N-linked Protein Glycosylation in Helicobacter Species

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Life Sciences

2012

Alison Grace Wood

Contents

Abstract: ……9
Declaration 10
Chapter 1 15
General Introduction 15
1.0 Introduction 16
1.1 Eukaryotic Protein Glycosylation 16
1.2 Protein Glycosylation in Archaea 19
1.3 O-linked Protein Glycosylation in Bacteria 20
1.4 Bacterial Cytoplasmic N-glycosylation Systems 21
1.5 The *C. jejuni* N-linked Protein Glycosylation System 22
1.6 The Bacterial OST and Amino Acid Sequon Specificity 23
1.7 *C. jejuni* Glycosyltransferases and Assembly of the Heptasaccharide Glycan 26
1.8 Function of *C. jejuni* N-linked Glycoproteins 28
1.9 Exploitating the *C. jejuni* N-linked Protein Glycosylation System 29
1.10 Campylobacter N-linked Protein Glycosylation (*pgl*) Genes 32
1.11 The N-linked Protein Glycosylation System of *W. succinogenes* 34
1.12 N-Linked Glycosylation Systems in *Helicobacter* Species 35
1.13 Summary 38
1.14 Aims of This Study 39
Chapter 2 40
Materials and Methods 40
2.1 Bacterial Strains and Growth Conditions 41
2.2 Isolation of Plasmids and Genomic DNA 41
2.3 PCR Amplification 42
2.4 Reverse Transcriptase PCR 42
2.5 Agarose Gel Electrophoresis 42

2.6 DNA Manipulation 43
2.7 E. coli Transformation 43
2.7 DNA Sequencing. 44
2.8 Site Directed Mutagenesis 44
2.9 Helicobacter Transformation 44
2.10 Purification of Glycoproteins from *H. pullorum* 45
2.11 SDS-PAGE and Western Blotting. 46
2.12 Mass Spectrometery for Characterisation of Proteins 47
2.13 Preparation of Bacterial Cell Membranes 48
2.14 Oligosaccharyltransferase Assay 48
2.15 Tricine SDS-PAGE 49
2.16 Purification of Biotinylated Peptide from the *In vitro* OST Assay 49
Chapter 3 51
The Second *Helicobacter pullorum* Oligosaccharyltransferase (PglB2) is Neither Essential nor Required for N-linked Protein Glycosylation 51
3.1 Background 52
3.2 Results 54
3.2.1 Construction of suicide plasmids for site specific integration of antibiotic resistance cassettes onto the *H. pullorum* chromosome. 54
3.2.2 Integration of antibiotic resistance cassettes onto the *H. pullorum* chromosome within the 23S rRNA gene. 56
3.2.3 Natural Transformation of *H. pullorum* NCTC 12824^{T} with genomic DNA from *H. pullorum* 23::E and 23::Kn strains. 58
3.2.4 Expression of *H. pullorum pgl*B1 and *pgl*B2 genes 59
3.2.5 Insertional mutagenesis of *H. pullorum pgl*B1 and *pgl*B2 genes 61
3.2.6 Role of PglB2 in the N-glycosylation of a small peptide *in vitro* 64
3.2.7 The N-glycan structure of glycopeptide derived from the *H. pullorum pgl*B2 mutant. 66

3.3 Discussion 69

Chapter 4 72

Identification of Putative *H. pullorum* Glycoproteins and Their Glycosylation in *E. coli* with the *C. jejuni* Heptasaccharide 72

4.1 Background 73

4.2 Results 75

4.2.1 Bioinformatic approach to predicting N-linked glycoproteins in *Helicobacter* species 75

4.2.2 Production of *H. pullorum* HgpA and Hp0114 putative glycoproteins in *E. coli* 78

4.2.3 Glycosylation of *H. pullorum* HgpA and Hp0114 in *E. coli* with a *C. jejuni* heptasaccharide glycan. 80

4.2.4 Localisation of the site of glycosylation of Hp0114 84

4.2.5 Amino acid site of glycosylation of *H. pullorum* HgpA 86

4.3 Discussion 88

Chapter 5 91

Glycosylation of Hp0114 and HgpA in *H. pullorum* 91

5.1 Background 92

5.2 Results 93

5.2.1 Production of hexa histidine tagged Hp0114 and HgpA proteins in *H. pullorum* 93

5.2.3 Site of glycosylation of Hp0114 in *H. pullorum* 98

5.2.4 Intact mass analysis of HgpA glycoprotein derived from *H. pullorum* ..100

5.2.5 MALDI Mass Spectrometry of the HgpA glycoprotein derived from wild type *H. pullorum.* 103

5.3 Discussion 110

Chapter 6 113

Characterization of the *H. pullorum* PglB1-Dependent N-Linked Protein Glycosylation System 113

6.1 Background 114

6.2 Results 118

6.2.1 The effect of insertional mutagenesis of putative glycosyltransferase encoding genes on the electrophoretic mobility of *H. pullorum* HgpA protein 118

6.2.2 Intact mass analysis of HgpA derived from *H. pullorum* glycosyltransferase-encoding gene mutants. 120

6.2.3 MALDI Mass Spectrometry of trypsin digested HgpA peptides from *H. pullorum* glycosyltransferase-encoding gene mutants. 123

6.3 Discussion 126

Chapter 7 130

The *H. pullorum* Biosynthetic Enzymes WbpO and WbpS are Required for PglB1-Dependent N-Linked Protein Glycosylation with a Pentasaccharide Glycan 130

7.1 Background 131

7.2 Results 134

7.2.1 Construction of *H. pullorum wbp*O::Kn and *wbp*S::Kn mutants. 134

7.2.2 Role of WbpS in the glycosylation of a small peptide *in vitro* 136

7.2.3 The role of WbpS in the glycosylation of a small peptide *in vitro* 138

7.2.4 Glycosylation of HgpA from *H. pullorum wbp*O::kn and *wbp*S::kn mutants. 140

7.2.5 Intact mass analysis of HgpA from *H. pullorum wbp*O::Kn and *wbp*S::Kn mutants. 142

7.2.6 Tandem MALDI Mass Spectrometry of HgpA peptides derived from *H. pullorum wbp*O::Kn and *wbp*S::Kn mutants. 144

7.3 Discussion 145

Chapter 8 149

General Discussion 149

Figures

Figure 1.1 N-linked Protein Glycosylation in Eukaryotes 18
Figure 1.2 Structural representation of *C. lari* PglB and its interaction with acceptor peptide 25
Figure 1.3 Model for assembly of the lipid-linked heptasaccharide of the *C. jejuni* N-glycosylation system. 27
Figure 1.4 Phylogenetic tree of PglB proteins from Delta and Epsilon Proteobacteria. 31
Figure 1.5 The *pgl* loci from Campylobacter species. 33
Figure 1.6 Gene map of *pgl* loci from *C. jejuni, W. succinogenes* and *H. pullorum* . 37
Figure 3.1 Verification of pHp23::E, pHp23::Kn and pHp23::Cm plasmid structures. 55
Figure 3.2 PCR verification of insertion of the kanamycin and erythromycin resistance cassettes into the *H. pullorum* 23S rRNA gene. 57
Figure 3.3 Expression of *H. pullorum* NCTC 12824^{T} *pgl*B1 and *pgl*B2 genes during *in vitro* growth. 60
Figure 3.4 Verification of plasmid structure of pHp*pgl*B1::Kn and pHp*pgl*B2::Kn . 62
Figure 3.5 Confirmation of insertional mutagenesis of *H. pullorum pgl*B1 and *pgl*B2 63
Figure 3.6 *In vitro* peptide glycosylation 65
Figure 3.7 Tandem MALDI Mass Spectrometry analysis of *H. pullorum* 67
Figure 3.8 Tandem MALDI Mass Spectrometry analysis of N-linked glycan from H. pullorum pglB2::Kn. 68
Figure 4.1 Sequence similarities of *C. jejuni* and *H. pullorum* potential glycoproteins. 77
Figure 4.2 Detection of *H. pullorum* Hp0114 and HgpA proteins produced in *E. coli*. 79
Figure 4.3 Glycosylation of *H. pullorum* HgpA expressed in *E. coli* with a *C. jejuni* heptasaccharide glycan. 81
Figure 4.4 Glycosylation of *H. pullorum* Hp0114 expressed in *E. coli* with a *C. jejuni* heptasaccharide glycan 82

Figure 4.5 Glycosylation of *H. pullorum* HgpA expressed in *E. coli* with a *C. jejuni* heptasaccharide glycan by *H. canadensis* PglB's. 83
Figure 4.6 Sites of glycosylation of Hp0114 expressed in *E. coli* with a *C. jejuni* heptasaccharide glycan..... 85
Figure 4.7 Site of glycosylation of HgpA expressed in *E. coli* with a *C. jejuni* heptasaccharide glycan..... 87
Figure 5.1 Construction of plasmids pHp23E::0114 and pHp23E::*hgp*A. 94
Figure 5.2 PCR verification of *H. pullorum* 23E::HgpA and 23E::0114 strains. 95
Figure 5.3 PglB1 but not PglB2, dependent glycosylation of *H. pullorum* HgpA and Hp0114 proteins. 97
Figure 5.4 Identification of the glycosylation site of the Hp0114 protein. 99
Figure 5.5 Intact mass analysis of HgpA derived from *H. pullorum* wild type (A) and a *pgl*B1::Kn mutant (B)..... 102
Figure 5.6 MALDI Mass Spectrometry of trypsin digested HgpA from *H. pullorum*. 105
Figure 5.7 Fragmentation of a modified HgpA derived tryptic peptide from wild type *H. pullorum*. 107
Figure 5.8 Glycan structure of HgpA derived from the *H. pullorum pgl*B2::Kn mutant..... 109
Figure 6.1 Schematic representations of the *C. jejuni* glycosyltransferase-encoding gene locus and arrangement of orthologues present in *H. pullorum* and *W. succinogenes*. 117
Figure 6.2 Glycosylation of *H. pullorum* HgpA derived from *H. pullorum* glycosyltransferase-encoding gene mutants. 119
Figure 6.3 Intact mass analysis of HgpA derived from *H. pullorum pgl*C::Kn (A), *pgl*A::Kn (B), *pgl*J::Kn (C) and *pgl*H::Kn (D) knock-out mutants. 122
Figure 6.4 Fragmentation of a modified HgpA derived tryptic peptide from *H. pullorum* glycosyltransferase-encoding gene mutants. 125
Figure 6.5 Model comparing the assembly of the *C. jejuni* lipid linked heptasaccharide glycan with the proposed assembly of the *H. pullorum* pentasaccharide glycan..... 129
Figure 7.1 *H. pullorum wbp*O and *wbp*S gene locations and the function of WbpO and WbpS in *P. aeruginosa* O6 and *E. coli* O121 strains. 133

Figure 7.2 Construction and verification of the *H. pullorum wbpS*::Kn mutant.....135
Figure 7.3 The *in vitro* modification of FAM and FITC labelled peptides co-incubated with membrane preparations from a *H. pullorum wbp*S::Kn strain 137
Figure 7.4 Tandem MALDI Mass Spectrometry of a peptide co-incubated with membrane preparations from the *H. pullorum wbp*S::Kn mutant.....139
Figure 7.5 Glycosylation of *H. pullorum* HgpA derived from *H. pullorum wbp*O::Kn and *wbp*S::Kn mutants. 141
Figure 7.6 Intact mass analysis of HgpA derived from *H. pullorum wbp*S::Kn (A) and *wbp*O::Kn (B) mutants. 143
Figure 7.7 Glycan structure of HgpA derived from *H. pullorum wbp*O::Kn and *wbp*S::Kn mutants 145

Final Word count: 38,675

N-linked Protein glycosylation in Helicobacter species

Alison Grace Wood, Doctor of Philosophy in the Faculty of Life Sciences, the University of Manchester (2012)

N-linked protein glycosylation involves the transfer of a glycan onto an Asparagine residue (N) of a polypeptide chain. It is common in Eukaryotes and has recently been observed in Prokaryotes, most notably in *Campylobacter jejuni*. The *C. jejuni* N-linked glycosylation system is encoded on a single *pgl* gene locus that also functions when expressed in *Escherichia coli*. The key enzyme involved in N-linked protein glycosylation is encoded by the *pgl*B gene and transfers lipid-linked glycan onto N residues of glycoproteins in the periplasm. It is clear from accumulating genome sequence data that *pgl*B orthologues are present in all *Campylobacter* species and in related species such as *Wolinella succinogenes, Desulfovibrio vulgaris* and *Desulfovibrio desulfuricans.* Most Helicobacter species, including *Helicobacter pylori*, lack the *pgl*B gene but three related Helicobacter species *Helicobacter pullorum, Helicobacter canadensis and Helicobacter winghamensis* have two distinct *pgl*B genes. These and other orthologues of *C. jejuni pgl* genes are located not within a single locus but rather at five distinct loci. One of the two *pgl*B genes, termed *pgl*B1, is required for *in vitro* N-glycosylation of peptides (Jervis *et al*., 2010). In this thesis I present data on the role of further *pgl* gene orthologues and previously uncharacterized genes in *H. pullorum* N-glycosylation. Furthermore I have also identified a number of *H. pullorum* glycoproteins and provide data comparing N-glycosylation processes in *C. jejuni* and *H. pullorum*. These data expand our preliminary observations on the first Helicobacter N-linked glycosylation system, and provide important information on the similarities and differences between the well characterised *C. jejuni* system and these more recently identified pathways.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the "Copyright") and s/he has given The University of Manchester the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.

ii. Copies of this thesis, either in full or in extracts, may be made **only** in accordance with the regulations of the John Rylands University Library of Manchester. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

iii. The ownership of any patents, designs, trade marks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and exploitation of this thesis, the Copyright and any Intellectual Property Rights and/or Reproductions described in it may take place is available from the Head of School of (*insert name of school*) (or the Vice-President) and the Dean of the Faculty of Life Sciences, for Faculty of Life Sciences' candidates.

Symbols and Abbreviations

Symbol	Definition
Ω	Ohms
ATCC	American type culture collection
AMBIC	Ammonium bicarbonate
APS	Ammonium persulphate
bac	bacillosamine (2,4-diamino-2,4,6-trideoxy-d-glucose)
bp	Base pairs
Cm	Chloramphenicol
DHB	2,5-dihydroxybenzoic acid
E	Erythromycin
ER	Endoplasmic Reticulum
FAM	6-carboxyfluorescein
FITC	fluorescein isothiocyanate
g	Gravitational acceleration
Gal	Galactose
GalNAc	N-acetylgalactosamine
GalNAcAN	2-acetamido-2-deoxy-D-galacturonamide
GlcNAc	N-acetylglucosamine
Glc	Glucose
Hex	Hexose
HexA	Hexuronic acid
HexNAc	N-acetylhexuronic acid
Kn	Kanamycin
LB	Luria-Bertani
LPS	lipopolysaccharide
MALDI	Matrix Assisted Laser Desorption Ionisation
Man	Mannose
NCTC	National collection of type cultures

OST	Oligosaccharyltransferase
OD	Optical density
PBS	Phosphate buffered saline
PBST	PBS with 0.1% Tween
PCR	Polymerase chain reaction
rpm	Revolutions per minute
RT	Reverse Transcriptase
SBA	Soybean agglutinin
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
TEMED	Tetramethylethylenediamine
T4SS	Type IV Secretion Systems
UDP	Uridine diphosphate
Und-PP	Undecaprenyl-Pyrophosphate
wt	wild type

Acknowledgements

I would like to acknowledge Dr. Dennis Linton who provided me with the opportunity to conduct this PhD research project and for his help with the writing process and career advice over the last three years. I also thank Professor Ian Roberts for his advice and help throughout my doctoral training period.

I would like to thank Dr Adrian Jervis and greatly appreciate all his help and training in the laboratory and advice in the analysis of data obtained throughout my postgraduate training.

I also acknowledge everyone in the Biomolecular Analysis core Facility at the University of Manchester for their help and training in Mass Spectrometry which formed a large part of this work.

I acknowledge the BBSRC as the source of funding for this studentship and thank everyone in the Microbiology department for providing support and friendship during this project, in particular Kathryn Bailey and Elizabeth Lord for their technical support over the last couple of years and Dr Jonathan Butler for his technical advice.

I also acknowledge Professor Brendan Wren and members of his laboratory in the London School of Hygiene and Tropical Medicine for their collaboration and advice.

Finally I would like to thank my family for being there for me and giving me the encouragement to embark on this postgraduate degree and my partner Paul for his support and motivation over the last few years.

Chapter 1

General Introduction

1.0 Introduction

Protein glycosylation is the process of adding a glycan structure to a protein. The glycan structure can be linked to an asparagine residue (N-linked glycosylation) or serine and threonine residues (O-linked glycosylation). The N-linked glycosylation of proteins was thought to be limited to Eukaryotes and Archaea until relatively recently when N-linked protein glycosylation was described in a member of the Epsilon Proteobacteria, *Campylobacter jejuni* (Szymanski *et al*., 1999; Linton *et al*., 2002; Wacker *et al.,* 2002; Young *et al.,* 2002; Linton *et al*., 2005). This introduction will review N-glycosylation systems in Eukaryotes and Archaea and will focus on N-glycosylation in Bacterial species.

1.1 Eukaryotic Protein Glycosylation

Carbohydrates in Eukaryotes are attached to a peptide backbone via N- or O-linkages. In the case of N-linked systems, glycan structures are usually attached to proteins at an N-X-S/T consensus sequence (where X represents any amino acid except proline, N represents asparagine, S represents serine and T represents threonine), via the amide nitrogen of the asparagine residue. O-linked systems generate glycans attached to serine and threonine residues via hydroxyl groups (Szymanski and Wren, 2005; Upreti *et al*., 2003; Benz and Schmidt, 2002). The Golgi complex is the main site for eukaryotic O-linked protein glycosylation.

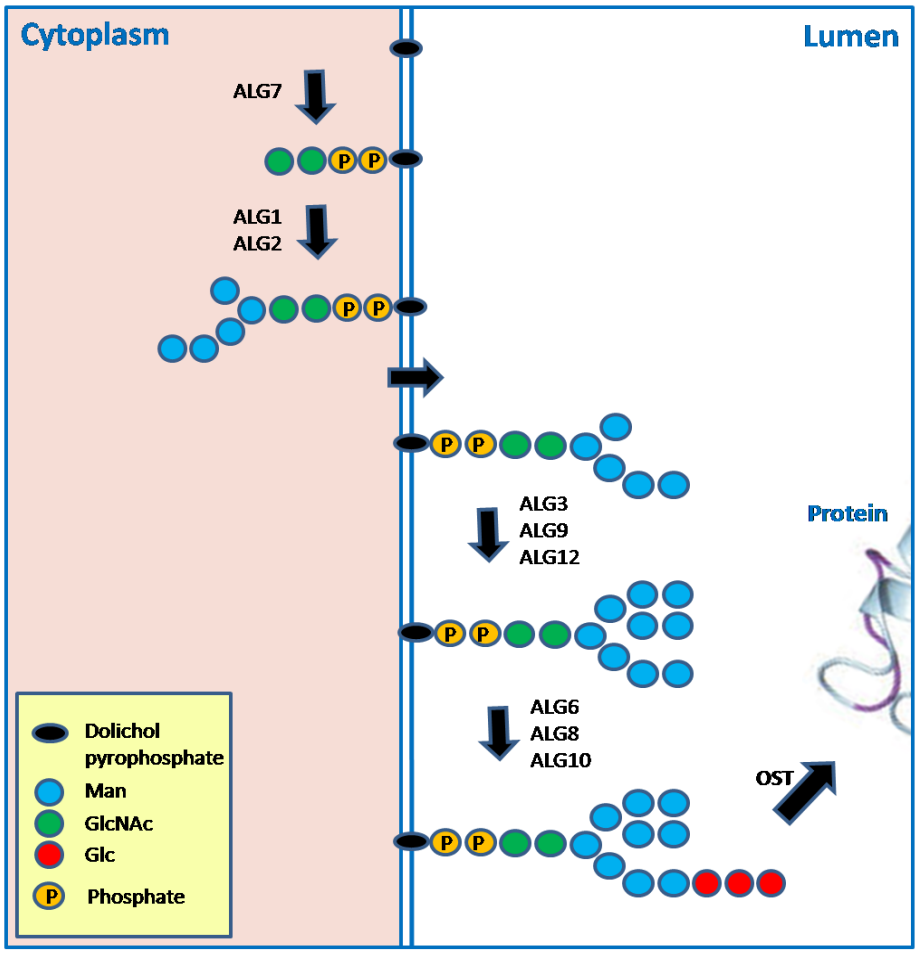
N-linked glycosylation is the most frequent protein modification in Eukaryotes and takes place across the Endoplasmic Reticulum (ER) membrane. In this process, an oligosaccharyltransferase (OST) complex catalyzes the transfer of a pre-assembled oligosaccharide from the lipid carrier dolichyl pyrophosphate to an asparagine residue on a polypeptide chain that is being translocated into the lumen of the rough ER (Figure 1.1). The oligosaccharide is assembled at the cytoplasmic side of the rough ER membrane as a heptasaccharide containing one branched sugar, flipped across the membrane into the lumen and then further processed to the tetradecasaccharide, $Glc_3Man_9GlcNAc_2$ (Burda and Aebi, 1999). The OST complex in the yeast *Saccharomyces cerevisiae* was shown to be made up of 8 different

subunits which are similar to those purified from higher Eukaryotes (Silberstein and Gilmore, 1996).

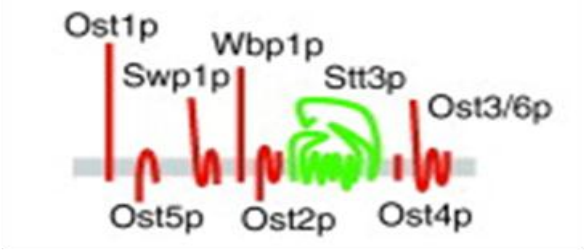
One subunit of the eukaryotic OST complex is the staurosporine- and temperature-sensitive yeast protein 3 (STT3). This protein contains the amino acid sequence WWDYG which is thought to be the active site necessary for transfer of the oligosaccharide to the polypeptide chain (Lennarz, 2007). STT3 protein homologues containing this conserved catalytic domain are found in many different species from *Homo sapiens* to Yeast and are characteristically large proteins with up to 12 potential membrane spanning regions.

Figure 1.1 N-linked Protein Glycosylation in Eukaryotes (Szymanski *et al.*, 2003). (A) Model of the Eukaryotic N-linked protein glycosylation. Two N-acetylglucosamine (GlcNAc) and five Mannose sugars (Man) are assembled on the membrane-bound lipid carrier dolichol pyrophosphate at the cytoplasmic side of the rough Endoplasmic Reticulum (ER) membrane. The heptasaccharide structure is flipped into the luminal side and further processed, by the addition of Glucose (Glc) and Man residues, to create the tetradecasaccharide and transferred to the amide nitrogen of asparagine within the N-X-S/T sequon of the protein by the OST. (B) A representation of the subunits of the *S. cerevisiae* OST complex taken from Schwarz and Aebi (2011).

(A)



(B)



1.2 Archaeal Protein Glycosylation

The earliest examples of prokaryotic protein glycosylation were found in Archaea and the first chemically characterized example of an archaeal glycoprotein was the cell surface (S-layer) protein of *Halobacterium halobium* (Mescher and Strominger, 1976; Benz and Schmidt, 2002). Halobacterial glycoproteins are involved in maintaining the structural integrity of the rod-shaped cells (Mescher and Strominger, 1975; Upreti *et al*., 2003). Although O-linked protein glycosylation is thought to occur in Archaea, there have been relatively few studies and this review will focus on N-linked protein glycosylation systems.

N-linked protein glycosylation systems have been identified in archeal species such as *Methanococcus voltae* and *Haloferax volcanii*, in which *agl* genes encode enzymes involved in glycan assembly (Voisin *et al*., 2005; Yurist-Doutsch *et al*., 2008b). In *M. voltae* and *H. volcanii* the STT3 homologue AglB appears to act as a single subunit OST and deletion of the AglB encoding gene disrupts N-linked protein glycosylation (Chaban *et al*., 2006; Abu-Qarn *et al*., 2007). As with the Eukaryotic OST, these Archeal OSTs also require the peptide acceptor sequence N-X-S/T (Abu-Qarn and Eichler, 2006; Chaban *et al*., 2006). The crystal structure of an OST from *Pyrococcus furiosus* was recently determined and the WWDYG catalytic site identified as the site of N-glycosylation (Igura *et al*., 2007).

The halophile *Haloferax volcanii* was originally isolated from the Dead Sea and has an S-layer for protection in this extreme environment (Sumper *et al*., 1990). This S-layer protein is N-glycosylated with a pentasaccharide composed of a Hexose (Hex), two Hexuronic acids (HexA), a methyl ester of HexA and a Man residue (Abu-Qarn *et al*., 2007; Magidovich *et al*., 2010). The glycan structure is assembled at the inner surface of the cell membrane on a dolichol phosphate carrier. This Archeal system uses more than one dolichol phosphate carrier (Guan *et al*., 2010). The glycosyltransferases AglJ, AglG, AglI and AglE sequentially transfer the first four sugars (Hex, HexA, HexA and the HexA methyl ester) to the dolichol phosphate carrier whilst AglD transfers the Man sugar to a separate pool of dolichol phosphate carriers (Abu-Qarn *et al*., 2007; Abu-Qarn *et al*., 2008; Yurist-Doutsch *et al*., 2008a;

Kaminski *et al*., 2010). The OST, AglB, then transfers these tetrasaccharide and monosaccharide glycans to the S-layer protein to produce a glycoprotein decorated with a pentasaccharide (Abu-Qarn *et al*., 2007; Guan *et al*., 2010). A halophile isolated from the Dead Sea, *Haloarcula marismortui* has an S-layer protein glycosylated with the same pentasaccharide but in this system the complete glycan is assembled on a single dolichol phosphate carrier before being transferred to the S-layer protein by the OST (Calo *et al*., 2011).

The N-linked tetrasaccharide glycan of *Methanococcus voltae* not only decorates S-layer protein but also flagellin (Voisin *et al*., 2005). The mechanism of this N-glycosylation system has not been determined but it is thought that the glycosyltransferase AglH is required for attachment of a GlcNAc to the dolichol phosphate carrier at the cytoplasmic face of the cell membrane (Shams-Eldin *et al*., 2008), and AglA transfers the third sugar residue, a modified mannuronic acid with the amino acid threonine covalently attached at position 6 (Chaban *et al*., 2006). Two further glycosyltransferases, AglC and AglK, have been implicated in the biosynthesis of this tetrasaccharide glycan and are thought to be required for the addition of a diacetylated glucuronic acid, but their precise roles have not yet been determined (Chaban *et al*., 2009).

1.3 Bacterial O-linked Protein Glycosylation

Protein glycosylation systems in which glycans coupled to S/T residues of target proteins have been discovered in several bacterial species including *C. jejuni*, *Helicobacter pylori*, *Streptococcus sanguis*, *Pseudomonas aeruginosa* and *Neisseria meningitides* (Erickson and Herzberg, 1993; Castric, 1995; Virji, 1997; Szymanski *et al*., 1999). Many O-linked protein glycosylation systems are involved in pilin glycosylation. For example, the pili of *N. meningitidis* are glycosylated with a trisaccharide consisting of a diacetamidotrideoxyhexose linked to two Galactose (Gal) residues (Stimson *et al*., 1995; Power and Jennings, 2003). The pilin protein of *Neisseria gonorrhoea* is also glycosylated at the same region but with a disaccharide containing Gal and GlcNAc. At least eight different genes, named *pgl*, are thought to encode this protein glycosylation system (Power and Jennings, 2003).

The flagella of *P. aeruginosa* are O-glycosylated and a locus of fourteen genes adjacent to the flagellin biosynthesis genes encode the glycosylation apparatus (Arora *et al*., 2001). Genes named *flm* that encode proteins associated with flagella glycosylation have been described in *Caulobacter crescentus,* (Leclerc *et al*., 1998) *H. pylori* (Josenhans *et al*., 2002) and *Aeromonas caviae* (Tabei *et al*., 2009). The O-linked glycosylation system in *C. jejuni* is involved in flagellar glycosylation where flagellins are modified with several monosaccharide analogues of the nine-carbon sugar pseudaminic acid (Thibault *et al*., 2001). The *C. jejuni* genome contains a flagellar glycosylation cluster of around fifty genes which includes genes encoding the flagellin filament structural proteins. More recently, OSTs involved in O-linked protein glycosylation were identified in *Vibrio cholerae* and *Burkholderia thailandensis* and named PglL(Vc) and PglL(Bt) respectively due to similarities with the *N. meningitidis* PglL (Gebhart *et al*., 2012). O-linked protein glycosylation was also recently described in *Francisella tularensis* in which the OST, named PglA, is involved in pilin glycosylation with a pentasaccharide glycan of the structure N-acetylhexuronic acid (HexNAc)-HexNAc-Hex-Hex-HexNAc-HexNAc (Egge-Jacobsen *et al*., 2011).

1.4 Bacterial Cytoplasmic N-glycosylation Systems

Recently an N-linked protein glycosylation system was discovered in the Gram negative pathogenic bacterium, *Haemophilus influenzae* (Grass *et al*., 2003*).* In this system, Hex residues are transferred to a glycoprotein (HMW1) in the cytoplasm rather than at the periplasmic face of the inner membrane (Gross *et al*., 2008). The glycosyltransferase HMW1C transfers Glc or Gal to asparagine residues found within an N-X-S/T sequon (Grass *et al*., 2003; Grass *et al*., 2010). Modification of the HMW1 adhesin prevents its premature degradation and promotes localization to the cell surface enabling HMW1 mediated adherence to human epithelial cells (Grass *et al*., 2003).

A HMW1C homologue from *Actinobacillus pleuropneumoniae* (ApHMW1C) can complement a deficiency in the *H. influenzae* HMW1C and mediate HMW1 glycosylation and adhesive activity of cells (Choi *et al*., 2010). In a

glycosyltransferase assay, ApHMW1C was able to transfer Glc (and Gal with less efficiency) but not Man, GlcNAc or N-acetylgalactosamine (GalNAc) to asparagine sites within N-X-S/T sequons of the *H. influenzae* glycoprotein, HMW1 (Choi *et al*., 2010). *In vitro,* ApHMW1C required the N-X-S/T sequon for peptide glycosylation, but did not require the presence of an acidic residue at the N-2 position which is essential for N-glyosylation in *C. jejuni* as discussed in section 1.6 (Schwarz *et al*., 2011a).

The open reading frame downstream of and in the same transcriptional orientation as the ApHMW1C gene encodes a putative glycosyltransferase and a mobility shift was observed when the purified gene product was incubated with HMW1 modified by the addition of a Glc residue by ApHMW1C (Schwarz *et al*., 2011a). This shift was observed in the presence of UDP-Glc but not UDP-Gal, UDP-GalNAc or UDP-GlcNAc (Schwarz *et al*., 2011a). Mass Spectrometry revealed the ApHMW1C dependent transfer of between two and six Glc moieties to HMW1 therefore it is thought that this putative glycosyltransferase encoding gene located directly downstream of the ApHMW1C gene encodes a polymerizing glucosyltransferase that elaborates the product of the ApHMW1C reaction (Schwarz *et al*., 2011a).

1.5 The *C. jejuni* N-linked Protein Glycosylation System

Experiments have shown that the lectin soybean agglutinin (SBA), which is known to interact with terminal GalNAc, binds strongly to a number of *C. jejuni* proteins (Linton *et al*., 2002). Two of these putative *C. jejuni* glycoproteins were purified, identified and analysed by Mass Spectrometry (Linton *et al*., 2002). One was PEB3, a previously identified antigenic surface protein (Pei *et al*., 1991) and the other was previously uncharacterised and therefore named Campylobacter glycosylation protein A or CgpA. The N-linked glycosylation of PEB3 was subsequently shown to be PglB dependent (Linton *et al*., 2005). A further 22 *C. jejuni* N-linked glycoproteins were identified using soybean lectin affinity chromatography and two-dimensional gel proteomics (Young *et al*., 2002). More recently Scott *et al*. identified 75 N-linked glycosylation sites in over 50 *C. jejuni* glycoproteins (Scott *et al*., 2011). Mass Spectrometry and nuclear magnetic resonance spectroscopy were

used to structurally characterize N-linked glycan from purified *C. jejuni* glycoproteins (Young *et al*., 2002; Scott *et al*., 2011). The structure consists of a heptasaccharide containing a single bacillosamine (2, 4-diacetamido-2, 4, 6-trideoxyglucopyranose) residue, five GalNAc residues and a single branching Glc residue. This heptasaccharide glycan is assembled sequentially on the lipid carrier Undecaprenyl-Pyrophosphate (Und-PP) at the cytoplasmic side of the bacterial inner membrane and flipped into the periplasm where it is transferred to proteins by the action of PglB (Linton *et al*., 2005).

1.6 The Bacterial OST and Amino Acid Sequon Specificity

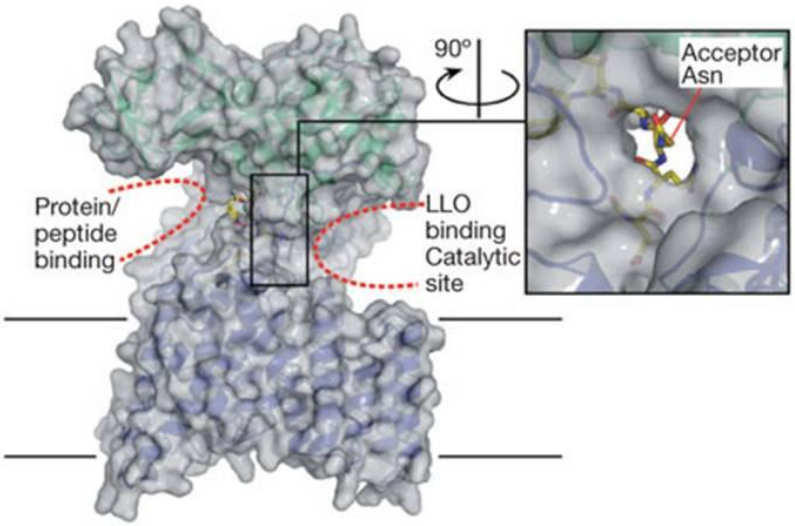
In all characterised *C. jejuni* N-linked glycoproteins, the site of glycosylation is an asparagine residue within an N-X-S/T sequon, but not all such sites in a glycoprotein are glycosylated (Wacker *et al*., 2002; Young *et al*., 2002). The *C. jejuni* glycoprotein AcrA was identified by comparing two dimensional electrophoresis gels from wild type and *pgl*B mutant *C. jejuni* (Young *et al*., 2002). In this *C. jejuni* glycoprotein there are five potential glycosylation sites but only two are glycosylated (Wacker *et al*., 2002; Nita-Lazar *et al*., 2005). It is now known that although the N-X-S/T sequence is required for N-glycosylation, it is not sufficient for glycosylation in *C. jejuni* (Nita-Lazar *et al*., 2005).

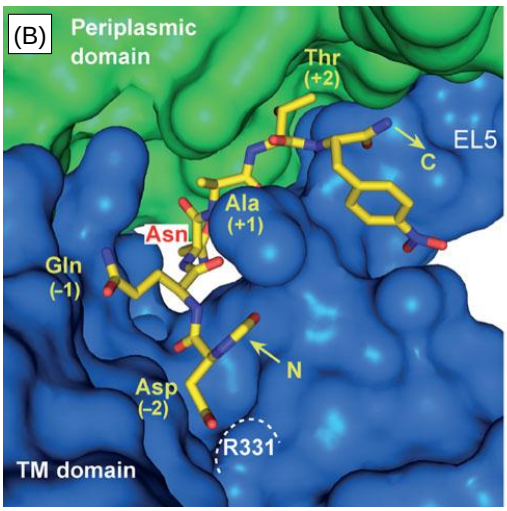
The N-X-S/T sequon required for N-glycosylation in *C. jejuni* was expanded to D/E-X1-N-X2-S/T (Kowarik *et al*., 2006). More recent studies investigated the specificity of PglB for the acidic residues at the N-2 region, the hydroxyamino acid residues at the N+2 position, and different amino acids at the X1 and X2 positions of the sequon (Chen *et al*., 2007). Although PglB glycosylation occurred with any amino acid except proline at the X1 and X2 positions, the enzyme favours alanine, serine and positively charged residues at position X2 and aromatic and amido residues at position X1 (Chen *et al*., 2007). Threonine was preferred over serine at position N+2 and aspartic acid was favoured over glutamic acid at position N-2 (Chen *et al*., 2007).

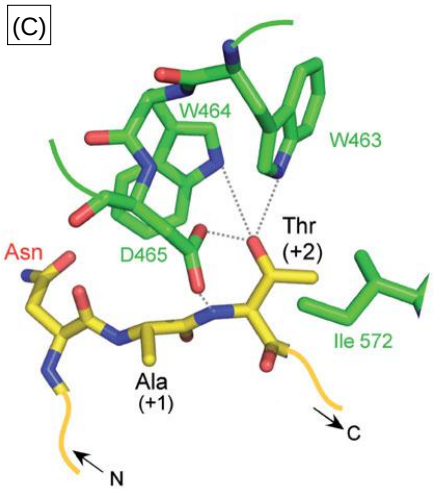
The X-ray structure of PglB from *Campylobacter lari*, which has 56% amino acid sequence identity with *C. jejuni pgl*B, was recently determined by Lizak *et al.* (2011) and provides structural explanation for the specificity of PglB for the D/E-X-N-X-S/T acceptor sequon. Lizak *et al.* (2011) co-crystalized the PglB with the peptide DQNATF which is readily glycosylated by *C. jejuni* PglB. The PglB structure contained a transmembrane domain containing 432 amino acids with thirteen transmembrane segments connected by cytoplasmic or external loops and a 279 amino acid periplasmic domain (Lizak *et al*., 2011). When the peptide is bound, PglB forms two large cavities at opposite sides of the protein above the membrane surface (Figure 1.2 A). One cavity allows access for an acceptor protein or peptide and the other contains the catalytic site and is probably where the lipid linked oligosaccharide binds (Lizak *et al*., 2011). The side chain of the acceptor asparagine connects the two sites. Figure 1.2 B is taken from Lizak *et al.* (2011) and represents the structure of the peptide embedded within the *C. lari* PglB. The majority of the peptide surface is embedded in the PglB structure between the transmembrane and periplasmic domain and as the binding site is small, only flexible loops of proteins are able to fit in (Lizak *et al*., 2011). The negative charge of the acceptor peptide at the N-2 position allows interaction with R331 via a salt bridge, resulting in tighter binding of the peptide, thus explaining a preference for aspartic or glutamic acid at this position (Lizak *et al*., 2011). Figure 1.2 C (Lizak *et al*., 2011) illustrates the interaction between the threonine residue of the acceptor peptide and the WWD of the WWDYG catalytic motif. Threonine also forms Van der Waals interactions with I572 of an MXXI motif but serine would not interact with this I572, thus explaining why threonine is preferred at the N+2 position and is more efficient than serine (Lizak *et al*., 2011).

Figure 1.2 Structural representation of *C. lari* PglB and its interaction with acceptor peptide (Images taken from Lizak *et al*., 2011). (A) A structural model of the *C. lari* PglB, containing two cavities at opposite sides (indicated by red dashed lines) where the lipid linked oligosaccharide and the acceptor peptide are able to bind. The cavities are connected by a tunnel which accommodates the acceptor asparagine. The acceptor peptide is shown in stick representation and coloured yellow. (B) Periplasmic and transmembrane domains of *C. lari* PglB are green and blue and the acceptor peptide is in stick representation with N and C denoting amino and carboxy termini. (C) The threonine residue of the acceptor peptide interacting with the WWD catalytic motif of *C. lari* PglB. Dashed lines indicate hydrogen bonds formed between the threonine β-hydroxyl group and the WWD motif.

(A)







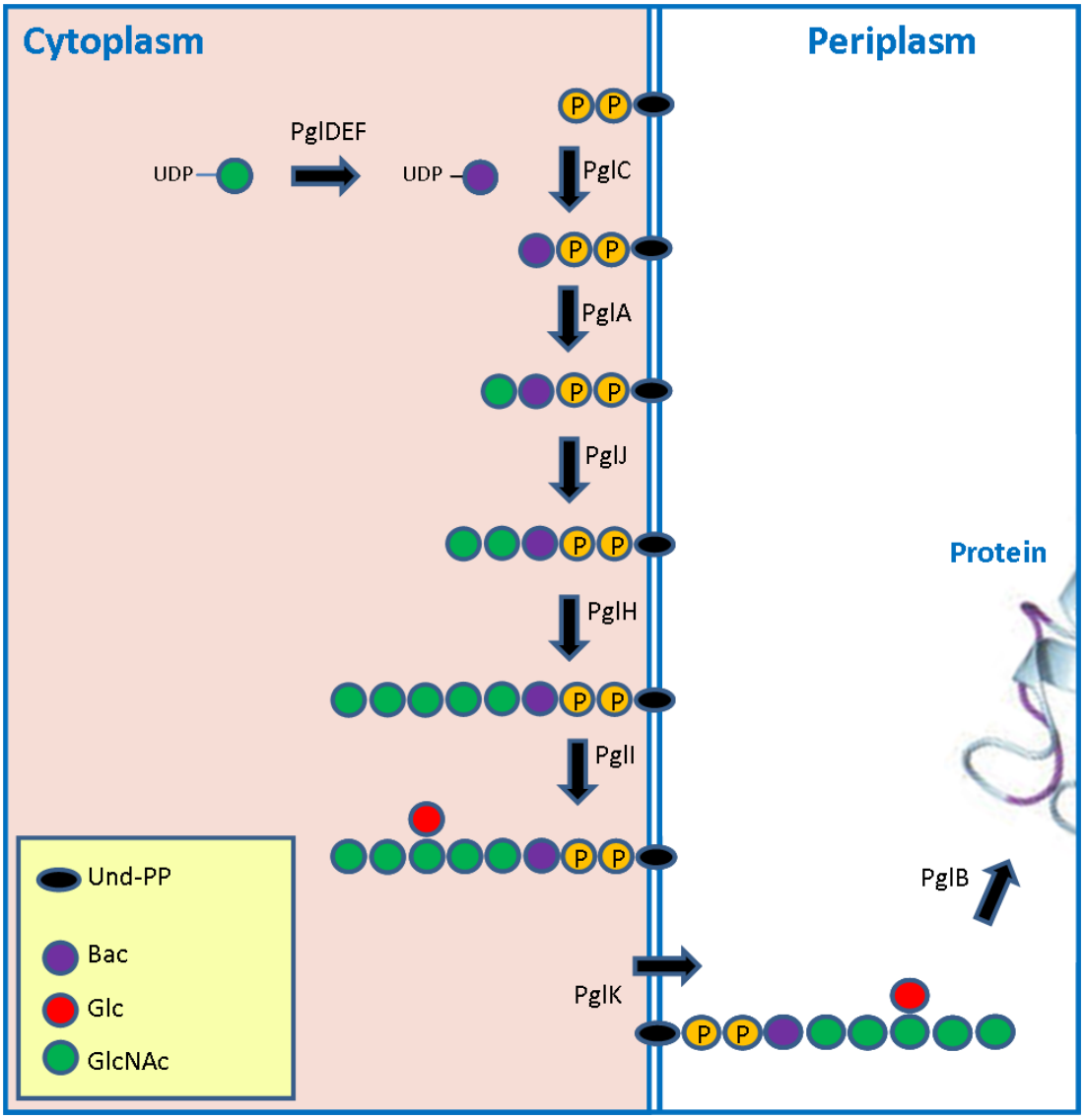
1.7 *C. jejuni* Glycosyltransferases and Assembly of the Heptasaccharide Glycan

The genes *pgl*A, *pgl*C, *pgl*H, *pgl*I and *pgl*J all encode glycosyltransferases. PglC is required for the transfer of the first sugar bacillosamine (bac) to the carrier lipid and PglA then adds the first GalNAc sugar (Linton *et al*., 2005). The genes *pgl*H, *pgl*I and *pgl*J partially overlap with each other and are required for the addition of the terminal GalNAc residues and the branching Glc residue which completes the lipid linked heptasaccharide structure (Linton *et al*., 2005). The glycosyltransferase PglJ transfers a single GalNAc residue, PglH then transfers the remaining three GalNAc residues to create a hexasaccharide and PglI transfers the branching Glc residue (Glover *et al*., 2005; Weerapana *et al*., 2005). We now have a complete understanding of how the *C. jejuni* N-linked heptasaccharide structure is assembled on a lipid carrier at the cytoplasmic side of the inner membrane (Figure 1.3).

The *C. jejuni pgl* glycosylation locus also contains genes involved in the biosynthesis of sugars of the N-linked heptasaccharide glycan (Linton *et al*., 2005). PglD is an acetyl transferase, PglE is an aminotransferase and PglF is a dehydratase. Together, these enzymes are involved in the formation of bac which is the first sugar of the N-glycan (Szymanski *et al*., 2003). Biochemical evidence has confirmed this role and indicates that PglD catalyzes the final step in the conversion from GlcNAc (Olivier *et al*., 2006).

Before the heptasaccharide can be transferred onto proteins, the lipid linked structure must be flipped from the cytoplasm into the periplasm. ATP-binding cassette (ABC) transporters are transmembrane proteins that use ATP to translocate substrates across a membrane. In *C. jejuni* the proposed ABC transporter, PglK, is involved in flipping the assembled lipid linked heptasaccharide across the inner membrane (Szymanski *et al*., 2003). Experiments performed in *E. coli* lacking all known flippases have confirmed that an ATP binding domain in the C-terminus of PglK is essential for flippase activity (Alaimo *et al*., 2006).

Figure 1.3 Model for assembly of the lipid-linked heptasaccharide of the *C. jejuni* N-glycosylation system. GlcNAc is converted to bac by the *C. jejuni* PglFED enzymes and attached to the membrane-bound lipid carrier, Und-PP (shown as black ovals within the membrane), by PglC. The glycosyltransferase, PglA, transfers the first GalNAc residue and PglH and PglJ transfer the next four. PglI catalyses the addition of the Glc branch and the whole heptasaccharide structure is flipped across the periplasmic membrane by PglK. Finally, the OST, PglB, transfers the entire structure to an asparagine residue of a polypeptide chain. Image adapted from Szymanski *et al*., 2003.



