The study of candidate sialometabolism genes and sialometabolism gene regulation in *Haemophilus influenzae*

CHEN HSUAN SHERRY TSAI

A thesis submitted for the degree of

DOCTOR OF PHILOSOPHY

in the Department of Paediatrics, Division of Medical Sciences

University of Oxford

Hilary, 2013

Acknowledgements

I would like to thank Prof Richard Moxon and Dr Derek Hood for allowing me to work in their lab.

To every lab colleague that has worked with me, thank you all for the great atmosphere and vibrance in the lab. It is something that I really miss. To Dr Rosie Young, thank you for sharing your knowledge and experience with me. I will remember all the fun moments we have had by our benches and all the great musicals we went to watch together. To Mrs Mary Deadman, thank you for your teaching, your guidance not only in science, but also in life, throughout my DPhil years. To Dr Derek Hood, in addition to acknowledge your extreme patience with me, I am also truly grateful for your critical appraisals and excellent advice that you have generously provided me at every stage of my research.

To past lab members who have worked so hard to lay the foundation of my current work, thank you all for your prior contributions.

To Prof Derrick Crook and Prof Enzo Cerundolo's lab and lab members, thank you for generously sharing your equipment with me. You are all truly great neighbours and I have enjoyed working alongside you. Thank you for your friendship and all the fun moments. A special thank to Dr Esther Robinson in Prof Crook's lab for your assistance and guidance in my qPCR work.

To all my friends in the UK, thank you all for making my life in Oxford such a wonderful and memorable experience.

To my parents Harrison and Vicky Tsai, my sister Leona Tsai, my husband Gary Huang, and my daughter Pierra Huang, thank you all for your unconditional love, constant encouragement, and your belief in me. To Gary, thank you for always being there when I am happy and down. It is a great privilege to be able to share every moment of my life with you. And to our daughter Pierra, thank you too for always being there (either inside or outside of me) during the last and toughest part of my DPhil years. Although looking after you and simultaneously writing a thesis has been really challenging, but I want you to know that you are the motivation that helped me carry on to finish this work. I love you.

Abstract

The study of candidate sialometabolism genes and sialometabolism gene regulation in *Haemophilus influenzae*

Chen Hsuan Sherry Tsai, Department of Paediatrics, Division of Medical Sciences, The Queen's College, DPhil thesis (Hilary, 2013)

Sialic acid (SA) is a known major virulence factor of *Haemophilus influenza* (*Hi*). This study aims to analyse the functions of some candidate sialometabolism genes, and to further our current understanding on the *Hi* sialometabolism gene regulation.

Two candidate sialometabolism genes (*HI0227* and *HI0148*) and their adjacent ORFs (*HI0228*, *HI0148.1* and *HI0149)* were studied. *HI0148.1* and *HI0149* are transcribed as a single gene in screened NT*Hi* strains, and we refer to the combined ORF as *NTHI0236* (the designation in strain 86-028NP). Across *Hi* strains screened, the sequences of *HI0227, HI0148* and *NTHI0236* are conserved. However, the sequence of *HI0228* is heterogeneous. Mutants that lack the functions of HI0227, HI0228, HI0148 and NTHI0236 were compared to their respective wild type parent strains for ability to grow on SA (in aerobic and microaerophilic conditions), their ability to sialylate LPS and their ability to resist complement mediated killing. The mutants did not exhibit major differences in the tested aspects of sialometabolism compared to their respective wild type strains. Changes observed in some of the mutants in serum bactericidal assays and LPS profiles were due to the effect of phase variable genes. The sialometabolism functions of *HI0227*, *HI0148*, and *NTHI0236* remain obscure, and we postulate that *HI0228* is a pseudogene.

Investigation of *Hi* sialometabolism gene regulation was conducted using mutants that lack different steps of the Neu5Ac catabolism pathway and the Neu5Ac activation pathway. The expression of *nanE* and *siaP*, respectively representing the Neu5Ac catabolism and transport operons, were assessed using RT-PCR and qPCR. We investigated a temporal/concentration effect of Neu5Ac on the expression of sialometabolism operons, which highlights the importance of studying the *Hi* sialometabolism gene regulation as a dynamic process. We further demonstrated that GlcN-6P, a Neu5Ac intermediate from the catabolism pathway, is likely the SIS sugar that interacts with SiaR, the repressor protein of the *Hi* sialometabolism operons. We postulate that upon binding of GlcN-6P to SiaR, the SiaR-mediated repression on the *Hi* sialometabolism operons is relieved, resulting in the induction of the expression of Neu5Ac catabolism and transport genes.

Abbreviations

amp	ampicillin
BHI	Brain Heart Infusion
BLAST	Basic local alignment search tool
BLASTN	nucleotide BLAST
BLASTP	Protein-protein BLAST
bp	base pair
CMAH	cytidine monophosphate-N-acetylneuramic acid hydroxylase
CMP-Neu5Ac	cytidine monophosphate-N-acetyl-neuramic acid
CMP-Neu5Gc	cytidine monophosphate-N-glycolyl-neuramic acid
CRP	cAMP receptor protein
Ct	cycle threshold
E. coli	*Escherichia coli*
EMSA	electrophoretic shift mobility assay
ery	erythromycin
eryC	erythromycin resistance cassette
ESR	Extracytoplasmic solute receptor
fH	factor H
Fru	Fructose
Fru-6P	Fructose 6-phosphate
FS	Fisher Syndrome
Gal	Galactose
GalNAc	N-acetyl-galactosamine
GBS	Guillain Barré Syndrome
Glc	Glucose
GlcN-6P	Glucosamine 6-phosphate
GlcNAc	N-acetyl glucosamine
GlcNAc-6P	N-acetyl glucosamine 6-phosphate
GT	Glycosyl transferase
Hap	*Haemophilus* adhesion and penetration protein
Hep	heptose
Hex	hexose
Hi	*Haemophilus influenzae*
*Hi*b	*H. influenzae* type b
Hia	*Hi* autotransporter adhesin
HMG	high molecular weight glycoform
HMW	high molecular weight protein

Hsf	*Haemophilus* surface fibrils
HTH	helix-turn-helix DNA binding domain
ICMT	Isoprenylcysteine carboxyl methyltransferase
Ig	Immunoglobulin
IPC	inter-plate calibration for qPCR
IPTG	isopropylthio-β-D-galactoside
kan	kanamycin
kanC	kanamycin resistance cassette
Kdo	2-keto-3-deoxyoctulosonic acid
L. interrogans	*Leptospira interrogans*
LOS	lipo-oligosaccharide
LPS	lipopolysaccharide
ManNAc	N-acetyl mannosamine
ManNAc-6P	N-acetyl mannosamine 6-phosphate
min	minutes
MLEE	multilocus enzyme electrophoresis
MLST	multilocus sequence typing
N. gonorrhoeae	*Neisseria gonorrhoeae*
N. meningitidis	*Neisseria meningitidis*
NagA	N-acetyl glucosamine 6-phosphate deacetylase
NagB	glucosamine 6-phosphate deaminase
NanA	N-acetyl neuraminate aldolase
NanE	N-acetyl mannosamine-6-phosphate epimerase
NanK	N-acetyl mannosamine kinase
Neu5Ac	N-acetyl-neuraminic acid
Neu5,4Ac2	4-O-acetyl *N*-acetylneuraminic acid
Neu5,9Ac2	9-O-acetyl *N*-acetylneuraminic acid
Neu5Gc	N-glycolyl-neuraminic acid
NT*Hi*	non-typable *Haemophilus influenzae*
OD	optical density
OMP	outer membrane protein
ORF	open reading frame
P. multocida	*Pasteurella multocida*
PCho	phosphocholine
PCR	polymerase chain reaction
PEtn	phosphoethanolamine
PEP	phosphoenolpyruvate
P..PEtn	pyrophosphoethanolamine

qPCR	quantitative real time polymerase chain reaction
RT-PCR	reverse transcription polymerase chain reaction
S. aureus	*Staphylococcus aureus*
SA	sialic acid
SBA	serum bactericidal assay
SIS	sugar isomerase domain
T. forsythia	*Tannerella forsythia*
TE	Tris EDTA
TEMED	tetramethylethylenediamine
TRAP	Tripartite ATP-independent periplasmic
V. cholerae	*Vibrio cholerae*
V. vulnificus	*Vibrio vulnificus*
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

Table of Contents

1 Introduction: 1

1.1 Characterisation of *Haemophilus influenzae* 1

1.1.1 The genus *Haemophilus* 1

1.1.2 The organism *Hi* 2

1.1.3 Pathogenesis of *Hi* 4

1.2 Major virulence determinants of *Hi* 8

1.2.1 Polysaccharide capsule 8

1.2.2 Surface proteins 10

1.2.3 IgA1 proteases 12

1.2.4 Lipopolysaccharide 13

1.3 General overview of SA 21

1.4 Sialometabolism of *Hi* 25

1.4.1 Acquisition of environmental Neu5Ac 25

1.4.2 Sialylation of the *Hi* LPS 29

1.4.3 Catabolism of Neu5Ac in *Hi* 34

1.4.4 Regulation of the Neu5Ac catabolism and transport operons 39

1.4.5 SA and biofilm 44

1.5 Candidate sialometabolism genes in *Hi* 47

2 Materials and Methods 48

2.1 Materials 48

2.1.1 Bacterial strains and plasmids used in this study 48

2.1.2 Enzymes 51

2.1.3 Oligonucleotides 51

2.2 Methods 54

2.2.1 Growth media preparation and culture conditions 54

2.2.2 Preparation of the *Hi* chromosomal DNA 55

2.2.3 Preparation of plasmid DNA 56

2.2.4 Agarose gel electrophoresis and extraction 57

2.2.5 Polymerase chain reaction (PCR) 57

2.2.6 *Hi* transformation 58
2.2.7 *E. coli* transformation 59
2.2.8 Tricine LPS gel 60
2.2.9 Serum bactericidal assay 62
2.2.10 RNA extraction, reverse transcriptase PCR, and quantitative PCR 63
2.2.11 DNA and plasmid sequencing 68
2.2.12 Growth assay of *Hi* in BHI and CDM 68
2.2.13 Fermentation Assay 69
2.2.14 Mutation of loci and transformation of *Hi* 70
3 Evaluation of the repertoire of known genes, and identification and analyses of candidate novel sialometabolism genes 84
3.1 Key sialometabolism genes are conserved across all *Hi* strains screened 84
3.2 Selection of candidate sialometabolism genes in *Hi* 86
3.2.1 Identification of sialometabolism associated genes in *E. coli* and screening their respective homologues in *Hi* 86
3.2.2 Study of adjacent ORFs of selected *Hi* candidate sialometabolism genes 94
3.3 Analyses of the distribution and sequences of selected candidate sialometabolism genes 94
3.3.1 *HI0227* is conserved in length across screened strains 94
3.3.2 *HI0228* and the *HI0227-HI0228* intergenic region are variable in length across strains screened 96
3.3.3 *HI0148* and *NTHI0236* are conserved across screened strains 101
3.4 Database search of the candidate sialometabolism genes 102
3.4.1 HI0227 belongs to the YhcH/YiaL/YjgK family 102
3.4.2 HI0228 contains domains of glycosyl transferase family 8 proteins 102
3.4.3 HI0148 belongs to the YjhT family 104
3.4.4 NTHI0236 belongs to the isoprenylcysteine carboxyl methyltransferase family 104
3.5 Discussion 105
4 Phenotypic characterisation of the candidate and peripheral sialometabolism genes 107
4.1 Mutants of *HI0227, HI0228* and *HI0227/HI0228* were created in group 2, 3 and 4 representative strains 107

4.2 The alternative Neu5Ac catabolism pathway is not present in *Hi* 108

4.3 Assessment of whether HI0227, HI0228, HI0148 and NTHI0236 are involved in the catabolism of Neu5Ac 114

4.4 Fermentation assessment of the *HI0227*-, *HI0228*-, *HI0227/HI0228*-, *HI0148*- and *NTHI0236*- mutants in strain Rd 118

4.5 LPS profiles of the *HI0227*-, *HI0228*-, *HI0227/HI0228*- mutants in strains 375, Rd, R2846 and 723 120

4.6 SBA of the *HI0227*-, *HI0228*-, *HI0227/HI0228*-, *HI0148*- and *NTHI0236*- mutant strains.. 131

4.7 Discussion 136

5 Investigation of a role of candidate novel sialometabolism genes in the *Hi* sialometabolism gene expression 140

5.1 Candidate novel sialometabolism genes do not affect the expression of the *Hi* sialometabolism operons and Neu5Ac has a temporal/concentration effect on the expression of the *Hi* sialometabolism operons 140

5.2 Discussion 151

6 Further evaluation on the regulation of the *Hi* sialometabolism operons 153

6.1 The temporal/concentration effect of Neu5Ac on the expression of *nanE* and *siaP* is strain independent 154

6.2 The CMP-Neu5Ac biosynthesis pathway does not affect the expression of *nanE* and *siaP* 157

6.3 Upregulation of *nanE* and *siaP* following growth in Neu5Ac for 10 min is likely due to reduced SiaR activity by binding to a Neu5Ac metabolite 164

6.4 Exploration of the combinatorial effects exerted through CRP and SiaR and confirmation that the expression of *nanE* and *siaP* is not strain dependent 169

6.5 Downregulation of *nanE* and *siaP* following growth in Neu5Ac for 60 min is due to the depletion of the SIS sugar 173

6.6 The up- and downregulation of *nanE* and *siaP* following growth in Neu5Ac for 10 or 60 min respectively do not affect LPS sialylation 176

6.7 Discussion 178

7 Final discussion 186

8 References 197

1 Introduction:

1.1 Characterisation of *Haemophilus influenzae*

1.1.1 **The genus *Haemophilus***

The *Haemophilus* genus belongs to the Pasteurellaceae family, and bacteria that belong to this genus are small, pleomorphic, non-motile, Gram-negative coccobacilli. In most cases, bacteria in the *Haemophilus* genus are commensals to human or animal hosts, and examples of species that form part of the normal human flora are: *H. influenzae (Hi*), *H. parainfluenzae*, *H. parahaemolyticus*, *H. paraphrophilus*, *H. segnis*, and *H. paraphrohaemolyticus*. Although in general *Hi* behaves as an obligate human commensal, however, it can cause significant diseases which will be discussed later (Fink D. & Geme J., 2006) (see section 1.1.3).

Species that belong to the *Haemophilus* genus are aerobes or facultative anaerobes, and *Hi* is an example of a bacterium that can survive in both aerobic or microaerophilic conditions. Growth of *Hi* in aerobic condition requires two accessory growth factors: X factor (hemin) and V factor (nicotinamide adenine dinucleotide, NAD), and this is due to the fact that *Hi* is unable to synthesise these growth factors, hence *Hi* is dependent on the host for the acquisition of these growth factors (Granick & Gilder, 1946, Evans *et al.*, 1974). Growth of *Hi* under microaerophilic conditions does not require X factor (White, 1963, Raghunathan *et al.*, 2004).

1.1.2 **The organism *Hi***

Hi was first isolated by Pfeiffer in 1892 during an influenza pandemic and was subsequently named Pfeiffer's bacillus. The bacterium was originally mistaken as the causative organism for the flu epidemic. In 1933 this error was rectified by Smith after his successful isolation of the influenza virus (Barry, 2004, Olitsky P.K., 1921, Smith, 1933).

To highlight a few contributions that *Hi* has provided in the field of science: 1) *Hi* was the organism from which the first restriction enzyme (endonuclease) was isolated, which made a huge contribution to the advent of molecular biology research (Smith & Wilcox, 1970, Kelly & Smith, 1970). 2) *Hi* strain Rd was the first free-living organism to have its whole genome sequenced (Fleischmann *et al.*, 1995). The 1830 kb genome of *Hi* strain Rd was predicted to comprise of a total of 1749 open reading frames (ORFs). 3) *Hi* was the second organism discovered which could undergo natural transformation, a mechanism by which a bacterium exchanges its genetic material with organisms that reside within the same niche (Voll & Goodgal, 1961). 4) Clinically, *Hi* is the first bacterium to demonstrate resistance against the antibiotic streptomycin through genetic means (Alexander & Leidy, 1953), and was documented as one of the first bacteria for which sulphonamide was prescribed in treating invasive infection (Carithers, 1974). 5) The first glycoconjugate vaccine was made from the capsule of *Hi* type b (*Hi*b) strains

(Anderson *et al.*, 1981), and the introduction of this vaccine successfully prevented invasive diseases caused by these strains (Peltola, 2000) (see section 1.1.3).

Hi is an obligate human commensal, normally residing in the nasopharynx. *Hi* adheres to respiratory epithelium that lines the respiratory tract (Lampe *et al.*, 1982) and interacts with this surrounding mucous layer (Lethem, 1997) which is produced by goblet cells and submucosal glands. Furthermore, *Hi* can also invade and survive in epithelial cells (Johnson *et al.*, 1996, St Geme & Falkow, 1990) and in mononuclear, macrophage-like cells that are present in human adenoid tissue (Forsgren *et al.*, 1994). This intracellular survival ability of *Hi* can potentially prolong colonisation of the host.

The prevalence of *Hi* is estimated to be approximately 20% amongst children in their first year of life. This number rises to over 50% in children at five years of age, and remains high throughout adulthood (Howard *et al.*, 1988). Although in most cases, *Hi* resides as a human commensal, it can nonetheless cause severe and invasive diseases that lead to high morbidity and mortality, adding a high cost and burden to our health system.

Hi is broadly classified into two groups: encapsulated strains and unencapsulated/ nontypeable *Hi* (NT*Hi*) strains. Pittman and colleagues described six antigenetic distinct polysaccharide

capsules (a to f), which forms the basis of capsular strain typing system (Pittman, 1931). The majority of *Hi* invasive disease is caused by *Hi*b strains. Currently, there are several classification systems for NT*Hi*, and they are based on biotyping (Kilian, 1991), outer membrane proteins (OMPs) (Murphy *et al.*, 1983, Murphy & Apicella, 1985), lipopolysaccharide (LPS) antigenic heterogeneity (Campagnari *et al.*, 1987), multilocus enzyme electrophoresis (MLEE) (Musser *et al.*, 1986, Porras *et al.*, 1986b), multilocus sequence typing (MLST) (Platonov *et al.*, 2000), restriction fragment length polymorphism (Groeneveld *et al.*, 1988, Loos *et al.*, 1989, Mencarelli *et al.*, 1993), and PCR amplification of repeat sequences (van Belkum *et al.*, 1994). Using MLEE and MLST, it has been shown that typeable and NT*Hi* strains are from separate lineages (Porras *et al.*, 1986a, Porras et al., 1986b) and that NT*Hi* is genetically more diverse than typeable strains.

1.1.3 **Pathogenesis of *Hi***

Colonization is a prerequisite for the survival of *Hi*, and diseases are infrequent events that follow successful colonization. The colonization process requires the bacterium to overcome host clearance mechanisms, such as the mucociliary escalator and local immunity. Mucociliary escalator traps respiratory foreign particles by mucous and sweeping them up to the epiglottis by the action of cilia. Adherence to host epithelial cells, and invasion into or penetration between host cells, are means by which *Hi* evades mucociliary clearance. Various bacterial factors

facilitate this adherence to and invasion into epithelial cells, and they include OMPs, pili and LPS (Reddy *et al.*, 1996, Hardy *et al.*, 2003, Swords *et al.*, 2000).

The carriage of *Hi* in the nasopharynx is a dynamic process, i.e. *Hi* can be transient or reside for a period of time. A strain of *Hi* is often replaced by another strain after a period of time (Garcia-Rodriguez & Fresnadillo Martinez, 2002). Efficient transmission of a strain between hosts requires direct contact with airborne droplets or contaminated secretions from the host respiratory system (Murphy *et al.*, 1989).

Invasive disease caused by *Hi,* such as septicaemia, meningitis, and epiglottitis, although being a minority of *Hi* disease cases, is often severe and life-threatening. Invasive disease is almost exclusively caused by *Hi*b strains (Riley, 1977). The ribosyl-ribitol phosphate capsule of *Hi*b strains is the major virulence determinant of *Hi*b associated invasive disease (Roberts *et al.*, 1981, Turk, 1967). The mechanism through which invasive disease is caused by *Hi* first requires the invasion of *Hi* through nasopharyngeal cells into the bloodstream, and is followed by systemic dissemination and deposition of the bacterium onto internal body parts (Rubin & Moxon, 1983). The exact mechanism by which *Hi* invades through the host nasopharyngeal cells remains to be determined. Once in the bloodstream, the ability of *Hi* to survive and replicate rapidly is a key factor that determines its ability to cause disseminated disease. It has been shown that only a

single bacterium is required to cause systemic infection (Moxon & Murphy, 1978). This single *Hi* bacterium, after replicating for 48 hours in the bloodstream, can produce up to 10^5 bacteria per ml.

In 1982, a glycoconjugate vaccine became available for the prevention of invasive disease caused by *Hi*b strains. The wide use of this vaccine not only successfully reduced the incidence of *Hi*b associated disease (by 78% to 100% for all invasive *Hi*b associated disease in 47 countries in the WHO European Region) (WHO, 2011), but also significantly decreased the carriage rate of *Hi*b in children (Barbour *et al.*, 1995). Studies have shown that after the introduction of the *Hi*b vaccine, the incidence of disease caused by other capsular strains and by NT*Hi* has increased slightly (LaClaire *et al.*, 2003, CDC, 2002, MacNeil *et al.*, 2011), however, these figures are still very low compared to the incidence of *Hi*b infection during the pre-vaccine era: in the USA, the disease incidence for non-type b and for NT*Hi* strains after the introduction of *Hi*b vaccine are 0.38 and 0.99 per 100,000 people respectively, whereas the *Hi*b incidence during the pre-vaccine era was approximately 37 per 100,000 people (Adams *et al.*, 1993, MacNeil et al., 2011).

NT*Hi* strains lack a polysaccharide capsule and most strains generally do not cause invasive disease. However, systemic infection associated with NT*Hi* has been documented but is usually

limited to neonates and to immune compromised patients (Rao *et al.*, 1999, O'Neill *et al.*, 2003). NT*Hi* can cause a variety of non-invasive diseases, which include upper and lower respiratory tract infections, sinusitis and otitis media (Jokipii & Jokipii, 1977, Everett *et al.*, 1977, Turk, 1984). NT*Hi* has also been well documented to frequently exacerbate underlying lung diseases in patients suffering from chronic obstructive pulmonary disease, bronchiectasis and cystic fibrosis (Sethi & Murphy, 2008, Murphy & Sethi, 1992). NT*Hi* causes disease following contiguous spread from its normal habitat, the nasopharynx, to areas such as the sinus, upper and lower respiratory tract, and middle ear via the Eustachian tube (Vergison, 2008). In developing countries, respiratory tract infection caused by NT*Hi* is a major health concern and can lead to high morbidity and mortality (Munson & Hunt, 1989).

Otitis media is inflammation of the middle ear and is a common paediatric illness. It accounts for approximately one third of all paediatric GP visits (Massa *et al.*, 2009, Teele *et al.*, 1983). Studies in the USA have shown that by the age of two, 90% of children would have experienced at least one episode of middle ear effusion (Paradise *et al.*, 1997). NT*Hi* is the most common bacterium to cause otitis media and accounts for approximately 20-30% of all acute otitis media documented cases (Leibovitz *et al.*, 2010). The cost of treating otitis media in the USA is estimated to be several billion dollars a year (Berman *et al.*, 1994). Complications of otitis media

can be serious, and they include impaired hearing, hearing loss, delayed speech development and intellectual retardation (Teele *et al.*, 1990).

Treating NT*Hi* associated infection requires prescription of antibiotics. Antibiotic prescription selects for antibiotic resistant NT*Hi* strains, thus allowing dangerous, antibiotic-resistant strains to prosper (Granoff *et al.*, 1978). Development of a successful vaccine that targets NT*Hi* strains is therefore needed to eradicate NT*Hi* infection. To date, no specific vaccine is available to target NT*Hi*. In recent years, a 10-valent pneumococcal vaccine containing protein D of NT*Hi*, which functions as the polysaccharide carrier protein, has been developed (Prymula *et al.*, 2006). Protein D of NT*Hi* is an outer membrane lipoprotein that facilitates bacterial adherence onto airway respiratory cells and phosphocholine (PCho) acquisition (Johnson *et al.*, 2010). The inclusion of NT*Hi* protein D in the pneumococcal vaccine has been reported to decrease the incidence of NT*Hi* induced otitis media (Prymula et al., 2006). However, a more specific NT*Hi* vaccine is required should one wish to successfully eradicate all NT*Hi* induced-infection.

1.2 Major virulence determinants of *Hi*

1.2.1 **Polysaccharide capsule**

A polysaccharide capsule can form a protective layer around the bacterium, and *Hi* can produce one of six types of capsules: namely a to f (Pittman, 1931). The *Hi* capsule is made of

intercalating, repeating disaccharide units that are linked to each other via glycosidic and/or phosphodiester bonds. Strains expressing the *Hi*b capsule are most virulent (Roberts *et al.*, 1981, Moxon & Vaughn, 1981). The *Hi*b capsule is a polymer made of ribose and ribitol-5-phosphate (Crisel *et al.*, 1975). Increasing the copy number of the *Hi*b capsule synthesis locus causes a higher level of capsule production (Kroll *et al.*, 1988). This results in a decreased susceptibility of the bacterium to complement-mediated lysis and opsonisation, suggesting that the *Hi*b capsule plays an important role in resisting killing by complement-mediated host defence mechanisms (Noel *et al.*, 1996).

Biosynthesis and transport of *Hi* capsule are encoded by the *cap* locus (Catlin *et al.*, 1972). The *cap* loci for all serotypes consist of three regions: region I, II, and III (Kroll *et al.*, 1991, Kroll *et al.*, 1989). Homologues of genes in region I can be found in a number of Gram-negative bacteria, such as *Escherichia coli (E. coli), Neisseria meningitidis (N. meningitidis), Actinobacillus pleuropneumoniae*, *Pasteurella multocida* (*P. multocida)* and *Mannheimia haemolytica* (Lo *et al.*, 2001). *bexDCBA* belong to region I of the *Hi cap* locus and encode an ATP-dependent capsule export system (Kroll *et al.*, 1990). Region II contains serotype-specific polysaccharide biosynthesis genes and is unique to each of the six capsule types (Van Eldere *et al.*, 1995). Region III genes are associated with capsule postpolymerization (Frosch & Muller, 1993, Lo et al., 2001).

1.2.2 **Surface proteins**

Several *Hi* surface proteins have been identified to play a role in assisting bacterial adherence onto respiratory cells, thus preventing clearance of the organism by mucociliary escalator. These surface proteins include haemagglutinating pili (Guerina *et al.*, 1982, Pichichero *et al.*, 1982), *Haemophilus* adhesion and penetration protein (Hap) (St Geme *et al.*, 1994a), high molecular weight (HMW1, HMW2) adhesins (Barenkamp & Leininger, 1992b, St Geme *et al.*, 1993), *Hi* autotransporter adhesin (Hia) (Barenkamp & St Geme, 1996), and *Haemophilus* surface fibrils (Hsf) (Geme & Cutter, 1995).

The types of surface proteins expressed between encapsulated and NT*Hi* strains can be different. The haemagglutinating pilus is a common adhesin expressed on encapsulated strains (64% of 170 *Hi*b strains studied), but is only expressed in 28% of NT*Hi* strains (Ecevit *et al.*, 2004). The HMW adhesins are expressed in 75% of NT*Hi* strains (St Geme *et al.*, 1993) and are absent in all screened *Hi*b strains (Ecevit et al., 2004). The Hsf protein appears to be ubiquitous among *Hi*b and other typeable strains, but is absent in NT*Hi* strains (St Geme *et al.*, 1996).

The *Hi* haemagglutinating pili, also known as fimbriae, are hair-like appendages that are formed by polymeric helical structures. They extend from bacterial surfaces and assist in bacterial adherence to human oropharyngeal epithelial cells (through binding to sialylated

lactosylceramide) and to respiratory mucous (Kubiet *et al.*, 2000, van Alphen *et al.*, 1991). The presence of pili on bacterial surfaces elicits an adaptive immune response. Interestingly, *Hi* strains isolated from blood and CSF have all been found to be non-piliated (Pichichero *et al.*, 1982, Weber *et al.*, 1991).

The Hap protein is a surface protein found in both encapsulated and NT*Hi* strains, and is highly homologous to bacterial IgA protease. There are two known functions of Hap: 1) to promote bacterial adherence and invasion of epithelial cells, and 2) to mediate bacterial aggregation and microcolony formation (St Geme *et al.*, 1994b). It is the aggregation of bacteria that results in development of bacterial resistance against bacteriostatic compounds (Hendrixson & St Geme, 1998).

Adhesion proteins that are more specific to NT*Hi* strains include the HMW1, the HMW2 and the Hia proteins (Barenkamp & Leininger, 1992b, Barenkamp & St Geme, 1994). 90% of NT*Hi* strains deficient in HMW1 and HMW2 proteins express the Hia protein, and strains that express the Hia protein do not express the HMW proteins. In approximately 5% of NT*Hi* strains, neither the HMW nor the Hia proteins are present. The HMW1 and HMW2 proteins are encoded by two separate loci: *hmw1A* and *hmw2A* (Barenkamp & Leininger, 1992a), and although these two proteins appear to be homologues of each other, they seem to recognise different host receptors.

The HMW1 protein recognises a glycoprotein receptor containing sialyl-α2,3 hexose (St Geme, 2002), and the exact receptor which HMW2 recognises remains unknown. Studies using the HMW1, the HMW2 and the Hia proteins showed that these adhesins are capable of inducing mucosal and systemic immunity, and are potential vaccine targets that can be used to prevent NT*Hi* induced otitis media: in a study using adenovirus as a vector to express the HMW1, HMW2, or Hia adhesins, it was shown that chinchillas immunised intraperitoneally with a single dose of either the HMW2 or the Hia adenoviral construct developed high level of anti-HMW2 or anti-Hia serum antibody within 4 weeks of immunisation, and that chinchillas immunised intranasally with a single dose of the Hia adenoviral construct also developed high anti-Hia serum antibody titers within 8 weeks of immunisation (Winter & Barenkamp, 2010). These results suggest that recombinant adenoviruses expressing the HMW1, HMW2, or Hia adhesins may be a promising means to induce mucosal and systemic immunity to protect against *Hi* induced otitis media.

1.2.3 **IgA1 proteases**

Immunoglobulin (Ig) A is the most prevalent antibody type found on human mucosal surfaces, and subtype IgA1 is the dominant antibody found in the upper respiratory tract (Kett *et al.*, 1986). IgA can bind to microbes and interfere with bacterial adherence to host cells

(Kraehenbuhl & Neutra, 1992). Capsular and NT*Hi* strains secrete IgA1 proteases (Kilian *et al.*, 1979), which assist in bacterial evasion from antibodies (Lomholt *et al.*, 1993).

97% of *Hi* strains contain the *iga* gene, which encodes the IgA protease. *Hi* can produce three distinct types of IgA protease, which cleave IgA at different amino acid positions (Kilian et al., 1996, Mulks et al., 1982, Fernaays *et al.*, 2006). IgA protease cleaves the heavy chain of IgA at the hinge region (Kilian *et al.*, 1996, Mulks *et al.*, 1982), and although the Fab component of cleaved IgA is still capable of binding to its cognate epitope on the bacterial surface, but without the presence of one or both Fc component, the antibody is non-functional (Poulsen et al., 1992). The *Hi* IgA protease only recognises IgA amino acid sequences from human and great ape (Johnson *et al.*, 2009).

1.2.4 **Lipopolysaccharide**

The LPS molecule is a complex surface glycolipid that is found exclusively in Gram-negative bacteria. LPS helps maintain bacterial membrane integrity and plays an important role in bacterial colonization, persistence and survival within hosts, and is therefore a major virulence determinant of *Hi* (Zwahlen *et al.*, 1986).

The general structure of *Hi* LPS consists of a conserved inner core covalently linked to a heterogeneous outer core (Fig. 1). Heterogeneity in the *Hi* LPS outer core can be found between and within strains (Kimura & Hansen, 1986). Studies on various LPS glycoforms from different *Hi* strains have shown that both major and minor LPS glycoforms can be present at the same time within a population of bacteria of a *Hi* strain (Masoud *et al.*, 1997). Most enteric Gram-negative bacteria contain a third LPS component, the O-antigen, which is a repeating polysaccharide side chain (Levine *et al.*, 1983). However, the O-antigen is absent in many non-enteric Gram-negative bacteria, such as *Hi* (Flesher & Insel, 1978); as a result, some literature refers to the LPS of *Hi* as lipo-oligosaccharide (LOS). Although the O-antigen is absent in *Hi*, this organism can nonetheless synthesise a high molecular weight LPS glycoform (HMG), which is similar to O-antigen, except the single tetrasaccharide unit added to the LPS *en bloc* is never polymerised (Hood *et al.*, 2004). An example of HMG can be found in strain Rd, where two different but related units are found (Fig. 3).

The conserved inner core of a *Hi* LPS molecule is made up of the following components: a triheptosyl backbone linked to a lipid A moiety via a phosphorylated 2-keto-3-deoxyoctulosonic acid (Kdo) residue (Zamze *et al.*, 1987, Zamze & Moxon, 1987, Risberg *et al.*, 1999), a glucose residue linked to the first heptose sugar (Phillips *et al.*, 1993, Masoud *et al.*, 1997), and a phosphoethanolamine moiety linked to the second heptose sugar (Phillips *et al.*, 1996) (Fig. 1).

The hydrophobic lipid A molecule, embedded within the bacterial outer membrane, is the anchor to which the inner core is attached, and is responsible for the endotoxic effects associated with Gram-negative bacterial infection (Flesher & Insel, 1978, Raetz, 1990). The lipid A molecule of *Hi* consists of a two glucopyranose rings and contains phosphate groups at the C1 and C4' positions (Fig. 2). Amide-linked and ester-linked 3-hydroxytetradecanoic acid are substituted at the C2/C2' and C3/C3' positions, respectively. The fatty acid chains at the C3' and C2' positions are further esterified by tetradecanoic acid chains (Helander *et al.*, 1988) (Fig. 2). Fatty acid chains of lipid A molecule are responsible for the endotoxic effects; therefore the number, positions, lengths and constituents of fatty acid chains are all important and contribute to varying degrees of toxicity.

Linked to the lipid A molecule is the phosphorylated Kdo residue, a molecule to which the heptose (Hep) trisaccharide backbone is attached (Fig. 1). The *kdkA* and *kdtA* are genes that respectively encode proteins that phosphorylate Kdo and add Kdo to lipid A (Fig. 3) (White *et al.*, 1997, White *et al.*, 1999). An attempt to create a *kdtA* inactivated mutant was unsuccessful, suggesting that this mutant is not viable (Hood *et al.*, 2001b). Heptosyltransferases that transfer each of the three heptoses in the core region have been identified and characterised in *Hi*, and they are encoded by *opsX*, *rfaF*, and *orfH* (Gronow *et al.*, 2005, Nichols *et al.*, 1997) (Fig. 3).

Phosphate-based molecules that respectively substitute the Kdo and HepII are pyrophosphoethanolamine (P..PEtn) and phosphoethanolamine (PEtn) (Fig. 1 and Fig 4). A PCho can be found attached to a hexose (Hex) or a heptose (Hep) in the oligosaccharide extension (Fig. 3) (Risberg et al., 1999, Lysenko *et al.*, 2000). Other common but variable substitutions of the LPS inner core include glycine (Li *et al.*, 2001) and O-acetyl groups (Fig. 3) (Schweda *et al.*, 2000, Mansson *et al.*, 2001, Mansson *et al.*, 2003).

The variable outer core of *Hi* LPS is represented by R1, R2 and R3 in Fig. 1. One of three possibilities can arise at each of the Hep, giving rise to the heterogeneity observed in the LPS outer core. The three possibilities are: 1) no sugar extension, 2) a monosaccharide extension, or 3) an extension consisting of oligosaccharide of variable composition. Various sugars that make up the LPS outer core include Hex, Hep, N-acetylhexosamines, and sialic acid (SA) in the form of 5-acetyl-neuraminic acid (Neu5Ac). Substitutions with PCho and/or O-acetyl groups further diversify the LPS outer core (Weiser *et al.*, 1997, Risberg *et al.*, 1999, Schweda et al., 2000).

Several enzymes that synthesise the *Hi* LPS outer core are phase variably expressed, and in each case, this involves a tetranucleotide repeat mediated mechanism (Weiser *et al.*, 1989). The ability to switch surface-expressed structures on or off provides further LPS heterogeneity within any given population of *Hi* strain. Examples of phase variable genes in *Hi* include sialyltransferase

genes *lic3A* and *lic3B* (Fig. 3). *lic3A* is present in all NT*Hi* strains screened, while *lic3B* is only found in approximately 60% of NT*Hi* strains screened (Hood *et al.*, 2001a, Fox *et al.*, 2006) (see section 1.4.2). The tetranucleotide repeats (5'-CAAT-3') of *lic3A* and *lic3B* are found at the 5' end of the reading frames (Weiser *et al.*, 1990, Hood et al., 2001a, Fox et al., 2006). The presence of tetranucleotide repeats can result in a high mutation rate, due to an increased likelihood of mispairing. Slipped strand mispairing is a process where short repeat sequences between mother and daughter DNA strands mispair during DNA replication, which results in a change in the number of repeats. When this happens, frame-shift occurs, leading to the loss of correct translation of reading frames (van Belkum *et al.*, 1998, Levinson & Gutman, 1987). In many organisms, phase variation has been adopted to allow the bacteria to change the expression of cell surface structures in a stochastic, high frequency and reversible manner. This makes it possible for a population of bacteria to adapt quickly through selection for individuals expressing the most optimal phase variable structures under a changed environment.

Previous study has shown that the *Hi* LPS outer core can mimic human antigens such as glycosphingolipids and some glycoproteins (Mandrell *et al.*, 1992), and that mimic epitopes are often the site for LPS sialylation. The term 'molecular mimicry' has been used to describe structural similarities found between microbial and host cell surfaces. LPS sialylation is a type of molecular mimicry, and this process seems to be adopted by bacteria that interact with

vertebrates (Lewis *et al.*, 2009, Vimr *et al.*, 2004). The presence of SA on the *Hi* LPS outer core can dampen down the host immune response, promote bacterial survival, and contribute to the pathogenesis of *Hi* (Schauer, 2009, Pangburn *et al.*, 2000).

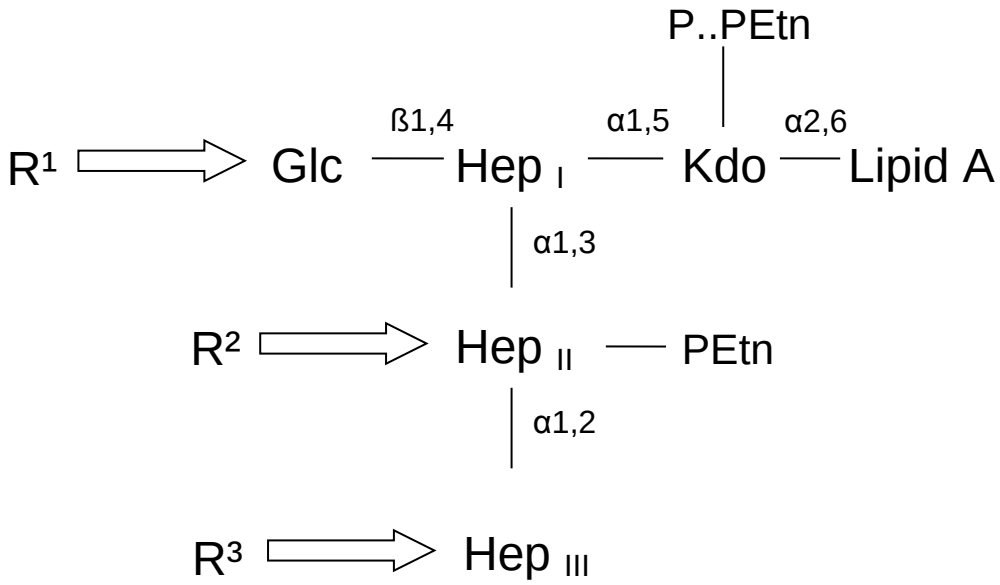


Fig. 1 Representation of the *Hi* LPS molecule.

The *Hi* LPS inner core consists of a triheptosyl backbone linked to a lipid A moiety via a phosphorylated Kdo residue, a phosphoethanolamine moiety linked to the second heptose sugar, and a glucose residue linked to the first heptose sugar.

R1, R2 and R3 represent the variable LPS outer core. One of three possibilities can arise at R1, R2 and R3: 1) no sugar extension, 2) a monosaccharide extension, or 3) an extension of oligosaccharide consisting of variable composition.

Represented in the figure are: Kdo, 2-keto-3-deoxyoctulosonic acid; Hep, L-glycerol-D-manno-heptose; Glc, glucose; PEtn, phosphoethanolamine; PPEtn, pyrophosphoethanolamine.

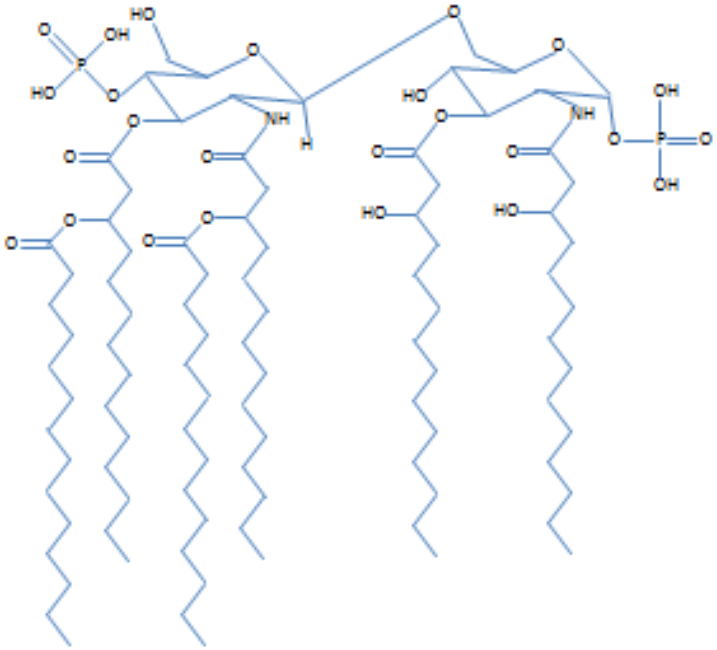


Fig. 2 General structure of the *Hi* lipid A molecule.

The *Hi* lipid A molecule consists of two glucopyranose rings, and phosphate residues at C1 and C4' positions. Ester and amide chains are linked at C2/C2' and C3/C3' positions. Six fatty acid chains extend from C2/C2' and C3/C3' positions.

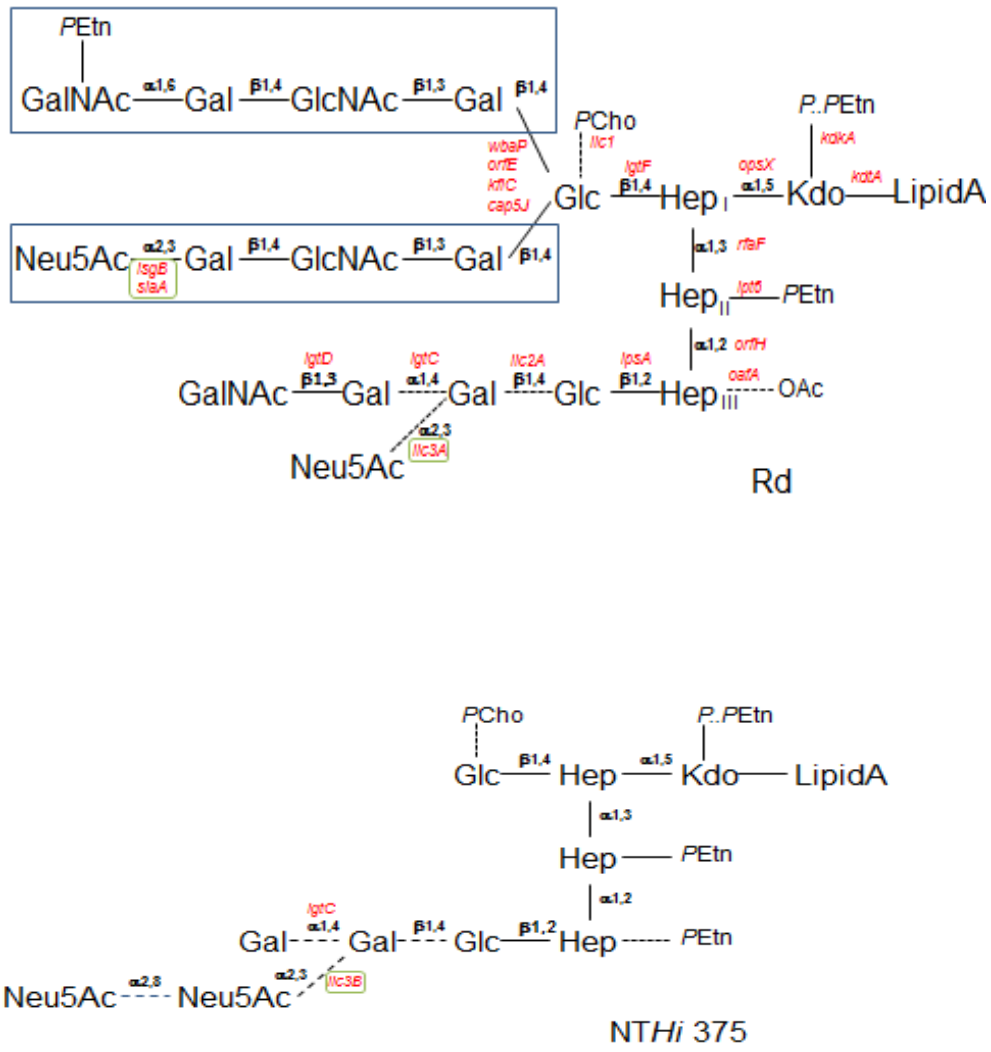


Fig. 3 LPS structures of *Hi* strains Rd (top) and 375 (bottom).

Genes that encode each transferase in *Hi* strain Rd are shown in red. Dotted lines represent phase variable LPS extensions. In strain Rd, one of two tetranucelotide units can be added *en bloc* (shown in black boxes). Sialyltransferase genes (*lic3A*, *lic3B*, *lsgB* and *siaA*) are drawn in green circles. *lic3A* and *lic3B* are phase variable genes. *lic3A* encodes a sialyltransferase possessing an α-2,3- activity, and *lic3B* encodes a sialyltransferase possessing both α-2,3- and α-2,8- activity. *lic3B* is not present in strain Rd. *lsgB* and *siaA* in strain Rd encode sialyltransferases possessing an α-2,3- activity for HMG synthesis.

1.3 General overview of SA

SA in the form of Neu5Ac can form part of the oligosaccharide chain of the *Hi* LPS outer core (Fig. 3). SA is a family of nine carbon alpha-keto sugar acid (Warren, 1963, Traving & Schauer, 1998, Angata *et al.*, 2001). There are more than 50 types of naturally occurring SAs and most are derived from the modification of the molecule Neu5Ac (also known as N-acetylneuraminic acid or 2-keto-3-deoxy-5-acetamido-D-glycero-D-galacto-nonulosonic acid) (Roseman, 1970). There are two important biochemical and physical properties that SA possesses: the net electronegative charge and the relatively large size of the molecule.

The general structure of a SA sugar consists of a nine-carbon ring backbone with various substitutions at the amino and/or the hydroxyl groups (Fig. 4) (Varki, 1992, Schauer, 2004, Angata & Varki, 2002). The numbering system of carbon atoms of SA starts at the carboxyl group and continues around the chain (Fig. 4). Configurations of SA whereby the carboxyl group is placed at the axial or at the equatorial positions are respectively the α- or β-anomers (Fig. 4A). Deprotonatation of the carboxyl group at C1 at physiological pH results in the net electronegative charge of the sugar. The 5-amino group is generally coupled with an acetyl or a glycolyl group (Fig. 4B). When the 5-amino group is replaced by a hydroxyl group, the molecule 2-keto-3-deoxy-nonulosonic acid (KDN) is formed (Fig. 4B) (Nadano *et al.*, 1986, Kanamori *et al.*, 1990). N-glycolylneuraminic acid (Neu5Gc), another common SA sugar that only differs

from Neu5Ac by an oxygen atom, is formed by hydroxylation of the N-acetyl group of Neu5Ac (Fig. 4B) (Jourdian & Roseman, 1962, Schoop *et al.*, 1969). Substitutions at one or more hydroxyl groups by acetyl, methyl, lactyl, phosphate or sulphate groups of Neu5Ac, Neu5Gc and KDN further diversify the available forms of SAs (Warren, 1964, Schauer, 1987, Iwasaki *et al.*, 1990, Manzi *et al.*, 1990). Diacetylated SAs are produced by substituting a hydroxyl group with a second acetyl group (Fig. 4B). Examples of such sugars include 9-O-acetyl N-acetylneuraminic acid (Neu5,9Ac2) and 4-O-acetyl *N*-acetylneuraminic acid (Neu5,4Ac2), which respectively contain an O-acetyl groups at the 9 or 4 carbon position (Fig. 4B). Neu5,9Ac2 and Neu5,4Ac2 are found in the lower intestinal tract of human, and in the brain, the intestine and the liver of mice (Rinninger *et al.*, 2006).

Neu5Ac and Neu5Gc are the most common SAs found in mammalian tissues. Neu5Gc is generally absent in normal adult human tissues (Gottschalk, 1960), due to the absence of the enzyme required for the biosynthesis of Neu5Gc in man. The enzyme, cytidine monophosphate-N-acetylneuramic acid hydroxylase (CMAH), is inactivated due to the deletion of an exon in the *CMAH* gene. This gene encodes the hydroxylation of cytidine monophosphate-N-acetyl-neuramic acid (CMP-Neu5Ac) to cytidine monophosphate-N-glycolylneuramic acid (CMP-Neu5Gc), the essential step in the biosynthesis of Neu5Gc (Shaw & Schauer, 1989, Takematsu *et al.*, 1994, Irie *et al.*, 1998, Chou *et al.*, 1998).

The absence of Neu5Gc likely brings an evolutionary advantage in man as pathogens that preferentially recognise and bind to surface-expressed Neu5Gc would no longer pose a threat (Campanero-Rhodes *et al.*, 2007, Kyogashima *et al.*, 1989, Schwegmann-Wessels & Herrler, 2006, Rolsma *et al.*, 1998, Martin *et al.*, 2005). The presence of a basal level of circulating anti-Neu5Gc antibodies found in normal human serum is thought to be associated with dietary intake of food rich in Neu5Gc (such as meat and fish) (Zhu & Hurst, 2002, Tangvoranuntakul *et al.*, 2003, Nguyen *et al.*, 2005).

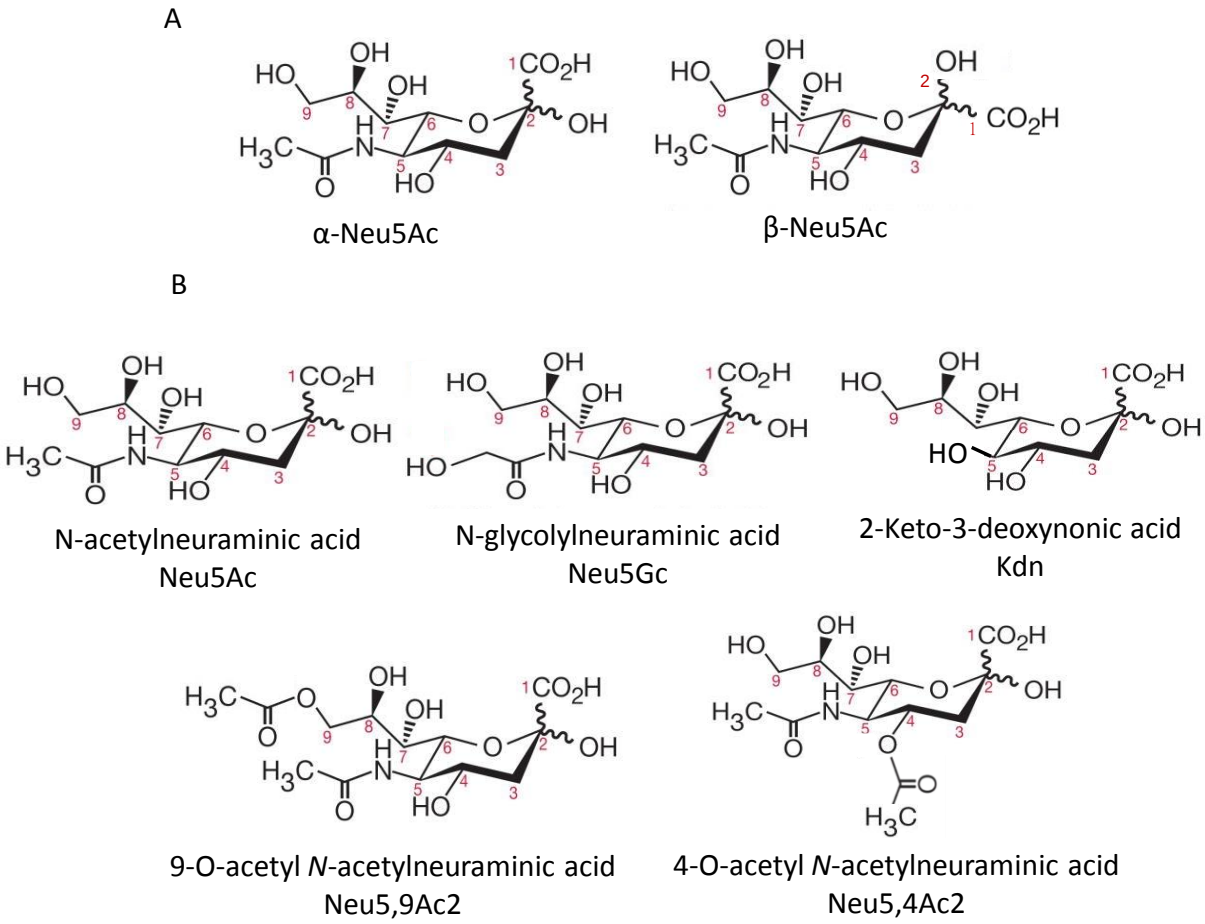


Fig. 4 Structures of common SAs.

In panel A: α- and β-anomers of Neu5Ac.

In panel B: The two most common SAs, Neu5Ac and Neu5Gc, differ from each other only by a hydroxyl group at the 5-amino terminus. KDN is a derivative of Neu5Ac: the 5-amino terminus is replaced by a hydroxyl group. Neu5,9Ac2 and Neu5,4Ac2 are diacetylated SAs, and contain an O-acetyl groups at the 9 and 4 carbon positions, respectively.

The numbers indicate the carbon atoms of SA, which starts from the carboxyl group and continues around the chain.

1.4 Sialometabolism of *Hi*

1.4.1 **Acquisition of environmental Neu5Ac**

Acquisition of Neu5Ac by *Hi*, as well as by many other microbial organisms such as some *E. coli* strains, depend on their successful scavenging of exogenous, free Neu5Ac (Vimr *et al.*, 2000). Neuraminidases (also known as sialidases) are enzymes that release free Neu5Ac from host cell glycoconjugates, and are secreted by organisms that reside within the same niche as *Hi* and by host cells (Fig. 6) (Bouchet *et al.*, 2003, Vimr & Troy, 1985a, Sohanpal *et al.*, 2004, Sohanpal *et al.*, 2007, Shakhnovich *et al.*, 2002). Secretion of neuraminidases by host cells is increased during inflammation, for example when host cells encounter bacterial LPS endotoxins.

Free Neu5Ac is transferred from the environment to the periplasm of *Hi*, and this is likely carried out by general porins such as OmpF and OmpC (Fig. 6) (Condemine *et al.*, 2005). Porins are beta-barrel proteins that cross the outer membrane of bacteria and act as pores through which molecules diffuse (Cowan *et al.*, 1992). In *E. coli*, a Neu5Ac specific porin has been identified, and this Neu5Ac specific porin, NanC, is induced by Neu5Ac when the general porins are not expressed (Condemine *et al.*, 2005). Since the homologue of *nanC* is not found in the *Hi* Rd genome sequence (see section 3.2.1, Table 4), it is likely that *Hi* utilises general porin/s for the transfer of Neu5Ac from the environment to its periplasm (Vimr, Fleischmann et al., 1995).

In *Hi*, transport of Neu5Ac from the periplasm to the cytoplasm is carried out by a transmembrane ATP-independent periplasmic (TRAP) transport system, which is a highly efficient and specific transport system (Fig. 6) (Severi *et al.*, 2005, Allen *et al.*, 2005). Other microbial organisms that utilise the Neu5Ac TRAP transport system include *P. multocida*, *Vibrio cholerae (V. cholerae)*, and *Fusobacterium nucleatum* (Chowdhury *et al.*, 2012, Steenbergen *et al.*, 2005). There are two other types of Neu5Ac transport systems, namely the Neu5Ac secondary and ABC transport systems. In general, a secondary transport system transports molecules via an ion- or a proton-coupled symporter or antiporter; and an ABC transport system relies on a periplasmic binding component and ATP hydrolysis. Both transport systems require a transmembrane component (Krulwich, 1983, Kelly & Thomas, 2001).

The Neu5Ac TRAP transport system consists of two major components: a periplasmic binding protein (also known as extracytoplasmic solute receptor [ESR]) and a transmembrane component (Forward *et al.*, 1997, Kelly & Thomas, 2001). The ESR of the Neu5Ac TRAP transport system is structurally similar to the periplasmic binding component of the ABC transport system (Muller *et al.*, 2006); however, the ESR does not require ATP hydrolysis (Severi et al., 2005), but instead depends on the electrochemical gradient of sodium (Kelly & Thomas, 2001, Severi *et al.*, 2010). The ESR of the Neu5Ac TRAP transport system possesses a high affinity and specificity for Neu5Ac, which are necessary and advantageous for neuraminidase-deficient organisms (such as

Hi) to efficiently scavenge the scarce amount of free Neu5Ac (Severi *et al.*, 2005). The transmembrane component of the Neu5Ac TRAP transport system is analogous to a secondary transporter. In contrast to a typical secondary transporter, which only consists of a single membrane unit, the transmembrane component of a TRAP system consists of two membrane subunits: a 12 transmembrane helix protein (Rabus *et al.*, 1999), and a four-transmembrane helix protein (Wyborn *et al.*, 2001). In many microbial organisms, this transmembrane component is encoded by two separate genes, namely *siaQ* and *siaM* (Forward et al., 1997, Kelly & Thomas, 2001, Rabus et al., 1999); however, in *Hi*, the two genes are fused and together they encode a single polypeptide (Allen *et al.*, 2005). The ESR and the transmembrane components of the *Hi* Neu5Ac TRAP transport system are encoded by *siaP* and *siaQM*, respectively (Fig. 7) (Steenbergen *et al.*, 2005, Allen et al., 2005, Severi et al., 2005). *SiaQM* has been referred to as *siaT* in some literature (Allen et al., 2005). Previous studies using RT-PCR has shown that *siaP* and *siaQM* are co-transcribed, confirming that the two genes belong to the same operon (Jenkins *et al.*, 2010). We shall refer to this operon as the *siaPQM* or the Neu5Ac transport operon from here onward.

Previous studies have shown that the ability to take up Neu5Ac is lost in *Hi* mutants deficient in either SiaP or SiaQM, and that the LPS profiles in these mutants are deficient in Neu5Ac (Allen et al., 2005, Severi et al., 2005). Furthermore, the *Hi siaP* and *siaQM* mutants have also become

more serum sensitive (Allen et al., 2005, Severi et al., 2005). When grown in the presence of Neu5Ac, the *Hi siaP* mutant is unable to utilise this sugar as a carbon and energy source (Severi et al., 2005).

The Neu5Ac transport system (a symporter) found in *E. coli* is another well studied Neu5Ac transport system. NanT, the specific Neu5Ac transporter in *E. coli*, belongs to the major facilitator superfamily (Vimr et al., 2004). Free Neu5Ac enters the *E. coli* cytoplasm via symport with a proton through NanT, and NanT functions as the sole Neu5Ac transporter in *E. coli* (Vimr et al., 2004). A *nanT* homologue has been identified in *Hi* (*HI1104*); however, deletion of this gene does not seem to affect Neu5Ac uptake or LPS sialylation (Severi et al., 2005, Allen et al., 2005). It is interesting to note that the *E. coli* YjhB protein is also a member of the major facilitator superfamily, and is a homologue of NanT. The exact function of YjhB remains unknown; although it has been postulated to be involved in sialometabolism of *E. coli*, possibly via the uptake of a SA related molecule (Steenbergen *et al.*, 2009).

Recently, a Neu5Ac mutarotase (NanM) has been identified in *E. coli*. It has been shown *in vitro* that NanM accelerates the conversion of α-Neu5Ac (the anomeric form of newly cleaved Neu5Ac) to β-Neu5Ac (the thermodynamically stable form of Neu5Ac, and the predominant form in solution), even though spontaneous mutarotation of Neu5Ac (from α- to β-Neu5Ac) can

occur without NanM, but the rate is slow (at $t_{0.5}$ of 80 min at pH 5.5) (Severi et al., 2008). Severi and colleagues suggested that rapid mutarotation of Neu5Ac is required because it is likely the NanT Neu5Ac transport system of *E. coli* transports Neu5Ac in its β-anomeric form and bacteria possessing the Neu5Ac mutarotase would be at a fitness advantage to successfully compete against other organisms for the limited amount of newly-cleaved, free Neu5Ac (Severi et al., 2008). A homologue of *nanM* has been identified in *Hi* (*HI0148*) (Fig. 6 and Fig. 7), which is one of the main focuses of our study (see chapters 3 and 4).

1.4.2 **Sialylation of the *Hi* LPS**

Sialylation of LPS is a structural feature found in many mucosal pathogens, and that includes species in the genera *Haemophilus* (Mandrell et al., 1992, Mansson et al., 2001, Bauer *et al.*, 2001, Hood *et al.*, 1999a), *Neisseria* (Kahler & Stephens, 1998) and *Campylobacter* (Gilbert *et al.*, 2000). Decoration of *Hi* LPS with Neu5Ac aids immune evasion in this organism by interfering with complement activation, and is associated with virulence. It has been shown that all NT*Hi* otitis media isolates tested to date have the ability to incorporate Neu5Ac into their LPS (Bauer et al., 2001), although the level and pattern of LPS sialylation may be quite different among strains (Hood et al., 1999a, Hood et al., 2001a). Strains that express sialylated LPS become more resistant to the bactericidal activity of normal human serum (Bouchet et al., 2003).

Sialylation of LPS in *Hi* (which shall be referred to as the anabolism of Neu5Ac) begins with the activation of cytoplasmic Neu5Ac, and is carried out by the enzyme CMP-Neu5Ac synthetase, encoded by *siaB* (Fig. 6). β-Neu5Ac is likely the anomer that is recognised by SiaB (Mizanur & Pohl, 2008). The activated form of Neu5Ac, CMP-Neu5Ac, contains a cytidine monophosphate attached to Neu5Ac, and is the substrate for sialyltransferases. Sialyltransferases are responsible for transferring Neu5Ac onto appropriate LPS acceptors via specific linkages. Four sialyltransferases have been identified in *Hi*, and they are Lic3A, Lic3B, LsgB and SiaA (Fox *et al*., 2006, Jones *et al.*, 2002, Hood et al., 2001a). The four sialyltransferases determine the frequency and pattern of LPS sialylation in this organism. In most *Hi* strains, three different patterns of LPS sialylation have been documented: 1) sialylated lactose, 2) disialylated lactose, and 3) sialylated N-acetyllactosamine (Fig. 3) (Fox et al., 2006, Hood et al., 1999a, Hood et al., 2001a, Cox *et al.*, 2002).

