The Lipopolysaccharide of *Haemophilus parainfluenzae*

Rosanna Young

Green Templeton College

Submitted for the degree of
DOCTOR OF PHILOSOPHY
in the Department of Paediatrics, University of Oxford

2011

The Lipopolysaccharide of *Haemophilus parainfluenzae*

Rosanna Young, Green Templeton College, University of Oxford

Submitted for the degree of Doctor of Philosophy in Paediatrics
Hilary Term 2011

Abstract

Haemophilus parainfluenzae (*Hp*) and *H. influenzae* (*Hi*) are closely related members of the Pasteurellaceae family and are common commensal bacteria of the human nasopharynx. Whilst *Hi* is frequently implicated in meningitis, otitis media and respiratory tract infections, reports of pathogenic behaviour by *Hp* are very rare. Lipopolysaccharide (LPS) is a key component of the Gram negative cell wall, and its structure influences the ability of *Haemophilus* to interact with the host and evade immune clearance. A better understanding of the differences in LPS structure between *Hi* and *Hp* could help to ascertain which parts of the molecule are important for commensal and pathogenic behaviour.

Hi LPS comprises lipid A, a conserved oligosaccharide inner core, and an oligosaccharide outer core that differs between strains. The latter is partly phase variable by the slipped strand mispairing during replication of DNA repeat tracts within several LPS biosynthesis genes. Very little was known about LPS in *Hp* so we investigated its biosynthesis and structure in a panel of 20 *Hp* carriage isolates. Using PCR, DNA sequencing and Southern analysis we demonstrated that *Hp* possesses homologues of the *Hi* lipid A and inner core LPS synthesis genes and a few of the genes for outer core synthesis; however, homologues of the *Hi* phase variable outer core genes were largely absent and did not contain repeat tracts. The results of immunoblotting and collaborative structural analysis were consistent with this data. Phosphocholine, a phase variable *Hi* LPS epitope that has been implicated in otitis media, was found to be absent in *Hp* LPS due to the lack of four genes required for its biosynthesis and incorporation. The introduction of these genes into *Hp* led to the phase variable addition of phosphocholine to the LPS, indicating that there is no fundamental reason why *Hp* could not use a similar mechanism of variation to *Hi* if it was advantageous to do so.

SDS-PAGE data suggested the presence of O-antigens (repeated chains of sugars) in many of the *Hp* strains, an unusual feature for *Haemophilus*, and all of the strains were found to contain a potential O-antigen synthesis locus. Each locus encodes homologues of several glycosyltransferases in addition to either the Wzy polymerase- or ABC transporter-dependent mechanisms of O-antigen synthesis and transport. Comparisons of wild type and isogenic mutant strains showed that the O-antigen enhances resistance to complement-mediated killing and appears to affect adhesion to epithelial cells in vitro. *Hp* is a successful commensal organism but lacks the flexibility of adapting its LPS using repeat-mediated phase variation, potentially limiting its range of host niches.

Table of Contents

1 Introduction 1
1.1 The genus *Haemophilus* 1
1.2 The commensal and pathogenic behaviour of *Hi* and *Hp* 4
1.3 Factors affecting the colonisation and virulence levels of Gram negative bacteria 12
1.4 Lipopolysaccharide 18
1.5 *Haemophilus* genome sequences and their use in this project 29
1.6 *Hp* strains used in this study 34
1.7 Collaborative structural analysis: mass spectrometry and NMR 36
1.8 Aims of the project 37

2 Materials and Methods 38
2.1 *Haemophilus* strains, culture and storage 38
2.2 *Haemophilus* genomic DNA (gDNA) extraction 38
2.3 Standard polymerase chain reaction (PCR) 39
2.4 DNA restriction digestion 39
2.5 Long-range PCR (LR-PCR) and digests 40
2.6 DNA sequencing 40
2.7 Southern analysis 40
2.8 Reverse transcriptase PCR 41
2.9 Tricine SDS-PAGE LPS gels 42
2.10 Construction of plasmids in *E. coli* 43
2.11 Transformation of *Hp* 46
2.12 Cloning and sequencing OAg loci 47
2.13 Resistance to the bactericidal effects of human sera 48
2.14 Epithelial cell association assay 48
2.15 Assay to detect the inhibition of *Hp* growth by *Hi* and vice versa 49
2.16 Detection of phosphocholine and digalactoside by immunoblot 50
2.17 Preparation of cell pellets for LPS structural analysis 51
2.18 Bioinformatic analysis 52

3 Results: Prediction of *H. parainfluenzae* core LPS structures using genetic and phenotypic analyses 60
3.1 Lipid A and inner core LPS 61
3.2 Initial glycose extensions from the triheptosyl inner core LPS 66
3.3 Other *Hi* LPS outer core substituents 74
3.4 Visualisation of LPS glycoforms by tricine SDS-PAGE 87
3.5 Discussion 87

4 Results: Initial studies on O-antigens and single O-units 93
4.1 The *Hp* T3T1 genome encodes the enzymes required for O-antigen synthesis 94
4.2 All strains could have an OAg locus between *glnA* and *pepB* 106
4.3 Many *Hp* strains exhibit OAg-like LPS patterns 110
4.4 Mutation of the *Hp lgtF* gene by transformation: the OAg or O-unit is attached via Glc on HepI 112

4.5 Discussion 117

5 Results: Further investigation of O-antigen genes and their predicted functions 122
5.1 Categorisation of *Hp* OAg loci 123
5.2 Further work on group 1 strains 145
5.3 Identification of an OAg ligase gene 152
5.4 Discussion 159

6 Results: Interactions between *H. parainfluenzae* and the host environment 163
6.1 Examination of the role of LPS structure in determining serum resistance 164
6.2 Phosphocholine as a component of the *Haemophilus* LPS molecule 179
6.3 The association of *Hp* with epithelial cells 186
6.4 Interactions between *Hi* and *Hp* 190
6.5 Discussion 193

7 Results: The variation of LPS structure in *H. parainfluenzae* 200
7.1 The capacity for phase variation in *Hp* 201
7.2 The regulation of OAg chain length in group 1 *Hp* strains 216
7.3 The regulation of OAg production in group 2 *Hp* strains 220
7.4 Discussion 223

8 Final discussion 228

9 References 246

Abbreviations

aa	amino acids
AAT	2-acetamido-4-amino-2,4,6-trideoxy-D-galactose
ABC	ATP-binding cassette
Amp	ampicillin
BHI	brain heart infusion
BLAST	basic local alignment search tool
BLASTN	nucleotide BLAST
BLASTP	protein-protein BLAST
bp	base pairs
cAMP	cyclic adenosine monophosphate
cDNA	complementary DNA
CMP	cytidine monophosphate
COPD	chronic obstructive pulmonary disease
CRP	C-reactive protein
Crp	cAMP receptor protein
DDHep	D-*glycero*-D-*manno*-heptose
EDTA	ethylene diamine tetraacetic acid
Ery	erythromycin
Fuc	fucose
Gal	galactose
gDNA	genomic DNA
Glc	glucose
GT	glycosyltransferase
Hep	L-*glycero*-D-*manno*-heptose
Hi	*Haemophilus influenzae*
Hib	*Hi* type b
HMG	high molecular weight LPS glycoform
Hp	*Haemophilus parainfluenzae*
HSP	highest scoring sequence pair
Kan	kanamycin
Kdo	3-deoxy-D-manno-octulosonic acid
LPS	lipopolysaccharide
LR-PCR	long range PCR
MW	molecular weight
NAD	nicotinamide adenine dinucleotide
NCBI	National Center for Biotechnology Information, U.S.
NMR	nuclear magnetic resonance
nt	nucleotides
NTHi	non-typeable *Hi*
OAc	O-acetyl
OAg	O-antigen
OD	optical density
OM	otitis media
OMP	outer membrane protein

ORF	open reading frame
PBS	phosphate-buffered saline
PCho	phosphocholine
PCR	polymerase chain reaction
Pers. comm.	personal communication
PEtn	phosphoethanolamine
PHS	pooled human serum
rRNA	ribosomal RNA
RT-PCR	reverse transcriptase PCR
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SSC	saline sodium citrate buffer
sv.	serovar
TBLASTN	translated BLAST
Tet	tetracycline
UDP	uridine diphosphate
UndP	undecaprenyl phosphate
USS	uptake signal sequence
wrt	with respect to

Acknowledgements

I would like to thank my supervisor, Derek Hood, for making this project possible. Over the past three years you have given me the freedom and confidence to explore ideas whilst always providing guidance and expertise when I needed them.

I am grateful to past and present members of the Molecular Infectious Diseases Group for their friendship and for guidance with laboratory techniques, in particular Mary Deadman, Kathy Makepeace and Sherry Tsai. Thanks are also due to Richard Moxon for helpful discussions at the beginning of the project, and to Peter Power for facilitating and advising on bioinformatic analyses.

Derrick Crook and Abdel Elamin kindly supplied the bacterial strains upon which this work is based, whilst Elke Schweda, Brigitte Twelkmeyer and Varvara Vitiazeva provided invaluable structural LPS data to support our genetic analysis.

This research was funded by the Medical Research Council. Conference grants were generously supplied by Green Templeton College and the Society for General Microbiology.

Finally, I would like to thank my parents and James for all of their support and encouragement during the course of my D.Phil. studies.

1 Introduction

1.1 The genus *Haemophilus*

The genus *Haemophilus* is part of the Gram negative gammaproteobacterial Pasteurellaceae family. The type species of the genus, *Haemophilus influenzae* (*Hi*), is a facultatively anaerobic coccobacillus requiring factor X (haemin) and factor V (NAD) for its growth and was originally studied due to the erroneous assumption that it was the causative agent of influenza in humans. 21 other species have subsequently been published as *Haemophilus* at various times but nine of these have now been reclassified into other genera (Euzéby, 2010). *Haemophilus* currently encompasses species isolated from a variety of hosts including humans, cats, dogs, rabbits and pigs, and their bacterial-host interactions range from asymptomatic carriage to fatal disease.

In the early 1920s T. M. Rivers performed biochemical and growth tests on two bacterial isolates taken from the throat and lung of two influenza patients (Rivers, 1922). He observed that like influenza bacilli (*Hi*) they were Gram negative, non-motile, pleomorphic and non-sporebearing, but whilst the growth of normal influenza bacilli required media containing both hematin and yeast extract, these two isolates required only the latter supplement. The component of yeast extract that was necessary for growth was later identified as NAD (Lwoff, 1937). Rivers proposed that this new group of NAD-dependent but haemin-independent organisms could be named *Bacillus para-influenzae* or *Hemophilus para-influenzae*, and the name *Haemophilus parainfluenzae* (*Hp*) was adopted. Whilst some workers argued that there was simply a gradient of haemin dependency across *Haemophilus* strains (Biberstein & Gills, 1961), Evans' study of 80 isolates subsequently found a clear distinction between *Hi*, which required a minimum of 2-10

μg haemin per ml for optimal growth, and *Hp* which did not need any haemin for optimal growth (Evans et al., 1974). Evans also noted that the two *Haemophilus* species differed in their NAD requirements: when grown on complete basal medium with haemin, *Hi* colonies were the same size (maximum growth) whether 1 μg/ml or 5 μg/ml NAD was used, whereas 14/30 *Hp* isolates had severely reduced or no growth at the lower concentration of supplement.

Prokaryotic genera and species have traditionally been delineated on the basis of their phenotype: their host pathologies, colonial appearance, bacterial morphology, growth requirements and biochemical characteristics. Genotypic analysis has subsequently revolutionised our ability to determine the true evolutionary relationships between isolates and to categorise them into species more appropriately. In combination with phenotypic data, the sequencing of housekeeping gene fragments and 16S rRNA loci enabled Norskov-Lauritsen and Kilian to demonstrate that four *Haemophilus* species were sufficiently distant from the others to be given their own genus (*Aggregatibacter*), and that a number of haemolytic *Haemophilus* isolates formed a distinct lineage which they renamed *H. pittmaniae* (Norskov-Lauritsen et al., 2005; Norskov-Lauritsen & Kilian, 2006).

Norskov-Lauritsen and Kilian's phylogenetic analysis (Fig. 1.1) indicates that the three X and V factor dependent *Haemophilus* species isolated from man (*Hi*, *H. aegyptius* and *H. haemolyticus*) are very closely related. 16S rRNA analysis places *Hp* and *H. pittmaniae* as a separate cluster (Fig. 1.1b), but concatenated sequence data from four housekeeping genes groups all five of these species as a monophyletic lineage with strong bootstrap support (Fig. 1.1a). Together with DNA-DNA hybridisation data, this suggests that the inclusion of *Hp* in the genus *Haemophilus* is

justified. *H. ducreyi*, which causes chancroid in the human genital tract, was excluded from the figure as it was shown to be only distantly related to *Hi* by DNA-DNA hybridisation and 16S rRNA analysis.

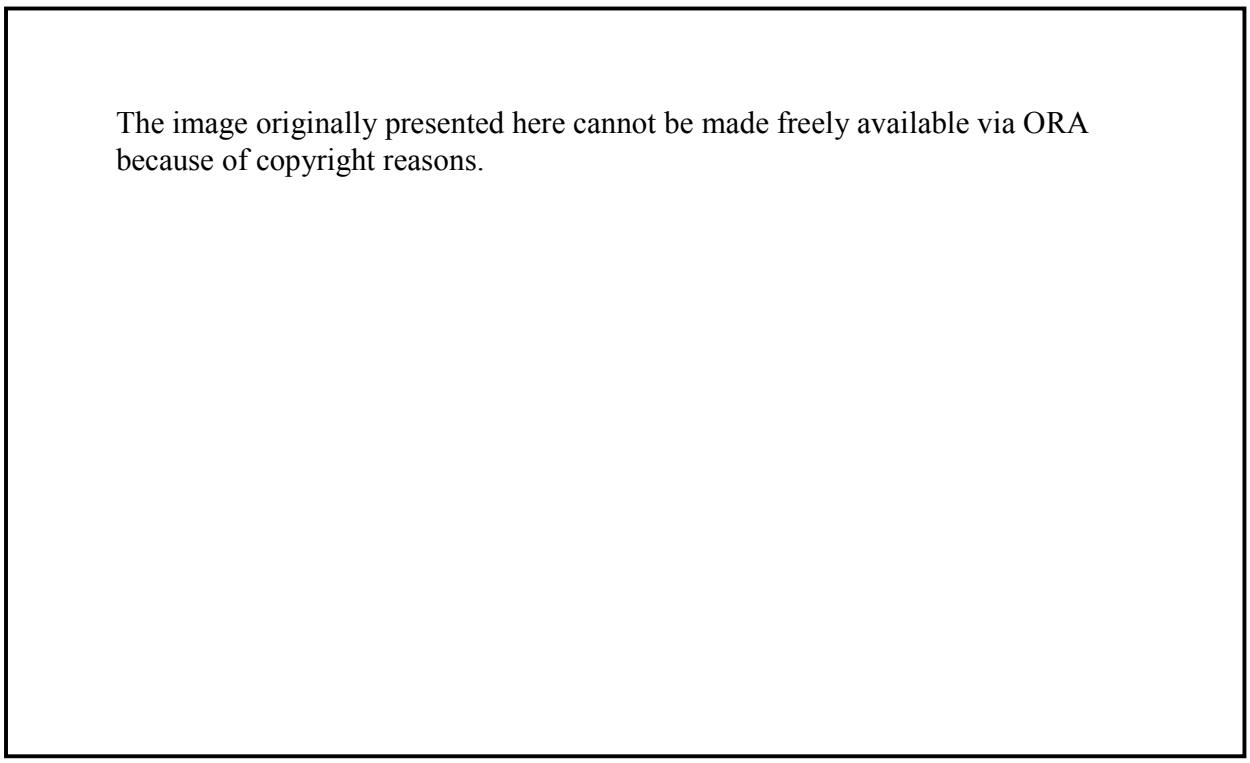


Figure 1.1. Inferred phylogenies of *Haemophilus* species and related bacteria. (a) Neighbour-joining tree based on concatenated sequences from four housekeeping genes *infB*, *adk*, *pgi* and *recA* (total length 1746 nt). Values to the left of nodes are bootstrap replications (%) supporting the node. Bar, 5 substitutions per 100 nt. **(b)** Neighbour-joining tree based on 1450–1454 nt 16S rRNA gene sequences from type strains. Nucleotide sequence accession numbers are given. Bar, 1 substitution per 100 nt.
Genus abbreviations are A, *Actinobacillus*; H, *Haemophilus*; P, *Pasteurella*. Note that *A. actinomycetemcomitans*, *H. aphrophilus*, *H. paraphrophilus* and *H. segnis* have since been reclassified as the genus *Aggregatibacter*. "*H. parainfluenzae*" indicates the *Hp* type strain HK409. Three other *Hp* isolates (HK102, HK755 and HK23) are included; for interpretation of other strain numbers see Norskov-Lauritsen et al. (2005). Image adapted from Norskov-Lauritsen et al. (2005).

Combining my own observations with those of others (Norskov-Lauritsen et al., 2005; Rivers, 1922; Taylor et al., 1992; Windsor et al., 1991), I would advocate the following description of *Haemophilus parainfluenzae* (for biochemical characteristics see Rivers (1922)):

A Gram negative, facultatively anaerobic, non-motile, non-haemolytic bacterium most frequently isolated from the human respiratory tract, ranging in form from coccobacillary to filamentous. Colonies grown on supplemented brain heart infusion agar are whitish and usually smaller and more opaque than those of *Hi*. Does not require haemin for growth. Most strains require NAD but occasional strains harbour a 5 kb plasmid that confers NAD independence. Urease, indole and ornithine decarboxylase production are dependent on biotype. Forms a heterogeneous but distinct phylogenetic cluster within the main *Haemophilus* lineage using housekeeping gene sequences and appears more closely related to *H. pittmaniae* than to other *Haemophilus* species using some genotypic methods.

1.2 The commensal and pathogenic behaviour of *Hi* and *Hp*

1.2.1 *Haemophilus* carriage in the upper airways

Hi and *Hp* are primarily commensal organisms residing in the human upper respiratory tract, and their ability to colonise and persist in this niche is crucial for their survival. The detected carriage rate of *Haemophilus* species in the human nose and throat depends on many factors including age, exposure to other carriers, vaccinations received and the sampling and culturing techniques employed. In most studies, *Hi* is considerably more common in infants than in adults. Carriage rates of type b *Hi* (Hib) have dropped since the introduction of the Hib vaccine but those of non-typeable (unencapsulated) *Hi* (NTHi) have not been affected since the vaccine is designed to recognise the type b polysaccharide capsule (Klein, 1997). Three examples of *Hi* carriage studies will be cited in order to represent the wide range of carriage rates in the literature.

Firstly, of nasopharyngeal swabs taken from 300 children aged 7-12 years, Ayyildiz et al. (2003) isolated Hib from 3%, non-type b *Hi* from 10% and *Hp* from 4%. Secondly, Hashida and colleagues (2008) found a nasopharyngeal *Hi* carriage rate of 47% in children aged 0-6 years attending day-care centres. Thirdly, Greenberg et al. (2004) swabbed the nasopharynx and oropharynx of 216 infants under the age of five and their mothers: the overall pharyngeal carriage rates of non-type b *Hi* were found to be 70% in the children and 36% in the adults. In the children, non-type b *Hi* was carried equally in the nasopharynx and oropharynx, whereas in the mothers 81% of positive cultures were positive from the oropharyngeal swab only; this suggests a changing distribution with age, perhaps due to an altered immunological and anatomical environment.

Fewer groups have examined carriage rates of *Hp*. As part of a study into chronic bronchitis, Taylor et al. (1992) took twelve monthly oropharyngeal swabs from each of 36 subjects aged 46-75 years with no recognised lung disease and found that *Hp* could be isolated from 100% of the individuals at least once, with 231/432 (53%) of the total samples proving positive. Another group, considering *Hp* as a potential source of antibiotic resistance, found ampicillin resistant *Hp* in throat swabs from 78% of 130 children under five presenting for day surgery (Scheifele & Fussell, 1981).

Many studies of bacterial carriage in the upper respiratory tract include only one species of *Haemophilus* and it is unwise to compare figures for *Hi* and *Hp* obtained in different settings. However, the overall picture that emerges is that both species have high carriage rates and can

colonise both the nasopharynx and oropharynx to some degree; and that the carriage of *Hi* but not necessarily *Hp* diminishes with age.

NTHi commonly causes otitis media and upper and lower respiratory tract infections, whilst encapsulated *Hi* can cause invasive disease including meningitis and septicaemia. In the year 2000 Hib is estimated to have caused 8.1 million serious illnesses worldwide in children under five (Watt et al., 2009). Conversely *Hp* infections are considered rare enough to merit the publication of individual case reports. However, several recent studies investigating new drugs against bacterial infections have included *Hp* in sensitivity testing, suggesting that drug developers consider this species to have pathogenic tendencies (Farrell et al., 2010; Jones et al., 2009). The following sections will compare the nature and frequency of *Hp* infections with those of *Hi*.

1.2.2 Invasive disease

Before the introduction of conjugate vaccines against Hib, encapsulated *Hi* was a leading cause of invasive disease including bacteraemia and meningitis. It is estimated that in 2000, the incidence of Hib meningitis per 100,000 children under five was 46 in Africa and 16 in Europe, reflecting widespread Hib vaccination in the latter continent (Watt et al., 2009). By comparison NTHi is rarely isolated from invasive disease, but occasional cases of NTHi bacteraemia and meningitis have been reported in recent years (Erwin et al., 2005). The isolation of *Hp* from invasive disease situations is also very infrequent, especially in adults. Black et al. (1988) found that 31 episodes of childhood meningitis and four of childhood septicaemia attributed to *Hp* had been reported in the English-language literature between 1965 and 1986. One of the rare *Hp*

adult meningitis cases occurred in a patient with inherited C7 deficiency, suggesting that the complement system usually plays an important role in the prevention of invasive *Hp* disease (Raoult et al., 1987).

Haemophilus is a rare cause of endocarditis, identified in around 1% of adult cases, although it may also account for some culture-negative cases (Darras-Joly et al., 1997; Hamed et al., 1994). Darras-Joly and colleagues (1997) reviewed 42 instances of *Haemophilus* endocarditis in adults and reported the causative species as 62% *Hp*, 7% *Hi* and 31% *H. aphrophilus* or *paraphrophilus* (now collectively *Aggregatibacter aphrophilus*). Risk factors included pacemaker implantation, dental procedures and tongue piercing, suggesting infection by surgical or oral routes. The low frequency of *Hi* endocarditis may be due to the pharyngeal rather than oral/dental location of this species as a commensal, or may simply reflect a lower carriage rate than *Hp*.

Pericarditis, usually as a complication of pneumonia or occasionally endocarditis, has been attributed to *Hi* in a number of cases (McDonald et al., 1994) although its incidence has declined in developed countries since the introduction of the Hib vaccine. No reports of *Hp*-associated pericarditis were found in the literature.

1.2.3 Respiratory tract and lung infections

Pneumonia is a major cause of mortality worldwide, especially in young, elderly or immunocompromised individuals. Until recently the bulk of *Haemophilus* pneumonia was due to Hib, but other capsular serotypes and NTHi are accounting for an increasing percentage of cases since the introduction of Hib vaccines. Hib is estimated to have caused 1304 cases of pneumonia

per 100,000 children under five globally in 2000, with figures of 1724 for Africa and 283 for Europe (Watt et al., 2009).

Hp is not generally reported to be an aetiological agent of pneumonia. One study detected *Hp* from the sputum of 3.8% of 921 paediatric pneumonia or bronchitis patients but it is not clear whether these were part of mixed infections (Numazaki et al., 2004). Another group do report a pure and prolific growth of *Hp* from the sputum of an adult pneumonia patient, although the methods used to classify the bacterium as *Hp* are not stated (Pillai et al., 2000). The fact that a case report was published reflects the perceived rarity of *Hp* as a cause of pneumonia; its true frequency is unknown.

The role of bacteria in chronic obstructive pulmonary disease (COPD) has been controversial, but a recent review cites *Hi* as a factor in 20-30% of COPD exacerbations with frequent involvement also in stable disease (Sethi & Murphy, 2008). In contrast, *Hp* is thought unlikely to play a role in either the stable disease or exacerbations, despite its frequent isolation from the sputum of COPD patients (Sethi & Murphy, 2008; Taylor et al., 1992). Marin and colleagues (2010) showed that the presence of *Hi* in COPD was associated with sputum neutrophilia and higher levels of IL-1β and IL-12 whereas no inflammatory response was found in the presence of *Hp*. In another study, *Hp* was isolated from sputum at the same frequency during acute episodes of bronchitis as was found in routine sputum samples from the same COPD patients (Taylor et al., 1992), again consistent with *Hp* behaving as a commensal organism in this clinical setting.

A case of thoracic empyema (the collection of pus in the pleural cavity) was reported in 1981 in which *Hp* was isolated from blood cultures and empyema fluid, two normally sterile sites (Cooney et al., 1981). The authors concluded that alcoholism and COPD may have predisposed the patient to infection.

1.2.4 Otitis media

Otitis media (OM) is responsible for a large proportion of paediatric GP visits, particularly in children aged under 36 months. In a study of 8198 bacterial culture-positive episodes of OM in children under five, *Hi* (usually NTHi) was isolated as the single species in 3108 cases and as part of a mixed infection in 1874 cases (Broides et al., 2009). It is difficult to ascertain whether *Hp* is ever capable of causing middle ear infections. Many authors include otitis media in a list of putative *Hp*-related diseases, but there are no reliable primary sources of information to confirm this. OM patients do not always undergo tympanocentesis to identify bacterial pathogens so it is likely that *Hp* OM would be missed if it is a rare occurrence. One recently reported case of *Hp* adult meningitis was preceded by acute OM but no culture of the middle ear fluid was performed (Cardines et al., 2009).

1.2.5 The digestive tract

Hi and *Hp* are generally regarded as members of the upper respiratory tract flora which can act as opportunistic pathogens in other sites of the body. However, Palmer (1981) isolated *Haemophilus* species from 63/525 (12%) of faecal samples from healthy adults, suggesting that they may also constitute part of the flora of the small intestine. When Palmer performed X and V

factor testing on these isolates along with 159 faecal *Haemophilus* isolates from patients with gastrointestinal symptoms, he classified 6% overall as *Hi*, 81% as *Hp* and the remainder as having variable X factor dependence. Some of the isolates were then plated on nutrient agar with a streak of ox bile; whilst the bile allowed normal growth of *Hp*, the *Hi* isolates grew weakly compared to their growth on chocolate (lysed blood) agar with bile. Palmer concluded that bile is a good source of V factor but that haemin levels are low and variable; this provides an explanation for the frequent isolation of *Hp* from the faecal samples, as the nutrient requirements of this species may be fulfilled by bile in the digestive tract.

Frankard and colleagues (2004) summarised 13 reported cases of biliary tract (liver or gall bladder) infections from which *Hp* had been isolated; in ten of these cases, *Hp* was the only pathogen isolated from the hepatic abscess or bile specimen. Several groups have hypothesised that *Hp* can ascend from the gastrointestinal tract to the biliary tract, resulting in hepatic abscesses, acute cholecystitis or peritonitis (Bottone & Zhang, 1995; Frankard et al., 2004). *Streptococcus* species and *Escherichia coli* appear to be responsible for most biliary tract infections, and the isolation of *Hp* from this site is rare.

1.2.6 Issues regarding the reported pathogenicity of *Hp*

A number of authors have expressed concerns over the validity of methods used to identify and record the occurrence of *Hp* in a clinical setting and suggest that levels of pathogenic behaviour may be under- or over-reported. Unless *Haemophilus* infection is already suspected, the specimen for identification may not be plated on an appropriate medium such as chocolate agar and *Hp* could be missed (Frankard et al., 2004). Even when *Hp* is identified from a clinical

sample it is often dismissed as background commensal flora especially when found alongside other bacterial species regarded as being more pathogenic. In these two instances the pathogenicity of *Hp* could be underestimated.

Under the current National Standard Method used by NHS clinical microbiologists, cultured species can be classified as a presumptive identification "if appropriate growth characteristics, colonial appearance and Gram stain of the culture are demonstrated" (HPA, 2009). Although additional biochemical and X/V factor dependence tests are often used to identify and distinguish *Haemophilus* species, 16S rRNA sequence typing is not routine so it is possible that many isolates are misidentified. According to one researcher in the USA, many clinical laboratories report all strains that only require factor V as *Hp* (Hamed et al., 1994); some of these isolates are likely to actually be *A. aphrophilus* or *Aggregatibacter segnis*, or even *Hi* if there has been carry-over of haemin into the factor X test. Despite these problems there is strong evidence that true *Hp* isolates are sometimes associated with disease including endocarditis and meningitis, as the more recent case studies use reliable 16S ribotyping for species confirmation (Cardines et al., 2009; Hamed et al., 1994).

Hp strains can be classified into eight biotypes based on their production of indole, urease and ornithine decarboxylase, with biotypes I and II typically accounting for around 80-90% of isolates (Palmer, 1981; Taylor et al., 1992). For respiratory isolates the percentage that are biotypes I and II are roughly equal, but amongst strains isolated from (but not necessarily causing) genitourinary or neonatal infections, biotype II was nine times more frequent than

biotype I (Quentin et al., 1989). No other associations have been reported between particular *Hp* biotypes and pathogenicity.

In conclusion, *Hp* is rarely reported as a pathogen compared to *Hi*, but their relative isolation rates vary depending on the type of infection. Otitis media, meningitis and pneumonia are frequently caused by *Hi* whereas *Haemophilus* endocarditis and biliary tract infections are rarer and appear to be largely *Hp*-related. Other disease situations from which *Hp* has occasionally been isolated include urinary tract infections, septic arthritis and conjunctivitis. It is unclear whether pathogenic behaviour is beneficial for *Haemophilus*: diseases such as meningitis would appear to be evolutionary dead ends, as they do not aid bacterial transmission and the bacteria would die with their hosts. Instead, virulence may be an accidental by-product of selection for traits which also enhance commensal survival in the upper respiratory tract.

1.3 Factors affecting the colonisation and virulence levels of Gram negative bacteria

The high carriage rates of *Hi* and *Hp* reflect their success at colonising the human respiratory tract. Colonisation requires a bacterial strain to transmit to a new host, survive and replicate using the nutrients available in the new environment, avoid mucociliary clearance, and avoid attacks by components of the immune system such as antimicrobial peptides, the membrane attack complex, macrophages and antibody responses. The persistence of the strain is ultimately determined by the balance of the replication rate versus the rate of clearance and death. The virulence of an organism then depends on its ability to reach a new niche (e.g. the middle ear or

nervous system) and upon the subsequent host-microbe interactions. Some of the factors which determine colonisation and virulence behaviour are reviewed here; most of these molecules are expressed at the cell surface, i.e. the physical interface between a bacterium and its host, or are released into the host environment.

1.3.1 Lipopolysaccharide

Lipopolysaccharide (LPS) is a major component of the outer leaflet of the outer membrane in Gram negative bacteria, playing a role in cell wall integrity and many other surface properties. As LPS is the main focus of this thesis, its structure and function are discussed in greater detail in Section 1.4.

1.3.2 Polysaccharide capsule

Capsules comprise high molecular weight chains of sugars and form a protective layer around the bacterial cell. The ability of an *Hi* strain to synthesise one of the six capsule types, designated a to f, is a major virulence determinant for invasive disease. Strains with a type b capsule are more resistant to phagocytosis than isogenic non-encapsulated strains (Noel et al., 1992), and an increased copy number of the *cap* (capsule synthesis) locus is associated with higher capsule production levels and decreased susceptibility to complement-mediated lysis (Noel et al., 1996). These effects are likely due in part to the shielding of invariant surface components from recognition by antibodies and to decreased deposition of complement component 3 (C3). The presence of a capsule is also associated with virulence in several other Gram negative and Gram positive invasive pathogens that initially colonise the respiratory tract, such as *Neisseria*

meningitidis (13 serotypes), *Staphylococcus aureus* (2 main serotypes) and *Streptococcus pneumoniae* (91 serotypes) (Arbeit et al., 1984; Hill et al., 2010; Hyams et al., 2010).

Each molecule of a polysaccharide capsule is a series of repeating units in which the sugars are linked by glycosidic and/or phosphodiester bonds. Structural features such as repeat unit composition, sugar configuration and linkage positions together help to determine the host's immune response to the polysaccharide, and for many species only a few highly virulent serotypes are responsible for the majority of the disease burden. The *Hi* capsule repeat units are mostly disaccharides and include sugars such as GlcNAc, GalNAc, ManANAc and ribitol.

Capsule synthesis and transport mechanisms have been well characterised in *E. coli* and appear to be reasonably conserved across other species. The repeat units are assembled on a carrier lipid, undecaprenyl phosphate (UndP), on the cytoplasmic face of the inner membrane using nucleotide sugar precursors and glycosyltransferase enzymes. In *E. coli*, the translocation and polymerisation of the UndP-linked capsular polysaccharides is then carried out by either a Wzy-dependent or ABC transporter mechanism (see Section 1.4.1) depending on the capsular serotype (Whitfield, 2006). In encapsulated *Hi*, this step always utilises the ABC transporter system, which is encoded by the *bexA* and *bexB* genes in the capsule locus (Kroll et al., 1990; Satola et al., 2003). Capsules must then be translocated through the outer membrane to the surface, requiring the *wza* gene in *E. coli* and the *hcsA* and *hcsB* genes in *Hi*. It is currently unclear whether capsular polysaccharides are anchored to the outer membrane by a lipid component or are fully released upon translocation.

Whilst there is an assumption in the literature that *Hp* strains do not typically produce a capsule, it appears that this hypothesis has never been examined by techniques such as electron microscopy, antiserum agglutination or genetic analysis as it has in *Hi*. Kilian (1976) reported that one of 85 *Hp* isolates was encapsulated but does not state the methods used to reach this conclusion.

1.3.3 Surface proteins and glycoproteins

Bacteria colonising the nasopharynx must adhere to epithelial cells, mucins and/or an extracellular matrix if they are to avoid mucociliary clearance, and in *Hi* several surface proteins and glycoproteins are known to mediate this process. These include high molecular weight proteins (HMW1 and HMW2), *Haemophilus* surface fibrils (Hsf), the *Hi* autotransporter adhesin (Hia), *Haemophilus* adhesion and penetration protein (Hap), and haemagglutinating pili (Cotter et al., 2005; St Geme et al., 1998). Several of these highly immunogenic adhesins are being analysed as potential vaccine targets for NTHi (Winter & Barenkamp, 2010). For some adhesins the receptor molecule has been characterised: for example, there is a specific interaction between P5 fimbriae on *Hi* cells and ICAM-1 on human epithelial cells (Avadhanula et al., 2006). The surface proteins of *Hp* have not been studied in such detail, but one strain was found to express at least three major outer membrane proteins (Morton & Williams, 1989) and another to express a 34 kDa protein that clustered on the cell surface and played a major role in adherence to saliva-coated hydroxyapatite and *Streptococcus sanguis* in an in vitro model of dental plaque (Lai et al., 1990; Liljemark et al., 1992). Kahn and Gromkova (1981) found that only one of 20 *Hp* isolates appeared to be piliated when studied using electron microscopy.

Other cell surface proteins can aid bacterial survival by disrupting specific aspects of the host's response. In the human immune system, C4BP downregulates the classical and lectin complement pathways by degrading C3 convertase and C4. Some NTHi strains (but not typeable *Hi*) specifically bind C4BP, resulting in reduced opsonisation and increased serum resistance (Hallstrom et al., 2007). The mechanism of binding is uncharacterised but is presumed to involve a surface protein. It is not known whether *Hp* possesses a similar system for complement evasion.

1.3.4 IgA protease

IgA is the most prevalent immunoglobulin on mucosal surfaces such as the lining of the respiratory tract, and has been associated with bacterial aggregation and the prevention of bacterial adhesion to host cells (Kraehenbuhl & Neutra, 1992). Kilian et al. (1979) incubated human IgA with supernatant from bacterial cell suspensions and found that IgA1 was partly cleaved by all of 19 *Hi* strains but by none of five *Hp* strains tested. A similar study showed no IgA protease activity in another seven *Hp* strains (Senior & Ip, 1999). The cleavage process separates the Fab and Fc fragments of IgA1, decoupling its recognition and effector functions. It has since been confirmed that all *Hi* strains contain an *iga* gene, which encodes an IgA protease: the C-terminal region of this autotransporter forms a pore in the outer membrane through which the rest of the protein is passed, and cleavage leads to secretion of the active protease (Poulsen et al., 1989; Vitovski et al., 2002). Approximately a third of *Hi* strains also have a second gene (*igaB*) which cleaves IgA at a different amino acid position (Fernaays et al., 2006).

The amino acid sequence that is recognised by the *Hi* IgA protease is only conserved in the IgA of humans and great apes (Johnson et al., 2009), so IgA from other species cannot be inactivated by this enzyme. This makes it difficult to study the role of the protease in commensal and pathogenic behaviour using animal models. However, one study found a significantly higher median level of IgA protease activity in vitro in nontypeable *Hi* isolates from normally sterile sites or sputum from patients compared to isolates from the throats of asymptomatic carriers (Vitovski et al., 2002), suggesting a possible correlation between *iga*/*igaB* expression and virulence.

1.3.5 Interactions with other microbes

In a complex polymicrobial community such as that found in the nasopharynx, bacteria must compete with each other for nutrients and for space on host surfaces. This is likely to drive the evolution not only of more efficient nutrient uptake systems and surface attachment mechanisms, but also of more direct strategies that eliminate the competition. For example *S. pneumoniae*, which is an upper respiratory tract pathogen, produces hydrogen peroxide that kills *Hi* in co-culture experiments; it also releases neuraminidase that can cleave sialic acid from *Hi* LPS, potentially leaving the bacteria more vulnerable to complement-mediated killing (Pericone et al., 2000; Shakhnovich et al., 2002).

A wide range of bacteria produce proteins known as bacteriocins which kill or prevent the replication of closely related species or strains in order to promote their own survival (Riley & Wertz, 2002). Bacteriocins have been identified in gut microbes such as *E. coli* and in several oral bacteria including *Streptococcus mutans* and *Lactobacillus paracasei* (Cascales et al., 2007;

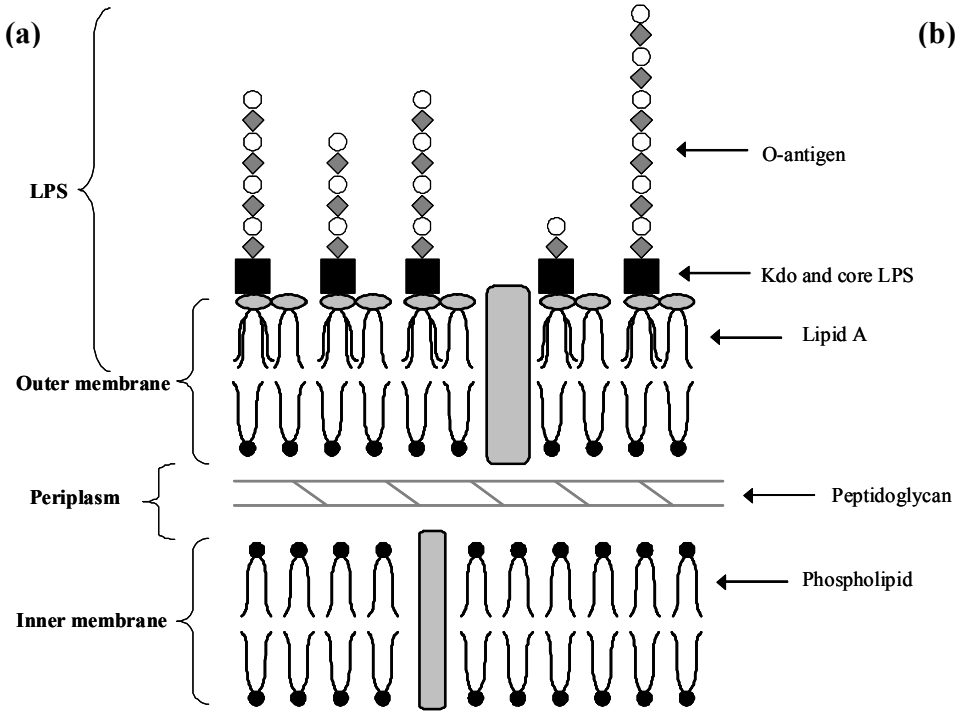
Hojo et al., 2009). Their modes of action range from creating pores in the inner membrane to attacking DNA or RNA with nucleases (Cascales et al., 2007). Most strains of Hib produce a bacteriocin named haemocin which prevents the growth of all other *Hi* capsular serotypes and NTHi by an unknown mechanism (Venezia & Robertson, 1975); as *Hp* is closely related to *Hi* it may also be susceptible to haemocin, depending upon whether it has close homologues of the relevant surface receptor and/or target enzyme/molecule. Whether *Hp* itself produces a bacteriocin, or indeed an antibiotic substance with a broader spectrum of target species, has yet to be ascertained.

1.4 Lipopolysaccharide

1.4.1 The LPS of Gram negative bacteria

The LPS structure of most Gram negative bacteria comprises three components: lipid A, core oligosaccharide and a repetitive polysaccharide termed the O-antigen (OAg). The hydrophobic lipid A moiety anchors the molecule in the membrane, whilst the sugar components of LPS are exposed on the cell surface (Fig. 1.2a) and confer various properties on the bacterium depending on their structure, potentially including adhesion and immune evasion.

The lipid A part of the LPS molecule, sometimes referred to as 'endotoxin', is composed of a phosphorylated glucosamine dimer and several acyl chains (Raetz & Whitfield, 2002). In the host, the release of LPS during bacterial colonisation and lysis induces the production of proinflammatory cytokines through the recognition of lipid A by a complex of Toll-like receptor 4 (TLR4) and its co-receptor MD2 (Bryant et al., 2010). Whilst this is an important mechanism



The image originally presented here cannot be made freely available via ORA because of copyright reasons.

Figure 1.2. **(a)** Model of the inner and outer membranes of a Gram negative bacterium. Grey rectangles represent transmembrane proteins. Based on Raetz and Whitfield (2002). **(b)** Electrophoretic LPS profiles of three *Bradyrhizobium* strains, demonstrating different OAg banding patterns due to variation in O-unit size and modal OAg length. Taken from Santamaria et al. (1998).

for local bacterial clearance, systemic overstimulation of this pathway can result in fatal septic shock. The exact composition of lipid A can vary with respect to the degree of phosphorylation and the number, position and length of the acyl chains. These variations can affect the binding affinity between lipid A and TLR4-MD2, altering activation of the downstream signalling cascade; for example, a *Salmonella enterica* sv. Typhimurium mutant with pentaacylated rather than hexaacylated lipid A grew at the same rate as the parental strain in mice but continued to replicate to ten times the normal lethal load before 10% of the mice died (Khan et al., 1998).

Lipid A is elaborated with between one and three 3-deoxy-D-manno-octulosonic acid (Kdo) residues. Whilst mutation of the Kdo transferase enzyme typically results in an attenuation of the

host immune response to LPS (Zughaier et al., 2004), Kdo addition is required for complete lipid A acylation and is the acceptor for the polysaccharide moiety of the LPS, so it is difficult to evaluate its own contribution to this effect. In most bacteria the LPS molecule is then extended by two to three residues of L-*glycero*-D-*manno*-heptose (Hep). The Kdo and Hep residues, together with any substitutions such as phosphoethanolamine whose positions are conserved within a species or genus, are collectively defined as inner core LPS. The fact that Kdo and Hep are found in the LPS of almost all Gram negative bacteria reflects their importance in outer membrane integrity (Raetz & Whitfield, 2002).

The remaining parts of the LPS molecule, namely outer core sugars and OAg, are more exposed to environmental stresses and host responses and consequently display greater structural diversity. Outer core LPS components vary between species and strains but can include hexose and heptose sugars, N-acetylhexosamines, sialic acid, glycine, phosphocholine, phosphoethanolamine and O-acetyl groups. In *E. coli* there are five known outer core types, each comprising a branched structure of five sugars from a selection of Glc, Gal, GlcNAc and Hep (Amor et al., 2000), and the glycosyltransferases involved in outer core assembly are encoded by the *waa* locus. Specific outer core sugars can play important roles in the pathogenic behaviour of a bacterium: for example, *S. enterica* sv. Typhimurium mutants that lack an outer core Gal residue from their LPS are less able to adhere to and invade epithelial cells (Bravo et al., 2010).

The outermost part of the LPS molecule, the OAg, is a long polysaccharide chain composed of repeat units (O-units) of one to seven sugars. LPS that contains OAg displays a characteristic regular ladder pattern when fractionated by tricine SDS-PAGE, with each ‘rung’ containing a

glycoform with a particular number of O-units (Fig. 1.2b). The enzymes required for OAg synthesis and assembly are usually encoded by a distinct gene cluster termed the OAg locus, facilitating co-regulation, although in *Helicobacter pylori* the relevant genes are spread throughout the genome (Hug et al., 2010). The combination of OAg enzymes synthesised by a particular bacterium determines the nature, order and linkages of the sugars in its O-unit, which can be further elaborated with non-stoichiometric substituents including acetyl and methyl groups; tens of different OAg serotypes can be observed for a single species, with 193 classified for *Vibrio cholerae* alone (Samuel & Reeves, 2003). The presence of OAg-containing LPS is widespread across Gram negative species, occurring for example in *Aeromonas hydrophila* isolated from diseased fish (Jimenez et al., 2008) and *Bradyrhizobium* from legume root nodules (Santamaria et al., 1998), in addition to many species isolated from humans such as those belonging to the genera *Escherichia*, *Klebsiella*, *Salmonella*, *Vibrio*, *Pseudomonas*, *Helicobacter* and *Francisella*.

OAgs appear to play numerous roles in the modulation of bacterial-host interactions. As a highly expressed OAg can effectively coat the bacterium, it is thought to act as a barrier to prevent host molecules from accessing the cell surface. For example, wild type *Bordetella parapertussis* is more efficient at colonising the mouse respiratory tract than isogenic OAg-deficient mutants because its OAg inhibits deposition of the complement component C3, protecting the bacteria from complement-mediated killing (Goebel et al., 2008). C3 deposition is required for the formation of the membrane attack complex (MAC) and for targeting bacteria to phagocytes. OAg can also prevent antibodies that recognise conserved membrane proteins from reaching their target: in a vaccine study, antibodies generated against an OAg-deficient *E. coli* mutant did

not confer protection against the wild type strain because OAg impeded antibody binding (Russo et al., 2009).

In addition to these general properties, the diversity of OAg structures within a species provides antigenic variation so that antibodies generated against the OAg of one strain are unlikely to recognise and target a second strain that may subsequently attempt to colonise the same host. The exact structure of an OAg, along with its length and the presence of any modifications, also influences its interactions with host cells and molecules. The gut bacterium *H. pylori* can express OAg with the same structure as the Lewis X or Y antigens that are found on human gastric epithelial cells, and this molecular mimicry prevents the formation of antibodies against these epitopes (Appelmelk et al., 2000). In other cases a specific OAg can serve as a ligand for a particular host molecule such as a surface receptor, affecting bacterial adhesion (John et al., 2002). Some OAgs, for example in *S. enterica*, act as receptors for bacteriophage through the binding of the tailspike protein (Steinbacher et al., 1997).

Whereas the core LPS oligosaccharides are assembled onto lipid A-Kdo through the sequential transfer of each sugar from its nucleotide sugar precursor, the OAg polysaccharide is added *en bloc*. OAgs are assembled in the same way as polysaccharide capsules (Section 1.3.2): an UndP-sugar phosphotransferase transfers the first sugar of the O-unit onto an UndP carrier lipid, and further glycosyltransferase enzymes add the subsequent sugars from their nucleotide sugar precursors to complete the O-unit. One of two alternative mechanisms is then used to polymerise and translocate the units (reviewed by Samuel and Reeves (2003)). In the Wzy-dependent system, the OAg flippase enzyme (Wzx) flips each UndP-linked O-unit from the cytoplasmic

face to the periplasmic face of the inner membrane. The units are then polymerised by the OAg polymerase, Wzy, and the resulting OAg chain is ligated to the LPS core by the OAg ligase, WaaL. In this system, the modal chain length is determined by an unknown mechanism involving a fourth enzyme named Wzz. Most *E. coli* OAgs are assembled using the Wzy-dependent mechanism.

The alternative system requires an ABC transporter, comprising two permease subunits (Wzm) for translocation and two ATPase subunits (Wzt) to drive the process. In this case the entire OAg chain is assembled on the cytoplasmic face of the inner membrane using glycosyltransferases before its translocation to the periplasmic side. The OAg is then ligated to the LPS core by WaaL as before. The ABC transport mechanism is traditionally associated with homopolymeric OAg, but OAg loci containing *wzm* and *wzt* in serotypes containing heteropolymeric OAg have recently been reported in species such as *Actinobacillus pleuropneumoniae* and *E. coli* (Perepelov et al., 2009; Xu et al., 2010).

The key structural distinction between capsule and OAg polysaccharides is the anchoring of OAg to the outer membrane through attachment to the LPS core. This is not always clear-cut: the seven different polysaccharides that determine serotype in *A. actinomycetemcomitans* were originally called capsules but were later discovered to be attached to LPS, hence their reclassification as OAg. In *E. coli*, serotypes that use the Wzy-dependent system for capsule synthesis also add the same polysaccharide, or a single repeat unit of it, to some LPS molecules; in several serotypes this is the only OAg structure found (Whitfield, 2006). Connections have also been observed between OAg synthesis and protein glycosylation in a few bacterial species.

For example, a single O-unit is added to *Pseudomonas aeruginosa* pilin using the enzymes encoded by the OAg locus, whilst flagellin subunits in the same species are glycosylated with a heterogeneous glycan chain whose synthesis also requires one of the same enzymes, namely the dehydrogenase WbpO (DiGiandomenico et al., 2002; Miller et al., 2008).

1.4.2 The LPS of *Hi*

Over the past twenty years, the LPS structures of many *Hi* strains have been elucidated and most of the genes required for their assembly have been identified and studied in detail (Deadman et al., 2006; Deadman et al., 2009; Schweda et al., 2007; Weiser et al., 1989a). Helander and colleagues (1988) demonstrated that the lipid A moiety of *Hi* strain I-69 Rd^-/b^+ is similar to that of LPS in other Gram negative bacteria, with a diglucopyranosamine headgroup phosphorylated at the C4' and C1 positions and linked to six acyl chains, the same number as in *E. coli*. This was later shown to be the major lipid A species in all of 24 other *Hi* strains, with minor species including variants with alkane heterogeneity or O-acetyl substitutions (Mikhail et al., 2005).

However, *Hi* LPS is unusual in several respects when compared to that of other Gram negative bacteria (Fig. 1.3). Firstly, whilst *E. coli* LPS contains two adjacent Kdo residues, *Hi* LPS contains only a single Kdo which is phosphorylated at the C4 position (Zamze et al., 1987). The genes required for Kdo phosphorylation and for the addition of Hep specifically to KdoP are also present in several other Pasteurellaceae genera (e.g. *Mannheimia* and *Aggregatibacter*) and in *Vibrio* species (BLASTP search, data not shown). A single KdoP is considered characteristic of *V. cholerae* LPS (Chatterjee & Chaudhuri, 2003). The biological implications of KdoP- versus Kdo_2-containing LPS epitopes are currently unclear.

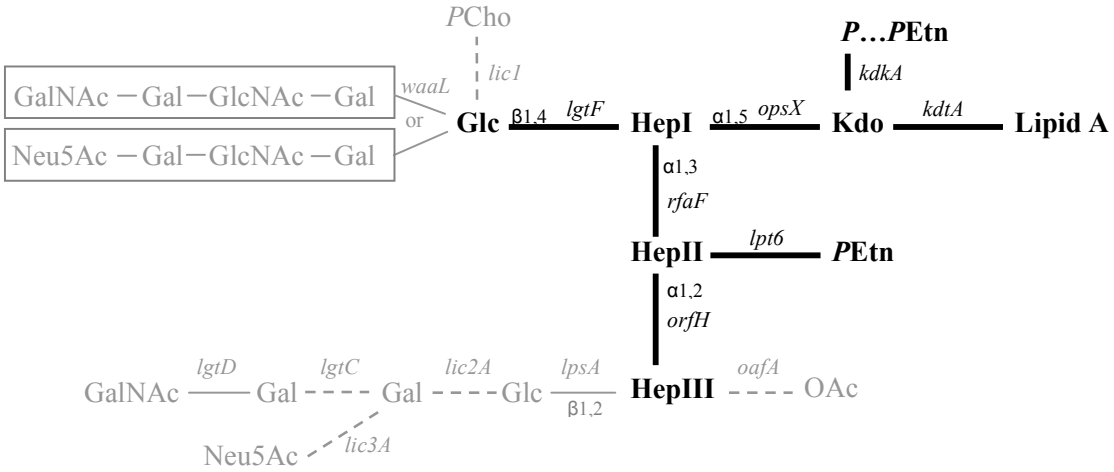


Figure 1.3. LPS structure of *Hi* strain Rd. *Hi* LPS contains lipid A, KdoP, a conserved inner core and a partly phase variable outer core; in some strains a tetrasaccharide HMG unit is added, but polymerised OAg is never observed. Bold type represents the region that is conserved across all *Hi* strains. Outer core LPS (Risberg et al., 1999) is shown in grey, with phase variable additions indicated by dotted lines. The gene encoding each transferase (Hood et al., 2001b) is given in italics. The two alternative HMG units are shown in rectangles (not all LPS glycoforms in *Hi* Rd contain HMG). Additions by *lgtC* and *lic3A* are mutually exclusive.

Secondly, many of the enzymes responsible for outer core LPS assembly in *Hi* are phase variably expressed using a tetranucleotide repeat mediated mechanism (Power et al., 2009; Weiser et al., 1989b). Simple sequence repeat tracts (microsatellites) have a higher mutation rate than other regions of DNA due to slipped-strand mispairing during DNA replication, resulting in a change in the repeat number. The slippage of intragenic repeats can cause a frameshift, leading to the loss of correct translation, whilst in other cases repeat tracts before the start codon help to control the level of promoter activity. A number of bacteria take advantage of this mechanism to alter the expression of various cell surface structures in a stochastic, high frequency and reversible manner, allowing them to adapt quickly through selection for individuals expressing the optimum combination of phase variable structures for a particular environment.

Six CAAT repeats: Lic1A is expressed
at gaa tac aaa aat gct atg caa tca atc aat caa tca atc aat caa att gta gga ttt gtt …

E Y K N A M **Q S I N Q S I N** Q I V G F V…

Gain one CAAT repeat (n = 7): Lic1A is expressed
atg aat aca aaa atg cta tgc aat caa tca atc aat caa tca atc aat caa att gta gga ttt gtt …

M N T K M L C **N Q S I N Q S I N** Q I V G F V…

Lose one CAAT repeat (n = 5): Lic1A is NOT expressed
a tga ata caa aaa tgc tat gca atc aat caa tca atc aat caa att gta gga ttt gtt …

***** (*I Q K C Y A* ***I N Q S I N*** *Q I V G F V*...)

Figure 1.4. Phase variation of protein expression due to the slippage of tetranucleotide repeat tracts during DNA replication. This mechanism is employed by *Hi* to vary the expression of LPS synthesis enzymes; the *lic1A* gene, which encodes a choline kinase, is shown as an example. The 5' end of the gene sequence is given with its translation below. The second and third DNA sequences illustrate the gain and loss of one tetranucleotide repeat unit, respectively. $(CAAT)_n$ repeats are highlighted in black, and potential start codons in grey. For *lic1A* to give the correct translation, at least one start codon must be in frame with the rest of the gene, as is seen for n = 6 and for n = 7. However, for n = 5 (or 8, 11, 14 etc.) there is no potential start codon in the correct reading frame, so Lic1A will not be expressed and no phosphocholine will be added to the LPS.

Although a large proportion of the phase variably expressed genes in *Hi* are LPS-related, repeat tracts have also been identified in locations such as the *hgp* (haemoglobin and haemoglobin-haptoglobin binding protein) genes and in the promoter region of the *hif* locus that controls the expression of fimbriae (Power et al., 2009; van Ham et al., 1993). Phase variation of outer membrane components has also been demonstrated in *N. meningitidis*, *H. pylori* and *Campylobacter jejuni*, but in these species it is usually mediated by the slippage of mono- or dinucleotide repeat tracts (Berrington et al., 2002; Hendrixson, 2006; Salaun et al., 2004).

The third unusual feature of *Hi* LPS is that it lacks the long polysaccharide OAg chains that constitute the most variable part of the molecule in many other Gram negative species. At first this was regarded as a key distinction, with some workers referring to the molecule as lipooligosaccharide (LOS), but the boundaries recently became blurred when an *Hi* gene cluster similar to an OAg locus was identified and shown to encode enzymes required for the addition of a single tetrasaccharide unit to the LPS *en bloc* (Hood et al., 2004b). Two different but related unit structures were observed within LPS isolated from a single *Hi* strain (Fig. 1.3); these are known as HMG as they result in high molecular weight LPS glycoforms. Homology searches suggested that the 10 kb *hmg* locus encoded an UndP-sugar phosphotransferase, several glycosyltransferases, an OAg flippase homologue and an OAg ligase homologue, with vestigial OAg polymerase and chain length determinant genes that contain many frameshifts and premature stop codons (Hood et al., 2004b).

The *hmg* locus and the corresponding HMG glycoforms are found in approximately 60% of NTHi strains, but it appears that the tetrasaccharide units are never polymerised into true OAg in *Hi*. Several other mucosal pathogens, including *N. meningitidis*, *Bordetella pertussis* and *C. jejuni*, also lack OAg on their LPS.

1.4.3 The LPS of *Hp*

Two studies have examined *Hp* LPS to date. In 1986, Roberts and colleagues analysed the silver-stained SDS-PAGE profiles of LPS from 25 *Hp* strains isolated from various sites of the body including the throat, urethra, eye and cerebrospinal fluid (Roberts et al., 1986). The LPS of 17 of the strains was found to contain only low molecular weight bands (rough LPS) but the remaining

eight strains produced ladder-like LPS profiles indicative of molecules containing OAgs of varying chain length (smooth LPS). This was the first observation of O-sidechains in *Haemophilus*, which had previously been considered as having a fundamentally different LPS structure to other Gram negative genera. Roberts et al. examined the relatedness of the strains using protein profiling and whole genome hybridisation, and considered the date, geographical location and body site of isolation, but did not find any variable which correlated with the two types of LPS observed.

More recently, Pollard et al. (2008) used mass spectrometry and NMR spectroscopy to determine the structure of the carbohydrate portion of LPS from two COPD isolates of *Hp*. In both strains a single KdoP was found attached to a heptose trisaccharide, with PEtn extending from HepII and Glc extending from HepI; this correlates with the conserved part of *Hi* LPS as shown in bold in Fig. 1.3. The Glc residue on HepI was augmented with a linear two or three sugar extension in each strain, specifically Glc-(β1,4)-DDHep- in *Hp* strain 4201 and Gal-(β1,6)-Glc-(β1,4)-DDHep- in *Hp* strain 4282, where DDHep is D-*glycero*-D-*manno*-heptose; similar extensions are found at this position in some *Hi* isolates. No other outer core substituents were found in the LPS from either *Hp* strain, and no OAg appeared to be present by electrophoretic or structural analyses. The LPS structures of these two *Hp* isolates are fairly simple in comparison to those of typical *Hi* strains or the putative OAg-containing *Hp* isolates studied earlier by Roberts.

Both of these studies considered only isolates from clinical situations, although it was not always proven that *Hp* was the causative agent of disease; in particular, the role of *Hp* in COPD is

debated (Section 1.2.3). The LPS of commensal *Hp* isolates from healthy individuals appears never to have been investigated.

1.5 *Haemophilus* genome sequences and their use in this project

1.5.1 Early *Hi* genome sequencing

Hi strain Rd, a non-encapsulated laboratory strain derived from a serotype d isolate, was the first free-living organism to undergo full genome sequencing (Fleischmann et al., 1995). The genome was found to be 1.83 Mb in length with a GC content of 38.2%. Interestingly, only four to five two-component regulators were identified from the translated sequence, compared to 40 found in a typical *E. coli* strain. These systems allow bacteria to respond to environmental signals by altering the phosphorylation of transcription factors, thereby controlling the expression of defined sets of genes. In particular, *Hi* Rd appears to lack homologues of the *E. coli* sigma-54 RNA polymerase subunit and the regulators that usually interact with it to control many operons (Fleischmann et al., 1995). Genome analysis also revealed 1465 copies of a 9 bp uptake signal sequence, or USS, previously identified for its importance in DNA uptake in *Hi*; these were distributed largely at random across the genome (Smith et al., 1995). In the current project the annotated *Hi* Rd sequence aided our comparison of LPS synthesis genes in *Hi* and *Hp*.

The genomes of 15 other *Hi* isolates have subsequently been sequenced; all of these strains encode the enzymes necessary for the synthesis of lipid A and inner core LPS, but in some strains novel genes containing tetranucleotide repeats were found, expanding the repertoire of putative outer core LPS glycosyltransferases (Hogg et al., 2007; Power et al., 2009).

1.5.2 The *Hp* strain T3T1 genome

The genome of *Hp* strain T3T1 has been sequenced (using an Illumina platform) and assembled by the Wellcome Trust Sanger Institute, Hinxton, Cambridge (EMBL accession number FQ312002; Power et al., manuscript in preparation). This is the first *Hp* strain to be sequenced and was isolated from the throat of a healthy seven year old girl in Gambia in 2001. At 2.09 Mb, the *Hp* T3T1 genome is larger than that of *Hi* Rd and contains approximately 200-300 more predicted genes than a typical *Hi* strain. Although around three quarters of *Hp* T3T1 genes have homologues in *Hi*, there is considerable genomic rearrangement between the two species. The overall GC content, 39.6%, is slightly higher than that of the sequenced *Hi* genomes. The presence of genes encoding the enzymes of a porphyrin biosynthesis pathway, which have no homologues in the *Hi* Rd genome, is consistent with the conclusion of Evans and colleagues (1974) that haemin dependency in *Haemophilus* is a discrete rather than continuous variable. These ten enzymes are thought to allow *Hp* to synthesise heme from glutamate, whereas *Hi* must obtain either heme or its precursor protoporphyrin IX from the host.

Genome analysis also reveals that like *Hi*, *Hp* strain T3T1 lacks a recognisable *de novo* synthesis pathway for NAD and has no homologue of NadV, which some Pasteurellaceae species use to convert nicotinamide into NAD. This explains its dependence upon an exogenous source of NAD, which is a coenzyme and substrate for many essential cellular reactions. The NAD utilisation pathway of *Hi* has been characterised (Andersen et al., 2003; Herbert et al., 2003; Kemmer et al., 2001). NAD and nicotinamide mononucleotide (NMN) can both be taken up through the OmpP2 porin in the outer membrane; the substrate is then converted to nicotinamide

riboside (NR) by a periplasmic enzyme, NadN. NR can also be taken up directly but the mechanism for this is not clear. NR is then translocated through the inner membrane by PnuC, and is converted (back) to NAD in the cytoplasm by NadR. An additional outer membrane protein, P4, can convert NMN (and possibly NAD) into NR on the cell surface. The *Hp* T3T1 genome encodes close homologues of P4, PnuC and NadR, but its two putative porins (Hp1965 and Hp0379) share only 43% and 26% amino acid identity respectively with *Hi* OmpP2 and there is no homologue of NadN. This suggests that *Hp* must either obtain NR directly from the host or rely on P4 to convert NAD or NMN to NR before it enters the cell. Cynamon et al. (1988) demonstrated that like *Hi*, *Hp* can grow on media containing either NAD, NMN or NR; however, Evans et al. (1974) observed that *Hp* requires higher NAD concentrations than *Hi* for optimal growth in vitro. Subtly different factor V requirements between the two species could play a role in their ability to colonise various niches within the host.

Throughout this thesis, *Hp* T3T1 genes are referred to in their original format "*Hpxxxx*" (e.g. *Hp0277*) and the proteins they encode in the format "Hpxxxx". The Sanger Institute recently made the *Hp* T3T1 genome sequence publicly available through the NCBI database, and have renamed the genes and proteins as "PARA_xxxxx". The original and new gene numbers do not correspond (e.g. *Hp0277* has been renamed as PARA_02790) so a conversion table for the genes detailed in the thesis is included as Appendix A.

We used a BLAST algorithm (http://blast.ncbi.nlm.nih.gov) to compare *Hp* T3T1 with *Hi* in terms of its genetic capacity to produce factors that may influence colonisation and virulence. As might be expected from the results reviewed in Section 1.3.4, the *Hp* T3T1 genome did not

contain a homologue of the *Hi* IgA protease gene. It also lacked any homologues of the *Hi* capsule locus genes and the Hib haemocin synthesis genes. In terms of surface proteins, *Hp* T3T1 encodes homologues of the four peptides PilA-PilD that are required for the assembly of a type IV pilus in *Hi* (49-76% amino acid identity). However, it appears to synthesise a different set of adhesins to *Hi*. The *Hp* T3T1 gene with the closest identity to each of the *Hi* adhesin genes *hmw1a*, *hmw2a*, *hia* and *hsf* is a 6.7 kb ORF designated *Hp0861*. There are no BLAST matches for the fifth *Hi* adhesin gene, *hap*. Besides *Hp0861*, other *Hp* T3T1 genes that encode YadA domains (a characteristic of adhesins) include *Hp0077*, *Hp1654* and *Hp1668*.

Our initial BLAST analyses also revealed striking similarities and differences in the complement of LPS biosynthesis genes present in *Hp* T3T1 and *Hi*. These findings provided the basis for examining the biosynthesis and structure of this important virulence factor in a panel of *Hp* isolates, and investigating its role in bacterial-host interactions.

In the last few years, genome sequences have become available via the NCBI database for individual strains of several more Pasteurellaceae species; these include *Actinobacillus pleuropneumoniae*, *Aggregatibacter actinomycetemcomitans*, *Aggregatibacter aphrophilus*, *Haemophilus ducreyi*, *Histophilus* (formerly *Haemophilus*) *somni*, *Mannheimia haemolytica* and *Pasteurella multocida*. This allows us to compare the genetics of LPS biosynthesis in *Hp* with that in other members of the same family and/or genus, some of which produce OAg. However, no genome sequences have been published for *H. pittmaniae*, which is the only species to cluster with *H. parainfluenzae* (*Hp*) using 16S rRNA analysis (Norskov-Lauritsen et al., 2005).

1.5.3 High throughput *Haemophilus* genome sequencing

In a recent collaboration between the Sanger Institute and our laboratory, genomic DNA was prepared from 96 *Haemophilus* strains and sequenced in a pilot project to test a new high throughput Illumina sequencing facility. Although most of the chosen strains were encapsulated or non-typeable *Hi*, 12 commensal *Hp* isolates were also included. Upon contig assembly it was found that for all genomes, areas with low % GC content had extremely poor coverage; this was attributed to unforseen problems with the Illumina reagents used at the Sanger Institute. However, short contigs were assembled for the regions with higher % GC content and a BLAST server was set up to allow homology searches against DNA and translated DNA sequences from each strain (Peter Power, pers. comm.). This resource (known as 'Genomophilus') was available during the last 12 months of the research presented in this thesis and is referred to throughout the text.

During preparation of this manuscript, the data from a second high-throughput *Haemophilus* genome sequencing project became available. The Crook Laboratory (Nuffield Department of Clinical Laboratory Sciences, John Radcliffe Hospital, Oxford) used the Illumina platform at the Wellcome Trust Centre for Human Genetics, University of Oxford to obtain genome sequences for 96 strains, overlapping with the strain collection used for Genomophilus. The data was assembled into large contigs, some spanning over 100 kb and containing entire OAg or HMG loci. The Crook group kindly provided us with access to this data set (Tanya Golubchik, pers. comm.) but unfortunately its value is limited due to the incorrect assignment of species and strain names to each genome sequence. Nevertheless the data was of use on a few occasions and is referred to in the text as the Crook group sequencing project.

1.6 *Hp* strains used in this study

The purpose of this project is to survey the genetics, structure and properties of cell surface LPS in a range of commensal *Hp* isolates. The 20 strains that were chosen for this study had been isolated from the throats of healthy children in 2001 in either Oxfordshire (UK) or Kulari (Gambia) and were kindly made available by Derrick Crook at the University of Oxford. For convenience we have designated each strain a number: for example, *Hp* strain O/OM30/1/A1 is referred to as *Hp* strain 2 in this thesis. Strain names, numbers and place of isolation are given in Table 2.1. The study panel includes the genome sequence strain, *Hp* T3T1, which is also referred to as *Hp* strain 26.

The 20 study strains are part of a larger collection of isolates that were originally classified as *Hp* by the Crook Laboratory using the standard diagnostic criteria, including growth on appropriate media, NAD dependence, haemin independence and colonial morphology. Later, 16S rRNA analysis and multi-locus sequence typing (MLST) were performed to check the species classification and the relatedness of the strains; most of the isolates were confirmed as *Hp* by these methods, but a number were reclassified either as *Hi* or as 'hybrid' strains that were phenotypically *Hp*-like but had *Hi*-like 16S rRNA sequences (Abdel Elamin, pers. comm.). The 20 selected study strains comprise 18 true *Hp* isolates as judged by phenotypic characterisation and by 16S rRNA analysis, plus two of the hybrid strains (numbered *Hp* 6 and 11) for comparison. Most of the study strains are distinct from each other using MLST analysis; there are no consistent branches between the dendrograms for the six housekeeping gene sequences used for MLST, suggesting that horizontal DNA transfer may be common in *Hp* (data not

shown). At least three *Hp* biotypes (I, II and III) are represented in the study panel. *Hp* strains 2, 8, 15, 18, 20 and 22 and the hybrid strain *Hp* 6 were chosen for inclusion in the Genomophilus sequencing project.

The enzymes required for capsule synthesis in *Hi* are encoded by the *cap* gene cluster, which comprises the six genes *bexA-D*, *hcsA* and *hcsB* plus a number of capsule type-specific ORFs such as glycosyltransferase genes (Hoiseth et al., 1986; Satola et al., 2003). BLAST searches against the *Hp* data in Genomophilus revealed homologues of all six of the non-variable capsule genes within the contigs that were available for *Hp* strain 18, and PCR analysis confirmed the presence of *bexD* in this strain (data not shown). However, the genes which flank the *cap* locus in *Hi* and in *Hp* strain 18 were found to be adjacent in the other 19 *Hp* study strains by PCR analysis, and *bexD* could not be amplified from these strains. These results suggest that only one of the 20 *Hp* study strains, namely strain 18, may have the capacity to synthesise a polysaccharide capsule using an *Hi*-like *cap* locus.

After selecting the study strains, each was grown on BHI agar, resuspended in PBS and examined using phase contrast microscopy. Bacterial morphology was strain-dependent and ranged from short coccobacilli to long bacilli (Appendix B). Many strains formed clumps that were difficult to disperse. Some strains, including *Hp* T3T1, formed long filaments during log phase in BHI broth culture. In general, *Hp* appeared to show much greater pleomorphism than *Hi*, which is usually coccobacillary and easily disperses into single bacteria in PBS. Together these observations suggest that the two species are likely to differ in their growth characteristics

and cell surface properties, e.g. adhesion, which could influence their ability to survive and replicate in particular niches within the human host.

1.7 Collaborative structural analysis: mass spectrometry and NMR

Although genetic analysis, electrophoretic LPS profiling and and immunoblotting can provide information regarding the potential LPS epitopes in a certain *Haemophilus* strain, it is important to corroborate this evidence with structural data. The LPS structures of several of the *Hp* strains used in the current study were elucidated using mass spectrometry and nuclear magnetic resonance techniques by the laboratory of Elke Schweda at the Karolinska Institute, Huddinge, Sweden, as part of a long-standing collaboration between our research groups. In addition to identifying each inner core, outer core and OAg component of the LPS, these methods allow the position, configuration and linkage of each to be defined.

The first of the LPS structures to be examined was that of the genome sequence strain, T3T1. Later, after preliminary work on the study collection, cell pellets were prepared from a few *Hp* strains (*Hp* 13, 16, 20, 22 and 30) and sent to the Schweda Laboratory for analysis. These were chosen to represent the diversity of electrophoretic profiles and OAg locus sequences that had been observed by that stage of the project. Samples from three *Hp* mutant strains were also sent for structural LPS analysis at a later stage to confirm that the deletion or addition of a gene had led to the expected change in LPS structure.

1.8 Aims of the project

Lipopolysaccharide is a key virulence factor for Gram negative bacteria, and the structure and synthesis of *Hi* LPS have been studied in great detail. However, very little is known about the LPS of its close relative, *Hp*. This study aims to survey a panel of 20 commensal *Hp* strains to determine the similarities and differences between LPS in the two species and the level of variation between *Hp* isolates. Firstly, as the genes required for *Hi* LPS biosynthesis are well characterised, we will begin by investigating the distribution of homologues of these genes in *Hp* to assess which regions of the molecule are likely to be conserved. Secondly, this project will endeavour to examine whether some or all of the panel strains have the genetic capacity to synthesise an OAg; and if so, whether such a molecule is made under laboratory conditions and affects bacterial interactions with host cells and molecules. Thirdly, we will evaluate the mechanisms used by *Hi* and *Hp* to vary cell surface molecules such as LPS in order to adapt quickly to a changing environment. As *Hp* occasionally causes disease, it is worthy of study in its own right, but the main purpose of this study is to highlight any fundamental differences between the LPS structures of *Hi* and *Hp* which could relate to their different levels of pathogenicity.

2 Materials and Methods

2.1 *Haemophilus* strains, culture and storage

Hp strains used in this study (see Section 1.6) are detailed in Table 2.1. *Hi* strains used for comparison purposes are from the Richard Moxon culture collection.

Hi strains were grown in brain heart infusion broth (BHI) (Merck) supplemented with 10 μg/ml haemin and 2 μg/ml NAD, and incubated at 37 ºC for 16 h shaking at 200 rpm. For *Hp* the haemin was omitted, except in experiments where *Hi* and *Hp* growth was being compared directly. For growth on solid medium, strains were plated on BHI agar (1%) supplemented with 10% Levinthals base (McLinn et al., 1970) and 66 μg/ml NAD and incubated at 37 ºC for up to 24 h.

Frozen stocks were prepared by mixing 750 μl log phase *Haemophilus* culture and 250 μl 80% (vol/vol) glycerol, and were stored at -80 ºC.

2.2 *Haemophilus* genomic DNA (gDNA) extraction

Bacteria from 3 ml log phase culture were pelleted by centrifuging at 13,000 g for 2 min then washed in PBS and resuspended in 200 μl TNE (100 mM NaCl, 10 mM Tris pH 8, 10 mM EDTA). SDS was added to 1%. Cells were lysed at 65 ºC for 10 min then treated with proteinase K (500 μg/ml) at 37 ºC for 2 h. The sample was then mixed with 1 vol phenol and centrifuged at 13,000 g for 5 min; the top layer (containing DNA) was taken into a fresh tube and mixed with 1

vol phenol/chloroform/isoamyl alcohol (25:24:1). After centrifugation (13,000 g for 5 min) the top layer was again taken into a fresh tube and the DNA was precipitated with 2 vol ethanol and 0.1 vol 3 M NaAc. The DNA was pelleted by centrifugation (13,000 g for 10 min) and washed in 70% ethanol, then air dried and resuspended in 200 µl TE buffer (10 mM Tris pH 8, 1 mM EDTA) with 50 µg/ml RNase.