1.8 Function of *C. jejuni* N-linked Glycoproteins

The *C. jejuni pgl* system is involved in the glycosylation of a wide variety of glycoproteins (Scott *et al*., 2011) and therefore may affect a range of cellular functions. A knock out mutation in the *C. jejuni* OST-encoding gene *pgl*B leads to a significant reduction in adherence to and invasion of intestinal cells *in vitro* and reduced ability to colonize the intestinal tract of mice (Szymanski *et al*., 2002). Mutation of other *pgl* genes such as *pgl*H, *pgl*I and *pgl*J in *C. jejuni* strains also resulted in reduced attachment to and invasion of cells and reduced colonization efficiency in a chick model (Karlyshev *et al*., 2004). It is therefore thought that some of the proteins glycosylated by the *pgl* system play important roles in adherence to intestinal cells and therefore intestinal colonization.

As previously described, Young *et al.* (2002) identified several *C. jejuni* glycoproteins. More recently, 22 strains of *C. jejuni* were constructed, each with a knock-out mutation in a gene encoding an N-linked glycoprotein (Kakuda and DiRita, 2006). Strains with a mutation in Cj1496c and Cj1565c, which encodes a paralyzed flagellin protein (PflA), invaded Eukaryotic cells with at least 10-fold less efficiency than wild type strains (Kakuda and DiRita, 2006). Chick colonization experiments revealed the significantly reduced colonization levels of the Cj1496c mutant most likely due to an adherence defect.

C. jejuni strain 81-176 is one of the best characterized strains of *C. jejuni*, and was shown to be virulent in a human volunteer study (Black *et al*., 1988). This strain has two plasmids, one of which, pVir, is involved in both virulence and natural competence (Bacon *et al*., 2000). This pVir plasmid contains several genes encoding homologues of proteins involved in Type IV Secretion Systems (T4SSs) (Bacon *et al*., 2000). These secretion systems are involved in DNA export, conjugation, and protein secretion and have been described in many Bacteria (Bacon *et al*., 2002). In *C. jejuni* 81-176 the T4SS contains at least one glycoprotein, VirB10, that is glycosylated at two N-X-S/T sites by the *pgl* system (Larsen *et al*., 2004). Mutation of the *pgl* locus and also lack of glycosylation at one of these sites both result in a natural transformation defect (Larsen *et al*., 2004). The *C. jejuni* N-linked protein

glycocosylation system is therefore a general system of protein glycosylation that is essential for the glycosylation of many proteins with diverse functions including cell adherence/colonisation and competence.

1.9 Exploiting the *C. jejuni* N-linked Protein Glycosylation System

Conjugate vaccines consist of capsular or lipopolysaccharide (LPS) antigens from pathogenic Bacteria covalently bound to carrier proteins and are often more efficient than polysaccharides alone as they can induce a long lasting T-lymphocyte dependent immunological memory (Ihssen *et al*., 2010). Examples of conjugate vaccines include the *H. influenzae* type b capsular polysaccharide (Anderson, 1983), the *Shigella dysenteriae* type O1 polysaccharide, the *Salmonella flexneri* type 2a polysaccharide and the *Shigella sonnei* polysaccharide which were conjugated to a carrier protein such as *Pseudomonas aeruginosa* exoprotein A (Taylor *et al*., 1993).

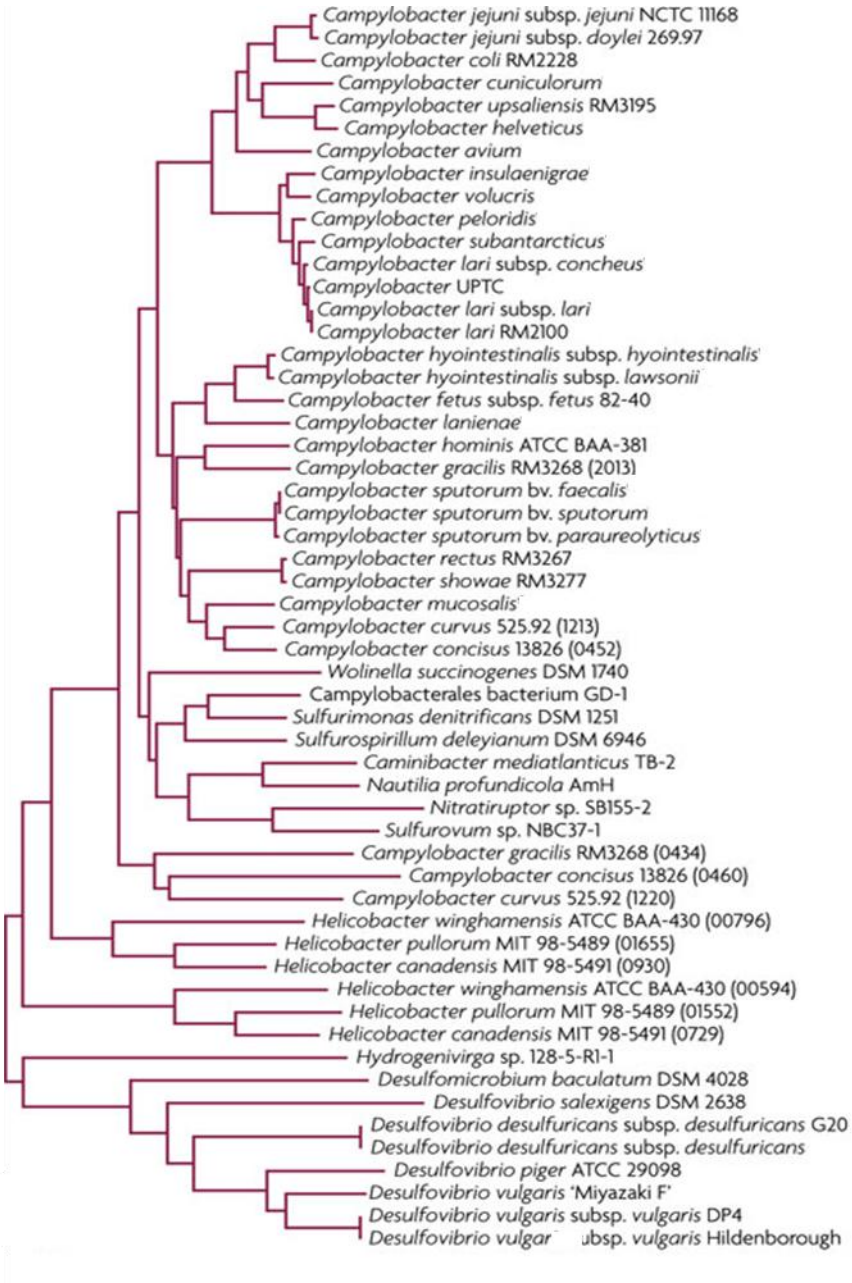
A major drawback in the production of conjugate vaccines is that it is a complex multistep process involving purification of the carrier protein and LPS antigen from cultivated bacterial strains, cleavage and purification of the LPS from the lipid A core, coupling of the LPS to the carrier protein and further purification (Ihssen *et al*., 2010). These processes can be time consuming and expensive therefore alternative methods of producing glycoconjugate vaccines were sought. The *C. jejuni* locus contains all of the genes required for the assembly of a lipid linked heptasaccharide glycan, the flipping of this structure across the membrane and its transfer to the asparagine residue of selected glycoproteins (Wacker *et al*., 2002). When this entire locus is cloned and expressed in *E. coli* along with a gene encoding a *C. jejuni* glycoprotein, the glycoprotein is glycosylated with the *C. jejuni* heptasaccharide glycan (Wacker *et al*., 2002). Current research is now focussed on exploiting this N-glycosylation system to produce specific glycoconjugates in a more timely and cost effective way (Ihssen *et al.,* 2010).

Feldman *et al.* (2005) demonstrated that *C. jejuni* PglB had a relaxed specificity toward the LLO as it can transfer structures that differ in sugar composition. This is in contrast to the eukaryotic OST which requires a terminal α1-2 linked Glc residue

for efficient glycosylation (Feldman *et al*., 2005). The *C. jejuni* protein glycosylation system was able to glycosylate the *C. jejuni* AcrA protein with *E. coli* O7 or O16 polysaccharides which have a GlcNAc as the reducing end sugar, and can also transfer *E. coli* O16 and *P. aeruginosa* O11 polysaccharide, with a FucNAc reducing end sugar, to AcrA (Feldman *et al*., 2005). PglB can therefore transfer glycan structures with GlcNAc, GalNAc or FucNAc as the reducing sugar. The *E. coli* K30 capsule and *Salmonella enterica* LT2 O antigen contain a Gal residue at the reducing end and could not be transferred to AcrA protein indicating that the acetamido group at the C-2 position of the reducing end sugar is important for PglB activity (Wacker *et al*., 2006). This acetamido group is thought to play a role in the catalytic mechanism for the transfer of the LLO to the asparagine residue of a glycoprotein (Lizak *et al*., 2011).

Based on the *C. jejuni* N-glycosylation system in *E. coli*, Ihssen *et al.* (2010) have recently established an efficient process for the *in vivo* production of two glycoconjugates where the carrier protein is either AcrA glycoprotein from *C. jejuni* or exotoxin A from *P. aeruginosa* and the polysaccharide is the *S. dysenteriae* serotype 1 O antigen. As with other polysaccharides that can be transferred by *C. jejuni* PglB, *S. dysentariae* O1 antigen contains a GlcNAc sugar at the reducing end (Falt *et al*., 1995). Not all polysaccharides have a C-2 acetamido reducing end sugar therefore the *C. jejuni* system may not be suitable for the biosynthesis of specific glycoconjugates. N-Glycosylation systems have recently been discovered in other species such as *Wolinella succinogenes* (Dell *et al*., 2010) and *H. pullorum* (Jervis *et al*., 2010) that are yet to be fully characterized and may have PglB's that differ in specificity to the *C. jejuni* PglB. Figure 1.4 is a phylogenetic tree showing predicted OSTs from sequenced Delta and Epsilon Proteobacteria species. Most of these species have only one predicted OST but a few, such as *Campylobacter gracilis* and *Helicobacter pullorum* have two.

Figure 1.4 Phylogenetic tree of PglB proteins from Delta and Epsilon Proteobacteria. The phylogenetic tree was constructed by multiple sequence alignment with the neighbour-joining method using MEGA 4 software (Northaft and Szymanski, 2010). For PglB protein sequences from genomes encoding multiple putative PglB orthologues, the number of the open reading frame is indicated in brackets after the strain name. (Image adapted from Northaft and Szymanski, 2010)

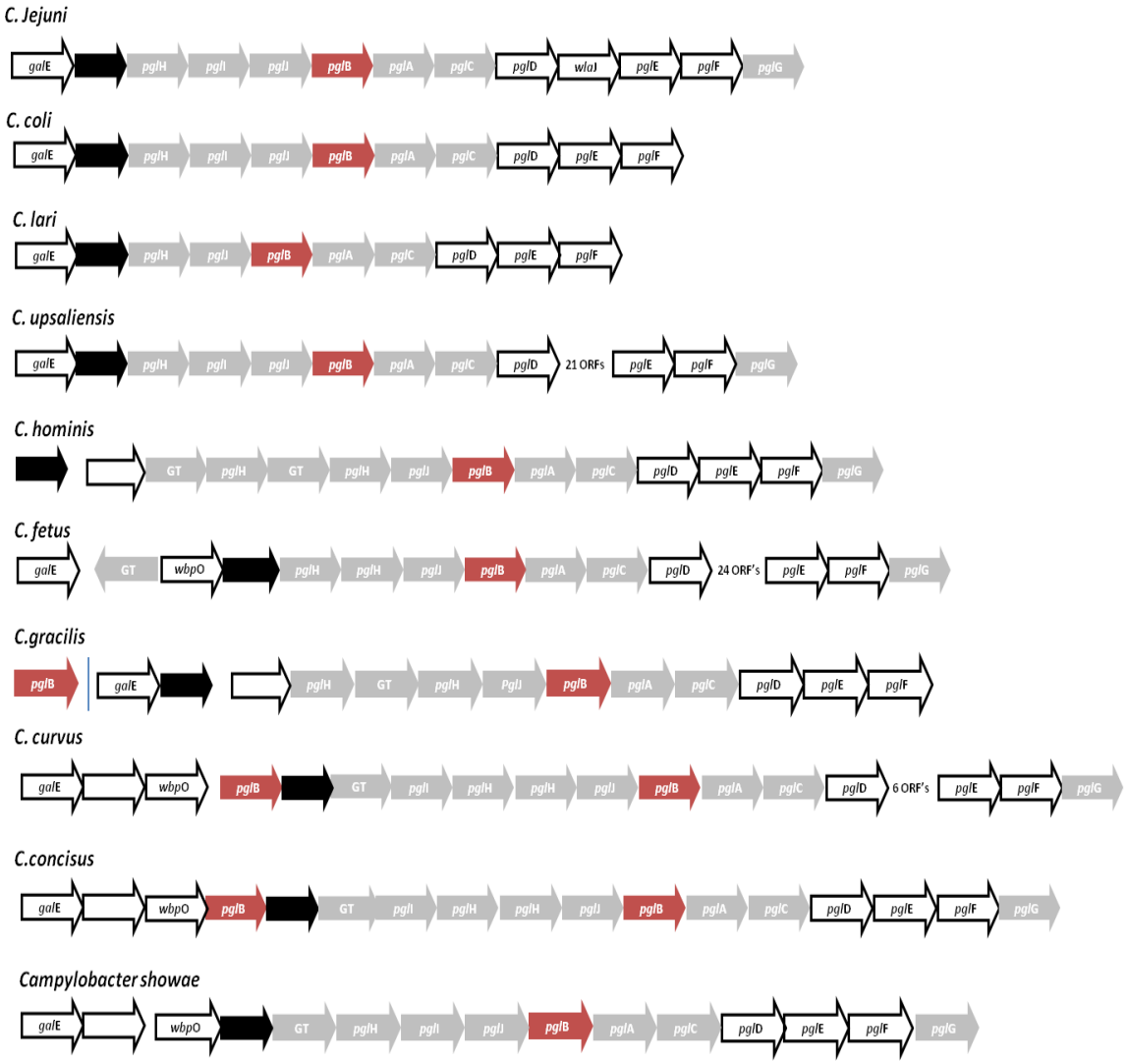


1.10 Campylobacter N-linked Protein Glycosylation (*pgl*) Genes

Figure 1.5 is a schematic diagram of *pgl* genes from different *Campylobacter* species. These genes encode putative glycosyltransferases and proteins involved in the biosynthesis of sugars required for N-linked protein glycosylation in *C. jejuni* and related bacterial species. The similarity of other Campylobacter species *pgl* loci to that of *C. jejuni* varies. The *Campylobacter coli, Campylobacter lari and Campylobacter gracilis pgl* loci lack a *pgl*G and *C. lari, Campylobacter hominis, Campylobacter fetus* and *C. gracilis* lack a *pgl*I. The *pgl*A and *pgl*C genes are located directly downstream of the *pgl*B for each species and *pgl*DEF genes are clustered with the exception of *Campylobacter upsaliensis*, *C. fetus* and *Campylobacter curvus* which have an insertion of up to 24 open reading frames between the *pgl*D and *pgl*EF genes. The *pgl* loci from many of these *Campylobacter* species also contain putative glycosyltransferase-encoding genes not found in *C. jejuni*. Interestingly, the *pgl* loci of *C. curvus* and *Campylobacter concisus* each contain two *pgl*B genes and the genome of *C. gracilis* contains a second *pgl*B gene located outside of the *pgl* locus.

The N-glycan structure of nine different *Campylobacter* species was recently identified (Jervis *et al*., 2012). The structure of N-glycan from *C. coli*, *Campylobacter helveticus*, and *C. upsaliensis* was comparable to the N-glycan from *C. jejuni* (Jervis *et al*., 2012). The N-glycan structure of *Campylobacter hyointestinalis*, *C. fetus* and *Campylobacter lanienae* contained only four HexNAc sugars and the structure of N-glycan from *C. lari*, which does not contain a *pgl*I gene, lacked a branching Hex (Jervis *et al*., 2012). Despite variations in structure, N-glycan from all strains were composed of a reducing end 228 Da sugar with attached HexNAc and Hex sugars with the exception of *C. concisus* which also contained a 217 Da residue (Jervis *et al*., 2012).

Figure 1.5 The *pgl* loci from Campylobacter species. Genes predicted to encode proteins involved in *Campylobacter* N-glycosylation systems are indicated by arrows. Genes encoding glycosyltransferases and sugar biosynthesis genes are designated according to *C. jejuni* orthologues or GT for glycosyltransferase if there is no definite *C. jejuni* orthologue. Grey arrows indicate genes encoding putative glycosyltransferases, red arrows indicate genes encoding PglB enzymes and black arrows indicate genes encoding transporters.



1.11 The N-linked Protein Glycosylation System of *W. succinogenes*

Along with *Campylobacter* and *Helicobacter*, *Wolinella* is also a member of the Epsilon Proteobacteria and was shown to be more closely related to *H. pylori* than *C. jejuni* based on 16S rRNA phylogeny (Baar *et al*., 2003). As with *C. jejuni*, the *W. succinogenes pgl* genes encoding glycosyltransferases and sugar biosynthesis enzymes are found at a single locus with one *pgl*B orthologue (Figure 1.6). The *W. succinogenes pgl* gene locus however is quite distinct from that of *C. jejuni* and has additional glycosyltransferase encoding genes and predicted sugar biosynthesis protein encoding genes. In *W. succinogenes*, the *pgl* gene cluster contains two separate gene insertions, five genes are inserted between *pgl*A and *pgl*B and four between *pgl*H and *pgl*J. Also *pgl*H and *pgl*I have exchanged their positions (Baar *et al*., 2003). Two of the four genes between *pgl*H and *pgl*J encode for a galactosyl transferase and a glycosyl transferase which were indicative of the formation of a glycan that differs from the *C. jejuni* heptasaccharide (Baar *et al*., 2003). The *W. succinogenes pgl* locus also contains two genes encoding asparagine synthetase B (*asn*B) enzymes (recently renamed *wbp*S) which may be involved in the biosynthesis of sugars that make up the *W. succinogenes* N-glycan (see Chapter 7 for details). The whole *W. succinogenes pgl* locus was reconstituted in *E. coli* and glycosylated the *C. jejuni* glycoprotein PEB3 with a hexasaccharide containing a reducing end bac followed by three 216 Da sugar residues and a 232 Da residue, with a branching Hex attached to the first 216 Da sugar (Dell *et al*., 2010).

1.12 N-Linked Glycosylation Systems in *Helicobacter* Species

Helicobacters are naturally competent, Gram negative, microaerophilic, spiral-shaped Bacteria related to the genus Campylobacter. *H. pylori* was first isolated from a gastric biopsy in 1982 and is the best characterized species of the Helicobacter genus. Long-term *H. pylori* infection of the human stomach is associated with the development of gastric and duodenal ulcers, gastric adenocarcinoma and chronic gastritis. The genome sequence of *H. pylori* contains genes encoding an O-linked glycosylation system that is involved in flagellin glycosylation (Josenhans *et al*., 2002) but lacks homologues of the genes involved in *C. jejuni* N-linked glycosylation, in particular the N-linked OST encoding gene *pgl*B.

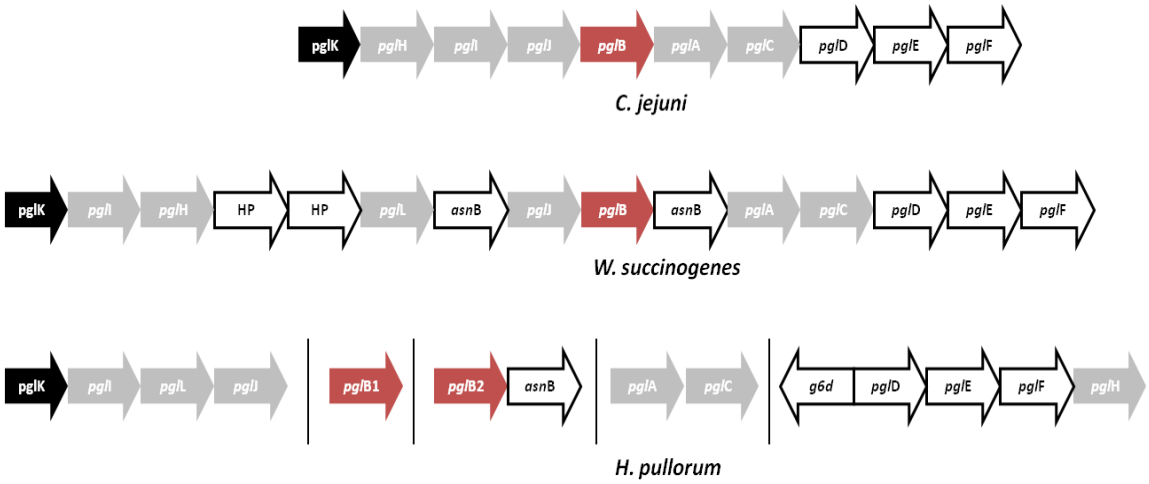
H. pullorum was first isolated from the caeca, liver and duodenum of broiler chickens, laying hens and also from human patients suffering with gastroenteritis (Stanley *et al*., 1994). *H. pullorum* resides in the host's gastrointestinal tract and although there is evidence to show that the organism may have the ability to cause human disease, this is not conclusive (Ceelen *et al*., 2005). It is possible that *H. pullorum* resides in the normal gut flora of humans without causing disease. A closely related species, *H. canadensis*, has also been isolated from humans with diarrhoea in Canada (Fox *et al*., 2000).

Analysis of genome sequence data indicates that *H. pullorum*, *Helicobacter canadensis* and the closely related *Helicobacter winghamensis* species each contain two distinct homologues of the *C. jejuni pgl*B gene (named *pgl*B1 and *pgl*B2) with a conserved WWDYG motif. These strains all have a similar arrangement of *pgl* genes which are likely involved in the synthesis and transfer of glycans but unlike *C. jejuni pgl* genes, they are found not at a single locus but as at least five separate clusters throughout the genome (Figure 1.6). As in *C. jejuni*, the *pgl*D, *pgl*E, *pgl*F and *pgl*H orthologues are clustered together as are *pgl*A and *pgl*C. The *H. pullorum pgl*I orthologue is situated downstream of an ABC transporter and is followed by a putative glycosyltransferase gene (named *pgl*L) and a *pgl*J gene. There is also a homologue of the *pgl*L gene in the *W. succinogenes pgl* N-glycosylation locus. The

*pgl*B1 gene is on a separate region of the chromosome to the other *pgl* homologues. Interestingly, *pgl*B2 is situated upstream of an *asn*B gene (renamed *wbp*S) on the *H. pullorum*, *H. canadensis* and *H. winghamensis* chromosomes and also at the *W. succinogenes* N-glycosylation locus.

Recently the *H. pullorum* N-linked glycan was identified as a pentasaccharide with a 203 Da reducing end HexNAc sugar attached to a sugar of 216 Da, two sugars of 217 Da and a non reducing end 203 Da HexNAc sugar (Jervis *et al*., 2010). This N-glycosylation system was detected *in vitro* using an OST assay in which *H. pullorum* membrane preparations were incubated with a biotinylated peptide with the sequence ADQNATAK (Jervis *et al*., 2010). Glycosylated peptide was purified and analysed by Matrix-assisted laser desorption ionization (MALDI) tandem Mass Spectrometry to identify the glycan structure (Jervis *et al*., 2010). MALDI tandem Mass Spectrometry methods are ideal for analysing oligosaccharide structures and involve the isolation of specific ion species which are then probed further to cleave glycosidic bonds between sugars and thus identify glycan structures. Experiments combining this technique with the insertional mutagenesis of *H. pullorum* genes indicated that *H. pullorum pgl*DEF genes are not required for *H. pullorum* PglB1 dependent N-glycosylation (Jervis *et al*., 2010). In *C. jejuni*, the products of these genes are involved in the biosynthesis of the sugar bac, which is not present in the *H. pullorum* pentasaccharide glycan. When asparagine or aspartic acid residues of the ADQNATAK sequence were substituted, the peptide remained unglycosylated indicating that *H. pullorum* PglB1 has the same or similar amino acid sequon requirements to the *C. jejuni* PglB (Jervis *et al*., 2010).

Figure 1.6 Gene map of *pgl* loci from *C. jejuni, W. succinogenes* and *H. pullorum*. Schematic representation of the *pgl* gene N-glycosylation locus from *C. jejuni* and *W. succinogenes* and orthologues in *H. pullorum*. Arrows representing genes are not to scale. Arrows labelled HP represent genes encoding hypothetical proteins. Genes encoding putative glycosyltransferases are coloured grey, genes encoding putative ABC transporters are black and genes encoding OSTs are shown in red. The *pgl* loci of *H. canadensis* and *H. winghamensis* (not shown) have the same arrangement of genes as the *H. pullorum pgl* loci.



1.13 Summary

Many different types of glycosylation systems have been described in Eukaryotes and Prokaryotes. The bacterial glycosylation systems so far discovered are either N-linked or O-linked and vary in terms of glycan structure and protein specificity. The *C. jejuni* N-linked protein glycosylation system is well characterized and involves the biosynthesis of a Und-PP linked heptasaccharide glycan at the cytoplasmic side of the inner bacterial membrane. This structure is then transferred to the periplasm by a flippase enzyme and the complete heptasaccharide structure is attached to a protein containing a D/E-X-N-X-S/T amino acid sequon by the OST PglB. In *C. jejuni*, all the genes required for N-linked protein glycosylation are found at a single *pgl* locus and are functional when reconstituted in *E. coli*. This *C. jejuni* N-linked glycosylation system has the potential to be exploited for the production of therapeutic proteins and glycoconjugate vaccines. The major drawback of this system however is that the glycan to be transferred must have an acetamido group at the carbon-2 position of the reducing sugar and therefore the range of glycans that can be transferred by *C jejuni* PglB is limited.

Other bacterial species have homologues of the gene encoding PglB and some, such as *H. pullorum*, have two *pgl*B gene orthologues situated at distinct locations on the chromosome. One of the *H. pullorum* PglB's is able to transfer a unique pentasaccharide glycan to a peptide (Jervis *et al*., 2010) but the genes involved in the biosynthesis of the lipid-linked pentasaccharide had not been identified. Many *C. jejuni* N-linked glycoproteins have been discovered but there was no published data on *H. pullorum* glycoproteins.

1.14 Aims of This Study

(1) Investigate the feasibility of genetic manipulation of *H. pullorum* and establish methods for the insertion of antibiotic resistance cassettes onto the chromosome to allow insertional mutagenesis of specific genes or expression of recombinant proteins.

(2) Construct a *H. pullorum pgl*B2 knock out mutant strain by insertional mutagenesis and investigate the involvement of PglB2 in the biosynthesis of the PglB1-dependent N-linked pentasaccharide glycan using a previously established OST assay and tandem MALDI Mass Spectrometry.

(3) Determine potential *Helicobacter* N-linked glycoproteins by bioinformatic analysis of open reading frames and prediction of N-glycosylation sites and signal peptide sequences.

(4) Confirm the N-glycosylation of at least one *H. pullorum* putative glycoprotein in *E. coli* producing the *C. jejuni* glycosylation machinery and either *Campylobacter* or *Helicobacter* PglB OSTs.

(5) Investigate the N-glycosylation of a glycoprotein in *H. pullorum* and identify the N-glycan structure.

(6) Determine the effect of insertional mutagenesis of predicted *H. pullorum* glycosyltransferase-encoding genes, *pgl*A, *pgl*C, *pgl*H and *pgl*J and thereby construct a model of glycosyltransferase mediated biosynthesis of the *H. pullorum* lipid-linked pentasaccharide glycan.

(7) Investigate genes involved in the biosynthesis of the unusual 216 and 217 Da sugars present in the *H. pullorum* pentasaccharide glycan.

Chapter 2

Materials and Methods

2.1 Bacterial Strains and Growth Conditions

H. pullorum strain NCTC 12824^T was used in this study to investigate N-linked protein glycosylation systems (for strain list see Appendix II). Chemically competent NovaBlue GigaSingles™ *E. coli* (Merck, USA) were used for cloning, plasmid manipulation and gene expression. Plasmids used are listed in Appendix III.

E. coli was cultured using Luria-Bertani (LB) broth or LB agar (Oxoid, UK) and antibiotics (Sigma, UK) were added as required at concentrations of 100 μg/ml ampicillin, 50 μg/ml kanamycin, 32 μg/ml chloramphenicol, 300 μg/ml erythromycin and 100 μg/ml trimethoprim. As required, IPTG and *x*-gal (Bioline, UK) were used at final concentrations of 50 and 80 μg/ml respectively. *E. coli* cultures were incubated at 37ºC overnight and with agitation at 200 revolutions per minute (rpm) for broth cultures.

H. pullorum was cultured on blood agar plates consisting of Columbia agar (Oxoid, UK) supplemented with 5% (v/v) horse blood (TCS Biosciences, UK). Antibiotics were added at concentrations of 50 μg/ml kanamycin, 16 μg/ml chloramphenicol, 100 μg/ml trimethoprim and 10 μg/ml erythromycin. Helicobacter strains were incubated at 42°C in a microaerobic MACS VA500 workstation (Don Whitley Scientific, UK) under atmospheric conditions of 85% nitrogen, 10% carbon dioxide and 5% oxygen.

2.2 Isolation of Plasmids and Genomic DNA

Plasmids were purified from *E. coli* using a QIAprep spin miniprep kit (Qiagen, UK). Chromosomal DNA was isolated from Helicobacter strains using an ArchivePure DNA cell/tissue kit (SLS, UK) according to manufacturer's instructions.

2.3 PCR Amplification

PCR reactions contained 1 ng/µl chromosomal or plasmid DNA and 1x NH4 reagent buffer system (Bioline, UK). Deoxynucleotides, $MgCl_2$ and *Taq* polymerase (Bioline, UK) were added at final concentrations of 0.8 µM/ml, 1.5 µM/ml and 25 U/ml respectively. Primers (MWG, Germany) were used at a final concentration of 1 µM each and are described in Appendix I. For high fidelity PCR reactions the Expand High Fidelity PCR Kit (Roche, UK) was used according to manufacturer's instructions. A Thermo Electron Corporation (USA) PxE 0.2 Thermal Cycler PCR machine was used for PCR reactions with denaturing at 94°C for 30 seconds, annealing at 1°C below predicted melting temperature of primers for 30 seconds and 72°C extension of 1 minute per kilobase pair (kbp) length of predicted PCR amplification product.

2.4 Reverse Transcriptase PCR

RNA was extracted from *H. pullorum* cells using a Qiagen RNeasy kit essentially according to manufacturers' instructions but with an additional DNAse I digestion step. The concentration of RNA was estimated using a nanodrop ND-1000 spectrophotometer (Thermo Scientific, UK). Reverse transcriptase PCR was performed on RNA using a Qiagen one step RT-PCR kit according to manufacurers' instructions. Each 25 µl RT-PCR reaction contained dNTPs at a final concentration of 400 µM each, primers at concentrations of 0.6 µM each, 1 x Qiagen OneStep RT-PCR buffer, 2 µl Qiagen OneStep RT-PCR enzyme mix and 1 µg RNA.

2.5 Agarose Gel Electrophoresis

Agarose gels (1% v/v) for electrophoretic separation of DNA were prepared with SeaKem LE agarose powder (Lonza, UK) and 1x TAE buffer (40 mM Tris-acetate, 1 mM EDTA). Ethidium bromide (Sigma, UK) was added at 0.1 mg/ml. DNA hyperladder I (Bioline, UK) was used as a molecular weight marker and samples were run with 1x DNA loading dye (Bioline, UK). Electrophoresis was performed at

100 volts for one hour in a Bio-Rad sub-cell gel tank (California, USA) with an EV265 electrophoresis power supply (Consort, Belgium).

2.6 DNA Manipulation

PCR products were directly cloned into pGEM T easy vectors (Promega, UK) according to manufacturer's instructions using the supplied DNA ligase and buffer. For cloning into all other plasmids, 1-2 µg DNA was digested with 1 µl restriction enzyme and 4 µl of corresponding 10x buffer (Roche, UK) per 40 µl reaction for 1 hour at 37°C and enzymes inactivated by incubation at 65°C for 10 minutes. Restricted plasmid was phosphatised by adding 1 µl Antarctic Alkaline phosphatase with 6 µl buffer (New England Biolabs, UK) and incubating at 37°C for 30 minutes. DNA digests were purified using a Qiagen (UK) PCR purification kit according to the manufacturer's instructions, DNA concentration was estimated using a nanodrop ND-1000 spectrophotometer and ligations were carried out at room temperature for 1 hour or at 4°C overnight with DNA ligase and ligase buffer (Roche, UK) according to the manufacturer's instructions. Ligation products were transformed into chemically competent *E. coli* cells and grown overnight on LB agar plates containing appropriate antibiotics.

2.7 *E. coli* Transformation

Plasmids were transformed into *E. coli* cells using two different methods: electroporation and heat shock. Electroporation of NovaBlue GigaSingles *E. coli* (Merck, UK) with plasmid DNA was carried out using the heat shock protocol described in the manufacturer's instructions. For the electroporation method, *E. coli* cells were grown to an OD_{600} of approximately 0.6 then kept at 4°C throughout. Cells were centrifuged at 8000 rpm for 20 minutes, washed and resuspended in sterile distilled water, then centrifuged and washed in 40 ml sterile distilled water a further five times. Electrocompetent cells were stored in 10% (v/v) glycerol at –80°C and electroporations with plasmid DNA were carried out in 2 mm cuvettes (Bio-Rad, UK) using a Bio-Rad gene pulser coupled to a Pulse controller II and Capacitance Extender, at 2.5 kV with a resistance of 200 Ω and a capacitance of 25

μF. LB broth (Oxoid, UK) was mixed with the cell solution, and spread onto the appropriate selective agar plate which was incubated at 37°C overnight.

2.7 DNA Sequencing.

DNA sequencing reactions were performed using Big Dye Terminator v1.1 (Applied Biosystems, UK) according to manufacturers' instructions. Sequencing reactions were either run on an Applied Biosystems Prism 3100 Genetic Analyser in the University of Manchester (UK) sequencing facility or data were obtained commercially (Geneservice, UK). Sequence data analysis was performed using Chromas Pro software (Technelysium Pty Ltd, Australia) and sequence alignments were made using ClustalW or the National Centre for Biology Information Basic Local Alignment Search Tool.

2.8 Site Directed Mutagenesis

Complementary primers were designed containing the mutation flanked by unmodified nucleotide sequences. Site directed mutagenesis reactions contained 1 ng/μl plasmid DNA, 1x reaction buffer and deoxynucleotides at a final concentration of 0.25 μM/ml. Primers were used at a final concentration of 0.5 μM each. Agilent Technologies (UK) Pfu Ultra high fidelity DNA polymerase was added at 62.5 U/ml. A Thermo Electron Corporation (USA) PxE 0.2 Thermal Cycler machine was used for reactions with denaturing at 95°C for 30 seconds, annealing at 55°C or 3°C below T_m of primers for 1 minute and 72°C extension of 1 minute per kbp. Between 12 and 16 cycles were performed for point mutations or single amino acid mutations. Remaining original plasmid DNA was digested by incubation with 0.5 μl Dpn1 (New England Biolabs, UK) at 37°C for between 1 and 2 hours. Plasmid was transformed into *E. coli* Novablue competent cells (Novagen, UK).

2.9 Helicobacter Transformation

Two methods were used to transform *Helicobacter* spp: electroporation and natural transformation. For electroporation, *Helicobacte*r spp were grown on blood agar

plates at 37°C in microaerobic conditions for two to three days. The cells were suspended in 2 ml ice cold wash buffer (272 mM sucrose, 15% (v/v) glycerol), centrifuged at 8000 g for five minutes at 4°C and resuspended in a further 2 ml of ice cold wash buffer. This centrifugation and wash was repeated three times. Approximately 50 µl of competent *Helicobacter* cells were added to 10 µl plasmid DNA sample and transfered to a 2 mm ice cold electroporation cuvette (Bio-Rad, UK). The sample was electroporated, using a Bio-Rad gene pulser coupled to a Pulse controller II and Capacitance Extender, at 2500 Volts with a resistance of 200 Ω and a capacitance of 25 µF. SOC media (Novagen, UK) was mixed with the cell solution, and spread onto a blood agar plate. After overnight incubation at 37°C in microaerobic conditions, cells were harvested using Mueller-Hinton broth (Oxoid, UK) and plated onto blood agar plates containing selective antibiotics. Plates were incubated for up to five days and examined for growth of single *Helicobacter* colonies.

For natural transformation, genomic DNA with a kanamycin or erythromycin antibiotic resistance cassette insertion was extracted from *Helicobacter* strains following 48 hours growth. Wild type *Helicobacter* strains were grown for 48 hours, harvested and resuspended to an OD_{600} of approximately 1.0. These suspensions were incubated microaerobically for three hours then 3 µg of genomic DNA was added and the mixtures further incubated for five hours before plating onto blood agar containing either kanamycin or erythromycin as appropriate. Plates were examined for growth between 3 and 5 days incubation in microaerobic conditions.

2.10 Purification of Glycoproteins from *H. pullorum*

Approximately 1 g of *H. pullorum* cells grown at 37°C on blood agar plates under microaerobic condidtions were resuspended in 3 ml binding buffer (Na_2HPO4, NaCl, Imidazole) with protease inhibitors phenylmethanesulfonylfluoride and benzamidine at final concentrations of 0.1 mM and 1 mM respectively. Cells were lysed using a French pressure cell (Thermo Scientific, UK) and insoluble material removed by centrifugation at 8000 g for 20 minutes. The supernatant was incubated with 150 µl nickel coated magnetic beads (Qiagen, UK) for 1 hour at room temperature with

mixing. Beads were washed with binding buffer according to manufacturers' instructions and protein eluted from the beads in binding buffer containing 500 mM imidazole.

2.11 SDS-PAGE and Western Blotting.

All SDS-PAGE gels consisted of a lower resolving gel, containing 12% (v/v) acrylamide:bisacrylamide 37:1 (National Diagnostics, UK), and an upper stacking gel, containing 4% (v/v) acrylamide:bisacrylamide 37:1, and were prepared in Invitrogen (UK) gel cassettes. Ammonium persulphate (APS) and tetramethylethylenediamine (TEMED) were added to resolving and stacking gels to initiate polymerization.

Proteins were denatured in SDS PAGE sample buffer (250 mM tris-HCl containing 10% (w/v) SDS, 50% (v/v) glycerol, 0.5% (w/v) bromophenol blue and 1.5% (v/v) mercaptoethanol) by heating to 100°C for 10 minutes and loaded onto gels along with between 2 and 5 µl of broad range pre-stained protein standard markers (New England Biolabs, UK). Electrophoresis was performed in an Invitrogen Xcell SureLock® Mini-Cell at 160 V for one hour with 1x reservoir buffer (25 mM tris-HCl, 192 mM glycine, 3.5 mM SDS, pH 8.3). Following electrophoresis, gels were stained using Instant blue (expedeon, UK) or used for western blotting.

For western blotting, proteins were transferred from SDS-PAGE gels to nitrocellulose membrane (Protran, UK) using an Invitrogen Xcell SureLock® Mini-Cell blot module according to manufacturer's recommendations using transfer buffer (2.4 g/l tris, 11.26 g/l glycine, 20% (v/v) methanol). Blotted membranes were incubated at 4°C overnight in phosphate buffered saline (PBS) with 5% (w/v) powdered milk. To detect histidine tagged proteins, blots were incubated for 1 hour at room temperature with anti-penta histidine antibody (Qiagen, UK) at a dilution of 1:2000 in PBS containing 0.1% (v/v) tween 20 (PBST) and 3% (w/v) milk. Blots were washed three times in PBST and incubated for 1 hour at room temperature in PBST with 3% (w/v) milk and infrared (680 nm) dye labelled anti mouse antibody (LI-COR Biotechnology, UK) at a 1:2000 dilution. To detect the *C. jejuni* heptasaccharide glycan, a rabbit polyclonal antiserum (hR6) was used as the primary

antibody at a dilution of 1:30,000 and infrared (800 nm) dye labelled anti rabbit (LI-COR Biotechnology, UK) as the secondary antibody at a dilution of 1:15,000. Western blots were visualized on an Odyssey infrared imaging system (LI-COR Biotechnology, UK)

2.12 Mass Spectrometry for Characterisation of Proteins

Purified proteins were dialysed against a 25 mM tris-HCL, 25 mM NaCl buffer to remove excess Imidazole and analysed by Electrospray Ionisation Mass Spectrometry to provide an accurate mass to within 1 Da for a 50,000 Da protein. Intact mass analysis was performed by staff at the Biomolecular Analysis core facility within the Faculty of Life Sciences at the University of Manchester. In this facility, the sample was applied to the Mass Spectrometer as a liquid and the protein signal was divided between several peaks reflecting the multiple charge states. Computer software was used to deconvolute these peaks back to an uncharged mass.

For Matrix Assisted Maser Desorption Ionisation (MALDI) Mass Spectrometry analysis, purified proteins were stained with Instant blue (expedeon) following SDS-PAGE, excised from the gel, cut into 1 mm sized cubes. The gel cubes were incubated at room temperature with 200 µl of 50 mM ammonium bicarbonate pH 8.4 (AMBIC) and 200 µl acetonitrile for five minutes to remove Instant blue stain. The liquid was then removed and gel pieces were dried for 15 minutes using a Heto Speedvac vacuum centrifuge then incubated at room temperature for 15 minutes in 20 µl of 50 mM AMBIC containing 25 ng/µl trypsin (Promega, UK). Additional 50 mM AMBIC was then added to cover the gel pieces, incubated overnight at 37°C and the supernatant containing hydrophilic peptides was removed. Gel pieces were immersed in 50 µl of 0.1% (v/v) trifluoroacetic acid at 37°C for five minutes followed by 100 µl acetonitrile for 15 minutes and this supernatant was added to the previous supernatant. This step was repeated and the sample volume reduced to between 10 and 20 µl using a Speedvac. This sample was acidified by adding 0.1% (v/v) formic acid and purified using a C18 ZipTip (Millipore, UK).

The ZipTip was wet with 50% (v/v) acetonitrile and equilibrated with 0.1% (v/v) formic acid and the sample was bound and washed with 0.1% (v/v) formic acid according to manufacturer's instructions. Purified peptide was eluted directly into between 4 and 10 µl of 50% (v/v) acetonitrile, 0.1% (v/v) formic acid. 20 mg/ml 2,5-dihydroxybenzoic acid (DHB) matrix was prepared in 30% (v/v) acetonitrile with 0.1% (v/v) trifluoroacetic acid and 1 µl was spotted onto a metal MALDI target plate with 1 µl of purified peptide. MALDI time of flight (TOF/TOF) Mass Spectrometry and tandem Mass Spectrometry spectra were acquired using a Bruker Ultraflex II Mass Spectrometer in the positive-ion reflection mode. Spectra were viewed and analyzed with FlexAnalysis 3.0 software (Bruker Daltronics).

2.13 Preparation of Bacterial Cell Membranes

Membrane preparations of *H. pullorum* and *C. jejuni* cells were prepared by suspending cells grown on blood agar plates in 50 mM tris-HCL (pH 7.0), 25 mM NaCl. Suspensions were centrifuged at 8000 g, the supernatant removed and the cell pellet weighed. The cell pellet was then resuspended in 50 mM tris-HCl (pH 7.0), 25 mM NaCl, cells disrupted using a French pressure cell at 13000 psi and unbroken cells removed by two rounds of centrifugation at 8000 g for 20 minutes at 4°C. The resultant cell lysate was ultracentrifuged at 100,000 g in a Beckman OptiMAX ultracentrifuge for one hour at 4°C, the supernatant discarded and the pelleted material was the total membrane fraction. This pellet was resuspended in 50 mM tris-HCl (pH 7.0), 25 mM NaCl, 1% (v/v) triton X-100. Resuspended membrane preparations were ultracentrifuged at 100,000 g for one hour at 4°C and the resulting supernatant or total membrane fraction was used in the oligosaccharyltransferase (OST) assay described below.

2.14 Oligosaccharyltransferase Assay

Fluorescent peptides (Invitrogen) were resuspended at 30 µM and 3 µl was added to 26 µl of total membrane preparation along with 1 µ of 150 mM $MnCl_2$ and incubated overnight in the dark at 30°C with agitation. The reaction was stopped by the addition of 10 µl 4x tris-tricine loading buffer (12% (w/v) SDS, 30% (v/v) glycerol,

0.05% (w/v) Coomassie blue, 150 mM tris-HCl pH 7.0) and samples heated to 95°C before loading onto a tricine SDS-PAGE gel (see below).

2.15 Tricine SDS-PAGE

Gels were prepared essentially as in Schagger (2006) with acrylamide/bisacrylamide at a 20:1 ratio (Sigma) and consisted of a 16.5% (v/v) separating gel, a 10% (v/v) spacer gel and a 4% (v/v) stacking gel. 10% (w/v) APS and TEMED were added to gels for polymerization. Samples from the OST assay (described above) were loaded into the wells and run at 4°C in an Invitrogen cell surelock electrophoresis tank at 30 V for approximately 30 minutes and then at 200 V for approximately 1 hour with cathode and anode buffers described in Schagger (2006). Gels were visualised on a Typhoon Trio plus imaging system (GE Healthcare, UK) using the green (532 nm) laser.

2.16 Purification of Biotinylated Peptide from the *In vitro* OST Assay

Biotinylated peptide was purified from OST assay mixtures using streptavidin-coated magnetic beads (Pierce, UK). Before use, 10 µl beads were washed three times with 150 µl of 50 mM Tris-HCl (pH 7.4) containing 0.1% (w/v) BSA using a magnetic separator. The beads were resuspended in the assay mixture containing biotinylated peptide and incubated for at least 30 minutes with agitation to bind biotinylated peptides. Beads were washed in 400 µl of buffer A (50 mM Tris-HCl pH 7.4, 1 mM DTT, 0.1 % (w/v) BSA) then twice in 400 µl of buffer A containing 0.1% (w/v) SDS. Beads were then washed twice with 400 µl buffer A containing 1 M NaCl and three times with 1mM DTT. The beads were then resuspended in 25 µl of 28 µM biotin, incubated at room temperature for 15 minutes with agitation, washed three times in 1 ml H_2O and peptide eluted in 10 µl of 70% (v/v) acetonitrile, 5% (v/v) formic acid and 125 µM biotin for 30 minutes at room temperature. Using a magnet, the liquid was separated from the beads, evaporated in a Speedvac to a volume of between 1 and 2 µl and acidified by the addition of between 10 and 20 µl

of 0.1% (v/v) formic acid. This sample was then purified using a C18 ZipTip and analysed by tandem MALDI Mass Spectrometry as described in 2.14.