Lic3A is an α2-3 sialyltransferase that adds Neu5Ac to a lactosyl acceptor to create the sialylated lactose glycoform. An example of this LPS extension is found in strain Rd (Fig. 3). The action of Lic3A appears to be influenced by another phase variable gene, *lgtC*, which encodes a galactosyltransferase that competes for the same lactosyl acceptor as Lic3A (Hood et al., 2001a) (Fig. 3). Lic3B is a homologue of Lic3A, and is present in approximately 60% of NT*Hi* strains screened (Fox et al., 2006). Lic3B is as a bifunctional sialyltransferase that adds Neu5Ac onto a

1) lactosyl, and a 2) sialyllactose acceptor, via α2-3 and α2-8 linkages, respectively (Fox et al., 2006). Examples of sialyllactose and disialyllactose LPS moieties can be found in NT*Hi* strain 375 (Fig. 3). SiaA and LsgB are α2-3 sialyltransferases that are involved in the sialylation of N-acetyllactosamine in the *Hi* HMG (Fig. 3) (Hood et al., 2004). Approximately 60% of NT*Hi* strains screened contain the HMG locus (Hood et al., 2004). The *siaA* gene, located in the *hmg* locus, is replaced by *orfY* in strain Rd (Hood et al., 2004). *SiaA* and *orfY* are 64% similar to each other (Hood et al., 2004), and are both homologues of the sialyltransferase gene *lst* from *Haemophilus ducreyi* (Bozue *et al.*, 1999). The *lsgB* gene is not located in the *hmg* locus, and is a homologue of the sialyltransferase gene *lst* from *N. meningitidis* (Gilbert *et al.*, 1996).

Inactivation of sialyltransferases produced mutants that were significantly less resistant against the killing effect of human serum than their respective wild type strains (Bouchet et al., 2003, Hood et al., 2001a, Fox et al., 2006). In a chinchilla model of otitis media, mutants that were deficient in Lic3A or SiaB became profoundly attenuated (Bouchet et al., 2003), and SiaB inactivated strains became fully virulent again when infecting complement-depleted chinchillas (Figueira et al., 2007). It is likely that the *Hi* LPS possesses target site/s for complement protein C3 binding, and that sialylation of LPS masks this site and prevents interaction of complement protein with the bacterium (Figueira et al., 2007). Although sialylation of the *Hi* LPS is a major

virulence determinant, the expression of Neu5Ac on LPS has no effect on the attachment to, and the invasion of, cultured human epithelial cells or neutrophils (Hood et al., 1999a).

Similar to sialylation of the *Hi* LPS molecule, sialylation of the *Neisseria gonorrhoeae (N. gonorrhoeae*) LPS also inhibits the alternative pathway of complement, although the mechanism is quite different (Ram *et al.*, 1998). Sialylation of the gonococcal LPS causes an increase in the binding of factor H (fH), an anti-activator of the complement alternative pathway, to the bacteria (Ram et al., 1998). fH downregulates the insertion of membrane attack complex (the terminal step of the complement cascade), prevents complement activation and increases the resistance of *N. gonorrhoeae* against the killing effect of human serum (Ram et al., 1998).

Previous studies have shown that pathogens expressing sialylated LPS glycoforms that mimic human gangliosides (glycosphingolipids with a SA residue linked to the oligosaccharide chain) are aetiological factors for post-infective auto-immune disorders, such as Guillain Barré Syndrome (GBS) and Fisher Syndrome (FS) (Jacobs *et al.*, 1998, Koga et al., 2005, Yuki, 2005, Willison, 2005). GBS and FS are characterised by the production of anti-ganglioside antibodies that target the gangliosides of peripheral nerves (for GBS) (Willison, 2005) and oculomotor and primary sensory nerves (for FS) (Yuki, 2005). The *Hi* strain DH1, a strain that contains disialylated LPS that mimics the terminal trisaccharide unit of the ganglioside GQ1b, was

isolated from the serum of a patient suffering from FS (Koga *et al.*, 2005). The LPS structure of *Hi* strain DH1 is shown in Fig. 5. Experimental studies suggest that the disialylated LPS glycoform of *Hi* strain DH1 may be a precipitating factor for the development of anti-GQ1b antibody, leading to FS, in this patient (Fig. 5) (Houliston *et al.*, 2007).

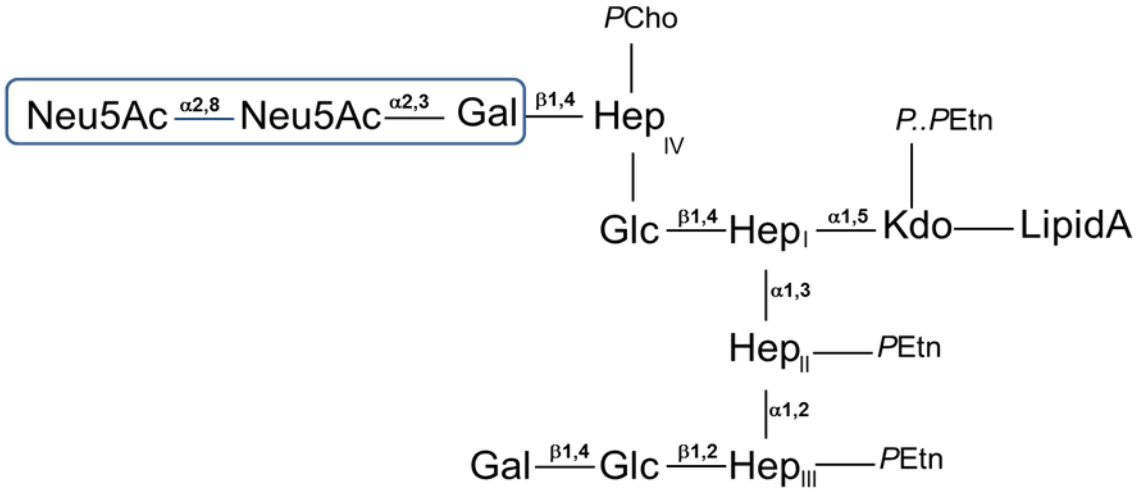


Fig. 5 LPS structures of *Hi* strains DH1.
Shown in the box is the trisaccharide unit found in all disialosyl-containing gangliosides.

1.4.3 **Catabolism of Neu5Ac in *Hi***

Many microbial organisms that are closely associated with vertebrates possess the ability to catabolise Neu5Ac. Phylogeny studies have shown that the genetic arrangement of Neu5Ac catabolism genes is diverse among both Gram-negative and Gram-positive species (Vimr et al., 2004).

The ability to catabolise Neu5Ac by *Hi* seems to be a protective mechanism for the organism to prevent excessive build-up of toxic Neu5Ac within the cell (Johnston *et al.*, 2007). Previous study using a NanA inactivated mutant in *Hi* (a mutant defective in the first step of Neu5Ac catabolism, Fig. 6) showed that the mutant is still capable of colonising the infant rat nasopharynx, indicating that Neu5Ac catabolism is not obligatory for colonisation within the host and that Neu5Ac is unlikely a major nutritional source for the bacterium (Vimr et al., 2000). In the *nanA* mutant, a significant increase in the percentage of sialylated LPS and an increase in survival with respect to serum killing was observed (Vimr et al., 2000).

Catabolism of Neu5Ac by *Hi* begins with the conversion of Neu5Ac to N-acetyl mannosamine (ManNAc) by the NanA aldolase (Fig. 6) (Vimr et al., 2000, Vimr & Troy, 1985a, Vimr & Troy, 1985b). This process produces pyruvate as a by-product, which may enter the phosphoenolpyruvate (Embden-Meyerhof-Parnas) pathway or the acetyl coenzyme A

(tricarboxylic or citric acid) pathway. Phosphorylation of ManNAc to N-acetyl mannosamine 6-phosphate (ManNAc-6P) and epimerization of ManNAc-6P to N-acetyl glucosamine 6-phosphate (GlcNAc-6P), respectively catalysed by NanK and NanE, then follow (Plumbridge & Vimr, 1999) (Fig. 6). GlcNAc-6P can then enter an amino sugar catabolism pathway where deacetylation of GlcNAc-6P to glucosamine 6-phosphate (GlcN-6P) and deamination of GlcN-6P to fructose 6-phosphate (Fru-6P), respectively catalysed by NagA and NagB, occur (Plumbridge & Vimr, 1999). Fru-6P can then enter the glycolysis pathway to produce carbon and energy for the bacterium. We shall refer to this Neu5Ac catabolism pathway carried out by *Hi* as the classic Neu5Ac catabolism pathway (Fig. 6).

An alternative Neu5Ac catabolism pathway has been documented, involving an epimerase that converts ManNAc to N-acetyl glucosamine (GlcNAc) and a kinase that phosphorylates GlcNAc to GlcNAc-6P (Fig. 6) (Biswas *et al.*, 1979). The epimerisation and the phosphorylation processes, respectively catalysed by the N-acetyl-D-glucosamine 2-epimerase and RokA kinase, bypass the phosphorylation step of ManNAc to ManNAc-6P and the epimerisation step of ManNAc-6P to GlcNAc-6P (the second and third steps of the classic Neu5Ac catabolism pathway, respectively) (Fig. 6) (Biswas et al., 1979).

In *Hi*, genes that make up the *Hi* Neu5Ac catabolism operon (also known as the *nagnan* operon) are adjacent to, and divergently transcribed from, the Neu5Ac transport operon (Fig. 7). Genes

that belong to the *Hi nagnan* operon include: *nagA, nagB, nanA, nanK, and nanE,* which respectively encode the GlcNAc-6P deacetylase, GlcN-6P deaminase, Neu5Ac aldolase, ManNAc kinase, and ManNAc-6P epimerase. Arrangement of the *Hi nagnan* operon can be very different to the arrangement of Neu5Ac catabolism operons in other organisms, such as *E. coli*, where the *nagA* and *nagB* genes are located at a distance to the rest of the Neu5Ac catabolism genes (Kalivoda *et al.*, 2003). Since the *Hi nagnan* and Neu5Ac transport operons are adjacent to, and divergently transcribed from, each other; we shall refer to the combined *Hi nagnan* and Neu5Ac transport operons as the *Hi* sialometabolism operons or the *Hi* sialometabolism gene cluster.

Previous studies in *E. coli* have shown that this bacterium is capable of utilising the diacetylated form of SA- Neu5,9Ac2 (but not Neu4,9Ac2) as a sole sugar source (Fig. 4) (Plumbridge & Vimr, 1999). Neu5,9Ac2 is the predominant diacetylated form of SA found in the human gut (Robbe *et al.*, 2003, Robbe-Masselot *et al.*, 2009). Deacetylation of the carbon-9 acetyl group of Neu5,9Ac2 would result in the production of Neu5Ac. This process is encoded by *nanS* in *E. coli* and likely occurs in the periplasm of the bacterium (Plumbridge & Vimr, 1999).

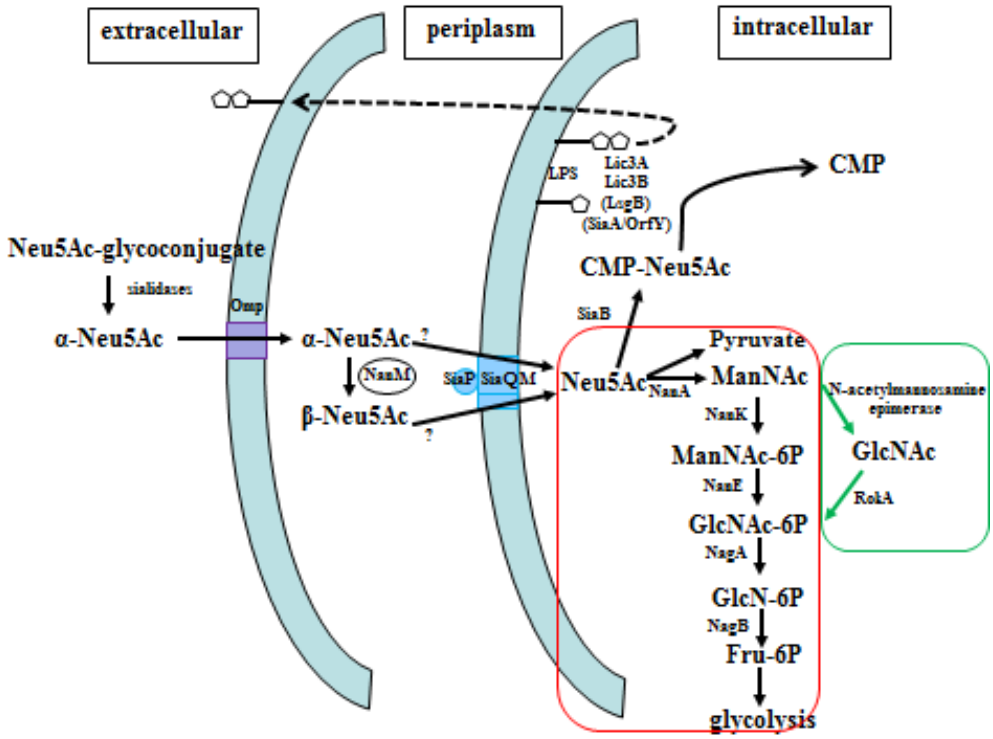


Fig. 6 Overview of the major pathways of Neu5Ac metabolism in *Hi*.

This figure does not represent all known Neu5Ac metabolic pathways of prokaryotes.

The putative Neu5Ac mutarotase, NanM (circled in black), is one of the genes studied. It has not yet been determined whether α-Neu5Ac or β-Neu5Ac or both is the substrate for the putative *Hi* NanM mutarotase. Catabolism of Neu5Ac in *Hi* is via the classical pathway (represented in the red block). An alternative Neu5Ac catabolism pathway exists (indicated in the green block), which bypasses the phosphorylation of ManNAc and the epimerisation of ManNAc- 6P to GlcNAc-6P. The sialyltransferases LsgB and SiaA/OrfY, shown in brackets, are responsible for the sialylation of N-acetyllactosamine in the *Hi* HMG.

Shown in the diagram are: Neu5Ac-glycoconjugate, sialoglycoconjugate; α-Neu5Ac, the α anomeric form of Neu5Ac; β-Neu5Ac, the β anomeric form of Neu5Ac; Omp, outer membrane protein; NanM, Neu5Ac mutarotase; SiaP, the ESR component of the *Hi* Neu5Ac TRAP transporter; SiaQM, the transmembrane component of the *Hi* Neu5Ac TRAP transporter; CMP-Neu5Ac, cytidine monophosphate-N-acetyl-neuramic acid; SiaB, CMP-Neu5Ac synthetase; Lic3A, α2-3 sialyltransferase; Lic3B, α2-3 and α2-8 sialyltransferase; LsgB and SiaA/OrfY, α2-3 sialyltransferases for the *Hi* HMG; NanA, N-acetyl neuraminate aldolase; NanK, N-acetyl mannosamine kinase; NanE, N-acetyl mannosamine-6-phosphate epimerase; NagA, N-acetyl glucosamine 6-phosphate deacetylase; NagB, glucosamine 6-phosphate deaminase; Neu5Ac, 5-acetyl neuraminate; ManNAc, N-acetyl mannosamine; ManNAc-6P, N-acetyl mannosamine 6-phosphate; GlcNAc, N-acetyl glucosamine; GlcNAc-6P, N-acetyl glucosamine 6-phosphate; GlcN-6P, glucosamine 6-phosphate; Fru-6P, fructose 6-phosphate; RokA, GlcNAc kinase.

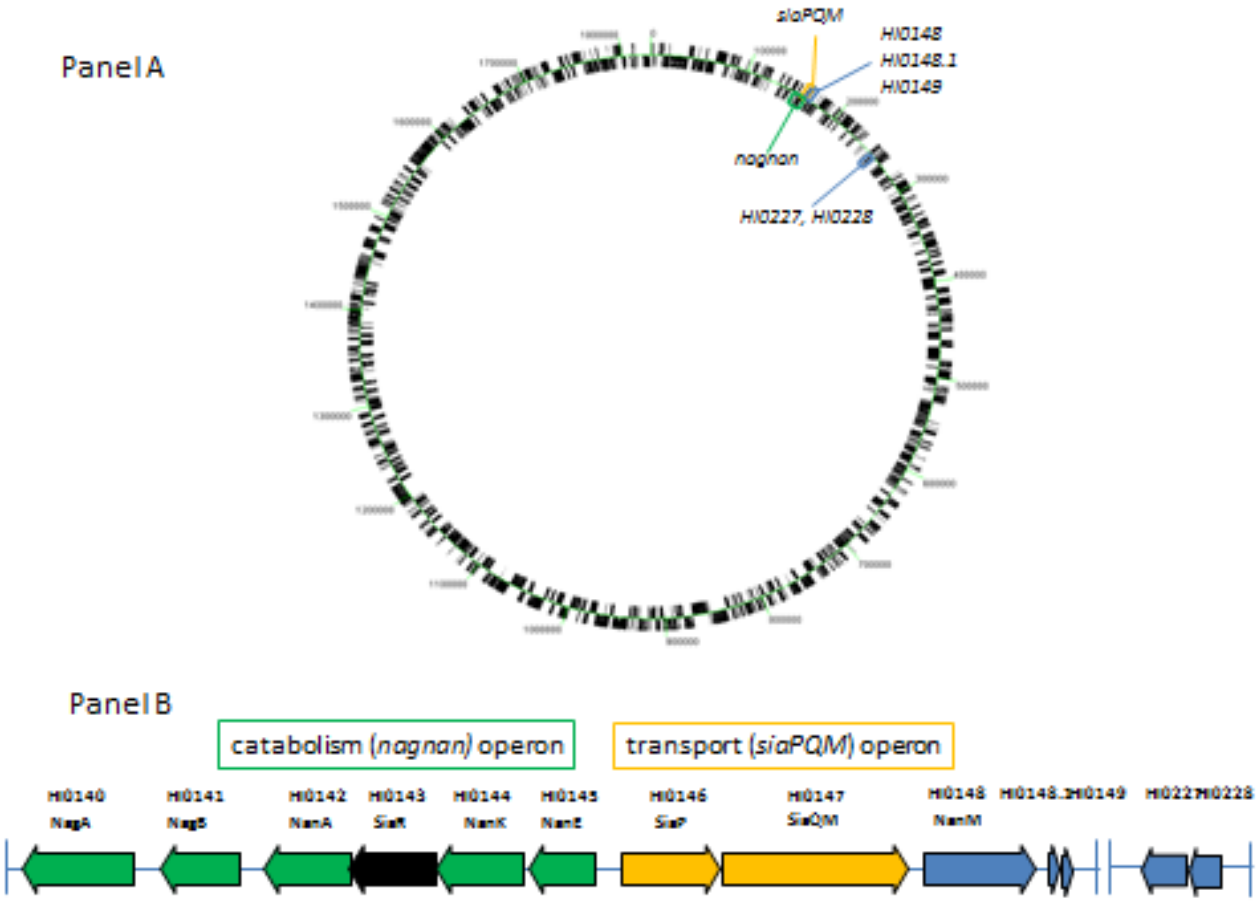


Fig. 7 Circular genome of *Hi* Rd and the locations of the *Hi* Rd Neu5Ac catabolism, transport and the *HI0227-HI0228* operons.

Shown in panel A is the circular genome of all known and putative ORFs in *Hi* Rd. The positions of the *Hi* Neu5Ac catabolism (*nagnan*) and transport *(siaPQM)* operons are respectively shown in green and orange boxes. The positions of the candidate (*HI0148* and *HI0227*) and peripheral candidate sialometabolism genes (*HI0148.1*, *HI0149* and *HI0228*) are shown in blue boxes.

Shown in panel B is the genetic arrangement of the Neu5Ac catabolism (represented by green arrows) and transport (represented by orange arrows) operons and the positions and transcriptional directions of the candidate (*HI0148* and *HI0227*, represented by blue arrows) and peripheral candidate sialometabolism genes (*HI0148.1*, *HI0149* and *HI0228*, represented by blue arrows).

The *Hi* Neu5Ac catabolism and transport operons are adjacent to, and divergently transcribed from, each other. *SiaR* (represented by the black arrow) encodes the transcriptional regulator of the Neu5Ac catabolism and transport operons, and is located within the Neu5Ac catabolism operon. Directions of the arrows represent the transcriptional directions of the ORFs. Loci numbers and gene names are stated above the arrows.

1.4.4 **Regulation of the Neu5Ac catabolism and transport operons**

The *Hi* sialometabolism gene cluster is complex, in order for this region to work efficiently, regulation is required. In *Hi,* two transcriptional regulators have been shown to play a role in the regulation of the *Hi* sialometabolism operons: SiaR and cAMP receptor proteins. cAMP receptor protein is also known as catabolite repressor protein (CRP) (Johnston et al., 2007, Jenkins et al., 2010). SiaR, encoded by *HI0143*, is located within the *Hi nagnan* operon (Fig. 7) and is a transcriptional repressor that downregulates the expression of the *Hi* sialometabolism operons (Johnston et al., 2007, Jenkins et al., 2010). The CRP protein, on the other hand, when bound to cAMP (cAMP-CRP complex), is a transcriptional activator of the *Hi* Neu5Ac transport operon, i.e. the expression of the *Hi* Neu5Ac transport operon, and not the *nagnan* operon, is upregulated by the cAMP-CRP complex (Johnston et al., 2007, Jenkins et al., 2010). cAMP-CRP is also known to affect the expression of genes involved in the metabolism of other sugars (such as galactose, ribose, xylose, and fucose) in *Hi* (Macfadyen *et al.*, 1996).

The SiaR repressor protein belongs to the RpiR family, and is characterised by the presence of a helix-turn-helix (HTH) DNA binding domain and a sugar isomerase (SIS) domain (Johnston et al., 2007, Bateman, 1999). The DNA binding site of SiaR spans 53 bp, and is located within the *nanE-siaP* intergenic region (113 bp and 185 bp upstream from the start codons of *nanE* and *siaP*, respectively) (Fig. 8) (Johnston et al., 2007). cAMP-CRP binds to a site within the

nanE-siaP intergenic region (Fig. 8) (Johnston et al., 2007, Jenkins et al., 2010). This CRP binding site was identified at 182 bp upstream from the start codon of *siaP*, and overlaps with the SiaR binding site by 3 bp (Fig. 8) (Johnston et al., 2007).

Johnston suggested that the SIS domain of SiaR binds to GlcN-6P, an intermediate phosphosugar in the Neu5Ac catabolism pathway (Fig. 6) (Johnston *et al.*, 2010). Using electrophoretic mobility shift assays, Johnston showed that the affinity of SiaR to its DNA binding site (*nanE-siaP* intergenic region, Fig. 8) is enhanced in the presence of GlcN-6P, and not in the presence of Neu5Ac, ManNAc, GlcNAc or GlcNAc-6P (Johnston et al., 2010, Johnston et al., 2007). Johnston suggested that SiaR and CRP co-operate with each other to modulate the expression of both the *nagnan* and the Neu5Ac transport operons (Johnston et al., 2007, Johnston et al., 2010). It was suggested that SiaR switches to become a transcriptional activator upon binding to GlcN-6P (which acts as a co-activator), and that the SiaR-GlcN-6P complex together with cAMP-CRP induce the expression of both *nanE* and *siaP* (genes used to represent the *Hi nagnan* and Neu5Ac transport operons, respectively) (Johnston et al., 2010). Furthermore, Johnston proposed that in the absence of SiaR, cAMP-CRP cannot influence the expression of the *nagnan* operon but can induce the *siaPQM* operon; and in the absence of GlcN-6P, SiaR returns to function as a transcriptional repressor, repressing both the *nagnan* and Neu5Ac transport operons (Johnston et al., 2010). It is interesting to note that this model proposed for the

regulation of the *Hi* sialometabolism operons by Johnston contradicts his previous finding, where he inferred that Neu5Ac is the sugar that is responsible for the induction of the expression of *nanE* and *siaP*, and that the conversion of Neu5Ac into CMP-Neu5Ac is responsible for inhibiting the induction of the *Hi* sialometabolism operon (Johnston et al., 2007).

It is important to mention here that the dual functions proposed for SiaR (the ability to function as both a transcriptional repressor and an activator) is atypical for a member of the RpiR regulator family, as all RpiR transcriptional regulators known to date (except for SiaR in *Hi*) function as repressors (see section 6.7) (Bateman, 1999, Sorensen & Hove-Jensen, 1996, Jaeger & Mayer, 2008, Kohler *et al.*, Daddaoua *et al.*, 2009). Furthermore, previous studies have shown that the repressor activity of RpiR-type transcriptional repressors is removed upon binding to their respective SIS sugars (Sorensen & Hove-Jensen, 1996, Jaeger & Mayer, 2008, Kohler et al., Daddaoua et al., 2009, Kim *et al.*, 2011, Olson *et al.*, 2013). The NanR proteins of *Vibrio vulnificus (V. vulnificus*) and *Staphylococcus aureus* (*S. aureus*) are both members of the RpiR-type transcriptional regulator, and function to regulate the expression of their respective sialometabolism operons (Kim et al., 2011, Olson et al., 2013). ManNAc-6P, an intermediate phosphosugar in the Neu5Ac catabolism pathway, is the SIS sugar that binds to the NanR regulators of *V. vulnificus* and *S. aureus*. In the absence of this SIS sugar, the NanR repressors of both organisms bind to their respective operator sites to represses Neu5Ac transport and

catabolism; and in the presence of ManNAc-6P, the binding of NanR repressors to their operator sites is weakened (Kim et al., 2011, Olson et al., 2013).

To determine whether inactivating *siaR* would affect LPS profiles and serum sensitivity, Johnston used a 2019 *siaR* mutant and showed that reduced incorporation of Neu5Ac into LPS and increased serum sensitivity were exhibited by this mutant (Johnston et al., 2007). Johnston suggested that in the 2019 *siaR* mutant, disrupting the regulation of Neu5Ac catabolism leads to an unregulated increase in the catabolism of Neu5Ac, which caused the observed decrease in LPS sialylation and increase in serum sensitivity (Johnston et al., 2007). However, these observations were not consistent with Jenkin's study where mutants that lack the SiaR function in *Hi* strains Rd, 375 and 486 were used. Jenkins showed that the *siaR* mutants from these strains exhibited different phenotypes: in the Rd *siaR* mutant, a similar LPS phenotype was exhibited by the mutant strain as in the wild type strain, but the Rd *siaR* mutant seemed to be slightly more serum sensitive than the wild type strain; in the 486 *siaR* mutant, a truncated LPS phenotype was exhibited, and this fits with the more serum sensitive phenotype observed in this mutant compared to the wild type strain; in the 375 *siaR* mutant, similar LPS phenotype and serum sensitivity were exhibited by the mutant strain as the wild type strain (Jenkins et al., 2010). These findings suggest that the regulation of the *Hi* sialometabolism operons may be more

complex than previously thought, and that the model provided by Johnston to explain the *Hi* sialometabolism gene regulation is probably not adequate to cover all aspects of it.

Furthermore, an inconsistent finding was observed with regard to the expression of *nanE* and *siaP* when Neu5Ac at 100 μM (30.9 μg/ml) was supplemented in the growth media: in the *Hi* 2019 wild type strain, Johnston observed no significant change in the expression of either *nanE* or *siaP* when cells were grown in the presence or absence of 100 μM Neu5Ac (Johnston et al., 2007), while in Jenkins's study, a two- to three-fold downregulation was observed in *nanE* and *siaP* when a similar amount of Neu5Ac (30 μg/ml) was added (Jenkins et al., 2010).

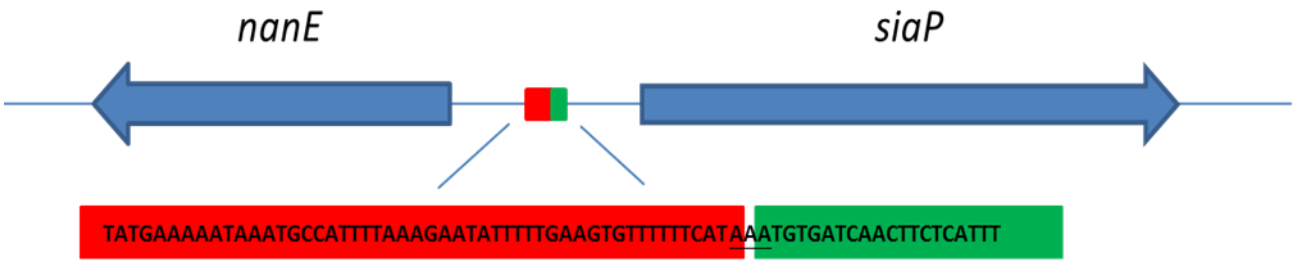


Fig. 8 The *nanE- siaP* intergenic region of *Hi*.
Block arrows indicate transcriptional directions of *nanE* and *siaP*. The SiaR and CRP binding sites are shaded in red and green, respectively. The three underlined nucleotides (AAA) found between the red and green boxes are the overlapping nucleotides of SiaR and CRP binding sites.

1.4.5 **SA and biofilm**

The term biofilm refers to a community of bacteria consisting of one or more species adhering to each other or to a non-living surface, and is surrounded by self-produced extracellular matrix (also known as slime or glue) (Costerton *et al.*, 1995). This matrix, produced by a programmed switch in gene expression, is made up of extracellular DNA, proteins and polysaccharides. It has been suggested that bacteria capable of producing a biofilm are at a fitness advantage as biofilm protects bacteria from harsh environments (such as in the presence of antibiotics), facilitates transport of nutrients through special channels, and prevents dislodgement of adherent bacteria from pipes or streams (Nickel *et al.*, 1985, Massol-Deya *et al.*, 1995) (Moxon *et al.*, 2008).

Many organisms, including *Hi*, have been documented to produce biofilms (Rayner *et al.*, 1998). The increase in antibiotic resistance and the failure to culture NT*Hi* from middle-ear effusions were original reasons put forward to suggest that NT*Hi* can produce a biofilm (Rayner *et al.*, 1998, Post, 2001, Ehrlich *et al.*, 2002). However, through detailed analyses of publications and *in vitro* experimental observations, it was later questioned whether NT*Hi* can truly produce a biofilm (Moxon *et al.*, 2008). Instead of the biofilm-specific genetic programme that signals the production of extracellular matrix, it has been suggested that the components of the so-called matrix may actually be made up of bacterial membrane, cytoplasmic contents (through blebbing and lysis of cells) and host cell debris generated from inflammation (Schooling & Beveridge,

2006, Brinkmann & Zychlinsky, 2007). What is also interesting is that this NT*Hi* biofilm disintegrates in the presence of antibiotics (Greiner *et al.*, 2004, Jurcisek & Bakaletz, 2007), an observation that contradicts to the conventional theory of biofilm which states that bacteria in a biofilm are often highly resistant against antibiotics (Moxon et al., 2008).

Although there is doubt regarding whether *Hi* can truly produce a biofilm; nonetheless, studies on this presumptive NT*Hi* biofilm have suggested that Neu5Ac may be an important component of it (Greiner *et al.*, 2004). The amount of biofilm produced by NT*Hi* strain 2019 in a continuous flow chamber is less when a chemically defined medium lacking Neu5Ac was used compared to the same medium containing 20 μM Neu5Ac (Greiner et al., 2004). Furthermore, LPS profiles of bacteria obtained from growth *in vitro* under conditions reported to be planktonic or biofilm showed an overall shift to lower phosphorylation states and an increase in HMG and sialylated glycoforms (Greiner et al., 2004). There are doubts regarding the interpretation of the LPS data because factors such as differences in media type (solid versus liquid) and composition have not been properly taken into account.

In lectin binding studies, where simultaneous analysis of two different lectins was performed, Greiner and colleagues have identified an α-2,6 linked Neu5Ac molecule from this presumptive biofilm produced by NT*Hi* strain 2019 (Greiner et al., 2004, Jurcisek *et al.*, 2005, Jurcisek &

Bakaletz, 2007). An α-2,6 Neu5Ac linkage has not been identified in *Hi* before, and this finding suggests a novel sialylated molecule may be associated with the formation of a NT*Hi* biofilm (Greiner et al., 2004). The presence of this α-2,6 linked Neu5Ac molecule needs to be confirmed by sensitive structural studies, such as nuclear magnetic resonance spectroscopy (Moxon et al., 2008).

In vitro study has suggested that sialylated LPS glycoforms promote the formation of biofilm, evidenced by less biofilm production in NT*Hi* 2019 *siaB* and *siaA* mutants compared to the wild type strain (Greiner et al., 2004). Ironically, inactivation of other sialyltransferase genes (*lic3A* and *lsgB*) did not produce such an effect, and the amount of biofilm produced was similar to the wild type strain (Greiner et al., 2004). These inconsistent findings question the likelihood of whether sialylated LPS glycoforms can truly promote biofilm formation. *In vivo* studies using the chinchilla model of otitis media have shown that the biofilm produced by NT*Hi* 2019 *siaA* and *lsgB* mutants were largely composed of dead organisms with dense and compact structures, while the NT*Hi* 2019 *siaB* mutant was unable to produce a biofilm (Jurcisek et al., 2005). It appears that clear evidence is lacking to substantiate that *Hi* can truly produce a biofilm, and we maintain our scepticism as to whether Neu5Ac plays an active part in promoting the formation of this presumptive biofilm.

1.5 Candidate sialometabolism genes in *Hi*

Sialometabolism is important in *Hi*, and contributes to the pathogenicity of the organism. Although the topic of sialometabolism has been studied for many years by different research groups, there are still aspects within this research area that have not been fully elucidated. The overall sialometabolism processes of *Hi* may encompass gene/s, or homologues of gene/s from another organism, that are located within sialometabolism gene clusters in either organisms but whose functions remain to be explored. Such genes, which can be predicted to possess sialometabolism function/s, will now be referred to as candidate sialometabolism genes. In this study, identification of candidate sialometabolism genes in *Hi* and the exploration of the possible role/s that they play have been conducted. Furthermore, an area of sialometabolism that is not fully understood, namely the regulation of the *Hi* sialometabolism operon, has been investigated. Reviewing these aspects of *Hi* sialometabolism will help further our current understanding of this complex process in *Hi*.

Two candidate sialometabolism genes that we focus on are the *HI0227* and *HI0148* genes from *Hi* strain Rd. There is evidence from literature to support that these genes play a role in sialometabolism of *Hi*. By studying candidate sialometabolism genes, such as *HI0227* and *HI0148*, we aim to further our current understanding of the complex processes of sialometabolism in *Hi*.

2 Materials and Methods

2.1 Materials

2.1.1 **Bacterial strains and plasmids used in this study**

In this study, the only *Hi* strain used that was non-NT*Hi* was RM118 (Rd), which is a capsular deficient derivative of a type d capsular strain, and is derived from the same source of strain used for the original sequencing project (http://www.tigr.org).

All other *Hi* strains described in this study are NT*Hi* strains. 25 NT*Hi* strains, designated 162, 176, 285, 375, 432, 477, 486, 667, 723, 981, 1008, 1124, 1158, 1159, 1180, 1181, 1200, 1207, 1209, 1231, 1232, 1233, 1247, 1268, and 1292, come from a Finnish otitis media study and were isolated by tympanocentesis from the middle ear of Finnish children who had otitis media. 25 NT*Hi* strains were selected based on ribotyping analysis of a population of 102 strains from the study to represent genetic diversity (Cody *et al.*, 2003).

NT*Hi* strain R2866 is an invasive NT*Hi* isolated from the blood of a 30-month-old infant and NT*Hi* strain R2846 is a middle ear isolate. These two strains were obtained from Arnold Smith's lab, Seattle Biomedical Research Institute, Seattle, Washington.

NT*Hi* strain 2019 was provided by Jim Richard's lab at Ottawa, Ontario, Canada. The strain was originally obtained from the sputum of an adult at Buffalo, New York (Murphy *et al.*, 1986).

E. coli DH5α strain was used to propagate plasmids.

Table 1 Bacterial strains and plasmids used in this study

Strains or plasmid	Genotype, relevant phenotype or selection marker	Source or Reference
Strains		
E. coli DH5α	Laboratory strain	Hanahan, 1983
Hi RM118 (Rd)	Laboratory strain	New York
Hi 162	Middle ear isolate	Cody, 2003
Hi 176	Middle ear isolate	Cody, 2003
Hi 285	Middle ear isolate	Cody, 2003
Hi 375	Middle ear isolate	Cody, 2003
Hi 432	Middle ear isolate	Cody, 2003
Hi 477	Middle ear isolate	Cody, 2003
Hi 486	Middle ear isolate	Cody, 2003
Hi 667	Middle ear isolate	Cody, 2003
Hi 723	Middle ear isolate	Cody, 2003
Hi 981	Middle ear isolate	Cody, 2003
Hi 1003	Middle ear isolate	Cody, 2003
Hi 1008	Middle ear isolate	Cody, 2003
Hi 1124	Middle ear isolate	Cody, 2003
Hi 1158	Middle ear isolate	Cody, 2003
Hi 1159	Middle ear isolate	Cody, 2003
Hi 1180	Middle ear isolate	Cody, 2003
Hi 1181	Middle ear isolate	Cody, 2003
Hi 1200	Middle ear isolate	Cody, 2003
Hi 1207	Middle ear isolate	Cody, 2003
Hi 1209	Middle ear isolate	Cody, 2003
Hi 1231	Middle ear isolate	Cody, 2003
Hi 1232	Middle ear isolate	Cody, 2003
Hi 1233	Middle ear isolate	Cody, 2003
Hi 1247	Middle ear isolate	Cody, 2003
Hi 1268	Middle ear isolate	Cody, 2003
Hi 1292	Middle ear isolate	Cody, 2003
Hi 2019	Clinical respiratory isolate	New York
Hi R2846	Middle ear isolate	Seattle
Hi R2866	Invasive NT*Hi* isolated from the blood of a 30-month-old infant	Seattle
375 *HI0227*-	NT*Hi* 375 *HI0227* mutant, kanamycin resistant	this thesis
Rd *HI0227*-	*Hi* Rd *HI0227* mutant, kanamycin resistant	this thesis
R2846 *HI0227*-	NT*Hi* R2846 *HI0227* mutant, kanamycin resistant	this thesis

723 *HI0227*-	NT*Hi* 723 *HI0227* mutant, kanamycin resistant	this thesis
375 *HI0228*-	NT*Hi* 375 *HI0228* mutant, kanamycin resistant	this thesis
Rd *HI0228*-	*Hi* Rd *HI0228* mutant, kanamycin resistant	this thesis
R2846 *HI0228*-	NT*Hi* R2846 *HI0228* mutant, kanamycin resistant	this thesis
723 *HI0228*-	NT*Hi* 723 *HI0228* mutant, kanamycin resistant	this thesis
375 *HI0227/HI0228*-	NT*Hi* 375 *HI0227* and *HI0228* double mutant, kanamycin resistant	this thesis
Rd *HI0227/HI0228*-	*Hi* Rd *HI0227* and *HI0228* double mutant, kanamycin resistant	this thesis
R2846 *HI0227/HI0228*-	NT*Hi* R2846 *HI0227* and *HI0228* double mutant, kanamycin resistant	this thesis
723 *HI0227/HI0228*-	NT*Hi* 723 *HI0227*and *HI0228* double mutant, kanamycin resistant	this thesis
375 *HI0148*-	NT*Hi* 375 *HI0148* mutant, erythromycin resistant	this thesis
981 *HI0148*-	NT*Hi* 981 *HI0148* mutant, erythromycin resistant	this thesis
Rd *HI0148*-	*Hi* Rd *HI0148* mutant, erythromycin resistant	this thesis
R2846 *HI0148*-	NT*Hi* R2846 *HI0148* mutant, erythromycin resistant	this thesis
981 *NTHI0236*-	NT*Hi* 981 *NTHI0236*- mutant, erythromycin resistant	this thesis
Rd *NTHI0236*-	*Hi* Rd *HI0148.1* and *HI0149* mutant, erythromycin resistant	this thesis
R2846 *NTHI0236*-	NT*Hi* R2846 *NTHI0236*- mutant, erythromycin resistant	this thesis
Rd*nagB*	*Hi* Rd *nagB* mutant, erythromycin resistant	this thesis
2019*nagB*	NT*Hi* 2019 *nagB* mutant, erythromycin resistant	this thesis
375*nagB*	NT*Hi* 375 *nagB* mutant, erythromycin resistant	this thesis
Rd *cyaAnagB*	*Hi* Rd *cyaA* and *nagB* double mutant, *cyaA* kanamycin resistant, *nagB* erythromycin resistant	this thesis
Rd *cyaA*	*Hi* Rd *cyaA* mutant, kanamycin resistant	Jenkins
2019 cyaA	*Hi* 2019 *cyaA* mutant, kanamycin resistant	this thesis
2019 cyaAnagB	*Hi* 2019 *cyaA* and *nagB* double mutant, *cyaA* kanamycin resistant, *nagB* erythromycin resistant	this thesis
Plasmids		
pscA	PCR cloning vector	Stratagene
pT7blue	PCR cloning vector	Novagen
puc4kan	*kanC* carrying vector	Pharmacia Biotech
pER2	*eryC* carrying vector	Eileen Rouppe van der Voort
pRdB	pSC-A, Rd *HI0227-HI0228* region	this thesis
p375B	pSC-A, 375 *HI0227-HI0228* region	this thesis
p1231B	pSC-A, 1231 *HI0227-HI0228* region	this thesis
p723B	pSC-A, 723 *HI0227-HI0228* region	this thesis
pRdBK	pSC-A, Rd *HI0227*-, kan resistant	this thesis
p375BK	pSC-A, 375 *HI0227*-, kan resistant	this thesis

p1231BK	pSC-A, 1231 *HI0227*-, kan resistant	this thesis
p723BK	pSC-A, 723 *HI0227*-, kan resistant	this thesis
pRdNK	pSC-A, Rd *HI0228*-, kan resistant	this thesis
p375NK	pSC-A, 375 *HI0228*-, kan resistant	this thesis
p723NK	pSC-A, 723 *HI0228*-, kan resistant	this thesis
p1231PK	pSC-A, 1231 *HI0228*-, kan resistant	this thesis
pRdBNK	pSC-A, Rd *HI0227- and HI0228*-, kan resistant	this thesis
p375BNK	pSC-A, 375 *HI0227- and HI0228*-, kan resistant	this thesis
p1231BNK	pSC-A, 1231 *HI0227- and HI0228*-, kan resistant	this thesis
p723BNK	pSC-A, 723 *HI0227- and HI0228*-, kan resistant	this thesis
pscA1482-2	pSC-A, Rd *HI0148* ORF	this thesis
pscA1482-2ER3	pSC-A, *HI0148*-, ery resistant	this thesis
pscA149	pSC-A, R2846 *NTHI0236* ORF	this thesis
pscA149ER	pSC-A, R2846 *NTHI0236*-, ery resistant	this thesis
p21	pT7Blue, Rd *HI0141- HI0142* region	Jenkins
p31	pT7Blue, Rd *cyaA*-, kan resistant	Jenkins

Note: Table 1 contains bacterial strains used and plasmids constructed during the course of this study. Bacterial strains were stored at -80 °C, and plasmids were stored at -20 °C.

2.1.2 **Enzymes**

Restriction endonucleases were purchased from New England Biolabs and Roche.

Taq DNA polymerase and SuperScript III reverse transcriptase were purchased from Invitrogen.

The manufacturers' instructions regarding enzyme usage were followed accordingly.

2.1.3 **Oligonucleotides**

Oligonucleotides used in this study (see Table 2) for polymerase chain reaction (PCR), quantitative real time PCR (qPCR) and DNA sequencing were obtained from Sigma-Aldrich.

Table 2 Oligonucleotides used in this study

Primer name	Sequence 5' to 3'	ORF	Gene name	Orientation with respect to ORF
0140for	CTGCAATTAAATGGCTGTGG	*HI0140*	*nagA*	Forward
0140forA	ACTAACTTAAAATCAGGTTGTACA	*HI0140*	*nagA*	Forward
0140rev	GCAATTGTGTCATTCGCATC	*HI0140*	*nagA*	Reverse
0140revA	TGCAATAGAGTCAGTCACAA	*HI0140*	*nagA*	Reverse
0141forA	TTATCAAGAACTGATTCGACT	*HI0141*	*nagB*	Forward
0141rev	CAGCAACTGCGCCTTCTA	*HI0141*	*nagB*	Forward
0141revA	TTAGCTCTTGTTGAGATTATAGG	*HI0141*	*nagB*	Reverse
nanAfor	TCCGCCATAATATCGACAAA	*HI0142*	*nanA*	Forward
nanArev	TTTGCTTTTGCAAGCTGTTC	*HI0142*	*nanA*	Reverse
0143for	AATTGCCGATACGATTTTGC	*HI0143*	*siaR*	Forward
0143rev	TATCTTCTTCGCCCTGCACT	*HI0143*	*siaR*	Reverse
0144for	TGCGTTGTTTAGCACTAG	*HI0144*	*nanK*	Forward
0144rev	GCTAATCCCACACTGCCA	*HI0144*	*nanK*	Reverse
0145for	TTGCCAACCTGTCGATGA	*HI0145*	*nanE*	Forward
0145rev	CCCTCAGCCATCACAAAACA	*HI0145*	*nanE*	Reverse
HI0145L	CAAGAAGTGCTTTCTCAAATTCAA	*HI0145*	*nanE*	Forward
0145revA	GTAAGCCAGAACGTCCGC	*HI0145*	*nanE*	Reverse
0146for	TGTTCTTGCCGCTGATTATG	*HI0146*	*siaP*	Forward
0146rev	CATTTTCGGCAGCATCTTTT	*HI0146*	*siaP*	Reverse
0146revA	ACCTAATTGTGAACTTGGATAAAGTGA	*HI0146*	*siaP*	Reverse
0147for	GGAGTGAAGAACTCGCCAAG	*HI0147*	*siaQM*	Forward
147forA	CACAAATCATCACATTTGTGC	*HI0147*	*siaQM*	Forward
0147rev	TCACGCATTGCTTTGATTTC	*HI0147*	*siaQM*	Reverse
0148for	TTTTTCAGCGAACGCACA	*HI0148*	*nanM*	Forward
148forA	AAACAGCATTATGTACTGCAC	*HI0148*	*nanM*	Forward
148forB	TCAAAGCAATAAGCTATGACG	*HI0148*	*nanM*	Forward
148revA	TACCGTCATAGCTTATTGC	*HI0148*	*nanM*	Reverse
0148rev	CAGTTTCACCGCCAATCA	*HI0148*	*nanM*	Reverse
0148.1for	CACCACCACTACTCTGTT	*HI0148.1*	-	Forward
0149rev	GTAACGACGATATTCATC	*HI0149*	-	Reverse
149for2	CGC ATGCA TCCTCCATTAATCCTCGTG	*HI0149*	-	Forward
149rev1	CTT CCACCGCG GTGGAACGCTACTTAAAGCAATC	*HI0149*	-	Reverse
150revA	TTCCGCTTTATGCAGGCAC	*HI0150*	*hflC*	Reverse
227forA	TTAGCAGTTTAACCAACCC	*HI0227*	-	Forward

227forC	CGTTGCTTGACGTGCAATC	*HI0229*	-	Forward
227revA	GGCTTGTGCGGTTCATAAG	*HI0227*	-	Reverse
227revB	CATCTGAAGGAATGGCTTG	*HI0226*	-	Reverse
Atpg up	ATGGCAGGTGCAAAAGAGAT	*HI0480*	*atpG*	Forward
Atpg dwn	TTGTACAACAGGCTTTTGCG	*HI0480*	*atpG*	Reverse
Atpg1	AGGATCGTATGGCTGCATCT	*HI0480*	*atpG*	Reverse
Atpg2	CATCCCACGATCTGTTGAAA	*HI0480*	*atpG*	Forward
16SF	CAAACAGGATTAGATACCCTGG	-	*16s rRNA*	Forward
16SR	GTGACGGGCGGTGTGTAC	-	*16s rRNA*	Reverse
frdB1	TGAACGCGACTTAGTGGTTG	*HI0834*	*frdB*	Forward
frdB2	CTGTGCTGGCGTTTGTTTAG	*HI0834*	*frdB*	Reverse
cyaA	GCTCAATAAGTCAAACCT	*HI0604*	*cyaA*	Forward
cyaB	GCTTCATTACAGACCGAT	*HI0604*	*cyaA*	Reverse
siaC	GAGAGCGGGTTCTAAAGG	*HI1279*	*siaB*	Forward
siaD	ATGGAACGATAAGTTGGC	*HI1279*	*siaB*	Reverse
lic3AA	CAAAAATTACTAACACAAGTTG	*HI0352*	*lic3A*	Forward
Lic3Z	CAAAGTGAATAACCGAGT	*HI0352*	*lic3A*	Reverse

Note: the underlined sequences in primers 149for2 and 149rev1 are restriction sites for NsiI and BstXI, respectively.

2.2 Methods

2.2.1 **Growth media preparation and culture conditions**

Brain heart infusion (BHI) broth was prepared by dissolving 37 g BHI powder (Merck) in 1 litre de-ionised water and autoclaved. The supplements for BHI broth are growth factors X (hemin, 10 μg/ml) and V (nicotinamide adenine dinucleotide, NAD, 2 μg/ml). For BHI plate preparation, 1% agar (w/vol) was added to BHI broth, and the mixture was supplemented with 10% Levinthals base (vol/vol) before pouring (Barbour *et al.*, 1993). Levinthals base contains both growth factors X and V. *Hi* was cultured at 37 °C in supplemented BHI (sBHI) broth with shaking at 180 rpm (Kuhner ShakerX), or on sBHI plates at 37 °C overnight.

Luria-Bertani (LB) broth and plates for *E. coli* culture were prepared from LB powder (Merck) and LB agar (Merck). *E. coli* was cultured in LB broth at 37 °C overnight with shaking at 180 rpm (Kuhner ShakerX), or on LB agar plates at 37 °C overnight.

Where necessary, antibiotics were added to culture, and their concentrations for *Hi* were: 10 μg/ml for kanamycin (kan), and 20 μg/ml for erythromycin (ery). The concentrations of antibiotics for *E. coli* were: 100 μg/ml for ampicillin (amp), 50 μg/ml for kan, and 300 μg/ml for ery.

If required, 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) and isopropylthio-β-D-galactoside (IPTG) were added. The concentration of X-gal for *Hi* and *E. coli* is 40 μg/ml, and the concentration of IPTG for *E. coli* is 40 μg/ml.

Coleman defined medium (CDM) was prepared by dissolving 2.08 g RPMI-1640 powder (Sigma-Aldrich, R1383) in 160 ml de-ionised water, followed by adding: 1.74 ml MEM sodium pyruvate (100 mM, Invitrogen), 1.74 ml β-NAD (1 mg/ml, Sigma, N7004), 3.5 ml haem-histidine stock solution (see below for preparation), 8.8 ml uracil stock solution (2 mg/ml, dissolved in 0.1 N NaOH), 17.4 ml inosine stock solution (20 mg/ml, dissolved in de-ionised water), and 5.34 ml $NaHCO_3$. The pH of RPMI CDM was adjusted to 7.18-7.20 using 1 M HCl. Haem-histidine stock used in CDM preparation was obtained by dissolving 0.2 g L-histidine (Sigma, H8000) in 200 ml de-ionised water, followed by adding 0.2 g hemin HCl (Sigma, H5533) and 4 ml of 1 N NaOH. The solution was heated at 75 ºC for 15 min and cooled at room temperature before filter sterilisation through 0.2 μm pore filter. Haem-histidine stock was stored at 4 °C.

2.2.2 **Preparation of the *Hi* chromosomal DNA**

The *Hi* genomic DNA was prepared using the phenol extraction method. 3 ml late-log culture of bacteria were harvested, washed in PBS (120 mM NaCl, 2.7 mM KCl, 10 mM NaH_2PO_4, 10 mM Na_2HPO_4), and resuspended in 250 μl TNE (100 mM $MgCl_2$, 10 mM Tris pH 8, 100 mM EDTA). Cells were lysed by adding 25 μl of 10% SDS (w/vol), and heated at 65 ºC for 10 min. 8 μl proteinase K (18-25 mg/ml, Roche) was then added and the mixture was incubated at 37 °C for 4-5 hours. DNA was extracted twice by an equal volume of phenol, followed by adding one tenth

volume of 3 M sodium acetate pH 5.4. The DNA was precipitated by adding 2 volumes of ethanol, and spun at 13,000 g for 15 min. DNA pellet was washed in 70% ethanol, and excess alcohol was removed by air or vacuum drying. 200 µl Tris EDTA (TE, 10 mM Tris pH 8, 1 mM EDTA pH 8) containing RNase (20 µg/ml) were added to resuspend DNA.

2.2.3 **Preparation of plasmid DNA**

Plasmid DNA was prepared by the alkaline-lysis method (Sambrook *et al*., 1989). 3 ml late-log culture of cells was spun down at 13,000 g for 5 minutes. The cell pellet was resuspended in 150 µl GTE (50 mM glucose, 10 mM EDTA pH 8, 25 mM Tris pH 8), and the mixture was incubated at room temperature for 5 min. 200 µl freshly prepared solution containing 0.2 M NaOH and 1% SDS was added, and the mixture was kept on ice for 5 min. 150 µl ice-cold potassium acetate (3 M potassium, 5 M acetate) at pH 4.8 was added and the mixture was again placed on ice for 5 min. The suspension was centrifuged at 13,000 g for 15 minutes, and the supernatant containing plasmid DNA was transferred into new tubes. Two volumes of ethanol were added to precipitate DNA. Plasmid DNA was pelleted down by centrifugation at 13,000 g for 15 minutes, followed by 70% ethanol washing. Excess alcohol was removed by air or vacuum drying, and 50 µl TE containing RNase (20 µg/ml) were added to resuspend plasmid DNA.

2.2.4 **Agarose gel electrophoresis and extraction**

One sixth volume of 6 × agarose gel loading buffer (0.25% Bromophenol blue (w/vol), 0.25% xylene cyanol (w/vol), 30% glycerol (vol/vol) in 1 × TE) was added to DNA sample. DNA was separated by electrophoresis on an agarose gel containing ethidium bromide (0.5 μg/ml). Agarose gels were made from TBE buffer (0.89 M Tris, 0.89 M Boric acid, 0.02 M EDTA), and 0.8% agarose (w/vol). The gel was viewed under UV. Where needed, DNA was extracted from agarose gels by using the QIAEX II Agarose Gel Extraction Kit (Qiagen), following the manufacturer's protocol.

2.2.5 **Polymerase chain reaction (PCR)**

Each PCR reaction contains a total volume of 50 μl, and consists of 42.5 μl de-ionised sterile water, 5 μl of 10 × PCR buffer with $MgCl_2$ (500 mM KCl, 200 mM Tris-HCl pH 8.4, 25 mM $MgCl_2$), 0.8 μl dNTPs (100 mM), 0.5 μl forward primer (20 μM), 0.5 μl reverse primer (20 μM), 0.2 μl Taq DNA polymerase (5 U/μl, Invitrogen), and 0.5-4 μl DNA template (~10 ng chromosomal DNA and ~15 to 20 ng cDNA). Unless otherwise specified, PCR reactions were performed at the following condition: 1 min denaturation (94 °C), 1 min annealing (50 °C), and 1 min polymerisation (72 °C) for 30 cycles in Mastercycler gradient PCR machine (Eppendorf). PCR products were electrophoresed on a 0.8% agarose gel containing ethidium bromide and viewed under UV.

An extended PCR reaction using Bio-X-Act Long DNA polymerase (Bioline) was used for the creation of the *NTHI0236*- mutants. Again, total volume of each PCR reaction was 50 µl, consisting of the following components: 5 µl 10 × OptiBuffer (Bioline), 1 µl $MgCl_2$ (50 mM), 0.5 µl dNTPs (100 mM), 2.5 µl DMSO, 0.3 µl primer mix (100 µM), 1 µl Bio-X-Act Long polymerase (4 U/µl), 1 µl DNA template (50 ng/µl), and 38.4 µl sterile water. The extended PCR reactions were performed under the following condition: 30 sec denaturation (94 °C), 30 sec annealing (60 °C), and 10 min polymerisation (68 °C) for 30 cycles.

2.2.6 ***Hi* transformation**

Transformation of *Hi* was mostly carried out using the MIV method, and performed as follows. *Hi* strains were cultured on sBHI plate overnight, and 1 µl loopful of bacteria were transferred to sBHI broth. Bacteria were cultured until the OD_{650} reached 0.2-0.4. Cells were pelleted down at 3000 g in a bench top centrifuge for 10 min at room temperature and gently resuspended in an equal volume of warm (37 °C) minimum medium IV (MIV) solution (Herriott *et al.*, 1970). MIV starves the bacteria and increases bacterial competency. Cells were re-pelleted, and resuspended in ¼ volume of warm MIV solution, followed by incubation at 37 °C with slow shaking (90 rpm) in an orbital shaker (Kuhner ShakerX) for 100 min. 3-4 µg of linearised DNA (5 µl plasmid DNA linearised by restriction endonuclease in a total volume of 15 µl) was added to 250 µl competent cells, and the mixture was incubated at 37 °C for 30 min with shaking at 180 rpm. 1

ml of warm BHI broth was added and samples were incubated for 30 min at 37 °C with shaking (180 rpm). Cells were pelleted in a microfuge for 1 min, and resuspended in 100 µl of the supernatant. Bacteria were plated out onto appropriate antibiotic sBHI plates, and cultured overnight. Transformants that grew up were re-plated onto new antibiotic sBHI plates, and cultured overnight. Chromosomal DNA was extracted, and correct transformants were confirmed by PCR analysis. Digestion of mutants' chromosomal DNA by restriction endonuclease (usually MfeI) was performed to confirm the digestion pattern corresponded to that of the parent strain.

Mutants that were not successfully created by the MIV procedure were then attempted via the static aerobic transformation procedure. *Hi* was cultured on sBHI plates overnight, and a 1 µl loopful of bacteria was inoculated into 10 ml sBHI broth. Bacteria were transferred to a petri dish, and incubated without shaking at 37 °C for 18 hours. Cells were gently resuspended, and 100 µl competent cells together with 0.5 µg DNA (linearised plasmid DNA, chromosomal DNA, or PCR product) were added to 500 µl sBHI broth. Samples were incubated at 37 °C with shaking (180 rpm) for 5 hours, pelleted down and plated onto appropriate antibiotic plates at 37 °C overnight.

2.2.7 ***E. coli*** **transformation**

E. coli DH5α cells were cultured in LB broth overnight. 200 µl of culture was transferred to 20 ml of fresh LB broth and grown until the OD_{600} was approximately 0.4. Cells were placed into

pre-cooled tubes and on ice for 10-30 min, before being pelleted down at 3000 g for 10 min in a pre-cooled 4 ºC bench top centrifuge. Cells were gently resuspended in an equal volume of ice-cold 0.1 M $MgCl_2$, and kept on ice for 10 min. Cells were then pelleted again as described above, and then resuspended in an equal volume of 0.1 M $CaCl_2$ and kept on ice for 10 min. Again, cells were pelleted, but resuspended in 1/20 volume of 0.1 M $CaCl_2$, and kept on ice for at least 30 min. DH5α cells were left on ice for at least one hour before use. The competency of cells will increase over a period up to 24 hours. 10 µl ligation mixture was added to 200 µl competent cells and the mixture was kept on ice for at least 40 min, before being heat shocked at 42 ºC for 2 min. Cells were placed back on ice for 2 min, and 1 ml pre-warmed LB broth was added. The mixture was incubated at 37 ºC for 30-60 min to allow expression of the antibiotic resistance phenotype. Appropriate dilutions of cells were performed and plated out onto antibiotic LB plates and cultured overnight.

2.2.8 **Tricine LPS gel**

Hi was cultured on sBHI plates supplemented with Neu5Ac (25 µg/ml) overnight and 1 µl loopful of cells were resuspended in 1 ml PBS at pH 5.8. A bacterial lysate was prepared, and this was done by adding 20 µl suspended cells to a 980 µl solution containing 1% SDS and 0.1 M NaOH. The optical density of bacterial lysate was read at 260 nm. The starting cell suspension was diluted to an optical density of 1 at 260 nm by adding PBS (pH 5.8). PBS at pH 5.8 is the

optimal buffer for *Clostridium perfringens* purified neuraminidase (Boehringer Mannheim), an enzyme that cleaves bound Neu5Ac at the α(2-3), α(2-6), α(2-8) linkages from its attached oligosaccharide.

For each sample, two aliquots of 25 µl diluted bacterial cells (at OD_{260} = 1) were incubated overnight at 37 ºC. One aliquot is treated with 2.5 µl neuraminidase (4 U/µl), while the other contains 2.5 µl PBS at pH 5.8. An equal volume of 2 × dissociation buffer (2.5 ml 1 M Tris pH 6.8, 4 ml glycerol, 0.78 g SDS, 2 ml mercaptoethanol, 0.8 ml 1% bromophenol blue, 10.7 ml de-ionised water) was added to each sample. The samples were then boiled for 5 min, and 25 µl of samples loaded into the well of the tricine LPS gel (Lesse *et al.*, 1990).

The tricine LPS gel, previously described by Lesse and colleagues (Lesse et al., 1990), is composed of a separating component and a stacking component. 15 ml is taken from the mixture (16.6 ml acrylamide, 10 ml gel buffer, 3.12 ml glycerol, 0.3 ml water, 6 µl TEMED and 60 µl 10% APS) to make the separating component, and 6 ml is taken from the mixture (1.7 ml acrylamide, 3.1 ml gel buffer, 7.7 ml water, 15 µl TEMED, 150 µl APS) to make the stacking component. The gel buffer consists of 3 M Tris, 0.3% SDS and pH adjusted to 8.45 with HCl. Anode and cathode buffers were used in running the gels. The anode buffer consisted of 0.2 M Tris and pH was adjusted to 8.9 with HCl, the cathode buffer consisted of 0.1 M Tris, 0.1 M Tricine, 0.1% SDS (vol/vol) and pH adjusted to 8.25 with NaOH. LPS gels were electrophoresed

at 20 mA for 16-17 hours at 4 °C, and LPS bands were visualised by staining with silver (GE Healthcare), following the manufacturer's instructions.

2.2.9 **Serum bactericidal assay**

The serum bactericidal assay (SBA) was used to determine the resistance of bacteria against the killing effect of pooled normal human sera, and is a technique previously described by Hood and colleagues (Hood et al., 1999a).

Hi were grown on Neu5Ac (25 μg/ml) sBHI plates at 37 ºC overnight. A small loop of bacteria were suspended in 750 μl PBS-BG (PBS containing 0.05 mM $MgCl_2$, 0.09 mM $CaCl_2$ and 0.1% glucose). A bacterial lysate was prepared, and this was achieved by adding 20 μl suspended cells to a 980 μl solution containing 1% SDS and 0.1 M NaOH. The optical density of the lysate was obtained at 260 nm. The initial bacteria suspension was diluted to an absorbance of 0.8 at 260 nm by PBS-BG, which is equivalent to ~10^9 organisms per ml. Bacterial suspension was further adjusted by serial dilution to yield ~10^3 cells in a 50 μl volume.

A 20% solution (vol/vol) of pooled normal human sera in PBS-BG was serially diluted, and used at concentration of 10%, 5%, 2.5%, 1.25%, 0.63%, 0.31%, 0.16% across a 96-well tissue culture plate in a volume of 50 μl per well. The 50 μl diluted bacteria (~10^3 cells) were added to 50 μl pooled human sera, and samples were incubated in a 96-well tissue culture plate at 37 ºC for one hour. A 10% decomplemented serum control was included. The serum was decomplemented by

heating at 56 ºC for 30 minutes. Survival of bacteria at each serum concentration was determined by plating out 25 µl of cells from each well onto sBHI plates and culturing at 37 ºC overnight. Bacterial colonies from overnight growth were counted to determine the level of bacterial survival at each serum concentration. The inoculating dose of bacteria was determined by plating out 25 µl diluted bacterial suspension onto sBHI plates and culturing at 37 ºC overnight. Bacterial survival was plotted as a percentage against the decomplemented serum control. The experiment was repeated and results were combined and averaged. Standard deviations were calculated using Microsoft Excel.