2.3 Standard polymerase chain reaction (PCR)

For expected product sizes of up to 6 kb, 50 µl PCRs were prepared using 1 U Taq DNA Polymerase (Invitrogen). Each reaction also included 1X PCR Buffer (Invitrogen), approximately 40 ng template gDNA, 0.4 µM each primer (Sigma), 0.4 mM each dNTP and 2.5 mM $MgCl_2$. DNA was amplified for 30 cycles comprising 1 min each of denaturation (94 ºC), annealing (50 ºC) and extension (72 ºC); for expected products of >1.5 kb the extension time was increased to 3 min. PCR products were electrophoresed on 0.8% agarose gels containing 0.5 µg/ml ethidium bromide at 100 V for approximately 1 h, and visualised under ultraviolet light. Primers used for PCR analysis are listed in Tables 2.2, 2.3 and 2.4.

2.4 DNA restriction digestion

Restriction enzymes were obtained from Roche and New England Biolabs. All restriction digestions were performed following the manufacturer's instructions in the appropriate 1X buffer and incubated for 2 h at the recommended temperature (usually 37 ºC). Digestion products were analysed by agarose gel electrophoresis where necessary.

2.5 Long range PCR (LR-PCR) and digests

50 µl LR-PCRs were performed using the Expand long range PCR kit (Roche) following the manufacturer's instructions. OAg loci were amplified from *Hp* gDNA using primers 286a (5'-GAGACTGCGGTAGTCGATCC-3') and 269a (5'-CCATCACTTGGTTTGATGCT-3'), which are specific for the locus-flanking genes *glnA* and *pepB* respectively. An extension time of 15 min in cycle 1, rising to 22 min by cycle 30, was found to be sufficient for the amplification of products of up to 20 kb. 5 µl of each LR-PCR product was digested with MfeI. LR-PCR products and digests were run on 0.7% agarose gels at 20 V for 36 h, and were stained and visualised as in Section 2.3.

2.6 DNA sequencing

Sequencing of plasmids and PCR products was carried out by the Weatherall Institute of Molecular Medicine Sequencing Service, John Radcliffe Hospital, Oxford, using an ABI-3730 DNA analyser with BigDye Terminator v3.1 (Applied Biosystems). Electropherograms were viewed using Chromas 1.62 software (Technelysium Pty Ltd, Queensland).

2.7 Southern analysis

MfeI-restricted gDNA was fractionated on a 0.8% agarose gel and transferred to a nylon membrane (Hybond N+, GE Healthcare) by Southern blotting (Sambrook et al., 1989).

For the analysis of LPS genes, internal gene fragments (approx. 600 bp) were amplified from *Hi* or *Hp* gDNA using standard PCR and purified using a QIAEX II gel extraction kit (Qiagen). Blots were hybridised with horseradish peroxidase (HRP) labelled PCR products for 16 h at 42 ºC, washed stringently at 42 ºC according to the manufacturer's instructions, and visualised on Hyperfilm by HRP-catalysed luminol production (ECL Direct kit, GE Healthcare).

For the analysis of tetranucleotide repeats, radiolabelled probes were used. 10 pmol of a 20-base oligonucleotide (Sigma) was phosphorylated with [γ-^{32}P]-dATP (10 µCi; Perkin Elmer) using 10 U bacteriophage T4 polynucleotide kinase (Roche). The labelled probe was purified through a NAP-5 column (Amersham). The membrane was incubated in 15 ml hybridisation buffer (6X SSC, 0.01M EDTA, 5X Denhardt's reagent (5X = 0.1% wt/vol each of Ficoll, polyvinylpyrrolidone and bovine serum albumin), 100 µg/ml salmon sperm DNA, 0.5% SDS) for 4 h at 45 ºC; the probe was then added and the incubation was continued for a further 16 h to allow hybridisation. Two wash steps were performed (each for 15 min at 45 ºC): firstly in 6X SSC, 0.2% SDS and secondly in 2X SSC, 0.2% SDS. Hybridised blots and photographic film were placed together at -80 ºC in a radiographic cassette containing an intensifying screen, and the film was developed after 20-24 h.

2.8 Reverse transcriptase PCR

1 ml of a mid-log phase *Hp* broth culture in BHI (OD_{600} = 0.5 for Fig. 7.6 and OD_{600} = 0.8 for Fig. 7.8) was harvested into two volumes of RNA Protect reagent (Qiagen). After centrifugation the bacterial pellet was resuspended in 100 µl TE buffer, and RNA was isolated using the SV Total RNA Isolation System (Promega) following the manufacturer's instructions, which include

DNase treatment. The RNA concentration was measured using a NanoDrop spectrophotometer (Thermo Scientific). 1 µg total RNA was reverse transcribed to make complementary DNA (cDNA) using the SuperScript III First-Strand Synthesis System (Invitrogen) with random primers. 1.5 µl of the cDNA reaction (from a total volume of 70 µl) was used as the template in standard PCR using Taq polymerase (Section 2.3). Primers designed to a housekeeping gene, *atpG*, were used in an additional PCR to control for any differences in total RNA or cDNA concentrations; these were atpGup (5'-ATGGCAGGTGCAAAAGAGAT-3') and AtpRS (5'-CGTGCTCGTAAAGRTAATCCC-3').

2.9 Tricine SDS-PAGE LPS gels

Colonies grown on BHI agar were harvested and resuspended in PBS. A 1:20 dilution was prepared in 1% SDS, 0.1 M NaOH to measure the optical density at 260 nm, and the resulting calculations were used to adjust the original PBS suspension to a concentration appropriate for electrophoresis (equivalent to OD_{260} = 1.0, which is approximately 10^9 cells/ml). The standardised suspensions were then diluted 1:1 in 2x dissociation buffer (125 mM Tris pH 6.8, 20% glycerol, 4% SDS, 10% mercaptoethanol, 0.004% bromophenol blue). Following proteinase K treatment (50 µg/ml) at 60 ºC for 3 h and denaturation at 100 ºC for 5 min, 30 µl samples were fractionated on tricine SDS-PAGE gels (Lesse et al., 1990). Each gel consisted of approximately 15 ml separating gel (16.6% acrylamide, 1 M Tris, 0.1% SDS, 10.4% glycerol, 0.03% TEMED, 0.03% APS) topped with 6 ml stacking gel (4.0% acrylamide, 0.73 M Tris, 0.073% SDS, 0.12% TEMED, 0.12% APS). Gels were run for 16 h at 30 mA and 6 ºC in a tank containing 0.2 M Tris pH 8.9 in the main tank (anode) and 0.1 M Tris, 0.1 M tricine and 0.1% SDS, pH 8.25 in the top

reservoir (cathode). LPS was visualised by staining with silver (GE Healthcare) following the manufacturer's instructions.

Bacterial samples to be neuraminidase treated to test for the presence of sialic acid in the LPS were suspended in PBS pH 5.8 and adjusted to OD_{260} = 1.0 as above. 15 µl of the adjusted suspension was incubated with 1.5 µl neuraminidase from *Clostridium perfringens* (4U/µl; Boehringer Mannheim) for 24 h at 37 ºC. An untreated control was also set up using 15 µl lysate and 1.5 µl PBS pH 5.8. Samples were then diluted in dissociation buffer, proteinase K treated and fractionated using the protocol above.

2.10 Construction of plasmids in *E. coli*

2.10.1 Cloning of DNA fragments

The target region was amplified from *Haemophilus* gDNA by standard PCR (Section 2.3) and cloned in *E. coli* using the pSCA or pSCA-amp-kan vector system (Stratagene) following manufacturer's guidelines for ligation and transformation. Correct transformants were selected by blue/white screening on LB agar (1% tryptone, 0.5% yeast extract, 0.5% NaCl, 1% agar) + 100 µg/ml ampicillin with 20 µl 4% X-Gal (5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside) spread on each plate. Colonies in which the *lacZ* gene on the plasmid is disrupted by an insert are white due to the lack of X-Gal hydrolysis.

2.10.2 Isolation of plasmid DNA by alkaline lysis

Single white colonies were picked into LB broth with 100 µg/ml ampicillin and incubated for 16 h at 37 ºC with shaking. 3 ml of the culture was centrifuged at 13,000 g for 2 min; the pelleted bacteria were resuspended in GTE (50 mM glucose, 10 mM EDTA, 25 mM Tris pH 8) and lysed by the addition of NaOH to 0.11 N and SDS to 0.6%. Following the addition of 0.4 vol potassium acetate (original concentration 3 M wrt K, 5 M wrt Ac) and incubation on ice for 5 min, the precipitated proteins and gDNA were pelleted by centrifugation (5 min at 13,000 g) and discarded. 2 vol ethanol was added to the supernatant and the sample was left on ice for 5 min to precipitate the plasmid DNA. After further centrifugation (5 min at 13,000 g) the plasmid DNA pellet was washed in 70% ethanol, air-dried and resuspended in 50 µl TE buffer containing 50 µg/ml RNase. The presence and orientation of the plasmid insert was determined using appropriate restriction digests.

2.10.3 Insertion of drug resistance cassette

Drug resistance cassettes (and uptake signal sequences (USS), where present) were taken from plasmids pUC4KanUSS, pUC4HinUSS and pER2. The pUC4KanUSS plasmid is based on pUC4Kan (GE Healthcare) but contains two USS after the kanamycin resistance gene. pUC4HinUSS is also based on pUC4Kan, but the kanamycin resistance gene has been replaced with a tetracycline resistance cassette (*tetR* and *tetA*) flanked by a USS on each side. pER2 contains the *ermC* gene, which confers erythromycin resistance. These plasmids were obtained from Mary Deadman.

The desired drug resistance cassette was isolated using appropriate restriction digestion (Section 2.4), excision of the correct band from an agarose gel, and purification using a QIAEX II gel extraction kit (Qiagen). Meanwhile, the plasmid containing the insert was cut with a restriction enzyme specific for a sequence in the 5' half of the reading frame to be disrupted, and was treated with antarctic phosphatase (NEB) following the manufacturer's instructions in order to prevent recircularisation. The restriction enzymes were chosen so that the vector and the drug resistance cassette fragment had blunt ends or compatible sticky ends. The cut vector and drug resistance cassette were ligated in a 1:3 molar ratio using T4 DNA ligase (Roche) in 1X ligase buffer at 16 ºC for 24 h, in a total volume of 10 µl; the ligase was then inactivated at 65 ºC for 10 min.

E. coli DH5α cells were made competent by growing to an OD_{600} of 0.4, cooling on ice for 30 min, washing in 0.1 M $MgCl_2$ and 0.1 M $CaCl_2$ sequentially at 4 ºC, resuspending in 0.1 M $CaCl_2$ to 1/20 of the original culture volume, and leaving on ice for a further 30 min. For the transformation, the 10 µl ligation reaction was added to 200 µl competent cells, and after a further 40 min on ice the cells were heatshocked (2 min at 42 ºC) and allowed to recover in 1 ml LB for 1 h at 37 ºC with shaking. Transformants were selected on LB agar with the appropriate antibiotic (12 µg/ml tetracycline, 50 µg/ml kanamycin or 300 µg/ml erythromycin) and incubated at 37 ºC for 24 h. Colonies were picked and plasmids extracted and checked as above (Section 2.10.2). The completed plasmid constructs used in the project are described in Table 2.5, and the primers used in their construction are listed in Table 2.6.

2.11 Transformation of *Hp*

2.11.1 Static aerobic method

Hp cells were made competent using a method based on that of Gromkova and Goodgal (1979). Briefly, 10 ml BHI broth was inoculated with a 1 µl loopful of cells from a BHI agar plate and incubated in a petri dish at 37 ºC for 18 h without shaking. The bacterial film was resuspended and 100 µl cells were added to a 1.5 ml tube containing 500 µl BHI, 2 µg/ml NAD, 20 mM $MgSO_4$ and 0.5 µg plasmid DNA that had been linearised with KpnI (DNA concentrations were estimated from an agarose gel). After shaking at 37 ºC for 5 h, the transformation mixture was plated onto BHI with appropriate antibiotic selection (2 µg/ml tetracycline, 15 µg/ml kanamycin or 20 µg/ml erythromycin) and incubated at 37 ºC for 24 h. Putative transformant colonies were picked and grown up as BHI broth cultures, which were used for freezer stock preparation and gDNA extraction. The mutant genotype was checked using PCR analysis, and the strain background was confirmed by comparing the mutant's gDNA MfeI restriction pattern to that of the parent strain. Details of the mutants are given in Table 2.7.

2.11.2 Electroporation method

Hp T3T1 was transformed with pRY12 using a protocol adapted from that of Mason et al. (2003). 50 ml BHI broth inoculated with an overnight BHI broth culture (3 ml) to give a starting OD_{600} measurement of 0.10 was incubated at 37 ºC with shaking. When the culture had reached an OD_{600} of 0.35 (150 min) it was chilled on ice for 30 min. All further steps were carried out at 4 ºC. Cells were pelleted for 10 min at 4200 g and washed 3 times with 0.5X SG (1X = 15% glycerol, 272 mM sucrose, pH 7.4) to increase their competence. After the final centrifugation,

cells were resuspended in 500 µl 1X SG. 40 µl competent cells were mixed with either 1 or 2 µg circular plasmid DNA and were subjected to electroporation at 2.5 kV, 200 Ω and 25 µF (BioRad Gene Pulser), with recovery in 1 ml BHI for 90 min at 37 ºC with shaking. Transformations were plated on BHI agar with 15 µg/ml kanamycin and incubated at 37 ºC for 24 h. Putative transformants were checked as above (Section 2.11.1).

2.12 Cloning and sequencing OAg loci

OAg loci were amplified from *Hp* strains 13, 17, 20 and 30 using LR-PCR (Section 2.5). Each 10-20 kb product was digested using EcoRV and HaeIII in separate reactions then cleaned by ethanol precipitation and dissolved in distilled H_2O. The digested fragments were ligated to HincII-cut, phosphatase-treated pBluescript (Stratagene) with T4 DNA ligase (Roche) and cloned in *E. coli* DH5α. Colonies were selected on LB agar + 100 µg/ml ampicillin with X-Gal to allow blue/white screening of inserts (see Section 2.10.1). Successful clones were picked into 1 ml LB + 100 µg/ml ampicillin and grown for plasmid extraction. Double digests using EcoRI and XhoI allowed clones to be categorised according to their insert size; one clone containing each EcoRV or HaeIII restriction fragment was then sequenced using primers M13-for-20 (5'-GTAAAACGACGGCCAGT-3') and M13-rev-24 (5'-AACAGCTATGACCATG-3') which bind to the pBluescript part of each construct and read into the insert region. The overlapping EcoRV and HaeIII fragment sequences were assembled into contigs using Vector NTI ContigExpress (Invitrogen), with additional PCR and sequencing across the gaps between contigs enabling full assembly of the OAg loci. For completion of the *Hp* strain 30 locus sequence, a second round of cloning was carried out using XmnI and SspI fragments, which were then sequenced as above.

2.13 Resistance to the bactericidal effects of human sera

The survival of *Haemophilus* strains in human sera pooled from 15-18 donors was analysed using a method similar to that of Hood et al. (1999). Colonies grown on BHI agar were suspended in PBS-BG (PBS + 0.1% glucose (wt/vol) + 0.05 mM $MgCl_2$ + 0.09 mM $CaCl_2$), and a 1:20 dilution was prepared in 1% SDS, 0.1 M NaOH to measure the optical density at 260 nm. The starting suspension was adjusted to the equivalent of OD_{260} = 0.8 then serially diluted to a final concentration equivalent to OD_{260} = $4x10^{-5}$ (i.e. 1/20,000 dilution). 20% pooled human serum (PHS) in PBS-BG was serially diluted 1:1 across seven wells of a flat bottomed polystyrene 96 well plate so that each well contained 50 µl of 20% to 0.32% PHS; the eighth well contained 50 µl of 20% PHS that had been decomplemented by heating at 56 ºC for 30 min. 50 µl of the cell suspension (approximately 2000 bacteria) was added to each well, giving final PHS concentrations of 10% to 0.16%. After 1 h at 37 ºC, 25 µl from each well was spread on BHI agar and incubated for 24 h at 37 ºC. The resulting colonies were counted to determine the level of bacterial survival in each concentration of PHS. Survival was plotted as a percentage of that in the decomplemented PHS control sample.

2.14 Epithelial cell association assay

The ability of *Hp* to adhere to an SV40-transformed human bronchial epithelial cell line, 16HBE14o$^-$, was studied using a protocol similar to that of Hood et al. (1999). This cell line has previously been used to study other respiratory tract bacteria. 16HBE14o$^-$ cells were grown to a confluent monolayer in a flat bottomed polystyrene 96 well plate and washed three times with

Hank's balanced salt solution (HBSS; Gibco). 50 µl of 4% decomplemented PHS in Dulbecco's modified Eagle medium (DMEM; Gibco) was added to each well. Bacteria grown on BHI agar were resuspended in 1.5 ml PBS-BG and left to settle for 2 min before the top 1 ml was taken into a fresh tube; this avoids large clumps of cells. A suspension equivalent to OD_{260} = 0.8 was made using the method outlined in Section 2.13, except that the cells were diluted with DMEM. 50 µl of the bacterial suspension was added to three monolayer wells and to three control wells containing no epithelial cells, and the plate was incubated for 2.5 h at 37 ºC in the presence of 5% CO_2. To measure association, all wells were washed with HBSS three times to remove non-adherent bacteria. 1% saponin was added to release the remaining bacteria (10 min at 37 ºC), which were then serially diluted and plated on BHI agar. Colonies were counted after overnight incubation. Association was plotted as a percentage of the total number of bacteria present in a control monolayer well after the 2.5 h incubation period (i.e. adhered plus non-adhered bacteria).

To measure bacterial uptake by the epithelial cells, monolayers were incubated with the bacteria as above, then the medium was replaced with 250 µl of 200 µg/ml gentamicin. After a further 1.5 h at 37 ºC the cells were washed, treated with saponin and plated as before. Gentamicin kills any bacteria exposed on the surface of the epithelial monolayer (this was confirmed by a sensitivity test) whilst internalised bacteria are protected from the antibiotic.

2.15 Assay to detect the inhibition of *Hp* growth by *Hi* and vice versa

The following method, based on that of LiPuma et al. (1990), was used primarily to examine whether *Hp* is sensitive to the Hib bacteriocin (haemocin) and whether *Hp* produces any substances that kill *Hi*. *Haemophilus* strains were grown on BHI agar and the strain to be tested

for sensitivity was resuspended in BHI broth to an OD_{650} of $1.5x10^{-3}$. 60 µl of the suspension (approx. 10^5 colony forming units) was spread evenly across a BHI agar plate and left to dry for 10 min. The strains to be tested for bacteriocin/toxin production were resuspended in BHI broth to an OD_{650} of 0.35 and a 3 µl spot was pipetted onto the bacterial lawn. Plates were incubated for 24 h at 37 ºC and examined for the presence of a zone of inhibition around each test spot.

2.16 Detection of phosphocholine and digalactoside by immunoblot

2.16.1 Colony immunoblots

Hp and *Hi* strains diluted from overnight broth cultures were spread onto BHI agar to give approximately 750 colony forming units per plate. Following incubation at 37 ºC for 24 h, the colonies were lifted onto a PROTRAN nitrocellulose blotting membrane (Whatman). After air-drying for 15 min, the membrane was covered with PBST-BSA (1X PBS, 0.05% Tween-20, 2% bovine serum albumin) for 1 h with shaking to wash off excess cells and block non-specific antibody binding. The membrane was incubated with the PCho-specific monoclonal antibody TEPC15 (Sigma; 1:2000 dilution in PBST-BSA) for 2 h then with Anti-mouse IgA alkaline phosphatase antibody produced in goat (Sigma; 1:20,000 dilution in PBST-BSA) for 1 h. Blots were developed in BCIP/NBT Substrate Solution (Perkin Elmer) for 10 min, which stained phosphatase-containing sites purple. The whole procedure was carried out at room temperature with 3 x 2 min washes in PBST (1X PBS, 0.05% Tween-20) between each step.

To isolate phase variants, the original BHI plates were reincubated for 24 h at 37 ºC and the appropriate colonies were picked into 3 ml BHI broth. Freezer stocks and gDNA were then prepared as above (Sections 2.1 and 2.2).

2.16.2 Dot immunoblots

Bacteria grown on BHI agar were suspended in PBS, adjusted to OD_{260} = 1 as described in Section 2.9 and boiled for 5 min to lyse. 3 µl of each suspension was pipetted onto a gridded PROTRAN membrane. PCho detection was then carried out exactly as for colony immunoblots. For the detection of digalactoside, the same protocol was followed but using the monoclonal antibody 4C4 (1:200 dilution in PBST-BSA; provided by E.J. Hansen, University of Texas) instead of TEPC15. 4C4 recognises digalactoside epitopes extending from HepI or HepII of the LPS.

2.17 Preparation of cell pellets for LPS structural analysis

Ten 14 cm BHI agar plates were inoculated with an overnight BHI broth culture of *Hp* and incubated at 37 ºC for 24 h. The resulting confluent growth was suspended in 50 ml PBS containing 0.5% phenol and incubated at room temperature for 30 min. Cells were pelleted at 7000 g for 10 min at 4 ºC. Cell pellets were resuspended sequentially in ethanol, acetone and ether with centrifugation between each step (conditions as above), then freeze-dried in a vacuum dessicator for 2 h. Dry cell pellets were sent to the laboratory of Elke Schweda, Karolinska Institute, Huddinge, Sweden where the LPS was isolated using a hot phenol method and structurally analysed by mass spectrometry and nuclear magnetic resonance techniques.

2.18 Bioinformatic analysis

The Artemis genome browser and annotation tool (Rutherford et al., 2000) was used to examine *Haemophilus* genomes and OAg locus sequences. Sequence homology analysis was performed using the NCBI basic local alignment search tool (BLAST) (Altschul et al., 1997) with the default algorithm parameters. Conserved protein domains were detected using CDD (Marchler-Bauer et al., 2009). DNA and protein sequences were aligned with ClustalW (www.ebi.ac.uk/clustalw) and presented using GENEDOC 2.7 (www.nrbsc.org/gfx/genedoc); in all alignments, darker shading indicates higher sequence conservation and the numbers above the alignment show the base or amino acid position. Analysis for the neighbour-joining tree in Fig. 5.3 was conducted in MEGA5 (Tamura et al., 2007); the evolutionary distances were computed using the Poisson correction method. Transmembrane domains within proteins were predicted using the TMHMM V2.0 tool at www.cbs.dtu.dk/services/TMHMM. Simple DNA sequence repeats and inverted repeats were identified using the programs msatfinder (www.genomics.ceh.ac.uk/cgi-bin/msatfinder/msatfinder.cgi) and einverted (http://emboss.bioinformatics.nl) respectively. Plasmid maps were drawn in Vector NTI (Invitrogen). Graphs and charts were produced using Microsoft Office Excel.

Table 2.1. *Hp* strains used in this study. Each strain was isolated from the throat of a healthy child in Oxfordshire, UK ('UK') or Kulari, Gambia ('G'). The age of the child, where known, is given in years. All strains have been classified as *Hp* using both phenotypic and genotypic methods, apart from strains 6 and 11 which are classified as 'hybrid' *Hi/Hp* (hyb) (see Section 1.6).

Strain number	Strain name	Origin	Age of child
2	O/OM30/1/A1	UK	-
8	MO/009/6/S	UK	-
10	O/OMA1/2/B4	UK	-
13	T5T1	G	6
14	O/OMD4/1/A1	UK	-
15	T6A1	G	4
16	O/OM41/1/A4	UK	-
17	G/T33T2	G	-
18	T7B2	G	7
19	T7A1	G	7
20	MO/009/1/L	UK	-
22	MO/009/4	UK	-
24	T14A1	G	6
26	**T3T1 (genome)**	G	7
30	T1A1	G	4
31	T23A1	G	7
34	T16B1	G	4
35	MO/099/61L	UK	-
6 (hyb)	O/OM33/2B3	UK	-
11 (hyb)	MO/099/5/S	UK	-

Table 2.2. Primers used to detect inner and outer core LPS synthesis genes. The target gene for each primer is listed in Table 3.3, except for the last eight primers, whose targets are explained in Sections 3.2, 3.3 or 7.1. Primer sequences are given in the 5' to 3' orientation.

Primer	Sequence
468a	TCACGTTTACAACGGTCACG
468b	TCAATTAAATCCGCCACAGG
255a	TTGAATTCCAACAAGATAATC
255b	CTTATCGAAATCAATGAGCC
253a	AGGCAATTCGCTTGAAGG
253b	ACGGCATTTCGTGCTTCG
383a	GATGATACCCTTCCGAGC
383b	ATGCCGCTCGGACATGGC
841a	TGACCGCACTTGTGCCGA
841b	CAATGGTCCAGTAGAGCC
380a	ATACTTTGCCGGCATTGAC
380b	CCAATGCTTTTCATCTCTTG
469a	ATGCCAACAATTAGCGTTGC
469b	GGCATGACTGACACCTTGC
ksgArev	GCATATCTTGAATCACATC

infrev	CCACGTGACCGTTTTCCAAC
6018D	GCACAATTATGTTATCAG
6018E	TTTTACTCGTCCAAATTC
HpLpsA-F	GTTAGAAGTCGATTATTTGCC
HpLpsA-R	CGTTTTACTCTCATAAATTCG
lex5'	GAGTGACTTAAATAGTAC
lex3'	ATCTCGCCATCATCTTC
933b	GTAAGGGCTGATGGCGTAAC
934b2	TCCCAAGCGTTCAATCAAC
losA-62	CTAAACGCCTTTCTGGTTTG
losA-640	CTGAGTCACTGTTTAATCCTC
lic3g	ATATACAAACTTCGGCGG
O52560	GCAGGTAATGGAACAAGTTT
286lic3	ATTGGCATTAAAAGCATATCG
287lic3	AAATGTACCGGATTGTTGAAG
234a	AGTCTGGCTGGATTCATTGC
234b	GCTAATTGATGCCGGATAGC
lpt3midF	CTAAGCCTATGAAATCGGG
lpt3midR	GGAATATGACTCCCTCTATG
lic1b-1	ATGCGTGGCTATCTCTTTGG
lic1b-773	CCAAGTCCAATTGCCCATAC
5'oafA	TGTCTTTCATTGCGTCTGTTATTT
oafE	ATTTTGTGATGTGGTAT
hepfor	GGTGTCATCTGCACCAAAC
heprev	GATCTCATTAACATTATCG
159a	TTGCCATAGAACCGGATC
163a	CAAATTCAATGGACTGGAC
883b	GATATGGGCTTAGTGAATACTG
884b	ATTCAATGGCTTGCTCTTG
934b	GGCCACGATACCTTGAATC
lic1a-250	GCGAAACTCATCGAATACG
Lic2BA	CATTTAGCGATGAGTTCC
RYlic2C	CATCAGAAGTCTCACTTACCAC

Table 2.3. Primers for OAg locus analysis (see Table 5.2) and amplification of the *Hp* 15-like OAg ligase gene (see Fig. 5.7). The target genes are shown in Table 5.2 and Fig. 5.7. Primer sequences are given in the 5' to 3' orientation.

Primer	Sequence
269a	CCATCACTTGGTTTGATGCT
271b	GGCGTTTTGGCTAGCATCC
271d	TTCAGATGTAGTGGTACTTTC
272a	CGTGCAGTCTTGCGTTACAT
272b	CCACCAGCTCCTGTTACCAT
272c	GTTCTCGAAATTCCAAATAC

274a	ATGCAGCGATTGGCTTAG
274b	TTGAAGATTACCCGCAAGC
275a	CAAAACTTGGAGAAATAGACATC
275b	GATGATATAGCTGAAAGAGACC
276a	TGGTGGACTGTCTAGTGC
276b	AACTAATACTTCCTTCCC
277a	CATTCACAGTGGCCTTTGG
277b	CGAATGCAATACAAACAAGAGC
277c	GTGCGAAATTTTCATTTGAAC
278a	TGCACAACCCCTTTACAAAC
278b	CACTCCATCGACAACAAAGC
279a	CGCCTTACTGATAGGTTCC
279b	AATATGCTAGCAAAGGGG
280a	GCATAAAAGGATATTTCGTAAG
280b	TGAAAGTATTAAAACCAACAAC
283a	AATACCATTTCCTGGACG
283b	GCTCCGAAAGCGGAATACC
285a	TCAGCCTTCGTTTGTACC
285b	TAATATCCTCTGACAGCC
285d	GATTTCGCATAGCCACC
286F4	ATATTCCAGCGGTAGCAAGC
841j	TCGTTACGAATGACTAAAATC
842c	GAAGCAATTAGCTATGGTGC
FLP17a	ACGTTTCCTTATGTTTCACG
FLP17b	ACCCCTAAAGATGTAACCAC
gr2-ABC1	CCGATTTTATGCATTCGATG
gr2-AC1	CCTTCAGGTCCTATTTGGAC
gr2-GT1	TTAGATGAGGCATTAGAACC
gr2-GT2	TCTTATGCTAAGAAAGGCTC
gr2-GT3	GCCTGCTCAAACTCGTTAAG
gr2-GT4	AAGTATAAATGCAGAAACATGC
gr2-GT5	TTATTGGGCTTGGCAAAATG
gr2-GT6	TCGCCTTCCAAATTAACTAC
gr2-mut1	ACGTGATGCTGAAACAAATG
gr2-mut2	TTCATTGATTCAGCAAATGG
hmg13D	CGATAGAGTAAAATGATAAGG
hmg13E	CAGATTGTCTAGATAAGTTACTCC
hmg13G	TGCATAAAATTACGTGGTG
hmg13H	CCAACGAAAGATACACAGG
hmg13M	TCAGTTTATTCTGGCGATTG
hmg13Q	CCAAGAAAACTCAGCATTAG
hmg13R	ATTGCGCCATTAATTGAG
hmg13S	CTGCATCGCCCACAGC
hmg13V	GATTGATTTGCCACATAGG
hmg13W	GTGCCAAATAAGCCTAATCAC
hmg18B	CATCAGTATAGAAACATTTCTC
hmg30N	GATAAAGACTCCATTGACAAAC
hmg30P	CAGCCTATCTTCTTACAGGAC

LIG15B	CCCTCGAATTATGCATATTC
LIG15C	TATGTACCAAGAAACCTTGGC
OAG17H	ATCCTACACTACTGCGACC
OAG17R	ACTTCCCATTTAAAGATAGTC
POL17a	GGAGATAGGCAATTAATAACG
POL17b	ATAACCTAGAATCAAGCGC
rmlA-F	GAAAGGTATTATTCTTGCAGG
rmlA-R	TTCATCAAATTCCACTACACC

Table 2.4. Alternative primers for OAg locus analysis (see Table 5.2). Primer sequences are given in the 5' to 3' orientation.

Gene	Alt. primers	Sequence
Hp0285	hmg19A	TACTTTTTCTGCGGAAACTCC
	hmg19B	ATCCTTAATCATAGGATAATCAGG
Hp0279	FLP20a	CCTATTTGGTACTTTCAAGG
	FLP20b	AATAGCCGATAGAGGAACTC
Hp0277	POL20a	TTCATATGTTTAGGTTGGGC
	POL20b	CTATGCCAGAAACTCCTTGC
Hp0275	275a	As before
	275bALT	GATGATATTTCTAATCCAGAGC
Hp0271	LIG20a	CACATATCTAGGCGCTCTCC
	LIG20f	GAAGTGGTGAAGCAGAATAAAGC

Plasmid	Insert	PCR template strain	Primers (insertion site)	Backbone
pRY11	*lgtF* interrupted by Tet^R+USS	*Hp* T3T1	Hp469-UP and Hp469-DN3 (BglII)	pSCA
pRY12	*lgtF* interrupted by Kan^R+USS	*Hp* T3T1	Hp469-UP and Hp469-DN3 (BglII)	pSCA
pRY27	*Hp1877* pseudogene interrupted by *lic1ABCD* and *ermC*	*Hp* T3T1 for *Hp1877*; *Hi* Rd for *lic1*	P1877a and P1877b for *Hp1877* (BsaBI); lic1-CLF and lic1-CLR for *lic1*	pSCA-amp-kan
pRY29	*tetB* interrupted by *lic1ABCD* and *ermC*	*Hp* T3T1 for *tetB*; *Hi* Rd for *lic1*	TetB-CLF and TetB-CLR for *tetB* (MfeI); lic1-CLF and lic1-CLR for *lic1*	pSCA-amp-kan
pRY34	*wecA* interrupted by Kan^R+USS	*Hp* T3T1	wecA-CLF and wecA-CLR (NruI)	pSCA
pRY35	*wcfS* interrupted by Kan^R+USS	*Hp* T3T1	272c and 274a (PsiI)	pSCA
pRY36	*wbaP* interrupted by Kan^R+USS	*Hp* 13	n/a: HaeIII fragment in clone 5B (Bpu10I)	pBluescript
pRY39	*wbaP* partial deletion interrupted by Kan^R+USS	*Hp* 13	HaeIII fragment; deletion by amplification with primers wba13F and wba13R	pBluescript
pRY41	*wcfS* partial deletion interrupted by Kan^R+USS	*Hp* T3T1	272c and 274a; deletion by amplification with primers wba26F and wba26R	pSCA
pRY46	*waaL* (*Hp*15-like) interrupted by *ermC*	*Hp* 19	842b and 841j (BsaBI)	pSCA-amp-kan

Table 2.5. Plasmid constructs for the interruption or insertion of genes in *Hp*. In general, a PCR product was amplified using gDNA template from the strain listed in column 3 with the primers given in column 4, and was ligated to the vector in column 5. A drug resistance gene/cassette was then used to interrupt the gene within its 5' end, at the restriction site given in brackets in column 4. Tet^R, tetracycline resistance cassette; Kan^R, kanamycin resistance cassette; ermC, erythromycin resistance gene; USS, uptake signal sequence.

Table 2.6. Primers used for the construction of recombinant plasmids. The target genes are listed in Table 2.5. Sequences are given in the 5' to 3' orientation.

Primer	Sequence
Hp469-UP	CATTCGCCACAATGAAAGG
Hp469-DN3	GTACGCCAAACGTCATTACC
P1877a	ATTGAGAGCAGGCTGCAAAC
P1877b	CCGCAATGGCAACAGTTAAG
lic1-CLF	ACTCATAAGATTCAGAGCCTTAC
lic1-CLR	GAATTAAATCTGCCCTTTCG
TetB-CLF	GTTCGACAAAGATCGCATTG
TetB-CLR	TGCCCTCTTGGGTTATCAAG
wecA-CLF	GGAGTTTACGCCACCAGG
wecA-CLR	GTTACGACGTGCACGACG
272c	GTTCTCGAAATTCCAAATAC
274a	ATGCAGCGATTGGCTTAG
wba13F	GTAGTACAATTGGTCCAATAGTGAAAAAGG
wba13R	GTAGTACAATTGCCATCCCTTGTAATGAT
wba26F	GTAGTACAATTGTCCTCGTCCAGAAGTAAG
wba26R	GTAGTACAATTGCAATCTTAATCCAAATGG
Hp842b	GTCAATTAGGTAACCCTGAAGG
Hp841j	TCGTTACGAATGACTAAAATC

Table 2.7. *Hp* mutants produced during the project. The strain listed under 'Parent' was transformed using either electroporation (Elec) or the static aerobic method (SA). The donor DNA for transformation was either a plasmid (pRYxx or 5BA2), or chromosomal DNA (chr) from another *Hp* strain as listed. The drug(s) upon which the transformants were selected is given as kanamycin (Kan), erythromycin (Ery) or tetracycline (Tet). Note that in this table, *waaL* refers to the OAg ligase gene that is outside the OAg locus.

Hp mutant	Parent	Genotype	Method	Donor	Drug
10.1	*Hp*10	*ΔwbaP*	SA	pRY39	Kan
10.2	*Hp*10	*ΔwbaP*	SA	pRY39	Kan
13.1	*Hp*13	*lgtF*	SA	20.5 chr	Kan
13.2	*Hp*13	*lgtF*	SA	20.5 chr	Kan
13.3	*Hp*13	*lgtF*	SA	20.5 chr	Kan
13.4	*Hp*13	*wecA*	SA	19.7 chr	Kan
13.5	*Hp*13	*wecA*	SA	19.7 chr	Kan
13.7	*Hp*13	*wbaP*	SA	pRY36	Kan
13.8	*Hp*13	*wbaP*	SA	pRY36	Kan
15.1	*Hp*15	*lgtF*	SA	19.1 chr	Kan
15.4	*Hp*15	*lgtF*	SA	20.5 chr	Kan
15.5	*Hp*15	*lgtF*	SA	20.5 chr	Kan
15.6	*Hp*15	*ΔwbaP*	SA	10.1 chr	Kan
15.7	*Hp*15	*ΔwbaP*	SA	10.1 chr	Kan

15.8	*Hp*15	*ΔwbaP*	SA	35.9 chr	Kan
15.10	*Hp*15	*waaL*	SA	19.19 chr	Ery
15.11	*Hp*15	*waaL*	SA	19.19 chr	Ery
15.12	*Hp*15	*waaL*	SA	19.19 chr	Ery
17.1	*Hp*17	*lgtF*	SA	pRY12	Kan
17.2	*Hp*17	*lgtF*	SA	pRY12	Kan
19.1	*Hp*19	*lgtF*	SA	pRY12	Kan
19.2	*Hp*19	*lgtF*	SA	pRY12	Kan
19.4	*Hp*19	*lic1+*	SA	pRY29	Ery
19.5	*Hp*19	*lic1+*	SA	pRY29	Ery
19.7	*Hp*19	*wecA*	SA	pRY34	Kan
19.8	*Hp*19	*wecA*	SA	pRY34	Kan
19.10	*Hp*19	*wcfS*	SA	pRY35	Kan
19.11	*Hp*19	*wcfS*	SA	pRY35	Kan
19.16	*Hp*19	*ΔwcfS*	SA	pRY41	Kan
19.17	*Hp*19	*ΔwcfS*	SA	pRY41	Kan
19.19	*Hp*19	*waaL*	SA	pRY46	Ery
19.20	*Hp*19	*waaL*	SA	pRY46	Ery
19.21	*Hp*19	*waaL*	SA	pRY46	Ery
19.24	*Hp*19	*losB1*	SA	5BA2	Kan
19.25	*Hp*19	*losB1*	SA	5BA2	Kan
20.1	*Hp*20	*tet+*	SA	T3T1 chr	Tet
20.4	*Hp*20	*lgtF*	SA	pRY11	Tet
20.5	*Hp*20	*lgtF*	SA	pRY12	Kan
20.6	*Hp*20	*lgtF*	Elec	pRY12	Kan
20.12	*Hp*20.1	*lic1+*	SA	pRY29	Ery
20.13	*Hp*20.1	*lic1+*	SA	pRY29	Ery
20.14	*Hp*20.1	*lic1+*	SA	pRY29	Ery
20.15	*Hp*20.1	*lic1+*	SA	pRY29	Ery
20.16	*Hp*20.1	*wcfS*	SA	pRY35	Kan
20.17	*Hp*20.16	*wcfS lic1+*	SA	pRY29	Kan Ery
20.18	*Hp*20	*ΔwcfS*	SA	pRY41	Kan
20.19	*Hp*20	*ΔwcfS*	SA	pRY41	Kan
20.21	*Hp*20.18	*ΔwcfS lic1+*	SA	pRY27	Kan Ery
20.22	*Hp*20.18	*ΔwcfS lic1+*	SA	pRY27	Kan Ery
26.2	*Hp*26 (T3T1)	*lgtF*	Elec	pRY12	Kan
26.3	*Hp*26 (T3T1)	*lgtF*	Elec	pRY12	Kan
35.2	*Hp*35	*lgtF*	SA	20.4 chr	Tet
35.7	*Hp*35	*lgtF*	SA	pRY12	Kan
35.9	*Hp*35	*ΔwbaP*	SA	pRY39	Kan
35.10	*Hp*35	*ΔwbaP*	SA	pRY39	Kan

3 Results: Prediction of *H. parainfluenzae* core LPS structures using genetic and phenotypic analyses

This chapter aims to address whether the LPS of our study strains (18 true *Hp* isolates and two hybrid *Hi/Hp* isolates) is composed of the same lipid A, inner core and partially phase variable outer core components as *Hi* LPS. The results were obtained from three main sources. Firstly, a BLAST search facility for the assembled *Hp* T3T1 draft genome sequence was available via www.sanger.ac.uk; the completed genome has now been transferred to NCBI (www.ncbi.nlm.nih.gov). The Artemis sequence viewer was used for further analysis of the *Hp* T3T1 genome sequence. Secondly, genomic DNA prepared from all of the study strains was used for PCR amplification and Southern analysis. Thirdly, cell lysates prepared from all of the study strains were used for immunoblots and LPS profiling. Limited sequence information later became available from the Genomophilus project and was added to the analysis where possible. These sources allowed a detailed comparison of the genetics of core LPS synthesis between *Hi* and *Hp* and also between different strains in the *Hp* study collection. The results have been integrated with structural LPS data from the Schweda Laboratory where available to confirm predictions of *Hp* core LPS structure.

3.1 Lipid A and inner core LPS

3.1.1 The *Hp* T3T1 genome contains homologues of the *Hi* lipid A synthesis genes

The lipid A moiety of LPS anchors the molecule in the outer membrane and is found in all Gram negative bacteria. The biosynthetic pathway for lipid A in *Escherichia coli* has been summarised by Wang and Quinn (2010). The first six enzymes in the pathway (LpxA-LpxD, LpxH and LpxK) catalyse the conversion of UDP-GlcNAc to the tetraacylated molecule disaccharide-1-P, and phosphorylate it to form lipid IV_A. KdtA then adds two 3-deoxy-D-manno-octulosonic acid (Kdo) residues before two final acyl chains are incorporated by LpxL and LpxM.

The lipid A structure of *Hi* LPS is very similar to that of *E. coli* (Helander et al., 1988) and homologues of all eight Lpx enzymes have been annotated in the *Hi* Rd genome sequence on the JCVI Comprehensive Microbial Resource website (www.cmr.jcvi.org). These *Hi* Rd amino acid (aa) sequences were used in TBLASTN searches against the *Hp* T3T1 genome on the Sanger Institute BLAST server (previously available at www.sanger.ac.uk). Intact homologues were identified in the translated *Hp* T3T1 genome for all of the enzymes required for lipid A synthesis in *Hi* (Table 3.1), with aa identity ranging from 70 to 92%. It is therefore likely that *Hp* T3T1 is capable of synthesising a lipid A molecule with a similar structure to that of *Hi*, based on a hexaacylated disaccharide, although it is possible that the length and degree of saturation of their acyl chains may differ. Structural analysis has subsequently confirmed that *Hp* T3T1 lipid A comprises a diglucosamine headgroup and six acyl chains, with an identical configuration to that of *Hi* (Elke Schweda, pers. comm.).

Process	Gene name	Protein function	*Hi* Rd ORF	*Hp* T3T1 E-value	*Hp* T3T1 homologue
Lipid A synthesis	*lpxA*	UDP-N-acetylglucosamine acyltransferase	HI_1061	1.2E-127	Hp1018
	lpxC	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	HI_1144	4.1E-147	Hp1256
	lpxD	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	HI_0915	1.7E-154	Hp1016
	lpxH	UDP-2,3-diacylglucosamine hydrolase	HI_0735	8.9E-87	Hp0867
	lpxB	lipid-A-disaccharide synthase	HI_1060	2.8E-162	Hp1019
	lpxK	tetraacyldisaccharide 4'-kinase	HI_0059	1.1E-140	Hp1343
	lpxL (*htrB*)	laurate transferase	HI_1527	9.0E-147	Hp1882
	lpxM (*msbB*)	myristate transferase	HI_0199	2.3E-143	Hp1901
Kdo synthesis and transfer	*kdsD*	arabinose-5-phosphate isomerase	HI_1678	7.9E-148	Hp1432
	kdsA	2-dehydro-3-deoxyphosphooctonate aldolase	HI_1557	2.9E-142	Hp1447
	kdsC (*yrbI*)	Kdo 8-phosphate phosphatase	HI_1679	1.7E-74	Hp1431
	kdsB	Kdo cytidylyltransferase	HI_0058	1.0E-115	Hp1345
	kdtA (*waaA*)	Kdo transferase	HI_0652	8.2E-161	Hp0468
	kdkA (*orfZ*)	Kdo kinase	HI_0260.1	1.7E-92	Hp0255
Regulation of lipid A and Kdo synthesis	*ftsH* (*hflB*)	ATP-dependent metallopeptidase / cell division protease	HI_1335	5.1E-311	Hp1713
Triheptosyl core synthesis and transfer	*gmhA*	phosphoheptose isomerase	HI_1181	1.7E-92	Hp1223
	hldE (*rfaE*)	bifunctional heptose 7-phosphate kinase/ heptose 1-phosphate adenyltransferase	HI_1526	1.0E-230	Hp1881
	gmhB	D,D-heptose 1,7-bisphosphate phosphatase	HI_0621.1	5.8E-83	Hp0830
	hldD (*rfaD*)	ADP-L-glycero-D-mannoheptose 6-epimerase	HI_1114	3.6E-154	Hp0384
	opsX	ADP-heptose:LPS heptosyltransferase I	HI_0261	6.2E-148	Hp0253
	rfaF	ADP-heptose:LPS heptosyltransferase II	HI_1105	7.5E-154	Hp0383
	orfH	ADP-heptose:LPS heptosyltransferase III	HI_0523	9.4E-156	Hp0841

Table 3.1. Identification in the *Hp* T3T1 genome of genes required for the synthesis of lipid A, Kdo and triheptosyl inner core LPS. The genes required for each process in *Hi* are listed (with alternative names in brackets) along with their protein functions as discussed in the text. Results of TBLASTN searches using *Hi* Rd amino acid sequences against the *Hp* T3T1 genome are shown in the last two columns: the closest homologue (i.e. the translated ORF representing the highest scoring HSP) is specified alongside the E-value for the match.

3.1.2 *Hp* contains homologues of the genes required for Kdo and triheptosyl inner core LPS synthesis in *Hi*, plus an additional HepI transferase

The biosynthesis of CMP-Kdo from arabinose-5-P requires the four enzymes KdsA-KdsD (Cipolla et al., 2009), which have been characterised in *E. coli* and have strong homologues in *Hi* (Table 3.1). Kdo is transferred from CMP-Kdo to lipid IV_A by KdtA, also known as WaaA. KdtA in *E. coli* and most other Gram negative bacteria is a bifunctional enzyme, adding two sequential Kdo residues to the lipid A. In contrast, *Hi* KdtA adds only a single Kdo residue (White et al., 1997), which is then phosphorylated by the Kdo kinase KdkA (White et al., 1999).

Intact homologues of the *Hi* Rd enzymes KdsA-KdsD, KdtA and KdkA were identified in the translated *Hp* T3T1 genome (Table 3.1). The translated product of gene *Hp0468* shares 71% aa identity with *Hi* Rd KdtA, whilst that of *Hp0255* shares 70% aa identity with *Hi* Rd KdkA. Together with the lipid A genetic data, this suggests that *Hp* is able to synthesise the same lipid A-KdoP structure as *Hi*. Primers designed to the *Hp* T3T1 *kdtA* and *kdkA* genes amplified PCR products of the expected size from the gDNA of 18/18 *Hp* strains used in this study (see the summary chart, Table 3.3, later in the chapter). No products were amplified for the two hybrid strains, probably due to sequence divergence.

Most Gram negative species elaborate the lipid A-Kdo moiety of their LPS with two or three residues of L-*glycero*-D-*manno*-heptose (Hep) to complete the inner core. The conserved biosynthetic pathway for the precursor nucleotide sugar, ADP-Hep, requires the enzymes GmhA, HldE, GmhB and HldD (Table 3.1) (Valvano et al., 2002). *Hi* adds three Hep residues to lipid A-KdoP sequentially using the ADP-Hep:LPS heptosyltransferases OpsX, (HepI), RfaF (HepII) and OrfH (HepIII). *Hp* T3T1 encodes homologues of the four ADP-Hep biosynthesis enzymes and also of all three *Hi* Hep transferases, the latter with aa identities of 78%, 83% and 80% respectively (Table 3.1). Southern analysis of MfeI restriction digested gDNA from the study strains, using a fragment of *Hp* T3T1 *orfH* as a probe, confirmed that an *orfH* homologue is present in all 20 strains (Fig. 3.1): as this is the last enzyme required in LPS triheptosyl core synthesis, its presence suggests that this structure is conserved across *Hp* strains as it is in *Hi*.

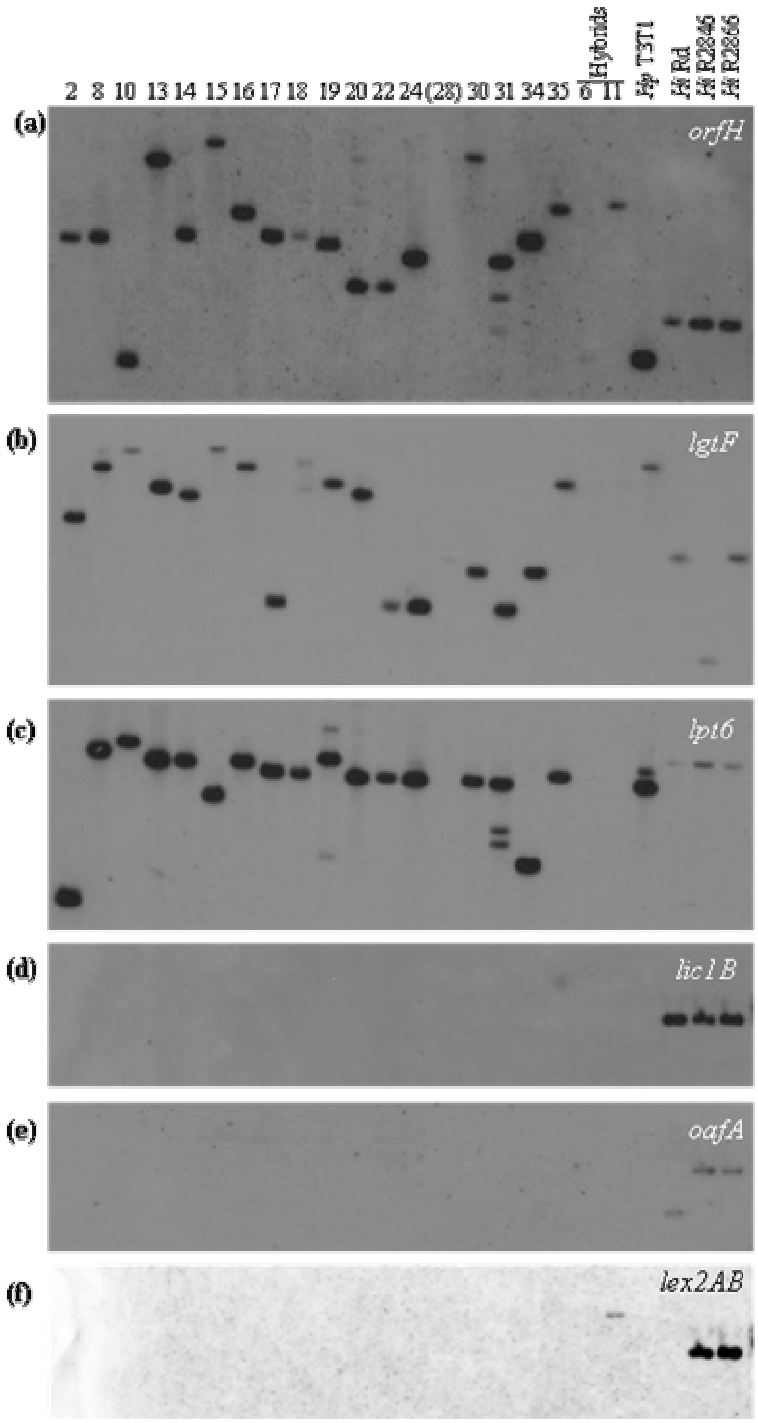


Figure 3.1. Southern analysis of MfeI-digested gDNA hybridised with PCR fragments from inner and outer core genes. *Hp* strain numbers are listed above each lane; three *Hi* controls are shown on the right. Note that *Hp* strain 28 was later reclassified as *Actinobacillus pleuropneumoniae* and was excluded from further analysis. Probes were as follows: **(a)** *Hp* T3T1 *orfH*, **(b)** *Hp* T3T1 *lgtF*, **(c)** *Hp* T3T1 *lpt6*, **(d)** *Hi* Rd *lic1B*, **(e)** *Hi* Rd *oafA*, **(f)** *Hi* Eagan *lex2A/lex2B*.

Due to the *Hi* inner core LPS containing lipid A-KdoP instead of the more common lipid A-Kdo_2, *Hi* OpsX is unusual both in sequence and in substrate specificity when compared to Hep I transferases in other Gram negative bacteria (Gronow et al., 2005). A TBLASTN search with the *E. coli* Hep I transferase WaaC (GenBank: ACT27183.1) revealed an *Hp* T3T1 gene, *Hp0380*, encoding a protein with 58% aa identity. *Hp0380* also has homologues in several Pasteurellaceae species including *Histophilus somni* (72% aa identity), *Pasteurella multocida* (69%), *Actinobacillus pleuropneumoniae* (61%) and *Mannheimia succiniciproducens* (61%), but not in *Hi*. This raises the possibility that *Hp* is capable of synthesising two different lipid A-Kdo structures and can extend these with HepI using different substrate-specific transferases encoded by *opsX* and *waaC*. This situation has already been observed in *P. multocida* LPS isolated from experimental fowl cholera infections, where 20% of the LPS contains Kdo_2 rather than KdoP (Harper et al., 2007). *P. multocida* mutants in which *waaC* was inactivated were still capable of causing lethal infections, so Kdo_2-containing glycoforms are not essential for virulence in this instance.

Primers 380a and 380b, designed to the *Hp* T3T1 *waaC* homologue, amplified a product from all 18 true *Hp* strains in the study panel but not from the two hybrid strains, providing strong evidence that this gene is conserved across *Hp*. Structural analysis of LPS from *Hp* T3T1 has so far identified only KdoP, not Kdo_2 (Elke Schweda, pers. comm.), but the latter may be a minor form and difficult to detect.

FtsH (HflB) is an ATP-dependent protease that regulates the concentration of LpxC and KdtA in order to ensure that the production and elaboration of lipid A are balanced, as demonstrated in *E.*

coli (Katz & Ron, 2008). The *Hp* T3T1 gene *Hp1713* encodes a protein with 91% aa identity to the *Hi* Rd HflB enzyme, HI1335. The high level of conservation between the *E. coli*, *Hi* and *Hp* HflB homologues suggests that the ability to regulate LPS production is under strong selection pressure.

3.2 Initial glycose extensions from the triheptosyl inner core LPS

3.2.1 Extension from HepI: *lgtF* is present in all of the study strains

In *Hi* LPS the proximal Hep is substituted with Glc via a β1-4 linkage by the glycosyltransferase LgtF (Hood et al., 2004a). The *Hp* T3T1 gene *Hp0469* encodes a protein with 86% aa identity and 92% similarity to LgtF from *Hi* Rd (Table 3.2); in both cases the gene is adjacent to that which encodes KdtA. A 600 bp fragment of *Hp0469* was amplified using the primers 469a and 469b, and used as a probe in Southern analysis of the study strains. Hybridising bands were observed for all true *Hp* strains and the three *Hi* control strains (Fig. 3.1 and Table 3.3), indicating that all of the *Hp* strains encode an LgtF homologue. Only very weak hybridising bands were detected for the hybrid strains *Hp* 6 and 11; *Hp* 6 data from Genomophilus revealed a reading frame that encodes an LgtF homologue with closer identity to the *Hi* Rd version (96% aa) than to Hp0469 (86%).

Table 3.2. *Hi* transferases required for LPS synthesis and their homologues in *Hp* T3T1. A search of the *Hi* LPS literature, cited in the text, enabled the compilation of a list of 33 known or potential transferases for LPS synthesis. Their substrate specificity is listed under 'Addition' where known (the function of the *Hi lsg* locus has only been characterised in transgenic *E. coli* strains). The frequency of each transferase amongst *Hi* strains was obtained from the literature or from work performed in our own group, except those listed as out of 11 strains which were checked using TBLASTN searches of the 11 NTHi genomes listed by Hogg et al. (2007). The *Hp* T3T1 genome was searched using each *Hi* transferase aa sequence ('ORF in *Hi*' column) with a TBLASTN algorithm; the highest-scoring match is reported in the two right hand columns. Genes for which the highest-scoring match had a strong E-value (<1E-60) are shaded in grey. Genes *Hp0270* to *Hp0285* comprise the OAg locus.

Table 3.2

	Data from *H. influenzae*					TBLASTN against *Hp* T3T1	
Role	**Gene name**	**Addition**	**ORF in *Hi***	**Frequency in *Hi***	**Repeat Unit Seq**	***Hp* T3T1 E-value**	**Hp T3T1 match**
Initial glycosyltransferases	*lgtF*	Glc to HepI	HI_0653	25/25 NTHi	-	3.80E-118	Hp0469
	lic2C	Glc to HepII	RM7004_U36398 orf3	12/25 NTHi	-	no hit	-
	lpsA	Glc or Gal to HepIII	HI_0765	25/25 NTHi	-	no hit	-
Other glycosyltransferases	*lex2A*	Needed for Lex2B function	R2846(v6)_1683	20/27 capsular	GCAA	no hit	-
	lex2B	Glc or Gal to Glc on HepI	NTHI0913	23/25 NTHi	-	no hit	-
	lgtC	Distal Gal of digalactoside	HI_0258	25/25 NTHi	GACA	2.30E-21	Hp0862
	lgtD	GalNAc to globotriose	HI_1578	11/11 NTHi	-	4.60E-19	Hp0317
	lic2A	Proximal Gal of digalactoside	HI_0550	25/25 NTHi	CAAT	0.210	Hp1755 (*lsgD*)
	lic2B	Gal (putative)	RM7004_U36398	12/25 NTHi	-	1.000	Hp1755 (*lsgD*)
	losA1	Glc to HepIV	R2846(v6)_38c	6/28 NTHi	CGAGCATA	0.064	Hp1755 (*lsgD*)
	losA2	Gal to HepIV	R2846(v6)_1560	10/28 NTHi	-	0.027	Hp1755 (*lsgD*)
	lsgC	Unknown	HI_1698	11/11 NTHi	-	no hit	-
	lsgD	Gal to GlcNAc (in *E. coli*)	HI_1697	11/11 NTHi	-	1.90E-95	Hp1755
	lsgE	GlcNAc to Gal (in *E. coli*)	HI_1696	11/11 NTHi	-	6.80E-121	Hp1753
	lsgF	Gal to GlcNAc (in *E. coli*)	HI_1695	11/11 NTHi	-	4.60E-124	Hp1752
	pgt1 putative GT	Unknown	R2846(v6)_916	4/11 NTHi	GACA	4.00E-38	Hp0862
	Putative GT (*lpt3* region)	Unknown	R2846(v6)_1528c	8/28 NTHi	AGTC	no hit	-
	Putative GT (HMG region)	Unknown	HI_0868	17/28 NTHi	-	9.80E-18	Hp0276
	Putative GT (HMG region)	Unknown	HI_0869	17/28 NTHi	-	1.30E-01	Hp1753 (*lsgE*)
	Putative HMG ligase	HMG to Glc on HepI	HI_0874	17/28 NTHi	-	1.80E-61	Hp0271
	Putative GT	Unknown	HI_1386		CAAC	no hit	-
Sialyltransferases	*lic3A*	Neu5Ac to lactose	HI_0352	25/25 NTHi	CAAT	no hit	-
	lic3B	Neu5Ac to lactose and sialyllactose	NTHI_1034	14/25 NTHi	CAAT	no hit	-
	lsgB	Neu5Ac (to undefined posn)	HI_1699	11/11 NTHi	-	1.90E-43	Hp0278
	siaA (*orfY*) (HMG region)	Neu5Ac to HMG	HI_0871	17/28 NTHi	-	no hit	-
Phosphoethanolamine transferases	*lpt6*	PEtn to HepII	HI_0275	70/70 Hi	-	6.10E-222	Hp0234
	lpt3	PEtn to HepIII	NTHI_1053	9/28 NTHi	TGTT	1.90E-05	Hp1006
	Putative PEtn transferase	PEtn to GalNAc of HMG	HI_1005	11/11 NTHi	-	4.30E-58	Hp1006
	Putative PEtn transferase	Unknown	HI_1064	11/11 NTHi	-	9.70E-181	Hp1006
Phosphocholine transferase	*lic1D*	PCho to Hex or HepIV	HI_1540	25/25 NTHi	CAAT (*lic1A*)	no hit	-
O-acetylases	*oafA*	OAc to HepIII	NTHI_0512	25/25 NTHi	GCAA	no hit	-
HepIV transferases	*losB1*	DD-Hep to Glc on HepI	R2846(v6)_37c	6/28 NTHi	-	no hit	-
	losB2	LD-Hep to Glc on HepI	R2846(v6)_1561	10/28 NTHi	-	5.20E-09	Hp380 (*waaC*)

					Hp strain #																			
Function	**Gene**	**Method**	**Template**	**Primers**	**2**	**8**	**10**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**22**	**24**	**30**	**31**	**34**	**35**	**6**	**11**	**T3T1**
Kdo transferase	*kdtA*	PCR	*Hp* T3T1	468a/468b	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+
Kdo kinase	*kdkA*	PCR	*Hp* T3T1	255a/255b	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+
Hep transferases	*opsX*	PCR	*Hp* T3T1	253a/253b	(+)	(+)	-	+	-	(+)	+	+	+	+	(+)	+	+	-	+	+	+	-	-	+
	rfaF	PCR	*Hp* T3T1	383a/383b	(+)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	+
	orfH	Southern	*Hp* T3T1	841a/841b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1	1	1	1	1
	waaC	PCR	*Hp* T3T1	380a/380b	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+
Initial glycosyltransferases	*lgtF*	Southern	*Hp* T3T1	469a/469b	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	0	0	1
	lic2C	Locus PCR	*Hi* Rd	ksgArev/infrev	-	-	-	+	-	-	+	-	+	-	-	-	+	-	+	-	-	-	-	-
	lpsA (*Hi*)	PCR	*Hi* Rd	6018D/6018E	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	lpsA (*Hp*)	PCR	*Hp* 20	HpLpsA-F/R	-	+	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-
Other glycosyltransferases	*lex2A/B*	Southern	*Hi* Eagan	lex5'/lex3'	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	lic2A	Locus PCR	*Hp* T3T1	933b/934b2	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	+	-	-	-	-
	losA1	PCR	*Hp* 17	losA-62/losA-640	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	+	-	-	-	-
Sialyltransferases	*lic3A* or *B*	PCR	*Hi* Rd	lic3g/O52560	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	lic3B	Locus PCR	*Hp* T3T1	286lic3/287lic3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PEtn transferases	*lpt6*	Southern	*Hp* T3T1	234a/234b	1	1	1	1	1	1	1	1	1	2	1	1	1	1	3	1	1	0	0	1
	lpt3	PCR	*Hi* 86-028NP	lpt3midF/R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Choline transport	*lic1B*	Southern	*Hi* Rd	lic1b-1/lic1b-773	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O-acetylase	*oafA*	Southern	*Hi* Rd	5'oafA/oafE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HepIV transferase	*losB1*	PCR	*Hi* R2846	hepfor/heprev	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	+	+	-	-	-

Table 3.3. Experimental detection of selected inner and outer core synthesis genes in the *Hp* study strains. Positive results are highlighted in black. *Hi* positive controls were included where necessary but are not shown. One of three methods was used per gene. 'Southern' indicates Southern analysis of gDNA digests using a ~600 bp probe which was amplified using the template gDNA and primers listed; the number of hybridising bands is stated. 'PCR' indicates the attempted amplification of an internal gene fragment using the primers listed, which were designed to the strain listed under 'Template'; '+' means that a product of the expected size was amplified, whilst '-' means that no product was amplified. Note that *lpsA* and *losA1* were later sequenced also from *Hp* strain 35. (+) indicates that no product was detected in PCR but that an ORF with >90% aa identity was identified in the Genomophilus data set. 'Locus PCR' indicates that the *Hi* location of this gene was amplified using primers designed to the two flanking genes; *Hp* strains with short PCR products (i.e. no intervening gene) were scored as negative, whilst strains with longer products were scored as positive and confirmed by sequencing.

3.2.2 Extension from HepII: *lic2C* was identified in a minority of the study strains

Approximately half of NTHi strains produce the glycosyltransferase Lic2C that is required for the addition of Glc to the second heptose of the LPS core (Hood et al., 2004a). When present, *lic2C* is found adjacent to a further LPS-related gene *lic2B*, which encodes a putative galactosyltransferase, and both genes are located between *ksgA* and *infA*. However, in other *Hi* strains, no genes or the *losA1* and *losB1* genes can be found between *ksgA* and *infA*. No *lic2B* or *lic2C* homologues were identified in the *Hp* T3T1 genome (Table 3.2). Primers ksgArev and infrev, which flank the locus, amplified one of three different sized PCR products from each of the 20 study strains (Fig. 3.2, PCR 1). For eleven strains including the hybrids *Hp* 6 and *Hp* 11 a 700 bp product indicated that no genes were present, as in the *Hp* T3T1 genome sequence. For five strains (*Hp* 13, 16, 18, 24 and 31) a 2 kb product was amplified: sequencing of the *ksgA-infA* region from *Hp* 13 revealed a single ORF, which shared 94% aa homology with Lic2C from *Hi* strain RM7004. Further PCR using primers 934b, RYlic2C, infrev and lic2BA indicated that this gene arrangement was conserved in the other four *Hp* strains listed, i.e. they contain only *lic2C* at this locus (data not shown). For the remaining four *Hp* strains (*Hp* 17, 19, 34 and 35) the ksgArev/infrev primers amplified a 2.6 kb product. The *Hp* 17 *ksgA-infA* intergenic region was sequenced and was found to encode homologues of LosA1 and LosB1 (see Section 3.3.5).

3.2.3 Extension from HepIII: a minority of the *Hp* strains contain *lpsB*

In *Hi* the enzyme responsible for hexose addition to the third heptose is encoded by a gene originally termed *lpsA* which is present in all studied *Hi* strains at the same locus. It was later noted that LpsA enzymes can be split into four allelic groups, each of which carries out a slightly

different glycosyltransferase function: these were renamed LpsA (directs the addition of Glc via a β1,2 linkage), LpsB (Gal β1,2), LpsC (Glc β1,3) and LpsD (Gal β1,3) (Deadman et al., 2006).

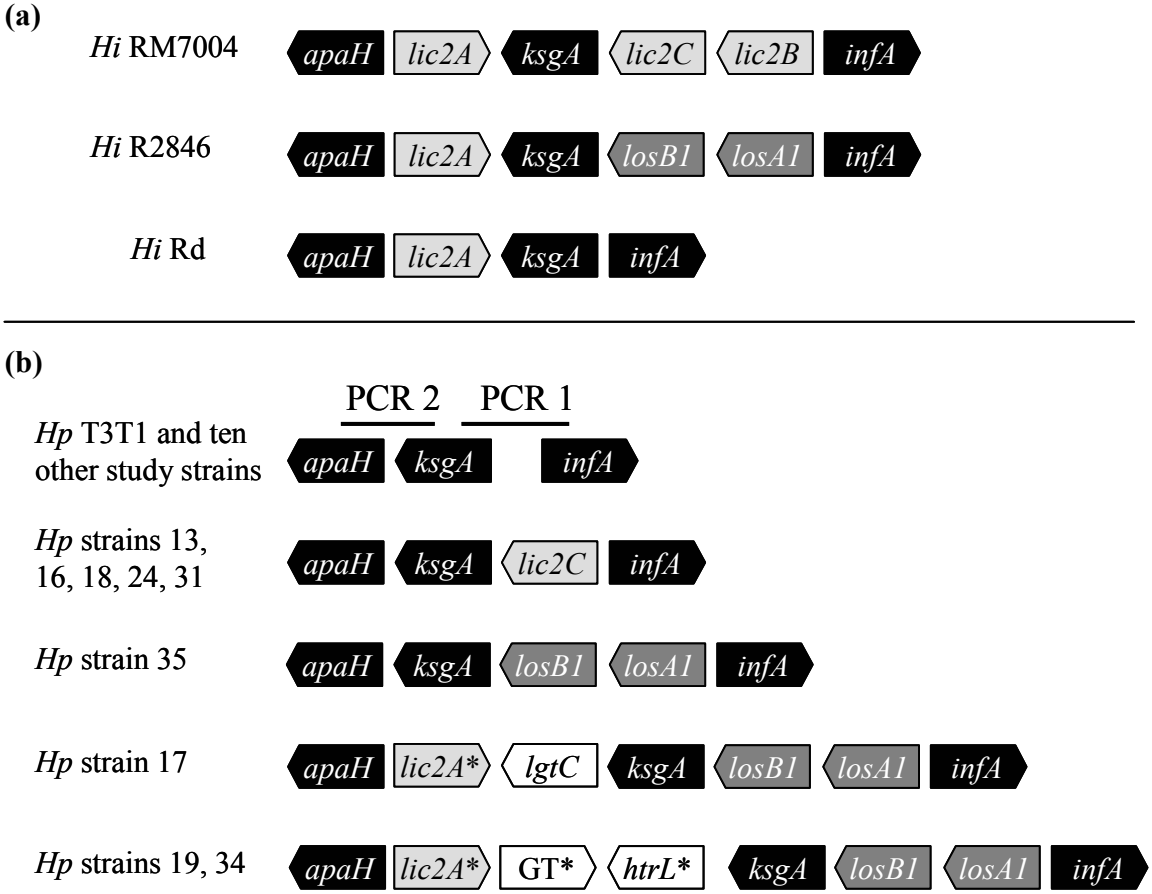


Figure 3.2. Organisation of the *lic2* locus in *Hi* and *Hp*. Genes flanking the compound locus are shown in black, and genes related to LPS synthesis in grey (typical *Hi lic2* locus genes) or white (genes not previously found at this locus). **(a)** The three known layouts of the locus in *Hi* are represented here by strains RM7004, R2846 and Rd. *Lic2A* is invariably present in this location in *Hi*; the second part of the locus contains either *lic2B* and *lic2C*, *losA1* and *losB1*, or no genes. **(b)** The organisation of the *lic2* locus in all 20 *Hp* study strains as deduced by PCR amplification and sequencing. *Hp* genes with closer homologues in *Aggregatibacter actinomycetemcomitans* than in *Hi* are indicated by asterisks. Not drawn to scale.

Although the *Hp* T3T1 genome does not encode homologues of these enzymes, the situation for other *Hp* strains was unknown. The *Hp* 20 contigs in the Genomophilus database were searched for homologues to *Hi* strain 375 *lpsA*, *Hi* strain 176 *lpsB*, *Hi* strain 486 *lpsC* and *Hi* strain 1292 *lpsD* using BLASTN and TBLASTN algorithms. For all eight searches the closest match was to the partial ORF at the beginning of *Hp* 20 contig 319, with the highest E-value being 4E-86 (TBLASTN using *Hi* 176 LpsB). The gene following the *lpsB* homologue in this contig has 95% aa homology to *Hp0884* in the *Hp* T3T1 genome.

Primers 883b and 884b were designed to the *Hp* T3T1 genes *Hp0883* (*menB*) and *Hp0884* (*menC*) and used in PCR on the *Hp* strain panel. For thirteen strains a 0.5 kb product was amplified, indicating that the genes were adjacent as in *Hp* T3T1. This is likely to be the ancestral gene arrangement, as MenB and MenC are involved in the same biosynthetic pathway and the genes are also contiguous in *E. coli* and *Hi*. For the remaining five strains (*Hp* 8, 10, 15, 20 and 35) the amplified products were 1.4 kb in length; DNA sequencing confirmed that intact *lpsB* homologues were present in this genomic location in all five strains.

In *Hi* the amino acid at position 151 of the different LpsA alleles determines whether the enzyme adds Gal (threonine) or Glc (cysteine, alanine or methionine) to the distal heptose of the LPS core. All five of the *Hp* homologues encode threonine at this position (Fig. 3.3). In addition, unidentified residues within the C-terminal region are known to determine the linkage specificity of these enzymes in *Hi*; all five of the *Hp* homologues have a generally Rd-like (β1,2) rather than *Hi* 486-like (β1,3) C-terminal region (Fig. 3.3). Together this information suggests that the LPS of *Hp* 8, 10, 15, 20 and 35 may be elaborated with Gal via a β1,2 linkage on the third heptose.

Structural analysis has confirmed that this is indeed the case for *Hp* 20 (Elke Schweda, pers. comm.). No extensions were found from the third heptose in *Hp* 26 or *Hp* 22 LPS during structural analysis, consistent with no *lpsA*, *lpsB*, *lpsC* or *lpsD* being detected in these strains.

The monoclonal antibody L6A9 detects the Gal-(β1,2)-HepIII epitope on *Haemophilus* LPS (Hood et al., 2004a). Dot immunoblots using boiled lysates of all 20 study strains showed that only *Hp* strains 8, 10, 15, 20 and 35 reacted with this antibody (Fig. 3.4), corresponding to the five strains containing an *lpsB* homologue as detected by PCR. In addition, the intense signal suggests that the epitope is unlikely to carry further substitutions in these strains. Combined with the data obtained from LPS structural analysis, the immunoblotting results indicate that the structure-function relationships elucidated for this glycosyltransferase family in *Hi* also hold true in *Hp*.

Hi lpsA alleles are located between *ribB* and *yibK*, but in *Hp* these two genes are distant and the *lpsA* alleles identified were at a different genomic location. It is possible that a single horizontal transfer of *Hi lpsB* into an *Hp* strain occurred and that this gene subsequently spread in *Hp* through either clonal expansion or further horizontal transfer and recombination events.

LpsA in one NTHi strain, 285, contains three tandem ACAGT repeats 5 bp after the start codon resulting in a truncated protein (Deadman et al., 2006). None of the five *Hp lpsA* sequences obtained contains tandem repeats in this region, suggesting that in *Hp* (as in the majority of *Hi* strains) LpsA is not phase variably expressed.

