Investigation of the role of a large serine rich repeat protein in *Streptococcus pneumoniae*

A thesis submitted to the University of Glasgow for the degree of Doctor of Philosophy

By

Alison Maren Saville B.Sc.

Submitted September 2008

Divison of Infection and Immunity
Faculty of Biomedical and Life Sciences
Biomedical Research Centre
Glasgow University
120 University Place
Glasgow G12 8TA

Acknowledgements

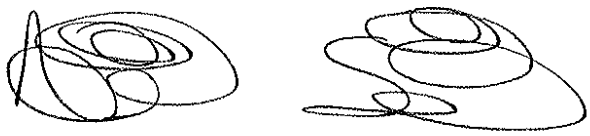
Firstly, I would like to thank my supervisor, Tim Mitchell, for giving me the opportunity to study for a PhD and for all the other opportunities I had throughout my studies. I would also like to thank all the members of the Mitchell laboratory, past and present, particularly those colleagues who contributed to the work in this Thesis.

Thanks to all my friends, especially those in Glasgow, who helped me in so many ways throughout my PhD.

Special thanks goes to Maren, Derek, Katya, Hollie, Chris, Norma and David for all their love and encouragement, and particularly Maren and Derek whose unquestioning support has meant the world to me.

And lastly, I would like to thank the Monster for living up to its name. Its elusiveness taught me far more than I expected.

This thesis is the original work of the author unless otherwise stated.



Alison Maren Saville

September 2008

1 Abstract

Streptococcus pneumoniae is a genetically diverse organism that varies substantially in its genomic content from one strain to another. Current therapeutic strategies in the management of pneumococcal disease include treatment with antibiotics and prevention by vaccination. However, due to the highly competent nature of the bacterium the prevalence of antibiotic resistance and vaccine escape is increasing. The pneumococcus causes a wide range of diseases, and this can be attributed to both the succeptibility of the human host and the genetic background of the infecting strain. The study of the contribution of variations in the genome of *S.pneumoniae* is clearly important in understanding the behaviour of this organism, and managing the burden of disease relating to this organism.

S.pneumoniae strains are able to acquire DNA from other strains, and also from other closely related species, who occupy the same niche in the human host. One region of genomic diversity in the pneumococcus encodes a large serine rich repeat protein, glycoysltransferases and secretion proteins, some of which are homologous to the Sec secretion pathway. Similar loci have been characterised and found to be important in the virulence of other gram positive bacteria, including *S.gordonii* and *S.parasanguinis*.

The presence of this locus was investigated in a diverse population of pneumococcal isolates, and shown to be present in a wide variety of isolates. The RNA of genes in the locus was found to be expressed. Expression of the SRR protein, encoded by *SP1772*, was investigated; a role in biofilm formation was identified utilising an isogenic mutant in *SP1772* of TIGR4. In addition, the gene encoding the SRR was found to be able to recombine within a single strain of *S.pneumoniae*, suggesting this region of the genome is not only variable in its presence in the pneumococcal population but also able to adapt to the environment it is in.

Contents

List of Tables

List of Figures

Abbreviations

α-	Anti
˚C	Degrees Celsius
Δ	Deletion
µl	micro litre
µm	micro metre
µM	micro molar
Alum	Aluminum phosphate
APS	Ammonium persulphate
ASL	Alternative secretion locus
BAB	Blood agar base
BHI	Brain heart infusion
BHY	Brain heart infusion + 5% yeast
BSA	Bovine serum albumin
CFU	Colony forming unit
CPS	Capsular polysaccharide
CSP	Competence stimulating peptide
D39	*Streptococcus pneumoniae* serotype 2 strain D39 (NCTC number: 7466)
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate

E.coli	*Escherichia coli*
FBS	Feotal bovine serum
g	Gram
g	Centrifugal force
gDNA	Genomic DNA
h	Hour
HRP	Horseradish peroxidase
Ig	Immunoglobulin
i.n	Intranasal
i.p	Intraperitoneal
IPTG	Isopropyl-ß-D-Thiogalactopyranoside
Kb	Kilobase
kDa	Kilodalton
L	Litre
LB	Luria broth
M	Molar
mg	Milligram(s)
min	Minute(s)
ml	Millilitre
MLST	Multi-Locus Sequence Typing

mM Millimolar

MW Molecular weight

n Group size

nm Nanometre

OD Optical density

PBS Phosphate buffered saline

PCR Polymerase chain reaction

Ply Pneumolysin

rpm revolutions per minute

RT Room temperature (~20˚C)

S.pneumoniae *Streptococcus pneumoniae*

SDS-PAGE Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis

SEM Standard error of the mean

sec Second(s)

SRR Serine rich repeat (protein)

ST Sequence type

TIGR The Institute for Genomic Research

TIGR4 *Streptococcus pneumoniae* serotype 4 strain TIGR4 (ATCC number: BAA-334)

V Volts

v/v	Volume/volume
wk	Week
WT	Wild type

Chapter 1

Introduction

1 Introduction

1.1 Streptococcus pnuemoniae

Streptococcus pneumoniae (the pneumococcus) is responsible for the death of up to a million children under the age of five worldwide per year (http://www.who.int/immunization_delivery/new_vaccines/pneumo/en/). It causes a wide range of diseases, those it is commonly associated with that can be fatal are septicaemia (infection of the blood stream by bacteria), meningitis (infection of the brain and spinal chord) and pneumonia (infection of the lungs); more benign diseases commonly caused by *S.pneumoniae* include otitis media (infection of the middle ear), sinusitis (infection of the paranasal sinuses) and bronchitis (infection of the bronchi) (Austrian, 1999). *S.pneumoniae* has also been associated with cases of arthritis (inflammation of the joints), osteomyelitis (infection of the bone or bone marrow), endocarditis (inflammation of the heart lining), endophthalmitis (inflammation of the eye), abscesses (formation of a cavity in the tissue) and necrotizing fasciitis (infection of subcutaneous tissue). Contact of *S.pneumoniae* in most instances will not cause disease in humans and can simply be carried by the host. Carriage of the bacterium is, however, often believed to be the first step in the pathogenesis of this organism. The treatment of pneumococcal disease is usually with antibiotics, however the burden and mortality of the diseases caused by *S.pneumoniae* remains high (Hausdorff *et al.*, 2005) and so together with the increasing prevalance of antibiotic resistant strains (Appelbaum, 2002; Jenkins *et al.*, 2008), a strategy of prevention by vaccination is clearly more desirable than treatment to reduce the burden of pneumococcal disease.

Current licensed vaccinations including Prevnar (Wyeth) and Pneumo 23 (Sanofi-Aventis) are composed of the capsular polysaccharide that coats *S.pneumoniae*. However 90 different variations of this polysaccharide exist within the species (Henrichsen, 1995) and even the most effective vaccination only protects against 7 of these (Black *et al.*, 2000). Studies have shown that a 90-valent vaccine, aside from being cost-inefficient, would actually be impossible to create.

The capsule of the pneumococcus is an indicator of the variation in the genome of this species. Vaccination strategies utilising other components of the

pneumococcal armoury are currently being developed, issues in developing these also include sequence variation, and in addition a lack of immunogenicity.

1.1.1 History and classification of the pneumococcus

S.pneumoniae is a gram positive facultative anaerobe. It was discovered simulataneously in 1880 by Chemist Louis Pasteur in France and Army physician George Sternberg in the US, and, since then several important scientific principles have been elucidated from the study of the pneumococcus. It is the organism in which genetic transformation was discovered (Griffith, 1928). It is also the organism in which the first non-protein antigen was found (Goebel and Adams, 1943) and in which DNA was shown to be the genetic material (Avery *et al.*, 1944).

1.1.2 MLST

MLST (Multi Locus Sequence Typing) was pioneered due to the perceived limitation of the historical classification system of *S.pneumoniae* by capsular serotyping in being sufficiently sensitive to adequately reflect population diversity and the ability of the organism to cause disease. MLST classification is assessed by the sequencing of 7 housekeeping genes of the pneumococcus: *aroE, gdh, gki, recP, spi, xpt* and *ddl* (Enright and Spratt, 1998). House keeping genes are relatively conserved genes within an organism given that they are generally required for essential function for the viability of the organism. They also are cytoplasmic and as such subject to less host immune pressure than cell-surface exposed factors. Each variant allele of each gene is given a number. The number for each gene then assigns a 'bar-code' of the seven numbers which denotes an MLST type of a particular isolate, commonly referred to as a sequence type (ST). The limitations of MLST have become apparent by the proven difference in genetic content of *S.pneumoniae* isolates of the same ST type by whole genome microarray (Silva *et al.*, 2006), which is a clear indicator of the massive genetic diversity extant within the species. Nonetheless MLST provides a more sensitive classification technique for the pneumococcus than capsular serotyping, and combined with an understanding of sequence diversity within virulence determinants of the bacterium could provide an extremely strong tool in understanding the implications of genetic diversity in the ability of a particular

species to cause disease (Hanage *et al.*, 2005). As such, understanding the contribution of sequence diversity seen in virulence factors of the pneumococcus will strengthen the current understanding of the ability of the organism to cause a wide range of diseases.

1.1.3 Comparative genomic hybridisation studies

Comparative genomic hybridisation studies enable the comparison of genomic content of a strain to a reference strain. These studies can identify how much the genetic content of a test strain varies from the reference strain, but also how different test strains compare to one another versus the test strain.

The pneumococcal genome is confirmed to be highly variable from isolate to isolate by these studies, one study found by comparative genomic hybridization of 19 strains, that 8 to 10% of genes were divergent from TIGR4, and that the nonconserved genes found across these strains was representative of approximately 20% of the TIGR4 genome (Hakenbeck *et al.*, 2001).

A comparative genomic hybridization study was carried out on pneumococcal strains of both the same serotype and sequence type. This identified that strains of serotype 14, ST124 do not have identical genomic content. This was consolidated by analysis of the virulence of these strains in an animal model of infection, with revealed that they behave differently (Silva *et al.*, 2006).

A further study sought to identify by comparative genomic hybridisation the genomic content of the pneumococcus which enables it to cause invasive disease. Strains that could be isolated from both invasive disease and carriage were selected, of serotypes 6A, 6B, and 14. The authors correlated the presence of certain regions of diversity within these serotypes with the ability of an isolate to cause disease (Obert *et al.*, 2006).

Comparative genomic hybridisation studies are valuable in understanding the variation that lies within genomes. However, this is by reference to a sequenced, reference strain. Whilst the study can confirm the presence of genes between a test and reference strain, the absence of genes could be due to a great divergence of that gene between test and reference strains, as opposed to its complete absence. In addition, these studies cannot characterise what is

present in the test strain and absent in the reference strain. Comparative genomic hybridisation studies of the pneumococcus have shown that there can be a high level of diversity within the population. Other studies have shown that genes within the pneumococcus can vary greatly in sequence and so the only way to completely characterise strains is by full sequencing.

Three strains of *S.pneumoniae* have been fully sequenced: R6 is an unencapsulated avirulent laboratory strain, derived from D39, a serotypes 2 invasive strain (Hoskins *et al.*, 2001); TIGR4 is a serotype 4, ST strain isolated from the blood of a patient with invasive pneumococcal disease (Tettelin *et al.*, 2001); G54 is a serotype 19F, ST 63 strain isolated from a respiratory sample, and is resistant to macrolides and tetracylcine (unpublished data, http://cmr.tigr.org/cgi-bin/CMR/GenomePage.cgi?org=ntsp05). In addition, there are currently up to 30 pneumococcal genomes currently being sequenced around the world. 4 strains currently being sequenced at the Sanger Center are discussed in Chapter 3.

1.2 Virulence factors of the pneumococcus

The pneumococcus produces a wide range of virulence factors. Virulence factors in the bacterium have been identified classically by associating a factor with a function of the bacterium's pathogenesis, however, in recent years whole genome sequencing has allowed the identification of putative virulence factors by homology to those identified in other species (Tettelin *et al.*, 2001), and whole genome signature-tagged mutagenesis has allowed the identification of genes of entirely unknown function in the disease process (Hava and Camilli, 2002; Lau *et al.*, 2001; Polissi *et al.*, 1998).

Major virulence factors are presented in Table 1, and discussed in the following sections.

Table 1 – Major virulence factors of *S.pneumoniae*

Virulence factor	Function	Reference
CbpA - Choline binding protein A (PspC - Pneumococcal surface protein C)	Binds secretory immunoglobulin	(Hammerschmidt *et al.*, 1997)
	Adhesion, virulence	(Rosenow *et al.*, 1997)
	Adhesion to nasopharyngeal and lung epithelia	(Cundell and Tuomanen, 1994)
	Adhesion to brain microvascular endothelium	(Ring *et al.*, 1998; Rosenow *et al.*, 1997)
	Allow host cell invasion	(Orihuela *et al.*, 2004; Zhang *et al.*, 2000)
	binds the secretory component of human secretory immunoglobulin A and human factor H	(Dave *et al.*, 2004; Quin *et al.*, 2005)
	enhances invasion of lungs *in vivo*	(Quin *et al.*, 2007)
	involved in colonisation	(LeMessurier *et al.*, 2006)
CbpD – Choline binding protein D, CbpE (Pce - Phosphorylcholine esterase), CbpG	CbpD and CbpE involved in colonisation, adherence	(Gosink *et al.*, 2000)
	CbpG is a protease and required for sepsis	(Mann *et al.*, 2006)
ClpC (Caseinolytic protease C)	Autolysis in some strains, pneumo growth in the lungs and bloodstream	(Ibrahim *et al.*, 2005)
ClpP (Caseinolytic protease P)	Thermotolerance, resistance to oxidative stress and virulence	(Ibrahim *et al.*, 2005)
Cps (Capsule)	Pivotal in allowing pneumococcus to cause disease	(Austrian, 1981)
	Expression variable at different stages of disease	(Hammerschmidt *et al.*, 2005)
	Allows pneumococcus to be resistant to complement mediated opsonophagocytosis and allows systemic dissemination	(Winkelstein, 1984)
Hyl (Hyaluronidase)	Contributes to the effects seen with pneumolysin including damage to the respiratory epithelium	(Feldman *et al.*, 2007; Paton *et al.*, 1993)
	Involvement in colonisation, tissue migration, produced by 90% isolates	(Feldman *et al.*, 2007; Paton *et al.*, 1993)
IgA	Cleaves human immunoglobulin A1 (IgA1)	(Bender and Weiser, 2006)

LytA – Autolysin A	Autolysin degrades the pneumococcal cell wall, resulting in lysis and release of intracellular and cell wall molecules Role in pneumonia and septicaemia Role in meninigitis	(Lopez *et al.*, 1997) (Berry and Paton, 2000) (Hirst *et al.*, 2008)
Lyt b – Autolysin b, Lytc- Autolysin c	Reduced colonisation in their absence	(Gosink *et al.*, 2000)
NanA – Neuraminidase A	Cleaves N-acetyl neuramic acid and breaks down mucin from mammalian cells allowing exposure of pneumococci to mammalian receptors Colonisation and persistence of the nasopharynx and middle ear Involved in the spread of pneumococci from the nasopharynx to the lung	(Tong *et al.*, 2000) (Tong *et al.*, 2000) (Orihuela *et al.*, 2003)
Ply – Pneumolysin	Allows host invasion and spread of disease, required for survival and replication in the lungs Interferes with phagocyte function Slows ciliary beating and disrupts integrity of human respiratory epithelium Meningitis	(Orihuela *et al.*, 2004) (Houldsworth *et al.*, 1994) (Feldman *et al.*, 1990) (Hirst *et al.*, 2008)
PotD – Polyamine transport protein D	Polyamine transport protein D Recombinant PotD protects mice against systemic pneumococcal infections	[Ware D 2006] [Shah p 2006]
PsaA - Pneumococcal surface adhesin A	Manganese ABC transporter required for competence and virulence Allows binding to host cells to protect pneumococci from oxidative damage Allows increased internalization of nasopharyngeal cells	(Dintilhac *et al.*, 1997) (Tseng *et al.*, 2002) (Rajam *et al.*, 2008)
PspA – Pneumococcal surface protein A	Cell surface, highly variable, in most strains, Role in complement receptor-mediated clearance of pneumococci Prevents complement mediated	(Briles *et al.*, 1988; Crain *et al.*, 1990) (Ren *et al.*, 2004) (Tu *et al.*, 1999)

	opsonization Binds to, and prevents killing by lactoferrin	(Hammerschmidt *et al.*, 1999; Shaper *et al.*, 2004)
RlrA, RgrA, RgrB	Pilus involved in adhesion and invasion of host cells	(Barocchi *et al.*, 2006; LeMieux *et al.*, 2006)
SpxB (Streptococcal pyruvate oxidase B)	Allows production of H_2O_2 which kills other bacteria in the upper respiratory tract Role in resistance to H_2O_2 Involved in colonisation	(Pericone *et al.*, 2000) (Pericone *et al.*, 2003) (Regev-Yochay *et al.*, 2007)
ZmpB (Zinc metalloprotease B)	Induces inflammation in the respiratory tract	(Blue *et al.*, 2003)

1.2.1 The pneumococcal capsule

The capsule of the pneumococcus is widely to be considered one of the most important factors in the ability of the pneumococcus to cause disease (Austrian, 1981; Henrichsen, 1995). Two key supporting arguments are that unencapsulated pneumococci are rarely isolated from patients, and that unencapsulated strains of pneumococci are unable to cause disease in animal models (Morona *et al.*, 1999).

The capsule is important in dissemination of the bacteria in the host, presumably mainly because it confers resistance on the bacterium to host complement mediated opsonophagocytosis (Winkelstein, 1984). Despite the fact that the ability of the pneumococcus is key in the ability of the organism to colonise and cause disease, high levels of capsule expression are not always desirable. It has been shown, for instance, that although capsular expression gives a considerable advantage in colonisation, less is required here than that required in invasive disease (Kim and Weiser, 1998; Magee and Yother, 2001; Ring *et al.*, 1998; Weiser *et al.*, 1994). So the capsule needs to be upregulated in instances where systemic dissemination is required, and down-regulated in the more intimate contact of the bacterium with host cells (Hammerschmidt *et al.*, 2005; Overweg *et al.*, 2000). This is also true for other organisms including *H.influenze* (Weiser, 1993) and *N.meningitidis* (Hammerschmidt *et al.*, 1996).

As previously mentioned, there are over 90 different variants of capsular serotype, i.e. antibodies raised against one of these 90 will not cross react with any of the other 90 types. This variation is likely to have arisen from intense immune pressure from the host (Bentley *et al.*, 2006) – in cases of systemic dissemination of the bacteria, the capsule may be one of the only virulence factors of the bacterium that is exposed to the host immune system, and as a highly immunogenic material it attracts a large immune response. Different serotypes, perhaps unsurprisingly, differ in their resistance to phagocytosis (Guckian *et al.*, 1980).

The rationale for capsular variation is clear. The mechanism by which this occurs has also been carefully studied. The majority of pneumococcal capsules are synthesised by a locus, *cps*, which lies between conserved genes *dexB* and *aliA* in the chromosome of pneumococcal strains (Garcia *et al.*, 2000; Kolkman *et al.*, 1998). Serotypes 3 and 37 are the exceptions to this rule, their capsule types are synthesised by the synthase pathway (Arrecubieta *et al.*, 1994; Cartee *et al.*, 2001; Dillard *et al.*, 1995; Llull *et al.*, 1999). Nonetheless, it is easy to imagine how both new capsule types are derived, since the recombination can occur between genes in the *cps* (Coffey *et al.*, 1998; Waite *et al.*, 2001; Waite *et al.*, 2003), but also how capsular switching – the ability of a strain to acquire an entirely different capsule type can occur (Brueggemann *et al.*, 2007; Coffey *et al.*, 1991). Capsular switching has already been observed in the post-vaccine era. The most common pneumococcal vaccine contains 7 capsular types conjugated to a protein. These 7 capsular types were the most common disease causing serotypes in the United States. It has now been observed that an ST type commonly associated with serotype 4 capsule has now been isolated from cases of invasive disease with a serotype 19A capsule, which is not included in the vaccine. Therefore, presumably this strain has maintained the genetic features allowing it to cause disease as a serotype 4 strain, but escapes the immunity afforded to the host by vaccination since it now has a 19A capsule. Presumably, capsule switching will increase over time and the effectiveness of the current vaccination will be diminished (Brueggemann *et al.*, 2007).

Capsule is important in determining the ability of an isolate to cause disease, however ST is also important (Brueggemann *et al.*, 2003) as are other genetic

factors that have not yet been utilised in a classification system (Mizrachi Nebenzahl *et al.*, 2004; Sandgren *et al.*, 2004; Sandgren *et al.*, 2005).

1.2.2 Virulence factors on the cell surface of S.pneumoniae

Many pneumococcal virulence factors are located on the cell surface. Their roles may be largely defined by capsule expression. When the capsule is upregulated, cell surface proteins may be 'masked' and may not carry out their function to the same extent as when the capsule is down-regulated (Hammerschmidt *et al.*, 2005). There is evidence that the capsule is up and down regulated at different stages of disease and in different disease types.

PspA, a choline binding protein, for example, has been shown to be responsible for adhesion to nasopharyngeal epithelia, lung epithelia and brain microvascular endothelium (Cundell and Tuomanen, 1994; Ring *et al.*, 1998; Rosenow *et al.*, 1997), and therefore allows the bacteria to invade host cells (Orihuela *et al.*, 2004a; Orihuela *et al.*, 2004b; Zhang *et al.*, 2000). This also correlates with where capsular expression is downregulated.

There are other choline binding proteins which are cell surface exposed and have been shown to be required in virulence. Additional important cell surface proteins include PspA and the pneumococcal pilus (see Table 1).

1.2.3 Other major virulence factors of S.pneumoniae

Autolysin, encoded by *LytA*, has been shown to have a major role in the virulence of the pneumococcus, in pneumonia, septicaemia and meningitis (Berry and Paton, 2000; Hirst *et al.*, 2008). It is responsible for the controlled lysis of the pneumococcal cell (Lopez *et al.*, 1997), and has implications in the larger evolution of the bacterium. However, it is also upon the lysis of the cell that *S.pneumoniae* is able to release virulence factors that are intracellular throughout the duration of the cell's viability. The best example of this is pneumolysin. Pnuemolysin is a pore-forming cytotoxin that is extremely important in the pathogenesis of all diseases caused by the pneumococcus (Feldman *et al.*, 1990).

1.3 Pathogenesis of *S.pneumoniae* induced disease

Large studies of serotype prevalence in invasive disease across Europe and the United states have revealed that most, if not all, serotypes are capable of causing disease (Nielsen and Henrichsen, 1992; Scott *et al.*, 1996; Verhaegen *et al.*, 1995), however, 85% of all invasive pneumococcal disease is caused by only 20 serotypes (Kalin, 1998) and more recent studies have suggested that invasive disease is caused by only some types of *S.pnuemoniae* (Sandgren et al., 2004). Host factors clearly play a role in the succeptibility of the host, for instance in Alaskan and Australian natives, the incidence of invasive pneumococcal disease is 10–50 times higher than other populations (Cortese *et al.*, 1992).

S.pneumoniae is usually found as a commensal organism in the human host. It resides in the nasopharynx of healthy individuals, and will likely never cause disease. Biofilms are one way in which the bacterium resides in the human host, and this is discussed further in section 1.5. However, where disease does occur, the first step in the pathogenesis of disease is the dissemination of the organism from the nasopharynx to other parts of the host. The organism must cross tissue barriers and be able to adapt to different niches within the host. This, predictably is a complex process, involving many of the virulence factors discussed above, and the organism must be able to maintain tight control of these to be successful (Finlay and Falkow, 1989). See figure 1.1 for a schematic overview of the pathogenic routes by which *S.pneumoniae* causes infection.

1.3.1 Carriage

Nasopharyngeal carriage of *S.pneumoniae* is most common in young children. Adults can also carry the bacteria, and although the incidence decreases with age, there are many factors which increase the likelihood of a person being a carrier. Children can act as reservoirs of the bacterium, and it has been noted that the carriage rate of the pneumococcus is 6% in adults without children, but 29% in those with children. Carriage of a strain usually lasts 3–4 months, but can last as long as 17 months (Gray *et al.*, 1982).

Figure 1.1 – The pathogenic routes for pneumococcal infection

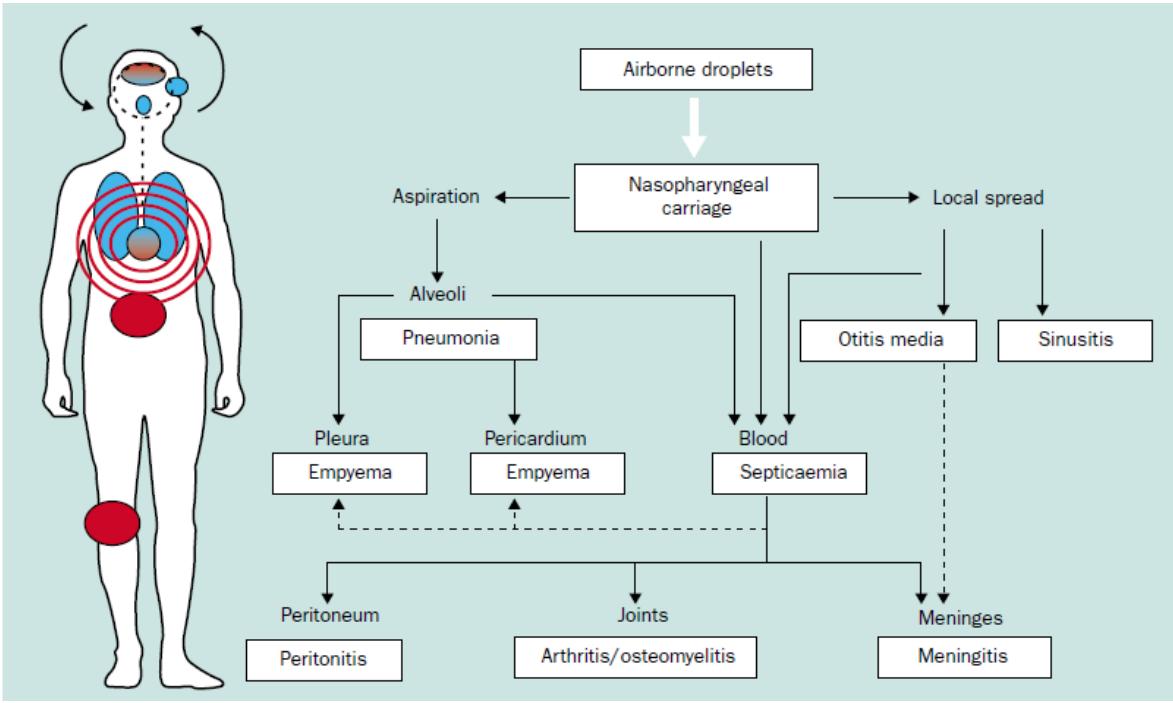


Diagram illustrating the pathogenic routes for pneumococcal infection. Organs infected through the airborne and haematogenic routes are depicted in blue and red, respectively. Figure from Boegart *et al* 2004a

Multiple serotypes of *S.pneumoniae* can be carried in the nasopharynx of healthy children and adults. The nasopharynx contains many bacterial species, including *Haempophilus influenzae, Moraxella cattarrhalis, Neisseria meningitidis, Staphylococcus aureus* and multiple streptococcal species (Bogaert *et al.*, 2004a), however the production of H_2O_2 by *S.pneumoniae* may confer a selective advantage on the organism, given that H_2O_2 can eliminate other bacteria (Pericone *et al.*, 2000) and *S.pneumoniae* is elicited protection from H_2O_2 by the gene product responsible for its production, SpxB (Pericone *et al.*, 2003). Nonetheless, it is of benefit to the pneumococcus to occupy this niche with other organisms, as a highly transformable bacterium it is within the nasopharynx that the organism acquires genes from other species, including penicillin resistance genes acquired from *S.mitis* (Dowson et al., 1994). The pneumococcus is able to acquire genes from other species due to the fact that it is naturally competent, and can uptake DNA from closely related organisms. This is explained further in section 1.4. Pneumococci in carriage do not necessarily cause disease, and it is believed that it is the introduction of a new serotype that often causes disease (Boulnois, 1992; Hardy *et al.*, 2001). Carriage is clearly a desirable state for the bacterium, especially considering the ease of dissemination from this location between human hosts. The most common serotypes isolated from carriage are 6A, 6B, 9V, 14, 18C, 19F and 23F (Bogaert *et al.*, 2004b) which are also those serotypes most commonly associated with invasive disease. However, S3 ST180, 6B, 19F and 23F are the isolates most strongly associated with carriage (Brueggemann *et al.*, 2003). This, together with the observation that disease most likely occurs in the host by a serotype they were already colonised with (Lloyd-Evans *et al.*, 1996) strongly suggests that carriage is the first step in the ability of the pneumococcus to cause invasive disease.

1.3.2 Pneumonia

Pneumonia is the infection of the lungs by bacteria. Following this, fluid accumulates in the lungs which prevents oxygen from easily transfusing to the bloodstream through the alveoli. There are 500,000 cases each year in the United States, with 50,000 resulting in deaths (Bogaert *et al.*, 2004a). In other developing countries the mortality rate has similarly been found to be 10-20%, seen particularly in infants, the elderly and immuno-comprimised patients

(Amdahl *et al.*, 1995; Johnston, 1991). The serotype of a strain is likely to determine whether it can cause pneumonia in a particular host (Sjostrom *et al.*, 2006).

Hava and Camilli identified in a STM screen that 1265 genes were involved in the ability of the pneumococcal strain TIGR4 to cause pneumonia (Hava and Camilli, 2002). Interestingly, STM screens of other pneumococcal strains do not overlap, which suggests that different genes are required by different strains to cause pneumonia. Other studies have identified LytA (Orihuela *et al.*, 2004a) CbpA (Rosenow *et al.*, 1997; Zhang *et al.*, 2000), NanA (Orihuela *et al.*, 2003), ply (Feldman *et al.*, 1990; Orihuela *et al.*, 2004b) and the competence system (Oggioni *et al.*, 2006) as contributing to the virulence of *S.pneumoniae* in pneumonia.

1.3.3 Meningitis

Meningitis is caused by the infection of the brain and spinal chord by a viral or bacterial pathogen. This infection causes fluid filled membranes to form around the brain and spinal chord. The mortality associated with this disease is high, and those that survive often present with learning disabilities and more extreme mental retardation, focal difficulties and hearing loss (Bohr *et al.*, 1985).

S.pneumoniae is now the leading cause of meningitis following the introduction of *H.influenze* type Ib conjugate vaccine (Bogaert *et al.*, 2004a), and is responsible for up to 75% if all cases in the developing world (http://www.who.int/immunization_delivery/new_vaccines/pneumo/en/).

Bacterial meningitis occurs following the translocation of bacteria across the blood-brain barrier (Ring *et al.*, 1998). Once the bacteria have entered the cerebral-spinal fluid, it is thought to be the huge immune response by the host, particularly leukocyte recruitment, that causes the hallmarks of the disease (Hirst *et al.*, 2004).

The capsule of the pnuemococcus is thought to contribute to the massive immune response by the host to the invading bacteria (Engelhard *et al.*, 1997), and the ability of a particular strain to cause meningitis can be attributed to their capsular type (Ribes *et al.*, 2008). It is likely that known proinflammatory

cell wall components such as lipoteichoic and teichoic acids also contribute to this immune response (Schneider *et al.*, 1999; Tuomanen *et al.*, 1985). NanC has also been implicated in pneumococcal meningitis by the observation that it is found to be in strains recovered from the cerebral-spinal fluid significantly more than those isolated from carriage (Pettigrew *et al.*, 2006). Pneumolysin has also been shown to have a role in meningitis (Hirst *et al.*, 2008).

1.3.4 Septicaemia

Septicaemia is the infection of the blood stream by bacteria. Pneumococcal septicaemia can follow infection of another site of the body, e.g. pneumonia. When the bacteria disseminate into the blood stream, the chances of survival of the host diminish rapidly.

One study identified that in 185 consecutive cases of invasive disease caused by *S.pneumoniae* in children, bacteraemia was present in 50% of cases, whilst 27% had pneumonia and 16% had meningitis (Myers and Gervaix, 2007).

The capsule is greatly associated with ability of the pneumococcus to cause septicaemia, particularly since it confers resistance upon the bacterium to complement mediated opsonophagocytosis, which greatly aids in systemic dissemination (Winkelstein, 1984). PotD is a cell surface protein that has been shown to contribute to pneumococal septicaemia in a murine model of infection (Ware *et al.*, 2006). In addition, mice immunised with PotD have a lower mortality mice than untreated mice (Shah and Swiatlo, 2006). A signature tagged mutagenesis screen carried out on *S.pneumoniae* identified the following virulence factors as contributing significantly to the ability of the organism to cause septicaemia: yha (hydrolysis of hylauronic acid), nanA, phoD (phosphate assimilation), msmK (sugar transport), mesD (unknown function) (Polissi *et al.*, 1998). LytA, ply, PspA and CbpG have also been shown to have a role in septicaemia (Berry and Paton, 2000; Mann *et al.*, 2006; Oggioni *et al.*, 2006).

1.3.5 Otitis media

Otitis media is a middle ear infection, which is in up to 50% cases caused by *S.pneumoniae,* by the passage of this organism to the Eustachian tube from the nasopharynx, where it resides in colonisation (Prellner et al., 1999). Otitis media

is usually diagnosed in children, and has a burden of disease of 7 million in the US each year (http://www.who.int/immunization_delivery/new_vaccines/pneumo/en/). Serotypes 3, 19 and 23 have been associated with otitis media in some studies (Hausdorff *et al.*, 2000), however other studies have failed to associate particular pneumococcal serotypes or STs with otitis media (Hanage *et al.*, 2004).

NanA, which encodes a neuraminidase, has been shown to contribute to the ability of the pneumococcus to cause otitis media (Tong *et al.*, 2000). A recent signature-tagged mutagenesis screen also identified 169 genes of *S.pnuemoniae* that allow the bacterium to cause otitis media. Interestingly, a good proportion of these were found not to be required for nasopharyngeal colonisation, and had also not been identified in previous STM screens in pneumonia and bacteraemia models (Chen *et al.*, 2007; Hava and Camilli, 2002; Lau *et al.*, 2001; Polissi *et al.*, 1998).