2.2.10 **RNA extraction, reverse transcriptase PCR, and quantitative PCR**

Hi strains were cultured on sBHI plate overnight. 1 colony of bacteria was transferred to sBHI broth and cultured overnight. The optical density of the culture at 600 nm was obtained. 1 ml of bacteria was pelleted down at 900 g in a microfuge for 10 min and resuspended in 1 ml fresh, pre-warmed CDM. The appropriate amount of resuspended bacteria was added to pre-warmed CDM to give an optical density of ~0.05 at 600 nm.

When appropriate, glucose, Neu5Ac, or cAMP was added to a final concentration of 2 mg/ml (for glucose), 30 or 500 µg/ml (for Neu5Ac), or 1 mM (for cAMP). Depending on the growth conditions, glucose, Neu5Ac and cAMP were either added at the start of bacterial cultures (when

optical density at 600 nm is ~0.05) or when bacteria reached mid-log growth phase (optical density at 600 nm is ~0.3).

Bacteria were grown to mid-log phase, and 1 ml of bacterial culture was added directly to 2 ml RNA Protect Bacterial Reagent (Qiagen) in a 5 ml bijoux. Using the SV Total RNA Isolation Kit (Promega) following the manufacturer's protocol, 100 µl bacterial RNA was extracted, and the same amount was treated with DNaseI using the Turbo DNA-free kit (Ambion) following the manufacturer's protocol. RNA was then quantified by Nanodrop (NanoDrop Spectrophotometer ND-1000). 2 µl DNase-treated RNA was used as template in PCR amplification to ensure that RNA samples are DNA free. RNA was then stored at -80 °C or reverse transcribed to make cDNA.

Approximately 300 ng RNA was used in each reverse transcription reaction. 2 µl random primers (500 µg/ml, Promega), and 2 µl dNTP (10 mM, Invitrogen) were added to the reaction and the mixture was made up to 26 µl with DEPC treated water. The mixture was then incubated at 65 °C for 5 min, placed on ice for at least 1 min, and then spun down briefly before adding 8 µl of 5 × first strand buffer (Invitrogen), 2 µl of DTT (0.1 M, Invitrogen), 2 µl of RNase out (40 U/µl, Invitrogen), and 2 µl Superscript III reverse transcriptase (200 U/µl, Invitrogen). The mixture was incubated at room temperature for 5 min, followed by 50 °C for ~45 min to allow for the synthesis of the first cDNA strand. The reaction mixture was then incubated at 70 °C for 15 min to inactivate the enzyme. The product is non-amplified cDNA. In order to generate the

complementary DNA strand and then amplify the cDNA template, 2 µl cDNA was used in a PCR reaction (see section 2.2.5). PCR reactions were carried out in a Mastercycler gradient PCR machine (Eppendorf). For reverse transcription PCR (RT-PCR), gel images were captured by ImageMaster VDS (Pharmacia Biotech) and the intensity of agarose gel bands was analysed using ImageQuant TL programme (GE Healthcare). This programme can assess the intensity of a band as a percentage against another selected band. RT-PCR was used as a semi-quantitative, screening method to determine the abundance of RNA molecules as a measure of gene expression.

Real-time quantitative PCR (qPCR) was performed in a DNA Engine Opticon 2 (MJ Research). Each q-PCR reaction contains 2 µl non-amplified cDNA, 0.5 µl forward primer (10 µM), 0.5 µl reverse primer (10 µM), 10 µl SYBR Green master mix (Finnzymes), and 7 µl sterile deionised water. The qPCR reactions were set up in a laminar vent hood.

Conditions for q-PCR were 15 min initial denaturation (95 °C) (a step to ensure complete reactivation of the hot-start DNA polymerase and denaturation of the template), followed by 35 cycles of 10 sec denaturation (94 °C), 30 sec annealing (55 °C), 30 sec polymerization (72 °C), and fluorescence data acquisition.

The melting curve (see below) was obtained to check the specificity of the amplified product. When the temperature is gradually increased, a sharp decrease in SYBR Green fluorescence is observed as the product undergoes denaturation. Specific products can be distinguished from

non-specific products by the difference in their melting temperatures. The condition used to obtain the melting curve was from 72 °C to 95 °C, with fluorescence detection at every 0.2 °C for 1 sec.

qPCR results were obtained from two or three biological replicates, two cDNA samples from each biological replicate, and two cDNA replicates. Analyses of q-PCR data were performed using the programme Opticon Monitor 3.1 (Bio-rad).

The threshold for detection of DNA-based fluorescence was set at 0.04. The number of cycles at which the fluorescence exceeded this threshold was the cycle threshold (Ct). Titration experiment using serial dilution of DNA template was performed to obtain a standard curve, and the slope of the standard curve was used to calculate the efficiency of a primer-template combination, using the equation $E=10^{[-1/slope]}$ (Bustin, 2004, Higuchi *et al.*, 1993). The correlation coefficient of a standard curve (automatically calculated by the programme Opticon Monitor 3.1) is a measure of how well Ct values from titration experiment fit the standard curve, which reflects the reliability of the standard curve. A correlation coefficient value above 0.990 was accepted.

Relative fold change of a target gene with respect to a reference gene was calculated using the equation: $Nx/Nc = (1+Ec)^{CTc} / (1+Ex)^{CTx}$, where Nx/Nc indicates the relative fold change of a target gene with respect to a reference gene, Ec indicates the primer efficiency of the reference

gene, Ex indicates the primer efficiency of the target gene, CTc indicates the Ct number of the reference gene, and CTx indicates the Ct number of the target gene (Applied-Biosystems, 1997). Results obtained were plotted in Microsoft Excel, and standard deviation calculated using the same programme.

Two reference genes were used for qPCR: *frdB* and *atpG*, which respectively encode a fumarate reductase iron-sulfur protein (involved in the tricarboxylic acid cycle), and an ATP synthase gamma chain (involved in ATP biosynthesis). *frdB* has been previously used as a reference gene in the assessment of the *Hi* sialometabolism gene expression (Jenkins et al., 2010). To quantify the *nanE* transcript in the *nanE* mutant background, we designed and used a qPCR primer pair (HI0145L and 0145revA) that binds to the DNA sequence proximal to the antibiotic resistance cassette insertion site.

Calibration between plates (also known as inter-plate calibration, IPC) was performed when experiments were done on different plates. Ct values of reference genes were used for IPC. The Ct values were averaged, and the difference between the averaged value and the Ct values from each of the plate was obtained. Ct values of target genes from each of the plates were adjusted accordingly.

The programme REST (relative expression software tool) 2009 (Qiagen), which compares treated and untreated groups, was used where necessary to help determine the fold change in expression of a target gene between treated and untreated groups. The mathematical model used

by REST is based on PCR efficiencies and Ct values between treated and untreated groups (the ΔΔCt method). The ΔΔCt method can be summarised by the following equations:

ΔCt of treated group = Ct of target gene from treated group – Ct of reference gene from treated group

ΔCt of untreated group = Ct of target gene from untreated group – Ct of reference gene from untreated group

ΔΔCt =ΔCt of treated group -ΔCt of untreated group

2.2.11 **DNA and plasmid sequencing**

DNA and plasmid sequences described in this thesis were obtained from the Weatherall Institute of Molecular Medicine Sequencing Service, John Radcliffe Hospital, Oxford. An ABI-3730 DNA analyser with BigDye Terminator v3.1 chemistry (Applied Biosystems) was used, and electropherograms were viewed using Chromas 1.62 software (Technelysium Pty Ltd, Queensland).

2.2.12 **Growth assay of *Hi* in BHI and CDM**

Hi was cultured on a sBHI plate at 37 °C overnight. A single colony was picked, transferred to sBHI broth, and cultured for 12-15 hours at 37 °C. Optical density of bacteria was obtained at 600 nm (Corning spectrophotometer 258) and bacteria were diluted to ~0.05 at OD_{600} using

pre-warmed BHI broth or glucose free CDM. Occasionally, an alternative wave length at 650 nm was used. Where appropriate, glucose (2 mg/ml) or Neu5Ac (0.5 mg/ml) was added. Bacteria were grown at 37 °C and optical density was obtained at appropriate time intervals (every 30 min, hourly, or two-hourly) and plotted accordingly. Optical density values at the exponential phase of growth (approximately OD_{650} 0.05 to 0.4) were used to calculate the doubling time of bacteria and this was done by logarithmic conversion of the values using Microsoft Excel. Where appropriate, experiments were repeated and results were combined and averaged. Standard deviation was calculated using Microsoft excel.

2.2.13 **Fermentation Assay**

Hi was cultured on a sBHI plate at 37 °C overnight. A single colony was picked, transferred to sBHI broth, and incubated at 37 °C for 12-15 hours. Optical density of bacteria was obtained at 600 nm (Corning spectrophotometer 258). 150 µl bacteria were added to 6 ml prewarmed glucose-free CDM and the optical density was obtained at 600 nm. 1 ml of diluted bacteria was placed in each well of a 48 well culture plate (Corning), and different sugars (glucose 0.05%, fructose 0.05%, Neu5Ac 0.05%) were added to different sets of wells. Photographs were taken by camera (Panasonic DMC-FX33) prior to incubation at 37 °C. For micro-aerophilic growth condition, culture plates were placed inside an anaerobic jar and oxygen concentration was

adjusted to 6% by Anoxomat. Photographs were taken hourly, and the optical density of bacteria at 600 nm was obtained after 8 hours of incubation.

2.2.14 **Mutation of loci and transformation of *Hi***

Construction of the *HI0227*- mutants (Fig. 9 and Fig. 10)

The *HI0227-HI0228* region of *Hi* strains Rd, 375, 1231 and 723 was amplified by standard PCR method (see section 2.2.5) using primers 227forC and 227revB (see Fig. 9 and Table 2). The PCR product (1495 bp in *Hi* strains Rd and 375; 1696 bp and 1900 bp in *Hi* strains 1231 and 723 respectively) (Fig. 9) was cloned into plasmid pSC-A (Stratagene, 240205) and transformed into StrataClone SoloPack *E. coli* competent cells (Stratagene, 240205), following the manufacturer's protocol. Correct transformants (white colonies) were selected using blue/white screening on LB agar plate supplemented with 100 μg/ml amp with 20 μl 4% X-Gal spread on top. White colonies indicated that the *lacZ* gene on the plasmid was disrupted by an insert, due to the lack of X-Gal hydrolysis. Single white colony was picked, transferred into LB broth containing 100 μg/ml amp and grew for 16 hours at 37 ºC with shaking, 3 ml of the culture was used to extract plasmid DNA (see section 2.2.3). The correct plasmids were verified using appropriate restriction digest for the presence and orientation of the PCR insert.

Vectors containing the PCR amplified *HI0227-HI0228* region from *Hi* strains Rd, 375, 1231 and 723 were respectively named pRdB, p375B, p1231B, p723B, and were digested with BsrGI,

which linearised the plasmids by cutting once at position 205 bp downstream from the start codon of *HI0227* and produced 5´ DNA overhangs. DNA overhangs of cut vectors were blunted using the Quick Blunting Kit (Biolabs), which contains T4 DNA polymerase with both 3´→ 5´ exonuclease and 5´→ 3´ polymerase activities.

The *kanC* isolated by digestion of puc4kan with HincII (produced a 1252 bp fragment) was ligated to cut vectors using the Quick Ligation Kit (Biolabs), following the manufacturer's instruction. The ligation product was transformed into *E. coli* DH5α competent cells (see section 2.2.7) and transformants were selected on LB plate containing kan from overnight culture at 37 ºC. Colonies were picked and plasmids were extracted. The correct constructs that contained the *HI0227* ORF and *kanC* in the same orientation (confirmed by PCR amplification and restriction analyses) were named pRdBK, p375BK, p1231BK, and p723BK, and were respectively used to transform *Hi* strains Rd, 375, R2846 and 723 by the MIV method (see section 2.2.6). The *kanC* was inserted in the same orientation as the ORF to ensure that the gene was properly inactivated and that *kanC* can be fully expressed.

Hi mutants that carried inactivated HI0227 were named Rd *HI0227*-A to -B; 375 *HI0227*-A to -E; R2846 *HI0227*-A to -I and 723 *HI0227*-A to -F respectively.

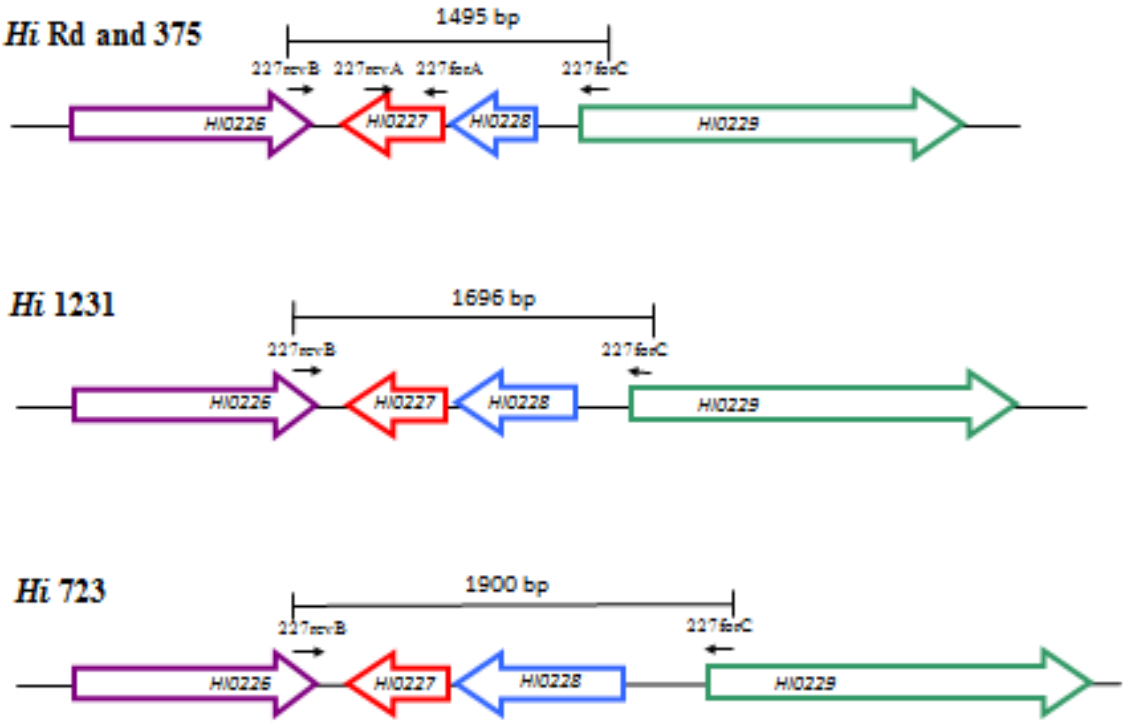


Fig. 9 DNA fragment sizes of PCR product using primers 227revB and 227forC of *Hi* strains Rd, 375, 1231 and 723.

The directions of the arrows represent the transcriptional directions of the ORFs. The purple, red, blue and green arrows respectively represent the *HI0226*, *HI0227*, *HI0228* and the *HI0229* ORFs. The small arrows shown above the *HI0226* and *HI0229* ORFs represent the 227revB and 227forC primers. DNA fragment sizes of PCR product amplified by using the 227revB and 227forC primers are stated above the primers. Locations of additional primers are shown in the *Hi* Rd and 375 panel.

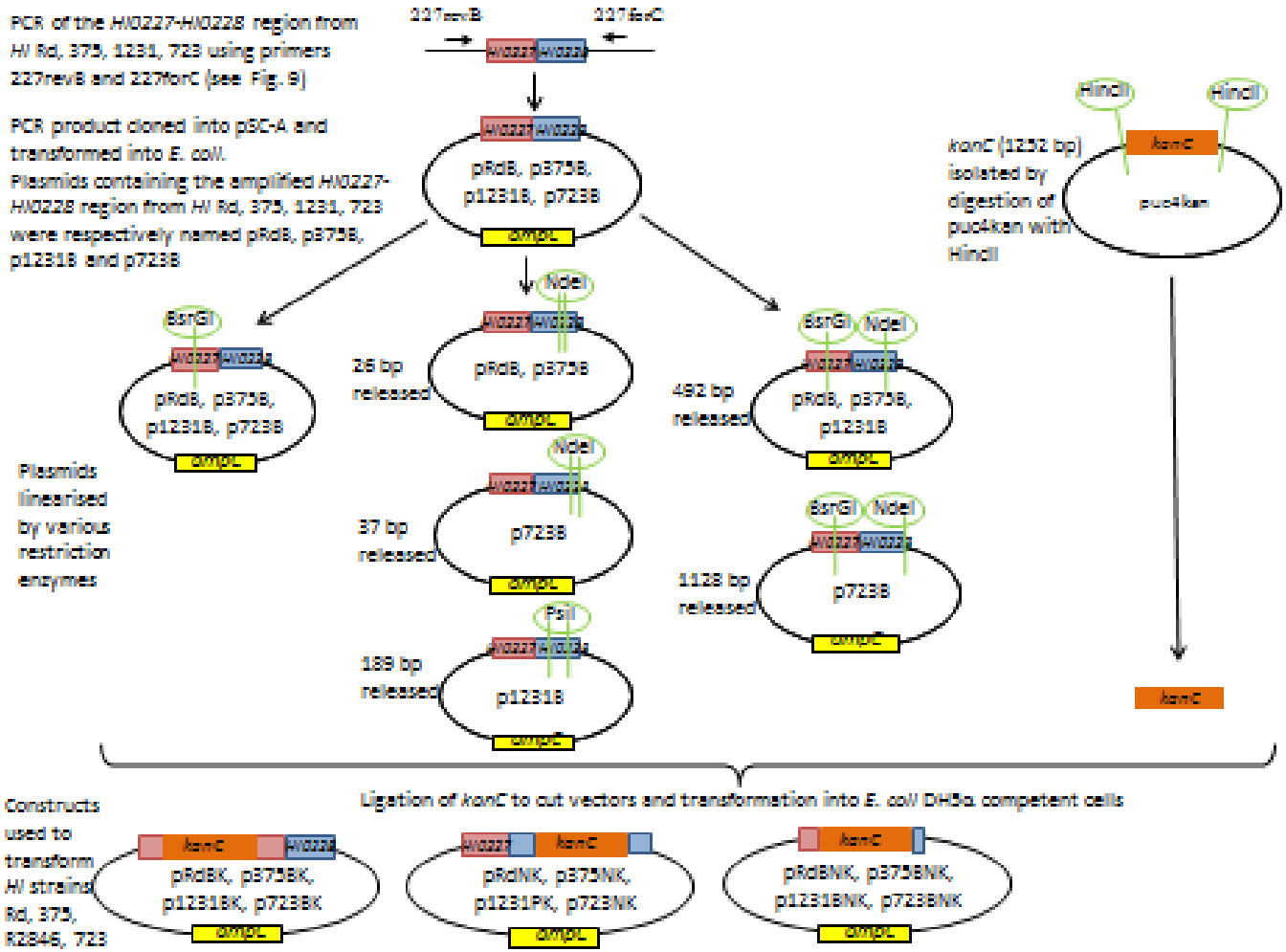


Fig. 10 Diagram showing the method used to construct the plasmids that carried inactivated *HI0227*, *HI0228*, and combined *HI0227/HI0228*.

Plasmids are represented by black circles. Name of the plasmids are stated within the circles. Restriction enzymes are circled in green and the green linear lines represent the sites at which the restriction enzymes cut. The red and blue boxes respectively represent the *HI0227* and *HI0228* ORFs. The orange and yellow boxes respectively represent the *kanC* and the ampicillin resistance cassette (*ampC*).

Construction of the *HI0228*- mutants (Fig. 9 and Fig. 10)

The same plasmids containing the amplified *HI0227-HI0228* region from *Hi* strains Rd, 375, 1231 and 723 (respectively named pRdB, p375B, p1231B and p723B, see 'Construction of the *HI0227*- mutants', stated above) were used (see Fig. 9 and Fig. 10).

Vectors pRdB, p375B and p723B were digested with NdeI, which released a 26 bp fragment from the cloned region of pRdB and p375B (NdeI cuts twice at positions 164 bp and 190 bp downstream from the start codon of the *Hi* Rd and 375 *HI0228* ORF), and a 37 bp fragment from the cloned region of p723B (NdeI cuts twice at positions 2 bp upstream and 35 bp downstream from the start codon of the *Hi* 723 *HI0228* ORF). 5′ DNA overhangs were produced from digestion with NdeI, and DNA overhangs of cut vectors were blunted using the Quick Blunting Kit (Biolabs), which contains T4 DNA polymerase with both 3′→ 5′ exonuclease and 5′→ 3′ polymerase activities.

The *kanC* isolated by digestion of puc4kan with HincII (produced a 1252 bp fragment) was ligated to cut vectors using T4 DNA Ligase (Biolabs) at 16°C overnight. The ligation product was transformed into *E. coli* DH5α competent cells (see section 2.2.7) and transformants were selected on LB plate containing kan from overnight culture at 37 ºC. Correct constructs that contained the *HI0228* ORF and the *kanC* in the same orientation (confirmed by PCR amplification and restriction analyses) were named p375NK, pRdNK, p723NK, which were respectively used to transform *Hi* strains 375, Rd and 723 by the MIV method.

Plasmid p1231B (the vector that contained the amplified *HI0227-HI0228* region from *Hi* strain 1231, see above) was digested with restriction enzyme PsiI (an endonuclease that produces blunt ends) to linearise the vector. PsiI cuts twice at positions 103 bp and 292 bp downstream from the start codon of the *Hi* 1231 *HI0228* ORF and released a 189 bp fragment.

The *kanC* isolated by digestion of puc4kan with HincII was ligated to cut p1231B using T4 DNA Ligase (Biolabs) at 16°C overnight, and the ligation product was transformed into *E. coli* DH5α

competent cells (see section 2.2.7). Transformants were selected on LB plate containing kan from overnight culture at 37 ºC, and the correct construct that contained the *Hi* 1231 *HI0228* ORF and the *kanC* in the same orientation (confirmed by PCR amplification and restriction analyses) was named p1231PK, which was used to transform *Hi* strain R2846 by the MIV method.

Hi mutants that carried inactivated HI0228 were named 375 *HI0228*-A to -D; Rd *HI0228*-A, -B; R2846 *HI0228*-A and 723 *HI0228*-A to -D.

Construction of the *HI0227/HI0228*- double mutants (Fig. 9 and Fig. 10)

The same vectors containing the amplified *HI0227-HI0228* region from *Hi* strains Rd, 375, 1231 and 723 (respectively named pRdB, p375B, p1231B and p723B, see 'Construction of the *HI0227*- mutants', stated above) were used (see Fig. 9 and Fig. 10).

Plasmids pRdB, p375B, p1231B and p723B were digested with BsrGI and NdeI, which released a 492 bp fragment from the cloned region in all cases except for p723B, which was 1128 bp.

5′ DNA overhangs were produced from digestion with BsrGI and NdeI, and cut vectors were blunted using the Quick Blunting Kit (Biolabs), which contains T4 DNA polymerase with both 3′→ 5′ exonuclease and 5′→ 3′ polymerase activities.

The *kanC* isolated by digestion of puc4kan with HincII (produced a 1252 bp fragment) was ligated to cut vectors using T4 DNA Ligase (Biolabs) at 16°C overnight. The ligation product was transformed into *E. coli* DH5α competent cells (see section 2.2.7) and transformants were selected on LB plate containing kan from overnight culture at 37 ºC. Correct constructs that contained the *HI0227-HI0228* region and *kanC* in the same orientation (confirmed by PCR amplification and restriction analyses) were named p375BNK, pRdBNK, p1231BNK, and p723BNK, which were respectively used to transform *Hi* strains 375, Rd, R2846 and 723 by the MIV method.

Hi mutants that carried inactivated HI0227 and HI0228 were named 375 *HI0227/HI0228*-A to -E; Rd *HI0227/HI0228*-A to -F; R2846 *HI0227/HI0228*-A, -B and 723 *HI0227/HI0228*-A to -C.

Construction of the *HI0148*- mutants (Fig. 11)

The *HI0148* ORF from *Hi* Rd was amplified by standard PCR method (see section 2.2.5) using primers 147forA and 148revA (see Table 2, Fig. 16). The PCR product (1323 bp in length) was cloned into plasmid pSC-A (Stratagene, 240205) and transformed into StrataClone SoloPack *E. coli* competent cells (Stratagene, 240205), following the manufacturer's protocol. Correct transformants (white colonies) were selected using blue/white screening on LB agar plate supplemented with 100 μg/ml amp with 20 μl 4% X-Gal spread on top (see 'Construction of the *HI0227*- mutants', stated above). Single white colony was picked, transferred into LB broth containing 100 μg/ml amp and grew for 16 hours at 37 ºC with shaking, 3 ml of the culture was used to extract plasmid DNA (see section 2.2.3). The correct plasmids were verified using appropriate restriction digest for the presence and orientation of the PCR insert.

The vector that contained the PCR amplified *HI0148* ORF from *Hi* Rd was named pscA1482-2, and was digested with BstAPI, which released a 113 bp fragment from the cloned region (BstAPI cuts twice at positions 65 bp upstream and 48 bp downstream from the start codon of the Rd *HI0148* ORF). 3´ DNA overhangs were produced from digestion with BstAPI, and DNA overhangs of cut vector were blunted using T4 DNA polymerase (New England Biolabs).

The *eryC* isolated by digestion of pER2 with PvuII (produced a 1596 bp fragment) was ligated to cut vector using T4 DNA Ligase (Biolabs). The ligation product was transformed into *E. coli* DH5α competent cells (see section 2.2.7) and transformants were selected on LB plate containing ery from overnight culture at 37 ºC. Colonies were picked and plasmids were extracted. The correct construct that contained the *HI0148* ORF and *eryC* in the same orientation (confirmed by PCR amplification and restriction analyses) was named pscA1482-2ER3, and was

used to transform *Hi* strains Rd, 375, 981 and R2846 by the MIV method (see section 2.2.6). The *eryC* was inserted in the same orientation as *HI0148* to ensure that the gene is properly inactivated and that *eryC* can be fully expressed. *Hi* mutants that carried inactivated HI0148 were named Rd *HI0148*-A to -F, 375 *HI0148*-A, 981 *HI0148*-A to -F, and R2846 *HI0148*-A to -D.

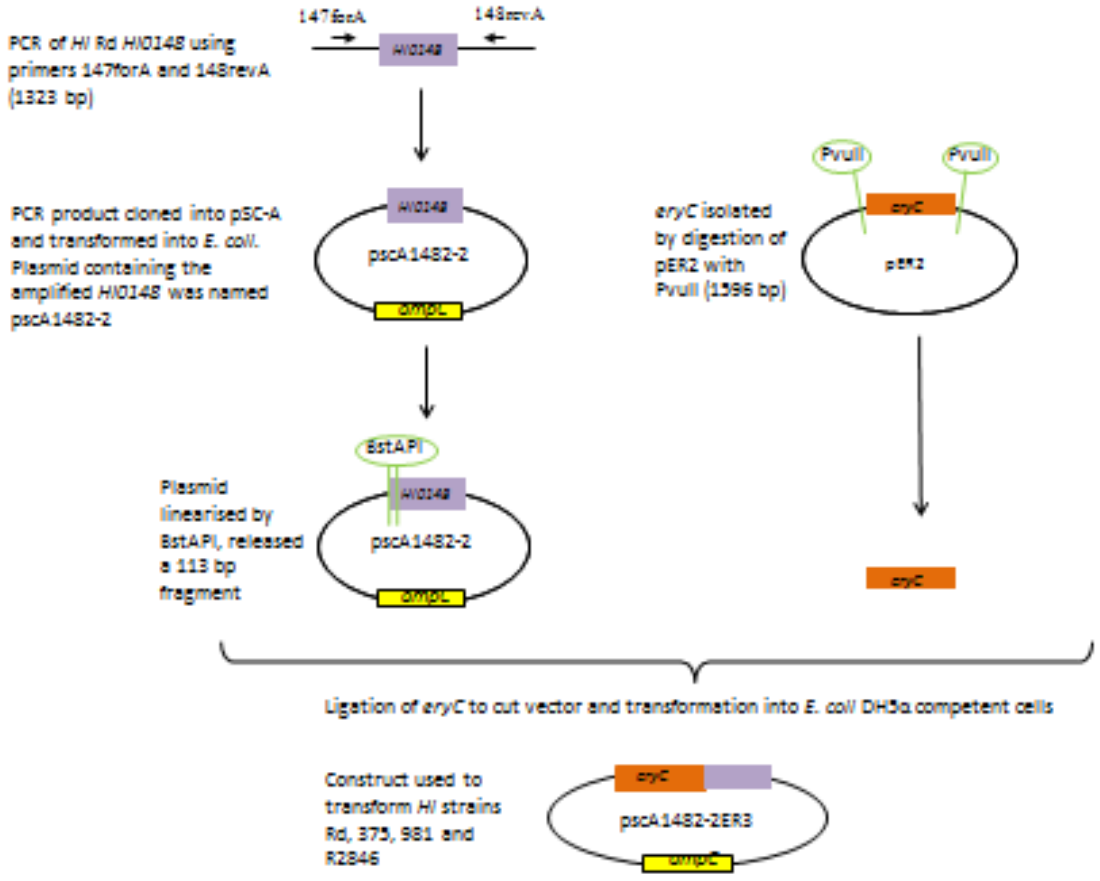


Fig. 11 Diagram showing the method used to construct the plasmid that carried inactivated *HI0148*.

Plasmids are represented by black circles. Name of the plasmids are stated within the circles. Restriction enzymes are circled in green and the green linear lines represents the sites at which the restriction enzymes cut. The purple box represents the *HI0148* ORF. The orange and yellow boxes respectively represent the *eryC* and the *ampC*.

Construction of the *NTHI0236*- mutants (Fig. 12)

The construction method used to create the *NTHI0236*- inactivated mutant is illustrated in Fig. 12. This method was used because the *NTHI0236* ORF lacked useful restriction sites for cloning. The *NTHI0236* ORF from *Hi* strain R2846 (denoted by the grey box in Fig. 12) corresponds to *HI0148.1* and *HI0149* in stain Rd. The *NTHI0236* ORF from *Hi* strain R2846 was amplified by PCR using primers 148forB and 150revA (see Table 2). The PCR product (710 bp in length) was cloned into pSC-A (Stratagene, 240205) and transformed into StrataClone SoloPack *E. coli* competent cells (Stratagene, 240205), following the manufacturer's protocol. Correct transformants (white colonies) were selected using blue/white screening on LB agar plate supplemented with 100 μg/ml amp with 20 μl 4% X-Gal spread on top (see 'Construction of the *HI0227*- mutants', stated above). Single white colony was picked, transferred into LB broth containing 100 μg/ml amp and grew for 16 hours at 37 ºC with shaking, 3 ml of the culture was used to extract plasmid DNA (see section 2.2.3). The correct plasmids were verified using appropriate restriction digest for the presence and orientation of the PCR insert. The vector that contained the PCR amplified *NTHI0236* ORF was named pscA149.

Primers 149for2 and 149rev1 (see Table 2) were designed to allow for the ligation of pscA149 to antibiotic resistance cassette, and this was carried out using inverse PCR reaction (see 'extended PCR reaction' in section 2.2.5). The 149for2 primer sequence consisted of (from 5' to 3'): 3 random nucleotides (CGC), the NsiI restriction site, and the sequence that corresponds to 166 bp to 183 bp downstream from the start codon of the *NTHI0236* ORF; and the 149rev1 primer sequence consisted of (from 5' to 3'): 3 random nucleotides (CTT), the BstXI restriction site, and the sequence that corresponds to 143 bp to 125 bp downstream from the start codon of the *NTHI0236* ORF.

The PCR product produced from the inverse PCR reaction (5000 bp) was digested with BstXI and NsiI. The resulting fragment was then ligated to a 1164 bp fragment containing *eryC* using

T4 DNA Ligase (Biolabs). *eryC* was isolated by digestion of pER2 with BstXI and NsiI. The ligation product was transformed into *E. coli* DH5α competent cells (see section 2.2.7) and transformants were selected on LB plate containing ery from overnight culture at 37 ºC. Colonies were picked and plasmids were extracted. The correct construct that contained the *NTHI0236* ORF and *eryC* in the same orientation (confirmed by PCR amplification and restriction analyses) was named pscA149ER, which was used to transform *Hi* strains 375, Rd and R2846 by the MIV method. *Hi* mutants that carried inactivated NTHI0236 were named Rd *HI0149*-A to -E, and R2846 *HI0149*-A to -F. Strain 375 could not be transformed upon this occasion.

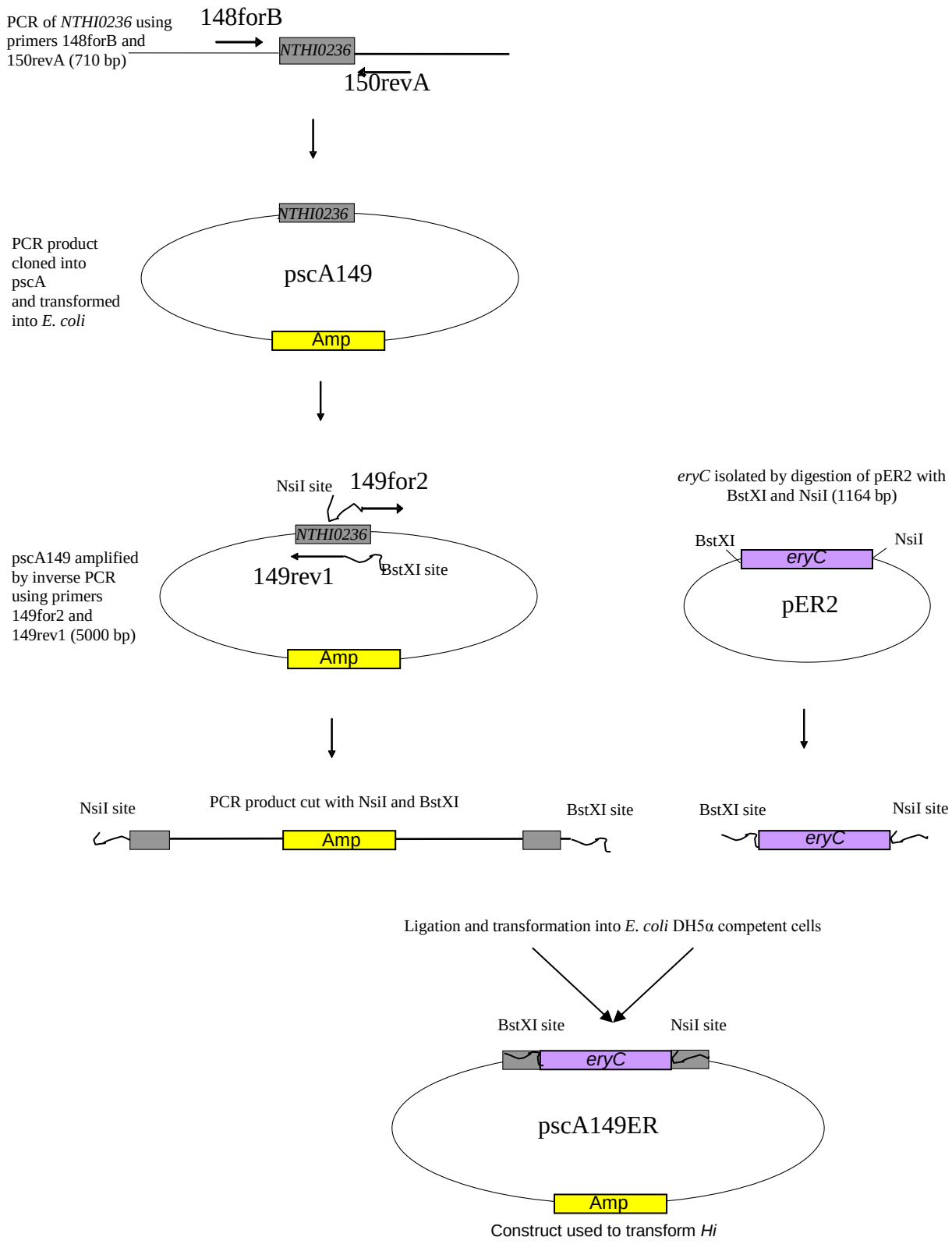


Fig. 12 Diagram showing the method used to construct plasmid pscA149ER

Construction of the *HI0141* (*nagB)* mutants (Fig. 13)

Plasmid p21, a pT7Blue vector (Novagen 800-526-7319) carrying the *Hi* Rd *HI0141-HI0142* region (amplified by PCR using primers 0141rev and nanAfor, which produced a 1868 bp fragment) had previously been created. Plasmid p21 was digested with BstEII, which linearised the vector by cutting once at position 117 bp downstream from the start codon of *HI0141*.

5´ DNA overhangs were produced from digestion with BstEII, and DNA overhangs of cut vector were blunted using Klenow (Roche), following the manufacturer's instructions.

The *eryC* isolated by digestion of pER2 with PvuII (produced a 1596 bp fragment) was ligated to cut vector using T4 DNA Ligase (Biolabs). The ligation product was transformed into *E. coli* DH5α competent cells (see section 2.2.7) and transformants were selected on LB plate containing ery from overnight culture at 37 ºC. Colonies were picked and plasmids were extracted. The correct construct that contained the *HI0141 (nagB)* ORF and *eryC* in the same orientation (confirmed by PCR amplification and restriction analyses) was named p21E and was used to transform *Hi* strains Rd, 2019, and 375 using the MIV method (see section 2.2.6). The *eryC* was inserted in the same orientation as *HI0141 (nagB)* to ensure that the gene is properly inactivated and that *eryC* can be fully expressed. *Hi* mutants that carried inactivated HI0141 (NagB) were named Rd *nagB*, 2019 *nagB*, 375 *nagB*, and Rd *cyaAnagB*.

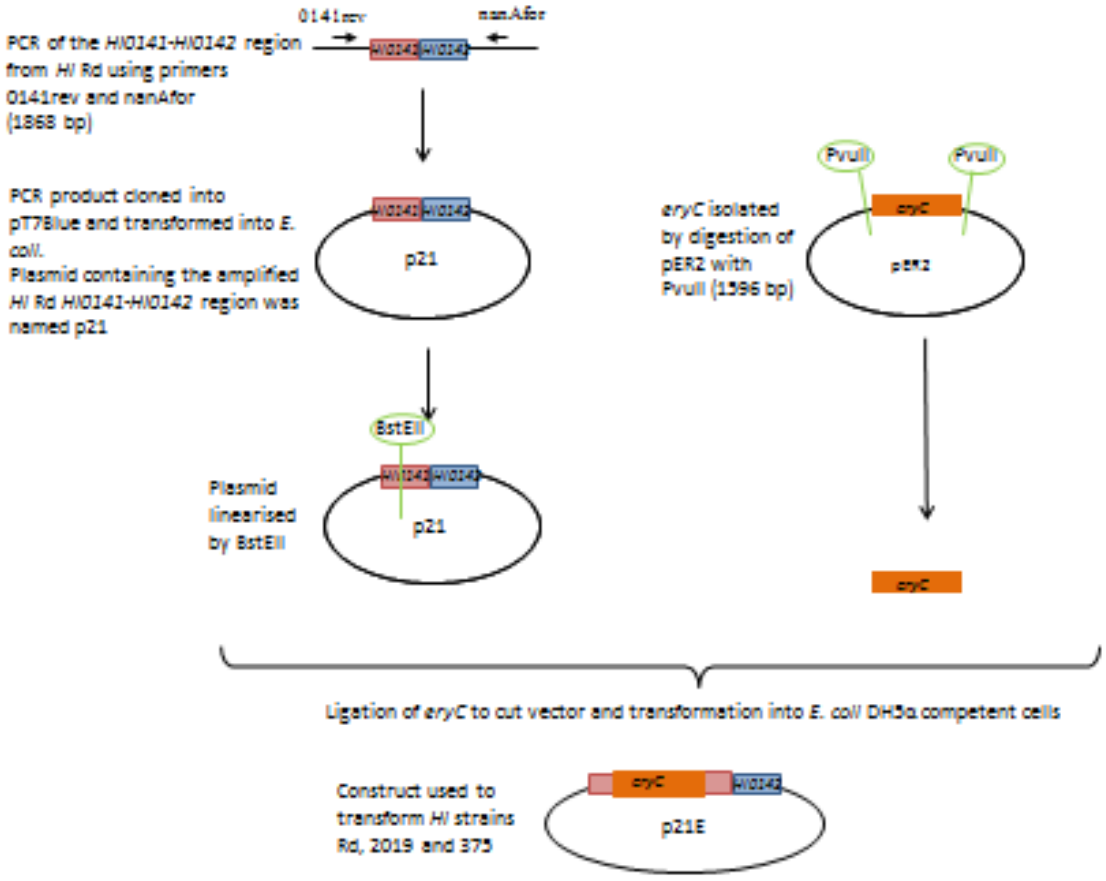


Fig. 13 Diagram showing the method used to construct the plasmid that carried inactivated HI0141 (NagB).

Plasmids are represented by black circles. Name of the plasmids are stated within the circles. Restriction enzymes are circled in green. The green linear lines represent the sites at which the restriction enzyme cut. The red and blue boxes respectively represent the *HI0141* and *HI0142* ORFs. The orange box represents the *eryC*.

Construction of the *cyaA* mutants

Plasmid p13, a pT7Blue vector (Novagen 800-526-7319) carrying the *Hi* Rd *cyaA* ORF and *kanC* in the same orientation, had been previously created. Plasmid p13 was used to transform *Hi* strains Rd and 2019 by the MIV method (see section 2.2.6). *Hi* mutants that carried inactivated CyaA were named Rd *cyaA*, 2019 *cyaA*, and 2019 *cyaAnagB.* The 2019 *cyaAnagB* double mutant could not be created by the MIV method, therefore the static transformation method (see section 2.2.6) was used instead, which required chromosomal DNA from the 2019 *cyaA* mutant and the 2019 *nagB* competent cells.

3 Evaluation of the repertoire of known genes, and identification and analyses of candidate novel sialometabolism genes

Sialometabolism in *Hi* involves different metabolic functions and these include Neu5Ac transport, catabolism, activation and sialyltransferase activity. In this chapter, we first examined whether sialometabolism is generally present in *Hi*. Key sialometabolism functions were investigated by screening strains for the presence of *siaP, nanA, siaB,* and *lic3A,* which respectively are involved in *Hi* Neu5Ac transport, catabolism, activation and sialyltransferase activity.

To further our understanding of *Hi* sialometabolism, genes that had not been previously studied in *Hi* but are predicted to be relevant to sialometabolism in this organism will be identified, characterised and their potential functions studied. These genes will now be referred to as candidate novel sialometabolism genes. Furthermore, genes adjacent, and are likely to be functionally linked, to the candidate novel sialometabolism genes will also be studied. These adjacent genes will now be referred to as peripheral candidate sialometabolism genes.

3.1 Key sialometabolism genes are conserved across all *Hi* strains screened

The presence of known sialometabolism genes (*nanA*, *siaP*, *siaB*, *lic3A*) across selected *Hi* strains was investigated. The selected strains include the 25 NT*Hi* strains from the Finnish otitis media study, 2019, R2846, R2866 and Rd (see Table 1). The 25 Finnish otitis media strains represent genetic diversity across NT*Hi.* Strains 2019, R2846, R2866 are clinical isolates and Rd is a standard laboratory *Hi* strain. The known sialometabolism genes were selected as representative genes involved in each of the key aspect of sialometabolism: *siaP* for Neu5Ac

transport, *nanA* for Neu5Ac catabolism, *siaB* for Neu5Ac activation, and *lic3A* for the transfer of Neu5Ac onto the LPS molecule.

Results obtained from PCR amplification of chromosomal DNA isolated from these strains showed that each of these key sialometabolism gene was present and conserved in length (Table 3). Primer pairs used to detect the presence of *nanA*, *siaP*, *siaB*, and *lic3A* were respectively nanAfor/nanArev, 0146for/0146rev, siaC/siaD, and lic3AA/lic3Z (see Table 2). Out of these primer pairs, only lic3AA/lic3Z was an internal and external primer pair (lic3AA anneals 99 to 120 bp downstream of the start codon of *lic3A*, and lic3Z anneals 396 to 413 bp downstream of the stop codon of *lic3A*) and the rest are internal primer pairs that anneal within the ORF of their respective genes. Internal primer pairs do not indicate the context of the genes. The presence of these genes among *Hi* strains screened suggests that sialometabolism is conserved across *Hi.*

Table 3 PCR amplification of known sialometabolism genes (*siaP, nanA, siaB, lic3A*).

Strains	Genes and primers			
	siaP 0146for/ 0146rev	*nanA* nanAfor/ nanA rev	*siaB* siaC/ siaD	*lic3A* lic3AA/ lic3Z
162	~700 bp	~750 bp	~600 bp	~1000 bp
176	~700 bp	~750 bp	~600 bp	~1000 bp
285	~700 bp	~750 bp	~600 bp	~1000 bp
375	~700 bp	~750 bp	~600 bp	~1000 bp
432	~700 bp	~750 bp	~600 bp	~1000 bp
477	~700 bp	~750 bp	~600 bp	~1000 bp
486	~700 bp	~750 bp	~600 bp	~1000 bp
667	~700 bp	~750 bp	~600 bp	~1000 bp
723	~700 bp	~750 bp	~600 bp	~1000 bp
981	~700 bp	~750 bp	~600 bp	~1000 bp
1008	~700 bp	~750 bp	~600 bp	~1000 bp
1124	~700 bp	~750 bp	~600 bp	~1000 bp
1158	~700 bp	~750 bp	~600 bp	~1000 bp
1159	~700 bp	~750 bp	~600 bp	~1000 bp
1180	~700 bp	~750 bp	~600 bp	~1000 bp
1181	~700 bp	~750 bp	~600 bp	~1000 bp
1200	~700 bp	~750 bp	~600 bp	~1000 bp
1207	~700 bp	~750 bp	~600 bp	~1000 bp
1209	~700 bp	~750 bp	~600 bp	~1000 bp
1231	~700 bp	~750 bp	~600 bp	~1000 bp

1232	~700 bp	~750 bp	~600 bp	~1000 bp
1233	~700 bp	~750 bp	~600 bp	~1000 bp
1247	~700 bp	~750 bp	~600 bp	~1000 bp
1268	~700 bp	~750 bp	~600 bp	~1000 bp
1292	~700 bp	~750 bp	~600 bp	~1000 bp
2019	~700 bp	~750 bp	~600 bp	~1000 bp
R2846	~700 bp	~750 bp	~600 bp	~1000 bp
R2866	~700 bp	~750 bp	~600 bp	~1000 bp
Rd	~700 bp	~750 bp	~600 bp	~1000 bp

The approximate (~) product sizes are shown in bp. PCR amplification product of each of the genes seemed to run at the same positions on the gel.

3.2 Selection of candidate sialometabolism genes in *Hi*

To try to further our knowledge in the complex sialometabolism process in *Hi*, we decided to study candidate novel sialometabolism genes in this organism. Candidate novel sialometabolism genes that were selected to be studied are homologues of genes associated with sialometabolism in *E. coli* or genes located beside sialometabolism genes in *E. coli*. The study of adjacent ORFs to the selected candidate novel sialometabolism genes in *Hi* (the peripheral candidate sialometabolism genes) has also been included. The peripheral candidate sialometabolism genes are likely transcribed in the same operon as, and thus may be functionally linked to one of the selected *Hi* candidate novel sialometabolism genes (see section 3.2.2).

3.2.1 **Identification of sialometabolism associated genes in *E. coli* and screening their respective homologues in *Hi***

Sialometabolism has been extensively studied in *E. coli*. Therefore, we decided to screen for the homologues of the *E. coli* sialometabolism associated genes in *Hi*, and select some homologues to study as our candidate novel sialometabolism genes in *Hi.*

We first identified sialometabolism genes in *E. coli* from the Comprehensive Microbial Resource website by looking up gene symbols and common names containing words with sia-, neura-, nan-, neu- and nag- (CMR, http://cmr.jcvi.org/tigr-scripts/CMR/CmrHome

Page.cgi) (Table 4). Amino acid sequences of proteins associated with the *E. coli* sialometabolism were obtained from CMR, and using the BLASTP search method, we screened for their respective homologues in *Hi* (Table 4). We also searched for published literature on sialometabolism in *Hi* and in *E. coli* using Pubmed (http://www.ncbi.nlm.nih.gov/pubmed/). Combining the published literature information with our search result for sialometabolism homologues in *Hi*, we identified the candidate novel sialometabolism genes to be studied in *Hi*.

Sialometabolism proteins in *E. coli*			Homologues in *Hi*					
Sialometabolism process	Protein name	Protein function	86028NP	Rd	PittGG	PittEE	R2846	R2866
Sialidase	Sia	Predicted endo-alpha-sialidase	NP	NP	NP	NP	NP	NP
Outer membrane transport	NanC	Neu5Ac porin	NP	NP	NP	NP	NP	NP
Transport	NanT	Neu5Ac transporter	NTHI0390	HI1104	CGSHiGG_09170	CGSHiEE_06540	Hflu103000056	Hflu203000577
SA mutarotase	NanM	Putative Neu5Ac mutarotase	NTHI0235	HI0148	CGSHiGG_03330	CGSHiEE_02545	r2846v6.511	r2866v6.460
Catabolism	NanA	Neu5Ac lyase (aldolase)	NTHI0228	HI0142	CGSHiGG_03300	CGSHiEE_02575	r2846v6.517	r2866v6.466
	NanE	ManNAc-6P epimerase	NTHI0231	HI0145	CGSHiGG_03315	CGSHiEE_02560	r2846v6.514	r2866v6.463
	NanK	ManNAc-6P kinase	NTHI0230	HI0144	CGSHiGG_03310	CGSHiEE_02565	r2846v6.515	r2866v6.464
	NagA	GlcNAc-6P deacetylase	NTHI0226	HI0140	CGSHiGG_03280	CGSHiEE_02585	r2846v6.519	r2866v6.468
	NagB	GlcN-6P deaminase	NTHI0227	HI0141	NP	CGSHiEE_02580	r2846v6.518	r2866v6.467
Regulation	NanR	Neu5Ac transcriptional regulator	NTHI0229	HI0143	CGSHiGG_03305	CGSHiEE_02570	r2846v6.516	r2866v6.465
SA activation	NeuA/SiaB	CMP-Neu5Ac synthetase	NTHI1891	HI_1279	CGSHiGG_02800	CGSHiEE_04135	Hflu103001146	Hflu203001032
Capsular polysialic acid biosynthesis	KpsM	Polysialic acid capsule translocation and transport: inner membrane protein	NP	NP	NP	NP	NP	NP
	KpsT	Polysialic acid capsule translocation and transport: ATP binding component	NTHI0473	HI0354	CGSHiGG_04575	CGSHiEE_01235	Hflu103001368	Hflu203001643
	NeuE	Polysialic acid capsule biosynthesis protein	NP	NP	NP	NP	NP	NP
	NeuS	Poly-α-2,8 sialyltransferase	NP	NP	NP	NP	NP	NP
	KpsS	Polysialic acid capsule synthesis protein	NP	NP	NP	NP	NP	NP
	KpsC	Polysialic acid capsule translocation and transport: inner membrane protein	NP	NP	NP	NP	NP	NP
	KpsU	CMP-KDO synthetase	NTHI0068	HI0058	CGSHiGG_02800	CGSHiEE_03010	Hflu103001278	Hflu203001245
	KpsD	Polysialic acid capsule translocation and transport: outer membrane protein	NP	NP	NP	NP	NP	NP
	KpsE	Polysialic acid capsule export inner-membrane protein	NP	NP	NP	NP	NP	NP
	KpsF	Arabinose-5-phosphate isomerase	NTHI1981	HI1678	CGSHiGG_02125	CGSHiEE_03670	Hflu103001461	Hflu203001119

Neu5,9Ac2 metabolism	NanS	Neu5,9Ac2 esterase	NP	NP	NP	NP	NP	NP
Neu5Ac bioynthesis	NeuC	UDP-GlcNAc 2 epimerase	NP	NP	NP	NP	NP	NP
	NeuB	Neu5Ac synthase	NP	NP	NP	NP	NP	NP
Unknown	YhcH	Predicted SA epimerase	NTHI0332	HI0227	CGSHiGG_03765	CGSHiEE_01940	R2846_0373	R2866_0355
	YjhC	Predicted oxidoreductase	NP	NP	NP	NP	NP	NP
	NeuD	Predicted polysialic acid capsule biosynthesis protein	NP	NP	NP	NP	NP	NP
	YjhB	Hypothetical metabolite transport protein	NTHI0390	HI1104	CGSHiGG_09170	CGSHiEE_06540	Hflu103000056	Hflu203000577
	NeuE	Predicted polysialic acid capsule biosynthesis protein	NP	NP	NP	NP	NP	NP

Table 4 Identification of the respective homologues of sialometabolism proteins from *E. coli* in *Hi*

Highlighted in red are the candidate novel sialometabolism genes in *Hi* that have been extensively studied.

Six *Hi* strains (86-028NP, Rd, PittGG, PittEE, R2846, R2866) were chosen to represent the different *HI0228* groups (see section 3.3.2).

Hi strain 86-028NP represents group 2A, strains Rd and PittGG represent group 2B, strains PittEE, R2846 and R2866 represent group 3 (see section 3.3.2).

Represented in the table: Neu5Ac, 5-acetyl neuraminate; ManNAc-6P, N-acetyl mannosamine 6-phosphate; GlcNAc-6P, N-acetyl glucosamine 6-phosphate; GlcN-6P, glucosamine 6-phosphate; CMP-Neu5Ac, Cytidine monophosphate N-acetylneuraminic acid; CMP-KDO, 3-deoxy-manno-octulosonate; Neu5,9Ac2, 9-O-acetyl N-acetylneuraminic acid; NP indicates not present.

The E-value cutoff was set at 0.001, which is the standard cutoff of NCBI BLAST.

It is clear from literature that the *E. coli* sialometabolism genes are often arranged in operons. These include the *nanATEK-yhcH* and the *yjhATS* operons (Fig. 14). The two operons were originally identified as sialometabolism operons because they were induced in the *nanR* inactivated mutant (Vimr et al., 2004). As previously mentioned, the *E. coli nanATEK-yhcH* operon encompasses sialometabolism transport (*nanT*) and catabolism (*nanAEK*) genes (Fig. 14). The *yhcH* gene is located within this operon, adjacent to the *nanK* gene (Fig. 14). *yhcH* and *nanK* are separated by 3 bp, which indicates that their expression is linked and that potentially their function is linked. A *yhcH* inactivated mutant created in *E. coli* has shown no difference to the wild type strain when grown in the presence of Neu5Ac (Kalivoda et al., 2003). The function of YhcH remains unknown. We have identified the homologue of YhcH in *Hi* (Table 4). The locus number in *Hi* Rd is *HI0227*. YhcH and HI0227 are almost the same length (154 and 155 amino acids for YhcH and HI0227, respectively), and they share 67 % amino acid sequence similarity. The crystal structure of HI0227 has been analysed by Teplyakov and colleagues and the function of HI0227 was postulated to be a SA epimerase (Teplyakov et al., 2005). An epimerase is an enzyme that catalyses the rearrangement of hydroxyl groups of substrates. One SA associated epimerase has been identified in the *Hi* Neu5Ac catabolism pathway: NanE (Fig. 6). It is unclear in which area of sialometabolism this putative SA epimerase could be functioning, therefore a series of experiments were designed to try elucidate the function in some aspect of sialometabolism for HI0227 (see chapter 4).

The *E. coli yjhATS* operon encompasses the *nanC* (*yjhA*), the *nanM* (*yjhT*) and the *nanS* (*yjhS*) genes, which respectively encode the Neu5Ac porin, the Neu5Ac mutarotase, and the Neu5,9Ac2 esterase (Fig. 14) (Condemine et al., 2005, Severi et al., 2008, Steenbergen et al., 2009). As already mentioned in sections 1.4.1, the Neu5Ac porin facilitates the diffusion of Neu5Ac, the Neu5Ac mutarotase functions to accelerate the conversion of α-Neu5Ac to the β-anomer

(Condemine et al., 2005, Severi et al., 2008), and the Neu5,9Ac2 esterase deacetylates the 9 carbon acetyl group of Neu5,9Ac2 to Neu5Ac (Steenbergen et al., 2009) (see section 1.4.3). A BLASTP search for the homologues of NanC, NanM and NanS in *Hi* reveals that *Hi* contains a homologue for NanM, but not for NanC or NanS (Table 4). It was hypothesised by Severi and colleagues that bacteria containing NanM, such as *E. coli*, would be at a physiological advantage: newly-cleaved Neu5Ac in the α-form (Holmquist & Ostman, 1975, Severi et al., 2008) could be rapidly converted to the ß-form (the likely anomeric form taken up by NanT, the Neu5Ac transporter of *E. coli*), so that limited amount of extracellular newly-cleaved Neu5Ac can be rapidly transported into the cell (Severi et al., 2008). The locus number given for *nanM* in *Hi* Rd is *HI0148*, and this gene is located within the sialometabolism gene cluster of *Hi*, downstream to transport genes *siaP* and *siaQM* (Fig. 7).

The homologues of HI0227 and HI0148 in *E. coli* strongly suggest that these proteins are involved in sialometabolism (Severi et al., 2008, Teplyakov et al., 2005). Furthermore, the positions of the *HI0227* and the *HI0148* homologues in *P. multocida* (a close relative of *Hi)* are located within sialometabolism gene clusters of that organism, which has an overall arrangement similar to that of *Hi* (Fig. 14). Given the fact that the positions of the *HI0227* and the *HI0148* homologues in *E. coli* and in *P. multocida* are linked with other sialometabolism genes, there is strong likelihood that the genes do possess some function related to sialometabolism.

Other candidate genes that may be associated with sialometabolism in *E. coli* but whose functions have not been fully characterised in *Hi* include: *nanT, yjhB, kpsT, kpsU,* and *kpsF*. These genes were excluded in our study due to various reasons.

NanT and YjhB are both homologues of *HI1104* in *Hi* Rd. In *E. coli*, NanT functions as the Neu5Ac transporter (Vimr & Troy, 1985a) (see section 1.4.1). The role of YjhB remains unclear. A previous study performed in *Hi* Rd showed that the growth rates of the wild type and the *HI1104* mutant strains were similar, and that all Neu5Ac supplemented (1 mg/ml) were consumed after about 3 hours of growth in both strains (Severi et al., 2005). These data indicate that the HI1104 protein has no role in the uptake and utilization of Neu5Ac under the conditions tested. We postulate that HI1104 in *Hi* may be involved in the transport of a related SA molecule/s. Since little is known about the availability of alternative SA molecule/s and the transport thereof in the upper respiratory tract of human, we had not included HI1104 in our study. The homologue of the *E. coli* KpsT protein is HI0354 in *Hi* Rd. HI0354 was excluded in our study because it was postulated to function as an ATP-binding protein for the uptake of taurine (an organic acid). Although both *kpsU* and *kpsF* belong to the same operon, which is involved in the translocation of capsular polysaccharide across periplasmic space and onto the cell surface of *E. coli*, the exact functions of these genes are unknown (Ferrero & Aparicio). Through amino acid sequence homology, KpsU and KpsF were respectively postulated to be a CMP-KDO synthetase and an arabinose-5-phosphate isomerase (Meredith & Woodard, 2003). KpsU and KpsF may actually play no role in sialometabolism.

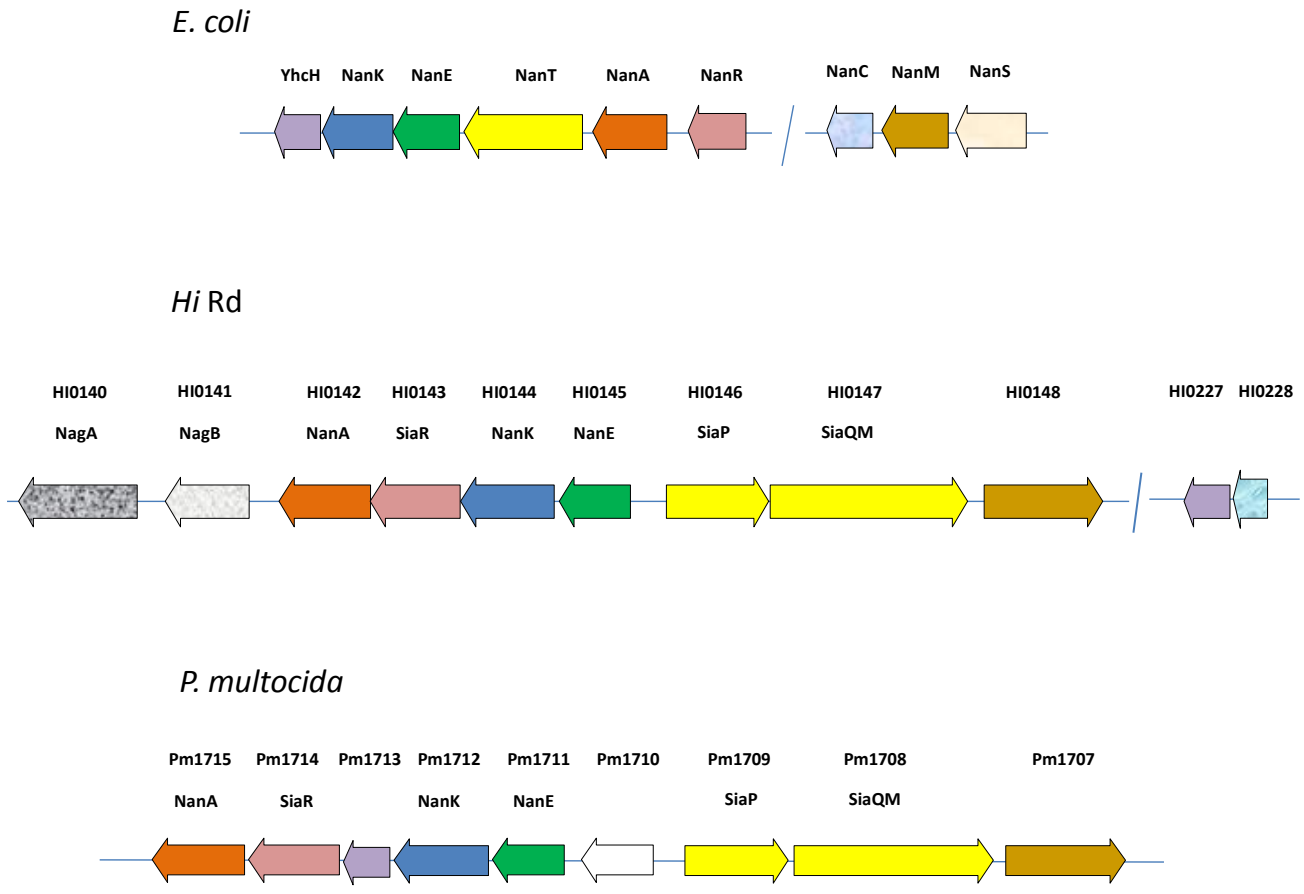


Fig. 14 Sialometabolism gene clusters of *E. coli*, *Hi* Rd, and *P. multocida* Pm70.
Transcriptional directions of genes are indicated by the direction of the arrows. Homologues are shaded in the same colours. The names of the proteins are stated directly above the arrows. Genetic organisation of the *P. multocida* sialometabolism operon is very similar to that of *Hi*. The *Hi HI0227* is located at a distance to the rest of the *Hi* sialometabolism genes; however, the *HI0227* homologues of *P. multocida* (*pm1713*) and *E. coli* (*yhcH*) are located within their respective sialometabolism gene clusters.
Represented in the diagram are: NanK, ManNAc-6P kinase; NanE, ManNAc-6P epimerase; NanT, *E. coli* Neu5Ac transporter; NanA, Neu5Ac aldolase; NanR, *E. coli* Neu5Ac repressor; NanC, Neu5Ac porin; NanM, Neu5Ac mutarotase; NanS, Neu5,9Ac2 esterase; NagA, GlcNAc-6P deacetylase; NagB, GlcN-6P deaminase; SiaR, *Hi* Neu5Ac transcriptional regulator; SiaP, Neu5Ac TRAP transporter ESR component; SiaQM, Neu5Ac TRAP transporter transmembrane units.