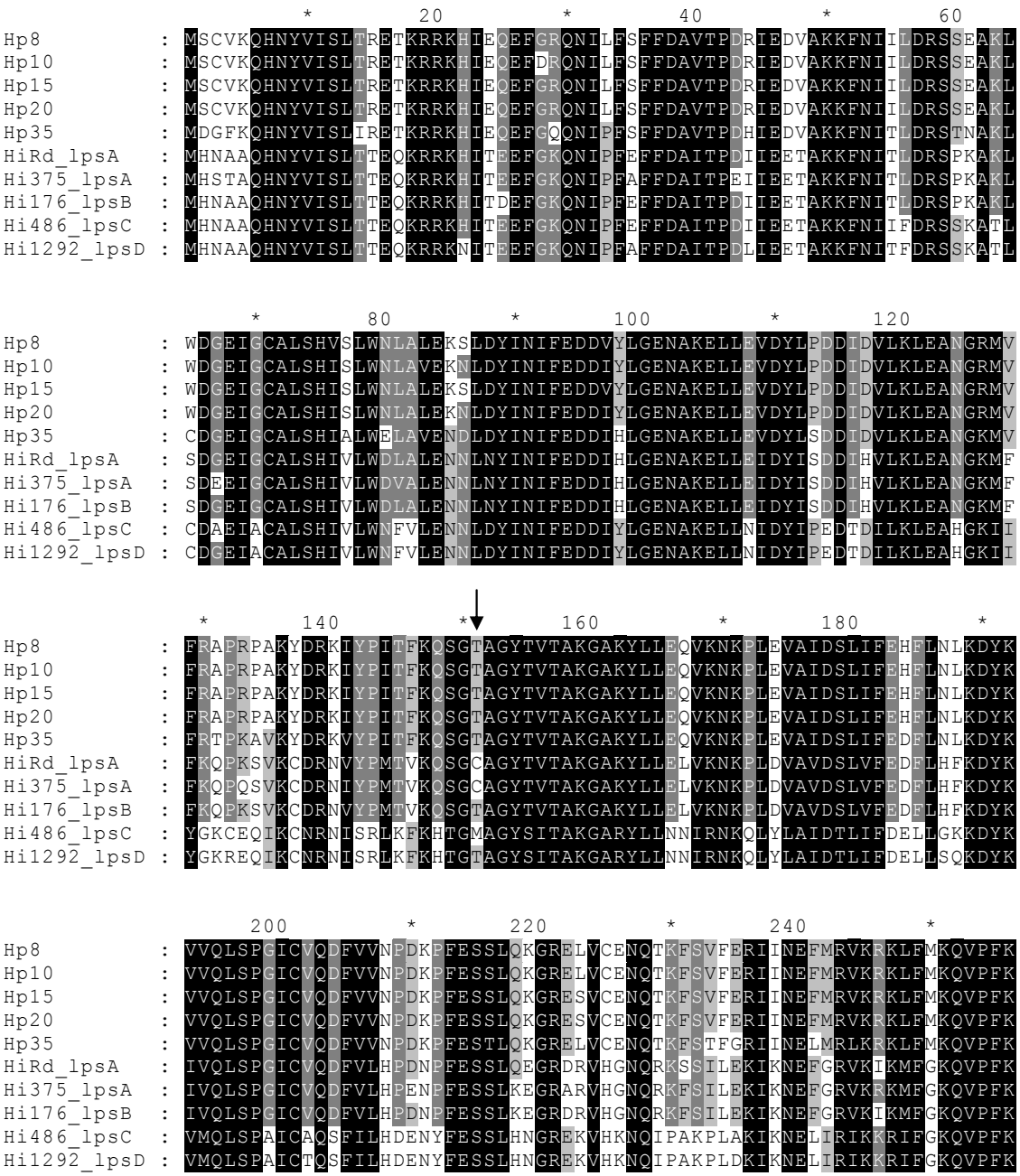


Figure 3.3. Amino acid alignment of LpsA alleles. The *Hp* LpsA alleles sequenced during this project are aligned against *Hi* sequences from the NCBI database to examine features which relate to sugar and linkage specificity. Darker shading indicates increased sequence conservation. Aa151, which determines specificity for Gal or Glc in the different *Hi* alleles, is indicated by an arrow. The *Hp* sequences encode threonine at this position, like LpsB and LpsD. The C-terminal half of the *Hp* sequences show more similarity to LpsA and LpsB than to LpsC and LpsD. Overall this suggests that the *Hp* sequences share the same function as *Hi* LpsB (the addition of Gal to the LPS via a β1,2 linkage).

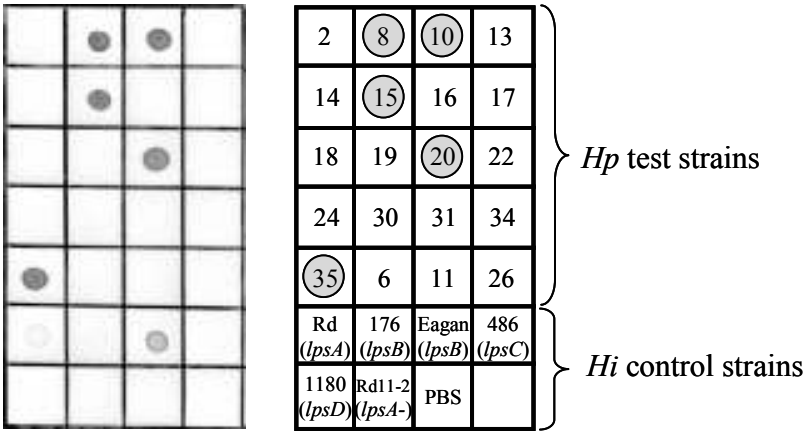


Figure 3.4. Dot immunoblot using L6A9 antibody to detect the Gal β1,2 LPS epitope. On the immunoblot (left), each square contains 3 µl boiled cells in PBS. *Hp* strains and *Hi* control strains are listed in the key on the right. Grey circles on the key indicate *Hp* strains in which *lpsB* was found; these correlate with L6A9 reactivity. Of the two *lpsB*-containing control strains, *Hi* Eagan gives a strong signal but L6A9 binding for *Hi* 176 may be inhibited by the presence of a substituent of the Gal, such as PCho.

Internal primers designed to *Hi* Rd *lpsA* did not give PCR products in any of the *Hp* panel of strains, and PCR using internal primers designed to *Hp* 20 *lpsB* did not amplify this gene from any *Hp* strain that had not already been identified as positive (Table 3.3). It is still possible that some additional *Hp* strains carry an undetected *lpsA* allele at a different locus, as there may be heterogeneity at the primer binding sites within the gene.

3.3 Other *Hi* LPS outer core substituents

The great variety of LPS structures found in *Hi* is largely dependent on differences in the presence and location of outer core substituents such as sialic acid, phosphoethanolamine, phosphocholine, O-acetyl groups, heptose, Glc and Gal. These vary between strains due to different complements of sugar transferase genes, but also within strains because many of these

transferases are phase variably expressed using tetranucleotide repeat-mediated phase variation (Table 3.2). Competition between enzymes for the same acceptor site introduces an additional level of complexity. Many outer core LPS substituents play key roles in the biology and virulence of *Hi* so it is important to establish whether *Hp* has a similar capacity to decorate its LPS with these molecules.

3.3.1 Sialic acid: Lic3A, Lic3B and other sialyltransferases

The sialyltransferase Lic3A is encoded by all tested NTHi strains, and 60% of these strains also encode an additional sialyltransferase, Lic3B (Fox et al., 2006; Hood et al., 2001a). Both proteins decorate lactose moieties of the LPS with outer core sialic acid in a phase variable manner. *Lic3A* and *lic3B* are highly homologous and both can be amplified from *Hi* gDNA using the primer pair lic3g and OS2560. However, these primers did not amplify any products from the *Hp* study strains and the *Hp* T3T1 genome does not contain any *lic3A* or *lic3B* homologues (Table 3.2).

The *Hp* T3T1 homologues of the genes which flank *Hi* Rd *lic3A* (*HI0352*) are in separate genomic locations and are not adjacent to any sialyltransferase genes. In contrast the *Hp* T3T1 homologues of the *lic3B* (*NTHI1034*) flanking genes, *glnA* and *typA*, are adjacent so this locus could be examined in the *Hp* study strains. PCR amplification using flanking primers 286lic3 and 287lic3 amplified a 600 bp product for all 20 study strains, indicating that in *Hp* this locus does not contain *lic3B* or indeed any other genes. *GlnA* and *typA* are adjacent in six out of seven diverse Pasteurellaceae genome sequences examined; the exception is *P. multocida* Pm70, in which the genes are separated by a *lic3B* homologue as in *Hi* (data not shown). The most

parsimonious explanation is that *glnA* and *typA* are ancestrally contiguous and that some *Hi* and *P. multocida* strains have subsequently acquired *lic3B*.

The *Hi* Rd genome encodes two further sialyltransferases. SiaA is encoded within the *hmg* locus and adds the terminal sugar of the HMG unit, whilst LsgB is a product of the *lsg* locus whose function is undefined. TBLASTN searches of the *Hp* T3T1 genome found no SiaA homologues and the only match for LsgB was the OAg locus sialyltransferase Hp0278 with 34% aa identity (Table 3.2).

In conclusion, a single sialyltransferase gene was identified in the *Hp* T3T1 genome and its location within the OAg locus suggests that it adds sialic acid to the LPS as part of the OAg unit rather than as an outer core substituent (see Section 4.1.2). No *lic3A* or *lic3B* homologues were detected in any of the *Hp* study strains; the presence of sialyltransferase genes in their OAg loci will be examined in Chapter 5.

3.3.2 Phosphoethanolamine: Lpt6 and Lpt3

All *Hi* strains encode the non-phase variable phosphoethanolamine (PEtn) transferase Lpt6, enabling them to decorate their LPS with PEtn via the 6 position on HepII (Wright et al., 2004). The closest *Hp* T3T1 homologue of this protein is Hp0234 which shares 71% aa identity with *Hi* Rd Lpt6 (HI0275). Southern analysis using a fragment of the *Hp0234* gene as a probe indicated that this gene is present in all 18 true *Hp* strains (Fig. 3.1 and Table 3.3). It is therefore likely that HepII of *Hp* LPS is substituted with PEtn-6. This has been confirmed for *Hp* T3T1 by structural analysis (Elke Schweda, pers. comm.).

32% of NTHi strains also encode a phase variable PEtn transferase, Lpt3, which adds PEtn to the distal heptose of the LPS (Derek Hood, pers. comm.). No *lpt3* homologues were detected in any of the *Hp* study strains by PCR with the *Hi* primers lpt3midF and lpt3midR (Table 3.3). In *Hi*, *lpt3* is located between *purD* and *glyA*; this locus contains three ORFs in *Hp* T3T1. Two of the ORFs are paralogues which appear to have arisen through gene duplication and share homology with a putative lipoprotein from *Aggregatibacter aphrophilus*, whilst the third encodes a hypothetical protein of unknown function. Primers 159a and 163a, designed to the flanking genes in *Hp* T3T1, were used to amplify the locus from all 18 *Hp* strains. The *Hp* T3T1 PCR product was 2.8 kb in length; thirteen other strains gave similar sized products whose PvuI restriction patterns were consistent with the *Hp* T3T1 gene layout (data not shown), and the remaining four strains gave a 2 kb product consistent with no duplication (this was confirmed for *Hp* 2, using Genomophilus data). Thus no *Hp* strain encodes any recognisable LPS-related protein at this position in the genome. To confirm that the three ORFs are not involved in LPS synthesis, gene deletion experiments and LPS analysis could be carried out in the future.

Two other genes in *Hi* Rd, *HI1005* and *HI1064*, are annotated as PEtn transferases; HI1005 adds PEtn to the HMG unit. The strongest *Hp* T3T1 homologue of Lpt3, HI1005 and HI1064 is Hp1006, which shares the most homology (76% aa identity) with HI1064 and is likely to perform a similar function (Table 3.2). The specific role of HI1064 is unknown: it may add PEtn to lipid A, to the sugar moiety of LPS or to another molecule such as a protein.

In conclusion, the *Hp* T3T1 genome includes a homologue of *lpt6* and one other putative PEtn transferase gene which may not be LPS-related.

3.3.3 Phosphocholine: Lic1A-Lic1D

In *Hi* the uptake of choline and its incorporation into the LPS as phosphocholine (PCho) is undertaken by the products of the *lic1* locus, Lic1A to Lic1D (Weiser et al., 1997). The expression of the choline kinase Lic1A, and therefore of the PCho epitope, is phase variable due to the presence of CAAT tandem repeats in *lic1A*. No homologues of the four Lic1 proteins were found in TBLASTN searches of the *Hp* T3T1 genome.

An 800 bp PCR product amplified from the *Hi* Rd *lic1B* gene, which encodes the choline transporter Lic1B, was used as a probe in Southern analysis against gDNA digests from the *Hp* study strains. No hybridising bands were detected for any of the 18 *Hp* strains or the two hybrids (Fig. 3.1 and Table 3.3), indicating that *lic1B* is either absent or highly diverged from the *Hi* Rd version. Further evidence for the lack of PCho in *Hp* LPS is presented in later chapters, and the absence of this molecule is the subject of Section 6.2.

3.3.4 O-acetyl groups: OafA

The phase variable gene *oafA* is ubiquitous across NTHi strains and its product O-acetylates the third heptose of the LPS core (Fox et al., 2005). Despite its apparent importance in *Hi*, no *oafA* homologue was detected in the *Hp* T3T1 genome in TBLASTN searches (Table 3.2) and no *oafA* hybridising bands were observed by Southern analysis for the 20 study strains (Fig. 3.1).

Additional TBLASTN searches using putative O-acetyl transferase sequences from a range of Gram negative bacteria against *Hp* T3T1 identified a single homologue, Hp0280, which is part of the OAg locus so is likely to acetylate only the O-unit rather than the LPS core (data not shown).

3.3.5 Fourth heptose and subsequent additions: LosA and LosB

The Glc residue on HepI of *Hi* LPS can be elaborated with a fourth heptose and a subsequent hexose. The transferases encoded by two paralogous pairs of genes are responsible for the addition of this epitope (Hood et al., 2010). The products of *losB1* and *losA1* add DDHep and Glc respectively; these ORFs are found in approximately 21% of NTHi strains and are always flanked by *ksgA* and *infA*. In contrast *losB2* and *losA2* are found in 35% of NTHi strains, are in a different genomic location, and encode enzymes which add LDHep and Gal respectively.

The *Hp* T3T1 genome does not encode any strong LosA1 and LosB1 homologues, and as explained in Section 3.2.2 the *ksgA-infA* intergenic region contains no ORFs in this strain. However, sequencing of the complete *ksgA-infA* PCR product from strain *Hp* 17 revealed two ORFs with homology to *Hi* R2846 *losA1* (81% aa identity) and *losB1* (92% aa identity) respectively. Further PCR analysis of the three other strains with the same size *ksgA-infA* PCR product as *Hp* 17 (*Hp* 19, 34 and 35) suggested that they also encode LosA1 and LosB1 at this locus (Fig. 3.2).

Work in our laboratory has indicated that particular residues in both *Hi* LosA and *Hi* LosB are likely to distinguish the sugar specificity of paralogues 1 and 2 (Mary Deadman, pers. comm.). LosA contains several motifs with sequence similarity to those in LpsA, and with respect to these

aa161 of LosA aligns to aa151 of LpsA which is known to confer specificity for Gal or Glc (Section 3.2.3). Across *Hi*, aa161 of the glucosyltransferase LosA1 is alanine, whereas the galactosyltransferase LosA2 consistently contains threonine at this position. Aligning the *Hp* 17 LosA with these sequences reveals an alanine residue at aa161, suggesting a specificity for adding Glc (Fig. 3.5). This is consistent with its higher overall sequence similarity to LosA1 than to LosA2 and with its LosA1-like genomic location. The central 75% of *losA1* was also sequenced from *Hp* 19 and *Hp* 34: both share 98% aa identity with the *Hp* 17 protein, and as in *Hp* 17 aa161 encodes alanine in these two strains.

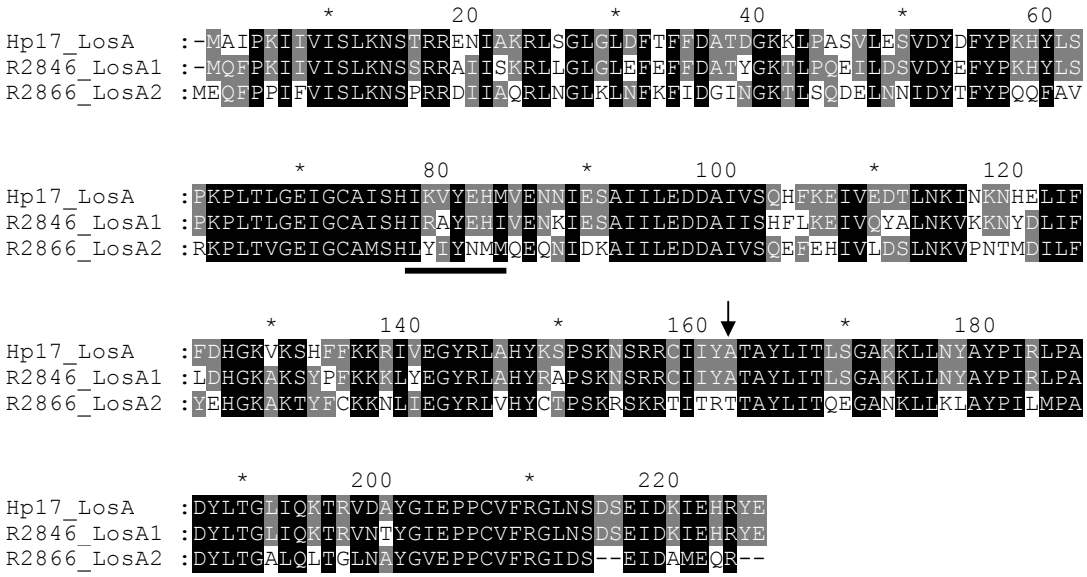


Figure 3.5. Amino acid alignment of the glycosyltransferase LosA. Residue 161, whose identity is thought to determine whether a Glc or Gal is added by LosA, is indicated by an arrow. The horizontal bar represents the octanucleotide repeat region in *Hi* R2846_LosA1, which is permissive for translation in this strain but introduces a frameshift in some *Hi* strains.

In a study of 30 NTHi strains, Erwin et al. (2006) identified an octanucleotide tandem repeat sequence, $(CGAGCATA)_n$, within *losA1*. In 24 of the strains there were only two repeats and the ORF was intact; however, the remaining 6 strains had between 3 and 10 repeats, resulting in a frameshift and premature termination of translation. This suggests that *losA1* with two repeats may 'phase vary' at an extremely low rate, although this may not involve the slipped strand mispairing mechanism that is implicated in tetranucleotide repeat phase variation. By aligning the *Hp* 17, *Hp* 19, *Hp* 34 and *Hi* R2846 *losA1* sequences over this region it was observed that in *Hp* the repeats are replaced by a non-repeating 16 bp sequence whilst still leaving the ORF intact (shown for *Hp* 17 in Fig. 3.5).

The *Hi* DD-heptosyltransferase, LosB1, and LD-heptosyltransferase, LosB2, are typically only 55% identical, making it difficult to predict which amino acids are responsible for their different functions. However, the protein encoded at the *Hi* 1158 *losB2* locus has a single amino acid difference from other *Hi* LosB2 sequences, and this appears to confer LosB1-like specificity as the LPS contains only DDHepIV (Mary Deadman, pers. comm.). The *Hp* 17 LosB sequence shares 92% aa identity with *Hi* R2846 LosB1 including the differentiating amino acid, so it is likely that it can add DDHep to the *Hp* 17 LPS (data not shown).

The only *Hp* T3T1 TBLASTN matches for LosA2 and LosB2 were LsgD and heptosyl transferases that had already been identified as those involved in triheptosyl core LPS synthesis, respectively (Table 3.2). Where present in *Hi*, *losA2* and *losB2* form part of a ten gene island which is also found in *P. multocida* (Hood et al., 2010). This entire island appears to be absent from the *Hp* T3T1 genome, whilst homologues of the genes flanking the island are present but

500 kb apart (data not shown). The other *Hp* study strains have not been checked for the presence of the island.

To summarise, 4/18 *Hp* strains (22%) encode LosA1 and LosB1 at the *ksgA-infA* locus; this is comparable to the frequency in NTHi (21%). In *Hi* these enzymes add a Glc-DDHep- epitope to the Glc on the proximal heptose of the inner core via an α1,6 linkage, and it is likely that they have the potential to perform the same function in *Hp*, especially as this exact extension was recently observed in the two COPD *Hp* isolates studied by Pollard et al. (2008).

3.3.6 Digalactoside: Lic2A and LgtC

Digalactoside epitopes have been found extending from a second Glc on HepI and HepII of *Hi* LPS, and from a Glc on HepIII. Lic2A adds the first Gal via a β1,4 linkage to the accepting Glc, and LgtC extends this with a second Gal via an α1,4 linkage (Griffin et al., 2005). The expression of both enzymes is phase variable.

There are no close homologues of either Lic2A or LgtC in *Hp* T3T1 (Table 3.2; Hp0862 shares only low homology with LgtC and is discussed below). *Lic2A* is present in all NTHi strains between the *apaH* and *ksgA* genes, i.e. adjacent to the *ksgA-infA* locus (Fig. 3.2). To examine whether *Hp* also encodes *lic2A* in the same genomic location, primers 933b and 934b2 were designed to the *Hp* T3T1 homologues of *apaH* and *ksgA* and used to amplify the locus from the study strains (Fig. 3.2, PCR 2). In *Hp* T3T1 there is no ORF between these two flanking genes; the PCR analysis indicated that indeed most *Hp* strains share this gene layout, with only *Hp* 17, 19 and 34 yielding larger products.

The *Hp* 17 *apaH-ksgA* region was fully sequenced and those of *Hp* 19 and *Hp* 34 were partially sequenced, using additional PCR amplification where necessary. The loci deduced are pictured in Fig. 3.2. The first ORF in the *Hp* 17 locus encodes a protein with 75% aa identity to D11S_0429 (Lex1) from *Aggregatibacter actinomycetemcomitans* (*Aa*) D11S-1 and 56% aa identity to HI0550 (Lic2A) from *Hi* Rd. The role of this protein in *Aa* is not clear but Lex1 is the former name of Lic2A in *Hi*. The second ORF encodes a protein with 50% aa identity to LgtC from *Hi* Rd (HI0258). The discovery of convergent *lic2A* and *lgtC* homologues in *Hp* 17 was striking, as these two digalactoside synthesis enzymes are always found at separate loci in *Hi*.

The *Hp* 19 and *Hp* 34 *apaH-ksgA* regions each contain three genes which are homologous to D11S_0429, D11S_0428 and D11S_0427 of *Aa* D11S-1; it is highly likely that these three contiguous genes have transferred between species *en bloc*. They are part of the 6 kb 'Genomic island G – LOS biosynthesis enzymes' region of the *Aa* D11S-1 genome, which has recently been sequenced (Chen et al., 2009). Precise functions have not been ascribed for these ORFs in *Aa*, but the D11S_0427 homologue belongs to the HtrL superfamily of unknown function (often found in LPS loci) and the D11S_0428 homologue contains a glycosyltransferase DXD motif and is related to capsule synthesis proteins.

The reaction of monoclonal antibody 4C4, which detects LPS digalactoside, with the *Hp* study strains was tested using a dot immunoblot (Fig. 3.6). The only reactive strain was *Hp* 17. This finding is consistent with the genetic evidence for *lic2A* and *lgtC* homologues in this single strain and suggests that the two genes are likely to express functional proteins. As the *lic2A/lex1*

homologue is conserved in *Hp* 17, 19 and 34 it is likely that all three of these strains can decorate their LPS with Gal; the other two proteins encoded by the *Hp* 19 and 34 loci could potentially add another one or two sugars to the LPS but their sugar specificity is currently unknown.

Given that the three *Hp* strains with a *lic2A* homologue (*Hp* 17, 19 and 34) correspond to a subset of the four strains that contain *losA1* and *losB1*, it is possible that Lic2A adds a Gal residue to the Glc-DDHep extension in these isolates. Two observations endorse, but do not prove, this hypothesis. Firstly, a Gal-Glc-DDHep- epitope has already been observed in *Hp* strain 4282 as the only outer core LPS structure (Pollard et al., 2008); the authors did not test for the presence of any LPS synthesis genes in this isolate. Secondly, *losB1* mutant strains described later in this thesis do not produce any detectable glycoforms with short outer core extensions (Section 6.1.5), suggesting that the action of Lic2A is dependent upon that of LosB1.

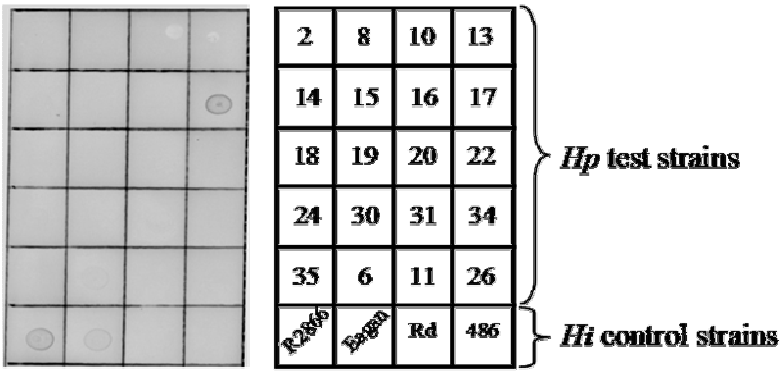


Figure 3.6. Dot immunoblot using 4C4 monoclonal antibody to detect the digalactoside LPS epitope. On the immunoblot (left), each square contains 3 µl boiled cells in PBS. *Hp* strains and *Hi* control strains are listed in the key on the right. *Hi* R2866, Eagan and Rd LPS contain the diGal epitope, but it is phase variable and in Rd it is masked by an additional GalNAc extension. *Hi* 486 does not express the diGal epitope (negative control). *Hp* 17 is the only 4C4-reactive *Hp* strain and is the only strain in which homologues of both *lic2A* and *lgtC* have been identified.

In *Hi* the digalactoside epitope can be extended through the addition of GalNAc by LgtD (Hood et al., 2001b). As *Hp* strain T3T1 does not appear to express digalactoside on its LPS we would not expect to find an *lgtD* homologue within its genome, and indeed the closest match (*Hp0317*) is only 32% identical on the amino acid level over 68% of the protein length. *Hp* strain 17, on the other hand, is thought to add digalactoside to its LPS so there is a possibility that it may synthesise the GalNAc transferase LgtD. This will be easier to test if the genome sequence or structural LPS data become available for this strain in the future.

3.3.7 Other glycosyltransferases

The phase variable *lex2A/lex2B* locus is present in the majority of NTHi strains and its products glycosylate the conserved Glc on HepI (Deadman et al., 2009). No Lex2A or Lex2B homologues were identified in the translated *Hp* T3T1 genome. In *Hi*, *lex2A/lex2B* are always located between *purL* and *dat1*; the *Hp* T3T1 homologues of these genes are 280 kb apart and are not flanked by any LPS-related genes. A *lex2A/lex2B* DNA probe did not hybridise to any bands in DNA digests from the 18 true *Hp* study strains (Fig. 3.1). Together this evidence suggests that the *lex2A/lex2B* locus is absent in *Hp*. It may, however, be present in the two hybrid strains included in this study: the *lex2A/lex2B* probe hybridised to an *Hp* 11 DNA fragment, and Genomophilus data for *Hp* 6 revealed a fragment at the end of contig DCO-OM33-2-B3ST-21_359 encoding the C-terminal 42 aa of a protein with 71% aa identity to NTHI0913 (Lex2B).

The *Hi* R2846 gene R2846(v6)_916 encodes a putative glycosyltransferase, Pgt1, that is likely to be phase variably expressed but whose precise function has yet to be determined. The genes flanking *pgt1* (*holC* and *fumC*) are adjacent in the *Hp* T3T1 genome and also in four *Hp* strains

included in Genomophilus (*Hp* 8, 15, 20 and 22). An *Hp* T3T1 TBLASTN search with Pgt1 identified only a weak homologue, Hp0862: this protein has closer homology to family 8 glycosyltransferases from *Actinobacillus minor*, *H. ducreyi* and *Streptococcus pneumoniae* and probably has different specificity to Pgt1 (see also Section 7.1.3).

3.3.8 *Lsg* locus

The 7.4 kb *Hi lsg* locus contains the seven genes *lsgA* to *lsgG*. Phillips et al. (1996) used a random mutagenesis strategy to show that mutations in the Hib strain A2 *lsg* locus result in fewer LPS glycoforms visible by silver stained SDS-PAGE, and the same group later demonstrated that the *Hi* glycosyltransferases LsgC to LsgF were together capable of adding the extension Gal-(1,4)-GlcNAc-(1,3)-Gal-(1→ to the 3 position of GlcNAc on an undecaprenyl phosphate carrier in *E. coli* (Johansen et al., 2010). However, the enzymes necessary for synthesis of the known LPS structures in *Hi* have already been accounted for by genes elsewhere in the *Hi* genome, so the role of the *lsg* locus is unclear.

Hp T3T1 appears to have retained the central part of the *lsg* locus: genes *Hp1755*, *Hp1753* and *Hp1752* encode homologues of LsgD, LsgE and LsgF respectively, with aa identity to *Hi* in the range of 70-89%. An extra gene in the *Hp* T3T1 locus, *Hp1754*, has no homologues in *Hi*. All four genes have homologues with 93% aa identity in *Aggregatibacter aphrophilus* NJ8700 and the position of the extra gene is conserved between the two species and also in *H. ducreyi*. The closest *Hp* T3T1 TBLASTN matches for the flippase LsgA (HI1700) and the sialyltransferase LsgB (HI1699) are found in the OAg locus rather than the *lsg* locus. No homologues were found

for the glycosyltransferase LsgC, or for LsgG, a putative transcriptional regulator which may be involved in the regulation of *wecA* (Johansen et al., 2010).

3.4 Visualisation of LPS glycoforms by tricine SDS-PAGE

The relative sizes of *Haemophilus* LPS glycoforms can be visualised by tricine SDS-PAGE fractionation of proteinase K treated cell lysates. *Hi* LPS profiles typically display several glycoforms due to the cumulative action and phase variable expression of multiple outer core transferases. In addition there is a large degree of interstrain heterogeneity between profiles (Fig. 3.7a) as the complement of genes encoding outer core LPS related enzymes, and their predominant phase variation state in the population, varies between strains. In contrast, the LPS profiles of the *Hp* study panel (Fig. 3.7b) are much less varied and are dominated by glycoforms of two specific molecular weights. The LPS profiles of three typical *Hp* strains are included in Fig. 3.7a for comparison purposes.

3.5 Discussion

The data presented in this chapter enable us to start to understand the nature of the *Hp* LPS profiles shown in Fig. 3.7. Analysis of the *Hp* T3T1 genome revealed homologues of all the genes necessary for the synthesis and assembly of an LPS backbone comprising lipid A, KdoP and three Hep residues. The genome was then searched for homologues of all other known *Hi* LPS transferases, which add substituents such as hexose, sialic acid, PCho, PEtn and OAc. The only homologues identified outside the *lsg* and OAg loci with E-values under 1E-60 were LgtF, Lpt6 and a PEtn transferase of unknown specificity. In *Hi*, LgtF adds Glc to HepI, and Lpt6

elaborates HepII with PEtn. This LPS backbone, most of which is well conserved across Pasteurellaceae species, will be referred to as 'inner core LPS' from now on. The definition will include a PEtn residue which extends from KdoP in *Hi* but for which no transferase has been identified.

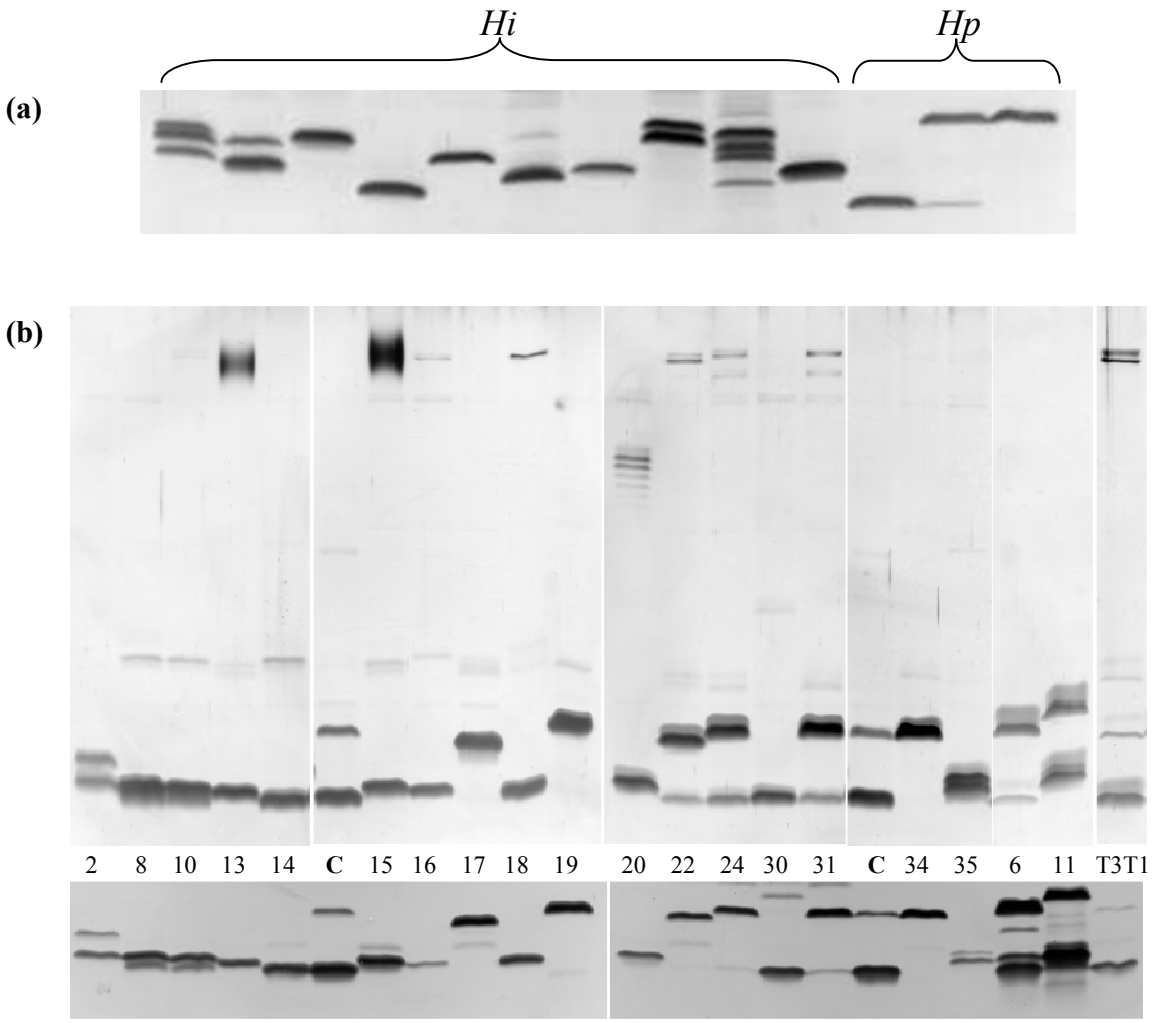


Figure 3.7. LPS profiles of *Haemophilus* strains. Proteinase K treated cell lysates were fractionated by tricine SDS-PAGE and silver stained. **(a)** LPS core region profiles of representative *Hi* and *Hp* strains. *Hi* strains from left to right are Rd, Eagan, and eight NTHi (176, 285, 486, 667, 1180, 1231, 1268, 1292). *Hp* strains are 14, 24 and 34. **(b)** Full (upper panel) and core region (lower panel) LPS profiles for the 20 *Hp* study strains. 30 µl of each sample was loaded on the upper panel, allowing the visualisation of OAgs in some strains; 20 µl of each sample was loaded on the lower panel, allowing clearer band resolution of the core region. Strain numbers are given across the centre and apply to both panels. An *Hi* Rd *lpsA lic1* mutant that was grown on BHI medium with added sialic acid is included for comparison (lane C): the lower band is inner core LPS and the upper band is inner core LPS plus HMG.

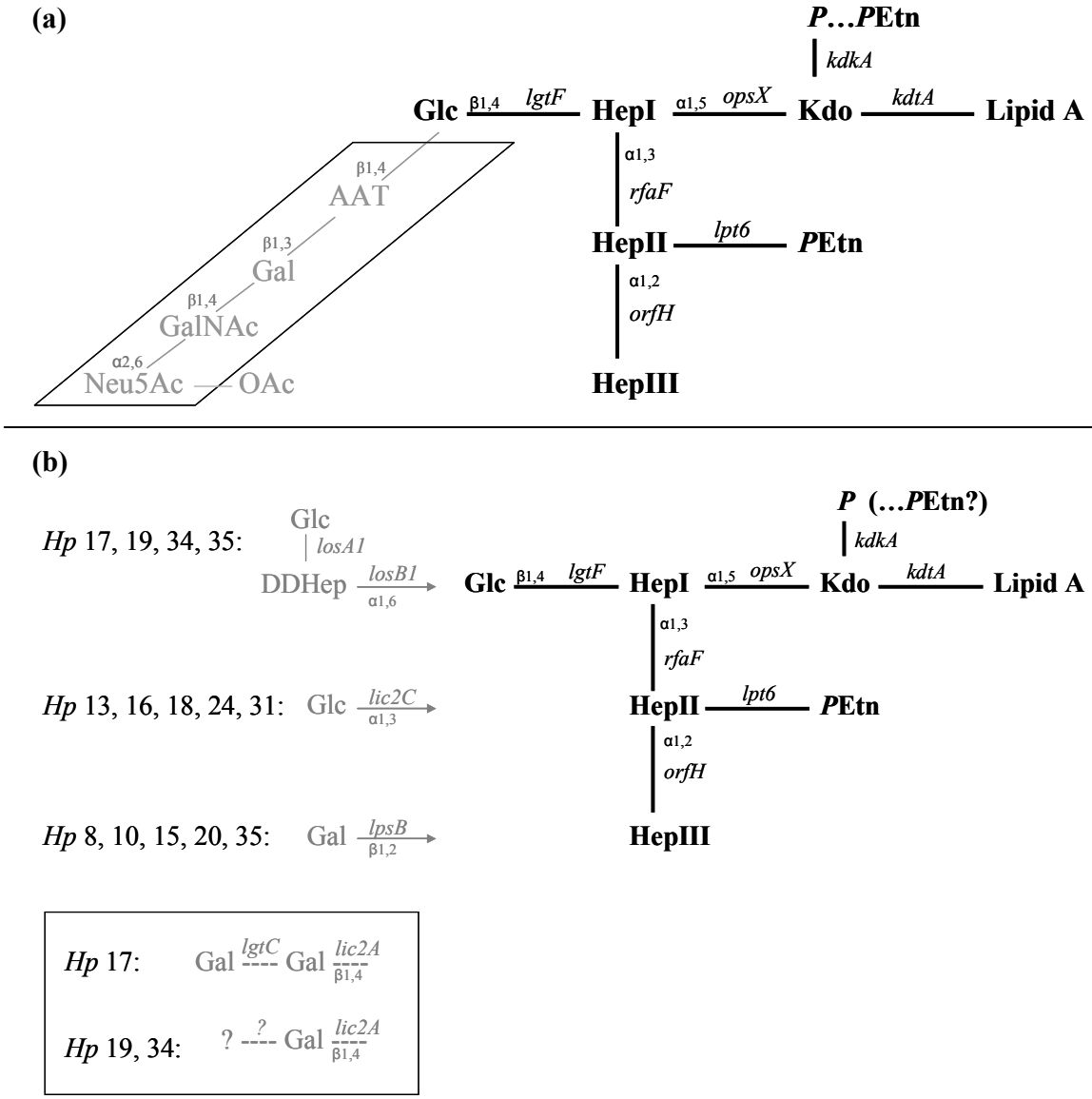


Figure 3.8. LPS structure of *Hp* T3T1 and putative core LPS structures of the study strains. Bold type represents the conserved inner core LPS. Hep residues are LDHep except where stated. **(a)** *Hp* T3T1 LPS structure as determined by mass spectrometry and NMR techniques (Elke Schweda, pers. comm.). A single O-unit is denoted by a parallelogram. Note the lack of outer core substituents. **(b)** Putative core LPS structures of the 17 other true *Hp* study strains based on genetic evidence. Outer core sugars and the genes encoding their transferases are shown in grey alongside their predicted points of addition; *Hp* strains with homologues to each gene are listed to the left. The position of the boxed outer core additions cannot be predicted. No outer core genes were detected for *Hp* strains 2, 14, 22 or 30.

Structural analysis using mass spectrometry and NMR techniques confirmed that the LPS of *Hp* T3T1 is composed only of inner core LPS and OAg (Fig. 3.8a; Elke Schweda, pers. comm.), in agreement with the genetic data presented here. As yet there is no structural evidence for the presence of Kdo_2-containing glycoforms. It seems unlikely that the putative glycosyltransferases encoded by the *Hp* T3T1 genes *lsgD*, *lsgE*, *lsgF*, *Hp0317* and *Hp0862* are involved in LPS assembly, as we have already accounted for all the glycosyltransferases necessary for synthesis of the observed core structure and the OAg enzymes are encoded separately (see Section 4.1).

To aid interpretation of the *Hp* LPS profiles, an *Hi* strain of known LPS structure was loaded on the gels in Fig. 3.7b. The lower band in lane C (an *Hi* Rd *lpsA lic1* mutant grown on media containing sialic acid) represents inner core LPS only, whereas the upper band represents inner core plus the four-sugar HMG. By comparison it is likely that the lower glycoform of the *Hp* T3T1 profile is inner core whilst the higher molecular weight glycoform is inner core plus approximately four sugars, which would correspond to one O-unit as shown in Fig. 3.8a and would agree with the structural data.

PCR and Southern analyses showed that the inner core LPS structure is likely to be conserved in all of the *Hp* study strains, suggesting evolutionary constraints on this part of the molecule. All of the *Hp* LPS glycoforms visualised by tricine SDS-PAGE are at least the molecular weight of *Hp* T3T1 inner core LPS (Fig. 3.7b), consistent with the genetic evidence.

In contrast, the lack of outer core sugars in the *Hp* T3T1 LPS does not appear to be typical of the species. 13 of the 18 true *Hp* study strains encode at least one putative outer core LPS

transferase, with potential additions including Gal, Glc and DDHep (Fig. 3.8b). Immunoblotting was performed to detect some of these additions and was consistent with the genetic evidence. Notably none of the *Hp* transferases appear to be phase variable; this will be explored in detail in Chapter 7. Whilst successful PCR amplification of an outer core LPS gene is a good indicator of its presence, a negative PCR result can mean either that the gene is absent or that it lacks enough homology to the primers which have been designed to a different strain or species. Some *Hp* strains may also contain novel outer core LPS genes, so the list of *Hp* outer core glycosyltransferases may not be complete.

In *Hi*, LPS outer core transferases are encoded throughout the genome with a few clustered and many single genes. In several instances, extensive rearrangements of the *Hi* genome with respect to that of *Hp* T3T1 made it impossible to examine the equivalent loci in *Hp*, as the flanking gene homologues were often several hundred kilobases apart. However, the positions of *apaH*, *ksgA* and *infA* were conserved and the genes present in the intervening loci were examined in all 20 study strains. This region of the *Haemophilus* genome appears to be favourable for many different LPS genes; perhaps sequence similarity between glycosyltransferase genes facilitates recombination, or the genes could be co-regulated more easily at a single locus (although some of the genes are convergent in orientation).

It is possible to correlate the presence of outer core glycosyltransferases with the LPS profiles of the *Hp* study strains to some extent. For example, no outer core genes were found in *Hp* 14, whose LPS consists of a single glycoform of a similar molecular weight as the inner core LPS control; *Hp* 13 and 15 contain *lic2C* and *lpsB* homologues respectively and their LPS appears to

be one sugar larger than that of *Hp* 14, as would be expected. Four putative outer core glycosyltransferases were identified in each of strains *Hp* 17, 19 and 34, which are the three *Hp* strains displaying a single glycoform at a position consistent with inner core LPS plus four sugars. Conversely the highest band in the LPS profiles of *Hp* 22, 24 and 31 cannot be accounted for by their identified outer core glycosyltransferases, but their similarity to the 'single O-unit' band of the *Hp* T3T1 profile suggests an alternative explanation.

At least four of the *Hp* LPS profiles (*Hp* strains 13, 15, 20 and T3T1) show evenly-spaced ladders of bands typical of a repeating oligosaccharide unit (OAg) as seen in other bacterial species, e.g. *E. coli* and *Salmonella enterica* (Wang et al., 2007), but not usually in *Haemophilus*. The composition and genetics of these OAgs are the subjects of the following two chapters.

4 Results: Initial studies on O-antigens and single O-units

Hi employs phase variation of outer core LPS components to adapt to changes in the host environment and to avoid immunological clearance. As outlined in Chapter 3, genome analysis suggests that *Hp* T3T1 does not use this mechanism and indeed does not add an outer core to its LPS at all. Instead, like many Gram negative species, it adds an OAg chain composed of a repeating pattern of sugars. This may provide an alternative system for adaptation and immunological protection.

Genes related to OAg synthesis are normally grouped together in a single locus to enable co-regulation; this may also facilitate the exchange of these genes between bacteria. In the *Hp* T3T1 genome, the ORFs between the *glnA* and *pepB* genes appear to constitute such an OAg locus. This is the same genomic location in which the *hmg* locus is found in *Hi*. Functional OAg loci typically include genes encoding the following enzyme groups:

(i) Nucleotide sugar synthesis, except for some molecules such as UDP-GlcNAc which are already synthesised for other cellular functions;

(ii) Sugar- and linkage-specific glycosyltransferases – these include WbaP (or a related enzyme) to transfer the first sugar of the O-unit to the isoprenoid lipid carrier, undecaprenyl phosphate (UndP);

(iii) One of the two OAg assembly and transport systems – either the Wzy system (OAg flippase Wzx, OAg polymerase Wzy and chain length determinant Wzz) or the ABC transporter system (ABC permease and ATPase subunits);

(iv) OAg ligase to transfer the OAg chain from the UndP carrier to LPS core.

In this chapter we examine how the proteins encoded by the *Hp* T3T1 OAg locus compare to other polysaccharide synthesis enzymes and how they relate to the known OAg structure. Some preliminary results regarding the OAg and OAg loci of other *Hp* study strains are then presented, followed by experiments to determine the position of OAg attachment within the LPS molecule through the disruption of an inner core LPS gene.

4.1 The *Hp* T3T1 genome encodes the enzymes required for O-antigen synthesis

The *Hp* T3T1 OAg locus (ORFs *Hp0285* to *Hp0270*) comprises 16 genes over 16.77 kb and has some notable characteristics when compared to the genome as a whole (Fig. 4.1). Whilst the GC content of the *Hp* T3T1 genome is 39.6%, that of the OAg locus is only 32.1%. In addition, the 9 bp *Hi* uptake signal sequence (USS) (Redfield et al., 2006) is distributed throughout the *Hp* T3T1 genome at an average frequency of one every 1.4 kb but is absent from the OAg locus. These observations support the hypothesis that *Hp* has acquired this region of DNA from another bacterium relatively recently.

BLASTP analysis of the *Hp* T3T1 OAg locus (Table 4.1) revealed genes encoding homologues of enzymes in all four of the categories listed above. The putative role of each protein encoded

by the locus was explored by comparison to homologues of known function, conserved domain searches and in some cases simple tertiary structure modelling. Using these methods it was possible to predict parts of the *Hp* T3T1 OAg structure through bioinformatics alone. This could then be compared with structural detail obtained once full chemical analysis had been performed.

4.1.1 Nucleotide sugar synthesis and glycosyltransferases

The *Hp* T3T1 OAg locus genes *Hp0274* and *Hp0273* are homologues of the *Bacteroides fragilis* genes *wcfR* and *wcfS* respectively, which are thought to be involved in the biosynthesis of the unusual sugar 2-acetamido-4-amino-2,4,6-trideoxygalactose (AAT) and its attachment to UndP (Coyne et al., 2001). This suggested that the first sugar of the *Hp* T3T1 O-unit could be AAT, a hypothesis that was confirmed by subsequent structural analysis (Fig. 3.8).

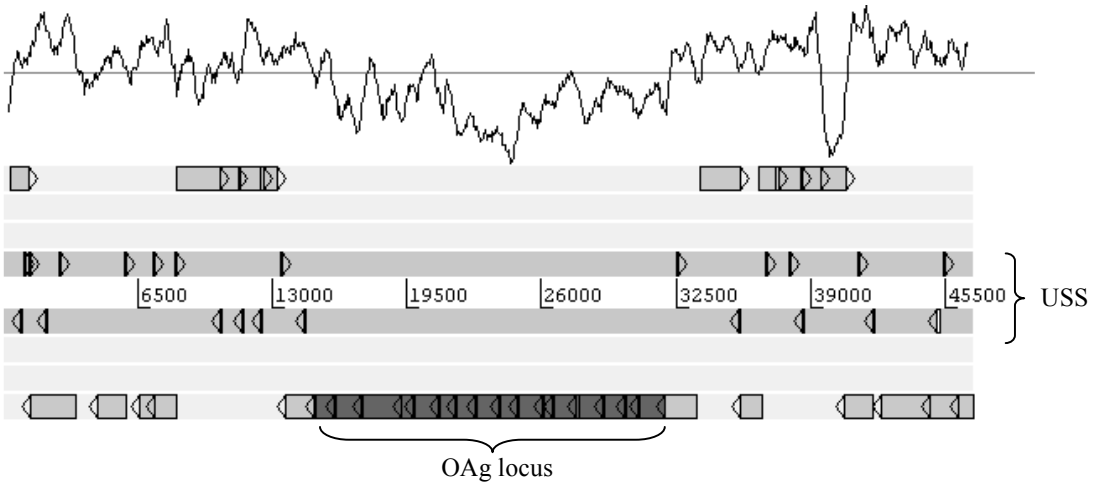


Figure 4.1. Features of the *Hp* T3T1 OAg locus. The locus is shown with 15 kb flanking sequence each side to demonstrate its atypical GC content and uptake signal sequence (USS) distribution. ORFs are represented by rectangles with arrowheads to show orientation; those forming the OAg locus are dark grey. GC content, calculated using a 500 bp window, is displayed as a graph; for the region shown, the mean value is 37.2% (shown by a horizontal line) and the range is 21.4% to 49.0%. Individual black triangles show the distribution of the USS (AAGTGCGGT), which is absent from the OAg locus.

Table 4.1. BLASTP analysis of the 16 ORFs in the *Hp* T3T1 O-antigen locus. Original gene names are listed in the left hand column, along with the locus tags (PARA numbers) that were later assigned and published in the NCBI database. The second column lists the putative size of the encoded protein in amino acids (aa), based on the positions of start and stop codons within the locus and on the annotated start codons of homologues. The species containing the three highest scoring BLASTP matches are stated, but only the top match for each genus is included so that the breadth of genera can be shown. For each match the % aa identity is given, with % query coverage in brackets. The putative function of each *Hp* protein is based on the known functions of homologues, domain model searches and in some cases transmembrane domain modelling (see text). The annotated functions of individual BLASTP matches are discussed in the text where relevant; most are either unconfirmed automated annotations or are in general agreement with the putative function listed in the third column. For proteins that may be related to AAT synthesis or transfer, additional specific BLASTP searches were carried out against species/strains which are known to produce AAT (shaded in grey).

Gene			**Highest scoring BLASTP matches**	
(NCBI locus tag)	**aa**	**Putative function**	**Species**	**%aa ID (QC)**
Hp0285	377	OAg chain length determinant (Wzz)	*Mannheimia succiniciproducens*	60 (94)
(PARA_02870)			*Actinobacillus minor*	46 (95)
			Haemophilus influenzae	47 (58)
Hp0284	209	N-acetylneuraminate synthase / O-acetyltransferase (NeuD)	*Mannheimia haemolytica*	74 (99)
(PARA_02860)			*Simonsiella muelleri*	66 (99)
			Bacteroides finegoldii	53 (98)
Hp0283	345	N-acetylneuraminate synthase (NeuB)	*Mannheimia haemolytica*	87 (99)
(PARA_02850)			*Simonsiella muelleri*	86 (99)
			Escherichia coli	75 (100)
Hp0282	419	N-acylneuraminate cytidylyltransferase (NeuA)	*Mannheimia haemolytica*	72 (100)
(PARA_02840)			*Simonsiella muelleri*	59 (100)
			Fusobacterium nucleatum	52 (100)
Hp0281	389	UDP-GlcNAc epimerase (NeuC)	*Mannheimia haemolytica*	70 (98)
(PARA_02830)			*Simonsiella muelleri*	71 (95)
			Escherichia coli	55 (98)
Hp0280	166	OAc transferase	*Sulfurospirillum deleyianum*	40 (74)
(PARA_02820)			*Idiomarina baltica*	41 (75)
			Vibrio harveyi	40 (79)
Hp0279	397	OAg flippase (Wzx)	*Pasteurella dagmatis*	57 (99)
(PARA_02810)			*Haemophilus influenzae*	57 (99)
			Mannheimia haemolytica	58 (99)
Hp0278	300	α-2,3-sialyl transferase (GT family 52)	*Streptococcus agalactiae*	40 (98)
(PARA_02800)			*Mannheimia haemolytica*	40 (99)
			Pasteurella multocida	38 (90)
Hp0277	357	OAg polymerase (Wzy)	*Shigella dysenteriae*	26 (98)
(PARA_02790)			*Bacteroides stercoris*	32 (88)
			Escherichia coli	27 (100)

continued overleaf

Gene (NCBI locus tag)	aa	Putative function	Highest scoring BLASTP matches Species	%aa ID (QC)
Hp0276 (PARA_02780)	323	GT (GTA superfamily)	*Streptococcus thermophilus*	39 (99)
			Bacteroides cellulosilyticus	34 (98)
			Clostridium sp.	33 (89)
Hp0275 (PARA_02770)	271	GT (GTA superfamily)	*Haemophilus parasuis*	58 (100)
			Bacteroides cellulosilyticus	39 (98)
			Eubacterium rectale	39 (97)
Hp0274 (PARA_02760)	403	Aminotransferase (AAT synthesis)	*Bacteroides fragilis* NCTC 9343, gene *wcfR*	79 (99)
			Porphyromonas endodontalis	73 (99)
			Pedobacter heparinus	60 (98)
			Streptococcus pneumoniae R6, gene *aatB*	44 (99)
			Shigella sonnei 53GI, gene *wbgX*	38 (97)
			Plesiomonas shigelloides C27, gene *wbgX*	38 (97)
			Bordetella parapertussis 12822, gene *bplF*	34 (97)
Hp0273 (PARA_02750)	195	UndP-AAT phosphotransferase (WcfS)	*Bacteroides fragilis* NCTC 9343, gene *wcfS*	73 (100)
			Porphyromonas endodontalis	64 (98)
			Fusobacterium nucleatum	62 (98)
			Shigella sonnei 53GI, gene *wbgY*	54 (98)
			Plesiomonas shigelloides C27, gene *wbgY*	54 (98)
			Bordetella parapertussis 12822, gene *bplG*	56 (98)
Hp0272 (PARA_02740)	631	UDP-GlcNAc dehydratase (AAT synthesis)	*Vibrio parahaemolyticus*	57 (99)
			Marinomonas sp.	56 (99)
			Photobacterium profundum	57 (99)
			Bacteroides fragilis NCTC 9343, gene BF2848	39 (91)
			Streptococcus pneumoniae R6, gene *aatA*	40 (74)
			Shigella sonnei 53GI, gene *wbgZ*	54 (99)
			Plesiomonas shigelloides C27, gene *wbgZ*	54 (99)
			Bordetella parapertussis 12822, gene *bplL*	48 (94)
Hp0271 (PARA_02730)	414	OAg ligase (WaaL)	*Aggregatibacter aphrophilus*	38 (96)
			Mannheimia succiniciproducens	33 (97)
			Haemophilus influenzae	37 (96)
Hp0270 (PARA_02720)	337	dTDP-glucose 4,6-dehydratase (RmlB)?	*Haemophilus influenzae*	91 (98)
			Actinobacillus minor	70 (100)
			Neisseria meningitidis	69 (99)

In *B. fragilis*, AAT is the first sugar of the repeat unit for the polysaccharide A (PS-A) capsule (Baumann et al., 1992). AAT (also called AATGal or 4nFucNAc) has been identified as part of various structures in only seven other bacterial species to date. These include *Streptococcus pneumoniae* serotype 1 capsule (Bentley et al., 2006), *S. pneumoniae* and *Streptococcus mitis* lipoteichoic acid (Bergstrom et al., 2000; Draing et al., 2006); *Shigella sonnei* (*E. coli* Sonnei) and *Plesiomonas shigelloides* OAg (Kenne et al., 1980; Shepherd et al., 2000); and *Bordetella pertussis*, *B. parapertussis* and *B. bronchiseptica* LPS as part of a linker between the core and

OAg constituents, where it is partially methylated (Preston et al., 2006). Xu and colleagues (2002) proposed a pathway for the biosynthesis of UndP-linked AAT in *S. sonnei*, using UDP-GlcNAc as the precursor (Fig. 4.2): this involves dehydration and reduction steps carried out by WbgZ, the transfer of an amino group by WbgX, and finally attachment to UndP by WbgY so that further sugars can be added to the chain.

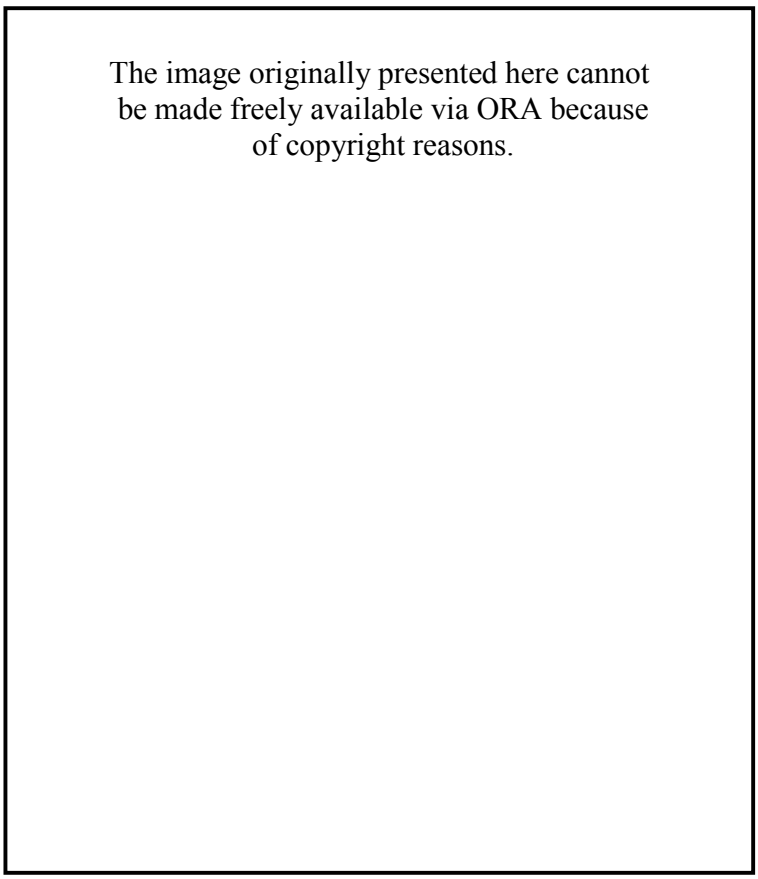


Figure 4.2. Proposed pathway for the biosynthesis of undecaprenyl phosphate linked AAT (also called 4n-FucNAc) in *S. sonnei* 53G, based on the predicted enzymatic activities of proteins encoded by the OAg locus. The *Hp* T3T1 homologues are noted in brackets beside each *S. sonnei* enzyme. Adapted from Xu et al. (2002).

The *Hp* T3T1 OAg locus encodes homologues of all three of these enzymes (Table 4.1, *Hp0274* to *Hp0272*); WbgX and WbgY are homologues of *B. fragilis* WcfR and WcfS. Indeed, BLASTP searches of the other known AAT-containing species (shaded in Table 4.1) suggest that the pathway is highly conserved, and one might predict that other bacteria with these genes such as *Porphyromonas endodontalis* and *Fusobacterium nucleatum* may also synthesise AAT as part of some glycoconjugate.

The proposed UndP-sugar phosphotransferase Hp0273 shares some homology with the C-terminal end of the UndP-Gal phosphotransferase (WbaP) enzymes from *Salmonella* and *Hi*, but as it appears to add AAT rather than Gal as the initial sugar of the O-unit we will refer to it as WcfS, after the UndP-AAT phosphotransferase from *B. fragilis*.

In addition to Hp0273, which contains a predicted transmembrane domain, the *Hp* T3T1 OAg locus encodes two putative cytoplasmic glycosyltransferases (GTs): Hp0275 and Hp0276. Their closest known homologues are uncharacterised GTs from *H. parasuis* and *Streptococcus* respectively. Hp0275 and Hp0276 are likely to add the second and third sugars of the O-unit, Gal and GalNAc, but the sugar specificity of each enzyme is unclear. Gal and GalNAc are also found in the *Hi* HMG unit; however, the linkages are different and the two *Hp* T3T1 OAg GTs share little homology with those encoded by the *Hi hmg* locus, so they are likely to have been derived independently.

Gal is typically provided for LPS synthesis by the conversion of the common cell wall precursor UDP-Glc to UDP-Gal. In *Hi* this reaction is catalysed by GalE (HI0351) (Masoud et al., 2008),

of which *Hp* T3T1 has a homologue with 91% aa identity (Hp0131). Some species require a very similar but separate enzyme (Gne) to epimerise UDP-GlcNAc into UDP-GalNAc (Canals et al., 2006), whilst others are thought to use a bifunctional epimerase for both conversions. It is unclear whether the *Hi* and *Hp* GalE can use UDP-GlcNAc as a substrate. If not, a putative candidate for *gne* in the *Hp* T3T1 OAg locus is *Hp0270*: its uncharacterised homologue in *Hi* Rd is annotated as *rffG*/*rfbB* (*rmlB*) but it shares 25% aa identity with *Hp* T3T1 *galE* and fits the GalE domain model with an E-value of 2.8E-28. There is no other obvious role for Hp0270 in the synthesis of the known *Hp* T3T1 LPS structure.

The proteins with highest homology to Hp0280 are all annotated as serine acetyltransferases (SAT), enzymes required for bacterial cysteine metabolism which use the substrates serine and acetyl-CoA. Hp0280 shares many of the amino acid residues indicated to be important for CoA binding in the C-terminal beta helix conserved domain model (cd03354), but lacks the N-terminal conserved domain SATase_N (Pfam PF06426). *Hp* T3T1 has another ORF elsewhere (*Hp0621*) which encodes both domains and which shares 94% aa homology with the well-characterised *Hi* serine acetyltransferase (Olsen et al., 2004). These findings suggest that Hp0280 may use acetyl-CoA to acetylate a substrate other than serine; as the protein is encoded within the OAg locus, the substrate is likely to be an OAg sugar or precursor.

4.1.2 Sialic acid synthesis and transfer

Several proteins encoded by the *Hp* T3T1 OAg locus appear to relate to the metabolism and transfer of sialic acid (Neu5Ac), a common component of OAg and capsular polysaccharides. The substrate for sialic acid addition is usually the nucleotide sugar CMP-Neu5Ac.

The biosynthetic pathway for CMP-Neu5Ac has been well characterised for *E. coli* K1, which uses this precursor to produce its polysialic acid capsule (Vimr et al., 2004). The process requires the four enzymes NeuA-NeuD. Firstly NeuC converts UDP-GlcNAc to N-acetylmannosamine (ManNAc); NeuB then condenses ManNAc and phosphoenolpyruvate to produce Neu5Ac. The CMP-sialic acid synthetase, NeuA, must then activate Neu5Ac so that it can be used as the substrate for sialyltransferases. NeuD plays an unidentified role in sialic acid synthesis but has also been shown to O-acetylate monomeric Neu5Ac (Steenbergen et al., 2006).

NeuA-neuD homologues are present within the *Hp* T3T1 OAg locus (*Hp0281-Hp0284*; Table 4.1), but they are more similar to genes found in *Mannheimia haemolytica* serotype A2 strains than to those of *E. coli*. *M. haemolytica* colonises the ruminant respiratory tract and is a member of the Pasteurellaceae family; serotype A2 strains are defined by their polysialic acid capsule (Adlam et al., 1987). Interestingly, the expression of CMP-Neu5Ac synthesis enzymes in *M. haemolytica* is downregulated at high incubation temperatures and no capsule is detected above 40 °C (Barrallo et al., 1999). It is possible that temperature may be a factor in the regulation of OAg expression in *Hp*, although the temperature within its obligate human host is unlikely to vary by more than a few degrees.

Hp0278 is a Family 52 GT with homology to sialyltransferases from a wide range of bacterial genera. It shares 40% aa identity with proteins from two species: firstly the capsule sialyltransferase of *M. haemolytica* serotype A2; and secondly Cps8K, which is involved in capsule synthesis in *Streptococcus agalactiae* serotype VIII strains, adding α-D-NeuAc via a 2,3

linkage to β-D-Gal (Cieslewicz et al., 2005). Given this relatively low level of homology, it is not possible to predict the acceptor sugar and linkage specificity of Hp0278, but it is likely to be a sialyltransferase that adds the product of the NeuA-NeuD biosynthetic pathway to the OAg unit. In agreement with this prediction, the fourth sugar of the *Hp* T3T1 O-unit was found to be Neu5Ac (Fig. 3.8). The acetylation of the sialic acid residue might be carried out by either NeuD (Hp0284) or the putative O-acetyltransferase (Hp0280) described above.

Hi can also decorate its LPS with Neu5Ac, which it obtains from the environment using a tripartite ATP-independent periplasmic (TRAP) transporter (Severi et al., 2005). This is encoded by genes *siaP* and *siaQ/M* which are located next to the six genes of the *nag/nan* operon for sialic acid catabolism. The *Hp* T3T1 genome sequence includes homologues of the *nag/nan* genes, but not of *siaP* and *siaQ/M* or of the *Hi* LPS sialyltransferase genes. Instead, *Hp* T3T1 encodes enzymes for the synthesis and transfer of Neu5Ac within the OAg locus. Whereas *Hi* requires exogenous sialic acid in order to produce the Neu5Ac-containing HMG unit, *Hp* T3T1 can add Neu5Ac-containing OAg to its LPS without this supplement.

4.1.3 OAg assembly, transport and ligation to core LPS

The *Hp* T3T1 OAg locus encodes homologues of the three enzymes required for Wzy-dependent OAg assembly and transport: Wzx, Wzy and Wzz.

Gene *Hp0279* encodes a protein with two homologues in *Hi* Rd: the *lsg* locus putative flippase (HI1700; 57% aa identity) and the *hmg* locus flippase (HI0867; 55% aa identity). A conserved domain search predicts Hp0279 to belong to the RfbX family of membrane proteins involved in

OAg export. The TMHMM transmembrane modelling tool, which uses a hidden Markov model to determine whether each residue of a protein is most likely to be cytoplasmic, periplasmic or embedded in the membrane, predicts that Hp0279 contains twelve transmembrane domains. Togther this evidence suggests that Hp0279 is an OAg flippase (Wzx), transporting UndP-linked O-units across the inner membrane.

The highest scoring BLASTP matches for Hp0277 are OAg and capsular polysaccharide polymerases from a range of bacterial families. OAg polymerases typically show little amino acid similarity to each other but their tertiary structure is more conserved, with at least ten transmembrane domains anchoring the protein in the inner membrane (Kim et al., 2010). The TMHMM algorithm predicts the Hp0277 protein to contain ten transmembrane helices (Fig. 4.3).

The OAg structures are known for some of the *E. coli*, *Shigella dysenteriae* and *Lactobacillus delbrueckii* serotypes whose OAg polymerases share 25-30% aa identity with Hp0277, but the range of substrate and acceptor sugar moieties is too wide to enable any prediction of the initial and final *Hp* T3T1 O-unit components in this instance. This is not surprising since the specificity of an OAg polymerase is likely to be determined by a few unidentified periplasmic amino acid residues.

The *Hp* T3T1 OAg chain length determinant Wzz, encoded by gene *Hp0285*, is most closely related to those of other Pasteurellaceae genera that produce OAg, including several *Mannheimia* and *Actinobacillus* species. The topology predicted for Hp0285 by the TMHMM algorithm (see above) is a long periplasmic loop flanked by two transmembrane domains (Fig. 4.3), fitting the

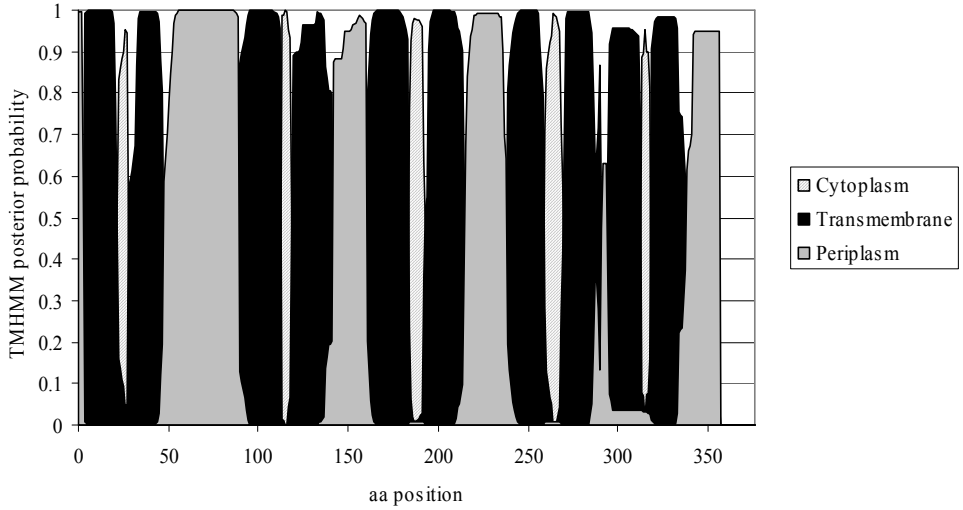
structure that has been determined for several polysaccharide co-polymerases including Wzz of other species (Morona et al., 2009). The atomic structure of *E. coli* O157:H7 FepE (Wzz) reveals a nonameric bell-shaped complex, with a ring of positive charge around the top created by two lysine residues from each of the nine Wzz monomers; it is speculated that these sidechains may interact with a partner in the outer membrane (Morona et al., 2009). The first of these residues is conserved in *Hp* T3T1 Wzz, but this is immediately followed by an extra 18 aa which are not present in FepE so perhaps the interactions of the protein differ between the two species.

In both Wzy- and ABC-dependent OAg synthesis, the completed OAg chain is removed from its UndP carrier and added to the LPS core by an OAg ligase (WaaL). It is likely that in *Hp* T3T1 this task is performed by Hp0271, which shares around 35% aa identity with putative OAg or HMG ligases from a range of other Pasteurellaceae species. TMHMM analysis predicts Hp0271 to contain 12 membrane-spanning domains and an 82 aa periplasmic loop. This topology is typical for an OAg ligase: the periplasmic loop regions of *Salmonella enterica* sv. Typhimurium and *E. coli* WaaL are predicted to be 73 and 84 aa respectively (Abeyrathne & Lam, 2007). An artificial H303A substitution in *Pseudomonas aeruginosa* WaaL (NP_253686) has been shown to completely abolish OAg ligase activity (Abeyrathne & Lam, 2007); this histidine residue is conserved in Hp0271 and is presumably necessary for the enzyme to function.

In the *Hi hmg* locus, the OAg flippase and ligase genes are intact but only vestigial OAg polymerase and chain length determinant genes are present; this results in the addition of a single 'O-unit' (HMG) to the LPS. In contrast, the *wzx*, *wzy*, *wzz* and *waaL* ORFs in the *Hp* T3T1 OAg

locus are intact, and are likely to encode functional enzymes which can be used to synthesise OAg with a tightly regulated modal chain length distribution as observed by tricine SDS-PAGE.

(a) *Hp* T3T1 OAg polymerase



(b) *Hp* T3T1 OAg chain length determinant

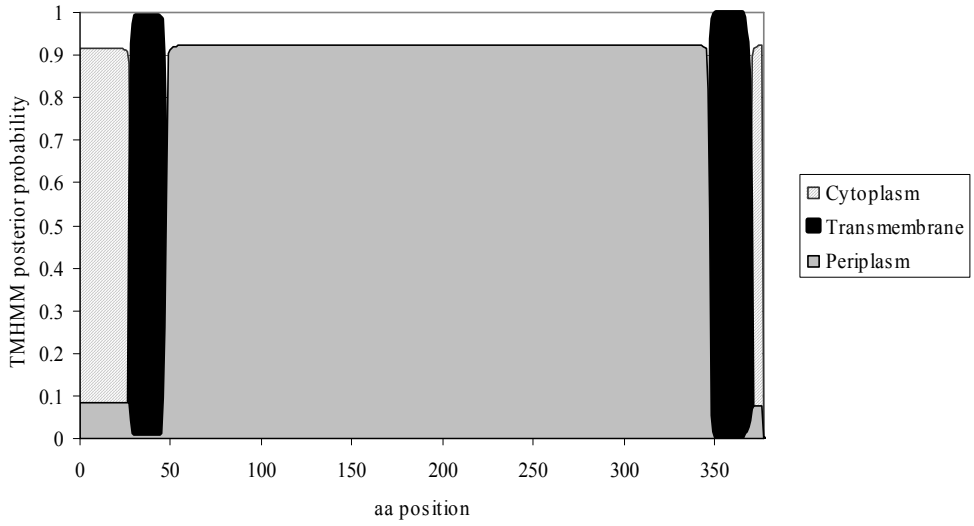


Figure 4.3. Topology predictions for *Hp* T3T1 OAg processing proteins. Data was generated using TMHMM and plotted using Excel. The x axis shows amino acid position; the probability of each residue being cytoplasmic, periplasmic or part of a transmembrane helix is plotted along the y axis. **(a)** OAg polymerase (Wzy), Hp0277; ten transmembrane helices (black peaks) are predicted, in agreement with Wzy from other species. **(b)** OAg chain length determinant (Wzz), Hp0285; two transmembrane helices and a periplasmic loop are predicted, as has been structurally determined for other Wzz proteins.

4.1.4 Summary of the *Hp* T3T1 OAg locus

The structure of *Hp* T3T1 OAg, as deduced by mass spectrometry and NMR techniques, is shown in Fig. 3.8 (Elke Schweda, pers. comm.). As predicted, the first sugar in the O-unit is AAT. This rare sugar has never been observed in *Haemophilus* LPS before. The presence of two other GT genes and one sialyltransferase gene is consistent with the structural evidence for Gal, GalNAc and sialic acid as the remaining components of the O-unit. Most heteropolymeric OAgs are thought to use the Wzy-dependent mechanism for assembly and the *Hp* T3T1 OAg is no exception.

The list of species in which the highest-scoring BLASTP matches were found (Table 4.1) gives an indication of the extraordinary level of reassortment that occurs between the polysaccharide synthesis loci of different bacteria. Whilst it is not possible to deduce the exact phylogeny of each gene in the *Hp* T3T1 OAg locus, it is obvious that there have been multiple DNA uptake and recombination events during its evolution, either in *Hp* itself or in an unidentified donor species.