1.4 Recombination in *S.pneumoniae*

The variation seen in the genome of different isolates of *S.pneumoniae* is thought to arise largely from horizontal gene transfer, i.e. the transfer of genetic material from one cell to another. *S.pneumoniae* will usually reside in the nasopharynx of humans, along with other species of bacteria including *H.influenze, M.cattarrhalis, N.meningitidis, S.aureus, S.mitis, S.oralis,* and other streptococcal species. It is thought that much genetic exchange between the different species occurs here. An example of this is the acquisition of penicillin binding resistance, which confer resistance to penicillin on the host strain, thought to have been acquired by pneumococcal species from *S.mitis* (Dowson *et al.*, 1993).

S.pneumoniae is a naturally competent species, and so can take up DNA relatively easily. Incoming DNA is then incorporated into the genome of *S.pneumoniae* by the process of homologous recombination. This occurs when the incoming DNA bears similarity to DNA in the chromosome of the recipient strain. Two regions of the incoming DNA anneal to the recipient DNA and the genetic content between the two regions of the annealing DNA effectively flips.

S.pneumoniae sharing a niche with organisms that are similar genetically is clearly advantageous in the acquisition of new genetic material therefore, since it is more likely that incoming DNA will bear similarity to areas of the pneumococcal chromosome, and therefore it is more likely that recombination will occur. The ability of several different serotypes of *S.pneumoniae* to colonise the same host also allows intra-species genetic exchange which has allowed capsule switching. Capsule switching is a major advantage to the organism, given that current vaccine formulations target the capsule, and do not include all pneumococcal capsule types.

The success of *S.pneumoniae* in changing to avoid host elimination by the immune system or elimination by antibiotics resides in its ability to acquire external DNA. Previously, it had been assumed that incoming DNA into pneumococcal cells was liberated from other cells by the lysis of other bacterial cells in its vicinity, however it has recently been shown that lysis is not required for the liberation of DNA from other species (Johnsborg *et al.*, 2008).

In 1995, Havarstein and colleagues identified a short peptide, released by pneumococci, which dramatically increased their ability to integrate DNA into the chromosome (Havarstein *et al.*, 1995). This peptide, named competence stimulating peptide (CSP), has radically changed pneumococcal molecular biology, as it can now be synthesised, and utilised to induce competence in laboratory strains of *S.pneumoniae* to induce defined genetic changes in the genome of a particular strain. This allows relatively easy production of mutant strains and isogenic variants. Other streptococcal species also produce CSPs, and there are several different types of CSP, that are not restricted by species. Strains are able to communicate with other strains with the same CSP type, and therefore genetic exchange between strains and species with the same CSP type avoids the requirement of lysis.

CSP is only one of many gene products involved in transformation. CSP is secreted from the cell via ComAB (Hui and Morrison, 1991). This is regulated by a two-component signal transduction system of ComD and ComE (Pestova *et al.*, 1996). The concentration of CSP external to the cell is detected by ComD, which is a membrane-bound receptor of CSP (Havarstein *et al.*, 1996). Competence induction is reliant on sufficient saturation of ComD. It is assumed, although it

has not been shown, that the competence signal transduction system functions in the same way as other signal transduction systems, and therefore that CSP binding to ComD autophosphorylates ComD which in turn phosphorylates ComE. Phosphorylated ComE is then able to activate the transcription of many genes, including alternative sigma factor ComX (Claverys and Havarstein, 2002; Lee and Morrison, 1999). ComX then regulates the expression of the machinery required for the uptake and integration of DNA by the pneumococcus. A limited number of these genes have been shown to be essential for recombination to occur (Dagkessamanskaia *et al.*, 2004; Peterson *et al.*, 2004). Some other genes under the control of ComX have been shown to be involved in a process where pneumococci are able to induce lysis in non-competent cells (Claverys and Havarstein, 2007; Guiral *et al.*, 2005; Moscoso and Claverys, 2004; Steinmoen *et al.*, 2002; Steinmoen *et al.*, 2003). LytA, LytC and CbpD have been implicated in the ability of competent cells to lyse their non-competent counterparts (Guiral *et al.*, 2005; Kausmally *et al.*, 2005; Steinmoen *et al.*, 2002). Cells can be afforded protection against lysis by their competent counterparts by ComM and other immunity proteins (Havarstein *et al.*, 2006). The ability of a cell to become competent has been shown to be under the same control as the ability of a cell to induce lysis to facilitate DNA release in another cell. It is therefore postulated that the systems evolved together to increase the ability of a cell to acquire exogenous DNA from bacteria of the same and closely related species (Claverys and Havarstein, 2007; Johnsborg *et al.*, 2007; Johnsborg *et al.*, 2008).

1.5 Biofilm formation

Microbiology, understandably given the huge diversity of species and the massive intra-species variation, has historically commonly involved the study of a single species, or strain of a species. However, this is thought to give us a distorted picture of the contribution of a strain's own genetic background in its ability to successfully survive and proliferate, given that there is increasing evidence that a particular strain will rarely occupy a niche alone (Donlan and Costerton, 2002). A similar paradigm in human models of infection is that while the study of *in vitro* tissue culture models is informative, it cannot be entirely representative since the human host is a multi-tissue type, multi-organ system and therefore one cell type cannot be representative of the activity that could occur at that site in a 'real' situation.

Biofilms exist in many niches in the human host, and can confer an advantage to the host, as in the case of biofilm formation of the human intestine and female GU tract (Habash and Reid, 1999). However, it has been suggested that biofilms are involved in 60% of all human infection (Parsek and Singh, 2003) and are responsible for many chronic and persistent infections (Habash and Reid, 1999). Indeed biofilms have been shown to be associated with with dental caries, periodontitis, otitis media, musculoskeletal infections, cystic fibrosis pneumoniae, native valve endocarditis and bacterial prostatitis (Froeliger and Fives-Taylor, 2001). Furthermore, biofilm formation has been shown to actually be required for dental caries, periodontal disease and infective endocarditis to occur (Donlan and Costerton, 2002; Kolenbrander, 2000). It is not just human tissue that bacteria can form biofilms on during the pathogenesis of disease. Bacteria can form biofilms on any surface that contains naturally occurring materials. This includes medical devices implanted into the human body (Habash and Reid, 1999). When bacteria infect these devices they are well placed to replicate within the human host, and can disseminate from here to cause systemic disease.

Biofilms can be comprised of one species, however, many biofilms require the presence of multiple species if they are to cause disease in the human host. A good example of this is the formation of dental plaque. The first step in the formation of dental plaque is the adhesion of 'pioneer' species of bacteria, for example *S.gordonii*, to the tooth (Gibbons and Houte, 1975; Kuboniwa *et al.*, 2006; Rosan and Lamont, 2000). Other bacterial species will then adhere to this biofilm, and later species can actually facilitate a disease causing process that earlier pioneer species were unable to facilitate.

Biofilm formation confers several advantages to the bacteria involved in their formation. It has been noted the growth of bacteria in a biofilm is not identical to that of their planktonic counterparts (Donlan and Costerton, 2002; Oggioni *et al.*, 2006; Stickler, 1999). Biofilms allow a secure and stable environment for cells involved, and can act as a pool of the bacteria for dissemination when appropriate circumstances arise (Hall-Stoodley and Stoodley, 2005). In biofilms, it is likely that bacteria are exposed to sub-lethal doses of antibiotics and the host immune devices, since the polysaccharide matrix is difficult to permeate (Stoodley *et al* 2004). In addition, some strains in biofilms are able to exist in a

sessile state, where they are not proliferating. Since many antibiotics target machinery involved in the replication of bacterial cells, sessile strains are afforded protection to these antibiotics when involved in biofilms. Due to this, a greater understanding of biofilm formation and ways to target it specifically and separately from planktonic growth is required (Donlan and Costerton, 2002; Oggioni *et al.*, 2006; Stickler, 1999).

Transposon mutagenesis has been useful in identifying virulence factors that are involved in biofilm formation in a variety of bacterial species. These studies have been carried out on the following bacteria: *E.coli* (Aanensen *et al.*, 2007), *V.cholerae* (Watnick and Kolter, 1999), *S.gordonii* (Loo et al., 2000) and *P.aeruginosa* (O'Toole, 2004).

1.5.1 Pneumococcal biofilm formation

1.5.1.1Genes involved in pneumococcal biofilm formation

Pneumococcal biofilm formation has also recently been studied. Genes particularly associated with biofilm formation in the pneumococcus as found by gene expression studies in biofilm models include metalloproteases, neuraminidases, oxidative stress and competence genes (Oggioni *et al.*, 2006). This correlates with the *in vitro* study of mutants in specific genes in a biofilm model. Moscoso and colleagues demonstrated that LytA, LytB, LytC, CbpA, PcpA, and PspA all have a role in biofilm development (Moscoso *et al.*, 2006). Allegrucci and colleagues also noted that gene expression is variable in biofilm models from liquid culture. Interestingly, they also noted the number of types of proteins expressed increases greatly in pneumococcal biofilms (Allegrucci *et al.*, 2006). The authors suggest therefore that the ability of the pneumococcus to form biofilms is important to the proliferation of the species. A greater diversity and general upregulation of the proteome has been noted in other species in a biofilm mode of growth (Bjarnsholt *et al.*, 2005; Donlan and Costerton, 2002; Hassett *et al.*, 1999; Yoon *et al.*, 2002).

This general proteomic upregulation presumably requires tightly controlled regulation of gene expression in response to changes in the environment of the bacterium, however the genetic content of a strain will also presumably dictate

its ability to switch into biofilm growth mode, and this may determine its ability to be virulent (O'Toole, 2004).

1.5.1.2The role of capsule in pneumococcal biofilm formation

Capsule has been shown not to be required by *S.aureus* for *in vitro* biofilm formation (Toledo-Arana *et al.*, 2005). It has also been shown in some strains of the pneumococcus that biofilm development is inhibited in the presence of capsule by more than 60% (Moscoso *et al.*, 2006).However, where capsule is present, studies have demonstrated that gene expression doesn't vary in a biofilm model of growth (Oggioni *et al.*, 2006), and also that the expression of a protein involved in capsular synthesis is upregulated (Allegrucci *et al.*, 2006). It is worth noting that whilst invasion of host cells is far more efficient in the absence of capsule production, the bacteria needs to maintain the ability to produce capsule if it is to be successful following invasion of host cells (Hammerschmidt *et al.*, 2005; Weiser and Kapoor, 1999). Indeed, capsule regulation is key to colonisation by the pneumococcus also, however not at high levels (Magee and Yother, 2001).

However, its has been identified that single colony variants lacking capsule do arise in pneumococcal biofilms (Allegrucci and Sauer, 2007). These colonies, presumably, given the requirement for capsule in later stages of disease, will remain in the biofilm and can successfully compete here. The conditions within a biofilm such as availability of oxygen, are associated with less capsule production (Weiser *et al.*, 1994). Also, increased H_2O_2 production increases mutations arising within the pneumococcal genome (Pericone *et al.*, 2000).

1.5.1.3Association of the biofilm mode of growth with the pathogenesis of pneumococcal diseases

One study by Oggioni and colleagues noted that gene expression in liquid culture was highly similar to gene expression in bacteria infecting the blood stream, and that gene expression in a biofilm model was highly similar to gene expression in tissue models of infection (Oggioni *et al.*, 2006). They also showed that strains grown in liquid culture were more able to cause infection of the blood stream, and strains grown in a biofilm model were more able to cause infection of the

tissue. They note that when considering the pathogenesis of these diseases, the mode of growth associated with them seems logical. This therefore suggests that strains causing pneumonia and meningitis require the ability to form biofilms.

1.6 Glycosylation

Glycosylation is the process by which proteins are modified and greatly diversified by carbohydrate groups. In eukaryotes, where glycosylation has been more extensively studied, it has been shown to be involved in the molecule recognition – either between cells or between molecules of the same cell, changing the properties of a cell to suit its environment - such as solubility, viscosity and surface charge and control of proteolysis.

In eukaryotes, there are five different types of linkages between carbohydrate groups and peptides: N-glycosides, O-glycosides, phosphorus-linked glycosides, S-glycosides and C-glycosides.

Two of these types of glycosylation have been observed in prokaryotes: (i) N-linked: Glycans are attatched to the amide nitrogen of an Asn residue in a Asn-X-Ser/Thr consensus sequence. (ii) O-linked: Glycans are attatched to the functional hydroxyl group of Ser/Thr residues.

1.6.1 Bacterial glycosylation

Prokaryotic glycosylation was believed not to occur until studies confirmed the existence of glycoproteins in the S-layer of halobacteria (Mescher and Strominger, 1976). However, the organisms containing the only known prokaryotic glycoproteins were of little medical importance, and prokaryotic glycobiology remained obscure until the discovery and characterisation of a glycoenzyme in pathogenic bacterium *Streptococcus faecium* that contributed to its virulence (Kawamura and Shockman, 1983). Since then, various prokaryotic glycoproteins have been studied, many of which have been implicated in the virulence of important pathogens.

Due to the limited study of prokaryotic glycoproteins, much of what is known about eukaryotic glycoproteins has been extrapolated in theories relating to the function of the former, however, although bacteria carry out the

aforementioned types of glycosylation, they do so in a much altered way and thus this practice is fast diminishing. This has led to the discovery of many types of glycosylation that are unique to the prokaryotic kingdom (Inga Benz, 2002; Messner, 2004; Raj K. Upreti, 2003; Szymanski and Wren, 2005). Prokaryotic glycoproteins can be classified according to their cellular location (Raj K. Upreti, 2003) and examples are given to illustrate the role of glycosylation in each of the following sections.

1.6.1.1S-layer glycoproteins

Surface layer glycoproteins are found present on the outer layer of the cell. In archea, most S-layer proteins are glycosylated, and this aids the organisms in surviving extreme environments (Schaffer *et al.*, 2001).

1.6.1.2Membrane associated glycoproteins

Chlamydia trachomatis, which causes blindness and sexually-transmitted diseases, expresses an abundant membrane protein which when unglycosylated prevents the bacteria being able to cause disease (Kuo *et al.*, 1996).

1.6.1.3Cell-surface glycoproteins

Some *Campylobacter jejeuni* and *Campylobacter coli* strains have been found to have glycosylated flagella, which are essential virulence factors to both species (Szymanski *et al.*, 2003; Young *et al.*, 2002). Flagella are proteinacious elements that extend beyond the surface of the cell, comprised of multiple subunits. They allow the motility of organisms within the host. Their glycosylation may aid immune evasion, or may endow the flagellum with a more suitable charge for the host environment.

1.6.1.4Secreted glycoproteins and exo-enzymes

In Mycobacterium *tuberculosis* three secreted proteins that are glycosylated have been shown to illicit an impaired immune response when deglycosylated (Horn *et al.*, 1999). In the absence of glycosylation, the ability of *M.tuberculosis* to stimulate T-lymphocyte production of the host is severely limited, and this is thought to be due to glycosylation interfering with recognition of glycosylated

peptides by the MHC recognition molecules which are recognised by T-lymphocytes (Romain *et al.*, 1999).

1.6.1.5Cellular glycoproteins

Bacillus thuringiensis produces sporangia whose glycosylation protects them from degradation by proteases through the inability of the protease to interact fully with its target (Garcia-Patrone and Tandecarz, 1995). Two glycoproteins (205 and 72 kDa) were found in Bacillus thuringiensis sporangia.

1.7 Bacterial secretion pathways

Secretion is clearly a vital part of the pathogenesis of any disease causing organism. The majority of bacterial factors that are important in causing disease are either secreted or located on the cell surface, which is logical given that the host cell is external to the bacterial cell. The systems bacteria utilise to export factors required for their virulence, are themselves thererfore, unsurprisingly, required for the virulence of any organism (Finlay and Falkow, 1989).

Much of the research into secretion systems in bacteria has taken place in *Escherichia coli*, but these findings are believed to be relevant to all bacteria due to the proteins involved being well conserved throughout the kingdom. In a bacterial cell, all proteins are synthesised in the cytoplasm, however some are also needed in other parts of the cell, and often these proteins are important in virulence. They are therefore made in a precursor form containing a signal that directs the protein to its required location in the cell through interactions with a number of secretion proteins. This sequence is the protein to be exported varies dependent on the proteins required to secrete it. This sequence is then cleaved off to give a mature form of the protein (von Heijne *et al.*, 1991).

1.7.1 The Sec pathway

The Sec-dependent protein export pathway is conserved throughout bacteria and it is responsible for translocating precursor proteins containing classical amino-terminal signal sequences across the cytoplasmic membrane (Danese and Silhavy, 1998)

Pathogenic and non-pathogenic bacteria utilize this pathway to localize proteins to the cell surface or as the first step in secreting proteins out of the cell. Many bacterial pathogens also possess specialized protein export systems dedicated to the secretion of virulence factors (Hoskins *et al.*, 2001).

A protein involved in secretion in *E.coli*, SecB (Secretion protein B), was discovered by fusion of specific signal sequences with reporter genes (Kumamoto and Beckwith, 1985). Its function was deduced to be the most common chaperone of periplasmic and outer membrane proteins to the inner membrane to prepare for their export, since mutational studies of certain fusion proteins suggested that SecB interacts with their immature forms, and in its absence they quickly adopt their mature form (Collier *et al.*, 1990). It is believed to chaperone them to the inner membrane because it can interact with the C-terminal of another secretion protein, SecA (Secretion protein A), when SecA is found on the inner membrane (Bayer, 1994). Sec A is a homodimer, found in the cytoplasm and inner membrane, and each subunit contains a low and a high affinity ATP binding. In the cytoplasm, SecA functions as its own transcriptional regulator (Schmidt and Oliver, 1989) and also as a minor chaperone of proteins for export to the inner membrane (Fekkes *et al.*, 1997). In the inner membrane, SecA binds to a complex of other secretory proteins, Sec E, G and Y (Secretion proteins E, G and Y) (Eichler *et al.*, 1997). Upon binding of ATP to the high-affinty site on SecA, both SecA and the molecule it is chaperoning, the protein for export, are taken up into the membrane due to a conformational change in SecA and they are protected from the hydrophobic environment a complex of Sec E, G and Y. However, the signal sequence of the protein for export remains in the cytoplasm where it is cleaved off. When the ATP is hydrolysed, SecA is released from the membrane, leaving the protein for export behind. The more ATP available for the translocation, the further the protein for export is translocated by SecA (Danese and Silhavy, 1998). SecE and Y have been shown to be essential for secretion, and the absence of SecG has been shown to significantly reduce secretion levels of secreted proteins (Ito *et al.*, 1983; Nishiyama *et al.*, 1994). SecY, the largest protein in the complex, requires the presence of SecE, in its absence, SecY is destroyed by a membrane protease (Kihara *et al.*, 1995). Electron microscopy studies have identified a eukaryotic homologue of SecY that forms a hollow cylinder in the membrane, and thus this is postulated for SecY also (Hanein *et al.*, 1996; Joly and Wickner, 1993). SecY

has also been shown to check the signal sequences of the proteins for export presented to the SecE/G/Y complex by SecA and reject proteins with incorrect sequences (Osborne and Silhavy, 1993). Three further proteins have been implicated in protein translocation across the inner membrane, SecD, F (Secretion protein D, F) and YajC (Preprotein translocase subunit). Although they have been shown to be non-essential, their presence does improve efficiency of the process, and it is speculated to stabilize the inner-membrane form of SecA (Danese and Silhavy, 1998) or be involved in clearing the secretion system channel (Veenendaal et al., 2004), or regulate SecA (Duong and Wickner, 1997). All of the secretory proteins above are required for efficient secretion of most of the proteins in the bacterial cell, however, recently it has emerged that several species of gram-positive bacteria contain additional homologues of some of the essential secretory proteins in their genome that are non-essential, but often required for virulence. See figure 1.2 for an overview of the Sec pathway.

1.7.2 Alternative secretion pathways

It has recently been found that some bacteria possess two copies of genes involved in the sec secretion system. It appears that there is a complete set of genes dedicated to the ordinary secretion of factors from the cytoplasm that is essential for the viability of the cell, and that the second, often incomplete set, is dedicated to the export of unusual proteins, usually required for the full virulence of the bacterium but not essential for survival. The most studied of alternative secretion systems contributing to the virulence of an organism are outlined below, however it is worth noting that in Corynebacterium glutamicum, which encodes two secA homologues, both have been found to be essential for the survival of the cell (Caspers and Freudl, 2008).

There are two main types of alternative secretion systems that have been characterised in the prokaryotic kingdom thus far. Figure 1.3, from (Rigel and Braunstein, 2008) outlines the two most characterised types of alternative secretion. One contains a SecA2 homologue which is presumably capable of interacting with the main Sec pathway of the cell. This system has been shown of being capable of transporting a variety of products from the cell, it is chromosomally associated with one of these products. The other type of alternative secretion system has been shown to contain homologues of both SecA

Figure 1.2 – Protein translocation through the SecYEG complex

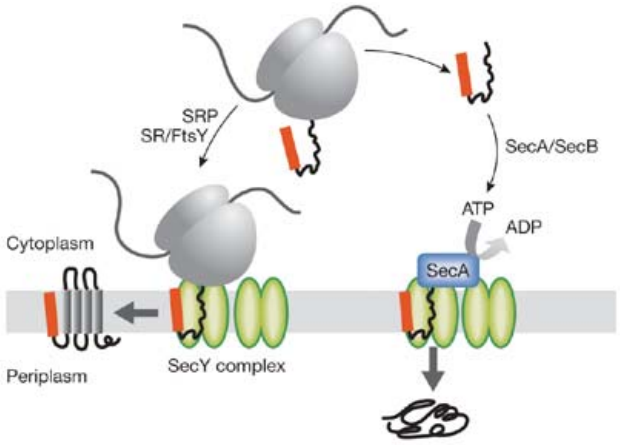


Figure from (Robson and Collinson, 2006). Schematic overview of protein translocation and membrane protein insertion through the SecY complex in bacteria. Nascent chains having a signal sequence (red) emerge from the ribosomes (grey) and are localized to the SecY complex (green) co-translationally by the signal recognition particle (SRP) and FtsY/SRP receptor (SR; left). Alternatively, they can be presented post-translationally through SecA or SecB (right).

Figure 1.3 – Models for SecA2 dependent transport

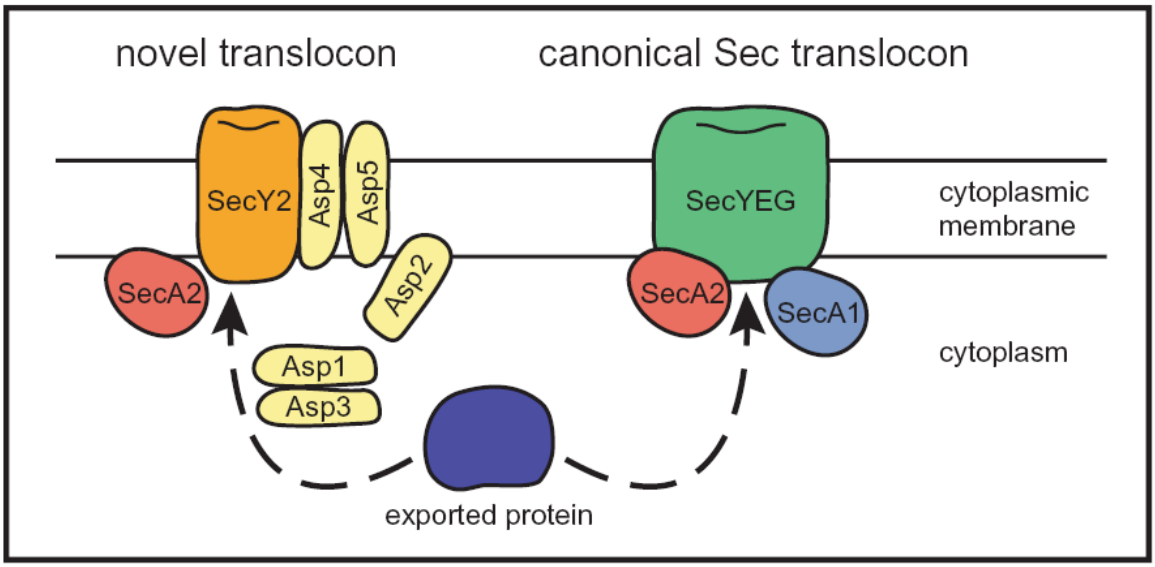


Figure from (Rigel and Braunstein, 2008). A secA2 protein may be exported through either the canonical Sec system of the cell (right hand side), for example in *M.tuberculosis*. Alternatively, the genome may encode an entirely separate novel translocon (left hand side), for example in *S.gordonii*.
this contributes to the bacterium being able to avoid elimination in the human

and SecY (named SecA2 and SecY2, respectively), and is chromosomally associated with a variety of other alternative secretion genes. These systems have been found to be associated with large serine-rich repeat proteins and glycosyl-transferases.

1.7.2.1M.tuberculosis

Mycobacterium tuberculosis causes tuberculosis. There were 9.2 million new incidences of tuberculosis in the world in 2006, and 1.7 million deaths from tuberculosis (http://www.who.int/tb/publications/global_report/2008/pdf/fullreport.pdf). The bacterium is able to survive and replicate within human macrophages, and host, even in the presence of antibiotics. The secretion of virulence and other factors is postulated to allow the survival of *M.tuberculosis* in the human macrophage, and as such, many studies have focussed on elucidating the essential secretion systems of the bacterium (Braunstein *et al.*, 2003).

M.tuberculosis has been found to contain a homologue of SecA in its genome, denoted SecA2. While its SecA is essential to the viability of the cell, SecA2 has been shown to be required for full virulence, probably due to its function as an exporter of superoxide dismutase A, used by M.tuberculosis to protect against attack from host macrophages (Braunstein *et al.*, 2003).

1.7.2.2L.monocytogenes

Listeria monocytogenes is a gram positive bacterium that is able to colonise a multitude of surfaces and survive in circumstances that many other non-spore forming organisms cannot. *L.monocytogenes* colonises the human host without causing disease, however it has also been shown to be involved in highly invasive disease, and is acquired through ingestion of contaminated food (Cossart, 1998).

L.monocytogenes contains a homologue of SecA alongside its essential copy, again denoted SecA2, which has been implicated in virulence and the phenotypic variation of the bacteria. Mutants of SecA2 in *L.monocytogenes* are unable to secrete 17 proteins. Interestingly, only some of these proteins have been shown to have unusual signal sequences (Lenz and Portnoy, 2002).

1.7.2.3S.gordonii

S.gordonii contains homologues of SecA and SecY, and chromosomally associated with these genes are five additional genes, *asp1-5* which have also been shown to be involved in secretion (Bensing and Sullam, 2002; Takamatsu *et al.*, 2004, , 2005b). All seven gene products have been shown to be exclusively involved in the export of GspB, a serine rich repeat protein that is involved in the adhesion of *S.gordonii* to platelets, an essential step in the ability of the organism to cause infective endocarditis. GspB and its associated pathway are further discussed in section 1.8.1. Other alternative secretion systems that are associated with the export of serine rich repeat (SRR) proteins are also discussed in section 1.8.

1.8 Serine rich repeat proteins

Various serine rich repeat (SRR) proteins that have been characterised in gram positive bacteria are detailed below. See Figure 1.4 for a diagram of the loci harbouring these SRRs.

1.8.1 GspB

Bensing and colleagues identified GspB, a SRR in *Streptococcus gordonii* in a study with the intent of identifying what allowed the organism to bind to platelets, which is thought to be an essential step in the ability of the organism to cause infective endocarditis (Bensing and Sullam, 2002). Infective endocarditis, which is the infection of the cardiac valves by bacteria, occurs when endocardial lesions occur in the human host. Under normal circumstances bacteria are unable to infect normal vascular endothelia (Alanee *et al.*, 2007). Following the appearance of lesions in the endocardium platelets accumulate at the site (Durack, 1975). *S.gordonii* is normally found in the mouth and is associated with dental plaque. However, the bacterium is also frequently isolated in cases of infective endocarditis. As a group, viridans streptococci are associated with 50% of all cases of infective endocarditis (Douglas *et al.*, 1993) and *S.gordonii* is the streptoccal species isolated most frequently (Rosenow *et al.*, 1997). The bacteria are believed to be able to bind to the platelets that have accumulated on the endocardial surface and form colonies there. When the

Figure 1.4 – Comparison of the ASL in different species of gram positive bacteria

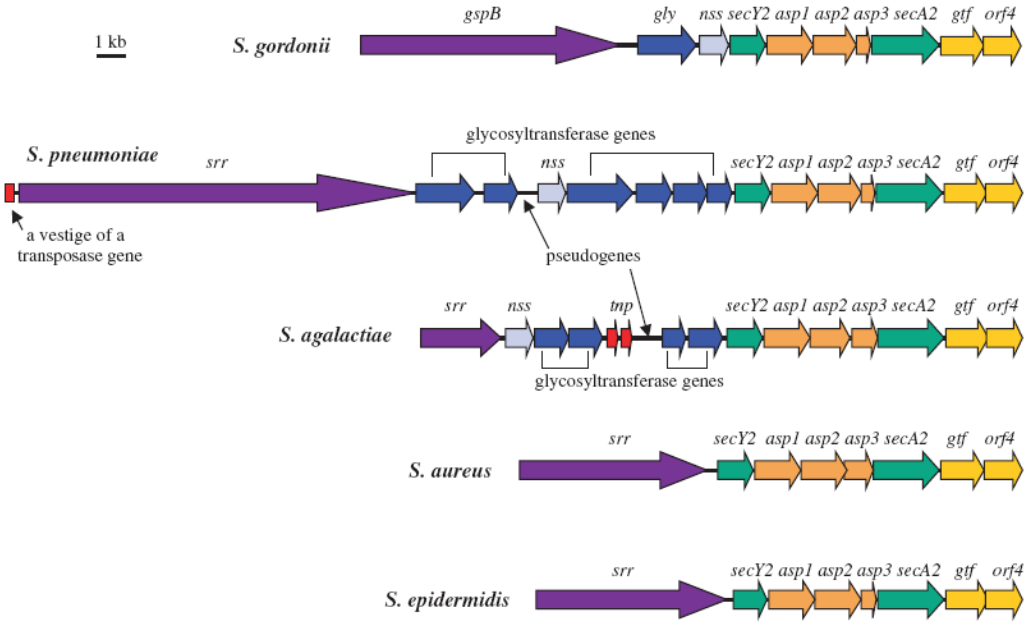


Figure from (Takamatsu *et al.*, 2004a). ASLs in several species of gram positive bacteria are annotated. Purple indicates the SRR, blue indicates a glycosyl-transferase, green are SecA2 and SecY2 homologues, orange indicate alternative secretion genes and yellow indicates essential glycosyl-transferases.

population reaches a critical mass, fibrin and platelets cover the colonies and as such a bacterial community is secured on the surface (Sullam *et al.*, 1985). These infections are difficult to treat and the morbidity and mortality of the disease remains high in spite of treatment with antibiotics and surgery (Durack, 1995).

Multiple proteins which mediate interactions between the human and bacterial cell have been identified in *S.gordonii*, however none had been implicated in platelet binding. Bensing and colleagues therefore sought to determine what allowed the organism to bind platelets by generating a library of mutants using an integrative vector (Bensing and Sullam, 2002) and assessing their ability to bind platelets. The mutant strain that was most severely attenuated in its ability to bind platelets was compared to the parent wild type strain and found to have an 8kb deletion, containing nine genes encoding glycosyl-transferases, homologues of SecA and SecY and several genes of unknown function. SecA and SecY have been shown to be involved in secretion out of the cell (Danese and Silhavy, 1998), and thus cell surface extracts of the mutant and wild type were compared and a protein far larger than the highest molecular weight marker of 215 kDa was identified. A gene encoding 3072 amino acids giving a product of 286 kDa was found to be present immediately upstream from the genes deleted in the mutant strain. Analysis of the sequence revealed it encoded a large serine rich repeat protein, and it was named GspB (Bensing and Sullam, 2002).

GspB binds to sialic acid residues on the O-linked carbohydrates of a receptor on platelets, Ibα (Bensing *et al.*, 2004). This binding is mediated by the central, basic region of GspB (Takamatsu *et al.*, 2005a). GspB has also been shown to bind salivary mucin MG2 and salivary agglutinin (Takamatsu *et al.*, 2006)

Subsequent studies have been carried out to identify the functions of the genes adjacent to GspB (Gordonii surface protein B) in the chromosome. Asp1-5 (alternative secretion proteins 1-5) were identified as required for the export of GspB, since strains deficient in these genes were unable to export GspB (Takamatsu *et al.*, 2004, , 2005b). GtfA and GtfB mutants were found not to produce GspB at all, in these mutants GspB was transcribed but not expressed. Gly and nss were found to alter the lectin binding properties of GspB, but were not essential for the expression or secretion of GspB, cell surface levels of GspB

are only slightly diminished in the absence of either gly or nss. However the alteration in lectin binding does translate to a reduction in platelet binding ability by mutants in gly or nss (Takamatsu *et al.*, 2004).

1.8.2 Hsa (Human sialic acid binding protein)

Hsa is a serine rich glycoprotein which has been idenitified in *S.gordonii* strain Challis (Kuo *et al.*, 1996). It is able to recognise receptors on erythrocytes and polymorphonuclear leukocytes [Takahishi *et al.,* 2004, 2006). Deletion of *hsa* reduces the ability of the bacteria to adhere to salivary agglutinin (Jakubovics *et al.*, 2005) which is a highly glycosylated, cysteine-rich glycoprotein (Prakobphol *et al.*, 2000). It has also been shown that Hsa binds salivary mucin MG2 and secretory immunoglobulin A (Takamatsu *et al.*, 2006).

1.8.3 Fap1 (Fimbral associated protein 1)

S.parasanguinis colonise the oral cavity and are involved in the formation of dental plaque. Fimbriae have been shown to mediate the adhesion of *S.parasnaguinis* to teeth in a *in vitro* tooth model (Burnette-Curley *et al.*, 1995; Quin *et al.*, 2007). Fap1, a serine-rich glycoprotein, has been shown to be the major subunit of fimbriae (Quin *et al.*, 2007). It is essential for fimbral formation and its deletion results in a reduced ability of *S.parasanguinis* to adhere to *in vitro* models (Quin *et al.*, 2007).

S.parasanguinis contains a locus that is dedicated to Fap1, which is fairly similar to that seen in *S.gordonii*, associated with GspB. The genes encoded here include alternative secretion genes, and glycosylation genes(Chen *et al.*, 2004; Chen *et al.*, 2006; Peng *et al.*, 2008; Quin *et al.*, 2007).