3.2.2 **Study of adjacent ORFs of selected *Hi* candidate sialometabolism genes**

HI0228 is closely juxtapositioned to *HI0227* and the two ORFs are transcribed in the same orientation. The intergenic spacing between the two ORFs in *Hi* Rd is only 75 bp, indicating that it is possible *HI0227* and *HI0228* are functionally linked. For these reasons, *HI0228* was included in our study.

HI0148.1 and *HI0149* are two ORFs downstream to, and transcribed in the same direction as, *HI0148* in *Hi* Rd. The spacings between *HI0148* and *HI0148.1*, and between *HI0148.1* and *HI0149* are 157 bp and 43 bp in *Hi* Rd, respectively. Previous studies on the expression of the *Hi* sialometabolism operon included these two small open reading frames (Johnston et al., 2007); therefore, *HI0148.1* and *HI0149* were also included in our study along with *HI0148*.

3.3 Analyses of the distribution and sequences of selected candidate sialometabolism genes

3.3.1 ***HI0227* is conserved in length across screened strains**

Across *Hi* strains screened, PCR amplifications of *HI0227* (using an internal primer pair: 227forA/227revA) and the *HI0226-HI0227* intergenic region (using an internal/external primer pair: 227forA/227revB) (Fig. 15) showed that the lengths of the *HI0227* ORF and the *HI0226-HI0227* intergenic region are overall conserved (Fig. 15 and Table 5). Furthermore, sequencing of *HI0227* and the *HI0226-HI0227* intergenic region showed that the lengths are conserved with 98 % sequence homology among strains. *HI0227* and the *HI0226-HI0227* intergenic region are 468 bp and 108 bp, respectively (Fig. 15).

Table 5 PCR amplification product from genomic DNA from the 25 NT*Hi* strains, 2019, R2846, R2866 and Rd across the *HI0226-HI0229* and the *HI0148-NTHI0236* regions.

Strains	Primers and regions amplified				
	(227forA 227revA) *HI0227*ORF	(227forA 227revB) *HI0227* ORF and *HI0226-HI0227* intergenic region	(227forC 227revB) *HI0229-HI0228 intergenic region*, *HI0228 ORF*, *HI0227* ORF and *HI0226-HI0227* intergenic region	(148forA 148revA) *HI0148* ORF	(148forB 150revA) *HI0148-HI0148.1* intergenic region, *NTHI0236* (*HI0148.1* and *HI0149*), and *HI0149-HI0150* intergenic region
162	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
176	~390 bp	~660 bp	~1650 bp	~1100 bp	~700 bp
285	~390 bp	~660 bp	~1650 bp	~1100 bp	~700 bp
375	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
432	~390 bp	~660 bp	~900 bp	~1100 bp	~700 bp
477	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
486	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
667	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
723	~390 bp	~660 bp	~1900 bp	~1100 bp	~700 bp
981	~390 bp	~660 bp	~1650 bp	~1100 bp	~700 bp
1008	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
1124	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
1158	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
1159	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
1180	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
1181	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
1200	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
1207	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
1209	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
1231	~390 bp	~660 bp	~1650 bp	~1100 bp	~700 bp
1232	~390 bp	~660 bp	~1650 bp	~1100 bp	~700 bp
1233	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
1247	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
1268	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
1292	~390 bp	~660 bp	~1500 bp	~1100 bp	~700 bp
2019	ND	ND	ND	~1100 bp	~700 bp
R2846	ND	ND	ND	~1100 bp	~700 bp
R2866	ND	ND	ND	~1100 bp	~700 bp
Rd	~390 bp	~660 bp	1500 bp	~1100 bp	~700 bp

The approximate (~) product sizes are shown in bp.

Shown in brackets are the primer pairs used. Locations of primer pairs can be found in Fig. 9 and Fig. 16.

ND indicates not done. Product sizes obtained from the primer pair 227forC/227revB are variable in length: approximately 900 bp (group 1), 1500 bp (group 2), 1650 bp (group 3), and 1900 bp (group 4).

3.3.2 *HI0228* and the *HI0227-HI0228* intergenic region are variable in length across strains screened

PCR amplification across the entire *HI0227-HI0228* region using an external primer pair (227forC/227revB) produced different product sizes (Fig. 15 and Table 5). We classified the screened strains into 4 different groups according to their PCR amplification product sizes. Groups 1, 2, 3, 4 were approximately 900 bp, 1500 bp, 1650 bp and 1900 bp in size, respectively. Strain 432 is the only strain in group 1. Out of the 25 Finnish otitis media NT*Hi* strains screened, 18 strains (72 %) belong to group 2, and 5 strains (20 %) to group 3. Strain 723 is the only strain in group 4.

To define the differences in the sequence of the *HI0227-HI0228* region across various groups, PCR amplification products (using primers 227forC and 227revB) of selected strains from each group were sequenced. Strains 375, 1247, 1233 were selected from group 2, and strains 981, 1231, 1232 from group 3. Strains 432 and 723, the only strains from group 1 and 4, respectively, were also selected for DNA sequencing.

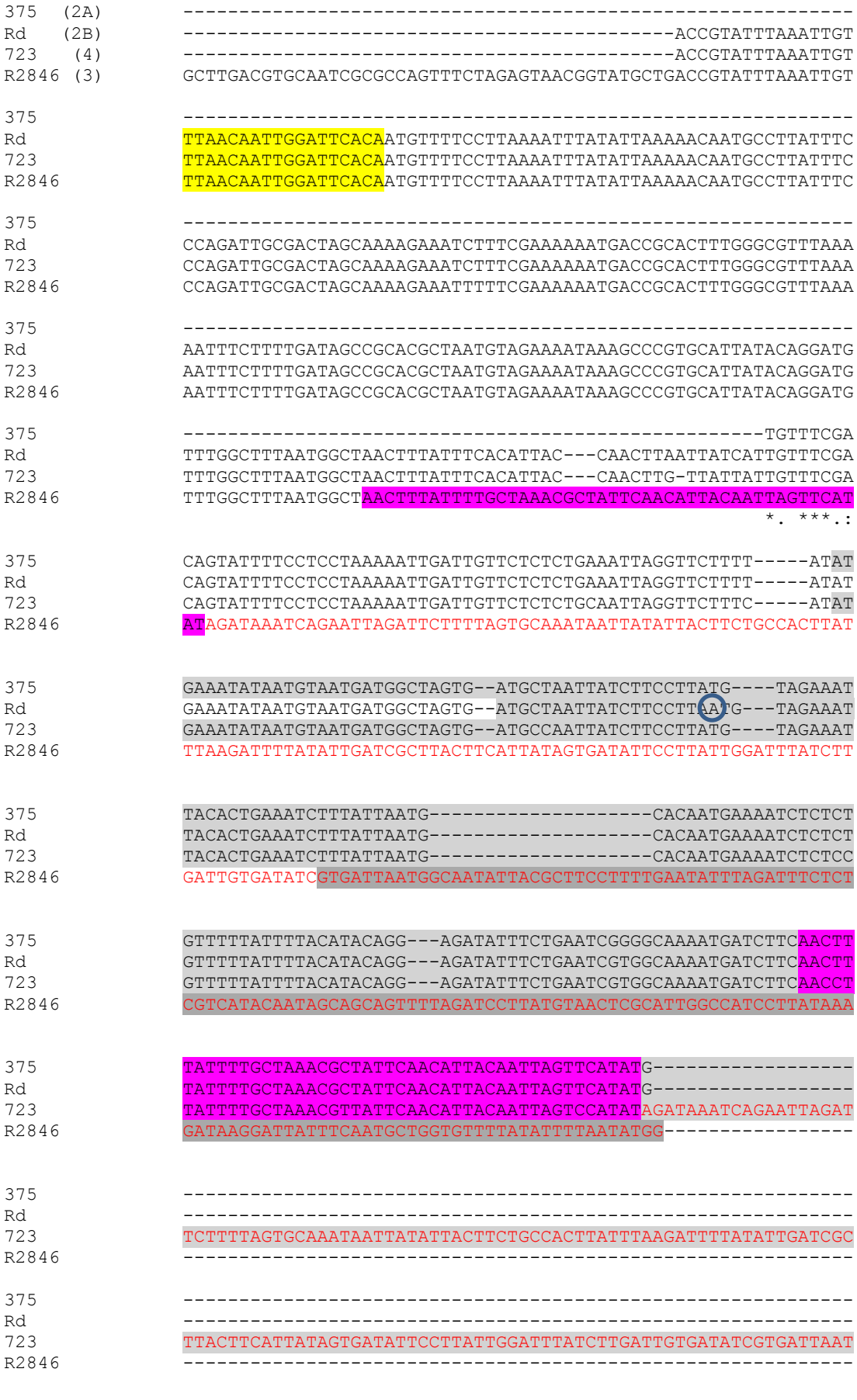
The DNA sequence from the group 1 strain (NT*Hi* 432) showed that the *HI0228* ORF is absent in this group, and the DNA sequence from the group 4 strain (NT*Hi* 723) showed that the *HI0228* ORF is the largest among all groups (891 bp) (Fig. 15). The DNA sequences of group 2 strains showed that this group can be further classified into two subgroups by the absence (2A) or presence (2B) of an adenine (Fig. 15). The adenine, at position 47 bp downstream from the start codon in strain 1233 (group 2B), is absent in strains 375 and 1247 (group 2A) (Fig. 15, shown in a blue circle). The presence of the adenine results in a frame-shift of the *HI0228* ORF, which would affect its translation. The length of the group 2A *HI0228* ORF is 448 bp. DNA

sequence of available completed *Hi* genome projects show that strain 86-028NP belongs to group 2A, while Rd and PittGG belong to group 2B.

The DNA sequence of group 3 strains showed that the entire *HI0228* of group 3 is conserved (with 97 % sequence homology) to the group 4 *HI0228* ORF from 325 bp to 842 bp (Fig. 15). The length of the group 3 *HI0228* ORF is 518 bp. NT*Hi* strains that belong to this group are 981, 1231, 1232, and DNA sequences from available completed *Hi* genome projects indicated that strains R2846, R2866 and PittEE also belong to this group.

The *HI0228* sequence of group 4 strain (NT*Hi* 723) is composed of *HI0228* sequences from different groups: the proximal *HI0228* sequence (1 bp to 189 bp) of strain 723 shares 98 % sequence homology to the *HI0228* sequence of group 2, and the distal *HI0228* sequence (190 bp to 700 bp) of strain 723 shares 97 % sequence homology to the *HI0228* sequence of group 3 (Fig. 15). The entire *HI0228* sequence of strain 723 is 889 bp.

These genetic rearrangements identified in the *HI0228* ORFs and the *HI0228-HI0229* intergenic region indicates that recombination and or deletion/insertion of DNA had probably occurred in the past to give rise to such sequence variations among strains.


375 (2A)
Rd (2B)
723 (4)
R2846 (3)
375
Rd
723
R2846
375
Rd
723
R2846
375
Rd
723
R2846
375
Rd
723
R2846
375
Rd
723
R2846
375
Rd
723
R2846
375
Rd
723
R2846
375
Rd
723
R2846
375
Rd
723
R2846
375
Rd
723
R2846

```
375  (2A)        ------------------------------------------------------------
Rd   (2B)        ------------------------------------------------------------
723   (4)        GGCAATATTACGCTTCCTTTTGAATATTTAGATTTCTCTCGTCATACAATAGCAGCAGTT
R2846 (3)        ------------------------------------------------------------

375              ------------------------------------------------------------
Rd               ------------------------------------------------------------
723              TTAGATCCTTATGTAACTCGCATTGGCCATCCTTATAAAGATAAGGATTATTTCAATGCT
R2846            ------------------------------------------------------------

375              ------------------------------------------------------------
Rd               ------------------------------------------------------------
723              GGTGTTTTATATTTTAATATGGATAAGTATCAATTAGGAATTTCATCATTTTCTAAAGAA
R2846            ----------------------ATAAGTATCAATTAGGAATTTCATCATTTTCTAAAGAA

375              ------------------------------------------------------------
Rd               ------------------------------------------------------------
723              TTAATTACATTACATATCCAATTAAAAGAAAGTTTAATTTACGGCGATCAAGATATTCTA
R2846            TTAATTACATTACATATCCAATTAAAAGAAAGTTTAATTTACGGCGATCAAGATATTCTA

375              ------------------------------------------------------------
Rd               ------------------------------------------------------------
723              AATTACTATTTTGAGGATCAGTGGATTCCATTAGACAAGCGTTATAATTTCCAATTAGAT
R2846            AATTACTATTTTGAGGATCAGTGGATTCCATTAGACAAGCGTTATAATTTCCAATTAGAT

375              ------ATTTCAATTAATACTCTTGATATTTCTCTTGTTATTTTTCATTTTACAGGTCCT
Rd               ------ATTTCAATTAATACTCTTGATACTTCTCCTAATATTTTTCATTTTACAGGTCCT
723              CATATGATTTCAATTAATACTCTTGATATTTCTCCTGTTATTTTTCATTTTACAGGTCCT
R2846            CATATGATTTCAATTAATACTCTTGATATTTCTCCTGTTATTTTTCATTTTACAGGTCCT

375              CATAAGCCTCTTGATAATATTTTTTCTGAAAATGCTTGTGTTAATGCCGTAATTTCATTG
Rd               CATAAGCCTCTTGATAATATTTTTTCTGAAAATGCTTGTGTTAATGCCGTAATTTCATTG
723              CATAAGCCTCTTGATAATATTTTTTCTGAAAATGTTTGTGTTAATGCCGTAATTTCATTG
R2846            CATAAGCCTCTTGATAATATTTTTTCTGAAAATGTTTGTGTTAATGCCGTAATTTCATTG

375              TTTAGACTTTATGCGTCAATAAGTTG-GCAAATATTTGCTCTCTTTCACTCGGTACTACA
Rd               TTTAGACTTTATGCGTCAATAAGTTGGCAAGATATTTGCTCTCTTCCACTCGGTACTACA
723              TTTAGACTTTATGCGTCAATAAGTTG-GCAAATATTTGCTCTCTTTCACTCGGTACTACA
R2846            TTTAGACTTTATGCGTCAATAAGTTGGCAAGATGTTTGCTCTCTTCCACTCGGCACTATA

375              AGAGCTAATTGGATTAATCAAGAGAGATGATGTAATCGGTGGCTTTATTTCATTTAACCG
Rd               AGAGCTAATTGGATTAATCAAGAGAGATGATGTAATCGGTGGCTTTATTTCATTTAACTG
723              AGAGCTAATTGGATTAATCAAGAGAGATGATGTAATCGGTGGCTTTATTTCATTTAACCG
R2846            AGAGCTAATTGGATTAATCAAGAGAGATGATGTAATCGGTGGCTTTATTTCATTTAACCG

375              ACTTGGTATCATAGTGTAACTTTTATCCACTATAAGGAAAATTCTATGATTATTAGCAGT
Rd               ACTTGGTATCATAGTGTAACTTTTATCCACTATAAGGAAAATTCTATGATTATTAGCAGT
723              ACTTGGTATCATAGTGTAACTTTTATCCACTATAAGGAAAATTCTATGATTATTAGCAGT
R2846            ACTTGGTATCATAGTGTAACTTTTATCCACTATAAGGAAAATTCTATGATTATTAGCAGT

375              TTAACCAACCCAAATTTTAAAGTTGGTTTACCAAAAGTGATTGCTGAAGTATGTGATTAT
Rd               TTAACCAACCCAAATTTTAAAGTTGGTTTACCAAAAGTGATTGCTGAAGTATGTGATTAT
723              TTAACCAGCCCGAATTTTAAAGTTGGTTTACCAAAAGTGATTGCTGAAGTATGTGATTAT
R2846            TTAACCAGCCCGAATTTTAAAGTTGGTTTACCAAAAGTGATTGCTGAAGTATGTGATTAT

375              TTAAATACGTTAGATTTGAATGCGTTAGAGAATGGTCGCCATGATATTAACGATCAAATT
Rd               TTAAATACGTTAGATTTGAATGCGTTAGAGAATGGTCGCCATGATATTAACGATCAAATT
723              TTGAATACGTTAGATTTGAATGCGTTAGAAAATGGTCGCCATGATATTAACGATCAAATT
R2846            TTGAATACGTTAGATTTGAATGCGTTAGAAAATGGTCGCCATGATATTAACGATCAAATT
```

```
375     (2A)      TATATGAACGTAATGGAACCAGAAACTGCAGAACCAAGCAGTAAAAAAGCAGAATTACAT
Rd      (2B)      TATATGAACGTAATGGAACCAGAAACTGCAGAACCAAGCAGTAAAAAAGCAGAATTACAT
723     (4)       TATATGAACGTAATGGAACCAGAAACTGCAGAATCAAGCAGTAAAAAAGCAGAATTACAT
R2846   (3)       TATATGAACGTAATGGAACCAGAAACTGCAGAATCAAGCAGTAAAAAAGCAGAATTACAT

375               CATGAATACTTAGATGTACAAGTGCTTATTCGTGGCACGGAAAATATCGAAGTGGGTGCA
Rd                CATGAATACTTAGATGTACAAGTGCTTATTCGTGGCACGGAAAATATCGAAGTGGGTGCA
723               CATGAATACTTAGATGTACAAGTGCTTATTCGTGGCACGGAAAATATCGAAGTGGGCGCA
R2846             CATGAATACTTAGATGTACAAGTGCTTATTCGTGGCACGGAAAATATCGAAGTGGGCGCA

375               ACCTATCCGAATTTATCTAAGTATGAAGATTACAACGAAGCGGATGATTATCAACTTTGT
Rd                ACCTATCCGAATTTATCTAAGTATGAAGATTACAACGAAGCGGATGATTATCAACTTTGT
723               ACCTATCCGAATTTATCTAAGTATGAAGATTACAACGAAGCGGATGATTATCAACTTTGT
R2846             ACCTATCCGAATTTATCTAAGTATGAAGATTACAACGAAGCGGATGATTATCAACTTTGT

375               GCAGATATTGATGATAAATTCACCGTCACAATGAAACCAAAAATGTTCGCTGTATTCTAT
Rd                GCAGATATTGATGATAAATTCACCGTCACAATGAAACCAAAAATGTTCGCTGTATTCTAT
723               GCAGATATTGATGATAAATTCACCGTCACAATGAAACCAAAAATGTTCGCTGTATTCTAT
R2846             GCAGATATTGATGATAAATTCACCGTCACAATGAAACCAAAAATGTTCGCTGTATTCTAT

375               CCTTATGAACCGCACAAGCCTTGCTGTGTAGTGAATGGAAAAACAGAAAAAATTAAAAAA
Rd                CCTTATGAACCGCACAAGCCTTGCTGTGTAGTGAATGGAAAAACAGAAAAAATTAAAAAA
723               CCTTATGAACCGCACAAGCCTTGCTGTGTAGTGAATGGAAAAACAGAAAAAATTAAAAAA
R2846             CCTTATGAACCGCACAAGCCTTGCTGTGTAGTGAATGGAAAAACAGAAAAAATTAAAAAA

375               TTAGTCGTAAAAGTACCTGTTAAATTAATTTAATTTTTGAAAGAGTGCGGTTAGAAATTT
Rd                TTAGTCGTAAAAGTACCTGTTAAATTAATTTGATTTTTGAAAGAGTGCGGTTAGAAATTT
723               TTGGTCGTAAAAGTACCTGTTAAATTAATTTAATTTTTGAAAAAGTGTGGTTAGAAATTT
R2846             TTGGTCGTAAAAGTACCTGTTAAATTAATTTGATTTTTGAAAAAGTGTGGTTAGAAATTT

375               ATCAGCCCTGTGCCTTATATCTCCGCCCTGAAGGACGGAGTTTT----------------
Rd                ATCAGCCCTGTGCCTTATATCTCCGCCCTGAAGGACGGAGTTTT----------------
723               ATCAGCCCTGTGCCTTATATCTCCGCCCTGAAGGACGGAGTTTTACGGCGCAGGGCTGAT
R2846             ATCAGCCCTGTGCCTTATATCTCCGCCCTGAAGGACGGAGTTTTACGGCGCAGGGCTGAT

375               -------------------------
Rd                -------------------------
723               AAATAAAACACTTTGTGATATTAAG
R2846             AAATAAAACACTTTGTGATATTAAG
```

Fig. 15 The *HI0227-HI0228* region of strains 375 (group 2A), Rd (group 2B), R2846 (group 3) and 723 (group 4).
Highlighted in green, grey and yellow are respectively the *HI0227*, *HI0228* and *HI0229* sequences. The *HI0229* ORF is in the opposite direction as *HI0228* and *HI0227*. Highlighted in pink is a 46 bp region located within the R2846 *HI0228-HI0229* intergenic region that is identical in sequence to an area found within the *HI0228* ORF of strains Rd, 375 and 723. Sequences shown in red for strains 723 and R2846 are conserved but are located at different positions. Shown in the blue circle in strain Rd is the adenine that results in the frame-shift of the *HI0228* ORF.

3.3.3 *HI0148* and *NTHI0236* are conserved across screened strains

PCR amplifications of *HI0148* and the *HI0148.1-HI0149* region using primer pairs 148forA/148revA and 148forB/150revA respectively, across the 25 NT*Hi* strains, showed that the products are conserved in length (Fig. 16 and Table 5).

The *HI0148.1* and *HI0149* ORFs in Rd are presented as a single gene in other *Hi* strains screened (NT*Hi* 86-028NP, PittEE, PittGG, R2846, R2866). This difference observed between *Hi* Rd and other NT*Hi* strains screened is attributable to the absence of a thymine at position 151 bp downstream from the translational start of the *HI0148.1* ORF, giving rise to a frame-shift of the gene and resulting in two ORFs in Rd. The combined *HI0148.1* and *HI0149* ORFs in Rd will now be called *NTHI0236*, the homologue of the two ORFs in strain 86-028NP.

Sequences of *HI0148* and *NTHI0236* obtained from available *Hi* genome projects (strains Rd, 86-028NP, PittEE, PittGG, R2846, R2866) showed that the two genes are conserved in length. The lengths of *HI0148* and *NTHI0236* are 1139 bp and 435 bp, respectively. The *HI0148-NTHI0236* intergenic region among screened strains is also conserved in length and consists of 161 bp (Fig. 16).

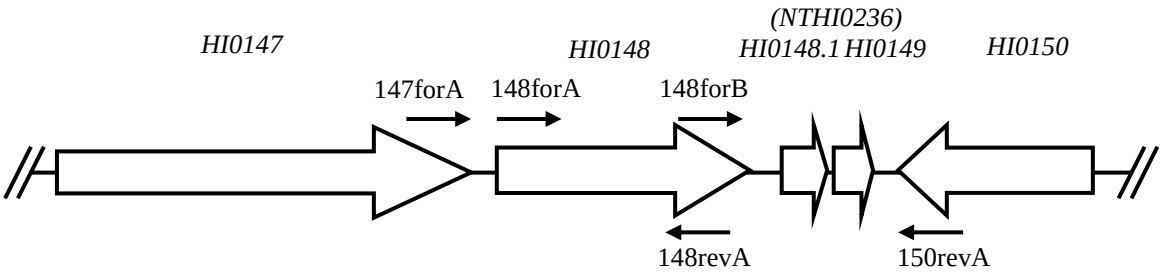


Fig. 16 *HI0147-HI0150* region
Transcriptional directions of ORFs are shown by the directions of the large arrows. PCR amplification primers are indicated by the small arrows. Primer names are shown directly above (forward) or below (reverse) the small arrows. The names of the ORFs are shown in italic. *NTHI0236* is the combined ORF of *HI0148.1* and *HI0149* in strain 86-028NP. The lengths of *HI0148* and *NTHI0236* are 1139 bp and 435 bp, respectively.

3.4 Database search of the candidate sialometabolism genes

3.4.1 **HI0227 belongs to the YhcH/YiaL/YjgK family**

The database search using HI0227 as the query sequence, as well as YhcH from *E. coli* and Pm1713 from *P. multocida*, showed that the three homologues belong to the YhcH/YiaL/YjgK family, under the Domain of Unknown Function (DUF) 386 superfamily (http://pfam.sanger.ac.uk/family?acc=PF04074). As the name suggests, this superfamily consists of proteins with unidentified function.

3.4.2 **HI0228 contains domains of glycosyl transferase family 8 proteins**

The database search using the group 4 HI0228 sequence showed that HI0228 belongs to the glycosyltransferase (GT) 8 protein family. The general function of a GT is to transfer an activated sugar moiety to specific acceptor molecules and enzymes that belong to the GT family 8 are involved in LPS and glycogen biosynthesis (http://pfam.sanger.ac.uk/family/ PF01501). These enzymes have broad substrate specificities, and known functions found in this family include: α-1,4-galactosyltransferase, α-1,3-D-galactosyltransferase,

α-1,2-glucosyltransferase, α-1,2-galactosyltransferase, α-1,2-glucosyltransferase, glycogenin glucosyltransferase, inositol 1-a-galactosyltransferase (galactinol synthase), homogalacturonan α-1,4-galacturonosyltransferase, α-glucuronyltransferase (http://www.cazy.org/GT8.html).

The *Hi* LgtC is an α-1,4-galactosyltransferase that belongs to the GT family 8 protein family, and its function is to add an α-galactose from the donor UDP-Gal to a galactose acceptor (Fig. 3). The crystal structure of LgtC from *N. meningitidis* was previously analysed by Persson and colleagues (Persson *et al.*, 2001). Persson and colleagues have shown that LgtC is made up of a N-terminal domain, containing active sites for donor sugar binding, manganese binding (for stability and activity of the protein) and acceptor molecule binding; and a C-terminal domain, whose function is thought to mediate bacterial membrane attachment (Persson et al., 2001). The responsible residues for the *Hi* LgtC protein that function in donor sugar binding (Ser 44, Asp 46, Tyr 49, Tyr 116, Arg 124, Val 142, Gln 225, Asp 226, Gly 274, Lys 277), manganese binding (Asp 141, Asp 143, His 271), and acceptor molecule binding (Asp 168, Phe 170, Gln 227) have been identified.

Comparing the amino acid sequence of the *Hi* LgtC to the amino acid sequence of the group 4 HI0228, we found that the two proteins share 47 % amino acid sequence similarity. LgtC and the group 4 HI0228 are respectively 296 and 330 amino acids in length. We also found that the group 4 HI0228 contains both the N-terminal domain (donor sugar binding, manganese binding, acceptor molecule binding) and the C-terminal domain of a functional GT family 8 protein. Aligning the *Hi* LgtC amino acid sequence to the group 2 HI0228 showed that the proximal half of HI0228 in this group contains the truncated N-terminal domain of the GT 8 family, and that HI0228 in this group contains some active residues for donor sugar binding (Ser 44, Asp 46, Tyr

49). Aligning the *Hi* LgtC amino acid sequence to the group 3 HI0228 showed that the proximal two thirds of HI0228 in this group contain the C-terminal domain of GT family 8.

3.4.3 **HI0148 belongs to the YjhT family**

The database search using HI0148 as the query sequence showed that it belongs to the YjhT family, and contains seven kelch repeats (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?RID=TFJZKD9E014&mode=all). The Kelch motif is widely found in eukaryotic and bacterial species, and known functions of kelch-containing proteins include: scruin, an actin cross-linking protein; galactose oxidase, which catalyses the oxidation of the hydroxyl group at the C6 position in D-galactose; neuraminidase, which cleaves glycoketosidic linkage of SA by hydrolytic reaction; and kelch may also be involved in cytoskeletal function (Adams *et al.*, 2000).

As previously mentioned, the homologue of HI0148 in *E. coli* (YjhT) is the founding member of the YjhT family, and is encoded within the *E. coli* sialometabolism *yjhATS* operon (Severi et al., 2008). Experiments conducted using the *E. coli* YjhT revealed that it functions as a Neu5Ac mutarotase (Severi et al., 2008). The function of a Neu5Ac mutarotase has been described in sections 1.4.1 and 1.5.

3.4.4 **NTHI0236 belongs to the isoprenylcysteine carboxyl methyltransferase family**

The BLAST algorithm search using NTHI0236 showed that the distal half of NTHI0236 contains the truncated domain of the isoprenylcysteine carboxyl methyltransferase (ICMT) superfamily (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?RID=TTUBHYXV01N&mode=all). Protein Sterile-14p, from yeast *Saccharomyces cerevisiae*, is the founding member of this superfamily (Romano & Michaelis, 2001).

Sterile-14p, an eukaryotic endoplasmic reticulum membrane protein, functions to carboxyl methylate cleaved proteins that terminate in the CaaX motif (Romano & Michaelis, 2001). Protein LmtA from the spirochete *Leptospira interrogans (L. interrogans)*, a homologue of Sterile-14p, has been found to be involved in the methylatation of the 1-phosphate group of the *L. interrogans* lipid A, creating a distinct, methylated lipid A structure (Boon Hinckley *et al.*, 2005). At present, the significance of methylated lipid A remains unknown. Although the lengths of LmtA and NTHI0236 are quite different: 250 and 144 amino acids respectively, the two proteins share 44 % amino acid sequence similarity. Since *NTHI0236* may be associated with sialometabolism operon in *Hi*, we included this ORF in our study.

3.5 Discussion

Selected sialometabolism genes (*nanA, siaP, siaB, lic3A*), which represent different processes of *Hi* sialometabolism, are present in all strains screened. These genes were chosen to represent the *Hi* Neu5Ac catabolism (*nanA*), transport (*siaP*), activation (*siaB*), and sialyltransferase (*lic3A*) processes. Using these representative genes, it is consistent that sialometabolism is a conserved process across NT*Hi* strains.

Sialometabolism is a common process in NT*Hi* strains and can play a major role in bacteria-host interaction. Studying candidate novel sialometabolism genes in *Hi* can help us elucidate and understand unknown aspects of sialometabolism in this organism. Two primary *Hi* candidate novel sialometabolism genes were selected to be studied, and they are *HI0227* and *HI0148*. Published literature suggests that *HI0227* and *HI0148* are involved in different aspects of sialometabolism: HI0227 may serve as a potential SA epimerase, and HI0148 is likely a SA mutarotase. The homologues of *HI0227* and *HI0148* in *E. coli* and in *P. multocida*, and the

HI0148 gene itself in *Hi*, are located within sialometabolism gene clusters. Circumstantial evidence for including these two primary genes is strong, and for proximity and other reasons (such as *NTHI0236* was included in previous gene expression study), the adjacent ORFs, *HI0228* and *NTHI0236*, were included in our study.

PCR and sequencing results showed that *HI0227* and *HI0148* are present and conserved in length across all screened strains. However, this is not the case for *HI0228*. Although the BLAST search of HI0228 indicated that it contains domains of the GT family 8 protein, which supports its role in sugar metabolism, not all strains screened contain the *HI0228* ORF, and among the strains that do contain *HI0228*, the lengths of *HI0228* is variable. In our study, it appears that strain 723 is the only strain that likely carries a functional *HI0228* gene. The sequence of *HI0228* is the longest in strain 723 and the amino acid sequence of HI0228 in this strain contains all conserved residues that are required for a functional GT family 8 protein. Although *HI0228* contains truncated GT 8 motifs in many strains, there may still be some residual activity in these proteins, and careful experiments would need to be conducted to examine whether or not HI0228 is involved in the *Hi* sialometabolism.

The *NTHI0236* ORF is adjacent to, and is transcribed in the same direction as, the *HI0148* gene. NTHI0236 contains the domain of ICMT whose role may be associated in LPS biosynthesis. Although *NTHI0236* is present and conserved in length across all screened strains, it however exists as two ORFs in strain Rd due to the deficiency of an adenine within the sequence. The frame-shift mutation in *Hi* Rd would most likely result in non-functioning of the combined ORFs. However, *NTHI0236* may still encode a functional protein in other *Hi* strains. We would include *NTHI0236* in our investigation to confirm whether or not NTHI0236 affects sialometabolism through a potential role in LPS biosynthesis.

4 Phenotypic characterisation of the candidate and peripheral sialometabolism genes

We have established in the previous chapter that the candidate novel sialometabolism genes *HI0227* and *HI0148*, as well as the peripheral candidate sialometabolism genes *HI0228* and *NTHI0236*, may be involved in sialometabolism of *Hi*. In this chapter, we have created isogenic strains with mutations in the *HI0227*, *HI0148*, *HI0228* and *NTHI0236* genes to allow us to compare various features of sialometabolism phenotypes with the respective parent strains.

Phenotypic characteristics associated with various aspects of sialometabolism, such as the ability to 1) utilise Neu5Ac as the only sugar source available in aerobic and microaerophilic conditions, 2) to sialylate LPS molecules following growth in the presence of Neu5Ac, and 3) to resist serum complement mediated killing have been assessed in the mutants to determine whether or not these aspects of *Hi* sialometabolism are altered.

4.1 Mutants of *HI0227*, *HI0228* and *HI0227/HI0228* were created in group 2, 3 and 4 representative strains

Mutant strains with *HI0227*, *HI0228* or the combined *HI0227*/*HI0228* loci inactivated were created in strains 375, Rd, R2846 and 723, which respectively represent groups 2A, 2B, 3, and 4. These mutants were designated *HI0227*-, *HI0228*-, and *HI0227/HI0228*- in the respective strain backgrounds. The mutants were created by either 1) directly inserting a *kanC* within the first half of the ORF (for the *HI0227*- mutants), or by 2) deleting some of the sequences within the ORF and inserting a *kanC* thereof (for the *HI0228*- and *HI0227/HI0228*- mutants) (see section 2.2.14). For all mutant strains, the antibiotic resistance cassette was inserted in the same orientation as the reading frame of the gene. The *HI0227/HI0228*- double mutant was created to eliminate the

possibility that the double mutant could produce a different effect compared to the single mutant, as *HI0227* and *HI0228* may be functionally linked due to the position of the ORFs.

Mutant strains with HI0148 and NTHI0236 inactivated were also created in various strain backgrounds (see section 2.2.14), and they are designated *HI0148*- and *NTHI0236*- in their respective strain backgrounds. The *HI0148*- mutant was created by deleting some of the sequence within the ORF and inserting an *eryC* (see section 2.2.14). Due to the lack of appropriate restriction sites within the first half of the *NTHI0236* ORF, the *NTHI0236*- mutant was created by PCR amplification of the pscA149 recombinant plasmid, and ligation of the plasmid to an *eryC* (see section 2.2.14).

Several independent mutants were created in the same strain background for each gene to allow us to test multiple mutants and minimise the potential effect of phase variation influencing the bacterial cell surface. In an experiment where only one mutant from a strain background was required to be tested, the mutant designated with -A would be used. However, occasionally where multiple mutants are tested, the same -A mutant, as well as the -B and -C mutants would also be selected to be used.

4.2 The alternative Neu5Ac catabolism pathway is not present in *Hi*

The HI0227 protein was predicted to encode a SA epimerase (Teplyakov et al., 2005). At present, two epimerases involved in sialometabolism have been identified in the Neu5Ac catabolism pathways: NanE and N-acetyl-D-glucosamine 2-epimerase (see section 1.4.3, Fig. 6).

NanE is involved in the third step of the classic Neu5Ac catabolism pathway, converting ManNAc-6P to GlcNAc-6P (see section 1.4.3 and Fig. 6). Catabolism of Neu5Ac in *Hi* is via the classic Neu5Ac catabolism pathway (see section 1.4.3 and Fig. 6).

N-acetyl-D-glucosamine 2-epimerase is involved in an alternative Neu5Ac catabolism pathway and functions to convert ManNAc to GlcNAc (Fig. 6). Following epimerisation of ManNAc to GlcNAc, phosphorylation of GlcNAc to GlcNAc-6P occurs and is carried out by the enzyme RokA, encoded by *HI0182* in *Hi* Rd (Fig. 6). Classes of bacteria that catabolise Neu5Ac via this alternative pathway include Spirochaetes, Bacteroidetes, and Cyanobacteria. *T. forsythia* is a bacterium that belongs to the Bacteroidetes family, and catabolises Neu5Ac via the alternative Neu5Ac catabolism pathway. This bacterium is implicated in periodontitis and there has been emerging interest recently in the study of sialometabolism in *T. forsythia*. Gene *TF0031* from *T. forsythia* strain 43037 encodes the N-acylglucosamine 2-epimerase, and TF0031 consists of 394 amino acids (Roy et al., 2010). BLASTP search using the *T. forsythia* TF0031 amino acid sequence against *Hi* does not show up any protein with reasonable match. Although *Hi* does not possess a homologue for the N-acylglucosamine 2-epimerase found in *T. forsythia*; however, since it has never been tested whether the alternative Neu5Ac catabolism pathway is present in *Hi* or not, there still remain a possibility that our candidate novel sialometabolism gene *HI0227* could be involved in this alternative Neu5Ac catabolism pathway to assist balancing the intracellular level of different Neu5Ac metabolites. Therefore, it would still be worth testing whether HI0227 is involved in the alternative Neu5Ac catabolism pathway or not.

Previous studies have shown that *Hi* is capable of utilising Neu5Ac as a main carbon and energy source (Macfadyen & Redfield, 1996, Severi et al., 2005). In order to test whether the alternative Neu5Ac catabolism pathway is present in *Hi* or not (by using growth assays where Neu5Ac is

added as the only sugar source available for growth), we needed to first re-evaluate the appropriate amount of Neu5Ac required to elicit a clear difference in growth between cultures with or without added Neu5Ac. CDM has been chosen to be used for all subsequent growth assays as this is a defined growth medium. Based on previous studies (Jenkins et al., 2010), we have chosen to use Neu5Ac at an initial concentration of 30 μg/ml. Higher concentrations of Neu5Ac (at 100 and 500 μg/ml) were then chosen to ascertain the optimal concentration of sugar at which clear defined difference in growth could be detected when compared to CDM with no added sugar.

The doubling times of *Hi* Rd with added Neu5Ac at 30 μg/ml and 100 μg/ml were respectively 44 ± 3 min and 42 ± 2 min, which were not significantly different from the doubling time of the same culture without sugar supplementation (45 ± 1 min, p=0.613 and 0.081 respectively, unpaired t-test, Fig. 17, Table 6A). Using strain Rd, we showed that supplementing growth media with Neu5Ac as the only sugar source available at either 30 μg/ml or 100 μg/ml did not significantly improve growth, compared to the culture without added sugar (Fig.14).

In contrast, the doubling times of *Hi* Rd with added Neu5Ac at 500 μg/ml and with added glucose at 2 mg/ml were respectively 39 ± 2 min and 38 ± 3 min, each significantly shorter than the wild type strain with no added sugar (45 ± 1 min; p=0.0186 and 0.01, respectively, unpaired t-test). Supplementing Neu5Ac at 500 μg/ml produced a more pronounced growth difference compared to the culture that has not received added Neu5Ac (Fig. 17). In fact, supplementing Neu5Ac at this concentration produced a similar growth rate, and achieved a similar final OD, as did supplementing glucose at 2 mg/ml (Fig. 17, Table 6A). This finding helped us identify a concentration of Neu5Ac (500 μg/ml) that was sufficient to allow us to identify growth

differences that were dependent upon Neu5Ac. Neu5Ac was supplemented at this concentration (500 μg/ml) in all subsequent growth assays.

Strains and mutants		Media type	Doubling time in min
A) *Hi* Rd	WT	CDM	45 ± 1
		CDM, Glc 2mg/ml	38 ± 3
		CDM, Neu5Ac 30 μg/ml	44 ± 3
		CDM, Neu5Ac 100 μg/ml	42 ± 2
		CDM, Neu5Ac 500 μg/ml	39 ± 2
B) *Hi* Rd	WT	CDM	45 ± 1
		CDM, Neu5Ac 500 μg/ml	38 ± 3
	nanK mutant	CDM	72 ± 1
		CDM, Neu5Ac 500 μg/ml	99 ± 2
	nanE mutant	CDM	62 ± 2
		CDM, Neu5Ac 500 μg/ml	124 ± 2
C) *Hi* Rd	WT	CDM	47 ± 4
		CDM, Neu5Ac 500 μg/ml	41 ± 3
	HI0227-A	CDM	40 ± 3
		CDM, Neu5Ac 500 μg/ml	36 ± 3
	HI0228-A	CDM	43 ± 1
		CDM, Neu5Ac 500 μg/ml	38 ± 2
	HI0227/HI0228-A	CDM	41 ± 2
		CDM, Neu5Ac 500 μg/ml	38 ± 2
D) *Hi* Rd	WT	CDM	50 ± 3
		CDM, Neu5Ac 500 μg/ml	41 ± 1
	HI0148-A	CDM	53 ± 2
		CDM, Neu5Ac 500 μg/ml	43 ± 2
	NTHI0236-A	CDM	55 ± 3
		CDM, Neu5Ac 500 μg/ml	44 ± 4
E) *Hi* 375	WT	CDM	65
		CDM, Neu5Ac 500 μg/ml	52
	HI0227-A	CDM	59
		CDM, Neu5Ac 500 μg/ml	52
	HI0148-A	CDM	64
		CDM, Neu5Ac 500 μg/ml	50

Table 6 Summary of growth indicated by doubling times of *Hi* from various experiments
WT represents wild type. The doubling times of strains obtained from one experiment do not have a standard deviation.

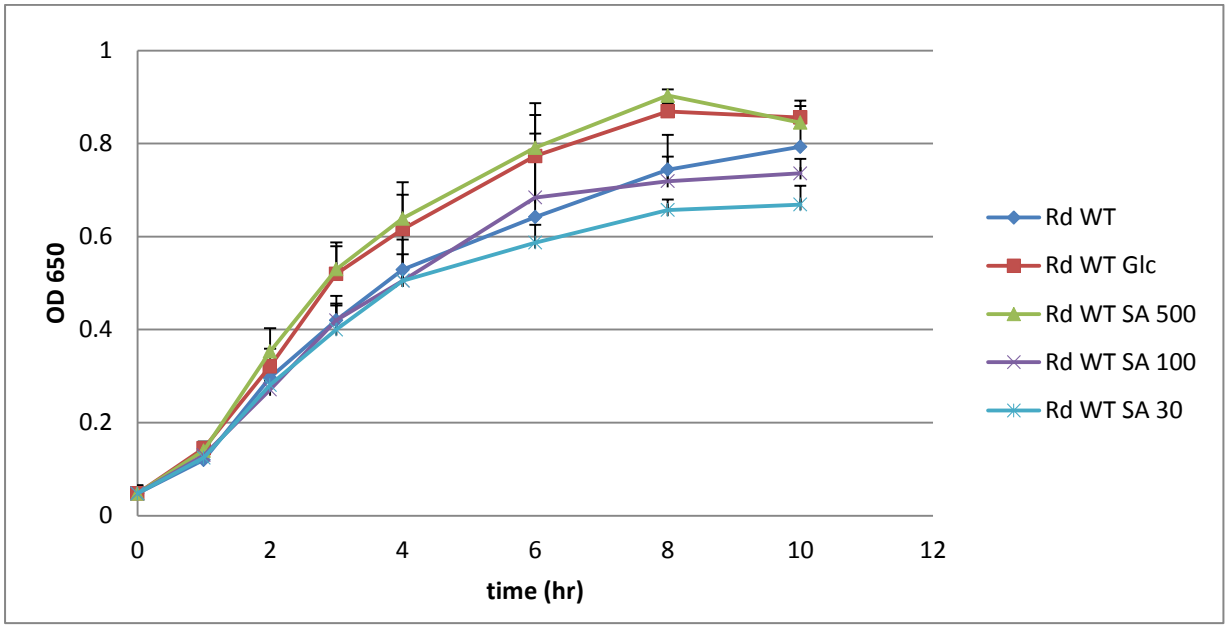


Fig. 17 Growth of *Hi* Rd in CDM with added glucose (Glc) at 2 mg/ml and Neu5Ac (SA) at 30 µg/ml, 100 µg/ml and 500 µg/ml.
Supplementation with glucose was included as a positive control as it is known to be a good source of carbon and energy for the growth of *Hi* (Macfadyen & Redfield, 1996). Results shown are the average of three independent experiments with standard deviations.

To determine whether the alternative Neu5Ac catabolism pathway exists in *Hi* or not, we grew the Rd *nanK* and *nanE* mutants in the absence or presence of 500 µg/ml Neu5Ac and compared their growths to the wild type strain (Fig. 18, Table 6B). NanK and NanE are involved in different steps of the classic Neu5Ac catabolism pathway (the second step for NanK and the third step for NanE). They respectively catalyse the chemical reactions ManNAc to ManNAc-6P and ManNAc-6P to GlcNAc-6P (Fig. 6), the steps that convene to the Neu5Ac alternative pathway. Determining whether the *nanK* and *nanE* mutants grow in the presence of added Neu5Ac or not would help us to define whether the Neu5Ac alternative exists in *Hi* or not.

Results obtained from three independent growth experiments showed that supplementing the culture media with Neu5Ac for growth of the Rd *nanK* and *nanE* mutants produced prolonged doubling times compared to the cultures without added Neu5Ac and to the wild type strain with

or without added Neu5Ac (Fig. 18, Table 6B). The doubling time of the Rd *nanK* mutant with added Neu5Ac was 99 ± 2 min, which was significantly longer than the same culture without added Neu5Ac (72 ± 1 min, $p<0.0001$, unpaired t-test, Table 6B) and the wild type strain without added Neu5Ac (45 ± 1 min, $p<0.0001$, unpaired t-test, Table 6B). Similarly, the doubling time of the Rd *nanE* mutant with added Neu5Ac (124 ± 2 min) was significantly longer than the same culture with added Neu5Ac (62 ± 2 min, $p<0.0001$, unpaired t-test, Table 6B) and the wild type strain without added Neu5Ac (45 ± 1 min, $p<0.0001$, unpaired t-test, Table 6B).

The prolonged doubling times observed for the Rd *nanK* and *nanE* mutants indicated that Neu5Ac is not being catabolised in these mutants, and thus the characterised alternative Neu5Ac catabolism pathway cannot be present in *Hi*. We conclude that HI0227 can therefore not play a role in the alternative Neu5Ac catabolism pathway. Interestingly, supplementing Neu5Ac in the cultures of the Rd *nanK* and *nanE* mutants produced even longer doubling times compared to cultures without Neu5Ac supplementation (Table 6B). This is likely due to the excess amount of Neu5Ac accumulated in these bacteria, as build-up of Neu5Ac (acidity) could be toxic to bacteria and thus inhibited growth of *Hi*.

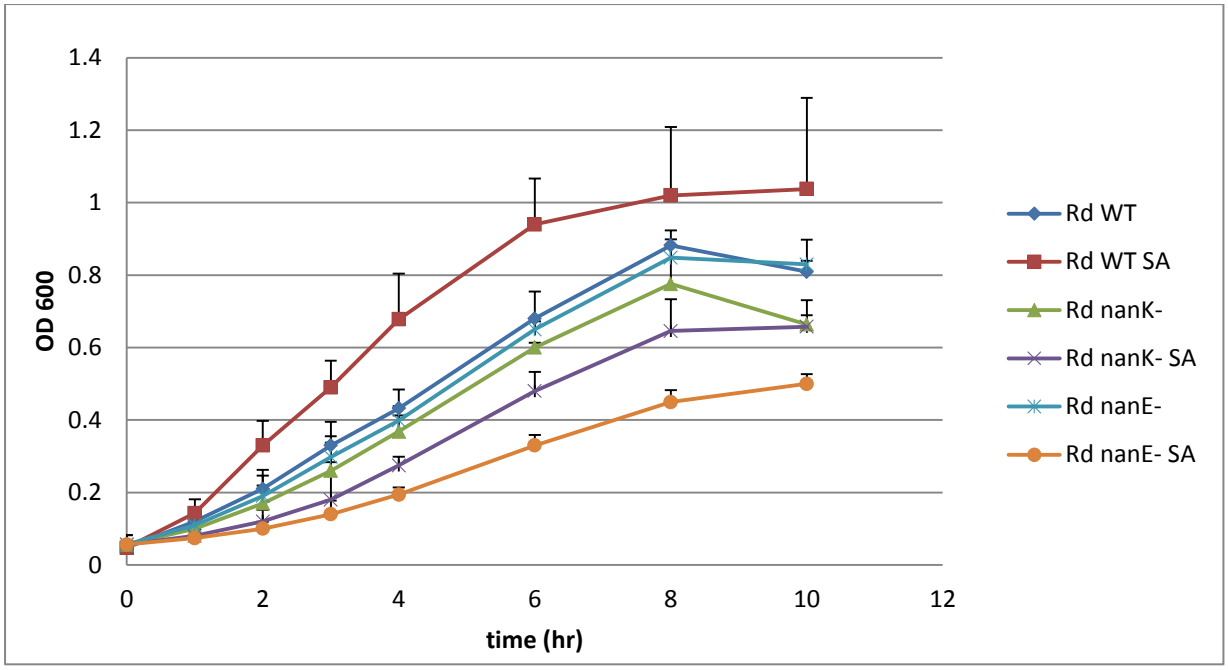


Fig. 18 Growth of the Rd wild type, *nanK* and *nanE* mutant strains in the absence or presence of 500 µg/ml Neu5Ac (SA) in CDM.
Results shown are the average of three independent experiments with standard deviations.

4.3 Assessment of whether HI0227, HI0228, HI0148 and NTHI0236 are involved in the catabolism of Neu5Ac

Although the alternative Neu5Ac catabolism pathway is not present in *Hi*, it remained unknown whether *HI0227* could encode an epimerase involving in an uncharacterised aspect of the Neu5Ac catabolism pathway. We therefore wish to determine whether the ability to utilise Neu5Ac as the only sugar source is affected in the *HI0227*- mutants or not. Furthermore, as HI0228 may contain sugar binding properties that could be implicated in sugar metabolism, we also wish to determine whether growth on Neu5Ac in the *HI0228*- mutant could be different from the wild type strain or not.

Many of the known sialometabolism functions (such as Neu5Ac transport, activation, catabolism and sialyltransferase activities) have been studied in detail in strain Rd and we chose to use *Hi* Rd for this purpose. Results obtained from three independent growth experiments showed that

the Rd mutant strains (*HI0227-A*, *HI0228-A*, *HI0227/HI0228-A*) exhibited similar growth characteristics as their parent strain in CDM without and with added Neu5Ac (Fig. 19A). The doubling times of the Rd *HI0227-A*, *HI0228-A*, *HI0227/HI0228-A* mutant strains in CDM without added sugar were respectively 40 ± 3 min, 43 ± 1 min, 41 ± 2 min, which were not significantly different from the parent strain without added sugar (47 ± 4 min; p=0.074, 0.168 and 0.081 respectively, unpaired t-test, Table 6C). Similarly, the doubling times of the Rd *HI0227-A*, *HI0228-A*, *HI0227/HI0228-A* mutant strains in CDM with added Neu5Ac (36 ± 3 min, 38 ± 2 min and 38 ± 2 min, respectively) were not significantly different from the parent strain in CDM with added Neu5Ac (41 ± 3 min; p=0.111, 0.223 and 0.223, respectively, unpaired t-test, Table 6C). These results suggest that HI0227 and HI0228 play no direct role in the catabolism of Neu5Ac by *Hi*.

HI0148 was postulated to be a Neu5Ac mutarotase, which may affect growth on Neu5Ac and NTHI0236 may be associated with sialometabolism of *Hi* due to its proximity to the *Hi* sialometabolism gene cluster. We therefore determined the growth characteristic of the Rd *HI0148*- and the Rd *NTHI0236*- mutants with Neu5Ac as the only sugar source available. Growth assays of the Rd *HI0148*-A and *NTHI0236*-A mutants showed that growth on Neu5Ac was not affected in these mutants compared to the wild type strain (Fig. 19B): the doubling times of the Rd *HI0148*-A and *NTHI0236*-A mutant strains in the presence of 500 μg/ml Neu5Ac were 43 ± 2 min and 44 ± 4 min, respectively, which were not significantly different from their parent strain supplemented with 500 μg/ml Neu5Ac (41 ± 1 min; p=0.196 and 0.276, respectively, unpaired t-test). Similarly, the doubling times of the Rd *HI0148*-A and *NTHI0236*-A mutant strains without added sugar were 53 ± 2 min and 55 ± 3 min, respectively, which were not significantly different from their parent strain without added sugar (50 ± 3 min; p=0.223 and 0.111, respectively, unpaired t-test).

To validate our results obtained from strain Rd, we performed a single growth experiment using a different strain background (*Hi* 375) to assess whether the *HI0227*- and *HI0148*- mutants did indeed exhibit similar growth phenotypes as the wild type strain. Strain 375 was selected because many of the sialometabolism functions are also well studied in this strain background. Our single growth experiment showed that the 375 *HI0227*-A and the 375 *HI0148*-A mutants did indeed exhibit similar growth characteristic as the wild type strain in the presence of Neu5Ac as the only sugar source available (Fig. 20). The doubling times of the 375 *HI0227*-A and 375 *HI0148*-A mutants were similar to the 375 wild type strain in the presence or absence of added Neu5Ac (Table 6E): 65 min, 59 min, and 64 min respectively for *Hi* 375 wild type, the 375 *HI0227-A*, and the 375 *HI0148-A* mutants in the absence of Neu5Ac; and 52 min, 52 min and 50 min respectively for the same cultures with added Neu5Ac (Table 6E). These findings obtained for strain 375 validated our results from strain Rd: that inactivation of HI0227 and HI0148 does not affect growth of *Hi* on Neu5Ac when it is the only sugar available for the bacterium.

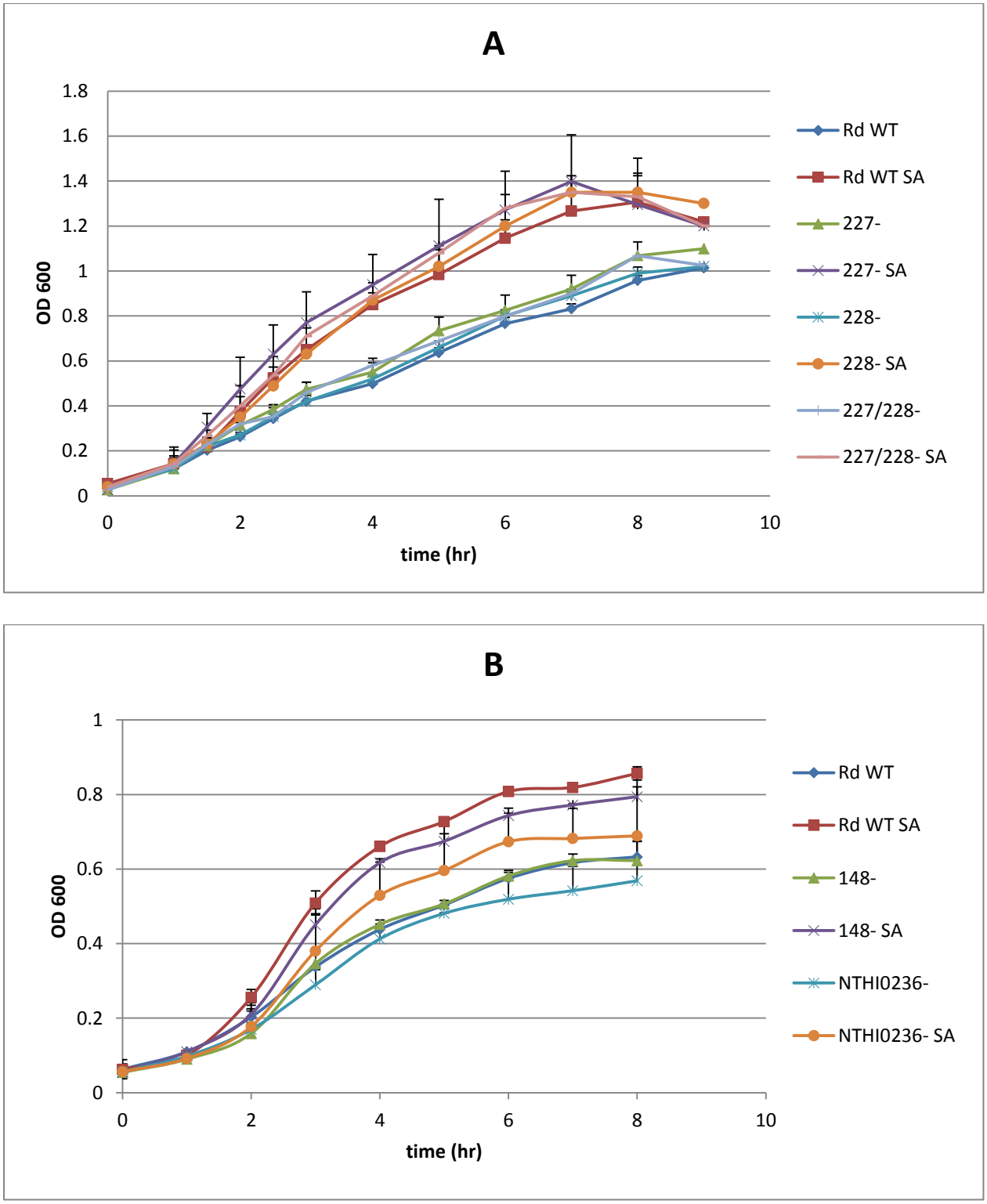


Fig. 19 Growth of *Hi* Rd wild type (WT), *HI0227*-A, *HI0228*-A mutant strains (panel A), and *HI0148*-A, *NTHI0236*-A mutant strains (panel B) in the absence or presence of 500 µg/ml Neu5Ac (SA) in CDM.
Results shown are the average of three independent experiments with standard deviations.

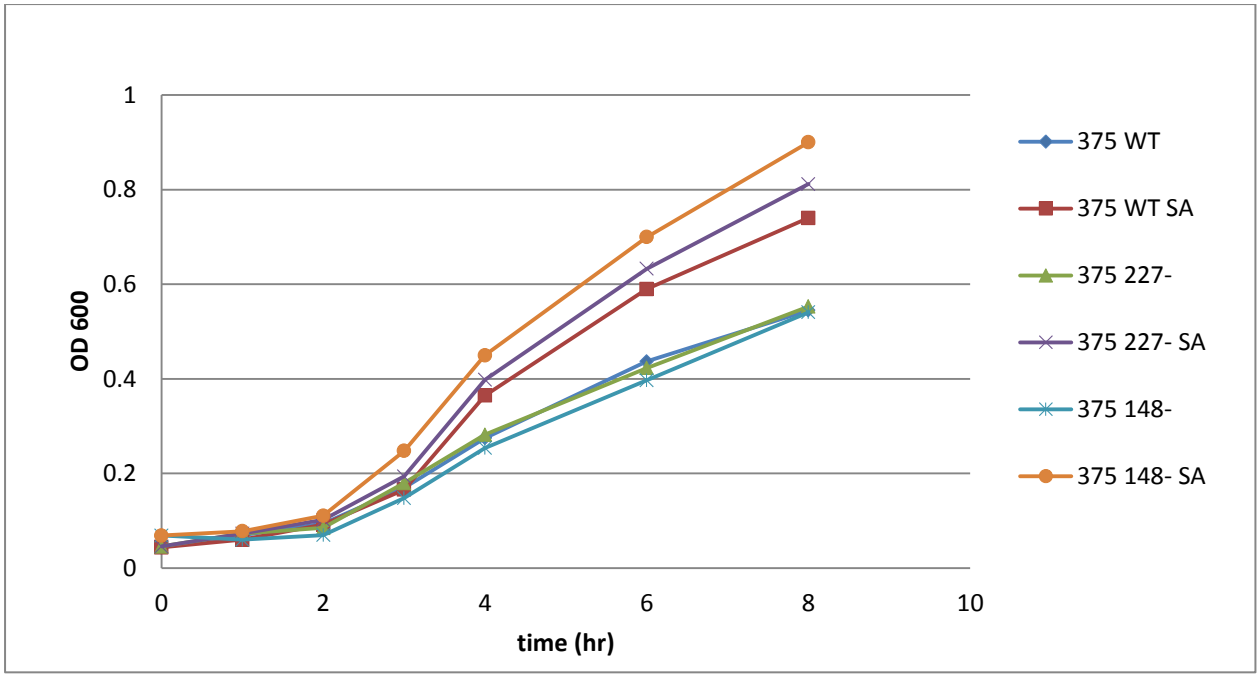


Fig. 20 Growth of *Hi* 375 wild type (WT), *HI0227*-A and *HI0148*-A mutant strains in the absence or presence of 500 µg/ml Neu5Ac (SA) in CDM.
Result shown was obtained from a single growth experiment.

4.4 Fermentation assessment of the *HI0227*-, *HI0228*-, *HI0227/HI0228*-, *HI0148*- and *NTHI0236*- mutants in strain Rd

Hi is a facultative anaerobe that can survive in both aerobic or microaerophilic conditions. In the presence of abundant oxygen, *Hi* produces ATP by aerobic respiration. However, when oxygen concentration becomes low, the bacterium switches to ferment sugars and grows as a facultative anaerobe. Fermentation is therefore a biological process by which sugars (such as glucose and fructose) are converted into cellular energy in the absence of, or in the presence of limited amount of, oxygen; and metabolic by-products such as lactic acid are produced.

The ability of *Hi* to ferment Neu5Ac has previously been shown by other research group (Macfadyen & Redfield, 1996). Under anaerobic condition, it is possible that the ability of *Hi* to ferment Neu5Ac occurs via a shunt that is not present under aerobic conditions. We therefore

wish to determine whether any of our target genes (*HI0227*, *HI0228*, *HI0148* and *NTHI0236*) in strain Rd could play a role related to fermentation.

Glucose and fructose were included in our fermentation assay as they are sugars that can be fermented by *Hi* (Macfadyen et al., 1996). The output of this assay is the pH dependent-change in colour, which is dependent upon the amount of lactic acid released into the growth media. This pH change is more obvious for bacteria grown on glucose and Neu5Ac, compared to those grown without sugar and fructose (Fig. 21). Our fermentation assay showed that Neu5Ac was indeed being fermented, but this was to the same degree by each strain (Fig. 21).

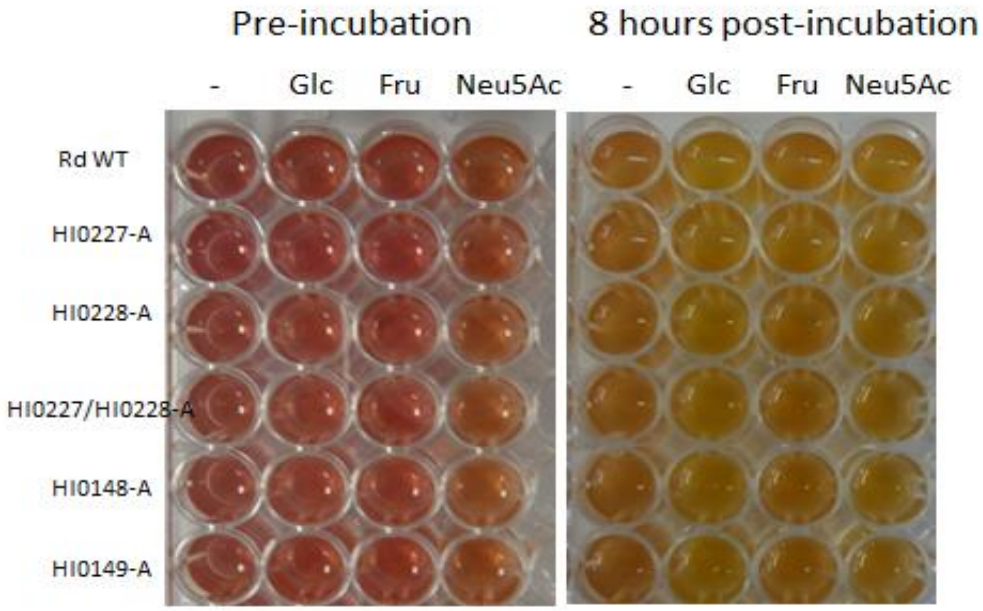


Fig. 21 Fermentation of glucose (Glc), fructose (Fru) and Neu5Ac by the Rd wild type (WT), *HI0227*-A, *HI0228*-A, *HI0227/HI0228*-A, *HI0148*-A and *NTHI0236*-A mutant strains.
Bacteria were grown in an anaerobic jar in the presence of 6% oxygen concentration. - indicates no sugar added.