Chapter 3

The Second *Helicobacter pullorum* Oligosaccharyltransferase (PglB2) is Neither Essential nor Required for N-linked Protein Glycosylation

3.1 Background

Helicobacters are naturally competent, Gram-negative, microaerophilic, spiral-shaped Bacteria related to the genus Campylobacter. The best characterized species of the Helicobacter genus, *H. pylori*, was first isolated from a gastric biopsy in 1982 (Marshall and Warren, 1984). Since then at least 28 further Helicobacter species have been described including the phylogenetically related species *H. pullorum* (Stanley *et al*., 1994), *Helicobacter canadensis* (Fox *et al.*, 2000) *and Helicobacter winghamensis* (Melito *et al.*, 2001). *H. pullorum* was first isolated from the caeca, livers and duodena of broiler chickens and laying hens. Isolates of *H. pullorum, H. canadensis* and *H. winghamensis* have also been recovered from the faeces of patients with gastroenteritis (Stanley *et al*., 1994).

Techniques for the genetic manipulation of *H. pylori* and *C. jejuni* are well established (Ferrero *et al.*, 1992; Haas *et al*., 1993). Gene transfer between *E. coli* and Campylobacter species was first described using recombinant shuttle vectors (Labigne-Roussel *et al*., 1987). Later methods involved the cloning of a *C. jejuni* 16S rRNA gene fragment into a suicide vector, disruption of the gene with a kanamycin resistance cassette and natural transformation into *C. jejuni* cells (Labigne-Roussel *et al*., 1988). Integration of the resistance cassette involved a double crossover between the 16S rRNA gene flanking regions present on the plasmid and the homologous 16S rRNA gene on the *C. jejuni* chromosome. Higher transformation efficiencies can be achieved by high voltage electroporation of plasmid DNA into cells (Miller *et al*., 1988). These techniques have been used with *C. jejuni* and *H. pylori* to insertionally inactivate many genes including urease and flagellin genes (Ferrero *et al*., 1992; Haas *et al*., 1993). The first aim of the work described in this chapter was to investigate the feasibility of constructing *H. pullorum* knock out mutants through electroporation and insertional mutagenesis with various antibiotic resistance cassettes and to inactivate specific genes putatively involved in N-linked glycosylation processes in *H. pullorum.*

In contrast to *C. jejuni* and most Campylobacter species that have a single N-linked OST encoding *pgl*B gene, *H. pullorum* has two distinct *pgl*B genes (Hp*pgl*B1 and

Hp*pgl*B2) located separately on the chromosome. Both HpPglB1 and HpPglB2 OST's contain multiple predicted membrane spanning regions within the N terminal domain and a conserved WWDYG sequence within the C terminal domain. The role of HpPglB1 in N-linked protein glycosylation was recently confirmed (Jervis *et al.*, 2010). As there are no characterised *H. pullorum* N-linked glycoproteins, an alternative approach for investigating this system was required whereby a 6-carboxyfluorescein (FAM)-labelled peptide, based on the known *C. jejuni* acceptor sequon (FAM-ADQNATA), was glycosylated *in vitro* by PglB's present in *C. jejuni* or *H. pullorum* detergent solubilised membrane preparations (Jervis *et al*., 2010). Following peptide purification, glycan structure was determined using MALDI tandem Mass Spectrometry (Jervis *et al*., 2010). The HpPglB1 protein was required for peptide N-glycosylation with a pentasaccharide glycan structure consisting of a 203 Da HexNAc sugar at the reducing end followed by a 216 Da sugar, two 217 Da sugars and a terminal 203 Da HexNAc (Jervis *et al.*, 2010). Attempted insertional mutagenesis of the Hp*pgl*B2 gene was unsuccessful and the role of this predicted OST in N-glycosylation is unknown.

In this chapter the expression of Hp*pgl*B1 and Hp*pgl*B2 genes was investigated using Reverse Transcriptase (RT)-PCR and methods developed for the insertion of genes onto the *H. pullorum* chromosome in order to further attempt the insertional inactivation of the Hp*pgl*B2 gene. The *in vitro* peptide assay developed by Jervis *et al*. (2010) was used in conjunction with MALDI tandem Mass Spectrometry to determine whether mutation of the *pgl*B2 gene affects the PglB1 dependent N-glycosylation system of *H. pullorum*.

3.2 Results

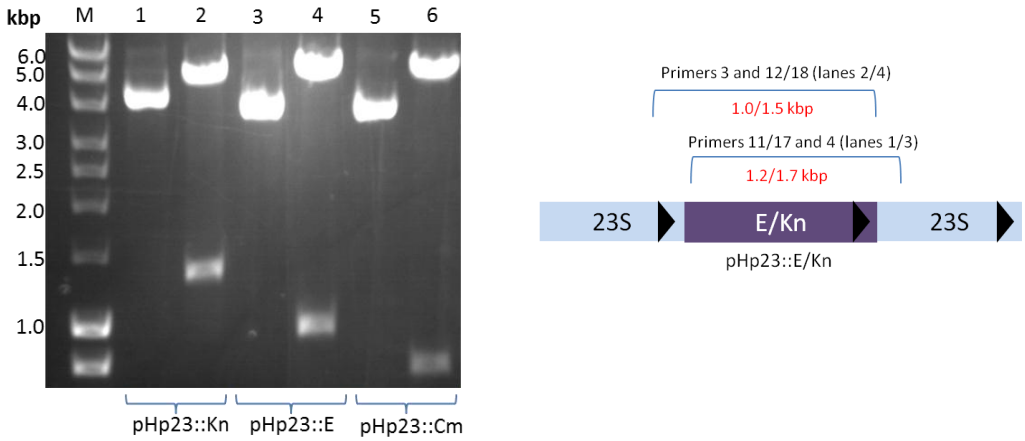
3.2.1 Construction of suicide plasmids for site specific integration of antibiotic resistance cassettes onto the *H. pullorum* chromosome.

In order to investigate the feasibility of constructing *H. pullorum* knock-out mutants, a region of the *H. pullorum* chromosome was chosen that would likely be permissive for insertion of antibiotic resistance cassettes. The *H. pullorum* genome contains three copies of the 23S rRNA gene and this was chosen for the site of insertion of antibiotic resistance cassettes. A 1.8 kbp fragment of the *H. pullorum* 23S rRNA gene was amplified by PCR using primers 1 and 2 (See Appendix I for primer list) and ligated into the pGEM T-easy vector to create plasmid pHp23S. The *H. pullorum* 23S rRNA gene contains two HindIII restriction sites (300 bp apart) in the central portion of the cloned 23S rRNA gene which were used to linearise the plasmid.

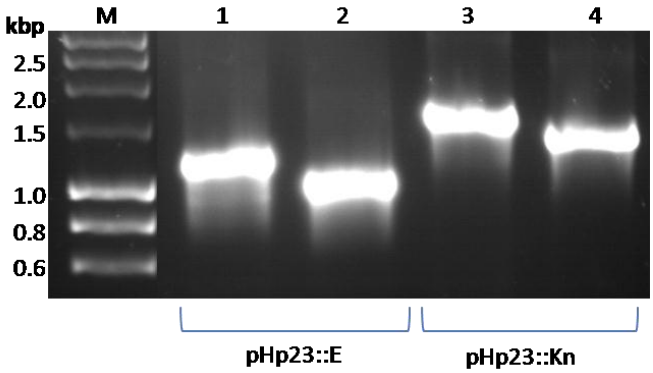
Three antibiotic resistance cassettes encoding kanamycin, erythromycin and chloramphenicol resistance (sized 1.5, 0.95 and 0.8 kbp respectively) were amplified by PCR with primers containing HindIII restriction sites and products ligated into pHp23S cut with HindIII. Primers 11/12, 17/18 and 7/8 were used for amplification of the kanamycin, erythromycin and chloramphenicol resistance cassettes respectively (See Appendix I). So that the antibiotic resistance cassette is less likely to interfere with the transcription of other genes on the *H. pullorum* chromosome, cassettes were introduced in the same transcriptional orientation as the 23S rRNA gene. This was confirmed by PCR analysis of plasmid constructs with primers specific for the 23S rRNA gene and the relevant antibiotic resistance cassette (Figure 3.1 B and C). The resultant plasmids were termed pHp23::Kn, pHp23::E and pHp23::Cm.

Figure 3.1 Verification of pHp23::E, pHp23::Kn and pHp23::Cm plasmid structures. (A) Analysis of plasmids pHp23::Kn (B) and 2), pHp23::E (lanes 3 and 4) and pHp23::Cm (lanes 5 and 6). The lane labelled M contains a DNA molecular weight marker. Plasmids in lanes 1, 3 and 5 are uncut, those in lanes 2, 4 and 6 are cut with HindIII and generate appropriately sized bands of 1.5 kbp for the kanamycin resistance cassette, 0.95 kbp for the erythromycin resistance cassette and 0.8 kbp for the chloramphenicol resistance cassette as well as a band at approximately 5 kbp corresponding to the vector pHp23. (B) Schematic diagram of primers used to confirm orientation of antibiotic resistance cassettes with respect to the flanking 23S rRNA gene. Blue arrowheads indicate transcriptional orientation. Predicted PCR product sizes are indicated in red. (C) PCR products generated from pHp23::Kn and pHp23::E confirming antibiotic resistance cassette orientation. The orientation of pHp23::Cm was also checked by PCR with primers 3, 4, 7 and 8 (image not shown).

(A)



(C)

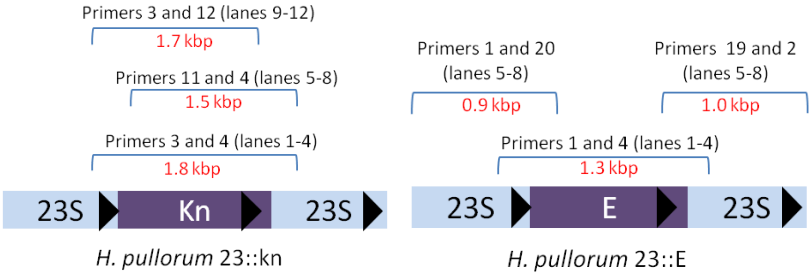


3.2.2 Integration of antibiotic resistance cassettes onto the *H. pullorum* chromosome within the 23S rRNA gene.

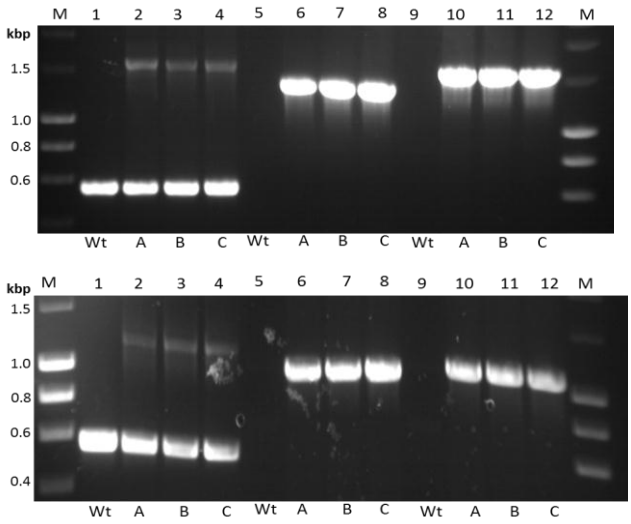
Wild-type *H. pullorum* NCTC 12824^T electrocompetent cells were electroporated with plasmids pHp23S::Kn, pHp23S::Cm and pHp23S::E as described in Methods and cell suspensions plated on appropriate selective agar plates. Following five days incubation, kanamycin and erythromycin resistant colonies were observed but no chloramphenicol resistant colonies suggesting that of the three antibiotic resistance cassettes, only kanamycin and erythromycin function in *H. pullorum*. Following three separate plasmid electroporations with different batches of electrocompetent cells, between 100 and 300 individual kanamycin resistant colonies were observed with the pHp23Kn plasmid, compared to between 11 and 39 individual erythromycin resistant colonies with the pHp23E plasmid. Genomic DNA was prepared from three colonies of each transformation and insertion of the antibiotic resistance cassette into the 23S rRNA gene verified (Figures 3.2 and 3.3). Strains with integrated kanamycin and erythromycin resistance cassettes at the 23S rRNA gene were designated *H. pullorum* NCTC 12824^T 23::Kn and 23::E respectively. These experiments demonstrate that both kanamycin and erythromycin resistance cassettes can be used to select for insertion of DNA onto the *H. pullorum* NCTC 12824^T genome.

Figure 3.2 PCR verification of insertion of the kanamycin and erythromycin resistance cassettes into the *H. pullorum* 23S rRNA gene. (A) Schematic diagram of primers used to screen for the presence of a kanamycin or erythromycin resistance cassette in the same transcriptional orientation as a *H. pullorum* 23S rRNA gene. Blue arrowheads indicate transcriptional orientation. Predicted PCR product sizes are indicated in red. (B) PCR analysis of genomic DNA of *H. pullorum* 23::Kn/23::E using primers from (A). Lane M contains a DNA molecular weight marker. Lanes 1 to 4 contain PCR amplification products from wild-type *H. pullorum* (wt) and three *H. pullorum* 23::Kn or 23::E strains (A-C) with primers 3 and 4. The high molecular weight bands of approximately 1.6 kbp (23::Kn strain) or 1.1 kbp (23::E strain) in lanes 2-4 indicate the presence of a kanamycin or erythromycin resistance cassette in at least one of three *H. pullorum* 23S genes. Bands at approximately 0.5 kbp are derived from 23S rRNA genes that do not contain an inserted resistance cassette. Lanes 5-8 and 9-12 contain PCR products from the same DNA generated with primers, one of which is specific for the kanamycin or erythromycin resistance cassette and the other for the *H. pullorum* 23S gene. These data confirm correct orientation of the cassette with respect to the transcriptional orientation of the 23S rRNA gene.

(A)



(B)



H. pullorum 23::Kn

H. pullorum 23::E

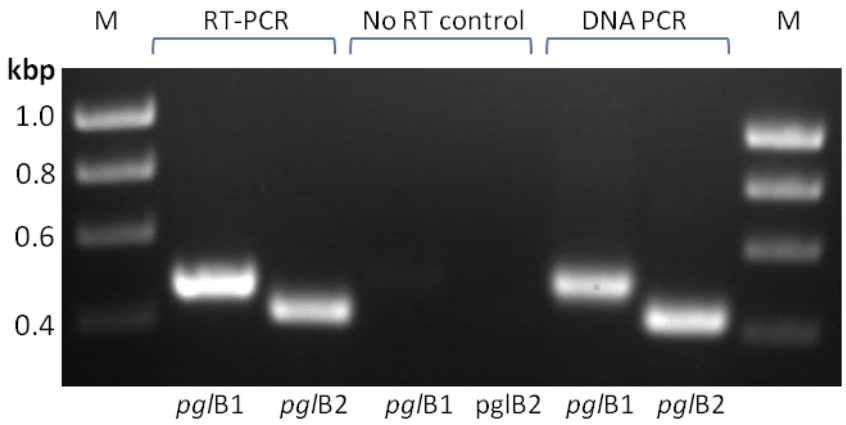
3.2.3 Natural Transformation of *H. pullorum* NCTC 12824^T with genomic DNA from *H. pullorum* 23::E and 23::Kn strains.

In order to determine whether insertional knock-out mutations could be 'moved' from one strain of *H. pullorum* NCTC 12824^T to another, a series of natural transformation experiments was carried out (see Methods for details). Genomic DNA (1-3 µg) was isolated from *H. pullorum* NCTC 12824^T 23::E and 23::Kn strains and incubated for five hours with a suspension of wild type *H. pullorum* NCTC 12824^T (grown for 48 hours in microaerobic conditions) to allow natural transformation to occur. Between 24 and 58 *H. pullorum* colonies were observed when transformants were plated on blood agar plates containing the appropriate antibiotics, indicating that natural transformation was successful in *H. pullorum* strain NCTC 12824^T. Genomic DNA from these colonies was isolated and PCR confirmed the presence of the antibiotic resistance cassettes within the 23S rRNA genes. Three further *H. pullorum* strains, *H. pullorum* NCTC 13153^T, *H. pullorum* NCTC 13155^T and *H. pullorum* 12826^T (See Appendix II for strain list) all gave rise to erythromycin resistant colonies following incubation with genomic DNA from *H. pullorum* NCTC 12824^T 23::E. All *H. pullorum* erythromycin resistant strains were checked by PCR of genomic DNA for insertion of the cassette at the 23S rRNA site as previously described. These experiments indicate that *H. pullorum* strains are naturally transformable. All future analyses were preformed with the *H. pullorum* strain NCTC 12824^T.

3.2.4 Expression of *H. pullorum pgl*B1 and *pgl*B2 genes

Having established that insertional knock out mutagenesis strategies could be employed to investigate the function of genes involved in N-linked protein glycosylation, the next experiment aimed to determine whether both of the oligosaccharyltransferase encoding *pgl*B genes of *H. pullorum* were expressed under *in vitro* growth conditions. Expression was investigated using Reverse Transcriptase PCR. Briefly RNA was extracted from cells of *H. pullorum* NCTC 12824^{T} grown on blood agar plates incubated microaerobically for 48 hours using a Qiagen RNeasy kit (See Methods for details), total RNA was reverse transcribed to cDNA and PCR performed with primers specific for an internal portion of the Hp*pgl*B1 (primers 22 and 23) and Hp*pgl*B2 genes (primers 24 and 25). Predicted sizes of the expected PCR products were 0.506 kbp for *pgl*B1 and 0.416 kbp for *pgl*B2. Data obtained clearly demonstrated that both *pgl*B1 and *pgl*B2 were expressed under *in vitro* growth conditions (Figure 3.3).

Figure 3.3 Expression of *H. pullorum* NCTC 12824^{T} *pgl*B1 and *pgl*B2 genes during *in vitro* growth. Gel electrophoresis of products generated with primers specific for *pgl*B1 and *pgl*B2 in Reverse Transcription (RT) PCR reactions. The lanes labelled M contain DNA molecular weight markers. No products were generated without reverse transcriptase (no RT control) and PCR of *H. pullorum* DNA (DNA PCR) generated products of the predicted sizes (0.506 kbp and 0.416 kbp for *pgl*B1and *pgl*B2 respectively).



3.2.5 Insertional mutagenesis of *H. pullorum pgl*B1 and *pgl*B2 genes

In previous attempts the *H. pullorum pgl*B1 gene but not the *pgl*B2 gene could be 'knocked-out' by standard insertional mutagenesis (Jervis *et al*., 2010). I therefore aimed to confirm whether it was possible to knock-out *pgl*B2 using kanamycin and/or erythromycin resistance cassettes known to function in *H. pullorum* (Figure 3.2B). Suicide plasmids designated pHp*pgl*B1::Kn and pHp*pgl*B2::Kn were previously constructed with portions of the *H. pullorum pgl*B1 and *pgl*B2 genes flanking the kanamycin resistance cassette. The arrangement of these plasmids was confirmed by restriction digest and PCR (Figure 3.4).

Plasmids pHp*pgl*B1::Kn and pHp*pgl*B2::Kn were electroporated into competent *H. pullorum* NCTC 12824^{T} cells and between 40 and 150 kanamycin resistant colonies were obtained following each of six separate electroporations with pHp*pgl*B1::Kn as previously described (Jervis *et al*., 2010). In contrast, electroporations with pHp*pgl*B2::Kn were much less efficient with only two kanamycin resistant colonies observed among six separate transformations although these were the first putative *pgl*B2 gene knock out mutants obtained in the laboratory. All electroporations with vectors pHp*pgl*B1::Kn and pHp*pgl*B2::Kn were carried out using the same preparations of electrocompetent cells and 3 µg of vector DNA per electroporation. Genomic DNA from *H. pullorum* strains was checked for insertion of the antibiotic resistance cassette by PCR with primers specific for the cassette and regions flanking the *pgl*B genes (Figure 3.5).

These data indicate that the insertional mutagenesis of *H. pullorum pgl*B1 and *pgl*B2 was successful and there was no obvious reason for the low efficiency of mutagenesis of the *H. pullorum pgl*B2 gene. These knock out mutants were taken forward to analyse the role of *H. pullorum* PglB2 in the PglB1-dependent N-linked protein glycosylation system.

Figure 3.4 Verification of plasmid structure of pHp*pgl*B1::Kn and pHp*pgl*B2::Kn (A) Map of pHp*pgl*B1::Kn and pHp*pgl*B2::Kn suicide plasmids showing restriction enzyme sites. Predicted sizes are indicated in red. (B) Plasmids from (A) confirming the insertion of kanamycin antibiotic resistance cassettes at specific restriction enzyme sites. Lanes 1 and 3 contain uncut pGEM*pgl*B1::Kn and pGEM*pgl*B2::Kn plasmids respectively. Lane 2 is pHp*pgl*B1::Kn cut with BamHI and lane 4 is pHp*pgl*B2::Kn cut with HindIII. Lane M contains a DNA molecular weight marker. (C) Schematic diagram of the *pgl*B gene regions with primers used to confirm the orientation of the kanamycin resistance cassette within *pgl*B1 and *pgl*B2 genes. Blue arrowheads indicate transcriptional orientation. (D) Products of a PCR reaction using primers from (C) to confirm the orientation of the kanamycin resistance cassette insertion within the *pgl*B1 and *pgl*B2 genes. Lane M contains a DNA molecular weight marker

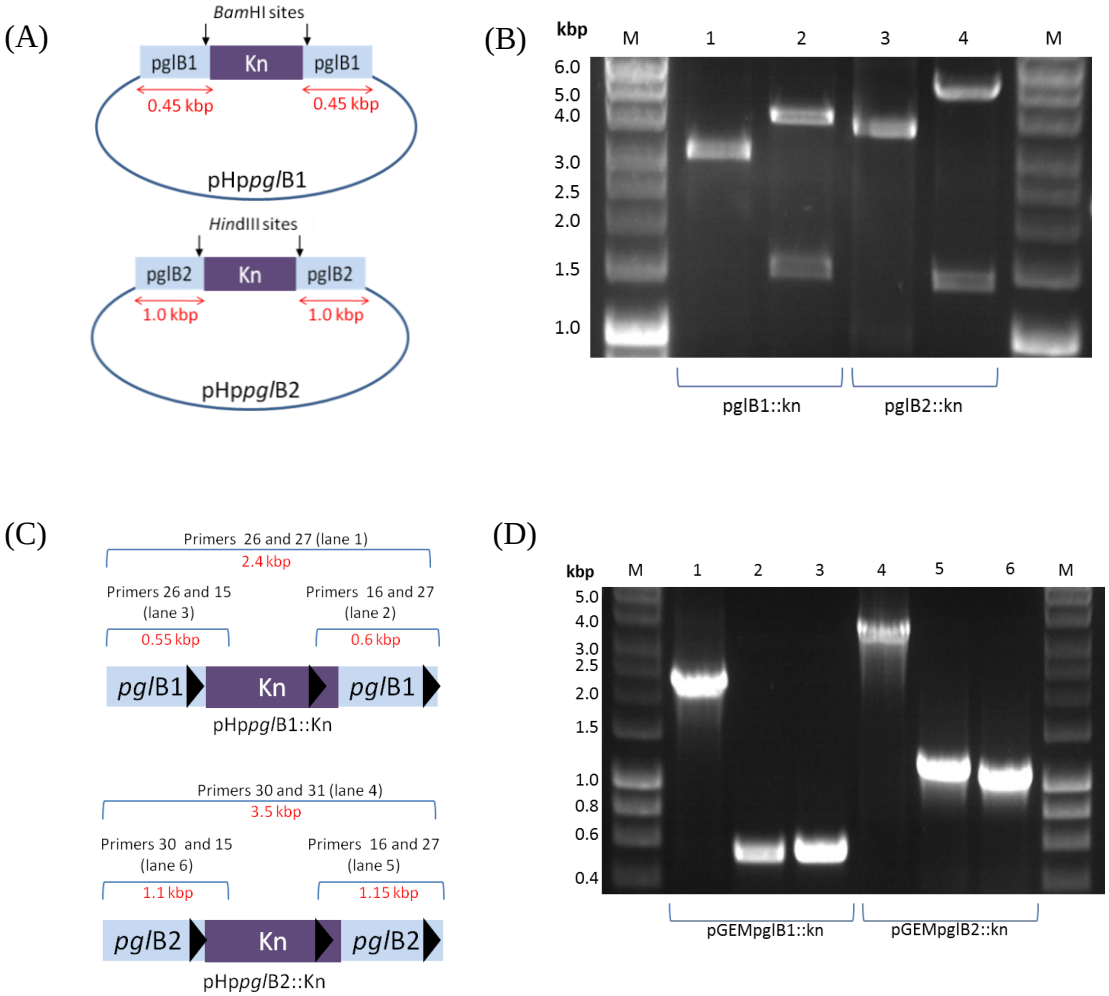
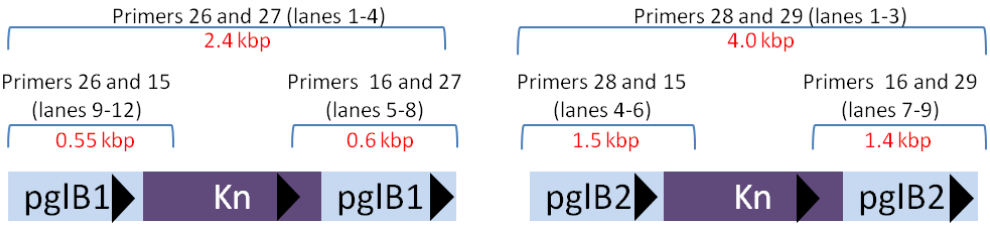
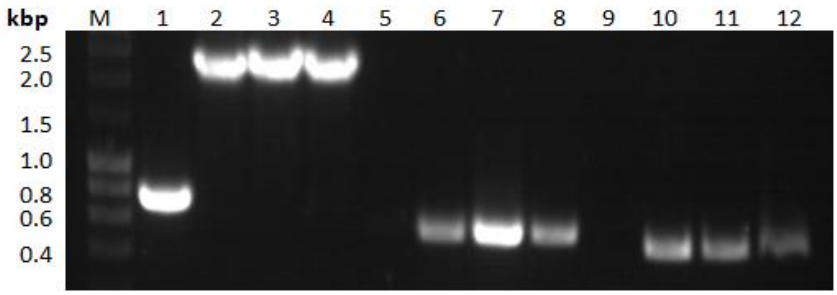


Figure 3.5 Confirmation of insertional mutagenesis of *H. pullorum pgl*B1 and *pgl*B2. (A) Schematic diagram of *H. pullorum pgl*B1 and *pgl*B2 genes indicating primers used to confirm the insertion of a kanamycin cassette within the *pgl*B1 or *pgl*B2 gene. Primers are labelled according to their primer number (see Appendix I). (B) PCR products amplified from chromosomal DNA of wild type *H. pullorum* and three *H. pullorum pgl*B1::Kn strains. Lanes 1 to 4 contain PCR products following amplification with primers 26 and 27, lanes 5 to 8 with primers 16 and 27 and lanes 9 to 12 with primers 26 and 15. (C) PCR products amplified from chromosomal DNA of wild type *H. pullorum* and two *H. pullorum pgl*B2::Kn strains. Lanes 1 to 3 contain PCR products following amplification with primers 28 and 29, lanes 4 to 6 with primers 28 and 15 and lanes 7 to 9 with primers 16 and 29.

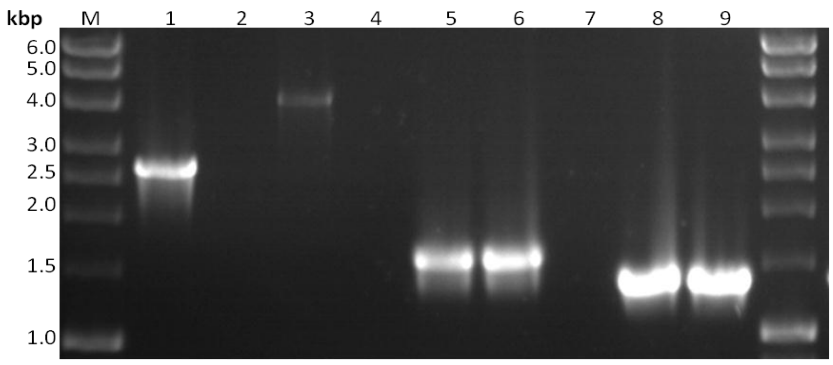
(A)



(B).



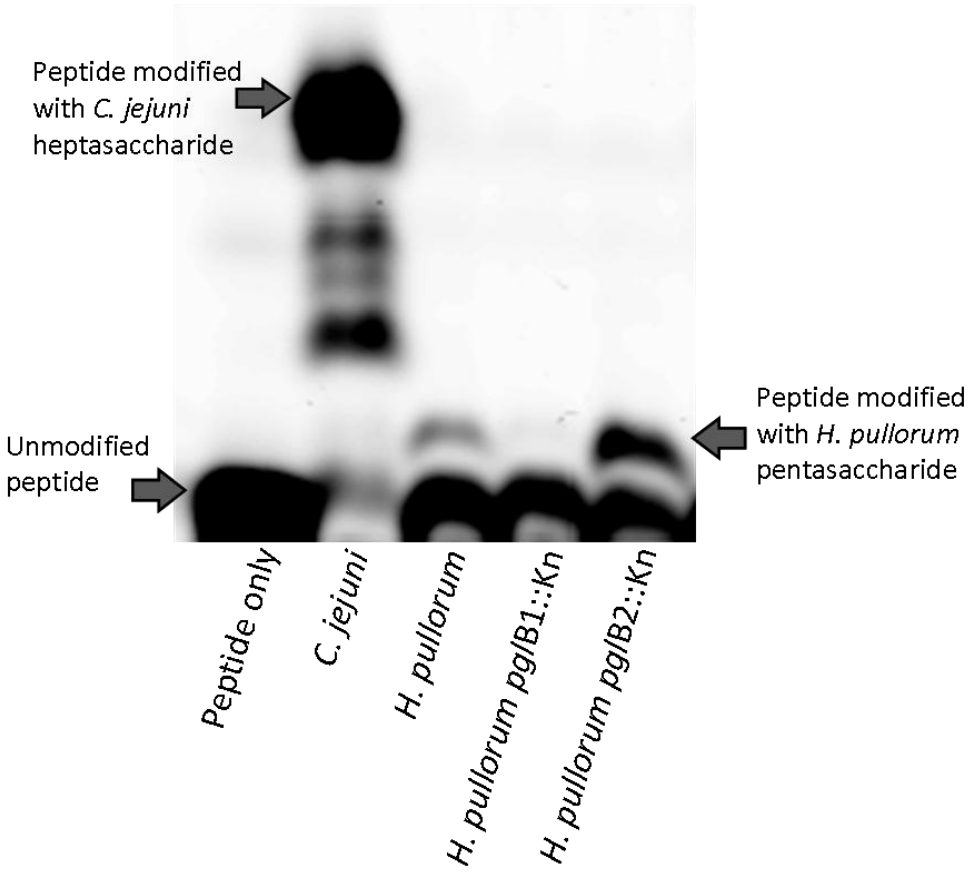
(C)



3.2.6 Role of PglB2 in the N-glycosylation of a small peptide *in vitro*

An *in vitro* OST assay (Kohda *et al.*, 2007) has previously shown that *H. pullorum pgl*B1 is essential for the transfer of a pentasaccharide glycan to an asparagine residue of a peptide based on the optimal *C. jejuni* glycosylation sequon (Jervis *et al.*, 2010). In order to determine whether the *H. pullorum pgl*B2 gene product was required for *in vitro* peptide glycosylation, cell membranes were prepared from the *H. pullorum pgl*B2::Kn mutant. As controls, membranes were also prepared from the *H. pullorum* wild-type and *pgl*B1::Kn mutant. The peptide was observed via its fluorescent tag following Tricine gel electrophoresis and there was evidence of peptide glycosylation when co-incubated with wild type and *pgl*B2::Kn mutant derived membranes but not membranes derived from the *pgl*B1::Kn mutant (Figure 3.6). This was evidence that PglB2 plays no role in the PglB1 dependent N-glycosylation pathway.

Figure 3.6 *In vitro* peptide glycosylation. The products of the *in vitro* peptide glycosylation assay were analysed by Tricine SDS-PAGE. Unmodified peptide is detected at the very bottom of the gel whilst modified products have reduced electrophoretic mobility. Products from *in vitro* assays with *C. jejuni* NCTC 11168^{T}, *H. pullorum* wild type, *H. pullorum pgl*B1::Kn and *H. pullorum pgl*B2::Kn membrane preparations are presented as indicated below the gel. There is no reduced electrophoretic mobility of peptide products with *H. pullorum pgl*B1::Kn membranes. Peptide from the *H. pullorum* wild type and *pgl*B2::Kn OST assays had reduced electrophoretic mobility compared to unmodified peptide suggesting glycosylation with the *H. pullorum* pentasaccharide glycan.

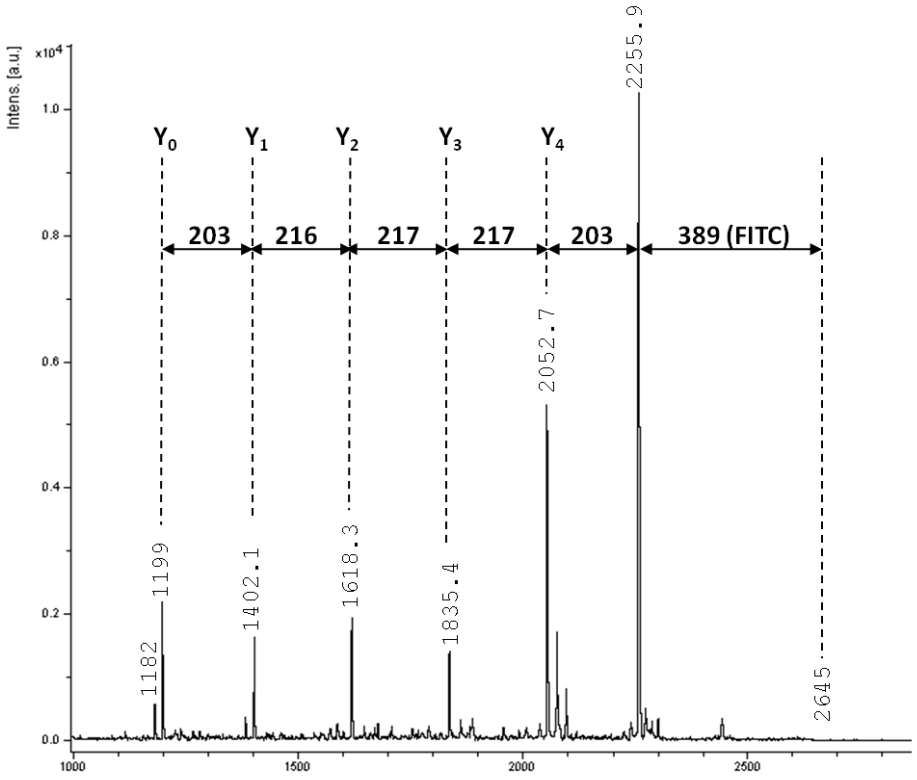


3.2.7 The N-glycan structure of glycopeptide derived from the *H. pullorum pgl*B2 mutant.

Membranes prepared from *H. pullorum* wild-type, *pgl*B1::Kn and *pgl*B2::Kn mutants were incubated overnight with the fluorescein isothiocyanate (FITC) labelled biotinylated peptide, ADQNATA. The biotin-labelled peptide was then purified using streptavidin coated magnetic beads (See Methods for details) and MALDI Mass Spectrometry performed. Peptide reaction products from the *in vitro* OST assay with membrane preparations from *H. pullorum* wild type and *pgl*B2::Kn mutant strains generated an *m/z* 2645 precursor ion which corresponded to glycosylated peptide as previously described (Jervis *et al*., 2010). This precursor ion was not observed following MALDI Mass Spectrometry with peptides incubated with membranes from the *pgl*B1::Kn mutants (data not shown). Fragmentation of the *m/z* 2645 precursor ion produced by both *H. pullorum* wild type and *pgl*B2::Kn derived glycopeptides generated a fragmentation series corresponding to the sequential loss of sugars, more specifically a non reducing end 203 Da HexNAc residue, two 217 Da residues, a 216 Da residue and a 203 Da reducing end HexNAc residue (Figure 3.7 and 3.8) as described in Jervis *et al*. (2010).

Figure 3.7 Tandem MALDI Mass Spectrometry analysis of *H. pullorum* N-linked glycan. (A) Tandem Mass Spectrometry spectrum of the *m/z* 2645 precursor ion generated following incubation of *H. pullorum* membranes with a biotin-labelled fluorescent peptide (FITC-ADQNATAK-biotin, 1587 Da in mass). Fragment ions resulting from the sequential loss of sugar residues are indicated in the spectrum. A peak corresponding to the Y ion of the biotinylated peptide lacking FITC was present at *m/z* 1199. The peak at *m/z* 1182, 17 Da less than the *m/z* 1199 peak, is characteristic of the fragmentation of a side-chain amide bond of an N-linked glycan (B) N-linked glycan structure of the *H. pullorum* pentasaccharide consistent with the fragmentation pattern seen in (A).

(A)



(B)

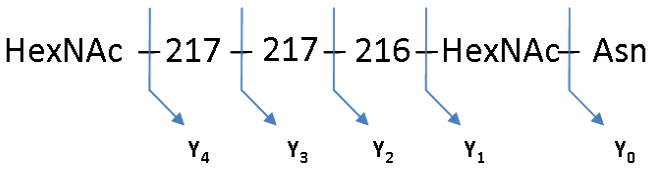
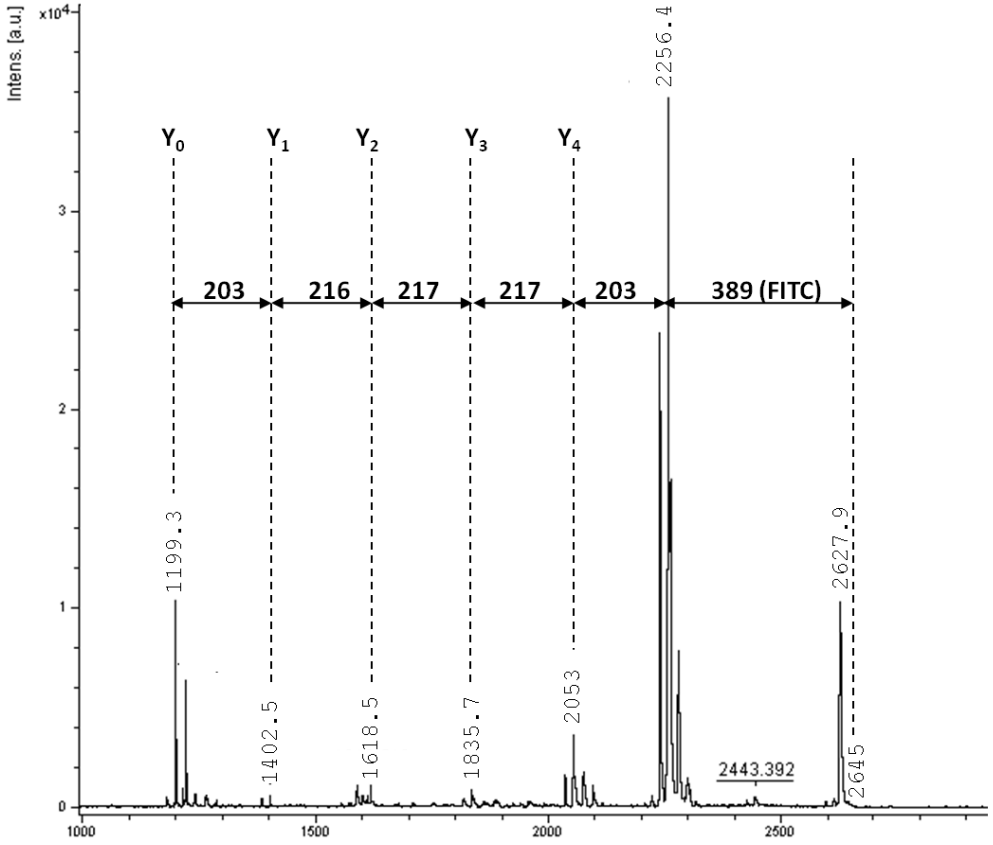
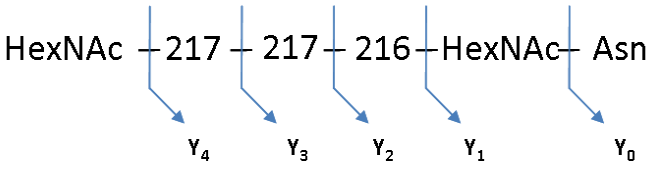


Figure 3.8 Tandem MALDI Mass Spectrometry analysis of N-linked glycan from H. pullorum pglB2::Kn. Tandem Mass Spectrometry spectrum of the *m/z* 2645 precursor ion generated following incubation of *H. pullorum pgl*B2::Kn membranes with a biotin-labelled fluorescent peptide (FITC-ADQNATAK-biotin, 1587 Da in mass). Fragment ions resulting from the sequential loss of sugar residues are indicated in the spectrum. A peak corresponding to the Y ion of the biotinylated peptide lacking FITC was present at *m/z* 1199. (B) N-linked glycan structure of the *H. pullorum pgl*B2::Kn pentasaccharide consistent with the fragmentation pattern seen in (A).

(A)



(B)



3.3 Discussion

H. pullorum is a poorly characterized species compared to *H. pylori* and *C. jejuni* and initial experiments were designed to determine whether genes could be introduced into the chromosomal DNA by electroporation with *E. coli* derived plasmids containing *H. pullorum* 23S rRNA gene fragments flanking an antibiotic resistance cassette. This method relies on recombination between homologous regions of plasmid and chromosomal DNA with replacement of the chromosomal DNA by plasmid-derived DNA as occurs in *C. jejuni* and *H. pylori (*Ferrero *et al*., 1992; Haas *et al*., 1993). In this way kanamycin and erythromycin but not chloramphenicol antibiotic resistance cassettes were inserted into the 23S rRNA gene on the *H. pullorum* chromosome. It is therefore possible to introduce DNA onto the chromosome by electroporation and homologous recombination. This provides a way to 'knock-out' certain *H. pullorum* genes and observe the effect on phenotype and also to insert genes into the *H. pullorum* chromosome in order to express genes of interest. We now have available two genetic markers encoding kanamycin and erythromycin resistance for future genetic approaches. Attempts to insert the chloramphenicol resistance cassette were unsuccessful. The chloramphenicol resistance cassette encodes an acetyltransferase gene, the product of which is able to inactivate chloramphenicol. It is possible that either this acetyltransferase is not functional in *H. pullorum* or it is detrimental to the *H. pullorum* cells.

H. pullorum was the first bacterial species in which two distinct genes encoding oligosaccharyltransferases (*pgl*B1 and *pgl*B2) were identified (Jervis *et al*., 2010). Most Campylobacter species have only one *pgl*B gene but recently a second *pgl*B gene was identified in *C. gracilis*, *C. curvus*, and *C. concisus* (Jervis *et al*., 2010). In this chapter, reverse transcriptase PCR demonstrated that both *H. pullorum pgl*B1 and *pgl*B2 genes are expressed under *in vitro* growth conditions and the role of *H. pullorum pgl*B2 in the *pgl*B1 dependent N-glycosylation system was investigated. The *H. pullorum pgl*B1 gene was previously knocked-out by insertional mutagenesis using an antibiotic resistance cassette but it was thought that the *H. pullorum pgl*B2 gene was essential due to difficulties creating a knock-out mutant (Jervis *et al.*,

2010). In this chapter I have now shown that it is possible to knock-out the *H. pullorum pgl*B2 gene by insertional mutagenesis but interestingly this was over 100 times less efficient than constructing a knock out mutation of *pgl*B1. The reason for this is unclear as the same experimental conditions were used for *pgl*B1 and *pgl*B2 mutagenesis with the same kanamycin resistance cassette and the *pgl*B2 flanking regions were actually larger than those of *pgl*B1. It is possible that the effect is region specific or perhaps due to an as yet undefined suppressor mutation in another gene.

The role of *H. pullorum* PglB2 in the *H. pullorum* PglB1 dependent N-linked protein glycosylation system was explored in this chapter. An *in vitro* assay was used to observe the N-glycosylation of a small peptide by *H. pullorum* membranes derived from a strain without a functional PglB2. Tricine gel electrophoresis and analysis of this peptide revealed a glycopeptide with electrophoretic mobility similar to that of glycopeptide derived from wild type *H. pullorum*. Purification and MALDI tandem Mass Spectrometry of this peptide indicated a pentasaccharide glycan structure identical to the known *H. pullorum* N-glycan (Jervis *et al*., 2010). This suggests that the *H. pullorum* PglB2 is not involved in the *H. pullorum* PglB1 dependent N-glycosylation system.

In summary, identification of suitable genetic markers and genetic manipulation techniques has laid the platform for an in depth mutational investigation of the N-linked glycosylation process in *H. pullorum*. Initial data indicated that knock out of the *H. pullorum pgl*B2 gene which was previously thought to be essential, is in fact possible. However, insertional mutagenesis of the *pgl*B2 gene was much less efficient than *pgl*B1, resulting in over 100-fold fewer kanamycin resistant transformants. It is possible that although the *pgl*B2 gene is normally essential, perhaps in approximately 1 in 100 cells, another protein can compensate for the absence of PglB2, but this remains to be determined. As glycopeptide from *in vitro* experiments with *pgl*B2 mutant membranes was comparable to that of glycopeptide from wild type strains, this suggests another role for PglB2. PglB2 may be involved in a second N-glycosylation system either completely separate from or overlapping with the PglB1 dependent system in terms of glycan and protein targets. Apart from

H. pullorum, *H. canadensis* and *H. winghamensis*, the only other bacterial species identified with two pglB OSTs are *C. curvus*, *C. concisus* and *C. gracilis* (Jervis *et al*., 2012) and work is currently ongoing to characterize these systems which may lead to predictions about the role of the Helicobacter PglB2 in N-linked protein glycosylation.