1.8.4 SraP (Serine-rich adhesin for platelets)

Staphylococcus aureus is able to cause infective endocarditis. Platelet binding, as in the case of *S.gordonii,* is assumed to be an important step in the pathogensis of the disease. *S.aureus* contains many factors which have been shown to allow it to bind to proteins on platelets. These are protein A (Chen *et al.*, 2007; Nguyen *et al.*, 2000), clumping factors A and B (Chen *et al.*, 2007) and fibronectin-binding protein (Heilmann *et al.*, 2004). Many of these are multi-

functional and functionally redundant in the binding of human platelets (Siboo *et al.*, 2005).

SraP was originally identified in a study that characterized novel LPXTG proteins by analysing available genome sequences, and it was designated SasA (Roche *et al.*, 2003). It is in the genomes of all *S.aureus* genomes published to date and is predicted to be 227 kDa without glycosylation.

SraP and its associated genes vary from GspB and its associated genes in several important ways. Two genes that are required for optimal export of GspB in *S.gordonii* are absent in *S.aureus* (Siboo et al., 2005). These are gly and nss, which have been shown not to be essential for GspB expression, but the detectable levels of GspB on the cell surface of *S.gordonii* are reduced in their absence (Takamatsu *et al.*, 2004). Asp4 and asp5 are also absent in *S.aureus,* suggesting a difference in the secretion of SraP compared to GspB (Siboo *et al.*, 2008). Indeed, disruption of SecA2 in *S.aureus* results in the near complete loss of SraP surface expression, but importantly does not completely abolish SraP export – the protein could no longer be detected by western blotting however it could be by a more sensitive technique of lectin blotting of the cell surface (Siboo *et al.*, 2008). Examination of the sequence of SraP reveals some differences which may explain the slightly less stringent export requirements of the protein. The signal peptide of SraP is greatly altered from that of GspB, and also SraP only contains 2 glycine residues, which are in different positions from any of the 3 glycine residues in GspB that have been shown to be essential for its export by the alternative secretion system (Bensing *et al.*, 2007). Interestingly, SraP has a pI that is acidic while GspB has a more basic pI. The combination of altered and possibly diminished glycosylation, absence of secretion machinery and changes in the SraP sequence by comparison to GspB suggest that the loci do not operate in an identical manner, and indeed SraP may, under certain circumstances, be suitable for export by the canonical Sec system of the cell (Siboo *et al.*, 2008). The absence of genes encoding two glycosyl-transferases and two genes involved in alternative secretion could also suggest that the functions of products encoded by these genes are intertwined.

The receptor for SraP has not been identified (Siboo *et al.*, 2008). Deletion of SecA2 also resulted in less lipase and more amidase, but fluctuating transcripts

detected by microarray in the case of lipase suggest this is not due directly to the absence of SecA2 transport. There is no explanation for the additional amidase in the sample (Siboo *et al.*, 2008).

1.8.5 SrpA (Serine-rich protein A)

S.cristatus is found in oral biofilms (Mouton *et al.*, 1980). SrpA is a SRR in *S.cristatus.* SrpA disruption causes irregular fibrilar tuft morphology, and fibrilar tufts are required for interaction of *S.cristatus* with *Corynebacterium matruchotii* (Lancy et al., 1980) and *Fusobacterium nucleatum* (Lancy *et al.*, 1983) and human cells (Handley *et al.*, 1991). SrpA disruption negates the ability of *S.cristatus* to interact with *C.matruchotii* and *F.nucleatum* (Handley *et al.*, 2005). This implies a role for SrpA in biofilm formation which is crucial in the formation of dental plaque.

1.8.6 SrpA (Serine-rich protein A)

Sanguis streptococci are primary colonizers of the tooth surface (Kolenbrander and London, 1993). They make up a major part of dental plaque and allow other plaque bacteria to adhere to them. These adhering bacteria cause caries and periodontal diseases (Gibbons and Houte, 1975). *S.sanguinis* is also able to cause infective endocarditis, the disease which GspB of *S.gordonii* is implicated as a major factor in. SrpA is a homologue of GspB and Plummer and colleagues have shown that SrpA mediates the adhesion of *S.sanguinis* to platelets by binding GPIb which presumably is also a primary step in the pathogenesis of the organism in causing infective endocarditis (Plummer *et al.*, 2005).

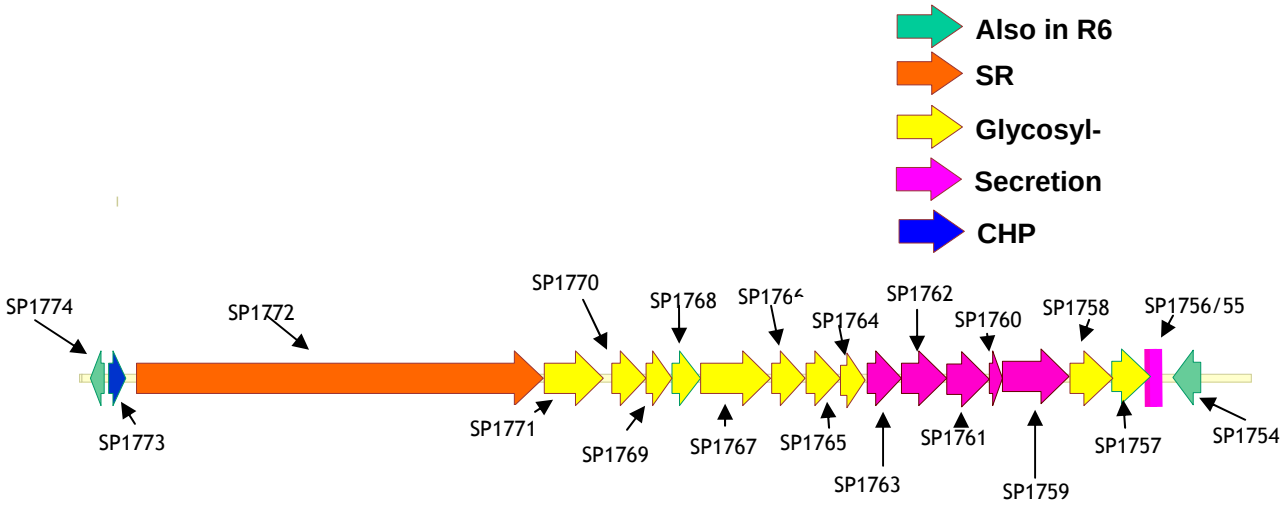
1.8.7 SP1772

SP1772 was first identified when the genome of virulent isolate TIGR4 was sequenced (Tettelin *et al.*, 2001). A signature tagged mutagenesis screen of TIGR4 identified SP1772 and 3 further genes in the locus SP1755-SP1773 as required for full virulence in a competitive model of lung infection (Hava and Camilli, 2002). Many studies have speculated as to the function of the SP1755-SP1773 locus by comparison to work in other strains, discussed above (Bensing and Sullam, 2002; Quin *et al.*, 2007). In addition, two comparative hybridisation

studies have identified this locus as a region of diversity of among pneumococcal strains (Obert *et al.*, 2006; Silva *et al.*, 2006). One of these studies found a correlation between the presence of this region of diversity and the propensity of a clone to cause invasive disease, and showed that SP1772 is required for full virulence in a pneumonia model of infection (Obert *et al.*, 2006).

SP1772 in the TIGR4 genome is 14.3kb. It has an LPXTG motif which presumably mediates its attachment to the cell surface by sortases, enzymes which allow the covalent attachment of proteins to the cell wall of the bacterium by interacting with the LPXTG motif of a protein when its hydrophobic domain and charged residues slow down its transport through the cell wall. Sortase cleaves the LPXTG motif and covalently bonds the protein to the cell wall (Paterson and Mitchell, 2004). SP1772 contains two serine rich repeat regions, SRR-1 and SRR-2. SRR-1 is closer to the N-terminal of SP1772 and is much smaller than SRR-2 which is closely followed by the C-terminal and LPXTG motif. A signal sequence is presumably encoded in the region upstream of SRR-1 in *SP1772*. See figure 1.5 for a diagram of RD20 and figure 1.6 for a diagram of SP1772.

Figure 1.5 – RD20 in TIGR4



Streptococcus pneumoniae TIGR4

Each of the genes in RD20 of *S.pneumoniae* TIGR4 is annotated here.

Figure 1.6 – SP1772 in TIGR4

Annotation of *SP1772* in TIGR4, with the SRRs and LPXTG motif highlighted. *SP1772* is 14.3kb in TIGR4.

Chapter 2

Materials and Methods

2 Materials and Methods

2.1 Preparation of bacterial strains

2.1.1 Streptococcus pneumoniae

Strains were streaked on a BAB (Oxoid) agar plate, with antibiotic selection as appropriate. Plates were inverted and incubated statically o/n at 37˚C. One colony was selected from the plate and inoculated in 10 ml Bovine heart infusion (BHI), with antibiotic selection as appropriate, and incubated at 37˚C until OD_{600} reached 0.6 (mid-log), determined on a spectrophotometer. 10% glycerol was then added to the culture and 1ml aliquots were taken and stored at -80˚C. A loop of the culture was taken simultaneously and streaked on a BAB plate, with antibiotic selection as appropriate, and incubated o/n at 37˚C with an optichin disc. Cultures were deemed *S.pneumoniae* if they had the correct morphology and were optichin sensitive. *S.gordonii* strains were prepared in the same fashion, but Bovine heart infusion plus 5% yeast (BHY) was used in place of BHY and strains were grown aerobically (in 10ml of culture in a 20ml universal). Additionally, strains were used if optichin resistant.

2.1.2 Escherichia coli

Strains were streaked on an Luria broth (LB, Oxoid) plate, with appropriate antibiotic selection. Plates were inverted and grown statically at 37˚C overnight. A colony was selected and inoculated into LB broth with antibiotic selection as appropriate. Cultures were incubated at 37˚C and shaking at 220 rpm until OD_{600} reached 1.0. 10% glycerol was added to the cultures and aliquots of 1 ml were taken and stored at 80˚C. Simultaneously, a loop of the culture was taken and streaked on an LB plate with appropriate antibiotic selection to ensure the purity of the glycerol stocks.

Table 2.1 – Strains used in this study

Name	Species	Genetic background	Intentional mutation	Study (chapters)
TIGR4	*S.pneumoniae*	S4, ST205		All
Fp23a	*S.pneumoniae*	ST205	TIGR4Δ*cps*	3,4,5
TIGR4Δ*SP1772*	*S.pneumoniae*	S4, ST205	TIGR4Δ*SP1772*	5
TIGR4Δ*cps*Δ*SP1772*	*S.pneumoniae*	ST205	TIGR4Δ*cps*Δ*SP1772*	5
TIGR4Δ*cps*Δ*srt*	*S.pneumoniae*	ST205	TIGR4Δ*cps*Δ*srt*	3
R6	*S.pneumoniae*	ST128	D39Δ*cps*	3,4,5
Inv104b	*S.pneumoniae*	S1, ST227		3
Inv200	*S.pneumoniae*	S14, ST9		3,6
OXC141	*S.pneumoniae*	S3, ST180		3
23F	*S.pneumoniae*	S23F, ST81		3
SK120	*S.gordonii*			4
BL21	*E.coli*			4
Top-10	*E.coli*			4,5,6

Table 2.2 – Plasmids used in this study

Name	Plasmid backbone	Insertion	Study (chapters)
pET33bSP1772A	pET33b(+)	SP1772A	4
pET33bSP1772B	pET33b(+)	SP1772B	4
pET33bSP1772C	pET33b(+)	SP1772C	4
pET33bSP1772D	pET33b(+)	SP1772D	4 (Figure 4.7a)
pET33bgatewayply	pET33bgatewa y	Ply	4 (Figure 4.7b
pET33bplySP1772	pET33bgatewa y	SP1772	4
pCR4-Topo-srtA	pCR4-Topo	srtA	4
Topo-erm	TopoTA	Erm	5
TopoΔSP1772	TopoTA	Erm+ SP1772 flanking DNA	5
TopoInv200SP1772-1	TopoTA	Inv200 SP1772 clone 1	6
TopoInv200SP1772-5	TopoTA	Inv200 SP1772 clone 5	6
TopoInv200SP1772-6	TopoTA	Inv200 SP1772 clone 6	6
TopoInv200SP1772-8	TopoTA	Inv200 SP1772 clone 8	6
TopoInv200SP1772-16	TopoTA	Inv200 SP1772 clone 16	6

2.2 Isolating DNA

2.2.1 Genomic DNA from S.pnuemoniae

A loop of glycerol stocks of the strain from which genomic DNA was to be acquired was streaked onto a BAB plate and incubated o/n at 37°C. An optichin disk was placed in an area of high growth to ensure the strain was pure. 20ml BHI was then inoculated with 1 colony from the plate and grown o/n at 37°C. Cultures were centrifuged (4,000 rpm, 15 mins, 4°C) and the pellet was resuspended in 1ml lysis buffer (10mM Tris, 100mM EDTA, 0.5% SDS in sterile double distilled H_2O), and incubated at 37°C for 1hr. 100μg/ml Proteinase K was added before incubating at 50°C for 3 hrs. Following this, 20μg/ml RnaseA was added before incubating at 37°C for 30 min. Then 1ml phenol:chloroform:IAA (25:24:1) was added and samples sharply inverted and centrifuged (13,000 rpm, 3 min, RT). The upper phase was removed to a fresh tube and 0.2 volumes 10M ammonium acetate was added, followed by 600μl EtOH before centrifugation (13,000 rpm, 3 min, RT). The pellet was then air dried for 15min and resuspended in 200μl TE buffer (10mM Tris.HCl, 1mM EDTA). Samples were incubated at 65°C for 10min to encourage resuspension of DNA then stored at -20°C, in aliquots of 20μl.

2.2.2 Plasmid DNA from E.coli

Strains of *E.coli* harbouring the plasmid of interest were grown in 3ml LB with appropriate antibiotic selection, shaken at 220rpm o/n. Plasmid DNA was extracted from the culture using the QIAprep Spin Miniprep kit according to the manufacturer's instructions.

2.3 Transformation of *S.pneumoniae*

BHI (1mM $CaCl_2$) was aliquoted to give 10ml and inoculated with 100μl glycerol stock of *S.pneumoniae* (cells grown to OD_{600} 0.6 in 15% glycerol and stored at -80ºC). This was incubated statically at 37°C until OD_{600} reached 0.6. Two 1ml

aliquots were removed to sterile tubes and 100ng/ml CSP-2 (competence stimulating peptide-2) (Pozzi *et al.*, 1996) was added to each. Competence stimulating peptides are naturally occurring and are able to induce competence in the pneumococcus. There are two major alleles, CSP-1 and CSP-2. Each allele will only induce competence in strains that are naturally sensitive to that allele. TIGR4 is naturally sensitive to CSP-2 and as such this allele is used in the transformation of TIGR4. Following the addition of CSP-2, cultures were incubated at 37°C for 15 min. No DNA was added to one whilst 1μg/ml plasmid DNA was added to the other. The tubes were incubated at 37°C for 1hr 15min. 10μl, 50μl and 200μl of culture were then spread on the surface of both selective and non-selective BAB plates using a sterile spreader. 3 different volumes of culture were used to allow for differing transformation efficiencies, to allow single colonies to be isolated from one of the plates. The plates were incubated o/n at 37°C and then observed for transformants.

2.4 Isolating RNA and creating cDNA

RNA was extracted from 15ml TIGR4 grown to OD_{600} 0.6 using RNeasy Mini Kit according to the standard protocol in the RNeasy Mini Handbook (3rd Edition, June 2001). Outwith the protocol,15mg/ml lysozyme was used in TE buffer, and initial incubation time was 15min at RT.

cDNA was synthesised from RNA using Invitrogen ThermoScript RT-PCR System according to the standard protocol.

2.4.1 Mapping the operon structure of SP1755-SP1773

PCR using cDNA synthesised from RNA isolated from TIGR4 (2.4) was used to assess the operon structure of the ASL. The reaction volume was 25μl (200μM dNTPs, 2mM $MgSO_4$, 2μM Primers, 0.1μl Invitrogen Platinum Taq Polymerase, 1μl cDNA/DNA, 2.5μl Invitrogen Platinum Taq 10x buffer; made up to 25μl with d.H_2O) The conditions were: Initial denaturation 94°C, 2min; Cycle x 30 (Denaturation 94°C, 30s; Annealing see appendix II for all primers 30s; Extension 72°C 40s); Final extension 72°C 5min. Samples were diluted 5:1 sample:loading buffer and run on a 1% agarose gel.

2.5 Microarray analysis

2.5.1 *Microarray construction, DNA labelling and hybridization*

All microarray experiments were carried out by Dr. Donald Inverarity. Microarrays were constructed by robotic spotting of PCR amplicons onto poly-L-lysine-coated glass microscope slides (MicroGrid II; BioRobotics, Huntingdon, United Kingdom). Amplicons were designed to represent each of the annotated open reading frames (ORFs) present in *S.pneumoniae* strain TIGR4 in addition to those unique to the other sequenced strain, R6. The process designed multiple amplicons using Primer3 for all TIGR4 ORFs, as determined by automated analysis of BLASTN comparisons. A single amplicon was selected to represent each ORF based on its lack of similarity to other ORFs on the array using BLASTN analysis to ensure minimal cross-hybridization. DNA was fluorescently labeled and hybridized to the microarray chips. Fluorchrome Cy3 or Cy5 dCTP (Amersham Pharmacia Biotech) was incorporated into whole-genomic DNA by a randomly primed polymerization reaction using large fragment DNA polymerase I (Invitrogen). Whole-genomic DNA comparisons were carried out by competing DNA from the test strains listed in Table 1 with a standard reference DNA from TIGR4. Microarray slides were prehybridized for 20 min at 65°C in a buffer containing 3.5x SSC (1x SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 0.1% sodium dodecyl sulfate (SDS), and 10 mg/ml bovine serum albumin. Labeled DNAs from TIGR4 and the test strains were mixed and purified using a Mini-elute kit (QIAGEN), after which 4x SSC and 0.3% SDS were added. After denaturation at 95°C for 2 min, the DNA mixture was applied to the microarray and hybridized during 18 h at 65°C. Before analysis, the slides were washed once in 1x SSC buffer with 0.06% SDS for 5 min at 65°C and twice in 0.06x SSC buffer at room temperature. Comparative hybridizations were performed between a Cy5-labeled genomic DNA of the reference strain (TIGR4) in competition with Cy3-labeled genomic DNA. Reciprocal dye swap experiments were performed to minimize labelling artefacts.

2.5.2 Data generation and analysis.

Hybridized slides were scanned using a ScanArray (Packard BioScience) according to the manufacturer's instructions, and the median pixel intensity values for

each element on the array were quantified using Quantarray (Packard BioScience). The data were further analyzed using Bluefuse. On the basis of preliminary PCR validations, the cutoff was determined for the normalized intensity ratio of >1.5 indicated the absence of a gene, unless the intensity was >1,500 in both channels (TIGR4 and test strain), in which case we decided that the gene is present. For intensities of <600 for both channels, the result was regarded as ambiguous and therefore was checked by PCR. Genes with a ratio of 1.4 gave poor agreement between the array and PCR results, and so the presence or absence of all such genes was determined by PCR.

2.6 Sequencing

Sequencing of PCR products, specific regions of genomic DNA and plasmid DNA was carried out by a sequencing unit (MBSU, University of Glasgow or DBS genomics, University of Durham). Whole genome sequencing of *S.pneumoniae* was carried out at the Sanger Centre.

2.7 Protein analysis in *S.pneumoniae*

2.7.1 Preparation of S.pneumoniae cultures for protein analysis

100μl glycerol stocks of pneumococcal strains were inoculated into 10ml BHI and grown to OD_{600} 0.6. The bacterial growth was the recovered by centrifugion (4,200 rpm, 10 min, 4°C), followed by resuspension in 1ml PBS. The supernatant, containing culture medium was discarded at this point. The pellet was then sonicated (4 x 30s pulse, 30s hold) to break open the bacterial cells and centrifuged again (13,000rpm; 5 min, 4°C). Supernatant (containing the soluble fraction of cells) and pellet (containing the insoluble fraction of cells) were separated, pellets resuspended in 1ml PBS and stored at -20°C. To analyse, 15μl sample was added to 5μl sample buffer, boiled for 5 min and separated by SDS-PAGE electrophoresis through 3%, 5% or 10% gels. The protocol was later modified as follows: a larger starting volume of culture was used (100ml, 1L); Sonication, resuspensions and centrifugations were modified accordingly; Amicon 100,000mW cutoff centrifugation filters and Millipore 300,000mW cutoff ultrafiltration membranes were used to concentrate the sample.

2.7.2 SDS PAGE

Samples for protein analysis (obtained as described in 2.7.1) were diluted 3:1 with sample buffer and boiled for 5 min. Appropriate gels were used for the size of protein to be detected. Gels were run at 100 Volts for between 80 min to 16 hours, dependant on the size of protein to be detected. Gels were then stained according to the manufacturer's instructions using Commassie stain and destain, carbohydrate stain (Sigma), silver stain (Invitrogen) or transferred to Hybond-C nitrocellulose membranes for western blotting for protein visualisation, all according to standard protocols. Coomassie stain is used to visualise proteins, while easy to use, its detection limit is fairly high. Carbohydrate stain is used to visualise proteins that have carbohydrate associated with them, for the purposes of this study it was used to detect glycosylated proteins. Silver stain is used to visualise proteins that are in lower copy number since it is highly sensitive.

2.7.3 Western blots

In order to detect protein samples seaparated on SDS PAGE gels using specific antibodies, SDS PAGE gels were transferred to Hybond-C nitrocellulose membrane (Amersham Biosciences) and blotted for 90 min at 80 Volts. Blots were then blocked overnight at RT in 3% skimmed milk in Tris-NaCl pH 7.4 with shaking and incubated at 37˚C with shaking for 2h in 3% skimmed milk with an appropriate dilution of primary antibody. Membranes were washed in Tris-NaCl pH 7.4 x 4 and incubated for 1h with 1:1000 HRP-linked secondary antibody (a general antibody raised against the animal that the primary antibody was raised in; in this study that was mouse, rabbit or goat) in 3% skimmed milk, washed x 4 and developed in developing solution (Amersham Bioscience). The reaction was stopped with distilled water.

2.7.4 Dot blots

50µl samples were pipetted onto Hybond-C nitrocellulose membrane (Amersham Biosciences) in 2µl drops, and dried. They were then blocked Blots were then blocked overnight at RT in 3% skimmed milk in Tris-NaCl pH 7.4 with shaking and incubated at 37˚C with shaking for 2h in 3% skimmed milk with an appropriate dilution of primary antibody. Membranes were washed in Tris-NaCl pH 7.4 x 4

Table 2.3 – Primary antibodies used in this study

Antibody	Immunogen	Source	Raised in
α-ply	*S.pneumoniae* ply	Tim Mitchell	Rabbits
α-GspB	GspB, raised by adsorption	Paul Sullam	Goats
α-SP1772D	SP1772D	This study	Mouse
α-SP1772Dply	SP1772Dply	This study	Mouse

and incubated for 1h with 1:1000 HRP-linked secondary antibody in 3% skimmed milk, washed x 4 and developed in developing solution. The reaction was stopped with distilled water.

2.8 Raising antibodies against SP1772

2.8.1 Cloning Fragments of SP1772

Primers were designed (and made by Sigma Genosys) to clone selected fragments of SP1772 into plasmid pET33b(+) (Novagen) to be expressed in *E.coli* DE3 BL21 cells (Novagen). The reaction volume for the PCR was 100μl (400μM dNTPs, 2.5mM MgCl2, 2.4μM primers, 0.5μl Expand High Fidelity PCR system (Roche) Enzyme, 2μl 1μg/μl TIGR4Δcps genomic DNA, 10μl 10x buffer; made up to 100μl with d.H2O) The conditions were: Initial denaturation 94°C, 2min; Cycle x 30 (Denaturation 94°C, 30s; Annealing 50°C - 65°C, 30s; Extension 72°C, 1 min 15 secs); Final extension 72°C 5min. The sample and loading buffer was run in a 1:1 ratio on a 1% agarose gel. To isolate the piece of DNA to be cloned the remaining PCR amplified material was run on a 1% agarose gel 1:5 loading buffer:sample and the band excised from the gel. The DNA was purified using QIAquick gel extraction kit protocol. This purified DNA was then cut with restriction enzymes (50μl DNA, 20μl multicore buffer, 1μl SacI, 1μl BamHI, 128μl s.d.H2O. Incubated 1hr, 37°C) and then purified again using QIAquick gel extraction kit protocol. 3μl DNA was then ligated into the plasmid (1μl cut pET33b(+), 1μl 10x ligase buffer, 1μl ligase, 4μl sdH2O; incubated at RT for 4 hrs). Plasmids were then electroporated into *E.coli* BL21s (Invitrogen). 1μl plasmid was added to 100μl competent BL21s thawed on ice and incubated on ice for 20min. The mixture was removed to a 0.1cm BioRad electroporation cuvette and electroporated using a GenePulser Eletroporator (Voltage 1.5kV, Capacito 25μF, Resistor 200Ω). 200μl LB was added and the mixture was removed to a fresh tube. Samples were incubated for 1hr at 37°C shaking. They were then plated out on selective plates (LBkan 50μg/ml) and incubated o/n at 37°C. 5 colonies were then selected from each sample and grown o/n @ 37°C, shaking. Plasmids were then purified using QIAprep Spin Miniprep kit. PCR was carried out to confirm an insertion, using primers designed for the plasmid outside the cloning site. The reaction volume for the PCR was 25μl (400μM

Table 2.4 Designing fragments of SP1772 for cloning

Fragment name	Primer A	Primer B	Predicted Size (Kb/Kda)
SP1772A	48L	48N	1.1 / 42.8
SP1772B	48M	48N	0.09 / 7.5
SP1772C	48L	48O	1.2 / 45.7
SP1772D	48M	48O	0.16 / 10.4

dNTPs, 2.5mM MgCl2, 2.4μM Primers, 0.125μl Expand High Fidelity PCR system (Roche) Enzyme, 2μl plasmid DNA, 2.5μl 10x buffer; made up to 25μl with d.H2O) The conditions were: Initial denaturation 94°C, 2min; Cycle x 30 (Denaturation 94°C, 30s; Annealing 50°C - 65°C, 30s; Extension 72°C, 1 min 15 secs); Final extension 72°C 5min. 1:1 sample:loading buffer was run on a 1% agarose gel. The sample that gave the strongest signal on the agarose gel at the correct size was sent off for sequencing to MBSU, University of Glasgow.

2.8.2 Creating SP1772ply in a Gateway system

SP1772D was amplified from the TIGR4 chromosome using primers 56L and 56M (see appendix II), containing attB sites. The reaction volume was 50μl (200μM dNTPs, 2mM $MgSO_4$, 2μM 56L and 56M, 0.2μl Invitrogen Platinum Taq Polymerase, 1.5μl TIGR4 DNA, 5μl Invitrogen Platinum Taq 10x buffer; made up to 50μl with d.H_2O) The conditions were: Initial denaturation 94°C, 2min; Cycle x 30 (Denaturation 94°C, 30s; Annealing 58°C) 30s; Extension 72°C 3min); Final extension 72°C 8min. This PCR product was then used in a one-step LR-BP cloning reaction (Invitrogen), according to the manufacturer's instructions and utilising the pET33bgatewayply vector as the entry vector (see figure 4.6).

2.8.3 Protein expression

Strains harbouring plasmids containing the constructs for expression were grown, shaking at 37˚C to OD_{600} 0.4 in TB plus LBkan 50μg/ml. To induce expression of protein, IPTG was added to a final concentration of 1mM, cultures were then incubated shaking at 30˚C over time ranges of 4 – 8 hours to determine the optimal amount of time for incubation, which was determined by SDS PAGE analysis. Cultures were centrifuged at 13,000 rpm for 3 min. Pellets were resuspended in PBS plus Dnase I (Sigma) and Benxamidine, an inhibitor of serine proteases, (Sigma) and cells were sonicated (4 x 1 min pulse, 1 min hold). Cell lysates were centrifuged at 12,000 rpm at 4˚C for 30 min. Following confirmation by western blot that the desired protein was in the supernatant, supernatants were then purified using 0.2μm syringe filters.

2.8.4 FPLC purification of proteins

All of the purified proteins utilised in this study were his-tagged, and were purified from crude extracts of the cell by FPLC on a Nickel-charged NTA colum (Qiagen Superflow resin). Proteins were eluted using a gradient of 0-300 mM imidazole in PBS on the FPLC. Resulting fractions were analysed by SDS page to determine which fractions contain the protein required. These fractions were then pooled for dialysis.

2.8.5 Dialysis of FPLC purified proteins

Fractions containing the relevant protein following FPLC were pooled and dialysed to remove salt. Dialysis tubing (Medicell International Ltd, UK) was boiled in 2% bicarbonate solution with 1mM EDTA for 10min, washed in distilled water and stored in ethanol at 4˚C. Tubing was then washed in distilled water, protein fractions added and the tubing sealed at both ends. The tubing was then placed in 1L phosphate buffer and placed on a magnetic stirrer overnight at 4˚C. The buffer was changed twice further. The resulting solution in the tubing was then concentrated using Amicon ultra-15 centrifugal tubes according to the manufacturer's instructions (Millipore, UK) and stored at -20˚C.

2.9 Construction of mutants

2.9.1 Creating TIGR4ΔcpsΔsrt

TIGR4ΔcpsΔsrt , a mutant with deleted capsule and sortase A genes was generated for experiments outlined in Chapter 5, was generated in the following manner: SrtA (spr1098) and flanking DNA in the R6 genome was amplified using primers 19H and 19I and cloned into pCR4-TOPO (Invitrogen) according to the manufacturer's instructions. Primers 19W and 19X were then used to delete an internal fragment of srtA and introduce an AscI site between the 5' and 3' flanking sequences required for homologous recombination. Once confirmed by sequencing, the plasmid was digested using the AscI enzyme and an erythromycin cassette from pGhost5 (Biswas et al., 1993) was ligated into the plasmid. This plasmid was then transformed into unencapsulated TIGR4, since unencapsulated TIGR4 is both easier to transform and to utilise in protein

expression visualisation experiments given the lack of heavily expressed capsule, to give TIGR4ΔcpsΔsrtA.Transformants were analysed for presence of correct deletions by PCR of their extracted genomic DNA. The reaction volume was 25μl (200μM dNTPs, 2mM NEB Vent $MgSO_4$, 2μM Primers, 0.15μl Promega Taq Polymerase, 2μl DNA, 2.5μl NEB Vent 10x buffer; made up to 25μl with d.H_2O) The conditions were: Initial denaturation 95°C, 5min; Cycle x 32 (Denaturation 95°C, 30s; Annealing 55°C 30s; Extension 72°C 2 min); Final extension 72°C 10min. 5:1 sample:loading buffer was run on a 1% agarose gel.

2.9.2 Creating TIGR4ΔSP1772 and TIGR4ΔcpsΔSP1772

The left flank PCR for generating TIGR4ΔSP1772: The reaction volume was 50μl (200μM dNTPs, 2mM $MgSO_4$, 2μM 50J and 49W, 0.2μl Invitrogen Platinum Taq Polymerase, 1.5μl T4 DNA, 5μl Invitrogen Platinum Taq 10x buffer; made up to 50μl with d.H_2O) The conditions were: Initial denaturation 94°C, 2min; Cycle x 30 (Denaturation 94°C, 30s; Annealing 55ºC) 30s; Extension 72°C 1min10s); Final extension 72°C 5min. RF PCR: Conditions are identical to above, but primers used were 49X and 50K. Erm PCR: Conditions are identical to above, but primers used were 49Z and 50A, and template DNA was Topo erm (Created by amplifying erm cassette from TJMGL0496 using 49Z and 50A and cloning into Topo TA). Crossover PCR: The reaction volume was 50μl (200μM dNTPs, 2mM $MgSO_4$, 2μM 50J and 50K, 0.2μl Invitrogen Platinum Taq Polymerase, 1μl LF, 1μl RF, 1.5μl erm, 5μl Invitrogen Platinum Taq 10x buffer; made up to 50μl with d.H_2O) The conditions were: Initial denaturation 94°C, 5min; Cycle x 30 (Denaturation 94°C, 30s; Annealing 56.4ºC) 30s; Extension 72°C 3min10s); Final extension 72°C 10min. Product is SP1772 P. This was cloned into TopoTA (Invitrogen) according to the manufacturer's manual Version J. Colonies were grown and verified by PCR then selected for sequencing. Plasmids were then used to transform unencapsulated TIGR4, an unencapsulated derivative of TIGR4. Colonies were selected and genomic DNA isolated and used to transform TIGR4 to create T4ΔSP1772.

2.10 *In vitro characterisation of strains*

2.10.1 Biofilm model

The biofilm model used was developed by Moscoso and colleagues (Moscoso *et al.*, 2006) and some experiments were carried by Dr. Patricia Romero-Fernandez. Strains were grown in C+Y (a minimal, chemically predefined) medium until an A_{550} of 0.4 was reached. A volume of cells equivalent to A_{550} of 0.6 was centrifuged and resuspended in 1 ml of prewarmed C+Y. The culture was then diluted and 200 µl were dispensed in a 96-well polystyrene plate. Plates were incubated at 34˚C for 4 to 6 hours and bacterial growth was determined by measuring the A_{595} using a plate reader. Biofilm formation was measured by staining with 1% crystal violet (CV) at room temperature for 15 min, solubilization of the biofilm with 200 µl of 95% ethanol, and determination of the A_{595}.

2.11 In vivo experiments

All *in vivo* experiments were carried out in accordance with the UK Animals (Scientific Procedures Act) 1986. All animal handling was carried out by Dr. Gill Douce, Dr. Gavin Paterson and Dr. Clare Blue.

2.11.1 Preparation of standard innocula

Strains for *in vivo* experiments were streaked onto blood agar plates. One colony was selected and grown to OD_{600} 0.6 at 37˚C. FCS was added to give a final concentration of 20% and strains were stored at -80˚C. Viable counts were then carried out and aliquots of bacteria at 10^7CFU/ml (colony forming units/ml) were prepared. Viable counts were carried out again to ensure the correct concentration of bacteria. The relevant strain of mouse was then infected i.p with 100µl 10^7CFU/ml (to give a final dose of10^6CFU). The animals were then sacrificed at 20 h and blood was inoculated into 10 ml BHI and grown o/n at 37˚C. Cultures were centrifuged at 3,000 rpm and re-suspended in 1ml BHI. 100 ul was used to inoculate BHI plus 20% heat inactivated FBS (Gibco BRL, UK) and incubated at 37˚C until OD_{600} was 0.6. Aliquots of 1ml of the cultures were taken

and stored at -80˚C. Where appropriate, antibiotic selection was used to ensure the growth of mutant strains.