4.5 LPS profiles of the *HI0227*-, *HI0228*-, *HI0227/HI0228*- mutants in strains 375, Rd, R2846 and 723

Although the ability to catabolise Neu5Ac is not affected by inactivating HI0227, HI0228, HI0148 and NTHI0236 in the Rd and 375 strain backgrounds, independent changes such as LPS sialylation could still occur in these mutants.

LPS sialylation can influence bacterial-host interaction (Figueira et al., 2007) (section 1.4.2). In order for us to determine whether or not the *HI0227*-, *HI0228*-, *HI0148*- and *NTHI0236*- mutants affect LPS sialylation, we need to first examine the LPS glycoforms of the wild type strains. Previous studies have shown that the *Hi* strains 375, Rd, R2846 and 723 contain sialylated LPS glycoforms when grown in the presence of 30 μg/ml Neu5Ac (Lundstrom *et al.*, 2008, Li *et al.*, 2007, Fox et al., 2006, Yildirim *et al.*, 2005). In our LPS gel, we could only detect sialylated LPS glycoforms in the Rd, R2846 and 723 wild type strains, and not in the 375 strain background, when grown on BHI agar in the presence of 30 μg/ml supplemented Neu5Ac (Fig. 22). The level of sialylated LPS glycoform in the 375 wild type strain is therefore below the level of detection by our LPS gel assay system.

Only one band was observed for the 375 wild type strain in our LPS gel, which would correspond to the sialyllactosyl [α-Neu5Ac(2–3)-β-Gal(1–4)-β-Glc-] or the globotriose [α-Gal(1–4)-β-Gal(1–4)-β-Glc-] LPS extensions (Fig. 3 and Fig. 22). Since neuraminidase treatment of the LPS of strain 375 did not affect the intensity of the band in the LPS gel, it is likely that the globotriose containing LPS glycoform formed the predominant LPS structure in this strain background (Fig. 22).

Three bands were observed for the LPS of the Rd wild type strain, which we respectively referred to as the high, intermediate and low molecular weight bands (Fig. 22). The high, intermediate and low molecular weight bands respectively correspond to the globotetraose [β-GalNAc(1-3)-α-Gal(1–4)-β-Gal(1–4)-β-Glc-], the sialyllactosyl or the globotriose LPS containing glycoforms, and the lactosyl [β-Gal(1→4)-β-Glc] containing LPS glycoforms (Fig. 3, Fig 18). Treatment with neuraminidase cleaves Neu5Ac off the sialyllactosyl LPS glycoform, which eliminated the sialyllactosyl LPS glycoform (indicated by the decrease in intensity of the intermediate molecular weight band, shown by the top arrow in lane 4 of Fig. 22), and increased the number of LPS glycoform with terminal lactose (indicated by the increase in intensity of the low molecular weight band, shown by the bottom arrow in lane 4 of Fig. 22).

Strain R2846 expresses a very heterogeneous mixture of LPS glycoforms (Fig. 23) (Lundstrom et al., 2008). There are three possible HepI extensions in this strain (Fig. 23) (Lundstrom et al., 2008). A major LPS glycoform contained a *β*-Glc-(1-4)-α-Hep-(1-6)-β-Glc-(1-4) motif extending off HepI and a β-Gal-(1-2) unit substituting the distal heptose (HepIII) (Fig. 23) (Lundstrom et al., 2008). Supplementing Neu5Ac in strain R2846 promotes the formation of sialylated HMG glycoform (Fig. 23) (the top band of the R2846 LPS), which loses a SA residue upon treatment with neuraminidase (Fig. 22 see arrow, Fig. 23) (Lundstrom et al., 2008).

Two bands were observed for strain 723. The top, less intense band corresponds to the sialyllactose LPS glycoform, while the bottom, more intense band corresponds to the inner LPS core (Fig. 22, Fig. 23). In strain 723, Glc is added to HepIII at a relatively low level, this is due to a less efficient initiation codon (ATT) of the *Hi* glycosyltransferase LpsA, which is responsible for the addition of a Hex to the distal heptose (HepIII) of the inner core (Deadman *et*

al., 2006). Treatment with neuraminidase in strain 723 cleaves Neu5Ac off the sialyllactose LPS glycoform, resulting in a LPS glycoform with terminal lactose (Fig. 22 see arrow, Fig. 23).

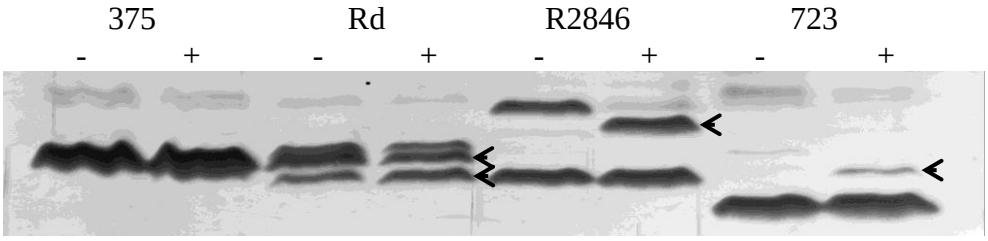
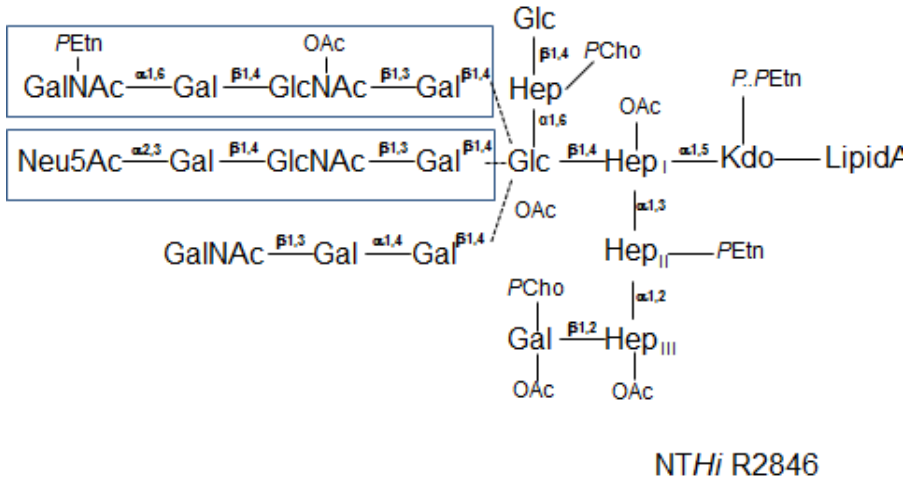


Fig. 22 LPS profiles of wild type strains 375, Rd, R2846, and 723 without (-) and with (+) neuraminidase treatment. The small arrows indicate the change in intensity of the bands following neuraminidase treatment.



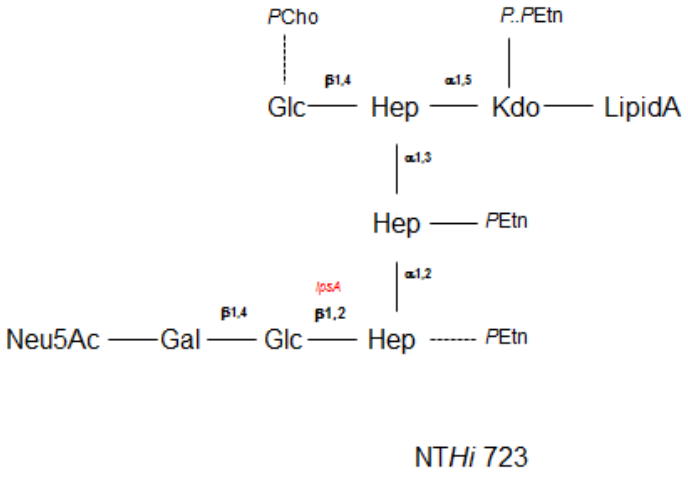


Fig. 23 LPS structures of NT*Hi* strains R2846 (top) and 723 (bottom).

To determine whether LPS glycoforms, particularly those that are sialylated, would be different in mutants that lack candidate novel sialometabolism genes and peripheral candidate novel sialometabolism genes, we analysed LPS by SDS-PAGE and compared the LPS profiles of the *HI0227*-, *HI0228*-, *HI0227/HI0228*-, *HI0148*- and *NTHI0236*- mutants to their respective wild type strains (Fig. 24, Fig. 25, Fig. 26).

Checking LPS profiles of multiple mutants from the same strain background was performed to ascertain that no independent changes in LPS structure through phase variation have taken place. For strain Rd, LPS profiles of two *HI0227*-, two *HI0228*-, and three *HI0227/HI0228*- mutant strains were obtained and analysed (Fig. 24). LPS profiles of a single Rd *HI0148*- and *NTHI0236*- mutant strain was analysed (Fig. 26). For strain R2846 and 723, LPS profiles of a single *HI0227*-, *HI0228*-, and *HI0227/HI0228*- mutant was analysed (Fig. 25). LPS profiles of one R2846 *HI0148*- and two R2846 *NTHI0236*- mutant strains were obtained and analysed (Fig. 26). For strain 375, the LPS of a single *HI0148*- mutant was analysed (Fig. 26).

Despite the fact that several *Hi* strains (Rd, R2846, 375 and 723) showing different LPS sialylation profiles have been used (Fig. 22), we observed that the LPS of our tested mutant strains exhibited similar LPS profiles as their respective wild type strains before neuraminidase treatment, and that the LPS of our tested mutant strains showed similar changes in the LPS as their respective wild type strains following neuraminidase treatment (Fig. 24, Fig. 25, Fig. 26). This indicates that the candidate novel sialometabolism genes and the peripheral candidate novel sialometabolism genes do not affect LPS sialylation of *Hi* (Fig. 24, Fig. 25, Fig. 26).

The LPS profile of the Rd *lgtC* mutant was included in Fig. 24 as a known molecular weight marker. The phase variable glycosyl transferase LgtC functions to transfer an α-galactose onto

the Hep III lactosyl extension, producing the globotriose containing LPS glycoform (Fig. 3). In the Rd *lgtC* mutant, the globotriose containing LPS glycoform cannot be synthesised, which results in the loss of the top two bands following neuraminidase treatment (Fig. 24). Interestingly, the LPS profile of this Rd *lgtC* mutant showed an obvious band corresponding to the HMG glycoform (Fig. 24). Following neuraminidase treatment, Neu5Ac is cleaved off from the sialyllacto-N-neotetraose extension of HMG, resulting in the production of the glycoform with the paraglobotriose [β-Gal(1-4)-β-GlcNAc(1-3)-β-Gal(1-4)-β-Glc] extension (Fig. 3 and Fig. 24).

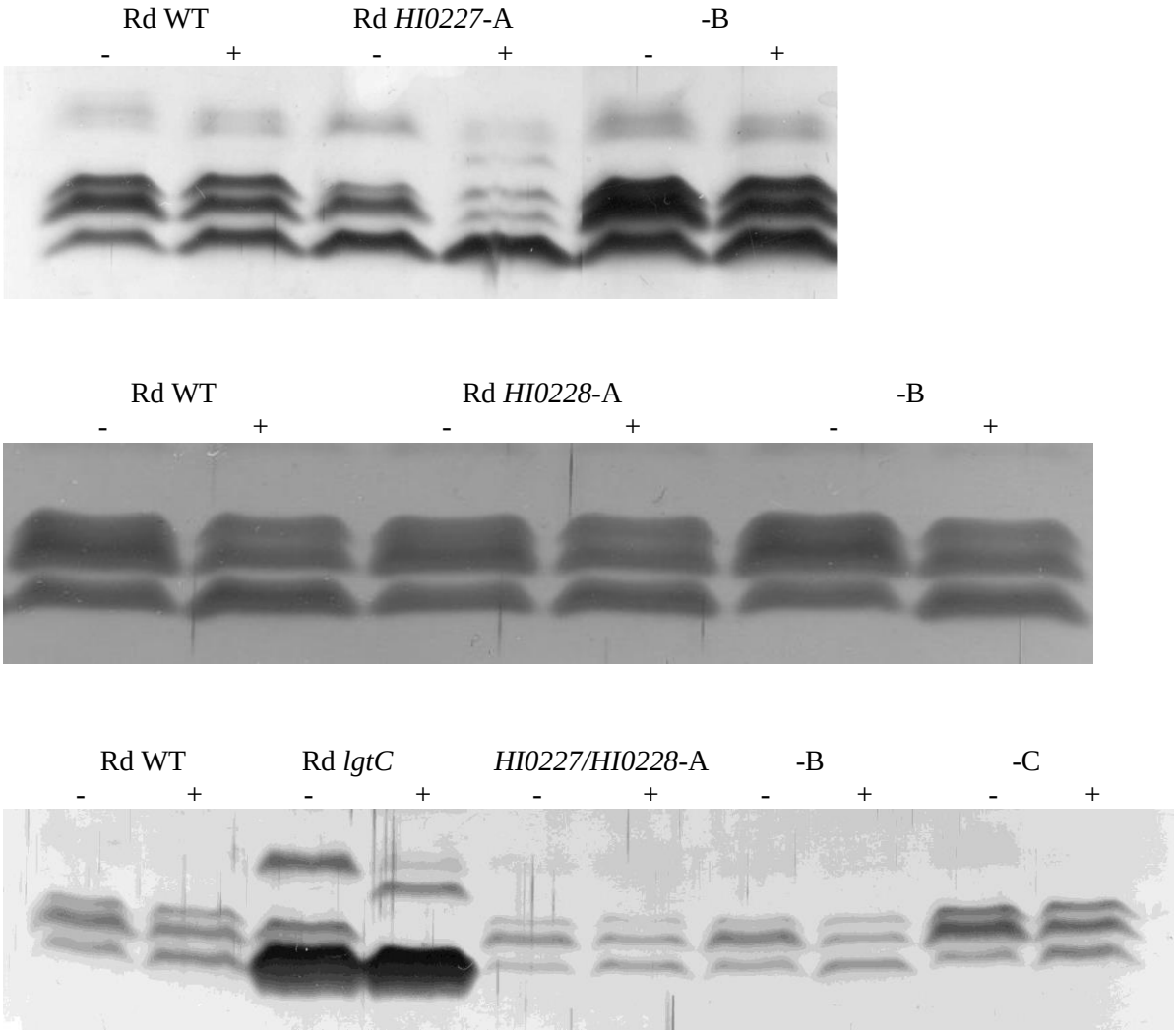


Fig. 24 LPS profiles of Rd wild type (WT), *HI0227-*, *HI0228-*, and *HI0227/HI0228-* mutant strains.

LPS glycoforms of two *HI0227*- mutants (Rd *HI0227*-A, -B), two *HI0228*- mutants (Rd *HI0228*-A, -B), and three *HI0227/HI0228*- mutants (Rd *HI0227/HI0228*-A to -C) are shown. The Rd *lgtC* mutant was included as a known molecular weight marker.

- and + indicate without and with neuraminidase treatment, respectively.

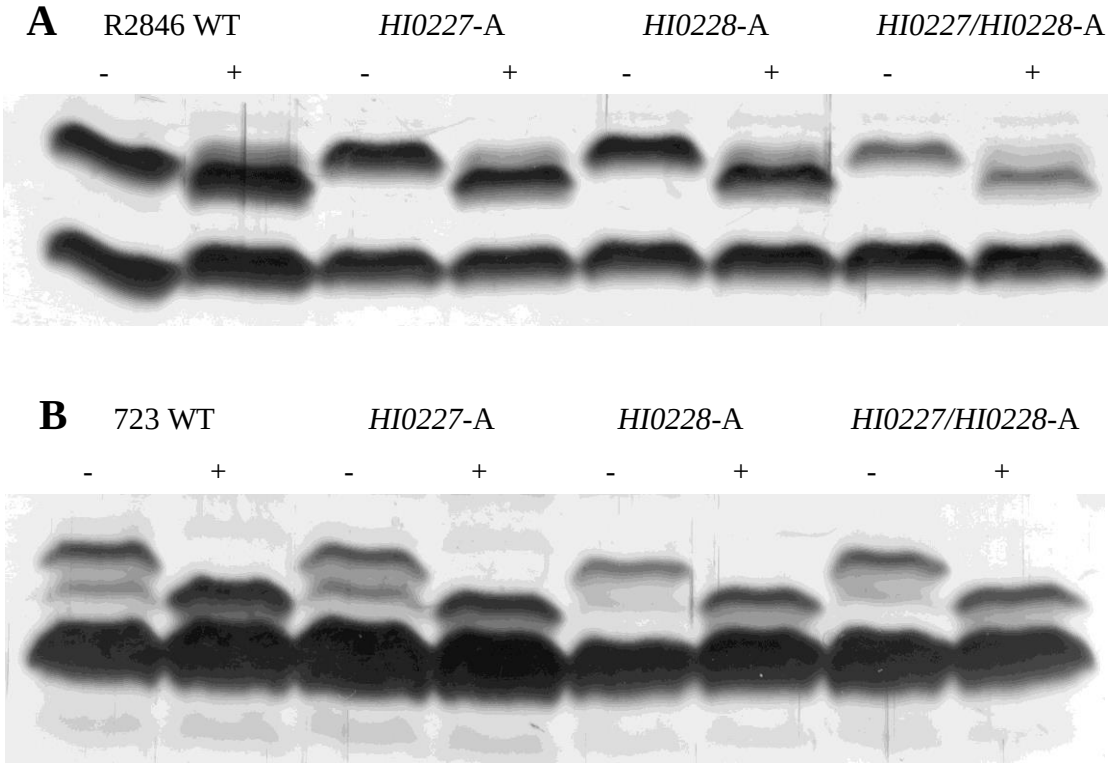


Fig. 25 LPS profiles of R2846 and 723 wild type (WT) and their respective *HI0227-*, *HI0228-*, and *HI0227/HI0228-* mutant strains.
Shown in panels A and B are the R2846 and 723 wild type and mutant strains, respectively.
- and + respectively indicate without and with neuraminidase treatment.

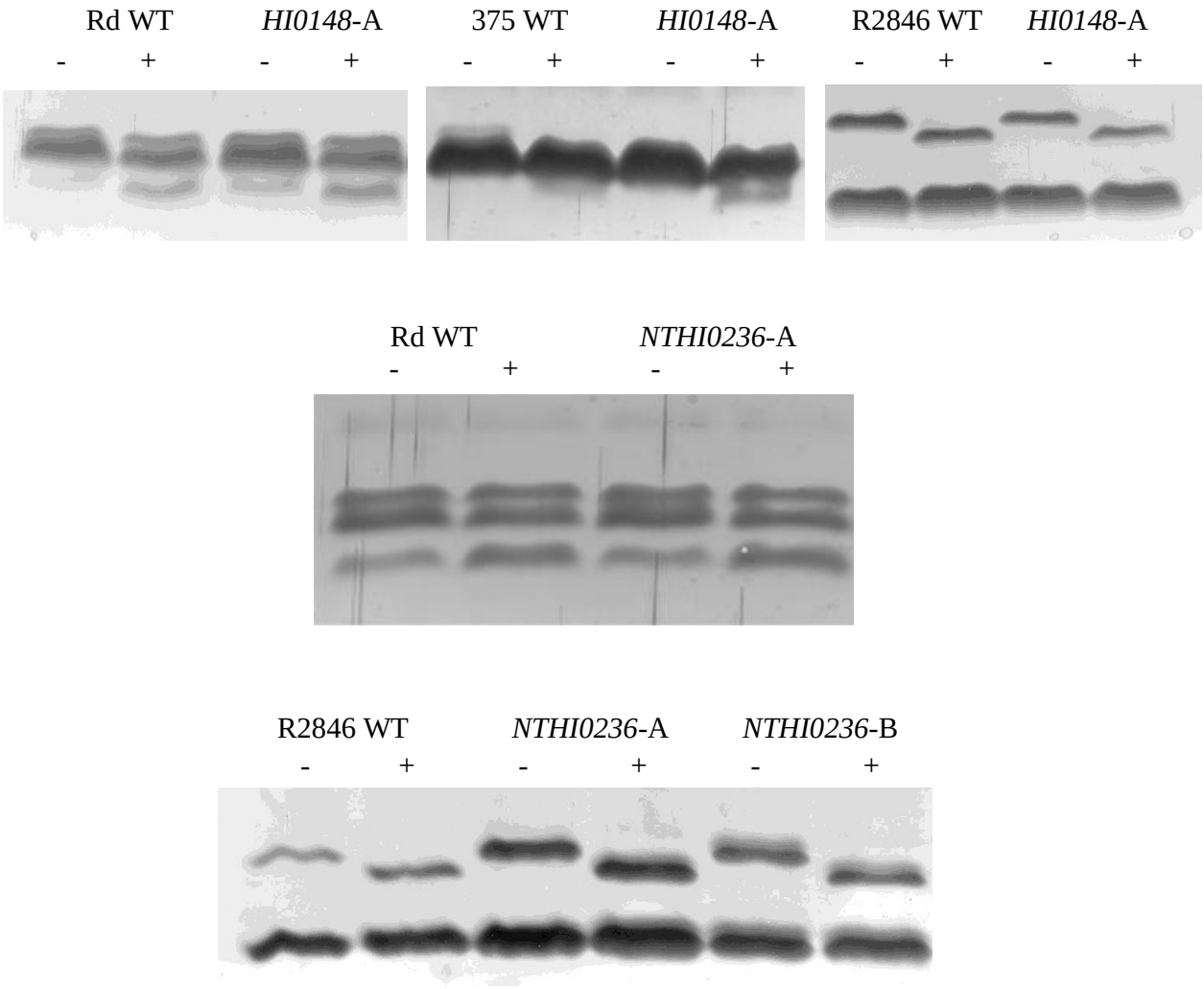


Fig. 26 LPS profiles of the Rd, 375 and R2846 wild type (WT), and respective *HI0148*- and *NTHI0236*- mutant strains.
LPS glycoforms of two R2846 *NTHI0236*- mutants (R2846 *NTHI0236*-A and -B) are shown.
- and + indicate without and with neuraminidase treatment, respectively.

The LPS of one *HI0227*-, two *HI0228*-, and two *HI0227/HI0228*- mutant strains from *Hi* 375 were analysed (Fig. 27). The 375 *HI0227*-A mutant exhibited a similar LPS profile as the wild type strain (Fig. 27A). However, the 375 *HI0228*-A mutant had an apparent dramatic effect upon LPS sialylation (Fig. 27). The 375 *HI0228*-A mutant was exhibiting disialylated LPS glycoform as both the top and middle bands were removed upon neuraminidase treatment (Fig. 27 panels A and B). To eliminate the possibility that other independent phase variable events were responsible for this LPS profile, we sequenced the repeat tracts of the phase variable *lic3A*, *lic3B* and *lgtC* genes in this mutant. Contrary to the wild type strain, the phase variable *lic3B* gene was in-frame, and the *lgtC* gene was out-of-frame in the 375 *HI0228-A* mutant (see section 1.2.4, Fig. 3 and Fig. 27). Lic3B is responsible for the production of disialylated LPS glycoforms (Fig. 3 and Fig. 27) (Fox et al., 2006), and LgtC is the glycosyl transferase that functions to transfer an α-galactose onto the Hep III lactose extension (Fig. 3).

lic3B as well as *lic3A* (another sialyltransferase gene which adds monosialic acid onto the lactosyl moiety off HepIII), are phase variable genes that contain CAAT tetranucleotide repeats (see section 1.2.4). The CAAT repeats can change in number, placing two possible translational start codons of each gene in- or out-of-frame. We shall refer to the two possible translational start codons of *lic3B* and *lic3A* as ATG1 and ATG2, which are respectively the start of the longer or the shorter amino acid sequences, i.e. ATG1 is further upstream than ATG2. It has been previously suggested that ATG2 in *lic3A* is associated with a stronger start signal (Szabo et al., 1992).

Prior to neuraminidase treatment, disialylated LPS glycoform was observed for the 375 *HI0228*-A mutant. This required *lic3B* to be in-frame. Indeed, sequencing the repeat tract of *lic3B* in the 375 *HI0228*-A mutant showed that *lic3B* was in-frame, and that Lic3B was

translated using ATG1 (Fig. 27). Since a lower molecular weight band was observed for the 375 *HI0228*-A mutant following neuraminidase treatment, which corresponded to a 3Hex-3Hep LPS glycoform, this would require *lgtC* to be out-of-frame. Sequencing the repeat tract of *lgtC* in this mutant confirmed that *lgtC* was indeed out-of-frame (Fig. 27). Without a functioning LgtC, Gal cannot compete with Neu5Ac for attachment onto the lactosyl extension off HepIII. Due to the out-of-frame *lgtC* and the in-frame *lic3B* gene, fully sialylated LPS glycoforms could be achieved in the 375 *HI0228*-A mutant (Fig. 3 and Fig. 27).

The LPS profile of the 375 *HI0227/HI0228*-A mutant prior to neuraminidase treatment showed a single band in panel A of Fig. 27, but two bands in panel B. This was likely due to under-staining of the LPS gel in panel A. Following neuraminidase treatment, a lower band, which corresponded to a 3Hex-3Hep LPS glycoform, was observed for the 375 *HI0227/HI0228*-A mutant (Fig. 27). Sequencing the repeat tract of *lic3A*, *lic3B* and *lgtC* in this mutant showed that *lic3B* and *lgtC* were both out-of-frame, but *lic3A* was in-frame and was translated using the ATG2 start codon (Fig. 27B). This indicated that the two bands observed for the 375 *HI0227/HI0228*-A mutant prior to neuraminidase treatment corresponded to the sialyllactosyl and the lactosyl LPS extensions, while the single band observed following neuraminidase treatment corresponded to the lactosyl LPS extension (Fig. 3 and Fig. 27).

Analyses of the LPS of two additional 375 mutants (the 375 *HI0228*-B and *HI0227/HI0228*-B mutants) were conducted. LPS profiles of these two mutants showed similar LPS phenotypes as the 375 wild type strain before and after neuraminidase treatment (Fig. 27 panel B). Sequencing the repeat tracts of *lic3A*, *lic3B* and *lgtC* in these two mutants showed that *lic3A* and *lgtC* were both in-frame. However, contrary to the wild type strain whereby *lic3A* was translated using

ATG2, *lic3A* of both the 375 *HI0228*-B and *HI0227/HI0228*-B mutants was translated using ATG1 (Fig. 27B).

LPS from the 375 *lgtC* mutant was included in Fig. 27 as a molecular weight marker. A lower molecular weight glycoform with respect to the 375 wild type strain was observed for the 375 *lgtC* mutant, and was due to the inability of the mutant to add α-Gal in a 1,4 linkage to the lactosyl extension off HepIII (Fig. 3).

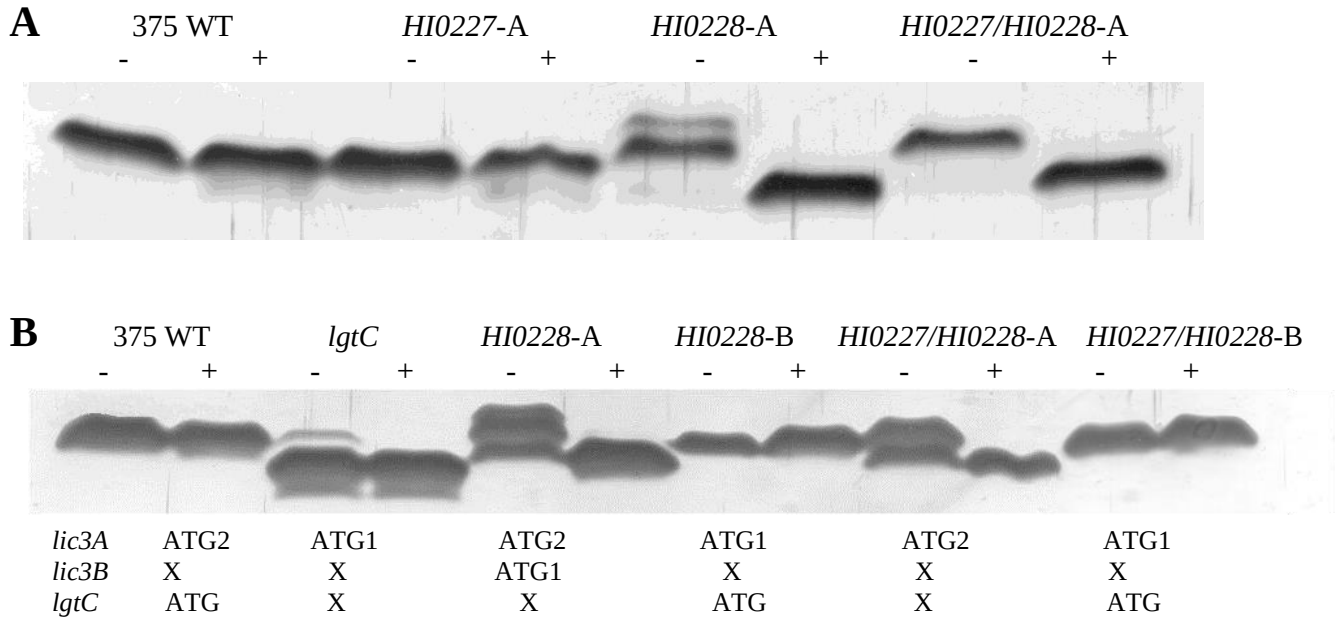


Fig. 27 LPS profiles of *Hi* 375 wild type (WT), *HI0228-*, and *HI0227/HI0228-* mutant strains.
Shown in panel A are LPS from 375 WT, *HI0227*-A, *HI0228*-A and *HI0227/HI0228*-A mutant strains.
Shown in panel B are LPS from 375 WT, *lgtC* mutant, *HI0228*-A, -B, *HI0227/HI0228*-A, -B mutant strains. The 375 *lgtC* mutant was included as a known molecular weight marker. Indicated at the bottom of panel B is whether the phase variable genes *lic3A, lic3B* and *lgtC* were in- or out-of-frame. ATG indicates that the phase variable genes were in-frame, and X indicates that the genes were out-of-frame. ATG1 and ATG2 indicate the more upstream or downstream start codons of *lic3A* and *lic3B*, respectively.

4.6 SBA of the *HI0227*-, *HI0228*-, *HI0227/HI0228*-, *HI0148*- and *NTHI0236*- mutant strains

It has previously been shown that an increase in LPS sialylation is closely linked to enhanced bacterial ability to resist serum complement mediated killing (Bouchet et al., 2003). Although we could not detect any clear difference in LPS sialylation based upon LPS phenotypes in our *HI0227*-, *HI0228*-, *HI0227/HI0228*-, *HI0148*- and *NTHI0236*- mutant strains; however, SBA was still performed as it can be more sensitive in detecting LPS sialylation and other subtle changes in LPS glycoform structure. SBA was performed on selected mutant strains from section 4.6 (whose LPS profiles are shown in Fig. 24, Fig. 25, Fig. 26, Fig. 27).

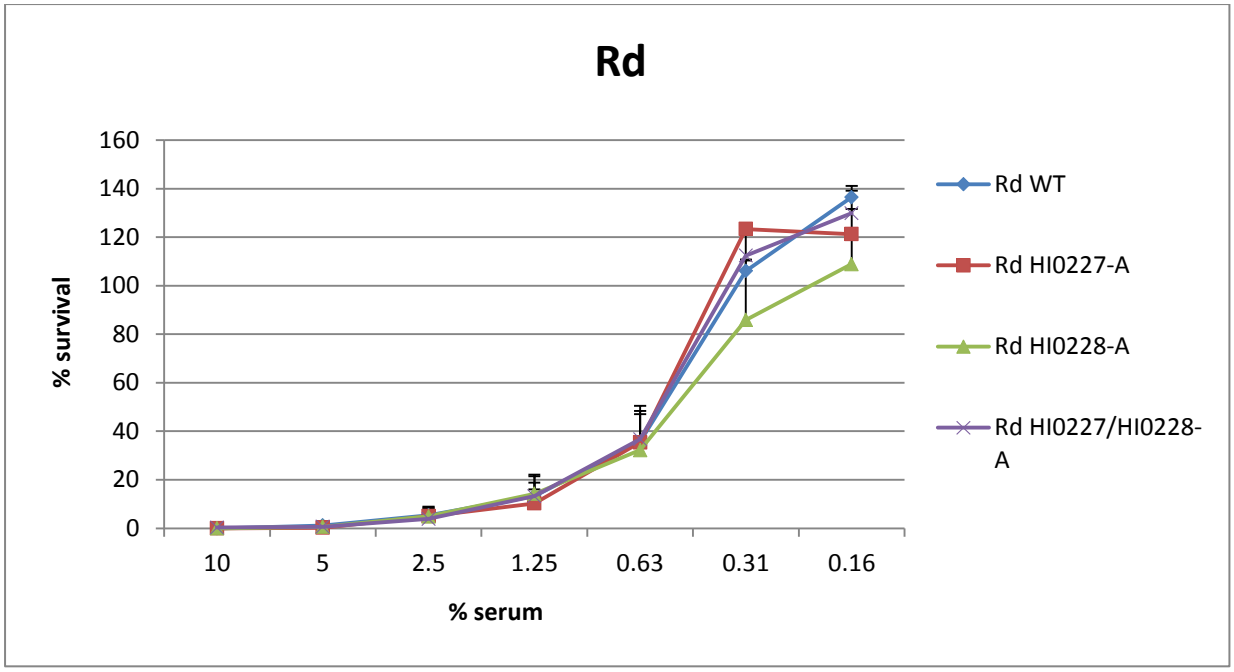
Using the SBA, serum resistance of the *HI0227*-, *HI0228*-, and *HI0227/HI0228*- mutants from strains Rd, R2846 and 723 was similar to their respective parent strains (Fig. 28). Furthermore, the Rd *HI0148*- mutants (-A and –B), the R2846 *HI0148*- mutants (-A and -B), and the R2846 *NTHI0236*-A mutant also showed a similar ability to resist the killing effect of pooled human sera as their respective parent strains (Fig. 29). These results obtained from SBA correlated to LPS phenotypes, where a similar level of LPS sialylation was observed between these mutant and the wild type strains (Fig. 24, Fig. 25, Fig. 26).

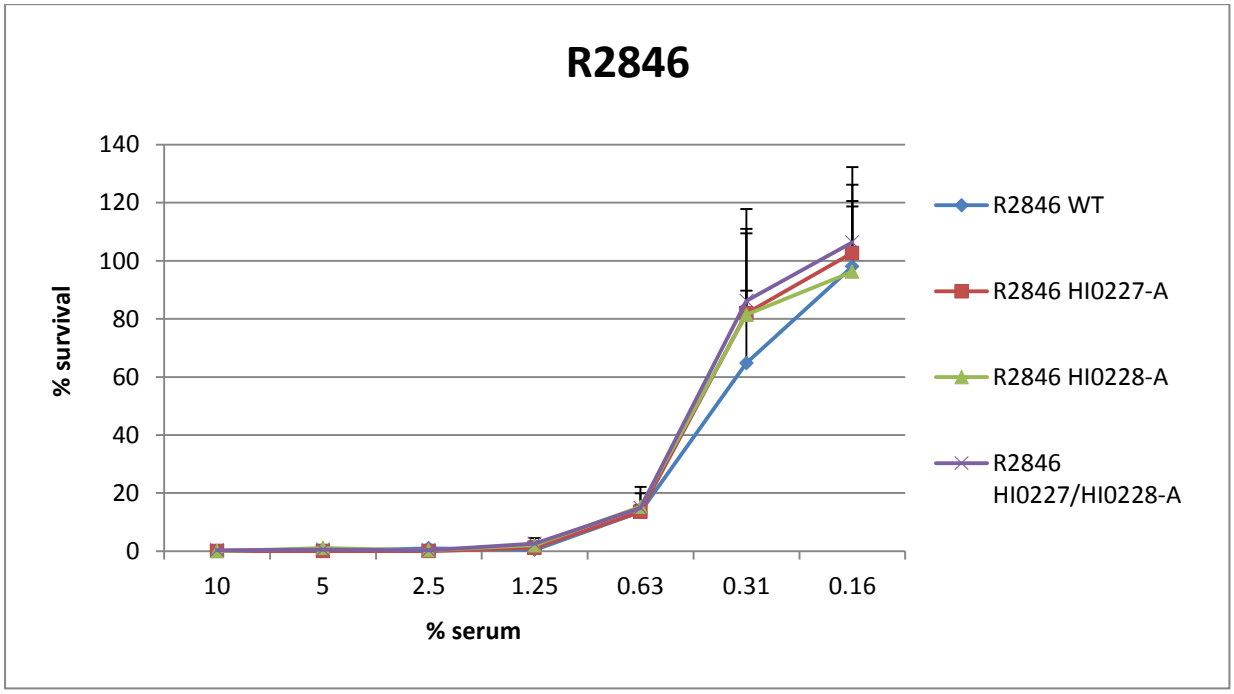
It is interesting to note that an increase in serum sensitivity was observed for the 375 *HI0227*-A, *HI0228*-B, and *HI0227/HI0228*-B mutants compared to the 375 wild type strain (evidenced by the shifting of the curves to the right in Fig. 28), even though the LPS of these mutants did not show any change in profile from the LPS of the 375 wild type strain (Fig. 27). It was not obvious what caused the changes in these mutants' ability to resist the killing effect of pooled human sera. It may be that small differences in LPS profiles exist for these mutants, which could not be detected by our LPS gel silver staining, which underlie the SBA changes. Examples of such

small LPS differences could be the presence or absence of phase variable PCho and O-acetyl groups. Further detailed structural analyses would be required to confirm whether or not these small LPS differences did indeed exist.

Contrary to the above, an increase in serum resistance was observed for the 375 *HI0228*-A mutant compared to the 375 wild type strain (evidenced by the shifting of the curve to the left in Fig. 28). Although the increase in serum resistance observed for the 375 *HI0228*-A mutant was consistent with the disialylated [α-Gal(1–4)-β-Gal(1–4)-] containing LPS phenotype exhibited by this mutant (Fig. 27); however, it is surprising that the magnitude of increase in serum resistance was not as great as one might expect from the significantly enhanced sialylation profile of the 375 *HI0228*-A mutant. It may be that the absence of the digalactoside containing LPS glycoform in the 375 *HI0228*-A, due to the out-of-frame *lgtC*, itself influenced the ability to resist complement mediated killing, as the presence of digalactoside containing LPS glycoform has been shown to be also implicated in virulence (Griffin *et al.*, 2005).

Similarly, an increase in serum resistance was observed for the 375 *HI0227/HI0228*-A mutant compared to the 375 wild type strain (evidenced by the shifting of the curve to the left in Fig. 28). From the LPS profile and the sequencing of the repeat tract of *lgtC* in the 375 *HI0227/HI0228-A* mutant, we have already established that *lgtC* in this mutant was out-of-frame. Because *lgtC* was out-of-frame in the 375 *HI0227/HI0228-A* mutant, this mutant was unable to synthesise digalactoside containing LPS glycoform, which predisposed the mutant to produce more sialylated LPS glycoform. With an increase in sialylated LPS glycoform, the 375 *HI0227/HI0228-A* mutant would become more serum resistant consistent with the observed in the SBA in Fig. 28.


Rd
160
140
120
100
80
60
40
20
0
% survival
10
5
2.5
1.25
0.63
0.31
0.16
% serum
Rd WT
Rd HI0227-A
Rd HI0228-A
Rd HI0227/HI0228-A


R2846
140
120
100
80
60
40
20
0
% survival
10
5
2.5
1.25
0.63
0.31
0.16
% serum
R2846 WT
R2846 HI0227-A
R2846 HI0228-A
R2846 HI0227/HI0228-A

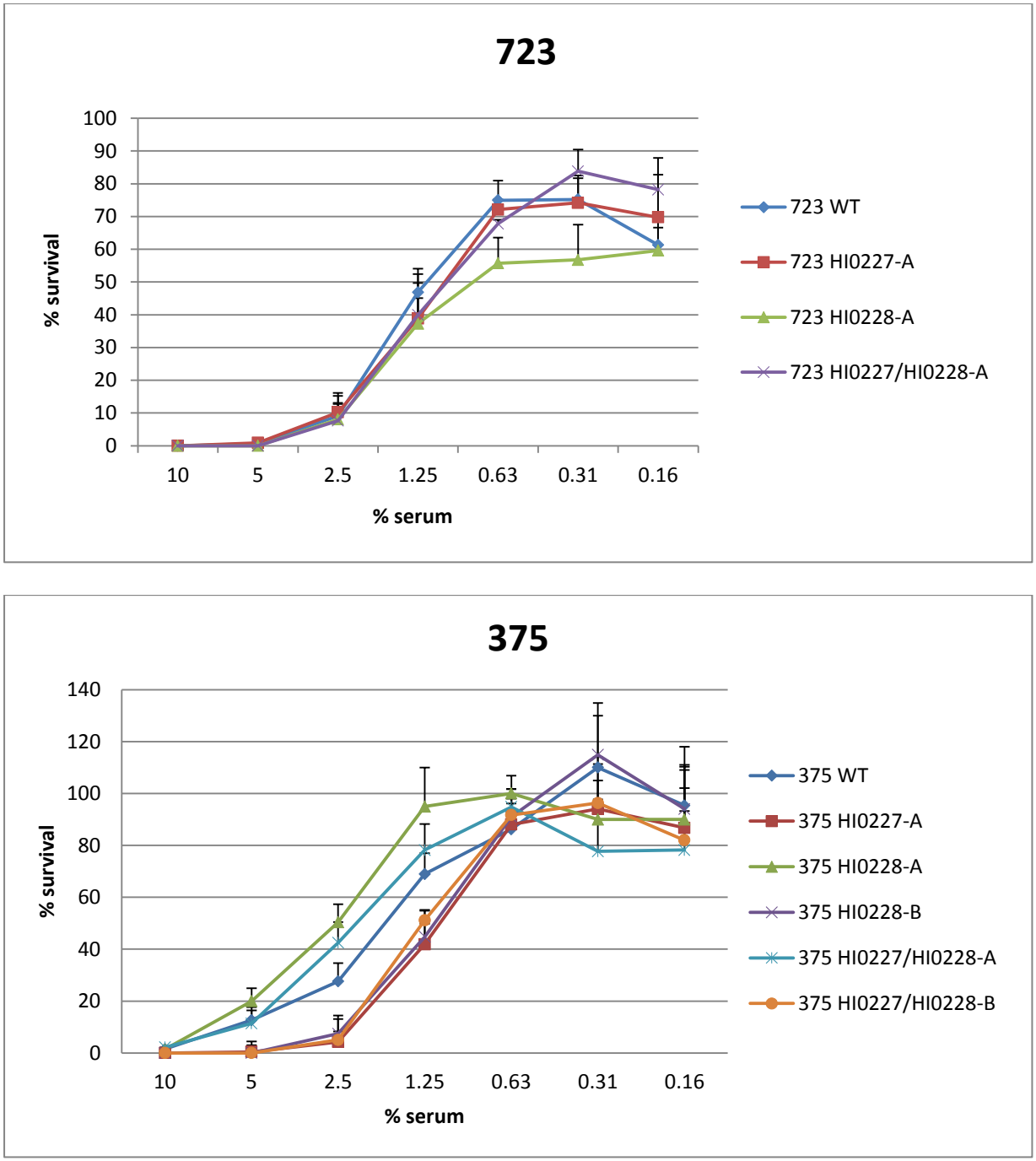


Fig. 28 SBA of the Rd, R2846, 723 and 375 wild type (WT), and their respective *HI0227-*, *HI0228-*, and *HI0227/HI0228-* mutants.
Results shown are the average of two independent experiments with standard deviations shown.

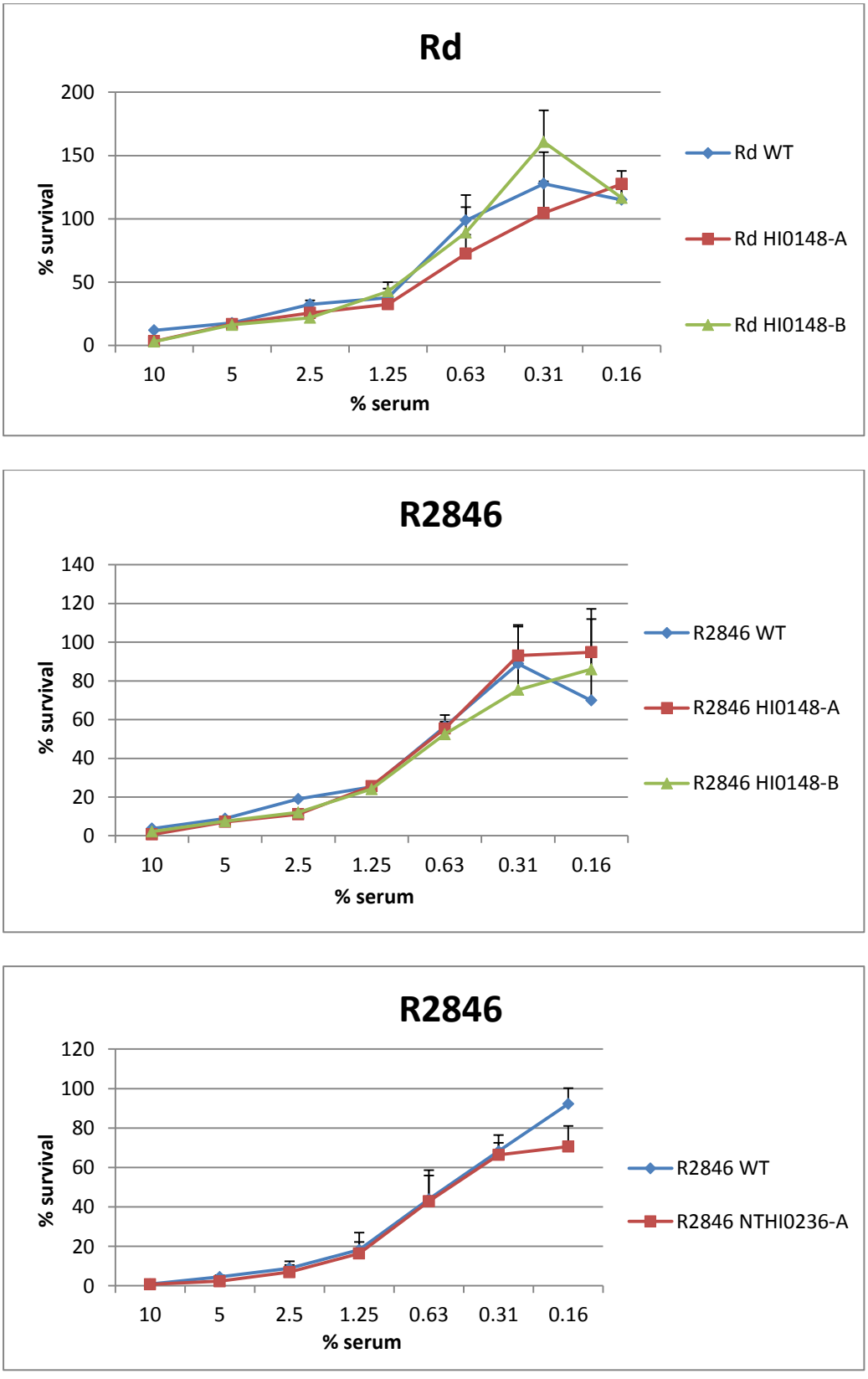


Fig. 29 SBA of the Rd and R2846 wild type (WT) and respective *HI0148*- and *NTHI0236*-mutant strains.
Results shown are the average of two independent experiments with standard deviations shown.

4.7 Discussion

Mutants that lack the function of a candidate sialometabolism gene or peripheral candidate sialometabolism gene were created in various strain backgrounds (*Hi* 375, Rd, R2846 and 723). From the experiments carried out, there is no evidence to suggest that these genes (*HI0227, HI0228, HI0148* and *NTHI0236*) are involved in a direct way in the catabolism of Neu5Ac or LPS sialylation.

From our growth assays which looked at the Neu5Ac catabolism of *Hi*, we found no evidence to suggest that the candidate sialometabolism genes and peripheral candidate sialometabolism genes are involved in either the classic or the alternative Neu5Ac catabolism pathways. In the classic Neu5Ac catabolism pathway (the Neu5Ac catabolism pathway documented for *Hi*), Neu5Ac is first converted to ManNAc by NanA, so phosphorylation of ManNAc to ManNAc-6P (the second step) by NanK could take place (see section 1.4.3 and Fig. 6). It has been noted in other organisms that an alternative Neu5Ac catabolism pathway, which bypasses the production of ManNAc-6P, and instead involved the production of GlcNAc (Fig. 6), exists. Since RokA, which encodes GlcNAc kinase (converts GlcNAc to GlcNAc-6P) and is involved in the second part of the alternative pathway is present in *Hi* (Fig. 6), it seemed plausible to predict that this alternative Neu5Ac catabolism pathway could be present in *Hi*, and that the postulated SA epimerase HI0227 could play a role in the alternative Neu5Ac catabolism pathway. However, from our growth assays we observed poorer growth of the Rd *nanK* and *nanE* mutant strains (which respectively lack the second and third steps of the classic Neu5Ac catabolism pathway) compared to the wild type strain in the presence of Neu5Ac, which indicated that particular alternative Neu5Ac catabolism pathway is not present in *Hi*, and that HI0227 does not play a role in it. Moreover, the Rd *nanK* and *nanE* mutants exhibited poorer growth in the presence of added Neu5Ac compared to the culture without Neu5Ac, which suggests that accumulation of Neu5Ac

or another Neu5Ac derivative caused a toxic effect on the growth of the Rd *nanK* and *nanE* mutants (Vimr et al., 2004) (see section 6.7).

A further possibility for HI0227 function was that under anaerobic condition, the potential to ferment Neu5Ac could occur via a shunt that is not present in aerobic condition. We performed a fermentation study to determine whether our target genes (*HI0227, HI0228, HI0148* and *NTHI0236*) play a role in the interaction of the products of the pathway. Our results showed that there was no difference in the ability to ferment Neu5Ac between our mutant and wild type strains, suggesting that the target genes are not involved in the ability to ferment Neu5Ac in *Hi*.

In general, LPS profiles and the ability to resist the killing effect of human sera in mutants that lack the function of candidate novel sialometabolism genes and peripheral candidate sialometabolism genes were no different to that of their parent strains. However, two mutants (375 *HI0228*-A and 375 *HI0227/HI0228*-A) exhibited an increase in their ability to resist the killing effect of human serum compared to the parent strain. The increase in the 375 *HI0228*-A mutant's ability to resist complement mediated killing could be explained by the disialylated LPS glycoform expressed by this mutant, due to the in-frame *lic3B* gene (translated using ATG1) (Fox et al., 2006). Sequencing of the repeat tract of *lgtC* in the 375 *HI0228*-A mutant showed that *lgtC* was out-of-frame. As LgtC is responsible for the attachment of Gal onto the lactosyl extension off HepIII (Fig. 3), creating the digalactoside [α-Gal(1–4)-β-Gal(1–4)-] containing LPS glycoform, the out-of-frame *lgtC* in the 375 *HI0228*-A would result in the lack of digalactoside containing LPS glycoform in this mutant. The combination effect of the in-frame *lic3B* and the out-of-frame *lgtC* resulted in maximal LPS sialylation in the 375 *HI0228*-A mutant. Interestingly, the magnitude of increase in serum resistance was not as great as one have predicted from the sialylation profile of the 375 *HI0228*-A mutant, and this could be explained

by the absence of the digalactoside containing LPS glycoform in the 375 *HI0228*-A mutant (due to the out-of-frame *lgtC*). The presence of digalactoside containing LPS glycoform has a role in SBA resistance and is implicated in virulence (Griffin et al., 2005).

An increase in serum resistance was also observed for the 375 *HI0227/HI0228*-A mutant compared to the *Hi* 375 wild type strain. Similarly, this would be explained by the out-of-frame *lgtC* in this mutant, which renders the mutant unable to synthesise the digalactoside containing LPS glycoform, and predisposes the mutant to produce more sialylated LPS glycoform. With an increase in sialylated LPS glycoform, the mutant would become more serum resistance in the SBA as observed (Fig. 28).

Certain mutant strains showed similar LPS profiles to their respective wild type strains, but exhibited at different level of resistance against the killing effect of human serum. These mutant strains were: 375 *HI0227*-A, 375 *HI0228*-B, 375 *HI0227/HI0228*-B, and Rd *NTHI0236*-A. It is likely that small differences in LPS structures (such as the presence of PCho or O-acetyl groups), which are below the level of sensitivity for detecting small changes in molecular weight in our LPS gels, were present in these mutant strains and such small LPS changes can cause differences in the level of resistance against the killing effect of human serum.

In this chapter, we have assessed various phenotypic characteristics relating to sialometabolism in our mutants that lack the function of candidate and peripheral sialometabolism genes. We observed no major differences in the phenotypes tested in our mutants when compared to the respective wild type strain. The lack of any phenotypic change in these mutants may be due to the assays used not being sensitive enough to detect all phenotypic changes. Despite the lack of evidence of sialometabolism related function for HI0227 and HI0148, it is possible that they may

be involved in some biochemical process that influences the regulation of the *Hi* sialometabolism operons as there remain aspects regarding the details of the *Hi* regulation of sialometabolism operons that are uncertain. We therefore investigated the potential roles of these genes on the expression of other sialometabolism genes (see section 5.1).

5 Investigation of a role of candidate novel sialometabolism genes in the *Hi* sialometabolism gene expression

We have established in the previous chapter that the candidate novel sialometabolism genes were not directly involved in growth or LPS phenotypes associated with sialometabolism under the conditions tested. Since the identity and provision of the SIS sugar had not been elucidated when our study was initiated, it was unknown whether HI0227 or HI0148 could be involved in biochemical processes that may influence the regulation of the *Hi* sialometabolism operons. In this chapter, we addressed whether HI0227 and HI0148 could affect gene expression of the *Hi* sialometabolism operons. During these studies, we found that both the effect of time of exposure and concentration of Neu5Ac influence gene expression, we refer to this as the temporal/concentration effect of Neu5Ac.

5.1 Candidate novel sialometabolism genes do not affect the expression of the *Hi* sialometabolism operons and Neu5Ac has a temporal/concentration effect on the expression of the *Hi* sialometabolism operons

As previously mentioned in section 1.4.4, the SiaR regulator protein of *Hi* possesses a SIS sugar binding domain (Johnston et al., 2007, Bateman, 1999). Binding of a sugar to the SIS domain could alter the regulatory property of a regulator protein (Bateman, 1999). We shall refer to the sugar that interacts with the SIS domain of SiaR as the SIS sugar. Considering that the identity of the SIS sugar remains unclear, HI0227 and HI0148 may play some role in providing the exact conformation of sugar that is required to interact with SiaR to modify its function. HI0227 was postulated to function as a SA epimerase (Teplyakov et al., 2005). It is possible that HI0227 has

the potential to influence the provision of the SIS sugar that interacts with SiaR. HI0148 was postulated to function as a Neu5Ac mutarotase, whose function is to interconvert the anomeric configuration of Neu5Ac (see section 1.4.1) (Severi et al., 2008). Should Neu5Ac be the SIS sugar that interacts with SiaR, the availability of the Neu5Ac isomer (α- or ß-Neu5Ac) that interacts with SiaR would be dependent upon the action of the Neu5Ac mutarotase.

The expression of *nanE* and *siaP* (see Fig. 7) was used to represent the *Hi nagnan* and the *siaPQM* operons, respectively. The two genes are the first to be transcribed in their respective operons, and are also the most studied genes for this purpose. *Hi* strain Rd was selected for our investigation as this strain has been used to characterise sialometabolism gene expression in previous studies (Jenkins et al., 2010).

We performed RT-PCR as a preliminary assessment to determine whether the loss of the HI0227 or the HI0148 functions affects the expression of the *Hi* sialometabolism operons. Our RT-PCR results showed that compared to the cultures with no sugar supplementation, there was downregulation of *nanE* and *siaP* in the Rd wild type, the Rd *HI0227*-A and the Rd *HI0148*-A mutant strains when grown in the presence of 30 μg/ml Neu5Ac, added from the start of culture (growth for approximately three hours) (Fig. 30). Neu5Ac was used at a concentration of 30 μg/ml in line with previous studies (Jenkins et al., 2010).

Using qPCR for quantitating results, we found that *nanE* was respectively downregulated by 1.7, 2.3 and 2.1 fold in the Rd wild type, *HI0227*-A and *HI0148*-A mutant strains following growth in the presence of 30 μg/ml Neu5Ac added from the start of the culture, compared to cultures with no sugar supplementation (Fig. 31). *siaP* was also downregulated by 1.6, 1.9 and 2.7 fold respectively in the Rd wild type, *HI0227*-A and *HI0148*-A mutant strains when 30 μg/ml of

Neu5Ac was added from the start of the culture (Fig. 31). These findings are consistent with previous published data on the Rd wild type strain, which showed that the expression of *nanE* and *siaP* were downregulated by approximately two- to three fold when grown in the presence of 30 μg/ml Neu5Ac added from the start of the culture (Jenkins et al., 2010). Since the expression of *nanE* and *siaP* were similar in the Rd *HI0227*- and *HI0148*- mutant strains compared to the Rd wild type, this indicates that HI0227 and HI0148 do not affect sialometabolism gene expression following growth in the presence of 30 μg/ml Neu5Ac added from the start of the culture (Fig. 30, Fig. 31).

It is interesting to note from our qPCR results that the relative expression of *nanE* was less than the expression of *siaP* by approximately 35 fold (Fig. 31). As such differential gene expression of *nanE* and *siaP* was not observed in our RT-PCR results (Fig. 30), it is likely that it was due to an artefact caused by the primers that were used. We suspect that the primer pair used to amplify *nanE* in qPCR does not work as efficiently as the primer pair for *siaP*, resulting in the differential expression of the *Hi* sialometabolism operons.

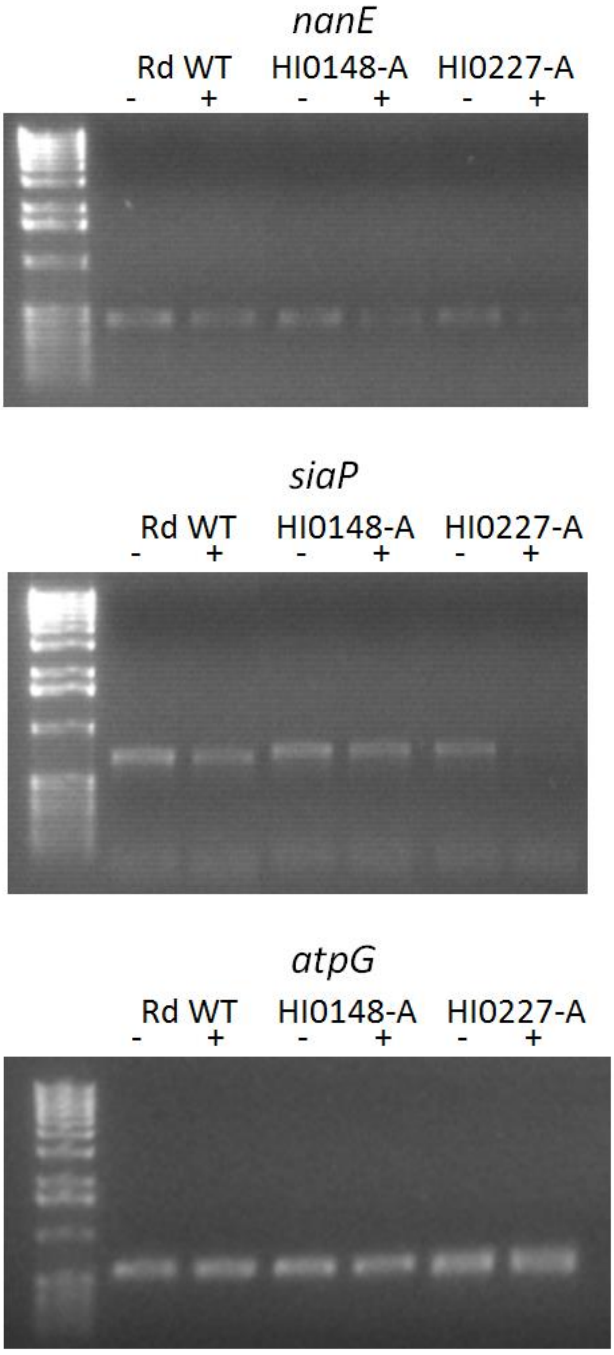


Fig. 30 RT-PCR products for the *nanE*, *siaP* and *atpG* genes in the Rd wild type, the *HI0148*-A, and the *HI0227*-A mutant strains without (-) or with (+) 30 µg/ml of Neu5Ac added from the start of the culture.
nanE and *siaP* respectively represent the *Hi nagnan* and the *siaPQM* operons. Even band intensities were observed in the *atpG* reference gene.