4.2 All strains could have an OAg locus between *glnA* and *pepB*

The OAg locus of *Hp* T3T1 is flanked by the *glnA* and *pepB* genes. In many species of Gram negative bacteria the OAg locus is at the same genomic location in different strains, and indeed the *hmg* (high molecular weight glycoform) locus carried by many *Hi* isolates is always located between *glnA* and *pepB*. Therefore if OAg loci are present in other *Hp* strains they are likely to be in this location. To investigate this, long range PCR (LR-PCR) was attempted using gDNA

from each of the 20 study strains with primers designed to *Hp* T3T1 *glnA* and *pepB*. During initial attempts the only product amplified was in the *Hp* T3T1 reactions, but by lowering the annealing temperature from 60ºC to 51ºC, products were obtained for all 18 true *Hp* strains. The different annealing temperature requirements are likely to reflect the level of nucleotide polymorphism between the primers and template DNA across strains. No products were obtained for the two hybrid strains (*Hp* 6 and *Hp* 11).

It was immediately obvious that the PCR products amplified from all 18 *Hp* strains were of high molecular weight (above the 10 kb marker on an agarose gel), suggesting that a series of genes fell between *glnA* and *pepB* in all cases. Refinement of electrophoresis conditions revealed a range of product sizes: the smallest (those of *Hp* strains 13 and 18) were approximately 12 kb, and the largest included the *Hp* T3T1 product, which can be calculated as 18.8 kb (Fig. 4.4a). Only *Hp* strains 19, 22, 24 and 31 gave LR-PCR products as large as that of *Hp* T3T1. The ends of the PCR products contain part of the flanking genes, *glnA* and *pepB*, so the sizes of the potential OAg loci themselves are estimated to range from 10 to 17 kb.

As the *Hp* T3T1 OAg locus is 99.2% coding, containing only 128 bp of intergenic DNA, the majority of the other *Hp* strains cannot have homologues of all the T3T1 OAg genes in this genomic location: they may contain a subset of these genes or a different set of ORFs entirely. MfeI restriction digest profiles of the 18 LR-PCR products were almost all unique, suggesting a high level of nucleotide divergence between the loci (Fig. 4.4b). Only *Hp* strains 24 and 31 had identical profiles; *Hp* strains 14 and 16 also shared three restriction fragments with each other so their loci may be closely related.

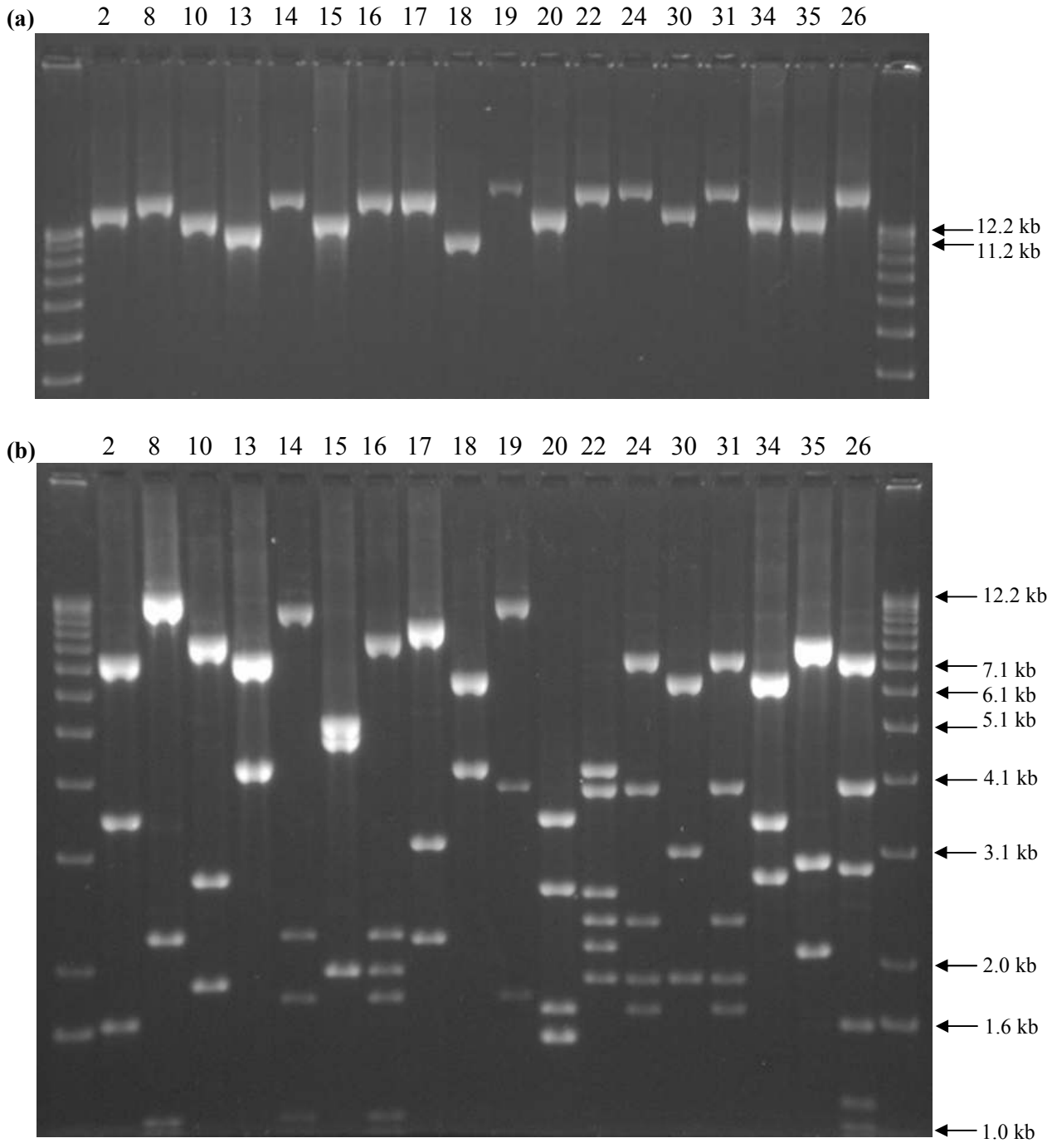


Figure 4.4. Long range PCR products obtained using primers to *glnA* and *pepB* which flank the OAg locus. The outside lanes contain 2 µl 1 Kb DNA Ladder, with kilobase sizes as indicated. Numbers above each lane indicate the *Hp* strain of the template gDNA (strain 26 = *Hp* T3T1). **(a)** 0.4 µl of each PCR reaction was separated by agarose gel electrophoresis. Products of a high molecular weight are visible for all 18 true *Hp* strains; hybrids *Hp* 6 and *Hp* 11 did not yield products and are not shown. **(b)** MfeI restriction digests of the LR-PCR products, with fragments >1 kb visible.

Short intragenic PCRs were performed on the 20 study strains to test for the presence of homologues of some of the *Hp* T3T1 OAg locus genes. PCR for the genes encoding the OAg polymerase, flippase, ligase and chain length determinant enzymes each amplified products for only 0 to 3 strains aside from *Hp* T3T1 (data incorporated into Table 5.2), and Southern analysis gave similar results (Fig. 4.5). However, PCR using primers designed to the putative UDP-GlcNAc dehydratase gene, *Hp0272*, amplified products for 7/20 strains. This was the first indication that several strains may be capable of decorating their LPS with at least one of the same sugars as *Hp* T3T1.

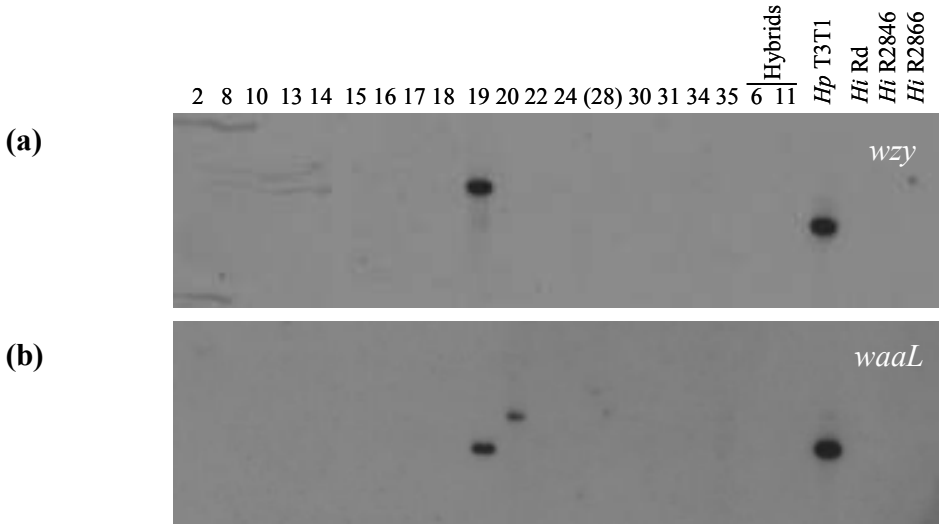


Figure 4.5. Southern analysis of MfeI-digested gDNA hybridised with PCR amplification products from *Hp* T3T1 OAg locus genes. *Hp* strain numbers are listed above each lane; three *Hi* controls are shown on the right. Probes were as follows: **(a)** *Hp* T3T1 *wzy* (OAg polymerase), **(b)** *Hp* T3T1 *waaL* (OAg ligase).

4.3 Many *Hp* strains exhibit OAg-like LPS patterns

In *Hi*, strains with an *hmg* locus correspond to those which produce an HMG unit visible by tricine SDS-PAGE. As OAg was initially observed in only four *Hp* LPS profiles (*Hp* strains 13, 15, 20 and 26; Fig. 3.7b) it was considered at first that these may be the only strains with an OAg locus, mirroring the situation in *Hi*. However, the LR-PCR results dispute this hypothesis and suggest that all *Hp* strains may have the capacity to synthesise an OAg. To determine whether any strains produced weak OAg that had not been detected, new tricine SDS-PAGE gels were prepared using ten times as much of each cell lysate as before (12.5 µl of OD_{260} = 10 rather than 12.5 µl of OD_{260} = 1). Upon silver staining, high molecular weight smears or 'ladder' patterns were observed for *Hp* strains 2, 8, 10, 14, 16, 17, 18, 30 and hybrid strain *Hp* 6 (Fig. 4.6). It is likely that these bands correspond to LPS core plus OAg of varied chain length. This brings the total number of study strains for which putative OAg has been observed to 13/20 (65%). In addition, we can now see the full extent of the *Hp* T3T1 OAg (lane 26); its chain length distribution is bimodal, with most LPS molecules having either 0-2 or around 20 O-units.

The intensity, apparent O-unit size and average molecular weight of the putative OAg glycoforms are reproducible for each strain under laboratory conditions but vary greatly between strains, suggesting differences in O-unit composition, chain length regulation and enzyme efficiency.

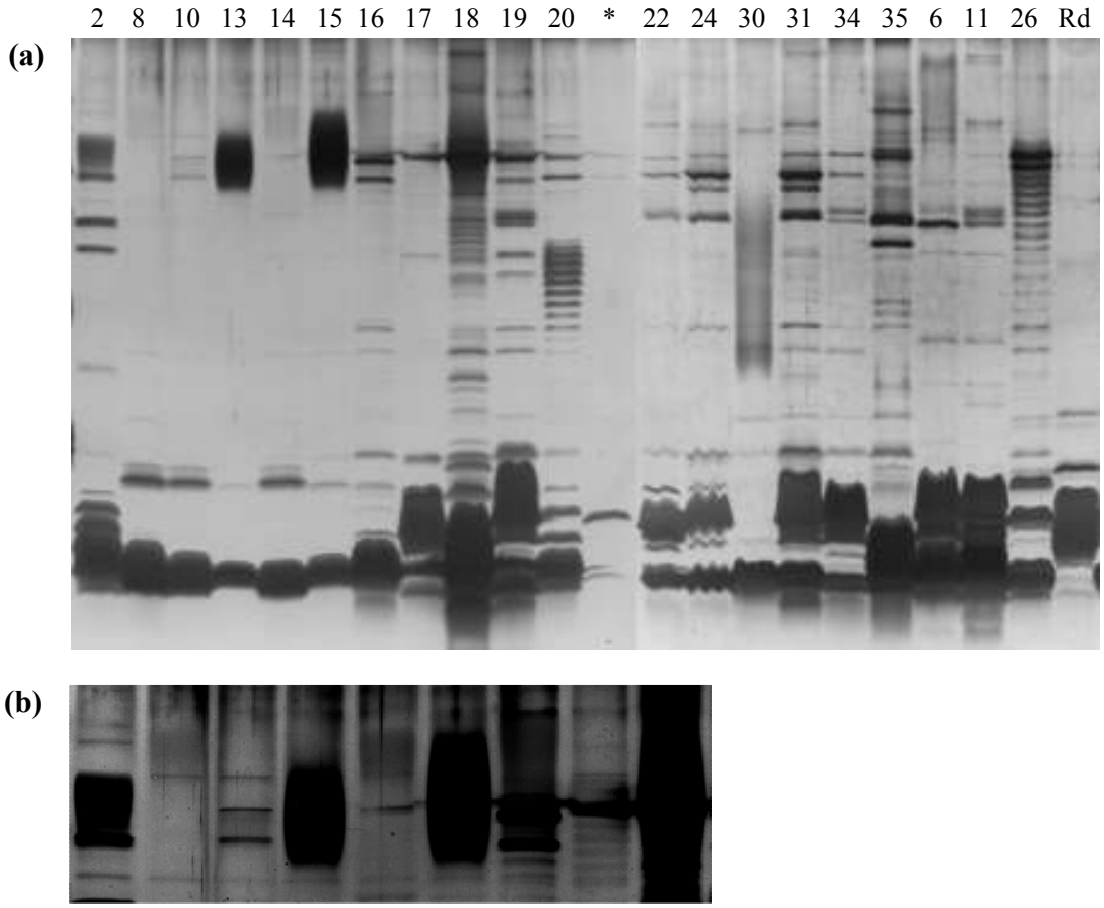


Figure 4.6. LPS profiles following tricine SDS-PAGE using concentrated *Hp* cell lysates. (a) LPS profiles for all 20 *Hp* strains. Strain numbers are listed above each lane. 12.5 µl of lysate at OD_{260} = 10 (rather than the standard OD_{260} = 1) were loaded per lane, in order to visualise any weak OAg that may be present. For *Hp* 13 and 15, lysates at OD_{260} = 5 were used as these strains were already known to produce large quantities of OAg. *Hp* strain 24 lysate at the standard concentration was loaded in the lane marked * as a reference. All lysates were treated with proteinase K, but this has not been 100% effective so some protein bands remain. OAg-like ladder patterns are now visible for *Hp* strains 2, (8), (10), 13, (14), 15, (16), (17), 18, 20, 30, 6 and 26 (strains with faint patterns are listed in brackets). **(b)** Magnified image of the upper region of the first nine lanes of panel (a), with the contrast digitally enhanced so that weak putative OAg can be visualised (e.g. for *Hp* strains 8 and 17).

4.4 Mutation of the *Hp lgtF* gene by transformation: the OAg or O-unit is attached via Glc on HepI

In *Hi*, the four-sugar HMG unit is added to the Glc residue on HepI of the LPS (Hood et al., 2004b). Structural analysis showed that the equivalent Glc in *Hp* T3T1 LPS carries the OAg (Fig. 3.8). As this Glc is probably present in the LPS of all the *Hp* study strains (see Section 3.2.1), it is possible that it may be the attachment site for all *Hp* OAgs. To test this hypothesis a plasmid construct was designed to interrupt *lgtF*, the gene which encodes the glucosyltransferase responsible for the Glc-(β1,4)-HepI linkage. As *lgtF* is highly conserved, a single construct using target sequence amplified from *Hp* T3T1 gDNA was used in transformation trials with several *Hp* strains. This also presented an opportunity to examine the transformability of different strains and to test various potential transformation methods.

The standard method used to disrupt genes in *Hi* by transformation is to incubate the cells in a starvation medium (MIV) to induce the competence genes, then add linearised plasmid DNA for uptake and recombination into the genome (Herriott et al., 1970; Poje & Redfield, 2003). Our own and other laboratories have attempted to use this method to transform *Hp* without success. However, two groups have previously reported the successful transformation of *Hp* using non-MIV methods. Gromkova and Goodgal (1979) used a static aerobic method to induce competence, growing BHI broth cultures of *Hp* statically in a petri dish at 37 °C overnight, then mixed 100 µl cells and 0.5 µg DNA in fresh BHI for a five hour incubation. The initial *Hp* inoculum density was not stated. This protocol was used to demonstrate the transfer of both chromosomal drug resistance markers and autonomously replicating plasmids. However, in a later paper, Gromkova and colleagues (1998) reported that only 19 of the 73 *Hp* strains studied

were transformable by this method and that none of the 30 clumping strains could be transformed.

In a paper published during the current project, Pollard et al. (2008) made a construct containing the *pgm* gene from *Hp* strain 4201 interrupted by a spectinomycin resistance cassette. 1 μg of this plasmid was pipetted onto a circle of confluent *Hp* 4201 growth on a BHI agar plate and incubated for 5 h, and spectinomycin selection was carried out after overnight broth growth. The chromosomal *pgm* gene was knocked out successfully. This method is similar to protocols regularly used for the transformation of *Neisseria*.

To construct the *lgtF* knockout plasmids pRY11 and pRY12, a 1.5 kb PCR product including the entire *lgtF* gene (*Hp0469*) was amplified from *Hp* T3T1 and ligated into the vector pSC-A. A 460 bp fragment of *lgtF* was replaced with either a Tet (pRY11) or Kan (pRY12) resistance cassette, leaving flanking regions of approximately 550 bp either side for recombination. Uptake signal sequences (USS) were also included with the aim of facilitating uptake of the plasmid DNA by *Hp* (see Section 2.10.3). The inclusion of USS in a plasmid appears to increase transformation rates in *Hi* (Mitchell et al., 1991), presumably by binding an unknown cell surface receptor which recognises this 9 bp sequence, and the same USS is also found throughout the *Hp* T3T1 genome.

In the first experiment, the transformation of five strains (*Hp* 2, 8, 20, 22 and 35) with 0.5 μg linearised pRY11 plasmid (as estimated from an agarose gel) was attempted using the static aerobic method of Gromkova and Goodgal (1979). A single colony was obtained for *Hp* strain

20 only and proved to be a successful transformant when tested by PCR and Southern analyses; it was named *Hp* 20.4. In PCR analysis using primers 469a and 469b, wild type *Hp* strain 20 yields a 620 bp product, whereas for *Hp* 20.4 the product was 2.6 kb due to the insertion of the Tet resistance cassette; other PCR reactions also gave the expected shifts in band migration. For the Southern analysis, five different restriction digests of gDNA from *Hp* strains 20 and 20.4 were fractionated, blotted and probed with fragments of *lgtF* and the Tet resistance cassette; it was observed that the two probes hybridised to the same restriction fragments for *Hp* 20.4, and that the size of the *lgtF*-hybridising fragments differed between the two strains as expected (data not shown).

Hp 20 cells clump when grown in BHI broth or when growth from solid medium is resuspended in PBS; this disputes the assertion of Gromkova et al. (1998) that clumping strains are untransformable by the static aerobic method, although the transformation was clearly of low efficiency. Later, the *lgtF* gene was successfully knocked out in *Hp* strains 17, 19, 20 and 35 using pRY12 and the static aerobic method. However, no transformants could be obtained for three key OAg-producing strains: *Hp* T3T1, *Hp* 13 and *Hp* 15. Some *Hp* genes will obviously be essential for survival and therefore impossible to knock out, but it is unlikely that *lgtF* would fall into this category in some strains and not others. One explanation for low transformation efficiency in some strains could be low homology between the plasmid insert and chromosomal regions for recombination, but as the *lgtF* plasmids were constructed using *Hp* T3T1 PCR products this can be ruled out for *Hp* T3T1. Increasing the DNA concentration in the transformation mix by up to five fold or using uncut plasmid was not beneficial, and no transformants were obtained for any strain using the method of Pollard et al. (2008). It was

important to develop methods to manipulate the major OAg-producing strains if possible, to allow further examination of OAg genetics and function.

In *Hi*, using chromosomal rather than plasmid donor DNA can increase the efficiency of transformation. We found that *Hp* strains 13 and 15 could be transformed by the static aerobic method using 3 μg chromosomal DNA prepared from an *Hp* 20 *lgtF* mutant, but *Hp* T3T1 could not. This showed that the transformation of *Hp* strains 13 and 15 is possible as long as another strain has been transformed with the relevant plasmid first. To reduce the risk that extra recombination events between the donor and recipient genomes could affect phenotypic results, at least three independent clones were analysed whenever chromosomal donor DNA was used for transformation. A second disadvantage of using chromosomal DNA is that depending on the position of the recombination end-points in each transformant, a significant length of DNA either side of the target gene may also be replaced with donor alleles during recombination, and the effects of this are difficult to predict.

Electroporation can be used to introduce plasmids into *Hi* by increasing the permeability of the cell membrane (Mason et al., 2003). The cells are first washed in sucrose-glycerol solution, then electroporated in the presence of circular plasmid DNA and incubated at 37 °C to recover. Using this protocol with the pRY12 construct, 16 kanamycin resistant *Hp* T3T1 colonies were obtained. Upon subculture, 13 of these gave PCR results that suggested single crossover events, i.e. they contain both wild type and interrupted copies of *lgtF*. This is a particular problem of using circular donor DNA. The remaining three colonies gave PCR results consistent with the desired double crossover event and were true *Hp* T3T1 *lgtF* mutant clones. This indicated that

homologous recombination is possible in *Hp* T3T1 and suggested that the difficulty may lie with DNA uptake.

Overall, mutants in which *lgtF* was interrupted were made for *Hp* strains 13, 15, 17, 19, 20, 35 and T3T1. The transformation efficiency varied greatly between strains: for example, static aerobic transformation with plasmid DNA yielded 2, 18 and 884 transformants per μg donor DNA for *Hp* strains 20, 19 and 35 respectively. Despite repeated attempts we could not obtain *lgtF* mutants for *Hp* strains 2, 8, 22 or 34 so the final success rate was 7/11 strains (64%). The LPS profiles of wild type *Hp* strains 13, 15, 20 and T3T1 contain OAg-like ladders of bands; the LPS profiles of *lgtF* mutants derived from these four strains lacked the putative OAg bands and the 'core LPS' band appeared lower in molecular weight, reflecting the loss of a core Glc as expected (Fig. 4.7). This suggests that in these four *Hp* strains, OAg is attached via the Glc on HepI of the core LPS. This position has now been confirmed by mass spectrometry and NMR methods for *Hp* strains 13, 20, 22 and T3T1 (Elke Schweda, pers. comm.).

The LPS profiles of *Hp* 17 and 19 *lgtF* mutants comprise a single glycoform of low molecular weight, predicted to correspond to core LPS minus Glc (Fig. 4.7). This is a considerably smaller glycoform than is seen for the respective parental strains and is consistent with all LPS outer core substitutions in *Hp* 17 and 19 being dependent on the addition of Glc to HepI. The most likely explanation is that all outer core sugars in these strains extend from this Glc, but it is also possible that some of these outer core sugars may be HepII or HepIII extensions whose transferase enzymes require the LPS molecule to be in a certain conformation that is dependent on initial HepI extension.

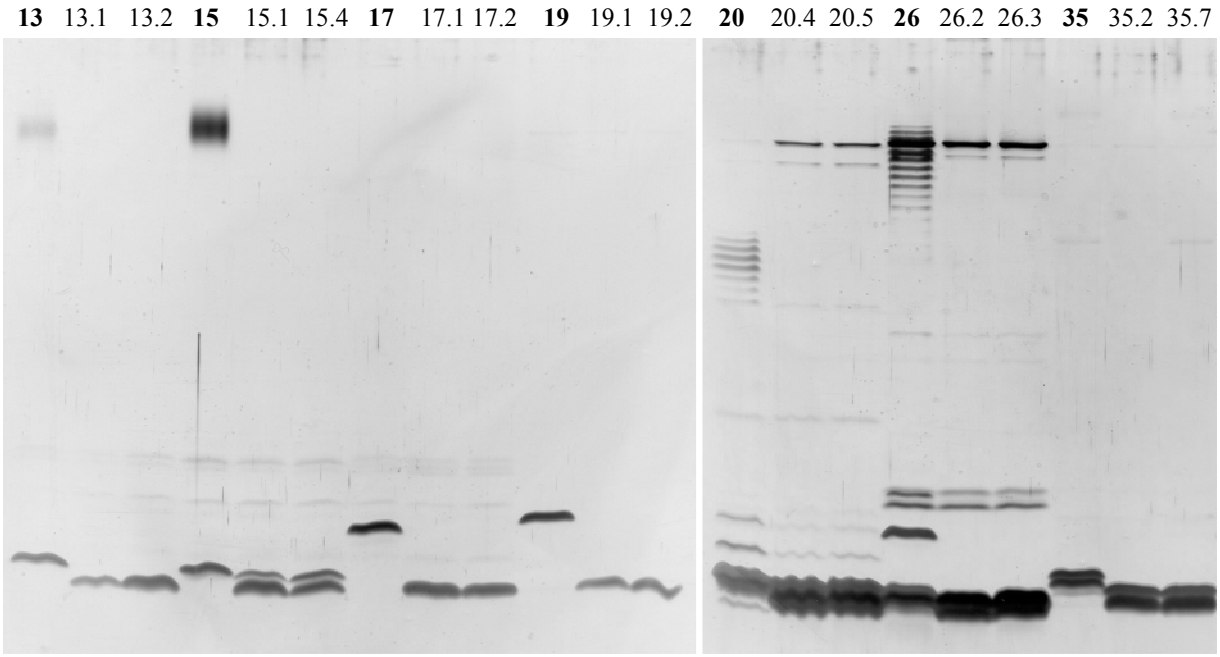


Figure 4.7. LPS profiles of *Hp lgtF* mutants. Proteinase K treated cell lysates of wild type and *lgtF* mutant strains were separated by tricine SDS-PAGE and silver stained. Numbers above each lane are the *Hp* strain (in bold) or *lgtF* mutant clone numbers. For *Hp* 20 and *Hp* 26 (= T3T1) and their derivative clones, 12.5 µl of lysate at OD_{260} = 5 (rather than the standard OD_{260} = 1) was loaded per lane to visualise the OAg ladder more clearly.

4.5 Discussion

The construction of bacterial polysaccharides such as OAgs and capsule requires nucleotide sugar synthesis proteins, transferases and assembly enzymes, all of which are typically encoded in a single operon. The OAg locus of *Hp* T3T1 encodes all the genes that would be required for the synthesis of its four sugar O-unit, its polymerisation, and its attachment to the LPS core. As LPS is one of the most abundant cell surface molecules, the nature of the specific components of the O-unit is likely to significantly influence host-microbe interactions.

The first sugar of the *Hp* T3T1 O-unit, AAT, has previously been identified in a few polysaccharides from diverse bacterial phyla including Firmicutes, Bacteroidetes and Proteobacteria. Despite the evolutionary distance between these species, the AAT biosynthetic pathway appears to be conserved, suggesting that the genes have either spread by horizontal transfer (clearly the case between *Shigella sonnei* and *Plesiomonas shigelloides* (Shepherd et al., 2000)) or undergone convergent evolution. The recent interest in this sugar has sparked efforts to synthesise it *de novo* so that its properties can be studied (Pragani et al., 2010), particularly in relation to its presence in zwitterionic polysaccharides. These contain balanced positive and negative charges, resulting in a polar but overall neutral molecule: for example, the *Bacteroides fragilis* PS-A capsule is zwitterionic because the amino group of AAT is positively charged whilst the pyruvate component is negatively charged. Polysaccharides with this charge profile are thought to be TLR2 agonists and to provoke T cell dependent responses such as intra-abdominal abscess formation (Tzianabos et al., 1995; Wack & Gallorini, 2008), arguing against the perceived T cell independence of all polysaccharide antigens.

According to the structure presented in Fig. 3.8, the *Hp* T3T1 O-unit is also potentially zwitterionic due to positively charged AAT and negatively charged sialic acid: the carboxyl group of the latter may lose a proton at physiological pH. The OAg alone might share the properties of other zwitterionic polysaccharides, but in practice the rest of the LPS molecule would also contribute to the overall charge profile and Toll-like receptor signalling capacities.

In contrast to AAT, the remaining three sugars in the *Hp* T3T1 O-unit (Gal, GalNAc and sialic acid) are commonly found in the LPS of *Hi* strains. All NTHi have the capacity to incorporate

sialic acid into their LPS as a terminal sugar; this addition increases resistance to the bactericidal effects of normal human serum and is a crucial virulence factor in an experimental chinchilla model of otitis media (Bouchet et al., 2003; Hood et al., 1999). If Neu5Ac-containing OAg emerges as a common constituent of *Hp* LPS, we would need to consider alternative explanations for the rarity of *Hp* otitis media in the human host.

Although the individual characteristics of the four sugars in the *Hp* T3T1 O-unit have been studied in other bacteria, it is the properties of the O-unit, OAg and LPS as a whole that must be taken into account when considering host interactions. These include its conformation, charge motif, length, ability to adhere to host cells or mucus, and immunological properties. In some polysaccharides the latter can be changed by a single glycosidic linkage alteration (Avci & Kasper, 2010). The particular O-unit found in *Hp* T3T1 has never been observed in bacteria before, so as yet its properties are largely unknown.

The analysis of the *Hp* T3T1 OAg locus demonstrated that it is possible to make meaningful predictions regarding OAg structure from the genes that are present. Currently there is very little information about the OAg of the other *Hp* study strains, but the presence of large *glnA-pepB* LR-PCR products and some OAg-like ladders is promising and suggests that further investigation of the *glnA-pepB* intergenic region may provide clues as to the composition of each OAg.

No OAg has yet been observed by SDS-PAGE for seven of the study strains. These strains may not encode an intact OAg locus, or perhaps the production of OAg or its addition to the LPS is

regulated according to environmental conditions with in vitro growth on BHI agar being unfavourable for its expression. These possibilities will be explored in later chapters.

The standard laboratory method of inducing competence in *Hi* strain Rd cells involves incubation in the starvation medium MIV, which upregulates the 25 genes of the competence regulon by increasing cAMP levels (Redfield et al., 2005). These genes encode DNA uptake, transport and processing proteins and share a particular promoter sequence that is upregulated by the cAMP receptor protein (Crp) and a transcription factor named Sxy. The *Hp* T3T1 genome encodes full length homologues of Crp (Hp0067) and Sxy (Hp0662) with 88% and 61% aa identity, respectively, and has Crp binding sites in the promoter regions of the *com* locus (*Hp0803-Hp0808*) and the *pil* locus (*Hp0106-Hp0109*). *Hp* therefore appears to use a similar method of competence regulation to *Hi*, but a more detailed comparison of Sxy sequence differences and of other Crp-regulated genes in the future might help to explain the low transformability of *Hp* T3T1.

The MIV protocol resulted in a 2-log decrease in viability in Grace Leidy's *Hp* 'Strain 14' when tested by Barany and Kahn (1985) and also failed to yield transformants in *Hp* T3T1 in our laboratory. An alternative method of inducing DNA uptake in *Hp* has been developed involving the overnight incubation of broth cultures in static petri dishes (Gromkova & Goodgal, 1979), which was successfully used to interrupt the *lgtF* gene in several of the *Hp* study strains (Section 4.4). It is not known whether this procedure induces the competence regulon or whether DNA uptake is mainly due to cell membrane permeabilisation by the divalent cations (20 mM $MgSO_4$) added during transformation. Although *Hp* T3T1 in vitro could only be transformed by

electroporation, the lineage from which this strain derived must have been naturally competent at some stage if the hypothesis of OAg locus evolution by genetic exchange is correct. *Hp* T3T1 may still be competent in the nasopharynx, where conditions differ from in vitro growth. Alternatively, DNA exchange may have occurred through other routes such as transduction or conjugation.

Transformation efficiencies were found to vary widely between *Hp* strains, a common phenomenon in 'competent' species of bacteria. Many *Hi* strains cannot be transformed efficiently by the standard (MIV) method, and the adoption of *Hi* Rd as a laboratory strain was largely based on its very high transformation levels. In a comparison of 34 *Hi* strains, Maughan and Redfield (2009a; 2009b) found a weak positive correlation between DNA uptake and recombination levels but struggled to find any relationships between competence and either *com* gene mutations or phylogeny. Gromkova et al. (1998) observed that the success of static aerobic transformation in *Hp* was related to the biotype and clumpiness of the strain. The reasons behind differing levels of competence in *Hp* are likely to be complex, but by experimenting with several transformation methods we have determined protocols which yield transformants in each key strain. This work lays the foundation for further analysis of OAg genetics and function.

5 Results: Further investigation of O-antigen genes and their predicted functions

The OAg is one of the most diverse structures found in Gram negative bacteria, differing on both the species and strain levels. It is the basis of typing schemes for many species, using antisera raised specifically to each OAg structure to test for reactivity, and in some cases a correlation can be seen between serogroup and clinical symptoms. For *Escherichia coli,* over 180 different OAgs have been proposed (Lundborg et al., 2010): each contains one to seven sugars per repeat unit, with extra variation added by different sugar conformations, linkages, branching patterns and modifications. Other species are more conservative; for example, there are only seven known serotypes for *Aggregatibacter actinomycetemcomitans* (Kaplan et al., 2001; Takada et al., 2010). As *Hp* OAgs have never been studied in detail, the number of serogroups is completely unknown and no antiserum is available to test for the *Hp* T3T1 OAg structure described earlier. In this chapter we aim to determine whether each *Hp* strain contains the genes necessary for OAg production, and how the OAgs of different strains are related.

As seen in the previous chapter, sequence analysis of an OAg locus can aid prediction of the OAg structure of a particular bacterial isolate. *Hp* strains 8, 15, 20 and 22 were chosen for inclusion in the Genomophilus genome sequencing project in the hope that it would provide intact OAg locus DNA sequences. However, for all four strains, the contigs covering the OAg locus flanking genes (*glnA* and *pepB*) terminated immediately before the locus, and no BLAST hits were found in the database for the majority of the *Hp* T3T1 OAg locus genes. OAg related

genes in *Hp* T3T1 and other Gram negative species tend to exhibit a low % GC content, and it was observed that regions with this attribute generally had very little or no coverage in Genomophilus (Section 1.5.3; Peter Power, pers. comm.).

Here we present OAg locus sequences obtained by an alternative method and use this information to compare the OAg loci of all *Hp* study strains. BLASTP analysis is used to predict OAg components where possible; this evidence is integrated with further LPS structural data obtained from the Schweda Laboratory. Finally, as several of the OAg loci appeared to lack an OAg ligase gene, the results of a search for a ligase encoded outside the OAg locus are described.

5.1 Categorisation of *Hp* OAg loci

5.1.1 Cloning and sequencing of the *Hp* strain 13 OAg locus

No information was yet available about the sequence and organisation of the putative OAg loci in most *Hp* strains. We therefore employed a sequencing strategy that does not initially require internal locus-specific primers. *Hp* 13 was chosen as the first strain for analysis because its OAg appears to have a different chain length distribution and O-unit size to that of *Hp* T3T1 (Fig. 4.6), its *glnA-pepB* LR-PCR product is intriguingly short (Fig. 4.4a), and it was negative by PCR amplification for gene *Hp0272*. The LR-PCR product from *Hp* 13 was digested, ligated into a pBluescript vector, cloned in *E. coli* and sequenced (Fig. 5.1). The sequences obtained from different cloned fragments of the OAg locus were assembled into contigs; further PCR amplification and sequencing enabled the contigs to be ordered and assembled into a complete in

silico 10.14 kb OAg locus sequence. The organisation of this locus, which contains ten ORFs, is summarised in Fig. 5.10 alongside others that were sequenced later.

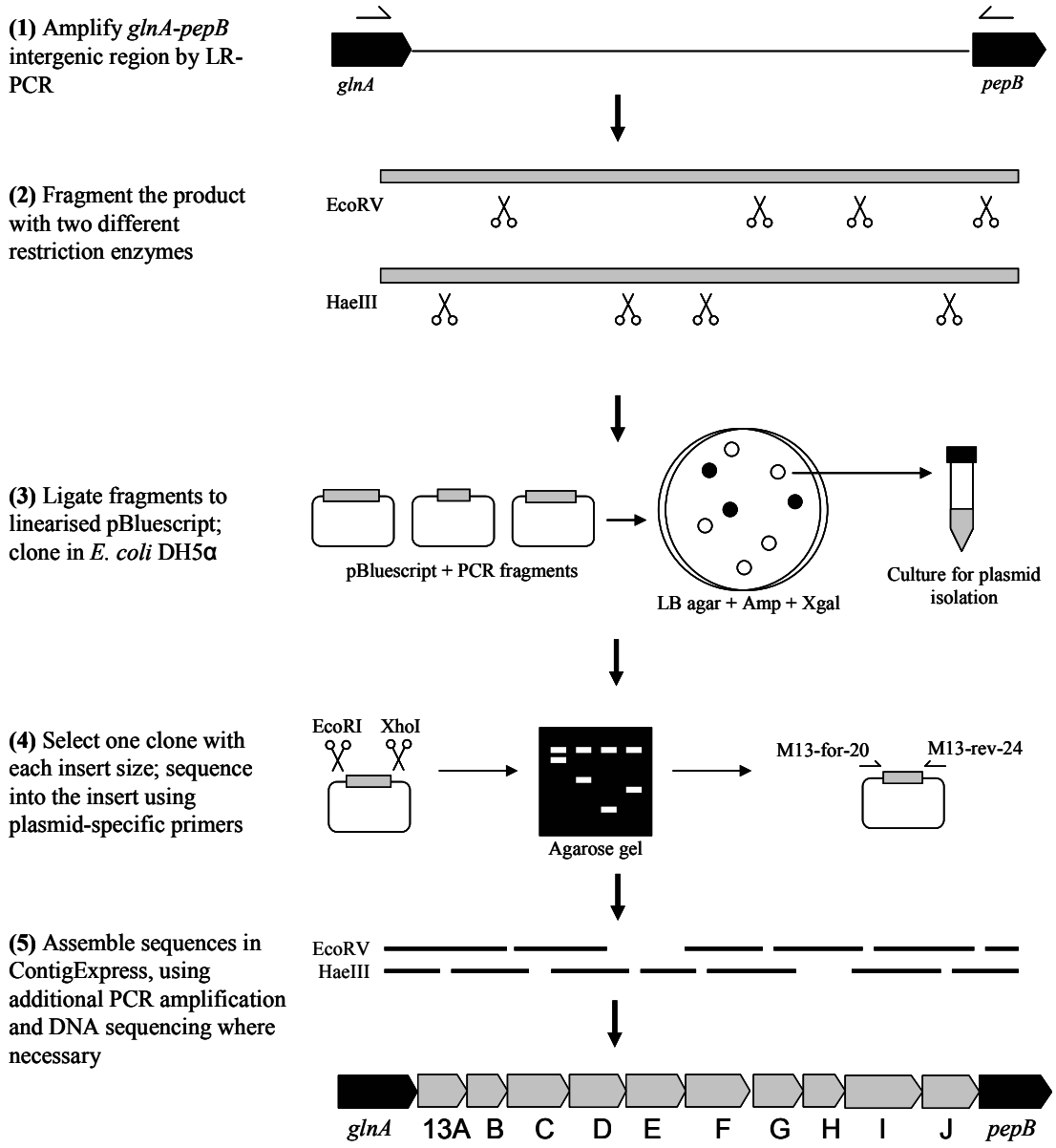


Figure 5.1. Scheme for cloning and sequencing the *Hp* O-antigen loci. This general method was used to clone and sequence the OAg loci of *Hp* strains 13, 17, 30 and 20. Further details can be found in the Materials and Methods chapter.

BLASTP searches of the ORFs in the *Hp* 13 *glnA-pepB* intergenic region predicted an almost entirely different set of enzymes to those encoded by the *Hp* T3T1 locus, with only two genes bearing any similarity between the strains. However, the *Hp* 13 ORFs still encode typical OAg synthesis proteins including nucleotide-sugar synthesis enzymes, glycosyltransferases (GTs) and an OAg transport system (Table 5.1).

The first two ORFs (*13A* and *13B*) encode the permease and ATPase subunits of an ABC transporter. Other bacterial species use ABC transporters to transfer completed OAg or capsule polysaccharide chains from the cytoplasmic to the periplasmic face of the inner membrane, as an alternative to the Wzx/Wzy/Wzz system. The transporter encoded by the *Hp* 13 OAg locus shares particular homology with the ABC2 subfamily, which also includes the *Hi* capsule locus transporter composed of BexA/BexB/BexC (Reizer et al., 1992). The closest homologues to *13A* and *13B* are found in *Actinobacillus pleuropneumoniae* and *Aggregatibacter aphrophilus*: it is likely that transfer of these genes has occurred between the Pasteurellaceae species.

Genes *13C* and *13D* encode proteins in the WcaA (COG0463) family of cell envelope and outer membrane GTs. As the closest matching proteins in the NCBI database are <40% homologous and of unknown function it is difficult to predict the sugar specificity of these two putative GTs.

Amino acids 9-321 of the 367 aa protein encoded by gene *13E* correspond to the acyltransferase family Pfam01757 with an E-value of 3.8E-06. Some of the homologues of protein 13E in other species are O-acetyltransferases encoded by ORFs near GT genes, suggesting that they may acetylate polysaccharides (e.g. OAg or capsule); the most likely role of 13E is OAg acetylation.

Table 5.1. BLASTP analysis of the ten ORFs in the *Hp* strain 13 O-antigen locus. Gene designations are given in the left hand column. The second column lists the putative size of the encoded protein in amino acids (aa). The species containing the three highest scoring BLASTP matches are shown, but only the top match for each genus is included. For each match the % aa identity (%aa ID) is given, with % query coverage (QC) in brackets. The putative function of each *Hp* protein is based on the known functions of homologues and on domain model searches (see text).

			Highest scoring BLASTP matches	
Gene	**aa**	**Putative function**	**Species**	**%aa ID (QC)**
13A	263	ABC transporter, permease subunit (Wzm)	*Actinobacillus pleuropneumoniae*	80 (100)
			Aggregatibacter aphrophilus	81 (98)
			Roseburia intestinalis	56 (100)
13B	243	ABC transporter, ATPase subunit (Wzt)	*Actinobacillus pleuropneumoniae*	94 (99)
			Aggregatibacter aphrophilus	84 (99)
			Coprococcus comes	67 (99)
13C	338	Glycosyltransferase (likely family 2)	*Ruminococcus gnavus*	34 (96)
			Eubacterium rectale	35 (92)
			Clostridium sp.	34 (93)
13D	334	Glycosyltransferase (likely family 2)	*Lactobacillus delbrueckii*	30 (91)
			Dyadobacter fermentans	36 (58)
			Bacteroides thetaiotaomicron	37 (65)
13E	367	O-acetyltransferase	*Erwinia amylovora*	46 (99)
			Actinobacillus succinogenes	48 (99)
			Acidovorax sp.	32 (99)
13F	382	UDP-galactopyranose mutase (Glf)	*Actinobacillus pleuropneumoniae*	91 (100)
			Aggregatibacter aphrophilus	88 (100)
			Simonsiella muelleri	74 (98)
13G	308	Glycosyltransferase (family 2)	*Actinobacillus pleuropneumoniae*	58 (99)
			Aggregatibacter aphrophilus	52 (98)
			Nocardioides sp.	43 (98)
13H	256	Glycosyltransferase	*Actinobacillus pleuropneumoniae*	57 (100)
			Aggregatibacter aphrophilus	54 (95)
			Clostridium botulinum	51 (98)
13I	471	UndP-sugar phosphotransferase (WbaP)	*Actinobacillus succinogenes*	59 (100)
			Mannheimia succiniciproducens	59 (100)
			Haemophilus influenzae	56 (100)
13J	336	dTDP-glucose 4,6-dehydratase	*Haemophilus parainfluenzae* T3T1	93 (99)
			Actinobacillus minor	70 (99)
			Neisseria meningitidis	69 (99)

Gene *13F* encodes a protein belonging to the Glf multidomain family of UDP-galactopyranose mutases (Pfam03275) with an E-value of 1E-95. Glf catalyses the conversion of UDP-galactopyranose to UDP-galactofuranose so that Gal*f* can be incorporated into structures such as the mycobacterial cell wall (Pan et al., 2001) or *E. coli* OAg (Nassau et al., 1996). The different conformations of Gal*p* (6-membered ring) and Gal*f* (5-membered ring) may confer different biological properties on the resulting structures; Gal*f* is thermodynamically less stable and occurs much less frequently in nature. The C-terminal half of the 13F protein represents a predicted NAD-binding Rossmann domain, although structural characterisation of Glf from other species has shown that the usual cofactor is flavin adenine dinucleotide (FAD) (Sanders et al., 2001). The presence of a *glf* homologue within the *Hp* strain 13 OAg locus strongly suggests that one of the sugars in the O-unit is Gal*f* or a derivative.

The third predicted GT encoded by the locus is 13G, which belongs to the GT2 family. Its closest homologues in the NCBI database are uncharacterised proteins from *A. pleuropneumoniae* (serotypes 1, 5b and 10) and *A. aphrophilus*. However, 13G also shares 36% aa identity with an *E. coli* serotype O167 GT, WfdJ, which is predicted to catalyse the addition of Gal*f* in a β1,3 linkage to Glc*p*NAc as part of OAg synthesis (Lundborg et al., 2010).

Protein 13H shares 57% aa identity with a protein encoded within the OAg loci of *A. pleuropneumoniae* serotypes 10 and 12 (annotated as EpsF). The repeat unit of serotype 10 is simply one Gal*f* residue whilst that of serotype 12 is a branched trisaccharide comprising Gal*f* and Gal*p*. This suggests a possible role for EpsF in the synthesis or transfer of Gal*f*. Although

some of the lower BLASTP matches for protein 13H are annotated as GTs, no putative conserved domains were detected in 13H.

Gene *13I* encodes a protein with similarity to the UndP sugar phosphotransferase (WbaP) domain family, TIGR03022. Unlike the *Hp* T3T1 WbaP homologue (WcfS; Hp0273), 13I is a full-length protein with respect to other members of this family. It shares 56% aa identity with *Hi* strain Rd WbaP (HI0872) which is understood to transfer Gal to UndP as the first step of HMG synthesis.

The final gene in the *Hp* 13 OAg locus, *13J*, shares significant homology with the dTDP-Glc dehydratase-like gene found in *Hi* HMG loci (*HI0873*) and in the *Hp* T3T1 OAg locus (*Hp0270*), whose role in HMG and OAg synthesis is unclear (see Section 4.1.1).

After the *Hp* 13 OAg locus had been sequenced and analysed, the LPS structure of this strain was partly elucidated by mass spectrometry and NMR techniques (Fig. 5.2; Brigitte Twelkmeyer, pers. comm.). The core LPS was found to be the same as that of *Hp* T3T1 except that HepII was elaborated with Glc via an α1,3 linkage, as predicted from the detection of a *lic2C* homologue in this strain (Section 3.2.2). The OAg is based on a repeating disaccharide of →6)-Gal*f*-(β1,3)-Glc*p*NAc-(β1→. It is likely that Gal*f* is synthesised by the putative Glf enzyme (13F) as discussed above and that it is added to GlcNAc by 13G, as this is the exact predicted sugar and linkage specificity of its *E. coli* homologue WfdJ. As the hexose linking the repeating OAg chain to the inner core LPS has yet to be identified, we cannot deduce the substrate specificity of the WbaP homologue 13I. Two more GTs would probably be required, one to add GlcNAc to the hexose and another to add GlcNAc to Gal*f* in the repeating chain, and these are most likely the

GTs encoded by genes *13C* and *13D* (or possibly *13H*). It is unusual for the ABC transporter dependent mechanism of OAg synthesis and transfer to be associated with heteropolymeric OAg (see Section 1.4.1). An OAg ligase is usually required to ligate the OAg to the LPS core, but no such enzyme was found in the *Hp* 13 OAg locus; it may be present in a different genomic location.

The *Hp* 13 OAg structure also contains PEtn and OAc sidechains. Presumably OAc is added by the predicted O-acetyltransferase 13E, but no PEtn transferase was identified in the OAg locus. The HMG unit in *Hi* is decorated with PEtn by an enzyme encoded outside the *hmg* locus (Table 3.2; Derek Hood, pers. comm.) and a similar situation may occur in *Hp* strain 13. In species such as *Bordetella bronchiseptica* and *Shigella flexneri*, certain OAgs are known to undergo late modification by OAc or PEtn transferases after the chain has been transported to the periplasm (Allison & Verma, 2000; King et al., 2009).

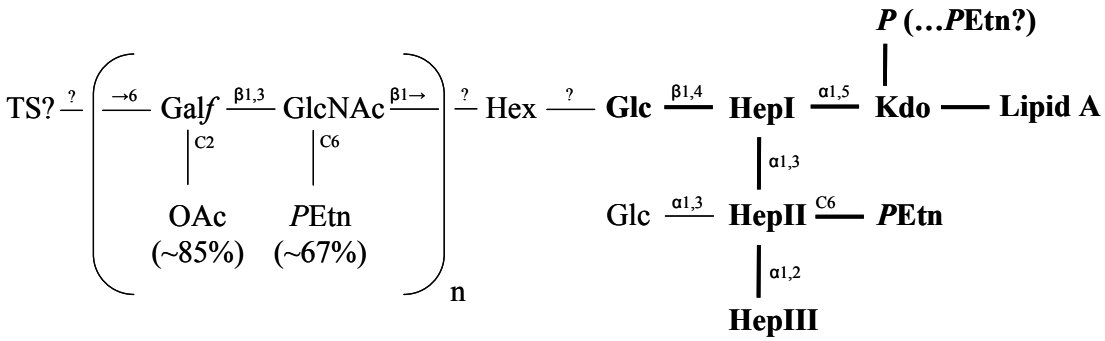


Figure 5.2. LPS structure of *Hp* strain 13. The structure was determined using mass spectrometry and NMR techniques (Brigitte Twelkmeyer, pers. comm.). Bold type represents the inner core LPS, which is the same as that of *Hp* strain T3T1. The repeating disaccharide O-unit is shown in brackets; it is attached to the inner core via a hexose residue (Hex) that has yet to be identified. It is not known whether the chain terminates with one of the O-unit sugars or a different residue (TS = terminal sugar). All sugars are in the pyranose conformation except the O-unit Gal, which is a furanose (Gal*f*).

5.1.2 Most *Hp* strains have an *Hp* T3T1-like (group 1) or *Hp* 13-like (group 2) OAg locus

To investigate whether other *Hp* strains have similar OAg loci to the two sequences that were now available, internal primers were designed to each of the genes in the *Hp* 13 OAg locus and to some of the genes that had yet to be studied from the *Hp* T3T1 OAg locus. The *Hp* study strains were tested for PCR amplification of each gene, and the results are summarised in Table 5.2 alongside data that was obtained later.

The PCR analysis was striking in that it immediately separated 15 of the 18 true *Hp* strains into two clear OAg locus categories. Seven strains, designated as group 1, contained homologues of the genes needed for AAT synthesis and transfer in *Hp* T3T1 (genes *Hp0274*, *Hp0273* and *Hp0272*). The high frequency of these genes suggests that AAT is particularly beneficial for successful colonisation. A subset of these seven strains also gave PCR products for the two putative *Hp* T3T1 GT genes and/or the sialic acid synthase gene *neuB*. The *neuB*-positive strains correspond to the five largest LR-PCR products in Fig. 4.4a.

Table 5.2. PCR analysis of the *Hp* study strains to test for homologues of known OAg locus genes. Primers were designed to most genes in the *Hp* strains 13 and 26 (T3T1) OAg loci and were used to test for the presence of each gene by PCR amplification in the *Hp* study strains. Primers designed to some genes from the sequenced but ungrouped OAg loci of *Hp* strains 17 and 30 (*17B*, *17C*, *17J*, *30B* and *30E*) were also included. *Hp* strain numbers are listed along the top, sorted according to the OAg locus groups which became apparent during this experiment. The origin of each strain is given below the strain number as either the Gambia (G) or the United Kingdom (UK).

PCR results are scored according to the products visible by agarose gel electrophoresis. '+' indicates a PCR in which a product of the expected size was amplified; '-' indicates that no product was detected. Boxes are left empty where no PCR was attempted. The *neuA-neuD* genes are usually found together, so primers were designed only to *neuB*; strains which were positive for *neuB* are shaded for the other *neu* genes (NT = not tested).

For some genes (indicated by an asterisk next to the primer names), a second primer pair was tested on strains which had given negative results. These alternative primers were designed to the *Hp* strain 19 or *Hp* strain 20 allele and are listed in Table 2.4; PCRs which gave a product only for the second primer pair are marked '19' or '20' respectively.

				Group 1 OAg loci							Group 2 OAg loci								Ungrouped			Hybrid	
			Hp strain #	2	19	20	22	24	26	31	8	10	13	14	15	16	18	35	17	30	34	6	11
	Gene	Putative function	Primers	UK	G	UK	UK	G	G	G	UK	UK	G	UK	G	UK	G	UK	G	G	G	UK	UK
Hp strain T3T1 (26) OAg locus	*Hp0285*	Chain length determinant	285a/285b*	-	19	-	19	19	+	19	-	-	-	-	-	-	-	-	-	-	-	-	-
	Hp0284	NeuD	NT																				
	Hp0283	NeuB	283a/283b	-	+	-	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
	Hp0282	NeuA	NT																				
	Hp0281	NeuC	NT																				
	Hp0280	O-acetyl transferase	280a/280b	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Hp0279	OAg flippase	279a/279b*	-	+	20	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Hp0278	Sialyltransferase	278a/278b	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Hp0277	OAg polymerase	277b/277c*	-	+	20	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Hp0276	Glycosyltransferase	276a/276b	-	+	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-		
	Hp0275	Glycosyltransferase	275a/275b*	-	+	20	20	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Hp0274	Aminotransferase	272c/274a	+	+	+	+	+	+	+										-		-		
	Hp0273	UndP-AAT P-transferase	Sequencing data	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Hp0272	UDP-GlcNAc dehydratase	272a/272b	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Hp0271	OAg ligase	271b/271d*	20	20	20	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
	Hp0270 (dTDP-Glc dehyd.) to *pepB*		hmg13W/269a	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
Hp strain 13 OAg locus	*glnA* to *13A* (ABC permease)		hmg13Q/286F4								+	+	+	+	+	+	+	-			-	-	-
	13A	ABC (permease subunit)	gr2-ABC1/hmg13Q	-	-	-	-	-	-	-	+	+	+	+	+	+	+	-	-	-	-	-	-
	13B	ABC (ATPase subunit)	hmg13R/hmg13S	-	-	-	-	-	-	-	-	+	+	-	+	-	+	-	-	-	-	+	-
	13C	Glycosyltransferase	gr2-GT1/hmg18B	-	-	-	-	-	-	-	-	-	+	-	+	-	+	-	-	-	-	-	-
	13D	Glycosyltransferase	gr2-GT2/gr2-GT3	-	-	-	-	-	-	-	-	-	+	-	+	-	+	-	-	-	-	-	-
	13E	O-acetyl transferase	gr2-AC1/hmg13D	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	-	-	-
	13F	UDP-Gal*p* mutase	gr2-mut1/gr2-mut2	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	-	-	-	+	+
	13G	Glycosyltransferase	hmg13E/gr2-GT4	-	-	-	-	-	-	-	-	-	+	-	-	-	+	+	-	-	-	+	-
	13H	Glycosyltransferase	gr2-GT5/gr2-GT6	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	-	-	-	+	+
	13I	UndP-sugar P-transferase	hmg13G/hmg13H	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	-	-	-	+	+
	13I to *13J* (dTDP-Glc dehyd.)		hmg13G/hmg13V								+	+	+	+	+	+	+	+				+	+
	17B	OAg polymerase	POL17a/POL17b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
	17C	OAg flippase	FLP17a/FLP17b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
	17J	Ribose-5P reductase	OAG17H/OAG17R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
	30B	Glc1P thymidylyltransferase	rmlA-F/rmlA-R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-
	30E	ABC (permease subunit)	hmg30N/hmg30P																	+	+		

A mutually exclusive set of eight *Hp* strains were positive for the *Hp* strain 13 UndP-sugar phosphotransferase gene *13I*, and for *13H* which we predict to encode an enzyme involved in the synthesis or transfer of Gal*f* (Section 5.1.1). All of these strains except *Hp* 35 also appear to encode a similar ABC transporter (permease subunit) to that of the *Hp* 13 OAg locus. The eight *Hp* 13-like loci were designated as group 2. In addition the two hybrid strains, *Hp* 6 and 11, share a few of the group 2 OAg locus genes. Both group 1 and group 2 OAg loci are found in a mix of Gambian and UK strains, but the group 2 loci from Gambian strains (*Hp* 13, 15 and 18) appear to be particularly closely related (Table 5.2).

OAg locus mapping by PCR amplification of various combinations of genes indicated that the order of the genes that are present is generally conserved within the groups (data not shown). It is evident from Fig. 4.4a that the OAg loci of *Hp* strains 14 and 16 are several kilobases longer than that of *Hp* strain 13; perhaps the central region comprises a novel set of GT genes and/or sugar synthesis genes instead of ORFs *13C* to *13E* (Table 5.2).

No strains were observed to give PCR products for both group 1 and group 2 genes: although *Hp* T3T1 and *Hp* 13 share the last ORF in the OAg locus (*Hp0270* / *13J*) there are no other genes in common, so the exchange of sections of the OAg locus between group 1 and group 2 strains by homologous recombination would be unlikely. The different assembly systems used by the two groups (Wzy-dependent or ABC transporter) would also impose constraints on reassortment, if the OAg locus is to remain functional.

For three strains (*Hp* 17, 30 and 34), no PCR products were obtained using any of the group-specific *Hp* T3T1 or *Hp* 13 primers. To investigate whether these strains contained novel sets of OAg genes, the *glnA-pepB* intergenic regions of *Hp* strains 17 and 30 were digested, cloned in *E. coli* and sequenced using the method explained in Fig. 5.1. Both of these strains produce unique OAg-like ladders visible on a heavily loaded SDS-PAGE gel, suggesting that their OAg loci are likely to be both novel and functional.

5.1.3 Sequencing and analysis of the *Hp* strain 17 OAg locus

The *Hp* 17 *glnA-pepB* intergenic region was successfully sequenced and the DNA sequence was assembled, revealing 14 ORFs over 14.8 kb. The mean GC content is 31.0% and all of the ORFs are in the same orientation. The 14 genes appear to encode functions that are typical of OAg synthesis and assembly (Table 5.3) but form a strikingly different OAg locus from those of *Hp* strains T3T1 and 13.

Although the *Hp* 17 OAg locus encodes the proteins of the Wzy-dependent OAg transport system (genes *17A* to *17C*), only the chain length determinant protein Wzz shares much homology with its *Hp* T3T1 counterpart (90% aa identity); the OAg polymerase and flippase of the two strains are only 25% and 19% identical respectively. This suggests that *Hp* T3T1 and *Hp* 17 have acquired *wzy*/*wzx* independently and perhaps from different species as indicated in Tables 4.1 and 5.3. There is no putative OAg ligase gene in the *Hp* 17 OAg locus: as for *Hp* 13, if the OAg is attached to the LPS there must be an OAg ligase gene elsewhere in the genome.

Table 5.3. BLASTP analysis of the 14 ORFs in the *Hp* strain 17 O-antigen locus. Gene designations are given in the left hand column. The second column lists the putative size of the encoded protein in amino acids (aa). The species containing the three highest scoring BLASTP matches are shown, but only the top match for each genus is included. The capsular serotype (for *S. pneumoniae*) or OAg serotype (for *A. pleuropneumoniae*) of strains which contain the protein are given in brackets to aid prediction of the *Hp* strain 17 OAg structure. For each match the % aa identity is given (%aa ID), with % query coverage (QC) in brackets. The putative function of each *Hp* protein is based on the known functions of homologues and on domain model searches (see text).

			Highest scoring BLASTP matches	
Gene	**aa**	**Putative function**	**Species (serotype)**	**%aa ID (QC)**
17A	374	OAg chain length determinant (Wzz)	*Hp* T3T1	90 (99)
			Mannheimia succiniciproducens	64 (97)
			Actinobacillus minor	47 (97)
17B	392	OAg polymerase (Wzy)	*Lactobacillus vaginalis*	29 (91)
			Eubacterium rectale	27 (97)
			Streptococcus pneumoniae	33 (71)
17C	476	OAg flippase (Wzx)	*Actinobacillus pleuropneumoniae*	57 (98)
			Catenibacterium mitsuokai	51 (99)
			Ruminococcus flavefaciens	50 (99)
17D	368	UDP-galactopyranose mutase (Glf)	*Streptococcus pneumoniae* (13, 29, 33B, 35B etc.)	88 (99)
			Lactobacillus plantarum	77 (99)
			Granulicatella elegans	79 (99)
			(*Hp* strain 13, gene 13F)	45 (100)
17E	263	GT (probable Gal*f* transferase)	*Granulicatella elegans*	66 (100)
			Streptococcus pneumoniae (17A, 20, 29, 35A, 35B)	67 (100)
			Mannheimia succiniciproducens	58 (99)
			(*Hp* strain 13, gene 13H)	39 (96)
17F	283	Phosphotransferase (WefL)	*Streptococcus pneumoniae*	64 (95)
			Catonella morbi	58 (95)
			Ruminococcus hydrogenotrophica	43 (93)
17G	325	Glycosyltransferase (likely family 2)	*Streptococcus mitis*	56 (97)
			Gemella haemolysans	58 (97)
			Granulicatella adiacens	54 (96)
17H	332	Glycosyltransferase (likely family 2)	*Streptococcus pneumoniae* (20, 33B, 33F)	50 (96)
			Clostridium sp. M62/1	39 (95)
			Ruminococcus gnavus	36 (98)
17I	235	Ribitol-5P cytidylyltransferase (RblB)	*Streptococcus pneumoniae* (all serotypes)	70 (99)
			Gemella haemolysans	72 (99)
			Bacillus subtilis	54 (99)
17J	340	Ribose-5P reductase (RblA)	*Gemella haemolysans*	64 (100)
			Streptococcus pneumoniae	63 (100)
			Bacillus subtilis	43 (100)
17K	321	Glycosyltransferase (likely family 14)	*Gemella haemolysans*	54 (94)
			Granulicatella elegans	50 (95)
			Streptococcus pneumoniae	50 (94)
17L	473	UndP-sugar phosphotransferase (WbaP)	*Actinobacillus pleuropneumoniae* (4, 7, 13)	69 (100)
			Haemophilus influenzae (HMG locus)	66 (100)
			Mannheimia succiniciproducens	67 (98)
17M	350	Acyltransferase	*Dialister invisus*	44 (98)
			Bacillus mycoides	31 (92)
			Bacteroides fragilis	30 (83)
17N	335	dTDP-glucose 4,6-dehydratase	*Hp* T3T1 and *Hi* Rd	91 (99)
			Actinobacillus minor	70 (99)
			Mannheimia haemolytica	69 (99)

The OAg structure of *Hp* strain 17 LPS has not been analysed, but the sugar synthesis and transfer genes sequenced here provide many clues to its possible constituents. Several of the gene products have homologues in particular *Streptococcus pneumoniae* (*Sp*) serotypes, where they are involved in capsule synthesis. The correlation between capsule structure and capsule locus genes has been intensively studied for the 90 pneumococcal serotypes (Aanensen et al., 2007; Bentley et al., 2006), allowing allocation of function to many of the gene products.

Two genes in the *Hp* 17 OAg locus provide evidence for the presence of Gal*f* in the OAg. Firstly, the product of gene *17D* shares 88% aa identity with Glf from *Sp*, where it is present in strains that include Gal*f* in their capsule. Thus the enzyme encoded by *17D* is likely to carry out the same function as the *Hp* 13 Glf, namely the conversion of UDP-Gal*p* to UDP-Gal*f*. Secondly, gene *17E* is a homologue of pneumococcal *wciB*, whose product has been categorised as a Gal*f* transferase and always adds the sugar via a β1,3 linkage (Aanensen et al., 2007).

BLASTP searches of genes *17F*, *17I* and *17J* indicate that another of the *Hp* 17 OAg sugars may be ribitol phosphate. 17I and 17J are homologues of the enzymes required for the two-step conversion of D-ribulose-5P to CDP-ribitol (Baur et al., 2009) which is used to make the polyribitol-phosphate component of teichoic acid in the Gram positive cell wall. *Hi* capsule types a and b also contain ribitol, but a bifunctional enzyme, Bcs1, performs both synthesis steps (Zolli et al., 2001); 17I and 17J are more closely related to the individual *Sp* enzymes RblB and RblA and the equivalent *Bacillus subtilis* enzymes TarI and TarJ. Ribitol phosphate is an unusual OAg component and would contribute negative charge to the O-unit.

The closest homologues of 17F are encoded within the capsule loci of several *Sp* serotypes whose capsule structures are unknown, but it also shares 61% aa identity with WefL in *S. oralis* strains C104 and SK144, which is proposed to transfer ribitol-5-phosphate to Gal*f* (Yang et al., 2009). Given the genetic evidence presented above, 17F may share this same sugar specificity.

Gene *17H* encodes a putative family 2 GT with a homologue (WciD) in the capsule loci of four *Sp* serotypes. Aanensen et al. (2007) could not assign an exact linkage to *Sp* WciD but hypothesised that the sugars involved were some combination of Gal*f*, Gal*p* and Gal*p*NAc. Gene *17G* also appears to encode a family 2 GT, similar to enzymes from *Enterococcus faecalis* and *Streptococcus* species, but none of its homologues have been functionally characterised. The protein encoded by gene *17K* shares homology with WcrG from *Sp* serotype 10A, which is a family 14 GT assigned to the role of adding Gal*p* to Gal*p*NAc via a β1-6 linkage (Aanensen et al., 2007). Gene *17M* encodes a protein with homology to putative acyltransferases from a variety of bacteria, but appears unrelated to the predicted acyltransferases from the OAg loci of *Hp* strains T3T1 and 13.

Protein 17L, the predicted UndP-sugar phosphotransferase for the *Hp* 17 OAg locus, is full length like its *Hp* 13 homologue. The 215 aa C-terminal domain of 17L (usually the region with phosphotransferase activity (Wang et al., 1996)) is 88% identical to that of *Hi* Rd WbaP, suggesting that the first sugar of the *Hp* 17 O-unit may be Gal. However, it is important to bear in mind that the sugar specificity of an enzyme can be altered by a single amino acid change (see Section 3.2.3).

In summary, the *Hp* 17 OAg locus includes genes encoding six putative GTs and one acyltransferase, suggesting an O-unit comprising up to six sugars and at least one acyl group. The closest homologues of each of the six GTs are three Gal transferases (including WbaP), one ribitol-P transferase and two GTs of unknown specificity. Other genes in the locus appear to encode synthesis enzymes for UDP-Gal*f* and CDP-ribitol, and the Wzy-dependent assembly and transport system.

5.1.4 Sequencing and analysis of the *Hp* strain 30 OAg locus

The *Hp* 30 *glnA-pepB* intergenic region was cloned and sequenced, and the DNA sequence was assembled; it contains 12 ORFs over a length of 12.1 kb. The first 11 of these ORFs correspond to the majority of the 12-gene *Aggregatibacter* (previously *Actinobacillus*) *actinomycetemcomitans* (*Aa*) serotype c OAg locus, with 71-94% aa identity between gene products (Table 5.4). This greatly aids our prediction of the *Hp* 30 OAg structure, as the *Aa* serotype c OAg has been studied in detail and is known to comprise →3)-6-deoxy-α-L-talose-(1,2)-6-deoxy-α-L-talose-(1→ with acetylation at the O-4 position of the first of these two sugars (Nakano et al., 1998; Shibuya et al., 1991).

Nakano and colleagues (2000) elucidated the functions of some of the *Aa* serotypes b and c OAg enzymes by overexpressing them in *E. coli* and assaying the purified proteins in vitro. They confirmed that RmlA, RmlB and RmlC from serotype c strains are the three enzymes needed to convert Glc-1-P and dTTP into dTDP-6-deoxy-L-*lyxo*-4-hexulose. This intermediate can then be used to synthesise either dTDP-L-rhamnose (using RmlD) or dTDP-6-deoxy-L-talose (using Tll) (Nakano et al., 2000). The authors were surprised to find that the serotype c RmlD homologue

synthesised dTDP-L-rhamnose, because the O-unit of this serotype does not appear to contain rhamnose. Perhaps the nucleotide sugar is made but cannot be transferred to the OAg due to a defective GT, or alternatively a single undetected rhamnose residue may initiate or cap the OAg chain. Either of these hypotheses would be consistent with the fact that the *Aa* serotype c locus contains three GT genes where only two GTs would seem necessary for assembly of the O-unit.

The *Hp* 30 OAg locus encodes homologues of RmlA-RmlD (30A-30D) and Tll (30H), so this strain may be able to synthesise both dTDP-6-deoxy-L-talose and dTDP-L-rhamnose. Genes *30G*, *30J* and *30K* are similar to the three GT genes in the *Aa* serotype c locus, therefore the hypotheses presented above would also apply to *Hp* strain 30. The twelfth and final gene of these two loci encodes a predicted UndP-sugar phosphotransferase; however, the *Hp* 30 allele shares more homology with *Hi* WbaP (which transfers Gal (Hood et al., 2004b)) and proteins from *A. pleuropneumoniae* serotype 7 and 13 (may transfer Gal or Rha (MacLean et al., 2004)) than with the *Aa* serotype c allele. This makes it difficult to predict the sugar that links the OAg to the LPS core in *Hp* 30.

Overall, the O-unit of *Hp* strain 30 is likely to be very similar to that of *Aa* serotype c, including one or more 6-deoxy-L-talopyranose residues and some degree of acetylation carried out by the product of gene *30I*. The OAg may also contain L-rhamnose, which differs from 6-deoxy-L-talose only in the stereochemistry of the C4 carbon. 6-deoxy-L-talose is not a common component of bacterial polysaccharides but it has been found in the OAg of three *E. coli* serotypes (Jann et al., 1995) and the *Mesorhizobium loti* type strain (Russa et al., 1995).

Table 5.4. BLASTP analysis of the *Hp* strain 30 O-antigen locus. Gene designations are given in the left hand column. The second column lists the putative size of the encoded protein in amino acids (aa). For the first 11 genes, the highest scoring BLASTP match is from *Aggregatibacter actinomycetemcomitans* (*Aa*) strain D11S-1, with gene order conserved between the two OAg loci (see locus tag numbers). For gene *30L* the highest scoring BLASTP matches were from *Actinobacillus pleuropneumoniae* (*Apl*) serotype 7 strain AP76 and *Hi* strain Rd. For each match the % aa identity (%aa ID) is given, with % query coverage (QC) in brackets. The putative function of each *Hp* protein is based on the known functions of homologues and on domain model searches (see text).

			Highest scoring BLASTP match	
Gene	**aa**	**Putative function**	**Species (Locus tag)**	**%aa ID (QC)**
30A	354	dTDP-Glc 4,6-dehydratase (RmlB)	*Aa* (D11S_1710)	92 (100)
30B	290	Glc-1-phosphate thymidylyltransferase (RmlA)	*Aa* (D11S_1709)	94 (99)
30C	294	dTDP-6-deoxy-L-*lyxo* -4-hexulose reductase (dTDP-L-rhamnose synthase; RmlD)	*Aa* (D11S_1708)	71 (100)
30D	180	dTDP-6-deoxy-D-*xylo* -4-hexulose-3,5-epimerase (RmlC)	*Aa* (D11S_1707)	94 (100)
30E	258	ABC transporter, permease subunit (Wzm)	*Aa* (D11S_1706)	90 (99)
30F	247	ABC transporter, ATPase subunit (Wzt)	*Aa* (D11S_1705)	83 (99)
30G	633	Glycosyltransferase (rhamnose?)	*Aa* (D11S_1704)	88 (100)
30H	270	dTDP-6-deoxy-L-*lyxo* -4-hexulose reductase (dTDP-6-deoxy-L-talose synthase; Tll)	*Aa* (D11S_1703)	87 (100)
30I	344	Acyl transferase	*Aa* (D11S_1702)	81 (100)
30J	299	Glycosyltransferase	*Aa* (D11S_1701)	76 (96)
30K	309	Glycosyltransferase (GTA superfamily; rhamnose?)	*Aa* (D11S_1700)	79 (98)
30L	472	UndP-sugar phosphotransferase (WbaP)	*Apl* (APP7_1480) *Hi* (HI_0872) *Aa* (D11S_1699)	61 (100) 58 (100) 47 (86)

It was recently discovered in an *Aa* serotype b strain that glycosylation of the adhesin EmaA is dependent upon the OAg synthesis and ligase enzymes (Tang & Mintz, 2010). This finding was unexpected because protein glycosylation pathways are usually separate from LPS synthesis in Gram negative bacteria. It raises the interesting possibility that *Hp* 'OAg' loci, in particular that of *Hp* 30 which is so closely related to an *Aa* locus, may also be responsible for the glycosylation of some cell surface proteins.

Primers designed to the *Hp* 30 genes *30B* (rhamnose/deoxytalose synthesis pathway gene) and *30E* (ABC transporter permease subunit gene) amplified PCR products from *Hp* strain 34 gDNA

(Table 5.2), and the MfeI digestion pattern of the LR-PCR product from this strain includes a fragment of approximately the same molecular weight as the *30G-30L* fragment from *Hp* 30 (6.5 kb; Fig. 4.4b). It therefore seems likely that *Hp* 34 contains an OAg locus of a broadly similar composition to that of *Hp* 30, although no OAg has been observed for the former strain. These results for *Hp* strain 34 mean that putative OAg related genes have now been detected in all 20 of the study strains.