Chapter 4

Identification of Putative *H. pullorum* Glycoproteins and Their Glycosylation in *E. coli* with the *C. jejuni* Heptasaccharide

4.1 Background

The previous chapter described how an *in vitro* peptide glycosylation assay can be used to generate and subsequently structurally characterize N-linked glycan structures. A limitation of this approach is that there is no formal evidence that these glycan structures correspond to the glycan N-linked to glycoproteins in the cell. Therefore I sought to provide structural data on the N-linked glycan linked to a *H. pullorum* glycoprotein and this required purification and most critically identification of a glycoprotein.

In *C. jejuni*, experiments involving sugar-binding proteins known as lectins were used to identify putative glycoproteins (Linton *et al*., 2002; Young *et al*., 2002). The *C. jejuni* N-linked glycan has a terminal GalNAc residue which binds to the lectin, SBA. Only three known *Helicobacter* species contain homologues of the *C. jejuni pgl*B gene: *H. pullorum*, *H. canadensis* and *H. winghamensis*. Despite screening with an array of lectins with varying specificities, none were shown to bind to N-glycosylated proteins in these species (unpublished data, Linton laboratory). I therefore initiated a bioinformatics-based approach to identify candidate *Helicobacter* N-linked glycoproteins that could be used as a tool to further characterize *H. pullorum* N-linked glycosylation.

The *C. jejuni* PglB spans the bacterial inner membrane and N-glycosylation occurs on folded proteins in the periplasm (Nita-Lazar *et al*., 2005). Therefore a common feature of *H. pullorum* glycoproteins is likely to be a signal peptide sequence to direct secretion across the inner membrane. For N-linked protein glycosylation in *C. jejuni* to occur, the acceptor protein must also have at least one conserved amino acid N-glycosylation sequon. As with Eukaryotic organisms, the sequon consists of an asparagine amino acid residue followed by any amino acid except proline and then either a serine or threonine residue (N-X-S/T). Unlike Eukaryotes, for efficient glycosylation, the *C. jejuni* acceptor protein must also have an extended D/E-X1-N-X2-S/T sequon with acidic residues at the N-2 region (Chen *et al.*, 2007; Kowarik *et al*., 2006). Although PglB glycosylation can occur when X1 and X2 is any amino acid except proline, the enzyme favours alanine, serine and positively charged

residues at position X2 and aromatic and amido residues at position X1 (Chen *et al*., 2007). Recent data has also demonstrated the requirement of a similar extended glycosylation sequon in an *in vitro* peptide OST assay with membranes from *H. pullorum* (Jervis *et al*., 2010). It was therefore predicted that glycoproteins in *H. pullorum* would also have this acceptor protein requirement for glycosylation with PglB1 and possibly also PglB2. In this chapter, the alternative method devised to predict glycoproteins in *Helicobacter* species was to examine all predicted proteins in *H. pullorum, H. canadensis* and *H. winghamensis* and identify those that have these features; namely a signal peptide and an extended glycosylation sequon.

A plasmid (p*pgl*$_{Cj}$) containing all the *C. jejuni* genes required for N-glycosylation with a heptasaccharide glycan was expressed in *E. coli* to produce functional enzymes that glycosylated a *C. jejuni* glycoprotein (Wacker *et al*., 2002). This model system provided a simple tool for identifying the roles of genes involved in *C. jejuni* N-linked protein glycosylation system by insertional mutagenesis of *pgl* genes and identification of glycan structure by Mass Spectrometry (Linton et al., 2005). The genes proposed to be involved in N-glycosylation in *Helicobacter* species however are not found at a single locus therefore it was more challenging to create a similar model system by reconstituting *H. pullorum pgl* genes in *E. coli*. The development of an alternative method to confirm the N-glycosylation of *Helicobacter* proteins using the p*pgl*$_{Cj}$ plasmid and plasmids expressing *C. jejuni*, *H. pullorum* or *H. canadensis pgl*B genes is described in this chapter.

4.2 Results

4.2.1 Bioinformatic approach to predicting N-linked glycoproteins in *Helicobacter* species

The *H. pullorum, H. canadensis* and *H. winghamensis* genomes each encode over 2000 predicted proteins. Each of these protein sequences were analysed for the presence of an extended N-glycosylation sequon (D/E-X-N-X-S/T) and each protein with at least one such sequon was also examined using bioinformatics software (SignalP 4.0) to find those with a signal peptide sequence. Membrane proteins were excluded to avoid the difficulties associated with purification of insoluble proteins. In total, 131 putative *Helicobacter* glycoproteins were identified with 67 of these present in *H. pullorum* NCTC 12824^{T} and *H. pullorum* 98-5489 (See Appendix IV for *H. pullorum* glycoprotein list). Only 15 putative *Helicobacter* glycoproteins were present in all three PglB-containing *Helicobacter* species. Of these, five also had sequence similarity to known *C. jejuni* N-linked glycoproteins (Scott *et al*., 2011). Four of these proteins were putative periplasmic proteins with unknown function (HPMG_00410, HPMG_01281, HPMG_01484 and HPMG_01601; nomenclature based on the *H. pullorum* 98-5489 genome sequence) and one was the paralyzed flagellum protein, PflA (HPMG_00374). Two of four *H. pullorum* periplasmic proteins of unknown function, HPMG_01281 and HPMG_00410, were named HgpA and Hp0114 respectively, and further analyzed.

The predicted Helicobacter glycoprotein HPMG_01281 has an overall level of sequence identity of 30% to the *C. jejuni* glycoprotein CgpA, (Linton *et al*., 2002) and was therefore named HgpA. The *H. pullorum, H. canadensis* and *H. winghamensis* HgpA proteins have a single predicted N-glycosylation sequon located at the same position on each protein (Figure 4.1) but *C. jejuni* CgpA has four N-glycosylation sequons, three of which were previously experimentally verified (Scott *et al*., 2011).

The predicted Helicobacter glycoprotein HPMG_00410 has an overall level of sequence identity of 30% to the *C. jejuni* glycoprotein, Cj0114 (Scott *et al*., 2011, Young *et al*., 2002) and was therefore names Hp0114. The predicted 0114

glycoproteins from *H. pullorum, H. canadensis, H. winghamensis* and *C. jejuni* have at least one extended N-glycosylation sequon (Figure 4.1). A protein named YbgF, which is conserved in most Gram negative Bacteria, including *E. coli*, contains a C-terminal tetratricopeptide repeat domain involved in protein-protein interaction (Krachler *et al*., 2010) that is also present in Helicobacter and Campylobacter 0114 proteins. The genes encoding YbgF in *E. coli* and 0114 in Helicobacter and Campylobacter species are located downstream of a gene encoding a peptidoglycan associated lipoprotein which is part of well conserved cluster of genes producing proteins involved in the Tol/Pal system. This system spans the inner and outer membranes and is important for maintaining the structural integrity of the bacterial cell (Vianney *et al*., 1996).

The two candidate *H. pullorum* glycoproteins, HgpA and Hp0114 are further investigated in this chapter to confirm that they are glycosylated and to investigate their glycosylation sites when they are produced in *E. coli* along with the *C. jejuni* glycosylation machinery and PglB's from *C. jejuni, H. pullorum* or *H. canadensis.*

Figure 4.1 Sequence similarities of *C. jejuni* and *H. pullorum* potential glycoproteins. (A) Alignment of part of the *C. jejuni* CgpA and *H. pullorum* HgpA amino acid sequences using bioinformatic software (ClustalW2). (B) Alignment of *C. jejuni* and *H. pullorum* 0114 amino acid sequences**.** Only one N-glycosylation sequon is present at the same position on both glycoproteins. Symbols indicating identical amino acids (*), conserved amino acid substitutions (:) and semi conserved substitutions (.) are shown below the amino acid. Predicted N-glycosylation sequons are underlined.

(A)

```
Cj            -----MKTRILAIFFIFTSLLYADEN---PFKTDQNITLVAPPEFQKEEVKFNSSARILK 52
Hp            IERIRTTMKILLPLLIPFFILYAQDSQTLPKIPDITIDTSTTPKEIPTPNENNDTKEIRN 60
                    . :**  ::*   :***::.   *  .* .*   :.*:      : *.: .* :

Cj            ----SITFNYINLDGSEDKIDLDVNKSIDWHDTYTISRFKSPDPSKVLDVSVTIPEKNSS 108
Hp            PFESVVTPKQSGQISNPPQLSLFTQTTLNLPSTARKIKKITLAYQNLDGSISTIEQELNG 120
                   :* :  .  ..  ::.* .:.:::  .*    :  :   .:: .   ** :: ..

Cj            KQESNS----TANVEIPLQVAKIYDF-----ISYAVYKNKIKLNTSDEMITDFSVGNPSK 159
              DIDWHFPLILSQEIKPEMQSQKLQDFNLGKIFDFHIDKQKITLKTPLNMIRDFTLASPTR 180
              . : :     : :::  :*  *: **     :.: : *:**.*:*. :** **::..*::

Cj            IVIDFRSKMISPTKNIRLSNS-IFKRIDFGSHKGYYRLVIYLDGTYNYNIQ-KDATGYMI 217
Hp            LILDFKSNSKKVLQDFIQTNLPIISKVDLQTHLDFYRITFTLDGQYKYSIKNLKEKGLEI 240
              :::**:*:  .  :::  :*  *:.::*: :* .:**:.: *** *:*.*:  . .*  *

Cj            NLL 220
Hp            ELF 243
              :*:
```

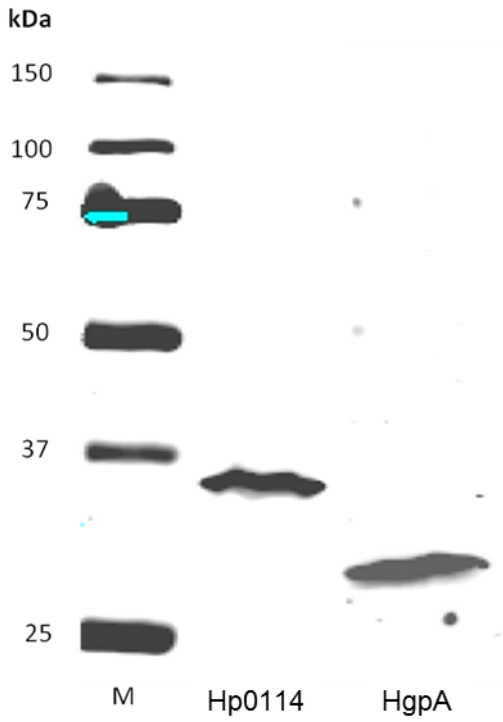
(B)

```
Cj            ---MKKIFTVALLGATLLYAESSAFGAGDITSNSSYGLTSNEKLFKEKLDNLNNENIQTN 57
Hp            MKLQKPIIASALLISFLYGVEPSAFNAGGDAP-----LKTETQILNERLFNLSN------ 49
                  * *:: *** : *  .*.***.**. :.     *.:: ::::*:* **.*

Cj            ARINEINERIEGLQSTLEGINSQYAKSNSRLSQVEENN----QNIENNFTSEIQKLKAYV 113
Hp            -KVKMLDESQEGLKSVFEGQIKRMQEFSSQLSQMADENNATISQMKEQVNTNFALQNENI 108
               ::: ::*  ***:*.:**  .:  : .*:***: ::*    .::::::.:::   :  :

Cj            EESRKIQEANNKQVKKVLAELSSLVDAINANYVSKNE---LNDANLSVKTITPSVVVSTT 170
Hp            EKLQASIVALGDLVQKVNAQTKVEIEALKKAILDDIESQTLTDETANKEVKKDTIKEDSV 168
              *: :    * .. *:** *: .  ::*::   :.. *   *.* . . :. . ::  .:.

Cj            DSNSTIENN--NTQNTQDDKAKQIDESWKKKKNNEILELAIKDVDKNAFEDSKAKLNFLI 228
Hp            ENNGSAPNANANIATIESAENNQKESKQPEKSLAEFFAEGEKLLEEKKYKLADEYLQTAI 228
              :.*.:  *   *  . :. : :* :..  :*.  *::  . * :::: :: :.  *:  *

Cj            TKQYKPARANFWLGEIEYKQKNYNNAIVYYKKSSSLSTKGDYFPKLLYHTAISLDKTGDT 288
Hp            KGHYKPARGNYLLGEIAFAQGRYEEAIYYYKTSATRYDKADYMPRLMLNSAKSFEKINDK 288
              . :*****.*: **** : * .*::** ***.*::   *.**:*:*: ::* *::* .*.

Cj            KTANGFYKALKTNYPNSPEAKASPNRK--- 315
Hp            ENAKKFLESLIALYPDSNEAKEAGNLLIKK 318
```

4.2.2 Production of *H. pullorum* HgpA and Hp0114 putative glycoproteins in *E. coli*

The *H. pullorum* Hp0114 and *hgp*A genes were amplified by PCR using Primers 34/35 and 36/37 respectively with SphI and BglII restriction sites introduced on the primers. The PCR products were digested and ligated into the multiple cloning site of pQE70. This resulted in addition of a hexa histidine tag onto the C terminus of the protein produced. Plasmids with inserts of the correct size were checked by sequencing. Plasmids, named pQE0114 and pQE*hgp*A, were transformed into *E. coli* and 50 ml cultures were grown overnight (16 hours). The production of Hp0114 or HgpA by these cells was induced by the addition of IPTG. Whole cell extracts were separated by SDS-PAGE and western blotted. Histidine tagged proteins were detected with mouse anti-histidine primary antibody and anti-mouse secondary antibody conjugated to infra-red dye (LI-COR biosciences). Both the Hp0114 and HgpA proteins were visualized using an Odyssey infrared imaging system. The predicted size of the Hp0114 and HgpA proteins corresponds to the observed bands in Figure 4.2 indicating that both proteins are produced.

Figure 4.2 Detection of Hp0114 and HgpA proteins produced in *E. coli*. Western blot analysis of Hp0114 and HgpA proteins expressed in *E. coli* following 16 hours incubation in broth cultures containing ampicillin. The first lane (M) contains a protein molecular weight marker. The predicted sizes of *H. pullorum* hexa histidine tagged Hp0114 and HgpA proteins lacking a signal peptide sequence are 33.1 kDa and 26.1 kDa respectively.



4.2.3 Glycosylation of *H. pullorum* HgpA and Hp0114 in *E. coli* with a *C. jejuni* heptasaccharide glycan.

In order to test if the Hp0114 and HgpA proteins could be glycosylated, a relatively simple system was used whereby the *C. jejuni* glycosylation genes were expressed in *E. coli* along with the *Hp*0114 or *hgp*A gene and glycoprotein detected by western blot. A version of the p*pgl*$_{Cj}$ plasmid termed p*pgl*ΔB_{Cj} contains all the genes required for the production of a *C. jejuni* lipid-linked heptasaccharide but has an insertionally inactivated *pgl*B gene (Linton *et al*., 2005). This plasmid was electroporated into electrocompetent *E. coli* along with plasmid pQE0114 or pQE*hgp*A. Plasmids were previously constructed containing *pgl*B genes sequences within a pMLBAD vector for the arabinose inducible expression of *C. jejuni pgl*B (pMAF10; Feldman *et al*., 2005), *H. pullorum pgl*B (pMLHP1 and pMLHP2; Jervis *et al*., 2010) and *H. canadensis pgl*B genes (pMLHC1 and pMLHC2; unpublished). These plasmids were electroporated into *E. coli* cells containing p*pgl*ΔB_{Cj} and *H. pullorum* pQE0114 or pQE*hgp*A and strains containing all three plasmids were selected. Each strain was grown separately and *pgl*B and *Hp*0114 or *hgp*A genes were induced overnight with IPTG and *pgl*B genes with arabinose. Protein from whole cell lysates was separated by SDS-PAGE and HgpA or Hp0114 protein detected by western blotting using anti penta histidine antibody. The *C. jejuni* N-linked glycan specific rabbit antiserum, hR6 (Schwarz *et al*., 2011b), was used to detect transfer of heptasaccharide glycan onto protein.

Co-expression of *pgl*B_{Cj} and *pgl*$B1_{Hp,}$ but not *pgl*$B2_{Hp}$ with plasmid p*pgl*ΔB_{Cj} was associated with production of hR6 reactive and decreased mobility form(s) of both HgpA (Figure 4.3) and Hp0114 (Figure 4.4). This provides preliminary evidence that both HgpA and Hp0114 are N-linked glycoproteins. These data indicate that $PglB1_{Hp}$ but not $PglB2_{Hp}$ is capable of transferring the *C. jejuni* heptasaccharide onto these two putative *H. pullorum* glycoproteins. It was also demonstrated that HgpA could similarly be glycosylated by the *H. canadensis* PglB1 but not PglB2 (Figure 4.5) and neither the *H. canadensis* PglB1 nor PglB2 were able to modify the Hp0114 protein.

Figure 4.3 Glycosylation of *H. pullorum* HgpA expressed in *E. coli* with a *C. jejuni* heptasaccharide glycan. *H. pullorum* HgpA protein from *E. coli* producing *C. jejuni* Pgl proteins and one of $PglB1_{Hp}$, $PglB2_{Hp}$ or $PglB_{Cj}$. (A) Green bands indicate *C. jejuni* glycan detected with glycan-specific antiserum (hR6) whilst HgpA was detected with anti penta histidine antibody (anti-his) and is red. Glycosylated HgpA protein bands appear yellow (green and red co localised). (B) HgpA protein detected with anti penta histidine antibody only. (C) Glycan detected with hR6 antiserum. HgpA protein is glycosylated by $PglB1_{Hp}$ and $PglB_{Cj}$ but not $PglB2_{Hp}$.

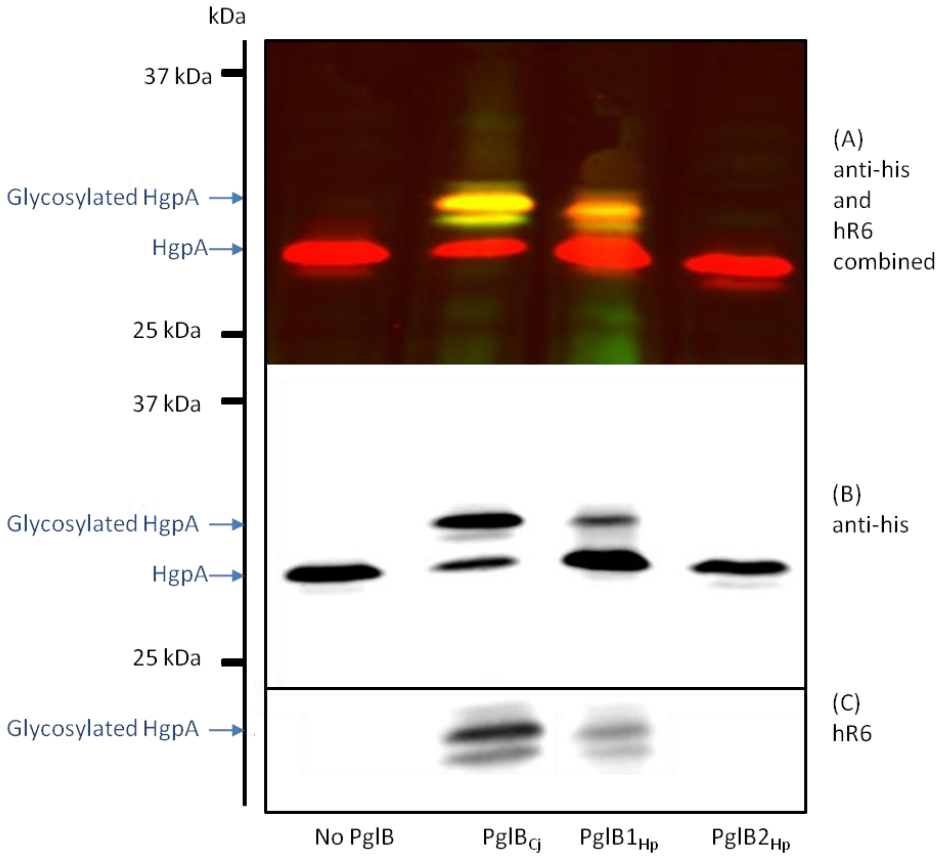


Figure 4.4 Glycosylation of Hp0114 expressed in *E. coli* with a *C. jejuni* heptasaccharide glycan. Hp0114 protein from *E. coli* producing *C. jejuni* Pgl proteins and one of $PglB1_{Hp}$, $PglB2_{Hp}$ or $PglB_{Cj}$. (A) Green bands indicate *C. jejuni* glycan detected with glycan-specific antiserum (hR6) whilst Hp0114 protein was detected with anti penta histidine antibody (anti-his) and is red. Glycosylated Hp0114 protein bands appear yellow (green and red co localised). (C) Glycan detected with hR6 antiserum. Hp0114 protein is glycosylated by $PglB1_{Hp}$ and $PglB_{Cj}$ but not $PglB2_{Hp}$.

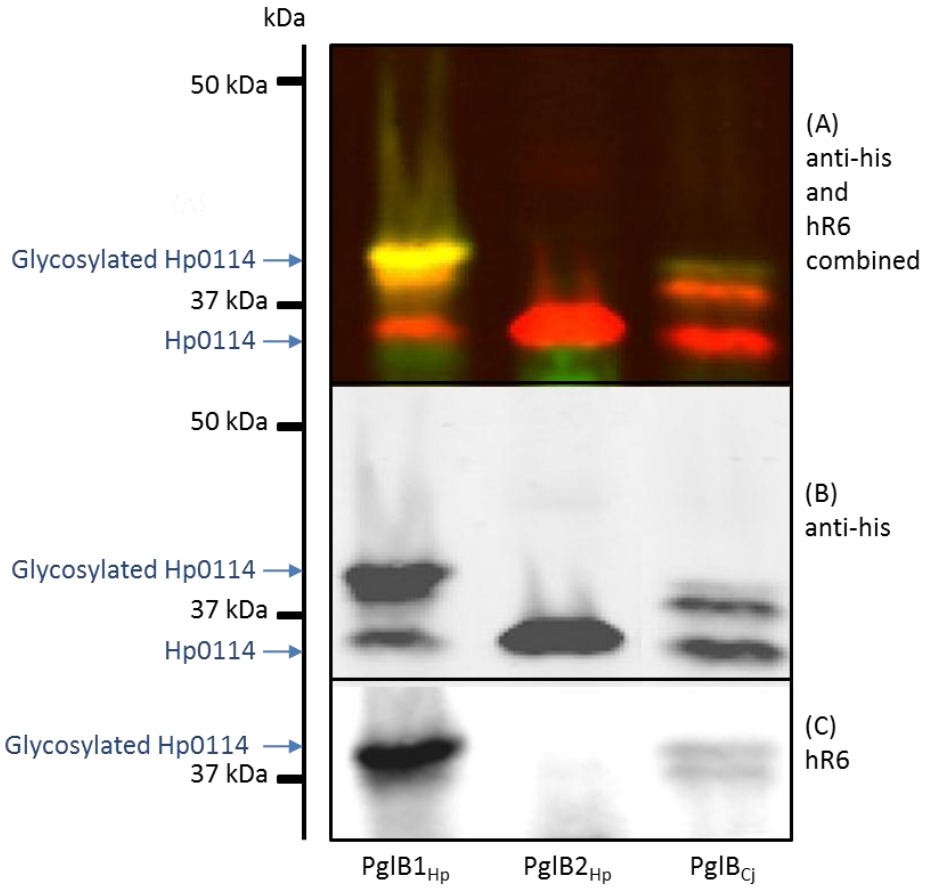
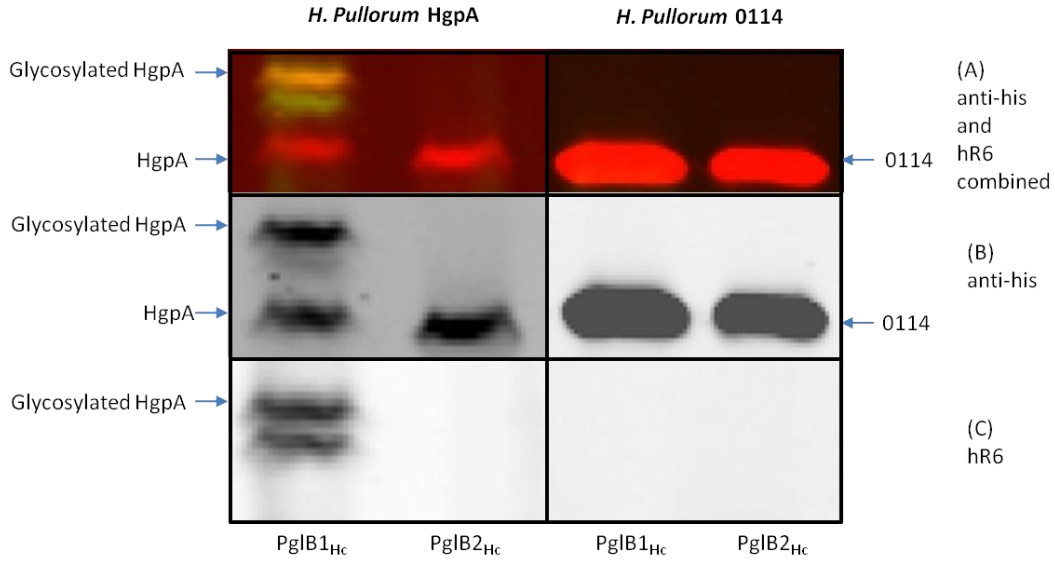


Figure 4.5 Glycosylation of *H. pullorum* HgpA expressed in *E. coli* with a *C. jejuni* heptasaccharide glycan by *H. canadensis* PglB's. *H. pullorum* HgpA and Hp0114 protein from *E. coli* producing *C. jejuni* Pgl proteins and $PglB1_{Hc}$ or $PglB2_{Hc.}$ (A) Green bands indicate *C. jejuni* glycan detected with glycan-specific antiserum (hR6) whilst HgpA and Hp0114 proteins were detected with anti penta histidine antibody (anti-his) and are red. Glycosylated HgpA protein bands appear yellow (green and red co localised). (B) HgpA and Hp0114 protein detected with anti penta histidine antibody only. (C) Glycan detected with hR6 antiserum. HgpA protein is glycosylated by $PglB1_{Hc}$ but not $PglB2_{Hc}$ and Hp0114 protein is not glycosylated by either *H. canadensis* PglB.



4.2.4 Localisation of the site of glycosylation of Hp0114

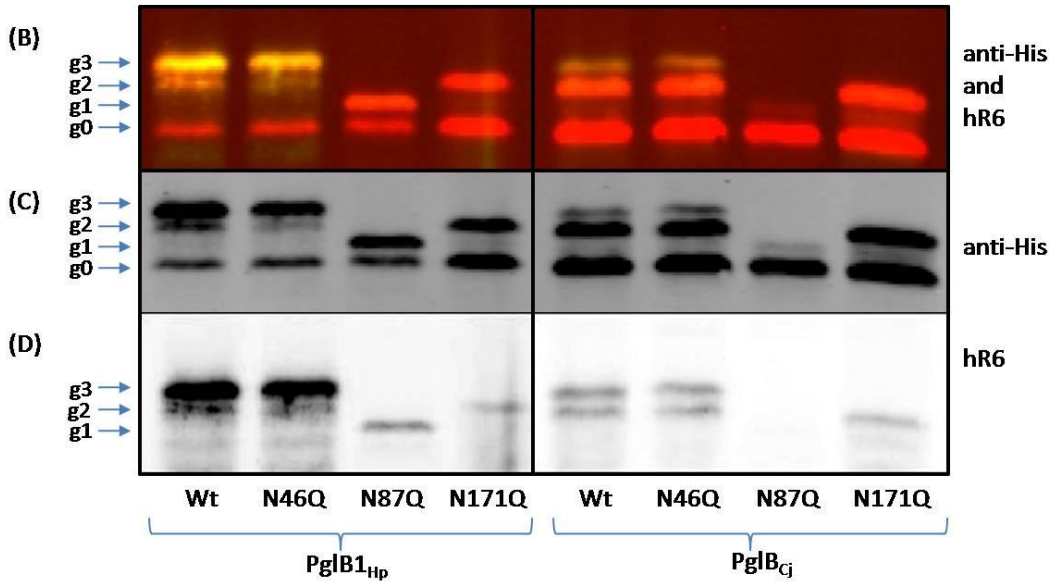
The ease of manipulation of the *E. coli* based glycosylation system allowed the identification of the N-glycosylation sites of the *H. pullorum* HgpA and Hp0114 glycoproteins by site directed mutagenesis. The Hp0114 protein has two predicted D/E-X-N-X-S/T N-glycosylation sequons (Figure 4.1 B) and one further N-X-S/T sequon but it was not known if each of these sites were glycosylated. In comparison, Cj0114 glycoprotein has four extended glycosylation sequons (Figure 4.1 B) but only two of these sequons are glycosylated in *E. coli* expressing the *C. jejuni pgl* gene cluster (Jervis *et al*., 2010). In order to investigate which asparagine residue(s) of Hp0114 (See Figure 4.1) can be glycosylated when expressed in *E. coli*, all three located within N-X-S/T sequons were changed to glutamine (which is similar in structure to asparagine) by site directed mutagenesis of pQE0114. These three plasmids (pQE0114N46Q, pQE0114N87Q and pQE0114N171Q) were each transformed into *E. coli* strains carrying plasmids encoding p*pgl*ΔB_{Cj} and either *pgl*B_{Cj} or *pgl*$B1_{Hp}$. The Hp0114 proteins from these *E. coli* strains were analysed by western blot with an antiserum specific for the hexa histidine tag of Hp0114 and the hR6 antiserum specific for the *C. jejuni* glycan (Figure. 4.6).

Mutation of the N46 asparagine residue resulted in a glycosylated form of the Hp0114 protein with the same mobility as non-mutated Hp0114 when produced with both $PglB_{Cj}$ and $PglB1_{Hp}$ indicating that this site is not essential for N-glycosylation. In contrast, mutation of either of the extended N-glycosylation sequons (N83 and N171) of Hp0114, resulted in electrophoretic mobility changes when compared to non-mutated Hp0114 produced with either $PglB_{Cj}$ and $PglB1_{Hp}$ (Figure 4.6). This indicated that both of these sites are glycosylated by *C. jejuni* PglB and *H. pullorum* PglB1. The N87Q mutation in the presence of either $PglB1_{Hp}$ or $PglB_{Cj}$ resulted in a monoglycoaylated form of the Hp0114 protein whereas mutation of N171Q resulted in a diglycosylated form. Slight differences were noted in efficiency of glycosylation with the two PglB enzymes. Monoglycosyation of the N87Q protein in the presence $PglB1_{Hp}$ was much more efficient than glycosylation by $PglB_{Cj.}$ In contrast, diglycosylation of N171Q was efficient with both PglB enzymes.

Figure 4.6 Sites of glycosylation of Hp0114 expressed in *E. coli* with a *C. jejuni* heptasaccharide glycan. (A) Amino acid sequence of wild type and mutated Hp0114 N-glycosylation sequons. (B) Hp0114 protein from *E. coli* producing *C. jejuni* Pgl proteins and $PglB1_{Hp}$ or $PglB_{Cj.}$ Green bands indicate *C. jejuni* glycan detected with glycan-specific antiserum (hR6) whilst Hp0114 protein was detected with anti penta histidine antibody (anti-his) and is red. Glycosylated Hp0114 protein bands appear yellow (green and red co localised). (C) Hp0114 protein detected with anti penta histidine antibody only. (D) Glycan detected with hR6 antiserum. Glycosylation of Hp0114 protein (g1-g3) with $PglB1_{Hp}$ or $PglB1_{Cj}$ was affected by mutation of N87 and N171 asparagine residues but not N46.

(A)

	Wild type sequon	Mutated sequon
N46Q	RLFNLS	RLFQLS
N87Q	DENNAT	DENQAT
N171Q	VENNGS	VENQGS



4.2.5 Amino acid site of glycosylation of *H. pullorum* HgpA

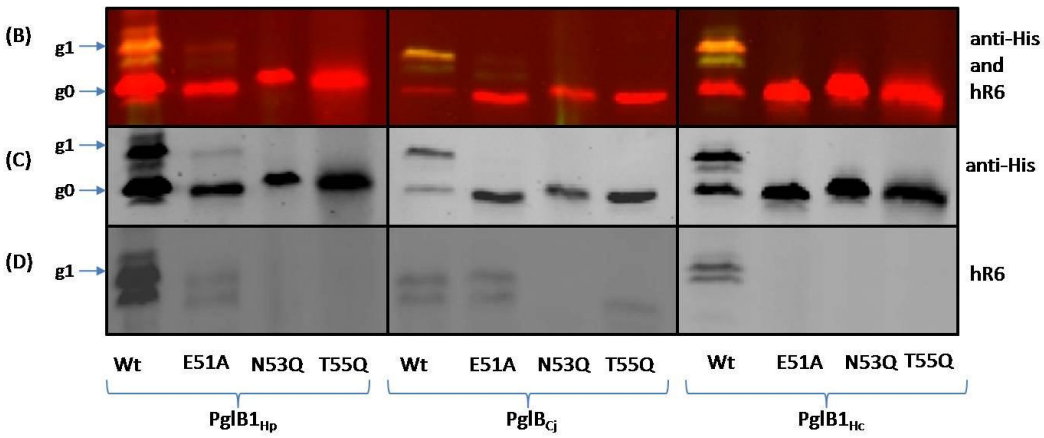
The *H. pullorum* HgpA amino acid sequence has only one D/E-X-N-X/S/T glycosylation sequon located at N53 on the protein ($_{51}$ENNDT$_{55}$). To confirm that this was the sequon that was glycosylated in *E. coli*, the N53 asparagine residue of the *hgp*A gene from the pQEHgpA plasmid was changed to glutamine, (N53Q, Figure 4.7) by site directed mutagenesis. As previously, the *hgp*A on this plasmid was then co-expressed in *E. coli* with p*pgl*ΔB$_{Cj}$ and *pgl*B1$_{Hp,}$ *pgl*B1$_{Hc}$ *or pgl*B$_{Cj}$ and western blotting with antisera to HgpA and the *C. jejuni* glycan indicated that the HgpA N53Q protein had mobility consistent with unglycosylated HgpA (Figure. 4.7) confirming that this sequon contains the N-glycosylated asparagine residue.

Based on previous work (Kowarik *et al*., 2006; Chen *et al.*, 2007), it was likely that the glutamic acid and threonine residues of this HgpA N-glycosylation sequon were also required for glycosylation. To confirm this, the glutamic acid residue (E51) was changed by site directed mutagenesis to alanine and the threonine residue (T55) to glutamine. The HgpA T55Q protein was no longer glycosylated in *E. coli* by PglB1$_{Hp}$, PglB1$_{Hc}$ or PglB$_{Cj}$ enzymes indicating that threonine was essential for N-glycosylation (Figure 4.7). Some HgpA E51A protein was still glycosylated by PglB1$_{Hp}$ and PglB$_{Cj}$ though clearly these enzymes have a preference for a glutamic acid amino acid at this site (Figure 4.7).

Figure 4.7 Site of glycosylation of *H. pullorum* HgpA expressed in *E. coli* with a *C. jejuni* heptasaccharide glycan. (A) Amino acid sequence of the wild type and mutated HgpA N-glycosylation sequons. (B) *H. pullorum* HgpA protein with glutamic acid, asparagine and threonine mutations from *E. coli* producing *C. jejuni* Pgl proteins and one of $PglB1_{Hp}$ $PglB_{Cj}$ or $PglB1_{Hc.}$ Green bands indicate *C. jejuni* glycan detected with glycan-specific antiserum (hR6) whilst HgpA protein was detected with anti penta histidine antibody (anti-his) and is red. Glycosylated HgpA protein bands appear yellow (green and red co localised). (C) HgpA protein detected with anti penta histidine antibody only. (D) Glycan detected with hR6 antiserum. Glycosylation of *H. pullorum* HgpA protein (g1) is abolished by mutation of asparagine or threonine residues and reduced by mutagenesis of glutamic acid.

(A)

	Wild type sequon	Mutated sequon
E41A	NENNDT	NANNDT
N53Q	NENNDT	NENQDT
T55Q	NENNDT	NENNDQ



4.3 Discussion

In order to further investigate the *H. pullorum* N-glycosylation system, it was important to identify at least one *H. pullorum* glycoprotein. Although many N-linked glycoproteins from *C. jejuni* have been identified, none were known in any *Helicobacter* species. A bioinformatic based approach was used to identify potential glycoproteins from the PglB-containing *H. pullorum, H. canadensis* and *H. winghamensis*. These predictions were based on the protein having a signal peptide sequence for secretion across the inner membrane and also at least one extended D/E-X-N-X-S/T glycosylation sequon. From this bioinformatic analysis, 131 putative *Helicobacter* glycoproteins were identified and 67 of these were present in *H. pullorum* (See Appendix IV for *H. pullorum* glycoprotein list), but only 15 of these were present in *H. pullorum, H. canadensis* and *H. winghamensis*. Of the five proteins that also had sequence similarity to previously characterized *C. jejuni* N-linked glycoproteins, two were chosen to further investigate glycosylation, *H. pullorum* HgpA and Hp0114 (Figure 4.1).

Histidine-tagged versions of these two predicted *H. pullorum* N-glycoproteins were produced in *E. coli* with *pgl* genes for the biosynthesis of a *C. jejuni* lipid-linked heptasaccharide and either *H. pullorum, H. canadensis* or C. *jejuni* PglB's. These experiments indicated that HgpA and Hp0114 proteins were glycosylated by $PglB1_{Hp}$ and $PglB_{Cj}$ but not $PglB2_{Hp}$ (Figure 4.3 and 4.4). These results were consistent with those from Chapter 3 which show that the $PglB1_{Hp}$ is essential for the N-glycosylation of a peptide with an extended N-glycosylation sequon but mutation of the gene encoding PglB2 does not affect this glycosylation. Interestingly, $PglB1_{Hc}$ but not $PglB2_{Hc}$ was also able to glycosylate the *H. pullorum* HgpA glycoprotein in *E. coli* (Figure 4.5). This was the first evidence that the *H. canadensis* PglB1 functions as an OST that can N-glycosylate proteins. The function of the *H. pullorum* and *H. canadensis* PglB2 remains unknown and requires further investigation.

The *C. jejuni* PglB targets an extended N-glycosylation sequon but it was unclear whether this sequon was also essential for *Helicobacter* PglB activity. The Hp0114

glycoprotein has two extended D/E-X-N-X-S/T sequons and one N-X-S/T sequon. Data presented herein confirmed that in *E. coli*, the two Hp0114 D/E-X-N-X-S/T extended N-glycosylation sequons were modified by $PglB1_{Hp}$ and $PglB_{Cj}$ but the N-X-S/T sequon remained unmodified (Figure 4.6), indicating that $PglB1_{Hp}$ has similar amino acid sequon requirements to $PglB_{Cj}$. Interestingly, although *H. pullorum* and *C. jejuni* PglB's could both glycosylate Hp0114, $PglB1_{Hp}$ and $PglB_{Cj}$ appeared to have different preferences for the site of N-glycosylation. Previous studies have also indicated a difference in specificity of the two enzymes as when $PglB1_{Hp}$ was used to complement $PglB_{Cj}$ in *E. coli*, a *C. jejuni* glycoprotein was modified at two sites compared to four sites with the $PglB_{Cj}$ (Jervis *et al*., 2010). When the N87 asparagine residue was substituted for glutamate (Figure 4.6), the Hp0114 protein was predominantly glycosylated in strains with $PglB1_{Hp}$ but most Hp0114 protein remained unglycosylated in strains with the $PglB_{Cj}$. This indicates that $PglB1_{Hp}$ is more efficient at glycosylation of the N171 sequon than $PglB_{Cj}$. When the N171 asparagine residue was substituted (Figure 4.6), leaving the N87 sequon, both $PglB_{Cj}$ and $PglB1_{Hp}$ were still able to efficiently glycosylate the Hp0114 protein. Interestingly however, the modified protein had decreased mobility compared to when the N87 sequon was substituted. The N87 sequon ($_{84}$DENNAT$_{89}$) contains two asparagine residues and also an aspartic acid residue at the N-2 position from the second asparagine. It is possible therefore that this sequon is glycosylated at both asparagine residues which would explain the differences in modified protein mobility.

To confirm that the $_{51}$ENNDT$_{55}$ amino acid sequon of *H. pullorum* HgpA was glycosylated, the asparagine residue (N53) was substituted with glutamate (Figure 4.7). This altered HgpA protein was no longer modified when produced in *E. coli* with the *C. jejuni* heptasaccharide and $PglB1_{Hp}$, $PglB1_{Hc}$ or $PglB_{Cj}$. Further experiments focussed on determining the importance of glutamic acid and threonine residues. When threonine (T55) was substituted with glutamine, the HgpA protein had mobility consistent with unglycosylated protein indicating the importance of this threonine residue (Figure 4.7). A small proportion of HgpA protein was still modified by $PglB1_{Hp}$ and $PglB_{Cj}$ following substitution of the glutamic acid with

alanine but this glycosylation was significantly reduced, indicating that *H. pullorum* and *C. jejuni* both prefer this acidic residue but it is not absolutely essential.

In summary, Chapter 4 describes the identification of a number of potential *H. pullorum* N-linked glycoproteins and the confirmation that at least two of these proteins, Hp0114 and HgpA, are glycosylated in *E. coli* producing *C. jejuni* Pgl enzymes and either *H. pullorum* PglB1 or *C. jejuni* PglB but not *H. pullorum* PglB2. The amino acid sequon requirements of target *H. pullorum* glycoproteins were similar to those of *C. jejuni* but where the glycoprotein had multiple glycosylation sequons differences in the efficiency of glycosylation between the C. *jejuni* PglB and *H. pullorum* PglB1 were noted at each site. Interestingly, the *H. canadensis* PglB1 OST was also able to glycosylate the HgpA protein. This is the first evidence that the *H. canadensis* PglB1 may be involved in an N-linked protein glycosylation system. The identification of these *H. pullorum* N-linked glycoproteins was the basis for work presented in Chapter 5 investigating N-linked glycosylation in *H. pullorum* itself.

Chapter 5

Glycosylation of Hp0114 and HgpA in *H. pullorum*

5.1 Background

In Chapter 3, the *H. pullorum pgl*B1 and *pgl*B2 genes were insertionally inactivated and *in vitro* N-linked glycosylation of a peptide investigated using an OST assay (Jervis *et al*., 2010). This assay is particularly useful for the identification of N-linked glycan structures from PglB-producing Bacteria when there are no characterized glycoproteins (Jervis *et al*., 2010; Jervis *et al*., 2012). Chapter 4 describes the identification of two potential *H. pullorum* glycoproteins, HgpA and Hp0114, and demonstrates their glycosylation in *E. coli* by *C. jejuni*, *H. pullorum* and *H. canadensis* PglB's with the *C. jejuni* heptasaccharide glycan. With the identification of two putative *H. pullorum* glycoproteins this chapter aims to determine whether these proteins are glycosylated in *H. pullorum* and whether this is dependent on either or both of the *H. pullorum* oligosaccharyltransferases PglB1 and PglB2. The *H. pullorum* HgpA protein only has one N-glycosylation sequon but Hp0114 has two, and both of these sites are glycosylated in *E. coli*. In this chapter, Hp0114 N-glycosylation sites identified in *E. coli* (Chapter 4) are compared to those in *H. pullorum*. These would be the first definitively identified Helicobacter N-linked glycoproteins and would also be extremely useful for identification of the genes involved in biosynthesis of the N-linked glycan in Helicobacter species.

5.2 Results

5.2.1 Production of hexa histidine tagged Hp0114 and HgpA proteins in *H. pullorum*

The aim of this chapter was to use the techniques described in Chapter 3 for the genetic manipulation of *H. pullorum* along with the two potential *H. pullorum* glycoproteins identified in Chapter 4 to investigate N-linked protein glycosylation in *H. pullorum.* In order to identify Hp0114 and HgpA in *H. pullorum*, strains of *H. pullorum* were modified to produce additional hexa histidine tagged versions of Hp0114 and HgpA. The genes encoding these proteins were PCR amplified with primers (38/39 and 40/41 respectively) containing BamHI restriction sites (Figure 5.1 A) and reverse primers encoded a hexa histidine tag. The vector pHp23E does not replicate in *H. pullorum* and is designed to allow integration of genes, through recombination of DNA cloned between flanking 23S rRNA gene sequences, onto the *H. pullorum* chromosome (see Chapter 3). PCR and cloning techniques were used to create a version of the pHp23E vector with a BamHI restriction site downstream of the erythromycin resistance cassette allowing insertion of *Hp*0114 and *hgp*A genes to create pHp23E::0114 and pHp23E::*hgp*A respectively. Integration of the *Hp*0114 and *hgp*A genes into pHp23E in the same transcriptional orientation as the erythromycin and 23S rRNA genes was confirmed by PCR (Figure 5.1 B and C) and sequencing. These plasmids were electroporated into *H. pullorum* wild type, *pgl*B1::Kn and *pgl*B2::Kn mutants (Chapter 4). In this way both the *hgp*A and *Hp*0114 genes were integrated onto the chromosome downstream of the promoter associated with the erythromycin resistance gene. Insertion of both genes within a *H. pullorum* 23S rRNA sequence was confirmed by PCR (Figure 5.2).

Figure 5.1 Construction of plasmids pHp23E::0114 and pHp23E::*hgp*A. (A) PCR amplification of full length *Hp*0114 and *hgp*A genes. (B) Schematic diagram of a region of pHp23E::0114 and pHp23E::*hgp*A and location of the primers used to confirm the transcriptional orientation of *Hp*0114 and *hgp*A genes within these plasmids. Predicted PCR product sizes are indicated in red. (C) Products of a PCR reaction with primers from (B). Lane M contains a DNA molecular weight marker.

