Infect Immun,* 75**,** 3074-9.

WIERSINGA, W. J., DESSING, M. C. & VAN DER POLL, T. 2008c. Gene-expression profiles in murine melioidosis. *Microbes Infect,* 10**,** 868-77.

WIERSINGA, W. J. & VAN DER POLL, T. 2009. Immunity to Burkholderia pseudomallei. *Curr Opin Infect Dis,* 22**,** 102-8.

WIERSINGA, W. J., VAN DER POLL, T., WHITE, N. J., DAY, N. P. & PEACOCK, S. J. 2006. Melioidosis: insights into the pathogenicity of Burkholderia pseudomallei. *Nat Rev Microbiol,* 4**,** 272-82.

WIERSINGA, W. J., WIELAND, C. W., DESSING, M. C., CHANTRATITA, N., CHENG, A. C., LIMMATHUROTSAKUL, D., CHIERAKUL, W., LEENDERTSE, M., FLORQUIN, S., DE VOS, A. F., WHITE, N., DONDORP, A. M., DAY, N. P., PEACOCK, S. J. & VAN DER POLL, T. 2007b. Toll-like receptor 2 impairs host defense in gram-negative sepsis caused by Burkholderia pseudomallei (Melioidosis). *PLoS Med,* 4**,** e248.

WIERSINGA, W. J., WIELAND, C. W., ROELOFS, J. J. & VAN DER POLL, T. 2008d. MyD88 dependent signaling contributes to protective host defense against Burkholderia pseudomallei. *PLoS One,* 3**,** e3494.

WIERSINGA, W. J., WIELAND, C. W., VAN DER WINDT, G. J., DE BOER, A., FLORQUIN, S., DONDORP, A., DAY, N. P., PEACOCK, S. J. & VAN DER POLL, T. 2007c. Endogenous interleukin-18 improves the early antimicrobial host response in severe melioidosis. *Infect Immun,* 75**,** 3739-46.

WILLIAMS, A. H. & RAETZ, C. R. 2007. Structural basis for the acyl chain selectivity and mechanism of UDP-N-acetylglucosamine acyltransferase. *Proc Natl Acad Sci U S A,* 104**,** 13543-50.

WILSON, I. B. 2002. Glycosylation of proteins in plants and invertebrates. *Curr Opin Struct Biol,* 12**,** 569-77.

WOESE, C. R. & FOX, G. E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A,* 74**,** 5088-90.

WOPEREIS, S., LEFEBER, D. J., MORAVA, E. & WEVERS, R. A. 2006. Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: a review. *Clin Chem,* 52**,** 574-600.

WU, T., MCCANDLISH, A. C., GRONENBERG, L. S., CHNG, S. S., SILHAVY, T. J. & KAHNE, D. 2006. Identification of a protein complex that assembles lipopolysaccharide in the outer membrane of Escherichia coli. *Proc Natl Acad Sci U S A,* 103**,** 11754-9.

WUHRER, M., DEELDER, A. M. & HOKKE, C. H. 2005. Protein glycosylation analysis by liquid chromatography-mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci,* 825**,** 124-33.

WUHRER, M., KOELEMAN, C. A., DEELDER, A. M. & HOKKE, C. H. 2004. Normal-phase nanoscale liquid chromatography-mass spectrometry of underivatized oligosaccharides at low-femtomole sensitivity. *Anal Chem,* 76**,** 833-8.

WUORIMAA, T., KAYHTY, H., ESKOLA, J., BLOIGU, A., LEROY, O. & SURCEL, H. M. 2001. Activation of cell-mediated immunity following immunization with pneumococcal conjugate or polysaccharide vaccine. *Scand J Immunol,* 53**,** 422-8.

WUTHIEKANUN, V., ANUNTAGOOL, N., WHITE, N. J. & SIRISINHA, S. 2002. Short report: a rapid method for the differentiation of Burkholderia pseudomallei and Burkholderia thailandensis. *Am J Trop Med Hyg,* 66**,** 759-61.

WUTHIEKANUN, V., MAYXAY, M., CHIERAKUL, W., PHETSOUVANH, R., CHENG, A. C., WHITE, N. J., DAY, N. P. & PEACOCK, S. J. 2005. Detection of Burkholderia pseudomallei in soil within the Lao People's Democratic Republic. *J Clin Microbiol,* 43**,** 923-4.