2.11.2 Viable counting of bacteria

A glycerol stock of 900µl of standard innocula were centrifuged at 13,000rpm. Cells were resuspended in 900µl sterile PBS (Sigma-Aldrich). Samples were then diluted 1/10 in a 96-well U-bottomed plate with sterile PBS in the range of 10^{-1} to 10^{-6}. 20µl of each dilution in triplicate was then plated on BAB plates, and each triplicate was also plated in duplicate and left to dry by a flame. Plates were then incubated o/n at 37˚C. Plates were examined and the sector where 10-100 colonies per spot were then counted to give an average CFU/ml per strain. Strains were then diluted to an appropriate dilution for the latter infection studies in animals.

2.11.3 Intraperitoneal model of infection

Viable counts were carried out and aliquots of bacteria at 10^7CFU/ml were prepared. Viable counts were carried out again to ensure the correct concentration of bacteria. MF-1 mice were then infected i.p with 100µl 10^7CFU/ml (to give 10^6CFU/dose). The animals were tail bled at 24h to get blood counts. They were then continued to be monitered for disease progression and survival. Viable counts were then carried out on the blood samples.

2.11.4 Pneumonia model of infection

Viable counts were carried out and aliquots of bacteria at 10^7CFU/ml were prepared. Viable counts were carried out again to ensure the correct concentration of bacteria. The relevant strain of mouse (MF-1 or Balb/c) was then infected i.n with 50µl 10^6CFU/dose (to give 5 x 10^5CFU/ml – MF-1) or 50µl 10^7CFU/ml (to give 5 x 10^6CFU/ml – Balb/c). Bacterial counts were taken at 12h and 24h (MF-1) or 2 and 3 days post infection (dpi) (Balb/c) from the blood and lungs. Animals were then continued to be monitored for disease progression and survival. Viable counts of the harvested bacterium were then carried out on the blood and lung samples, as per the methodology of viable counting.

2.11.5 Colonisation model of infection

Viable counts were carried out and aliquots of bacteria at 10^6CFU/ml were prepared. Viable counts were carried out again to ensure the correct concentration of bacteria. The relevant strain of mouse (MF-1) was then infected i.n with 10µl 10^6CFU/ml (to give 10^4 CFU/dose). Bacterial counts were taken at 5 days from nasopharyngeal washes and tissue and viable counts of bacteria taken.

2.11.6 Immunisation of mice

4 female mice were injected subcutaneously at the base of the neck with 10µg of protein (Chapter 2) mixed with 7µl aluminium hydroxide and made up to 100µl with s.PBS. Initial immunisations were given on day 1. Sample bleeds were taken from the tail at days 13, 27 and 38. The sera were in all cases unreactive with SP1772 and hence repeat boosts were given on the day following the sample bleed.

2.12 Inv200 *SP1772* genomics

2.12.1 Southern blotting

Southern blotting was carried out using the DIG (Digoxigenin) DNA labelling and detection kit (Roche), according to the manufacturer's instructions. gDNA was prepared (Section 2.2.1), and the probe utilised was amplified using primers 56O and 56N (See appendix II). The reaction volume was 50µl (200µM dNTPs, 2mM $MgSO_4$, 2µM 56O and 56N, 0.2µl Invitrogen Platinum Taq Polymerase, 1.5µl DNA, 5µl Invitrogen Platinum Taq 10x buffer; made up to 50µl with d.H_2O) The conditions were: Initial denaturation 94°C, 2min; Cycle x 30 (Denaturation 94°C, 30s; Annealing 56°C) 30s; Extension 72°C 15min); Final extension 72°C 25min.

2.12.2 Cloning

Genomic DNA was extracted from Inv200 (Table 2.1). SP1772 in Inv200 was amplified: The reaction volume was 50µl (200µM dNTPs, 2mM $MgSO_4$, 2µM 56F

and 56G, 0.2μl Invitrogen Platinum Taq Polymerase, 1.5μl Inv200 DNA, 5μl Invitrogen Platinum Taq 10x buffer; made up to 50μl with d.H_2O) The conditions were: Initial denaturation 94°C, 2min; Cycle x 30 (Denaturation 94°C, 30s; Annealing 55°C) 30s; Extension 72°C 4min); Final extension 72°C 10min. This entire pcr mix was cloned into Topo TA and electroporated into Top10 cells. Transformants were picked from the plates and colony pcr was carried out: The reaction volume was 25μl (200μM dNTPs, 2mM $MgSO_4$, 2μM 56F and 07F, 0.2μl Invitrogen Platinum Taq Polymerase, 2.5μl Invitrogen Platinum Taq 10x buffer; made up to 25μl with d.H_2O) The conditions were: Initial denaturation 94°C, 2min; Cycle x 30 (Denaturation 94°C, 30s; Annealing 55°C) 30s; Extension 72°C 4min); Final extension 72°C 10min. Plasmids were then extracted from each strain, and the above PCR was repeated using plasmid DNA as confirmation and in preparation for sequencing.

Chapter 3

Results

Investigating the presence of the ASL in pneumococcal strains

3 Investigating the Presence of the ASL in Pneumococcal Strains

3.1 Introduction

The pneumococcus is able to cause a wide range of diseases, and there are many reasons for this, including host factors. However, sequencing of pneumococcal genes and genomes has revealed that variations between pneumococcal genomes are substantial, and likely to be one of the most important contributory factors in the diversity seen in disease outcomes of pneumococcal infections. It is therefore of interest to study genes that are present or absent in different isolates of *S.pneumoniae*, since they may have a role in the different behaviours of strains.

There have been several studies carried out to examine the diversity in the pneumococcus (Obert *et al.*, 2006). One microarray analysis idenitified 22 regions of diversity among 13 selected pneumococcal genomes (Silva *et al.*, 2006). In this analysis, one of the regions of diversity (RD), RD20, was found to be present in a minority of strains, including TIGR4 (Figure 3.1). Examination of the genome sequence of TIGR4 (Tettelin *et al.*, 2001) revealed that RD20 contained 18 genes, all of which are absent in R6, an unencapsulated avirulent sequenced strain of *S.pneumoniae* (Hoskins *et al.*, 2001), (Figure 3.2). R6 does contain the genes flanking RD18 in TIGR4 at either end (SP1755, SP1774).

Most of the genes in RD18 were found to have homologues in *S.gordonii* (See Figure 3.3). One of these genes, SP1772, encodes a serine rich repeat (SRR) protein. The other 17 genes fell into three categories (i) Glycosyl-transferases, (ii) Alternative secretion genes and (iii) Genes of unknown function (Figure 3.3). Work in *S.gordonii* has shown that the glycosyl-transferases glycosylate GspB, the SRR, whilst the alternative secretion genes secrete GspB via a pathway distinct from the main Sec system of the cell (Bensing and Sullam, 2002; Takamatsu *et al.*, 2004). SecA2 and SecY2 are homologous to SecA and SecY, many of the other genes have no homologues in the main pathway (Bensing *et al.*, 2005). Work in *S.gordonii* has also determined the functions of two of the

Figure 3.1 – Regions of diversity among pneumococcal strains

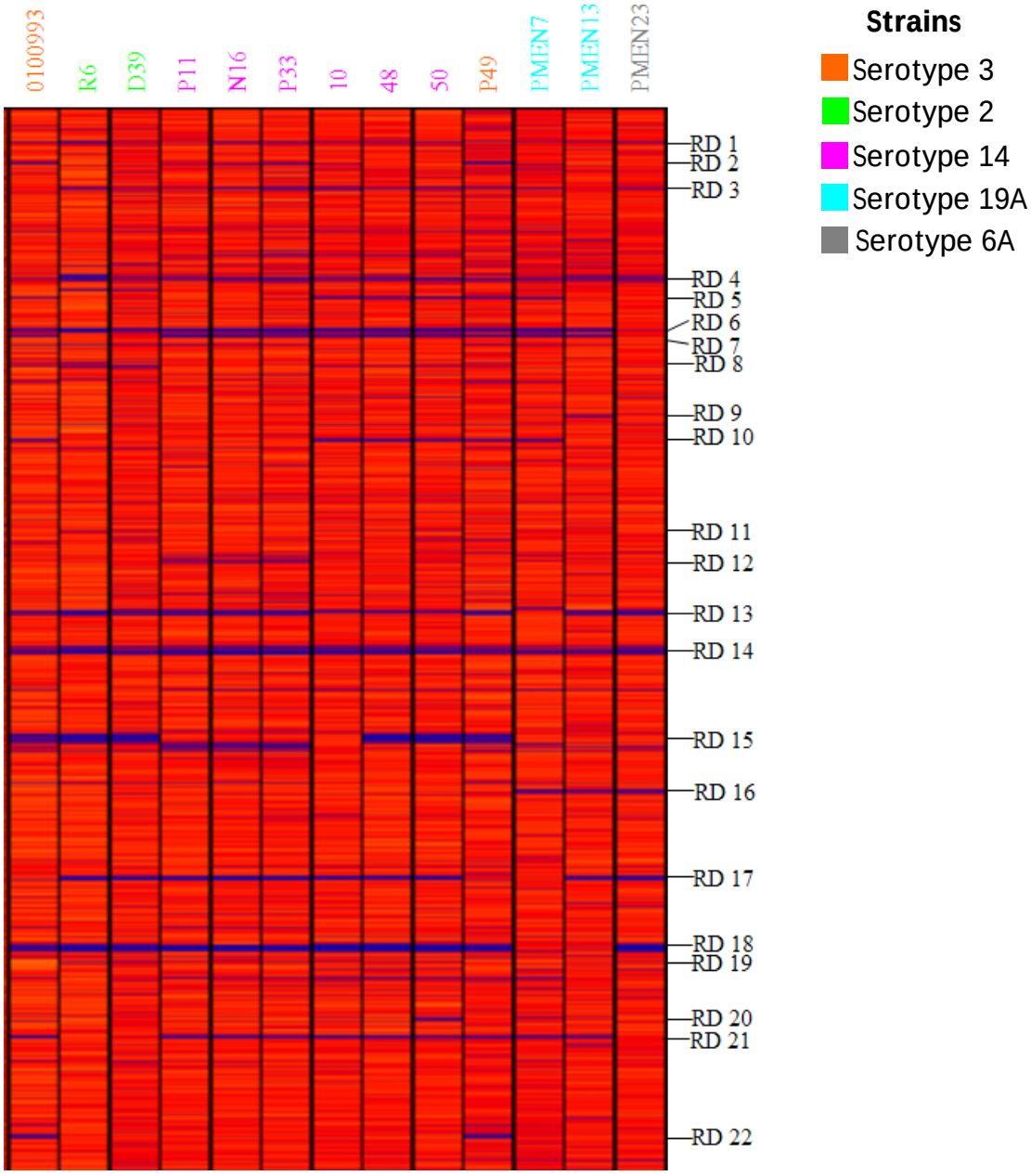


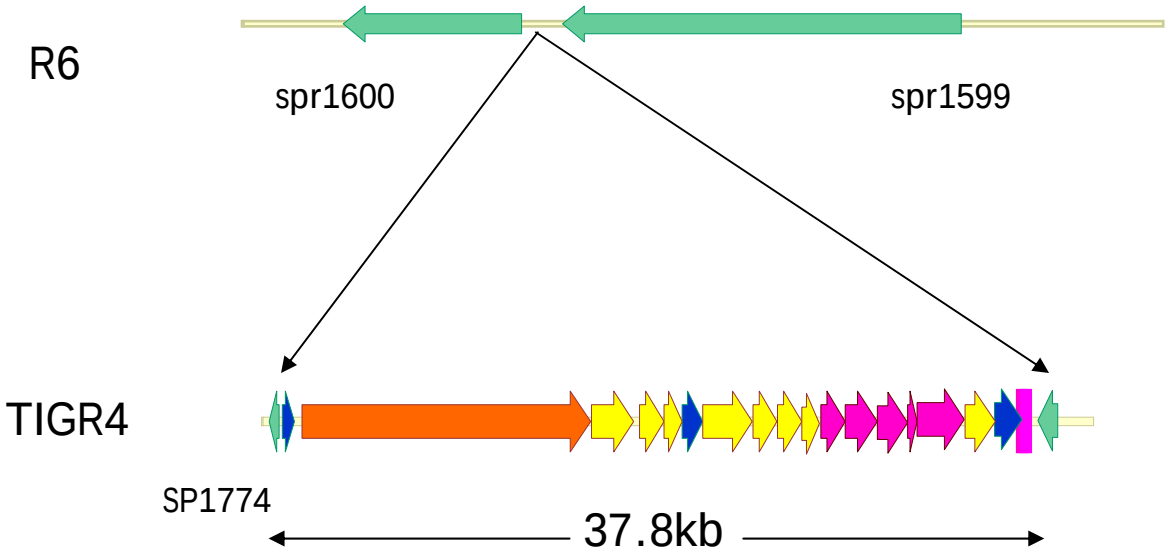
Figure from (Silva *et al.*, 2006). Comparative genomic hybridisation was carried out a panel of pneumococcal isolates. 22 regions of diversity were found amongst the 13 selected strains of varying serotypes. Red indicates the presence of a gene, and blue the absence. This study highlighted RD20 as a region that varies in its presence and absence between strains. The composition and size of each region of diversity is presented in Table 2.
3 further RDs were characterised beyond this analysis, and are outlined in Table 1.

Table 3.1 – Regions of diversity among pneumococcal strains

RD	Approximate size (kb)	SP no. of variable genes
1	9	SP0067-074
2	5.8	SP0109-0115
3	5.6	SP0163-0168
4	14.2	SP0346-0360
5	3.3	SP0378-0380
6	5.4	SP0394-0397
7	12.6	SP0460-0468
8	7.1	SP0473-0478
9	5.6	SP0531-0544
10	11	SP0643-0648
11	8	SP0664-0666
12	4.4	SP692-0700
13	1.7	SP0888-0891
14	7.9	SP0949-0954
15	11.9	SP1050-1065
16	9.2	SP1129-1147
17	33.7	SP1315-1352
18	12.1	SP1433-1444
19	10.3	SP1612-1622
20	34.8	SP1756-1773
21	5.3	SP1793-1799
22	3.2	SP1828-1830
23	3.2	SP1911-1918
24	9.4	SP1948-1955
25	5.3	SP2159-2166

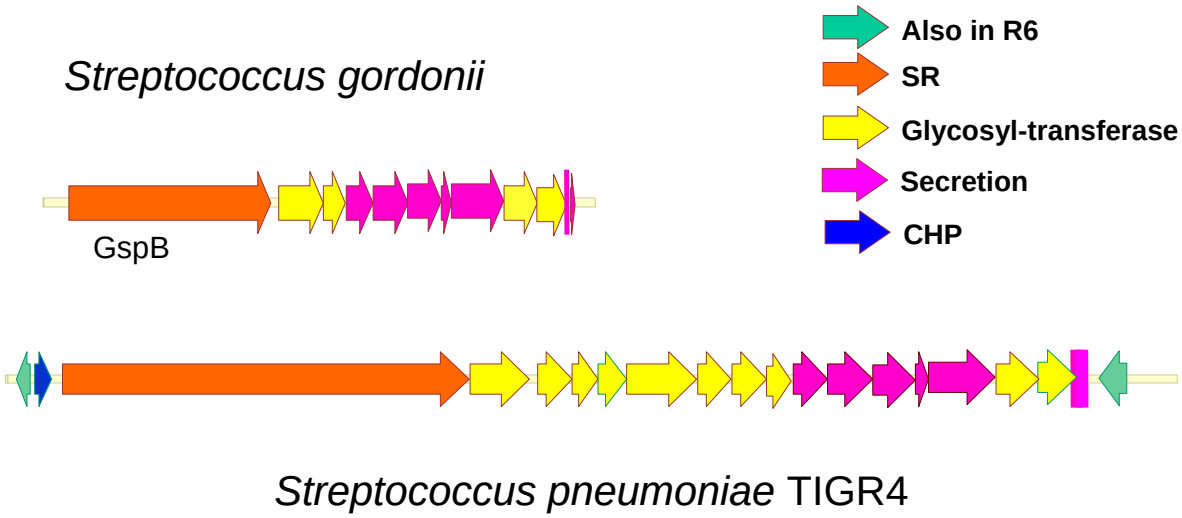
Table taken from (Silva *et al.*, 2006). Comparative genomic hybridisation was carried out a panel of pneumococcal isolates. 22 regions of diversity were found amongst the 13 selected strains of varying serotypes. The regions of diversity outlined in Figure 3.1 are detailed here.

Figure 3.2 – RD20 in R6 and TIGR4



The homologues of TIGR4 SP1754 and SP1774 in R6 are spr1600 and spr1599, respectively. Annotation of the genetic region between spr1599 and spr1600 in R6, SP1754 and SP1774 in TIGR4 reveals RD20 is present in TIGR4, and absent from R6. The intergenic region between spr1599 and spr1600 in R6 is found to contain an insertion in TIGR4, which contains 18 genes: SP1755 – SP1773.

Figure 3.3 – Functions of the genes in RD20



Functions are assigned to the genes of RD20 according to their homology with genes from *S.gordonii* (established by BLAST analysis). Aside from the gene encoding the SRR protein, these genes were found to be homologous to characterised and uncharacterised glycosyltransferases and alternative secretion machinery.

genes of unknown function as additional glycosyl-transferases (Takamatsu *et al.*, 2004).

The work presented in this chapter was carried out to further determine the distribution of RD20 amongst pneumococcal isolates by sequencing and comparative genomic hybridisation, to determine the operon structure of RD20 in TIGR4 and to investigate whether genes in RD20 of TIGR4 are transcribed.

3.2 Results

3.2.1 RD20 in strains currently being sequenced at the Sanger Centre

The Wellcome Trust Sanger Institute, in collaboration with our laboratory, are currently sequencing the genomes of sixteen pneumococcal isolates (http://www.sanger.ac.uk/Projects/S_pneumoniae/). Four of these isolates are utilised in this study as they were available at the time the research was carried out (23F, Inv200, Inv104b and OXC141). Given the diversity seen in comparative genomic hybridisation studies and individual gene sequencing, the whole availability of genome sequences is invaluable of our understanding of the bacterium to give us a more complete picture of the genetic diversity of this species than serotyping, sequence typing and comparative genomic hybridisation studies can provide.

The strains selected for sequencing were as follows: Inv104b is representative of the major clone of serotype 1, ST227, which is considered to be highly invasive. Inv200 is a serotype 14 ST9 clone, this clonal group is major disease causing group in children (Brueggemann *et al.*, 2003). OXC141 is a serotype 3, ST180 clone. This clonal group is associated with carriage in children (Brueggemann *et al.*, 2003). 23F is a major multi-antibiotic resistant clone that is associated with carriage (Coffey *et al.*, 1991; McDougal *et al.*, 1992; Munoz *et al.*, 1991).

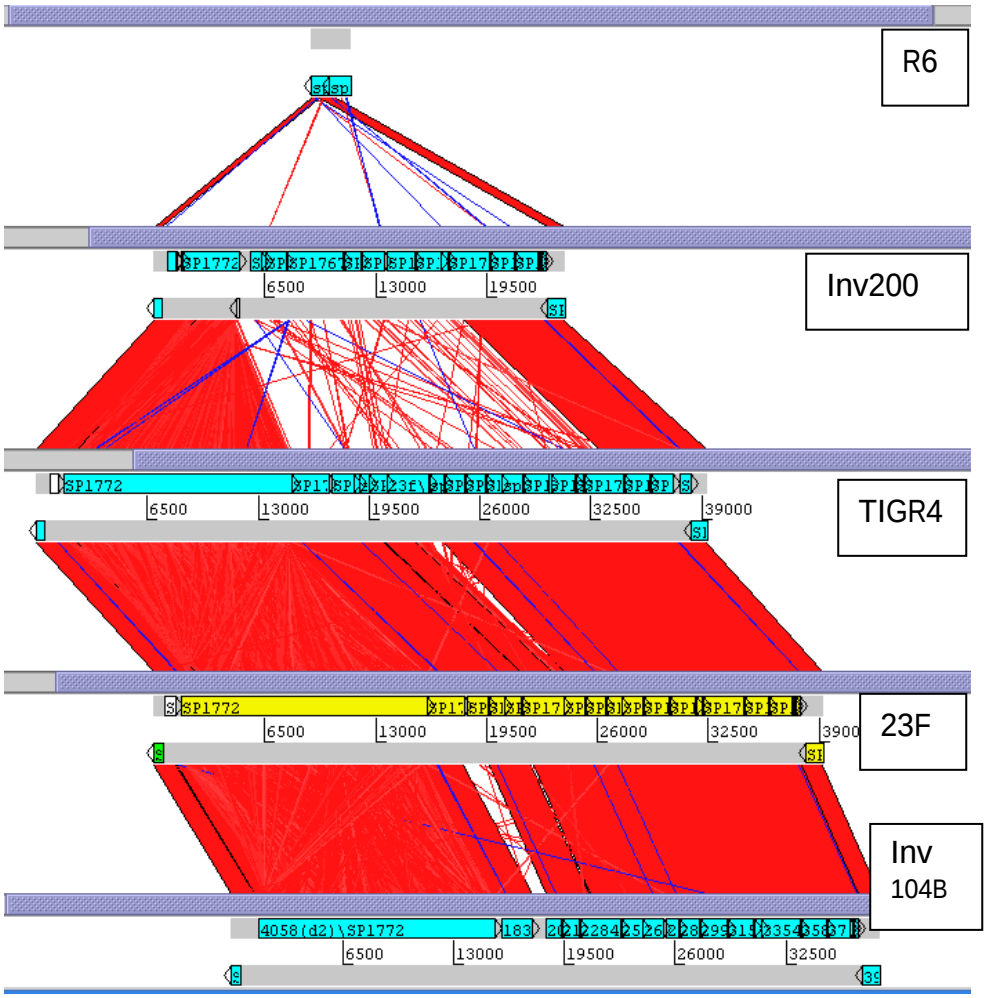
The full sequence of these strains therefore afforded an opportunity to establish not only whether the locus containing SP1772 was present, but also to look at the variation that exists in the locus between strains. Certainly, between different species of bacteria, the genomic organisation of the locus has been shown to differ, particularly in the truncation of the SRR and fewer numbers of

glycosyl-transferases (Takamatsu *et al.*, 2004). In order to determine whether each strain contained a homologue of SP1772, the nucleic acid sequence of SP1772 was blasted against assembled contigs of each strain. Three strains, 23F, Inv104B and Inv200 were all found to contain homologues of SP1772. The specific contig of each strain was then aligned with the TIGR4 genome using Webact, a whole genome alignment tool available from the Sanger Centre (http://www.webact.org/WebACT/home). The locus in 23F and Inv104b appears to be similar to that in TIGR4, with minor differences arising presumably from divergent evolution since the locus was transferred (Figure 3.4). The 23F genome encodes an additional glycosyl-transferase. The locus in Inv200 was markedly different from that in the other genomes. Not only is the locus much smaller in size, but the genes are arranged differently. Inv200 was then directly compared to each strain individually to see if it was most disparate from TIGR4, but closer related to the other strains. This was found not to be the case (Figure 3.5).

Loci were annoated using Vector NTI (Invitrogen) and BLAST. Initially, SP1754 and SP1774 (ORFs at the beginning and end of RD20) were both blasted against the subject genome, and as such the location of the ASL within the genome was identified. All of the ORFs between SP1754 and SP1774 were then predicted. Each ORF was translated into its amino acid sequence, and the sequence was then blasted against the full NCBI database. Functions were assigned to genes on the basis of homology (Figure 3.6). Interestingly Inv200 was found to contain a truncated version of SP1772, and as such presented an opportunity to overcome some of the difficulties, as discussed in the next results chapter, of working with SP1772. This work is outlined in chapter 6.

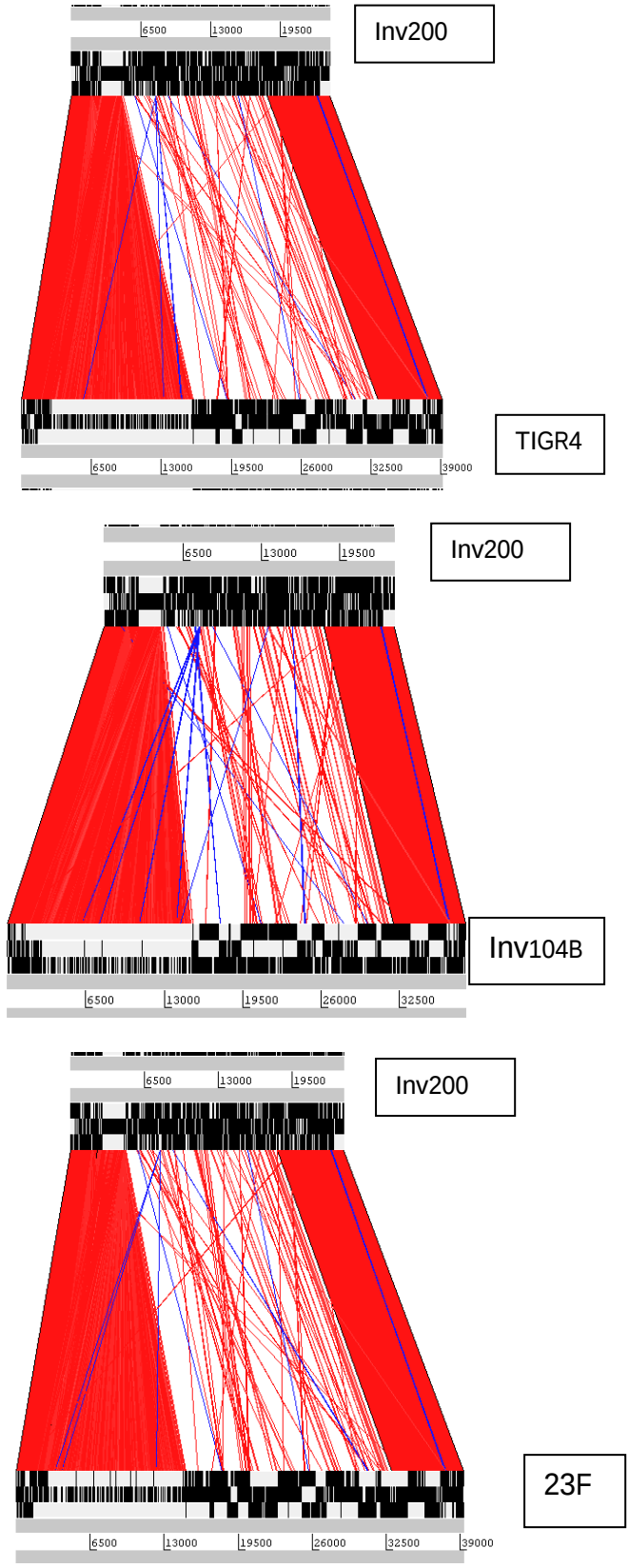
As previously mentioned, the ASLs identified in 23F and Inv104b were similar to that of TIGR4, however the sequence of the Inv200 ASL was divergent from these. The locus in Inv200 contains a truncated SRR, while *SP1772* in TIGR4 is 14.7kb, the *SP1772* homologue in Inv200 is 3.3kb. The alternative secretion machinery associated with *SP1772* is conserved between the two strains. However, while TIGR4 contains 8 different glycosyl-transferases associated with *SP1772*, Inv200 contains only 5. This is quite feasibly due to the truncated *SP1772* in Inv200 requiring less glycosylation.

Figure 3.4 – Webact analysis of sequenced strains



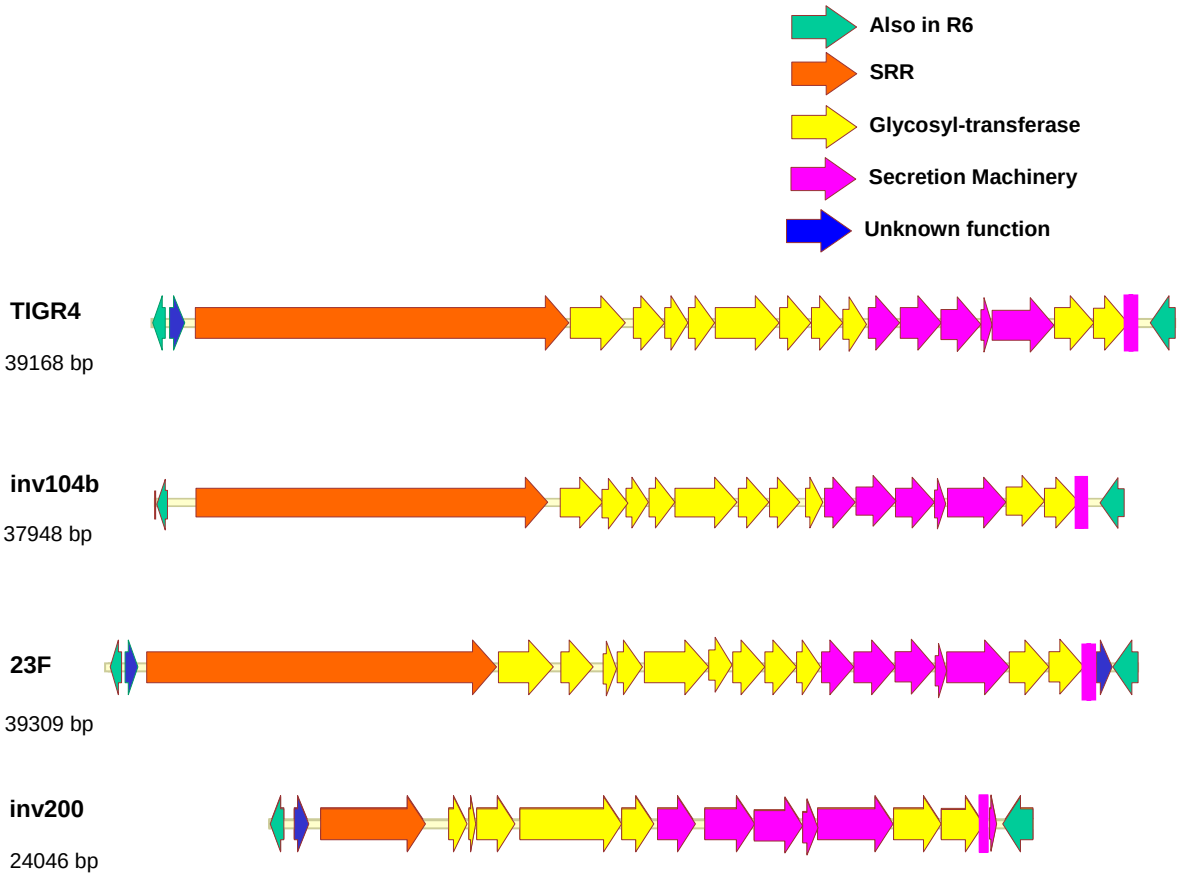
The region between the homologues of spr1599 and spr1600 in R6 in strains containing a homologue of SP1772 in TIGR4 were aligned using Webact (Sanger). Red lines represent areas of homology between the two strains, blue lines represent areas of inverted homology between the two strains and white areas represent areas of no homology. This analysis shows that the region (RD20 in TIGR4) is fairly well conserved between TIGR4, Inv104b and 23F but the sequence of the locus is somewhat divergent in Inv200.

Figure 3.5 – Webact comparison of Inv200 directly with other sequenced pneumococcal strains



SP1772 and associated genes appear divergent from other sequenced pneumococcal ASLs in a multi-strain comparison. A direct comparison of Inv200 with each strain is presented here, and shows that there is little similarity for the middle section in comparison with each ASL.

Figure 3.6 – Annotation of sequenced strains containing RD20



RD20 in Inv104b, 23F and Inv200. Each ORF was predicted and the translated protein product was blasted against TIGR4. Functions were assigned to genes on the basis of homology (Cut-off utilised was 50%).

This was interesting since the ASL is not found in all pneumococcal isolates, but is found in other bacteria, and supports the theory that the locus moves by horizontal gene transfer between different species of bacteria. The genomic organisation of the locus varies from species to species. A figure from Takamatsu and colleagues is presented in Figure 3.7. This figure shows that the number of genes associated with an SRR in any given species can vary significantly, mainly in the region between the SRR and the homologue of SecY. The Inv200 locus varied from other pneumococcal loci in this region, however, there were genes present in this region, unlike in *S.gordonii, S.aureus* or *S.epidermidis*. *S.algalactiae* similarly contained genes in this region, and to compare how related the locus in Inv200 and the locus in *S.algalactiae* were, an ACT comparison was done. This is presented in Figure 3.8. This revealed that the locus in Inv200 and *S.algalactiae* are also divergent, suggesting that a recent movement of the locus between these two species has not occurred. However, it does suggest that this location in the locus is able to vary significantly between strains and species.

OXC141, a serotype ST180 strain isolated from the nasopharynx of a healthy child (Brueggemann *et al.*, 2003), was found not to contain a homologue of SP1772. When SP1754 and SP1774 were blasted against oxc141 they were found, and the intergenic region examined it contained no predicted ORFs. Each additional gene in the ASL was blasted against the genome to verify that the ASL had inserted elsewhere in the genome, and the absence of these genes within OXC141 confirmed that this strain does not contain the ASL (Figure 3.9).

3.2.2 Microarray analysis of Pneumococcal strains

Previously, genomic hybridisation carried out in our laboratory identified the ASL as a region of diversity, RD20, which was found in a minority of strains (Silva *et al.*, 2006). This study was therefore expanded by looking at more pneumococcal strains, and looking at large groups of the same ST to determine if MLST was a sufficiently sensitive classification for the pneumococcus. MLST is a system that classifies the pneumococcus according to variations in the sequence of 7 housekeeping genes (Enright and Spratt, 1998), and it was sought to identify if in some instances these housekeeping genes can remain conserved while other parts of the genome can vary.