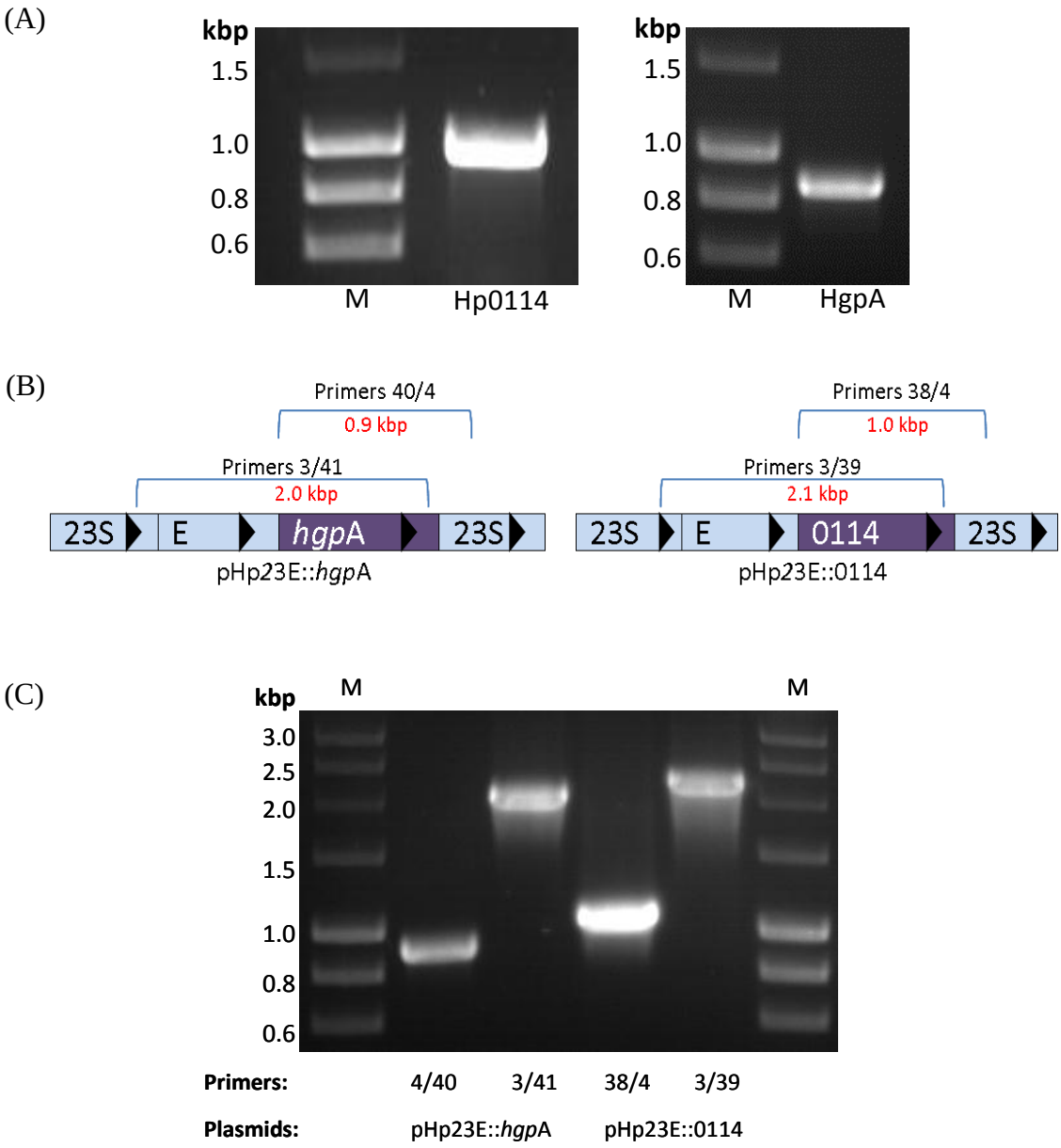
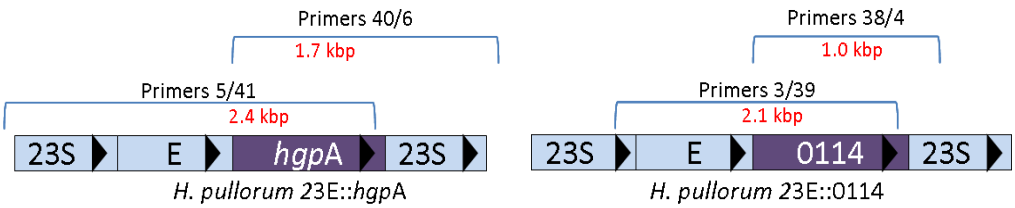
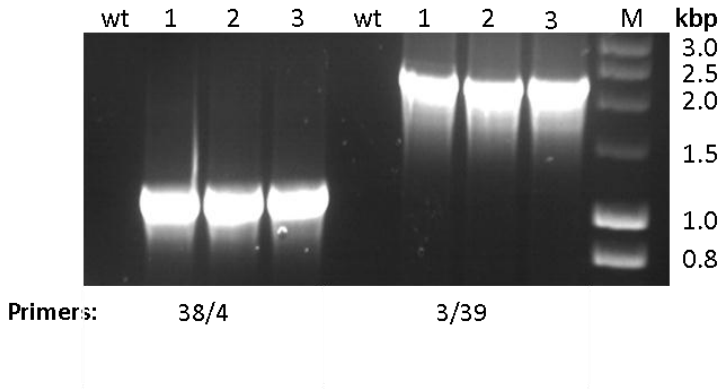


Figure 5.2 PCR verification of *H. pullorum* 23E::HgpA and 23E::0114 strains. (A) Schematic diagram of the arrangement of *hgp*A and *Hp*0114 genes in *H. pullorum* 23E::*hgp*A and 23E::0114 respectively with primers used to confirm that the *hgp*A or *Hp*0114 gene insert is in the same transcriptional orientation as the 23S rRNA gene. Predicted PCR product sizes are indicated in red. (B) PCR amplification of genomic DNA from *H. pullorum* wild type and three *H. pullorum* 23E::0114 strains (1-3) using primers from (A). Lane M contains a DNA molecular weight marker. (C) PCR amplification of genomic DNA from *H. pullorum* wild type and two *H. pullorum* 23E::*hgp*A strains (1 and 2) using primers from (A).

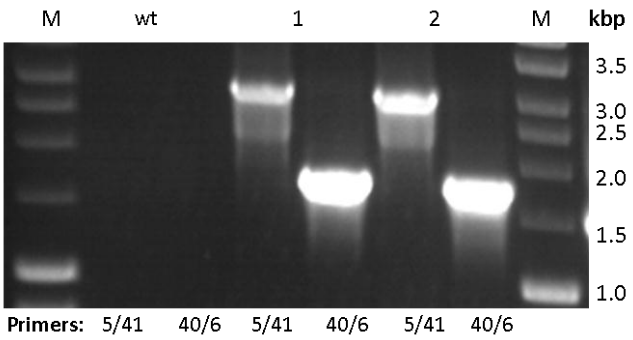
(A)



(B)



(C)

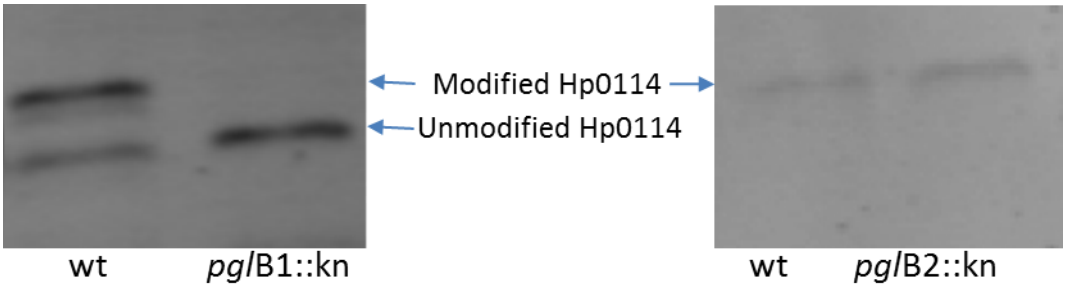


5.2.2 Detection of Hp0114 and HgpA proteins in *H. pullorum*

In order to purify histidine tagged Hp0114 and HgpA proteins, *H. pullorum* cells containing genes encoding these proteins were grown on erythromycin-containing blood agar plates, harvested and hexa histidine tagged Hp0114 and HgpA were purified (see Methods for details). Hp0114 and HgpA were also produced in and purified from *H. pullorum* *pgl*B1::Kn and *pgl*B2::Kn knock out mutants. The proteins were run on SDS-PAGE gels and analysed by western blot with anti histidine antiserum (Figure 5.3). Both HgpA and Hp0114 proteins were produced as lower mobility forms in the wild type and the *pgl*B2::Kn genetic background whilst in contrast, in a *pgl*B1::Kn mutant background, only the higher mobility forms are produced.

Figure 5.3 PglB1 but not PglB2, dependent glycosylation of *H. pullorum* HgpA and Hp0114 proteins. Western blot of Hp0114 (A) and HgpA (B) from *H. pullorum* wild-type, *pgl*B1::Kn and *pgl*B2::Kn mutants detected with anti histidine antiserum. The Hp0114 protein from the *H. pullorum pgl*B1::Kn mutant has increased electrophoretic mobility compared to Hp0114 from the wild type strain which is produced as two forms. In contrast, Hp0114 protein from the *pgl*B2::Kn mutant has comparable mobility to Hp0114 from the wild type strain. (B) *H. pullorum* HgpA from wild type, *pgl*B1::Kn or *pgl*B2::Kn mutant. As with Hp0114 protein, HgpA derived from a *pgl*B2::Kn mutant has comparable mobility to HgpA derived from wild type *H. pullorum* but HgpA derived from a *pgl*B1::Kn mutant has increased mobility suggesting this protein is unmodified.

(A)

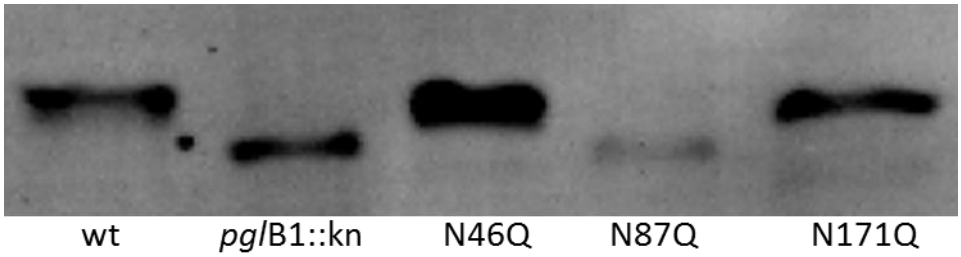


(B)

5.2.3 Site of glycosylation of Hp0114 in *H. pullorum*

The Hp0114 glycoprotein has two D/E-X-N-X-S/T glycosylation sequons, ENNAT (N87) and ENNGS (N171) and also one N-X-S/T sequon lacking an acidic group at the N-2 region (N46). When expressed in *E. coli*, *H. pullorum* PglB1 and *C. jejuni* PglB modified both N87 and N171 sites with the *C. jejuni* heptasaccharide (Chapter 5). In order to determine whether the same sites were glycosylated when the Hp0114 protein was produced in *H. pullorum*, the N46, N87 and N171 residues of Hp0114 within pHp23E::0114 were substituted with glutamine residues by site directed mutagenesis using primers 48/49, 50/51 and 52/53 respectively, creating plasmids pHp23E::0114N46Q, pHp23E::0114N87Q and pHp23E::0114N171Q. The introduction of glutamine for asparagine residues at each site was verified by sequencing and each plasmid was individually electroporated into *H. pullorum* cells. Following selection for recombination at the 23S rRNA gene by growth on erythromycin containing blood agar, the introduction of the mutated forms of the Hp0114 genes was verified by PCR and sequencing (data not shown). Strains, designated *H. pullorum* 23E::0114N46Q, 23E::0114N87Q and 23E::0114N171Q, along with the wild type strain, *H. pullorum* 23E::0114, were grown on blood agar plates and cells harvested. The hexa histidine tagged Hp0114 glycoproteins were partially purified using nickel coated magnetic beads (Qiagen) and detected by SDS-PAGE and western blotting (see Methods for details). The Hp0114 protein from *H. pullorum* 23E::0114N46Q and 23E::0114N171Q had comparable electrophoretic mobility to Hp0114 from *H. pullorum* 23E::0114 (Figure 5.4), suggesting that substitution of the N46 and N171 asparagine residues does not affect modification of the Hp0114 glycoprotein. In contrast, Hp0114 protein from 23E::0114N87Q had comparable electrophoretic mobility to Hp0114 from the *H. pullorum* 23E::0114 *pgl*B1::Kn strain (Figure 5.4), indicating that this N87 asparagine is the site of Hp0114 modification in *H. pullorum*.

Figure 5.4 Identification of the glycosylation site of the Hp0114 protein. Western blot of Hp0114 proteins with N46Q, N87Q and N171Q substitutions detected with anti histidine antiserum. Modified and unmodified Hp0114 glycoproteins from wild type *H. pullorum* (wt) and *pgl*B1::Kn mutants respectively are shown as controls. The N46Q and N171Q substituted Hp0114 protein had comparable electrophoretic mobility to the modified Hp0114 protein but the N87Q substituted Hp0114 protein had mobility consistent with unmodified Hp0114 from the *pgl*B1::Kn mutant.



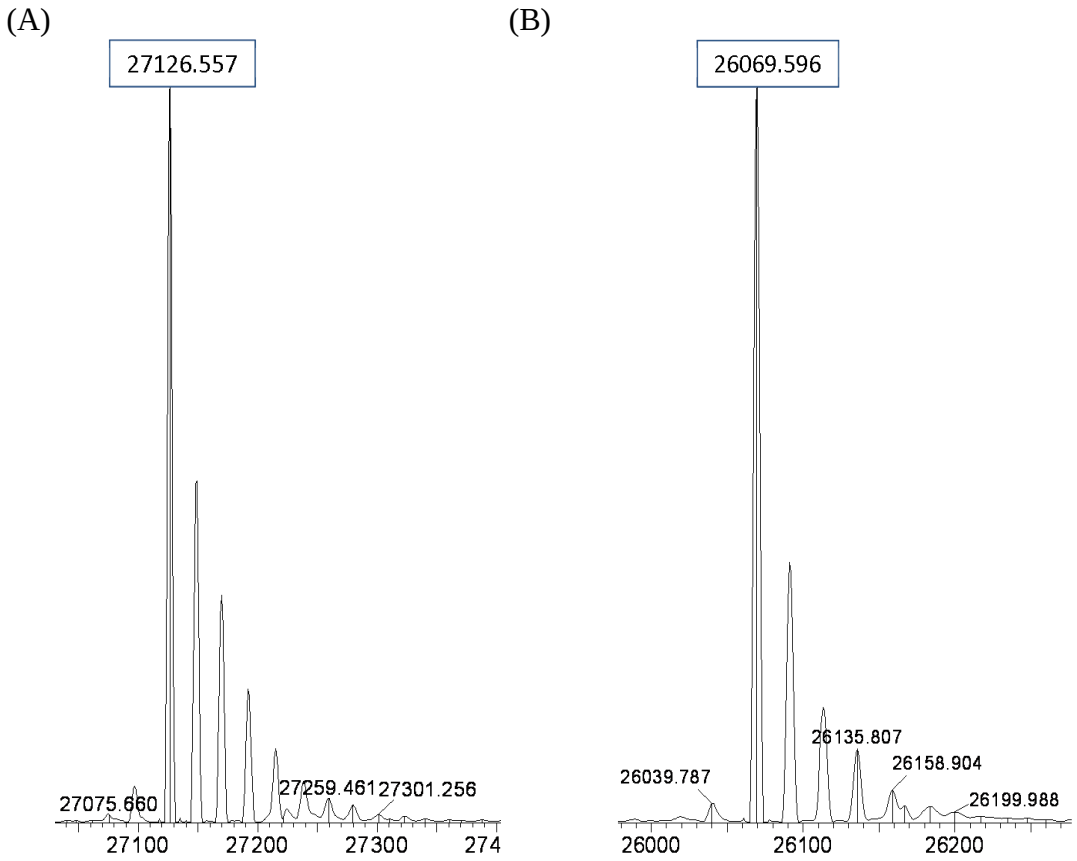
5.2.4 Intact mass analysis of HgpA glycoprotein derived from *H. pullorum*

Proteins were analysed using electrospray ionisation Mass Spectrometry to provide an accurate intact mass to within 1 Da for a 50,000 Da protein. The sample was applied to the Mass Spectrometer as a liquid and the protein signal divided between several peaks. These peaks reflect multiply charged states and the MassLynx computer software, MaxEnt1, was used to deconvolute these peaks back to an uncharged mass.

Data from Figure 5.3 indicates that Hp0114 and HgpA proteins derived from *H. pullorum* wild type strains are modified but Hp0114 and HgpA from the *H. pullorum pgl*B1::Kn mutant are unmodified. *H. pullorum* derived HgpA was readily purified in sufficient quantities to be visible on a coomassie stained SDS-PAGE gel, in contrast to the Hp0114 protein which proved more difficult to purify. For this reason, and also because *H. pullorum* HgpA has only one N-glycosylation sequon so that preparations would be more homogenous, the *H. pullorum* HgpA glycoprotein was chosen for further analysis using Mass Spectrometry. Intact mass analysis was employed to compare the masses of HgpA derived from *H. pullorum* wild type and *pgl*B1::Kn mutant strains (Figure. 5.5) and thus predict the mass of the associated glycan portion of the glycoprotein. Excess salt was removed from the HgpA sample by dialysis (see Methods for details) as any salt present may change the mass of the protein. The analysis was carried out by the Biomolecular analysis core facility within the Faculty of Life Sciences at the University of Manchester. The predicted mass of the unmodified histidine tagged HgpA protein is 26071 Da and this is in agreement with the observed mass of 26070 Da of the HgpA protein derived from the *pgl*B1::Kn mutant (Figure 5.5 B). The HgpA protein derived from wild type *H. pullorum* had an increased mass of 27127 Da (Figure 5.5 A) and this is consistent with the reduced electrophoretic mobility of HgpA derived from this strain (Figure 5.3). The difference in mass between HgpA proteins derived from the wild type and the *pgl*B1::Kn mutant of *H. pullorum* is 1057 Da. This corresponds to mass of the previously determined *H. pullorum* pentasaccharide glycan (Jervis *et al*., 2010) given that the intact mass analysis was only accurate to 1 Da. Intact mass analysis was also performed on HgpA derived from the *H. pullorum pgl*B2::Kn mutant but an

accurate mass value could not be obtained. These data provide strong evidence that the *H. pullorum* HgpA protein is N-glycosylated at N87 in a PglB1 dependent manner with a glycan likely to be the previously described pentasaccharide glycan. In order to confirm this, the N-linked glycan was further characterized by tandem MALDI Mass Spectrometry.

Figure 5.5 Intact mass analysis of HgpA derived from *H. pullorum* wild type (A) and a *pgl*B1::Kn mutant (B). The intact mass for each protein is the boxed value indicated. The difference in mass (to the nearest Da) between HgpA protein from wild type *H. pullorum* and *pgl*B1::Kn mutant strains is 1057 Da.



5.2.5 MALDI Mass Spectrometry of the HgpA glycoprotein derived from wild type *H. pullorum*.

Data from Figure 5.5 suggests PglB1-dependent glycosylation of the *H. pullorum* HgpA protein with a glycan of mass 1057 Da consistent with the mass of the N-linked glycan from an *in vitro* peptide glycosylation assay with *H. pullorum* membrane preparations (Jervis *et al*., 2010). In order to further investigate the nature of the modifications to *H. pullorum* HgpA, this protein was analysed by Tandem MALDI Mass Spectrometry. This involved trypsin digestion of the purified HgpA glycoprotein and Mass Spectrometry analysis of a tryptic glycopeptide. A theoretical trypsin digest of the *H. pullorum* HgpA amino acid sequence is presented (Table 2). The HgpA N-glycosylation sequon (ENNDT) is located on a tryptic peptide of 1371 Da.

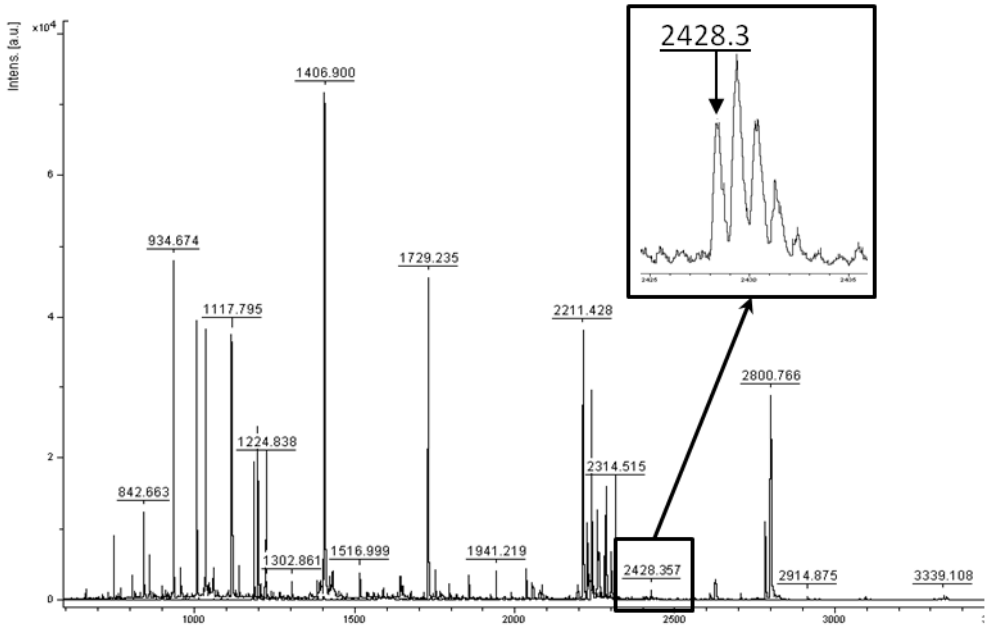
HgpA proteins derived from *H. pullorum* wild type and *pgl*B1::Kn mutant strains were excised from SDS-PAGE gels, digested overnight with trypsin and peptides extracted for analysis by MALDI Mass Spectrometry. Following MALDI Mass Spectrometry on trypsin-digested HgpA from *H. pullorum* wild type or *pgl*B1::Kn mutant strains, many peaks were observed (Figure 5.6). The *m/z* value of many of these peaks corresponded to the masses of peptides following the theoretical trypsin digest of HgpA (Table 2) including peptides of mass 934, 1007, 1117, 1407, 1729 and 2800 Da. A peak at *m/z* 2428.3 (approximately corresponding to the predicted 1371 Da peptide and a 1056 Da glycan) was observed for trypsin digested HgpA protein derived from the wild-type *H. pullorum* strain but not for HgpA derived from the *H. pullorum pgl*B1::Kn mutant. This peak was selected for further fragmentation by tandem MALDI Mass Spectrometry (Figure 5.7).

Table 2 Theoretical trypsin digest of *H. pullorum* HgpA. Bioinformatics software (Peptide cutter) was used to predict the peptide fragments resulting from digestion with trypsin based on the HgpA amino acid sequence. The peptide highlighted in blue has a mass consistent with the mass of the N-glycosylated peptide containing the glycosylation sequon. The peptide cleaved at site 243 is expected to be 840 Da larger than indicated below due to a terminal Da hexa histidine tag.

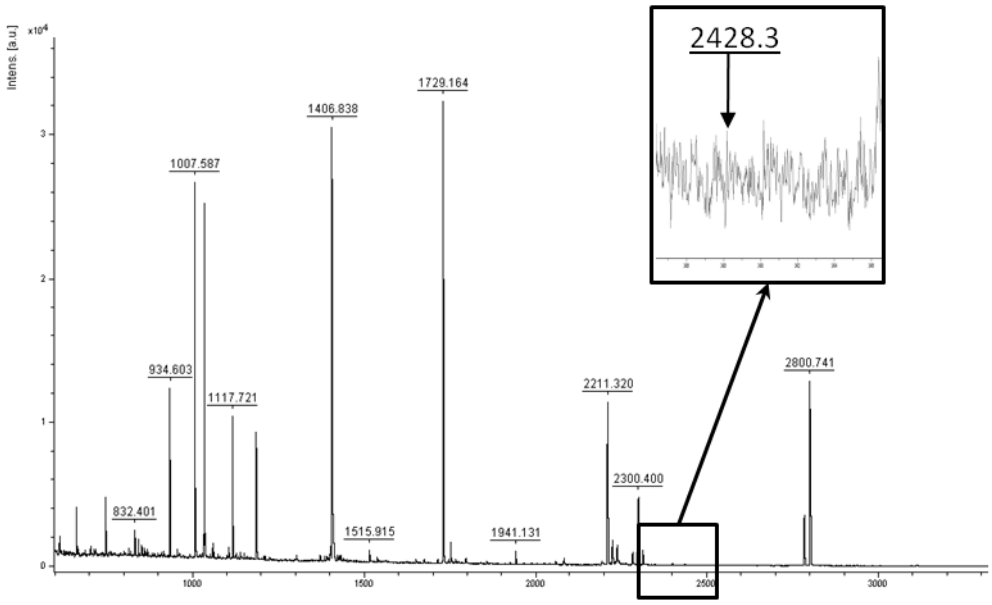
Position of cleavage site	Resulting peptide sequence	Peptide length	Peptide mass
96	K	1	146.189
99	K	1	146.189
191	K	1	146.189
98	IK	2	259.349
160	QK	2	274.320
236	EK	2	275.305
5	IR	2	287.362
234	NLK	3	373.453
3	IER	3	416.478
59	EIR	3	416.478
190	SNSK	4	434.450
164	ITLK	4	473.613
9	TTMK	4	479.592
231	YSIK	4	509.603
186	LILDFK	6	747.933
243	GLEIELF	7	819.953
171	TPLNMIR	7	844.040
150	LQDFNLGK	8	934.060
180	DFTLASPTR	9	1007.111
158	IFDFHIDK	8	1034.180
69	NPFESVVTPK	10	1117.267
227	ITFTLDGQYK	10	1185.342
56	EIPTPNENNDTK	12	1371.423
44	IPDITIDTSTTPK	13	1401.577
217	VDLQTHLDFYR	11	1406.561
206	VLQDFIQTNLPIISK	15	1729.049
31	ILLPLLIPFFILYAQDSQTLPK	22	2544.115
95	QSGQISNPPQLSLFTQTTLNLPSTAR	26	2800.121
142	ITLAYQNLDGSISTIEQELNGDIDWHFPLILSQEIKPEMQSQK	43	4944.544

Figure 5.6 MALDI Mass Spectrometry of trypsin digested HgpA from *H. pullorum*. (A) HgpA peptides derived from a wild type *H. pullorum* strain. (B) HgpA peptides derived from a *H. pullorum pgl*B1::Kn mutant. a.u, arbitary units. The *m/z* value of some peaks corresponds to the predicted masses of peptides following theoretical trypsin digest of HgpA (Table 2) including peptides of mass 934, 1007, 1117, 1407, 1729 and 2800 Da. A peak at 2426.3 *m/z* (boxed in A) was observed with HgpA from wild type *H. pullorum* but was absent in the *pgl*B1::Kn strain.

(A)



(B)

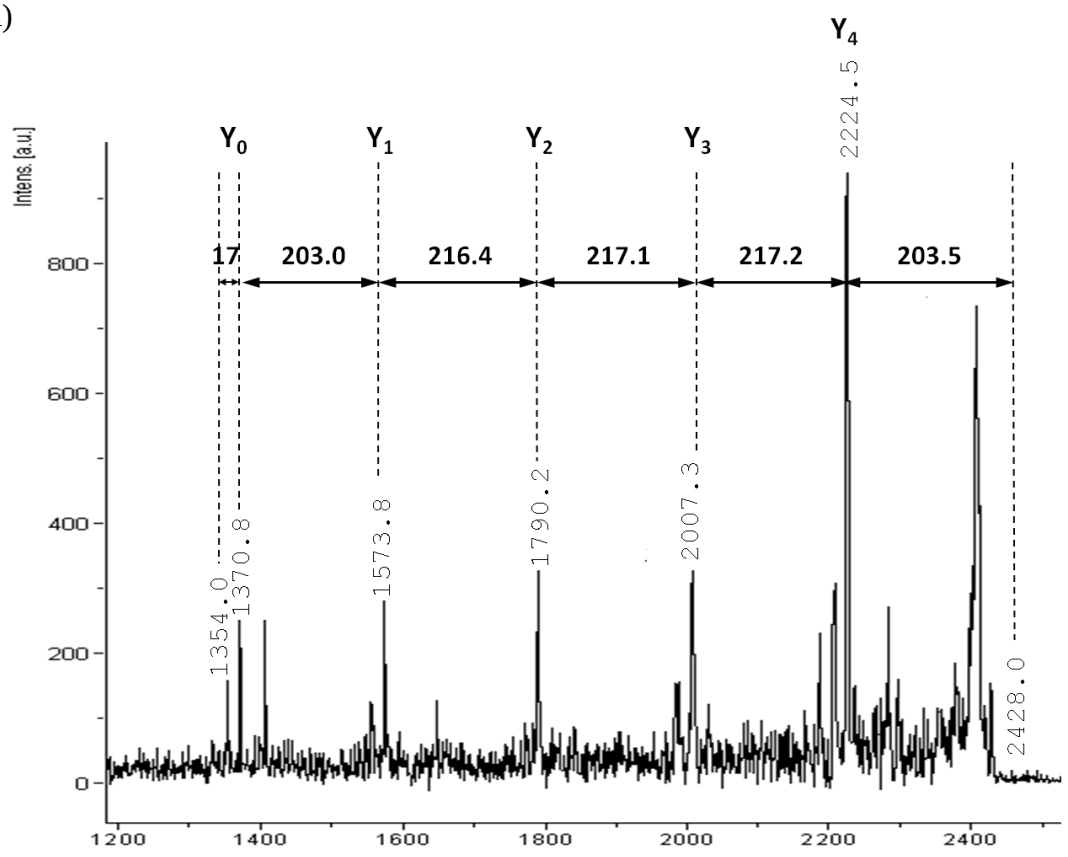


5.2.6 Tandem MALDI Mass Spectrometry of modified HgpA glycoprotein derived from wild type *H. pullorum*.

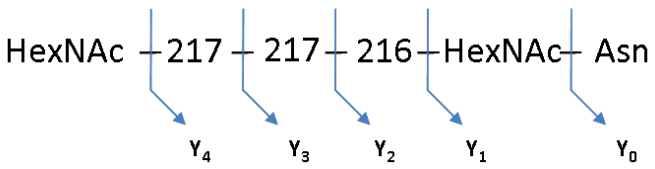
Following the generation of a series of peaks by MALDI Mass Spectrometry of trypsin-digested HgpA protein from wild type *H. pullorum*, a peak of *m/z* 2428 was selected for further fragmentation by tandem MALDI Mass Spectrometry. Fragmentation of the peptide corresponding to *m/z* 2428 resulted in a series of peaks (Figure 5.7) consistent with the previously characterized N-linked pentasaccharide structure of Asn-203 Da-216 Da-217 Da-217 Da-203 Da (Jervis *et al*., 2010). The *m/z* 2224 peak corresponded to the loss of a terminal 203 Da residue, the *m/z* 2007 and 1790 peaks corresponded to the sequential loss of two 217 Da residues, the *m/z* 1574 peak corresponded to the additional loss of a 216 Da residue and the *m/z* 1371 peak corresponded to the mass of the unmodified tryptic peptide EIPTPNENNDTK. An additional peak of *m/z* 1354, 17 Da less than the *m/z* 1371 peak, is characteristic of the cleavage of an amide bond during fragmentation and is further evidence that this glycosylation process is N-linked. These data confirm that HgpA is an N-linked glycoprotein glycosylated at N87 with a pentasaccharide glycan. This is the first Helicobacter N-linked glycoprotein to be identified.

Figure 5.7 Fragmentation of a modified HgpA derived tryptic peptide from wild type *H. pullorum*. (A) Tandem MALDI Mass Spectrum of the *m/z* 2428 precursor ion generated following trypsin digestion of *H. pullorum* HgpA glycoprotein. a.u., arbitary units. Fragment ions resulting from the sequential loss of sugar residues are indicated in the spectrum. A peak corresponding to the Y ion of the HgpA glycopeptide was present at *m/z* 1371. The peak at *m/z* 1354, 17 Da less than the *m/z* 1370.8 peak, is characteristic of the fragmentation of a side-chain amide bond of an N-linked glycan (B) N-linked glycan structure of the *H. pullorum* pentasaccharide consistent with the fragmentation pattern seen by Tandem MALDI Mass Spectrometry.

(A)



(B)



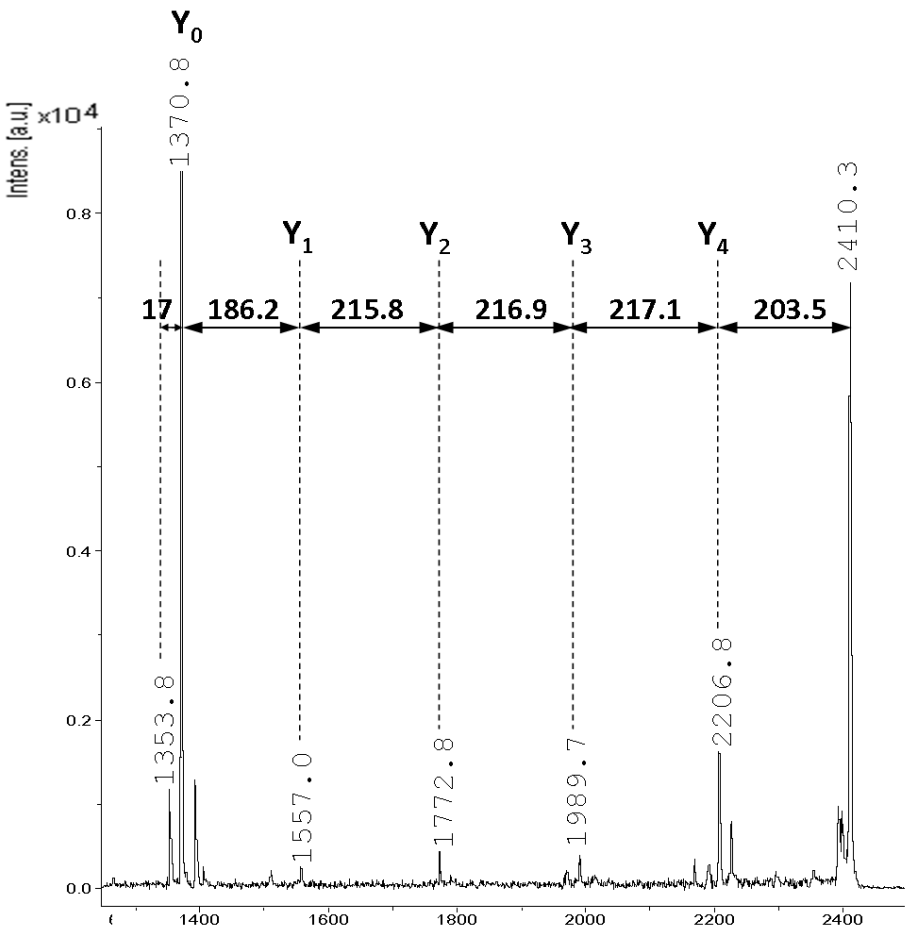
5.2.7 MALDI Mass Spectrometry of modified HgpA glycoprotein derived from the *H. pullorum pgl*B2::Kn mutant.

Data from Figure 5.3 suggests that glycosylation of the *H. pullorum* HgpA protein is PglB1, but not PglB2, dependent as HgpA protein from the *H. pullorum pgl*B2::Kn mutant has comparable electrophoretic mobility to HgpA from the wild type *H. pullorum* strain but the electrophoretic mobility of HgpA protein from the *H. pullorum pgl*B1::Kn strain is consistent with unglycosylated HgpA. In order to further investigate the nature of the modifications to HgpA derived from the *H. pullorum pgl*B2::Kn mutant, this protein was also analysed by MALDI Mass Spectrometry. A series of peaks corresponding to HgpA tryptic peptides were obtained (data not shown). A peak of *m/z* 2428 would indicate the well characterized *H. pullorum* pentasaccharide glycan attached to the 1371 Da peptide but was not observed in this sample. Instead, a peak of *m/z* 2410 was selected and produced a series of fragmented peaks following tandem MALDI Mass Spectrometry (Figure 5.8).

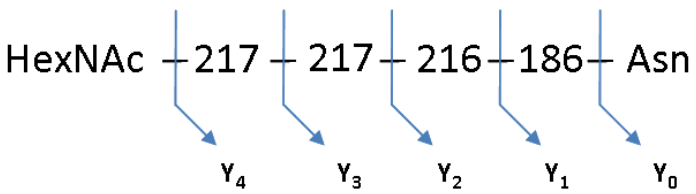
A *m/z* 2207 peak corresponded to the loss of a terminal 203 Da residue, the *m/z* 1990 and 1773 peaks corresponded to the sequential loss of two 217 Da residues, a *m/z* 1574 peak corresponded to the additional loss of a 216 Da residue and the *m/z* 1371 peak corresponded to the unmodified peptide. In contrast to the fragmentation series of HgpA peptide derived from wild type *H. pullorum* which contains a 203 Da reducing end HexNAc residue, the fragmentation of a HgpA peptide derived from the *pgl*B2::Kn mutant revealed a pentasaccharide structure with a 186 Da reducing sugar. An additional peak of *m/z* 1354, 17 Da less than the *m/z* 1371 peak, indicates an N-linked glycosylation process. These data indicate that HgpA glycosylation in *H. pullorum* is not dependent on the *pgl*B2 gene product but there are structural differences in the reducing end sugar of the pentasaccharide in this genetic background.

Figure 5.8 Glycan structure of HgpA derived from the *H. pullorum pgl*B2::Kn mutant. (A) Tandem MALDI Mass Spectrometry of the *m/z* 2412 precursor ion generated following purification of *H. pullorum* HgpA. a.u., arbitary units. Fragment ions resulting from the sequential loss of sugar residues are indicated in the spectrum. A peak corresponding to the Y ion of the HgpA glycopeptide was present at *m/z* 1372. The peak at *m/z* 1355, 17 Da less than the *m/z* 1372 peak, is characteristic of the fragmentation of a side-chain amide bond of an N-linked glycan. (B) N-linked glycan structure of the *H. pullorum* pentasaccharide on HgpA from the *H. pullorum pgl*B2::Kn mutant consistent with the fragmentation pattern seen by Tandem MALDI Mass Spectrometry.

(A)



(B)



5.3 Discussion

Chapter 3 described the *in vitro* glycosylation of a short peptide containing an optimal *C. jejuni* glycosylation sequon following incubation with *H. pullorum* membrane preparations. Mass spectrometry of the glycopeptide following incubation with membranes from *H. pullorum* wild type, *pgl*B1::Kn and *pgl*B2::Kn mutants confirmed that PglB1 was essential for glycosylation with a pentasaccharide glycan but PglB2 was not. Chapter 4 described the identification of two glycoproteins from *H. pullorum* and demonstrates that they can be glycosylated in *E. coli* with the *C. jejuni* heptasaccharide glycan. Although these approaches are informative the most direct method to study *H. pullorum* protein glycosylation is in the organism itself. The principle aim of Chapter 5 therefore was to develop a practical system to investigate *H. pullorum* N-linked protein glycosylation that would be used to investigate of the role of *H. pullorum* genes in glycan assembly and transfer.

In this chapter, *H. pullorum* strains were created that constitutively produce a hexa histidine-tagged copy of the Hp0114 and HgpA proteins. Both proteins had reduced electrophoretic mobility when produced in a wild type or *pgl*B2::Kn genetic background compared to proteins produced in a *pgl*B1::Kn genetic background. These data are in agreement with *in vitro* OST assay results from Chapter 4 and suggest that N-glycosylation in *H. pullorum* is a PglB1 dependent, PglB2 independent process.

The Hp0114 protein has two D/E-X-N-X-S/T glycosylation sequons, $_{85}$ENNAT$_{89}$ and $_{169}$ENNGS$_{173,}$ and one N-X-S/T sequon, $_{46}$NLS$_{48}$. Site directed mutagenesis experiments demonstrated that in *H. pullorum*, Hp0114 protein is glycosylated only at N87 (Figure 5.4). This is in contrast to results from Chapter 4 where Hp0114 produced in *E. coli* expressing *H. pullorum pgl*B1 was glycosylated with the *C. jejuni* heptasaccharide at N87 and N171 sites. The explanation for this difference in site specificity between HpPglB1 produced in *E. coli* and *H. pullorum* is unclear but may be related to differences in glycan structures produced in these different species or differing levels of lipid-linked oligosaccharide. Differences between these *H.*

pullorum and *E. coli* N-glycosylation systems emphasise the importance of, as far as possible, investigating the native systems.

In order to investigate the *H. pullorum* N-linked protein glycosylation system, a reasonable amount of purified glycoprotein was required, approximately the quantity that could be detectable by coomassie staining of an SDS-PAGE gel. Purification of Hp0114 protein via its hexa histidine tag was inefficient with very little glycoprotein recovered. However, purification of HgpA was much more efficient and this protein was used as a tool to further investigate N-glycosylation.

The intact masses of HgpA proteins from the *H. pullorum* wild type and *pgl*B1::Kn mutant strains were consistent with glycosylated and unglycosylated forms of HgpA respectively. Intact mass data from HgpA could not be obtained for HgpA derived from the *H. pullorum pgl*B2::Kn mutant. It is possible that this sample contained a mixture of HgpA glycoproteins with different glycan structures and intact mass is only accurate for determining the mass of a single protein. MALDI Mass Spectrometry of HgpA protein produced in the *H. pullorum* wild type strain was consistent with the previously predicted pentasaccharide structure of Asn-203 Da-216 Da-217 Da-217 Da-203 Da (Jervis *et al*., 2010).

Glycosylation of HgpA was clearly PglB1 but not PglB2 dependent. Tandem MALDI Mass Spectrometry of HgpA derived from the *H. pullorum pgl*B2::Kn mutant indicated a reducing sugar of 186 Da rather than 203 Da (Figure 5.8). The *C. jejuni pgl* locus contains genes encoding a dehydratase (PglF), an aminotransferase (PglE) and an acetyltransferase (PglD) that are involved in the conversion of a GlcNAc (203 Da) residue to form bac (228 Da). In this pathway, PglF converts GlcNAc to a 185 Da keto sugar which is converted to a 186 Da amino sugar by PglE and then to the 228 Da bac residue by PglD (Olivier *et al*., 2006). It is possible that in the *pgl*B2::Kn mutant the 186 Da reducing end sugar is derived from this pathway but this requires further investigation. The structural data obtained from the HgpA peptide derived from a *pgl*B2::Kn mutant was slightly different to data based on peptide *in vitro* glycosylation (Chapter 3). The glycopeptide derived from the OST assay following incubation with *H. pullorum pgl*B2::Kn membrane preparations

contains a 203 Da HexNAc reducing end sugar. It is possible that multiple N-linked glycan structures are present in this *pgl*B2::Kn mutant. Intact mass analysis is only useful for identifying the mass of a single protein or glycoprotein and the intact mass results of HgpA derived from the PglB2::Kn mutant were unreliable (data not shown), suggesting that glycoproteins of different masses were present in this sample. These results are interesting and suggest a link between PglB1 and a putative PglB2 N-glycosylation system that may have a bac type sugar at the reducing end but require more investigation beyond the scope of this thesis.

In summary, the HgpA glycoprotein identified in Chapter 4 is glycosylated *in H. pullorum* at a single site with the previously characterized pentasaccharide glycan but HgpA derived from *a H. pullorum pgl*B1::Kn strain is unmodified. This hexa histidine tagged HgpA protein was used to identify the N-glycan structure in a *H. pullorum pgl*B2::Kn mutant and can now be used to determine the role of other putative *H. pullorum pgl* genes in the assembly of the N-linked pentasaccharide.

Chapter 6

Characterization of the *H. pullorum* PglB1-Dependent N-Linked Protein Glycosylation System

6.1 Background

Chapter 5 describes the purification and structural characterisation of the N-linked glycan from the *H. pullorum* HgpA protein. There are currently no published data on the role of *H. pullorum* genes in the biosynthesis of Helicobacter N-linked glycans. The aim of Chapter 6 was therefore to use techniques described in Chapter 5 to identify genes involved in *H. pullorum* N-glycosylation and characterize the role of their gene products in the biosynthesis of a pentasaccharide glycan.

The *H. pullorum, H. canadensis, H. winghamensis* and *W. succinogenes* genomes contain genes encoding proteins with significant levels of amino acid sequence identity to the *C. jejuni* PglA, PglC, PglH, PglJ and PglI glycosyltransferases involved in biosynthesis of the *C. jejuni* N-linked heptasaccharide (Figure 6.1; Jervis *et al*., 2010). In *C. jejuni, W. succinogenes*, and other Campylobacter species, the genes encoding glycosyltransferases involved in N-linked protein glycosylation are located within a single genetic locus (Szymanski et al., 1999; Jervis *et al*., 2012). In contrast, in the *H. pullorum, H. canadensis and H. winghamensis* genomes, these genes are found at three distinct locations (Jervis *et al*., 2010).

An N-linked glycan was structurally characterised in *H. pullorum* (Jervis *et al*., 2010), *C. jejuni* (Linton *et al*., 2002) and *W. succinogenes* (Dell *et al*., 2010) but there are no published data on the role of glycosyltransferases in the assembly of the N-linked glycan in *W. succinogenes* or *H. pullorum*. Figure 6.1 outlines the arrangement of the glycosyltransferase-encoding genes from these three species. All three contain tandemly arranged genes encoding PglA and PglC enzymes and in *C. jejuni* and *W. succinogenes* these are located with other *pgl* genes. The PglA and PglC glycosyltransferases from *C. jejuni, W. succinogenes* and *H. pullorum* have a high level of identity and are therefore likely to have a similar role in N-glycosylation. In *C. jejuni*, PglC transfers the first sugar residue, bac, to the membrane-bound lipid carrier, Und-PP (Linton *et al*., 2005; Glover *et al*., 2006). The *C. jejuni* PglA glycosyltransferase then transfers a GlcNAc residue to generate a lipid-linked disaccharide (Glover *et al*., 2005; Linton *et al*., 2005).