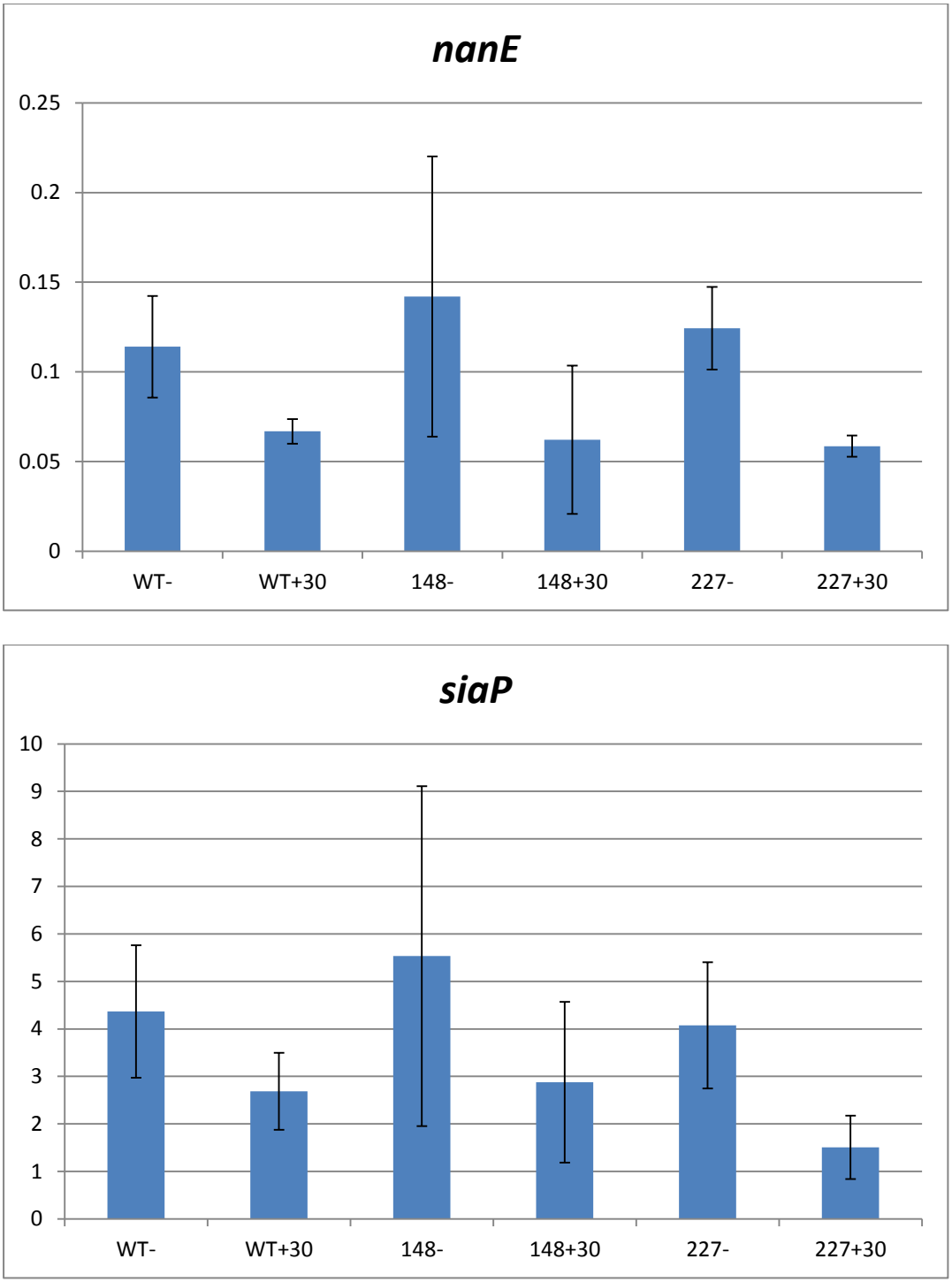
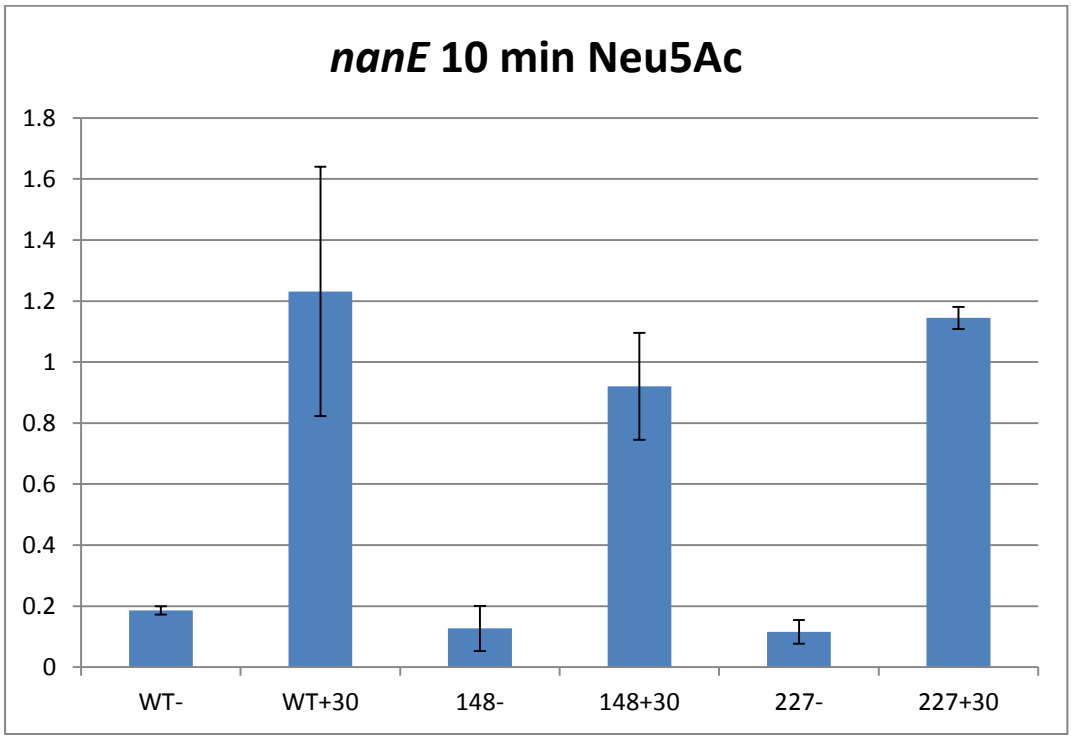


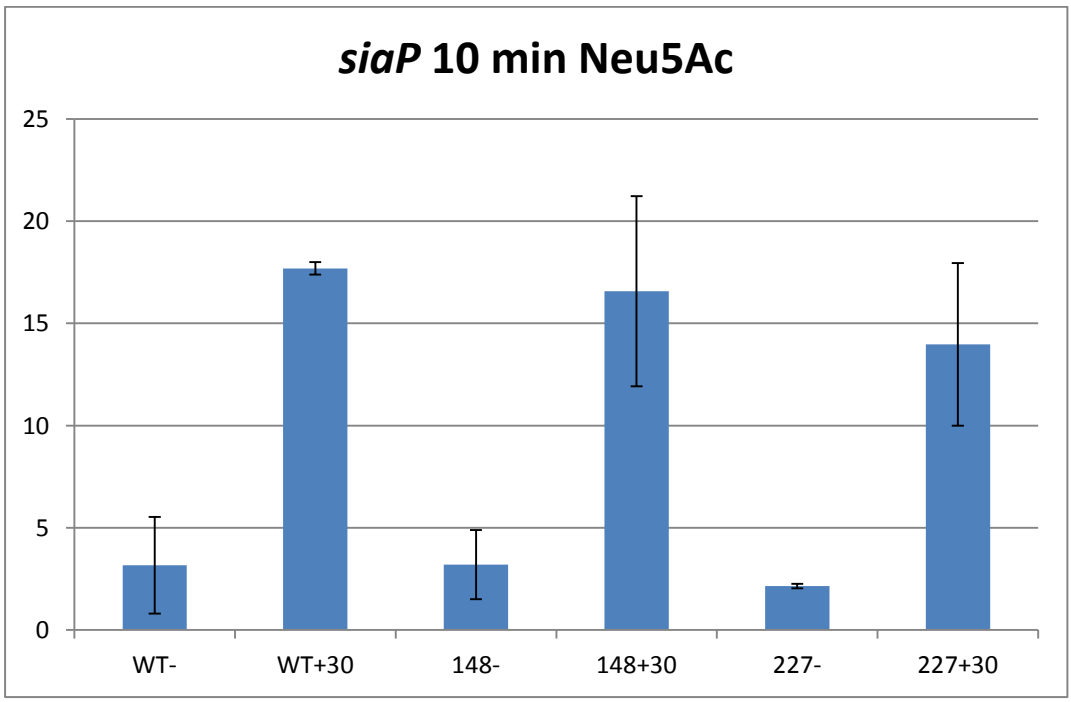
Fig. 31 qPCR analyses of *nanE* and *siaP* gene expression in the Rd wild type (WT), *HI0148*-A (148), and *HI0227*-A (227) mutant strains following growth in the absence (-) or presence (+) of 30 µg/ml Neu5Ac added from the start of the culture.
Represented in the y-axis is the fold change in mRNA amount relative to the *frdB* reference gene. qPCR results were the average of two biological replicates with standard deviations, and each biological replicate included two technical replicates.

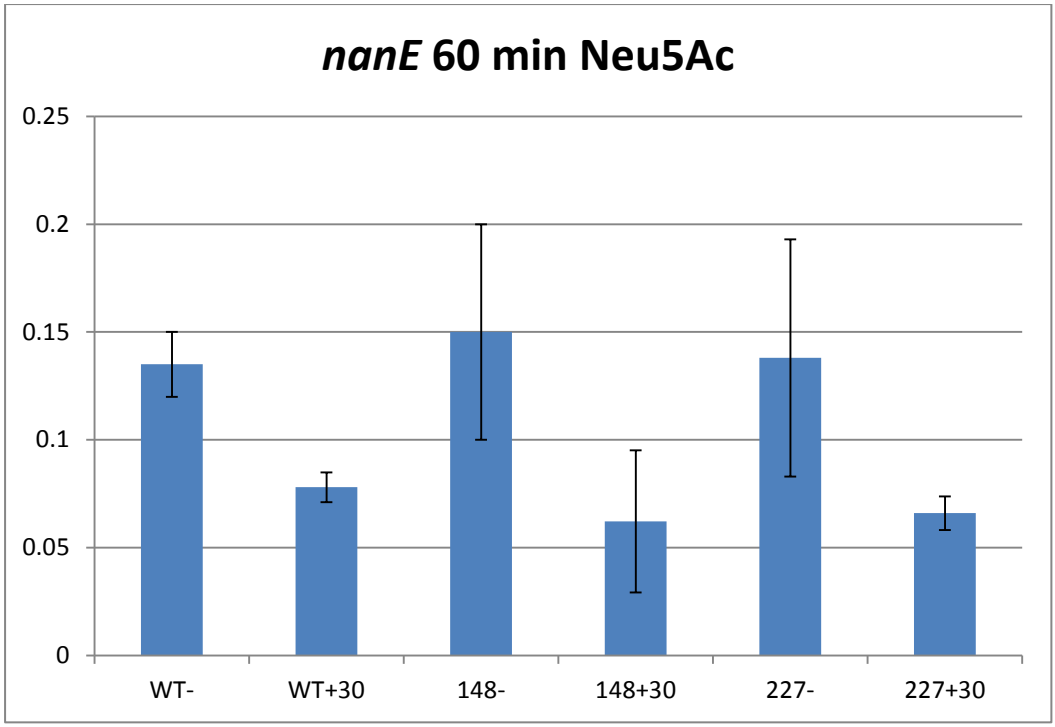
Although the expression of *nanE* and *siaP* was not changed in the Rd *HI0227*-A and *HI0148*-A mutants compared to the wild type strain following growth in the presence of 30 μg/ml Neu5Ac added from the start of the culture; however, it remained a possibility that any effect on the expression of *nanE* and *siaP* could be transient. We therefore examined the expression of *nanE* and *siaP* in the mutants compared to the wild type strains when grown in 30 μg/ml Neu5Ac for 10 min or 60 min. Expression of *nanE* and *siaP* obtained from the two shorter time points (10 min and 60 min) following growth in Neu5Ac were compared to growth in Neu5Ac added from the start of the culture.

We discovered that the expression of *nanE* and *siaP* were upregulated in the Rd wild type and mutant strains when grown in the presence of 30 μg/ml Neu5Ac for 10 min (Fig. 32). *nanE* was upregulated by 6.6, 9.9 and 7.2 folds respectively in the Rd wild type, *HI0227*- and *HI0148*- mutant strains (Fig. 32); and *siaP* was upregulated by 5.6, 6.5 and 5.1 folds respectively in the Rd wild type, *HI0227*- and *HI0148*- mutant strains (Fig. 32). This finding indicated that Neu5Ac can exert a temporal effect on the expression of the *Hi* sialometabolism operons.

Furthermore, the expression of *nanE* and *siaP* in the Rd wild type, *HI0227*- and *HI0148*- mutant strains following growth in Neu5Ac at 30 μg/ml for 60 min were downregulated compared to the cultures that had received no supplementation (Fig. 32): *nanE* was downregulated by 1.7, 2.1 and 2.4 fold respectively in the Rd wild type, *HI0227*- and *HI0148*- mutant strains, while *siaP* was downregulated by 1.9, 2.6 and 1.9 fold respectively in the Rd wild type, *HI0227*- and *HI0148*- mutant strains. The expression level of *nanE* and *siaP* obtained from growth in 30 μg/ml Neu5Ac for 60 min were similar to growth in 30 μg/ml Neu5Ac added from the start of the culture (Fig. 31, Fig. 32).


nanE 10 min Neu5Ac
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
WT-
WT+30
148-
148+30
227-
227+30


siaP 10 min Neu5Ac
25
20
15
10
5
0
WT-
WT+30
148-
148+30
227-
227+30



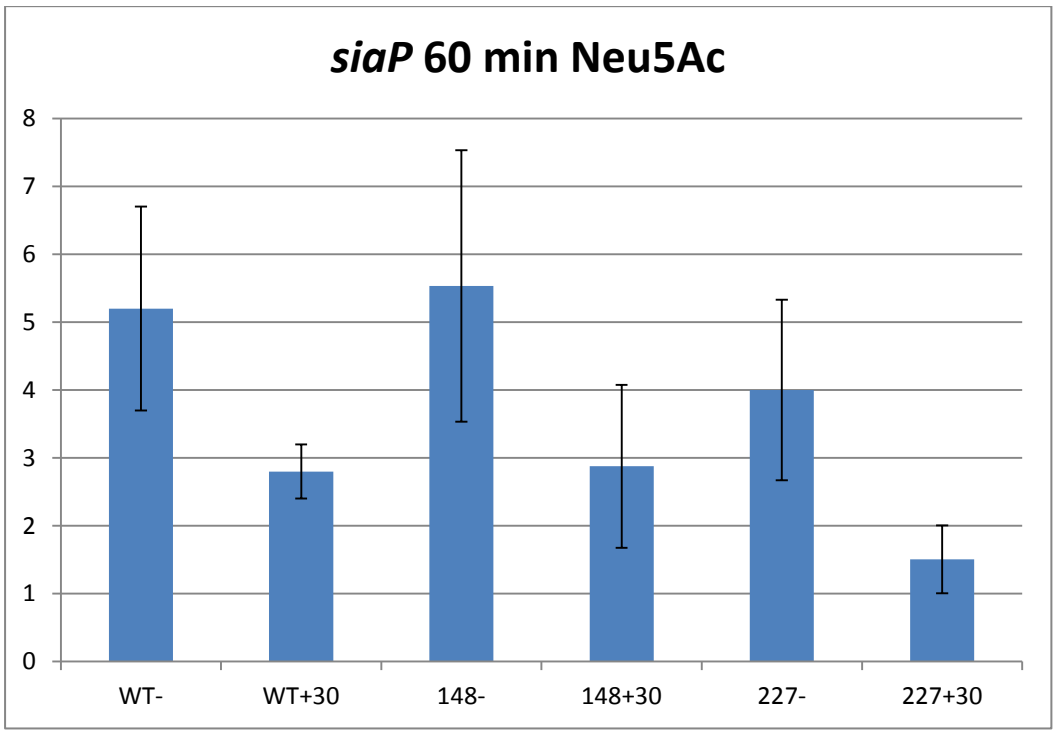


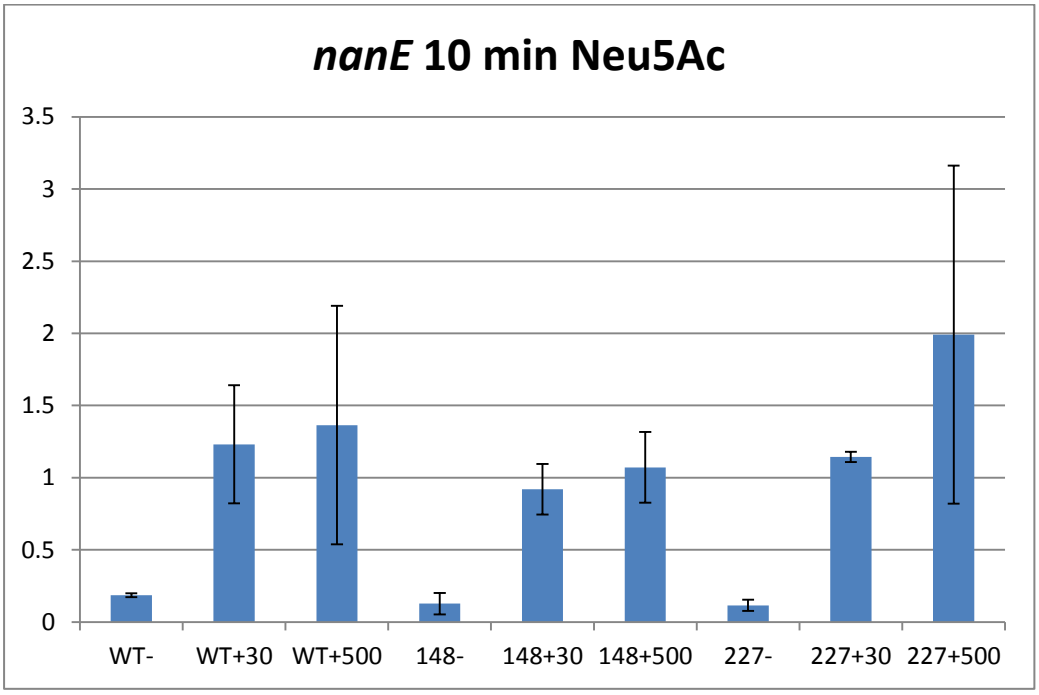
Fig. 32 qPCR analyses of *nanE* and *siaP* gene expression in the Rd wild type (WT), *HI0148*-A (148), and *HI0227*-A (227) mutant strains following growth in the absence (-) or presence (+) of 30 µg/ml Neu5Ac added for 10 or 60 min.
Represented in the y-axis is the fold change in mRNA amount relative to the *frdB* reference gene. qPCR results shown were the average of two biological replicates with standard deviations, and each biological replicate included two technical replicates.

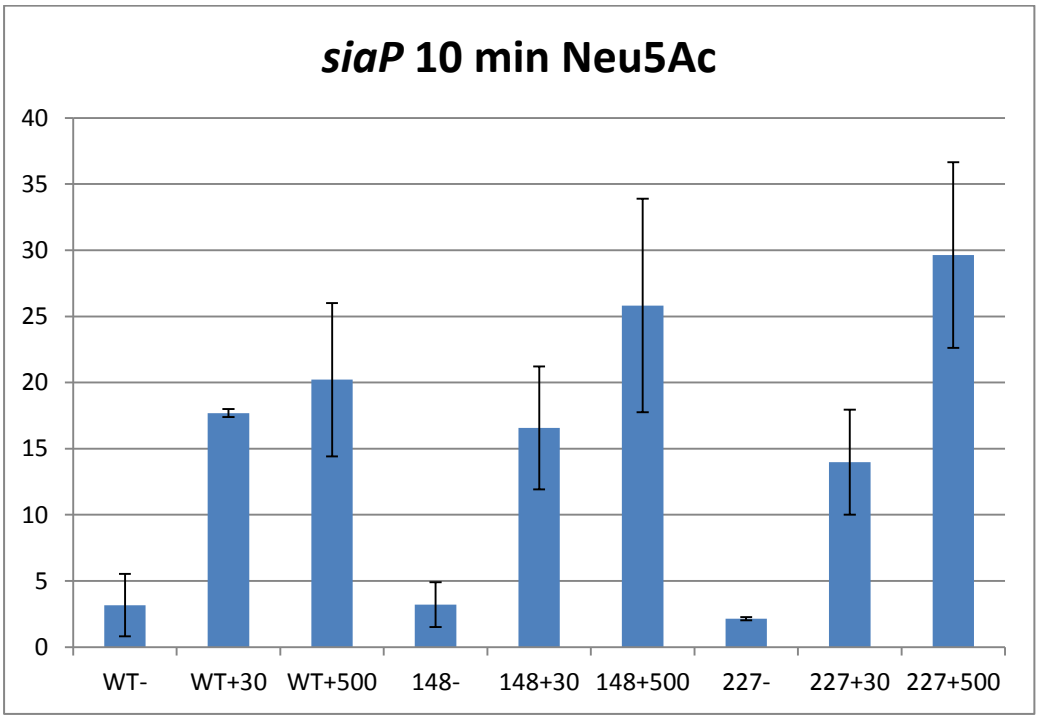
In section 4.4, we showed in strain Rd that growth of *Hi* had a reduced doubling time in the presence of 500 μg/ml Neu5Ac compared to the presence of 30 μg/ml Neu5Ac (Fig. 17). The increased growth rate of strain Rd in the presence of 500 μg/ml Neu5Ac infers that growth is perpetuated in that culture, which could likely result in enhanced gene expression for a longer period of time compared to the culture supplemented with 30 μg/ml Neu5Ac. Although it was difficult to detect small changes in our growth assays; however we would expect that the culture supplemented with 30 μg/ml Neu5Ac would have an enhanced early exponential phase (similar to the culture supplemented with 500 μg/ml Neu5Ac) and reduced growth rate for the rest of the growth.

We also decided to check whether the expression of *nanE* and *siaP* would be different in the Rd *HI0227*-A and *HI0148*-A mutant strains compared to the wild type strain when grown in the presence of this relatively higher concentration (500 μg/ml) of Neu5Ac. Similar to growth in the presence of 30 μg/ml Neu5Ac for 10 min, the expression of *nanE* and *siaP* in the Rd wild type and mutant strains were also upregulated following growth in the presence of 500 μg/ml Neu5Ac for 10 min: *nanE* was upregulated by 7.3, 17.2 and 8.4 fold respectively in the Rd wild type, *HI0227*- and *HI0148*- mutant strains; while *siaP* was upregulated by 6.4, 13.8 and 8.0 fold respectively in the Rd wild type, *HI0227*- and *HI0148*- mutant strains (Fig. 33A).

In contrast to growth in the presence of 30 μg/ml Neu5Ac for 60 min, the expression of *nanE* and *siaP* in the Rd wild type, *HI0227*- and *HI0148*- mutant strains following growth in 500 μg/ml Neu5Ac for 60 min remained upregulated: *nanE* was upregulated by 10.3, 7.4 and 8.8 fold respectively in the Rd wild type, *HI0227*- and *HI0148*- mutant strains; while *siaP* was upregulated by 3.3, 5.9 and 4.7 fold respectively in the Rd wild type, *HI0227*- and *HI0148*- mutant strains (Fig. 33B). This interesting effect of Neu5Ac concentration, which is likely due to

the availability of Neu5Ac over time, further signifies the importance of Neu5Ac temporal effect on the regulation of the *Hi* sialometabolism operon. We shall refer to this effect as the temporal/concentration effect of Neu5Ac.





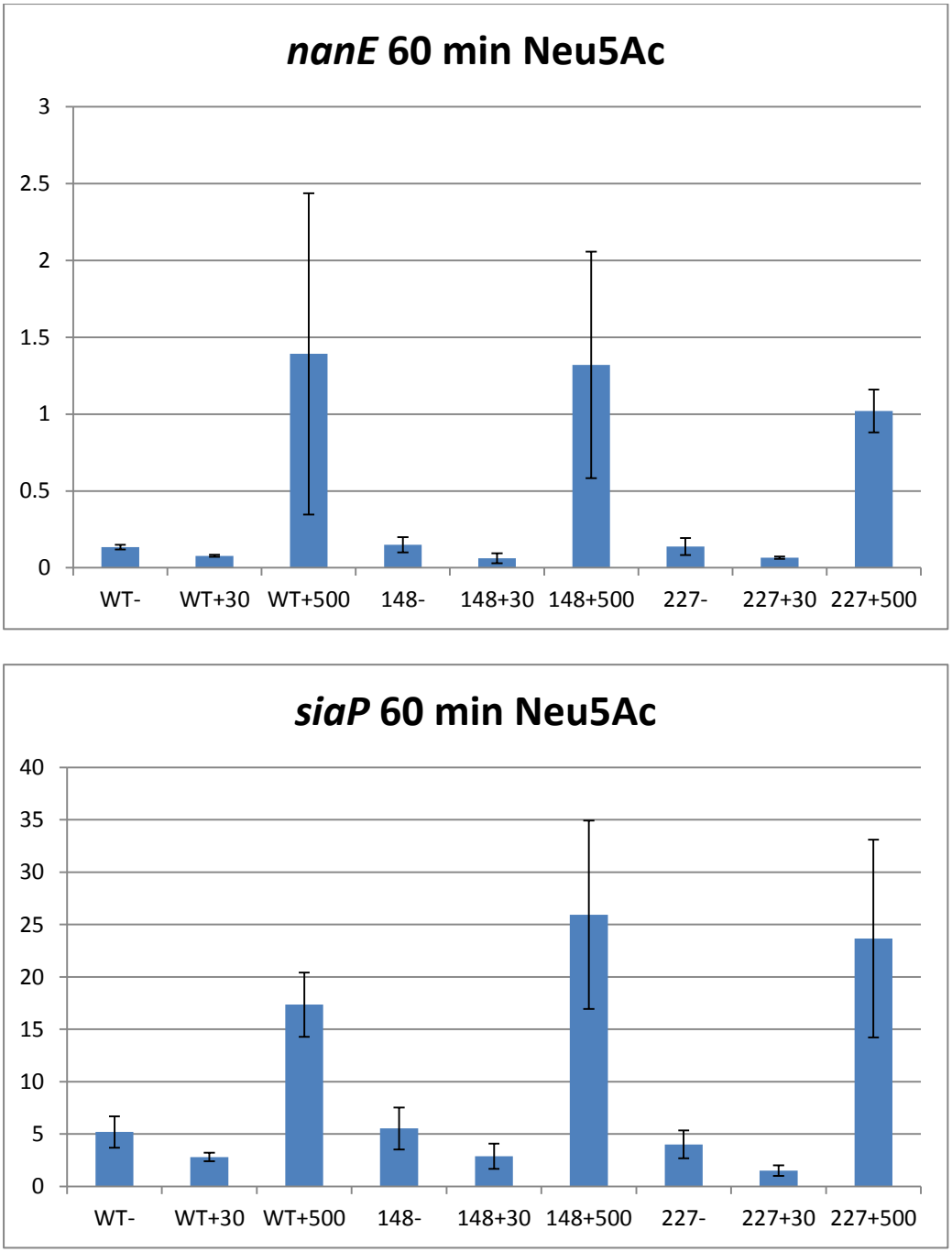


Fig. 33 qPCR analyses of *nanE* and *siaP* gene expression in the Rd wild type (WT), *HI0148*-A (148), and *HI0227*-A (227) mutant strains following growth in the absence (-) or presence (+) of 30 µg/ml (30) or 500 µg/ml (500) Neu5Ac for 10 or 60 min.
Represented in the y-axis is the fold change in mRNA amount relative to the *frdB* reference gene. qPCR results shown were the average of two biological replicates with standard deviations, and each biological replicate included two technical replicates.

5.2 Discussion

In this chapter, we investigated the effect of inactivating HI0227 and HI0148 on the expression of the *Hi* sialometabolism genes. The *nanE* and *siaP* transcripts were used to represent the *Hi nagnan* and *siaPQM* operons, respectively. RT-PCR and qPCR analyses showed that the expression of *nanE* and *siaP* in the Rd *HI0227*-A and Rd *HI0148*-A mutant strains were similar to the wild type strain following growth in the presence of 30 μg/ml Neu5Ac added from the start of the culture (growth for approximately three hours). Our result obtained from the *HI0148*-mutant was consistent with previous finding which showed that no significant change in sialometabolism gene expression was observed when the Rd *HI0148* gene was mutated (Jenkins et al., 2010).

Previous study had shown a two- to three-fold downregulation of *nanE* and *siaP* in strain Rd following growth in 30 μg/ml Neu5Ac added from the start of the culture (Jenkins et al., 2010). Since it has never been made clear why this downregulation occurs, we felt it was necessary to look at the expression of *nanE* and *siaP* in strain Rd following growth in Neu5Ac for shorter time periods (10 min and 60 min) and at a higher concentration (500 μg/ml). Assessment of *nanE* and *siaP* under these new conditions will help to provide further insight into the effect of Neu5Ac over time on the expression of the *Hi nagnan* and *siaPQM* operons.

Our results obtained from RT-PCR and qPCR analyses showed a finding that has never been previously reported - that the availability of Neu5Ac over time affects the expression of the *Hi* sialometabolism operon, an effect we refer to as the temporal/concentration effect of Neu5Ac. This was evidenced by the upregulation of *nanE* and *siaP* in the Rd wild type and mutant strains following growth in 30 μg/ml Neu5Ac for 10 min, downregulation of *nanE* and *siaP* in these strains following growth in 30 μg/ml Neu5Ac for 60 min, and continued upregulation of *nanE*

and *siaP* in these strains following growth in 500 μg/ml Neu5Ac for 60 min (due to the continued availability of Neu5Ac).

We believe that the up- and downregulation of *nanE* and *siaP* observed following growth in CDM with different times and concentrations of Neu5Ac were due to changes in the activity of the SiaR regulator protein. It is likely that the concentration of Neu5Ac affects the availability of the SIS sugar, which in turn modulates the activity of the SiaR protein. The amount of SiaR may also be important and that the intricate balance between the availability of the SIS sugar and the amount of SiaR likely determines the expression level of the *nagnan* and *siaPQM* operons. Our proposed mechanism for the *Hi* sialometabolism gene regulation will be outlined in the next chapter

The expression of *nanE* and *siaP* in the *HI0228*- mutant was not studied because *HI0228* is likely a pseudogene, which is unlikely to show an effect on the *Hi* sialometabolism (see chapter 4). Assessment of *nanE* and *siaP* expression was also not conducted in the *NTHI0236*- mutant because in our opinion, NTHI0236 is unlikely to affect the expression of *nanE* and *siaP* as the loss of the NTHI0236 function did not affect any of the tested aspects of *Hi* sialometabolism. NTHI0236 is also considered to be a less strong candidate of sialometabolism gene through its position (see chapter 4).

We have established that the time of bacteria grown in, and the concentration of, Neu5Ac exert an effect on the expression of *nagnan* and *siaPQM*. We want to gather as much information as we could about the nature of the SIS sugar and its role in the interaction with SiaR to modulate the function of the protein, which will be one of the focuses in the next chapter.

6 Further evaluation on the regulation of the *Hi* sialometabolism operons

To confirm that the temporal/concentration effect of Neu5Ac is a general trend of *Hi*, we assessed the expression of *nanE* and *siaP* in a different strain background- *Hi* 2019, which was chosen because it has been extensively used by others in the study of *Hi* sialometabolism and gene regulation (Johnston et al., 2007, Johnston et al., 2010). Previous study has indicated that GlcN-6P is the SIS sugar for SiaR in *Hi* (Johnston et al., 2010). However, since there are details surrounding this that remain unclear, we addressed these issues through investigation of a number of mutants that affect the anabolic and catabolic processes of Neu5Ac in strain Rd.

In previous study, the magnitude of induction of *nanE* and *siaP* using the 2019 *cyaAnagB* double mutant (a strain that cannot synthesise CyaA hence the effect of cAMP is abolished, and accumulates GlcN-6P following growth in Neu5Ac) was significantly higher than our finding in strain Rd (Johnston et al., 2010). We felt it was necessary to confirm this using the Rd and 2019 *cyaAnagB* double mutants created by us.

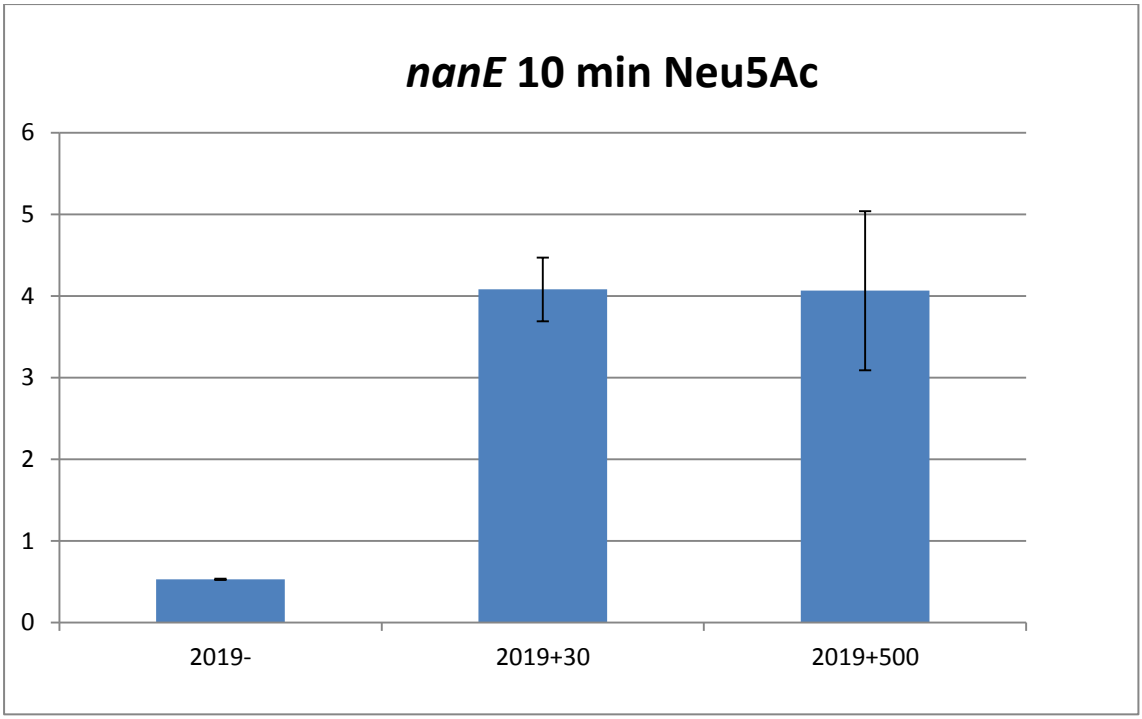
We have determined that the amount of Neu5Ac and time of exposure affect sialometabolism gene expression over short periods of time and we want to investigate whether this could reflect in LPS phenotype. We have chosen strain R2846 for this investigation because of its enhanced LPS sialylation profile.

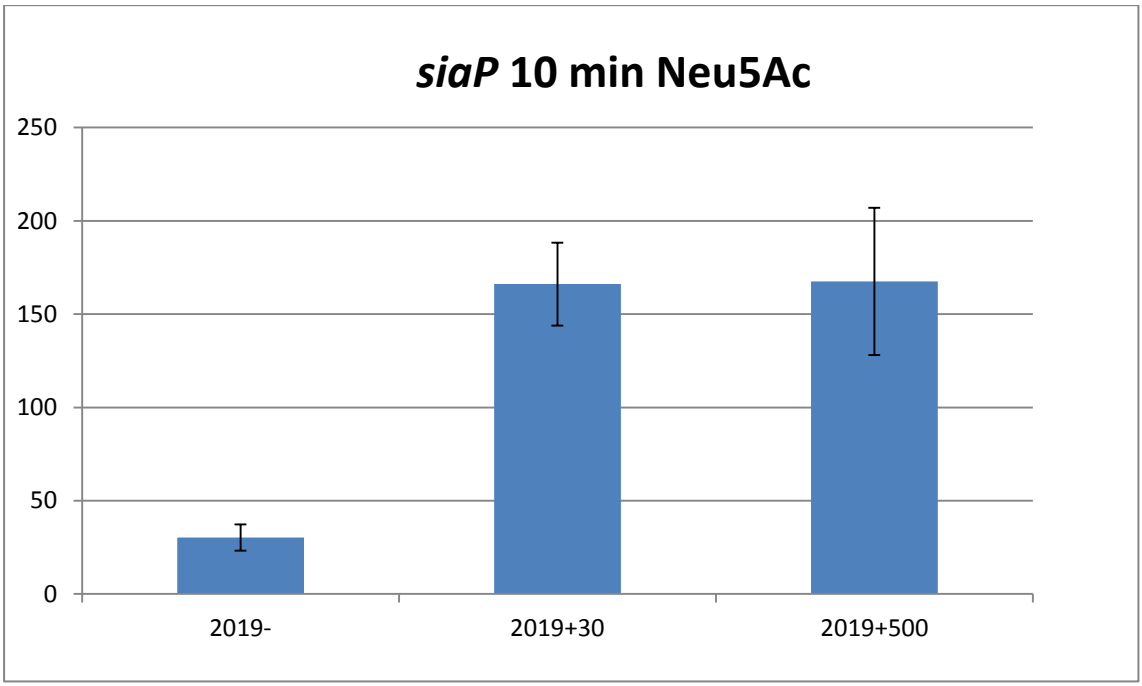
6.1 The temporal/concentration effect of Neu5Ac on the expression of *nanE* and *siaP* is strain independent

It is likely that the temporal/concentration effect of Neu5Ac is a general trend of *Hi* strains. To confirm this, we assessed whether or not the expression of *nanE* and *siaP* in *Hi* strain 2019 would be affected by the temporal/concentration effect of Neu5Ac.

The expression of *nanE* and *siaP* in *Hi* 2019 was upregulated by 7.7 and 5.5 fold, respectively, following growth in Neu5Ac at 30 µg/ml for 10 min compared to the culture with no supplementation (Fig. 34). Similar to strain Rd, this effect of Neu5Ac was largely reversed when *Hi* 2019 was grown in the presence of 30 µg/ml Neu5Ac for 60 min, i.e. compared to the culture with no sugar supplementation, the expression of *nanE* and *siaP* became downregulated by 2.1 and 2.0 fold respectively following growth in the presence of 30 µg/ml of Neu5Ac for 60 min (Fig. 34). The level of *nanE* and *siaP* downregulation observed in strain 2019 was similar to previous published results for strain Rd (Jenkins et al., 2010). Furthermore, similar to strain Rd, *nanE* and *siaP* remained upregulated following growth in Neu5Ac at a higher concentration (500 µg/ml) for 60 min compared to the culture with no supplementation (Fig. 34). The fold inductions were 8.3 and 5.8 fold respectively for *nanE* and *siaP* (Fig. 34).

The expression of *nanE* and *siaP* in *Hi* 2019 were up- and downregulated in a similar fashion to strain Rd (see above), which indicates that the dynamic control of Neu5Ac on the expression of the *Hi nagnan* and *siaPQM* operons is likely a general phenomenon of *Hi* strains.


nanE 10 min Neu5Ac
6
5
4
3
2
1
0
2019-
2019+30
2019+500


siaP 10 min Neu5Ac
250
200
150
100
50
0
2019-
2019+30
2019+500

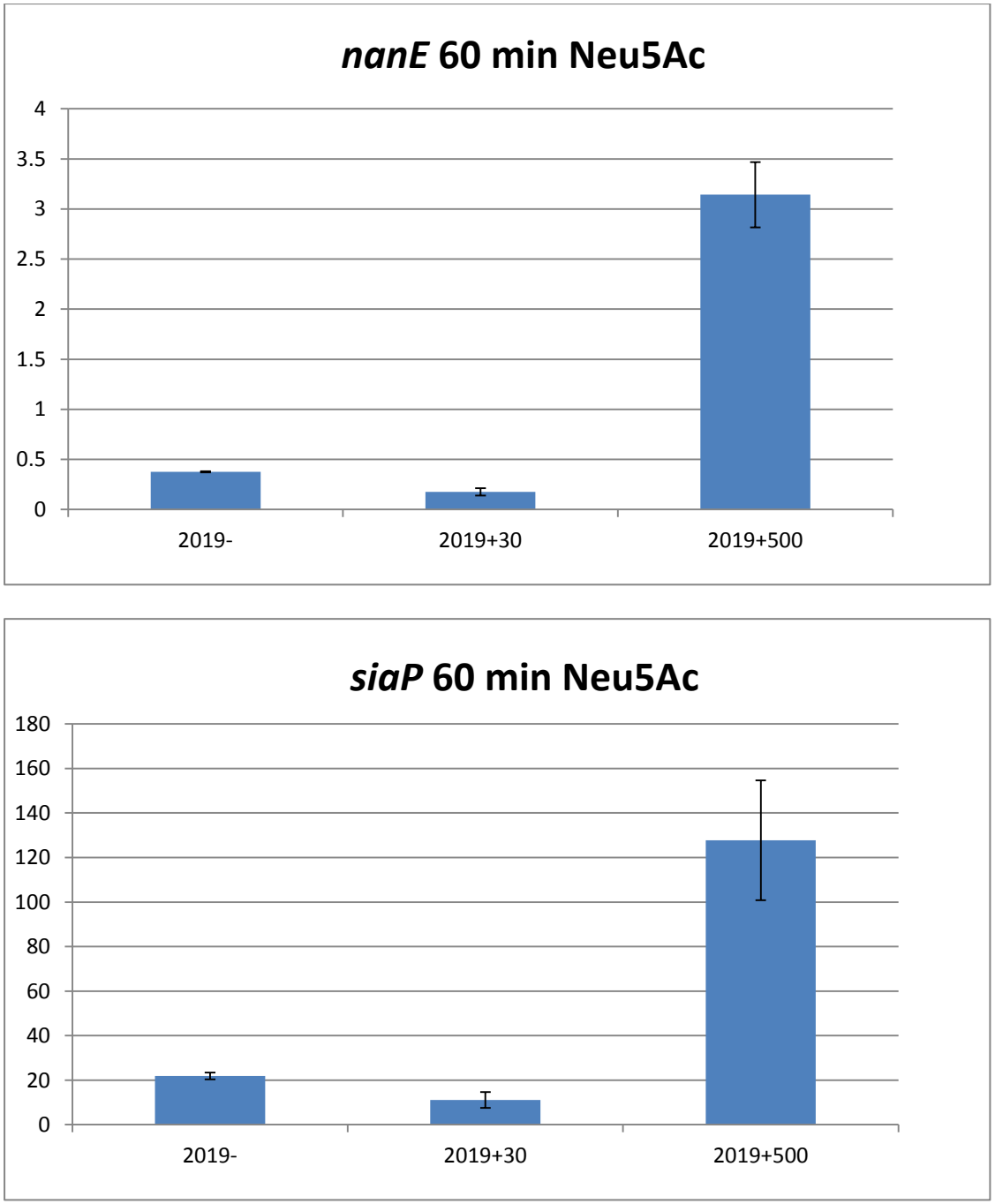


Fig. 34 qPCR analyses of *nanE* and *siaP* gene expression in *Hi* 2019 wild type strain following growth in the absence (-) or presence (+) of 30 µg/ml (30) or 500 µg/ml (500) Neu5Ac for 10 or 60 min.

Represented in the y-axis is the fold change in mRNA amount relative to the *frdB* reference gene. qPCR results were the average of two biological replicates with standard deviations, and each biological replicate included two technical replicates.

6.2 The CMP-Neu5Ac biosynthesis pathway does not affect the expression of *nanE* and *siaP*

The *Hi* SiaR repressor protein belongs to the RpiR transcriptional regulator family and contains a HTH DNA binding domain and a SIS sugar binding domain (see section 1.4.4). It has been shown in many repressors proteins belonging to the same family that the interaction of the repressor with the SIS sugar prevents the binding of the repressor to its operator site (see section 1.4.4.). Although previous study suggests that GlcN-6P (derived from Neu5Ac catabolism, Fig. 6) is the SIS sugar that interacts with SiaR (Johnston et al., 2010); however, we felt that the identity of the SIS sugar has not been rigorously tested and that the postulated function of GlcN-6P as a co-activator for SiaR, which facilitates the binding of SiaR to its operator site, is atypical for a SIS sugar (Johnston et al., 2010). In this section and the next, we reassessed whether an alternative sugar (Neu5Ac, another Neu5Ac catabolite, or CMP-Neu5Ac) could substitute GlcN-6P as the SIS sugar for SiaR.

Although it has not been shown before that a sugar nucleotide (e.g. CMP-Neu5Ac) could function as a SIS sugar, previous study has suggested that the removal of Neu5Ac through its conversion to CMP-Neu5Ac prevented the upregulation of sialometabolism genes (Johnston et al., 2007). To determine whether CMP-Neu5Ac affects sialometabolism gene expression directly, we examined the expression of *nanE* and *siaP* in the Rd *siaB* mutant (where CMP-Neu5Ac could not accumulate as SiaB is the CMP-Neu5Ac synthetase) following growth in 30 µg/ml Neu5Ac for 10 min. We have done so using RT-PCR.

Our RT-PCR results showed that similar to the Rd wild type strain, the expression of *nanE* and *siaP* in the Rd *siaB* mutant was upregulated by several fold following growth in 30 µg/ml Neu5Ac for 10 min (Fig. 35). Furthermore, the expression of *nanE* and *siaP* in the Rd *siaB*

mutant following growth in 30 µg/ml Neu5Ac for 60 min was again similar to the Rd wild type strain: both genes were downregulated by several fold (Fig. 35). Confirmation and quantification of the up- and downregulation of *nanE* and *siaP* were carried out using qPCR (Fig. 36): *nanE* was upregulated by 4.9 and 3.5 fold respectively in the Rd wild type and *siaB* mutant strains, and *siaP* was upregulated by 2.3 and 1.6 fold respectively in the Rd wild type and *siaB* mutant strains following growth in 30 µg/ml Neu5Ac for 10 min (Fig. 36). Furthermore, the expression of *nanE* following growth in 30 µg/ml Neu5Ac for 60 min was downregulated by 2.2 and 3.3 fold respectively in the Rd wild type and *siaB* mutant strains, and *siaP* was downregulated by 4.2 and 1.9 fold respectively in the Rd wild type and *siaB* mutant strains (Fig. 36). The up- and downregulation of *nanE* and *siaP* observed in the Rd *siaB* mutant were similar to the levels observed for the Rd wild type strain (Fig. 36 and see section 5.1). This data indicates that anabolism of Neu5Ac does not affect the expression of *Hi* sialometabolism genes under the conditions tested.

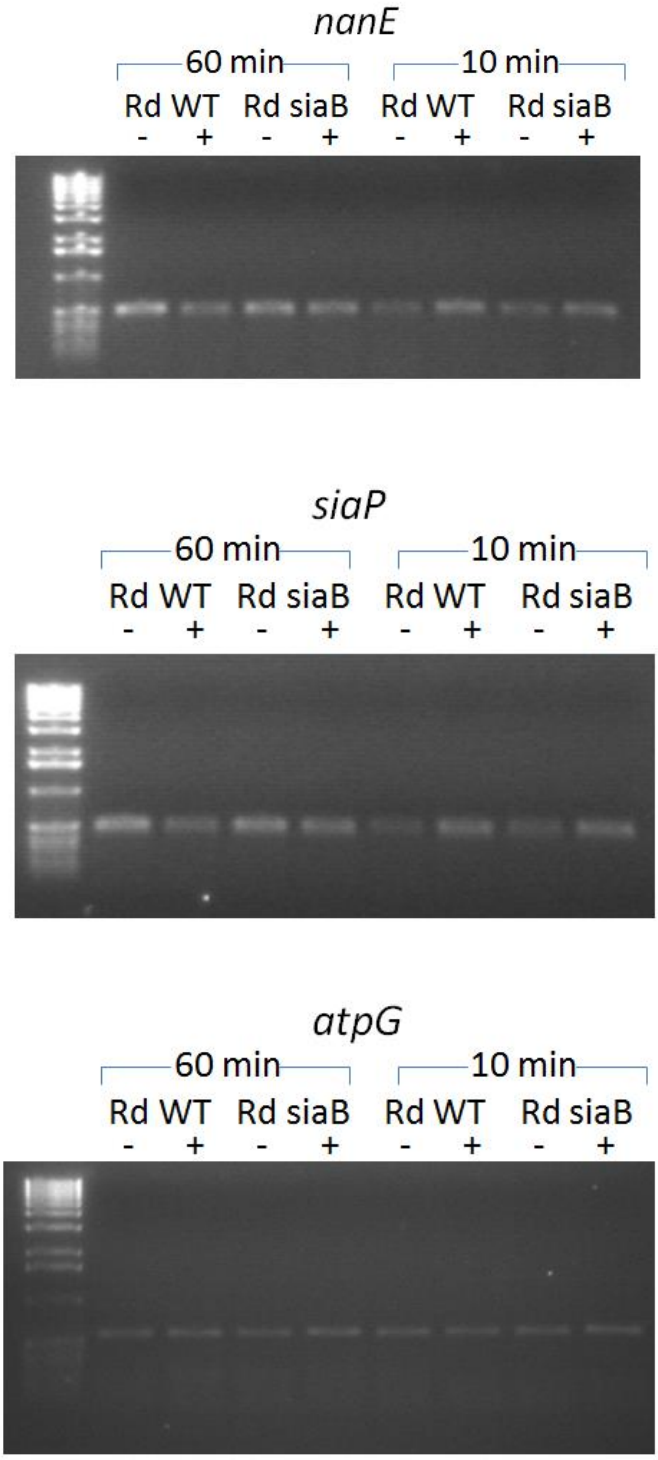
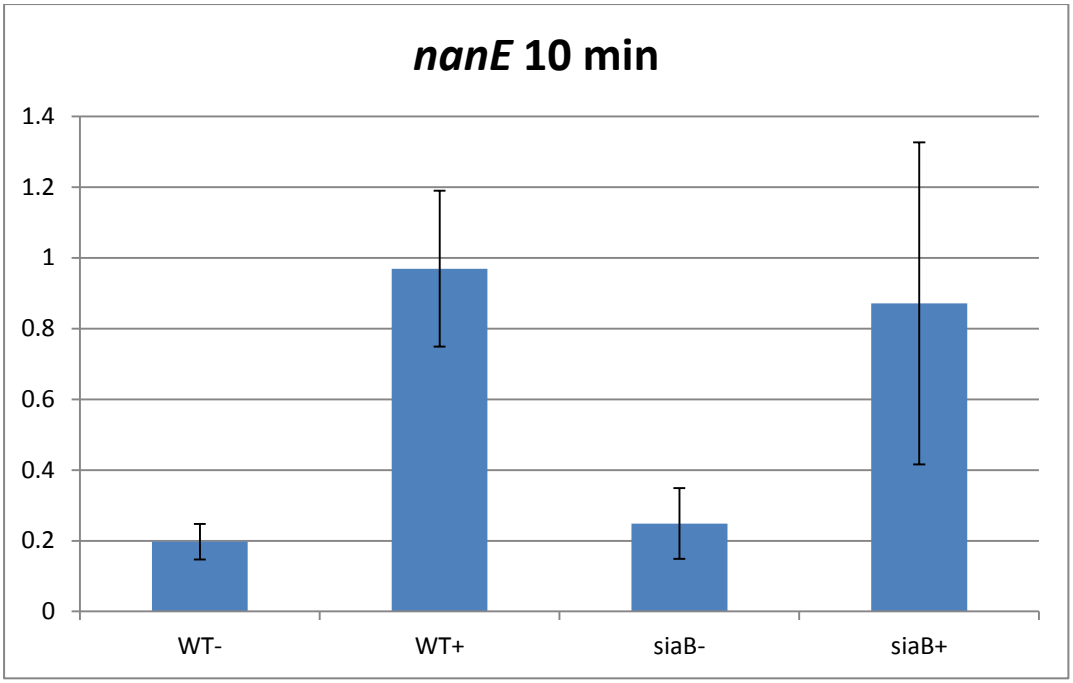
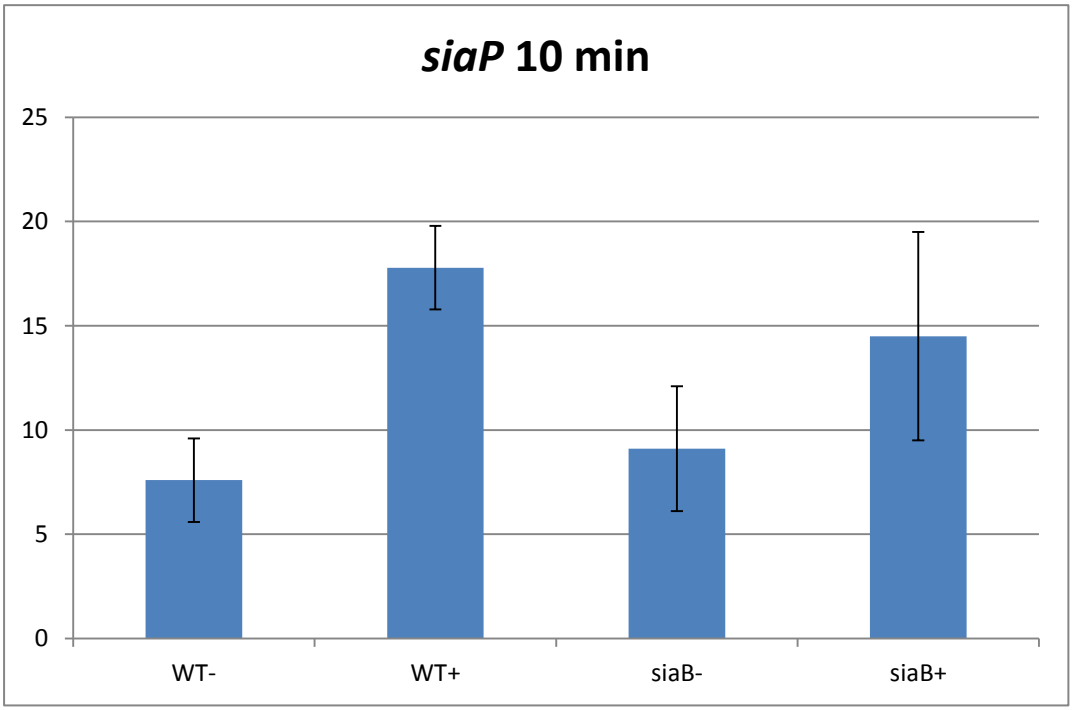


Fig. 35 RT-PCR products for the *nanE*, *siaP* and *atpG* genes in the Rd wild type (WT) and *siaB* mutant strains without (-) or with (+) 30 µg/ml Neu5Ac added for 60 or 10 min.
Even band intensities were observed for the *atpG* reference gene. Each lane contained 10 µl PCR reaction, except for *atpG* where only 2 µl was loaded.


nanE 10 min
1.4
1.2
1
0.8
0.6
0.4
0.2
0
WT-
WT+
siaB-
siaB+


siaP 10 min
25
20
15
10
5
0
WT-
WT+
siaB-
siaB+

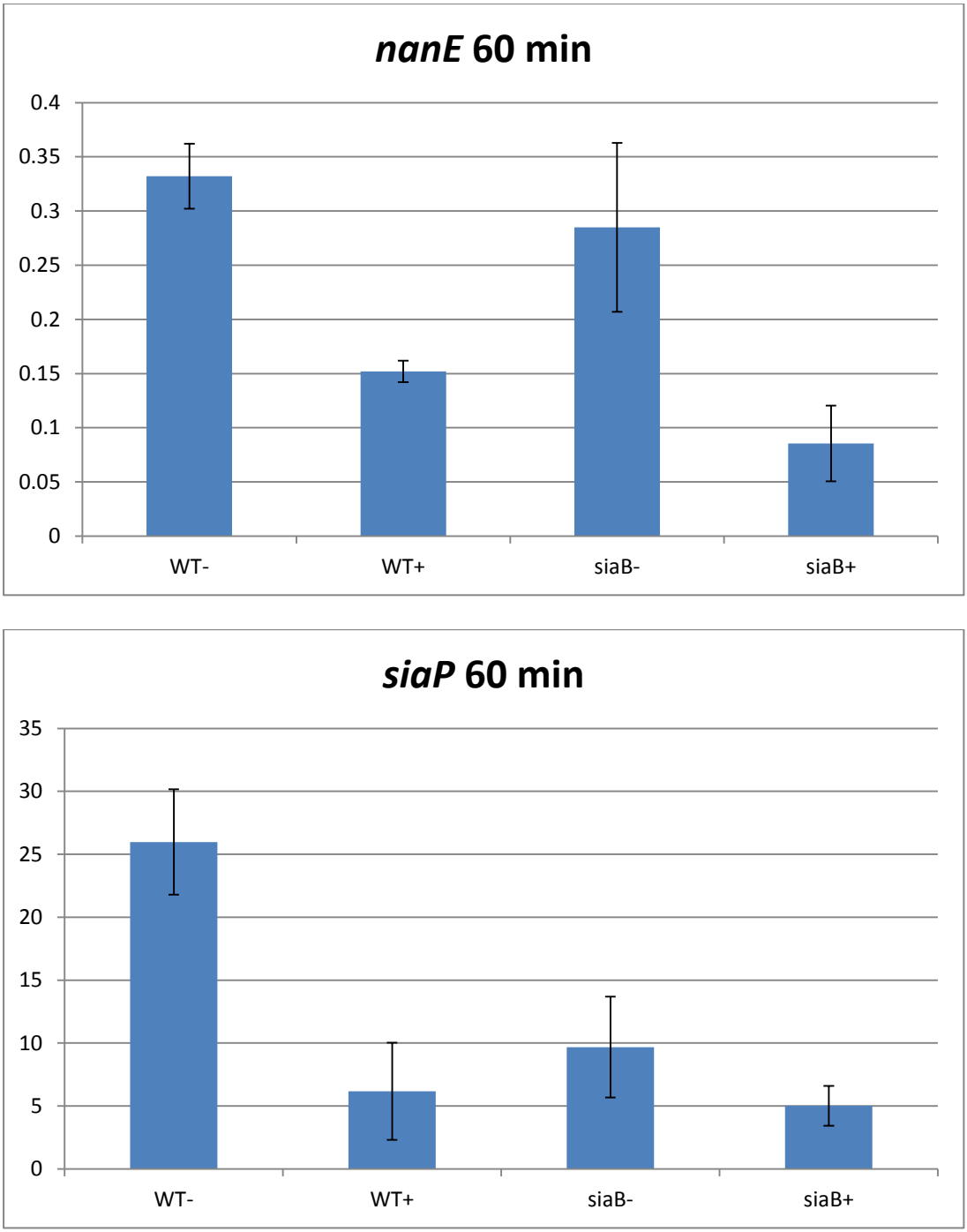
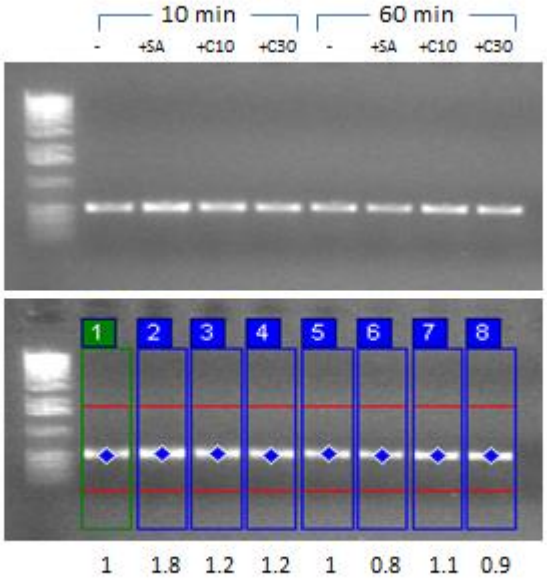


Fig. 36 qPCR analyses of *nanE* and *siaP* gene expression in the Rd wild type (WT) and *siaB* mutant strains following growth in the absence (-) or presence (+) of 30 µg/ml Neu5Ac for 10 or 60 min.
Represented in the y-axis is the fold change in mRNA amount relative to the *frdB* reference gene. qPCR results were the average of two biological replicates with standard deviations, and each biological replicate included two technical replicates.

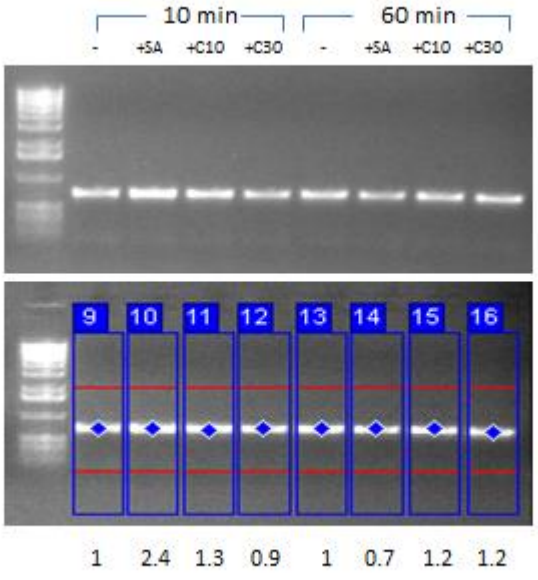
To further verify that CMP-Neu5Ac does not directly affect the expression of the *Hi nagnan* and *siaPQM* operons, we assessed the expression of *nanE* and *siaP* in the Rd wild type strain by adding exogenous CMP-Neu5Ac to the medium. It has previously been shown in some *Hi* strains that supplementing exogenous CMP-Neu5Ac (at 50 µg/ml) increases LPS sialylation (Hood et al., 1999a). This indicates that *Hi* is capable of transporting CMP-Neu5Ac and directly utilising it to sialylate its LPS molecules, albeit details of the mechanism of CMP-Neu5Ac transport are not known.

Semi-quantification of the level of transcription from RT-PCR using the ImageQuant TL programme (GE Healthcare) showed that supplementing exogenous CMP-Neu5Ac at 10 or 30 µg/ml for 10 or 60 min had no effect on the expression of *nanE* or *siaP* in strain Rd (Fig. 37). Semi-quantification was carried out as these PCR reactions were not quantified using qPCR. Although less accurate, nonetheless this data confirms that the expression of *nanE* and *siaP* were not altered in the presence of exogenous CMP-Neu5Ac. These results, combined with our findings for the Rd *siaB* mutant, showed that CMP-Neu5Ac does not affect the expression of the *Hi nagnan* and *siaPQM* operons.

nanE


10 min
60 min
- +SA +C10 +C30 - +SA +C10 +C30
1 2 3 4 5 6 7 8
1 1.8 1.2 1.2 1 0.8 1.1 0.9

siaP


10 min
60 min
- +SA +C10 +C30 - +SA +C10 +C30
9 10 11 12 13 14 15 16
1 2.4 1.3 0.9 1 0.7 1.2 1.2

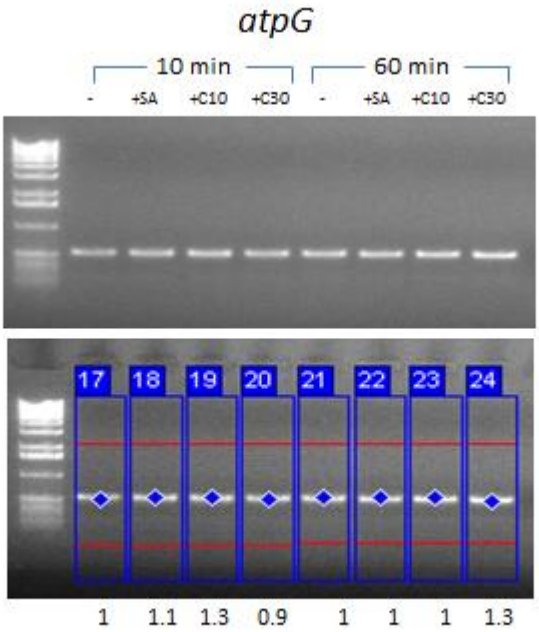


Fig. 37 RT-PCR products for the *nanE*, *siaP* and *atpG* genes in the Rd wild type strain following growth in the absence (-) or presence (+) of 30 µg/ml Neu5Ac (SA), 10 µg/ml CMP-Neu5Ac (C10), or 30 µg/ml CMP-Neu5Ac (C30) for 10 or 60 min.
The region of the gel used in the semi-quantitative analyses is indicated in the lower panel of each gene. The level of transcription (represented by the band in a lane) was semi-quantified, and this was done so by assessing the intensity of a band as a percentage against another selected band. Percentages were converted to decimal numbers, shown at the bottom of the lower panel of each gene.
Bands 2, 3, 4 were expressed as percentages against band 1; bands 6, 7, 8 were expressed as percentage against band 5; bands 10, 11, 12 were expressed as percentages against band 9; bands 14, 15, 16 were expressed as percentages against band 13; bands 18, 19, 20 were expressed as percentages against band 17; and bands 22, 23, 24 were expressed as percentages against band 21.

6.3 Upregulation of *nanE* and *siaP* following growth in Neu5Ac for 10 min is likely due to reduced SiaR activity by binding to a Neu5Ac metabolite

Having determined that CMP-Neu5Ac does not directly affect the expression of the *Hi nagnan* and *siaPQM* operons, it is likely that the SIS sugar is a molecule related to Neu5Ac catabolism. In *Hi*, GlcN-6P was postulated to be the SIS sugar that binds to SiaR to induce the expression of the *nagnan* and *siaPQM* operons: when GlcN-6P binds to SiaR, the affinity of SiaR to its

operator site is enhanced, suggesting that GlcN-6P may function as a co-activator for SiaR and SiaR converts from a transcriptional repressor to an activator (Johnston et al., 2010). The fact that GlcN-6P could enhance the binding of SiaR to its operator site and that SiaR could function as a transcriptional activator is atypical, as all RpiR-type regulation studied so far involves the SIS sugar to derepress gene expression through preventing RpiR-type regulators binding to their operator sites (Sorensen & Hove-Jensen, 1996, Jaeger & Mayer, 2008, Kohler et al., Daddaoua et al., 2009, Kim et al., 2011, Olson et al., 2013). Previous study on the *V. vulnificus* sialometabolism gene regulation system and a recent publication in *S. aureus* showed that ManNAc-6P (a Neu5Ac catabolite, see section 1.4.3 and Fig. 6) is the SIS sugar that interacts with the RpiR-type NanR repressors of both organisms to regulate sialometabolism operons (Olson et al., 2013, Kim et al., 2011). In these cases, ManNAc-6P causes derepression of sialometabolism genes through preventing NanR repressors binding to their operator sites (Olson et al., 2013, Kim et al., 2011). Since the model of sialometabolism gene regulation involving GlcN-6P and SiaR proposed for *Hi* is atypical, we decided to explore the possibility that another Neu5Ac catabolite could be involved with SiaR function in *Hi*.