Figure 3.7 – Comparison of the ASL in different species of gram positive bacteria

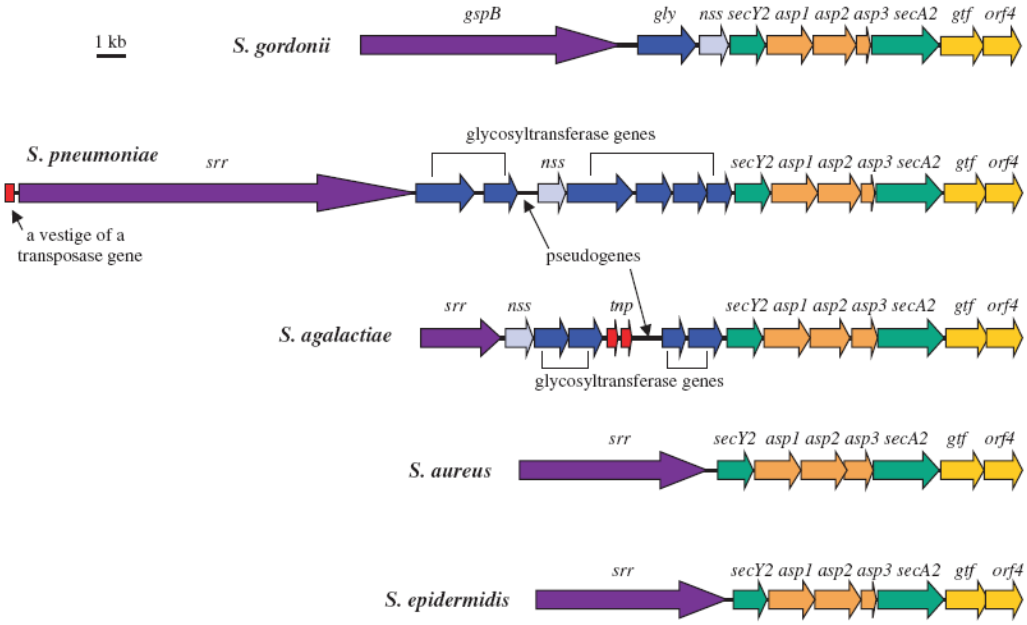
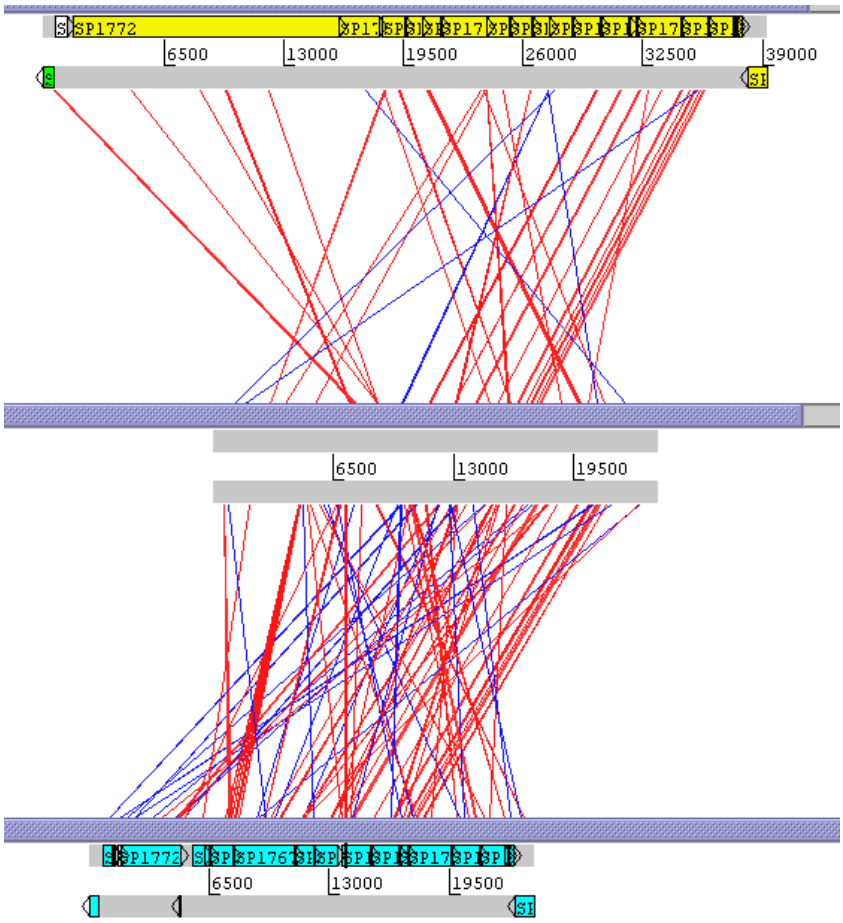


Figure from (Takamatsu *et al.*, 2004a). ASLs in several species of gram positive bacteria are annotated. Purple indicates the SRR, blue indicates a glycosyl-transferase, green are SecA2 and SecY2 homologues, orange indicate alternative secretion genes and yellow indicates essential glycosyl-transferases.

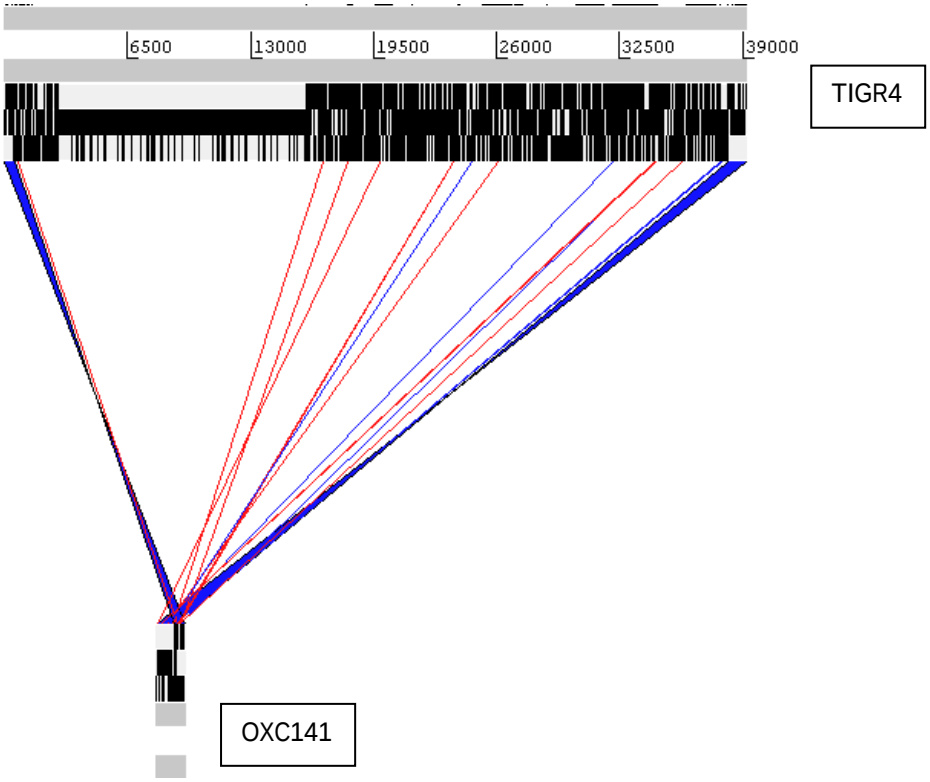
Figure 3.8 – Webact alignment of RD20 in Inv200 with homologous regions in other species



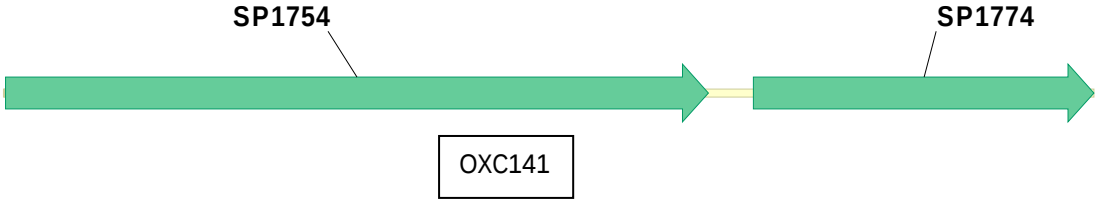
The sequence of RD20 in Inv200 was aligned in RD20 to available loci from other bacteria using webact. This resulted in less homology found between inv200 and other bacteria than with other pneumococcal strains. This figure presents an example of an alignment with *S.algalactiae*.

Figure 3.9 – RD20 in OXC141

(a) Webact alignment of RD20 in TIGR4 and OXC141

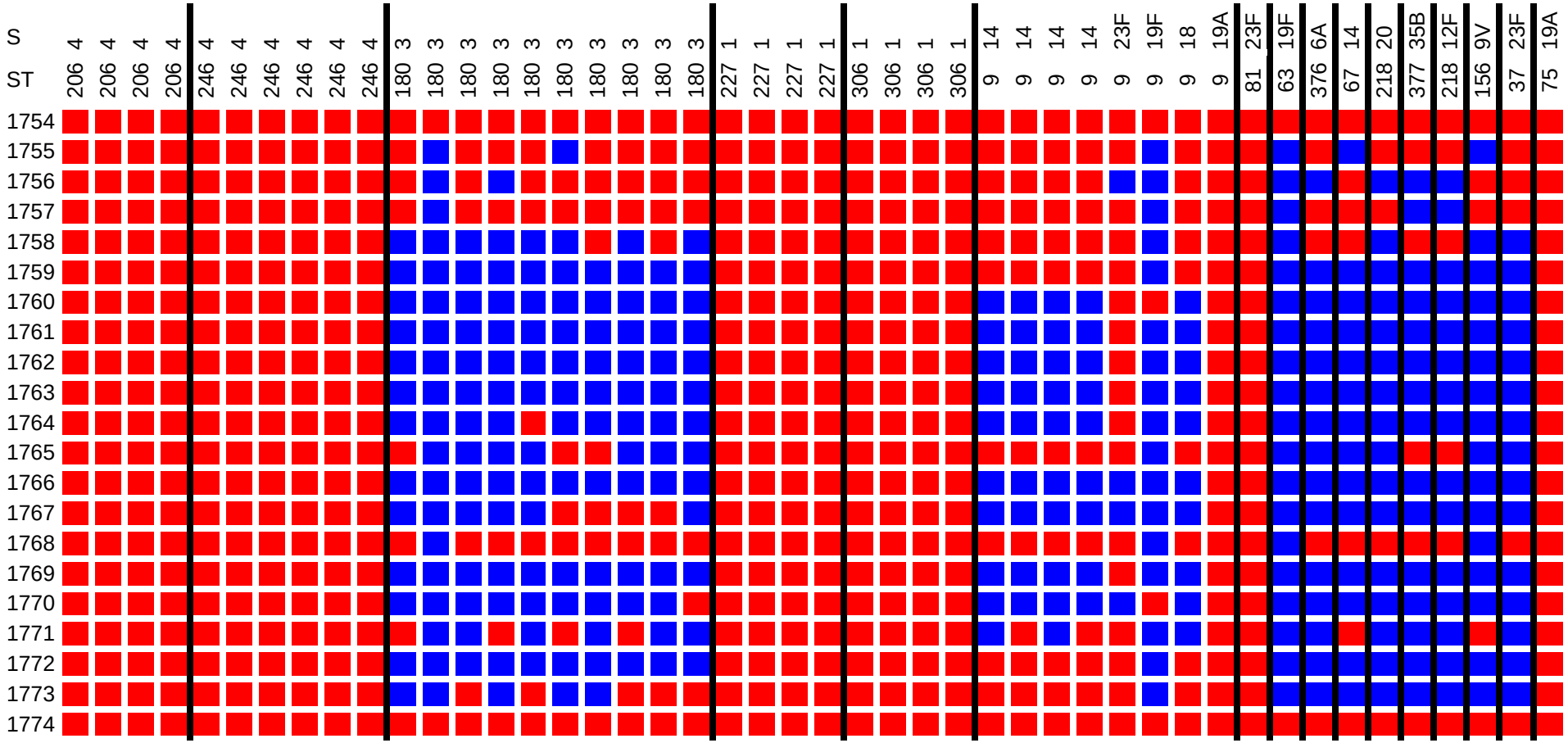


(b) The genomic region between homologues of SP1754 and SP1774 in OXC141



(a) The sequence of RD20 in TIGR4 was aligned to the region between homologues of SP1754 and SP1774 in OXC141. The locus does not appear to be present. (b) Annotation of the locus reveals the absence of the locus in OXC141.

Figure 3.10 – Microarray analysis of *S.pneumoniae* isolates



Microarray analysis was carried out to determine the presence and absence of genes in RD20 in TIGR4. The ORFs are listed. Red indicates the presence of the gene, while blue indicates the absence of the gene. Generally the locus is present or absent in keeping with ST and serotype, for example all serotype 4 strains examined contain the locus, and all serotype 3 strains do not.

Figure 3.10 presents the microarray data, analysed using Blue fuse software. In most serotypes and MLSTs examined, there was consistency in whether RD20 was present or absent. In all serotype 4, MLST 256 strains examined, for example, the locus was always present. In all serotype 3 strains, ST180 strains most genes in the locus were always absent. Where some genes from the locus did appear to present in serotype 3, ST180 strains did appear to be present, it can be assumed this is due to homology to other genes in the pneumococcus. For example SP1768 often appears to be present in serotype, ST180 strains. This is a gene of unknown function, and it is possible that the probe for this strain has homology to other genes in *S.pneumoniae.* There is no homology between this gene and any other in the OXC141 genome however, which shows there are differences at the genomic level between serotype 3, ST180 strains.

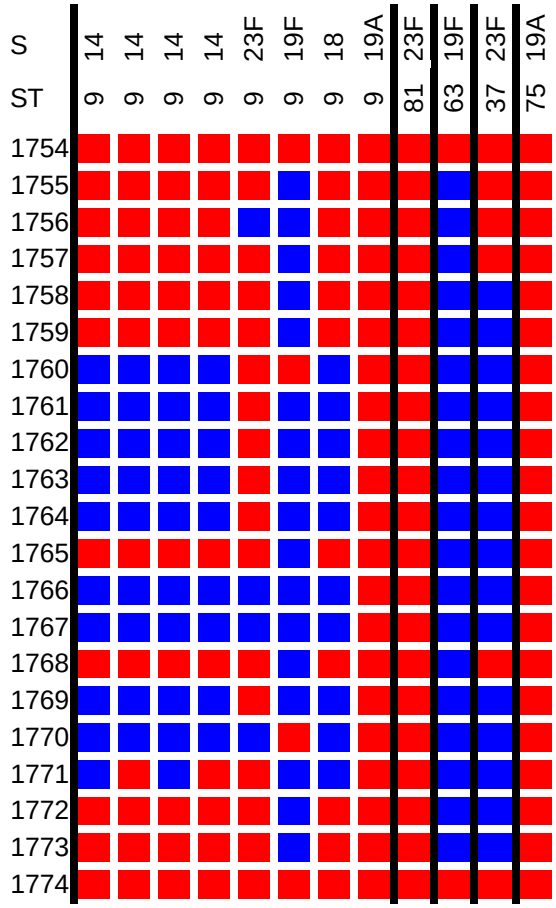
3.2.2.1 Presence of the locus does not always correlate with serotype or ST

In the previous section, the correlation between presence and absence of locus and the serotype and ST of the strain under examination was noted. However this study highlighted that there are exceptions to this rule.

All ST9 strains, the majority were also serotype 14, evaluated in the study were found to contain the locus with the exception of an ST9 strain with capsule type 19F. It is possible that capsule type 19F is not compatible with this locus, in that the cell is not able to produce both capsule and the product of this locus. This theory supported by the finding that a capsule type 19F ST 63 strain that also did not contain the locus (Figure 3.11). 19F is similar in capsule type to 19A, in fact the replacement of only 2 genes in the 19F capsular locus allows the locus to synthesis capsule of a 19A type (Morona *et al.*, 1999). Examination of 19A strains included in the array (ST9, ST41) revealed that they did contain the ASL, and as such suggests that the absence of the ASL in 19F strains examined is either due to a very subtle difference in capsule biosynthesis or, alternatively, due to the requirements of the strains harbouring the locus. However, due to the limited number of 19F strains included in the study, conclusions cannot be drawn.

One of the ST9 strains containing the locus was of capsule type 23F. Examination of other 23F capsule types, showed that an ST81 strain did contain the locus (validated by the occurrence of the locus in the 23F ST81 strain being sequenced

Figure 3.11 – Microarray analysis of specific *S.pneumoniae* isolates



Generally the presence of the locus is the same in keeping with ST or serotype. An ST9, 19F strain does not contain the locus. An ST63, 19F also does not contain most of the genes in the locus. This could possibly be because the capsule type of 19F is not compatible with the locus.
Strains of the same capsule type can also vary in their presence of the locus. Strains with capsule type 23F and ST9 and ST81 on the array contain the locus. A 23F ST37 does not contain the locus. This could signal capsule switching has occurred.

at the Sanger Centre), while a ST37 did not. This could indicate that capsule switching has occurred, and that the ST37 strain acquired the 23F locus in place of a capsule previously present. Capsular switching is able to occur due to the conserved location of the capsule across pneumococcal strains (Serotypes 3 and 37 are the exceptions to this). The genes encoding capsule are located between *dexB* (glucan 1,6-alpha-glucosidase) and *aliA* (oligopeptide ABC transporter, oligopeptide-binding protein AliA), and as such, a highly competent species such as the pneumococcus is able to exchange this genetic material relatively easily. There is a pressure on the pneumococcus to switch capsule, it is highly immunogenic and has been used in vaccine formulations directed against the pneumococcus. This has created a pressure on the pneumococcus to acquire different capsular types (Bentley *et al.*, 2006; Brueggemann *et al.*, 2007).

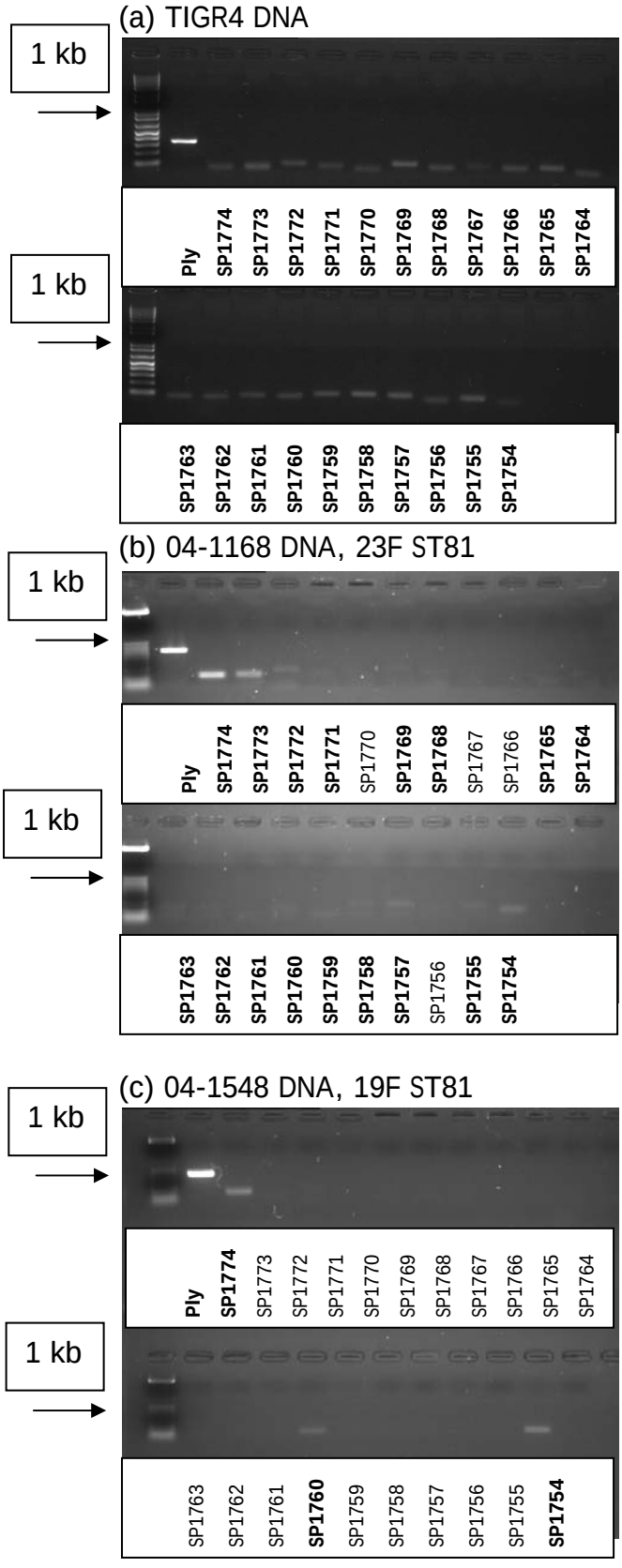
PCR confirmation of each gene in RD20 was carried out on two of the above mentioned strains, a 23F and 19F, both ST81 (Figure 3.12). These PCRs confirmed the results from the array, that the locus was present in the 23F strain and absent in the 19F strain.

Interestingly, the ST9 strains included in the array contain many genes absent in TIGR4. Inv200, the ST9 strain sequenced at the Sanger Centre, contains a divergent locus from that of TIGR4. The results from the array suggest that this locus is highly divergent in the ST9 population, and there is the possibility that this is due to their other genomic content, or a requirement of the strain, its behaviour, or alternatively the compatibility between the locus and the serotype 14 capsule type, since ST9 strains are predominantly serotype 14.

3.2.3 RNA expression of genes in the ASL

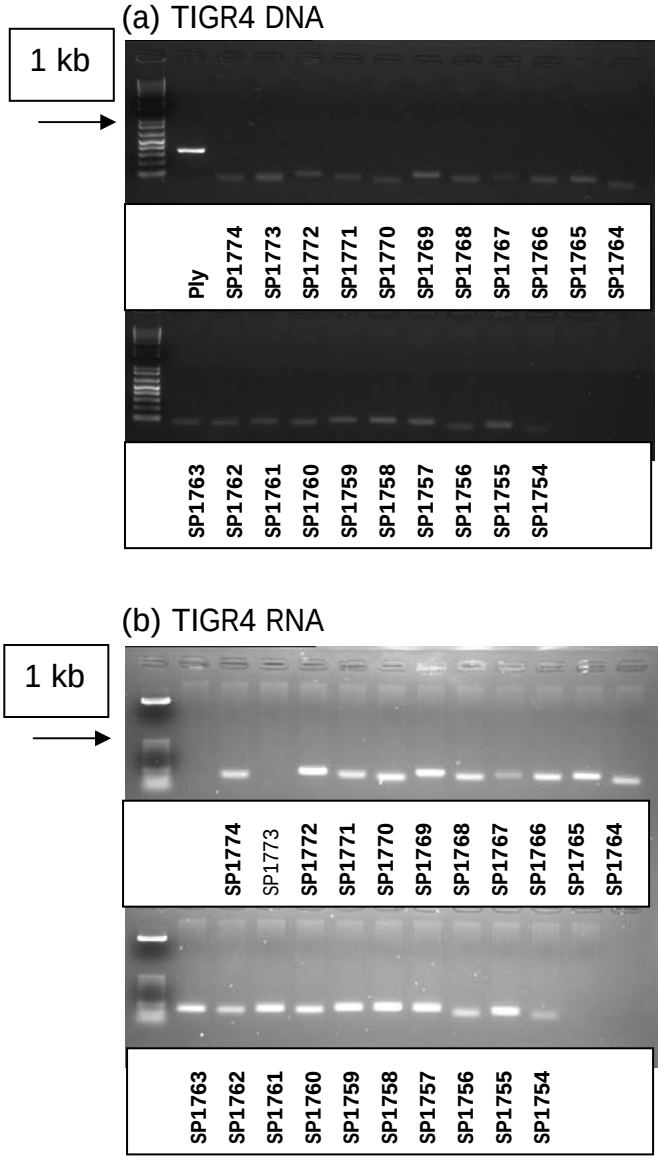
The DNA of the ASL is present in TIGR4, and so it was investigated whether the RNA of all of the genes within the locus is expressed. RNA was isolated from bacteria grown in BHI to midlog, and RT-PCR for each gene was carried out. Primers for each gene were designed against the TIGR4 genome, and each PCR was optimised using TIGR4 DNA. The experiments showed that mRNA from 17 of the 18 genes is expressed (Figure 3.13). The gene that was not shown to be expressed, SP1773 is a transposase. It was presumably involved in the translocation of this locus to TIGR4, but whose expression is no longer required.

Figure 3.12 – PCR validation of Microarray Data



PCR confirmation of microarray analysis results. (a) All of the primers were designed against TIGR4 DNA and are validated (b) 04-1168, the 23F ST81 strain, contains most of the genes in the locus (c) 04-1548, the 19F ST81 strain, does not contain most of the genes in the locus. Ply was used as a positive control. The presence of genes is dentoed in bold.

Figure 3.13 – RT-PCR of genes in RD20



RT PCR was carried out to determine if the genes in RD20 of TIGR4 are also expressed as RNA. (a) The primers are validated using TIGR4 DNA. (b) RT-PCR confirms that all of the genes in the locus are expressed, with the exception of SP1773. Ply was used as a positive control. The presence of genes is denoted in bold.

The operon structure of the locus was then determined. This was done by designing primers that flanked the start and end of each gene. RT-PCR was carried out on TIGR4 cDNA to determine how genes were co-expressed. Pairs of primer were used that gave a product of the end of one gene and the start of its adjacent gene and the intergenic region for each gene pair if the genes are transcribed together (Figure 3.14), (see appendix II for primers). The results for each gene pair allowed a map of the operon structure of RD20 to be drawn (Figure 3.15). This showed that three transcripts are expressed (The RNA of SP1773 is not expressed, as shown in Figure 3.13). Figure 3.15 also shows an analysis of the transcripts by putative functions of the genes, and membrane association (predicted using the GES – Goldman, Engelman and Steitz – hydrophobicity scale).

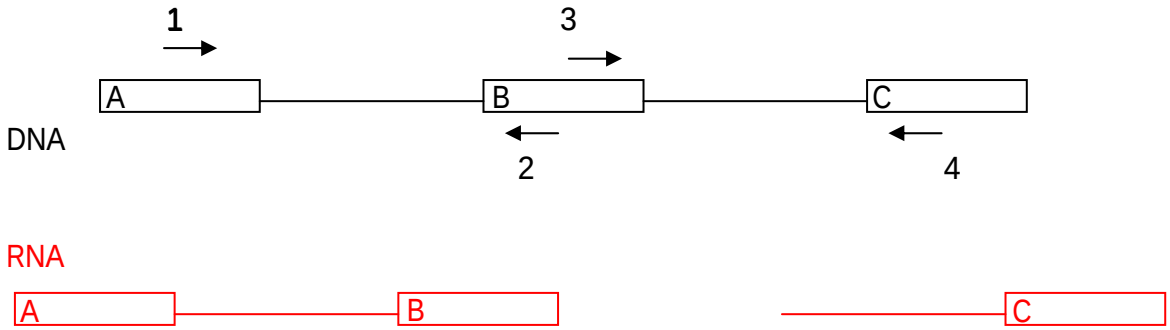
The transcript containing SP1772 is 16kb and also contains SP1771, a putative glycosyl-transferase. Therefore levels of this glycosyl-transferase will presumably be controlled at the same level as SP1772. This glycosyl-transferase is not predicted to be membrane bound.

SP1770 is predicted to be transcribed alone. It is possible that this glycosyl-transferase is required in very high, low or variable levels depending on the niche of the bacteria. It may therefore be unsuitable to be co-transcribed with either the preceding or following transcript. In order to establish the levels of transcription, qRT-PCR could be carried out.

The following transcript of genes SP1769-SP1759 is 15kb and contains the largest number of genes. SP1759-SP1763 are all predicted by homology to be genes involved in alternative secretion. Some are predicted to be membrane bound, and others not. SP1763-SP1767 and SP1769 are all predicted to be glycosyl-transferases that are not membrane bound. SP1768 encodes a glycosyl-transferase which is predicted to be membrane bound. SP1764-SP1769 encode glycosyl-transferases which are often absent in other species of bacteria with contain the locus (Figure 3.4). It is interesting that they are co-transcribed with genes that are commonly found associated with all SRRs.

The next transcript is 3kb and is comprised of genes SP1755-SP1758, which encode two glycosyl-transferase and two alternative secretion proteins. All four products are predicted to be membrane associated. The two glycosyl-

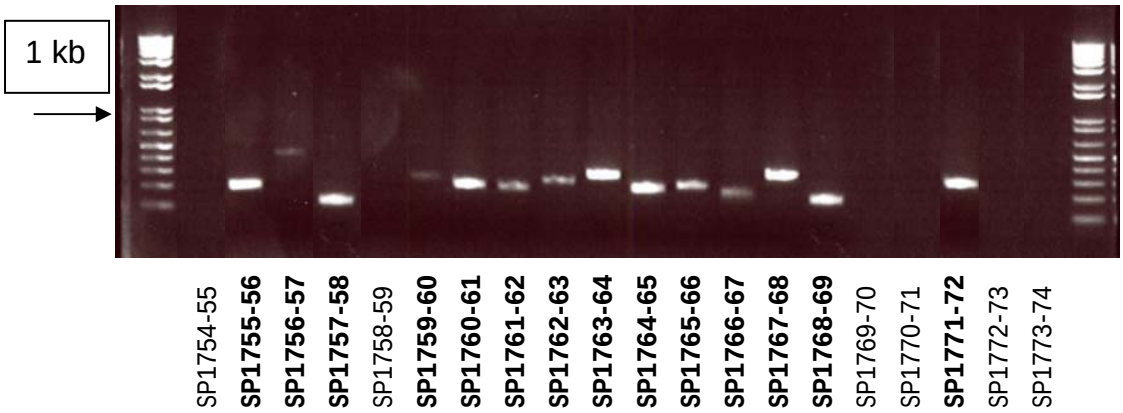
Figure 3.14 – Mapping the Operon Structure of RD20



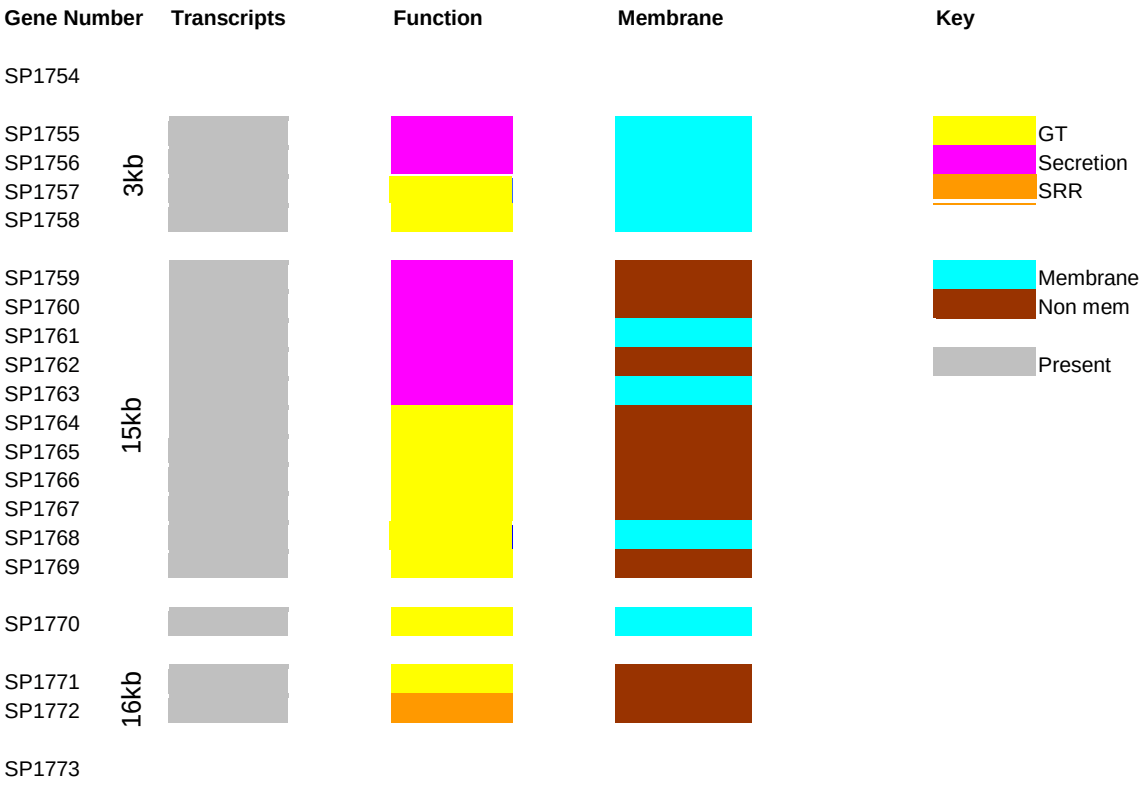
Primers were designed to flank the end of one gene and the beginning of the next gene in the genome of TIGR4. The primers were validated using TIGR4 genomic DNA. The primers were then used to amplify TIGR4 RNA. In the example above, Primers 1 and 2 give a product, and therefore gene A and B are expressed together. Primers 3 and 4 do not give a product, and therefore genes B and C are not expressed on the same transcript.

Figure 3.15 – The operon structure of RD20 in TIGR4

(a) TIGR4 RNA



(b) Operon Mapping



Primers were designed to map the operon structure of RD20 in TIGR4, and validated using TIGR4 DNA (Figure 3.12). The primers were then used to amplify TIGR4 RNA and the operon structure was found. Bold denotes the presence of the transcript (b) The operon structure was then correlated with the putative functions of genes in RD20, and the predicted membrane association.

transferases are found in all species containing the SRR, however, the two alternative secretion genes are not.

3.3 Discussion

RD20 was identified in TIGR4 and shown by comparative hybridisation studies to vary in its presence across the pneumococcal population (Silva *et al.*, 2006). Analysis of a larger number of strains suggests that the locus is associated with strains that cause disease, and it is notably absent from all serotype 3, ST180 strains which are strongly associated with carriage (Brueggemann *et al.*, 2003). Another comparative genomic hybrdisation study has also noted a correlation between the presence of this locus and the ability of an isolate to cause invasive disease (Obert *et al.*, 2006).