WUTHIEKANUN, V., SMITH, M. D., DANCE, D. A., WALSH, A. L., PITT, T. L. & WHITE, N. J. 1996. Biochemical characteristics of clinical and environmental isolates of Burkholderia pseudomallei. *J Med Microbiol,* 45**,** 408-12.

WUTHIEKANUN, V., SMITH, M. D., DANCE, D. A. & WHITE, N. J. 1995a. Isolation of Pseudomonas pseudomallei from soil in north-eastern Thailand. *Trans R Soc Trop Med Hyg,* 89**,** 41-3.

WUTHIEKANUN, V., SMITH, M. D. & WHITE, N. J. 1995b. Survival of Burkholderia pseudomallei in the absence of nutrients. *Trans R Soc Trop Med Hyg,* 89**,** 491.

YABUUCHI, E., KOSAKO, Y., OYAIZU, H., YANO, I., HOTTA, H., HASHIMOTO, Y., EZAKI, T. & ARAKAWA, M. 1992. Proposal of Burkholderia gen. nov. and transfer of seven species of the genus Pseudomonas homology group II to the new genus, with the type species Burkholderia cepacia (Palleroni and Holmes 1981) comb. nov. *Microbiol Immunol,* 36**,** 1251-75.

YAN, A., AHMED, E., YAN, Q. & LENNARZ, W. J. 2003. New findings on interactions among the yeast oligosaccharyl transferase subunits using a chemical cross-linker. *J Biol Chem,* 278**,** 33078-87.

YAN, Q. & LENNARZ, W. J. 2002. Studies on the function of oligosaccharyl transferase subunits. Stt3p is directly involved in the glycosylation process. *J Biol Chem,* 277**,** 47692-700.

YANG, L. L. & HAUG, A. 1979. Purification and partial characterization of a procaryotic glycoprotein from the plasma membrane of Thermoplasma acidophilum. *Biochim Biophys Acta,* 556**,** 265-77.

YOO, H. G., KWON, S. Y., KARKI, S. & KWON, H. J. 2011. A new route to dTDP-6-deoxy-l-talose and dTDP-L-rhamnose: dTDP-L-rhamnose 4-epimerase in Burkholderia thailandensis. *Bioorg Med Chem Lett,* 21**,** 3914-7.

YOUNG, N. M., BRISSON, J. R., KELLY, J., WATSON, D. C., TESSIER, L., LANTHIER, P. H., JARRELL, H. C., CADOTTE, N., ST MICHAEL, F., ABERG, E. & SZYMANSKI, C. M. 2002. Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, Campylobacter jejuni. *J Biol Chem,* 277**,** 42530-9.

YU, Y., KIM, H. S., CHUA, H. H., LIN, C. H., SIM, S. H., LIN, D., DERR, A., ENGELS, R., DESHAZER, D., BIRREN, B., NIERMAN, W. C. & TAN, P. 2006. Genomic patterns of pathogen evolution revealed by comparison of Burkholderia pseudomallei, the causative agent of melioidosis, to avirulent Burkholderia thailandensis. *BMC Microbiol,* 6**,** 46.

YURIST-DOUTSCH, S., ABU-QARN, M., BATTAGLIA, F., MORRIS, H. R., HITCHEN, P. G., DELL, A. & EICHLER, J. 2008. AglF, aglG and aglI, novel members of a gene island involved in the N-glycosylation of the Haloferax volcanii S-layer glycoprotein. *Mol Microbiol,* 69**,** 1234-45.

YURIST-DOUTSCH, S. & EICHLER, J. 2009. Manual annotation, transcriptional analysis, and protein expression studies reveal novel genes in the agl cluster responsible for N glycosylation in the halophilic archaeon Haloferax volcanii. *J Bacteriol,* 191**,** 3068-75.

YURIST-DOUTSCH, S., MAGIDOVICH, H., VENTURA, V. V., HITCHEN, P. G., DELL, A. & EICHLER, J. 2010. N-glycosylation in Archaea: on the coordinated actions of Haloferax volcanii AglF and AglM. *Mol Microbiol,* 75**,** 1047-58.

ZAHRINGER, U., MOLL, H., HETTMANN, T., KNIREL, Y. A. & SCHAFER, G. 2000. Cytochrome b558/566 from the archaeon Sulfolobus acidocaldarius has a unique Asn-linked highly branched hexasaccharide chain containing 6-sulfoquinovose. *Eur J Biochem,* 267**,** 4144-9.