We used different Neu5Ac catabolism mutants in strain Rd to examine the expression of *nanE* and *siaP* following growth in 30 µg/ml Neu5Ac for 10 min (Fig. 38). Different Neu5Ac catabolism mutants were used to interrupt the Neu5Ac catabolism pathway at alternative points. It would be presumed that this would lead to an accumulation of different Neu5Ac intermediates in alternative mutant strains.

qPCR analysis showed that the upregulation of *nanE* and *siaP* following growth in 30 µg/ml Neu5Ac for 10 min observed in the wild type strain was lost in the *nanA* mutant background (Fig. 38). This finding is consistent with, and extends, previous results where a *nanA* mutant in strain

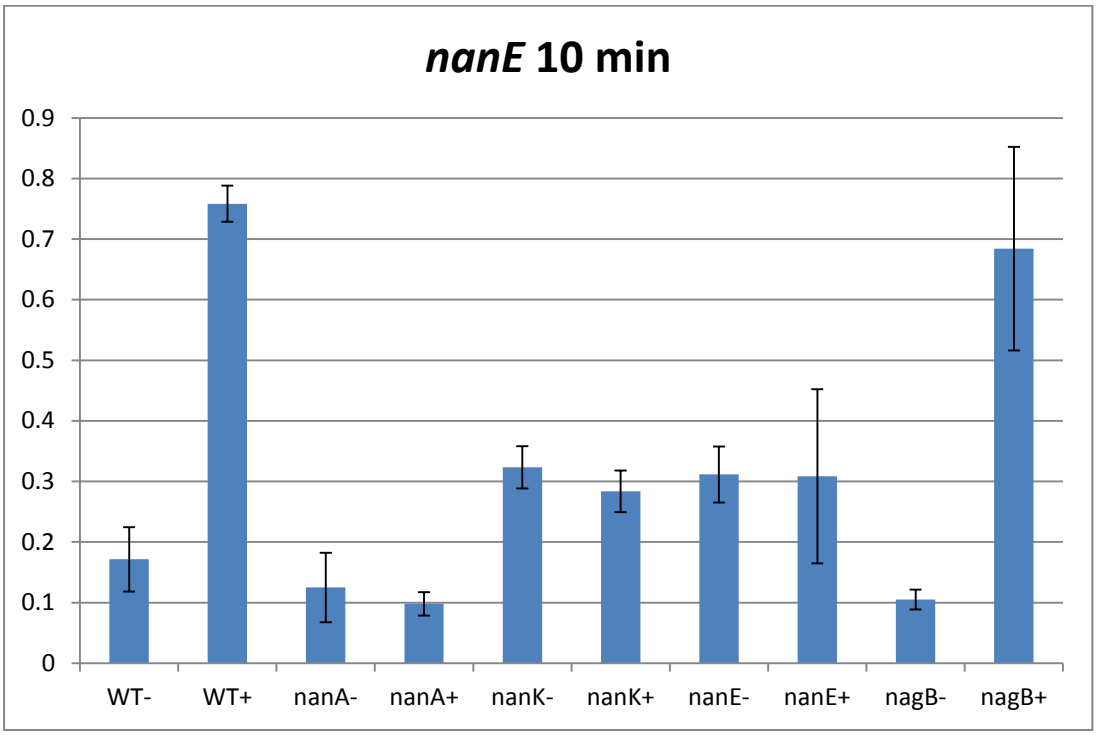
2019 was used (Johnston et al., 2007). In a *nanA* mutant, Neu5Ac cannot be catabolised and CMP-Neu5Ac could possibly accumulate. The loss of *nanE* and *siaP* upregulation in the *nanA* mutant background makes it clear that Neu5Ac is not the sugar responsible for causing the derepression of *nanE* and *siaP*, and is consistent with our hypothesis that a Neu5Ac catabolite may be involved in the regulation of the *Hi* sialometabolism gene cluster.

Although ManNAc-6P was proposed to function as the SIS sugar for NanR in both *V. vulnificus* and *S. aureus*, we found no evidence to suggest that ManNAc-6P plays a similar role in *Hi*. In the *nanE* mutant strain where ManNAc-6P could accumulate (Fig. 6), *nanE* and *siaP* were both upregulated in the presence and in the absence of added exogenous Neu5Ac at 30 μg/ml for 10 min, compared to the wild type strain where no sugar was added (Fig. 38): *nanE* was upregulated by 1.8 fold without and with added Neu5Ac, and *siaP* was upregulated by 2.7 and 1.9 fold respectively without and with added Neu5Ac (Fig. 38). This induction was independent of Neu5Ac, negating the possibility that ManNAc-6P functions as the SIS sugar for SiaR in *Hi*. It is unknown at present whether or not the upregulation of sialometabolism genes in the absence of added Neu5Ac in the *nanE* mutant background compared to the wild type strain was due to some polar effect on *siaR* expression through *nanE* interruption.

In the *nagB* mutant background, GlcN-6P accumulates (Fig. 6). We observed upregulation of *nanE* and *siaP* by 6.5 and 2.8 fold respectively in the *nagB* mutant background following growth in 30 μg/ml Neu5Ac for 10 min, compared to the culture without added Neu5Ac (Fig. 38). This upregulation pattern is similar to the Rd wild type strain (Fig. 38), which suggests that GlcN-6P is likely the SIS sugar that interacts with SiaR to cause derepression of genes dependent upon SiaR, consistent with published data in *Hi*. Interestingly, the magnitude of *siaP* induction was less (by two fold) than the magnitude of *nanE* induction (Fig. 38), and this could be accounted

for by the differential effect of cAMP-CRP on the *nagnan* and *siaPQM* operons (Johnston et al., 2007, Johnston et al., 2010, Jenkins et al., 2010).

In the *nanK* mutant, phosphorylation of ManNAc to ManNAc-6P cannot take place (Fig. 6), and ManNAc could accumulate. The upregulation of *nanE*, but not *siaP*, was observed in the *nanK* mutant background without and with added Neu5Ac at 30 µg/ml for 10 min, with respect to the Rd wild type strain with no sugar added: *nanE* was upregulated by 1.9 and 1.7 fold respectively without and with added Neu5Ac (Fig. 38). The differential induction of the *Hi nagnan* and *siaPQM* operon in the *nanK* mutant background remains to be investigated.



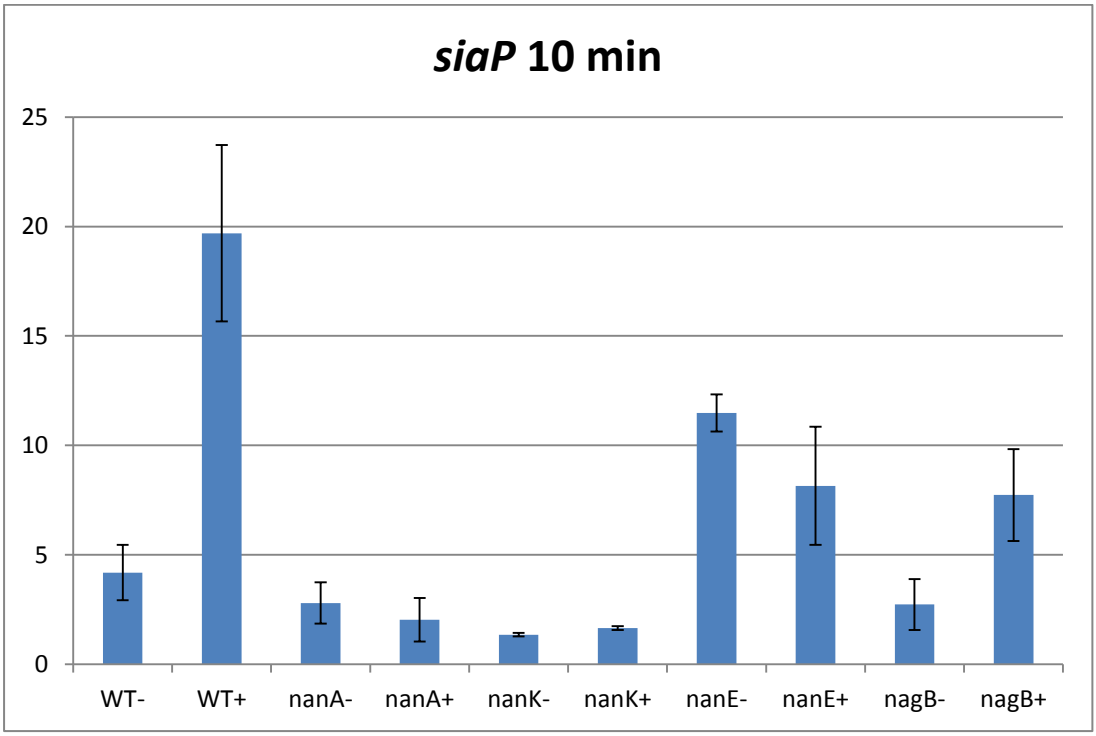


Fig. 38 qPCR analyses of *nanE* and *siaP* gene expression in the Rd wild type (WT), *nanA*, *nanK*, *nanE* and *nagB* mutant strains following growth in the absence (-) or presence (+) of 30 µg/ml Neu5Ac for 10 min.
Represented in the y-axis is the fold change in mRNA amount relative to the *frdB* reference gene. qPCR results were the average of two biological replicates with standard deviations.

6.4 Exploration of the combinatorial effects exerted through CRP and SiaR and confirmation that the expression of *nanE* and *siaP* is not strain dependent

Sialometabolism in *Hi* is complex and there is evidence to suggest that Neu5Ac influences the expression of *nanE* and *siaP* and this may be strain dependent: in the study of Johnston, *nanE* and *siaP* were upregulated by approximately 80- and 50-fold respectively in the 2019 *cyaAnagB* double mutant (a mutant that cannot synthesise cAMP and accumulates GlcN-6P) when both Neu5Ac and cAMP were supplemented for 30 min, compared to the culture that has received no supplementation (Johnston et al., 2010). Such an enhanced level of gene induction has never been observed by us in strain Rd. To assess whether the expression of *nanE* and *siaP* is strain dependent or not, we analysed the expression of *nanE* and *siaP* in the Rd *cyaAnagB* double mutant created by us.

Using the same growth condition described by Johnston (bacteria grown to early exponential phase, and either 100 µM Neu5Ac [equivalent to 30.9 µg/ml], or 1 mM cAMP, or both, added for 30 min) (Johnston et al., 2010), and normalised against the *frdB* reference gene, we saw a similar trend of induction of *nanE* and *siaP* in our Rd *cyaAnagB* double mutant compared to published results in the 2019 *cyaAnagB* double mutant; however the magnitude of induction was smaller: *nanE* was induced by 10 fold following growth in the presence of Neu5Ac alone, and by 15.7 fold following growth in the presence of both Neu5Ac and cAMP; and *siaP* was induced by 3.6 fold following growth in the presence of both Neu5Ac and cAMP (Fig. 39A). It may be that strain background determines the magnitude of change in sialometabolism gene expression; we therefore chose to investigate an equivalent 2019 mutant strain.

In our 2019 *cyaAnagB* double mutant, the trend of *nanE* expression was similar to that observed by Johnston: *nanE* was induced (by 5.3 fold) following growth in the presence of Neu5Ac alone and induced (by 6.8 fold) following growth in the presence of both Neu5Ac and cAMP (Fig. 39 panel B). However, in contrast to Johnston's finding where *siaP* was only induced in the presence of both Neu5Ac and cAMP, the expression of *siaP* in our 2019 *cyaAnagB* double mutant was similar across different growth conditions with added supplements (Fig. 39 panel B). For the Rd and 2019 *cyaAnagB* double mutants created by us, the fold inductions of *nanE* and *siaP* were similar and consistent between strains (Fig. 39 panels A and B).

16S rRNA was used as the reference gene by Johnston throughout his studies of the *Hi* sialometabolism gene regulation (Johnston et al., 2007, Johnston et al., 2010). Since 16S rRNA and *frdB* are very different controls in terms of quantity in gene expression experiments (there are multiple copies of 16S rRNA and only one of *frdB* gene), we assessed the expression of *nanE* and *siaP* in our Rd *cyaAnagB* double mutant using 16S rRNA as the reference gene, and compared the results to our existing data where *frdB* was used as the reference gene (Fig. 39). In the Rd *cyaAnagB* double mutant, the expression of *nanE* and *siaP* was similar in magnitude using either 16S rRNA or *frdB* as the reference gene (Fig. 39). When 16S rRNA was used as the reference gene, *nanE* was induced by 8.3 fold following growth in the presence of Neu5Ac and by 15.0 fold following growth in the presence of both Neu5Ac and cAMP; and *siaP* was induced by 3.3 fold following growth in the presence of both Neu5Ac and cAMP (Fig. 39).

The magnitude of *nanE* and *siaP* induction in both our Rd and 2019 *cyaAnagB* double mutants was similar, which is in contrast to Johnston's findings where a high level of gene induction (80- and 50-fold respectively for *nanE* and *siaP*) was observed. We suggest that the differences

between our results and Johnston's finding may be due to variations in experimental techniques and strains used between laboratories.

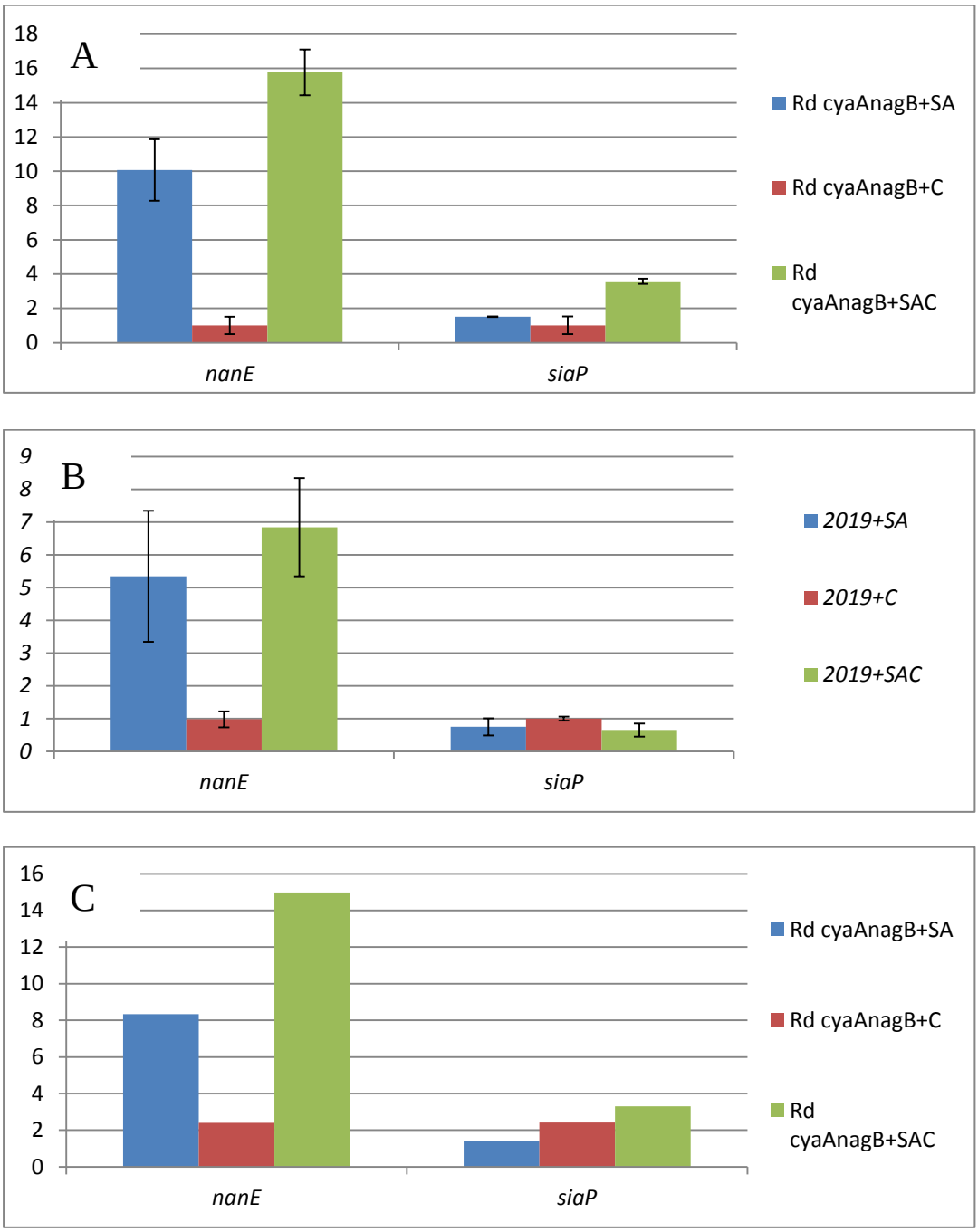


Fig. 39 Expression of *nanE* and *siaP* in the *cyaAnagB* double mutant in strains Rd (panels A and C) and 2019 (panel B) following growth in the presence (+) of 30 µg/ml Neu5Ac (SA), 1 mM cAMP (C) or both (SAC) for 30 min.

Represented in the y-axis is the fold change of gene expression. Cultures grown in Neu5Ac alone (blue box), in cAMP alone (red box), and in both Neu5Ac and cAMP (green box) were compared to cultures that received no supplementation. Shown in panels A and B are the expression of *nanE* and *siaP* in the Rd and 2019 *cyaAnagB* double mutants respectively, and normalised against *frdB*. Shown in panel C is the expression of *nanE* and *siaP* in the Rd *cyaAnagB* double mutant, normalised against 16S rRNA. qPCR results were obtained from two biological replicates for panels A and B, and from a single biological experiment for panel C.

6.5 Downregulation of *nanE* and *siaP* following growth in Neu5Ac for 60 min is due to the depletion of the SIS sugar

Having previously shown in the wild type strains that *nanE* and *siaP* were downregulated by two- to three-fold following growth in 30 µg/ml Neu5Ac for 60 min (see sections 5.1 and 6.1), we now went on to explore the expression of *nanE* and *siaP* using different Neu5Ac catabolism mutants (which accumulate different Neu5Ac catabolites) under the same condition. This downregulation of *nanE* and *siaP* expression is likely due to the depletion of the SIS sugar that interacts with SiaR, resulting in binding of SiaR to its operator site and causing repression of the *nagnan* and *siaPQM* operons.

In the *nanA* and *nanK* mutant backgrounds, no change in the expression of *nanE* and *siaP* was observed with respect to the wild type strain following growth in the absence or presence of 30 µg/ml Neu5Ac for 60 min (Fig. 40). This implies that Neu5Ac or ManNAc do not have a direct role on the expression of *nanE* and *siaP* under the condition tested, consistent with our result obtained from the 10 min time point (see section 6.3). In the *nagB* mutant background, *nanE* and *siaP* were upregulated by 12.6 and 6.4 fold respectively following growth in the presence of 30 µg/ml Neu5Ac for 60 min, compared to the culture without added sugar (Fig. 40). GlcN-6P would be the sugar that accumulates first in the *nagB* mutant background. The observed upregulation of *nanE* and *siaP* in the *nagB* mutant background is consistent with GlcN-6P being the SIS sugar that interacts with SiaR to cause derepression of the *Hi nagnan* and *siaPQM* operons.

Interestingly, in the *nanE* mutant background, upregulation of both *nanE* and *siaP* were observed without and with added Neu5Ac at 30 µg/ml for 60 min, compared to the wild type strain without added sugar: *nanE* was upregulated by 3.7 and 5.6 fold respectively in the absence and

presence of Neu5Ac, while *siaP* was upregulated by 4.6 and 9.2 fold respectively in the absence and presence of Neu5Ac (Fig. 40). The reason for the upregulation of *nanE* and *siaP* in the *nanE* mutant background without and with added Neu5Ac remained to be explored.

In sections 5.1 and 6.1, we observed that the expression of *nanE* and *siaP* in the Rd and 2019 wild type strains remained upregulated following growth at a high concentration (500 µg/ml) of Neu5Ac for 60 min, compared to the culture that has received no supplementation (Fig. 33 and Fig. 34). Since growth rates of *Hi* were comparable between the cultures where Neu5Ac at 500 µg/ml or glucose at 2 mg/ml were supplemented (Fig. 17), this indicates that Neu5Ac would still be available to be metabolised in the cell following growth at this high concentration (500 µg/ml) of Neu5Ac, therefore GlcN-6P could bind to SiaR for continued derepression of the *nagnan* and *siaPQM* operons in the wild type strains (see section 5.1 and 6.1 and Fig. 33 and Fig. 34).

Upregulation of *nanE* as well as *siaR* (which encodes the SiaR repressor protein) was observed following growth in 30 µg/ml Neu5Ac for 10 min, compared to the culture that has not received any supplementation (Fig. 33 and Fig. 41). This is consistent with previous finding which indicated *siaR* and *nanE* (both located within the *Hi nagnan* operon, Fig. 7) are co-transcribed (Jenkins et al., 2010). Since *nanE* and *siaP* were both upregulated at this 10 min time point, this implies that the SIS sugar (likely to be GlcN-6P) is available at 10 min to interact with SiaR and cause derepression of the *nagnan* and *siaPQM* operons.

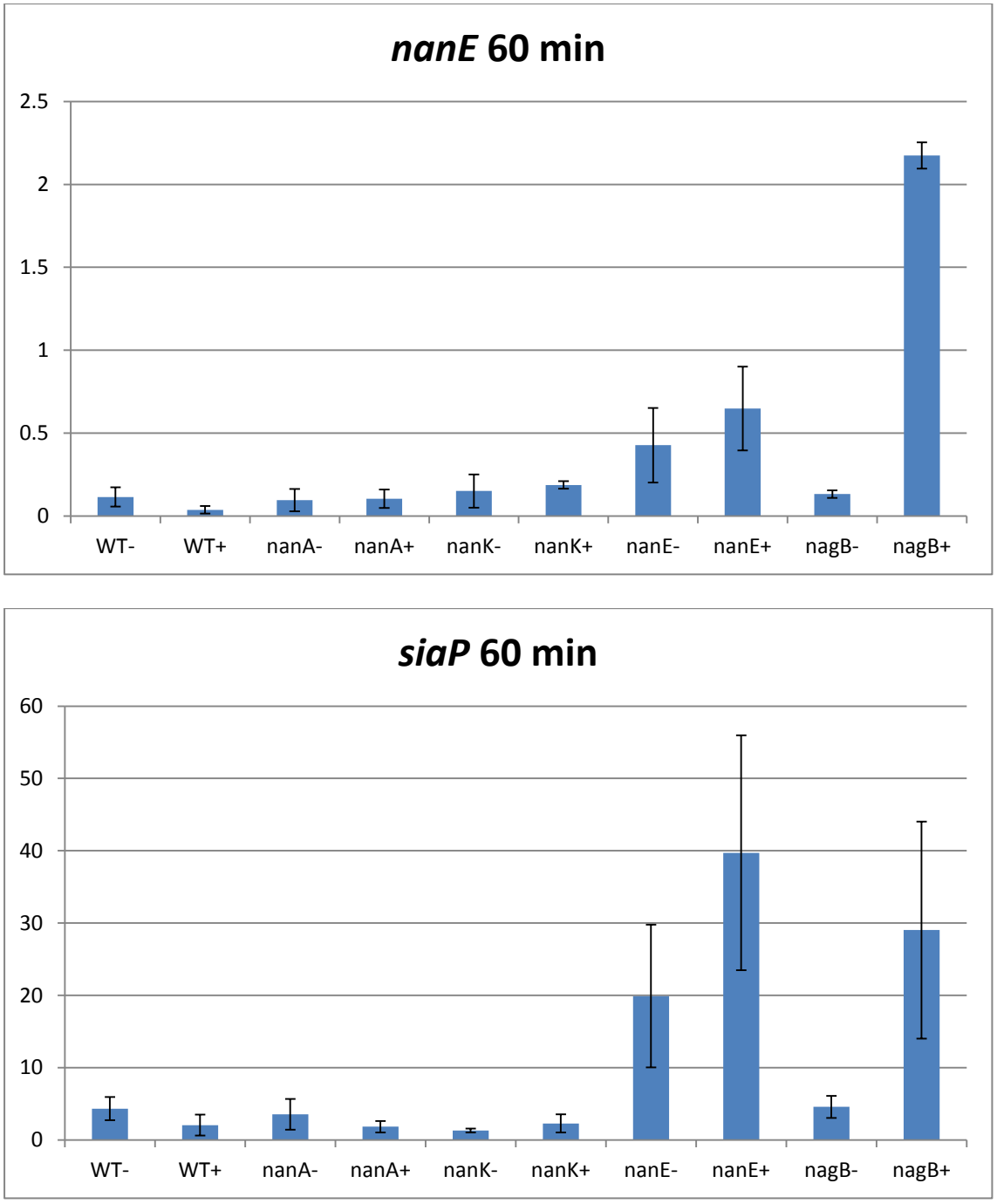


Fig. 40 qPCR analyses of *nanE* and *siaP* gene expression in the Rd wild type (WT), *nanA*, *nanK*, *nanE* and *nagB* mutant strains following growth in the absence (-) or presence (+) of 30 µg/ml Neu5Ac for 60 min.
Represented in the y-axis is the fold change in mRNA amount relative to the *frdB* reference gene. qPCR results were the average of two biological replicates with standard deviations.

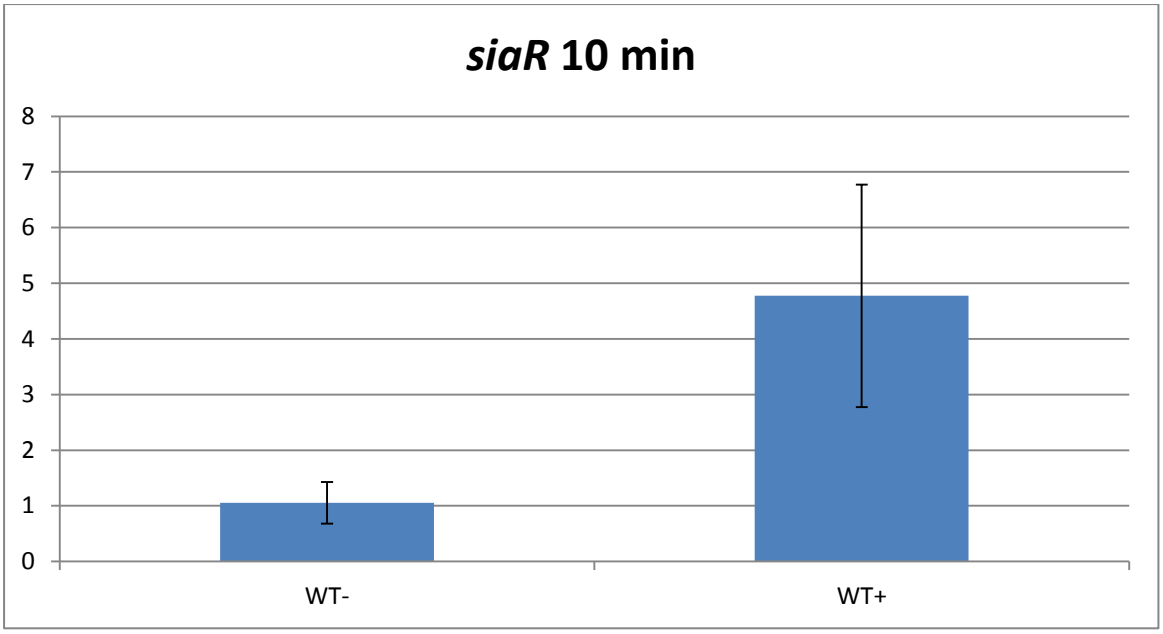


Fig. 41 qPCR analyses of *siaR* gene expression in the Rd wild type (WT) strain following growth in the absence (-) or presence (+) of 30 µg/ml Neu5Ac for 10 min.
Represented in the y-axis is the fold change in mRNA amount relative to the *frdB* reference gene. qPCR results were the average of two biological replicates with standard deviations.

6.6 The up- and downregulation of *nanE* and *siaP* following growth in Neu5Ac for 10 or 60 min respectively do not affect LPS sialylation

We have determined that the expression of the *Hi nagnan* and *siaPQM* operons is altered over short periods of time (10 or 60 min) by the availability of Neu5Ac. However, it is not known whether changes in Neu5Ac flux through the catabolic pathway could affect LPS sialylation over the same time period. Therefore, LPS sialylation was examined following growth of *Hi* in different times (10 versus 60 min) and concentrations (30 versus 500 µg/ml) of Neu5Ac. This analysis was carried out using strain R2846 because LPS sialylation in this strain is more evident by SDS PAGE (Fig. 22).

It is important to mention here that we analysed the LPS profiles of strain R2846 following growth in CDM broth (Fig. 22), i.e. the same growth condition used in the gene regulation study (bacteria grown to mid-log in CDM broth, and either 30 or 500 µg/ml Neu5Ac added for 10 or

60 min). This is different to the condition used for previous analysis of LPS sialylation, i.e. bacteria grown overnight on BHI plate (see sections 2.2.8 and 4.6).

The resulting LPS gel showed no evidence of LPS sialylation in the R2846 wild type strain following growth in CDM broth without or with added 30 or 500 µg/ml Neu5Ac for 10 or 60 min (Fig. 42). This is in contrast to the LPS profiles obtained following culture on BHI plate supplemented with 30 µg/ml Neu5Ac overnight, where LPS sialylation can be readily visualised (Fig. 42). These results suggest that although short incubation times in the presence of Neu5Ac affect gene expression of the *Hi nagnan* and *siaPQM* operons, detectable LPS sialylation does not occur under these same conditions.

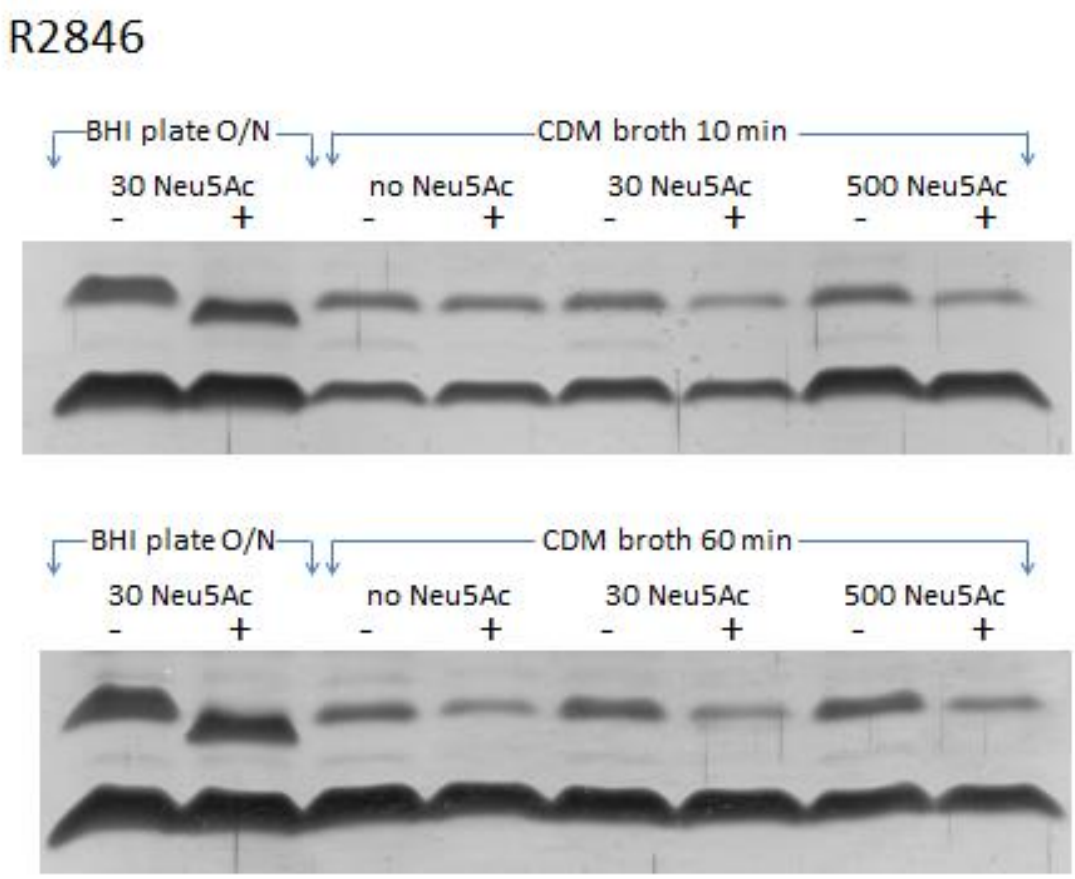


Fig. 42 LPS profiles of R2846 wild type strain following growth on Neu5Ac supplemented sBHI plate overnight (O/N) or in CDM broth without or with added Neu5Ac (at 30 or 500 µg/ml) for 10 or 60 min.
- and + respectively represent without and with neuraminidase treatment.

6.7 Discussion

In the previous chapter, we observed a temporal/concentration effect of Neu5Ac upon the expression of the *Hi* sialometabolism operons. Using genes *nanE* and *siaP* to represent the *Hi nagnan* and *siaPQM* operons respectively, we set out to determine whether the details of this were the same between strains. Our findings showed that exogenous Neu5Ac exerted the same temporal/concentration effect on the expression of *nanE* and *siaP* in strain 2019 as it did in Rd: *nanE* and *siaP* were both upregulated following growth in 30 or 500 µg/ml Neu5Ac for 10 min, both were downregulated following growth in 30 µg/ml Neu5Ac for 60 min, and the two genes remained upregulated following growth in 500 µg/ml Neu5Ac for 60 min. The temporal/concentration effect of Neu5Ac on the expression of *nanE* and *siaP* is likely a general feature seen across *Hi* strains.

We propose that the upregulation of *nanE* and *siaP* observed is due to the interaction of an SIS sugar with SiaR, causing derepression of sialometabolism gene expression through preventing SiaR binding to its operator site. If this derepression is specific for Neu5Ac, then it must involve one of three possibilities: Neu5Ac itself, the anabolic product (CMP-Neu5Ac), or a catabolic product of Neu5Ac. To determine which one of the three possible candidates was the most likely, we investigated the expression of *nanE* and *siaP* in the *siaB* mutant (where no CMP-Neu5Ac is made) and in various Neu5Ac mutants where the Neu5Ac catabolic pathway is inactivated.

In the Rd *siaB* mutant, expression of *nanE* and *siaP* was similar to the wild type stain, i.e. both were upregulated following growth in 30 µg/ml Neu5Ac for 10 min. This indicates that CMP-Neu5Ac is not the SIS sugar that interacts with SiaR to derepress sialometabolism gene expression. In accord to the finding, a sugar nucleotide has never been shown to interact with a RpiR-type regulator to influence the expression of genes. Also, when CMP-Neu5Ac at two

different concentrations was supplemented in the growth media, we observed no induction in the expression of either *nanE* or *siaP*, further evidence that CMP-Neu5Ac is not the SIS sugar. Previous study has shown that supplementing CMP-Neu5Ac in the growth media increases LPS sialylation (Hood *et al.*, 1999b), which indicates that CMP-Neu5Ac is taken up by the bacterium. Thus, we conclude that the *Hi* Neu5Ac activation pathway does not directly influence the expression of the *Hi nagnan* and *siaPQM* operons.

It is therefore likely that the SIS sugar is Neu5Ac or an intermediate from the Neu5Ac catabolism pathway. To investigate this, a number of steps in the catabolic pathway were inactivated, and *nanE* and *siaP* expression was investigated in the respective mutant strains. In a *nanA* mutant background, Neu5Ac is likely the only relevant molecule that could affect the expression of *nanE* and *siaP*. Our qPCR results showed that the upregulation of *nanE* and *siaP* observed in the wild type strains following growth in Neu5Ac for 10 min, was not seen in the *nanA* mutant background, which is consistent with published results (Johnston et al., 2007). Previous study showed through electrophoretic shift mobility assay (EMSA) that 1 mM of Neu5Ac did not bind to SiaR (Johnston et al., 2007). These findings are consistent with Neu5Ac not being the SIS sugar that interacts with SiaR, making it highly probably that an intermediate derived from its catabolism plays this role.

Evidence obtained using a number of other mutant strains affected at different steps of Neu5Ac catabolism favours GlcN-6P as the SIS sugar that interacts with SiaR. The strongest evidence supporting this hypothesis was the observed upregulation of both *nanE* and *siaP* following growth in Neu5Ac for 10 or 60 min in the *nagB* mutant background, a strain whereby GlcN-6P cannot be converted to Fru-6P, and GlcN-6P is the major sugar that accumulates (Fig. 6). This

finding is consistent with previous study whereby GlcN-6P was proposed to be the SIS sugar for SiaR (Johnston et al., 2010).

The model previously proposed for the *Hi* sialometabolism gene regulation stated that the interaction of GlcN-6P with SiaR enhances the binding of SiaR to its operator site; upon binding of GlcN-6P to SiaR in the presence of cAMP-CRP, the function of SiaR converts from a transcriptional repressor to a transcriptional activator (Johnston et al., 2010). We felt it was necessary to reconsider this model as it is atypical for SiaR, a member of the RpiR protein family, to possess both transcriptional repression and activation properties. In fact, none of the RpiR-type regulators studied so far has been shown to possess transcriptional activation properties and all are known to function as transcriptional repressors. It is also atypical that a SIS sugar such as GlcN-6P could enhance the binding of a RpiR-type regulator to its operator site, as all SIS sugars known to date prevent the binding of their respective RpiR-type regulators to the operator sites. For these reasons, we propose an alternative model for the *Hi* sialometabolism gene regulation: following growth of Neu5Ac, GlcN-6P interacts with SiaR to prevent it binding to its operator site, this results in derepression of the *nagnan* and *siaPQM* operons. When Neu5Ac becomes depleted from the culture medium over time (such as following growth in 30 µg/ml Neu5Ac for 60 min), we propose less GlcN-6P would be available to bind to SiaR, resulting in SiaR binding to its operator site and concomitant downregulation of *nanE* and *siaP*. We further propose that when a higher concentration of Neu5Ac (500 µg/ml) is added to the growth media, catabolism of Neu5Ac continues at and after 60 min; and the available GlcN-6P allows upregulation of *nanE* and *siaP* to continue. Our observed downregulation of *nanE* and *siaP* in the wild type strains following growth in 30 µg/ml Neu5Ac for 60 min, and upregulation of the two genes following growth in a higher concentration (500 µg/ml) of Neu5Ac for 60 min are consistent with our proposed model. However, it is likely that the model we proposed for the

Hi sialometabolism gene regulation is more complex than described, as other components, notably CRP, play a role (Johnston et al., 2007, Jenkins et al., 2010).

The *siaR* transcript, which encodes the SiaR protein, is located within the *nagnan* operon and is co-transcribed along with *nanE* and other *nagnan* genes (Jenkins et al., 2010). The upregulation of *siaR* at the 10 min time point, compared to the culture that has not received any supplementation, was similar to that for *nanE* and *siaP*. The fact the *nanE* and *siaP* were also upregulated at 10 min indicates that the SIS sugar is available to bind to SiaR, albeit *siaR* is upregulated at this time point.

Previous study on the sialometabolism gene regulation system of *V. vulnificus* and a recent publication in *S. aureus* identified ManNAc-6P as the SIS sugar that interacts with the RpiR-type NanR repressor in both organisms to regulate sialometabolism operons (Kim et al., 2011, Olson et al., 2013). However, since *nanE* and *siaP* were both upregulated in the Rd *nanE* mutant background (where ManNAc-6P would likely accumulate first) following growth in the presence and absence of added Neu5Ac for 10 or 60 min, compared to the wild type strain without added sugar (i.e. no Neu5Ac dependent effect on the expression of sialometabolism genes was seen), this implies that ManNAc-6P does not have a direct role on the expression of sialometabolism operons in *Hi*. It is unclear at present what could cause the upregulation of *nanE* and *siaP* in the *nanE* mutant in the absence of Neu5Ac when compared to the wild type, but some effect through inactivating SiaR in the *nanE* mutant may play a role.

Upregulation of *nanE* (and not *siaP*) was observed without and with added Neu5Ac for 10 min in the *nanK* mutant background (a strain whereby ManNAc cannot be converted to ManNAc-6P, and ManNAc is the first sugar to accumulate) (Fig. 6). This differential expression of *nanE* and

siaP observed through inactivating *nanK* demonstrates that altering the balance of intermediates in the Neu5Ac catabolic pathway can have some counter intuitive effect on the expression of *nagnan* and *siaPQM* operons. These results from the *nanE* and *nanK* mutants indicate the complexity in regulation of the *Hi* sialometabolism operons, some details of this remain to be elucidated.

As previously mentioned, one reason why sialometabolism gene regulation in *Hi* is complex is because in addition to the effect of SiaR, CRP also plays a role in the process (Johnston et al., 2007, Jenkins et al., 2010). There are two potential means by which CRP and the level of cAMP could affect the expression in the *nanE* and *nanK* mutant strains. Firstly, through pyruvate produced from the first step of Neu5Ac catabolism (Fig. 6). In *E. coli*, an increase in the Phosphoenolpyruvate (PEP) to pyruvate ratio is associated with an increase in cAMP level (Cunningham *et al.*, 2009). As the same effect is likely relevant in *Hi*, reduced catabolism of Neu5Ac to Fru-6P will produce less pyruvate as a by-product, so we could expect an increase in the ratio of PEP to pyruvate, which would result in an increase in cAMP level. Secondly, reduced Neu5Ac catabolism will impinge on the level of Fru-6P (the final product of Neu5Ac catabolism) (Fig. 6). As Fru-6P, through glycolysis, can give rise to ATP, it is likely that decrease in ATP production through inactivating Neu5Ac catabolism would lead to a lower energy state of the cell, which could result in a higher cAMP level and have an effect on the expression of the *siaPQM* operon. Perturbation of the *Hi* Neu5Ac catabolism pathway could also lead to changes in other metabolic processes in the cell, which may have an influence on the intracellular level of pyruvate, and hence the cAMP-CRP level. As the influence of cAMP-CRP is largely on the expression of the *siaPQM* operon, the availability of pyruvate could affect this expression.

In the work done by us, we observed upregulation of *nanE* and *siaP* in Rd and 2019 wild type strains by approximately seven fold following growth in the presence of 30 µg/ml Neu5Ac added for 10 min, and normalised to the *frdB* expression level (Fig. 33 and Fig. 34). However, in Johnston's study, a much higher fold induction was observed in strain 2019 than we did across different experiments in strain Rd: in Johnston's *Hi* 2019 *cyaAnagB* double mutant (a mutant that cannot synthesise cAMP, and accumulates GlcN-6P), *nanE* and *siaP* were upregulated by approximately 80- and 50-fold respectively following growth in the presence of both Neu5Ac and cAMP added for 30 min, and normalised to the 16S rRNA expression level (Johnston et al., 2010). Details relating to the strains that were used and the way that the mutants were created could quite logically contribute to the differences in the data that one obtained. In Johnston's study, the *cyaAnagB* double mutant was created in the *Hi* strain 2019 background. It may be that inter-strain differences are present regarding the regulation of sialometabolism genes. Furthermore, in Johnston's study, the *sacB* cassette, which contains the tetracycline promoter to allow for inducible expression of *sacB* in the presence of chlortetracycline, was used to create mutant strains; whereas in our study, *kanC* and *eryC* were used. The length of time that bacteria were grown in the presence of Neu5Ac were different in obtaining the two sets of data: in our gene expression study for 10 or 60 min, in Johnston's study for 30 min; and since we have shown that there is a temporal/concentration effect of Neu5Ac on the expression *nanE* and *siaP*, these different times could affect the expression levels. In Johnston's study, the times that bacteria were grown in the presence of Neu5Ac were not always stated, direct comparison with our data is not always possible. Also for the *cyaAnagB* double mutant used in Johnston's study, the medium was supplemented with 1 mM cAMP to activate CRP. Inactivating *cyaA* and artificially supplementing cAMP could result in a non-physiological level of cAMP in the bacterium, which may change gene expression dependent upon activated CRP. The expression of *nanE* and *siaP* in our gene expression analysis was normalised against *frdB*, in contrast to

Johnston's study where 16S rRNA was used, and this difference in reference gene copy number could potentially influence the magnitude of *nanE* and *siaP* expression.

To investigate the potential relevance of some of these factors, we created our own *cyaAnagB* double mutants in strains Rd and 2019 and assessed the expression of *nanE* and *siaP* following growth under the same conditions described by Johnston (bacteria grown to early exponential phase, and either 100 µM [equivalent to 30.9 µg/ml] Neu5Ac, or 1 mM cAMP, or both, were added for 30 min) and using either *frdB* or 16S rRNA as reference genes. We observed that the trend of *nanE* and *siaP* induction was similar between our Rd and 2019 *cyaAnagB* double mutant and Johnston's 2019 *cyaAnagB* double mutant (Johnston et al., 2010), but the magnitude of induction observed was far less (by approximately 5 and 15 fold for *nanE* and *siaP* respectively in our 2019 *cyaAnagB* double mutant with respect to Johnston's level). We suggest that the difference between our result and Johnston's finding may be due to variations in the details of the techniques used, such as the way in which the mutants were created. Furthermore, the 2019 strains that were used may have subtly different genotypes between labs that influence experiments in a way that is not easy to understand.

Having established that sialometabolism can show a temporal/concentration effect of Neu5Ac, we felt it is important to determine if changes in Neu5Ac availability can influence LPS sialylation and virulence over short periods of time. We assessed the LPS sialylation profile of strain R2846 grown under the same condition used for gene regulation studies, albeit that this is different to the growth condition used for our previous study of LPS sialylation (see chapter 4). When grown in CDM broth without and with 30 or 500 µg/ml Neu5Ac for 10 or 60 min, we found no evidence of LPS sialylation in strain R2846. The level of LPS sialylation can vary between solid and liquid growth using the same medium, with less LPS sialylation following

growth in broth compared to plate (Hood et al., 2004). Furthermore, CDM or BHI growth media can provide alternative nutrients and sources of carbon and energy. Such growth differences could confound a simple comparison of LPS sialylation between the experiments as observed for strain R2846.

7 Final discussion

SA can play a major role in determining *Hi* virulence. Extending our current understanding of sialometabolism in *Hi* could help us understand an aspect of the complex bacteria-host interaction that underlies human disease. Through PCR amplification of selected sialometabolism genes (*siaP, nanA, siaB, lic3A*, which respectively represent Neu5Ac transport, catabolism, activation and sialyltransferase processes in *Hi*), we demonstrated that sialometabolism is a process common to *Hi*. There are aspects within sialometabolism and its regulation in *Hi* that were unclear, so to gain further insight into sialometabolism of *Hi*, we studied genes whose functions are predicted or known to be involved in SA related function in other organism and further investigated factors that influence sialometabolism regulation.

Potential novel sialometabolism genes that were investigated were *HI0227* and *HI0148*. The crystal structure of HI0227 in *Hi* Rd predicted that the protein may function as a SA epimerase (Teplyakov et al., 2005). The homologue of the *Hi* Rd HI0148 protein in *E. coli* was proposed to function as a Neu5Ac mutarotase, whose role is to convert the anomeric configuration of Neu5Ac (mainly from α- to β-Neu5Ac) (Severi et al., 2008). Further evidence to suggest that *HI0227* and *HI0148* are associated with sialometabolism in *Hi* is that the two homologues in *E. coli* and in *P. multocida* (a bacterium closely related to *Hi*) are located within sialometabolism gene clusters in those organisms (Fig. 14). Therefore, *HI0227* and *HI0148* were considered to be two primary candidate novel sialometabolism genes for study.

Immediately adjacent to *HI0227* and *HI0148* are *HI0228* and *NTHI0236*, respectively; and due to their close proximity and information from predicted functions, we referred to *HI0228* and *NTHI0236* as peripheral candidate sialometabolism genes. DNA sequence indicates that *HI0228* may be linked to *HI0227*, as putative -35 sequences of *HI0227* were identified within the distal

end of *HI0228* of some NT*Hi* strains, which implies that transcription of *HI0227* and *HI0228* may be inter-dependent on each other due to the potential of overlapping transcripts. As *HI0227* was postulated to be involved in sialometabolism, it is reasonable to assume that *HI0228* may also encode a sialometabolism related function. *NTHI0236* had been included in previous studies of the *Hi* sialometabolism gene regulation, and for this reason, it was included in ours.

Results from our experiments showed that different aspects of sialometabolism in *Hi* (growth on Neu5Ac in aerobic and microaerophilic conditions, LPS sialylation, the ability to resist complement mediated killing which is correlated to the level of sialylated LPS, and the expression of sialometabolism genes as a means to understand sialometabolism gene regulation) in mutants that lack the functions of HI0227 and HI0148 were not different from those of the wild type strains (see chapters 4 and 5). The fact that we observed no growth difference in our *HI0227* mutants compared to the wild type strains following growth in Neu5Ac as the only sugar source available is consistent with published result in *E. coli*. In *E. coli*, deleting the *HI0227* homologue (the *yhcH* gene), unlike deleting other sialometabolism genes (*nanT*, *nanA*, *nanE*, *nanK*, which respectively are involved in transport [for *nanT*] and catabolism [for *nanA*, *nanE* and *nanK*] of Neu5Ac in *E. coli*), produced no apparent effect on sialometabolism in the presence of Neu5Ac as the only sugar source available (Kalivoda et al., 2003).

The fact that HI0227 was postulated to function as a SA epimerase was based on the similarity of the crystal structure to other epimerases and the association of the *E. coli* homologue with a sialometabolism operon (Teplyakov et al., 2005). Since the function of HI0227 has not been shown to be associated with sialometabolism in *E. coli* from published data (Kalivoda et al., 2003) or in *Hi* from our own data, we query that HI0227 is associated with sialometabolism at all.

Sialylated LPS glycoform has been shown to be a major virulence factor in the pathogenesis of experimental otitis media in chinchillas (Bouchet et al., 2003).

An *in vivo* experiment subsequently conducted in a mouse model of otitis media showed that mice inoculated with the *HI0227* mutant created in NT*Hi* strain 162 behaved no differently to mice inoculated with the wild type NT*Hi* strain (D Hood, personal communication). This result is consistent with HI0227 not having a role in the utilisation of SA *in vivo*. We propose that HI0227 may actually not function as a SA epimerase, albeit the homologues of *HI0227* in *E. coli* and *in P. multocida* are associated with sialometabolism gene clusters.

HI0148 is the other candidate novel sialometabolism gene that was studied, and the homologue of *HI0148* in *E. coli* was predicted to encode the Neu5Ac mutarotase function (see section 1.4.1) (Severi et al., 2008). To determine whether HI0148 affects some aspects of sialometabolism in *Hi* (the ability to utilise Neu5Ac in aerobic and microaerophilic conditions, and the ability to sialylate LPS molecule), we performed growth assays, SBA and LPS gel using mutants that lack the function of HI0148 and compared the results to their respective wild type strains. It is known that Neu5Ac in solution contains a mixture of α- and β-anomers and that Neu5Ac can convert to predominantly the β-form (to 90 %) over time, although this happens at a relatively slow rate ($t_{0.5}$ = 80 min at pH 5.4) (Jaques *et al.*, 1977). The fact that we observed no difference in the *HI0148* mutants in the tested aspects of sialometabolism indicates that the anomeric form of Neu5Ac used in our experiments may have been more toward the β-anomer than the α-anomer.

Previous study has documented that in the presence of a Neu5Ac mutarotase, mutarotation of Neu5Ac happens so quickly that complete mutarotation of α-Neu5Ac to β-Neu5Ac can occur within 7 min at pH 5.5 (Severi et al., 2008). In *E. coli*, mutarotation of Neu5Ac likely occurs in the periplasm and is a step that is required as it is likely that Neu5Ac is taken up in the β-form

(Severi et al., 2008). Although it is not known which anomeric form of Neu5Ac is preferentially taken up by the Neu5Ac TRAP transport system of *Hi*, since growth on Neu5Ac and LPS sialylation profiles of our *HI0148* mutants were similar to the wild type strains, this implies that β-Neu5Ac (the predominant form in solution) can be transported into the cell and be metabolised. Bacteria containing a Neu5Ac mutarotase are likely at a fitness advantage as a limited amount of newly-cleaved, free Neu5Ac presented in the α-anomeric form can be rapidly converted to the β-form and be transported into the cell for further metabolism (Severi et al., 2008). To elucidate the sialometabolism role of HI0148 in *Hi*, we suggest to perform *in vivo* experiments using the chinchilla model of otitis media with our *HI0148*- mutants, as Neu5Ac that are newly-cleaved during inflammation would be in its α-form, which means that potentially we could see an effect in the *HI0148*- mutant *in vivo* under inflammation conditions.

HI0228 is one of the two peripheral candidate sialometabolism genes that were studied. HI0228 contains part of or all of the domains required for a functional GT family 8 protein, and has a predicted role in sugar metabolism. We propose that *HI0228* is likely a pseudogene in most NT*Hi* strains because unlike *HI0227* and *HI0148*, not all strains screened contained this ORF, and among the strains screened that do contain *HI0228*, there are frame-shifts and missing sequences within the ORF.

Although changes were observed in LPS profiles and SBA by some of the *HI0228*- mutants (such as the 375 *HI0228*-A and the 375 *HI0227/HI0228*-A mutants), these changes were shown to be due to the effect of phase variable genes and not due to loss of the HI0228 function per se. The increase in serum resistance observed in the 375 *HI0228*-A mutant compared to the wild type strain was due to the disialylated LPS phenotype, resulting from the in-frame *lic3B* gene and the out-of-frame *lgtC*. Lic3B has been shown to be responsible for the production of disialylated

LPS glycoform (Fox et al., 2006) and without a functioning LgtC, there would be no competition for the production of the disialylated LPS glycoform. The magnitude of increase in serum resistance observed in the 375 *HI0228*-A mutant was not as great as one might expect from the sialylation profile, which could be due to the absence of digalactoside containing glycoform as a result of the out-of-frame *lgtC*, since this LPS glycoform has been shown to contribute to serum resistance and is implicated in virulence (Griffin et al., 2005). The 375 *HI0227/HI0228*-A mutant similarly exhibited an increase in serum resistance that was dependent upon the same change in the *lgtC* phase variable gene. It is worth mentioning here that although certain mutants (e.g. 375 *HI0227*-A, 375 *HI0228*-B, 375 *HI0227/HI0228-B,* and Rd *NTHI0236*-A) showed differences in their ability to resist complement mediated killing compared to the wild type strains, similar LPS phenotypes were exhibited by these mutants compared to the wild type strains. It is likely that small differences in LPS (such as PCho and O-acetyl groups) were present, which are below the level of detection in our LPS gels, caused the difference in the level of resistance against the killing effect of human serum in these mutants. Furthermore, modification of surface structures, such as the presence, absence or abundance of surface proteins and changes in glycosylation of surface proteins, may also contribute to the bacterium's ability to resist complement mediated killing.

The second peripheral candidate sialometabolism gene that was studied is *NTHI0236*, which is the designation given in strain 86-028NP and is the DNA sequence that contains the combined *HI0148.1* and *HI0149* ORFs in strain Rd. The length of the *NTHI0236* ORF is conserved across most NT*Hi* strains screened, except for *Hi* Rd, where an adenine is missing within the ORF, resulting in a frame-shift which causes the two ORFs to be formed. Although NTHI0236 contains a truncated domain of ICMT, which may play a role in methylation of lipid A during LPS biosynthesis and that the function of lipid A methylation remains unknown to date (Boon

Hinckley et al., 2005); however, from our LPS gel we could not detect any difference in LPS phenotype in the mutants that lack the function of NTHI0236 compared to the wild type strain, and the ability to resist serum complement mediated killing (a biological readout that could vary with LPS phenotype) were similar in mutants that lack the function of NTHI0236 compared to the wild type strain. We could not detect any obvious role that NTHI0236 plays in LPS biosynthesis and it is unclear how NTHI0236 could have any role in sialometabolism. More sensitive structural analytical techniques such as NMR would need to be employed to help us determine the function of NTHI0236 on the structure of lipid A.

Although we could not detect any roles in sialometabolism in the candidate and peripheral sialometabolism genes from the various assays that were performed; nonetheless, we demonstrated that the expression of sialometabolism genes can be influenced by the concentration and time of exposure to Neu5Ac, an effect we refer to as the temporal/concentration effect of Neu5Ac. This temporal/concentration effect of Neu5Ac is likely a general feature seen across *Hi* strains as the same phenomenon was observed in another *Hi* strain- 2019, which (like Rd) is a strain that has been extensively used by others in the study of sialometabolism and gene regulation (Johnston et al., 2007, Johnston et al., 2010). Having identified the importance of the temporal/concentration effect of Neu5Ac on the expression of sialometabolism genes, we investigated the relevance of this to the *Hi* sialometabolism gene regulation. We derived a revised model for the regulation of sialometabolism genes in *Hi* with respect to the function of SiaR. SiaR is a RpiR-type transcriptional regulator and functions as the transcriptional repressor of the *Hi* sialometabolism operon upon binding to its operator site, which is located within the intergenic region between *nanE* and *siaP* (Fig. 8). Our evidence indicates that GlcN-6P, produced as an intermediate in the catabolism of Neu5Ac, is most likely the SIS sugar that interacts with SiaR, and upon binding to SiaR results in the reversal of the

binding of the repressor protein to its operator site, and increase transcription of the *nagnan* and *siaPQM* operons. During our study, we found no evidence to support Johnston's hypothesis that SiaR could function as a transcriptional activator (Johnston et al., 2010) and other published data has not shown that RpiR-type regulators could function as transcriptional activators.

Our proposal that GlcN-6P functions as the SIS sugar that binds to SiaR is supported by the evidence that upregulation of both *nanE* and *siaP* were observed following growth in Neu5Ac for 10 or 60 min in the *nagB* mutant background (a strain whereby GlcN-6P cannot be converted to Fru-6P, and GlcN-6P is likely the major sugar that accumulates) (Fig. 6). Previous study on the sialometabolism gene regulation system of *V. vulnificus*, and a recent publication in *S. aureus*, identified ManNAc-6P as the SIS sugar that binds to the RpiR-type NanR repressors of both organisms to regulate their respective sialometabolism operons (Kim et al., 2011, Olson et al., 2013). Since we had not observed a Neu5Ac dependent effect on the expression of *nanE* and *siaP* in the *nanE* mutant background (where ManNAc-6P is likely the major sugar that accumulates), we have no evidence to suggest that ManNAc-6P plays a direct role in the regulation of the *Hi* sialometabolism operons under the condition tested.

The fact that the SiaR regulator of *Hi* interacts with a different SIS sugar to that of the NanR regulators of *V. vulnificus* and *S. aureus* is interesting, as all three regulators belong to the same family of proteins and all contain the HTH and the SIS domains of a RpiR-type regulator. To validate the identity of the SIS sugar that interacts with SiaR and to further our understanding on the interaction of cAMP-CRP with SiaR in the presence of the SIS sugar, we propose to perform an EMSA experiment using the SiaR protein, the cAMP-CRP complex, and the SIS sugar. Previous study has shown that the binding of SiaR to its operator site is enhanced in the presence of GlcN-6P, suggesting that GlcN-6P may function as a co-activator and SiaR as a

transcriptional activator (Johnston et al., 2010). It is worth considering the possibility and testing the hypothesis that GlcN-6P may function no differently to other SIS sugars in that it displaces the SiaR repressor from its operator site and induces gene expression, but may require the presence and contribution of cAMP-CRP. Since ManNAc-6P, the only intermediate sugar in the *Hi* Neu5Ac catabolism pathway that has not been tested before if it can affect the affinity of SiaR to its operator site, this sugar should also be included in the EMSA study to negate the possibility that ManNAc-6P has a potential role to alter the affinity of SiaR to its operator site. The role of GlcN-6P (and not of ManNAc-6P) would be made more apparent had we time to conduct the EMSA study using GlcN-6P and ManNAc-6P in the presence of SiaR, and in the presence or absence of cAMP-CRP.

From our gene expression study, we observed that *nanE* and *siaP* were both downregulated in the wild type strains compared to the culture with no added sugar following growth in 30 µg/ml Neu5Ac for an extended period of time (60 min compared to 10 min). We propose that the downregulation of *nanE* and *siaP* observed at the 60 min time point is likely due to the depletion of Neu5Ac, causing a significant reduction in the availability of the SIS sugar; and hence SiaR can bind to its operator site, causing repression of the *Hi* sialometabolism genes. Supplementing Neu5Ac at a higher concentration (500 µg/ml) would allow continued catabolism of Neu5Ac at 60 min. Consistent with our hypothesis that depletion of the SIS sugar caused the downregulation of *nanE* and *siaP* following growth in 30 µg/ml Neu5Ac for 60 min with respect to the culture without added Neu5Ac, we observed that *nanE* and *siaP* remained upregulated at the 60 min time point following growth in 500 µg/ml Neu5Ac for 60 min. Although *siaR*, located within the *Hi nagnan* operon and is co-transcribed in the same operon as *nanE* (Jenkins et al., 2010), was unregulated at the 10 min time point; the fact *nanE* and *siaP* were also unregulated at this time point implies that the SIS sugar is available to interact with SiaR and

cause derepression of the *nagnan* and *siaPQM* operons. The decrease in the availability of the SIS sugar at 60 min, and possibly also the increase in the amount of SiaR from the upregulation at 10 min, contribute to the downregulation of *nanE* and *siaP* in the wild type strains at the 60 min time point.

The fact that we observed differential expression of *nanE* and *siaP* through inactivating *nanK* (where ManNAc cannot be converted to ManNAc-6P, leading to ManNAc being the first sugar to accumulate) demonstrates that altering the balance of intermediates in the Neu5Ac catabolic pathway may cause some counter intuitive effect on the expression of *nagnan* and *siaPQM* operons. Since upregulation of *nanE* and *siaP* was observed in the *nanE* mutant background (where ManNAc-6P is likely the first sugar to accumulate) in both the absence and the presence of added Neu5Ac, this indicates that the *Hi* sialometabolism gene regulation is more complex than previously thought. The details that undermine the complexity of the *Hi* sialometabolism gene regulation are not properly understood and cAMP-CRP could contribute to the degree of complexity. To appreciate the role that cAMP-CRP plays on the regulation of the *Hi* sialometabolism operons, we need to first have a basic understanding of the possible means by which the level of cAMP could be altered in the cell. There are two potential and independent means by which the level of cAMP could be altered with respect to sialometabolism in *Hi*. Firstly, through the availability of pyruvate as a by-product produced from the first step of Neu5Ac catabolism (Fig. 6). Previous study in *E. coli* showed that an increase in the PEP to pyruvate ratio is associated with an increase in cAMP level (possibly via phosphorylation of a phosphotransferase system protein) (Cunningham et al., 2009), and it is likely that the same mechanism is present in *Hi*. Secondly, through reduced catabolism of Neu5Ac to Fru-6P, which would likely lead to less pyruvate produced, so one would expect an increase in the ratio of PEP to pyruvate to occur, resulting in an increase in cAMP level and affect the expression of the

siaPQM operon. Since the binding sites of SiaR and CRP overlap by three bps (see section 1.4.4. and Fig. 8), the two transcriptional regulators may work synergistically or antagonistically to each other. Further research in this area would be required to provide a better insight into the details of the mechanism of the *Hi* sialometabolism gene regulation.