Since the locus is large, and has been implicated in virulence this presents interesting questions about the function of the locus in the pathogenicity of *S.pneumoniae*. A study by Hava and Camilli identified SP1760, SP1770, SP1771 and SP1772 as virulence factors (Hava and Camilli, 2002). Work to elucidate the function of the genes in the locus is presented in the following chapters. Further analysis in this chapter has also shown that the locus is not identical in each strain of *S.pneumoniae* where it is present. This divergence lies not only in the sequence of individual genes, but also in the presence and absence of particular genes. The most striking difference in the number of genes is found in the comparison of TIGR4 and Inv200. RD20 in Inv200 contains only 5 glycosyltransferases, whilst in TIGR4 the locus contains 8 glycosyltransferases. The SRR in Inv200 is greatly truncated compared to that of SP1772. It is feasible therefore that the number of glycosyltransferases in the locus required by each strain is proportional to the size of the SRR. There could be some redundancy amongst the glycosyltransferases, however it could also be that the larger the SRR, the more variation in glycosylation is required, and thus it requires a greater number of diverse glycosyl-transferases to facilitate this. This could be to facilitate greater immune evasion, or alternatively to utitlise different sugars in the bacterial cell. The locus in *S.gordonii* also contains far fewer glycosyltransferases, and the SRR is much smaller (Bensing and Sullam, 2002).

However, *S.algalactiae* also contains a greatly truncated SRR, and yet encodes more glycosyl-transferases than *S.gordonii, S.aureus* or *S.epidermidis* (Takamatsu *et al.*, 2004). Where glycosyl-transferases are absent in Inv200 and other species, the function of these genes could be related, in part, to the size of the SRR. However, where these genes are present in Inv200, the function of the genes could be specific to the function of the locus in *S.pneumoniae*. This could be further investigated by sequencing of this locus in more strains, and comparing the number of glycosyltransferases with the size of the SRR. In addition, by elucidating the functions of the glycosyltransferases it could be determined what the function of variation in sequence and presence of these genes could be.

The variation seen in the size of SRR and the locus harbouring it across pneumococci and other species, and its absence from all strains examined of closely related species *Streptococcus pyogenes, Streptococcus mutans* and *Enterococcus faecalis* suggest that the locus has not come from a common ancestor, but has been disseminated by horizontal transfer. This is feasible particularly when the natural competence of *S.pneumoniae* and *S.gordonii* is considered. The transposases present in *S.pneumoniae* and *S.algalactiae* also suggest a recent transfer of this locus to the species.

The variation in genomic content of the locus also implies that the functions of the SRR can vary, not only between species but also between strains. Fap1 in *S.parasanguinis*, a member of the oral flora, has been shown to be responsible for biofilm formation and adhesion to teeth, both important in the formation of dental plaque (Froeliger and Fives-Taylor, 2001). SraP in *S.aureus* and GspB and Hsa of *S.gordonii* meanwhile have all been shown to aid in binding of bacteria to platelets, a first step in the organisms causing infective endocarditis (Bensing and Sullam, 2002; Siboo *et al.*, 2005). However, while Hsa has also been shown to bind secretory immunoglobulin A, salivary mucin MG2 and salivary agglutinin (Jakubovics *et al.*, 2005; Takamatsu *et al.*, 2006), GspB only binds the latter two. The only major difference in the loci associated with GspB and Hsa is the size of the SRR (Hsa is smaller than GspB), and a variant glycosyl-transferase (Takamatsu *et al.*, 2006). This implies that relatively small changes in the structure of the locus can alter the function of the resultant SRR, and as such, the diversity seen in the locus within the pneumococcal population could be due

to the differing behaviour of strains harbouring the strain, or alternatively due to the acquisition of the locus from different sources due to the occupation of differing niches by different strains.

Comparative genomic hybridisation studies presented here demonstrate that the presence or absence of the locus can generally be correlated with the serotype and ST of a strain. However, there are exceptions to this. Strains of ST9 generally were found to contain the locus, albeit with the absence of several genes. This could be expected when considering Inv200, the ST9 serotype 14 strain, which contains a smaller version of the locus with a smaller SRR and fewer glycosyltransferases.

However, one ST9 strain of serotype 19F was found not to contain the locus. Another 19F strain included in the study, of ST63, also did not contain the locus. SP1772 and capsular biosynthesis pathways are feasibly related in the cell of the pneumococcus. The full expression of both requires a large amount of sugars from the cell. It is possible that in some instances the cell either cannot afford the expenditure of all the sugars required to fully synthesize both, or alternatively that the pathways interfere with one another to the extent that they cannot co-exist within one strain. The idea that the 19F capsule and SP1772 are incompatible is possible, but diminished somewhat by the observation that all strains of capsule type 19A contain the locus. Capsule type 19A is very closely related to 19F (Morona *et al.*, 1999), and it is unlikely that the minor differences in both the synthesis and resultant capsule are enough to constitute a difference in the ability of a strain to host SP1772 and associated genes.

When strains of different serotypes are the same ST type, it is proposed to be due to capsule switching (Coffey *et al.*, 1998). However this study revealed that there are also other differences in the genomes of strains of the same ST. This could be because the locus is incompatible with certain capsule types, as discussed above, and so when the strain acquires a different capsule type, it cannot support the expression of the genes and so loses the locus. However, it is also possible that strains which are the same ST but different capsule types are not the result of capsule switching, but rather the result of genomic divergence. This can be further analysed by comparing the presence and absence of other genes these strains, or a more complete picture could be gained by sequencing

the strains and looking at sequence diversity at the gene level. In addition, by looking at expression of the locus under different conditions, the conditions under which expression of the locus in induced in nature may have been deduced, or indeed what has driven the presence or absence of the locus in natural selection could have been elucidated.

The variation in the presence of *SP1772* and associated genes in strains of *S.pneumoniae,* combined with the finding that several genes in the locus are virulence determinants demonstrates a role for this locus in the pathogenicity of *S.pneumoniae*. It has been shown here that the RNA for these genes is expressed in TIGR4, and we therefore sought to investigate if SP1772 is expressed by the cell, and what role the protein plays in the virulence of TIGR4.

Chapter 4

Results

Analysis of SP1772 Expression

4 Analysis of SP1772 Expression

4.1 Introduction

SP1772 has a homologue in *Streptococus gordonii*, GspB. GspB was identified in a transposon mutagensis screen, in which investigators attempted to identify the gene in *S.gordonii* responsible for binding platelets, an important step in the pathogenesis of the organism in causing infective endocarditis (Bensing and Sullam, 2002). Mutants significantly attenuated in their ability to bind platelets were sequenced around the insertion site of the transposon. It was found that the mutant most deficient in platelet binding had an 8 kb deletion in the chromosome of the parent strain. The 8 kb deletion was found by sequence analysis to contain nine contiguous genes which were found by homology to encode glycosyl-transferases, homologues of SecA and SecY, and genes of unknown function. SecA and SecY in *E.coli* have been shown to be essential to the survival of the cell as they have an essential role in protein secretion (Danese and Silhavy, 1998), and thus the investigators sought to determine the effect of this deletion on the proteome of the mutant strain. No differences were identified by SDS PAGE of cell extracts of the mutant compared to the wild type strain in the 8-200 kDa range, however above 216 kDa (the largest molecular weight marker) there was a band present in the wild type strain that was absent in the mutant strain. The investigators then sequenced upstream from the deletion in the mutant strain and found a gene encoding a product of 286 kDa (Bensing and Sullam, 2002). In addition to the previous three glycosyl-transferases immediately downstream of this gene, one of the genes of unknown function was additionally identified as responsible for glycosylating this gene (Takamatsu *et al.*, 2004a).

The predicted protein product of the *SP1772* gene, SP1772 (See Figure 4.2), is 412kDa without glycosylation. There are 8 glycosyl-transferases adjacent to SP1772 in the TIGR4 chromosome. It is unknown what the molecular weight of the native protein product SP1772 is although, as in the case of GspB it could be assumed to be larger than 412kDa due to glycosylation by the adjacent glycosyltransferases in the chromosome (Takamatsu *et al.*, 2004a). In addition, there are 8 glycosyl-transferases predicted to be adjacent to *SP1772* in TIGR4

whilst only 4 have been identified in adjacent to GspB in *S.gordonii* (Takamatsu et al., 2004b).

RT-PCR of all of the genes in the locus, presented in chapter 3 shows that the RNA of all of the genes in locus is expressed, with the exception of SP1773, a transposase. The work presented in this chapter was carried out to establish if the protein product of *SP1772* RNA, SP1772, is expressed in TIGR4.

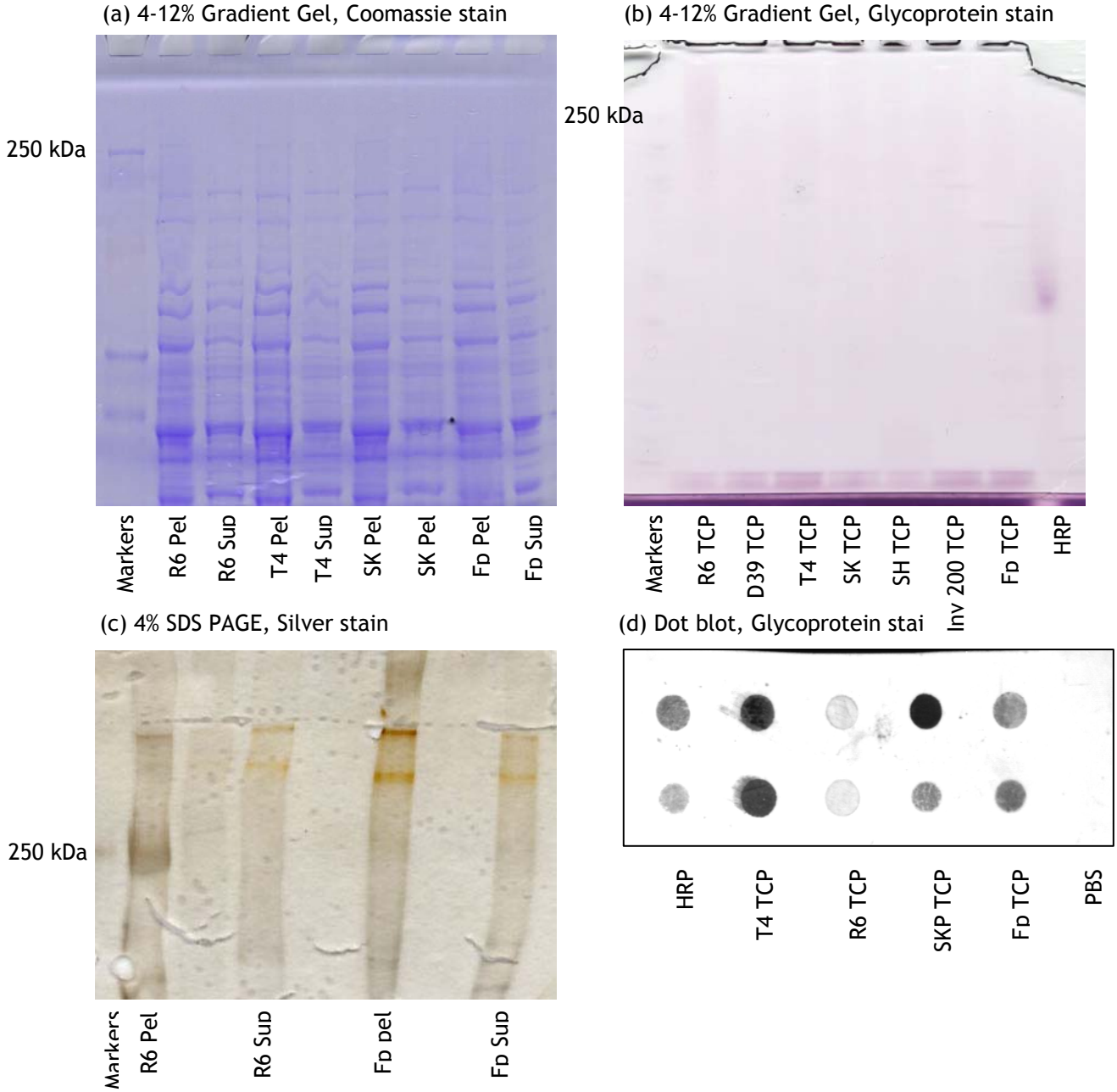
4.2 Results

4.2.1 Detection of SP1772 in TIGR4 cell extracts by SDS PAGE and Dot blot analysis

Cell extracts of TIGR4 were sonicated, fractions separated by centrifugation and then analysed to determine whether SP1772 is expressed by TIGR4. Due to the LPXTG motif in SP1772 it is predicted to be anchored to the cell surface by sortase, and so it is likely to be present in the insoluble fraction of centrifuged cell extracts. Regardless, in all instances either the insoluble and soluble fraction, or else total cell protein (the two combined) were analysed. An unencapsulated derivative of TIGR4, TIGR4Δcps (Pearce et al., 2002) was used in all cases, to ensure that resulting bands did not result from capsular protein aggregation. R6, an unencapsulated derivative of D39 (Hoskins et al., 2001) that has been sequenced was used as a negative control since it does not contain SP1772 or any of the other genes in the locus. SK120, a S.gordonii strain, was used as a positive control, since it contains GspB, the glycosylated homologue of SP1772.

Cell extracts were prepared as described in 2.7.1 and analysed on a variety of SDS-PAGE gels as described in 2.7.2. The percentage of an SDS-PAGE gel is defined by the ratio of acrylamide to water. Lower percentage gels contain less acrylamide and therefore there is more space in the gel for large proteins to migrate through. Since SP1772 is predicted to be 412 kDa without glycosylation low percentage gels or gradient gels were used. Gradient gels contain a gradient of acrylamide through the gel and so smaller proteins can be resolved at the bottom of the gel while larger proteins can still be resolved further up the gel. SDS-PAGE was utilised to identify a band migrating above 250 kDa in TIGR4Δcps

Figure 4.1 – Sonicated cell extracts of bacteria, separated by SDS PAGE and blotted, and stained with different stains, to detect SP1772



Sonicated cell extracts of *S.pneumoniae, S.gordonii* and *S.aureus*. SH – *S.aureus* SH100, SK – *S.gordonii* SK120. TCP – Total cell protein, Pel – Pellet, Sup – Supernatant. Extracts were separated by SDS PAGE or blotted onto nitrocellulose and stained with Coomassie blue, Silver stain or glycoprotein stain.
Figure 4.1(a) is a 4-12% gradient gel, coomassie stained and run for 16 hours. There was not a band that could be detected above 250 kDa in the insoluble (cell wall associated, pel) or soluble (cytoplasmic, sup) fraction
Figure 4.1 (b) is a 4-12% gradient gel, glycoprotein stained and run for 12 hours. The detection limit of the stain appears to be very low, and there was not a band detected above 250 kDa in any of the samples.
Figure 4.1 (C) is a 4% gel, silver stained. Silver staining is a far more sensitive detection method for proteins. There were bands of above 250kDa present using this stain, however they were also present in R6, which does not contain the locus.
Figure 4.1 (D) is a dot blot, glycoprotein stained. Each strain studied had duplicate samples. The stain reacted with all strains, indicating the presence of a glycoprotein in all strains, including R6, which does not contain SP1772. Therefore we could not confirm the presence of SP1772 since the reaction with TIGR4 may have been due to a reaction with the same glycoprotein as is present in R6.

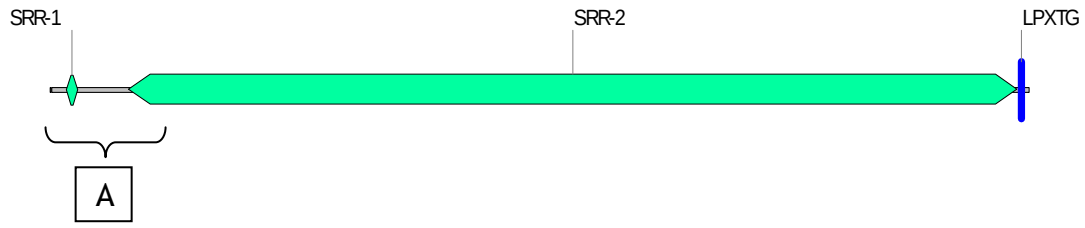
extracts that was absent in R6, since SP1772 is not present in R6. Figure 4.1a is an example of a 4-12% gradient gel stained with coommassie blue, run for 16 hours. There was not a band that fulfilled the criteria necessary to identify it as SP1772. The gel in figure 4.1b is an example of a gel stained with Glycoprotein detection kit (Sigma). This stain is more sensitive, and more specific, it will only react with glycoproteins. However, we could not identify a suitable band, even in cell extracts that had been concentrated using centrifugation filters (Amicon). There was a reaction with horse-radish peroxidase, the positive control for the stain, which shows the stain was working and so it appears that if glycol-proteins are present then they were not in high enough concentration to react with the stain. A highly sensitive silver stain (Silver Quest, Invitrogen) was then used on concentrated cell extracts (Figure 4.1c), on a 4% gel. This did not stain any high molecular weight bands that weren't also present in the negative control strain, R6. It was decided that if SP1772 couldn't be resolved on an SDS PAGE gel due to its size, whole cell extracts should be analysed by dot blot. A glycoprotein stain was used, and if SP1772 is the only glycosylated protein in TIGR4, then the stain would not react with R6. However, the stain did react with R6, showing that R6 contains glycoprotein. We therefore could not conclude that the glycoprotein strain was reacting with TIGR4 due to the presence of SP1772 (Figure 4.1d). Therefore it was decided that a more specific reagent to recognise SP1772 was required. An antibody to GspB had been used in previous experiments with *S.gordonii* (Bensing and Sullam, 2002). In this study, antiserum was raised against wild type S.gordonii. This antisera was then adsorbed using heat killed cells of a mutant deficient in GspB. The resulting mixture was then filtered for use. This antiserum was then able to detect GspB because the majority of the antibodies in the mixture were against GspB, the one antigen that was present in the strain used to create the antiserum and not present in the strain used to adsorb the antiserum. Since SP1772 and GspB are homologues, it was possible that the anti-GspB antibody would therefore react with TIGR4 whole cell extracts containing SP1772. There was, however, no reaction of the antibody with TIGR4. GspB and SP1772 may be too divergent for an antibody raised against GspB to cross-react with SP1772 and therefore attempts were made to raise an antibody by cloning, expressing and purifying a suitable antigen and using it to raise antisera in a mouse.

4.2.2 Raising an antibody against SP1772

In order to raise an antibody to SP1772 an attempt was made to clone the whole *SP1772* by PCR. This initial step was unsuccessful, and it was deemed that should the PCR work, it would be unlikely to be expressed by *E.coli*, which is the organism commonly used to express proteins. This is due to the size and number of repeats of SP1772. Additionally, studies in *S.gordonii* have shown that the glycosyl-transferases adjacent to GspB in the chromosome are required for the stability of GspB. In their absence GspB can no longer be detected on the surface of *S.gordonii* cells, and instead the protein is found to aggregate in the cytoplasm (Takamatsu *et al.*, 2004a). Whilst there are glycosyl-transferases present in *E.coli*, these are unlikely to be able to glycosylate pneumococcal proteins. It therefore seemed feasible to selected a part of *SP1772* to clone, express and purify.

There were several features of SP1772 that had to be taken into account in designing a suitable fragment of SP1772, against which to raise an antibody to SP1772. The features of SP1772 are outlined in Figure 4.2. SP1772 contains two serine-rich repeat regions. SRR-1 is a small repeat region near the N-terminal of the protein, and SRR-2 is a much larger domain, forming the bulk of the protein and spanning towards the C-terminal. There is an LPXTG motif downstream of SRR-2 which is predicted to allow SP1772 to be anchored to the cell surface by sortase. The N-terminal portion of the protein is predicted to contain a signal sequence which allows export to the cell surface from the cytoplasm through interacting with SecA2. It is unknown where this sequence falls in SP1772. It is unknown where the glycosylation sites on SP1772 are, however it is predicted that the repeat regions that are glycosylated, since truncates of GspB in the repeat region lead to a reduction in molecular weight of greater than predicted based upon the sequence (Bensing *et al.*, 2004). Given that the repeat regions are incredibly expensive for the cell to produce but form such a large part of the protein, it is likely that these are important in the function of the protein. Due to all these factors, several fragments of SP1772 were cloned (Figure 4.3), to incorporate these considerations. SP1772A contains the entire N-terminal, and no repeat region. SP1772B contains none of the N-terminal, and no repeat region. SP1772C contains the entire N-terminal, and one repeat from SRR-2. SP1772D contains none of the N-terminal, and one repeat from SRR-2.

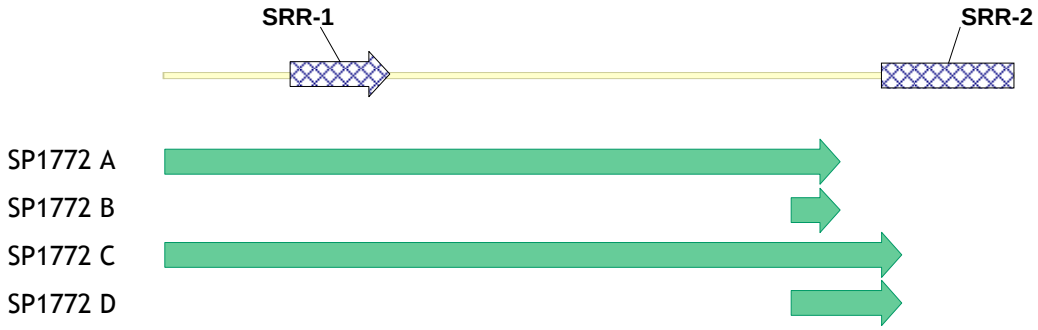
Figure 4.2 - Predicted features of SP1772



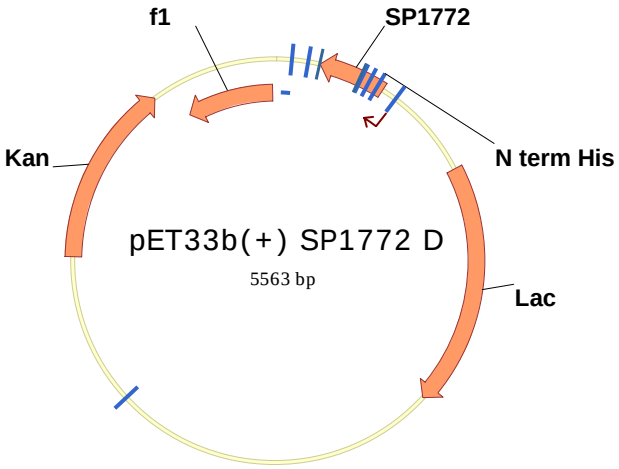
SP1772 protein predicted areas. SRR-1 and SRR-2 denote the serine rich repeat regions of the protein. The LPXTG denotes the motif recognised by sortase to attach proteins to the cell wall. It is therefore predicted that this end of the protein is attached to the cell surface and the sequence upstream of this is increasingly distant from the cell surface and therefore more exposed to the host environment. SP1772 will also contain a signal sequence allowing it to be recognised by SecA2 for export from the cytoplasm to the cell surface. This is predicted to be somewhere in A, most likely upstream of SRR-1.

Figure 4.3 Raising an antibody against SP1772

(a) Selection of fragments of SP1772 to clone, express and purify



(b) SP1772D in pET33b (2.8.1)



(a) Selection of fragments of SP1772 to clone in pET33b for expression and purification. (b) SP1772D was selected because it contained some of SRR-2, and very little of the N-terminal, which is predicted to be cleaved at an unknown site.

All four fragments were cloned into pET33b (Novagen) according to the manufacturer's instructions. See appendix II for a list of the primers used to amplify each of the fragments for cloning into pET. The proteins were expressed in BL21 cells in Terrific broth (TB). Cells were then sonicated and centrifuged, and the supernatant was collected and filtered. Each fragment was then purified using FPLC. SP1772B was found not to express well, whilst SP1772A and SP1772C did not purify as effectively as SP1772D. Therefore SP1772D was selected for further use.

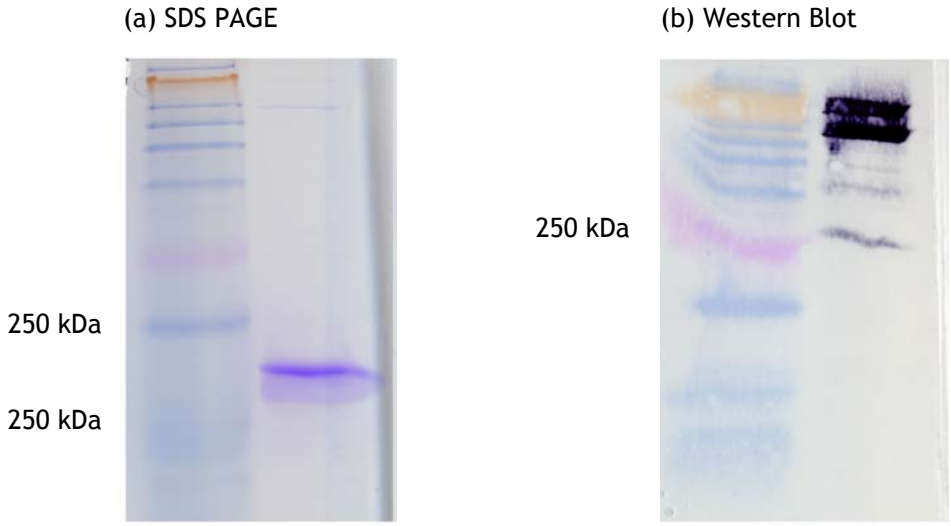
BL21 cells harbouring the pET33bSP1772D plasmid were then grown on a larger scale, and protein was harvested in the above manner. Samples were then purified using FPLC and dialysed.

The resulting protein from SP1772D cloning, expression and purification was analysed by mass-spec and determined to be the correct sequence, with no post-translational modifications, showing that the section chosen was not glycosylated in *E.coli.* The protein also migrated at the correct size on a 15% SDS PAGE gel, to correspond to its predicted weight of 10.5 kDa (Figure 4.4). The protein was then used to raise an antibody. The protein was injected subcutaneously in mice.

The antiserum was used in a western blot with purified SP1772D to determine if it would react with SP1772. The antisera did not react with SP1772D (Figure 4.4), there was a non-specific reaction with some minor contaminating bands. This may have been due to a lack of immunogenicity of SP1772D. This is potentially due to the sequence of the protein, although analysis using MIF bioinformatics: Predicting Antigenic Peptides (http://bio.dfci.harvard.edu/Tools/antigenic.pl) suggests that there are two antigenic regions in this peptide fragment (see Figure 4.5).

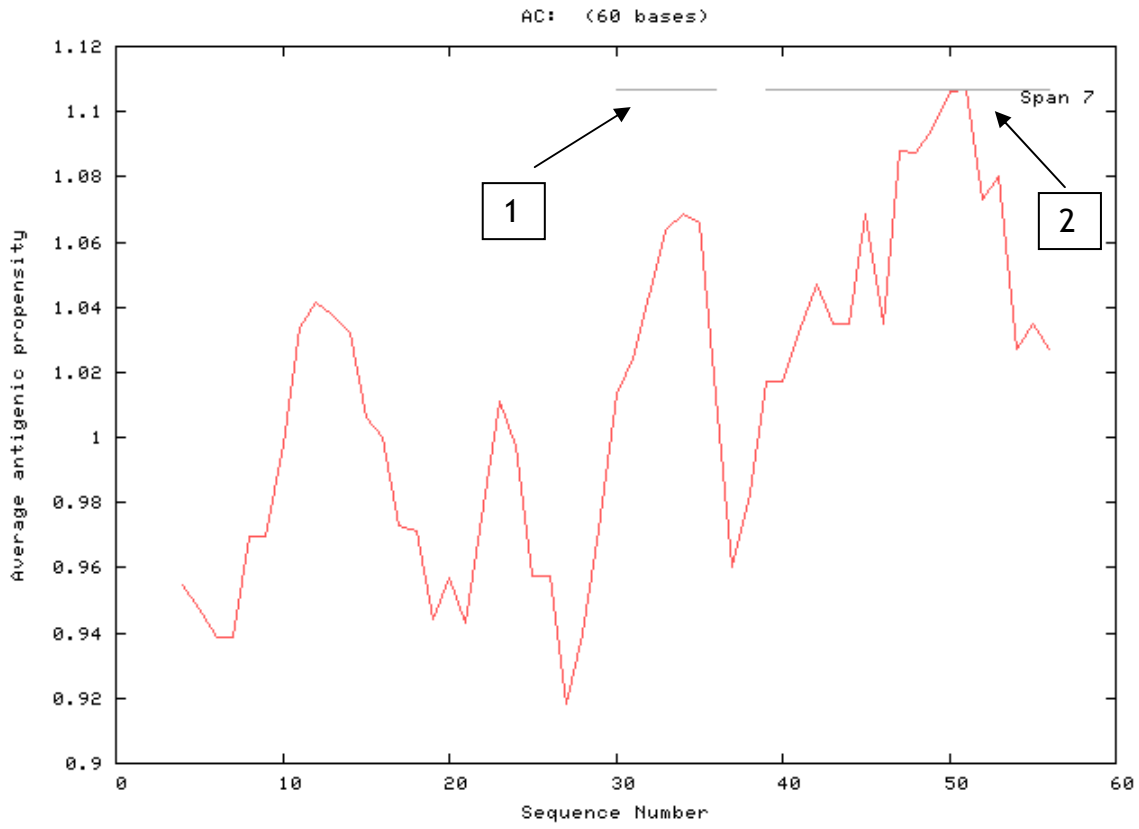
It appeared that this particular fragment of SP1772 was not sufficiently immunogenic alone to raise antibodies.

Figure 4.4 – Evaluation of Antisera raised against purified SP1772D



(a) SDS PAGE and (b) western blot of purified SP1772D using antisera raised against SP1772. SP1772 is predicted to be 10.5kDa. The serum did not react with the protein, however it did react with minor contaminating bands just visible at the top of the SDS-PAGE gel. This suggests the immunisation was successful, however antibodies weren't raised to SP1772D.

Figure 4.5 Immunogenicity of SP1772D



SP1772D sequence:

VPITGTDTSFTFTPYAARTDRIGINYFNGGGKVVESSTTSQSLSQSKSLSVSASQSASAS

SP1772D was analysed for immunogenicity using MIF bioinformatics: Predicting Antigenic Peptides (http://bio.dfci.harvard.edu/Tools/antigenic.pl). It is predicted that there are two antigenic regions in SP1772D. They are highlighted as 1 and 2 on the plot, and in red on the sequence of SP1772D below the plot.

4.2.3 Raising an antibody to SP1772 using technology fusing an adjuvant to SP1772D

Pneumolysin (ply) is a highly immunogenic virulence factor produced by *S.pneumoniae* (Feldman et al., 1990).The fusion of ply to other virulence factors of *S.pneumoniae* is predicted to make more efficient antisera against *S.pneumoniae* since ply is highly immunogenic, and the combination of two antigens from the ply will be more likely to elicit an immune response that would be effective against *S.pneumoniae*.

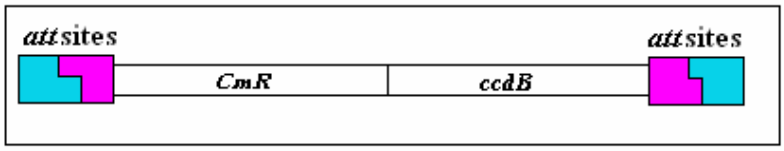
Unpublished studies from our laboratory have shown that green jellyfish protein (GFP) is not immunogenic alone, however, when fused to ply, it can elicit an immune response in mice. By combining SP1772D to ply, the same response might be seen, and antibodies to SP1772 could be derived. Therefore, in order to raise antisera to SP1772, Gateway cloning technology (Invitrogen) was utilised to fuse SP1772D to ply.

The Gateway technology is based on the bacteriophage lambda (λ) site-specific recombination system that facilitates the integration of bacteriophage λ into the chromosome of *E. coli*. Recombination occurs between specific attachment (att) sites on interacting DNA, which each possess a 15bp core region common to all, and differing flanking regions that provide binding sites for recombination proteins (Hartley *et al.*, 2000). There are four different att sites designated attB, attP, attL and attR and it is these sites which are utilised by the Gateway technology. All vectors used in Gateway cloning contain a gateway cassette. The cassette contains a chloramphenicol resistance gene (CmR) and control of cell death gene (ccdB), flanked by att sites (Bernard *et al.*, 1993).These are attP sites in the donor vector and attR sites in the destination vector. The presence of this cassette allows vectors with and without gene inserts to be distinguished as they only grow in DB3.1 E. coli cells which are resistant to the effects of the ccdB gene. The ‘BP’ and ‘LR’ recombination reactions constitute the Gateway technology and refer to the att sites being recombined in each. Figure 4.6 outlines Gateway technology.