5.1.5 Analysis of *Hp* undecaprenyl phosphate sugar phosphotransferase genes (*wbaP* or *wcfS*)

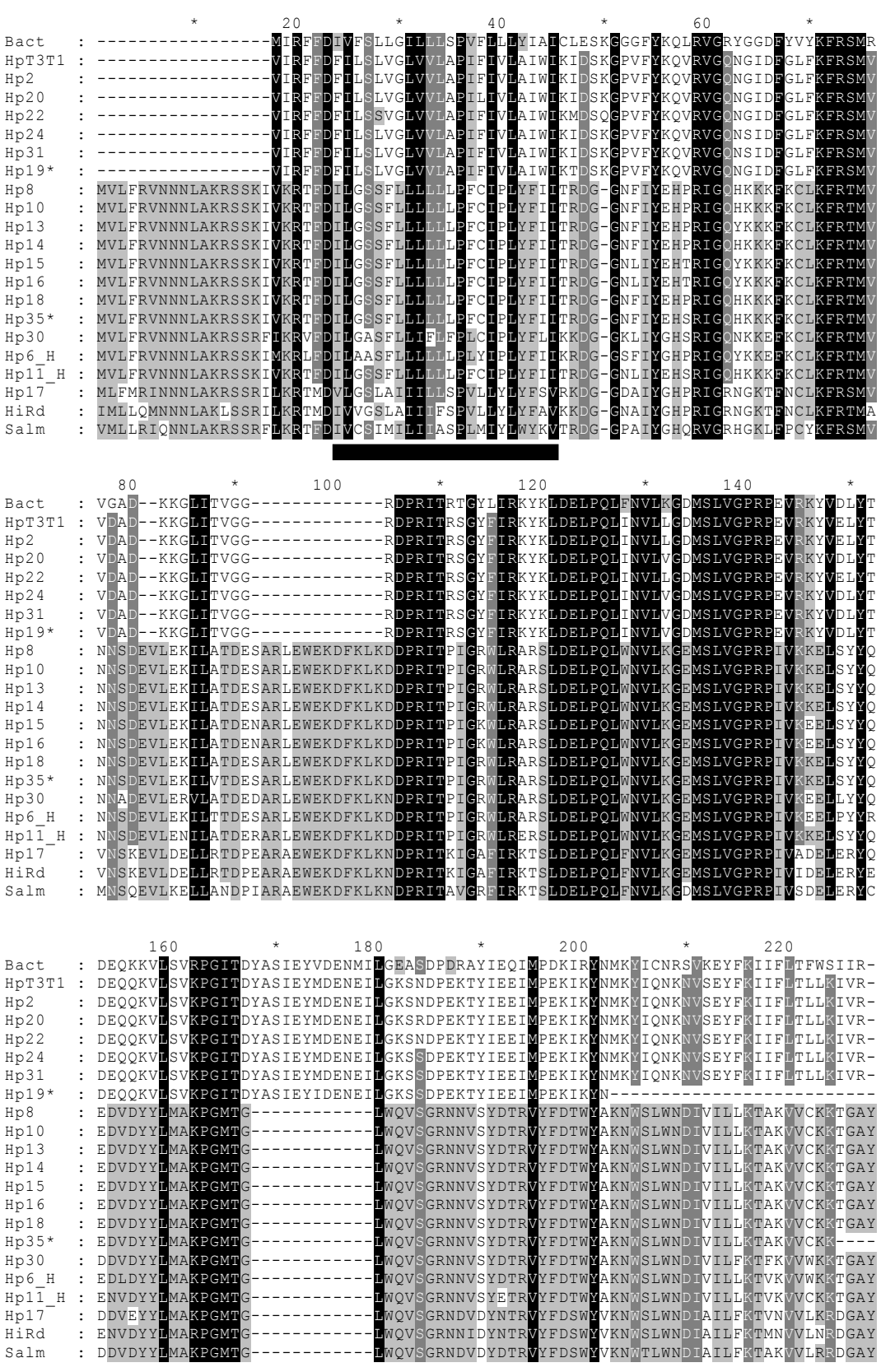
The first step of the OAg synthesis pathway is the transfer of a sugar from a nucleotide-sugar precursor (e.g. UDP-Gal or UDP-AAT) to an UndP lipid carrier molecule in the inner membrane, forming a structure upon which the rest of the O-unit or OAg can be assembled. Thus the substrate specificity of the UndP-sugar phosphotransferase that carries out this step determines the identity of the first sugar of the O-unit (for the Wzy-dependent system) or the adaptor sugar between inner core LPS and OAg (for the ABC transporter system, where the whole OAg is synthesised directly on one UndP-linked sugar). UndP-sugar phosphotransferase enzymes fall into two unrelated families: the polyisoprenyl-P N-acetylhexosamine-1-P transferases (PNPTs), of which the prototype is *E. coli* WecA, and the polyisoprenyl-P hexose-1-P transferases (and related enzymes), of which the prototype is *Salmonella enterica* WbaP (Lehrer et al., 2007).

The UndP-sugar phosphotransferase sequences encoded by the OAg loci of *Hp* strains T3T1 and 13 both fall into the second family and show sequence homology with *S. enterica* WbaP. Saldias

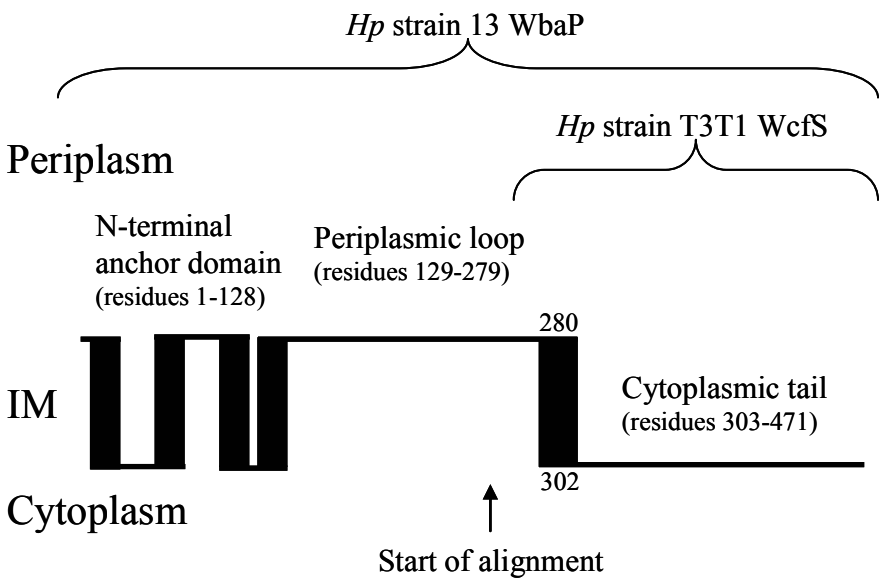
and colleagues (2008) characterised the domain structure and function of the *S. enterica* protein as an N-terminal domain putatively conferring stability, a central periplasmic loop involved in determining OAg chain length, and a C-terminal domain with independent Gal-P transferase activity. Alignment of the *S. enterica* WbaP sequence with the *Hp* alleles indicates that whilst the *Hp* strain 13 enzyme (13I) appears to include all three domains, the *Hp* strain T3T1 version (WcfS) is truncated and contains only the C-terminal domain. Wang and co-workers (1996) showed that in *S. enterica* both a full length protein and a 25 kDa C-terminus peptide are produced from *wbaP*, with the shorter peptide apparently being translated from a second start codon; the *Hp* T3T1 allele corresponds to this short peptide, beginning with a putative start codon (GTG) that aligns to a position 17 aa downstream of the second start codon of *S. enterica wbaP*. Therefore it is likely that the *Hp* T3T1 WcfS protein has UndP-sugar phosphotransferase activity but does not require stabilising by an N-terminal transmembrane region and does not regulate OAg chain length.

Since the C-terminal domain of WbaP (or the whole length of WcfS) is thought to determine the substrate specificity of the enzyme, obtaining the DNA sequence of this region for each *Hp* strain may help us to predict the first sugar of the relevant O-unit or the adaptor sugar between OAg and core LPS. The region was amplified by PCR and sequenced using primers 272c and 274b (or 274a) for group 1 strains and primers hmg13V and hmg13M (or hmg13G) for group 2 strains, and the translated sequences were aligned with those already known from OAg locus sequencing and with related sequences from other species (Fig. 5.3).

(a)


Bact
HpT3T1
Hp2
Hp20
Hp22
Hp24
Hp31
Hp19*
Hp8
Hp10
Hp13
Hp14
Hp15
Hp16
Hp18
Hp35*
Hp30
Hp6_H
Hp11_H
Hp17
HiRd
Salm

(b)


Hp strain 13 WbaP
Hp strain T3T1 WcfS
Periplasm
N-terminal
anchor domain
(residues 1-128)
Periplasmic loop
(residues 129-279)
280
IM
Cytoplasmic tail
(residues 303-471)
302
Cytoplasm
Start of alignment

(c)

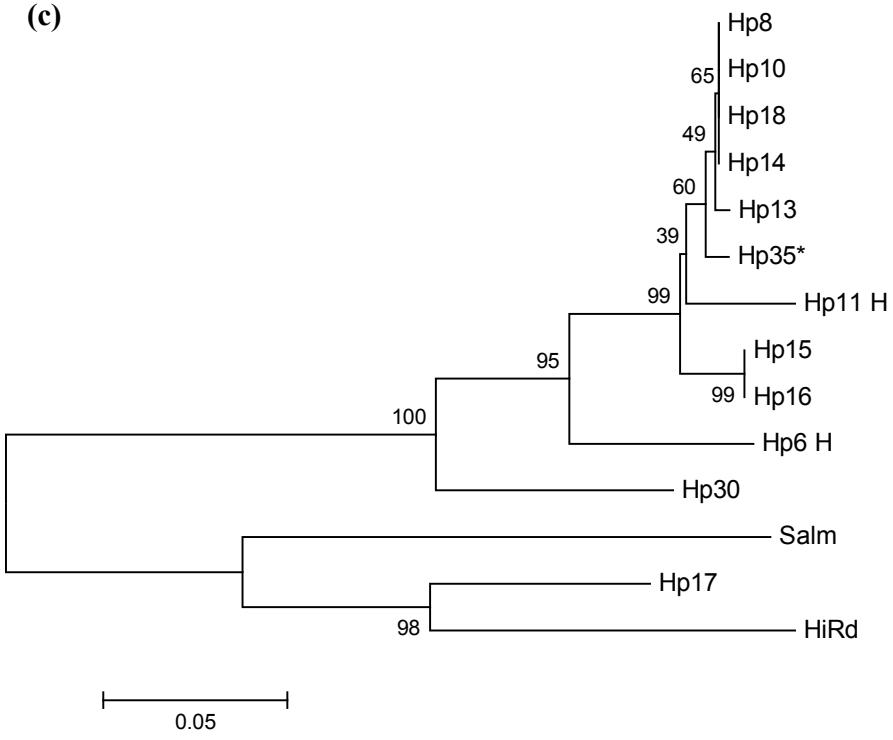

Hp8
65
Hp10
Hp18
49
Hp14
60
Hp13
39
Hp35*
99
Hp11 H
95
Hp15
99
Hp16
100
Hp6 H
Hp30
Salm
Hp17
98
HiRd
0.05

Figure 5.3. Analysis of WcfS and C-terminal WbaP amino acid sequences. The UndP-sugar phosphotransferase gene (or its 3' end) was amplified and sequenced from each study strain using group 1 or group 2 specific primers, except where the data was already available from OAg locus sequencing. Translated sequences were obtained for all study strains apart from *Hp* strain 34. H = hybrid strain; * = incomplete sequence. C-terminal WbaP sequences from *Hi* strain Rd and *Salmonella enterica* serovar Typhimurium strain LT2 (accession number CAA40130.1) are included for comparison and are labelled 'HiRd' and 'Salm' respectively: these are both UndP-Gal phosphotransferases. The *Bacteroides fragilis* strain YCH46 UndP-AAT phosphotransferase sequence (WcfS; accession number YP_098730.1) is labelled 'Bact'.

(a) ClustalW alignment. *Hp* strain numbers are given on the left. The horizontal black bar indicates the position (in both WbaP and WcfS) of the last transmembrane domain shown in panel (b). The top seven *Hp* strains in the alignment have WcfS-like UndP-sugar phosphotransferases: for these, the complete protein sequence is shown (approximately 195 aa), beginning with an alternative start codon (GTG) which normally encodes valine. The other strains have WbaP-like UndP-sugar phosphotransferases of which only the C-terminal region (215 aa) is shown, starting with the second start codon that is proposed for the *S. enterica* protein by Wang et al. (1996). Darker shading indicates a higher level of conservation between sequences.

(b) Predicted topology of *Hp* strain 13 WbaP. Drawn according to transmembrane predictions from the TMHMM algorithm. IM = inner membrane; the vertical black rectangles in this region represent transmembrane domains. Other parts of the protein are indicated by horizontal lines. Domain labels are as annotated for *S. enterica* WbaP by Saldias et al. (2008). The region covered by the *Hp* T3T1 WcfS homologue (i.e. the last transmembrane domain and the cytoplasmic tail) is indicated. The diagram is to scale, along the horizontal axis.

(c) Neighbour-joining tree of the C-terminal WbaP regions shown in the alignment. Positions containing missing data (i.e. the last 4 aa of the alignment) were eliminated. The percentage of replicate trees in which the associated sequences clustered together in the bootstrap test (500 replicates) are shown next to the branches. The scale bar indicates 0.05 amino acid substitutions per site. Analysis was conducted using MEGA5.

When compared to the *Hp* T3T1 WcfS amino acid sequence, the sequences from the other six group 1 strains (*Hp* 2, 19, 20, 22, 24 and 31) were 98-100% identical. This strongly suggests that they would add the same sugar, AAT, as the first component of the O-unit. Homologues of the two genes necessary for UDP-AAT synthesis, *Hp0272* and *Hp0274*, were detected in these strains earlier (Table 5.2).

WbaP from *Hi* strain Rd (encoded within the *hmg* locus) and *S. enterica* are both thought to add Gal as the first sugar of the O-unit. Twelve of the study strains have *wbaP* (rather than *wcfS*)

homologues: the 3' end of these genes all encode C-terminal domains with 65-67% aa identity to the same region of *Hi* Rd WbaP, apart from the *Hp* strain 17 sequence which shares 85% aa identity with the *Hi* Rd version (Fig. 5.3). The WbaP C-terminal domains of all group 2 *Hp* strains are 96-100% identical to each other and probably share the same specificity, suggesting that all of these isolates have the same adaptor sugar between OAg and inner core LPS. In *Hp* strain 13 this sugar is known to be a hexose (Fig. 5.2) and its identity is currently being investigated.

The *Hp* T3T1 gene *Hp1729* encodes a homologue of *E. coli* WecA (previously called Rfe), sharing 58% aa identity. In *E. coli* this enzyme is specific for GalNAc and GlcNAc, and it initiates the synthesis of the Enterobacterial common antigen (ECA) and of most Wzy-dependent or ABC transporter-dependent OAg (Alexander & Valvano, 1994; Meier-Dieter et al., 1992). Its role in *Hp* strain T3T1 (and in *Hi*, where it is also present) is unclear. However, the sequenced region of WbaP/WcfS does not include premature stop codons or frameshifts in any of the strains, indicating selection pressure to retain its function. This suggests that WecA is not the UndP-phosphotransferase responsible for OAg synthesis in *Hp*.

5.2 Further work on group 1 strains

5.2.1 Neuraminidase tests suggest that Neu5Ac is the last sugar in several group 1 OAgs

The *Hp* T3T1 OAg locus contains the genes required for the synthesis and transfer of sialic acid, and structural data shows that Neu5Ac is the terminal sugar of the *Hp* T3T1 O-unit. To determine if the O-units of other selected *Hp* strains also contain this sugar, protein-free cell

lysates were treated with neuraminidase from *Clostridium perfringens* and compared to untreated samples on a silver-stained tricine SDS-PAGE gel (Fig. 5.4). The LPS profiles of *Hp* strains 22, 24 and 31 gained a band just below the proposed O-unit band, suggesting that neuraminidase treatment results in a partial loss of the terminal sugar of the O-unit. This provides evidence that the terminal sugar of at least some *Hp* strain 22, 24 and 31 O-units is Neu5Ac.

There are several possible explanations for the partial retention of the original O-unit band:

(i) The neuraminidase treatment may not be 100% efficient;

(ii) Two different O-units may be present with different terminal sugars due to competing GTs, as is observed in *Hi* HMG;

(iii) The Neu5Ac on some LPS molecules may be modified, or attached via a different linkage, protecting against the action of neuraminidase. *C. perfringens* neuraminidase hydrolyzes terminal sialic acid attached through an α(2→3), α(2→6), or α(2→8) glycosidic linkage.

The LPS profiles of *Hp* strains 17, 19, 34 and T3T1 were unaffected by neuraminidase treatment. For *Hp* T3T1, structural analysis of the O-unit revealed that the terminal Neu5Ac is capped with an O-acetyl group: this may make it inaccessible to neuraminidase. For *Hp* strains 17, 19 and 34, four to five putative outer core GT genes were identified in Chapter 3 (Fig. 3.8), so the main LPS glycoform could contain outer core structures rather than an O-unit. An alternative explanation for the *Hp* strain 19 result would be as described above for *Hp* T3T1, as both share a putative O-acetyl transferase gene (*Hp0280*; Table 5.2). In summary, at least four *Hp* strains (T3T1, 22, 24 and 31) add sialic acid as the terminal sugar of the LPS O-unit.

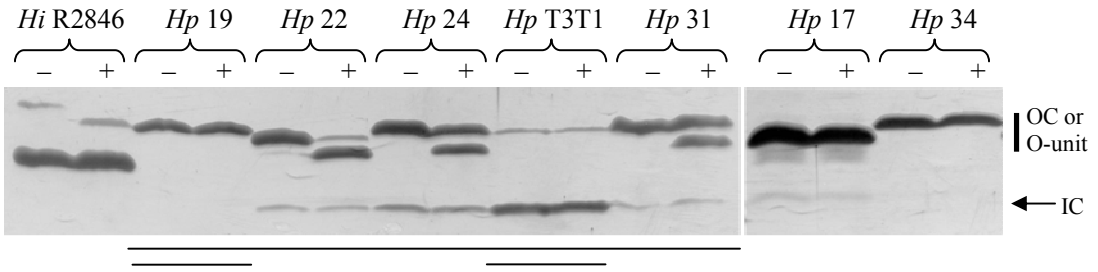


Figure 5.4. Tricine SDS-PAGE LPS profiles of selected *Hp* strains following neuraminidase treatment. Lysates were incubated with 1.5 µl of either neuraminidase (+) or PBS (-) at 37 °C for 24 h. Only the lower part of the gel is shown. *Hi* strain R2846 is included as a control: its highest molecular weight LPS glycoform contains HMG, which loses a sialic acid residue upon treatment with neuraminidase. *Hp* strains that are positive in PCR for *neuB* (sialic acid synthase gene) are underlined beneath their profile. Strains that are underlined twice are also positive in PCR for the putative OAc transferase gene *Hp0280*. IC = putative inner core LPS. OC or O-unit = LPS glycoforms proposed to include outer core substituents (OC; *Hp* strains 17, 19 and 34) or a single O-unit (*Hp* strains 19, 22, 24, T3T1 and 31).

5.2.2 Sequencing and analysis of the *Hp* strain 20 OAg locus

The O-unit of *Hp* 20 LPS appears to be a different size to that of *Hp* T3T1 when separated by electrophoresis (Fig. 4.6), and most of the *Hp* T3T1 OAg locus genes were not detected in *Hp* 20 by PCR. These findings indicated that the O-unit and OAg locus composition of at least some group 1 strains may differ from *Hp* T3T1, i.e. they may confer a different 'serotype'. To investigate this, the *Hp* 20 *glnA-pepB* LR-PCR product was cloned and sequenced using the same procedure as for *Hp* 13. Upon assembly, the DNA sequence was found to comprise ten intact genes spanning 11.7 kb. The products of seven of these ORFs share 69-99% aa identity with proteins encoded by the *Hp* T3T1 OAg locus (shown by the intensity of shading in Table 5.5), but the remaining three share no significant homology with *Hp* T3T1 proteins.

BLASTP analysis (Table 5.5) suggests that the first three genes of the *Hp* 20 OAg locus encode Wzz, Wzx and Wzy, the enzymes needed for Wzy-dependent OAg assembly. Transmembrane domain predictions made using the TMHMM algorithm are consistent with these putative enzyme functions. The predicted *Hp* 20 chain length determining protein, 20A, shares 83% aa identity with that of *Hp* T3T1. This high level of homology reflects the similar OAg modal chain length of the two species, which is approximately 17-20 repeat units when visualised on a heavily loaded LPS gel. However, the predicted Wzx (20B) and Wzy (20C) proteins share so little identity with the *Hp* T3T1 alleles that they cannot be aligned; they are more closely related to enzymes from *Bacteroides*. Since the closest known homologues of the AAT-related genes found in both *Hp* T3T1 and *Hp* 20 are also from this genus (Table 4.1) it is tempting to speculate that they may have been acquired together with 20B and 20C, and that later horizontal transfer events have altered the Wzx and Wzy alleles in *Hp* T3T1.

In the *Hp* T3T1 genome, the three genes of the Wzy-dependent OAg assembly system are interspersed with the six genes related to the synthesis of sialic acid and its transfer onto the OAg; these six ORFs are completely absent in the *Hp* 20 locus, hence the shorter LR-PCR product. Given this information, it seems unlikely that the *Hp* 20 OAg contains sialic acid.

In addition to the predicted UndP-AAT phosphotransferase (20G), which would determine the first sugar of the O-unit to be AAT, the *Hp* 20 OAg locus encodes two other transferases (20D and 20E). We can therefore predict that the O-unit of this strain probably comprises three sugars. The closest homologue of GT 20E in the NCBI database is Hp0275 from *Hp* strain T3T1, but since they share only 69% aa identity they may not necessarily transfer the same sugar.

Table 5.5. BLASTP analysis of the ten ORFs in the *Hp* strain 20 O-antigen locus. Gene designations are given in the left hand column. The second column lists the putative size of the encoded protein in amino acids (aa). The species containing the highest scoring BLASTP matches are stated for each gene; % aa identity (%aa ID) is given, with % query coverage (QC) in brackets. For seven of the ORFs, the highest scoring matches were to OAg locus genes from *Hp* strain T3T1, and in these cases the degree of homology is represented by the intensity of the shading. The putative function of each *Hp* strain 20 protein is based on the known functions of homologues and on domain model searches (see text).

			Highest scoring BLASTP matches	
Gene	**aa**	**Putative function**	**Species (protein)**	**%aa ID (QC)**
20A	380	OAg chain length determinant (Wzz)	*Hp* T3T1 (Hp0285)	84 (98)
20B	415	OAg flippase (Wzx)	*Bacteroides uniformis* (BACUNI_01900)	34 (98)
			Caminibacter mediatlanticus	33 (92)
			Psychroflexus torquis	35 (90)
			[*Hp* strain 17 (17C)]	22 (86)
20C	455	OAg polymerase (Wzy)	*Bacteroides uniformis* (BACUNI_01904)	33 (82)
			Sebaldella termitidis	29 (89)
			Leptotrichia goodfellowii	28 (82)
20D	343	GT (WcwK homologue)	*Haemophilus parasuis* (HAPS_0044)	56 (99)
			Escherichia coli serogroup O152	37 (99)
			Acinetobacter baumannii	36 (98)
20E	270	GT (GTA superfamily)	*Hp* T3T1 (Hp0275)	69 (99)
			Haemophilus parasuis (HAPS_0045)	57 (99)
			Vibrio alginolyticus	40 (99)
			Sulfurospirillum deleyianum	46 (80)
20F	403	Aminotransferase (AAT synthesis)	*Hp* T3T1 (Hp0274)	99 (100)
20G	195	UndP-AAT phosphotransferase (WcfS)	*Hp* T3T1 (Hp0273)	98 (100)
20H	631	UDP-GlcNAc dehydratase (AAT synthesis)	*Hp* T3T1 (Hp0272)	96 (100)
20I	418	OAg ligase (WaaL)	*Hp* T3T1 (Hp0271)	77 (100)
20J	337	dTDP-glucose 4,6-dehydratase	*Hp* T3T1 (Hp0270)	94 (100)

Protein 20D shares 37% aa identity with WfgC encoded by the OAg loci of *E. coli* serogroup O152 and *Shigella dysenteriae* group 12. These two OAg contain the moiety α-Glc*p*NAc-(1-P-6)-α-Glc*p* (Liu et al., 2008; Olsson et al., 2005), and WfgC is predicted to catalyse the unusual phosphodiester linkage between the two sugars (Lundborg et al., 2010).

In 2005, Sperisen and colleagues identified a novel family of proteins that was conserved across most eukaryotes and some prokaryotes (Sperisen et al., 2005). They termed it 'Stealth' because the bacterial members of the family appeared to be involved in immune evasion, and hypothesised that the proteins were hexose-1-phosphoryl transferases. When compared to the Stealth alignment published by Sperisen et al., it becomes clear that protein 20D belongs to this family. The conserved regions (CR) which define Stealth are present at the following positions within the 20D sequence: CR1 = aa 5-16; CR2 = aa 40-139; CR3 = aa 223-271; CR4 = aa 308-343. Together with its similarity to WfgC, this information strengthens the case for gene *20D* encoding a phosphoryl transferase (i.e. an enzyme that connects two sugars via a phosphodiester linkage).

Two *Haemophilus parasuis* genomes (strains 29755 and SH0165) also contain homologues of gene *20D* within 'LPS-related' gene clusters; the surface polysaccharides of this species have not been studied in detail but perhaps either the capsule or OAg contains a phosphodiester group. Likewise, the serotype-specific region of the *cap* locus in *Hi* serotype f contains a gene encoding a Stealth protein (*fcs1*; accession no. Q714U9) (Satola et al., 2003); this is consistent with the capsule f repeat unit structure which is known to contain a phosphodiester linkage (Egan et al.,

1980). However, the *Hp* strain 20 and *Hi* Stealth proteins are highly divergent and are likely to have been acquired independently.

Compared to the *Hp* T3T1 OAg locus, the *Hp* strain 20 version encodes the enzymes necessary to synthesise UDP-AAT and to add it to UndP as the first sugar, but lacks any sialic acid related genes. It encodes two other GTs; one shares homology with an *Hp* T3T1 OAg GT but the other is unique and is likely to generate a phosphodiester linkage. This genetic data correlates well with the results of structural analysis of purified *Hp* 20 LPS (Fig. 5.5; Elke Schweda, pers. comm.): the trisaccharide O-unit comprises AAT, glucose phosphate and GalNAc.

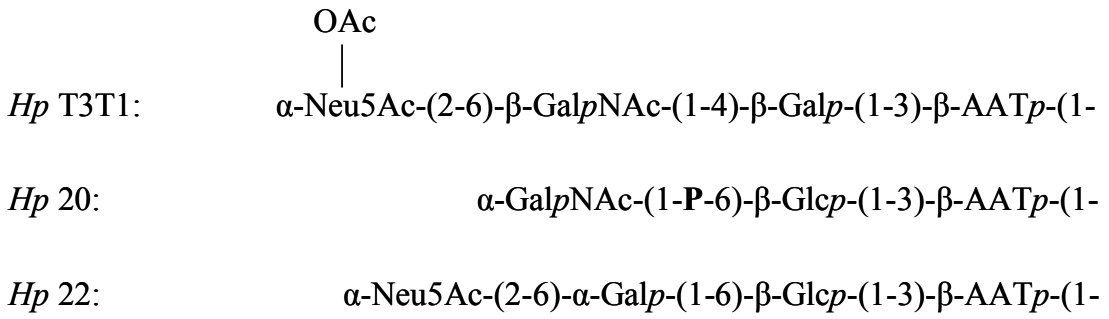


Figure 5.5. O-unit structures of three *Hp* strains that contain a group 1 OAg locus. Each structure has been determined by mass spectrometry and NMR studies, and is attached to the conserved inner core LPS via a β1,4 linkage between AAT and the Glc on HepI (Elke Schweda, pers. comm.). For *Hp* strains T3T1 and 22, only glycoforms containing no or one O-unit were detected by chemical methods, but for *Hp* strain 20, some glycoforms were found to contain multiple polymerised O-units (i.e. true OAg). **P** = phosphodiester linkage; AAT = 2-acetamido-4-amino-2,4,6-trideoxy-D-galactose.

The lack of Neu5Ac in the OAg explains the need for a different Wzy (OAg polymerase) sequence to that of *Hp* strain T3T1, as the acceptor sugar specificity of the enzyme will be altered. However, as both Neu5Ac and phosphate groups are generally anionic in character, the

Hp T3T1 and *Hp* 20 O-units may have similar overall charge profiles. The remaining LPS structure of *Hp* strain 20 comprises the same inner core as defined in Section 3.4, plus a single outer core substituent, Gal (as shown in Fig. 3.8), which is presumably added by the LpsB homologue identified in Section 3.2.3.

The LPS of another group 1 strain, *Hp* 22, was also structurally characterised (Elke Schweda, pers. comm.). The inner core LPS structure was found to be identical to that of *Hp* T3T1 (Fig. 3.8), with no outer core components. This was to be expected, as no outer core transferases were identified for this strain in Chapter 3. The *Hp* strain 22 LPS was decorated with a single four-sugar O-unit comprising AAT, Glc, Gal and Neu5Ac (Fig. 5.5). This correlates with the presence of *Hp* T3T1-like AAT and Neu5Ac related genes and of the proposed *Hp* strain 20 glucosyltransferase gene, *20E*, by PCR analysis (Table 5.2). Based on the O-unit structure, we would predict the *Hp* strain 22 OAg locus to include a galactosyltransferase gene that differs from the GT genes seen in *Hp* strains T3T1 and 20.

5.3 Identification of an OAg ligase gene

The transfer of OAg from the UndP carrier to core LPS is carried out by an OAg ligase. Whilst all seven group 1 *Hp* OAg loci include a putative OAg ligase gene (Table 5.2), the group 2 OAg locus of *Hp* strain 13 and the ungrouped OAg loci of *Hp* strains 17 and 30 do not (Section 5.1). We therefore investigated whether any of the unsequenced *Hp* OAg loci contain an OAg ligase gene in a particular position and whether *Hp* 13 and other strains encode an OAg ligase elsewhere in the genome.

5.3.1 The *Hp* study strains do not contain an OAg ligase gene between *rmlB* and *pepB*

In the *Hi hmg* locus, the ORF encoding the HMG ligase is at the distal end between *HI0873* (annotated as an *rmlB* homologue but of unknown function) and the locus-flanking gene *pepB* and is convergent in orientation with the rest of the locus (Hood et al., 2004b). Analysis of some *pepB-glnA* regions extracted from the Crook group sequencing project indicated that a few unidentified *Haemophilus* isolates contain both *Hp*-like OAg genes and a *Hi*-like ligase in its normal *Hi* position and orientation, so it was considered worthwhile to screen the *Hp* study strains for similar gene arrangements.

PCR amplification using primers designed to the *Hp rmlB* and *pepB* genes was attempted for all study strains and for *Hi* Rd. PCR products containing the HMG/OAg ligase gene were expected to be 2240 bp in length (e.g. from *Hi* Rd), whereas those without a gene in this position should be 850 bp (e.g. from *Hp* strains T3T1 and 13). Sixteen of the 20 study strains yielded 850 bp products (Fig. 5.6), ruling out the possibility that their OAg are ligated to the LPS by an enzyme encoded by the last gene of the OAg locus. *Hp* strain 30 did not give a PCR product because its *rmlB* gene is at the other end of the OAg locus and is divergent in sequence (Table 5.4); we already know that this locus lacks an OAg ligase. The *Hp* 34 OAg locus is likely to be similar (Section 5.1.4). The other two PCR-negative strains were the hybrids *Hp* 6 and 11, which we know contain both *rmlB* and *pepB* genes. Perhaps *pepB* is not one of the OAg locus flanking genes in *Hi*/*Hp* hybrid strains. Overall, the presence of a ligase gene between *rmlB* and *pepB* appears to be a *Hi*-specific gene arrangement and is not seen in *Hp*.

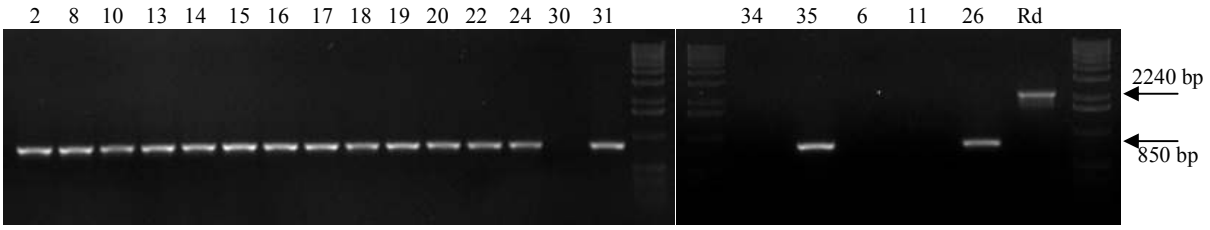


Figure 5.6. PCR amplification using primers hmg13W and 269a to test for the presence of an OAg ligase gene between the *rmlB* and *pepB* genes. *Hp* strain numbers are listed above each lane. Unmarked lanes contain 1 kb ladder. 26 = *Hp* strain T3T1 (*rmlB* and *pepB* are consecutive, giving an 850 bp product); Rd = *Hi* strain Rd (the HMG ligase gene is present between *rmlB* and *pepB*, giving a 2.2 kb product). None of the 16 *Hp* strains that give a PCR product appears to have a gene between *rmlB* and *pepB*.

5.3.2 Thirteen strains encode a putative OAg ligase between *fba* and *orfH*

If group 2 OAg loci do not include OAg ligase (*waaL*) genes, how do strains such as *Hp* 13 attach their OAg to the LPS core? In *E. coli*, genes associated with OAg are occasionally found outside the OAg locus, and this may also be the case with *Hp*. The *Hp* strain 13 genome has not been sequenced, but limited sequence data is available via Genomophilus for *Hp* strain 15, which produces an OAg of similar appearance and molecular weight and has a similar OAg locus to *Hp* strain 13. The *Hp* strain 15 contigs were searched for a homologue to the *Hp* T3T1 OAg ligase, Hp0271, using BLASTP. A significant match was identified at the beginning of contig 243: it encoded an amino acid sequence with 41% identity to the last 280 aa of Hp0271, and was followed by a homologue of gene *Hp0841* (*orfH*). No *Hp* 15 contig containing the 5' end of this potential OAg ligase gene could be identified using BLASTP searches.

If the gene order is conserved between *Hp* strains T3T1 and 15, we would expect the 5' end of the potential *waaL* gene to be flanked by a homologue of ORF *Hp0842* (fructose bis-phosphate aldolase gene, *fba*). Primers designed against *Hp0841* and *Hp0842* sequences amplified a 1.6 kb

product from *Hp* 15 gDNA in PCR, indicating that there was indeed an extra gene at this position compared to *Hp* T3T1, and by using these primers in combination with internal *Hp* 15 *waaL* primers it was confirmed as the putative ligase gene.

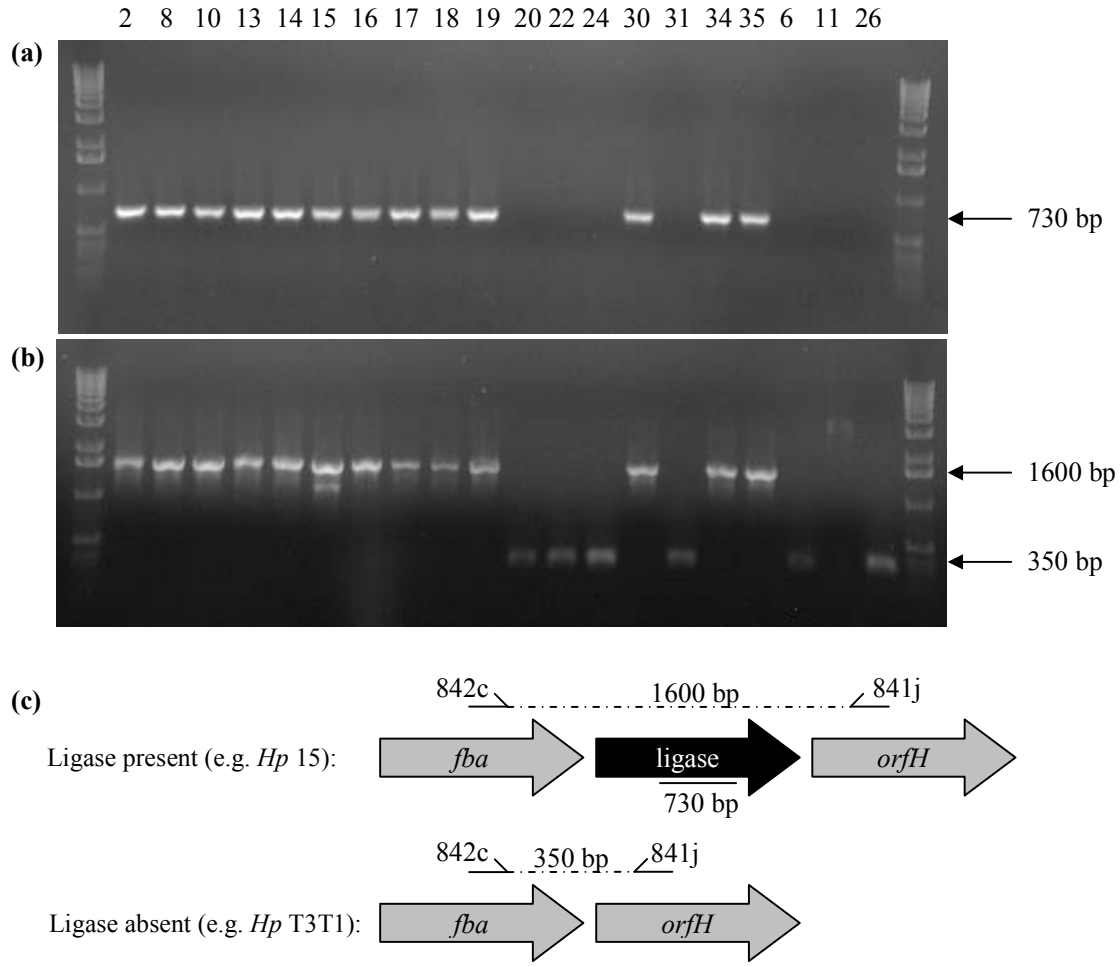


Figure 5.7. PCR amplification of the *Hp* 15-type OAg ligase. *Hp* strain numbers are listed above each lane. External lanes contain 1 kb ladder. **(a)** Internal primers LIG15B and LIG15C were used to test for the amplification of a 730 bp fragment of the ligase gene from each study strain. 13 of the 20 strains were positive. **(b)** Flanking primers 842c and 841j were used to amplify the ligase locus from each study strain. Products were either 1600 bp (consistent with the presence of a ligase gene at this locus) or 350 bp (consistent with *fba* and *orfH* being contiguous, as in the *Hp* T3T1 genome). Hybrid strain *Hp* 11 gave a weak 4 kb product which was not investigated further. **(c)** Representation of the *Hp* 15 ligase locus and the equivalent region in *Hp* T3T1. Not drawn to scale.

The OAg loci of *Hp* strains 13, 17 and 30 all lack recognisable OAg ligase genes (Section 5.1), so PCR analysis was performed to test whether these and other *Hp* study strains contain an *Hp* 15-like *waaL* gene instead. Using internal *Hp* 15 *waaL* primers LIG15B and LIG15C, a 730 bp product was amplified from 13 of the 20 study strains (Fig. 5.7a). These comprised all eight *Hp* strains with group 2 OAg loci, the three ungrouped strains (*Hp* 17, 30 and 34), and two strains with group 1 OAg loci (*Hp* 2 and *Hp* 19). PCR products obtained using the primers designed to *fba* and *orfH* were consistent with the presence of an intervening gene (i.e. *waaL*) at this locus in the same 13 strains (Fig. 5.7b and c). Using various primer combinations for PCR and DNA sequencing reactions, full length *Hp* 15-like *waaL* DNA sequences were obtained for *Hp* strains 13 and 19. When translated, these sequences share more homology with WaaL from *Aggregatibacter aphrophilus* (57% aa identity) than with the *Hi* Rd HMG ligase, HI0874 (42%) or the *Hp* T3T1 OAg ligase, Hp0271 (34%) (Fig. 5.8).

5.3.3 *Hp* 15 ligase mutants lack detectable OAg

To test the function of the putative WaaL protein, plasmids were constructed in which the *Hp* 15-like *waaL* was interrupted by a erythromycin resistance cassette. The *waaL* region for this purpose was amplified from either *Hp* strain 13 (for construct pRY45) or *Hp* strain 19 (for construct pRY46). The gene was successfully mutated in *Hp* strain 19, but not strains 13 or 15, using these plasmid constructs with static aerobic transformation. Chromosomal DNA from one of the *Hp* 19 *waaL* mutant clones (*Hp* 19.19) was subsequently used to transform *Hp* strains 13 and 15.

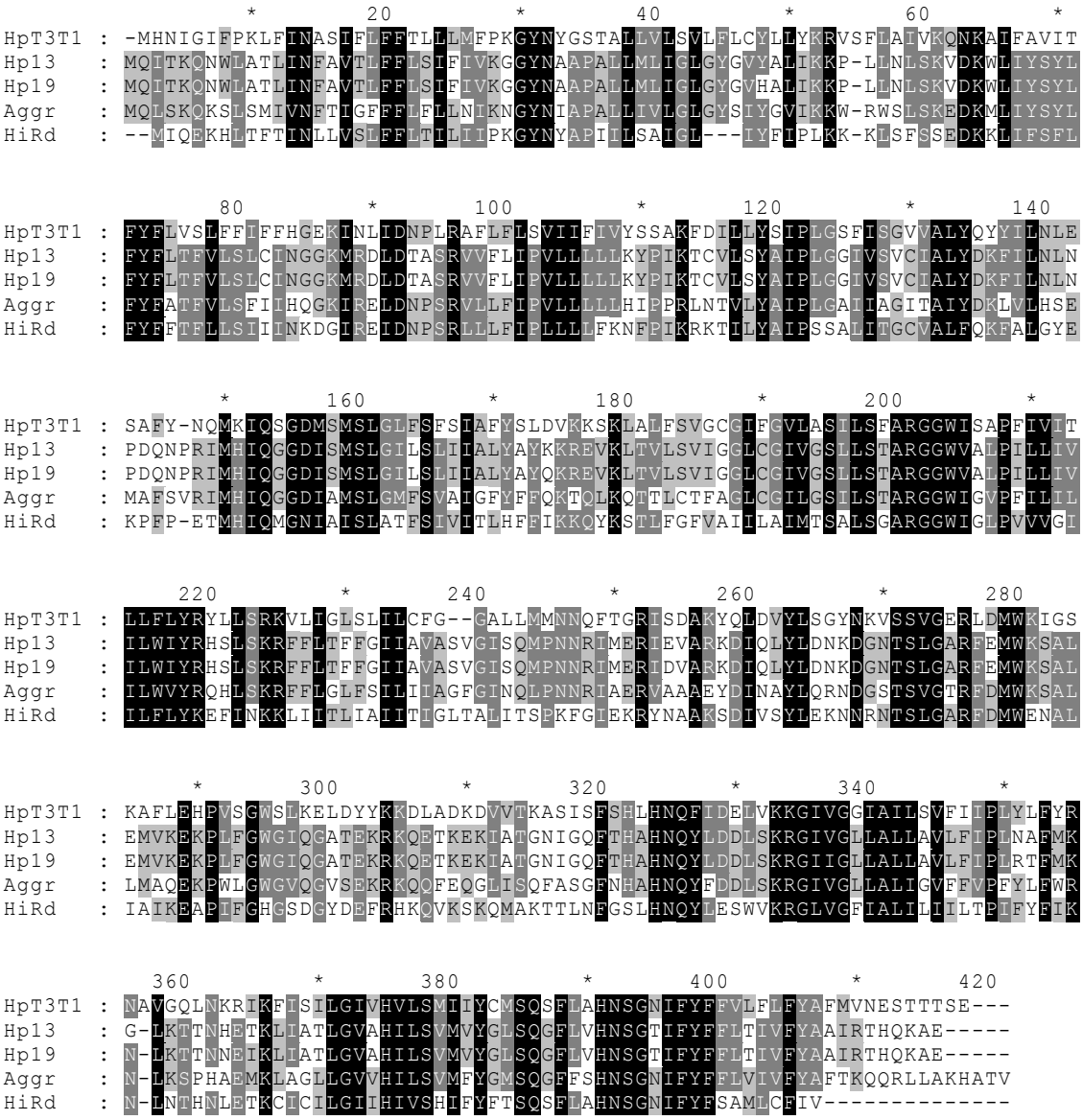


Figure 5.8. Amino acid alignment of OAg/HMG ligases. HpT3T1 = *Hp* strain T3T1 protein Hp0271; Hp13 = *Hp* strain 13 OAg ligase homologue (*Hp* 15-like); Hp19 = *Hp* strain 19 OAg ligase homologue (*Hp* 15-like); Aggr = *Aggregatibacter aphrophilus* strain NJ8700 protein NT05HA_1726; HiRd = *Hi* strain Rd protein HI0874. Darker shading indicates a higher level of sequence conservation.

WaaL mutant transformants were obtained for *Hp* strain 15 but not for strain 13. In Fig. 5.9, the LPS profiles of three of the *Hp* 15 *waaL* mutants are compared to that of the parental strain: they do not appear to synthesise any OAg-containing glycoforms, consistent with the hypothesis that

the gene identified in Section 5.3.2 is the ligase responsible for the addition of the OAg to the LPS core in this strain.

Whilst it is reassuring to identify a putative OAg ligase gene in the *Hp* group 2 and ungrouped strains, its presence in *Hp* strains 2 and 19 was unexpected as these strains are already thought to include genes related to *Hp* strain 20 *waaL* in their OAg loci (PCR analysis, Table 5.2). A possible scenario is that these two lineages recently exchanged a group 2 locus for a new group 1 locus, acquiring an extra OAg ligase gene in the process, and have not yet lost the original ligase gene. Alternatively, one might be a pseudogene, or the two ligases might act upon different donor (polysaccharide) and/or acceptor (LPS/protein) molecules. Since we have not been able to transform *Hp* strain 2, and the analysis of *Hp* strain 19 LPS is complicated (see Section 6.1.5), it is difficult to test these hypotheses.

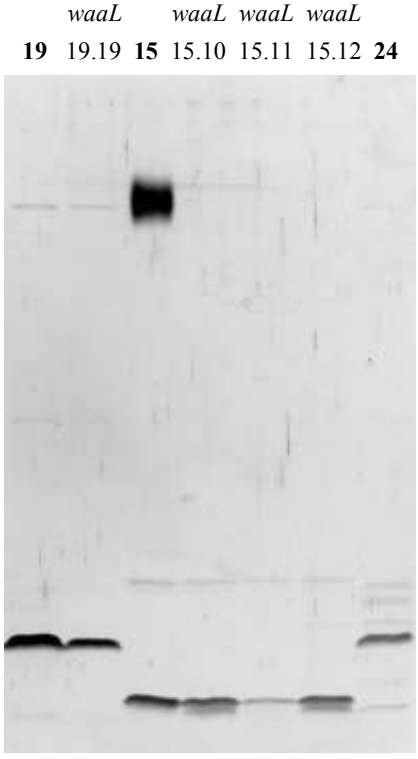


Figure 5.9. LPS profiles of *Hp* OAg ligase mutants. Proteinase K treated cell lysates were fractionated by tricine SDS-PAGE and stained with silver. Numbers above each lane are the *Hp* wild type strain (in bold) or the mutant clone number. Note that the putative OAg ligase gene interrupted in the mutants is outside of the OAg locus; *Hp* strain 19 has another OAg ligase gene within the OAg locus. *Hp* strain 24 is shown for reference.

5.4 Discussion

The results presented in this chapter demonstrate that all of the 18 true *Hp* strains included in the study contain a cluster of genes related to polysaccharide synthesis and transport flanked by *glnA* and *pepB*. This provides strong evidence that there is positive selection in *Hp* for the ability to synthesise a polysaccharide structure on an UndP carrier on the cytoplasmic face of the inner membrane and transfer it to the periplasmic face. For some, but not all, of the *Hp* strains, we have genetic and structural evidence that this polysaccharide is a true OAg, i.e. a repeated chain of O-units attached to the LPS core.

The five *Hp* OAg loci that have now been sequenced are illustrated in Fig. 5.10 for the purpose of comparison. Each comprises 10-16 ORFs, all in the same orientation. There is very little intergenic DNA, suggesting that in general the genes at each locus form an operon in which the genes are co-regulated and co-transcribed. There is no strict gene order, but the transport genes are usually found towards the 5' end and the *wbaP*/*wcfS* homologue is typically towards the 3' end. The sugars predicted or proven to form part of these *Hp* OAg structures include Gal*p*, Gal*f*, Gal*p*NAc, Glc*p*NAc, AAT*p*, Neu5Ac, ribitol and deoxytalose, with acetyl and PEtn additions. Six of these ten components are also found in the outer core LPS of various *Hi* strains. As predicted from analysis of *lgtF* mutants in Section 4.4, the *Hp* OAgs that have been structurally determined are linked to the Glc moiety on HepI of the core LPS.

The OAg loci of the 18 true *Hp* strains fall into two main categories based on their assembly/transport system and UndP-sugar phosphotransferase genes. In general, a *wcfS*-like initial transferase gene and at least one of the genes required for Wzy-dependent OAg assembly

were identified in group 1 loci, whereas group 2 loci encode an ABC transporter and a WbaP-like initial transferase. However, the transferases responsible for subsequent sugar additions and modifications seem to be much more variable: the OAg loci of *Hp* strains T3T1 and 20 (both group 1) encode different GTs, and PCR amplification using primers specific for the *Hp* strain 13 GTs was negative in many other group 2 strains. As GT genes are generally found in the central region of the locus, several could be replaced in one recombination event from the bacterial gene pool to create diversity within each group.

To estimate the number of potential *Hp* serogroups (different OAg structures), we can categorise the study strains according to their PCR results as follows:

- Group 1 OAg that are likely to include Neu5Ac: *Hp* strains 19, 22, 24, 26 (= T3T1) and 31
- Group 1 OAg that are not likely to include Neu5Ac: *Hp* strains 2 and 20
- Group 2 OAg that use *Hp* 13-like GTs: *Hp* strains 13, 15 and 18
- Group 2 OAg that use GTs unlike those of *Hp* 13: *Hp* strains 8, 10, 14, 16 and 35
- Ungrouped OAg, possibly containing ribitol and Gal*f*: *Hp* strain 17
- Ungrouped OAg, possibly containing deoxytalose: *Hp* strains 30 and 34.

Therefore, assuming that all of the OAg loci are functional, there would be at least six *Hp* serogroups represented just within the 18 study strains. In reality, the total number of *Hp* serogroups is likely to be much higher. If an individual host developed an adaptive immune response to the OAg of one *Hp* serogroup and eliminated it from the nasopharynx, the other serogroups could still avoid immune recognition. The diversity of *Hp* OAg structures contrasts

strongly with the *Hi* HMG unit, for which the biosynthetic enzymes are also encoded between *glnA* and *pepB*: 60% of NTHi strains synthesise the same two HMG structures, and the remaining 40% do not have the *hmg* locus (Hood et al., 2004b).

All group 2 and ungrouped *Hp* strains were found to encode a putative OAg ligase outside the OAg locus, between the genes *fba* and *orfH*. This was shown to be required for the synthesis of OAg-containing LPS glycoforms in *Hp* strain 15. A homologue of this ligase gene is also present next to *fba* (but not *orfH*) in the genome strains of several other Pasteurellaceae, including species of the genera *Actinobacillus*, *Aggregatibacter*, *Pasteurella* and *Mannheimia*, suggesting that it is either ancestral or has spread through the family very successfully by lateral gene transfer. When Tang and Mintz (2010) interrupted the ligase gene next to *fba* in *A. actinomycetemcomitans*, they observed not only the loss of OAg from the LPS profile but also a shift in the electrophoretic mobility of the adhesin EmaA, which is usually glycosylated with what is thought to be the same 'OAg'. It is possible that the *Hp* homologue is also bifunctional and could glycosylate LPS and certain outer membrane proteins with the same polysaccharide as each other.

In some species, particular OAg serogroups are associated with increased virulence: for example, most cases of Legionnaire's disease are caused by serogroup 1 strains of *Legionella pneumophila*, which are spread across diverse lineages but share the same OAg genes (Cazalet et al., 2008). Although *Hp* is rarely implicated in disease there could still be a correlation between OAg structure and pathogenicity. If suitable strains became available, it would be interesting to

test whether all *Hp* disease isolates (or those from a particular disease situation) are associated with group 1, group 2 or ungrouped OAg loci or with certain OAg components such as Neu5Ac.

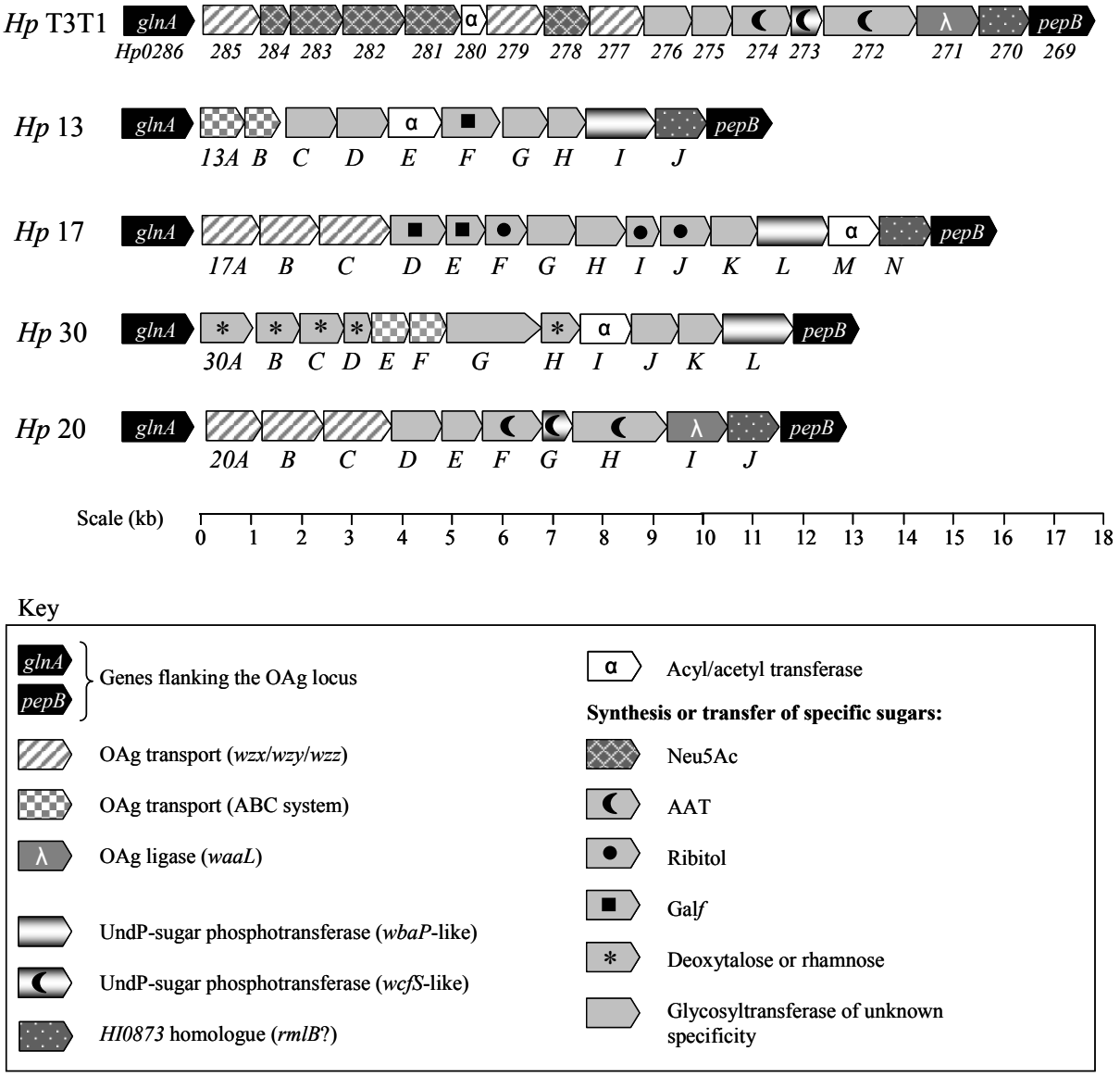


Figure 5.10. Organisation of the five sequenced *Hp* OAg loci. The *Hp* strain name is given to the left of each diagram. Each arrow represents an ORF, and its predicted function (as discussed in the text) is indicated by its shading and pattern as shown in the key. Each locus putatively encodes one of two transport systems (Wzy-dependent or an ABC transporter), an UndP-sugar phosphotransferase, and a series of enzymes for the synthesis and transfer of various OAg components. Drawn to the scale indicated.

6 Results: Interactions between *H. parainfluenzae* and the host environment

We have established in earlier chapters that *Hp* LPS lacks most of the outer core components that are found in *Hi*, but that all of the *Hp* study strains encode a potential OAg locus. To investigate whether the observed differences in LPS structure influence the pathogenicity of these two closely related species, several in vitro assays that are proxies for aspects of host interaction were performed using *Hp* LPS mutants in addition to wild type *Hi* and *Hp* strains. In the serum bactericidal assays, pooled human serum was used as a source of complement to test whether OAg can protect *Hp* from complement-mediated killing. This mechanism of host defence can be activated by the direct binding of complement component C3b to the bacterial cell surface (alternative pathway), via recognition of specific cell surface molecules by antibodies (classical pathway) or via the binding of lectins to surface carbohydrates. All three pathways lead to the formation of the C3 convertase, activating the complement cascade. In the host, this can result in (a) the production of pro-inflammatory molecules such as C3a and C5a, (b) opsonisation of bacteria by C3b followed by phagocytosis, and/or (c) insertion of the membrane attack complex (MAC, comprising C5b and C6-C9) and subsequent bacterial lysis.

Serum isolated from human blood contains complement components and some antibodies, so the serum bactericidal assay can detect both classical and alternative complement activation. It primarily measures killing that is mediated by MAC formation, because serum does not contain the macrophages and neutrophils that are required for opsonophagocytosis. Although the

complement system is found primarily in serum, complement components have also been identified in respiratory mucosa, especially during inflammation (Hallstrom & Riesbeck, 2010), so the ability of *Hp* to resist killing by complement is relevant to its commensal behaviour as well as its rare association with disease.

Genetic analysis suggests that *Hp* does not elaborate its LPS with the biologically important molecule phosphocholine (PCho), at least not using the same enzymes as *Hi* (Section 3.3.3). In the current chapter we present further evidence for the lack of PCho on the *Hp* cell surface and investigate whether this species is capable of expressing PCho-containing LPS upon the introduction of a transgenic *Hi lic1* locus.

Bacteria in the upper respiratory tract interact with a complex array of different host cells, mucins, immune molecules and other microbes. We finish this chapter by examining two such interactions: the association of *Hp* with epithelial cells, and competition between *Haemophilus* strains by the release of bactericidal molecules.

6.1 Examination of the role of LPS structure in determining serum resistance

6.1.1 *Hp* generally displays higher serum resistance than *Hi* Rd

Ten representative *Hp* strains were compared to *Hi* regarding their ability to survive in various concentrations of pooled human serum for 1 h (Fig. 6.1). After the incubation period, surviving bacteria were plated onto BHI agar, and colony counts were performed the following day. Typically, non-encapsulated *Hi* strains show less than 5% survival in 10% serum (Fig. 6.1). In

contrast, six of the ten *Hp* strains showed at least 85% survival in 10% serum, a similar level of resistance to that of encapsulated *Hi* strains (e.g. the type b strain RM7004). All three of the *Hp* strains with a group 2 OAg locus (*Hp* 8, 10 and 13) were serum resistant. Of the five strains with a group 1 OAg locus, the two with the greatest serum resistance were those that synthesised high levels of polymeric OAg (*Hp* strains 20 and T3T1).

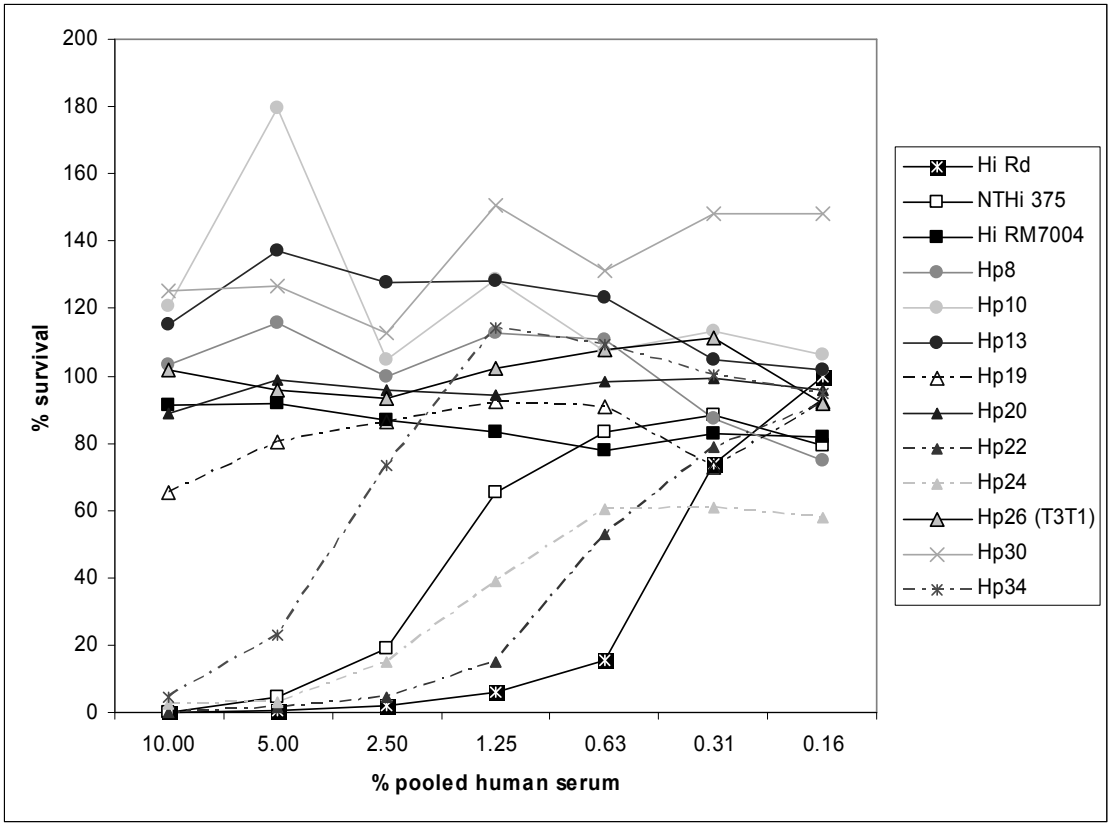


Figure 6.1. Resistance of wild type *Hi* and *Hp* strains to the killing effect of pooled human serum. Results are shown as the survival of inoculating bacteria as a percentage of survival in a 10% decomplemented serum control well. Ten representative *Hp* study strains are denoted by triangles (group 1 OAg locus), circles (group 2 OAg locus) or crosses (deoxytalose type OAg locus); *Hp* strains for which no polymeric OAg has been visualised are shown with dotted lines. Three representative *Hi* strains are shown by squares for comparison: *Hi* Rd (no capsule but derived from serotype d), *Hi* 375 (non-typeable) and *Hi* RM7004 (serotype b encapsulated).

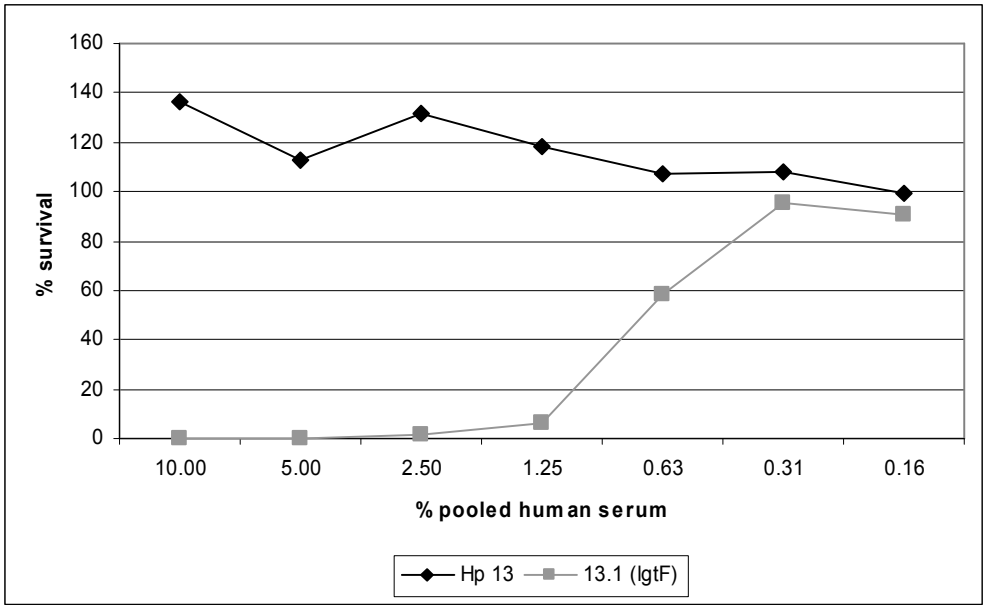
A similar pattern was seen for the two ungrouped strains with putative deoxytalose synthesis genes: *Hp* strain 30, which appears to produce polymeric OAg in Fig. 4.6, shows 100% survival in 10% serum whilst the equivalent survival rate for *Hp* 34 was only 4%. However, to assess the true impact of particular LPS structural features on the biological characteristics of *Hp*, it is important to compare isogenic strains rather than to rely solely on natural isolates which will differ in their genetic backgrounds and in their expression patterns of multiple cell surface components.

6.1.2 Interruption of *lgtF* lowers the serum resistance of several *Hp* strains

The first *Hp* mutants to be made during this project were the *lgtF* mutants in seven different strain backgrounds (see Fig. 4.7 for LPS profiles). The parental strains comprised three with a group 1 OAg locus, three with a group 2 OAg locus, and one ungrouped strain (*Hp* 17). The serum resistance of each group 1 or group 2 wild type strain and at least one of its *lgtF* mutants was assessed using a serum bactericidal assay. Whilst bacterial survival in 10% serum was above 60% for the six wild type strains, all of the *lgtF* mutants were extremely susceptible to serum at this concentration, with survival rates of only 0-3%. There was a strong negative correlation between bacterial survival rate and serum concentration for the *lgtF* mutants in all six strain backgrounds. Fig. 6.2 shows representative assay results for a group 1 strain (*Hp* 20) and a group 2 strain (*Hp* 13).

These results demonstrate that LPS structure is a key determinant of serum resistance in *Hp*. More specifically, they show that the LgtF-dependent Glc moiety that extends from HepI of the LPS core, and/or subsequent additions to this Glc (for example OAg or LosB1-dependent

extensions), confers a high level of protection against the killing effect of complement in human serum.



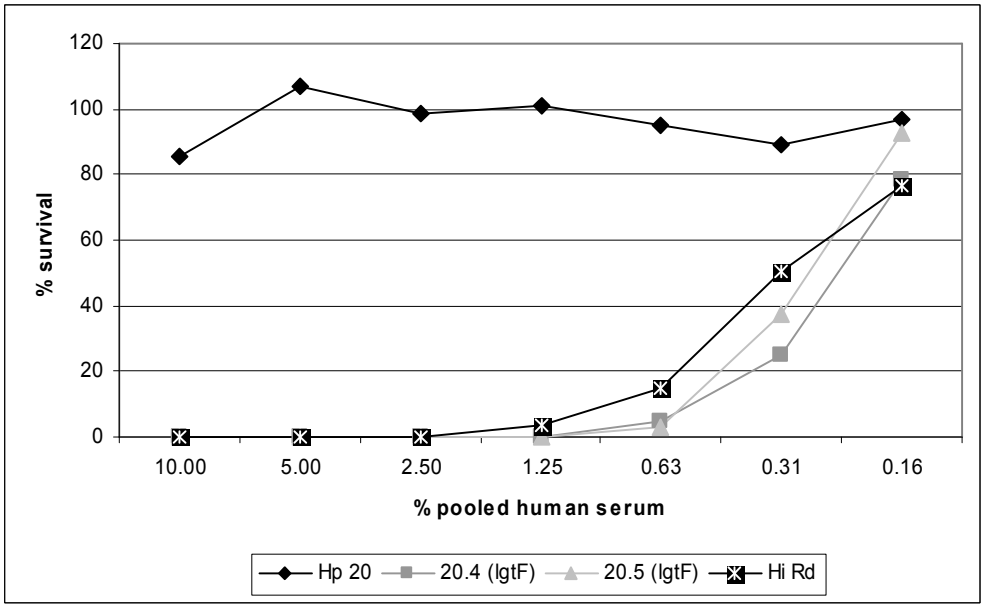


Figure 6.2. Resistance of *lgtF* mutant strains to the killing effect of pooled human serum. Results are shown as the survival of inoculating bacteria as a percentage of the survival in a 10% decomplemented serum control well. The parental strains are shown as black diamonds. The name of each mutant is given in the legend, with the gene that is knocked out shown in brackets. The *lgtF* gene is interrupted by different drug resistance markers in the *Hp* strain 20 mutants 20.4 and 20.5. *Hi* strain Rd is included for comparison.

6.1.3 *WcfS* and *wbaP* are required for the production of OAg-containing glycoforms in group 1 and group 2 *Hp* strains, respectively

The comparison of wild type *Hp* strains in Section 6.1.1, and the effect of *lgtF* mutation seen in Section 6.1.2, together suggest that OAg may be important for the resistance of *Hp* to complement-mediated killing. The sequencing of various *Hp* OAg loci has now increased our understanding of the genetics of OAg synthesis in different *Hp* strains, allowing us to test this hypothesis more directly by mutating putative OAg genes rather than *lgtF*. This will also confirm or disprove our prediction that these genes are required for OAg synthesis.

In order to create *Hp* mutants lacking only OAg from their LPS, plasmid constructs were designed that would interrupt the chromosomal UndP-sugar phosphotransferase gene (*wcfS* or *wbaP*) with a drug resistance cassette. If our functional predictions are correct, this should prevent the initiation of OAg synthesis on the UndP carrier molecule. The *Hp* strain T3T1 *wcfS* gene, *Hp0273*, was amplified by PCR to make a construct suitable for strains with a group 1 OAg locus; after the insertion of a kanamycin resistance cassette, the completed plasmid construct (pRY35) was used to interrupt *wcfS* in *Hp* strain 20 by static aerobic transformation. As predicted, the resulting clone (*Hp* 20.16) lacked detectable OAg, consistent with a requirement for *wcfS* in the OAg synthesis process (Fig. 6.3a). The construct was also used to disrupt *wcfS* in *Hp* strain 19, but in this strain background the LPS profiles of the mutants had the same banding pattern as the wild type (data not shown).

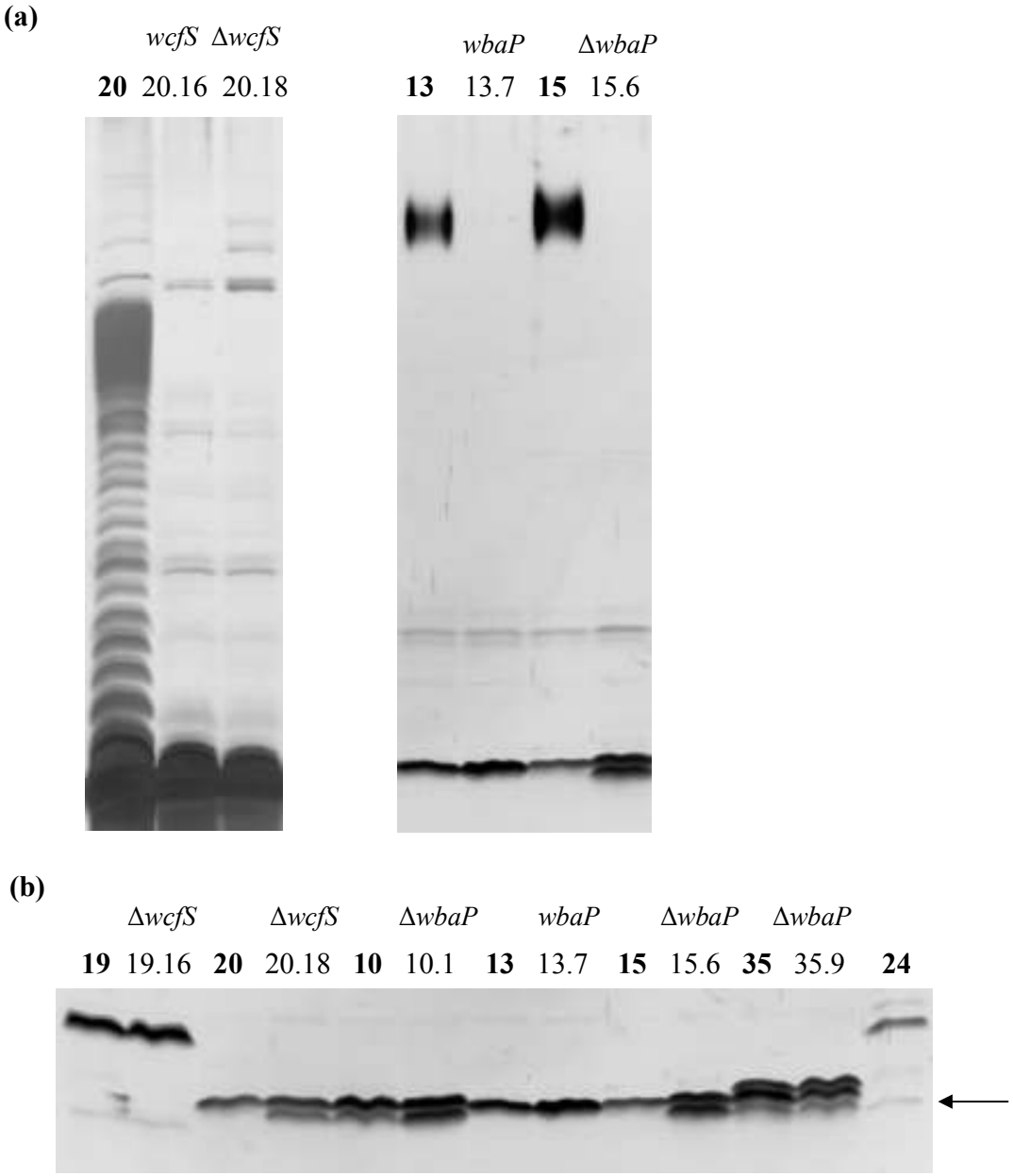


Figure 6.3. LPS profiles of *Hp wcfS* and *wbaP* mutants. The UndP-sugar phosphotransferase gene (*wcfS* for *Hp* group 1 and *wbaP* for *Hp* group 2) was mutated in six strain backgrounds. Proteinase K treated cell lysates were separated by tricine SDS-PAGE and stained with silver. Numbers above each lane are the *Hp* strain (in bold) or the mutant clone number. **(a)** Profiles for *Hp* strains in which OAg is normally visible, showing the loss of OAg-containing LPS glycoforms upon mutation of *wcfS* or *wbaP*. The *Hp* 20 wild type and mutant strains were grown on chocolate agar, and 12.5 µl of lysate at OD_{260} = 10 was loaded per lane so that the OAg was easily visible. Standard loading and growth conditions (i.e. 12.5 µl of lysate at OD_{260} = 1, grown on BHI agar) were used for the rest of the strains in panel (a) and for panel (b). **(b)** Core LPS profiles for mutants in all six strain backgrounds, with wild type *Hp* strain 24 shown for reference (the arrow shows the putative inner core LPS band in this strain).

The pRY35 construct interrupts *wcfS* at the position encoding …SKGP↓VFYK… (amino acid position 52 of the alignment in Fig. 5.3). This would be expected to prevent the production of an intact, functional WcfS enzyme. However, as the *Hp* 19 *wcfS* mutants had a wild type LPS profile we wished to ensure that WcfS function was fully eliminated, so a second construct (pRY41) was made in which 229 bp of the gene was deleted. This mutation results in the deletion of the regions of WcfS that are thought to interact with the UndP carrier and with the sugar (Wang et al., 1996).