Genes encoding PglH and PglJ glycosyltransferases are present on the *H. pullorum*, *C. jejuni* and *W. succinogenes* genomes. In *C. jejuni*, PglJ is involved in the transfer of a second GlcNAc sugar to the lipid linked disaccharide to produce a trisaccharide glycan (Glover *et al*., 2005; Linton *et al*., 2005). The *H. pullorum* gene encoding PglJ is located directly downstream of genes encoding PglK (a predicted ABC transporter) and PglI and PglL (predicted glycosyltransferases). In *C. jejuni*, the *pgl*J gene is located directly downstream of *pgl*K, *pgl*H and *pgl*I genes.

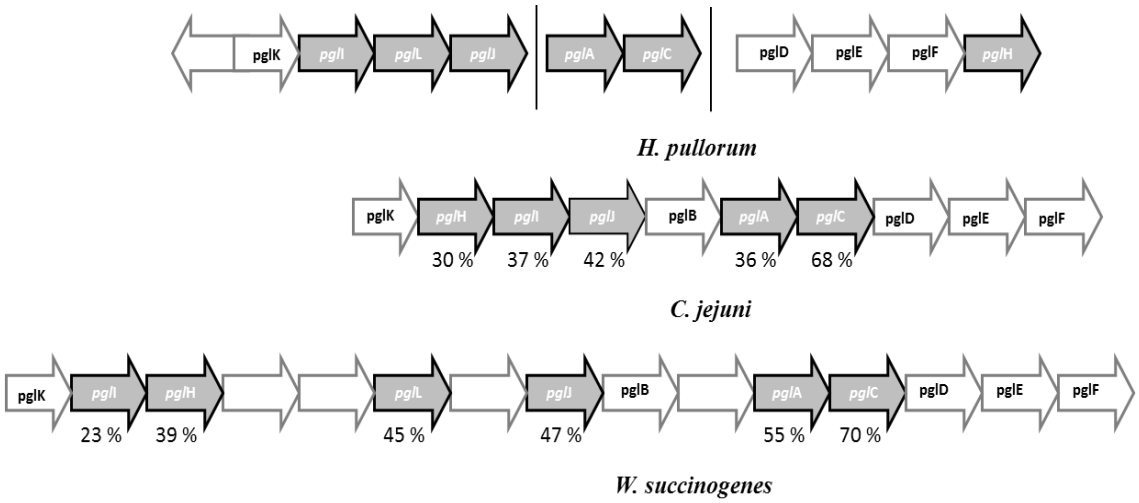
In *C. jejuni*, the PglH glycosyltransferase transfers three 203 Da GlcNAc residues to the lipid linked trisaccharide to produce a lipid-linked hexasaccharide (Glover *et al*., 2005; Linton *et al*., 2005; Troutman and Imperiali, 2009). On the *H. pullorum* genome, the *pgl*H gene is adjacent to *pgl*D, *pgl*E and *pgl*F genes. These genes in *C. jejuni* encode proteins that are involved in the biosynthesis of the N-linked glycan reducing end sugar, bac (Olivier *et al*., 2006). In *H. pullorum* however, the PglB1-dependent N-linked pentasaccharide does not contain a bac residue and mutation of the *pgl*DEF genes does not appear to affect the structure of this N-linked glycan (Jervis *et al*., 2010).

The genomes of *H. pullorum*, *C. jejuni* and *W. succinogenes* all contain a gene encoding PglI but only *H. pullorum* and *W. succinogenes* contain a gene encoding PglL. The *C. jejuni* PglI is involved in the transfer of a branching Glc residue to the lipid-linked hexasaccharide to produce the final lipid-linked heptasaccharide which is transferred to protein by the action of PglB (Glover *et al*., 2005; Linton *et al*., 2005). Interestingly, a Hex residue is not present in the *H. pullorum* PglB1-dependent N-linked glycan (Jervis *et al*., 2010) but as *H. pullorum* encodes two PglB enzymes the *pgl*I gene may be involved in a separate N-glycosylation system. The *H. pullorum* PglL has a relatively high sequence identity to the *W. succinogenes* PglL (42%) but the role of this glycosyltransferase in N-linked protein glycosylation systems has not yet been reported.

Strains of *H. pullorum* with insertionally inactivated *pgl*A, *pgl*C, *pgl*H or *pgl*I genes were recently constructed (Jervis and Linton, unpublished data) but the effect of these mutations on glycosylation has not been investigated. Attempts to ‘knock-out’

*pgl*I and *pgl*L by insertional mutagenesis were unsuccessful, indicating that they may be essential. In this chapter I describe the use of a histidine-tagged *H. pullorum* glycoprotein, HgpA, as a tool to investigate the effect of individual *pgl* gene mutations on the *H. pullorum* PglB1-dependent N-glycosylation system.

Figure 6.1 Schematic representations of the *C. jejuni* glycosyltransferase-encoding gene locus and arrangement of orthologues present in *H. pullorum* and *W. succinogenes*. Genes that encode glycosyltransferases and proteins involved in sugar biosynthesis are indicated by arrows and percentage values below the genes are the levels of identity between amino acid sequences encoded by *H. pullorum* and *C. jejuni* or *W. succinogenes*. Grey arrows indicate glycosyltransferase encoding genes.



6.2 Results

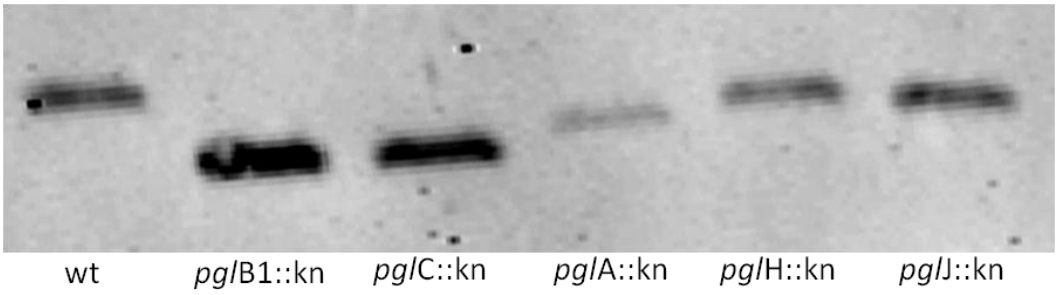
6.2.1 The effect of insertional mutagenesis of putative glycosyltransferase encoding genes on the electrophoretic mobility of *H. pullorum* HgpA protein

Chapter 5 describes how the hexa histidine tagged HgpA protein can be used to investigate the effect of gene knock-outs involved in protein glycosylation on protein glycosylation. A similar approach was used to investigate the role of putative *H. pullorum* glycosyltransferases in assembly of the *H. pullorum* N-linked pentasaccharide.

The *H. pullorum pgl*A::Kn, *pgl*C::Kn, *pgl*H::Kn and *pgl*J::Kn insertional knock-out mutants were available in the laboratory (Jervis and Linton, unpublished data) but attempts to similarly knock-out *H. pullorum pgl*I and *pgl*L genes were unsuccessful. *H. pullorum pgl*A::Kn, *pgl*C::Kn, *pgl*H::Kn and *pgl*J::Kn knock-out mutants were created in *H. pullorum* producing hexa histidine tagged HgpA glycoprotein by electroporation with plasmids pHp*pgl*C::Kn, pHp*pgl*A::Kn, pHp*pgl*H::Kn and pHp*pgl*J::Kn. HgpA glycoprotein was purified from these strains and detected in SDS-PAGE and western blots with anti histidine antiserum (Figure 6.9). Glycosylated HgpA derived from wild type *H. pullorum* and unglycosylated HgpA derived from *H. pullorum pgl*B1::Kn were similarly analysed for comparison.

The HgpA protein derived from the *H. pullorum pgl*C::Kn mutant had comparable electrophoretic mobility to the unmodified HgpA derived from the *H. pullorum pgl*B1::Kn mutant. HgpA derived from the *pgl*A::Kn mutant had reduced electrophoretic mobility compared to the HgpA from the *pgl*B1::Kn mutant but was more mobile than HgpA from the wild type *H. pullorum* strain indicating that HgpA from the *pgl*A::Kn mutant is modified but with a glycan of reduced size. HgpA derived from the *pgl*J::Kn and *pgl*H::Kn mutants had reduced electrophoretic mobility compared to HgpA from the *pgl*A::Kn mutant, indicating that the HgpA from these strains was modified with a larger glycan. These data are useful for identifying the involvement of *H. pullorum pgl* genes in the assembly of the N-linked glycan but in order to characterize the structure of the HgpA glycan from these strains, more complex Mass Spectrometry-based techniques were required.

Figure 6.2 Glycosylation of *H. pullorum* HgpA derived from *H. pullorum* glycosyltransferase-encoding gene mutants. Western blot of partially purified hexa histidine tagged *H. pullorum* HgpA from *H. pullorum* wild-type, *pgl*B1::Kn, *pgl*C::Kn, *pgl*A::Kn, *pgl*H::Kn and *pgl*J::Kn mutants. Compared to HgpA from wild type *H. pullorum*, the mobility of HgpA protein is increased in *H. pullorum* strains with *pgl*C and *pgl*A glycosyltransferase-encoding gene knock-out mutations and possibly also with *pgl*H and *pgl*J mutation.



6.2.2 Intact mass analysis of HgpA derived from *H. pullorum* glycosyltransferase-encoding gene mutants.

Western blotting indicated that *H. pullorum* PglC, PglA, PglH and PglJ glycosyltransferases are involved in the assembly of the PglB1-dependent *H. pullorum* N-linked pentasaccharide glycan present on HgpA (Figure 5.9). In order to further characterize the HgpA glycan structures produced by *H. pullorum pgl*C::Kn, *pgl*A::Kn, *pgl*H::Kn and *pgl*J::Kn mutants, HgpA protein was purified and the intact mass of the glycoprotein determined.

The intact mass of HgpA from the *H. pullorum pgl*C::Kn mutant was 26069 Da to the nearest Da (Figure. 6.3), which is almost identical to the predicted mass of 26071 Da for unglycosylated HgpA and the observed mass of HgpA derived from the *pgl*B1::Kn mutant (Chapter 5; Figure 5.5 B). These data indicate that HgpA derived from the *pgl*C::Kn strain lacks glycan.

The difference between the intact mass of HgpA from the *H. pullorum pgl*A::Kn mutant (26297 Da) and HgpA from the *pgl*B1::Kn mutant (26069 Da) was 228 Da and most likely corresponds to a protein with a glycan mass of 228 Da, presumably a bac residue. Interestingly, the HgpA reducing end sugar in this strain differs from the reducing end sugar of the HgpA pentasaccharide glycan from wild type *H. pullorum*, which is a 203 Da HexNAc residue.

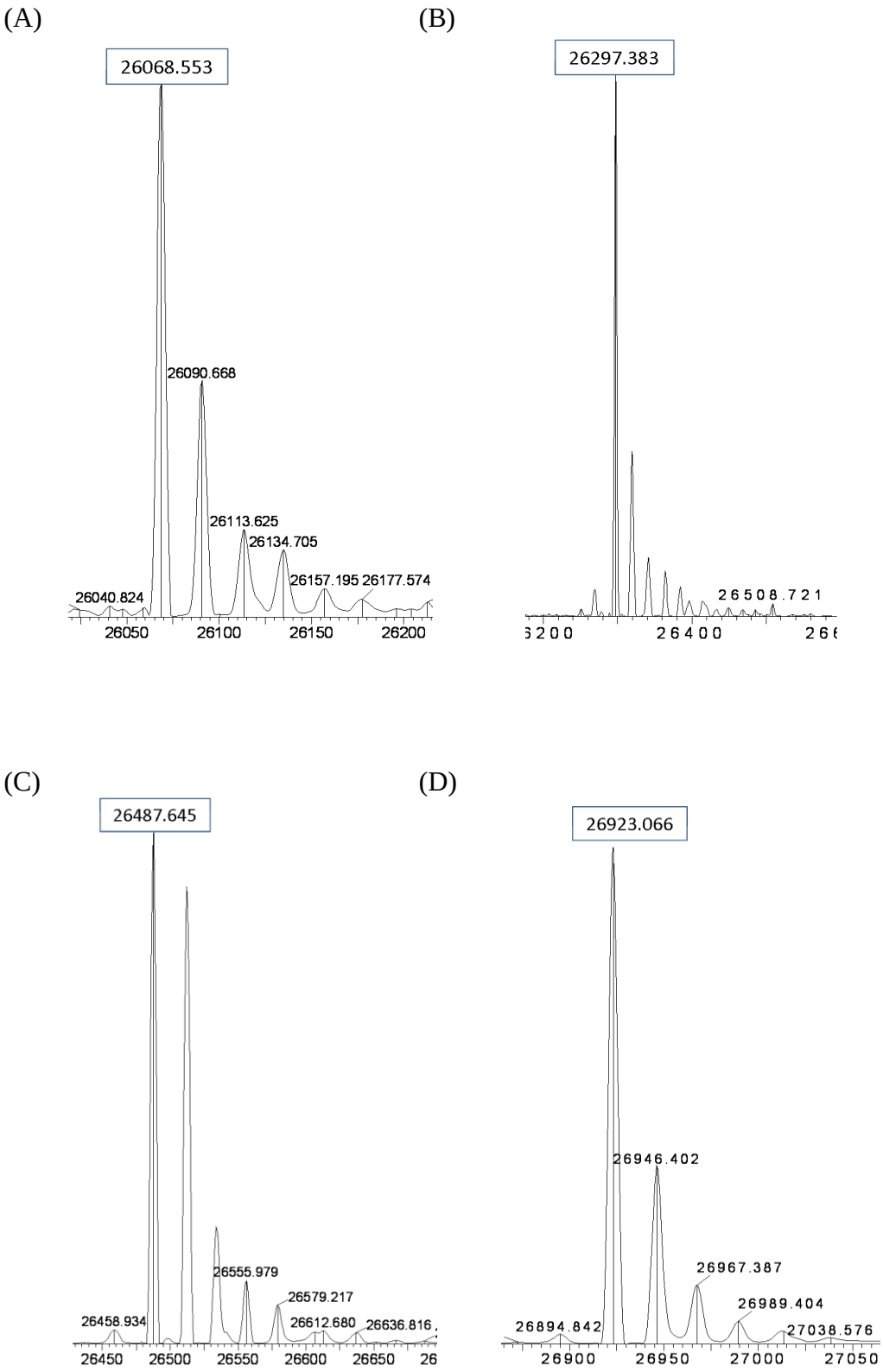
The *H. pullorum pgl*H gene is located immediately downstream of the *pgl*DEF genes which, in *C. jejuni*, are involved in the biosynthesis of the 228 Da reducing end sugar of the N-linked glycan, bac. A comparison of the intact mass of HgpA derived from the *H. pullorum pgl*H::Kn mutant (26923 Da) with the intact mass of HgpA derived from the *pgl*B1::Kn mutant (26069 Da) indicated a glycan of 854 Da consistent with a tetrasaccharide glycan structure containing a reducing end 203 Da HexNAc sugar, with a 216 Da and two 217 Da sugar residues.

The *H. pullorum pgl*J gene is located immediately downstream of *pgl*I (which transfers a branching Glc in *C. jejuni*), *pgl*L (which encodes a predicted

uncharacterized glycosyltransferase) and *pgl*K (which encodes a predicted ABC transporter). A comparison of the intact mass of HgpA derived from a *H. pullorum pgl*J::Kn mutant (26488 Da) with the intact mass of HgpA derived from the *H. pullorum pgl*B1::Kn mutant (26069 Da) indicated a glycan of 419 Da consistent with a disaccharide structure of a 203 Da reducing end sugar and a 216 Da terminal sugar.

Combined these data indicate specific roles for the glycosyltransferases encoded by the *pgl*A, *pgl*C, *pgl*H and *pgl*J genes in the biosynthesis of the *H. pullorum* N-linked pentasaccharide. More specifically, the *pgl*C gene product may be required for transfer of the reducing end sugar, *pgl*A the next sugar, *pgl*J the third sugar and *pgl*H the fifth non reducing end sugar residue of the pentasaccharide. These inferences were further tested by characterising the HgpA N-linked glycans from these various glycosyltransferase mutants in more detail.

Figure 6.3 Intact mass analysis of HgpA derived from *H. pullorum pgl*C::Kn (A), *pgl*A::Kn (B), *pgl*J::Kn (C) and *pgl*H::Kn (D) knock-out mutants. The intact mass for each protein is the boxed value indicated.



6.2.3 MALDI Mass Spectrometry of trypsin digested HgpA peptides from *H. pullorum* glycosyltransferase-encoding gene mutants.

Data from Figure 6.3 suggests glycosylation of the *H. pullorum* HgpA protein from *pgl*A::Kn, *pgl*H::Kn and *pgl*J::Kn strains with a glycan structure of mass 228, 854 and 419 Da respectively whilst HgpA from the *H. pullorum pgl*C::Kn strain was unmodified. The glycan structure of HgpA derived from *pgl*A::Kn, *pgl*H::Kn and *pgl*J::Kn strains was therefore further characterized using MALDI Mass Spectrometry techniques described in Chapter 5.

Following the generation of a series of peaks by MALDI Mass Spectrometry of trypsin-digested HgpA protein from *H. pullorum pgl*A::Kn, *pgl*H::Kn and *pgl*J::Kn, peaks of *m/z* 1599.3, 2224.0 and 1790.5 respectively were identified as potential glycopeptides. These were selected for further fragmentation by tandem MALDI Mass Spectrometry. Fragmentation of the peak corresponding to a 1599 Da glycopeptide of HgpA derived from the *pgl*A::Kn mutant resulted in one major peak corresponding to 1371 Da, the mass of the unmodified peptide (Figure 6.4 A). This is in agreement with data from intact mass analysis (Figure 6.3) and indicates that HgpA from a *pgl*A::Kn mutant is glycosylated with a single sugar residue of mass 228 Da. This was unexpected as the reducing end sugar of the HgpA N-linked glycan derived from wild type *H. pullorum* is a 203 Da HexNAc (Chapter 5).

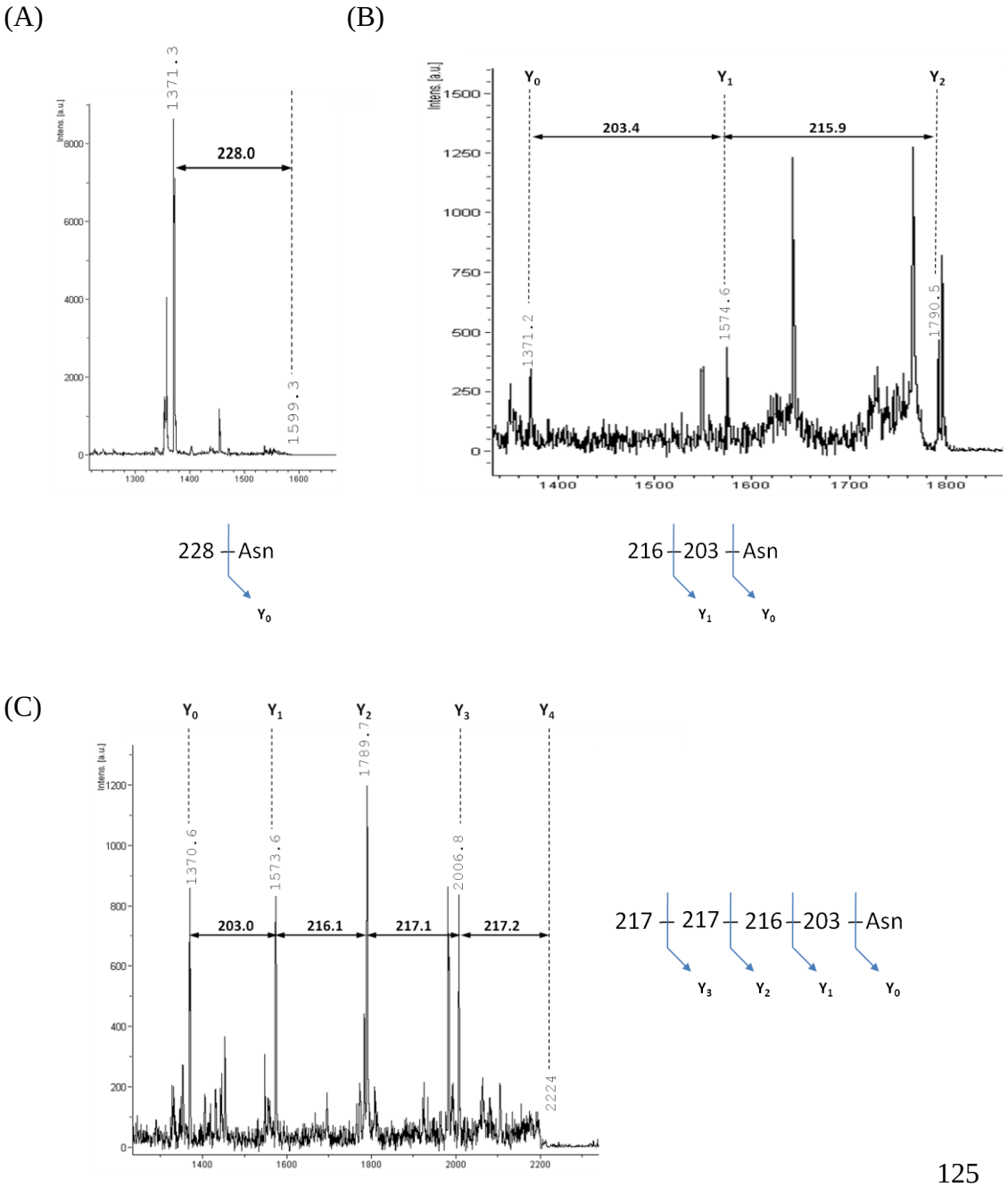
Fragmentation of the peak corresponding to a 1791 Da glycopeptide of HgpA derived from the *pgl*J::Kn mutant resulted in a series of peaks (Figure 6.4 B). The *m/z* peak corresponding to 1575 Da indicated loss of a 216 Da residue, and the *m/z* peak corresponding to 1372 Da indicated further loss of a 203 Da residue. These data are consistent with previous intact mass data (6.2.2).

Fragmentation of the peak corresponding to a 2224 Da glycopeptide of HgpA derived from the *pgl*H::Kn mutant resulted in a series of peaks that were consistent with a tetrasaccharide structure of Asn - 203 Da – 216 Da – 217 Da – 217 Da (Figure 6.4 C). This structure is in agreement with data from intact mass analysis of HgpA (6.2.2) and indicates a lack of the non reducing end 203 Da residue from the

previously described *H. pullorum* N-linked pentasaccharide glycan (Jervis *et al*., 2010).

Combining the data on N-glycan structures of HgpA derived from *H. pullorum* *pgl*C::Kn, *pgl*A::Kn, *pgl*J::Kn and *pgl*H::Kn insertional knock-out mutants enabled a model for the role of these glycosyltransferases in the assembly of the lipid linked pentasaccharide glycan to be proposed. Interestingly, it is clear from this data, that *H. pullorum* PglB1 will transfer mono, di, tetra and pentasaccharides to the asparagine residue of glycoproteins in the absence of glycosyltransferases required for the complete assembly of the *H. pullorum* lipid linked pentasaccharide glycan.

Figure 6.4 Fragmentation of a modified HgpA derived tryptic peptide from *H. pullorum* glycosyltransferase-encoding gene mutants. Tandem MALDI Mass Spectrometry spectrum of the *m/z* 1599.3, 1790.5 and 2224.0 precursor ions generated following purification of HgpA derived from *H. pullorum pgl*A::Kn (A), *pgl*J::Kn (B) and *pgl*H::Kn (C) mutants respectively. a.u., arbitary units. Fragment ions resulting from the sequential loss of sugar residues are indicated in the spectrum. A peak corresponding to the Y ion of the HgpA glycopeptide was present at *m/z* 1371. The N-linked glycan structure consistent with the fragmentation pattern seen by Tandem MALDI Mass Spectrometry is indicated for each spectrum.



6.3 Discussion

Methods for the purification of the *H. pullorum* N-linked glycoprotein, HgpA and the analysis of glycan structure were developed in Chapter 5. These methods were used in Chapter 6 to investigate the role of genes encoding predicted glycosyltransferases in assembly of the *H. pullorum* lipid-linked pentasaccharide glycan that is transferred to asparagine residues of *H. pullorum* proteins by the action of PglB1.

The *H. pullorum* genome contains genes encoding orthologues of the *C. jejuni* PglA, PglC, PglH and PglJ glycosyltransferases, but unlike the *C. jejuni pgl* genes, the genes encoding these *H. pullorum* orthologues are present at three distinct locations on the chromosome. Four *H. pullorum* strains were previously constructed with insertional knock out mutation strains in *pgl*A, *pgl*C, *pgl*H *and pgl*J genes. In this chapter, the plasmids used to create these glycosyltransferase mutants were electroporated into *H. pullorum* cells expressing an additional hexa histidine labelled copy of the HgpA protein as described in Chapter 5. HgpA from these strains was purified, analysed by SDS-PAGE, western blotting, intact mass determination and tandem MALDI Mass Spectrometry.

*H. pullorum pgl*A and *pgl*C are co-located on the *H. pullorum* genome. In *C. jejuni,* PglC catalyzes the transfer of the first sugar, bac, to a lipid carrier and *pgl*A catalyzes the transfer of the first GalNAc sugar to this lipid linked monosaccharide (Glover *et al*., 2005; Linton *et al*., 2005; Glover *et al*., 2006). The intact mass of HgpA protein derived from the *H. pullorum pgl*C::Kn mutant was almost identical to the intact mass of unmodified HgpA from the *pgl*B1::Kn mutant indicating that in *H. pullorum* PglC functions to transfer the 203 Da reducing terminal HexNAc sugar to the lipid carrier. The intact mass of HgpA from the *H. pullorum pgl*A::Kn strain is consistent with a single 228 Da sugar modification suggesting that *H. pullorum* PglA transfers the second sugar of the *H. pullorum* N-linked pentasaccharide, a 216 Da residue. Interestingly, the reducing end sugar attached to the asparagine of HgpA from the wild type strain is not 228 Da but rather 203 Da as previously described. It is possible that when a 216 Da residue fails to be transferred onto the 203 Da

residue, the reducing end 203 Da HexNAc residue attached on the lipid linked oligosaccharide is converted to a 228 Da trideoxydiacetamidohexose residue such as bac, possibly by the action of the PglDEF enzymes. This may be the point where biosynthesis of a PglB1 dependent N-linked glycan diverges from a putative 228 Da containing PglB2 dependent pathway although this will require further investigation. In agreement with the intact mass data, fragmentation of the HgpA glycopeptide derived from the *H. pullorum pgl*A::Kn mutant produced a spectrum consistent with the loss of a 228 Da residue (Figure 6.4 A).

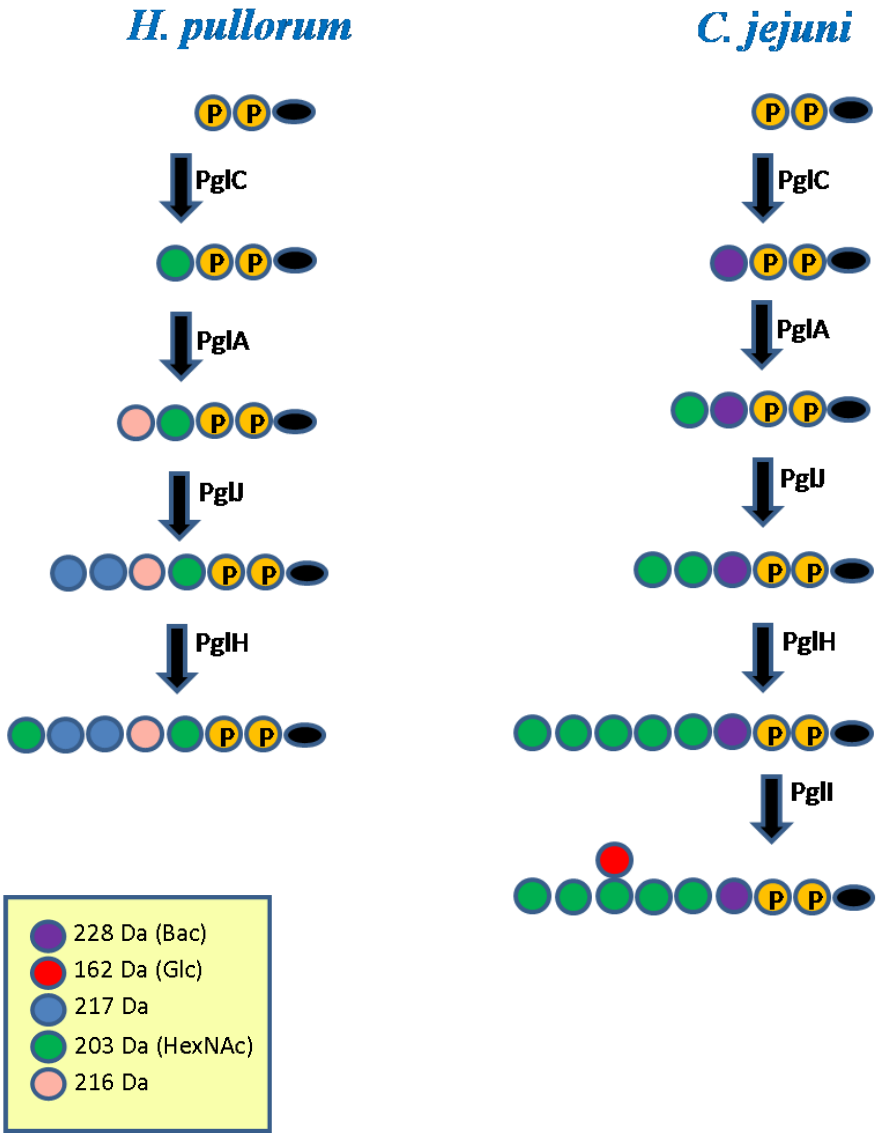
In *C. jejuni*, PglJ is required for the transfer of the third sugar, GalNAc, to the lipid linked disaccharide structure (Glover *et al*., 2005; Linton *et al*., 2005). Intact mass analysis of HgpA and tandem MALDI Mass Spectrometry of the HgpA peptide from a *H. pullorum pgl*J::Kn strain indicated a structure consistent with an N-linked disaccharide glycan containing a 203 Da reducing end HexNAc sugar and a 216 Da terminal sugar. These results indicate that *H. pullorum* PglJ is also involved in transferring a third sugar to the lipid linked disaccharide which, in the *H. pullorum* PglB1 dependent pentasaccharide glycan, is a 217 Da sugar. It is possible that *H. pullorum* PglJ also transfers the second 217 Da sugar residue.

The *C. jejuni* PglH glycosyltransferase is responsible for the addition of the three terminal 203 Da GalNAc residues of the *C. jejuni* lipid linked heptasaccharide (Glover *et al*., 2005; Linton *et al*., 2005; Troutman *et al*., 2009). Intact mass analysis of HgpA and analysis of tandem MALDI Mass Spectrometry of the HgpA peptide from a *H. pullorum pgl*H::Kn strain identified a tetrasaccharide N-linked glycan structure consisting of a 203 Da reducing end HexNAc sugar, a 216 Da sugar and two terminal 217 Da sugar residues. It is likely therefore that the *H. pullorum* PglH glycosyltransferase transfers the fifth and final 203 Da HexNAc sugar to complete the lipid linked pentasaccharide glycan. Biochemical analysis of purified PglH functioning *in vitro* provided further evidence that PglH is responsible for addition of the terminal sugar of the *H. pullorum* pentasaccharide glycan (Jervis and Linton, unpublished data).

Insertional mutagenesis of *H. pullorum* PglI and PglL was unsuccessful, resulting in no transformants. It remains to determined whether these glycosyltransferases are involved in assembly of the PglB1-dependent N-linked glycan and due to difficulties in creating insertional knock out mutants, these glycosyltransferases may also be involved in a separate essential *H. pullorum* PglB2-dependent N-linked protein glycosylation system. Future work beyond the scope of this thesis will aim to identify a second, PglB2 dependent, N-glycosylation system in *H. pullorum*.

In summary, this chapter describes how the HgpA glycoprotein can be used to determine the N-linked glycan structure of glycoproteins from *H. pullorum* wild type and glycosyltransferase mutants. Data presented in this chapter have allowed the proposal of a model for the assembly of *H. pullorum* lipid-linked pentasaccharide that is transferred to protein in a PglB1-dependent manner (Figure 6.5). In this model, *H. pullorum* PglC transfers the first 203 Da HexNAc sugar and PglA transfers the second sugar, a 216 Da residue. In the absence of PglA, the 203 Da reducing sugar is converted to a 228 Da trideoxydiacetamidohexose, possibly by the action of PglDEF. The *H. pullorum* PglJ glycosyltransferase is involved in transfer of the third, and possibly also the fourth 217 Da residue. The *H. pullorum* PglH glycosyltransferase then adds a terminal HexNAc residue to complete the *H. pullorum* lipid linked pentasaccharide which is transferred to protein by the action of PglB1.

Figure 6.5 Model comparing the assembly of the *C. jejuni* lipid linked heptasaccharide glycan with the proposed assembly of the *H. pullorum* pentasaccharide glycan. In *H. pullorum* a pentasaccharide is assembled on a lipid carrier at the cytoplasmic side of the inner bacterial membrane by the action of four glycosyltransferases. Assembly of the branched *C. jejuni* heptasaccharide also requires similar glycosyltransferases and an additional glucosyltransferase named PglI (Linton *et al*., 2005). Sugar residues appear as coloured circles and phosphate groups as yellow circles containing a P. Black oval shapes represent membrane bound lipid carriers.



Chapter 7

The *H. pullorum* Biosynthetic Enzymes WbpO and WbpS are Required for PglB1-Dependent N-Linked Protein Glycosylation with a Pentasaccharide Glycan

7.1 Background

Chapter 6 describes the role of four glycosyltransferases (PglA, PglC, PglH and PglJ) in the PglB1-dependent *H. pullorum* N-glycosylation system. Tandem MALDI Mass Spectrometry identified a *H. pullorum* pentasaccharide glycan consisting of 203 Da HexNAc reducing and non-reducing end sugars, with unusual 216 Da (one residue) and 217 Da (two residues) sugars. A 217 Da residue in the N-linked flagellar glycan of *Methanococcus maripaludis* was structurally characterized by Kelly *et al.* (2009) and identified as 2-acetamido-2,4-dideoxy-5-*O*-methyl-α-L-erythro-hexos-5-ulo-1,5-pyranose. There were no other characterized examples of sugars with these masses until 2010 when a 216 Da sugar from the LPS of *Pseudomonas aeruginosa* O6 and *E. coli* O121 was structurally characterized and identified as a 2-acetamido-2-deoxy-D-galacturonamide (GalNAcAN) (King *et al*., 2010). This 216 Da uronamide sugar is synthesized from a 217 Da sugar (GalNAcA) in turn synthesized from a GlcNAc (203 Da) sugar by the action of a WbpO enzyme (Figure 7.1). Interestingly, WbpO has 32% sequence similarity to a *H. pullorum* protein encoded by a gene located adjacent to the *pgl*KILG gene cluster (Figure 7.1; King *et al*., 2010). The similarity of these two genes along with the presence of 217 Da sugar residues in the *H. pullorum* N-glycan suggests a role for this gene product in the conversion of a 203 Da HexNAc residue to a 217 Da HexNAcA residue.

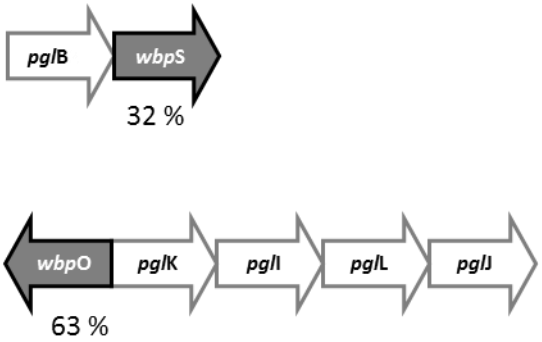
The WbpS enzyme of *P. aeruginosa* O6 and *E. coli* O121 catalyzes the transfer of an amide group to the 217 Da GalNAcA sugar, resulting in a 216 Da GalNAcAN residue (Figure 7.1; King *et al*., 2010). WbpS has 63% sequence similarity to the gene product encoded by a gene which is located directly downstream of the gene encoding the second putative *H. pullorum* N-linked OST, PglB2.

These observations led us to predict that these *H. pullorum* gene products may have similar enzymatic functions as homologues from *E. coli* and *P. aeruginosa*. More specifically they may function in biosynthesis of the uncharacterized 216/217 Da residues in the *H. pullorum* N-linked pentasaccharide. Due to the similarity of these *H. pullorum* gene products (previously annotated as G6D and AsnB), with *E. coli*

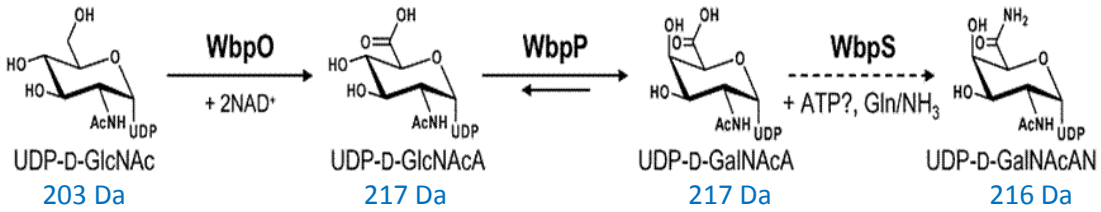
and *P. aeruginosa* homologues they were renamed WbpO and WbpS. In this chapter the genetic manipulation, protein expression and Mass Spectrometry techniques described in previous chapters were used to experimentally characterize the role of the *H. pullorum* WbpO and WbpS proteins in the biosynthesis of the PglB1-dependent N-linked glycan.

Figure 7.1 *H. pullorum wbp*O and *wbp*S gene locations and the function of WbpO and WbpS in *P. aeruginosa* O6 and *E. coli* O121 strains. (A) Schematic diagram of *H. pullorum wbp*O and *wbp*S genes showing their location on the *H. pullorum* chromosome with respect to other *pgl* genes. Genes indicated by arrows are not to scale and percentage values below the *wbp*O and *wbp*S genes are the pair wise levels of identity between amino acid sequences encoded by *H. pullorum* and *P. aeruginosa wbp*O and *wbp*S. (B) Pathway for the biosynthesis of UDP-2-acetamido-2-deoxy-d-galacturonamide (UDP-D-GalNAcAN). The *P. aeruginosa* O6 and the *E. coli* O121 LPS structures contain a 216 Da sugar that is synthesized from GlcNAc by the action of WbpO, WbpP and WbpS. The WbpO enzyme is involved in the addition of an oxygen molecule to the 203 Da GlcNAc sugar to create a carboxylic acid group resulting in a 217 Da GlcNAcA sugar. The WbpS enzyme converts a hydroxyl group to an amide group resulting in the 216 Da GalNAcAN sugar. Image (B) was taken from King *et al.*, (2010).

(A)



(B)



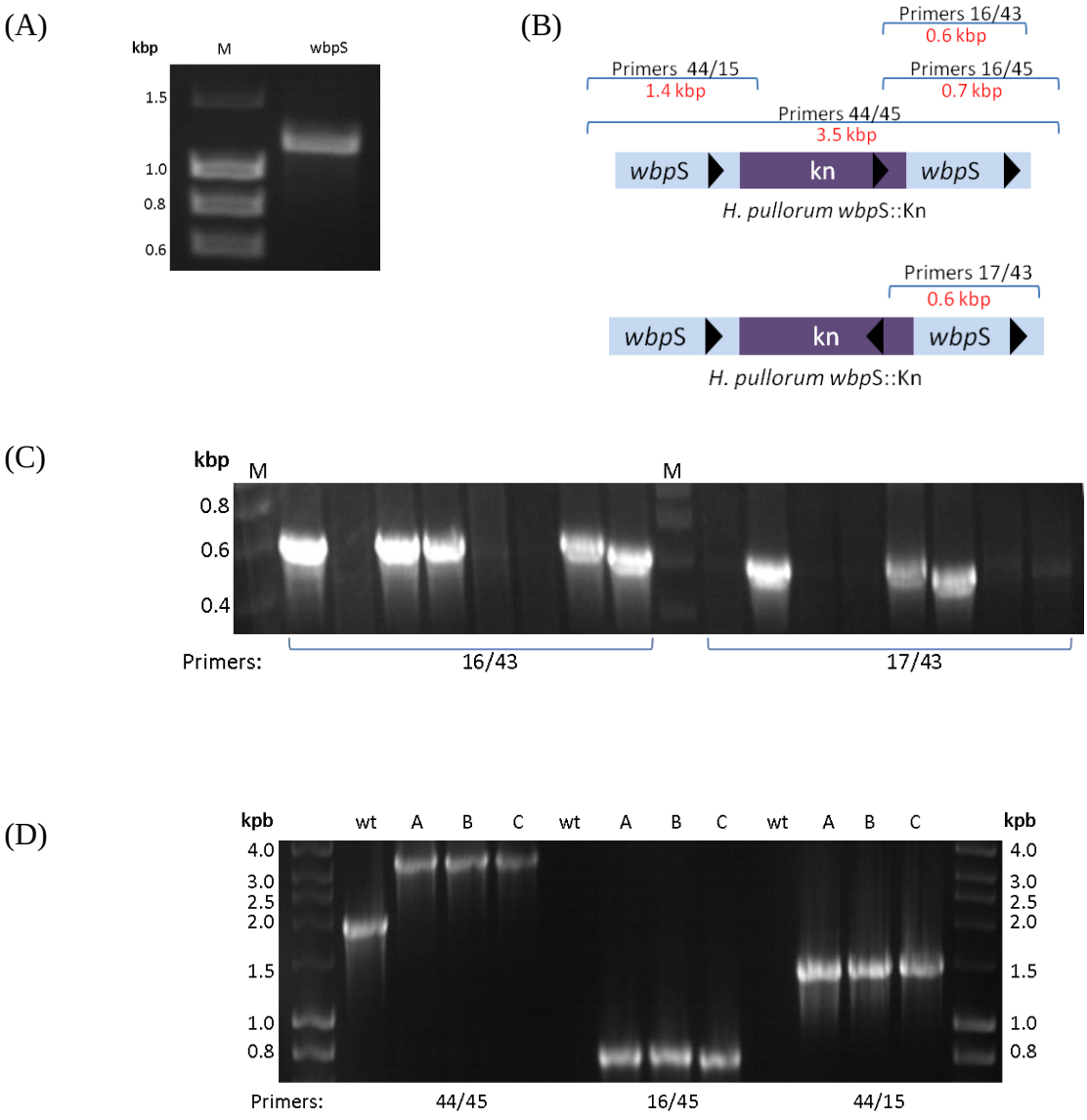
7.2 Results

7.2.1 Construction of *H. pullorum wbp*O::Kn and *wbp*S::Kn mutants.

A *H. pullorum wbp*O::Kn mutant was previously constructed by insertional mutagenesis (Jervis, Lord and Linton, unpublished). To construct the *H. pullorum wbp*S::Kn mutant, a 1.2 kbp central portion of the *H. pullorum wbp*S gene was amplified by PCR with primers 42 and 43 (Figure 7.3) and ligated into plasmid pGEM T-easy (Promega) to create pHp*wbp*S. A BamHI restriction site was created in the central portion of the *wbp*S gene by site directed mutagenesis (primers 46 and 47). A kanamycin resistance cassette was PCR amplified with flanking BamHI restriction enzyme sites (primers 13 and 14), and ligated into the pHp*wbp*S plasmid and cloned in *E. coli*. The orientation of the kanamycin resistance cassette within the *wbp*S gene was determined by PCR (Figure 7.1 C) and plasmid with the cassette in the same transcriptional orientation as *wbp*S was selected (Figure 7.3, lane 8). Following electroporation into wild-type competent *H. pullorum* cells and selection on blood agar containing kanamycin, genomic DNA was isolated from transformants and checked by PCR for insertion of the antibiotic resistance cassette within the *wbp*S gene (Figure 7.2).

Along with the previously constructed *wbp*O gene knock out mutant, the *wbp*S knock out mutant will allow functional characterization of the role of these two gene products in *H. pullorum* N-linked glycosylation. The first approach employed the *in vitro* OST assay and tandem MALDI Mass Spectrometry to identify the glycan structure transferred to an N-glycosylation sequon containing peptide by enzymes present in *H. pullorum* membrane preparations. The second approach involved the purification of a hexa histidine tagged HgpA glycoprotein from *H. pullorum* and identification of any modifications to the HgpA protein by intact mass analysis and tandem MALDI Mass Spectrometry.

Figure 7.2 Construction and verification of the *H. pullorum wbpS*::Kn mutant. (A) PCR of internal portion of the *wbp*S gene from *H. pullorum* genomic DNA (primers 42/43). See Appendix I for primer list. (B) Schematic diagram of the *wbp*S gene with a kanamycin resistance cassette insertion indicating primers used for PCR amplification. Black arrowheads indicate the direction of transcriptional orientation and the predicted sizes of PCR products are indicated in red. Agarose gel electrophoresis of PCR amplification products using primers from (B) indicates the orientation of the kanamycin resistance cassette within *wbp*S from pHp*wbp*S::Kn plasmids 1-8 (C) and three *H. pullorum wbp*S::Kn strains (A-C) (D). Plasmid verified in lane 1 (C) was used to create *H. pullorum wbp*S::Kn strains and clone A, verified in (D), was used in subsequent experiments.



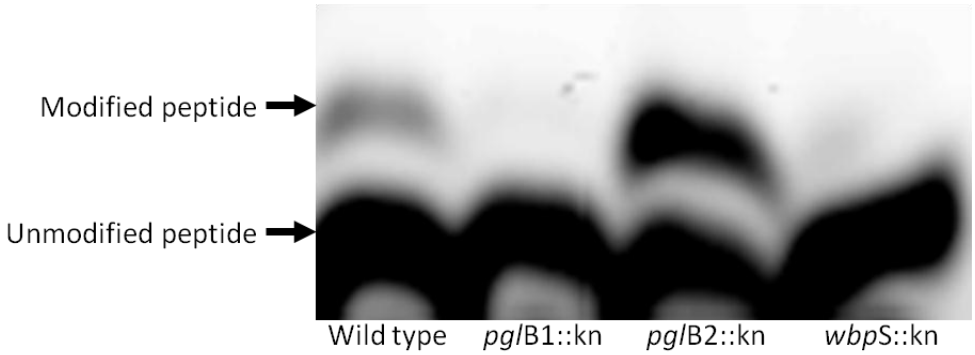
7.2.2 Role of WbpS in the glycosylation of a small peptide *in vitro*

An *in vitro* OST assay (Kohda *et al.*, 2007) has shown that *H. pullorum pgl*B1 is essential for the transfer of a pentasaccharide glycan to an asparagine residue of a peptide based on the optimal *C. jejuni* N-glycosylation sequon (Jervis *et al.*, 2010). Chapter 3 describes the use of this assay to investigate the requirement for the *H. pullorum* PglB2 putative OST in this PglB1-dependent system. In this chapter, the use of this OST assay to investigate N-glycosylation in the *wbp*S::Kn mutant is described.