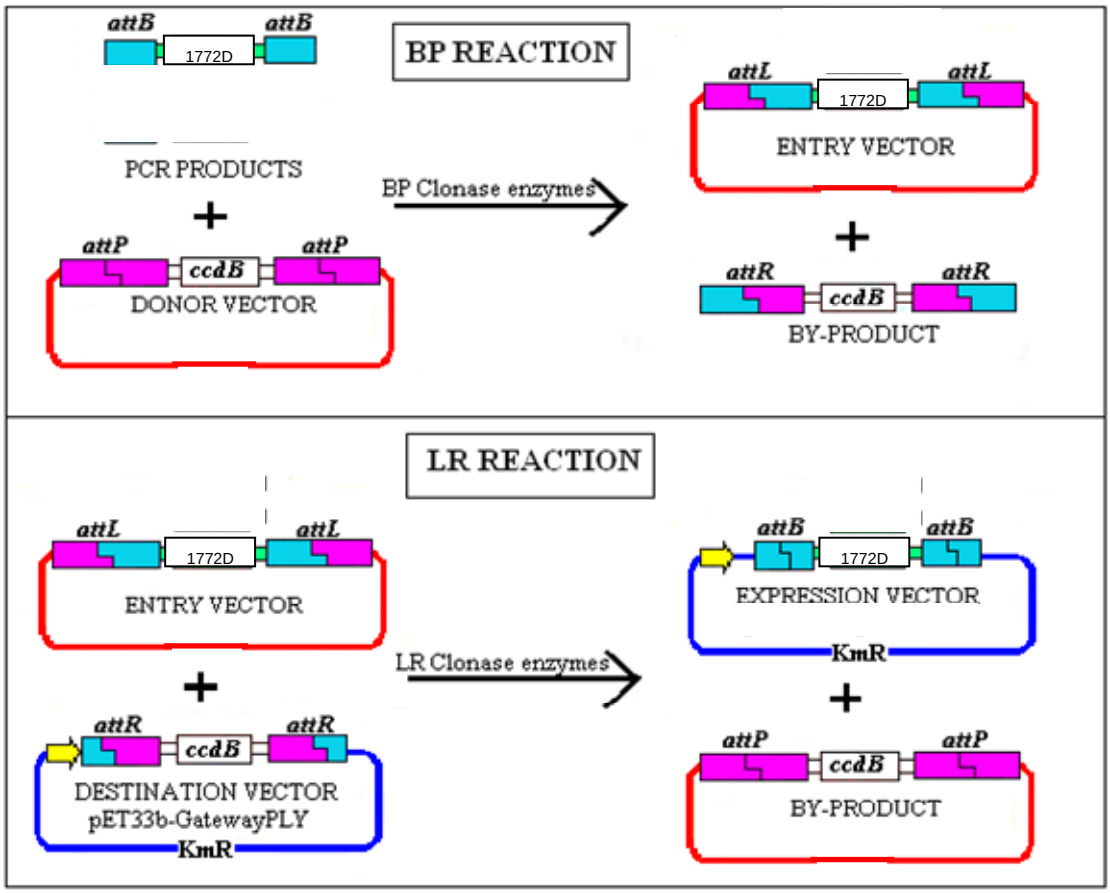
The BP reaction allows genes of interest with flanking attB sites tagged on via specifically designed PCR primers, to be inserted into a donor vector. These

Figure 4.6 – Gateway technology

(a) The gateway cassette



(b) The BP and LR gateway reactions



(a) Features of a Gateway cassette. The cassette contains a chloramphenicol resistance gene (CmR) and control of cell death gene (ccdB), flanked by att sites. These are attP sites in the donor vector and attR sites in the destination vector. (b) The BP and LR reactions constituting gateway technology. The att sites allow recombination between the gene of interest, entry vectors and destination vectors to give expression vectors. Figures adapted from Gateway Cloning technology instruction manual (Invitrogen)

donor vectors possess attP sites to facilitate recombination and, between these, a gateway cassette. During recombination, with the aid of a BP Clonase enzyme mixture (Invitrogen), the gene of interest replaces the gateway cassette. Resulting vectors, referred to as entry vectors, contain attL sites and can be used for the LR reaction.

The LR reaction allows recombination between the donor vector and a destination vector, which contains a gateway cassette and also a gene for the gene of interest to be fused to. The destination vector contains attR sites. During recombination, with the aid of a LR clonase enzyme mixture (Invitrogen), the gene of interest from the entry vector replaces the gateway cassette through the interaction of its attL sites and the destination vector's attR sites. This creates an expression vector containing the gene of interest fused to the gene.

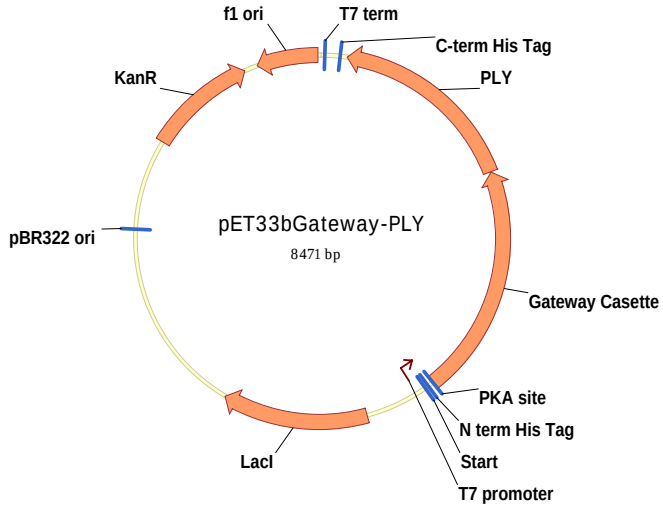
The BP and LR reactions can be carried out in one step, by the addition of all vectors and the gene of interest amplified to contain attB sites to the reaction mixture.

Primer 56L and primer 56M were used to amplify SP1772D from TIGR4 DNA. These primers were designed to contain the attB sites. An entry vector had previously been constructed. This vector, pET33bgatewayply (Figure 4.7) contained the pneumolysin gene fused to a gateway cassette. By amplifying SP1772D with appropriate primers, it could then be recombined into the pET33bgatewayply vector, to replace the gateway cassette giving pET33bSP1772Dply, the entry vector (Figure 4.7).

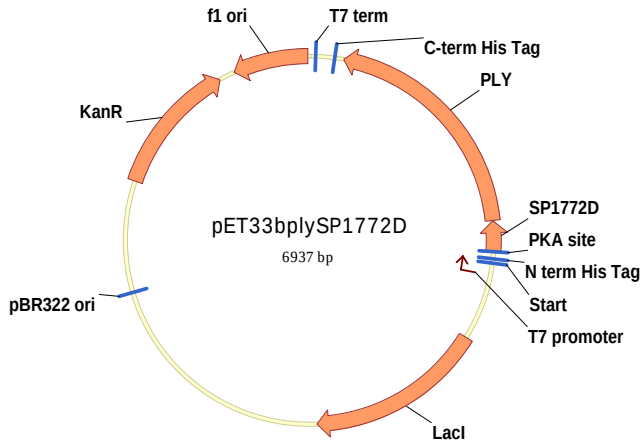
This vector was then used to express and purify the plySP1772D fusion protein. The protein was then injected into mice intranasally at a molarity of pneumolysin known to produce antiserum (Table 1). The administration of protein was repeated on four occasions. The antiserum was then used in a western blot, and it was shown that the antiserum did not cross react with SP1772D, the fragment that it was raised against in fusion with ply (Figure 4.8). However, the antibody did react with the SP1772Dply, and therefore the immunisation was successful, but SP1772D appeared not to be immunogenic.

Figure 4.7 – Entry and Expression Vectors for Gateway technology

(a) The entry vector

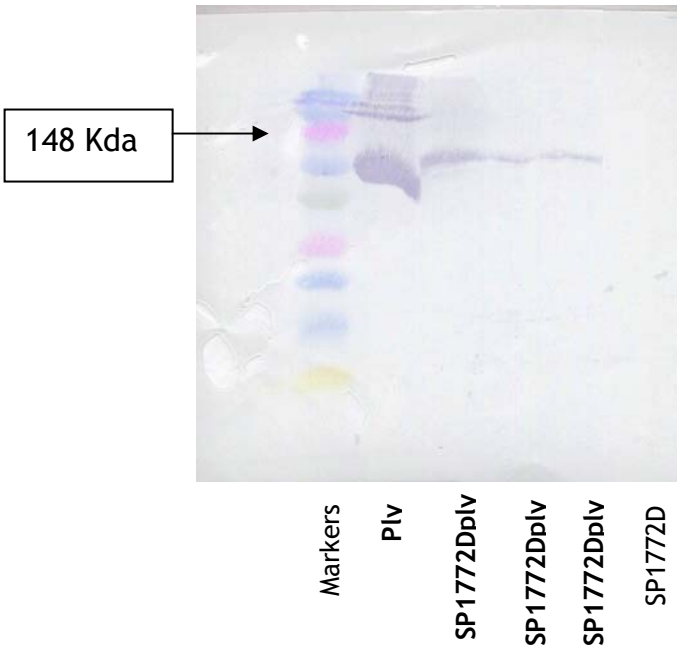


(b) The expression vector



(a) The entry vector utilised in gateway technology contains the gateway cassette and the pneumolysin gene. (b) Following BP and LR reactions, SP1772D is fused to ply in the expression clone.

Figure 4.8 – Evaluation of antisera raised against purified SP1772Dply



Western blot evaluating antisera raised against purified SP1772Dply. Samples where there was a reaction with the antibody are denoted in bold. The serum reacts with purified ply and SP1772Dply, but does not react with SP1772D.

Table 1 – Calculating the amount of protein for intranasal vaccination with SP1772DPly

Name of protein	Molecular weight of protein	Mass equivalent to 0.1 µg PLY	Concentration of purified protein (µg/µl)	Vol of prep containing mass of protein in 1 dose (µl)	Vol containing mass of protein in 100 µl for 5 doses (µl)
WT Ply (non-His tagged)	53006.41	0.1	0.417		
SP1772D	90532.73	0.170795815	0.454	0.38	1.88
Explanations	MW of proteins	(MW of protein/MW of Ply) x 0.1 ug		(Mass equivalent to 0.1 ug Ply/concentration)	Vol of prep x 5

4.2.4 Construction of a sortase mutant in the TIGR4 background

SP1772 contains the LPXTG motif, and so is predicted to be cell-wall anchored by sortase. A mutant was constructed to knockout the sortase gene, to analyse whether a difference could be seen in the cellular localisation of SP1772 i.e. that SP1772 would be seen in the soluble fraction. The *srtA* gene was deleted by allelic replacement, as described below.

SrtA (spr1098) and flanking DNA in the R6 genome was amplified using primers 19H and 19I and cloned into pCR4-TOPO (Invitrogen) according to the manufacturer's instructions. Primers 19W and 19X were then used to delete an internal fragment of *srtA* and introduce an AscI site between the 5' and 3' flanking sequences required for homologous recombination. Once confirmed by sequencing, the plasmid was digested using the AscI enzyme and an erythromycin cassette from pGhost5 (Biswas *et al.*, 1993) was ligated into the plasmid. This plasmid was then transformed into unencapsulated TIGR4 to give TIGR4Δ*cps*Δ*srtA*.

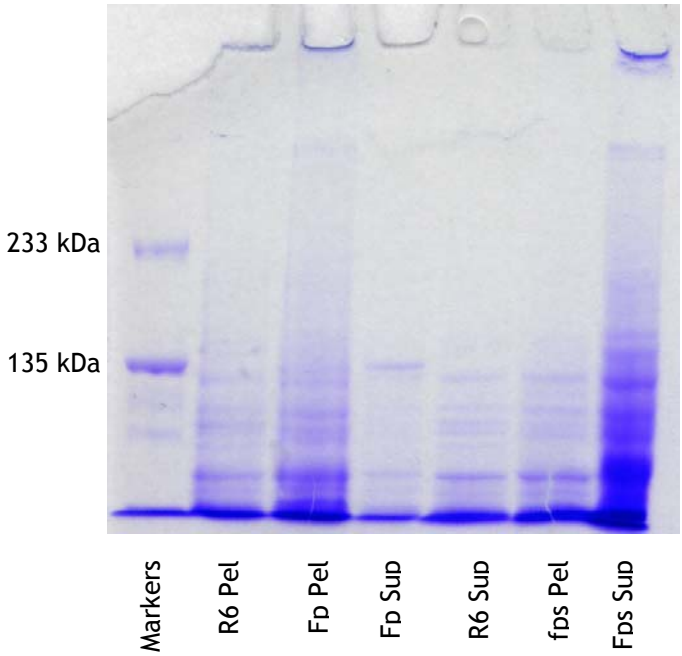
Evaluation of the sortase mutant showed that many proteins which are in insoluble fraction of fp23a (unencapsulated TIGR4) are in the soluble fraction of TIGR4ΔcpsΔsrt (Figure 4.9), however there was no band could conclusively be identified as SP1772, due to the limitations of the visualisation methods used. It would have been useful to repeat this analysis using a glycoprotein stain, since coomassie stains do not always visualise glycoproteins.

4.3 Discussion

The RNA from each of the genes in this locus, including the SP1772 gene is expressed, as demonstrated by the RT-PCR presented in the previous chapter. The work presented here was carried out in order to show that the full protein product of the SP1772 gene is expressed, however this was not possible. There are several reasons why the protein is not able to be detected by conventional methods. One is that the protein too large to be resolved by SDS-PAGE, it is predicted to be 412kDa without glycosylation, and SDS-PAGE is typically only useful for separating proteins up to 500kda in size. Additionally, since the SP1772 and associated genes are so expensive for the cell to make, the protein is

predicted to be found in very low copy number in the cell, and thus present in concentrations below the limit of detection. To overcome this, more sensitive stains were used, however a band that was present in TIGR4 derivatives and not R6 could not be identified. A more specific glycoprotein stain could also not identify SP1772 by dot blot analysis since R6 reacted with the stain and thus probably also contains glycoproteins. Incidentally, in work with GspB it was noted that conventional stains were not ideal in resolving GspB, including glycoprotein stains (Takamatsu *et al.*, 2004a). GspB also was not easily visualised using silver stain, the most successful methods used to identify GspB involved the use of a specific anti-GspB antibody (Bensing and Sullam, 2002). This antibody was raised by adsorbing antisera from a wild type strain against the

Figure 4.9 – Evaluation of TIGR4ΔcpsΔsrt



SDS PAGE analysis of TIGR4ΔcpsΔsrt (Fps). Cell extracts were diluted to contain the same amount of protein, determined by Bradford's assay; then separated on a 4% SDS PAGE gel. Many proteins that are present in the pellet of TIGRΔcps are present in the supernatant of TIGR4ΔcpsΔsrt. This was predicted to be the case, since sortase anchors many proteins with an LPXTG motif to the cell membrane, and thus in the absence of sortase, proteins are no longer cell wall associated. SP1772, however, could not be identified.

proteome of a GspB mutant strain. Given the homology between GspB and SP1772 it was anticipated that this antibody may react with SP1772 strains. However, given that the glycosylation of these two serine rich repeat proteins is probably different, it is quite plausible that there would be no cross-reactivity.

However, this was not the case. GspB is however, able to enter an SDS-PAGE gel and can be transferred by western blotting. As previously mentioned, the predicted size of SP1772 means that it is unlikely to even enter SDS-PAGE gels, and so dot blots of whole cell extracts were used with the antibody.

It was therefore attempted to raise a specific antibody to SP1772. Given the size and number of repeats in SP1772, it was not able to be amplified by PCR. It was also deemed that without the associated glycosyl-transferases in the TIGR4 chromosome, full SP1772 would not be stable in *E.coli*, as null mutants of GspB associated glycosyl-transferases result in either complete abolition of expression (GtfA and GtfB) or very limited expression (gly and nss) (Takamatsu *et al.*, 2004a). Therefore, selected parts of SP1772 were cloned, expressed and purified. A small fragment, entitled SP1772D encoding a portion of the N-terminal and one repeat region from SRR-2 was selected as it expression and purification was efficient. When used to raise antisera, no specific antiserum to this fragment was obtained, but antibodies were raised to minor contaminating bands from the protein purification, demonstrating that the immunisation had been successful. Analysis of the fragment for immunogenicity suggested it should be antigenic, however it was not.

A strategy was used to make SP1772D antigenic by gateway fusion to pneumolysin, a highly immunogenic pore-forming toxin encoded by *S.pneumoniae*. Previous unpublished studies had shown that green jelly fish protein, which is not immunogenic, could be made immunogenic by fusion to ply. SP1772D was therefore fused to ply and used to raise antisera. However, the resulting antiserum reacted with the SP1772D-ply fusion protein used to raise the antisera, but did not react with SP1772D alone. This strongly suggests that this portion of SP1772 is not immunogenic.

Further work that could be carried out would be to evaluate other sections of SP1772 to raise antibodies, or alternatively to utilise the method used by Bensing

and colleagues to raise an anti-GspB antibody, in which specific anti-GspB antisera was raised by adsorption.

One further strategy was utilised in an attempt to identify if SP1772 is expressed by TIGR4. A sortase mutant was constructed. Sortase has been shown to anchor proteins to the cell wall by their LPXTG motif. By constructing a strain of TIGR4 without sortase, it was hoped that SP1772 would be released from the cell wall of the pneumococcus, and a band could be identified in the soluble fraction of cell lysates in the mutant that was in the insoluble fraction of cell lysates in wild type TIGR4. Due to the inability of SP1772 to be detected by western blot or SDS-PAGE, no band could be identified as SP1772, although the mutant was clearly successful in that a great deal of the proteome that was detectable in the insoluble fraction of TIGR4 was now detectable in the soluble fraction of the mutant strain. Further work with this mutant strain could entail concentrating the soluble fraction with a high molecular weight cut-off filter, and mass-spectometry could be carried out to identifiy if SP1772 is present. The antibodies raised against SP1772D and SP1772Dply could also be evaluated using this strain, however given that they do not react with the proteins against which they were raised, it is unlikely that a reaction would occur.

In order to determine if the protein is expressed in the cell, a strategy to determine the function of SP1772 in TIGR4 was devised by examining the phenotype of an SP1772 null mutant. This work is outlined in the following results chapter.

Chapter 5

Results

Construction and analysis of SP1772 null mutants

5 Construction and Analysis of *SP1772* null Mutants

5.1 Introduction

In a chapter 3, it has been shown that *SP1772* is present in many, but not all, pneumococcal strains and *SP1772* RNA was shown to be expressed. Conventional methods were not able to determine if TIGR4 expresses *SP1772*, due to the highly unusual nature of the native gene product. We therefore sought to determine the function of SP1772 by constructing and examining the phenotypes of mutants of TIGR4 where the gene encoding SP1772 was knocked out.

The function of SraP, a SRR protein in *Staphylococcus aureus* was determined by disrupting the gene by allelic replacement in its native strain. A mutant of *S.aureus* in this gene was compared to its parent strain for its ability to bind platelets, and it was found that it was deficient in its ability to do this. The mutant was also assessed in a rabbit model of endocarditis, and was found in significantly lower levels (7-fold) than its parent strain (Siboo *et al.*, 2005).

Many methods have been used to generate mutants in *S.pneumoniae*. The method selected for use in this instance was that used by Throup and colleagues to study the role of two-component signal transduction systems in the virulence of *S.pneumoniae* (Throup *et al.*, 2000). In this method cross-over PCR is used to generate an insert for a plasmid which contains an antibiotic resistance cassette flanked by DNA which flanks the target gene in *S.pneumoniae*. Cross-over PCR is carried out by amplifying numerous fragments which are to be annealed to one another. Each fragment is amplified using primers with some sequence from the desired adjacent fragments in the final product. A PCR utilising the final 5' and 3' primers and each fragment will then anneal all of the desired fragments to one another. The plasmid is then transformed into *S.pneumoniae* and transformants, selected on the basis of antibiotic resistance, should contain an antibiotic resistance cassette in place of the target gene. By careful design of the transforming plasmid, the gene can be cleanly replaced with an antibiotic resistance cassette, avoiding downstream polar affects. The absence of downstream polar affects can be assessed by RT-PCR of adjacent genes.

5.2 Results

5.2.1 Constructing SP1772 deletion mutants in the TIGR4 background

5.2.1.1 Construction

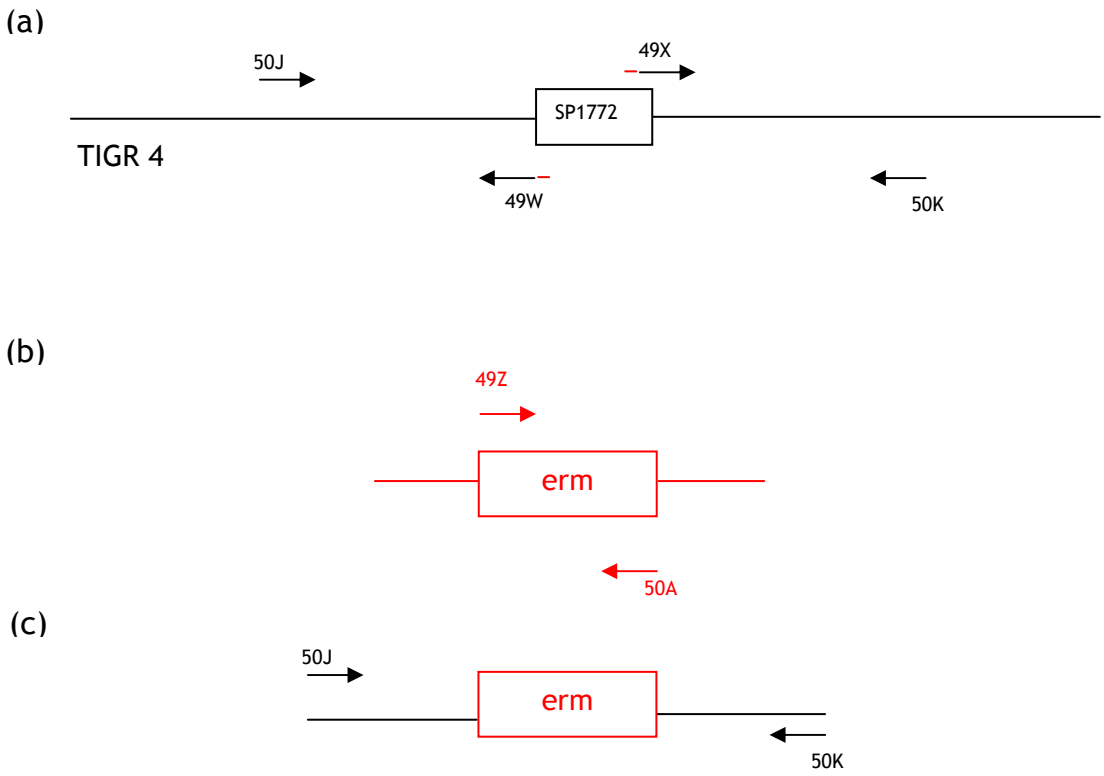
In order to delete *SP1772* in TIGR4, a method of allelic replacement was used to replace *SP1772* with an erythromycin resistance cassette (erm). This was based on a method developed to create a library of mutants deficient in two-component signal transduction systems (Throup *et al.*, 2000).

Primers 50J, 50K, 49W and 49X were designed to amplify approximately 1kb up and downstream of *SP1772* in the TIGR4 genome. The primers 49W and 49X (Figure 5.1) were designed to also contain regions homologous to an erythromycin cassette (Throup *et al.*, 2000). PCRs were carried out to amplify up the left and right flank of SP1772, and then a crossover PCR was carried out using primers 50J and 50K, and the erythromycin cassette, and left and right flanking DNA of SP1772 as template DNA (Figure 5.2). This construct was ligated into Topo TA (Invitrogen), and the plasmid was transformed into *E.coli* Top10 cells. The plasmid was sequenced and then used to transform an unencapsulated derivative of *S.pneumoniae* strain TIGR4. The genomic DNA of resulting transformants was checked by PCR and sequencing and then used to transform TIGR4.

5.2.1.2Confirmation of SP1772 deletion mutants

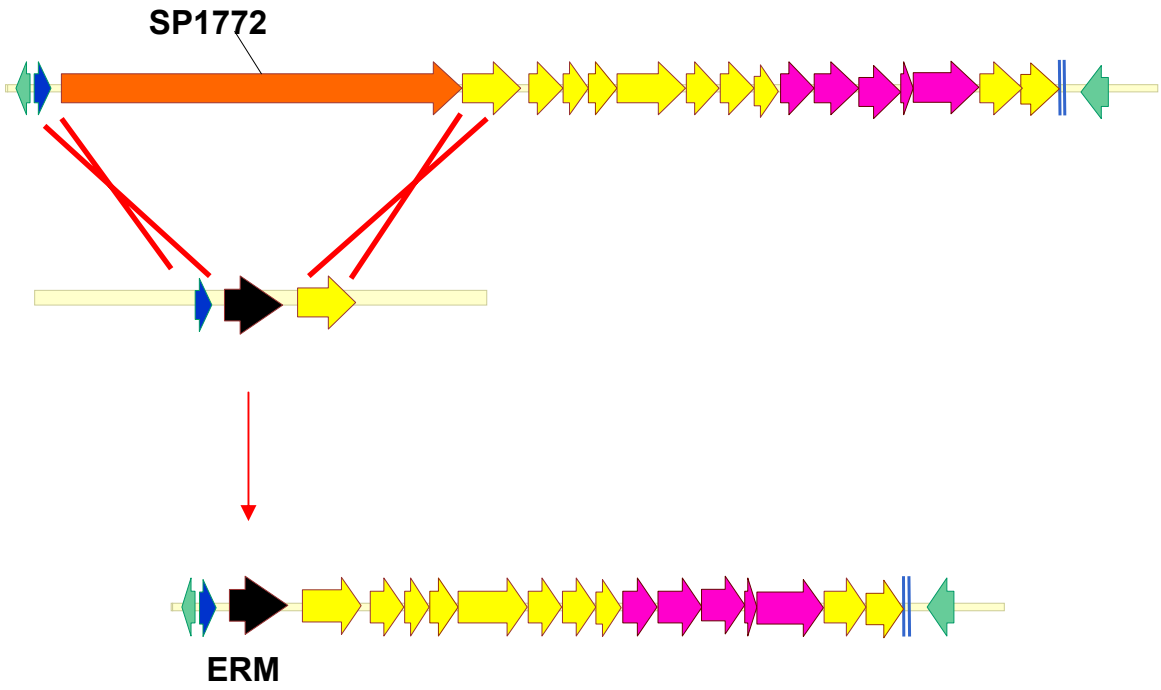
Mutants were selected by the phenotype of erythromycin resistance. The homologous recombination replacement of *SP1772* with the erythromycin cassette was confirmed by several PCRs. Primers 50X and 50Y were used to confirm the absence of *SP1772*, and primers were used to confirm the presence of genes adjacent to SP1772 in the chromosome that were not knocked out (51B and 51C – SP1771; 50V and 50W – SP1773). Specific primers were also used to

Figure 5.1 – Crossover strategy to replace SP1772 in the TIGR4 chromosome



(a) Primers are designed 1kb up and downstream of SP1772 in TIGR4 (50J and 50K). Primers 49W and 49X are designed to contain the sequence just up and downstream of SP1772, and also contain DNA homologous to the erythromycin cassette (marked in red). (b) Primers 49Z and 50A are used to amplify the erythromycin cassette from Topo erm. (c) Left flank and Right flank of PCR 1 and the erythromycin (erm) cassette are added to primers 50J and 50K. Tails on the LF and RF will anneal to the ery cassette and primers 50J and 50K will read through, making a PCR product that contains flanking DNA of SP1772 with the erm cassette in place of SP1772. This construct can then be used to transform TIGR4 to replace the erm cassette.

Figure 5.2 – Strategy to create TIGR4ΔSP1772



The strategy to knockout SP1772 from the TIGR4 chromosome was devised based on replacing SP1772 with an erythromycin resistance cassette by homologous recombination (Throup *et al.*, 2000). A plasmid is created by crossover PCR that contains the erythromycin cassette flanked by SP1772 flanking DNA. SP1772 is then replaced by homologus recombination between the plasmid and the TIGR4 genome.

verify that the erythromycin cassette had integrated into the chromosome of TIGR4 in the correct location, and that surrounding DNA had not been disrupted. Primers 49Z and 50A together amplify the erythromycin cassette and so primer 50A was used with primer 50X (SP1773 end) and primer 49Z was used with primer 51A (SP1771 start) to confirm the correct insertion. All primers are listed in (appendix II) and all PCRs are presented in Figure 5.3.

5.2.2 Characterisation of SP1772 deletion mutants

5.2.2.1*In vitro* growth Curves

In order to establish if the absence of SP1772 had an effect on the growth of TIGR4, the growth of the mutant was compared to WT. If the mutant strain did have aberrant growth from the wildtype, this has to be taken into account when looking at other phenotypes of the strain. Growth curves were carried out in BHI and by viable counting and it was confirmed that the deletion of *SP1772* had no effect on the growth of TIGR4. BHI is a rich medium however, and to confirm the absence of *SP1772* had no effect on growth of the strain, it would be desirable to also carry out a growth curve in a nutrient poor medium.

5.2.3 Investigating the role of SP1772 in biofilm formation

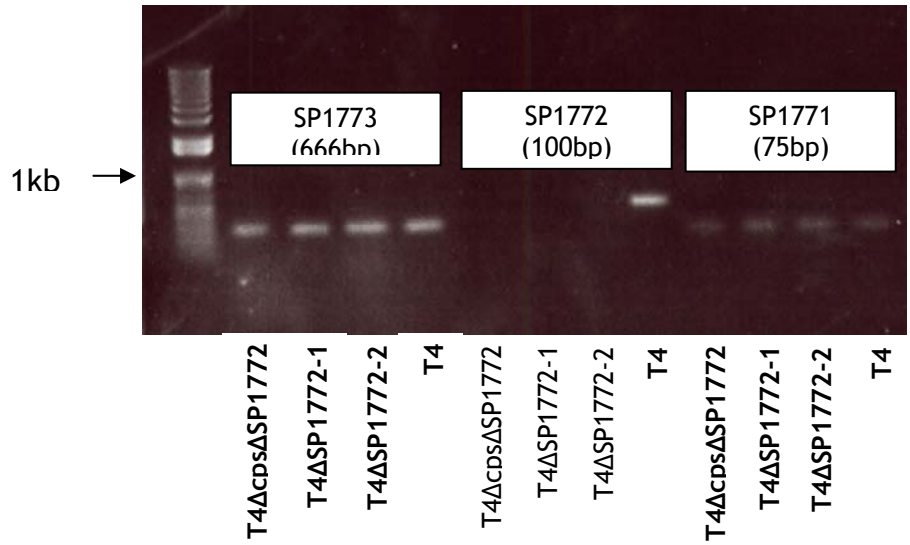
S.pneumoniae have been shown to produce biofilms *in vitro*. It has been noted that not only are many genes upregulated in a biofilm model of infection, but also that the number differing detectable proteins also differs significantly (Allegrucci *et al.*, 2006; Moscoso *et al.*, 2006; Oggioni *et al.*, 2006).

Homologues of SP1772 have been shown to be associated with biofilm formation. *S.parasanguis* contains a serine rich repeat protein, Fap1, that is required for biofilm formation (Froeliger and Fives-Taylor, 2001; Quin *et al.*, 2007), essential in the ability of the organism to contribute to dental plaque formation. *S.cristatus,* also involved in plaque formation, also conatins a serine-rich repeat protein, SrpA which is involved in biofilm formation (P. S. Handley, 2005).

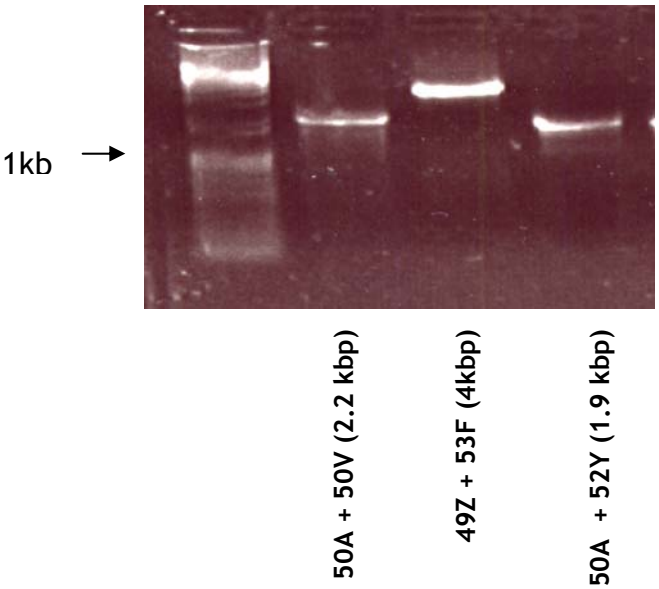
SP1772 was therefore evaluated in a biofilm model, the discontinuous-culture model of biofilm growth developed by Moscoso and colleagues for *S.pneumoniae*

Figure 5.3 – Confirming TIGR4ΔcpsΔSP1772

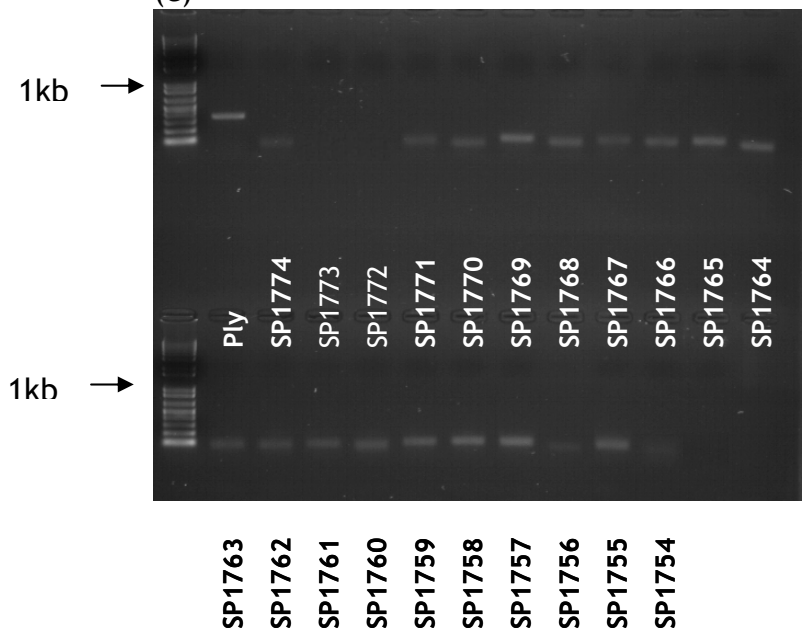
(a) Presence and absence of genes



(b) Localisation and orientation of erythromycin cassette in T4ΔcpsΔSP1772



(c)



(a) Mutants were confirmed by PCRs looking for the presence and absence of SP1772 and surrounding genes. Present genes are denoted in bold. (b) The localisation and orientation of the erythromycin cassette in the TIGR4 genome was confirmed by PCRs using primers internal to the erythromycin cassette and external to the area of recombination in the TIGR4 genome. (c) RT-PCR was carried out on mutants. Present genes are denoted in bold.

(Moscoso et al., 2006). This assay utilises the ability of the bacteria to form biofilms of polystyrene plates.

Strains are grown in media and then transferred to a 96-well plate. Plates are then incubated to allow growth and biofilm formation is determined by staining with 1% crystal violet at room temperature, solubilizing the biofilm with ethanol and determining the A595.

Strains evaluated in this model were: T4, T4Δcps, T4ΔSP1772, T4ΔCPSΔSP1772 and R6. All of the results are presented in Figure 5.4. Figure 5.5 presents the data anaylsed in pairs of strains.

In the absence of capsule, compared to the wild type, the ability of TIGR4 to form biofilms is slightly reduced (Figure 5.5a). This is in disagreement with previous studies where capsule has been shown not to be required in a variety of pneumococcal strains, including TIGR4 and D39 (Oggioni *et al.*, 2006) (Moscoso *et al.*, 2006). However it has also been noted that capsular expression is the same across sessile and planktonic growth models(Oggioni *et al.*, 2006) and that some capsular components expression is upregulated in biofilm models of growth (Allegrucci *et al.*, 2006). It is worth noting that the observation regarding TIGR4 was made in a different biofilm model and the observations of requirement of the capsule in this model was not in TIGR4 (Moscoso *et al.*, 2006).