Plasmid pRY41 was used to generate *wcfS* deletion mutants (Δ*wcfS*) in *Hp* strains 19 and 20. The LPS profiles of these clones were identical to those of the previous *wcfS* mutants, confirming that WcfS is required for OAg synthesis in *Hp* strain 20 (clone 20.18; Fig. 6.3a and b) but not for the generation of a wild type LPS profile in *Hp* strain 19 (clone 19.16; Fig. 6.3b). Attempts to produce *wcfS* or Δ*wcfS* mutants in the other group 1 *Hp* strains (22, 24, 31 and T3T1) using various transformation methods were unsuccessful.

To generate *wbaP* mutants in group 2 *Hp* strains, a plasmid construct was made in which the *Hp* strain 13 *wbaP* gene (*13I*; see Section 5.1.1) was interrupted by a drug resistance cassette. Conveniently, base pairs 108-1408 of this 1413 bp gene had already been cloned into the pBluescript vector as one of the HaeIII restriction fragments during the sequencing of the *Hp* 13 OAg locus described in Section 5.1.1, so this clone (5B) was simply cut with Bpu10I and blunt ligated to a HincII-cut kanamycin resistance cassette to interrupt *wbaP* in the region that encodes the periplasmic loop. *Hp* mutants made using this construct, pRY36, are described as *wbaP* mutants. The C-terminal domain of *Salmonella enterica* WbaP can apparently be translated from

a second start codon at the 3' end of the periplasmic loop-encoding region and the resulting protein has sugar phosphotransferase activity like the full length enzyme (Wang et al., 1996) (Section 5.1.5), so we made a second construct (pRY39) in which 271 bp of the region encoding the *Hp* 13 WbaP C-terminal domain was deleted (Δ*wbaP*).

The silver stained LPS profiles of *Hp* 13 *wbaP* mutants and *Hp* 15 Δ*wbaP* mutants are shown in Fig. 6.3a (we could not obtain any *Hp* 13 Δ*wbaP* or *Hp* 15 *wbaP* transformants). Both of these mutations result in the complete loss of the high molecular weight OAg-containing LPS glycoforms that are seen for wild type *Hp* 13 and 15. This provides evidence that the *wbaP* gene is indeed required for OAg synthesis in these strains, as predicted.

It is notable that *Hp wbaP* mutants (e.g. clone 13.7) do not appear to synthesise OAg-containing LPS even though the potential second start codon of *wbaP* remains intact. It seems likely that either (i) the C-terminus of *Hp* WbaP cannot be translated from this second 'start codon' or is not functional on its own, or (ii) the interruption of the region that encodes the periplasmic loop disrupts downstream transcription in these mutants.

Overall, *wcfS* was mutated in two strain backgrounds (*Hp* 19 and 20) and *wbaP* in four strain backgrounds (*Hp* 10, 13, 15 and 35). Although the most striking changes to the LPS occurs in the OAg region (Fig. 6.3a), the core regions of the wild type and mutant LPS profiles (Fig. 6.3b) provide additional information regarding outer core substituents in these strains. The LPS of *Hp* strain 20 is known to include one outer core sugar, Gal, which is attached to HepIII by LpsB. In the *Hp* 20 Δ*wcfS* mutant, half of the LPS molecules in the core region run at a molecular weight

consistent with inner core LPS only (see arrow on Fig. 6.3b): the most likely scenario is that they lack the outer core Gal. The same trend is seen between *Hp* strain 15 (which also contains *lpsB*) and its Δ*wbaP* mutant. The reason for this is unclear; perhaps disruption of the OAg genes affects the availability of nucleotide sugar precursors such as UDP-Gal.

6.1.4 *WcfS* and *wbaP* mutants have lowered serum resistance

The ability of the Δ*wcfS*, Δ*wbaP* and *wbaP* mutant strains shown in Fig. 6.3b to survive in serum concentrations of 0.16-10% was compared to that of their parent strains. Mutation of the UndP-sugar phosphotransferase gene resulted in a dramatic loss of resistance to complement mediated killing in all six strain backgrounds. This is especially obvious at high serum concentrations (Fig. 6.4a). Serum resistance profiles representative of group 1 Δ*wcfS* mutants and group 2 Δ*wbaP* mutants are shown in Figures 6.4 b and c respectively.

The heavily loaded LPS profiles in Fig. 4.6 showed that the vast majority of LPS molecules in *Hp* strain 10 do not contain OAg. Despite this, the WbaP enzyme appears to be crucial for the survival of this strain in 2.5-10% human serum (Fig. 6.4c). This implies that the long OAg chains present on a minority of the LPS molecules are enough to prevent the deposition of complement on the bacterial cell surface.

(a)

Wild type strain	***Hp* 19**	***Hp* 20**	***Hp* 10**	***Hp* 13**	***Hp* 15**	***Hp* 35**
% survival in 10% PHS	74	92	88	133	99	58
UndP-sugar PT mutant	**19.16**	**20.18**	**10.1**	**13.7**	**15.6**	**35.9**
% survival in 10% PHS	1	0	0	0	0	0

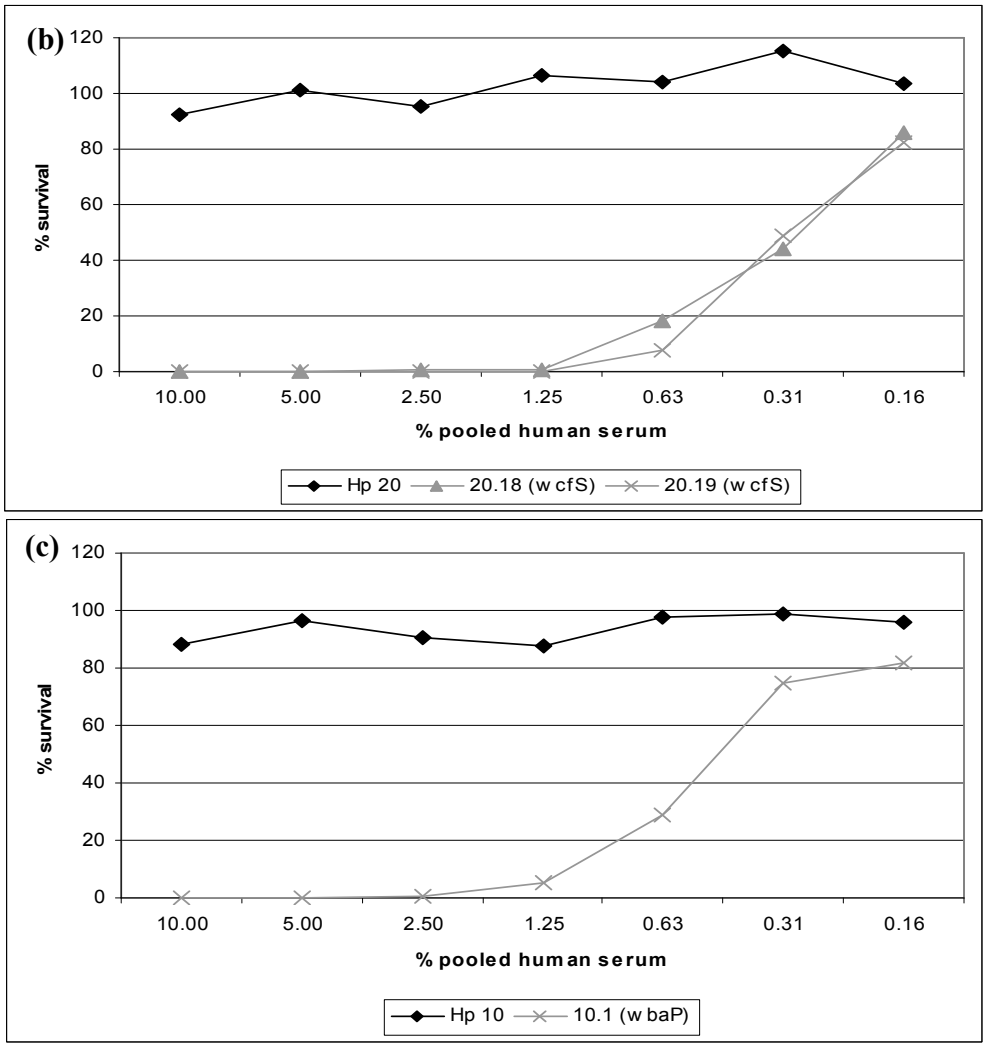


Figure 6.4. Resistance of UndP-sugar phosphotransferase mutant strains to the killing effect of pooled human serum. Results are shown as the survival of inoculating bacteria as a percentage of the survival in a 10% decomplemented serum control well. **(a)** Percentage survival in 10% serum (PHS) of each of the *Hp* mutant strains in Figure 6.3(b), compared to their parental strains. **(b)** A representative assay for a group 1 strain, *Hp* 20. The parental strain is shown in black diamonds. The name of each mutant is given in the legend, with the gene that is knocked out listed in brackets. 20.18 and 20.19 are independent transformants made using the same parental strain and plasmid. Each data point represents the mean of two replicates. **(c)** A representative assay for a group 2 strain, *Hp* 10. Annotated as for (b). Each data point represents a single colony-count.

The serum resistance of different *Hp* strain 15 LPS mutants produced in Chapters 4, 5 and 6 was compared directly in a triplicate assay (Fig. 6.5). The similarity between the serum resistance profiles of clones in which *lgtF*, the OAg ligase gene *waaL*, or *wbaP* have been mutated helps to clarify two points. Firstly, it suggests that the serum sensitivity of *lgtF* mutants is due to the loss of OAg rather than the loss of the core Glc residue. This Glc may play a different biological role or may be conserved simply as an attachment site for the addition of OAg and outer core sugars. Secondly, it argues against the possibility that the *Hp* 15 OAg synthesis/ligation pathway is also used to glycosylate outer membrane proteins: any such glycoproteins should still be present in an *lgtF* mutant and we would expect their 'OAg' to protect against complement-mediated killing, but this was not observed.

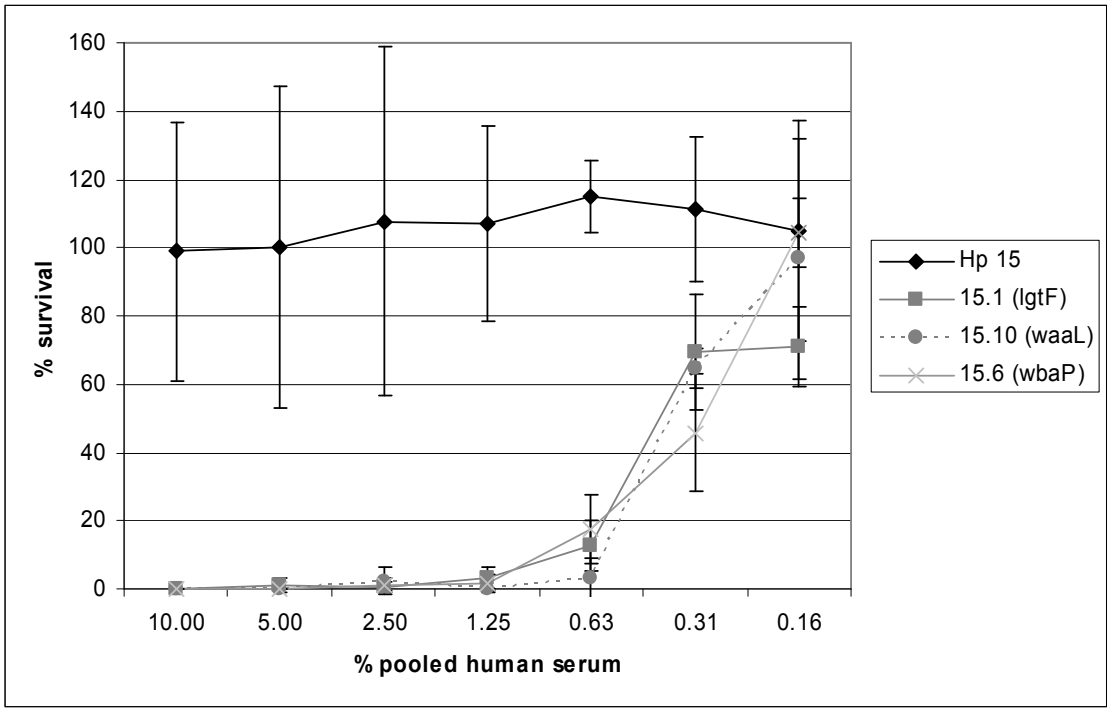


Figure 6.5. Resistance of *Hp* strain 15 and its isogenic mutants to the killing effect of pooled human serum. Results are shown as the survival of inoculating bacteria as a percentage of the survival in a 10% decomplemented serum control well. The name of each *Hp* strain 15 mutant is given in the legend, with the gene that is knocked out listed in brackets. The experiment was performed in triplicate so each data point represents the mean of three replicates. Error bars show ± one standard deviation.

6.1.5 Analysis of the LPS glycoforms of *Hp* strain 19

The apparent discrepancy between glycoform and serum resistance analyses for *Hp* 19.16 (Figures 6.3 and 6.4a) demanded an explanation, so we combined all the available data regarding *Hp* strain 19 and developed some additional mutants to help resolve the issue. The LPS profile of *Hp* 19 comprises a single band at a molecular weight suggestive of inner core LPS plus four sugars. This strain contains similar OAg locus genes to *Hp* T3T1 (Table 5.2), and the glycoform runs at the same position as the *Hp* T3T1 inner core LPS + 1 O-unit band (Fig. 3.7b), so we might expect it to contain one four-sugar O-unit. However, it also runs at the same position as the LPS of *Hp* strain 34, with which *Hp* 19 shares four putative outer core GT genes identified in Chapter 3 including *losB1*, *losA1* and *lic2A* (Fig. 3.8b), suggesting an alternative structure comprising inner core LPS plus four outer core sugars. OAg and LosB1-dependent outer core extensions are generally attached via the Glc on HepI of the inner core LPS, so either of these structures would be consistent with the observed *Hp* 19 *lgtF* mutant LPS profile, which appears to lose five sugars compared to wild type LPS.

To explore the structure and genetics of *Hp* 19 LPS further, we mutated five different genes that may be involved in its synthesis and compared the LPS profiles (Fig. 6.6) and serum resistance (Fig. 6.7) of the wild type and mutant strains. Three of these mutants have been described in earlier chapters. As reported in Section 6.1.3, the LPS profiles of *Hp* 19 Δ*wcfS* mutants are identical to that of the wild type strain, so we might assume that the OAg locus is non-functional or is not expressed in *Hp* 19 under laboratory conditions. However, there are two other striking phenotypic differences which dispute this. Firstly, whilst *Hp* 19 bacteria normally grow as an aggregate in BHI broth and are hard to resuspend in PBS after growth on BHI agar, the Δ*wcfS*

(and *wcfS*) mutants grew as a cloudy suspension in broth and easily formed a smooth suspension in PBS after growth on agar. This phenomenon was also observed in *Hp* strain 20 Δ*wcfS* mutants, but was less evident as the wild type cells do not aggregate as much as those of *Hp* 19. Secondly, *Hp* 19 Δ*wcfS* mutants cannot survive in 10% serum, whereas the wild type strain shows around 75% survival (Fig. 6.7). Perhaps any O-unit-containing glycoforms in the wild type *Hp* 19 LPS profile were being masked by an alternative outer-core-containing glycoform that runs at an equivalent molecular weight.

Figure 6.6. Tricine SDS-PAGE profiles of *Hp* strain 19 mutants. The gene that has been knocked out is stated above each profile (WT = wild type strain). The clone numbers are the same as in Figure 6.7.

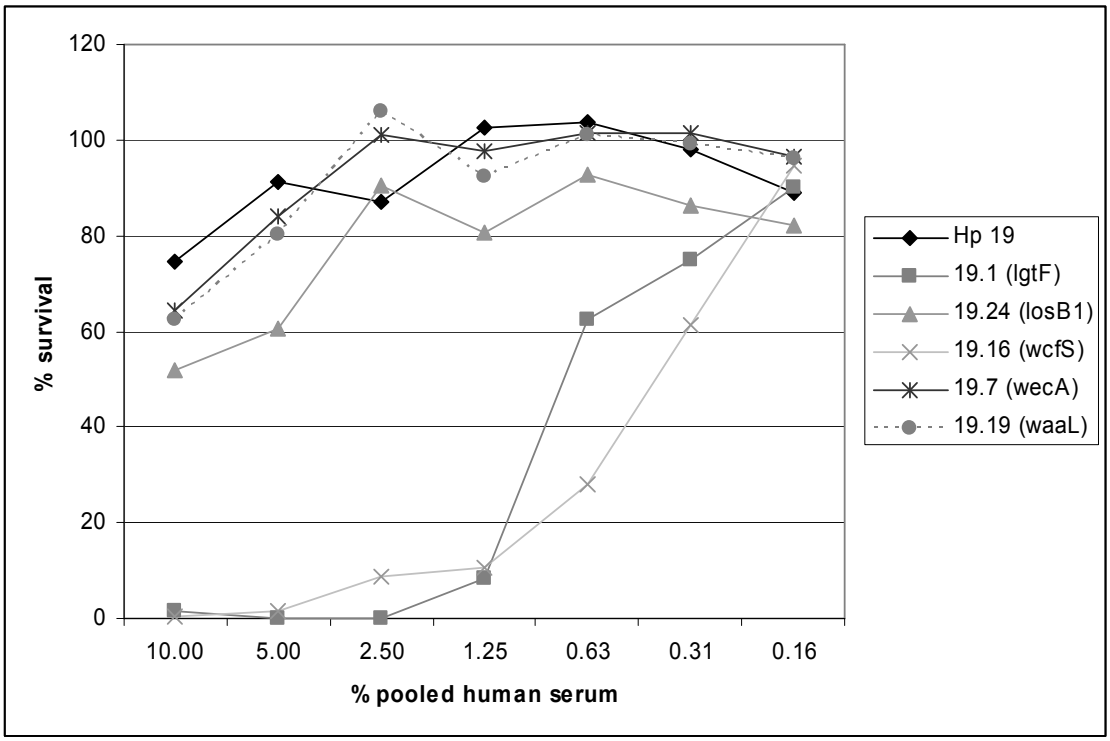


Figure 6.7. Resistance of *Hp* strain 19 and its isogenic mutants to the killing effect of pooled human serum. Results are shown as the survival of inoculating bacteria as a percentage of the survival in a 10% decomplemented serum control well. The name of each *Hp* strain 19 mutant is given in the legend, with the gene that is knocked out listed in brackets.

To test this hypothesis, *Hp* 19 *losB1* mutants 19.24 and 19.25 were generated using the plasmid construct 5BA2 (a gift from Mary Deadman); 5BA2 contains an *Hi losB1* gene interrupted by a kanamycin resistance cassette, and the homology between the *Hi* and *Hp losB1* alleles was evidently sufficient to allow recombination. The LPS profiles of the mutants comprised two bands: a strong glycoform corresponding to inner core LPS only, and a weaker glycoform corresponding to the molecular weight of the wild type band (Fig. 6.6). The weaker band most likely contains a single O-unit, although we cannot rule out the possibility that *Hp* strain 19 contains a second LosB paralogue or other GT that can initiate outer core addition in the mutant strains.

From this evidence we conclude that *Hp* strain 19 is probably capable of adding two mutually exclusive extensions to the Glc on HepI of its core LPS: a single four-sugar O-unit with a similar composition to the O-unit of strain T3T1, and a four-sugar extension for which LosB1 is the first transferase. The O-unit appears to confer resistance to complement-mediated killing, whilst the outer core extension does not. Presumably the latter must have other benefits to the survival of *Hp* in certain environmental conditions. *Hp* strain 19 may be able to adapt by regulating the expression of the OAg locus and/or the outer core GTs to alter the balance of LPS glycoforms with different biological properties.

The LPS profiles of *Hi* strains are often difficult to interpret, and one reason for their microheterogeneity is thought to be the presence of glycoforms with different degrees of a (non phase variable) extension: for example, an *Hi* strain with LosB1 and LosA1 may contain a mixture of LPS molecules with DDHep-Glc, DDHep only, or no sugars extending from the Glc on HepI. However, this idea is challenged by the simplicity of the *Hp* 19 LPS profile, in which no 'incomplete' glycoforms are observed. This suggests that either the GTs involved are extremely efficient or the outer core extension in this strain is added *en bloc* in a similar manner to an O-unit in *Hp* or HMG in *Hi*. One could distinguish between the two hypotheses by mutating the *losA1* or *lic2A* GT genes and examining the resulting LPS profiles for evidence of glycoforms with two or three outer core sugars.

In addition to *wcfS*, *Hp* strain 19 contains another putative UndP-sugar phosphotransferase gene: a homologue of *E. coli wecA* (see Section 5.1.5). To check whether WecA is involved in LPS synthesis and/or serum resistance in *Hp* strain 19, mutants were created using the construct

pRY34. These mutants displayed a wild type LPS profile (Fig. 6.6) and their serum resistance was not significantly different to that of wild type *Hp* 19 (Fig. 6.7). *Hp* strain 13 *wecA* mutants similarly had wild type LPS profiles and serum resistance levels (data not shown). From these results, it appears that *Hp* WecA is not involved in LPS synthesis and that any WecA-dependent surface antigens do not confer serum resistance.

In Section 5.3.3 we discussed possible explanations for the presence of two different putative OAg ligase genes (*waaL*) in *Hp* strain 19. Whilst mutating the OAg ligase in *Hp* strain 15 resulted in decreased serum resistance due to the loss of OAg (Fig. 6.5), the *Hp* 15-like OAg ligase gene in *Hp* strain 19 is not required for serum resistance (clone 19.19, Fig. 6.7). As with WecA, we can conclude that any surface antigen (e.g. a glycoprotein) whose synthesis requires this second 'OAg' ligase in *Hp* 19 is not necessary for protection against complement-mediated killing in vitro; in fact it seems most likely that this enzyme is only utilised in *Hp* strains with a group 2 or ungrouped OAg locus and not in *Hp* strain 19.

6.2 Phosphocholine as a component of the *Haemophilus* LPS molecule

Phosphocholine (PCho) is a component of cell surface structures in a number of respiratory tract bacteria including *Hi*, *Pseudomonas aeruginosa*, *Neisseria meningitidis* and *Streptococcus pneumoniae* (Rock et al., 2001; Weiser et al., 1998a). In *Hi* this molecule is found on the LPS (Weiser et al., 1997) and has been shown to mediate interactions with the host by several different mechanisms, as discussed later. The synthesis of PCho and its incorporation into *Hi* LPS requires four enzymes encoded by the *lic1* locus, which is found in over 90% of NTHi strains and in a majority of capsular strains (McCrea et al., 2010; Weiser et al., 1989a). In the

proposed pathway, choline is first taken up from the host environment by a permease, Lic1B; once inside the cell, it is phosphorylated by Lic1A and converted to a nucleotide sugar (possibly CDP-Cho) by Lic1C (Rock et al., 2001; Weiser et al., 1997). Finally the PCho is transferred to the LPS by Lic1D (Lysenko et al., 2000a). The *lic1A* gene contains a tetranucleotide repeat tract, so the expression of Lic1A and of the PCho epitope is phase variable (Weiser et al., 1989b). In this section we investigate differences in PCho expression between *Hi* and *Hp* and examine how these could relate to the different behaviour of the two species.

6.2.1 *Hp* LPS lacks phosphocholine but can express it upon the introduction of a transgenic *lic1* locus

Several lines of evidence suggest that *Hp* does not contain the *lic1A-lic1D* genes. No BLAST matches were identified for *lic1A-lic1D* in the *Hp* T3T1 genome sequence and no DNA fragments from the *Hp* study strains hybridised to a *lic1B* probe during Southern analysis (Section 3.3.3). Further Southern analysis using a probe comprising the *lic1A* repeat sequence $(CAAT)_5$ was also negative in all 20 study strains (Fig. 7.1). In addition, none of the strains reacted with the PCho-specific antibody TEPC15 on a dot immunoblot of cell lysates (Fig. 6.8a). Structural analysis of LPS from *Hp* strains T3T1, 13, 20 and 22 has not identified PCho as an LPS substituent in any of the four strains (Elke Schweda, pers. comm.). Together the data suggest that *Hp* does not decorate its LPS with PCho.

To explore possible evolutionary reasons for the difference in PCho expression between *Hi* and *Hp*, a plasmid construct (pRY29) was designed that would introduce the *lic1* locus from *Hi* strain Rd into the *Hp* genome (Fig. 6.8b). A number of the *Hp* study strains naturally contain a

tetracycline resistance gene cluster (*tet*): this was chosen as the target recombination site as it is well conserved but non-essential and its disruption would result in a defined phenotypic change. Part of the construct was sequenced to ensure that its *lic1A* gene contained a number of CAAT repeats that was permissive for Lic1A protein expression.

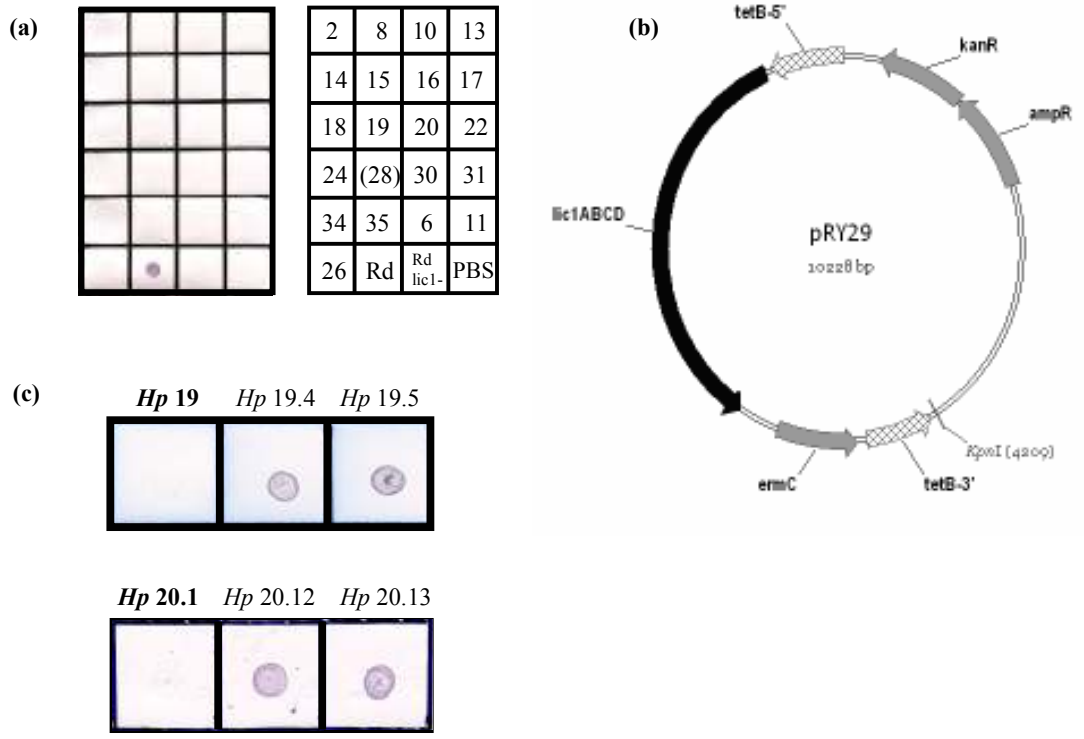


Figure 6.8. Transgenic phosphocholine expression in *H. parainfluenzae*. (a) Dot immunoblot using TEPC15 antibody to detect PCho on LPS. On the blot (left), each square contains 3 µl boiled cells in PBS. *Hp* strains are listed in the key on the right; the positive control is *Hi* Rd. **(b)** Representation of the plasmid construct pRY29, in which the *Hp* tetracycline resistance gene *tetB* is interrupted by *Hi* Rd genes *lic1A-lic1D* and an erythromycin resistance gene (*ermC*). The genes were cloned using the commercial cloning vector pSCA-amp-kan, so pRY29 also contains genes conferring kanamycin resistance (kanR) and ampicillin resistance (beta lactamase gene; ampR). *Lic1A-lic1D* are the genes required for the expression of PCho-containing LPS glycoforms in *Hi*. **(c)** TEPC15 dot immunoblots using lysates from parental *Hp* strains (in bold) and their derivative clones containing a transgenic *lic1* locus.

Hp strain 19 was successfully transformed with the pRY29 plasmid, resulting in the integration of genes *lic1A-lic1D* and *ermC* into the genomic *tetB* gene as confirmed by PCR. Boiled lysates from *Hp* 19 and two of its *lic1*$^{+}$ clones, *Hp* 19.4 and 19.5, were immunoblotted using TEPC15 as the primary antibody to detect the expression of PCho on the LPS (Fig. 6.8c). The antibody was found to bind to both of the *lic1*$^{+}$ clones. This demonstrates that *Hp* can express and utilise the proteins encoded by the *Hi lic1* locus and that there is no fundamental reason why the species could not include PCho as a constituent of its LPS if it was advantageous to do so.

The *lic1D* gene in *Hi* strain Rd (and hence in pRY29) directs the addition of PCho to the Glc on HepI of the LPS core. As this is the site of attachment for OAg in *Hp*, we wished to examine whether the addition of PCho might inhibit OAg addition, and whether long OAg chains might prevent interactions between PCho and host cells thereby making PCho addition futile. *Hp* strain 20 was chosen as a suitable candidate for these experiments because it synthesises OAg and can be transformed relatively efficiently. As it does not contain the target *tet* locus, this was first introduced by mixing *Hp* 20 cells with *Hp* T3T1 genomic DNA in a static aerobic transformation and selecting on tetracycline. PCR analysis was used to confirm the presence of the *tet* locus in transformant *Hp* 20.1; this clone was subsequently transformed with plasmid pRY29 to generate two *Hp* 20 *lic1*$^{+}$ clones, both of which bound TEPC15 (Fig. 6.8c). Structural analysis of LPS isolated from one of these clones (*Hp* 20.12) confirmed that a minority of glycoforms contained PCho, and that the molecule was found at the C-6 position of the Glc on HepI as expected (Brigitte Twelkmeyer, pers. comm.).

6.2.2 Competition between the addition of PCho and other LPS substituents in the *Hp lic1*$^{+}$ clones

In *Hp* strain 19, the only LPS glycoforms normally visible by tricine SDS-PAGE correspond to inner core LPS plus four sugars (Section 6.1.5). Upon the introduction of the *lic1* locus, a second band appears at approximately the position associated with inner core LPS (Fig. 6.9a), indicating that the addition of PCho to some of the LPS molecules has indeed inhibited the addition of a four-sugar extension. Without structural analysis we cannot confirm whether it is the O-unit or the LosB1-dependent extension (or both) that has been affected, but as we predict both the PCho and the fourth heptose to be added via C-6 of the same inner core Glc residue (Sections 3.3.5 and 6.2.1) there would be direct competition between cytoplasmic Lic1D and LosB1 enzymes for the same acceptor site.

The group 1 *Hp* OAgs that have been structurally analysed so far (including that of *Hp* strain 20) are attached to C-4 of the Glc on HepI, whereas PCho is added to C-6 of this Glc residue: theoretically, therefore, both substituents could be added to the same LPS molecule. However, the structural analysis of *Hp* 20.12 LPS found that OAg and PCho addition are mutually exclusive in this transformant (Brigitte Twelkmeyer, pers. comm.). This could be due to steric hindrance of the periplasmic OAg ligase by PCho, which would have been added earlier in the process because Lic1D is cytoplasmic.

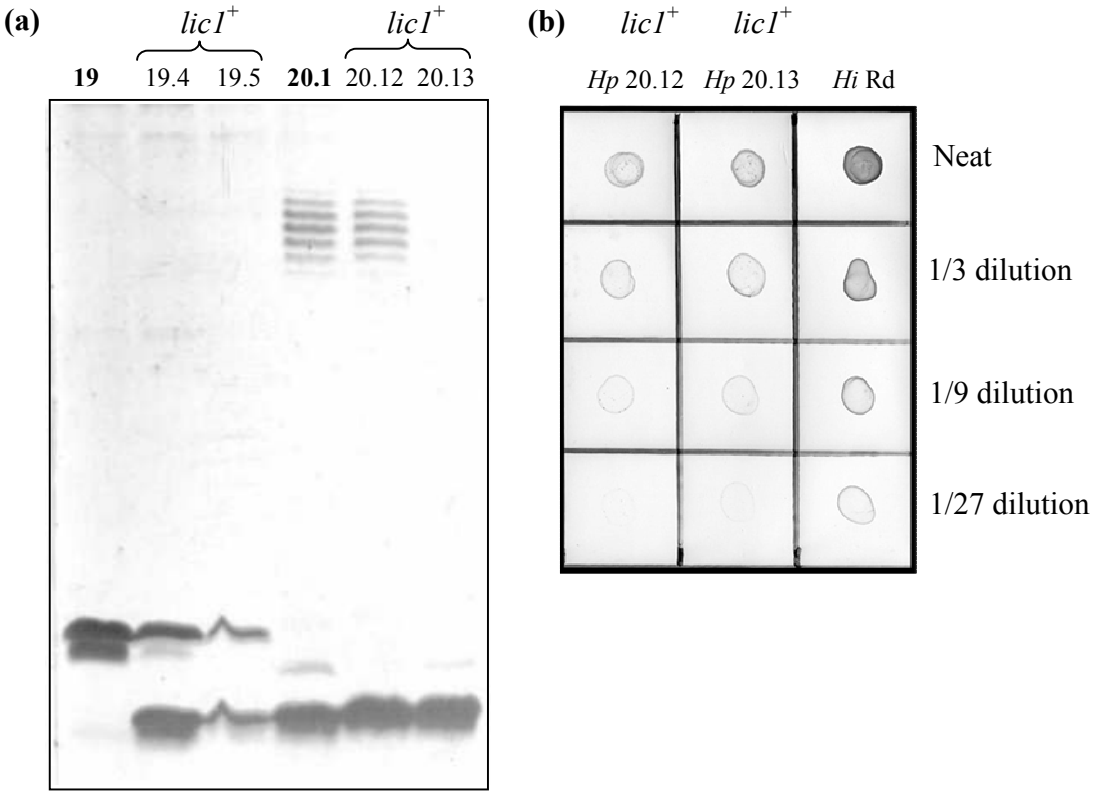
The LPS profile of *Hp* 20.12 was identical to that of the parental strain *Hp* 20.1, suggesting that the proportion of LPS molecules that contain OAg is not affected by PCho addition in this strain (Fig. 6.9a). In contrast, the second *lic1*$^{+}$ clone (*Hp* 20.13) consistently produced lower amounts

of OAg-containing LPS than *Hp* 20.1 (Fig. 6.9a). Both mutants bound TEPC15 equally well (Fig. 6.9b) and their *lic1A* genes contained the same number of CAAT repeats as the pRY29 construct, so the discrepancy did not seem to be a result of PCho phase variation. Two more *Hp* 20 *lic1*$^{+}$ clones, 20.14 and 20.15, were generated and were found to display similar LPS profiles to *Hp* 20.12 (data not shown). Thus clone *Hp* 20.13 is atypical and may have an unidentified mutation in an OAg-related gene.

Overall, our results for *Hp* strain 19 mutants indicate that transgenic Lic1D competes with other LPS outer core glycosyltransferases for the same acceptor molecule; this might be detrimental to the survival of *Hp* 19 in the host if the natural extensions are required for key biological roles. Meanwhile, our results for *Hp* strain 20 mutants suggest that PCho addition to Glc-HepI precludes OAg attachment to the same molecule but does not affect the overall amount of OAg on the cell surface, presumably because PCho was only found on a minority of the LPS molecules.

6.2.3 Serum resistance of *Hp lic1*$^{+}$ clones

The four *Hp* 20 *lic1*$^{+}$ transformants were assayed for their ability to resist the killing effect of pooled human serum (Fig. 6.9c). The three 'typical' clones retained wild type levels of serum resistance: the incorporation of PCho into *Hp* 20 LPS via an extension from HepI does not appear to make the strain more susceptible to complement-mediated killing. Clone *Hp* 20.13 was more sensitive to serum that the parental strain, consistent with the effects of a lower level of OAg expression seen in Fig. 6.4.


(a)
lic1+
lic1+
19
19.4
19.5
20.1
20.12
20.13
(b)
lic1+
lic1+
Hp 20.12
Hp 20.13
Hi Rd
Neat
1/3 dilution
1/9 dilution
1/27 dilution

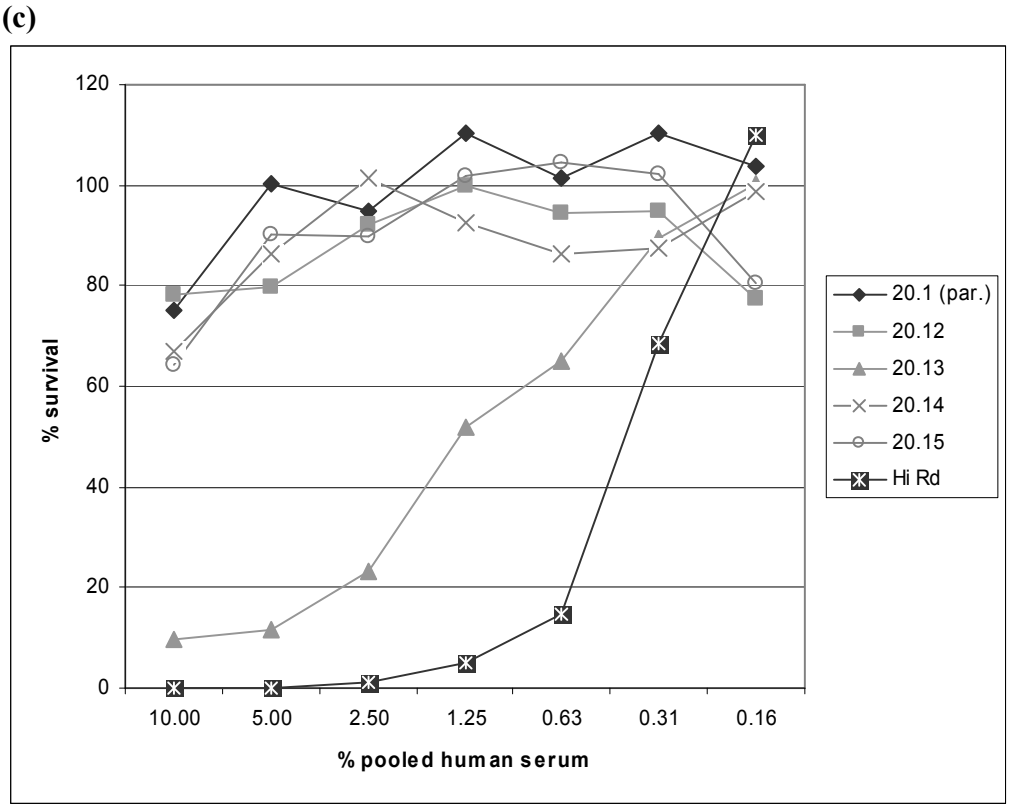

(c)
120
100
80
60
40
20
0
% survival
10.00
5.00
2.50
1.25
0.63
0.31
0.16
% pooled human serum
20.1 (par.)
20.12
20.13
20.14
20.15
Hi Rd

Figure 6.9. Phenotypic characterisation of *Hp lic1*$^+$ transformants. **(a)** Tricine SDS-PAGE profiles of parental (bold) and *lic1*$^+$ strains. **(b)** TEPC15 dot immunoblot of two *Hp* 20 *lic1*$^+$ strains using different lysate dilutions, showing that TEPC15 has approximately equal affinity for the two strains (*Hi* strain Rd is included as a positive control). **(c)** Resistance of *Hp* 20.1 (parental) and four *lic1*$^+$ transformants to the killing effect of pooled human serum. Results are shown as the survival of inoculating bacteria as a percentage of the survival in a 10% decomplemented serum control well.

6.3 The association of *Hp* with epithelial cells

Adherence to host cells is one of the first steps required for the colonisation of the respiratory tract, allowing bacteria to establish themselves in a relatively constant environment whilst avoiding ciliary clearance. For *Hi*, association with epithelial cells is thought to be a prerequisite for invasive disease, as the bacteria must move either through or between the cells of the epithelial barrier to reach the bloodstream. We performed some preliminary association and uptake assays to examine whether *Hi* and *Hp* interact differently with host cells.

A human bronchial epithelial cell line, 16HBE14, was grown in a 96-well plate to establish a monolayer. The medium in each well was then replaced with 100 µl DMEM containing $5x10^7$ bacteria. After an incubation of 2.5 h, the monolayer was washed to remove bacteria that were not associated; appropriate dilutions of the adherent bacteria were plated onto BHI agar and colonies were counted following overnight growth. The same procedure was followed for control wells containing no epithelial cells to check that the bacteria did not adhere to the plastic.

Two OAg-expressing *Hp* strains were compared to *Hi* strain Rd regarding their ability to associate with epithelial cells. A significantly higher proportion of bacteria adhered to the monolayer for *Hp* strains 13 and 20 than for *Hi* Rd (Fig. 6.10a). If these results are typical of *Hp* and *Hi* in general, *Hp* may be at an advantage during colonisation as it is more protected from

physical clearance if attached to a host surface rather than dispersed. However, adherence is only one of the factors involved in successful colonisation and should be considered alongside others such as immune responses and growth requirements.

A second set of wells in the same assay was reserved for the analysis of bacterial uptake (or invasion) into epithelial cells. Gentamicin was added to kill bacteria on the outside of the epithelial cells, whilst intracellular bacteria remained protected and could be plated onto BHI agar as before. The percentage of total bacteria that were found inside epithelial cells was $1.9x10^{-5}$ % for *Hp* strain 13, $3.3x10^{-5}$ % for *Hp* strain 20 and $9.3x10^{-5}$ % for *Hi* strain Rd (Fig. 6.10b). This data is consistent with the hypothesis that of the two species, *Hp* is less able to invade epithelial cells so may be less likely to behave pathogenically. This could be investigated further by using a wider range of *Hi* and *Hp* strains and by analysing cytokine production by the epithelial cell layer.

To examine the role of LPS structure in adhesion, association assays were performed using *lgtF* mutants of strains with group 1 or group 2 OAg (Fig. 6.11). Intriguingly, the effect of ablating the addition of OAg and Glc to the LPS (i.e. mutating *lgtF*) was strain-dependent. The *Hp* T3T1 *lgtF* mutant, *Hp* 26.2, showed fifteen times greater association with epithelial cells than wild type *Hp* T3T1, suggesting that the OAg may impair adherence. For *Hp* strain 13, which has a different OAg structure, the *lgtF* mutant appeared to adhere less than the wild type strain. OAg therefore appears to play opposite roles in adhesion in the two study strains.

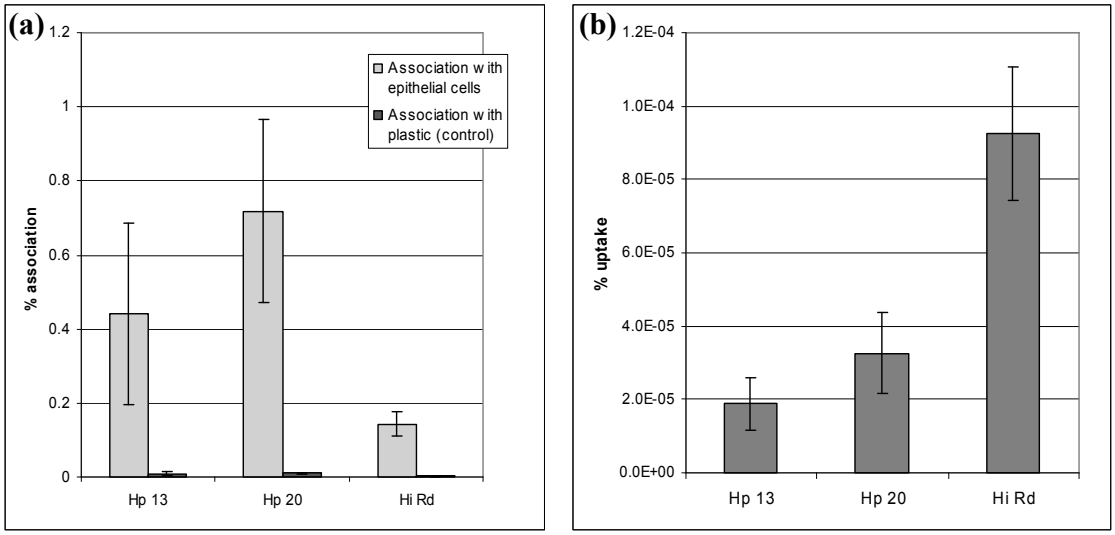


Figure 6.10. Association of *Haemophilus* with human bronchial epithelial cells. *Hp* strains 13 and 20 and *Hi* strain Rd were incubated in a 96-well plate with a monolayer of 16HBE14 cells in DMEM medium. The y axes record the number of bacteria that were **(a)** associated, or **(b)** inside the epithelial cells, after a 2.5 h incubation as a percentage of the total bacteria in a control well. Each bar represents the mean value for three replicates; the error bars show ± one standard deviation.

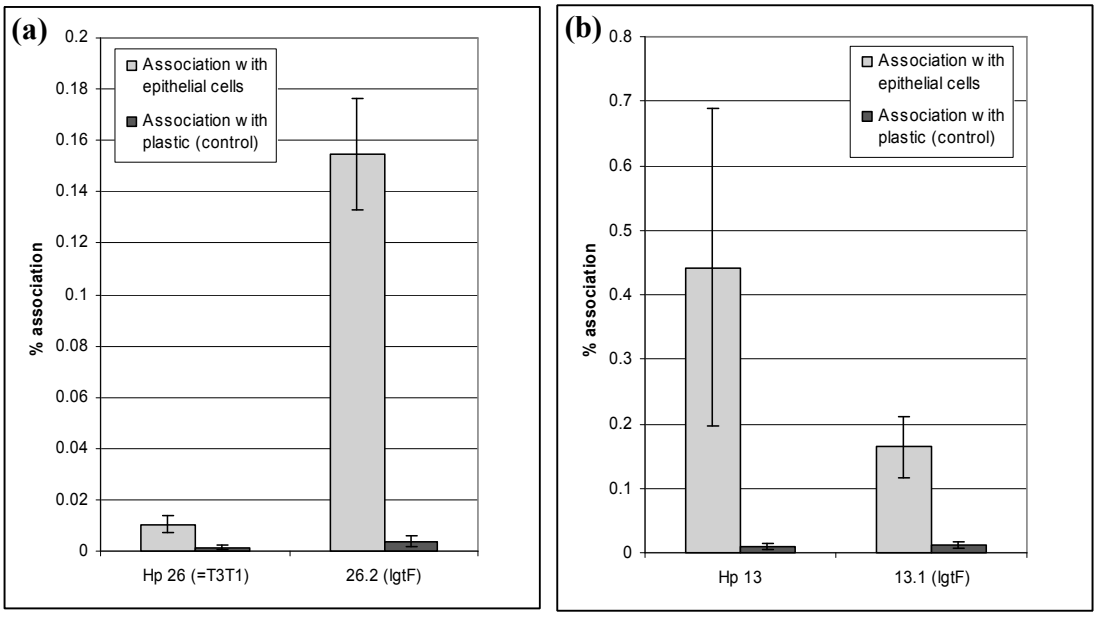


Figure 6.11. Association of *Hp lgtF* mutants with human bronchial epithelial cells. Wild type *Hp* strains T3T1 and 13 and their *lgtF* mutants (which lack OAg and Glc from the LPS) were incubated in a 96-well plate with a monolayer of 16HBE14 cells in DMEM medium. The y axes record the number of bacteria that were associated after a 2.5 h incubation as a percentage of the total bacteria in a control well. Each bar represents the mean value for three replicates; the error bars show ± one standard deviation. Note that the scales differ; both *lgtF* mutants show approximately the same level of association to epithelial cells.

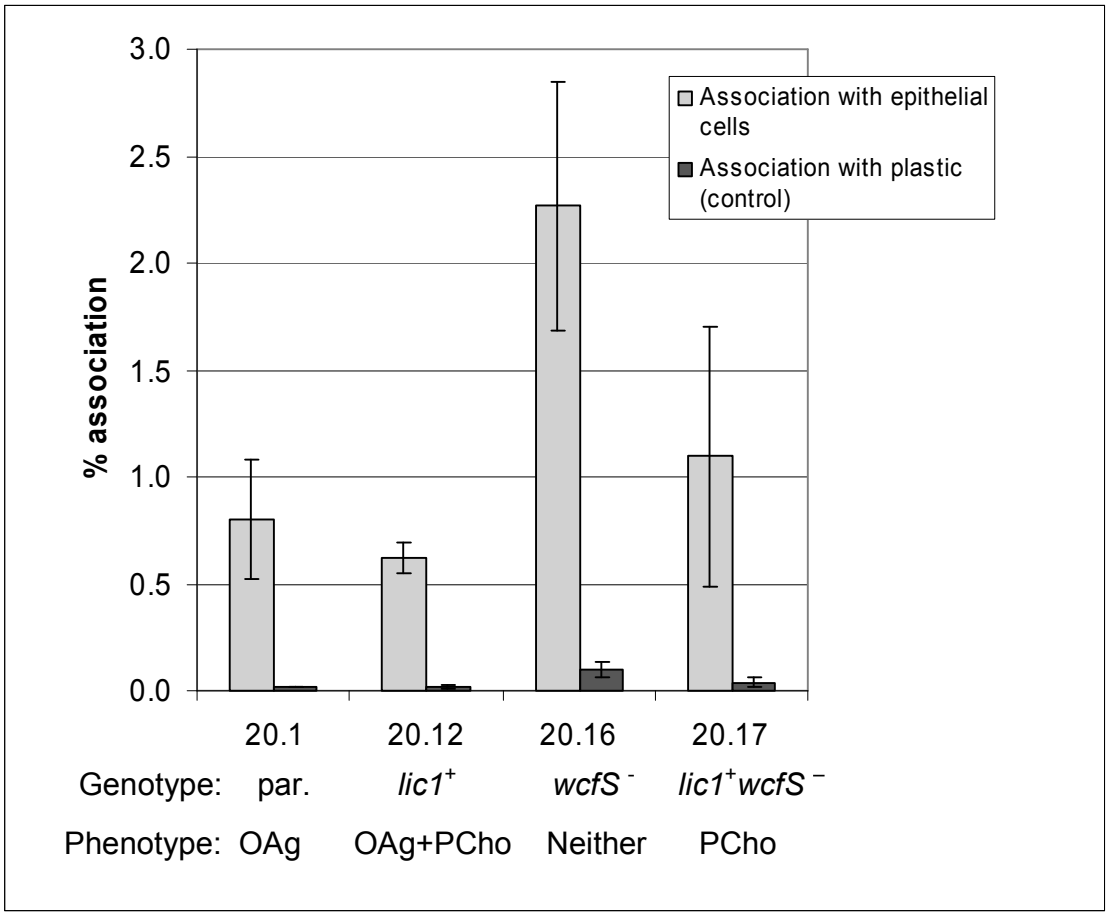


Figure 6.12. Association of *Hp* 20 LPS mutants with human bronchial epithelial cells. The *Hp* clone number and its genotype and phenotype (the presence of OAg and/or phosphocholine on the LPS) is given under each pair of bars. *Hp* 20.1 is the parental strain (par.) for the series of mutants. The y axis records the number of bacteria that were associated after a 2.5 h incubation as a percentage of the total bacteria in a control well. Each bar represents the mean value for three replicates; the error bars show ± one standard deviation.

Once further mutant strains had been developed, we could test the role of OAg in adhesion more directly using UndP-sugar phosphotransferase mutants. Four strains derived from *Hp* 20 were compared in an epithelial cell association assay (Fig. 6.12). In agreement with Fig. 6.10a, the parental *Hp* clone 20.1 (*Hp* 20 transformed with the *tet* locus) showed approximately 0.8% association. Upon the abolition of OAg synthesis (*wcfS* mutant, *Hp* 20.16) the proportion of cells

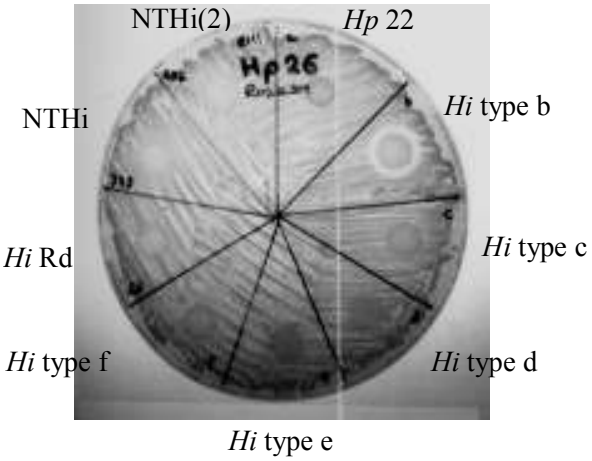
that associated with the epithelial layer almost tripled to 2.3%. An *Hp* 20 *lic1*$^{+}$ clone which synthesised both OAg and PCho (*Hp* 20.12) showed similar levels of association to the parent strain. However, when the OAg was abolished, the addition of PCho appeared to decrease epithelial cell association (compare *Hp* clone 20.16 with the double mutant, 20.17). In neither case was there evidence for a positive role for PCho in epithelial cell attachment as has been proposed in *Hi* (Swords et al., 2000).

6.4 Interactions between *Hi* and *Hp*

Hp and *Hi* are both common commensal bacteria of the human nasopharynx, and must compete with each other and with other flora for attachment sites and nutrients if they are to colonise. In addition to evolving competitive strategies such as stronger adhesion, more efficient uptake systems or faster replication rates, some bacteria directly inhibit the growth of other species or strains using microbial defence systems including the release of bacteriocin proteins.

To test whether the growth of *Hp* is inhibited by *Hi*, 3 µl circles of eight different live *Hi* suspensions at OD_{650} = 0.35 were pipetted onto a lawn of *Hp* strain T3T1 on BHI agar and the plate was incubated for 24 h at 37 °C (Fig. 6.13a). *Hp* T3T1 growth was not inhibited around NTHi, *Hi* Rd or *Hi* strains with capsular serotypes c, d, e or f, suggesting that in general *Hp* may be able to co-colonise with *Hi*. In contrast, a zone of inhibition was clearly observed around the *Hi* serotype b (Hib) strain 10810. Another Hib strain, Eagan, produced the same effect (data not shown). *LgtF* mutants of *Hp* strains 13 and T3T1 were no more or less susceptible to inhibition by Hib than the parental strains, indicating that OAg has no effect on this interaction (Fig. 6.13b and c).

(a) Test strain: *Hp* T3T1



(b) Test strains: *Hp* 13 and its *lgtF* mutant

(c) Test strains: *Hp* T3T1 and its *lgtF* mutant

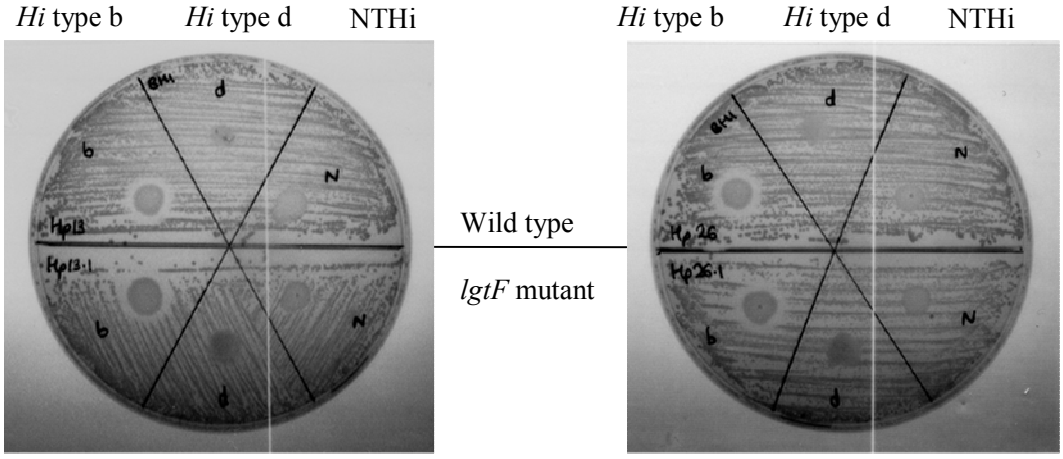


Figure 6.13. The effect of different *Hi* serotypes on the growth of *H. parainfluenzae*. The *Hp* test strain was spread onto BHI agar, then a 3 µl spot of an *Hi* cell suspension was pipetted onto each sector. Plates were incubated for 24 h at 37 °C and examined for zones of growth inhibition around each *Hi* strain. The *Hi* strains are labelled as follows: type b, 10810; type c, RM7422; type d, RM7033; type e, RM7066; NTHi, strain 375; NTHi(2), strain 486. **(a)** *Hp* strain T3T1 growth is inhibited by Hib. This panel also includes *Hp* strain 22 to check for inhibition between *Hp* strains. **(b-c)** *Hp* mutants lacking OAg appear no more or less susceptible to inhibition by Hib than wild type strains.

Hib is already known to inhibit the growth of other *Hi* strains (NTHi and non type b capsular serotypes) by releasing a bacteriocin termed haemocin (Venezia & Robertson, 1975). The mechanism of inhibition has not been well characterised but may involve prevention of DNA

synthesis (Streker et al., 1978). The results presented here likely indicate that *Hp* is also sensitive to haemocin.

The ability of Hib to synthesise and export haemocin is conferred by the genes *hmcA-hmcD*. An adjacent gene, *hmcI*, confers immunity to haemocin. The *Hp* T3T1 gene *Hp0778* encodes a protein that shares 35% aa identity with HmcI from Hib strain E1a and does not have any other close BLAST matches. *Hp0778* homologues were detected in all 18 true *Hp* strains, but not the two hybrid strains, by PCR analysis (data not shown). However, it is clear from Fig. 6.13 that this protein does not confer protection against haemocin, at least under laboratory conditions. The sequence similarity between HmcI and Hp0778 is intriguing and it is possible that the latter may protect *Hp* against a related bacteriocin or antibiotic produced by another species in the same niche, for example actinobacillicin which is produced by *Aggregatibacter actinomycetemcomitans* (Hammond et al., 1987).

It is not known whether *Hp* synthesises a bacteriocin to prevent the expansion of *Hi* strains that are attempting to colonise the same niche. To test this, 3 µl circles of five *Hp* strains (*Hp* 13, 20, 22, 26 and 35) were pipetted onto a lawn of *Hi* strain 375 or Rd. After overnight incubation, no zones of inhibition were observed around any of the ten circles of *Hp* growth (data not shown). This suggests that *Hp* generally does not express bacteriocins that are inhibitory to *Hi* under laboratory conditions.

6.5 Discussion

6.5.1 Part I: Serum resistance and epithelial association

The results presented here demonstrate that cell surface structures, in particular OAg, can have a profound effect on interactions between *Hp* and host cells and molecules. Most of the *Hp* study strains do not produce large amounts of polymeric OAg in vitro; for group 1 strains, it is more common to produce LPS apparently containing a single O-unit, whilst the majority of group 2 strains produce a barely detectable amount of OAg-containing glycoforms. If this is reflected in vivo it would suggest that the optimal length and amount of OAg is strain-specific, perhaps depending on the O-unit structure, the identity of any outer core LPS structures, and the existence and properties of other surface molecules such as adhesins and pili. We observed that even a small amount of OAg can significantly affect cell surface properties.

Roberts et al. (1986) found that 22 of their 25 *Hp* isolates were resistant to killing in pooled normal human serum, where resistance was defined as >90% survival after a 40 min incubation in 20% serum. Although the authors state that no correlation was found between serum resistance and the type of LPS produced, the LPS profiles of all three susceptible strains were listed as 'rough', indicating a lack of polymeric OAg. Our own observations are consistent with this trend: most of our *Hp* isolates are highly serum resistant, and the three most susceptible strains appear to synthesise a maximum of one O-unit. We detected polymeric OAg in a higher percentage of serum resistant *Hp* strains than Roberts and colleagues, probably because we checked for weak OAg expression using heavily loaded tricine SDS-PAGE gels. In addition, the genetic and structural characterisation of LPS in some of our strains allowed us to perform

assays with isogenic wild type and OAg mutant strains that demonstrated more directly that OAg can confer resistance to the killing effects of human complement.

There are at least two ways in which OAg might prevent MAC formation in the cell membrane. Firstly, long OAg chains could block antibodies (e.g. IgG) that recognise outer membrane proteins such as porins from accessing the cell surface. Secondly, antibodies binding to the end of the OAg chain could trigger complement deposition away from the bacterial membrane or again prevent the access of other antibodies to the membrane. High concentrations of anti-OAg IgG in HIV patients are known to inhibit the complement-mediated killing of *Salmonella enterica* sv. Typhimurium, making these individuals particularly susceptible to fatal *Salmonella* infections (MacLennan et al., 2010). We do not know whether the pooled human serum used in our bactericidal assays contained antibodies specific for any of the *Hp* OAg structures. *Hp* colonisation of the upper respiratory tract is very common, and OAg can be highly immunogenic, but the high level of structural variation between the OAg of different *Hp* strains would decrease the chance of specific antibodies being present. This could be tested in the future by incubating *Hp* cell lysates with pooled or individual human serum, separating the LPS by SDS-PAGE and hybridising the blotted gel with an anti-human secondary antibody to see if particular OAg-containing glycoforms are bound by human antibodies.

Although our experiments indicate that most *Hp* strains are highly resistant to lysis mediated by the MAC, we did not test their resistance to opsonophagocytosis. Both of these properties are required if a bacterium is to survive in the bloodstream and cause invasive disease in an immunocompetent patient.

For some Gram negative species, the presence of OAg on the LPS appears to decrease association with host cells. Holzer and colleagues (2009) found that mutation of the OAg ligase in *Salmonella enterica* sv. Typhimurium resulted in a five-fold increase in adhesion to HeLa cells and also increased invasion. In *Burkholderia cenocepacia*, a pathogen often isolated from cystic fibrosis patients, the expression of OAg did not affect adhesion to macrophages but decreased adhesion to bronchial epithelial cells ten-fold (Saldias et al., 2009). In cases such as these it is usually postulated that OAg may prevent attachment by masking surface adhesins; this could indeed be true for *Hp* strains T3T1 and 20, which attach more readily to epithelial cells in the absence of OAg.

In contrast, the OAg of other species have been shown to contribute to adhesion. Human epithelial cells secrete a multipurpose β-galactoside-specific lectin called galectin-3 which binds to the epithelial surface and the extracellular matrix (Dumic et al., 2006). Therefore LPS structures that contain terminal β-galactoside, such as certain OAg in *Helicobacter pylori* and the outer core in *Neisseria gonorrhoeae*, promote bacterial-host cell adhesion by binding to galectin-3 (Fowler et al., 2006; John et al., 2002).

The OAg of *Hp* strain 13, but not of *Hp* T3T1 or 20, contains a putatively terminal β-galactoside structure (Figs. 5.2 and 5.5; note that the GalNAc and Glc residues in the *Hp* 20 structure are linked through a phosphodiester bond so do not constitute a glycoside). Only in *Hp* 13 did the OAg appear to promote, rather than reduce, adhesion to epithelial cells. We therefore propose that galectin-3 on the epithelial surface may bind specifically to *Hp* strains whose OAg contains

terminal β-galactoside, such as *Hp* strain 13, thereby mediating bacterial-host cell adhesion. Indeed, the *Hp* 13 OAg backbone is poly-N-acetyllactosamine, which (albeit with different linkages) was the oligosaccharide that bound galectin-3 most strongly from a panel of 41 potential ligands in a study by Hirabayashi and colleagues (2002).

In summary, a wide range of *Hp* OAg structures appear able to confer resistance to complement-mediated killing in vitro, whilst the effect of OAg on the association of *Hp* with epithelial cells can be positive or negative depending on the OAg structure and is putatively related to the presence of a terminal β-galactoside epitope.

6.5.2 Part II: Phosphocholine

PCho is considered to be an important virulence factor in several respiratory tract colonising species. It has been identified as a component of the LPS in at least four species that are closely related to *Hp*: the cattle respiratory pathogen *Histophilus somni* (Elswaifi et al., 2009), the human periodontal pathogen *Aggregatibacter actinomycetemcomitans* (Schenkein et al., 2000), the human respiratory tract commensal species *Haemophilus haemolyticus* (McCrea et al., 2010), and *Hi* (Weiser et al., 1997). However, we found no evidence for PCho expression in any of the *Hp* study strains. Based on our knowledge of the role of PCho in *Hi*, this could potentially affect interactions with host cells and molecules in both positive and negative ways, as outlined below.

There is evidence to suggest that PCho allows *Hi* to associate with epithelial cells by binding to the receptor for a human PCho-containing molecule, platelet-activating factor (PAF). For certain *Hi* strains, $PCho^+$ bacteria adhered to and invaded 16HBE14 cells significantly more than $PCho^-$

mutants, and were found to co-localise with the PAF receptor on the epithelial surface (Swords et al., 2000). The context of the PCho epitope within the LPS structure, i.e. its attachment position and the other substituents that are present, may be important for PCho-PAF receptor interactions as a role in adhesion is not always observed (Fox et al., 2008). When we allowed *Hp* strain 20 to express PCho on its LPS using a transgenic *lic1* locus, attachment to epithelial cells remained unchanged, and in the absence of OAg it even appeared to decrease. This might suggest that at least in terms of host cell adhesion for colonisation of the respiratory tract, there is no advantage to be gained from *Hp* expressing PCho.

The invasion of epithelial cells can lead to pathogenic behaviour by *Hi*. In some instances, invasion is more frequent when *Hi* expresses PCho (Swords et al., 2000), but it is not clear whether this internalisation is simply a result of increased adhesion or specifically requires interaction with the PAF receptor. If the latter is true, *Hp* may be less able to invade epithelial cells than many *Hi* strains. This would agree with the data presented in Fig. 6.10b.

Lysenko et al. (2000b) demonstrated that the proportion of *Hi* PCho^{+} phase variants increases in the presence of cathelicidin LL-37/hCAP18, an antimicrobial peptide expressed by many host cells including the epithelial lining of the respiratory tract. PCho is zwitterionic and the positively charged choline moiety is predicted to project outwards from the cell, potentially repelling the cationic cathelicidin and preventing it from disrupting the cell envelope (Lysenko et al., 2000b). Although we did not test the resistance of *Hp* to antimicrobial peptides, it is plausible that OAg plays a similar protective role to that of PCho, altering the cell surface charge profile and/or blocking access. Of three *Proteus mirabilis* transposon mutants selected for their

sensitivity to the antimicrobial peptide polymixin B, two lacked OAg, suggesting that this structure can indeed provide resistance in some species (Gunn, 2001).

Despite the benefits of cell surface PCho for epithelial cell association and cathelicidin resistance, it can also be detrimental to survival due to its recognition by antibodies and C-reactive protein (CRP) (Weiser et al., 1998b). CRP is an acute phase reactant which activates the complement cascade by binding C1q; it is mainly found in serum but is also present in the respiratory tract.

For our experiments in *Hp* we chose to use a *lic1D* allele that would add PCho to a HepI extension because in *Hi* this position is twice as common as from HepIII (McCrea et al., 2010), making our results comparable to a larger proportion of *Hi* strains. We found no difference in serum resistance between wild type *Hp* strain 20 and the three 'typical' *Hp* 20 *lic1*$^{+}$ clones, suggesting that the incorporation of PCho into some of the LPS molecules has not made the bacteria more susceptible to CRP binding and serum killing in our assay. In *Hi*, the TEPC15 antibody can bind to any PCho within LPS, but CRP can only access PCho that is linked to a HepIII extension rather than a HepI extension (Lysenko et al., 2000a), possibly due to steric hindrance of the large CRP molecule. Our results for *Hp* therefore correlate with previous findings for *Hi* in this respect. To study the effect of CRP binding on a HepIII-bound PCho epitope in *Hp*, we would need to construct a plasmid with a different *lic1D* allele.

The successful introduction of a functional *lic1* locus into *Hp* illustrated a number of points. It showed that *Hp* is capable of taking up *Hi* DNA and integrating it into the genome, at least in the

laboratory using linearised plasmid DNA and targeted flanking sequences. Moreover, *Hp* can express *Hi* genes as functional proteins. This is not surprising, as it was already apparent from Chapter 5 that *Hp* can take up and express LPS-related genes from other species, but it can now be ruled out as an explanation for the lack of naturally occurring *lic1* genes in *Hp*.

It seems likely that the perceived advantages of PCho expression in *Hi*, namely host cell adhesion and resistance to particular antimicrobial compounds, can be obtained by the expression of certain OAg structures in *Hp* with the added benefit of resistance to complement-mediated killing. Our preliminary epithelial invasion assay indicated that *Hp* may be less able to invade epithelial cells than *Hi*; this could be a side effect of both PCho absence and OAg presence. This would still allow effective colonisation of the nasopharynx whilst reducing the virulence potential of the organism.

7 Results: The variation of LPS structure in *H. parainfluenzae*

Phase variation mediated by simple sequence repeats (SSR) is one of the key mechanisms used by *Hi* to adapt its cell surface to a changing host environment. In a study of 16 *Hi* genomes, our laboratory used an algorithm called *Hi*SSRfinder to show that each strain contained 8-20 tetranucleotide repeat tracts of at least 20 bp in length (Power et al., 2009); the longest tract was 204 bp (51 repeats). Conversely, the *Hp* T3T1 genome sequence does not contain any tetranucleotide repeat tracts longer than 12 bp (Peter Power, pers. comm.) and does not contain homologues of any of the phase variable outer core LPS genes found in *Hi* (Chapter 3). In the first half of this chapter, we analyse the potential for phase variable gene expression in the *Hp* study strains, in particular *Hp* 20 and T3T1.

The role of an OAg depends on three key features: its O-unit structure, modal chain length and abundance. We have already demonstrated the high level of interstrain heterogeneity in the O-unit structure of *Hp* LPS. The second half of this chapter explores possible reasons for the differences in OAg expression and chain length that we have observed between strains.

7.1 The capacity for phase variation in *Hp*

7.1.1 *Hp* does not use the same tetranucleotide repeats as *Hi* for phase variation

To test whether the *Hp* strains contain particular tetranucleotide repeats, Southern analysis was carried out on MfeI restriction digested gDNA from the 20 study strains using four different $(NNNN)_5$ oligonucleotide probes which between them represent 91% of the 199 tetranucleotide repeat tracts identified across 16 *Hi* strains. Initial attempts using the AlkPhos Direct detection system (GE Healthcare) were unsuccessful, perhaps because the alkaline phosphatase label could not be incorporated into such a short DNA probe in sufficient quantities to enable chemiluminescent detection above background levels. We therefore radiolabelled the probes with ^{32}P. No hybridising bands were found for *Hp* using the probes $(CAAT)_5$, $(GACA)_5$ or $(CAAC)_5$, contrasting with the strong signals obtained for all three *Hi* control strains (Fig. 7.1). For the $(GCAA)_5$ probe, multiple weak bands were found for several *Hp* strains, indicating the presence of short and/or imperfect GCAA repeat tracts (Fig. 7.1c).

In *Hi*, GCAA repeat tracts are found in two phase variable outer core LPS-related genes, *lex2A* and *oafA*. Repeatless fragments of these genes were used as probes in earlier Southern analysis on the same blot (Fig. 3.1). As described in Chapter 3, the *lex2A/lex2B* and *oafA* probes did not hybridise to the gDNA of any of the 18 true *Hp* strains, suggesting that the weak $(GCAA)_5$ signals do not represent repeat tracts within *lex2A* or *oafA* homologues. Whilst the *lex2A/lex2B* probe did bind to a gDNA restriction fragment from the hybrid strain *Hp* 11, this fragment did not hybridise to the $(GCAA)_5$ probe.

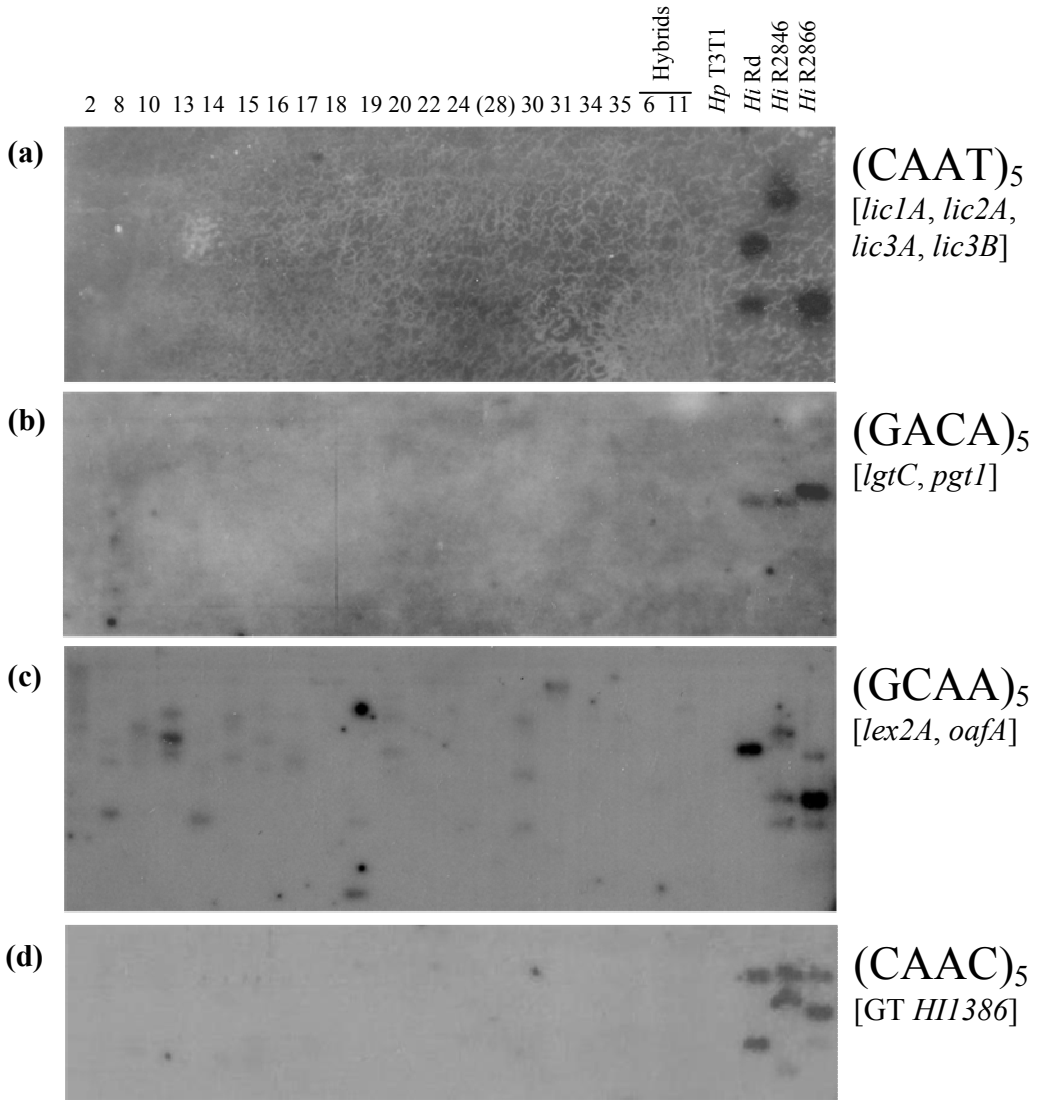


Figure 7.1. Southern analysis of MfeI-restricted gDNA hybridised with ^{32}P-labelled tetranucleotide repeat probes. *Hp* strain numbers are listed above each lane; three *Hi* control strains are shown on the right. Note that strain 28 is not *Hp* and should be ignored. Probe sequences used for each panel are given to the right of the figure, with *Hi* LPS genes that contain this repeat sequence listed in square brackets. In *Hi* strain Rd, $(CAAC)_n$ is also present in four *hgp* genes, which encode haemoglobin and haemoglobin-haptoglobin binding proteins.