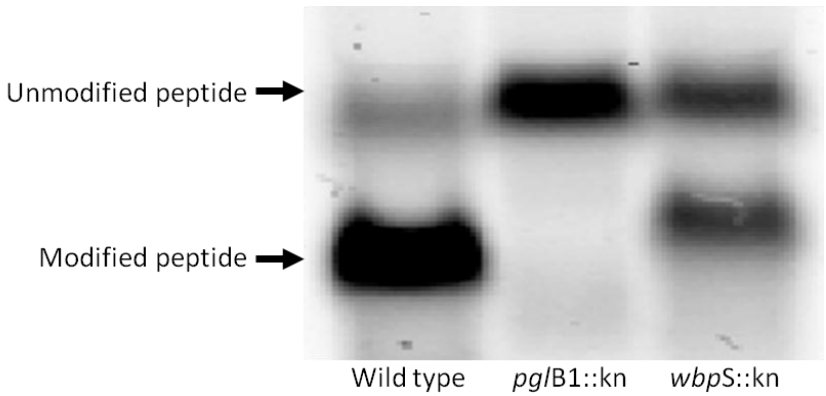
Cell membranes from *H. pullorum* wild-type, *pgl*B1::Kn and *wbp*S::Kn mutants were incubated overnight with FITC or FAM-labelled peptides (see Methods for details). Reaction products were run on Tricine SDS-PAGE gels and detected via the fluorescent tag. When a FAM labelled peptide was employed (Figure 7.3 A), insertional mutagenesis of *wbp*S appears to abolish glycosylation completely, resulting in a peptide consistent in electrophoretic mobility with that of unmodified peptide as seen for the *pgl*B1::Kn mutant. However, a similar approach using a FITC-labelled peptide indicated that *wbp*S::Kn mutation did not abolish glycosylation but rather led to production of a glycopeptide with altered electrophoretic mobility (Figure 7.3 B). This indicated that an N-linked glycan was present in the *wbp*S::Kn mutant but this glycan differed in mass to that of wild type *H. pullorum*. It seems likely that this altered mobility FAM-labelled glycopeptide co-migrated with unmodified peptide near the bottom of the gel (Figure 7.3 A). In order to further characterize N-linked glycan structure, the glycopeptide derived from glycosylation activity of the *H. pullorum wbp*S::Kn mutant was purified (see Methods for details) and analysed by tandem MALDI Mass Spectrometry.

Figure 7.3 The *in vitro* modification of FAM and FITC labelled peptides co-incubated with membrane preparations from a *H. pullorum wbp*S::Kn strain. (A) Tricine SDS-PAGE of FAM labelled peptide following incubation with membranes from wild type *H. pullorum* and *pgl*B1::Kn, *pgl*B2::Kn and *wbp*S::Kn mutants. Glycosylated peptide is indicated by bands showing reduced mobility. The difference in mobility between peptide from assays with *H. pullorum* wild type (pentasaccharide) and *pgl*B1::Kn mutant (unmodified peptide) is not as marked as expected due to the negative charge of the *H. pullorum* pentasaccharide glycan. The lack of lower mobility FAM-labelled glycopeptide with membranes from the *H. pullorum wbp*S::Kn mutant is likely due to co-migration of unmodified and modified peptide. (B) In contrast, FITC-labelled glycopeptide resulting from co-incubation of peptide with membranes derived from the *H. pullorum wbp*S::Kn mutant had an increased mobility compared to unmodified peptide but was less mobile than peptide incubated with the wild type strain.

(A)



(B)



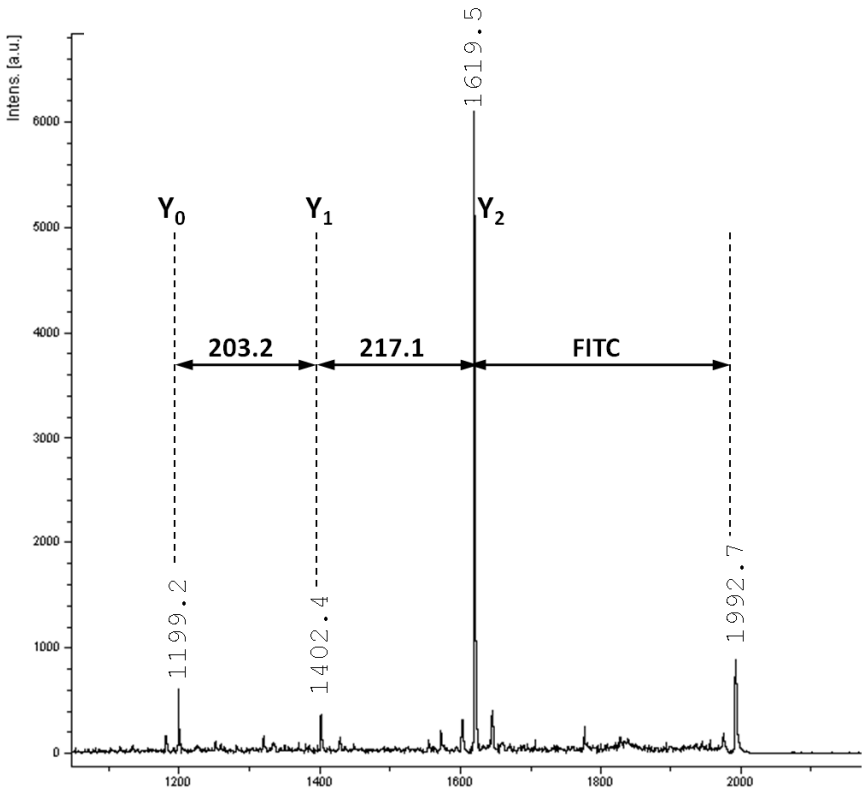
7.2.3 The role of WbpS in the glycosylation of a small peptide *in vitro*

To structurally characterize modified peptides generated by the OST assay, a biotin labelled fluorescent peptide was employed (FITC-ADQNATAK-biotin) with a mass of 1587 Da and products were purified by using streptavidin-coated magnetic beads prior to Mass Spectrometry (see methods for details).

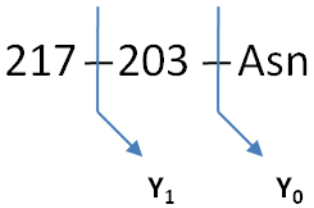
MALDI Mass Spectrometry analysis of the purified peptide from a reaction mixture containing *H. pullorum wbp*S::Kn membranes identified a peak at *m/z* 1992.7 that corresponds to the 1199 Da peptide modified with a glycan structure of 794 Da in mass. Tandem MALDI Mass Spectrometry of this peak produced a fragmentation series consistent with a disaccharide structure (Figure 7.4). The peak of *m/z* 1619.5 was most likely glycopeptide lacking the FITC moiety. The *m/z* 1402.4 and *m/z* 1199.2 peaks were consistent with the sequential loss of a 217 and a 203 Da residue respectively. The glycan structure of the glycopeptide co-incubated with *H. pullorum wbp*S::Kn membrane preparations was therefore identified as Asn – 203 – 217. This is further evidence that WbpS is essential for the assembly of the complete *H. pullorum* N-linked pentasaccharide. The absence of a 216 Da sugar suggests that WbpS may be involved in its biosynthesis, most likely from a 203 Da residue. In the absence of a 216 Da residue, a 217 Da residue was transferred to the monosaccharide glycan to generate a disaccharide glycan which was not modified further.

Figure 7.4 Tandem MALDI Mass Spectrometry of a peptide co-incubated with membrane preparations from the *H. pullorum wbp*S::Kn mutant. (A) Tandem Mass Spectrometry spectrum of the *m/z* 1992.7 precursor ion generated following incubation of *H. pullorum wbp*S::Kn membranes with a biotin-labelled FITC peptide (FITC-ADQNATAK-biotin, 1587 Da in mass). Fragment ions resulting from the sequential loss of 217 and 203 Da sugar residues are indicated in the spectrum. A peak corresponding to the Y ion of the biotinylated peptide lacking FITC was present at *m/z* 1199. (B) N-linked disaccharide structure of the *H. pullorum wbp*S::Kn mutant consistent with the fragmentation pattern seen by tandem MALDI Mass Spectrometry.

(A)



(B)



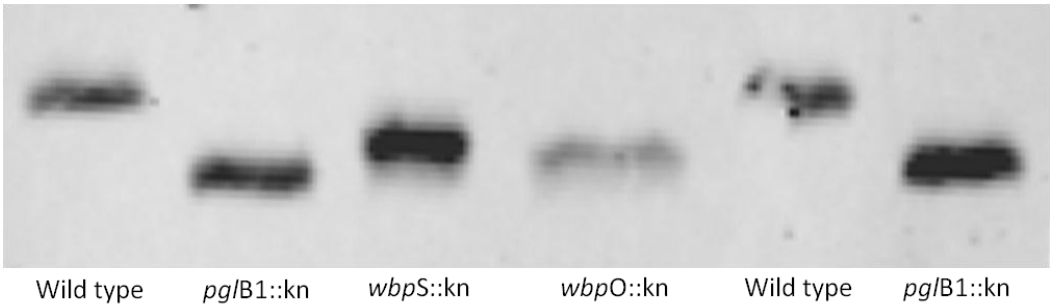
7.2.4 Glycosylation of HgpA from *H. pullorum wbp*O::kn and *wbp*S::kn mutants.

Chapter 6 describes the use of the *H. pullorum* glycoprotein, HgpA to identify the role of glycosyltransferases in the assembly of the PglB1-dependent N-linked pentasaccharide. This method was also used to investigate the role of the *H. pullorum* putative sugar biosynthesis genes, *wbp*O and *wbp*S.

The gene expressing a hexa histidine-tagged HgpA protein was introduced into *H. pullorum wbp*O::Kn and *wbp*S::Kn mutants using methods described in Chapter 5. Integration of this *hgp*A gene on the chromosome was confirmed by PCR (data not shown). The HgpA protein was purified from these strains and analysed by western blotting with anti histidine antibody (Figure 7.5).

The HgpA protein derived from *H. pullorum wbp*O::Kn and *wbp*S::Kn mutants had increased electrophoretic mobility compared to HgpA derived from the wild type *H. pullorum* strain but was less mobile than HgpA derived from the *H. pullorum pgl*B1::Kn mutant. These results indicate that both WbpO and WbpS are required for assembly of a complete *H. pullorum* pentasaccharide. Further analysis of the glycan structures of HgpA protein from these *H. pullorum* mutants was carried out using Mass Spectrometry.

Figure 7.5 Glycosylation of *H. pullorum* HgpA derived from *H. pullorum wbp*O::Kn and *wbp*S::Kn mutants. Western blot of purified, hexa histidine tagged *H. pullorum* HgpA derived from *H. pullorum wbp*S::Kn and *wbp*O::Kn mutants. Unmodified HgpA derived from *H. pullorum* wild type and glycosylated HgpA from the *pgl*B1::Kn mutant are included as controls. HgpA proteins derived from *H. pullorum wbp*O::Kn and *wbp*S::Kn mutants have increased mobility compared to HgpA from the wild type strain but appear less mobile than HgpA derived from the *pgl*B1::Kn mutant suggesting an intermediate level of glycosylation.

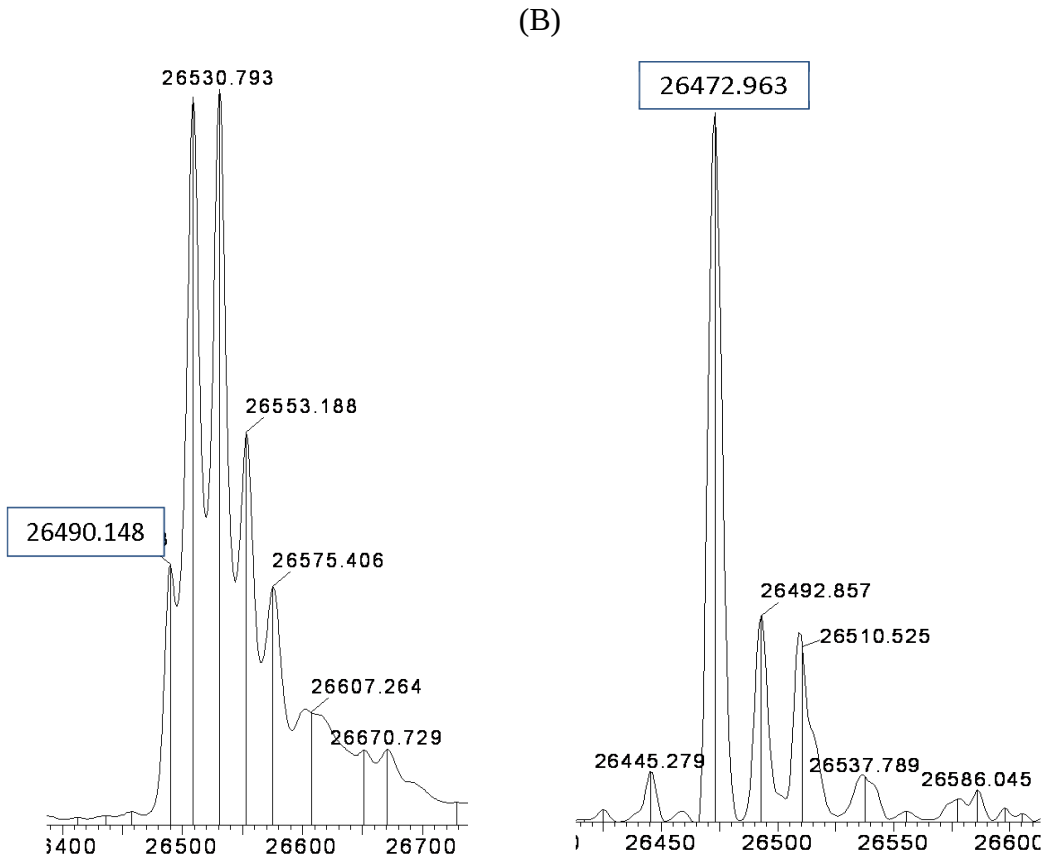


7.2.5 Intact mass analysis of HgpA from *H. pullorum wbp*O::Kn and *wbp*S::Kn mutants.

Intact mass analysis data presented in Chapter 6 determined that the mass of HgpA from *H. pullorum* wild type and *pgl*B1::Kn strains was 27127 Da and 26070 Da respectively. The intact mass of HgpA derived from the *H. pullorum wbp*S::Kn mutant was 26490 Da, consistent with a 420 Da modification likely composed of a disaccharide of 203 and 217 Da residues (Figure 7.6 A). These observations suggest that WbpS is indeed required for the biosynthesis of this 216 Da residue as predicted from its similarity to the *P. aeruginosa* O6 WbpS. Interestingly the 216 Da residue that is linked to the reducing end 203 Da residue of the N-linked pentasaccharide present on HgpA derived from the wild type strain is replaced by a 217 Da residue.

The intact mass of HgpA derived from the *H. pullorum wbp*O::Kn mutant was 26473 Da with a predicted glycan mass of 403 Da (Figure 7.6 B). It was unclear which sugars make up this 403 Da glycan therefore further investigation was required. Tandem MALDI Mass Spectrometry techniques were used to further characterize the glycan structures present on the HgpA glycoprotein derived from the *wbp*O::Kn and *wbp*S::Kn mutants.

Figure 7.6 Intact mass analysis of HgpA derived from *H. pullorum wbp*S::Kn (A) and *wbp*O::Kn (B) mutants. The intact mass for each protein is the boxed value indicated.



7.2.6 Tandem MALDI Mass Spectrometry of HgpA peptides derived from *H. pullorum wbp*O::Kn and *wbp*S::Kn mutants.

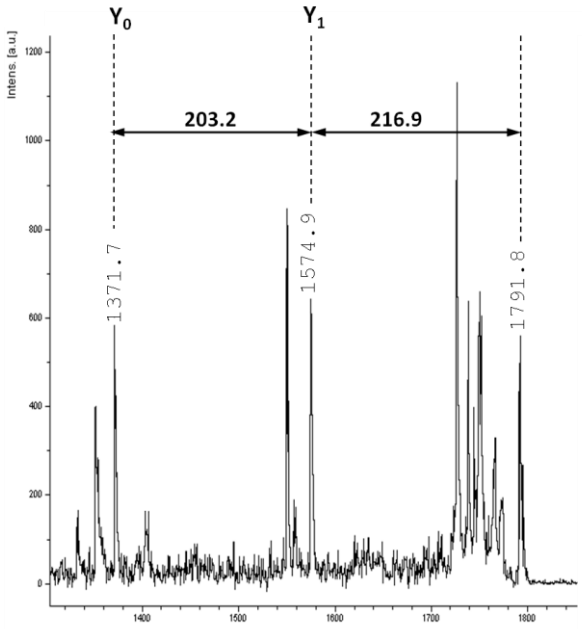
Tandem MALDI Mass Spectrometry was performed on trypsin digested HgpA peptides derived from *H. pullorum wbp*S::Kn and *wbp*O::Kn mutants (Figure 7.7). Peaks of *m/z* 1791.8 and 1803.1 derived from *wbp*S::Kn and *wbp*O::Kn strains respectively were fragmented to generate a series of peaks corresponding to the glycan structure.

Fragmentation of the 1791.8 Da peak from the *wbp*S::Kn mutant generated a peak of *m/z* 1574.9 consistent with the loss of a 217 Da residue and a peak of *m/z* 1371.7 consistent with the mass of the peptide following the sequential loss of the 217 and 203 Da residues (Figure 7.7 A). These data, combined with intact mass analysis data, indicate that the *H. pullorum* WbpS is involved in N-linked pentasaccharide glycan biosynthesis, more specifically synthesis of the 216 Da residue which is absent in HgpA from this strain. This observation is in agreement with the predicted role of *H. pullorum* WbpS based on *P. aeruginosa* and *E. coli* homologues.

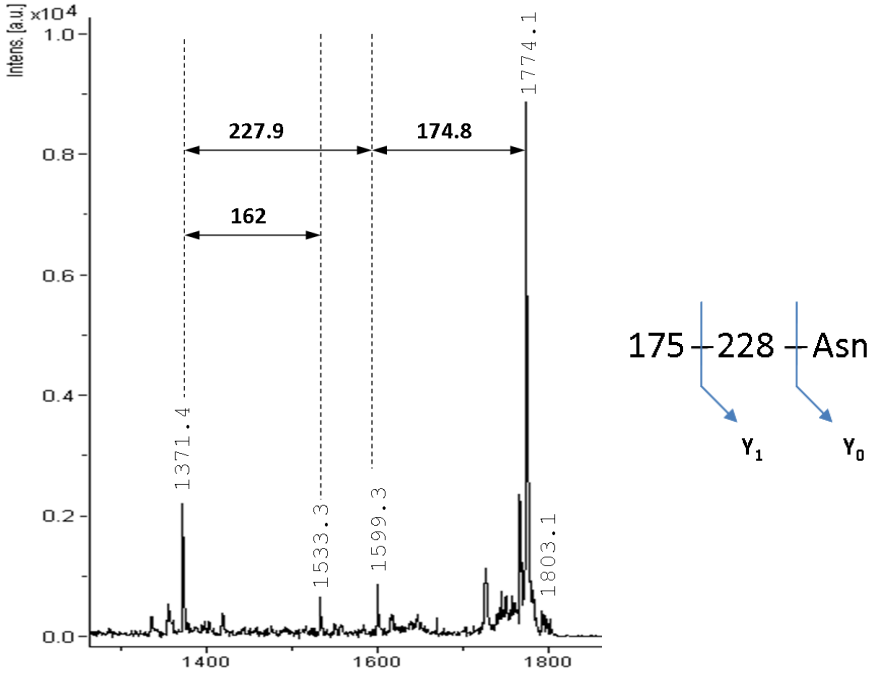
Intact mass analysis of HgpA from the *H. pullorum wbp*O::Kn mutant identified a modification of 403 Da. Tandem MALDI Mass Spectrometry was therefore performed to characterize this structure. Fragmentation of the HgpA *m/z* 1803.1 glycopeptide peak derived from this mutant revealed a structure consistent with a disaccharide glycan consisting of a 228 Da reducing end sugar and a terminal 175 Da sugar (Figure 2.7 B). An additional peak of m/z 1533 was consistent with a 162 Da Hex reducing end sugar but this result was not confirmed. Based on homology with WbpO from *P. aeruginosa*, it was predicted that *H. pullorum* WbpO was involved in the conversion of the 203 Da residue to a 217 Da residue. These data are in agreement with this prediction because in the *wbp*O::Kn mutant the glycan structure is deficient in 216 and 217 Da residues. However, the presence of 228 and 175 Da residues as a disaccharide is confusing and indicates the *wbp*O::Kn mutation has a complex effect on N-linked glycosylation possibly involving gene products involved in the uncharacterised PglB2-dependent

Figure 7.7 Glycan structure of HgpA derived from *H. pullorum wbp*O::Kn and *wbp*S::Kn mutants. Tandem mass spectra of fragmentation of the *m/z* 1791.8 and 1803.1 precursor ions generated following trypsin digestion of *H. pullorum* HgpA from the *wbp*S::Kn (A) and *wbp*O::Kn (B) mutants respectively. Fragment ions resulting from the sequential loss of sugar residues are indicated in the spectrum. A peak corresponding to the Y ion of the HgpA glycopeptide was present between *m/z* 1371 and 1372.

(A)



(B)



7.3 Discussion

By analogy to detailed analysis of eukaryotic and the bacterial *C. jejuni* N-linked glycosylation process, the *H. pullorum* N-linked pentasaccharide is likely to be assembled sequentially on a lipid carrier by the action of glycosyltransferase enzymes. This pentasaccharide structure has recently been characterized (Jervis *et al*., 2010) and contains HexNAc sugars commonly found in bacterial glycans but also two unusual sugars of 216 and 217 Da. The aim of this chapter was to investigate two genes potentially involved in the biosynthesis of these less common sugars. Chapter 6 describes the use of the hexa histidine tagged HgpA protein to identify glycosyltransferase encoding genes involved in assembly of the *H. pullorum* N-linked glycan structure. The same methods were employed in Chapter 7 to investigate the importance of putative sugar biosynthesis gene products in the assembly of this pentasaccharide in *H. pullorum.*

Interestingly, the *P. aeruginosa* O6 LPS antigen contains a 216 Da uronamide sugar which was identified as GalNAcAN (King *et al*., 2010). Figure 7.1 B illustrates how this sugar residue is synthesised in *P. aeruginosa* from a HexNAc precursor (GlcNAc). A dehydrogenase enzyme named WbpO catalyzes the conversion of GlcNAc to GlcNAcA (Zhao *et al*., 2000) which is then converted to GalNAcA by the epimerase, WbpP (Creuzenet *et al*., 2000). In *P. aeruginosa* O6, the 217 Da GalNAcA sugar is converted to the 216 Da sugar GalNAcAN by the action of the amidotransferase, WbpS (King *et al.*, 2010). The *H. pullorum* genome contains genes, previously of unknown function, that are now named *wbp*O and *wbp*S due to the amino acid sequence similarity of their gene products with WbpO and WbpS from *P. aeruginosa* O6 (Creuzenet *et al*., 2000; King *et al*., 2010). The *H. pullorum wbp*O gene is located directly upstream of the *pgl*KILG cluster and the *wbp*S gene is located directly downstream of *pgl*B2 (Figure 7.1 A).

A *H. pullorum wbp*O::Kn knock out mutant was previously constructed but not characterized and this chapter describes the construction of a *H. pullorum wbp*S::Kn mutant. In order to investigate their function in these strains a hexa histidine tagged HgpA protein was employed.

Results indicated that the *wbp*O gene was required for glycosylation of the HgpA protein with a pentasaccharide glycan. Intact mass analysis of HgpA from the *H. pullorum wbp*O::Kn mutant suggested a 403 Da glycan structure (Figure 7.6 B). In agreement with this, tandem MALDI Mass Spectrometry of HgpA from this strain indicated a disaccharide containing a reducing end 228 Da residue and an uncharacterized 175 Da residue. It is likely that in the absence of WbpO, HgpA proteins are glycosylated with a 228 Da trideoxydiacetamidohexose as the reducing end sugar and it was already shown in Chapter 6 that in the absence of the PglA glycosyltransferase, predicted to transfer the first 216 Da sugar, the reducing end sugar was a 228 Da residue rather than the usual 203 Da HexNAc. Interestingly, a small peak of m/z 1533 was consistent with a 162 Da Hex reducing end sugar which was unexpected as *C. jejuni* requires an acetamido group at the C-2 position of the reducing end sugar (Wacker *et al*., 2006) but further evidence would be required to confirm this result. Although the precise function of WbpO in the biosynthesis of glycan sugars was not confirmed, it is clear from these results that the N-glycan structure decorating proteins in a *wbp*O::Kn mutant does not contain either 216 Da or 217 Da sugar residues. This adds strength to the hypothesis that WbpO is required for the biosynthesis of the 217 Da sugar.

Based on *P. aeruginosa* WbpS function (King *et al*., 2010) it was predicted that the *H. pullorum* WbpS enzyme may also be involved in the conversion of a 217 Da sugar to create the 216 Da sugar forming part of the *H. pullorum* pentasaccharide glycan. In agreement with this prediction, intact mass analysis (Figure 7.6 A) of HgpA from the *H. pullorum wbp*S::Kn mutant and tandem Mass Spectrometry results (Figure 7.7 A) indicate that the *wbp*S gene, situated downstream of *pgl*B2, is required for the biosynthesis of a complete pentasaccharide N-glycan structure. The HgpA glycan structure from this *wbp*S::Kn mutant was identified as a 203 Da reducing end sugar attached to a 217 Da residue. It appears that in the absence of a 216 Da residue, the 217 Da residue is transferred to the 203 Da reducing end sugar but the glycan structure is not extended beyond this disaccharide likely due to glycosyltransferase specificities. The *W. succinogenes pgl* locus contains two *wbp*S orthologues, one of which is directly downstream of *pgl*L and one directly downstream of *pgl*B, which encode proteins with 55% and 34% overall sequence

similarity to *H. pullorum* WbpS respectively. As the *W. succinogenes* contains three 216 Da residues (Dell *et al*., 2010) it is likely that a similar mechanism exists for the conversion of a 203 Da HexNAc to the 216 Da HexNAcAN involving one or both *wbp*S genes but further investigation is required.

In summary, this chapter describes two genes required for assembly of the complete pentasaccharide glycan on HgpA proteins in the *H. pullorum* PglB1 dependent N-linked glycosylation system. We hypothesise, based on similarity to *P. aeruginosa* O6 WbpO and WbpS enzymes (King *et al*., 2010), and our observations on the glycan structures in these mutants that the *H. pullorum* WbpO enzyme is involved in the biosynthesis of a 217 Da sugar and the WbpS enzyme is required for the conversion of this sugar to a 216 Da residue.

Chapter 8

General Discussion

8.1 Established Knowledge Regarding Helicobacter N-glycosylation

The well characterized *C. jejuni* N-linked protein glycosylation system was the first bacterial N-glycosylation system discovered (Linton *et al*., 2002; Wacker *et al.*, 2002; Linton *et al*., 2005). *H. pullorum*, *H. canadensis* and *H. winghamensis* are the only *Helicobacter* species with *pgl*B orthologues and interestingly they each contain two *pgl*B genes. At the outset of this thesis N-linked glycosylation systems in *H. pullorum* were less well characterized than in *C. jejuni* but recent work from our laboratory used an OST assay to identify a *H. pullorum* pentasaccharide N-linked glycan (Jervis *et al*., 2010). In this assay, a peptide was glycosylated *in vitro* with a pentasaccharide glycan by wild type *H. pullorum* membrane preparations, but the peptide was not glycosylated when incubated with membrane preparations from a *H. pullorum pgl*B1::Kn mutant (Jervis *et al*., 2010). The *H. pullorum* N-linked glycan was structurally characterized and contains HexNAc sugars commonly found in bacterial glycans but also two unusual sugars of 216 and 217 Da (Jervis *et al*., 2010). The work described in this thesis has significantly expanded knowledge of *H. pullorum* N-linked glycosylation and draws out similarities and significant differences with the well established and extensively studied *C. jejuni* N-linked glycosylation system (Linton *et al*., 2002; Wacker *et al*., 2002; Linton *et al*., 2005). The work described has provided increased understanding in three distinct areas, namely (a) the clear demarcation between PglB1 and PglB2 dependent N-glycosylation systems; (b) the identification and characterization of the first *H. pullorum* N-linked glycoproteins and finally (c) their use to investigate the role of glycosyltransferases and biosynthetic enzymes in N-linked glycan assembly.

8.2 PglB1 and PglB2 Dependent N-glycosylation Systems

The *H. pullorum pgl*B1 gene was previously knocked-out but the *pgl*B2 gene was thought essential due to difficulties creating a knock-out mutant strain (Jervis *et al.*, 2010). Results described in Chapter 3 now describe insertional mutagenesis of the *pgl*B2 gene, which was over 100-fold less efficient than insertional mutagenesis of *pgl*B1. The reasons for this were not identified but it is possible that the effect is region specific (i.e. insertional mutagenesis is more or less efficient depending on the

target location on the bacterial chromosome) or alternatively a suppressor mutation is required to allow insertional mutagenesis of *pgl*B2. Given that there is no evidence for region specific effects on insertional knock out mutation frequency in the literature on *C. jejuni* or *H. pylori* we favour the latter explanation. Although beyond the scope of this thesis, it would be interesting to search for the location of proposed suppressor mutants.

Two independent approaches were used to compare the effect of insertional mutagenesis of *H. pullorum pgl*B1 and *pgl*B2 genes on N-linked protein glycosylation. The first approach involved the established *in vitro* OST assay (Jervis *et al*., 2010) and Mass Spectrometry based techniques and had the advantage that identification of a *H. pullorum* N-linked glycoprotein was not required. In this assay, the OST(s) present in wild type and *pgl*B2 mutant *H. pullorum* membrane preparations were able to transfer a glycan from lipid linked oligosaccharide onto peptides containing the D/E-X-N-X-S/T sequon, which is required for *C. jejuni* N-glycosylation. As previously described by Jervis *et al.* (2010) this peptide was not glycosylated following incubation with membrane preparations from the *pgl*B1::Kn mutant. Analysis of Tandem MALDI Mass Spectrometry data of glycopeptide produced following incubation with the *H. pullorum pgl*B2::Kn mutant membranes revealed a pentasaccharide glycan identical to that produced by wild type *H. pullorum* membranes (Jervis *et al*., 2010). This is clear evidence that PglB2 is not involved in the transfer of the N-linked pentasaccharide glycan.

The second approach involved expression, purification and tandem MALDI Mass spectrometry of a labelled copy of the *H. pullorum* HgpA protein (identified in Chapter 4) from *H. pullorum* wild type, *pgl*B1::Kn and *pgl*B2::Kn mutants. As expected, HgpA from the wild type strain was glycosylated with the previously characterized pentasaccharide glycan (Jervis *et al*., 2010). Interestingly, the HgpA N-linked pentasaccharide in the *pgl*B2::Kn mutant was similar in structure but had a 186 instead of a 203 Da reducing end sugar. These results are in contrast with *in vitro* data and highlight the importance of using complimentary approaches to investigate mutant phenotypes.

HgpA protein derived from wild type *H. pullorum* had an increased mass compared to HgpA from the *pgl*B1::Kn mutant and this difference in mass was 1057 Da which is consistent with the 1056 Da mass of the *H. pullorum* pentasaccharide glycan (Jervis *et al*., 2010). Data could not be obtained by intact mass analysis of HgpA derived from the *H. pullorum pgl*B2::Kn mutant despite three independent attempts with three separately purified HgpA protein samples. It is possible that these samples contained HgpA glycoproteins with different glycan structures and as intact mass is only accurate for determining the mass of a protein species of a single defined mass it would explain why data could not be obtained. The *C. jejuni pgl* locus contains genes encoding a dehydratase (PglF), an aminotransferase (PglE) and an acetyltransferase (PglD) that are involved in the conversion of a GlcNAc residue to form bac (Olivier *et al*., 2006). In this pathway, PglF converts the 203 Da GlcNAc residue to a 185 Da keto sugar which is converted to a 186 Da amino sugar by PglE and then to the 228 Da bac residue by PglD (Olivier *et al*., 2006). In the *H. pullorum pgl*B2::Kn mutant, the presence of the 186 Da sugar instead of the 203 Da HexNAc may indicate involvement of PglE and PglF although it is unclear why this would occur.

These data indicate that the *H. pullorum* PglB2 has a role that is most likely separate from the *H. pullorum* PglB1-dependent N-glycosylation system. It is possible that *H. pullorum* PglB2 modifies glycoproteins with a different glycan structure to the PglB1-dependent pentasaccharide in a pathway involving further, uncharacterized glycosyltransferases or the two systems could be overlapping in terms of the glycan structure and glycosyltransferases involved in assembly. This requires further investigation beyond the scope of this thesis. Further investigations may involve comparing *H. pullorum* glycoproteins from wild type and *pgl*B2::Kn mutants by two dimensional SDS-PAGE or identifying lectins that specifically bind to *H. pullorum* glycoproteins.

8.3 Identification and Characterization of the First *H. pullorum* N-linked Glycoproteins

In order to further characterize N-linked protein glycosylation in *H. pullorum*, putative N-linked glycoproteins were identified. The bioinformatic approach to analysing proteins from *H. pullorum, H. canadensis* and *H. winghamensis* described in Chapter 4 identified 131 putative glycoproteins defined as having a signal peptide sequence and at least one N-glycosylation sequon. Of the genes encoding putative glycoproteins, 67 were present in *H. pullorum* (See Appendix IV for full list) and two of these were chosen for further investigation due to their similarity with previously identified *C. jejuni* glycoproteins. These two *H. pullorum* proteins, named Hp0114 (HPMG_00410) and HgpA (HPMG_01281), were produced in *E. coli* along with the enzymes required for *C. jejuni* N-glycosylation as a simple method to determine whether they were N-glycosylated. The *C. jejuni* N-glycosylation system is able to glycosylate N-glycoproteins when reconstituted in *E. coli* (Wacker *et al*., 2002). Western blotting confirmed that both Hp0114 and HgpA were glycosylated with the *C. jejuni* heptasaccharide glycan (Chapter 4). This N-glycosylation occurred in the presence of either *C. jejuni* PglB or *H. pullorum* PglB1 but not *H. pullorum* PglB2 suggesting that *H. pullorum* PglB2 has different specificities with regards to glycan structure and/or protein targets. Another interesting observation described in Chapter 4 was the glycosylation of the *H. pullorum* glycoprotein, HgpA, at a single site, by the *H. canadensis* PglB1, but not PglB2, with the *C. jejuni* heptasaccharide. This is the first evidence that *H. canadensis* PglB1is a functional OST that can transfer oligosaccharide to a target glycoprotein. Now it is evident that *H. canadensis* PglB can transfer an N-linked glycan structure to a protein, future experiments beyond the scope of this thesis may involve characterisation of the *H. canadensis* N-linked glycan structure.

Having verified that the Hp0114 and HgpA proteins were glycosylated in the model *E. coli* system, their glycosylation in *H. pullorum* with the previously characterized pentasaccharide glycan was confirmed (Chapter 5) and predicted N-glycosylation sites were investigated.

The *H. pullorum* HgpA glycoprotein has a single predicted N-glycosylation site ($_{51}$ENNDT$_{55}$) and results from Chapter 4 indicated that as for the *C. jejuni* PglB, the asparagine and threonine residues were essential for modification in *E. coli* with the *C. jejuni* heptasaccharide. *C. jejuni* PglB and *H. pullorum* PglB1 glycosylation was more efficient when the target sequon contained a glutamic acid residue (E51) but some HgpA protein was still glycosylated in the absence of this amino acid. These results suggest that the *C. jejuni* and *H. pullorum* PglB OSTs have similar preferences with regards to the target amino acid sequon.

The Hp0114 glycoprotein has two D/E-X-N-X-S/T glycosylation sequons, $_{85}$ENNAT$_{89}$ and $_{169}$ENNGS$_{173}$ and one N-X-S/T sequon, $_{46}$NLS$_{48}$ and interestingly, although *H. pullorum* PglB1 and *C. jejuni* PglB could glycosylate both sites when produced in *E. coli* with the *C. jejuni* heptasaccharide, they had different preferences for the site of N-glycosylation (Chapter 4). The *H. pullorum* PglB1 was more efficient at glycosylation of the N171 site but both *C. jejuni* PglB and *H. pullorum* PglB1 could efficiently transfer the heptasaccharide glycan to the N87 site. These results also suggest that the Hp0114 protein was glycosylated with two glycans at the $_{85}$ENNAT$_{89}$ sequon but only one at the $_{169}$ENNGS$_{173}$ sequon. This may be due to the presence of two asparagine residues (N86 and N87) both with acidic residues at N-2 positions in this sequon ($_{84}$DENNAT$_{89}$) but further experiments involving mutation of the N86 asparagine residue would be required to confirm this. In contrast to Hp0114 produced in these *E. coli* strains, only the N87 asparagine residue of Hp0114 was glycosylated in *H. pullorum* (Chapter 5). There may be a limited supply of lipid-linked oligosaccharide in *H. pullorum* and under these conditions a single amino acid sequon is preferred. In the *E. coli* system, expression of glycosyltransferase-encoding genes may be at a higher level, resulting in increased biosynthesis of lipid-linked oligosaccharide and more efficient glycosylation of each Hp0114 N-glycosylation sequon. Together these results suggest differences in N-glycosylation efficiency between *E. coli* strains producing the *C. jejuni* heptasaccharide glycan and *H. pullorum* strains producing the characterized pentasaccharide glycan.

8.4 The role of Glycosyltransferase and Biosynthetic Enzymes in *H. pullorum* N-linked Glycan Assembly

Analysis of the *H. pullorum* glycoprotein HgpA produced in *H. pullorum* wild type and various mutant strains by western blotting, intact mass analysis and tandem MALDI Mass Spectrometry enabled the role of a number of gene products in N-glycan biosynthesis to be investigated.

The genes encoding *H. pullorum* PglC and PglA are found as a single cluster on the *H. pullorum* chromosome (Figure 6.1). PglC and PglA from *C. jejuni* are involved in the transfer of the first two sugars (bac and GalNAc respectively) that make up the *C. jejuni* N-linked heptasaccharide glycan (Linton *et al*., 2005). HgpA from the *H. pullorum pgl*C::Kn mutant had an intact mass consistent with unmodified protein suggesting that *H. pullorum* PglC is required for transfer of the first sugar of the N-glycan, the 203 Da HexNAc. Mass Spectrometry results based on HgpA derived from the *H. pullorum pgl*A::Kn mutant indicated glycosylation with a single sugar therefore PglA is predicted to transfer the 216 Da residue to create the lipid-linked disaccharide. Unexpectedly, the single residue was not 203 Da but rather 228 Da although there is no 228 Da sugar in the previously characterized *H. pullorum* pentasaccharide. It seems that either a 228 Da residue was transferred to the lipid carrier instead of a 203 Da HexNAc residue or the HexNAc residue was converted to the 228 Da residue in the absence of the 216 Da sugar. A sugar of 228 Da in mass is consistent with a trideoxydiacetamidohexose, which in *C. jejuni* is bac. *H. pullorum* contains orthologues of the *C. jejuni pgl*DEF genes that are involved in the biosynthesis of bac therefore it is possible that in the *H. pullorum pgl*A::Kn mutant the 203 Da HexNAc was converted to a bac by PglDEF enzymes in the absence of the 216 Da sugar.

Results from Chapter 6 also suggest a role for the *H. pullorum pgl*J from the *pgl*KILJ gene cluster in pentasaccharide biosynthesis. Mass Spectrometry analysis of HgpA derived from the *H. pullorum pgl*J::Kn mutant revealed that this glycoprotein was modified with a disaccharide structure containing a reducing end 203 Da HexNAc and a 216 Da sugar. *H. pullorum* PglJ was therefore predicted to transfer the third

217 Da sugar of the *H. pullorum* pentasaccharide glycan. In *C. jejuni*, PglJ is also required for the transfer of the third sugar (GalNAc) to the lipid linked disaccharide (Linton *et al*., 2005). In *H. pullorum* it is possible that PglJ also transfers the fourth sugar (a second 217 Da sugar) or the fourth sugar could be transferred by a predicted glycosyltransferase such as PglL or PglI. Difficulties in the insertional mutagenesis of PglI and PglL suggest these genes may be involved in a separate essential glycosylation system, possibly involving PglB2. PglK is a putative ABC transporter and may be involved in flipping the lipid linked glycan from the cytoplasm to the periplasm but further work will be required to confirm this.

The *H. pullorum* PglH has 30% sequence identity to the *C. jejuni* PglH and the gene encoding this predicted glycosyltransferase is located within a cluster on the *H. pullorum* chromosome containing *pgl*DEF genes. The PglH glycosyltransferase in *C. jejuni* is responsible for the addition of the three terminal 203 Da GalNAc residues of the *C. jejuni* lipid linked heptasaccharide. Results indicate that the N-glycan of HgpA derived from the *H. pullorum pgl*H::Kn mutant also lacks a terminal HexNAc residue suggesting that PglH transfers the fifth and final HexNAc to complete the *H. pullorum* pentasaccharide glycan.

Based on results presented in this thesis, a model was proposed for the assembly of the *H. pullorum* pentasaccharide glycan on the membrane-bound lipid carrier (Figure 8.1). In this model, *H. pullorum* PglC transfers the reducing end HexNAc sugar, PglA transfers the 216 Da residue, PglJ is involved in transfer of the third, and possibly also the fourth sugar which are 217 Da and finally, the *H. pullorum* PglH transfers the non reducing end HexNAc residue to complete the *H. pullorum* lipid linked pentasaccharide which is transferred to protein by PglB1.

HexNAc residues are commonly found in bacterial N-linked glycans but the unusual 216 and 217 Da sugars that also make up the *H. pullorum* pentasaccharide glycan are less common and a potential pathway for their biosynthesis was explored in this thesis. A 216 Da residue was recently identified in the LPS of *P. aeruginosa* O6 and structurally identified as a GalNAcAN (King *et al*., 2010). In *P. aeruginosa* O6, an enzyme WbpO catalyzes the conversion of GlcNAc to the 217 Da sugar GlcNAcA,

which is then converted to GalNAcA by WbpP and then to GalNAcAN by WbpS (Figure 7.1 B). As the *H. pullorum* N-linked pentasaccharide glycan also contains 217 Da and 216 Da residues, a similar mechanism was proposed for the biosynthesis of these sugars from a HexNAc precursor.

The *H. pullorum* genome contains orthologues of the *wbp*O gene, located upstream of the *pgl*KILJ cluster, and the *wbp*S gene, located directly downstream of *pgl*B2. Results indicate that HgpA derived from the *H. pullorum wbp*S::Kn mutant is modified with a disaccharide glycan containing a reducing end HexNAc and a 217 Da sugar. The absence of a 216 Da sugar in this glycan suggests that WbpS is required its biosynthesis and when the 216 Da sugar was unavailable a 217 Da sugar was transferred instead but the glycan structure was not extended further to create the pentasaccharide likely due to glycosyltransferases specificities. Intact mass analysis and tandem MALDI Mass Spectrometry of HgpA derived from the *H. pullorum wbp*O::Kn mutant results were in agreement that the HgpA protein was glycosylated with a 403 Da glycan. Results indicated a structure consisting of a reducing end 228 Da residue and a non-reducing end 175 Da residue which could not be identified. These analyses therefore indicate that WbpO is essential for assembly of the complete pentasaccharide N-glycan and the 217 and 216 Da sugars were no longer observed suggesting that WbpO is required for biosynthesis of the 217 Da residue and therefore also the 216 Da residue which is downstream in the pathway.

In summary, data from this thesis provides increased understanding of the *H. pullorum* N-glycosylation system and provides information on the involvement of glycosyltransferases and sugar biosynthesis enzyme encoding genes found at four separate locations on the *H. pullorum* chromosome.