From a biological point of view, it is likely that a limited amount of free Neu5Ac is present in the nasopharynx. Since it would benefit *Hi* to scavenge the limited amount of free Neu5Ac for decoration of its LPS molecule or for production of carbon and energy for the bacterium, it would make sense to regulate sialometabolism genes such that cells only express these genes when encountering free Neu5Ac. We postulate that in the absence of another preferred sugar source, such as glucose or fructose, Neu5Ac that is transported into the cell through the Neu5Ac TRAP transport system (encoded by *siaP* and *siaQM*) (Severi et al., 2005, Allen et al., 2005) would preferentially be utilised in catabolism to produce carbon and energy, and not in anabolism for LPS sialylation, as production of carbon and energy from Neu5Ac would likely benefit the bacterium in an environment that is scarce in other carbon sources. Our hypothesis is supported by the lack of LPS sialylation observed in strain R2846 following growth in CDM broth (a defined medium that lacks other sugar as carbon source) with Neu5Ac.

GlcN-6P that is produced as an intermediate phosphosugar from the catabolism pathway is likely the SIS sugar that interacts with the *Hi* SiaR repressor, causing derepression of the Neu5Ac transport and catabolism operons. Derepression of Neu5Ac transport operon would lead to further transport of Neu5Ac into the cell, which in turn allows more Neu5Ac to be catabolised and be utilised in LPS sialylation. Further catabolism of Neu5Ac could lead to more GlcN-6P becoming available to interact with SiaR, resulting in induction of Neu5Ac transport and catabolism operons: a positive feed-back cycle is created. When extracellular Neu5Ac becomes

depleted, there would be no more Neu5Ac transported into the cell and catabolism of Neu5Ac would come to a halt, resulting in less available GlcN-6P to interact with SiaR, and reduced gene expression. Since *siaR* (transcribed as part of the Neu5Ac catabolism operon) (Fig. 7) is induced during periods of time when Neu5Ac is available, it is likely that an increase in the amount of SiaR, in conjunction with the decrease in the availability of GlcN-6P, cause the downregulation of Neu5Ac transport and catabolism operons during times when Neu5Ac is depleted. The ability of *Hi* to create a feedback cycle to regulate the expression of sialometabolism genes appears to be an ideal way for the bacteria to only express sialometabolism genes when they encounter Neu5Ac.

Previous study has shown that sialylation of LPS is required to cause disease in the middle ear of the chinchilla (Figueira et al., 2007). The availability of Neu5Ac would be much higher in inflammatory conditions (e.g. in otitis media) than in non-inflammatory conditions (i.e. in the nasopharynx). We suggest that the acquisition of free Neu5Ac (one of the inflammatory markers produced by the host) by *Hi* during inflammation would be rapid since rapid removal of free Neu5Ac would bring fitness advantage of the bacterium through better growth from catabolism of Neu5Ac and dampen down host responses. Furthermore, sialylation of LPS would assist in bacterial survival through increasing the bacteria's resistance against complement mediated killing.

Our research presented in this thesis has demonstrated the importance of the flux of Neu5Ac on the regulation of sialometabolism genes and fine control of sialometabolism gene regulation is likely crucial for the survival of *Hi in vivo*.

8 References

Adams, J., R. Kelso & L. Cooley, (2000) The kelch repeat superfamily of proteins: propellers of cell function. *Trends Cell Biol* **10**: 17-24.

Adams, W. G., K. A. Deaver, S. L. Cochi, B. D. Plikaytis, E. R. Zell, C. V. Broome & J. D. Wenger, (1993) Decline of childhood Haemophilus influenzae type b (Hib) disease in the Hib vaccine era. *Jama* **269**: 221-226.

Alexander, H. E. & G. Leidy, (1953) Induction of streptomycin resistance in sensitive Hemophilus influenzae by extracts containing desoxyribonucleic acid from resistant Hemophilus influenzae. *The Journal of experimental medicine.* **97**: 17-31.

Allen, S., A. Zaleski, J. W. Johnston, B. W. Gibson & M. A. Apicella, (2005) Novel sialic acid transporter of Haemophilus influenzae. *Infect Immun* **73**: 5291-5300.

Anderson, P., R. A. Insel, D. H. Smith, T. R. Cate, R. B. Couch & W. P. Glezen, (1981) A polysaccharide-protein complex from Haemophilus influenzae type b. III. Vaccine trial in human adults. *The Journal of infectious diseases.* **144**: 530-538.

Angata, T. & A. Varki, (2002) Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. *Chem Rev* **102**: 439-469.

Angata, T., N. M. Varki & A. Varki, (2001) A second uniquely human mutation affecting sialic acid biology. *The Journal of biological chemistry.* **276**: 40282-40287.

Applied-Biosystems, (1997) Relative Quantitation Of Gene Expression.

Barbour, M. L., D. W. Crook & R. T. Mayon-White, (1993) An improved antiserum agar method for detecting carriage of Haemophilus influenzae type b. *Eur J Clin Microbiol Infect Dis* **12**: 215-217.

Barbour, M. L., R. T. Mayon-White, C. Coles, D. W. Crook & E. R. Moxon, (1995) The impact of conjugate vaccine on carriage of Haemophilus influenzae type b. *J Infect Dis* **171**: 93-98.

Barenkamp, S. J. & E. Leininger, (1992a) Cloning, expression, and DNA sequence analysis of genes encoding nontypeable Haemophilus influenzae high-molecular-weight surface-exposed proteins related to filamentous hemagglutinin of Bordetella pertussis. *Infection and immunity.* **60**: 1302-1313.

Barenkamp, S. J. & E. Leininger, (1992b) Cloning, expression, and DNA sequence analysis of genes encoding nontypeable Haemophilus influenzae high-molecular-weight surface-exposed proteins related to filamentous hemagglutinin of Bordetella pertussis. *Infect Immun* **60**: 1302-1313.

Barenkamp, S. J. & J. W. St Geme, 3rd, (1994) Genes encoding high-molecular-weight adhesion proteins of nontypeable Haemophilus influenzae are part of gene clusters. *Infect Immun* **62**: 3320-3328.

Barenkamp, S. J. & J. W. St Geme, 3rd, (1996) Identification of a second family of high-molecular-weight adhesion proteins expressed by non-typable Haemophilus influenzae. *Molecular microbiology.* **19**: 1215-1223.

Barry, J. M., (2004) *The Great Influenza*. Penguin Books, London.

Bateman, (1999) The SIS domain: a phosphosugar-binding domain.

Bauer, S. H., M. Mansson, D. W. Hood, J. C. Richards, E. R. Moxon & E. K. Schweda, (2001) A rapid and sensitive procedure for determination of 5-N-acetyl neuraminic acid in lipopolysaccharides of Haemophilus influenzae: a survey of 24 non-typeable H. influenzae strains. *Carbohydr Res* **335**: 251-260.

Berman, S., R. Roark & D. Luckey, (1994) Theoretical cost effectiveness of management options for children with persisting middle ear effusions. *Pediatrics* **93**: 353-363.

Biswas, M., B. Singh & A. Datta, (1979) Induction of N-acetylmannosamine catabolic pathway in yeast. *Biochim Biophys Acta* **585**: 535-542.

Boon Hinckley, M., C. M. Reynolds, A. A. Ribeiro, S. C. McGrath, R. J. Cotter, F. N. Lauw, D. T. Golenbock & C. R. Raetz, (2005) A Leptospira interrogans enzyme with similarity to yeast Ste14p that methylates the 1-phosphate group of lipid A. *J Biol Chem* **280**: 30214-30224.

Bouchet, V., D. W. Hood, J. Li, J. R. Brisson, G. A. Randle, A. Martin, Z. Li, R. Goldstein, E. K. Schweda, S. I. Pelton, J. C. Richards & E. R. Moxon, (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.

Bozue, J. A., M. V. Tullius, J. Wang, B. W. Gibson & R. S. Munson, Jr., (1999) Haemophilus ducreyi produces a novel sialyltransferase. Identification of the sialyltransferase gene and construction of mutants deficient in the production of the sialic acid-containing glycoform of the lipooligosaccharide. *The Journal of biological chemistry.* **274**: 4106-4114.

Brinkmann, V. & A. Zychlinsky, (2007) Beneficial suicide: why neutrophils die to make NETs. *Nat Rev Microbiol* **5**: 577-582.

Bustin, S. A., (2004) *A-Z of quantitative PCR*. California, USA.

Campagnari, A. A., M. R. Gupta, K. C. Dudas, T. F. Murphy & M. A. Apicella, (1987) Antigenic diversity of lipooligosaccharides of nontypable Haemophilus influenzae. *Infection and immunity.* **55(4)**: 882-887.

Campanero-Rhodes, M. A., A. Smith, W. Chai, S. Sonnino, L. Mauri, R. A. Childs, Y. Zhang, H. Ewers, A. Helenius, A. Imberty & T. Feizi, (2007) N-glycolyl GM1 ganglioside as a receptor for simian virus 40. *J Virol* **81**: 12846-12858.

Carithers, H. A., (1974) The first use of an antibiotic in America. *Am J Dis Child* **128**: 207-211.

Catlin, B. W., J. W. Bendler, 3rd & S. H. Goodgal, (1972) The type b capsulation locus of Haemophilus influenzae: map location and size. *J Gen Microbiol* **70**: 411-422.

CDC, (2002) Serotyping discrepancies in Haemophilus influenzae type b disease--United States, 1998-1999. *MMWR Morb Mortal Wkly Rep* **51**: 706-707.

Chou, H. H., H. Takematsu, S. Diaz, J. Iber, E. Nickerson, K. L. Wright, E. A. Muchmore, D. L. Nelson, S. T. Warren & A. Varki, (1998) A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence. *Proceedings of the National Academy of Sciences of the United States of America.* **95**: 11751-11756.

Chowdhury, N., J. Norris, E. McAlister, S. Y. Lau, G. H. Thomas & E. F. Boyd, (2012) The VC1777-VC1779 proteins are members of a sialic acid-specific subfamily of TRAP transporters (SiaPQM) and constitute the sole route of sialic acid uptake in the human pathogen Vibrio cholerae. *Microbiology* **158**: 2158-2167.

Cody, A. J., D. Field, E. J. Feil, S. Stringer, M. E. Deadman, A. G. Tsolaki, B. Gratz, V. Bouchet, R. Goldstein, D. W. Hood & E. R. Moxon, (2003) High rates of recombination in otitis media isolates of non-typeable Haemophilus influenzae. *Infect Genet Evol* **3**: 57-66.

Condemine, G., C. Berrier, J. Plumbridge & A. Ghazi, (2005) Function and expression of an N-acetylneuraminic acid-inducible outer membrane channel in Escherichia coli. *J Bacteriol* **187**: 1959-1965.

Costerton, J. W., Z. Lewandowski, D. E. Caldwell, D. R. Korber & H. M. Lappin-Scott, (1995) Microbial biofilms. *Annu. Rev. Microbiol.* **49**: 711-745.

Cowan, S. W., T. Schirmer, G. Rummel, M. Steiert, R. Ghosh, R. A. Pauptit, J. N. Jansonius & J. P. Rosenbusch, (1992) Crystal structures explain functional properties of two E. coli porins. *Nature* **358**: 727-733.

Cox, A. D., D. W. Hood, A. Martin, K. M. Makepeace, M. E. Deadman, J. Li, J. R. Brisson, E. R. Moxon, J. C. Richards & N. R. C. O. C. A. C. n. c. Institute for Biological Sciences,

(2002) Identification and structural characterization of a sialylated lacto-N-neotetraose structure in the lipopolysaccharide of Haemophilus influenzae. *European journal of biochemistry / FEBS.* **269(16)**: 4009-4019.

Crisel, R. M., R. S. Baker & D. E. Dorman, (1975) Capsular polymer of Haemophilus influenzae, type b. I. Structural characterization of the capsular polymer of strain Eagan. *The Journal of biological chemistry.* **250**: 4926-4930.

Cunningham, D. S., Z. Liu, N. Domagalski, R. R. Koepsel, M. M. Ataai & M. M. Domach, (2009) Pyruvate kinase-deficient Escherichia coli exhibits increased plasmid copy number and cyclic AMP levels. *J Bacteriol* **191**: 3041-3049.

Daddaoua, A., T. Krell & J. L. Ramos, (2009) Regulation of glucose metabolism in Pseudomonas: the phosphorylative branch and entner-doudoroff enzymes are regulated by a repressor containing a sugar isomerase domain. *The Journal of biological chemistry.* **284**: 21360-21368.

Deadman, M. E., S. L. Lundstrom, E. K. Schweda, E. R. Moxon & D. W. Hood, (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.

Ecevit, I. Z., K. W. McCrea, M. M. Pettigrew, A. Sen, C. F. Marrs & J. R. Gilsdorf, (2004) Prevalence of the hifBC, hmw1A, hmw2A, hmwC, and hia Genes in Haemophilus influenzae Isolates. *Journal of clinical microbiology.* **42**: 3065-3072.

Ehrlich, G. D., R. Veeh, X. Wang, J. W. Costerton, J. D. Hayes, F. Z. Hu, B. J. Daigle, M. D. Ehrlich & J. C. Post, (2002) Mucosal biofilm formation on middle-ear mucosa in the chinchilla model of otitis media. *Jama* **287**: 1710-1715.

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

Everett, E. D., A. E. Rham, Jr., R. Adaniya, D. L. Stevens & T. R. McNitt, (1977) Haemophilus influenzae pneumonia in adults. *Jama* **238**: 319-321.

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

Ferrero, M. A. & L. R. Aparicio, Biosynthesis and production of polysialic acids in bacteria. *Appl Microbiol Biotechnol* **86**: 1621-1635.

Figueira, M. A., S. Ram, R. Goldstein, D. W. Hood, E. R. Moxon & S. I. Pelton, (2007) Role of complement in defense of the middle ear revealed by restoring the virulence of nontypeable *Haemophilus influenzae siaB* mutants. *Infect. Immun.* **75**: 325-333.

Fink D. & Geme J., (2006) The Genus Haemophilus. In: The Prokaryotes (3rd edition), Proteobacteria. S. E. Rosenberg E., Thompson F., Lory S., DeLong E. (ed). Verlag Berlin Heidelberg: SpringerReference, pp.

Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick & et al., (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science.* **269**: 496-512.

Flesher, A. R. & R. A. Insel, (1978) Characterization of lipopolysaccharide of Haemophilus influenzae. *The Journal of infectious diseases.* **138**: 719-730.

Forsgren, J., A. Samuelson, A. Ahlin, J. Jonasson, B. Rynnel-Dagoo & A. Lindberg, (1994) Haemophilus influenzae resides and multiplies intracellularly in human adenoid tissue as demonstrated by in situ hybridization and bacterial viability assay. *Infection and immunity.* **62**: 673-679.

Forward, J. A., M. C. Behrendt, N. R. Wyborn, R. Cross & D. J. Kelly, (1997) TRAP transporters: a new family of periplasmic solute transport systems encoded by the

dctPQM genes of Rhodobacter capsulatus and by homologs in diverse gram-negative bacteria. *J Bacteriol* **179**: 5482-5493.

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

Frosch, M. & A. Muller, (1993) Phospholipid substitution of capsular polysaccharides and mechanisms of capsule formation in Neisseria meningitidis. *Mol Microbiol* **8**: 483-493.

Garcia-Rodriguez, J. A. & M. J. Fresnadillo Martinez, (2002) Dynamics of nasopharyngeal colonization by potential respiratory pathogens. *J Antimicrob Chemother* **50 Suppl S2**: 59-73.

Geme, J. W., 3rd & D. Cutter, (1995) Evidence that surface fibrils expressed by Haemophilus influenzae type b promote attachment to human epithelial cells. *Mol Microbiol* **15**: 77-85.

Gilbert, M., J. R. Brisson, M. F. Karwaski, J. Michniewicz, A. M. Cunningham, Y. Wu, N. M. Young & W. W. Wakarchuk, (2000) Biosynthesis of ganglioside mimics in Campylobacter jejuni OH4384. Identification of the glycosyltransferase genes, enzymatic synthesis of model compounds, and characterization of nanomole amounts by 600-mhz (1)h and (13)c NMR analysis. *J Biol Chem* **275**: 3896-3906.

Gilbert, M., D. C. Watson, A. M. Cunningham, M. P. Jennings, N. M. Young & W. W. Wakarchuk, (1996) Cloning of the lipooligosaccharide alpha-2,3-sialyltransferase from the bacterial pathogens Neisseria meningitidis and Neisseria gonorrhoeae. *The Journal of biological chemistry.* **271**: 28271-28276.

Gottschalk, A., (1960) Sialic acids: their molecular structure and characteristic reactions. *Arzneimittelforschung* **10**: 355-357.

Granick, S. & H. Gilder, (1946) The Porphyrin Requirements of Haemophilus Influenzae and Some Functions of the Vinyl and Propionic Acid Side Chains of Heme. *J Gen Physiol* **30**: 1-13.

Granoff, D. M., M. Basden & R. Rockwell, (1978) Ampicillin-resistant Hemophilus influenzae. *West J Med* **128**: 101-105.

Greiner, L. L., H. Watanabe, N. J. Phillips, J. Shao, A. Morgan, A. Zaleski, B. W. Gibson & M. A. Apicella, (2004) Nontypeable *Haemophilus influenzae* strain 2019 produces a biofilm containing N-acetylneuraminic acid that may mimic sialylated O-linked glycans. *Infect. Immun.* **72**: 4249-4260.

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

Groeneveld, K., L. van Alphen, P. P. Eijk, H. M. Jansen & H. C. Zanen, (1988) Changes in outer membrane proteins of nontypable Haemophilus influenzae in patients with chronic obstructive pulmonary disease. *The Journal of infectious diseases.* **158**: 360-365.

Gronow, S., W. Brabetz, B. Lindner & H. Brade, (2005) OpsX from Haemophilus influenzae represents a novel type of heptosyltransferase I in lipopolysaccharide biosynthesis. *J Bacteriol* **187**: 6242-6247.

Guerina, N. G., S. Langermann, H. W. Clegg, T. W. Kessler, D. A. Goldman & J. R. Gilsdorf, (1982) Adherence of piliated Haemophilus influenzae type b to human oropharyngeal cells. *The Journal of infectious diseases.* **146**: 564.

Hardy, G. G., S. M. Tudor & J. W. St Geme, 3rd, (2003) The pathogenesis of disease due to nontypeable Haemophilus influenzae. *Methods Mol Med* **71**: 1-28.

Helander, I. M., B. Lindner, H. Brade, K. Altmann, A. A. Lindberg, E. T. Rietschel & U. Zahringer, (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. & J. W. St Geme, 3rd, (1998) The Haemophilus influenzae Hap serine protease promotes adherence and microcolony formation, potentiated by a soluble host protein. *Mol Cell* **2**: 841-850.

Herriott, R. M., E. M. Meyer & M. Vogt, (1970) Defined nongrowth media for stage II development of competence in Haemophilus influenzae. *J Bacteriol* **101**: 517-524.

Higuchi, R., C. Fockler, G. Dollinger & R. Watson, (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology (N Y)* **11**: 1026-1030.

Holmquist, L. & B. Ostman, (1975) The anomeric configuration of N-acetylneuraminic acid released by the action of Vibrio cholerae neuraminidase. *FEBS Lett* **60**: 327-330.

Hood, D. W., A. D. Cox, M. Gilbert, K. Makepeace, S. Walsh, M. E. Deadman, A. Cody, A. Martin, M. Mansson, E. K. Schweda, J. R. Brisson, J. C. Richards, E. R. Moxon & W. W. Wakarchuk, (2001a) Identification of a lipopolysaccharide alpha-2,3-sialyltransferase from Haemophilus influenzae. *Mol Microbiol* **39**: 341-350.

Hood, D. W., A. D. Cox, W. W. Wakarchuk, M. Schur, E. K. Schweda, S. L. Walsh, M. E. Deadman, A. Martin, E. R. Moxon & J. C. Richards, (2001b) Genetic basis for expression of the major globotetraose-containing lipopolysaccharide from H. influenzae strain Rd (RM118). *Glycobiology* **11**: 957-967.

Hood, D. W., K. Makepeace, M. E. Deadman, R. F. Rest, P. Thibault, A. Martin, J. C. Richards & E. R. Moxon, (1999a) 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., G. Randle, A. D. Cox, K. Makepeace, J. Li, E. K. Schweda, J. C. Richards & E. R. Moxon, (2004) 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., J. C. Richards & E. R. Moxon, (1999b) Haemophilus influenzae lipopolysaccharide. *Biochem Soc Trans* **27**: 493-498.

Houliston, R. S., M. Koga, J. Li, H. C. Jarrell, J. C. Richards, V. Vitiazeva, E. K. Schweda, N. Yuki & M. Gilbert, (2007) A Haemophilus influenzae strain associated with Fisher syndrome expresses a novel disialylated ganglioside mimic. *Biochemistry* **46**: 8164-8171.

Howard, A. J., K. T. Dunkin & G. W. Millar, (1988) Nasopharyngeal carriage and antibiotic resistance of Haemophilus influenzae in healthy children. *Epidemiol Infect* **100**: 193-203.

Irie, A., S. Koyama, Y. Kozutsumi, T. Kawasaki & A. Suzuki, (1998) The molecular basis for the absence of N-glycolylneuraminic acid in humans. *The Journal of biological chemistry.* **273**: 15866-15871.

Iwasaki, M., S. Inoue & F. A. Troy, (1990) A new sialic acid analogue, 9-O-acetyl-deaminated neuraminic acid, and alpha -2,8-linked O-acetylated poly(N-glycolylneuraminyl) chains in a novel polysialoglycoprotein from salmon eggs. *The Journal of biological chemistry.* **265**: 2596-2602.

Jacobs, B. C., P. H. Rothbarth, F. G. van der Meche, P. Herbrink, P. I. Schmitz, M. A. de Klerk & P. A. van Doorn, (1998) The spectrum of antecedent infections in Guillain-Barre syndrome: a case-control study. *Neurology* **51**: 1110-1115.

Jaeger, T. & C. Mayer, (2008) The transcriptional factors MurR and catabolite activator protein regulate N-acetylmuramic acid catabolism in Escherichia coli. *J Bacteriol* **190**: 6598-6608.

Jaques, L. W., E. B. Brown, J. M. Barrett & W. S. J. W. W. Brey, Jr., (1977) Sialic acid. A calcium-binding carbohydrate. *The Journal of biological chemistry.* **252**: 4533-4538.

Jenkins, G. A., M. Figueira, G. A. Kumar, W. A. Sweetman, K. Makepeace, S. I. Pelton, R. Moxon & D. W. Hood, (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.

Johnson, P. D., F. Oppedisano, V. Bennett-Wood, G. L. Gilbert & R. M. Robins-Browne, (1996) Sporadic invasion of cultured epithelial cells by Haemophilus influenzae type b. *Infect Immun* **64**: 1051-1053.

Johnson, R. W., G. McGillivary, P. Denoel, J. Poolman & L. O. Bakaletz, (2010) Abrogation of nontypeable Haemophilus influenzae Protein D function reduces phosphorylcholine decoration, adherence to airway epithelial cells, and fitness in a chinchilla model of otitis media. *Vaccine*.

Johnson, T. A., J. Qiu, A. G. Plaut & T. Holyoak, (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., H. Shamsulddin, A. F. Miller & M. A. Apicella, (2010) Sialic acid transport and catabolism are cooperatively regulated by SiaR and CRP in nontypeable Haemophilus influenzae. *BMC Microbiol* **10**: 240.

Johnston, J. W., A. Zaleski, S. Allen, J. M. Mootz, D. Armbruster, B. W. Gibson, M. A. Apicella & R. S. Munson, Jr., (2007) Regulation of sialic acid transport and catabolism in Haemophilus influenzae. *Mol Microbiol* **66**: 26-39.

Jokipii, A. M. & L. Jokipii, (1977) Haemophilus influenzae in otitis media and sinusitis: serotypes and susceptibilty to ampicillin and amoxycillin in vitro. *Infection* **5**: 140-143.

Jones, P. A., N. M. Samuels, N. J. Phillips, R. S. Munson, Jr., J. A. Bozue, J. A. Arseneau, W. A. Nichols, A. Zaleski, B. W. Gibson & M. A. Apicella, (2002) *Haemophilus influenzae* type b strain A2 has multiple sialyltransferases involved in lipooligosaccharide sialylation. *J. Biol. Chem.* **277**: 14598-14611.

Jourdian, G. W. & S. Roseman, (1962) The sialic acids. II. Preparation of N-glycolylhexosamines, N-glycolylhexosamine 6-phosphates, glycolyl coenzyme A, and glycolyl glutathione. *The Journal of biological chemistry.* **237**: 2442-2446.

Jurcisek, J., L. Greiner, H. Watanabe, A. Zaleski, M. A. Apicella & L. O. Bakaletz, (2005) Role of sialic acid and complex carbohydrate biosynthesis in biofilm formation by nontypeable *Haemophilus influenzae* in the chinchilla middle ear. *Infect. Immun.* **73**: 3210-3218.

Jurcisek, J. A. & L. O. Bakaletz, (2007) Biofilms formed by nontypeable *Haemophilus influenzae* in vivo contain both double-stranded DNA and type IV pilin protein. *J. Bacteriol.* **189**: 3868-3875.

Kahler, C. M. & D. S. Stephens, (1998) Genetic basis for biosynthesis, structure, and function of meningococcal lipooligosaccharide (endotoxin). *Crit Rev Microbiol* **24**: 281-334.

Kalivoda, K. A., S. M. Steenbergen, E. R. Vimr & J. Plumbridge, (2003) Regulation of sialic acid catabolism by the DNA binding protein NanR in Escherichia coli. *J Bacteriol* **185**: 4806-4815.

Kanamori, A., S. Inoue, M. Iwasaki, K. Kitajima, G. Kawai, S. Yokoyama & Y. Inoue, (1990) Deaminated neuraminic acid-rich glycoprotein of rainbow trout egg vitelline envelope. Occurrence of a novel alpha-2,8-linked oligo(deaminated neuraminic acid) structure in O-linked glycan chains. *The Journal of biological chemistry.* **265**: 21811-21819.

Kelly, D. J. & G. H. Thomas, (2001) The tripartite ATP-independent periplasmic (TRAP) transporters of bacteria and archaea. *FEMS Microbiol Rev* **25**: 405-424.

Kelly, T. J., Jr. & H. O. Smith, (1970) A restriction enzyme from Hemophilus influenzae. II. *J Mol Biol* **51**: 393-409.

Kett, K., P. Brandtzaeg, J. Radl & J. J. Haaijman, (1986) Different subclass distribution of IgA-producing cells in human lymphoid organs and various secretory tissues. *Journal of immunology (Baltimore, Md. : 1950)* **136**: 3631-3635.

Kilian, M., (1991) Manual of clinical microbiology, 5th ed. In. W. J. H. A. Balows, Jr., K.L. Herrmann, H.D. Isenberg, H.J. Shadomy (ed). Washington, D.C.: American Society for Microbiology, pp. 463-470.

Kilian, M., J. Mestecky & R. E. Schrohenloher, (1979) Pathogenic species of the genus Haemophilus and Streptococcus pneumoniae produce immunoglobulin A1 protease. *Infect Immun* **26**: 143-149.

Kilian, M., J. Reinholdt, H. Lomholt, K. Poulsen & E. V. Frandsen, (1996) Biological significance of IgA1 proteases in bacterial colonization and pathogenesis: critical evaluation of experimental evidence. *APMIS* **104**: 321-338.

Kim, B. S., J. Hwang, M. H. Kim & S. H. Choi, (2011) Cooperative regulation of the Vibrio vulnificus nan gene cluster by NanR protein, cAMP receptor protein, and N-acetylmannosamine 6-phosphate. *The Journal of biological chemistry.* **286**: 40889-40899.

Kimura, A. & E. J. Hansen, (1986) Antigenic and phenotypic variations of Haemophilus influenzae type b lipopolysaccharide and their relationship to virulence. *Infect Immun* **51**: 69-79.

Koga, M., M. Gilbert, J. Li, S. Koike, M. Takahashi, K. Furukawa, K. Hirata & N. Yuki, (2005) Antecedent infections in Fisher syndrome: a common pathogenesis of molecular mimicry. *Neurology* **64**: 1605-1611.

Kohler, P. R., E. L. Choong & S. Rossbach, The RpiR-like repressor IolR regulates inositol catabolism in Sinorhizobium meliloti. *J Bacteriol* **193**: 5155-5163.

Kraehenbuhl, J. P. & M. R. Neutra, (1992) Molecular and cellular basis of immune protection of mucosal surfaces. *Physiol Rev* **72**: 853-879.

Kroll, J. S., B. Loynds, L. N. Brophy & E. R. Moxon, (1990) The bex locus in encapsulated Haemophilus influenzae: a chromosomal region involved in capsule polysaccharide export. *Mol Microbiol* **4**: 1853-1862.

Kroll, J. S., B. M. Loynds & E. R. Moxon, (1991) The Haemophilus influenzae capsulation gene cluster: a compound transposon. *Molecular microbiology.* **5**: 1549-1560.

Kroll, J. S., E. R. Moxon & O. U. U. K. Department of Paediatrics, (1988) Capsulation and gene copy number at the cap locus of Haemophilus influenzae type b. *Journal of bacteriology.* **170(2)**: 859-864.

Kroll, J. S., S. Zamze, B. Loynds & E. R. Moxon, (1989) Common organization of chromosomal loci for production of different capsular polysaccharides in Haemophilus influenzae. *J Bacteriol* **171**: 3343-3347.

Krulwich, T. A., (1983) Na+/H+ antiporters. *Biochim Biophys Acta* **726**: 245-264.

Kubiet, M., R. Ramphal, A. Weber & A. Smith, (2000) Pilus-mediated adherence of Haemophilus influenzae to human respiratory mucins. *Infect Immun* **68**: 3362-3367.

Kyogashima, M., V. Ginsburg & H. C. Krivan, (1989) Escherichia coli K99 binds to N-glycolylsialoparagloboside and N-glycolyl-GM3 found in piglet small intestine. *Arch Biochem Biophys* **270**: 391-397.

LaClaire, L. L., M. L. Tondella, D. S. Beall, C. A. Noble, P. L. Raghunathan, N. E. Rosenstein & T. Popovic, (2003) Identification of Haemophilus influenzae serotypes by standard slide agglutination serotyping and PCR-based capsule typing. *Journal of clinical microbiology.* **41**: 393-396.

Lampe, R. M., E. O. Mason, Jr., S. L. Kaplan, C. L. Umstead, M. D. Yow & R. D. Feigin, (1982) Adherence of Haemophilus influenzae to buccal epithelial cells. *Infect Immun* **35**: 166-172.

Leibovitz, E., A. Broides, D. Greenberg & N. Newman, (2010) Current management of pediatric acute otitis media. *Expert Rev Anti Infect Ther* **8**: 151-161.

Lesse, A. J., A. A. Campagnari, W. E. Bittner & M. A. Apicella, (1990) Increased resolution of lipopolysaccharides and lipooligosaccharides utilizing tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *J Immunol Methods* **126**: 109-117.

Lethem, R. a., (1997) Airway mucus: basic mechanisms and clinical perspectives. In. Switzerland, pp.

Levine, M. M., J. B. Kaper, R. E. Black & M. L. Clements, (1983) New knowledge on pathogenesis of bacterial enteric infections as applied to vaccine development. *Microbiol Rev* **47**: 510-550.

Levinson, G. & G. A. Gutman, (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* **4**: 203-221.

Lewis, A. L., N. Desa, E. E. Hansen, Y. A. Knirel, J. I. Gordon, P. Gagneux, V. Nizet & A. Varki, (2009) Innovations in host and microbial sialic acid biosynthesis revealed by phylogenomic prediction of nonulosonic acid structure. *Proceedings of the National Academy of Sciences of the United States of America.* **106**: 13552-13557.

Li, J., S. H. Bauer, M. Mêansson, E. R. Moxon, J. C. Richards, E. K. Schweda & N. R. C. o. C. O. O. C. K. A. O. R. Institute for Biological Sciences, (2001) Glycine is a common substituent of the inner core in Haemophilus influenzae lipopolysaccharide. *Glycobiology.* **11(12)**: 1009-1015.

Li, J., M. Dzieciatkowska, D. W. Hood, A. D. Cox, E. K. Schweda, E. R. Moxon & J. C. Richards, (2007) Structural characterization of sialylated glycoforms of H. influenzae by electrospray mass spectrometry: fragmentation of protonated and sodiated O-deacylated lipopolysaccharides. *Rapid Commun Mass Spectrom* **21**: 952-960.

Lo, R. Y., L. J. McKerral, T. L. Hills & M. Kostrzynska, (2001) Analysis of the capsule biosynthetic locus of Mannheimia (Pasteurella) haemolytica A1 and proposal of a nomenclature system. *Infect Immun* **69**: 4458-4464.

Lomholt, H., L. van Alphen & M. Kilian, (1993) Antigenic variation of immunoglobulin A1 proteases among sequential isolates of Haemophilus influenzae from healthy children and patients with chronic obstructive pulmonary disease. *Infection and immunity.* **61**: 4575-4581.

Loos, B. G., J. M. Bernstein, D. M. Dryja, T. F. Murphy & D. P. Dickinson, (1989) Determination of the epidemiology and transmission of nontypable Haemophilus influenzae in children with otitis media by comparison of total genomic DNA restriction fingerprints. *Infect Immun* **57**: 2751-2757.

Lundstrom, S. L., J. Li, M. E. Deadman, D. W. Hood, E. R. Moxon & E. K. Schweda, (2008) Structural analysis of the lipopolysaccharide from nontypeable Haemophilus influenzae strain R2846. *Biochemistry* **47**: 6025-6038.

Lysenko, E., J. C. Richards, A. D. Cox, A. Stewart, A. Martin, M. Kapoor & J. N. Weiser, (2000) 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.

Macfadyen, L. P., I. R. Dorocicz, J. Reizer, M. H. Saier, Jr. & R. J. Redfield, (1996) Regulation of competence development and sugar utilization in Haemophilus influenzae Rd by a phosphoenolpyruvate:fructose phosphotransferase system. *Mol Microbiol* **21**: 941-952.

Macfadyen, L. P. & R. J. Redfield, (1996) Life in mucus: sugar metabolism in Haemophilus influenzae. *Res Microbiol* **147**: 541-551.

MacNeil, J. R., A. C. Cohn, M. Farley, R. Mair, J. Baumbach, N. Bennett, K. Gershman, L. H. Harrison, R. Lynfield, S. Petit, A. Reingold, W. Schaffner, A. Thomas, F. Coronado, E. R. Zell, L. W. Mayer, T. A. Clark & N. E. Messonnier, (2011) Current epidemiology and

trends in invasive Haemophilus influenzae disease--United States, 1989-2008. *Clin Infect Dis* **53**: 1230-1236.

Mandrell, R. E., R. McLaughlin, Y. Aba Kwaik, A. Lesse, R. Yamasaki, B. Gibson, S. M. Spinola & M. A. Apicella, (1992) Lipooligosaccharides (LOS) of some Haemophilus species mimic human glycosphingolipids, and some LOS are sialylated. *Infect Immun* **60**: 1322-1328.

Mansson, M., S. H. Bauer, D. W. Hood, J. C. Richards, E. R. Moxon & E. K. Schweda, (2001) A new structural type for Haemophilus influenzae lipopolysaccharide. Structural analysis of the lipopolysaccharide from nontypeable Haemophilus influenzae strain 486. *Eur J Biochem* **268**: 2148-2159.

Mansson, M., D. W. Hood, E. R. Moxon & E. K. Schweda, (2003) Structural diversity in lipopolysaccharide expression in nontypeable Haemophilus influenzae. Identification of L-glycerol-D-manno-heptose in the outer-core region in three clinical isolates. *Eur J Biochem* **270**: 610-624.

Manzi, A. E., A. Dell, P. Azadi & A. Varki, (1990) Studies of naturally occurring modifications of sialic acids by fast-atom bombardment-mass spectrometry. Analysis of positional isomers by periodate cleavage. *The Journal of biological chemistry.* **265**: 8094-8107.

Martin, M. J., J. C. Rayner, P. Gagneux, J. W. Barnwell & A. Varki, (2005) Evolution of human-chimpanzee differences in malaria susceptibility: relationship to human genetic loss of N-glycolylneuraminic acid. *Proceedings of the National Academy of Sciences of the United States of America.* **102**: 12819-12824.

Masoud, H., E. R. Moxon, A. Martin, D. Krajcarski & J. C. Richards, (1997) Structure of the variable and conserved lipopolysaccharide oligosaccharide epitopes expressed by Haemophilus influenzae serotype b strain Eagan. *Biochemistry* **36**: 2091-2103.

Massa, H. M., A. W. Cripps & D. Lehmann, (2009) Otitis media: viruses, bacteria, biofilms and vaccines. *Med J Aust* **191**: S44-49.

Massol-Deya, A. A., J. Whallon, R. F. Hickey & J. M. Tiedje, (1995) Channel structures in aerobic biofilms of fixed-film reactors treating contaminated groundwater. *Appl Environ Microbiol* **61**: 769-777.

Mencarelli, M., C. Marsili, A. Zanchi, C. Pantini & C. Cellesi, (1993) Genomic DNA fingerprints and phenotypic characteristics of serotype B Haemophilus influenzae isolates from Italy. *Eur J Epidemiol* **9**: 353-360.

Meredith, T. C. & R. W. Woodard, (2003) Escherichia coli YrbH is a D-arabinose 5-phosphate isomerase. *The Journal of biological chemistry.* **278**: 32771-32777.

Mizanur, R. M. & N. L. Pohl, (2008) Bacterial CMP-sialic acid synthetases: production, properties, and applications. *Appl Microbiol Biotechnol* **80**: 757-765.

Moxon, E. R. & P. A. Murphy, (1978) Haemophilus influenzae bacteremia and meningitis resulting from survival of a single organism. *Proceedings of the National Academy of Sciences of the United States of America.* **75**: 1534-1536.

Moxon, E. R., W. A. Sweetman, M. E. Deadman, D. J. Ferguson & D. W. Hood, (2008) Haemophilus influenzae biofilms: hypothesis or fact? *Trends Microbiol* **16**: 95-100.

Moxon, E. R. & K. A. Vaughn, (1981) The type b capsular polysaccharide as a virulence determinant of Haemophilus influenzae: studies using clinical isolates and laboratory transformants. *The Journal of infectious diseases.* **143**: 517-524.

Mulks, M. H., S. J. Kornfeld, B. Frangione & A. G. Plaut, (1982) Relationship between the specificity of IgA proteases and serotypes in Haemophilus influenzae. *The Journal of infectious diseases.* **146**: 266-274.

Muller, A., E. Severi, C. Mulligan, A. G. Watts, D. J. Kelly, K. S. Wilson, A. J. Wilkinson & G. H. Thomas, (2006) Conservation of structure and mechanism in primary and secondary

transporters exemplified by SiaP, a sialic acid binding virulence factor from Haemophilus influenzae. *J Biol Chem* **281**: 22212-22222.

Munson, R., Jr. & A. Hunt, (1989) Isolation and characterization of a mutant of Haemophilus influenzae type b deficient in outer membrane protein P1. *Infection and immunity.* **57**: 1002-1004.

Murphy, T. F. & M. A. Apicella, (1985) Antigenic heterogeneity of outer membrane proteins of nontypable Haemophilus influenzae is a basis for a serotyping system. *Infection and immunity.* **50(1)**: 15-21.

Murphy, T. F., L. C. Bartos, A. A. Campagnari, M. B. Nelson & M. A. Apicella, (1986) Antigenic characterization of the P6 protein of nontypable Haemophilus influenzae. *Infect Immun* **54**: 774-779.

Murphy, T. F., K. C. Dudas, J. M. Mylotte & M. A. Apicella, (1983) A subtyping system for nontypable Haemophilus influenzae based on outer-membrane proteins. *The Journal of infectious diseases.* **147(5)**: 838-846.

Murphy, T. F. & S. Sethi, (1992) Bacterial infection in chronic obstructive pulmonary disease. *Am Rev Respir Dis* **146**: 1067-1083.

Murphy, T. V., J. F. Clements, M. Petroni, S. Coury & L. Stetler, (1989) Haemophilus influenzae type b in respiratory secretions. *The Pediatric infectious disease journal.* **8**: 148-151.

Musser, J. M., S. J. Barenkamp, D. M. Granoff & R. K. Selander, (1986) Genetic relationships of serologically nontypable and serotype b strains of Haemophilus influenzae. *Infect Immun* **52**: 183-191.

Nadano, D., M. Iwasaki, S. Endo, K. Kitajima, S. Inoue & Y. Inoue, (1986) A naturally occurring deaminated neuraminic acid, 3-deoxy-D-glycero-D-galacto-nonulosonic acid (KDN). Its unique occurrence at the nonreducing ends of oligosialyl chains in polysialoglycoprotein of rainbow trout eggs. *The Journal of biological chemistry.* **261**: 11550-11557.

Nguyen, D. H., P. Tangvoranuntakul & A. Varki, (2005) Effects of natural human antibodies against a nonhuman sialic acid that metabolically incorporates into activated and malignant immune cells. *J Immunol* **175**: 228-236.

Nichols, W. A., B. W. Gibson, W. Melaugh, N. G. Lee, M. Sunshine & M. A. Apicella, (1997) Identification of the ADP-L-glycero-D-manno-heptose-6-epimerase (rfaD) and heptosyltransferase II (rfaF) biosynthesis genes from nontypeable Haemophilus influenzae 2019. *Infect Immun* **65**: 1377-1386.

Nickel, J. C., J. B. Wright, I. Ruseska, T. J. Marrie, C. Whitfield & J. W. Costerton, (1985) Antibiotic resistance of Pseudomonas aeruginosa colonizing a urinary catheter in vitro. *Eur J Clin Microbiol* **4**: 213-218.

Noel, G. J., A. Brittingham, A. A. Granato & D. M. Mosser, (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.

O'Neill, J. M., J. W. St Geme, 3rd, D. Cutter, E. E. Adderson, J. Anyanwu, R. F. Jacobs & G. E. Schutze, (2003) Invasive disease due to nontypeable Haemophilus influenzae among children in Arkansas. *Journal of clinical microbiology.* **41**: 3064-3069.

Olitsky P.K., G. F. L., (1921) EXPERIMENTAL STUDIES OF THE NASOPHARYNGEAL SECRETIONS FROM INFLUENZA PATIENTS. *The Journal of experimental medicine.* **33**: 713-729.

Olson, M. E., J. M. King, T. L. Yahr & A. R. Horswill, (2013) Sialic Acid Catabolism in Staphylococcus aureus. *J Bacteriol*.

Pangburn, M. K., K. L. Pangburn, V. Koistinen, S. Meri & A. K. Sharma, (2000) Molecular mechanisms of target recognition in an innate immune system: interactions among factor

H, C3b, and target in the alternative pathway of human complement. *J Immunol* **164**: 4742-4751.

Paradise, J. L., H. E. Rockette, D. K. Colborn, B. S. Bernard, C. G. Smith, M. Kurs-Lasky & J. E. Janosky, (1997) Otitis media in 2253 Pittsburgh-area infants: prevalence and risk factors during the first two years of life. *Pediatrics* **99**: 318-333.

Peltola, H., (2000) Worldwide Haemophilus influenzae type b disease at the beginning of the 21st century: global analysis of the disease burden 25 years after the use of the polysaccharide vaccine and a decade after the advent of conjugates. *Clin Microbiol Rev* **13**: 302-317.

Persson, K., H. D. Ly, M. Dieckelmann, W. W. Wakarchuk, S. G. Withers & N. C. Strynadka, (2001) Crystal structure of the retaining galactosyltransferase LgtC from Neisseria meningitidis in complex with donor and acceptor sugar analogs. *Nat Struct Biol* **8**: 166-175.

Phillips, N. J., M. A. Apicella, J. M. Griffiss & B. W. Gibson, (1993) Structural studies of the lipooligosaccharides from Haemophilus influenzae type b strain A2. *Biochemistry* **32**: 2003-2012.

Phillips, N. J., R. McLaughlin, T. J. Miller, M. A. Apicella & B. W. Gibson, (1996) Characterization of two transposon mutants from Haemophilus influenzae type b with altered lipooligosaccharide biosynthesis. *Biochemistry* **35**: 5937-5947.

Pichichero, M. E., M. Loeb, Anderson & D. H. Smith, (1982) Do pili play a role in pathogenicity of Haemophilus influenzae type B? *Lancet* **2**: 960-962.

Pittman, M., (1931) Variation and Type Specificity in the Bacterial Species Hemophilus Influenzae. *The Journal of experimental medicine.* **53**: 471-492.

Platonov, A. E., G. A. Shipulin & O. V. Platonova, (2000) [Multilocus sequencing--a new method of genotyping bacteria and first results of its use]. *Genetika* **36**: 597-605.

Plumbridge, J. & E. Vimr, (1999) Convergent pathways for utilization of the amino sugars N-acetylglucosamine, N-acetylmannosamine, and N-acetylneuraminic acid by Escherichia coli. *J Bacteriol* **181**: 47-54.

Porras, O., D. A. Caugant, B. Gray, T. Lagergard, B. R. Levin & C. Svanborg-Eden, (1986a) Difference in structure between type b and nontypable Haemophilus influenzae populations. *Infection and immunity.* **53**: 79-89.

Porras, O., D. A. Caugant, T. Lagergard & C. Svanborg-Eden, (1986b) Application of multilocus enzyme gel electrophoresis to Haemophilus influenzae. *Infection and immunity.* **53**: 71-78.

Post, J. C., (2001) Direct evidence of bacterial biofilms in otitis media. *Laryngoscope* **111**: 2083-2094.

Poulsen, K., J. Reinholdt & M. Kilian, (1992) A comparative genetic study of serologically distinct Haemophilus influenzae type 1 immunoglobulin A1 proteases. *Journal of bacteriology.* **174**: 2913-2921.

Prymula, R., P. Peeters, V. Chrobok, P. Kriz, E. Novakova, E. Kaliskova, I. Kohl, P. Lommel, J. Poolman, J. P. Prieels & L. Schuerman, (2006) Pneumococcal capsular polysaccharides conjugated to protein D for prevention of acute otitis media caused by both Streptococcus pneumoniae and non-typable Haemophilus influenzae: a randomised double-blind efficacy study. *Lancet* **367**: 740-748.

Rabus, R., D. L. Jack, D. J. Kelly & M. H. Saier, Jr., (1999) TRAP transporters: an ancient family of extracytoplasmic solute-receptor-dependent secondary active transporters. *Microbiology* **145 (Pt 12)**: 3431-3445.

Raetz, C. R., (1990) Biochemistry of endotoxins. *Annu Rev Biochem* **59**: 129-170.

Raghunathan, A., N. D. Price, M. Y. Galperin, K. S. Makarova, S. Purvine, A. F. Picone, T. Cherny, T. Xie, T. J. Reilly, R. Munson, Jr., R. E. Tyler, B. J. Akerley, A. L. Smith, B. O.

Palsson & E. Kolker, (2004) In Silico Metabolic Model and Protein Expression of Haemophilus influenzae Strain Rd KW20 in Rich Medium. *OMICS* **8**: 25-41.

Ram, S., A. K. Sharma, S. D. Simpson, S. Gulati, D. P. McQuillen, M. K. Pangburn & P. A. Rice, (1998) A novel sialic acid binding site on factor H mediates serum resistance of sialylated Neisseria gonorrhoeae. *The Journal of experimental medicine.* **187**: 743-752.

Rao, V. K., G. P. Krasan, D. R. Hendrixson, S. Dawid & J. W. St Geme, 3rd, (1999) Molecular determinants of the pathogenesis of disease due to non-typable Haemophilus influenzae. *FEMS Microbiol Rev* **23**: 99-129.

Rayner, M. G., Y. Zhang, M. C. Gorry, Y. Chen, J. C. Post & G. D. Ehrlich, (1998) Evidence of bacterial metabolic activity in culture-negative otitis media with effusion. *Jama* **279**: 296-299.

Reddy, M. S., J. M. Bernstein, T. F. Murphy & H. S. Faden, (1996) Binding between outer membrane proteins of nontypeable Haemophilus influenzae and human nasopharyngeal mucin. *Infect Immun* **64**: 1477-1479.

Riley, H. D., Jr., (1977) The dilemma of invasive disease due to Hemophilus influenzae. *South Med J* **70**: 133-135.

Rinninger, A., C. Richet, A. Pons, G. Kohla, R. Schauer, H. C. Bauer, J. P. Zanetta & R. Vlasak, (2006) Localisation and distribution of O-acetylated N-acetylneuraminic acids, the endogenous substrates of the hemagglutinin-esterases of murine coronaviruses, in mouse tissue. *Glycoconj J* **23**: 73-84.

Risberg, A., G. Alvelius & E. K. Schweda, (1999) Structural analysis of the lipopolysaccharide oligosaccharide epitopes expressed by Haemophilus influenzae strain RM.118-26. *Eur J Biochem* **265**: 1067-1074.

Robbe-Masselot, C., E. Maes, M. Rousset, J. C. Michalski & C. Capon, (2009) Glycosylation of human fetal mucins: a similar repertoire of O-glycans along the intestinal tract. *Glycoconj J* **26**: 397-413.

Robbe, C., C. Capon, E. Maes, M. Rousset, A. Zweibaum, J. P. Zanetta & J. C. Michalski, (2003) Evidence of regio-specific glycosylation in human intestinal mucins: presence of an acidic gradient along the intestinal tract. *The Journal of biological chemistry.* **278**: 46337-46348.

Roberts, M., T. L. Stull & A. L. Smith, (1981) Comparative virulence of Haemophilus influenzae with a type b or type d capsule. *Infect Immun* **32**: 518-524.

Rolsma, M. D., T. B. Kuhlenschmidt, H. B. Gelberg & M. S. Kuhlenschmidt, (1998) Structure and function of a ganglioside receptor for porcine rotavirus. *J Virol* **72**: 9079-9091.

Romano, J. D. & S. Michaelis, (2001) Topological and mutational analysis of Saccharomyces cerevisiae Ste14p, founding member of the isoprenylcysteine carboxyl methyltransferase family. *Mol Biol Cell* **12**: 1957-1971.

Roseman, S., (1970) The synthesis of complex carbohydrates by multiglycosyltransferase systems and their potential function in intercellular adhesion. *Chem Phys Lipids* **5**: 270-297.

Roy, S., C. W. Douglas & G. P. Stafford, (2010) A novel sialic acid utilization and uptake system in the periodontal pathogen Tannerella forsythia. *J Bacteriol* **192**: 2285-2293.

Rubin, L. G. & E. R. Moxon, (1983) Pathogenesis of bloodstream invasion with Haemophilus influenzae type b. *Infect Immun* **41**: 280-284.

Schauer, R., (1987) Analysis of sialic acids. *Methods Enzymol* **138**: 132-161.

Schauer, R., (2004) Sialic acids: fascinating sugars in higher animals and man. *Zoology (Jena)* **107**: 49-64.

Schauer, R., (2009) Sialic acids as regulators of molecular and cellular interactions. *Curr Opin Struct Biol* **19**: 507-514.

Schooling, S. R. & T. J. Beveridge, (2006) Membrane vesicles: an overlooked component of the matrices of biofilms. *J. Bacteriol.* **188**: 5945-5957.

Schoop, H. J., R. Schauer & H. Faillard, (1969) [On the biosynthesis of N-glycolyneuraminic acid. Oxidative formation of N-glycolylneuraminic acid from N-acetylneuraminic acid]. *Hoppe Seylers Z Physiol Chem* **350**: 155-162.

Schweda, E. K., J. R. Brisson, G. Alvelius, A. Martin, J. N. Weiser, D. W. Hood, E. R. Moxon & J. C. Richards, (2000) Characterization of the phosphocholine-substituted oligosaccharide in lipopolysaccharides of type b Haemophilus influenzae. *Eur J Biochem* **267**: 3902-3913.

Schwegmann-Wessels, C. & G. Herrler, (2006) Sialic acids as receptor determinants for coronaviruses. *Glycoconj J* **23**: 51-58.

Sethi, S. & T. F. Murphy, (2008) Infection in the pathogenesis and course of chronic obstructive pulmonary disease. *N Engl J Med* **359**: 2355-2365.

Severi, E., A. H. Hosie, J. A. Hawkhead & G. H. Thomas, (2010) Characterization of a novel sialic acid transporter of the sodium solute symporter (SSS) family and in vivo comparison with known bacterial sialic acid transporters. *FEMS microbiology letters.* **304**: 47-54.

Severi, E., A. Muller, J. R. Potts, A. Leech, D. Williamson, K. S. Wilson & G. H. Thomas, (2008) Sialic acid mutarotation is catalyzed by the Escherichia coli beta-propeller protein YjhT. *J Biol Chem* **283**: 4841-4849.

Severi, E., G. Randle, P. Kivlin, K. Whitfield, R. Young, R. Moxon, D. Kelly, D. Hood & G. H. Thomas, (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., S. J. King & J. N. Weiser, (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. *Infection and immunity.* **70**: 7161-7164.

Shaw, L. & R. Schauer, (1989) Detection of CMP-N-acetylneuraminic acid hydroxylase activity in fractionated mouse liver. *Biochem J* **263**: 355-363.

Smith, H. O. & K. W. Wilcox, (1970) A restriction enzyme from Hemophilus influenzae. I. Purification and general properties. *J Mol Biol* **51**: 379-391.

Smith, W., Andrews, C., Laidlaw, P.P., (1933) A virus obtained from influenza patients. *Lancet* **ii**: 66-68.

Sohanpal, B. K., S. El-Labany, M. Lahooti, J. A. Plumbridge & I. C. Blomfield, (2004) Integrated regulatory responses of fimB to N-acetylneuraminic (sialic) acid and GlcNAc in Escherichia coli K-12. *Proceedings of the National Academy of Sciences of the United States of America.* **101**: 16322-16327.

Sohanpal, B. K., S. Friar, J. Roobol, J. A. Plumbridge & I. C. Blomfield, (2007) Multiple co-regulatory elements and IHF are necessary for the control of fimB expression in response to sialic acid and N-acetylglucosamine in Escherichia coli K-12. *Mol Microbiol* **63**: 1223-1236.

Sorensen, K. I. & B. Hove-Jensen, (1996) Ribose catabolism of Escherichia coli: characterization of the rpiB gene encoding ribose phosphate isomerase B and of the rpiR gene, which is involved in regulation of rpiB expression. *J Bacteriol* **178**: 1003-1011.

St Geme, J. W., 3rd, (2002) Molecular and cellular determinants of non-typeable Haemophilus influenzae adherence and invasion. *Cell Microbiol* **4**: 191-200.

St Geme, J. W., 3rd, D. Cutter & S. J. Barenkamp, (1996) Characterization of the genetic locus encoding Haemophilus influenzae type b surface fibrils. *Journal of bacteriology.* **178**: 6281-6287.

St Geme, J. W., 3rd, M. L. de la Morena & S. Falkow, (1994a) A Haemophilus influenzae IgA protease-like protein promotes intimate interaction with human epithelial cells. *Mol Microbiol* **14**: 217-233.

St Geme, J. W., 3rd, M. L. de la Morena & S. Falkow, (1994b) A Haemophilus influenzae IgA protease-like protein promotes intimate interaction with human epithelial cells. *Molecular microbiology.* **14**: 217-233.

St Geme, J. W., 3rd & S. Falkow, (1990) Haemophilus influenzae adheres to and enters cultured human epithelial cells. *Infection and immunity.* **58**: 4036-4044.

St Geme, J. W., 3rd, S. Falkow & S. J. Barenkamp, (1993) High-molecular-weight proteins of nontypable Haemophilus influenzae mediate attachment to human epithelial cells. *Proceedings of the National Academy of Sciences of the United States of America.* **90**: 2875-2879.

Steenbergen, S. M., J. L. Jirik & E. R. Vimr, (2009) YjhS (NanS) is required for Escherichia coli to grow on 9-O-acetylated N-acetylneuraminic acid. *J Bacteriol* **191**: 7134-7139.

Steenbergen, S. M., C. A. Lichtensteiger, R. Caughlan, J. Garfinkle, T. E. Fuller & E. R. Vimr, (2005) Sialic Acid metabolism and systemic pasteurellosis. *Infect Immun* **73**: 1284-1294.

Swords, W. E., B. A. Buscher, K. Ver Steeg Ii, A. Preston, W. A. Nichols, J. N. Weiser, B. W. Gibson & M. A. Apicella, (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.

Szabo, M., D. Maskell, P. Butler, J. Love & R. Moxon, (1992) Use of chromosomal gene fusions to investigate the role of repetitive DNA in regulation of genes involved in lipopolysaccharide biosynthesis in Haemophilus influenzae. *J Bacteriol* **174**: 7245-7252.

Takematsu, H., T. Kawano, S. Koyama, Y. Kozutsumi, A. Suzuki & T. Kawasaki, (1994) Reaction mechanism underlying CMP-N-acetylneuraminic acid hydroxylation in mouse liver: formation of a ternary complex of cytochrome b5, CMP-N-acetylneuraminic acid, and a hydroxylation enzyme. *J Biochem* **115**: 381-386.

Tangvoranuntakul, P., P. Gagneux, S. Diaz, M. Bardor, N. Varki, A. Varki & E. Muchmore, (2003) Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. *Proc Natl Acad Sci U S A* **100**: 12045-12050.

Teele, D. W., J. O. Klein, C. Chase, P. Menyuk & B. A. Rosner, (1990) Otitis media in infancy and intellectual ability, school achievement, speech, and language at age 7 years. Greater Boston Otitis Media Study Group. *The Journal of infectious diseases.* **162**: 685-694.

Teele, D. W., J. O. Klein, B. Rosner, L. Bratton, G. R. Fisch, O. R. Mathieu, P. J. Porter, S. G. Starobin, L. D. Tarlin & R. P. Younes, (1983) Middle ear disease and the practice of pediatrics. Burden during the first five years of life. *Jama* **249**: 1026-1029.

Teplyakov, A., G. Obmolova, J. Toedt, M. Y. Galperin & G. L. Gilliland, (2005) Crystal structure of the bacterial YhcH protein indicates a role in sialic acid catabolism. *J Bacteriol* **187**: 5520-5527.

Traving, C. & R. Schauer, (1998) Structure, function and metabolism of sialic acids. *Cell Mol Life Sci* **54**: 1330-1349.

Turk, D. C., (1967) *Hacemophiliis influcniae, Its Clinical Importance.* English Universities Press Ltd., London.

Turk, D. C., (1984) The pathogenicity of Haemophilus influenzae. *J Med Microbiol* **18**: 1-16.

van Alphen, L., L. Geelen-van den Broek, L. Blaas, M. van Ham & J. Dankert, (1991) Blocking of fimbria-mediated adherence of Haemophilus influenzae by sialyl gangliosides. *Infection and immunity.* **59**: 4473-4477.

van Belkum, A., B. Duim, A. Regelink, L. Moller, W. Quint & L. van Alphen, (1994) Genomic DNA fingerprinting of clinical Haemophilus influenzae isolates by polymerase chain

reaction amplification: comparison with major outer-membrane protein and restriction fragment length polymorphism analysis. *J Med Microbiol* **41**: 63-68.
van Belkum, A., S. Scherer, L. van Alphen & H. Verbrugh, (1998) Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev* **62**: 275-293.
Van Eldere, J., L. Brophy, B. Loynds, P. Celis, I. Hancock, S. Carman, J. S. Kroll & E. R. Moxon, (1995) Region II of the Haemophilus influenzae type be capsulation locus is involved in serotype-specific polysaccharide synthesis. *Molecular microbiology.* **15**: 107-118.
Varki, A., (1992) Diversity in the sialic acids. *Glycobiology* **2**: 25-40.
Vergison, A., (2008) Microbiology of otitis media: a moving target. *Vaccine* **26 Suppl 7**: G5-10.
Vimr, E., C. Lichtensteiger & S. Steenbergen, (2000) Sialic acid metabolism's dual function in Haemophilus influenzae. *Mol Microbiol* **36**: 1113-1123.
Vimr, E. R., Unified theory of bacterial sialometabolism: how and why bacteria metabolize host sialic acids. *ISRN Microbiol* **2013**: 816713.
Vimr, E. R., K. A. Kalivoda, E. L. Deszo & S. M. Steenbergen, (2004) Diversity of microbial sialic acid metabolism. *Microbiol Mol Biol Rev* **68**: 132-153.
Vimr, E. R. & F. A. Troy, (1985a) Identification of an inducible catabolic system for sialic acids (nan) in Escherichia coli. *J Bacteriol* **164**: 845-853.
Vimr, E. R. & F. A. Troy, (1985b) Regulation of sialic acid metabolism in Escherichia coli: role of N-acylneuraminate pyruvate-lyase. *J Bacteriol* **164**: 854-860.
Voll, M. J. & S. H. Goodgal, (1961) Recombination during transformation in Hemophilus influenzae. *Proceedings of the National Academy of Sciences of the United States of America.* **47**: 505-512.
Warren, L., (1963) The Distribution of Sialic Acids in Nature. *Comp Biochem Physiol* **10**: 153-171.
Warren, L., (1964) N-Glycolyl-8-O-Methylneuraminic Acid, a New Form of Sialic Acid in the Starfish Asterias Forbesi. *Biochim Biophys Acta* **83**: 129-132.
Weber, A., K. Harris, S. Lohrke, L. Forney & A. L. Smith, (1991) Inability to express fimbriae results in impaired ability of Haemophilus influenzae b to colonize the nasopharynx. *Infect Immun* **59**: 4724-4728.
Weiser, J. N., J. M. Love & E. R. Moxon, (1989) The molecular mechanism of phase variation of H. influenzae lipopolysaccharide. *Cell* **59**: 657-665.
Weiser, J. N., D. J. Maskell, P. D. Butler, A. A. Lindberg & E. R. Moxon, (1990) Characterization of repetitive sequences controlling phase variation of Haemophilus influenzae lipopolysaccharide. *J Bacteriol* **172**: 3304-3309.
Weiser, J. N., M. Shchepetov & S. T. Chong, (1997) Decoration of lipopolysaccharide with phosphorylcholine: a phase-variable characteristic of Haemophilus influenzae. *Infect Immun* **65**: 943-950.
White, D. C., (1963) Respiratory systems in the hemin-requiring Haemophilus species. *J Bacteriol* **85**: 84-96.
White, K. A., I. A. Kaltashov, R. J. Cotter & C. R. Raetz, (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., S. Lin, R. J. Cotter & C. R. Raetz, (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.
WHO, (2011) Haemophilus influenza (HiB) diseases, Facts and figures. In. R. Martin (ed). Denmark, pp.

Willison, H. J., (2005) The immunobiology of Guillain-Barre syndromes. *J Peripher Nerv Syst* **10**: 94-112.

Winter, L. E. & S. J. Barenkamp, (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.

Wyborn, N. R., J. Alderson, S. C. Andrews & D. J. Kelly, (2001) Topological analysis of DctQ, the small integral membrane protein of the C4-dicarboxylate TRAP transporter of Rhodobacter capsulatus. *FEMS microbiology letters.* **194**: 13-17.

Yildirim, H. H., J. Li, J. C. Richards, D. W. Hood, E. R. Moxon & E. K. Schweda, (2005) Complex O-acetylation in non-typeable Haemophilus influenzae lipopolysaccharide: evidence for a novel site of O-acetylation. *Carbohydr Res* **340**: 2598-2611.

Yuki, N., (2005) Carbohydrate mimicry: a new paradigm of autoimmune diseases. *Curr Opin Immunol* **17**: 577-582.

Zamze, S. E., M. A. Ferguson, E. R. Moxon, R. A. Dwek & T. W. Rademacher, (1987) Identification of phosphorylated 3-deoxy-manno-octulosonic acid as a component of Haemophilus influenzae lipopolysaccharide. *Biochem J* **245**: 583-587.

Zamze, S. E. & E. R. Moxon, (1987) Composition of the lipopolysaccharide from different capsular serotype strains of Haemophilus influenzae. *Journal of general microbiology.* **133 (Pt 6)**: 1443-1451.

Zhu, A. & R. Hurst, (2002) Anti-N-glycolylneuraminic acid antibodies identified in healthy human serum. *Xenotransplantation* **9**: 376-381.

Zwahlen, A., L. G. Rubin & E. R. Moxon, (1986) Contribution of lipopolysaccharide to pathogenicity of Haemophilus influenzae: comparative virulence of genetically-related strains in rats. *Microbial pathogenesis.* **1**: 465-473.