Whilst we cannot rule out the possibility that other tetranucleotide permutations may be present as SSR in the *Hp* study strains, the Southern analysis provides evidence that the four repeat sequences that are most commonly used by *Hi* to mediate phase variation are not prevalent in

Hp. Interestingly, only two of the *Hi* phase variable glycosyltransferase genes appear to have homologues in any of the *Hp* strains that we studied (see Section 3.3.6 for homology values), and the *Hp* versions lack tetranucleotide repeats. These two genes, *lic2A* and *lgtC*, were sequenced from *Hp* strain 17 and their translated sequences are shown in Figures 7.2 and 7.3. They are likely to be functional and to add sequential Gal residues of the digalactoside epitope as their *Hi* homologues do, because an antibody specific for this epitope binds to an *Hp* 17 lysate (Fig. 3.6).

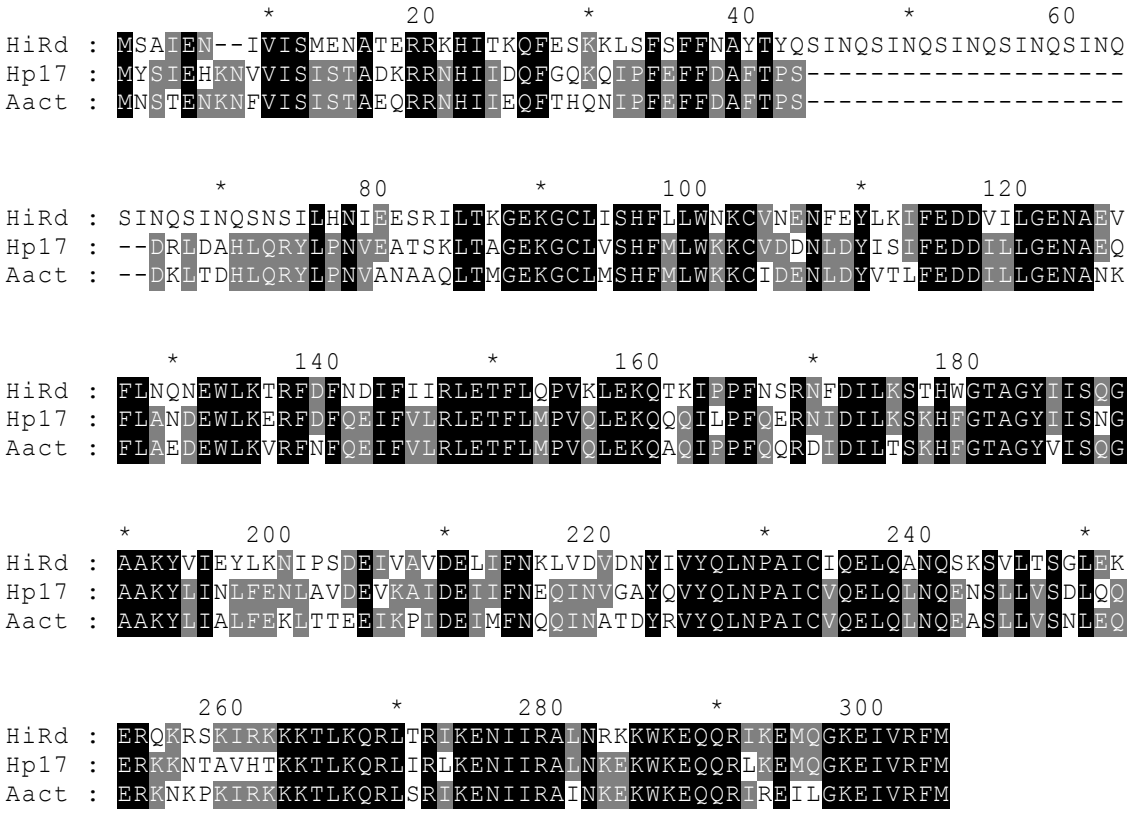


Figure 7.2. Amino acid alignment of Lic2A sequences from *Hi*, *Hp* and *Aggregatibacter actinomycetemcomitans*. HiRd = *Hi* strain Rd (protein HI0550); Hp17 = *Hp* strain 17; Aact = *A. actinomycetemcomitans* strain D11S-1 (protein D11S_0429). The *Hi lic2A* gene contains CAAT repeats, which translate as $(SINQ)_n$ and mediate phase variation of the enzyme; these are absent in the *Hp* 17 *lic2A* homologue.

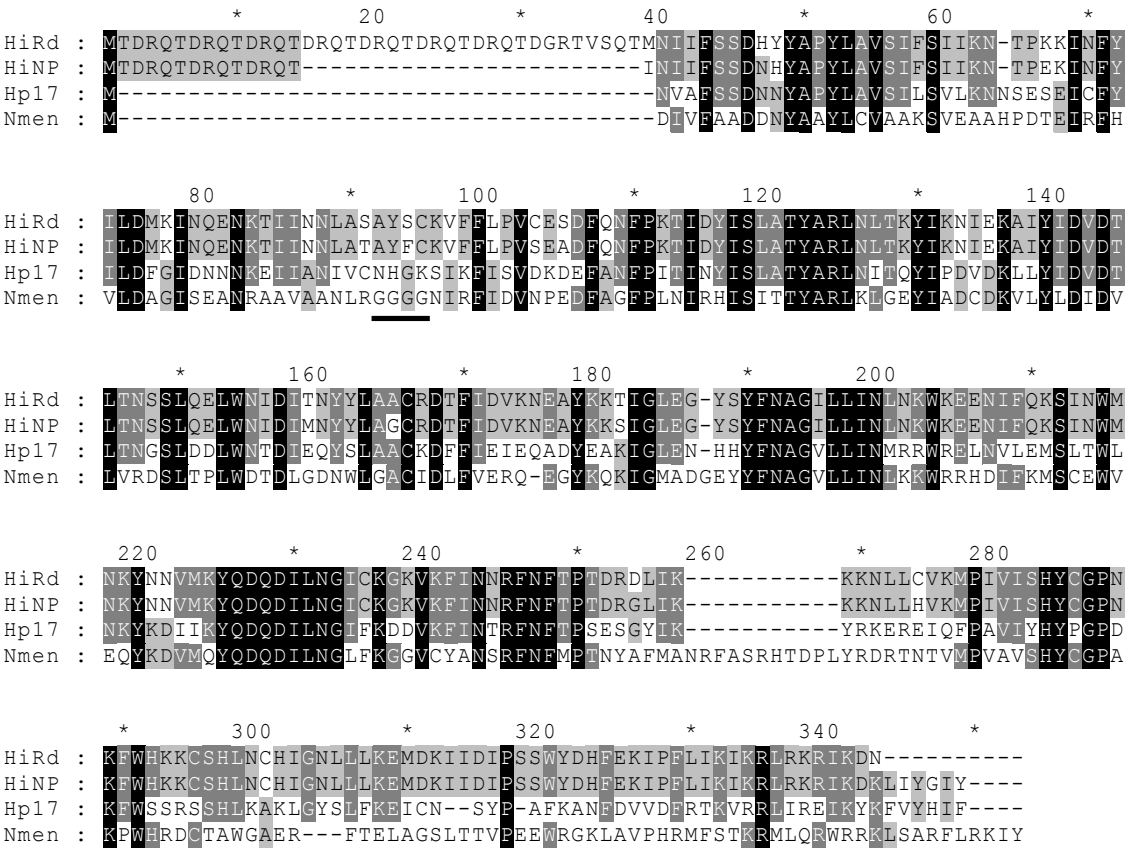


Figure 7.3. Amino acid alignment of LgtC sequences from *Hi*, *Hp* and *Neisseria meningitidis*. HiRd = *Hi* strain Rd (protein HI0258); HiNP = *Hi* strain 86-028NP (protein NTHI0365); Hp17 = *Hp* strain 17; Nmen = *N. meningitidis* strain M978 (protein AAL12839). In *Hi*, LgtC expression is phase varied by the slippage of GACA repeats that follow the start codon; these translate as $(TDRQ)_n$ repeats. In *N. meningitidis*, LgtC expression is phase varied by the slippage of a poly-G tract which translates as glycine repeats (underlined in the figure). The *Hp* strain 17 *lgtC* sequence contains neither of these repeat tracts.

In the *Hi* Rd *lic2A* gene, the $(CAAT)_n$ repeat tract begins 120 bp after the second potential start codon. The *lic2A* genes from *Hp* 17 and *Aggregatibacter actinomycetemcomitans* contain a shorter, non-repetitive sequence in place of these repeats (Fig. 7.2), suggesting that the expression of the enzyme is not phase variable in these two species.

The *Hp* strain 17 homologue of LgtC aligns to the *Hi* and *Neisseria meningitidis* LgtC sequences as shown in Fig. 7.3. The *Hi lgtC* sequence begins 5'-ATGACA$(GACA)_n$-3', so the start codon

is only in frame with the rest of the gene for certain values of n. *N. meningitidis* LgtC is also phase variable but this relies on the slippage of a homopolymeric, rather than tetranucleotide, repeat tract (underlined in Fig. 7.3) (Berrington et al., 2002). In contrast, the *Hp* 17 *lgtC* sequence contains no simple sequence repeats, indicating that its expression is not controlled by phase variation.

Whilst we cannot trace the exact phylogeny of the repeatless *Hp* strain 17 *lic2A* and *lgtC* genes, some scenarios are more likely than others. Two other *Hp* strains contain homologues of three contiguous genes from *A. actinomycetemcomitans*, one of which is the repeatless *lic2A* gene (Fig. 3.2). This was probably the ancestral gene arrangement for *Hp* 17, until two of the genes were replaced with *lgtC*. In some *Hi* strains, *lgtC* contains an ATG codon 21 bp after the repeats (e.g. *Hi* Rd in Fig. 7.3), so *Hp* 17 could have acquired a sufficiently complete *lgtC* gene from *Hi* without inheriting the tetranucleotide repeats. This ATG codon may be functional for translation initiation in *Hp* 17 even if it is not in *Hi* Rd, because the region upstream of it differs between the two species; the RNA secondary structure of the $(GACA)_n$ repeats in *Hi* may preclude the existence of an appropriate ribosome binding site.

Alternatively, *Hp* strain 17 may have acquired *lgtC* and its tetranucleotide repeats from *Hi* and subsequently lost the repeat tract, either because there was no selective advantage in phase varying the digalactoside epitope or because it could not do so in the same way as *Hi*. The latter suggestion is explored in the next section.

7.1.2 *Hp* is capable of phase varying a transgenic tetranucleotide repeat locus

The lack of detectable phase variable loci in *Hp* raised the hypothesis that this species may differ from *Hi* in its ability to replicate DNA repeat tracts accurately. If the replication machinery in *Hp* was so precise that slippage of tandem repeats occurred very rarely or never, the presence of repeat tracts would not mediate phase variation so would not be selectively maintained for environmental adaption. If there was a bias towards the shortening of tandem repeat tracts during replication slippage they would be lost over time; a bias towards increasing repeat tract lengths would also render this phase variation mechanism unfeasible, as the repeats would become increasingly unstable and in addition could alter protein structure and functionality.

To investigate the ability of *Hp* to alter the number of repeats in a tetranucleotide repeat tract, the expression of phosphocholine (PCho) was studied in an *Hp* strain containing a transgenic *lic1* locus. The *Hp* 20 *lic1*$^{+}$ transformants described in Chapter 6 contained the four genes *lic1A-lic1D* targeted into the *tetB* gene; *tetB* had been introduced from *Hp* T3T1 gDNA where it is part of the Tn10 transposon on an integrative and conjugative element (ICE) (Juhas et al., 2007). As the ICE may be capable of switching between chromosomal and plasmid forms, these clones were not ideal for phase variation experiments, where a single-copy stable chromosomal locus is desirable. A different clone was therefore developed in which *lic1A-lic1D* had been targeted to interrupt the *Hp* strain 20 homologue of gene *Hp1877*, known to be a pseudogene in *Hp* T3T1 (pRY27 plasmid construct details are given in Table 2.5). PCho expression in this strain was confirmed by a dot immunoblot using the PCho-specific mAb TEPC15.

The expression of the *lic1* locus was to be observed by colony immunoblotting. To avoid a high background signal, it is important to wash off the majority of the bacteria from each colony so that only the cells directly attached to the nitrocellulose membrane remain. The membrane can then be incubated with the primary antibody. In initial tests, *Hp* 20 wild type and *lic1*$^{+}$ colonies adhered very strongly to the membrane and did not wash off sufficiently during the one hour PBST-BSA blocking step. This was observed for all *Hp* strains tested, even when the step was extended to 18 h, and contrasts with *Hi* strain Rd colonies where the majority of cells dissociate from the membrane within a few minutes. This suggests a fundamental difference in the surface properties of *Hp* and *Hi* cells, at least when grown on agar: adherence to the membrane may be promoted by particular LPS structures, adhesins or pili in *Hp*. The *Hp* colonies could be wiped from the membrane, but on the resulting immunoblots, PCho^{+} and PCho $^{-}$ control colonies (as confirmed by TEPC15 dot immunoblots using cell lysates) bound TEPC15 at a medium level and were indistinguishable.

As described in Section 6.1.5, wild type *Hp* 20 cells remain aggregated in PBS solution whereas *Hp* 20 Δ*wcfS* bacteria form a smooth suspension. When *Hp* 20 Δ*wcfS* colonies were lifted onto nitrocellulose membranes and washed in PBST-BSA, they dissociated at a similar rate to *Hi* Rd. A double mutant strain, *Hp* 20.21, with the genotype *Hp* 20 Δ*wcfS lic1*$^{+}$ was therefore constructed and used in the following phase variation experiments. In this strain the *lic1* locus was again targeted to the *Hp1877* homologue (see Tables 2.5-2.7 for plasmid and strain details).

Approximately 3000 *Hp* 20.21 colonies (750 colonies from each of four BHI agar plates) were lifted onto membranes which were then subjected to immunoblotting using TEPC15 as the

primary antibody. The majority of colonies were reactive and stained dark purple; however, six non-reactive (pale purple) colonies were observed (Fig. 7.4a). These six colonies were designated as 20.21A-20.21F and were streaked onto new BHI plates. Primers lic1a-250 and lic1-CLF were used to PCR amplify and sequence the repeat region of the *lic1* locus from gDNA prepared from these putative phase variants. When the sequences were aligned against the $(CAAT)_{16}$ region of the pRY27 plasmid construct and of the parental *Hp* 20.21 strain, it was clear that all six colonies had gained one CAAT repeat, making the *lic1A* gene out of frame with either of the potential start codons (compare pRY27 and 20.21A in Fig. 7.5). This provided evidence that repeat-mediated slippage can occur in *Hp* 20.21.

One of the six ON to OFF phase variants, 20.21A, was plated and the resulting colonies were immunoblotted as above. 29/4200 colonies were TEPC15-reactive (Fig. 7.4b); the $(CAAT)_n$ repeat tract was sequenced for six of these. Three were found to have gained repeats and three had lost repeats (PV1-PV6 in Fig. 7.5).

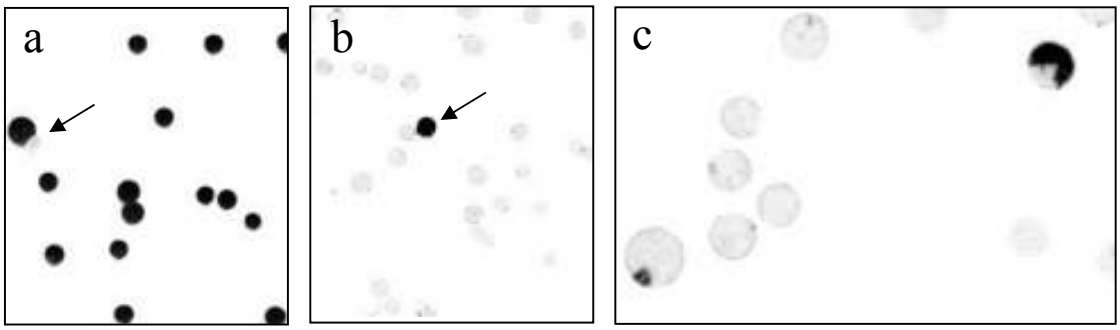


Figure 7.4. Detection of repeat-mediated phase variation in a transgenic *Hp* strain. Colonies were grown on BHI agar and immunoblotted using TEPC15 as the primary mAb to detect the presence of PCho. TEPC15-reactive (dark) colonies are assumed to have a permissive number of CAAT repeats in the transgenic *lic1* locus, whereas non-reactive (pale) colonies are likely to have a non-permissive number of repeats at this locus. **(a)** Clone *Hp* 20.21 (genotype *Hp* 20 Δ*wcfS lic1*$^+$) with phase variant *Hp* 20.21A indicated by an arrow. **(b)** *Hp* 20.21A with a phase variant colony indicated by an arrow. **(c)** Magnified view of sectored colonies derived from *Hp* 20.21A, showing phase variation within a colony.

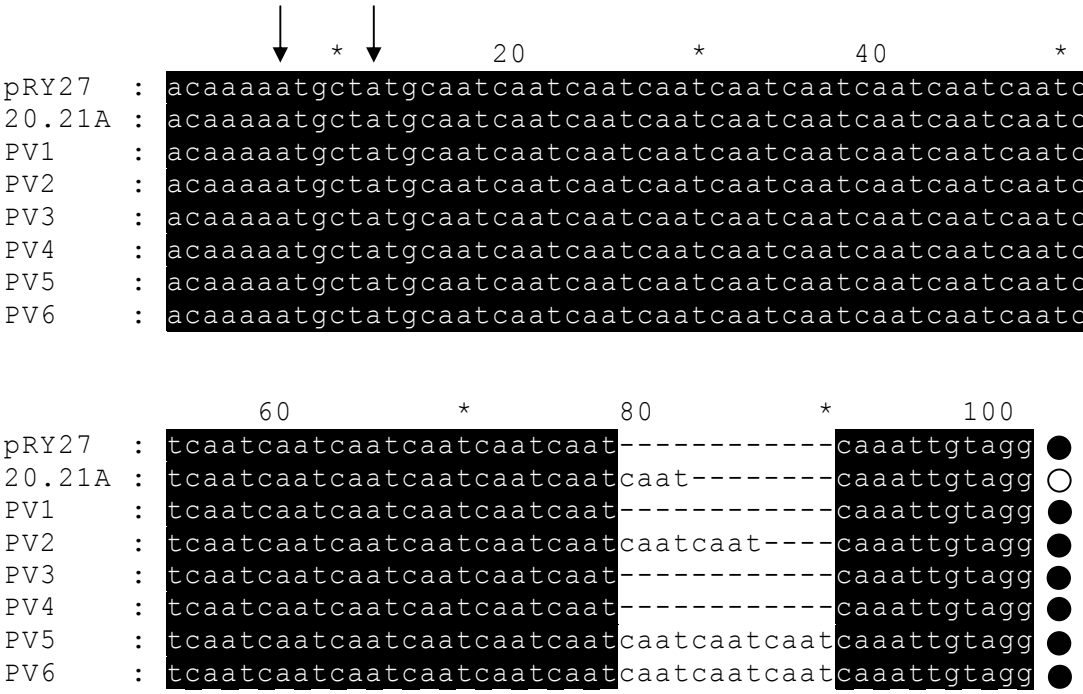


Figure 7.5. Repeat-mediated slippage of the $(CAAT)_n$ region of *lic1A* in phase variants derived from *Hp* clone 20.21. The construct pRY27 was used to introduce a transgenic *lic1* locus into an *Hp* 20 Δ*wcfS* clone to generate clone 20.21. In both pRY27 and clone 20.21, *lic1A* contains 16 CAAT repeats, which is permissive for protein expression. An ON to OFF phase variant of clone 20.21 contained 17 repeats (20.21A). Six OFF to ON phase variants of clone 20.21A contained 16, 18 or 19 repeats (PV1-PV6). The start codons for the two frames that allow Lic1A expression begin at the positions shown by arrows. Phenotypes are indicated by filled circles (TEPC15 reactive) or open circles (TEPC15 non-reactive).

In *Hi* strain Rd, three clear levels of TEPC15 reactivity (pale, medium and dark) can be observed in colony immunoblots, and the latter two are presumed to be the result of different translational efficiencies for the two permissive reading frames for *lic1A* expression. However, for *Hp* we found only two reactivity levels, pale (non-permissive reading frame) and dark (either of the permissive reading frames). Without structural LPS analysis we cannot tell whether these results correspond to two or three levels of PCho expression. PCho on the *Hp* LPS may be more accessible to TEPC15 due to the presence of fewer outer core substituents than in *Hi*.

These experiments show that *Hp* is capable of phase variable PCho addition to its LPS when provided with a transgenic *lic1* locus containing 16 CAAT tandem repeats. Although accurate slippage rates were not calculated, phase variants were detected at a similar frequency in *Hp* 20.21 and an *Hi* Rd control; both the gain and loss of repeats were observed, showing that the process is reversible. This suggests that there is no fundamental reason why *Hp* could not use tandem repeat slippage as a mechanism of phase variation.

7.1.3 Non-tetranucleotide simple sequence repeats in *Hp* T3T1

Although tetranucleotide repeat tracts are the most common form of simple sequence repeat associated with phase variation in *Hi*, at least one dinucleotide and one pentanucleotide repeat tract have been shown to direct phase variable gene expression in this species (van Ham et al., 1993; Zaleski et al., 2005) and it is thought that repeat motifs of other lengths may also be used occasionally (Power et al., 2009). Analysing the presence of non-tetranucleotide repeat tracts in *Hp* could help to establish whether this species ever employs a repeat-mediated slippage mechanism to phase variably express cell surface structures or if there is selection against repetitive DNA. The *Hi*SSRfinder algorithm was used to scan the *Hp* T3T1 genome for simple sequence repeats (Table 7.1; Peter Power, pers. comm.); in addition, the *Hp* T3T1 homologues of some of the putative phase variable genes from *Hi* were analysed to determine whether the repeats were intact, absent or replaced with non-repetitive DNA. The results for mono-, di-, penta- and heptanucleotide repeats are described here.

Table 7.1. Simple sequence repeats in the *Hp* T3T1 genome sequence. The number of repeat tracts with each unit length was determined using the minimum number of repeats given in the second column as a threshold. The third column lists the range in the number of repeat tracts for four *Hi* strains (Rd, R2846, R2866 and 86-028NP) to allow comparison between *Hi* and *Hp*. Data for repeat unit lengths of 1-2 was obtained from a manual search for repeats using Artemis. Data for repeat unit lengths of 3-9 was produced using the *Hi*SSRfinder algorithm by Peter Power. The threshold repeat numbers were as determined empirically by Power et al. (2009).

Repeat unit length	Min repeat number	No. in *Hi*	No. in *Hp* T3T1
1 (mono)	9	13-18	15
2 (di)	5	4-6	5
3 (tri)	4	8-13	11
4 (tetra)	3	12-14	0
5 (penta)	3	3-5	3
6 (hexa)	3	3-5	2
7 (hepta)	3	0-4	0
8 (octa)	3	0-1	0
9 (nona)	3	1-2	2

15 mononucleotide repeat sequences of 9 bp or longer are present in the *Hp* T3T1 genome. These were identified using a manual search in the Artemis sequence viewer and comprise one C_{11} tract, one A_{10} tract and 13 A_9 or T_9 tracts. The two longest repeats are unlikely to mediate phase variation, as the C_{11} tract is located between two convergently transcribed genes (*Hp1710* and *Hp1711*) and the A_{10} sequence is in the 3' half of gene *Hp0123* whose homologue in *Hi* contains no repeats. Despite the fact that 89.8% of the genome putatively encodes ORFs, only two of the 13 A_9 or T_9 tracts were found within reading frames: these were both at the 5' end of genes within the ICE (*Hp1147* and *Hp1150*), encoding a site-specific recombinase and a putative DNA binding protein respectively. The *Hp1147* homologues in *Hi* strains R2866 and 86-028NP contain a shorter repeat tract at the same position (T_7 or T_8 rather than T_9), which we would expect to result in a frameshift, although it is also possible that the real translational start codon is located after the repeats. There is no obvious reason why it would be beneficial to phase vary these ICE-encoded genes.

Whilst the slipped-strand mispairing of repeat tracts within ORFs causes frameshifts that determine whether a gene can be translated correctly, simple sequence repeats in a promoter or operator region can control the level of transcription. One of the A_9 tracts in *Hp* T3T1 is between the divergently transcribed ORFs *Hp0065* (GlcNAc-6-phosphate epimerase gene, *nanE*) and *Hp0066* (encoding a putative GlcNAc-specific transport protein not found in *Hi*). Another is between the divergently transcribed *glpT* and *glpA* genes, which are involved in glycerol metabolism; *glpQ*, which encodes a surface protein termed protein D, is also part of this regulon. However, both of these repeat tracts are surrounded by other, shorter poly-A:T sequences and both of the regulons are induced by sugar phosphates and repressed by specific transcription factors (Jenkins et al., 2010; Johnston et al., 2010), so it is possible that these methods of regulation simply require a particular DNA curvature that is provided by sequences that are rich in poly-A:T. Other reasons that poly-A:T repeat tracts are more common in non-coding regions than within genes include the overall lower % GC content of these regions, lack of functional constraint on the sequence, and the use of poly-uracil sequences after stem-loops to aid transcription termination (e.g. after gene *Hp1405*).

In *Hi*, the dinucleotide repeat tract which overlaps with the divergent *hifA* and *hifB* promoters is usually 9-12 repeats long and has been shown to mediate the phase variable expression of fimbriae (Bayliss et al., 2004; van Ham et al., 1993). The *hif* locus is absent in *Hp* T3T1. We identified five dinucleotide repeat tracts at least 10 bp long in the *Hp* T3T1 genome: four are of the minimum length (five repeats) and within essential genes, so are unlikely to mediate phase variation. The fifth tract, $(AT)_6$, is 180 bp into the OAg polymerase gene *Hp0277*. Whilst this

repeat tract is also relatively short, it is possible that slipped-strand mispairing may occur occasionally; this would result in a single O-unit being added to the LPS. As this phenotype is observed for several group 1 *Hp* strains, their *wzy* sequences were investigated (Section 7.2).

Power et al. (2009) identified five heptanucleotide or pentanucleotide repeat tracts in *Hi* that could potentially mediate phase variation. The *Hp* T3T1 genome does not contain homologues of the three *Hi* genes with heptanucleotide repeats (*hmw1a*, *hmw2a* and *HI1369*), but homologues of the two genes with pentanucleotide repeats (*wzx* and *hsdM*) are present and their sequences were examined.

The *Hi* HMG flippase gene, *HI0867* (*wzx*), contains $(TCAGC)_n$ repeats 350 bp into the ORF; a repeat length of n = 2 (as found in most of the genome sequences) is permissible for translation, whereas n = 3 (as found in the R2866 genome sequence) is not. The propensity of this short repeat tract to undergo slippage during replication has never been tested, but it could potentially cause phase variable expression of the entire HMG unit on the LPS. In the *Hp* T3T1 OAg flippase gene (*Hp0279*), which shares 65% identity with *HI0867*, the repeat region is replaced with a non-repetitive ten base pair sequence (data not shown) and there are no other repeat tracts within the ORF. This indicates that the OAg flippase cannot be phase variably expressed using a repeat-mediated mechanism.

HsdM, part of the HindI type I restriction/modification locus in *Hi*, contains phase variable pentanucleotide $(ACGAG)_n$ repeats 5 bp after the ATG start codon (Zaleski et al., 2005). The *Hp* T3T1 homologue of *hsdM* was identified by a BLASTP search and examined for simple

sequence repeats. Hp1588 has 93% aa identity to *Hi* 86-028NP HsdM3 (YP_249266.1); however, alignment of the DNA sequences reveals that in lieu of $(ACGAG)_n$ repeats *Hp1588* contains a non-repetitive ten base pair sequence, placing the start codon in frame with the rest of the coding sequence (data not shown). To investigate whether *Hp* T3T1 is typical of *Hp* in this respect, the *Hp1588* sequence was queried against the Genomophilus database. Whilst homologues were identified for most *Hi* strains in the database, showing that this region of DNA is not difficult to sequence, only two of the 12 *Hp* strains had significant matches. The *hsdM* homologues were from *Hp* strains MO-099-5L and MO-099-8M (two isolates that are not part of our study collection) and contained pentanucleotide repeats that differed at one position to the *Hi* motif: $(ACGAC)_n$. The number of repeat units is unknown as the contigs end within the repeat region, but is at least four.

Of the three pentanucleotide repeat tracts found in *Hp* T3T1, two were only three repeats long and were upstream of putatively essential genes. The third was an unusually long repeat tract, $(AAACT)_{10}(AAAACT)_9$, which was downstream of *Hp0861* (a putative adhesin gene) and ended 17 bp upstream of *Hp0862* (encoding a putative family 8 glycosyltransferase with 33-38% identity to proteins in Pasteurellaceae and Streptococcaceae species). We attempted to PCR amplify these genes and the region between them from the other *Hp* study strains but no products were obtained, suggesting that they are not highly conserved. It is tempting to speculate that the adhesin encoded by *Hp0861* may be glycosylated by Hp0862. Alternatively, Hp0862 may glycosylate the LPS, but this seems unlikely since we have accounted for all the transferases needed for the synthesis of the known *Hp* T3T1 LPS structure. In either case it is not clear whether slippage of the repeat tract could lead to phase variable transcription of this gene; it is

too long to fall between the -35 and -10 elements of a promoter, but it might influence the level of transcriptional readthrough from *Hp0861* or the binding of transcription factors.

7.1.4 Analysis of potential OAg phase variation mechanisms

A few Gram negative species have been found to phase vary their OAg expression using various strategies. The *Helicobacter pylori* fucosyltransferase gene *futC*, whose product converts the Lewis X OAg structure into Lewis Y, contains a homopolymeric repeat tract which undergoes slipped strand mispairing (Sanabria-Valentin et al., 2007). In contrast, seven of the ten OAg and capsule loci in *Bacteroides fragilis* strain NCTC 9343 begin with invertible promoters which can be flipped by an invertase to switch transcription on and off (Cerdeno-Tarraga et al., 2005). We therefore examined the five sequenced *Hp* OAg loci (those of *Hp* strains T3T1, 13, 17, 20 and 30) for any evidence of these phase variation mechanisms.

Analysis with an EMBOSS program, einverted, found that there were no inverted repeats between the upstream flanking gene and the first gene of the locus for any of the *Hp* OAg loci, suggesting that *Hp* does not use an invertible promoter mechanism to phase vary the OAg. The Msatfinder v.2.0.9 algorithm (similar to *Hi*SSRfinder) was then used to scan each OAg locus for mono-, di-, tetra- and pentanucleotide repeat tracts using the same repeat number cutoffs as in Table 7.1. Apart from the dinucleotide repeats in the *Hp* T3T1 OAg polymerase gene (Section 7.1.3), the only tracts identified were $(TTTA)_3$ repeats within the 3' end of genes *13G* and *20B*. These were found in areas of low % GC content and are very short compared to the tetranucleotide repeat tracts that usually mediate phase variation in *Hi*. During the sequencing of the *Hp* strain 13 OAg locus it was observed that gene *13G*, which encodes a putative

glycosyltransferase, was split across two reading frames; the frameshift occurred within a homopolymeric tract of nine guanine residues 200 bp into the gene. However, PCR and resequencing across the repeat tract revealed that in the original *Hp* 13 gDNA the repeat is only 8 bp long and the gene is intact, suggesting that the repeat-mediated slippage had occurred either during LR-PCR with Expand polymerase or during replication of the *E. coli* clone containing the plasmid insert with this sequence. Whilst in *N. meningitidis* the G_5-G_{17} tract in the *lgtA* gene is known to mediate phase variation (Berrington et al., 2002), we consider the G_8 tract in gene *13G* unlikely to do the same for two reasons. Firstly, the *Hp* T3T1 genome contains over 100 A_8 tracts, so widespread genomic instability would occur if homopolymeric 8 bp repeats were highly mutable in this species. Secondly, PCR amplification and sequencing revealed that the *13G* homologues in *Hp* strains 6 and 18 do not contain repeat tracts (data not shown). Our hypothesis could be checked in the future by screening *Hp* 13 colonies with an antibody against the OAg, should one become available.

7.2 The regulation of OAg chain length in group 1 *Hp* strains

Of the seven *Hp* strains with a group 1 type OAg locus, polymeric OAg is visible on an LPS gel for only three: *Hp* 2, 20 and T3T1. For the other four strains (*Hp* 19, 22, 24 and 31) the molecular weight of the major glycoform is consistent with inner core LPS plus a single four-sugar unit. We have structural evidence that *Hp* strain 22 synthesises LPS containing an O-unit similar to that of *Hp* T3T1 (Fig. 5.5). The alteration of LPS profiles upon neuraminidase treatment suggested that the major glycoforms of *Hp* strains 24 and 31 also contain an O-unit, because the only sialic acid synthesis/transfer genes found in these strains were within the OAg

locus (Section 5.2.1). As outlined in Section 6.1.5, we believe *Hp* strain 19 to also produce some LPS containing one O-unit.

To investigate whether alternative in vitro growth conditions could induce the production of longer OAg chains through the provision of a different range of nutrients, we grew *Hp* strains 19 and 24 on a range of solid media including BHI, chemically defined media (CDM) with glucose, blood agar and chocolate agar, each supplemented with NAD and haemin. In addition, strains 19, 22, 24 and 31 were grown on BHI agar with added Glc, Gal, GlcNAc and/or Neu5Ac as these sugars are likely to be precursors for OAg synthesis in group 1 strains. The tricine SDS-PAGE LPS profiles of the four *Hp* strains when grown on different media or sugars were very similar to the profiles from growth on BHI agar, i.e. the most prevalent glycoform was consistent with inner core LPS plus four sugars, and no polymeric OAg was visible (data not shown). Whilst it is still possible that these strains can produce polymeric OAg in vivo, an alternative hypothesis is that they are not genetically capable of polymerising O-units: this has already been observed for the *Hi* HMG (Hood et al., 2004b).

In order to produce an O-unit, most of the genes in the OAg locus would need to be intact and expressed. The two exceptions to this are the OAg polymerase gene, *wzy*, and the chain length determinant gene, *wzz*. To investigate whether *wzy* and *wzz* are present and intact in group 1 strains that only appear to elaborate their LPS with a single O-unit, we attempted to amplify and sequence these genes from *Hp* strains 19, 22, 24 and 31.

A variety of primer pairs designed to *wzy* and its flanking regions in *Hp* strains T3T1 or 20 were used for PCR amplification with template gDNA from the four test strains. Only *Hp* strain 19 yielded any PCR products, consistent with the Southern analysis in Fig. 4.5. Upon DNA sequencing and contig assembly, we found that the translated *Hp* 19 *wzy* sequence was full-length and differed from that of *Hp* T3T1 at only eight amino acid positions (98% aa identity). Reverse transcriptase PCR using cDNA synthesised from *Hp* strain 19 total RNA demonstrated that *wzy* is transcribed in this strain (data not shown).

The lack of a functional chain length determinant protein, or differences in its amino acid sequence, could be responsible for the observed LPS structures for *Hp* strains 19, 22, 24 and/or 31. To begin to test this hypothesis, six primers designed to *Hp* T3T1 *wzz* were used in PCR with *Hp* strain 19 template DNA. Only primer pair 285a/285d yielded a product; this was sequenced and used to design further primers which enabled the completion of a full-length, intact *Hp* 19 *wzz* sequence. *Hp* 19 Wzz shares only 54% aa identity and 71% similarity with *Hp* T3T1 Wzz (Fig. 7.6) but is predicted to have a similar topology using TMHMM software, with N-terminal and C-terminal transmembrane domains flanking a long periplasmic loop as is typical for polysaccharide co-polymerases (data not shown).

From the alignment in Fig. 7.6, it is clear that the available Wzz sequences from *Hi* and *Hp* separate into two clusters which correlate with the observation of a single O-unit/HMG (*Hi* Rd and *Hp* 19) or polymeric OAg (*Hp* strains 20, T3T1 and 17). The terminal 630 bp (210 aa) of *Hp* 22 *wzz* was also sequenced and falls into the first group, sharing 88% aa identity with the *Hp* 19 version.

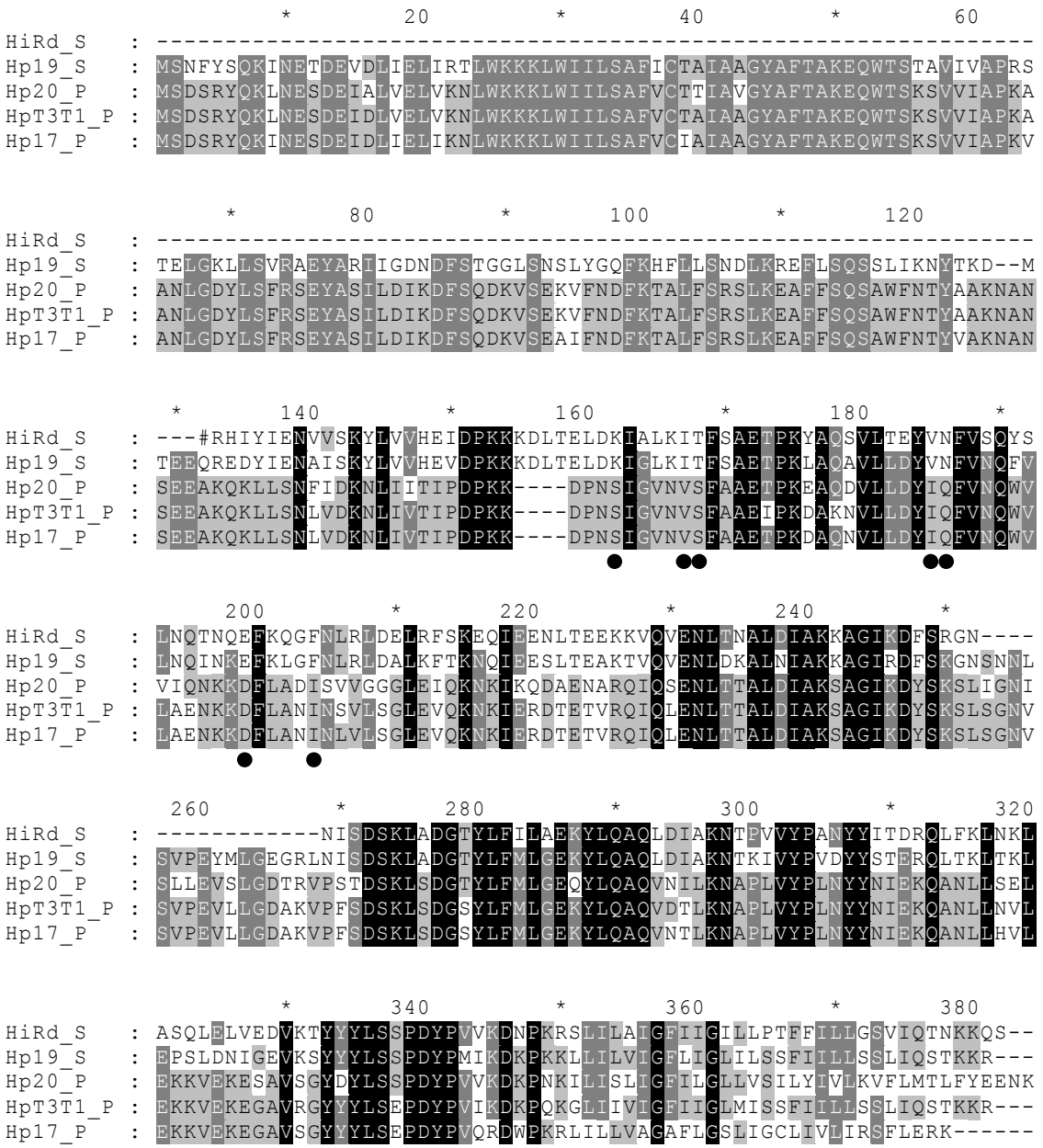


Figure 7.6. Amino acid alignment of chain length determinant (Wzz) proteins. The *wzz* gene was sequenced from *Hp* strains 17, 19 and 20 and the translated sequences were aligned against those of *Hi* strain Rd and *Hp* strain T3T1. The letter after each strain name indicates whether polymeric OAg (P) or only a single O-unit (S) has been observed on the LPS by SDS-PAGE. In *Hi* Rd, *wzz* is a pseudogene containing several frameshifts and premature stop codons, so only the largest fragment of the translated sequence is shown (# = stop codon). See text for the significance of residues marked with black dots.

The sequence differences between the two clusters are mainly in the predicted periplasmic domain; these include seven residues (marked with black dots on Fig. 7.6) that are conserved between Wzz from *E. coli* and from the second *Hp* cluster (strains with polymeric OAg), suggesting that they could be important for the function of the enzyme.

PCR analysis indicates that *Hp* strain 19 contains homologues of all the genes in the *Hp* T3T1 OAg locus, yet no polymeric OAg is observed even when likely competition from the LosB1-dependent extension is removed. This could be due to one or more of the polymorphisms found between their OAg polymerase sequences, but it is more likely that divergence in their chain length determinant enzymes is responsible.

For *Hp* strains 22, 24 and 31, we found no evidence for a *wzy* gene using PCR and Southern analysis; it could be absent, vestigial or simply divergent from the alleles found in *Hp* strains T3T1, 19 and 20. Primers designed to the *Hp* 19 *wzz* gene, but not those designed to the *Hp* T3T1 allele, amplified products for *Hp* strains 22, 24 and 31. This adds weight to the correlation that was observed between Wzz sequence and OAg phenotype above. For these three *Hp* strains, the lack of detectable polymeric OAg could be due to an absent or non-functional OAg polymerase gene and/or a chain length determinant enzyme that specifies the chain length as one unit.

7.3 The regulation of OAg production in group 2 *Hp* strains

Polymeric OAg-like ladder patterns have been observed in the LPS profiles of seven of the eight *Hp* strains with group 2 OAg loci, the exception being *Hp* strain 35. This indicates that the

majority of the group 2 loci are intact. However, the degree of OAg expression on the LPS varies widely between strains, with *Hp* 13 and 15 producing a much greater proportion of OAg-containing glycoforms than *Hp* 8, 10, 14, 16 or 18 (Figs. 3.7 and 4.6). There are several levels at which these differences might be regulated. Firstly, the OAg locus (or the separately encoded OAg ligase gene) might be transcribed more in *Hp* 13 and 15 than in other strains, for example because their promoters are stronger or their regulatory regions bind transcription factors more efficiently. Secondly, prokaryotic translation can be regulated by mRNA secondary structure or mRNA binding proteins, potentially affecting the amount of each of the OAg synthesis enzymes that is produced. Thirdly, different group 2 strains are likely to require different sets of sugars for OAg synthesis, so the amount of OAg may depend on the availability of these precursors during growth on BHI agar.

To test the hypothesis that strains whose LPS contains less OAg have lower levels of mRNA transcript for OAg-related genes, RNA was isolated from four group 2 *Hp* strains and used in reverse transcriptase (RT) PCR experiments. An equal amount of RNA extracted from log phase BHI broth cultures of *Hp* strains 10, 14, 15 and 35 was reverse transcribed to make cDNA, which was then used as the template for standard PCR analysis (Fig. 7.7). The first four primer pairs listed across the top of the figure should amplify internal fragments of four OAg-related genes, whilst the fifth amplifies part of the housekeeping gene *atpG* and serve as a control for equal cDNA concentration. As no product was observed when using the *atpG* primers with *Hp* strain 35 cDNA, we will consider only the results for *Hp* strains 10, 14 and 15.

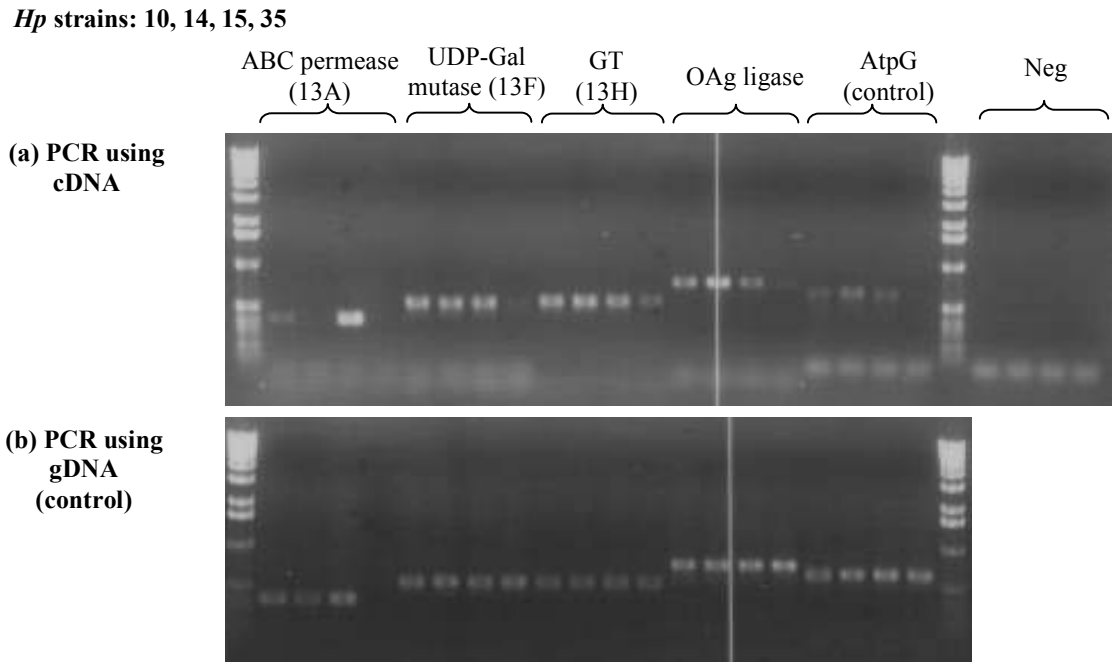


Figure 7.7. RT-PCR of OAg synthesis genes using RNA from four group 2 *Hp* strains. RNA was isolated from log phase cultures and reverse transcribed to make cDNA, which was used as the template in standard PCR. Each curly bracket represents a set of four reactions using **(a)** cDNA, or **(b)** genomic DNA, from (left to right) *Hp* strains 10, 14, 15 and 35. The putative function of the protein encoded by the amplified region is given above each bracket, along with the name of the homologue in *Hp* strain 13. Primers were as follows: ABC permease = gr2-ABC1/hmg13Q; UDP-Gal mutase = gr2-mut1/gr2-mut2; GT = gr2-GT5/gr2-GT6; OAg ligase = LIG15B/LIG15C; AtpG = atpGup/AtpRS. Negative control reactions (Neg) were performed using RNA and the AtpG primer set to confirm that the RNA was not contaminated with DNA. The PCRs using gDNA were included to control for differences in primer binding between strains.

Hp strains 10 and 14 add barely detectable amounts of OAg to their LPS, whilst the LPS profile of *Hp* 15 contains the most OAg of any strain in the study collection. The first three lanes of Fig. 7.7a show a striking correlation between OAg levels and the amount of mRNA transcript that is present for the first gene of the OAg locus, which encodes the permease component of the ABC transporter. Control reactions using genomic DNA as the template (Fig. 7.7b) show that this primer set has approximately equal affinity for the gene in *Hp* strains 10, 14 and 15, so the intense cDNA signal obtained for *Hp* strain 15 is likely to reflect a genuinely high level of

transcript. Conversely, transcript levels for two genes further along the OAg locus (encoding the UDP-Gal mutase and a glycosyltransferase) were equal for all three strains. Levels of transcript for the OAg ligase gene appeared to differ between strains but did not correlate with the amount of OAg.

As the genes between *glnA* and *pepB* in *Hp* 13 (and the other *Hp* strains presented in Fig. 5.10) all appear to be related to OAg synthesis and are all in the same orientation, we might expect them to constitute an operon, i.e. under the control of a single promoter and transcribed together. However, the results presented in Fig. 7.7 dispute this for group 2 OAg loci. Analysis of the *Hp* 13 OAg locus (the only group 2 OAg locus whose sequence is currently available) reveals that the intergenic distance between genes *13B* and *13C* is 111 bp, which is long enough to accommodate a second promoter and regulatory regions. The simplest explanation for our preliminary observations is that in group 2 OAg loci, the genes encoding the two components of the ABC transporter (equivalent to genes *13A* and *13B* in *Hp* strain 13) are transcribed together, with the level of transcript determining the number of ABC transporter complexes and consequently the amount of OAg that is added to the LPS; and that meanwhile the rest of the locus (genes *13C* to *13J* in *Hp* strain 13) is transcribed as an independent operon to the same level across all of the strains. Our preliminary RT-PCR data is a starting point for future detailed investigations into *Hp* OAg expression.

7.4 Discussion

Tetranucleotide repeat tracts are thought to mediate the phase variable expression of four main groups of proteins in *Hi*: outer core LPS synthesis enzymes, YadA proteins (putative adhesins), a

type III restriction/modification system methylase (Mod), and the haemoglobin and haemoglobin-haptoglobin binding proteins (Hgp). In contrast, the few outer core LPS synthesis genes that appear to be present in *Hp* all lack simple sequence repeats. The putative adhesin genes in *Hp* T3T1 are not preceded by repeat tracts, and there is no *Hp* T3T1 homologue of *mod*. The protein encoded by gene *Hp0192* shares 20% aa identity with *Hi* Hgp and 52% aa identity with *Neisseria meningitidis* HmbR (haemoglobin binding protein), but this gene contains neither the tetranucleotide repeats of *hgp* nor the mononucleotide repeats of *hmbR*. We did identify several simple sequence repeats in the *Hp* T3T1 genome sequence, but the tracts were usually short and upon examination were not considered likely to control gene expression. Overall it appears that *Hp* rarely, if ever, uses the repeat-mediated slipped-strand mispairing of DNA to phase vary cell surface structures.

Given that *Hp* and *Hi* both inhabit the upper respiratory tract and are closely related, it seems remarkable that *Hp* has not acquired any tetranucleotide repeat-containing genes for LPS synthesis. We demonstrated that the DNA replication machinery in *Hp* allows a transgenic tetranucleotide repeat tract to both shorten and lengthen at a low rate, as is required for a viable phase variation mechanism, although we did not calculate an exact mutation rate. The repertoire of DNA replication and repair genes is largely conserved between *Hi* and *Hp* T3T1 (Power et al., manuscript in preparation); the main exception is the presence in *Hp* of *dinB*, which encodes an error-prone DNA polymerase. In *N. meningitidis*, the mutation or over-expression of *dinB* does not alter the rate of tetranucleotide repeat-mediated phase variation (Martin et al., 2004), so the presence of this gene in *Hp* does not explain the lack of *Hi*-like phase variable LPS outer core synthesis genes. Perhaps genomic rearrangements and/or DNA divergence between *Hi* and *Hp*

preclude horizontal transfer. However, *Hp* appears perfectly capable of acquiring genes from more distantly related species in order to vary its LPS in another way: the addition of different OAgs.

In this chapter we investigated the reasons for differences in OAg abundance and chain length between *Hp* strains. In *E. coli*, deletion of the *wzz* gene from an OAg cluster results in an OAg of unregulated chain length, with the probability of chain extension decreasing with each extra repeat unit (Franco et al., 1998). Deletion of *wzy* prevents polymerisation of the OAg and results in the addition of a single O-unit to the LPS (Tao et al., 2004). We found differences in both *wzz* and *wzy* between *Hp* group 1 strains expressing polymeric OAg or single O-units. Experiments in which different alleles or fragments of these genes are exchanged between strains (e.g. *Hp* 19 and T3T1), with subsequent analysis of electrophoretic LPS profiles, could help to determine which residues of *wzy* and *wzz* are important for polymeric OAg expression in *Hp*. It would also be interesting to clone and sequence the *Hp* strain 22, 24 or 31 OAg locus to confirm whether *wzy* is absent, present as a pseudogene, or phase variable in these strains, as only a single O-unit has been observed on their LPS.

For group 2 *Hp* strains, we found a correlation between the amount of OAg and the transcription level of the ABC transporter permease gene. The three tested strains were grown under the same conditions, which suggests that genetic or epigenetic differences around the promoter region may be responsible for their different transcript levels. However, we may find upon further investigation that there are additional explanations for the differences in OAg expression between strains. Cuthbertson and colleagues (2007) demonstrated that in *E. coli*, the ATPase

subunit of the OAg ABC transporter (Wzt) determines the specificity of the transporter for a particular OAg structure, whilst the permease subunit (Wzm) is interchangeable between serotypes. If some group 2 *Hp* strains have acquired different OAg glycosyltransferase genes relatively recently, their ABC transporters may not yet be optimised for the new OAg structure, affecting the amount of OAg that is ligated to the LPS core.

We speculated that the *Hp* strain 13 OAg locus may be transcribed as two separate operons. A similar situation has been observed for an *Aeromonas hydrophila* OAg gene cluster, where the ORF at the 3' end (*wzz*) is transcribed separately from the other sixteen ORFs (Jimenez et al., 2008). The intergenic sequence preceding *wzz* contains a putative transcriptional terminator then a promoter. When grown at 20 °C, this bacterium produces abundant OAg with a long modal chain length. At 37 °C, transcription of *wzz* (but not the rest of the OAg gene cluster) is downregulated, and most of the LPS molecules carry 0, 1 or 2 O-units with only a small amount of long-chain OAg produced. This illustrates the concept that the production of OAg, or its addition to the LPS, can be controlled by the transcriptional regulation of a single component of the OAg machinery. Whilst OAgs are unlikely to be regulated by temperature in *Hp*, which is not thought to survive for long outside the human host, factors such as substrate availability might be able to regulate the transcription of a subset of the OAg synthesis genes (for example by the modulation of transcription factor activity through sugar binding).

In genera such as *Escherichia* and *Salmonella*, an elongation factor termed RfaH increases the expression of distal genes in certain long operons, particularly those acquired by horizontal transfer (Belogurov et al., 2010). As there is no RfaH homologue in the *Hp* T3T1 genome, *Hp*

might have difficulty in transcribing the distal genes of the OAg locus at a high level. Most of the genes in the *Hp* 13 OAg locus are very close together, but the previously mentioned 111 bp gap between genes *13B* and *13C* and another 77 bp gap between *13F* and *13G* may provide additional RNA polymerase binding sites to allow transcription of the full locus.

The results presented in this chapter suggest that whilst *Hp* is capable of phase variably expressing transgenic loci by repeat-mediated slippage, it has no need to employ this mechanism for the successful commensal colonisation of the upper respiratory tract. Instead, OAg structural variation between strains presumably helps it to evade the host immune response, whilst a combination of genetics and transcriptional control regulates the length and abundance of the OAg in order to modulate the properties of the bacterial cell surface.

8 Final discussion

LPS is one of the main structural components of the outer membrane in Gram negative bacteria, and this exposed position makes it a key determinant of interactions with the host. Whilst the structure and biosynthesis of LPS has been extensively characterised for *Hi*, very little was known about this molecule in *Hp*. Through studying 18 *Hp* strains we were able to establish a broad picture of the genetic blueprint for LPS biosynthesis in this species, assess the level of LPS structural variation, and gain some insight into the biological relevance of different components of the LPS molecule. Our results show that the LPS of a typical *Hp* isolate comprises an *Hi*-like lipid A and inner core, a few non-phase variable outer core sugars, and in some glycoforms an OAg chain or single O-unit extending from the Glc residue on HepI. Consequently there are major differences between the most exposed parts of the LPS molecule in *Hp* and *Hi*, as the latter lacks OAg and synthesises a highly phase variable outer core. These distinctions at the bacterial-host interface may have significant implications for the establishment and outcome of upper respiratory tract colonisation by these two closely related *Haemophilus* species.

The LPS inner core of most Gram negative bacteria contains Kdo and a number of heptose residues, indicating the important role of these components in outer membrane integrity, but the exact structure and presence of substituents vary (Raetz & Whitfield, 2002). The current study demonstrates that the biosynthetic pathway for the LPS inner core structure found in *Hi* is also conserved in *Hp*, suggesting that it was established before the divergence of the two species and evolved to be beneficial in the respiratory tract environment that they share. Genetic analysis

suggests that *Hp* may add a second Kdo residue to some LPS molecules, although this has not yet been detected using chemical and structural techniques.

Of the 13 other Pasteurellaceae species for which genome sequences are available, 11 have a homologue of *orfH* and 12 have a homologue of *lgtF*; this suggests that in most of these species the LPS contains a triheptosyl inner core with a Glc residue on HepI, as is found in *Hi* and *Hp*. The role of the Glc residue in virulence behaviour has not been established, but viable *Hp lgtF* mutants were made during this study, indicating that the gene is not essential for bacterial survival in vitro. Another substituent of the *Hi* and *Hp* inner core LPS, PEtn on HepII, is less likely to be conserved across Pasteurellaceae species as only *Histophilus somni* has a homologue of the PEtn transferase gene *lpt6*. In *Neisseria meningitidis*, PEtn substitution of the LPS enhances epithelial adhesion and resistance to the antimicrobial peptide LL-37 (Takahashi et al., 2008; Tzeng et al., 2005), although the PEtn residue that was studied was attached to lipid A rather than HepII so may not be directly comparable with the situation in *Hp*. PEtn on HepII does not appear to be involved in complement-mediated killing in *Hi*, as *lpt6* mutants survive as well as wild type strains during serum bactericidal assays (Wright et al., 2004).

According to our PCR and sequence analyses of glycosyltransferase genes, the majority of the *Hp* strains (12/18) are predicted to add a limited non-phase variable *Hi*-like outer core extension to a single inner core sugar through the following linkages: DDHep-(α1,6)-Glc (on HepI) using LosB1; Glc-(α1,3)-HepII using Lic2C; or Gal-(β1,2)-HepIII using LpsB. The latter two extensions have been structurally confirmed for *Hp* strains 13 and 20 respectively. The largely mutually exclusive distribution of the *losB1/lic2C* and *lpsB* loci is intriguing and contrasts with

the situation in *Hi*, where many strains are capable of extending from multiple inner core sugars. One possibility is that an *Hp* strain lacking the *losB1/lic2C* locus obtained *lpsB* through horizontal transfer from *Hi*, and that the subsequent spread of this gene was mainly through clonal expansion rather than genetic exchange. The greater prevalence of non-phase variable LPS outer core hexose residues in *Hi* compared to *Hp* reflects the importance of these sugars as attachment sites for key phase variable substituents. The only core hexose that appears to be conserved in all *Hp* strains is the Glc residue on HepI, which is required for the addition of OAg and single O-units.

One of the main findings of this research was the scarcity of simple sequence repeat tracts in *Hp* when compared to *Hi*, and in particular the apparent absence of phase variably expressed LPS genes. Previous work in our laboratory has shown that 25/25 NTHi isolates possessed (among others) the phase variable genes *lic3A*, *oafA* and *lic2A/lgtC*, whose products direct the addition of sialic acid, O-acetyl and digalactoside epitopes to the LPS respectively. All three of these epitopes are known to confer some level of resistance to the killing effects of human complement (Schweda et al., 2007); sialic acid and digalactoside are also virulence factors in chinchilla otitis media and rat bacteraemia models respectively (Bouchet et al., 2003; Griffin et al., 2005). *Hp* strain 17 is thought to express a non-phase variable digalactoside epitope, but we found no evidence for core sialic acid or O-acetyl expression in *Hp*. However, both epitopes were later found as part of OAg or O-unit structures in various *Hp* strains. As *Hp* is rarely found in disease situations, sialic acid and O-acetyl groups are presumably expressed in order to aid commensal colonisation. If *Hp* is capable of regulating OAg expression or chain length according to environmental signals, this could be an alternative way of modulating the presence of sialic acid

and O-acetyl groups when it is advantageous to do so, instead of relying on a stochastic phase variation mechanism like *Hi*.

Roberts et al. (1986) speculated that some *Hp* isolates may elaborate their LPS with OAg after observing ladder-like LPS banding patterns by SDS-PAGE. In this study we present the first genetic evidence to support this hypothesis and show that all (18/18) of our *Hp* strains possess a polysaccharide biosynthesis locus located between genes *glnA* and *pepB*, irrespective of whether or not OAg was detectable by SDS-PAGE. Mutation of the undecaprenyl phosphate gene in several strains, followed by electrophoretic and structural LPS analyses, demonstrated the involvement of the locus between *glnA* and *pepB* in OAg or O-unit synthesis. The analysis of *lgtF* mutants showed that OAg addition is dependent upon the presence of Glc on HepI.

We found an approximately equal distribution in *Hp* of the two known OAg synthesis and transport systems (Wzy-dependent and ABC transporter dependent). Both of these mechanisms are widespread amongst Gram negative bacteria; for example, most *E. coli* OAgs are synthesised using the Wzy-dependent system, whereas *Klebsiella pneumoniae* OAgs are mainly ABC transporter dependent (Samuel & Reeves, 2003). Although many species comprise a mixture of Wzy-dependent and ABC transporter dependent OAg serotypes, it is unusual for the two types of OAg loci to map to the same genomic location as they do in *Hp*.

It is clear that other bacteria found in the human respiratory tract are an important source of OAg synthesis genes for *Hp*. Many of the closest known homologues are from other human respiratory tract bacteria, including some fellow Pasteurellaceae species such as *Aggregatibacter*

actinomycetemcomitans and *Aggregatibacter aphrophilus*. However, genetic exchange does not appear to be limited to genes from related species or those involved in OAg synthesis, as some of the closest homologues to *Hp* OAg locus genes were capsule and teichoic acid synthesis genes from the Gram positive respiratory tract species *Streptococcus pneumoniae* and *Gemella haemolysans*. As *Hp* is reported to colonise the human digestive tract and can be cultured from faecal samples (Palmer, 1981), genetic exchange between *Hp* and species such as *Bacteroides fragilis* and *Shigella dysenteriae* may also occur: this was reflected in the homology analysis. OAg expression may in fact aid colonisation of the digestive tract by *Hp* as it does for other bacteria (Fowler et al., 2006). It is important to note that *Hp* is likely to act as a DNA donor as well as recipient so the direction of horizontal gene transfer events is difficult to ascertain.

Both the *Hi hmg* locus and the *Hp* OAg locus are flanked by the genes *glnA* and *pepB*, whilst the *lic2* locus (containing 0-5 outer core LPS genes) is between *apaH* and *infA* in both species (Fig. 3.2). We looked briefly at the regions around these flanking genes in the genome sequences of 13 other Pasteurellaceae species for any evidence of similar LPS-related loci. In all of the genomes, *glnA* and *pepB* were in separate genomic locations so clearly did not flank a polysaccharide synthesis locus together. However, *pepB* and *infA* are adjacent in *A. actinomycetemcomitans* and *A. aphrophilus*, are separated only by *rmlB* in *Pasteurella multocida*, and flank what appears to be a 12 gene polysaccharide synthesis locus in *Mannheimia succiniciproducens*. One possibility is that in the common ancestor of *Hi* and *Hp*, a polysaccharide synthesis locus was located between *pepB* and *infA*, and a genomic rearrangement split the genes into two groups which evolved into the *hmg*/OAg and *lic2* loci.

In our *Hp* study strains there was no correlation between the presence of particular *Hi*-like outer core LPS genes and the type of OAg locus. This emphasises the high degree of genetic exchange that must occur in *Hp*, particularly in the genes required for OAg synthesis. It also increases the likelihood that an *Hp* strain colonising a new host expresses an LPS structure that has not been encountered by the host's immune system before, thereby extending the average length of colonisation whilst specific antibodies are produced. Indeed, if we combine the genetic data for outer core and OAg related genes (Fig. 3.8 and Table 5.2), and assume that all strains can express at least one O-unit in vivo, it appears likely that around 16 of the 18 true *Hp* strains express unique LPS structures.

Some of the *Hp* strains that we studied do not synthesise any detectable OAg under laboratory conditions. It is well documented that the ability to synthesise an OAg can be selected against under certain circumstances. For instance, cystic fibrosis (CF) seems to select for mutants lacking OAg to some extent, perhaps because the increased adhesive properties of the bacteria prevent physical clearance. Three CF isolates of *Burkholderia cenocepacia* were unable to synthesise OAg due to an insertion sequence in a particular GT gene (Ortega et al., 2005); several *Pseudomonas aeruginosa* isolates from CF patients also lack OAg because of insertion sequences, frameshift mutations, or deletion of the OAg locus (Pier, 2007). As another example, when Pollock et al. (1994) selected for spontaneous bacitracin resistance in several Gram negative species, a high proportion of the resulting colonies had lost surface polysaccharides. Bacitracin sequesters undecaprenyl pyrophosphate and prevents its recycling as UndP, the lipid carrier required for the synthesis of both peptidoglycan and surface polysaccharides (e.g. OAg); the mutants presumably diverted all available UndP towards essential peptidoglycan synthesis. It

is not clear whether selective media containing bacitracin was used for the isolation of our *Hp* study strains ten years ago, as it occasionally is for *Hi*. If so, bacteria expressing less OAg may have been selected for during isolation.

As an OAg locus was present in all of our 18 true *Hp* study strains, long term evolution presumably favours an ability for OAg expression in this species, and strains deficient in OAg synthesis or that only express a single O-unit may be able to thrive only in a particular host environment or in vitro in the laboratory. In many species the amount and modal length of OAg varies depending on factors such as temperature, iron concentration, the presence of serum, or growth phase; the appropriate conditions for optimal OAg expression in some *Hp* strains may not be provided by BHI medium. It is difficult to test whether OAg is expressed in vivo as the bacteria would probably need to be subcultured after harvesting to obtain sufficient material for SDS-PAGE. However, bacteria grown in human serum or on epithelial cells in vitro could be examined more easily, and any changes in OAg expression would provide a starting point for the elucidation of OAg regulation mechanisms in *Hp*.

When *Hp* LPS is visualised by SDS-PAGE, the 'rungs' of the OAg patterns are much more closely spaced in strains with group 2 (ABC transporter dependent) rather than group 1 (Wzy-dependent) OAg, usually appearing as a smear in the former. In group 2 strains, the OAg chain is thought to be extended one sugar at a time (e.g. by the alternating action of a GlcNAc transferase and a Gal transferase for *Hp* strain 13), whereas the Wzy-dependent system adds completed O-units. This suggests that strains such as *Hp* 13 could potentially control their OAg chain length in a more precise and subtle manner. Modal chain length regulation in ABC transporter dependent

OAg systems is poorly understood; in some *E. coli* serotypes, chain termination and transport occurs upon methylation of the terminal sugar (Clarke et al., 2004). No methylated sugars were found in the *Hp* 13 OAg using structural analysis (Elke Schweda, pers. comm.) and none of the proteins encoded by the *Hp* 13 OAg locus are predicted to contain the coiled coil motif that is typical of methyltransferases, suggesting that the tight OAg chain length distribution in this strain is controlled by a different mechanism.

We have demonstrated that it is possible to predict OAg structure from the genes present in the *Hp* OAg locus, and conversely to predict gene presence from structure, to a certain extent. This has allowed us to generate a picture of the level of O-unit diversity within *Hp* without needing to analyse every structure using chemical methods. Many of the glycosyltransferase genes found in the *Hp* OAg loci do not have close homologues in other species in the NCBI database and may have novel donor sugar, acceptor site and/or linkage specificities, making them potentially interesting to glycobiologists attempting to synthesise particular oligosaccharides or creating panels of glycans for drug discovery screens. For example, researchers aiming to develop a treatment for *S. dysenteriae* infections used two well-characterised glycosyltransferase genes from *Neisseria meningitidis* to create an artificial globotriose extension on the cell surface LPS of *E. coli* (Paton et al., 2000). This extension mimics the shiga toxin receptor that is found on host cells, and sequesters the toxin so that it cannot bind to, enter, and inhibit protein synthesis in host cells.

If we are to understand the differences in pathogenicity between *Hp* and *Hi*, it is important to study how the host inflammatory response to each of the species might differ. During in vitro

experiments, heat-killed cells that we prepared from nine representative *Hp* study strains were all shown to induce lower levels of dendritic IL-12 production than nine heat-killed NTHi strains (Giorgio Napolitani, pers. comm.). IL-12 is a cytokine that stimulates T cell differentiation and is induced when TLR4 on the dendritic cells is activated by the lipid A component of LPS. In addition, Tuyau and Sims (1983) found that LPS isolated from an *Hp* strain was eight times less potent per microgram than that of an *Hi* strain in terms of inducing lethal endotoxic shock in mice. We might therefore expect the lipid A structures of *Hi* and *Hp* to differ, but our genetic analysis and the chemical structure data provided by our collaborators indicate that in fact these species synthesise very similar lipid A, with the same headgroups and the same six acyl chains (Chapter 3 and Elke Schweda, pers. comm.). A study in which a human epithelial cell line was incubated with *E. coli* LPS demonstrated that considerably less IL-8 was induced when the LPS contained OAg, although the two *E. coli* strains used were not isogenic (Backhed et al., 2003). This discovery was authenticated when a screen of 5000 *E. coli* transposon mutants for those with elevated IL-8 inducing properties identified a strain in which the OAg ligase gene had been interrupted, leading to the loss of OAg (Billips et al., 2008). The only other mutants which induced high IL-8 levels were those that released abnormally high levels of peptidoglycan, which is recognised as a bacterial molecule by host pattern recognition receptors. These results suggest that OAg somehow downregulates the innate immune response to LPS, perhaps by shielding lipid A or by altering LPS conformation so that it binds to the TLR4-MD2 complex less efficiently.

During an immune response in the upper respiratory tract, inflammatory mediators including neutrophils and complement must cross the epithelial barrier to reach bacteria on the mucosal

surface. This inevitably requires some loss of barrier integrity, for example the opening of tight junctions, which in turn may allow surviving bacteria to invade or cross the epithelium. In a murine model, either colonisation of the upper respiratory tract with *Hi* or nasal administration of *Hi* LPS led to TLR4 activation, the induction of cytokines such as IL-6, and a loss of epithelial barrier integrity (Beisswenger et al., 2009). If *Hp* colonisation of the upper respiratory tract induces a weaker immune reponse than *Hi*, either due to its OAg or because it sheds less LPS (Tuyau & Sims, 1983), it may be able to colonise for longer or in greater densities without total clearance, reflecting its success as a commensal organism. Conversely there may be less opportunity for pathogenic behaviour as the integrity of the epithelial barrier would be maintained. In reality the situation is likely to be more complex due to the interplay between responses to different bacterial species in a mixed population on the mucosal surface.

TLR4 is present on the cell surface of myeloid cells such as macrophages. However, in pulmonary and intestinal epithelial cells, TLR4 is restricted to an intracellular compartment, so its activation by lipid A requires the internalisation of LPS. Duerr et al. (2009) demonstrated that the presence of OAg in *Salmonella enterica* LPS significantly delayed LPS internalisation by intestinal epithelial cells, resulting in a delayed inflammatory response. Differences in OAg expression may therefore be one of the reasons that *Hi*, but not *Hp*, appears to exacerbate COPD symptoms despite both organisms being present in the sputum of COPD patients (Section 1.2.3).

Could the differences in LPS structure between *Hi* and *Hp* help to explain why the latter is rarely implicated in otitis media, or are other distinctions more important? This disease occurs when impairment of Eustachian tube function allows bacteria such as *Hi* to travel from the

nasopharynx to colonise the middle ear, triggering a local immune response and inflammation (Straetemans et al., 2001). There are several steps at which *Hp* could behave differently or induce dissimilar host responses from *Hi*: during nasopharyngeal colonisation, entrance and passage along the Eustachian tube, or middle ear colonisation.

At the point of entry into the respiratory tract, the nasopharyngeal tonsils (adenoids) sample antigens and use immune cells to survey and destroy incoming bacteria (van Kempen et al., 2000). Memory and effector lymphocytes are also generated here and migrate to nearby mucosal sites to target bacteria which evaded the original clearance. As *Hp* is frequently isolated from the nasopharynx, it must be able to evade this attack to some extent or at least replicate fast enough to avoid total clearance. Our preliminary epithelial association assays, performed in vitro, suggest that *Hp* is more adherent than *Hi* strain Rd and that *Hp* OAgs can either help or hinder this process depending on the strain. Adherence to epithelial cells may protect *Hp* from physical clearance by cilia in the respiratory tract.

Since *Hp* is a successful commensal organism, the scarcity of *Hp* otitis media is unlikely to be due to inefficient nasopharyngeal colonisation. The mechanism of subsequent migration along the Eustachian tube in otitis media has not been fully elucidated, but *Hi* is reported to adhere preferentially to mucus rather than epithelial cells during this stage (Miyamoto & Bakaletz, 1996). The ability of *Hp* to adhere to mucus remains to be tested. In *Hi* this property is mediated partly by OMP P5 fimbriae (Miyamoto & Bakaletz, 1996); the *Hp* T3T1 gene *Hp1961* encodes a protein with 75% identity to *Hi* OMP P5, but other factors including LPS outer core/OAg structure may also influence mucosal attachment.

As the Eustachian tube and middle ear are usually sterile sites, the detection of bacteria in these areas triggers a strong immune response including the recruitment of complement proteins, cytokines and polymorphonuclear leukocytes. We have demonstrated that OAg plays a major role in resisting complement-mediated killing in vitro for many *Hp* strains. However, migration through the Eustachian tube and colonisation of the middle ear may well require the flexibility provided by a phase variable LPS outer core in order to evade antigen recognition, adhere to a range of different host structures and promote continued Eustachian tube dysfunction. On rare occasions an *Hp* strain may alter its OAg structure by genetic exchange or point mutations in its glycosyltransferase genes, but these methods of adaptation do not have the advantages of phase variation: high frequency, reversible changes leading to a population of cells expressing dozens of slightly different LPS structures available for short term natural selection.

The expression of particular outer core substituents on the LPS also appears to make an *Hi* strain more likely to reach and/or survive in the middle ear. Pettigrew et al. (2002) found that the *lic2B* gene, which encodes a putative LPS galactosyltransferase, could be amplified from 52% of middle ear NTHi isolates but only 9% of nasopharyngeal and throat isolates. Interestingly, whilst five of our *Hp* study strains contained a *lic2C* homologue, the *lic2B* gene that flanks it in *Hi* was always absent from this position in *Hp*. PCho expression on the *Hi* LPS has also been linked to otitis media. Researchers using a chinchilla model showed that $PCho^{+}$ phase variants were associated with greater colonisation of the middle ear than $PCho^{-}$ phase variants (Tong et al., 2000), and in middle ear NTHi isolates from children with otitis media, greater PCho expression was related to the prolongation of the disease (Fujita et al., 2009). Our demonstration that *Hp*

does not add PCho to its LPS adds weight to the idea that PCho is a key virulence factor for *Hi* otitis media.