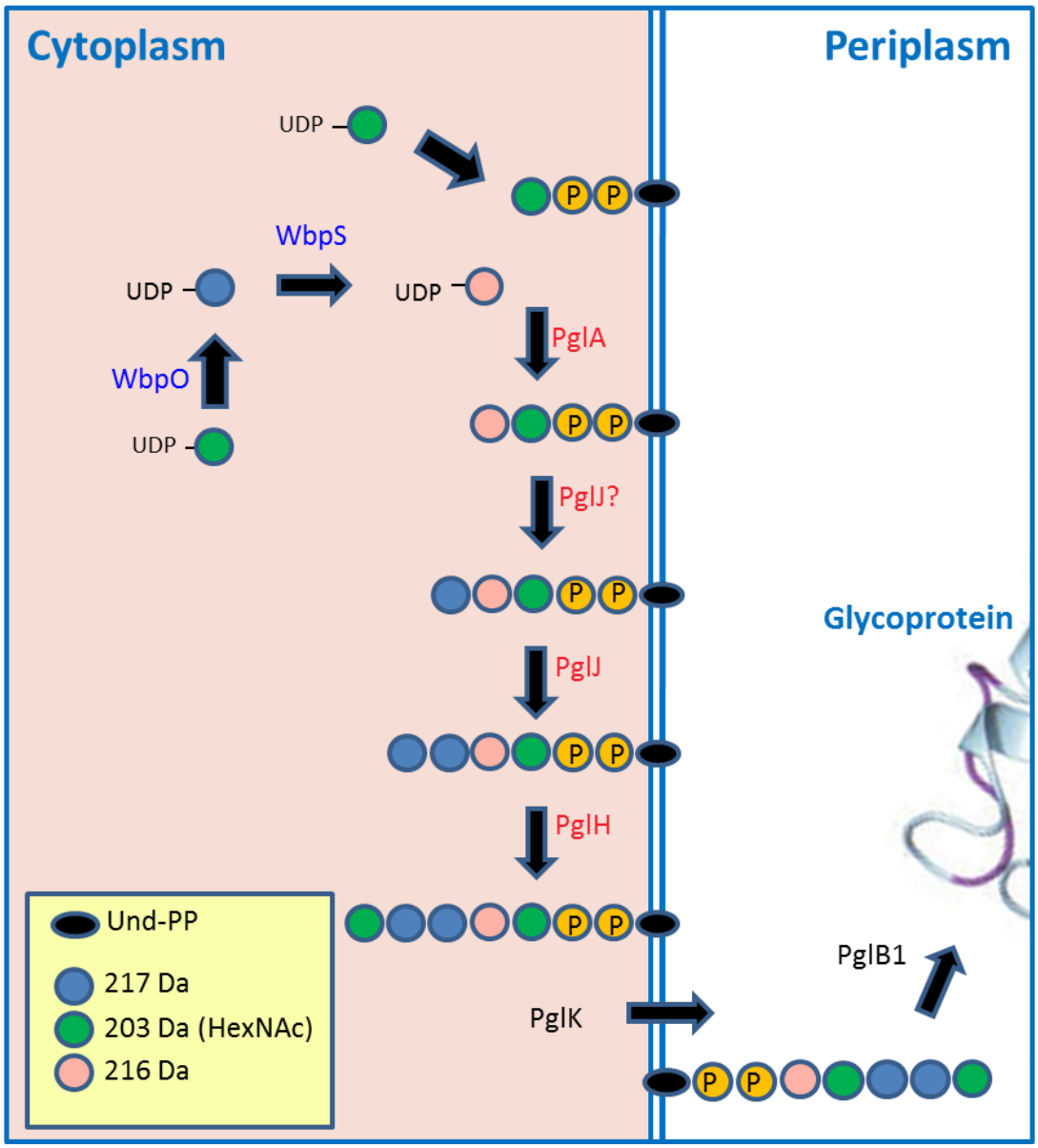
8.5 Conclusions

The *H. pullorum* N-linked pentasaccharide glycan was previously characterized and contains HexNAc residues and unusual 217 and 216 Da sugars (Jervis *et al.*, 2010). Prior to this study, no *H. pullorum* N-linked glycoproteins had been identified and the mechanism for the assembly of the *H. pullorum* PglB1-dependent N-linked pentasaccharide glycan was completely uncharacterized. In this thesis, bioinformatic analysis was used to identify putative *H. pullorum* N-glycoproteins and modification of two of these proteins was confirmed in *E. coli* producing the *C. jejuni* lipid linked heptasaccharide and either *C. jejuni* PglB or *H. pullorum* PglB1.

Mass Spectrometry techniques based on the *H. pullorum* glycoprotein, HgpA, allowed the identification of N-glycan structure in *H. pullorum* wild type and putative protein glycosylation gene mutants. Based on these data, a model for the assembly of the *H. pullorum* N-linked pentasaccharide glycan was proposed (Figure 8.1). In this model, *H. pullorum* PglC transfers the reducing end HexNAc residue to a lipid carrier and PglA transfers the second sugar, a 216 Da residue. PglJ then transfers at least one of the 217 Da sugar residues and PglH completes the pentasaccharide glycan by transferring the terminal HexNAc residue. The unusual 216 and 217 Da sugars of the *H. pullorum* pentasaccharide are synthesized sequentially from HexNAc by the action of WbpO and WbpS.

The *H. pullorum* N-linked protein glycosylation system described in this thesis is only the second such bacterial system to be characterized. This *H. pullorum* PglB1 dependent N-linked protein glycosylation system has similarities to *C. jejuni* N-glycosylation with regards to target protein sequons and glycan reducing end sugar structural requirements but appears to be more complex, perhaps due to the presence of a second PglB2 dependent N-glycosylation system that is likely to be essential. Further investigations beyond the scope of this thesis will focus on identifying a second, PglB2-dependent N-glycosylation system and investigating the role of the putative glycosyltransferases, PglI and PglL.

Figure 8.1 Proposed mechanism for the assembly of the *H. pullorum* lipid linked pentasaccharide. A pentasaccharide is assembled on a lipid carrier at the cytoplasmic side of the inner bacterial membrane by the action of glycosyltransferases as indicated, flipped into the periplasm by the predicted ABC transporter, PglK and transferred to a *H. pullorum* glycoprotein by the action of the OST, PglB. Sugar residues are represented as blue, green or pink circles and phosphate groups as yellow circles. Black ovals represent the membrane bound lipid carrier Und-PP.



References

ABU-QARN, M. AND EICHLER, J. (2006). Protein N-glycosylation in Archaea: defining *Haloferax volcanii* genes involved in S-layer glycoprotein glycosylation. *Molecular Microbiology,* **61:** 511-25.

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

ABU-QARN, M., YURIST-DOUTSCH, S., GIORDANO, A., TRAUNER, A., MORRIS, H. R., HITCHEN, P., MEDALIA, O., DELL, A. AND 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. *Journal of Molecular Biology,* **374:** 1224-36.

ALAIMO, C., CATREIN, I., MORF, L., MAROLDA, C. L., CALLEWAERT, N., VALVANO, M. A., FELDMAN, M. F. AND AEBI, M. (2006). Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. *European Molecular Biology Organisation Journal,* **25:** 967-76.

ANDERSON, P. (1983). Antibody responses to *Haemophilus influenzae* type b and diphtheria toxin induced by conjugates of oligosaccharides of the type b capsule with the nontoxic protein CRM197. *Infection and Immunity,* **39:** 233-8.

ARORA, S. K., BANGERA, M., LORY, S. AND RAMPHAL, R. (2001). A genomic island in *Pseudomonas aeruginosa* carries the determinants of flagellin glycosylation. *Proceedings of the National Academy of Science U S A,* **98:** 9342-7.

ATABAY, H.I., CORRY, J.E. AND ON, S.L. (1998). (1998). Identification of unusual Campylobacter-like isolates from poultry products as *Helicobacter pullorum. Journal of Applied Microbiology*, **84**: 1017-24.

BAAR, C., EPPINGER, M., RADDATZ, G., SIMON, J., LANZ, C., KLIMMEK, O., NANDAKUMAR, R., GROSS, R., ROSINUS, A., KELLER, H., JAGTAP, P., LINKE, B., MEYER, F., LEDERER, H. AND SCHUSTER, S. C. (2003). Complete genome sequence and analysis of *Wolinella succinogenes*. *Proceedings of the National Acadademy of Science U S A,* **100:** 11690-5.

BACON, D. J., ALM, R. A., BURR, D. H., HU, L., KOPECKO, D. J., EWING, C. P., TRUST, T. J. AND GUERRY, P. (2000). Involvement of a plasmid in virulence of *Campylobacter jejuni* 81-176. *Infection and Immunity,* **68:** 4384-90.

BACON, D. J., ALM, R. A., HU, L., HICKEY, T. E., EWING, C. P., BATCHELOR, R. A., TRUST, T. J. AND GUERRY, P. (2002). DNA sequence and mutational analyses of the pVir plasmid of *Campylobacter jejuni* 81-176. *Infection and Immunity,* **70:** 6242-50.

BENZ, I. AND SCHMIDT, M. A. (2002). Never say never again: protein glycosylation in pathogenic bacteria. *Molecular Microbiology,* **45:** 267-76.

BLACK, R. E., LEVINE, M. M., CLEMENTS, M. L., HUGHES, T. P. AND BLASER, M. J. (1988). Experimental *Campylobacter jejuni* infection in humans. *Journal of Infectious Diseases,* **157:** 472-9.

BURDA, P. AND AEBI, M. (1999). The dolichol pathway of N-linked glycosylation. *Biochimica et Biophysica Acta,* **1426:** 239-57.

BURNENS, A.P., STANLEY, J., MORGENSTERN, R. AND NICOLET, J. (1994). Gastroenteritis associated with *Helicobacter pullorum*. *Lancet*, **344(8936):** 1569-70

CALO, D., GUAN, Z., NAPARSTEK, S. AND 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. *Molecular Microbiology,* **81:** 1166-77.

CASTRIC, P. (1995). *pil*O, a gene required for glycosylation of *Pseudomonas aeruginosa* 1244 pilin. *Microbiology,* **141 (Pt 5):** 1247-54.

CEELEN, L., DECOSTERE, A., VERSCHRAEGEN, G., DUCATELLE, R. AND HAESEBROUCK, F. (2005). Prevalence of *Helicobacter pullorum* among patients with gastrointestinal disease and clinically healthy persons. *Journal of Clinical Microbiology,* **43:** 2984-6.

CHABAN, B., VOISIN, S., KELLY, J., LOGAN, S. M. AND 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. *Molecular Microbiology,* **61:** 259-68.

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*. *Journal of Bacteriology,* **191(1):** 187-195.

CHEN, M. M., GLOVER, K. J. AND IMPERIALI, B. (2007). From peptide to protein: comparative analysis of the substrate specificity of N-linked glycosylation in *C. jejuni*. *Biochemistry,* **46:** 5579-85.

CHOI, K. J., GRASS, S., PAEK, S., ST GEME, J. W., 3RD AND YEO, H. J. (2010). The *Actinobacillus pleuropneumoniae* HMW1C-like

glycosyltransferase mediates N-linked glycosylation of the *Haemophilus influenzae* HMW1 adhesin. *PLoS One,* **5:** e15888.

CREUZENET, C., BELANGER, M., WAKARCHUK, W. W. AND LAM, J. S. (2000). Expression, purification, and biochemical characterization of WbpP, a new UDP-GlcNAc C4 epimerase from *Pseudomonas aeruginosa* serotype O6. *Journal of Biological Chemistry,* **275:** 19060-7.

DELL, A., GALADARI, A., SASTRE, F. AND HITCHEN, P. (2010). Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *International Journal of Microbiology,* **2010:** 148178.

EGGE-JACOBSEN, W., SALOMONSSON, E. N., AAS, F. E., FORSLUND, A. L., WINTHER-LARSEN, H. C., MAIER, J., MACELLARO, A., KUOPPA, K., OYSTON, P. C., TITBALL, R. W., THOMAS, R. M., FORSBERG, A., PRIOR, J. L. AND KOOMEY, M. (2011). O-linked glycosylation of the PilA pilin protein of *Francisella tularensis*: identification of the endogenous protein-targeting oligosaccharyltransferase and characterization of the native oligosaccharide. *Journal of Bacteriology,* **193:** 5487-97.

ERICKSON, P. R. AND HERZBERG, M. C. (1993). Evidence for the covalent linkage of carbohydrate polymers to a glycoprotein from *Streptococcus sanguis*. *Journal of Biological Chemistry,* **268:** 23780-3.

FALT, I. C., SCHWEDA, E. K., KLEE, S., SINGH, M., FLODERUS, E., TIMMIS, K. N. AND LINDBERG, A. A. (1995). Expression of *Shigella dysenteriae* serotype 1 O-antigenic polysaccharide by *Shigella flexneri aro*D vaccine candidates and different *S. flexneri* serotypes. *Journal of Bacteriology,* **177:** 5310-5.

FELDMAN, M. F., WACKER, M., HERNANDEZ, M., HITCHEN, P. G., MAROLDA, C. L., KOWARIK, M., MORRIS, H. R., DELL, A., VALVANO, M. A. AND AEBI, M. (2005). Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. *Proceedings of the National Academy of Science U S A,* **102:** 3016-21.

FERRERO, R. L., CUSSAC, V., COURCOUX, P. AND LABIGNE, A. (1992). Construction of isogenic urease-negative mutants of *Helicobacter pylori* by allelic exchange. *Journal of Bacteriology,* **174:** 4212-7.

FOX, J. G., CHIEN, C. C., DEWHIRST, F. E., PASTER, B. J., SHEN, Z., MELITO, P. L., WOODWARD, D. L. AND RODGERS, F. G. (2000). *Helicobacter canadensis* sp. nov. isolated from humans with diarrhea as an example of an emerging pathogen. *Journal of Clinical Microbiology,* **38:** 2546-9.

GEBHART, C., IELMINI, M. V., REIZ, B., PRICE, N. L., AAS, F. E., KOOMEY, M. AND FELDMAN, M. F. (2012). Characterization of exogenous bacterial oligosaccharyltransferases in *Escherichia coli* reveals the potential for O-

linked protein glycosylation in *Vibrio cholerae* and *Burkholderia thailandensis*. *Glycobiology*.

GLOVER, K. J., WEERAPANA, E., CHEN, M. M. AND IMPERIALI, B. (2006). Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the *Campylobacter jejuni* N-linked glycosylation pathway. *Biochemistry,* **45:** 5343-50.

GLOVER, K. J., WEERAPANA, E. AND IMPERIALI, B. (2005). *In vitro* assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked glycosylation. *Proceedings of the National Academy of Science U S A,* **102:** 14255-9.

GRASS, S., BUSCHER, A. Z., SWORDS, W. E., APICELLA, M. A., BARENKAMP, S. J., OZCHLEWSKI, N. AND ST GEME, J. W., 3RD (2003). The *Haemophilus influenzae* HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis. *Molecular Microbiolpgy,* **48:** 737-51.

GRASS, S., LICHTI, C. F., TOWNSEND, R. R., GROSS, J. AND ST GEME, J. W., 3RD (2010). The *Haemophilus influenzae* HMW1C protein is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin. *Public Library of Science Pathogens,* **6:** e1000919.

GROSS, J., GRASS, S., DAVIS, A. E., GILMORE-ERDMANN, P., TOWNSEND, R. R. AND ST GEME, J. W., 3RD (2008). The *Haemophilus influenzae* HMW1 adhesin is a glycoprotein with an unusual N-linked carbohydrate modification. *Journal of Biological Chemistry,* **283:** 26010-5.

GUAN, Z., NAPARSTEK, S., KAMINSKI, L., KONRAD, Z. AND 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. *Molecular Microbiology,* **78:** 1294-303.

HAAS, R., MEYER, T. F. AND VAN PUTTEN, J. P. (1993). Aflagellated mutants of *Helicobacter pylori* generated by genetic transformation of naturally competent strains using transposon shuttle mutagenesis. *Molecular Microbiology,* **8:** 753-60.

IGURA, M., MAITA, N., OBITA, T., KAMISHIKIRYO, J., MAENAKA, K. AND KOHDA, D. (2007). Purification, crystallization and preliminary X-ray diffraction studies of the soluble domain of the oligosaccharyltransferase STT3 subunit from the thermophilic archaeon *Pyrococcus furiosus*. *Acta Crystallographica Section F Structural Biology and Crystalization Communications,* **63:** 798-801.

IHSSEN, J., KOWARIK, M., DILETTOSO, S., TANNER, C., WACKER, M. AND THONY-MEYER, L. (2010). Production of glycoprotein vaccines in *Escherichia coli. Microbial Cell Factories,* **9:** 61.

JERVIS, A. J., BUTLER, J. A., LAWSON, A. J., LANGDON, R., WREN, B. W. AND LINTON, D. (2012). Characterisation of the structurally diverse N-linked glycans of Campylobacter species. *Journal of Bacteriology*.

JERVIS, A. J., LANGDON, R., HITCHEN, P., LAWSON, A. J., WOOD, A., FOTHERGILL, J. L., MORRIS, H. R., DELL, A., WREN, B. AND LINTON, D. (2010). Characterization of N-linked protein glycosylation in *Helicobacter pullorum*. *Journal of Bacteriology,* **192:** 5228-36.

JOSENHANS, C., VOSSEBEIN, L., FRIEDRICH, S. AND SUERBAUM, S. (2002). The neuA/flmD gene cluster of *Helicobacter pylori* is involved in flagellar biosynthesis and flagellin glycosylation. *Federation of European Microbiological Societies Letter,* **210:** 165-72.

KAKUDA, T. AND DIRITA, V. J. (2006). Cj1496c encodes a *Campylobacter jejuni* glycoprotein that influences invasion of human epithelial cells and colonization of the chick gastrointestinal tract. *Infection and Immunity,* **74:** 4715-23.

KAMINSKI, L., ABU-QARN, M., GUAN, Z., NAPARSTEK, S., VENTURA, V. V., RAETZ, C. R., HITCHEN, P. G., DELL, A. AND EICHLER, J. (2010). AglJ adds the first sugar of the N-linked pentasaccharide decorating the *Haloferax volcanii* S-layer glycoprotein. *Journal of Bacteriology,* **192:** 5572-9.

KARLYSHEV, A. V., EVEREST, P., LINTON, D., CAWTHRAW, S., NEWELL, D. G. AND WREN, B. W. (2004). The *Campylobacter jejuni* general glycosylation system is important for attachment to human epithelial cells and in the colonization of chicks. *Microbiology*, **150:** 1957-64.

KELLY, J., LOGAN, S. M., JARRELL, K. F., VANDYKE, D. J. AND VINOGRADOV, E. (2009). A novel N-linked flagellar glycan from *Methanococcus maripaludis*. *Carbohydrate Research,* **344:** 648-53.

KING, J. D., VINOGRADOV, E., TRAN, V. AND LAM, J. S. (2010). Biosynthesis of uronamide sugars in *Pseudomonas aeruginosa* O6 and *Escherichia coli* O121 O antigens. *Environmental Microbiology,* **12:** 1531-44.

KOHDA, D., YAMADA, M., IGURA, M., KAMISHIKIRYO, J. AND MAENAKA, K. (2007). New oligosaccharyltransferase assay method. *Glycobiology,* **17:** 1175-82.

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. AND AEBI, M. (2006). Definition of the bacterial N-glycosylation site consensus sequence. *European Molecular Biology Organization Journal,* **25:** 1957-66.

KRACHLER, A. M., SHARMA, A., CAULDWELL, A., PAPADAKOS, G. AND KLEANTHOUS, C. (2010). TolA modulates the oligomeric status of YbgF in the bacterial periplasm. *J ournal of Molecular Biology,* **403:** 270-85.

LABIGNE-ROUSSEL, A., COURCOUX, P. AND TOMPKINS, L. (1988). Gene disruption and replacement as a feasible approach for mutagenesis of *Campylobacter jejuni*. *Journal of Bacteriology,* **170:** 1704-8.

LABIGNE-ROUSSEL, A., HAREL, J. AND TOMPKINS, L. (1987). Gene transfer from *Escherichia coli* to Campylobacter species: development of shuttle vectors for genetic analysis of *Campylobacter jejuni*. *Journal of Bacteriology,* **169:** 5320-3.

LARSEN, J. C., SZYMANSKI, C. AND GUERRY, P. (2004). N-linked protein glycosylation is required for full competence in *Campylobacter jejuni* 81-176. *Journal of Bacteriology,* **186:** 6508-14.

LECLERC, G., WANG, S. P. AND ELY, B. (1998). A new class of *Caulobacter crescentus* flagellar genes. *Journal of Bacteriology,* **180:** 5010-9.

LENNARZ, W. J. (2007). Studies on oligosaccharyl transferase in yeast. *Acta Biochimica Polonica,* **54:** 673-7.

LINTON, D., ALLAN, E., KARLYSHEV, A. V., CRONSHAW, A. D. AND WREN, B. W. (2002). Identification of N-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in *Campylobacter jejuni*. *Molecular Microbiology,* **43:** 497-508.

LINTON, D., DORRELL, N., HITCHEN, P. G., AMBER, S., KARLYSHEV, A. V., MORRIS, H. R., DELL, A., VALVANO, M. A., AEBI, M. AND WREN, B. W. (2005). Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway. *Molecular Microbiology*, **55:** 1695-703.

LIZAK, C., GERBER, S., NUMAO, S., AEBI, M. AND LOCHER, K. P. (2011). X-ray structure of a bacterial oligosaccharyltransferase. *Nature,* **474:** 350-5.

MAGIDOVICH, H., YURIST-DOUTSCH, S., KONRAD, Z., VENTURA, V. V., DELL, A., HITCHEN, P. G. AND EICHLER, J. (2010). AglP is a S-adenosyl-L-methionine-dependent methyltransferase that participates in the N-glycosylation pathway of *Haloferax volcanii. Molecular Microbiology,* **76: 190-9.**

MARSHALL, B. J. AND WARREN, J. R. (1984). Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet,* **1:** 1311-5.

MELITO, P. L., MUNRO, C., CHIPMAN, P. R., WOODWARD, D. L., BOOTH, T. F. AND RODGERS, F. G. (2001). *Helicobacter winghamensis* sp. nov., a novel Helicobacter sp. isolated from patients with gastroenteritis. *Journal of Clinical Microbiology,* **39:** 2412-7.

MESCHER, M. F. AND STROMINGER, J. L. (1975). Bacitracin induces sphere formation in Halobacterium species which lack a wall peptidoglycan. *J ournal of General Microbiology,* **89:** 375-8.

MESCHER, M. F. AND STROMINGER, J. L. (1976). Purification and characterization of a prokaryotic glucoprotein from the cell envelope of *Halobacterium salinarium*. *Journal of Biological Chemistry,* **251:** 2005-14.

MILLER, J. F., DOWER, W. J. AND TOMPKINS, L. S. (1988). High-voltage electroporation of bacteria: genetic transformation of *Campylobacter jejuni* with plasmid DNA. *Proceedings of the National Academy of Science U S A,* **85:** 856-60.

NITA-LAZAR, M., WACKER, M., SCHEGG, B., AMBER, S. AND AEBI, M. (2005). The N-X-S/T consensus sequence is required but not sufficient for bacterial N-linked protein glycosylation. *Glycobiology,* **15:** 361-7.

NOTHAFT, H.. AND SZYMANSKI, C.M. (2010). Protein glycosylation in bacteria: sweeter than ever. *Nature Reviews Microbiology,* **8:** 765-778.

OLIVIER, N. B., CHEN, M. M., BEHR, J. R. AND IMPERIALI, B. (2006). In vitro biosynthesis of UDP-N,N'-diacetylbacillosamine by enzymes of the *Campylobacter jejuni* general protein glycosylation system. *Biochemistry,* **45:** 13659-69.

PEI, Z. H., ELLISON, R. T., 3RD AND BLASER, M. J. (1991). Identification, purification, and characterization of major antigenic proteins of *Campylobacter jejuni*. *Journal of Biological Chemistry,* **266:** 16363-9.

POWER, P. M. AND JENNINGS, M. P. 2003. The genetics of glycosylation in Gram-negative bacteria. *Federation of European Microbiological Societies Letters,* **218:** 211-22.

SCHAGGER, H. (2006). Tricine-SDS-PAGE. *Nature Protocols,* **1:** 16-22.

SCHWARZ, F. AND AEBI, M. 2011. Mechanisms and principles of N-linked protein glycosylation. *Current Opinions in Structural Biology,* **21:** 576-82.

SCHWARZ, F., FAN, Y. Y., SCHUBERT, M. AND AEBI, M. (2011a). Cytoplasmic N-glycosyltransferase of *Actinobacillus pleuropneumoniae* is an inverting enzyme and recognizes the NX(S/T) consensus sequence. *Journal of Biological Chemistry,* **286:** 35267-74.

SCHWARZ, F., LIZAK, C., FAN, Y. Y., FLEURKENS, S., KOWARIK, M. AND AEBI, M. (2011b). Relaxed acceptor site specificity of bacterial oligosaccharyltransferase *in vivo*. *Glycobiology,* **21:** 45-54.

SCOTT, N. E., PARKER, B. L., CONNOLLY, A. M., PAULECH, J., EDWARDS, A. V., CROSSETT, B., FALCONER, L., KOLARICH, D., DJORDJEVIC, S.

P., HOJRUP, P., PACKER, N. H., LARSEN, M. R. AND CORDWELL, S. J. (2011). Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of *Campylobacter jejuni*. *Molecular Cell Proteomics,* **10:** M000031-MCP201.

SHAMS-ELDIN, H., CHABAN, B., NIEHUS, S., SCHWARZ, R. T. AND JARRELL, K. F. (2008). Identification of the archaeal *alg*7 gene homolog (encoding N-acetylglucosamine-1-phosphate transferase) of the N-linked glycosylation system by cross-domain complementation in *Saccharomyces cerevisiae*. *Journal of Bacteriology,* **190:** 2217-20.

SILBERSTEIN, S. AND GILMORE, R. 1996. Biochemistry, molecular biology, and genetics of the oligosaccharyltransferase. The Journal of the *Federation of Americal Societies for Experimental Biology Journal,* **10:** 849-58.

STANLEY, J., LINTON, D., BURNENS, A. P., DEWHIRST, F. E., ON, S. L., PORTER, A., OWEN, R. J. AND COSTAS, M. (1994). *Helicobacter pullorum* sp. nov.-genotype and phenotype of a new species isolated from poultry and from human patients with gastroenteritis. *Microbiology,* **140 (Part 12):** 3441-9.

STEINBRUECKNER, B., HAERTER, G., PELZ, K., WEINER, S., RUMP, J.A., DEISSLER, W., BERESWILL, S. AND KIST, M. (1997). Isolation of *Helicobacter pullorum* from patients with enteritis. Scandinavian Journal of Infectious Diseases, **29(3):**315-8.

STIMSON, E., VIRJI, M., MAKEPEACE, K., DELL, A., MORRIS, H. R., PAYNE, G., SAUNDERS, J. R., JENNINGS, M. P., BARKER, S., PANICO, M. AND ET AL. (1995). Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose. *Molecular Microbiology,* **17:** 1201-14.

SUMPER, M., BERG, E., MENGELE, R. AND STROBEL, I. (1990). Primary structure and glycosylation of the S-layer protein of *Haloferax volcanii*. *Journal of Bacteriology,* **172:** 7111-8.

SZYMANSKI, C. M., BURR, D. H. AND GUERRY, P. (2002). Campylobacter protein glycosylation affects host cell interactions. *Infection and Immunity,* **70:** 2242-4.

SZYMANSKI, C. M., LOGAN, S. M., LINTON, D. AND WREN, B. W. (2003). Campylobacter--a tale of two protein glycosylation systems. *Trends in Microbiology,* **11:** 233-8.

SZYMANSKI, C. M. AND WREN, B. W. (2005). Protein glycosylation in bacterial mucosal pathogens. *Nature Reviews Microbiology,* **3:** 225-37.

SZYMANSKI, C. M., YAO, R., EWING, C. P., TRUST, T. J. AND GUERRY, P. (1999). Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Molecular Microbiology,* **32:** 1022-30.

TABEI, S. M., HITCHEN, P. G., DAY-WILLIAMS, M. J., MERINO, S., VART, R., PANG, P. C., HORSBURGH, G. J., VICHES, S., WILHELMS, M., TOMAS, J. M., DELL, A. AND SHAW, J. G. (2009). An *Aeromonas caviae* genomic island is required for both O-antigen lipopolysaccharide biosynthesis and flagellin glycosylation. *Journal of Bacteriology,* **191:** 2851-63.

TAYLOR, D. N., TROFA, A. C., SADOFF, J., CHU, C., BRYLA, D., SHILOACH, J., COHEN, D., ASHKENAZI, S., LERMAN, Y., EGAN, W. AND ET AL. (1993). Synthesis, characterization, and clinical evaluation of conjugate vaccines composed of the O-specific polysaccharides of *Shigella dysenteriae* type 1, *Shigella flexneri* type 2a, and *Shigella sonnei* (*Plesiomonas shigelloides*) bound to bacterial toxoids. *Infection and Immunology,* **61:** 3678-87.

THIBAULT, P., LOGAN, S. M., KELLY, J. F., BRISSON, J. R., EWING, C. P., TRUST, T. J. AND GUERRY, P. (2001). Identification of the carbohydrate moieties and glycosylation motifs in *Campylobacter jejuni* flagellin. *Journal of Biological Chemistry,* **276:** 34862-70.

TROUTMAN, J. M. AND IMPERIALI, B. (2009). *Campylobacter jejuni* PglH is a single active site processive polymerase that utilizes product inhibition to limit sequential glycosyl transfer reactions. *Biochemistry,* **48:** 2807-16.

UPRETI, R. K., KUMAR, M. AND SHANKAR, V. (2003). Bacterial glycoproteins: functions, biosynthesis and applications. *Proteomics,* **3:** 363-79.

VIANNEY, A., MULLER, M. M., CLAVEL, T., LAZZARONI, J. C., PORTALIER, R. AND WEBSTER, R. E. (1996). Characterization of the *tol-pal* region of *Escherichia coli* K-12: translational control of *tol*R expression by TolQ and identification of a new open reading frame downstream of *pal* encoding a periplasmic protein. *Journal of Bacteriology,* **178:** 4031-8.

VIRJI, M. (1997). Post-translational modifications of meningococcal pili. Identification of common substituents: glycans and alpha-glycerophosphate--a review. *Gene,* **192:** 141-7.

VOISIN, S., HOULISTON, R. S., KELLY, J., BRISSON, J. R., WATSON, D., BARDY, S. L., JARRELL, K. F. AND LOGAN, S. M. (2005). Identification and characterization of the unique N-linked glycan common to the flagellins and S-layer glycoprotein of *Methanococcus voltae*. *Journal of Biological Chemistry,* **280:** 16586-93.

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. AND AEBI, M. (2006). Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proceedings of the National Academy of Science 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. AND AEBI, M. (2002). N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science,* **298:** 1790-3.

WEERAPANA, E., GLOVER, K. J., CHEN, M. M. AND IMPERIALI, B. (2005). Investigating bacterial N-linked glycosylation: synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. *Journal of the American Chemical Society,* **127:** 13766-7.

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. AND SZYMANSKI, C. M. (2002). Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. *Journal of Biological Chemistry,* **277:** 42530-9.

YURIST-DOUTSCH, S., ABU-QARN, M., BATTAGLIA, F., MORRIS, H. R., HITCHEN, P. G., DELL, A. AND EICHLER, J. (2008a). *agl*F, *agl*G and *agl*I, novel members of a gene island involved in the N-glycosylation of the *Haloferax volcanii* S-layer glycoprotein. *Molecular Microbiology,* **69:** 1234-45.

YURIST-DOUTSCH, S., CHABAN, B., VANDYKE, D. J., JARRELL, K. F. AND EICHLER, J. (2008b). Sweet to the extreme: protein glycosylation in Archaea. *Molecular Microbiology,* **68:** 1079-84.

ZHAO, X., CREUZENET, C., BELANGER, M., EGBOSIMBA, E., LI, J. AND LAM, J. S. (2000). WbpO, a UDP-N-acetyl-D-galactosamine dehydrogenase from *Pseudomonas aeruginosa* serotype O6. *Journal of Biological Chemistry,* **275:** 33252-9.

Appendix I Primers used for PCR amplification. Primers were used for the amplification and site directed mutagenesis of Helicobacter genes and to confirm the correct construction of plasmids and knock-out mutations. All primer sequences are in the 5' to 3' orientation

Primer Number	Primer name	Primer sequence (5' to 3')
1	Pull23SF	GTC AGA TGC TGC AGA CCC G
2	Pull23SR	GCA GGT AGT CTT CCT GCG ACC
3	Pull23SinsF	GGT TAG TCG GGT CCT AAG AC
4	Pull23SinsR	CGT TAT AGT TAC GGC CGC CG
5	Pull23SoutF	GTG AGC GGA GTG AAA TAG AAC C
6	Pull23SoutR	GCT TAC ACC CCT ACC CTA TC
7	CatHindIIIF	AGC AAG CTT TTA TTT ATT CAG CCA GTC TTG TA
8	CatHindIIIR	AGC AAG CTT TTA TTT ATT CAG CCA GTC TTG TA
9	Cmfout	GCA TGG ACT AAT GCT TGA AAC CC
10	Cmrout	CTT GGA AAG GAA CAC CGC CG
11	KanHindIIIF	AGC TAA GCT TGA TAA ACC CAG CGA ACC ATT TG
12	KanHindIIIR	ATG CAA GCT TTT TAG ACA TCT AA
13	KanBamHIF	GGC ATT AAT GGA TCC CCG GGT AC
14	KanBamHIR	CGA TTA ATG GAT CCT CTA GAG
15	Kanfout	CAT CCT CCT CGT CTT GGT AGC
16	Kanrout	TTG CCT TCT GCG TCC GGT CG
17	EryHindIIIF	AAG CTT AGG CCG GCC AGT ATA AAA CC
18	EryHindIIIR	AAG CTT AGG CGC GCC ATC TAC
19	Eryfout	GCA GGA ATT GAC GAT TTA AAC
20	Eryrout	GCA CGA GCT CTG ATA AAT ATG
21	EryBamHIHindIIIR	CGA AAA GCT TGG ATC CTT AGG CGC GCC ACT TAC
22	PullpglB1RTPCRF	GTC ATC TTT TGT GGG ATA TTA GCC
23	PullpglB1RTPCRR	GGA ATT TCT TGC AAT GCT TTG GCT TC
24	PullpglB2RTPCRF	GGT TGG TTC TTT TTG GCT TTT TGG G
25	PullpglB2RTPCRR	GCA CCA ATA TCC CTA ACC CAT G
26	PullpglB1F	GGG TGT TTA TTG TCG GTG C
27	PullpglB1R	CCT ACT ATG GTA GAT GTT TCC
28	PullpglB2outF	GCA GTT CGC GGT ATG CTT CC
29	PullpglB2outR	GGA AGC ATA CCG CGA ACT GC
30	PullpglB2F	GCT TTT TTG GGC TAG TTT TGC
31	PullpglB2R	CAC AAA TGC GGG AAT ACG CT
32	pQEF	CCC GAA AAG TGC CAC CTG

33	pQER	GTT CTG AGG TCA TTA CTG G
34	pQE0114sphIF	TAA GCA TGC TGA AAT TGC AAA AAC CAA TTA TTG C
35	pQE0114BglIIR	AGT AGA TCT TTT TTT TAT TAA AAG ATT CCC AGC
36	pQEHgpASphIF	GAA GCA TGC TGA AAA TTT TAC TAC CAC TGC
37	pQEHgpABglIIR	ACA AGA TCT GAA GAG TTC AAT CTC TAA GCC
38	Pull0114BamHIF	GAA TGG ATC CGT TAC CTT TAA AAT GTT GCC
39	Pull0114HisBamHIR	TGA GGA TCC CTA GTG GTG GTG GTG GTG GTG TTT TTT TAT TAA AAG ATT CCC AGC
40	PullHgpABamHIF	TTT GGA TCC AAC TCT CTA AGT GTA TTA TTT G
41	PullHgpAHisBamHIR	CTC GGA TCC TTA GTG GTG GTG GTG GTG GTG GAA GAG TTC AAT CTC TAA GCC
42	PullasnBF	CTC TCC AAA CAC ACA TTA GAG G
43	PullasnBR	GGG CTT TTG GGA TTC TTG ATA TTT C
44	PullasnBoutF	AGG GGA TCC CCT GCG ATG AAT CTT GTT TTT GAG
45	PullasnBoutR	CGA CAT ATG CGC ACA AAA TAA TCA AAA AGC CC
46	PullasnBSDMF	CTA CCA AAA GGA TCC TAG CAA ATA TTA TC
47	PullasnBSDMR	GAT AAT ATT TGC TAG GAT CCT TTT GGT AG
48	0114N46SDMF	CGA AAG GCT TTT TCA ACT TTC TAA TAA AG
49	0114N46SDMR	CTT TAT TAG AAA GTT GAA AAA GCC TTT CG
50	0114N87SDMF	GGC TGA TGA AAA TCA AGC GAC AAT CTC AC
51	0114N87SDMR	GTG AGA TTG TCG CTT GAT TTT CAT CAG CC
52	0114N171SDMF	GAT TCG GTA GAA AAT CAA GGA AGT GCT CC
53	0114N171SDMR	GGA GCA CTT CCT TGA TTT TCT ACC GAA TC
54	HgpAE51SDMF	CCC AAC TCC CAA CGC AAA TAA TGA TAC C
55	HgpAE51SDMR	GGT ATC ATT ATT TGC GTT GGG AGT TGG G
56	HgpAN53SDMF	CTT TGG TAT CTT GAT TTT CGT TGG GAG
57	HgpAN53SDMR	CTC CCA ACG AAA ATC AAG ATA CCA AAG
58	HgpAT55SDMF	CCC AAC GAA AAT AAT GAT GGC AAA GAA ATA AG
59	HgpAT55SDMR	CTT ATT TCT TTG CCA TCA TTA TTT TCG TTG GG

Appendix II *H. pullorum* Strain list

Strain	Description
NCTC 12824^{T}	Wild type strain (Stanley *et al*., 2004)
NCTC 13153^{T}	Isolated from chicken carcasses in Bristol, UK (Atabay *et al*., 1998)
NCTC 13155^{T}	Isolated from human faeces, Germany (Steinbrueckner *et al*., 1997)
NCTC 12826^{T}	First reported human faecal isolate, Switzerland (Burnens *et al*., 1994)
23::E	NCTC 12824^{T} with a erythromycin resistance cassette insertion in a 23S rRNA gene
23::Kn	NCTC 12824^{T} with a kanamycin resistance cassette insertion in a 23S rRNA gene
*pgl*B1::Kn	NCTC 12824^{T} with a disrupted *pgl*B1 gene
*pgl*B2::Kn	NCTC 12824^{T} with a disrupted *pgl*B2 gene
23E::*hgp*A	NCTC 12824^{T} expressing a hexa histidine tagged *hgp*A gene
23E::0114	*H. pullorum* 12824 expressing a hexa histidine tagged *Hp*0114 gene
*hgp*A *pgl*B1::Kn	23E::*hgp*A *with a disrupted pgl*B1 *gene*
*hgp*A *pgl*B2::Kn	23E::*hgp*A *with a disrupted pgl*B2 *gene*
*hgp*A *pgl*A::Kn	23E::*hgp*A *with a disrupted pgl*A *gene*
*hgp*A *pgl*C::Kn	23E::*hgp*A *with a disrupted pgl*C *gene*
*hgp*A *pgl*J::Kn	23E::*hgp*A *with a disrupted pgl*J *gene*
*hgp*A *pgl*H::Kn	23E::*hgp*A *with a disrupted pgl*H *gene*
*hgp*A *wbp*O::Kn	23E::*hgp*A *with a disrupted pgl*J *gene*
*hgp*A *wbp*S::Kn	23E::*hgp*A *with a disrupted pgl*H *gene*

Appendix III Plasmid list

Plasmid	Description
pHp23S	pGEM T easy plasmid containing a portion of the *H. pullorum* 23S rRNA gene
pHp23::E	pHp23S with a erythromycin resistance cassette insertion
pHp23::Kn	pHp23S with a kanamycin resistance cassette insertion
pHp23::Cm	pHp23S with a chloramphenicol resistance cassette insertion
pHp*pgl*B1::Kn	pGEM T easy plasmid containing a portion of the *H. pullorum pgl*B1 gene interrupted with a kanamycin resistance cassette
pHp*pgl*B2::Kn	pGEM T easy plasmid containing a portion of the *H. pullorum pgl*B2 gene interrupted with a kanamycin resistance cassette
pHp*pgl*A::Kn	pGEM T easy plasmid containing a portion of the *H. pullorum pgl*A gene interrupted with a kanamycin resistance cassette
pHp*pgl*C::Kn	pGEM T easy plasmid containing a portion of the *H. pullorum pgl*C gene interrupted with a kanamycin resistance cassette
pHp*pgl*J::Kn	pGEM T easy plasmid containing a portion of the *H. pullorum pgl*J gene interrupted with a kanamycin resistance cassette
pHp*pgl*H::Kn	pGEM T easy plasmid containing a portion of the *H. pullorum pgl*H gene interrupted with a kanamycin resistance cassette
pHp*wbp*O::Kn	pGEM T easy plasmid containing a portion of the *H. pullorum wbp*O gene interrupted with a kanamycin resistance cassette
pHp*wbp*S	pGEM T easy plasmid containing a portion of the *H. pullorum wbp*S gene
pHp*wbp*S::Kn	pHp*wbp*S with a kanamycin resistance cassette insertion
pHp23E::*hgp*A	Plasmid for the insertion of a hexa histidine tagged *hgp*A gene onto the *H. pullorum* chromosome
pHp22E::0114	Plasmid for the insertion of a hexa histidine tagged 0114 gene onto the *H. pullorum* chromosome
ppgl_{Cj}	Plasmid containing all the genes required for the production of a *C. jejuni* lipid-linked heptasaccharide in *E. coli*
p$pgl\Delta B_{Cj}$	$ppgl_{Cj}$ with an insertionally inactivated *pgl*B gene
pMAF10	Plasmid for the expression of *C. jejuni pgl*B in *E. coli*
pMLHP1	Plasmid for the expression of *H. pullorum* pglB1 in *E. coli*
pMLHP2	Plasmid for the expression of *H. pullorum pgl*B2 in *E. coli*
pMLHC1	Plasmid for the expression of *H. canadensis pgl*B1 in *E. coli*
pMLHC2	Plasmid for the expression of *H. canadensis pgl*B2 in *E. coli*
pQE*hgp*A	Plasmid for the production of hexa histidine tagged *H. pullorum* HgpA in *E. coli*
pQE0114	Plasmid for the production of hexa histidine tagged *H. pullorum* 0114 in *E. coli*

Appendix IV Predicted Helicobacter N-linked glycoproteins. Putative *H. pullorum* glycoproteins and proteins with amino acid sequence similarity from *H. canadensis* and *H. winghamensis*. Putative glycoproteins were identified on the basis of a signal peptide sequence and at least one D/E-X-N-X-S/T glycosylation sequon. 67 potential glycoproteins were identified in *H. pullorum*, 15 of these had homology to *H. canadensis* putative glycoproteins and 11 had homology to *H. winghamensis* putative glycoproteins. As the genome sequence from the *H. pullorum* NCTC 12824 strain is unpublished, the gene number corresponds to genes from *H. pullorum* NCTC 98-5489.

H. pullorum NCTC 98-5489			*H. canadensis* NCTC 98-5491		*H. winghamensis* ATCC BAA-430	
Gene number	Sequons	Protein name	Gene number	Sequons	Gene number	Sequons
HPMG_00016	1	conserved hypothetical protein				
HPMG_01076	4	conserved hypothetical protein	HCAN_0714	4	HWAG_01179	4
HPMG_01475	1	conserved hypothetical protein				
HPMG_00136	2	outer membrane permeability protein	HCAN_0748	1	HWAG_00555	1
HPMG_00175	1	lipoprotein	HCAN_0920	1		
HPMG_00207	1	conserved hypothetical protein				
HPMG_00236	1	conserved hypothetical protein				
HPMG_00257	1	methyl-accepting chemotaxis protein 2	HCAN_1106	1		
HPMG_00332	1	conserved hypothetical protein				
HPMG_00374	1	flagellar functional protein	HCAN_0847	1	HWAG_00847	2
HPMG_00408	2	translocation protein TolB	HCAN_0878	1	HWAG_01428	1
HPMG_00410	2	conserved hypothetical protein	HCAN_0880	2	HWAG_01426	2

HPMG_00420	1	conserved hypothetical protein	HCAN_0890	1	HWAG_00753	1
HPMG_00424	1	fibronectin type III domain-containing protein	HCAN_0894	1	HWAG_00749	1
HPMG_00433	1	predicted protein				
HPMG_00503	1	predicted protein				
HPMG_00567	1	predicted protein	HCAN_0513	1		
HPMG_00621	2	TonB-dependent receptor protein				
HPMG_00689	1	outer membrane phospholipase a	HCAN_0149	2		
HPMG_00730	1	conserved hypothetical protein	HCAN_0114	1	HWAG_00335	1
HPMG_00765	3	conserved hypothetical protein	HCAN_0496	3		
HPMG_00785	2	conserved hypothetical protein				
HPMG_00797	1	Mn2/Zn2/ ABC transporter				
HPMG_00827	2	conserved hypothetical protein	HCAN_1530	1		
HPMG_00831	1	predicted protein	HCAN_1534	1		1
HPMG_00854	1	conserved hypothetical protein				
HPMG_00874	1	pdp protein				
HPMG_00883	1	conserved hypothetical protein				
HPMG_00913	1	predicted protein				
HPMG_00918	1	flavoprotein				
HPMG_00934	1	macrolide-specific efflux protein macA				
HPMG_01038	3	conserved hypothetical protein	HCAN_0430	1		
HPMG_01076	4	conserved hypothetical protein	HCAN_0714, HCAN_0457	4, 2	HWAG_01179	4

HPMG_01090	1	predicted protein				
HPMG_01102	1	conserved hypothetical protein				
HPMG_01115	2	conserved hypothetical protein				
HPMG_01116	3	pilus assembly protein				
HPMG_01119	1	predicted protein				
HPMG_01124	1	conserved hypothetical protein				
HPMG_01125	2	predicted protein				
HPMG_01126	1	predicted protein				
HPMG_01196	1	conserved hypothetical protein	HCAN_0266	1		
HPMG_01202	1	predicted protein				
HPMG_01206	3	predicted protein				
HPMG_01232	1	preprotein translocase subunit	HCAN_0591	1	HWAG_00154	1
HPMG_01241	1	ubiquinol-cytochrome c reductase				
HPMG_01275	1	predicted protein				
HPMG_01281	1	conserved hypothetical protein	HCAN_0374	1	HWAG_00315	1
HPMG_01334	1	nitrite reductase				
HPMG_01348	3	response regulator				
HPMG_01354	1	sensor histidine kinase HpkA				
HPMG_01475	1	conserved hypothetical protein				
HPMG_01484	2	conserved hypothetical protein	HCAN_0521	2	HWAG_00178	1
HPMG_01544	1	abc transport system periplasmic binding protein	HCAN_0723	1		
HPMG_01569	1	conserved hypothetical protein				

HPMG_01601	3	conserved hypothetical protein	HCAN_0988	1	HWAG_01420	1
HPMG_01667	1	conserved hypothetical protein	HCAN_0946	1	HWAG_01532	1
HPMG_01680	3	predicted protein				
HPMG_01718	1	conserved hypothetical protein				
HPMG_01725	1	conserved hypothetical protein				
HPMG_01727	1	conserved hypothetical protein				
HPMG_01728	1	predicted protein				
HPMG_01801	1	conserved hypothetical protein	.			
HPMG_01927	1	conjugal transfer protein TraG				
HPMG_01983	2	lectin C-type domain-containing protein				
HPMG_01992	1	filamentous hemagglutinin domain-containing protein	.			
HPMG_01995	1	predicted protein				