A direct comparison of the role of LPS structure in determining the level of virulence behaviour by *Hi* and *Hp* is confounded by other distinctions between the species, such as differences in surface proteins (adhesins and pili), the presence of a capsule, IgA protease expression, and metabolic pathways. Using our increased understanding of *Hp* LPS biosynthesis and structure, it might be possible to engineer an *Hp* strain to express various combinations of *Hi*-like LPS components to examine whether these alone are sufficient to promote Eustachian tube ascension and/or the colonisation and inflammation of the middle ear in the chinchilla model of otitis media, assuming that wild type *Hp* strains do not do so. We are part way to achieving this with the development of OAg^-, $PCho^+$ and OAg^- $PCho^+$ strains during this project. This would also help us to determine whether OAg affects survival or inflammation in the middle ear. Similar experiments are currently being performed with gut bacteria, where a non-pathogenic strain of *E. coli* whose LPS had been truncated was transformed with the genes required for the expression of *Helicobacter pylori* Lewis Y OAg (Yavuz et al., 2011) in order to later examine host responses to the OAg. *Hp* is ideally suited to testing the biological effects of individual *Hi* LPS epitopes, as its LPS typically contains few outer core substituents (the OAg can be removed by interrupting a single gene) and the results would not be confounded by the phase variable expression of other epitopes as they are in *Hi*.

It was established over thirty years ago that several factor X independent but mainly factor V dependent '*Haemophilus* species' (now known as *Aggregatibacter segnis*, *A.*

actinomycetemcomitans and *A. aphrophilus*) colonise the human oral cavity and are frequently isolated from dental plaque, whereas the *Haemophilus* species that require both factors (*Hi* and *H. haemolyticus*) are associated more with pharyngeal colonisation (Kilian, 1976; Norskov-Lauritsen & Kilian, 2006). This suggests that NAD, but not haemin, is readily available on the dental surface. Although *Hp* is more closely related genetically to *Hi* than to the *Aggregatibacter* species, its ability to synthesise porphyrin appears to widen its range of possible host niches and it is frequently isolated from saliva and dental plaque as well as from the pharynx (Kilian et al., 1976; Salako & Kleinberg, 1989).

Whilst *A. actinomycetemcomitans* and *A. segnis* are associated with both commensal colonisation and periodontal disease, there is little evidence that *Hp* is a dental pathogen. One study found that antibody titres against *A. actinomycetemcomitans*, but not *Hp*, were raised in subjects with localised aggressive periodontitis (Genco et al., 1985), and *Hp* is being investigated for use as an antagonist to prevent reinfection after treatment for *Porphyromonas gingivalis* periodontitis (Van Hoogmoed et al., 2008). A comparison of LPS structures in *Hp* and *A. actinomycetemcomitans* might therefore be useful in establishing which components are likely to aid commensal dental colonisation or progression to periodontitis.

Like *Hp*, *A. actinomycetemcomitans* expresses OAg on its LPS. There are seven known OAg serotypes, designated a-g; typical sugars include rhamnose, mannose, deoxytalose and GalNAc (Perry et al., 1996; Takada et al., 2010). Some of the OAgs are O-acetylated, but none contains sialic acid. In some strains only one putative O-unit is observed on the LPS by SDS-PAGE (Masoud et al., 1991), mirroring our findings for *Hp*. Since the presence of OAg/O-units and O-

acetyl groups is common to both *Hp* and *A. actinomycetemcomitans*, these LPS features may be beneficial for non-symptomatic colonisation of tooth enamel; indeed, Fujise and colleagues (2008) found that *A. actinomycetemcomitans* serotype a OAg promotes adherence to glass, a proxy for tooth enamel. *A. actinomycetemcomitans* core LPS structures have not been studied in detail, but BLAST analysis of the strain D11S-1 genome sequence using *Hi* outer core LPS genes identifies homologues of *lic2A*, *losA1* and *losB1*, indicating the presence of heptose and hexose extensions that are also found in a subset of *Hp* strains.

Although there is no *lic1D* homologue in the *A. actinomycetemcomitans* strain D11S-1 genome sequence, some strains are known to express mAb TEPC15-reactive cell surface PCho (Schenkein et al., 2000). These strains are able to invade endothelial cells in vitro, and this ability is inhibited by the addition of a PAF receptor antagonist, suggesting that invasion is mediated at least in part by the binding of PCho to the PAF receptor. In contrast, mAb TEPC15 did not react with any of our *Hp* study strains until we inserted a transgenic *lic1* locus. PCho expression by *Hi* and *A. actinomycetemcomitans* may help the bacteria to gain access to host cells, increasing their virulence potential, and may trigger a local antibody-mediated inflammatory response, which is a symptom of both otitis media and periodontitis.

The genome announcement paper for *A. actinomycetemcomitans* strain D11S-1 (Chen et al., 2009) identifies LPS and OAg biosynthesis loci but does not mention phase variation. A quick BLASTN search of this genome using the sequence $(CAAT)_4$ reveals a family 2 glycosyltransferase gene (Genbank: CP001733.1) with this motif five base pairs after the start codon. Whilst a comprehensive survey of repeat tracts in *A. actinomycetemcomitans* was beyond

our remit, this finding hints that it may be able to phase vary the glycosylation of one or more cell surface components, perhaps LPS or an adhesin. In summary, a lack of simple sequence repeats and PCho expression distinguish *Hp* from its more pathogenic relatives, *Hi* and (putatively) *A. actinomycetemcomitans*, whilst porphyrin biosynthesis and OAg expression distinguish the species that can colonise dental and gingival surfaces (*Hp* and *A. actinomycetemcomitans*) from *Hi*. Further research is needed to test the significance of these observations. *Hi* strains that cause systemic infections are usually encapsulated and it would be interesting to test the rare systemic *Hp* and *A. actinomycetemcomitans* isolates for an *Hi*-like *cap* locus.

Hp is a largely commensal organism whose presence may confer some protection against colonisation with more pathogenic species, but it is also implicated in disease on rare occasions. Opinion is likely to be divided over whether it would be desirable to retain or reduce *Hp* colonisation, but in any case it is important that we understand the potential impact of any *Hi* vaccination programme on the respiratory flora. We found capsule synthesis genes in only one of the 18 *Hp* study strains, and the sugar-specific genes were not *Hi*-like, so the current capsule-based *Hi* vaccines are unlikely to target *Hp*. However, the NTHi vaccines that are now being developed are based on cell surface components that may also be present in *Hp*. A conjugate vaccine including *Hi* protein D showed some efficacy against *Hi* otitis media in a clinical trial, and *Hi* PilA was protective in a chinchilla model (Murphy et al., 2009). BLAST analysis of the *Hp* T3T1 genome sequence shows that both of these proteins have close homologues in *Hp*, so the carriage of this species could potentially decrease if either vaccine is introduced. Conversely, there would be no direct impact from a vaccine based on the *Hi* LPS HMG unit (Sundgren et al.,

2010): the results presented in this thesis show that although the *Hp* OAg locus is in some ways similar to the *Hi hmg* locus, the O-unit structures produced by *Hp* contain different sugar and linkage combinations to the HMG.

Despite their phylogenetic relatedness and physical proximity, *Hp* and *Hi* have tackled the problem of survival in the human respiratory tract in slightly different ways. *Hp* decorates its LPS with a limited number of outer core sugars and one of a wide range of OAg or O-unit structures. Together with other cell surface molecules, this must provide it with sufficient levels of antigenic variation, epithelial adhesion, protection against mucosal IgA and resistance to killing by antimicrobial peptides, membrane attack complexes and macrophages, because *Hp* is a highly successful commensal species. OAg may also be beneficial for *Hp* in other host niches such as dental surfaces and the digestive tract. Meanwhile, *Hi* LPS contains at most one 'O-unit' (HMG) and the structure of this moiety does not appear to vary between strains. Instead, the phase variable expression of several important outer core components provides enough antigenic and functional diversity that a few bacteria will probably avoid the antibody response and have suitable surface properties for their current environment. The ability of NTHi to ascend a compromised Eustachian tube and cause otitis media, with no obvious benefits in terms of transmission, may be an unintentional consequence of its unusual and highly adaptable mechanism of LPS variation along with its capacity to synthesise an IgA protease. The research presented in this thesis reveals key distinctions between the LPS of *Hi* and *Hp*, demonstrates some of the biological effects of these differences in vitro, and lays the groundwork for in vivo experiments using the wild type *Hp* study strains and their LPS mutants.

9 References

Aanensen, D. M., Mavroidi, A., Bentley, S. D., Reeves, P. R. & Spratt, B. G. (2007). Predicted functions and linkage specificities of the products of the *Streptococcus pneumoniae* capsular biosynthetic loci. *J Bacteriol* **189**, 7856-7876.

Abeyrathne, P. D. & Lam, J. S. (2007). WaaL of *Pseudomonas aeruginosa* utilizes ATP in in vitro ligation of O antigen onto lipid A-core. *Mol Microbiol* **65**, 1345-1359.

Adlam, C., Knights, J. M., Mugridge, A., Williams, J. M. & Lindon, J. C. (1987). Production of colominic acid by *Pasteurella haemolytica* serotype A2 organisms. *FEMS Microbiology Letters* **42**, 23-25.

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

Allison, G. E. & Verma, N. K. (2000). Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*. *Trends Microbiol* **8**, 17-23.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402.

Amor, K., Heinrichs, D. E., Frirdich, E., Ziebell, K., Johnson, R. P. & Whitfield, C. (2000). Distribution of core oligosaccharide types in lipopolysaccharides from *Escherichia coli*. *Infect Immun* **68**, 1116-1124.

Andersen, C., Maier, E., Kemmer, G., Blass, J., Hilpert, A. K., Benz, R. & Reidl, J. (2003). Porin OmpP2 of *Haemophilus influenzae* shows specificity for nicotinamide-derived nucleotide substrates. *J Biol Chem* **278**, 24269-24276.

Appelmelk, B. J., Monteiro, M. A., Martin, S. L., Moran, A. P. & Vandenbroucke-Grauls, C. M. (2000). Why *Helicobacter pylori* has Lewis antigens. *Trends Microbiol* **8**, 565-570.

Arbeit, R. D., Karakawa, W. W., Vann, W. F. & Robbins, J. B. (1984). Predominance of two newly described capsular polysaccharide types among clinical isolates of *Staphylococcus aureus*. *Diagn Microbiol Infect Dis* **2**, 85-91.

Avadhanula, V., Rodriguez, C. A., Ulett, G. C., Bakaletz, L. O. & Adderson, E. E. (2006). Nontypeable *Haemophilus influenzae* adheres to intercellular adhesion molecule 1 (ICAM-1) on respiratory epithelial cells and upregulates ICAM-1 expression. *Infect Immun* **74**, 830-838.

Avci, F. Y. & Kasper, D. L. (2010). How bacterial carbohydrates influence the adaptive immune system. *Annu Rev Immunol* **28**, 107-130.

Ayyildiz, A., Aktas, A. E. & Yazgi, H. (2003). Nasopharyngeal carriage rate of *Haemophilus influenzae* in children aged 7-12 years in Turkey. *Int J Clin Pract* **57**, 686-688.

Backhed, F., Normark, S., Schweda, E. K., Oscarson, S. & Richter-Dahlfors, A. (2003). Structural requirements for TLR4-mediated LPS signalling: a biological role for LPS modifications. *Microbes Infect* **5**, 1057-1063.

Barany, F. & Kahn, M. E. (1985). Comparison of transformation mechanisms of *Haemophilus parainfluenzae* and *Haemophilus influenzae*. *J Bacteriol* **161**, 72-79.

Barrallo, S., Reglero, A., Revilla-Nuin, B., Martinez-Blanco, H., Rodriguez-Aparicio, L. B. & Ferrero, M. A. (1999). Regulation of capsular polysialic acid biosynthesis by temperature in *Pasteurella haemolytica* A2. *FEBS Lett* **445**, 325-328.

Baumann, H., Tzianabos, A. O., Brisson, J. R., Kasper, D. L. & Jennings, H. J. (1992). Structural elucidation of two capsular polysaccharides from one strain of *Bacteroides fragilis* using high-resolution NMR spectroscopy. *Biochemistry* **31**, 4081-4089.

Baur, S., Marles-Wright, J., Buckenmaier, S., Lewis, R. J. & Vollmer, W. (2009). Synthesis of CDP-activated ribitol for teichoic acid precursors in *Streptococcus pneumoniae*. *J Bacteriol* **191**, 1200-1210.

Bayliss, C. D., Sweetman, W. A. & Moxon, E. R. (2004). Mutations in *Haemophilus influenzae* mismatch repair genes increase mutation rates of dinucleotide repeat tracts but not dinucleotide repeat-driven pilin phase variation rates. *J Bacteriol* **186**, 2928-2935.

Beisswenger, C., Lysenko, E. S. & Weiser, J. N. (2009). Early bacterial colonization induces toll-like receptor-dependent transforming growth factor beta signaling in the epithelium. *Infect Immun* **77**, 2212-2220.

Belogurov, G. A., Sevostyanova, A., Svetlov, V. & Artsimovitch, I. (2010). Functional regions of the N-terminal domain of the antiterminator RfaH. *Mol Microbiol* **76**, 286-301.

Bentley, S. D., Aanensen, D. M., Mavroidi, A. & other authors (2006). Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet* **2**, e31.

Bergstrom, N., Jansson, P. E., Kilian, M. & Skov Sorensen, U. B. (2000). Structures of two cell wall-associated polysaccharides of a *Streptococcus mitis* biovar 1 strain. A unique teichoic acid-like polysaccharide and the group O antigen which is a C-polysaccharide in common with pneumococci. *Eur J Biochem* **267**, 7147-7157.

Berrington, A. W., Tan, Y. C., Srikhanta, Y., Kuipers, B., van der Ley, P., Peak, I. R. & Jennings, M. P. (2002). Phase variation in meningococcal lipooligosaccharide biosynthesis genes. *FEMS Immunol Med Microbiol* **34**, 267-275.

Biberstein, E. L. & Gills, M. (1961). Catalase activity of *Haemophilus* species grown with graded amounts of hemin. *J Bacteriol* **81**, 380-384.

Billips, B. K., Schaeffer, A. J. & Klumpp, D. J. (2008). Molecular basis of uropathogenic *Escherichia coli* evasion of the innate immune response in the bladder. *Infect Immun* **76**, 3891-3900.

Black, C. T., Kupferschmid, J. P., West, K. W. & Grosfeld, J. L. (1988). *Haemophilus parainfluenzae* infections in children, with the report of a unique case. *Rev Infect Dis* **10**, 342-346.

Bottone, E. J. & Zhang, D. Y. (1995). *Haemophilus parainfluenzae* biliary tract infection: rationale for an ascending route of infection from the gastrointestinal tract. *J Clin Microbiol* **33**, 3042-3043.

Bouchet, V., Hood, D. W., Li, J. & other authors (2003). Host-derived sialic acid is incorporated into *Haemophilus influenzae* lipopolysaccharide and is a major virulence factor in experimental otitis media. *Proc Natl Acad Sci U S A* **100**, 8898-8903.

Bravo, D., Hoare, A., Silipo, A., Valenzuela, C., Salinas, C., Alvarez, S. A., Molinaro, A., Valvano, M. A. & Contreras, I. (2010). Different sugar residues of the lipopolysaccharide outer core are required for early interactions of *Salmonella enterica* serovars Typhi and Typhimurium with epithelial cells. *Microb Pathog*.

Broides, A., Dagan, R., Greenberg, D., Givon-Lavi, N. & Leibovitz, E. (2009). Acute otitis media caused by *Moraxella catarrhalis*: epidemiologic and clinical characteristics. *Clin Infect Dis* **49**, 1641-1647.

Bryant, C. E., Spring, D. R., Gangloff, M. & Gay, N. J. (2010). The molecular basis of the host response to lipopolysaccharide. *Nat Rev Microbiol* **8**, 8-14.

Canals, R., Jimenez, N., Vilches, S., Regue, M., Merino, S. & Tomas, J. M. (2006). The UDP N-acetylgalactosamine 4-epimerase gene is essential for mesophilic *Aeromonas hydrophila* serotype O34 virulence. *Infect Immun* **74**, 537-548.

Cardines, R., Giufre, M., Ciofi degli Atti, M. L., Accogli, M., Mastrantonio, P. & Cerquetti, M. (2009). *Haemophilus parainfluenzae* meningitis in an adult associated with acute otitis media. *New Microbiol* **32**, 213-215.

Cascales, E., Buchanan, S. K., Duche, D., Kleanthous, C., Lloubes, R., Postle, K., Riley, M., Slatin, S. & Cavard, D. (2007). Colicin biology. *Microbiol Mol Biol Rev* **71**, 158-229.

Cazalet, C., Jarraud, S., Ghavi-Helm, Y., Kunst, F., Glaser, P., Etienne, J. & Buchrieser, C. (2008). Multigenome analysis identifies a worldwide distributed epidemic *Legionella pneumophila* clone that emerged within a highly diverse species. *Genome Res* **18**, 431-441.

Cerdeno-Tarraga, A. M., Patrick, S., Crossman, L. C. & other authors (2005). Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science* **307**, 1463-1465.

Chatterjee, S. N. & Chaudhuri, K. (2003). Lipopolysaccharides of *Vibrio cholerae*. I. Physical and chemical characterization. *Biochim Biophys Acta* **1639**, 65-79.

Chen, C., Kittichotirat, W., Si, Y. & Bumgarner, R. (2009). Genome sequence of *Aggregatibacter actinomycetemcomitans* serotype c strain D11S-1. *J Bacteriol* **191**, 7378-7379.

Cieslewicz, M. J., Chaffin, D., Glusman, G., Kasper, D., Madan, A., Rodrigues, S., Fahey, J., Wessels, M. R. & Rubens, C. E. (2005). Structural and genetic diversity of group B *Streptococcus* capsular polysaccharides. *Infect Immun* **73**, 3096-3103.

Cipolla, L., Polissi, A., Airoldi, C., Galliani, P., Sperandeo, P. & Nicotra, F. (2009). The Kdo biosynthetic pathway toward OM biogenesis as target in antibacterial drug design and development. *Curr Drug Discov Technol* **6**, 19-33.

Clarke, B. R., Cuthbertson, L. & Whitfield, C. (2004). Nonreducing terminal modifications determine the chain length of polymannose O antigens of *Escherichia coli* and couple chain termination to polymer export via an ATP-binding cassette transporter. *J Biol Chem* **279**, 35709-35718.

Cooney, T. G., Harwood, B. R. & Meisner, D. J. (1981). *Haemophilus parainfluenzae* thoracic empyema. *Arch Intern Med* **141**, 940-941.

Cotter, S. E., Yeo, H. J., Juehne, T. & St Geme, J. W., 3rd (2005). Architecture and adhesive activity of the *Haemophilus influenzae* Hsf adhesin. *J Bacteriol* **187**, 4656-4664.

Coyne, M. J., Tzianabos, A. O., Mallory, B. C., Carey, V. J., Kasper, D. L. & Comstock, L. E. (2001). Polysaccharide biosynthesis locus required for virulence of *Bacteroides fragilis*. *Infect Immun* **69**, 4342-4350.

Cuthbertson, L., Kimber, M. S. & Whitfield, C. (2007). Substrate binding by a bacterial ABC transporter involved in polysaccharide export. *Proc Natl Acad Sci U S A* **104**, 19529-19534.

Cynamon, M. H., Sorg, T. B. & Patapow, A. (1988). Utilization and metabolism of NAD by *Haemophilus parainfluenzae*. *J Gen Microbiol* **134**, 2789-2799.

Darras-Joly, C., Lortholary, O., Mainardi, J. L., Etienne, J., Guillevin, L. & Acar, J. (1997). *Haemophilus endocarditis*: report of 42 cases in adults and review. *Haemophilus* Endocarditis Study Group. *Clin Infect Dis* **24**, 1087-1094.

De Bolle, X., Bayliss, C. D., Field, D., van de Ven, T., Saunders, N. J., Hood, D. W. & Moxon, E. R. (2000). The length of a tetranucleotide repeat tract in *Haemophilus influenzae* determines the phase variation rate of a gene with homology to type III DNA methyltransferases. *Mol Microbiol* **35**, 211-222.

Deadman, M. E., Lundstrom, S. L., Schweda, E. K., Moxon, E. R. & Hood, D. W. (2006). Specific amino acids of the glycosyltransferase LpsA direct the addition of glucose or galactose to the terminal inner core heptose of *Haemophilus influenzae* lipopolysaccharide via alternative linkages. *J Biol Chem* **281**, 29455-29467.

Deadman, M. E., Hermant, P., Engskog, M., Makepeace, K., Moxon, E. R., Schweda, E. K. & Hood, D. W. (2009). Lex2B, a phase-variable glycosyltransferase, adds either a glucose or a galactose to *Haemophilus influenzae* lipopolysaccharide. *Infect Immun* **77**, 2376-2384.

DiGiandomenico, A., Matewish, M. J., Bisaillon, A., Stehle, J. R., Lam, J. S. & Castric, P. (2002). Glycosylation of *Pseudomonas aeruginosa* 1244 pilin: glycan substrate specificity. *Mol Microbiol* **46**, 519-530.

Draing, C., Pfitzenmaier, M., Zummo, S., Mancuso, G., Geyer, A., Hartung, T. & von Aulock, S. (2006). Comparison of lipoteichoic acid from different serotypes of *Streptococcus pneumoniae*. *J Biol Chem* **281**, 33849-33859.

Duerr, C. U., Zenk, S. F., Chassin, C., Pott, J., Gutle, D., Hensel, M. & Hornef, M. W. (2009). O-antigen delays lipopolysaccharide recognition and impairs antibacterial host defense in murine intestinal epithelial cells. *PLoS Pathog* **5**, e1000567.

Dumic, J., Dabelic, S. & Flogel, M. (2006). Galectin-3: an open-ended story. *Biochim Biophys Acta* **1760**, 616-635.

Egan, W., Tsui, F. P. & Schneerson, R. (1980). Structural studies of the *Haemophilus influenzae* type f capsular polysaccharide. *Carbohydr Res* **79**, 271-277.

Elswaifi, S. F., St Michael, F., Sreenivas, A., Cox, A., Carman, G. M. & Inzana, T. J. (2009). Molecular characterization of phosphorylcholine expression on the lipooligosaccharide of *Histophilus somni*. *Microb Pathog* **47**, 223-230.

Erwin, A. L., Nelson, K. L., Mhlanga-Mutangadura, T. & other authors (2005). Characterization of genetic and phenotypic diversity of invasive nontypeable *Haemophilus influenzae*. *Infect Immun* **73**, 5853-5863.

Erwin, A. L., Bonthuis, P. J., Geelhood, J. L., Nelson, K. L., McCrea, K. W., Gilsdorf, J. R. & Smith, A. L. (2006). Heterogeneity in tandem octanucleotides within *Haemophilus influenzae* lipopolysaccharide biosynthetic gene *losA* affects serum resistance. *Infect Immun* **74**, 3408-3414.

Euzéby, J. P. (2010). List of Prokaryotic names with Standing in Nomenclature. http://www.bacterio.net.

Evans, N. M., Smith, D. D. & Wicken, A. J. (1974). Haemin and nicotinamide adenine dinucleotide requirements of *Haemophilus influenzae* and *Haemophilus parainfluenzae*. *J Med Microbiol* **7**, 359-365.

Farrell, D. J., Sader, H. S., Castanheira, M., Biedenbach, D. J., Rhomberg, P. R. & Jones, R. N. (2010). Antimicrobial characterisation of CEM-101 activity against respiratory tract pathogens, including multidrug-resistant pneumococcal serogroup 19A isolates. *Int J Antimicrob Agents* **35**, 537-543.

Fernaays, M. M., Lesse, A. J., Cai, X. & Murphy, T. F. (2006). Characterization of *igaB*, a second immunoglobulin A1 protease gene in nontypeable *Haemophilus influenzae*. *Infect Immun* **74**, 5860-5870.

Fleischmann, R. D., Adams, M. D., White, O. & other authors (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496-512.

Fowler, M., Thomas, R. J., Atherton, J., Roberts, I. S. & High, N. J. (2006). Galectin-3 binds to *Helicobacter pylori* O-antigen: it is upregulated and rapidly secreted by gastric epithelial cells in response to *H. pylori* adhesion. *Cell Microbiol* **8**, 44-54.

Fox, K. L., Yildirim, H. H., Deadman, M. E., Schweda, E. K., Moxon, E. R. & Hood, D. W. (2005). Novel lipopolysaccharide biosynthetic genes containing tetranucleotide repeats in *Haemophilus influenzae*, identification of a gene for adding O-acetyl groups. *Mol Microbiol* **58**, 207-216.

Fox, K. L., Cox, A. D., Gilbert, M., Wakarchuk, W. W., Li, J., Makepeace, K., Richards, J. C., Moxon, E. R. & Hood, D. W. (2006). Identification of a bifunctional lipopolysaccharide sialyltransferase in *Haemophilus influenzae*: incorporation of disialic acid. *J Biol Chem* **281**, 40024-40032.

Fox, K. L., Li, J., Schweda, E. K., Vitiazeva, V., Makepeace, K., Jennings, M. P., Moxon, E. R. & Hood, D. W. (2008). Duplicate copies of lic1 direct the addition of multiple phosphocholine residues in the lipopolysaccharide of *Haemophilus influenzae*. *Infect Immun* **76**, 588-600.

Franco, A. V., Liu, D. & Reeves, P. R. (1998). The Wzz (Cld) protein in *Escherichia coli*: amino acid sequence variation determines O-antigen chain length specificity. *J Bacteriol* **180**, 2670-2675.

Frankard, J., Rodriguez-Villalobos, H., Struelens, M. J. & Jacobs, F. (2004). *Haemophilus parainfluenzae*: an underdiagnosed pathogen of biliary tract infections? *Eur J Clin Microbiol Infect Dis* **23**, 46-48.

Fujise, O., Wang, Y., Chen, W. & Chen, C. (2008). Adherence of *Aggregatibacter actinomycetemcomitans* via serotype-specific polysaccharide antigens in lipopolysaccharides. *Oral Microbiol Immunol* **23**, 226-233.

Fujita, K., Hirano, T., Kodama, S. & Suzuki, M. (2009). Prognostic impact of phosphorylcholine expression in nontypeable *Haemophilus influenzae* in otitis media with effusion. *Acta Otolaryngol* **129**, 832-838.

Genco, R. J., Zambon, J. J. & Murray, P. A. (1985). Serum and gingival fluid antibodies as adjuncts in the diagnosis of *Actinobacillus actinomycetemcomitans*-associated periodontal disease. *J Periodontol* **56**, 41-50.

Goebel, E. M., Wolfe, D. N., Elder, K., Stibitz, S. & Harvill, E. T. (2008). O antigen protects *Bordetella parapertussis* from complement. *Infect Immun* **76**, 1774-1780.

Greenberg, D., Broides, A., Blancovich, I., Peled, N., Givon-Lavi, N. & Dagan, R. (2004). Relative importance of nasopharyngeal versus oropharyngeal sampling for isolation of *Streptococcus pneumoniae* and *Haemophilus influenzae* from healthy and sick individuals varies with age. *J Clin Microbiol* **42**, 4604-4609.

Griffin, R., Bayliss, C. D., Herbert, M. A., Cox, A. D., Makepeace, K., Richards, J. C., Hood, D. W. & Moxon, E. R. (2005). Digalactoside expression in the lipopolysaccharide of *Haemophilus influenzae* and its role in intravascular survival. *Infect Immun* **73**, 7022-7026.

Gromkova, R. & Goodgal, S. (1979). Transformation by plasmid and chromosomal DNAs in *Haemophilus parainfluenzae*. *Biochem Biophys Res Commun* **88**, 1428-1434.

Gromkova, R. C., Mottalini, T. C. & Dove, M. G. (1998). Genetic transformation in *Haemophilus parainfluenzae* clinical isolates. *Curr Microbiol* **37**, 123-126.

Gronow, S., Brabetz, W., Lindner, B. & Brade, H. (2005). OpsX from *Haemophilus influenzae* represents a novel type of heptosyltransferase I in lipopolysaccharide biosynthesis. *J Bacteriol* **187**, 6242-6247.

Gunn, J. S. (2001). Bacterial modification of LPS and resistance to antimicrobial peptides. *J Endotoxin Res* **7**, 57-62.

Hallstrom, T., Jarva, H., Riesbeck, K. & Blom, A. M. (2007). Interaction with C4b-binding protein contributes to nontypeable *Haemophilus influenzae* serum resistance. *J Immunol* **178**, 6359-6366.

Hallstrom, T. & Riesbeck, K. (2010). *Haemophilus influenzae* and the complement system. *Trends Microbiol* **18**, 258-265.

Hamed, K. A., Dormitzer, P. R., Su, C. K. & Relman, D. A. (1994). *Haemophilus parainfluenzae* endocarditis: application of a molecular approach for identification of pathogenic bacterial species. *Clin Infect Dis* **19**, 677-683.

Hammond, B. F., Lillard, S. E. & Stevens, R. H. (1987). A bacteriocin of *Actinobacillus actinomycetemcomitans*. *Infect Immun* **55**, 686-691.

Harper, M., Boyce, J. D., Cox, A. D., St Michael, F., Wilkie, I. W., Blackall, P. J. & Adler, B. (2007). *Pasteurella multocida* expresses two lipopolysaccharide glycoforms simultaneously, but only a single form is required for virulence: identification of two acceptor-specific heptosyl I transferases. *Infect Immun* **75**, 3885-3893.

Hashida, K., Shiomori, T., Hohchi, N., Muratani, T., Mori, T., Udaka, T. & Suzuki, H. (2008). Nasopharyngeal *Haemophilus influenzae* carriage in Japanese children attending day-care centers. *J Clin Microbiol* **46**, 876-881.

Helander, I. M., Lindner, B., Brade, H., Altmann, K., Lindberg, A. A., Rietschel, E. T. & Zahringer, U. (1988). Chemical structure of the lipopolysaccharide of *Haemophilus influenzae* strain I-69 Rd-/b+. Description of a novel deep-rough chemotype. *Eur J Biochem* **177**, 483-492.

Hendrixson, D. R. (2006). A phase-variable mechanism controlling the *Campylobacter jejuni* FlgR response regulator influences commensalism. *Mol Microbiol* **61**, 1646-1659.

Herbert, M., Sauer, E., Smethurst, G., Kraiss, A., Hilpert, A. K. & Reidl, J. (2003). Nicotinamide ribosyl uptake mutants in *Haemophilus influenzae*. *Infect Immun* **71**, 5398-5401.

Herriott, R. M., Meyer, E. M. & Vogt, M. (1970). Defined nongrowth media for stage II development of competence in *Haemophilus influenzae*. *J Bacteriol* **101**, 517-524.

Hill, D. J., Griffiths, N. J., Borodina, E. & Virji, M. (2010). Cellular and molecular biology of *Neisseria meningitidis* colonization and invasive disease. *Clin Sci (Lond)* **118**, 547-564.

Hirabayashi, J., Hashidate, T., Arata, Y. & other authors (2002). Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. *Biochim Biophys Acta* **1572**, 232-254.

Hogg, J. S., Hu, F. Z., Janto, B., Boissy, R., Hayes, J., Keefe, R., Post, J. C. & Ehrlich, G. D. (2007). Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. *Genome Biol* **8**, R103.

Hoiseth, S. K., Moxon, E. R. & Silver, R. P. (1986). Genes involved in *Haemophilus influenzae* type b capsule expression are part of an 18-kilobase tandem duplication. *Proc Natl Acad Sci U S A* **83**, 1106-1110.

Hojo, K., Nagaoka, S., Ohshima, T. & Maeda, N. (2009). Bacterial interactions in dental biofilm development. *J Dent Res* **88**, 982-990.

Holzer, S. U., Schlumberger, M. C., Jackel, D. & Hensel, M. (2009). Effect of the O-antigen length of lipopolysaccharide on the functions of Type III secretion systems in *Salmonella enterica*. *Infect Immun* **77**, 5458-5470.

Hood, D. W., Makepeace, K., Deadman, M. E., Rest, R. F., Thibault, P., Martin, A., Richards, J. C. & Moxon, E. R. (1999). Sialic acid in the lipopolysaccharide of *Haemophilus influenzae*: strain distribution, influence on serum resistance and structural characterization. *Mol Microbiol* **33**, 679-692.

Hood, D. W., Cox, A. D., Gilbert, M. & other authors (2001a). Identification of a lipopolysaccharide alpha-2,3-sialyltransferase from *Haemophilus influenzae*. *Mol Microbiol* **39**, 341-350.

Hood, D. W., Cox, A. D., Wakarchuk, W. W. & other authors (2001b). Genetic basis for expression of the major globotetraose-containing lipopolysaccharide from *H. influenzae* strain Rd (RM118). *Glycobiology* **11**, 957-967.

Hood, D. W., Deadman, M. E., Cox, A. D., Makepeace, K., Martin, A., Richards, J. C. & Moxon, E. R. (2004a). Three genes, *lgtF*, *lic2C* and *lpsA*, have a primary role in determining the pattern of oligosaccharide extension from the inner core of *Haemophilus influenzae* LPS. *Microbiology* **150**, 2089-2097.

Hood, D. W., Randle, G., Cox, A. D., Makepeace, K., Li, J., Schweda, E. K., Richards, J. C. & Moxon, E. R. (2004b). Biosynthesis of cryptic lipopolysaccharide glycoforms in *Haemophilus influenzae* involves a mechanism similar to that required for O-antigen synthesis. *J Bacteriol* **186**, 7429-7439.

Hood, D. W., Deadman, M. E., Engskog, M. K., Vitiazeva, V., Makepeace, K., Schweda, E. K. & Moxon, R. (2010). Genes required for the synthesis of heptose-containing oligosaccharide outer core extensions in *Haemophilus influenzae* lipopolysaccharide. *Microbiology*.

HPA (2009). Identification of *Haemophilus* species and the HACEK group of organisms. National Standard Method BSOP ID 12 Issue 2.

Hug, I., Couturier, M. R., Rooker, M. M., Taylor, D. E., Stein, M. & Feldman, M. F. (2010). *Helicobacter pylori* lipopolysaccharide is synthesized via a novel pathway with an evolutionary connection to protein N-glycosylation. *PLoS Pathog* **6**, e1000819.

Hyams, C., Camberlein, E., Cohen, J. M., Bax, K. & Brown, J. S. (2010). The *Streptococcus pneumoniae* capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. *Infect Immun* **78**, 704-715.

Jann, B., Shashkov, A., Torgov, V., Kochanowski, H., Seltmann, G. & Jann, K. (1995). NMR investigation of the 6-deoxy-L-talose-containing O45, O45-related (O45rel), and O66 polysaccharides of *Escherichia coli*. *Carbohydr Res* **278**, 155-165.

Jenkins, G. A., Figueira, M., Kumar, G. A., Sweetman, W. A., Makepeace, K., Pelton, S. I., Moxon, R. & Hood, D. W. (2010). Sialic acid mediated transcriptional modulation of a highly conserved sialometabolism gene cluster in *Haemophilus influenzae* and its effect on virulence. *BMC Microbiol* **10**, 48.

Jimenez, N., Canals, R., Salo, M. T., Vilches, S., Merino, S. & Tomas, J. M. (2008). The *Aeromonas hydrophila* wb*O34 gene cluster: genetics and temperature regulation. *J Bacteriol* **190**, 4198-4209.

Johansen, E. B., Szoka, F. C., Zaleski, A., Apicella, M. A. & Gibson, B. W. (2010). Utilizing the O-antigen lipopolysaccharide biosynthesis pathway in *Escherichia coli* to interrogate the substrate specificities of exogenous glycosyltransferase genes in a combinatorial approach. *Glycobiology* **20**, 763-774.

John, C. M., Jarvis, G. A., Swanson, K. V., Leffler, H., Cooper, M. D., Huflejt, M. E. & Griffiss, J. M. (2002). Galectin-3 binds lactosaminylated lipooligosaccharides from *Neisseria gonorrhoeae* and is selectively expressed by mucosal epithelial cells that are infected. *Cell Microbiol* **4**, 649-662.

Johnson, T. A., Qiu, J., Plaut, A. G. & Holyoak, T. (2009). Active-site gating regulates substrate selectivity in a chymotrypsin-like serine protease the structure of *Haemophilus influenzae* immunoglobulin A1 protease. *J Mol Biol* **389**, 559-574.

Johnston, J. W., Shamsulddin, H., Miller, A. F. & Apicella, M. A. (2010). Sialic acid transport and catabolism are cooperatively regulated by SiaR and CRP in nontypeable *Haemophilus influenzae*. *BMC Microbiol* **10**, 240.

Jones, R. N., Bell, J. M., Sader, H. S., Turnidge, J. D. & Stilwell, M. G. (2009). In vitro potency of doripenem tested against an international collection of rarely isolated bacterial pathogens. *Diagn Microbiol Infect Dis* **63**, 434-439.

Juhas, M., Power, P. M., Harding, R. M. & other authors (2007). Sequence and functional analyses of *Haemophilus* spp. genomic islands. *Genome Biol* **8**, R237.

Kahn, M. E. & Gromkova, R. (1981). Occurrence of pili on and adhesive properties of *Haemophilus parainfluenzae*. *J Bacteriol* **145**, 1075-1078.

Kaplan, J. B., Perry, M. B., MacLean, L. L., Furgang, D., Wilson, M. E. & Fine, D. H. (2001). Structural and genetic analyses of O polysaccharide from *Actinobacillus actinomycetemcomitans* serotype f. *Infect Immun* **69**, 5375-5384.

Katz, C. & Ron, E. Z. (2008). Dual role of FtsH in regulating lipopolysaccharide biosynthesis in *Escherichia coli*. *J Bacteriol* **190**, 7117-7122.

Kemmer, G., Reilly, T. J., Schmidt-Brauns, J. & other authors (2001). NadN and *e* (P4) are essential for utilization of NAD and nicotinamide mononucleotide but not nicotinamide riboside in *Haemophilus influenzae*. *J Bacteriol* **183**, 3974-3981.

Kenne, L., Lindberg, B. & Petersson, K. (1980). Structural studies of the O-specific side-chains of the *Shigella sonnei* phase I lipopolysaccharide. *Carbohydr Res* **78**, 119-126.

Khan, S. A., Everest, P., Servos, S. & other authors (1998). A lethal role for lipid A in *Salmonella* infections. *Mol Microbiol* **29**, 571-579.

Kilian, M. (1976). A taxonomic study of the genus *Haemophilus*, with the proposal of a new species. *J Gen Microbiol* **93**, 9-62.

Kilian, M., Prachyabrued, W. & Theilade, E. (1976). Haemophili in developing dental plaque. *Scand J Dent Res* **84**, 16-19.

Kilian, M., Mestecky, J. & Schrohenloher, R. E. (1979). Pathogenic species of the genus *Haemophilus* and *Streptococcus pneumoniae* produce immunoglobulin A1 protease. *Infect Immun* **26**, 143-149.

Kim, T. H., Sebastian, S., Pinkham, J. T., Ross, R. A., Blalock, L. T. & Kasper, D. L. (2010). Characterization of the O-antigen polymerase (Wzy) of *Francisella tularensis*. *J Biol Chem* **285**, 27839-27849.

King, J. D., Vinogradov, E., Preston, A., Li, J. & Maskell, D. J. (2009). Post-assembly modification of *Bordetella bronchiseptica* O polysaccharide by a novel periplasmic enzyme encoded by *wbmE*. *J Biol Chem* **284**, 1474-1483.

Klein, J. O. (1997). Role of nontypeable *Haemophilus influenzae* in pediatric respiratory tract infections. *Pediatr Infect Dis J* **16**, S5-8.

Kraehenbuhl, J. P. & Neutra, M. R. (1992). Molecular and cellular basis of immune protection of mucosal surfaces. *Physiol Rev* **72**, 853-879.

Kroll, J. S., Loynds, B., Brophy, L. N. & Moxon, E. R. (1990). The *bex* locus in encapsulated *Haemophilus influenzae*: a chromosomal region involved in capsule polysaccharide export. *Mol Microbiol* **4**, 1853-1862.

Lai, C. H., Bloomquist, C. & Liljemark, W. F. (1990). Purification and characterization of an outer membrane protein adhesin from *Haemophilus parainfluenzae* HP-28. *Infect Immun* **58**, 3833-3839.

Lehrer, J., Vigeant, K. A., Tatar, L. D. & Valvano, M. A. (2007). Functional characterization and membrane topology of *Escherichia coli* WecA, a sugar-phosphate transferase initiating the biosynthesis of enterobacterial common antigen and O-antigen lipopolysaccharide. *J Bacteriol* **189**, 2618-2628.

Lesse, A. J., Campagnari, A. A., Bittner, W. E. & Apicella, M. A. (1990). Increased resolution of lipopolysaccharides and lipooligosaccharides utilizing tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *J Immunol Methods* **126**, 109-117.

Liljemark, W. F., Bloomquist, C. G. & Lai, C. H. (1992). Clustering of an outer membrane adhesin of *Haemophilus parainfluenzae*. *Infect Immun* **60**, 687-689.

LiPuma, J. J., Richman, H. & Stull, T. L. (1990). Haemocin, the bacteriocin produced by *Haemophilus influenzae*: species distribution and role in colonization. *Infect Immun* **58**, 1600-1605.

Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. N., Wang, Q., Reeves, P. R. & Wang, L. (2008). Structure and genetics of *Shigella* O antigens. *FEMS Microbiol Rev* **32**, 627-653.

Lundborg, M., Modhukur, V. & Widmalm, G. (2010). Glycosyltransferase functions of *E. coli* O-antigens. *Glycobiology* **20**, 366-368.

Lwoff, A. & Lwoff, M. (1937). Studies on codehydrogenases. I. Nature of growth factor "V". *Proceedings of the Royal Society of London Series B, Biological Sciences* **122**, 352-359.

Lysenko, E., Richards, J. C., Cox, A. D., Stewart, A., Martin, A., Kapoor, M. & Weiser, J. N. (2000a). The position of phosphorylcholine on the lipopolysaccharide of *Haemophilus influenzae* affects binding and sensitivity to C-reactive protein-mediated killing. *Mol Microbiol* **35**, 234-245.

Lysenko, E. S., Gould, J., Bals, R., Wilson, J. M. & Weiser, J. N. (2000b). Bacterial phosphorylcholine decreases susceptibility to the antimicrobial peptide LL-37/hCAP18 expressed in the upper respiratory tract. *Infect Immun* **68**, 1664-1671.

MacLean, L. L., Perry, M. B. & Vinogradov, E. (2004). Characterization of the antigenic lipopolysaccharide O chain and the capsular polysaccharide produced by *Actinobacillus pleuropneumoniae* serotype 13. *Infect Immun* **72**, 5925-5930.

MacLennan, C. A., Gilchrist, J. J., Gordon, M. A. & other authors (2010). Dysregulated humoral immunity to nontyphoidal *Salmonella* in HIV-infected African adults. *Science* **328**, 508-512.

Marchler-Bauer, A., Anderson, J. B., Chitsaz, F. & other authors (2009). CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* **37**, D205-210.

Marin, A., Monso, E., Garcia-Nunez, M., Sauleda, J., Noguera, A., Pons, J., Agusti, A. & Morera, J. (2010). Variability and effects of bronchial colonisation in patients with moderate COPD. *Eur Respir J* **35**, 295-302.

Martin, P., Sun, L., Hood, D. W. & Moxon, E. R. (2004). Involvement of genes of genome maintenance in the regulation of phase variation frequencies in *Neisseria meningitidis*. *Microbiology* **150**, 3001-3012.

Mason, K. M., Munson, R. S., Jr. & Bakaletz, L. O. (2003). Nontypeable *Haemophilus influenzae* gene expression induced in vivo in a chinchilla model of otitis media. *Infect Immun* **71**, 3454-3462.

Masoud, H., Weintraub, S. T., Wang, R., Cotter, R. & Holt, S. C. (1991). Investigation of the structure of lipid A from *Actinobacillus actinomycetemcomitans* strain Y4 and human clinical isolate PO 1021-7. *Eur J Biochem* **200**, 775-781.

Masoud, H., Uhrin, D., Moxon, E. R. & Richards, J. C. (2008). Identification of a novel structural motif in the lipopolysaccharide of the *galE/galK* double mutant of *Haemophilus influenzae* strain Eagan. *Carbohydr Res*.

Maughan, H. & Redfield, R. J. (2009a). Extensive variation in natural competence in *Haemophilus influenzae*. *Evolution* **63**, 1852-1866.

Maughan, H. & Redfield, R. J. (2009b). Tracing the evolution of competence in *Haemophilus influenzae*. *PLoS One* **4**, e5854.

McCrea, K. W., Xie, J., Marrs, C. F. & Gilsdorf, J. R. (2010). Prevalence of genetic differences in phosphorylcholine expression between nontypeable *Haemophilus influenzae* and *Haemophilus haemolyticus*. *BMC Microbiol* **10**, 286.

McDonald, C. L., Crafton, E. M., Covin, F. A., Massey, C. V. & Ramsey, K. M. (1994). Pericarditis: a probable complication of endocarditis due to *Haemophilus influenzae*. *Clin Infect Dis* **18**, 648-649.

McLinn, S. E., Nelson, J. D. & Haltalin, K. C. (1970). Antimicrobial susceptibility of *Hemophilus influenzae*. *Pediatrics* **45**, 827-838.

Meier-Dieter, U., Barr, K., Starman, R., Hatch, L. & Rick, P. D. (1992). Nucleotide sequence of the *Escherichia coli rfe* gene involved in the synthesis of enterobacterial common antigen. Molecular cloning of the *rfe-rff* gene cluster. *J Biol Chem* **267**, 746-753.

Mikhail, I., Yildirim, H. H., Lindahl, E. C. & Schweda, E. K. (2005). Structural characterization of lipid A from nontypeable and type f *Haemophilus influenzae*: variability of fatty acid substitution. *Anal Biochem* **340**, 303-316.

Miller, W. L., Matewish, M. J., McNally, D. J., Ishiyama, N., Anderson, E. M., Brewer, D., Brisson, J. R., Berghuis, A. M. & Lam, J. S. (2008). Flagellin glycosylation in *Pseudomonas aeruginosa* PAK requires the O-antigen biosynthesis enzyme WbpO. *J Biol Chem* **283**, 3507-3518.

Mitchell, M. A., Skowronek, K., Kauc, L. & Goodgal, S. H. (1991). Electroporation of *Haemophilus influenzae* is effective for transformation of plasmid but not chromosomal DNA. *Nucleic Acids Res* **19**, 3625-3628.

Miyamoto, N. & Bakaletz, L. O. (1996). Selective adherence of non-typeable *Haemophilus influenzae* (NTHi) to mucus or epithelial cells in the chinchilla eustachian tube and middle ear. *Microb Pathog* **21**, 343-356.

Morona, R., Purins, L., Tocilj, A., Matte, A. & Cygler, M. (2009). Sequence-structure relationships in polysaccharide co-polymerase (PCP) proteins. *Trends Biochem Sci* **34**, 78-84.

Morton, D. J. & Williams, P. (1989). Characterization of the outer-membrane proteins of *Haemophilus parainfluenzae* expressed under iron-sufficient and iron-restricted conditions. *J Gen Microbiol* **135**, 445-451.

Murphy, T. F., Bakaletz, L. O. & Smeesters, P. R. (2009). Microbial interactions in the respiratory tract. *Pediatr Infect Dis J* **28**, S121-126.

Nakano, Y., Yoshida, Y., Yamashita, Y. & Koga, T. (1998). A gene cluster for 6-deoxy-L-talan synthesis in *Actinobacillus actinomycetemcomitans*. *Biochim Biophys Acta* **1442**, 409-414.

Nakano, Y., Suzuki, N., Yoshida, Y., Nezu, T., Yamashita, Y. & Koga, T. (2000). Thymidine diphosphate-6-deoxy-L-lyxo-4-hexulose reductase synthesizing dTDP-6-deoxy-L-talose from *Actinobacillus actinomycetemcomitans*. *J Biol Chem* **275**, 6806-6812.

Nassau, P. M., Martin, S. L., Brown, R. E., Weston, A., Monsey, D., McNeil, M. R. & Duncan, K. (1996). Galactofuranose biosynthesis in *Escherichia coli* K-12: identification and cloning of UDP-galactopyranose mutase. *J Bacteriol* **178**, 1047-1052.

Noel, G. J., Hoiseth, S. K. & Edelson, P. J. (1992). Type b capsule inhibits ingestion of *Haemophilus influenzae* by murine macrophages: studies with isogenic encapsulated and unencapsulated strains. *J Infect Dis* **166**, 178-182.

Noel, G. J., Brittingham, A., Granato, A. A. & Mosser, D. M. (1996). Effect of amplification of the Cap b locus on complement-mediated bacteriolysis and opsonization of type b *Haemophilus influenzae*. *Infect Immun* **64**, 4769-4775.

Norskov-Lauritsen, N., Bruun, B. & Kilian, M. (2005). Multilocus sequence phylogenetic study of the genus *Haemophilus* with description of *Haemophilus pittmaniae* sp. nov. *Int J Syst Evol Microbiol* **55**, 449-456.

Norskov-Lauritsen, N. & Kilian, M. (2006). Reclassification of *Actinobacillus actinomycetemcomitans*, *Haemophilus aphrophilus*, *Haemophilus paraphrophilus* and *Haemophilus segnis* as *Aggregatibacter actinomycetemcomitans* gen. nov., comb. nov., *Aggregatibacter aphrophilus* comb. nov. and *Aggregatibacter segnis* comb. nov., and emended

description of *Aggregatibacter aphrophilus* to include V factor-dependent and V factor-independent isolates. *Int J Syst Evol Microbiol* **56**, 2135-2146.

Numazaki, K., Chiba, S., Umetsu, M. & other authors (2004). Etiological agents of lower respiratory tract infections in Japanese children. *In Vivo* **18**, 67-71.

Olsen, L. R., Huang, B., Vetting, M. W. & Roderick, S. L. (2004). Structure of serine acetyltransferase in complexes with CoA and its cysteine feedback inhibitor. *Biochemistry* **43**, 6013-6019.

Olsson, U., Lycknert, K., Stenutz, R., Weintraub, A. & Widmalm, G. (2005). Structural analysis of the O-antigen polysaccharide from *Escherichia coli* O152. *Carbohydr Res* **340**, 167-171.

Ortega, X., Hunt, T. A., Loutet, S., Vinion-Dubiel, A. D., Datta, A., Choudhury, B., Goldberg, J. B., Carlson, R. & Valvano, M. A. (2005). Reconstitution of O-specific lipopolysaccharide expression in *Burkholderia cenocepacia* strain J2315, which is associated with transmissible infections in patients with cystic fibrosis. *J Bacteriol* **187**, 1324-1333.

Palmer, G. G. (1981). *Haemophili* in faeces. *J Med Microbiol* **14**, 147-150.

Pan, F., Jackson, M., Ma, Y. & McNeil, M. (2001). Cell wall core galactofuran synthesis is essential for growth of mycobacteria. *J Bacteriol* **183**, 3991-3998.

Paton, A. W., Morona, R. & Paton, J. C. (2000). A new biological agent for treatment of Shiga toxigenic *Escherichia coli* infections and dysentery in humans. *Nat Med* **6**, 265-270.

Perepelov, A. V., Li, D., Liu, B. & other authors (2009). Structural and genetic characterization of *Escherichia coli* O99 antigen. *FEMS Immunol Med Microbiol* **57**, 80-87.

Pericone, C. D., Overweg, K., Hermans, P. W. & Weiser, J. N. (2000). Inhibitory and bactericidal effects of hydrogen peroxide production by *Streptococcus pneumoniae* on other inhabitants of the upper respiratory tract. *Infect Immun* **68**, 3990-3997.

Perry, M. B., MacLean, L. M., Brisson, J. R. & Wilson, M. E. (1996). Structures of the antigenic O-polysaccharides of lipopolysaccharides produced by *Actinobacillus actinomycetemcomitans* serotypes a, c, d and e. *Eur J Biochem* **242**, 682-688.

Pettigrew, M. M., Foxman, B., Marrs, C. F. & Gilsdorf, J. R. (2002). Identification of the lipooligosaccharide biosynthesis gene *lic2B* as a putative virulence factor in strains of nontypeable *Haemophilus influenzae* that cause otitis media. *Infect Immun* **70**, 3551-3556.

Phillips, N. J., McLaughlin, R., Miller, T. J., Apicella, M. A. & Gibson, B. W. (1996). Characterization of two transposon mutants from *Haemophilus influenzae* type b with altered lipooligosaccharide biosynthesis. *Biochemistry* **35**, 5937-5947.

Pier, G. B. (2007). *Pseudomonas aeruginosa* lipopolysaccharide: a major virulence factor, initiator of inflammation and target for effective immunity. *Int J Med Microbiol* **297**, 277-295.

Pillai, A., Mitchell, J. L., Hill, S. L. & Stockley, R. A. (2000). A case of *Haemophilus parainfluenzae* pneumonia. *Thorax* **55**, 623-624.

Poje, G. & Redfield, R. J. (2003). Transformation of *Haemophilus influenzae*. *Methods Mol Med* **71**, 57-70.

Pollard, A., St Michael, F., Connor, L., Nichols, W. & Cox, A. (2008). Structural characterization of *Haemophilus parainfluenzae* lipooligosaccharide and elucidation of its role in adherence using an outer core mutant. *Can J Microbiol* **54**, 906-917.

Pollock, T. J., Thorne, L., Yamazaki, M., Mikolajczak, M. J. & Armentrout, R. W. (1994). Mechanism of bacitracin resistance in gram-negative bacteria that synthesize exopolysaccharides. *J Bacteriol* **176**, 6229-6237.

Poulsen, K., Brandt, J., Hjorth, J. P., Thogersen, H. C. & Kilian, M. (1989). Cloning and sequencing of the immunoglobulin A1 protease gene (*iga*) of *Haemophilus influenzae* serotype b. *Infect Immun* **57**, 3097-3105.

Power, P. M., Sweetman, W. A., Gallacher, N. J., Woodhall, M. R., Kumar, G. A., Moxon, E. R. & Hood, D. W. (2009). Simple sequence repeats in *Haemophilus influenzae*. *Infect Genet Evol* **9**, 216-228.

Pragani, R., Stallforth, P. & Seeberger, P. H. (2010). De novo synthesis of a 2-acetamido-4-amino-2,4,6-trideoxy-D-galactose (AAT) building block for the preparation of a *Bacteroides fragilis* A1 polysaccharide fragment. *Org Lett* **12**, 1624-1627.

Preston, A., Petersen, B. O., Duus, J. O., Kubler-Kielb, J., Ben-Menachem, G., Li, J. & Vinogradov, E. (2006). Complete structures of *Bordetella bronchiseptica* and *Bordetella parapertussis* lipopolysaccharides. *J Biol Chem* **281**, 18135-18144.

Quentin, R., Musser, J. M., Mellouet, M., Sizaret, P. Y., Selander, R. K. & Goudeau, A. (1989). Typing of urogenital, maternal, and neonatal isolates of *Haemophilus influenzae* and *Haemophilus parainfluenzae* in correlation with clinical source of isolation and evidence for a genital specificity of *H. influenzae* biotype IV. *J Clin Microbiol* **27**, 2286-2294.

Raetz, C. R. & Whitfield, C. (2002). Lipopolysaccharide endotoxins. *Annu Rev Biochem* **71**, 635-700.

Raoult, D., Drancourt, M., Gallais, H., Casanova, P. & Lesavre, P. (1987). *Haemophilus parainfluenzae* meningitis in an adult with an inherited deficiency of the seventh component of complement. *Arch Intern Med* **147**, 2214.

Redfield, R. J., Cameron, A. D., Qian, Q., Hinds, J., Ali, T. R., Kroll, J. S. & Langford, P. R. (2005). A novel CRP-dependent regulon controls expression of competence genes in *Haemophilus influenzae*. *J Mol Biol* **347**, 735-747.

Redfield, R. J., Findlay, W. A., Bosse, J., Kroll, J. S., Cameron, A. D. & Nash, J. H. (2006). Evolution of competence and DNA uptake specificity in the Pasteurellaceae. *BMC Evol Biol* **6**, 82.

Reizer, J., Reizer, A. & Saier, M. H., Jr. (1992). A new subfamily of bacterial ABC-type transport systems catalyzing export of drugs and carbohydrates. *Protein Sci* **1**, 1326-1332.

Riley, M. A. & Wertz, J. E. (2002). Bacteriocins: evolution, ecology, and application. *Annu Rev Microbiol* **56**, 117-137.

Risberg, A., Masoud, H., Martin, A., Richards, J. C., Moxon, E. R. & Schweda, E. K. (1999). Structural analysis of the lipopolysaccharide oligosaccharide epitopes expressed by a capsule-deficient strain of *Haemophilus influenzae* Rd. *Eur J Biochem* **261**, 171-180.

Rivers, T. M. (1922). Influenza-like bacilli. Growth of influenza-like bacilli on media containing only an autoclave-labile substance as an accessory food factor. *Bull Johns Hopkins Hosp* **33**, 429-431.

Roberts, M. C., Mintz, C. S. & Morse, S. A. (1986). Characterization of *Haemophilus parainfluenzae* strains with low-Mr or ladder-like lipopolysaccharides. *J Gen Microbiol* **132**, 611-616.

Rock, C. O., Heath, R. J., Park, H. W. & Jackowski, S. (2001). The *licC* gene of *Streptococcus pneumoniae* encodes a CTP:phosphocholine cytidylyltransferase. *J Bacteriol* **183**, 4927-4931.

Russa, R., Urbanik-Sypniewska, T., Shashkov, A. S., Kochanowski, H. & Mayer, H. (1995). The structure of the homopolymeric O-specific chain from the phenol soluble LPS of the *Rhizobium loti* type strain NZP2213. *Carbohydrate Polymers* **27**, 299-303.

Russo, T. A., Beanan, J. M., Olson, R., MacDonald, U. & Cope, J. J. (2009). Capsular polysaccharide and the O-specific antigen impede antibody binding: a potential obstacle for the successful development of an extraintestinal pathogenic *Escherichia coli* vaccine. *Vaccine* **27**, 388-395.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945.

Salako, N. O. & Kleinberg, I. (1989). Incidence of selected ureolytic bacteria in human dental plaque from sites with differing salivary access. *Arch Oral Biol* **34**, 787-791.

Salaun, L., Linz, B., Suerbaum, S. & Saunders, N. J. (2004). The diversity within an expanded and redefined repertoire of phase-variable genes in *Helicobacter pylori*. *Microbiology* **150**, 817-830.

Saldias, M. S., Patel, K., Marolda, C. L., Bittner, M., Contreras, I. & Valvano, M. A. (2008). Distinct functional domains of the *Salmonella enterica* WbaP transferase that is involved in the initiation reaction for synthesis of the O antigen subunit. *Microbiology* **154**, 440-453.

Saldias, M. S., Ortega, X. & Valvano, M. A. (2009). *Burkholderia cenocepacia* O antigen lipopolysaccharide prevents phagocytosis by macrophages and adhesion to epithelial cells. *J Med Microbiol* **58**, 1542-1548.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning*. New York: Cold Spring Harbor Laboratory Press.

Samuel, G. & Reeves, P. (2003). Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res* **338**, 2503-2519.

Sanabria-Valentin, E., Colbert, M. T. & Blaser, M. J. (2007). Role of *futC* slipped strand mispairing in *Helicobacter pylori* Lewis y phase variation. *Microbes Infect* **9**, 1553-1560.

Sanders, D. A., Staines, A. G., McMahon, S. A., McNeil, M. R., Whitfield, C. & Naismith, J. H. (2001). UDP-galactopyranose mutase has a novel structure and mechanism. *Nat Struct Biol* **8**, 858-863.

Santamaria, M., Gutierrez-Navarro, A. M. & Corzo, J. (1998). Lipopolysaccharide profiles from nodules as markers of *Bradyrhizobium* strains nodulating wild legumes. *Appl Environ Microbiol* **64**, 902-906.

Satola, S. W., Schirmer, P. L. & Farley, M. M. (2003). Genetic analysis of the capsule locus of *Haemophilus influenzae* serotype f. *Infect Immun* **71**, 7202-7207.

Scheifele, D. W. & Fussell, S. J. (1981). Frequency of ampicillin-resistant *Haemophilus parainfluenzae* in children. *J Infect Dis* **143**, 495-498.

Schenkein, H. A., Barbour, S. E., Berry, C. R., Kipps, B. & Tew, J. G. (2000). Invasion of human vascular endothelial cells by *Actinobacillus actinomycetemcomitans* via the receptor for platelet-activating factor. *Infect Immun* **68**, 5416-5419.

Schweda, E. K., Richards, J. C., Hood, D. W. & Moxon, E. R. (2007). Expression and structural diversity of the lipopolysaccharide of *Haemophilus influenzae*: implication in virulence. *Int J Med Microbiol* **297**, 297-306.

Senior, B. W. & Ip, C. L. (1999). An investigation into the influences of species and biotype on the type of IgA1 protease produced by isolates of *Haemophilus*. *J Med Microbiol* **48**, 389-394.

Sethi, S. & Murphy, T. F. (2008). Infection in the pathogenesis and course of chronic obstructive pulmonary disease. *N Engl J Med* **359**, 2355-2365.

Severi, E., Randle, G., Kivlin, P., Whitfield, K., Young, R., Moxon, R., Kelly, D., Hood, D. & Thomas, G. H. (2005). Sialic acid transport in *Haemophilus influenzae* is essential for lipopolysaccharide sialylation and serum resistance and is dependent on a novel tripartite ATP-independent periplasmic transporter. *Mol Microbiol* **58**, 1173-1185.

Shakhnovich, E. A., King, S. J. & Weiser, J. N. (2002). Neuraminidase expressed by *Streptococcus pneumoniae* desialylates the lipopolysaccharide of *Neisseria meningitidis* and *Haemophilus influenzae*: a paradigm for interbacterial competition among pathogens of the human respiratory tract. *Infect Immun* **70**, 7161-7164.

Shepherd, J. G., Wang, L. & Reeves, P. R. (2000). Comparison of O-antigen gene clusters of *Escherichia coli* (*Shigella*) sonnei and *Plesiomonas shigelloides* O17: sonnei gained its current plasmid-borne O-antigen genes from *P. shigelloides* in a recent event. *Infect Immun* **68**, 6056-6061.

Shibuya, N., Amano, K., Azuma, J., Nishihara, T., Kitamura, Y., Noguchi, T. & Koga, T. (1991). 6-Deoxy-D-talan and 6-deoxy-L-talan. Novel serotype-specific polysaccharide antigens from *Actinobacillus actinomycetemcomitans*. *J Biol Chem* **266**, 16318-16323.

Smith, H. O., Tomb, J. F., Dougherty, B. A., Fleischmann, R. D. & Venter, J. C. (1995). Frequency and distribution of DNA uptake signal sequences in the *Haemophilus influenzae* Rd genome. *Science* **269**, 538-540.

Sperisen, P., Schmid, C. D., Bucher, P. & Zilian, O. (2005). Stealth proteins: in silico identification of a novel protein family rendering bacterial pathogens invisible to host immune defense. *PLoS Comput Biol* **1**, e63.

St Geme, J. W., 3rd, Kumar, V. V., Cutter, D. & Barenkamp, S. J. (1998). Prevalence and distribution of the *hmw* and *hia* genes and the HMW and Hia adhesins among genetically diverse strains of nontypeable *Haemophilus influenzae*. *Infect Immun* **66**, 364-368.

Steenbergen, S. M., Lee, Y. C., Vann, W. F., Vionnet, J., Wright, L. F. & Vimr, E. R. (2006). Separate pathways for O acetylation of polymeric and monomeric sialic acids and identification of sialyl O-acetyl esterase in *Escherichia coli* K1. *J Bacteriol* **188**, 6195-6206.

Steinbacher, S., Miller, S., Baxa, U., Weintraub, A. & Seckler, R. (1997). Interaction of *Salmonella* phage P22 with its O-antigen receptor studied by X-ray crystallography. *Biol Chem* **378**, 337-343.

Straetemans, M., van Heerbeek, N., Tonnaer, E., Ingels, K. J., Rijkers, G. T. & Zielhuis, G. A. (2001). A comprehensive model for the aetiology of otitis media with effusion. *Med Hypotheses* **57**, 784-791.

Streker, R. G., Venezia, R. A. & Robertson, R. G. (1978). Mode of action of the *Haemophilus* bactericidal factor. *Antimicrob Agents Chemother* **13**, 527-532.

Sundgren, A., Lahmann, M. & Oscarson, S. (2010). Synthesis of 6-PEtN-alpha-D-Gal*p*NAc-(1->6)-beta-D-Gal*p*-(1->4)-beta-D-Glc*p*NAc-(1->3)-beta-D-Gal*p*-(1->4)-beta-D-Glc*p*, a *Haemophilus influenzae* lipopolysacharide structure, and biotin and protein conjugates thereof. *Beilstein J Org Chem* **6**, 704-708.

Swords, W. E., Buscher, B. A., Ver Steeg Ii, K., Preston, A., Nichols, W. A., Weiser, J. N., Gibson, B. W. & Apicella, M. A. (2000). Non-typeable *Haemophilus influenzae* adhere to and invade human bronchial epithelial cells via an interaction of lipooligosaccharide with the PAF receptor. *Mol Microbiol* **37**, 13-27.

Takada, K., Saito, M., Tsuzukibashi, O., Kawashima, Y., Ishida, S. & Hirasawa, M. (2010). Characterization of a new serotype g isolate of *Aggregatibacter actinomycetemcomitans*. *Mol Oral Microbiol* **25**, 200-206.

Takahashi, H., Carlson, R. W., Muszynski, A., Choudhury, B., Kim, K. S., Stephens, D. S. & Watanabe, H. (2008). Modification of lipooligosaccharide (LOS) with phosphoethanolamine by LptA in *Neisseria meningitidis* enhances meningococcal adhesion to human endothelial and epithelial cells. *Infect Immun*.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596-1599.

Tang, G. & Mintz, K. P. (2010). Glycosylation of the collagen adhesin EmaA of *Aggregatibacter actinomycetemcomitans* is dependent upon the lipopolysaccharide biosynthetic pathway. *J Bacteriol* **192**, 1395-1404.

Tao, J., Feng, L., Guo, H., Li, Y. & Wang, L. (2004). The O-antigen gene cluster of *Shigella boydii* O11 and functional identification of its *wzy* gene. *FEMS Microbiol Lett* **234**, 125-132.

Taylor, D. C., Cripps, A. W., Clancy, R. L., Murree-Allen, K., Hensley, M. J., Saunders, N. A. & Sutherland, D. C. (1992). Biotypes of *Haemophilus parainfluenzae* from the respiratory secretions in chronic bronchitis. *J Med Microbiol* **36**, 279-282.

Tong, H. H., Blue, L. E., James, M. A., Chen, Y. P. & DeMaria, T. F. (2000). Evaluation of phase variation of nontypeable *Haemophilus influenzae* lipooligosaccharide during nasopharyngeal colonization and development of otitis media in the chinchilla model. *Infect Immun* **68**, 4593-4597.

Tuyau, J. E. & Sims, W. (1983). Aspects of the pathogenicity of some oral and other haemophili. *J Med Microbiol* **16**, 467-475.

Tzeng, Y. L., Ambrose, K. D., Zughaier, S., Zhou, X., Miller, Y. K., Shafer, W. M. & Stephens, D. S. (2005). Cationic antimicrobial peptide resistance in *Neisseria meningitidis*. *J Bacteriol* **187**, 5387-5396.

Tzianabos, A. O., Kasper, D. L. & Onderdonk, A. B. (1995). Structure and function of *Bacteroides fragilis* capsular polysaccharides: relationship to induction and prevention of abscesses. *Clin Infect Dis* **20 Suppl 2**, S132-140.

Valvano, M. A., Messner, P. & Kosma, P. (2002). Novel pathways for biosynthesis of nucleotide-activated glycero-manno-heptose precursors of bacterial glycoproteins and cell surface polysaccharides. *Microbiology* **148**, 1979-1989.

van Ham, S. M., van Alphen, L., Mooi, F. R. & van Putten, J. P. (1993). Phase variation of *H. influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. *Cell* **73**, 1187-1196.

Van Hoogmoed, C. G., Geertsema-Doornbusch, G. I., Teughels, W., Quirynen, M., Busscher, H. J. & Van der Mei, H. C. (2008). Reduction of periodontal pathogens adhesion by antagonistic strains. *Oral Microbiol Immunol* **23**, 43-48.

van Kempen, M. J., Rijkers, G. T. & Van Cauwenberge, P. B. (2000). The immune response in adenoids and tonsils. *Int Arch Allergy Immunol* **122**, 8-19.

Venezia, R. A. & Robertson, R. G. (1975). Bactericidal substance produced by *Haemophilus influenzae* b. *Can J Microbiol* **21**, 1587-1594.

Vimr, E. R., Kalivoda, K. A., Deszo, E. L. & Steenbergen, S. M. (2004). Diversity of microbial sialic acid metabolism. *Microbiol Mol Biol Rev* **68**, 132-153.

Vitovski, S., Dunkin, K. T., Howard, A. J. & Sayers, J. R. (2002). Nontypeable *Haemophilus influenzae* in carriage and disease: a difference in IgA1 protease activity levels. *JAMA* **287**, 1699-1705.

Wack, A. & Gallorini, S. (2008). Bacterial polysaccharides with zwitterionic charge motifs: Toll-like receptor 2 agonists, T cell antigens, or both? *Immunopharmacol Immunotoxicol* **30**, 761-770.

Wang, L., Liu, D. & Reeves, P. R. (1996). C-terminal half of *Salmonella enterica* WbaP (RfbP) is the galactosyl-1-phosphate transferase domain catalyzing the first step of O-antigen synthesis. *J Bacteriol* **178**, 2598-2604.

Wang, W., Perepelov, A. V., Feng, L. & other authors (2007). A group of *Escherichia coli* and *Salmonella enterica* O antigens sharing a common backbone structure. *Microbiology* **153**, 2159-2167.

Wang, X. & Quinn, P. J. (2010). Lipopolysaccharide: Biosynthetic pathway and structure modification. *Prog Lipid Res* **49**, 97-107.

Watt, J. P., Wolfson, L. J., O'Brien, K. L. & other authors (2009). Burden of disease caused by *Haemophilus influenzae* type b in children younger than 5 years: global estimates. *Lancet* **374**, 903-911.

Weiser, J. N., Lindberg, A. A., Manning, E. J., Hansen, E. J. & Moxon, E. R. (1989a). Identification of a chromosomal locus for expression of lipopolysaccharide epitopes in *Haemophilus influenzae*. *Infect Immun* **57**, 3045-3052.

Weiser, J. N., Love, J. M. & Moxon, E. R. (1989b). The molecular mechanism of phase variation of *H. influenzae* lipopolysaccharide. *Cell* **59**, 657-665.

Weiser, J. N., Shchepetov, M. & Chong, S. T. (1997). Decoration of lipopolysaccharide with phosphorylcholine: a phase-variable characteristic of *Haemophilus influenzae*. *Infect Immun* **65**, 943-950.

Weiser, J. N., Goldberg, J. B., Pan, N., Wilson, L. & Virji, M. (1998a). The phosphorylcholine epitope undergoes phase variation on a 43-kilodalton protein in *Pseudomonas aeruginosa* and on pili of *Neisseria meningitidis* and *Neisseria gonorrhoeae*. *Infect Immun* **66**, 4263-4267.

Weiser, J. N., Pan, N., McGowan, K. L., Musher, D., Martin, A. & Richards, J. (1998b). Phosphorylcholine on the lipopolysaccharide of *Haemophilus influenzae* contributes to persistence in the respiratory tract and sensitivity to serum killing mediated by C-reactive protein. *J Exp Med* **187**, 631-640.

White, K. A., Kaltashov, I. A., Cotter, R. J. & Raetz, C. R. (1997). A mono-functional 3-deoxy-D-manno-octulosonic acid (Kdo) transferase and a Kdo kinase in extracts of *Haemophilus influenzae*. *J Biol Chem* **272**, 16555-16563.

White, K. A., Lin, S., Cotter, R. J. & Raetz, C. R. (1999). A *Haemophilus influenzae* gene that encodes a membrane bound 3-deoxy-D-manno-octulosonic acid (Kdo) kinase. Possible involvement of Kdo phosphorylation in bacterial virulence. *J Biol Chem* **274**, 31391-31400.

Whitfield, C. (2006). Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu Rev Biochem* **75**, 39-68.

Windsor, H. M., Gromkova, R. C. & Koornhof, H. J. (1991). Plasmid-mediated NAD independence in *Haemophilus parainfluenzae*. *J Gen Microbiol* **137**, 2415-2421.

Winter, L. E. & Barenkamp, S. J. (2010). Construction and immunogenicity of recombinant adenovirus vaccines expressing the HMW1, HMW2, or Hia adhesion protein of nontypeable *Haemophilus influenzae*. *Clin Vaccine Immunol* **17**, 1567-1575.

Wright, J. C., Hood, D. W., Randle, G. A., Makepeace, K., Cox, A. D., Li, J., Chalmers, R., Richards, J. C. & Moxon, E. R. (2004). *lpt6*, a gene required for addition of phosphoethanolamine to inner-core lipopolysaccharide of *Neisseria meningitidis* and *Haemophilus influenzae*. *J Bacteriol* **186**, 6970-6982.

Xu, D. Q., Cisar, J. O., Ambulos Jr, N., Jr., Burr, D. H. & Kopecko, D. J. (2002). Molecular cloning and characterization of genes for *Shigella sonnei* form I O polysaccharide: proposed biosynthetic pathway and stable expression in a live *Salmonella* vaccine vector. *Infect Immun* **70**, 4414-4423.

Xu, Z., Chen, X., Li, L., Li, T., Wang, S., Chen, H. & Zhou, R. (2010). Comparative genomic characterization of *Actinobacillus pleuropneumoniae*. *J Bacteriol*.

Yang, J., Ritchey, M., Yoshida, Y., Bush, C. A. & Cisar, J. O. (2009). Comparative structural and molecular characterization of ribitol-5-phosphate-containing *Streptococcus oralis* coaggregation receptor polysaccharides. *J Bacteriol* **191**, 1891-1900.

Yavuz, E., Maffioli, C., Ilg, K., Aebi, M. & Priem, B. (2011). Glycomimicry: display of fucosylation on the lipo-oligosaccharide of recombinant *Escherichia coli* K12. *Glycoconj J*.

Zaleski, P., Wojciechowski, M. & Piekarowicz, A. (2005). The role of Dam methylation in phase variation of *Haemophilus influenzae* genes involved in defence against phage infection. *Microbiology* **151**, 3361-3369.

Zamze, S. E., Ferguson, M. A., Moxon, E. R., Dwek, R. A. & Rademacher, T. W. (1987). Identification of phosphorylated 3-deoxy-manno-octulosonic acid as a component of *Haemophilus influenzae* lipopolysaccharide. *Biochem J* **245**, 583-587.

Zolli, M., Kobric, D. J. & Brown, E. D. (2001). Reduction precedes cytidylyl transfer without substrate channeling in distinct active sites of the bifunctional CDP-ribitol synthase from *Haemophilus influenzae*. *Biochemistry* **40**, 5041-5048.

Zughaier, S. M., Tzeng, Y. L., Zimmer, S. M., Datta, A., Carlson, R. W. & Stephens, D. S. (2004). *Neisseria meningitidis* lipooligosaccharide structure-dependent activation of the macrophage CD14/Toll-like receptor 4 pathway. *Infect Immun* **72**, 371-380.