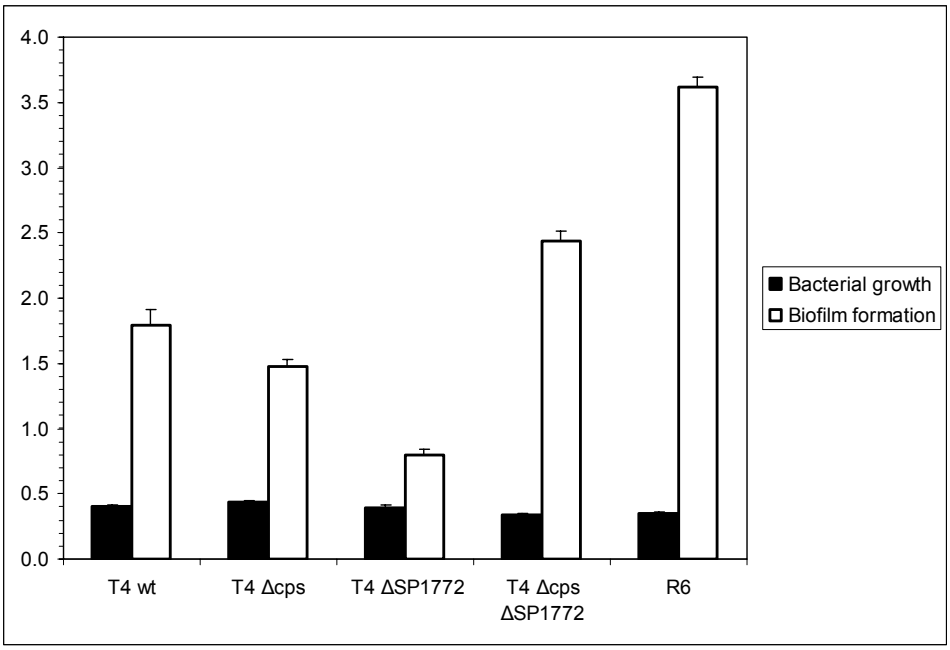
In the absence of SP1772, the ability of TIGR4 to form biofilms is greatly reduced (Figure 5.5b). However, comparison of strains reveals that a mutant strain lacking both capsule and SP1772, TIGR4ΔcpsΔSP1772, is much more able to produce biofilms than TIGR4, TIGR4Δcps or TIG4ΔcpsΔSP1772 (Figure 5.5c). R6, an unencapsulated strain that does not contain SP1772, is better able to form biofilms than any of the TIGR4 derivatives, including TIGR4ΔcpsΔSP1772 (Figure 5.4).

5.2.4 In vivo characterisation of SP1772 deletion mutants

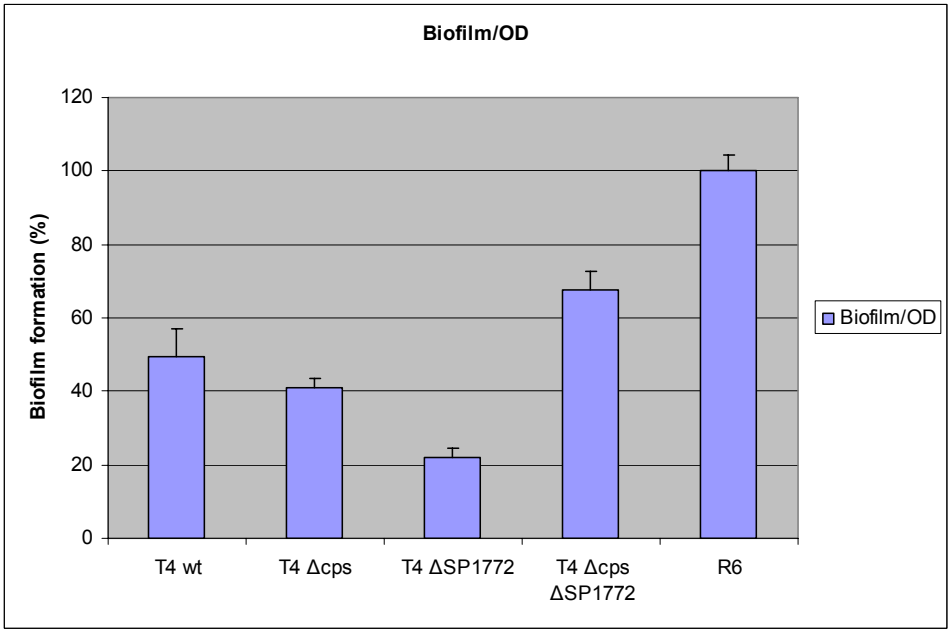
Studies were carried out to determine if SP1772 plays a role in the virulence of *S.pnuemoniae*. A signature tagged mutagenesis screen of TIGR4 had previously demonstrated that in a competitive model of pneumonia infection in mice, SP1772 had a role in virulence (Hava and Camilli, 2002). We therefore sought to

Figure 5.4 – Evaluating *SP1772* knockouts in a biofilm model

(a) Biofilm growth and bacterial growth in a biofilm model



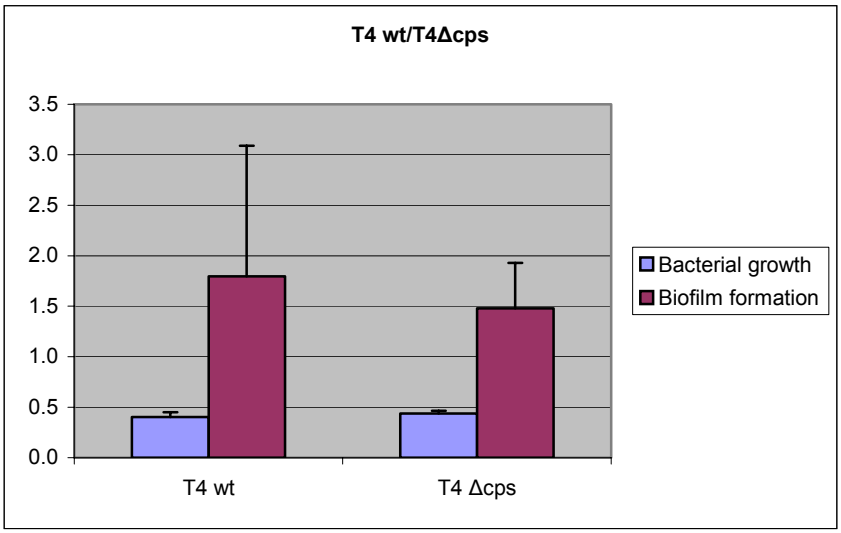
(b) Biolfilm formation



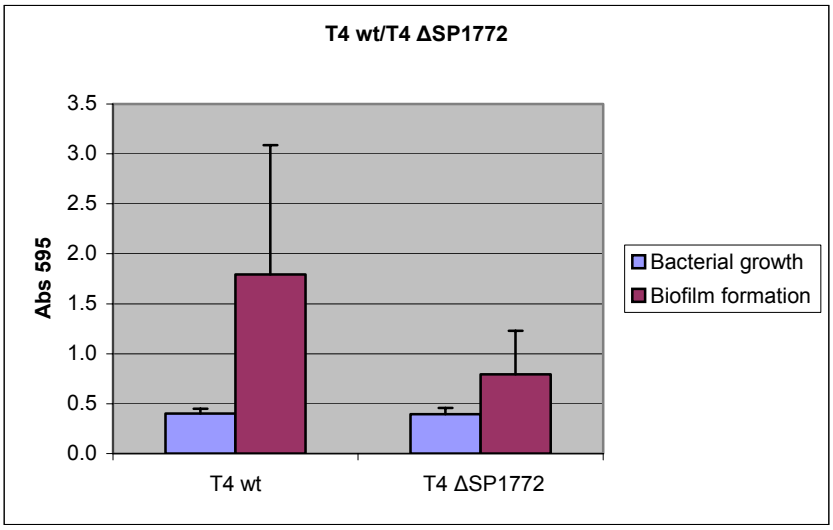
Summarized data of bacterial growth and biofilm formation of *S.pneumoniae* strains and mutants in a biofilm model. Strains studied were TIGR4, TIGR4Δcps, TIGR4ΔSP1772, TIGR4ΔcpsΔSP1772 and R6.
(a) The bacterial growth and biofilm formation of each strain studied. Filled and open bars indicate growth and biofilm formation, respectively.
(b) The ability for each strain to form biofilms is illustrated. The values have been normalized for absorbance, and the percentages calculated in relation to R6.Direct strain comparisons are presented in Figure 5.5.

Figure 5.5 – Evaluating *SP1772* knockouts in a biofilm model – comparison between strains

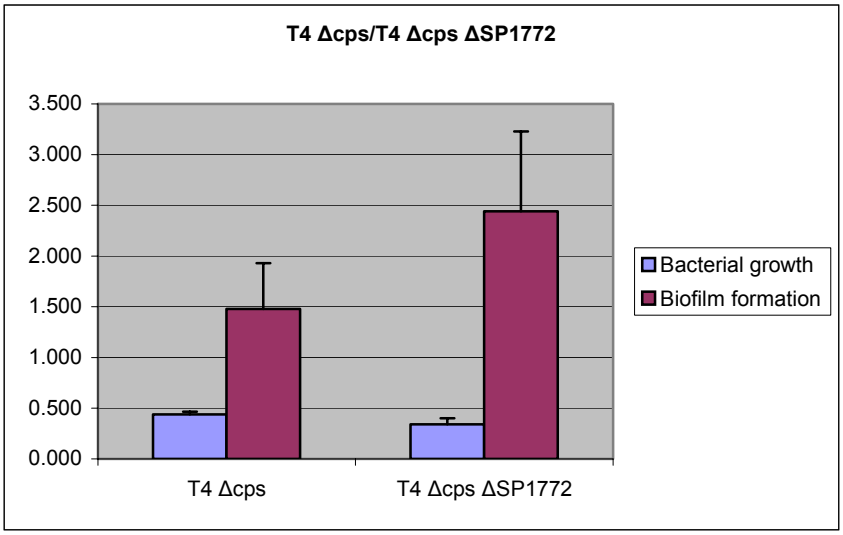
(a) TIGR4 and TIGR4Δcps



(b) TIGR4 and TIGR4ΔSP1772



(b) TIGR4 and TIGR4ΔcpsΔSP1772



Comparison in pairs of biofilm generated in TIGR4 and mutants in the capsule region and SP1772.
(a) Bacterial growth and biofilm formation of TIGR4 and TIGR4Δcps. TIGR4 is less able to form biofilms in the absence of capsule.
(b) Bacterial growth and biofilm formation of TIGR4 and TIGR4ΔSP1772. TIGR4 is less able to form biofilms in the absence of SP1772.
(c) Bacterial growth and biofilm formation of TIGR4Δcps and TIGR4ΔSP1772. In the absence of both capsule and SP1772, the TIGR4 is able to form more biofilm than either wt or TIGR4Δcps

compare the virulence of TIGR4 with TIGR4ΔSP1772 to see if the latter was attenuated in various single strain models of infection in mice. The relationship between capsule and SP1772 could not be evaluated in *in vivo* models, since unencapsulated strains of TIGR4 are unable to cause disease in animal models.

5.2.4.1 Intraperitoneal infection and generation of standard Innocula

The mutant, TIGR4ΔSP1772, and the parent wild type strain, TIGR4, were given intraperitoneally (i/p) to generate standard innocula for future experiments and to assess any difference in virulence by this route of infection. 100µl of 10^7 CFU/ml (colony forming units/ml) was injected into 5 MF-1 (outbred) mice for each strain. After 20 h mouse number one from each strain was culled and bled to prepare standard innocula. The rest were tail bled at 24 h to measure bacterial levels in the blood and monitored for disease progression and survival. The following blood counts were calculated, and showed no difference by this route of infection: Mean+/- SEM TIGR4 5.47 +/- 0.44 $\log_{10}$ CFU/ml TIGR4ΔSP1772 5.58 +/- 0.37 $\log_{10}$ CFU/ml. There was no difference in survivalof animals treated with the different strains. These data are presented in Figure 5.6.

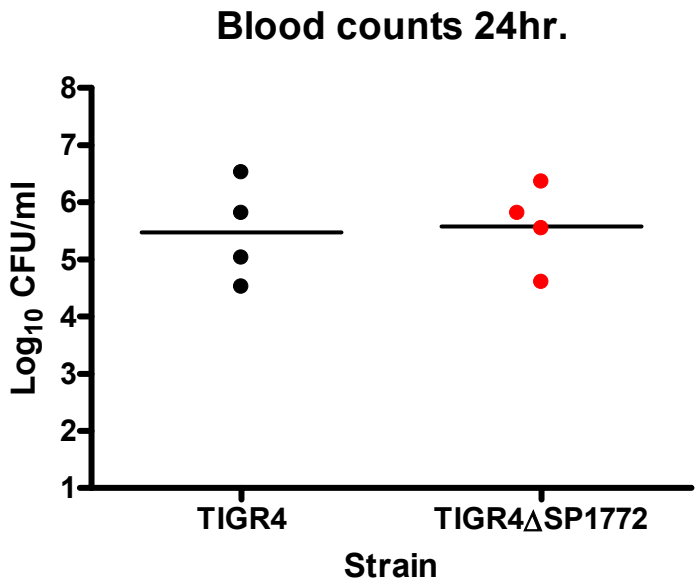
5.2.4.2 Investigating the role of SP1772 in a pneumonia model of infection

TIGR4ΔSP1772 and TIGR4 were then given intranasally (i/n) to assess any difference in virulence in a pneumonia model of infection. 50 µl of 10^6 CFU/ml was injected into the nares of 5 anaesatized MF-1 (outbred) mice for each strain. Bacterial counts were taken at 24 h from the lungs and blood. Survival was then monitored. The following blood counts were calculated, and showed no difference by this route of infection: Mean+/- SEM TIGR4 6.10 +/- 1.81 $\log_{10}$ CFU/ml; TIGR4ΔSP1772 5.40 +/- 2.06 log10 CFU/ml. There was no difference in survival either. These data are presented in Figure 5.7.

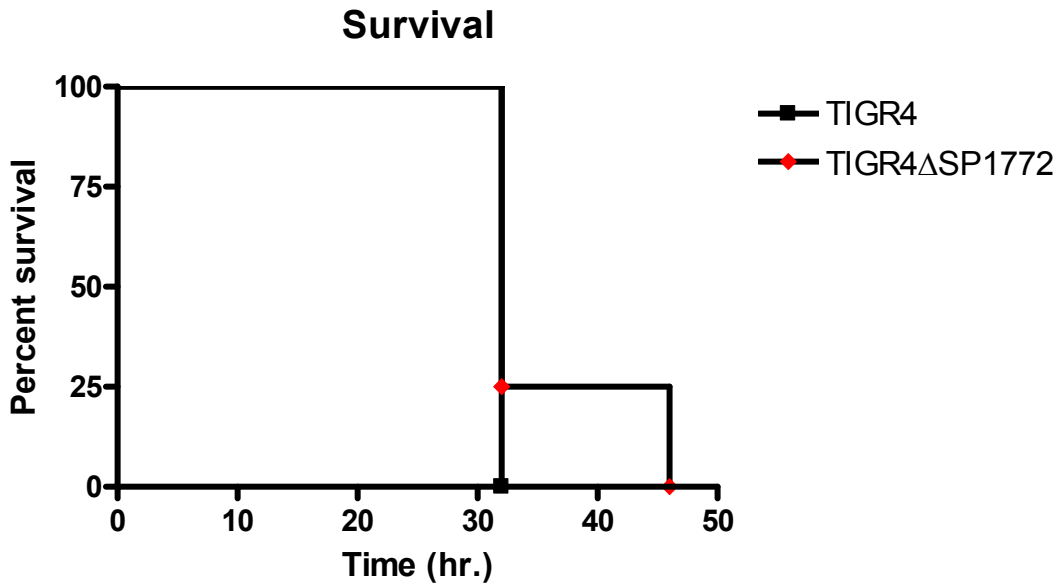
Data were then published by Obert *et al* in which a mutant in the TIGR4 background of SP1772 was generated by insertion-duplication (Obert *et al.*, 2006). When evaluated in a pneumonia model of infection, it was found that SP1772 was required for the full virulence of TIGR4 in mice. The experiments

Figure 5.6 – Intraperitoneal passage of TIGR4 and TIGR4ΔSP1772

(a) Blood counts



(b) Survival



(a) Log of counts of bacterial in the blood of MF-1 24 h post infection intraperitoneally with 10^6 CFU TIGR4 or TIGR4ΔSP1772. Each animal is represented by a dot, with the mean represented by the horizontal line. (b) Survival curve from the experiment described above.

had been done using Balb/c (inbred) mice, using 10^7 CFU/ml in a 50 µl volume. We therefore sought to repeat this model of infection, to determine if we could see an attenuation in virulence with our mutant by this model. We also repeated our model of infection for pneumonia in parallel. Survival data and blood counts are presented in Figure 5.8 (MF-1) and Figure 5.9 (Balb/c). In both strains of mice no difference in virulence was seen between TIGR4 and TIGR4ΔSP1772.

5.2.4.3 Investigating the role of SP1772 in a colonisation model of infection

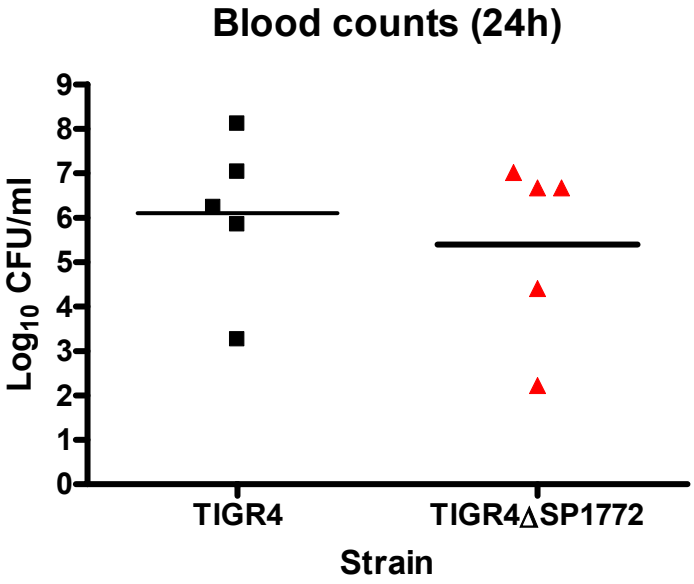
S.pneumoniae is normally found in a commensal state in the human host. It will not, in most instances of human contact, cause disease. However when it does cause disease, the bacteria has usually colonised the human host prior to this. Given that we had not seen a role for SP1772 in invasive disease by our models, we sought to investigate whether SP1772 has a role in colonisation. TIGR4ΔSP1772 and TIGR4 were given intranasally (i/n) to assess any difference in virulence in a colonisation model of infection. 10 µl of 10^6 CFU/ml was injected into the nares of 5 anaesatised MF-1 (outbred) mice for each strain. Nasopharyngeal washes (2 ml) and nospahryngeal tissue counts were taken at day 5. The following counts were calculated, and showed a difference in colonisation after 5 days, however this wasn't significant. These data are presented in Figure 5.10.

5.3 Discussion

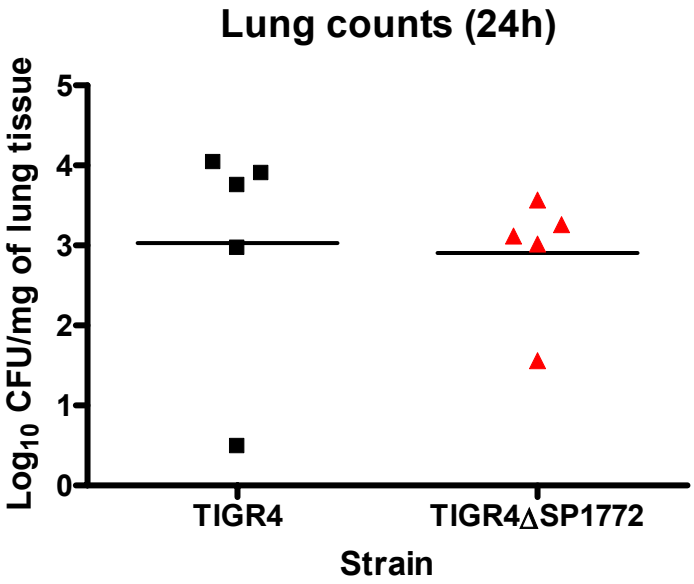
The work presented in this chapter shows SP1772 has an unusual role in biofilm formation, that is linked to the presence of capsule. Several studies have shown that capsule is not required for biofilm formation *in vitro* by examining the behaviour of isogenic mutants with the capsule knocked out (Donlan *et al.*, 2004; Moscoso *et al.*, 2006). However, other studies have noted that capsular expression is not down-regulated in a biofilm model (Oggioni *et al.*, 2006) and also that proteins involved in the synthesis of capsule are upregulated (Allegrucci *et al.*, 2006). These observations taken together suggest that whilst capsule is not required for biofilms to form, that in its presence, biofilm formation is not inhibited. However, Moscoso and colleagues suggest that in fact

Figure 5.7 – Pneumonia intranasal infection model of TIGR4 and TIGR4ΔSP1772

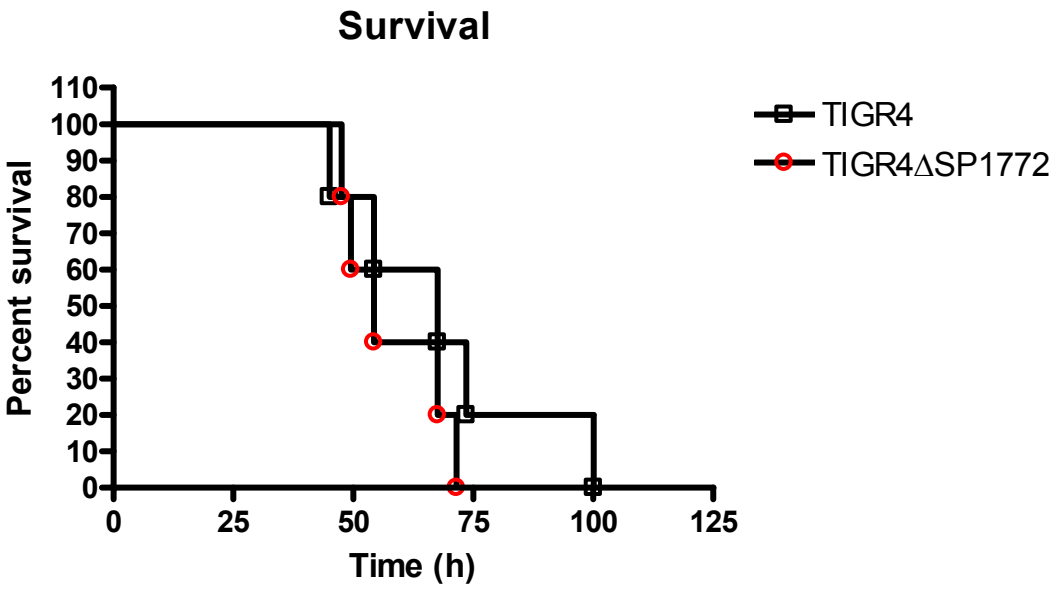
(a) Blood counts



(b) Lung counts



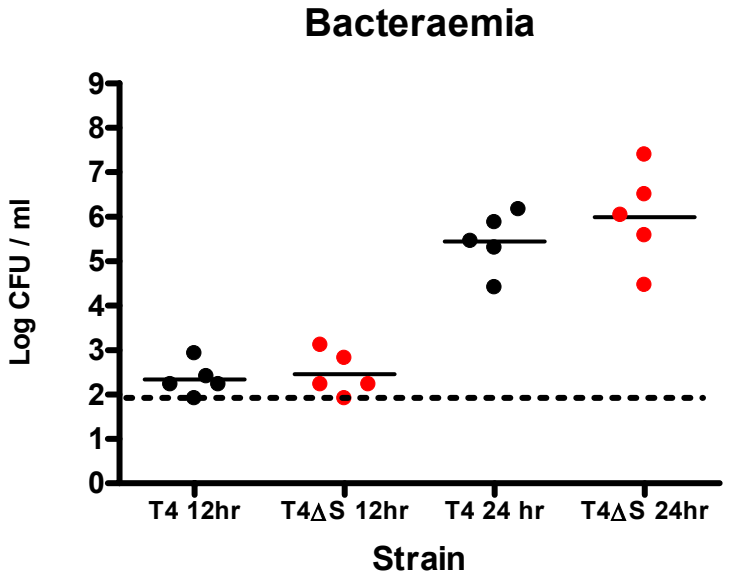
(c) Survival



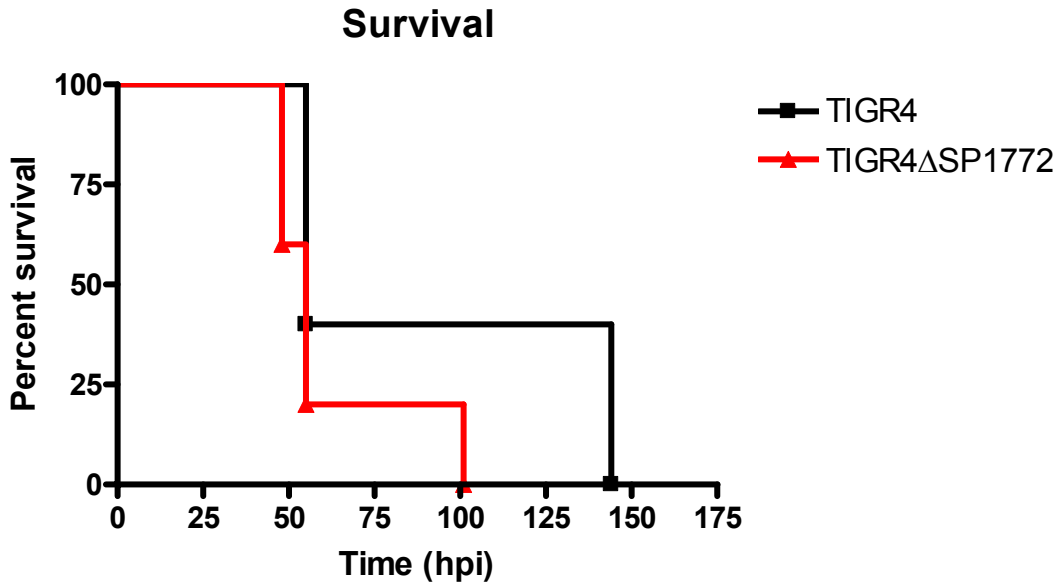
(a) Log of counts of bacteria in the blood of MF-1 24 h post infection intransally with 10^6 CFU/ml TIGR4 or TIGR4ΔSP1772. Each animal is represented by a symbol, with the mean represented by the horizontal line. (b) Log of counts of bacteria in the lungs of MF-1 mice from the experiment described above. (c) Survival curve from the experiment described above.

Figure 5.8 – MF-1 (CB) Pneumonia infection model of TIGR4 and TIGR4ΔSP1772

(a) Blood counts



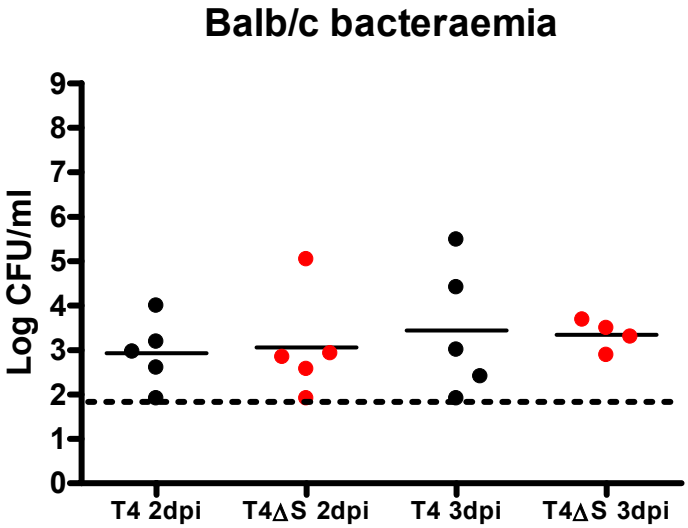
(b) Survival



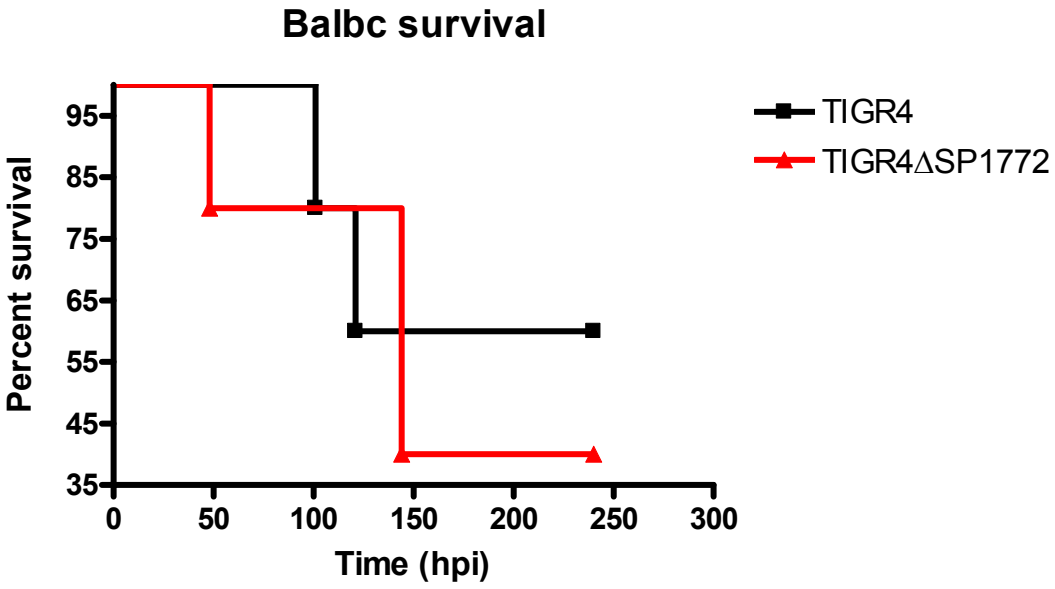
(a) Log of counts of bacteria in the blood of MF-1 12 h and 24 h post infection intransally with 10^6 CFU TIGR4 or TIGR4ΔSP1772. Each animal is represented by a symbol, with the mean represented by the horizontal line. (b) Survival curve from the experiment described above.

Figure 5.9 – Balb/c (CB) Pneumonia infection model of TIGR4 and TIGR4ΔSP1772

(a) Blood counts



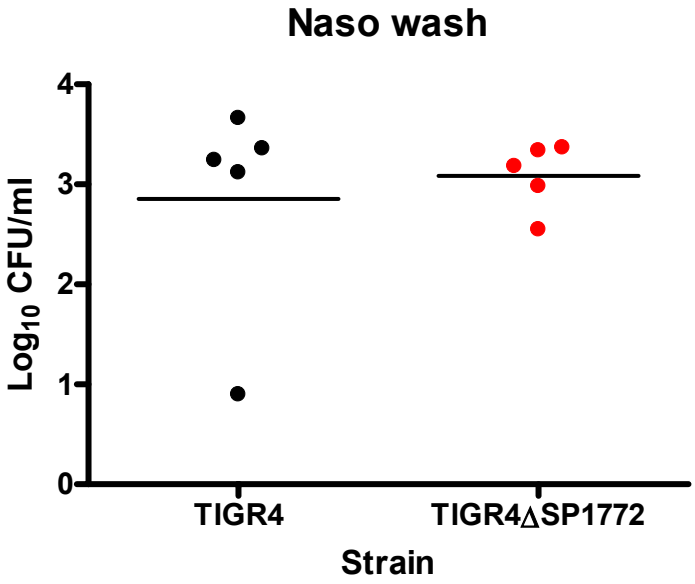
(b) Survival



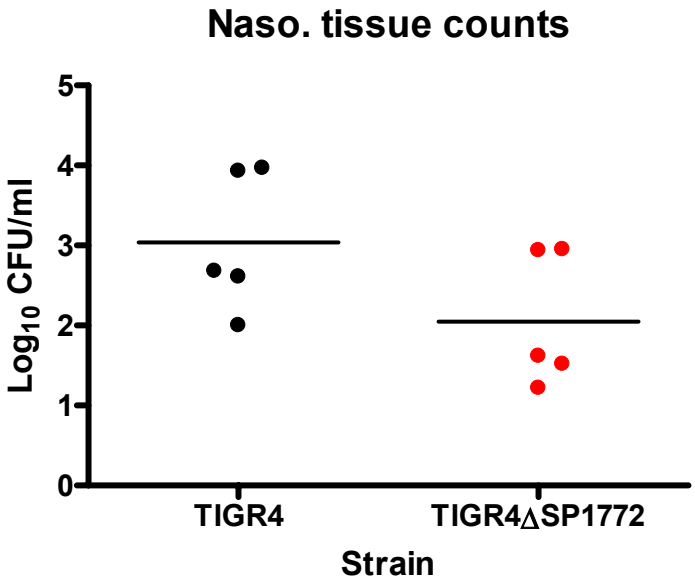
(a) Log of counts of bacteria in the blood of balb/c 2 d and 3 d post infection intransally with 10^7 CFU TIGR4 (T4) or TIGR4ΔSP1772 (T4ΔS). Each animal is represented by a symbol, with the mean represented by the horizontal line. (b) Survival curve from the experiment described above.

Figure 5.10 – Colonisation model of TIGR4 and TIGR4ΔSP1772

(a) Nasopharyngeal wash



(b) Nasopharyngeal tissue counts



(a) Log of counts of bacteria from a nasopharyngeal wash of MF-1 mice 5 d post infection intransally with 10^6 CFU TIGR4 or TIGR4ΔSP1772. Each animal is represented by a symbol, with the mean represented by the horizontal line. (b) Nasopharyngeal tissue counts from the experiment described above, also taken 5 d post infection.

biofilm formation can be reduced by more than 60% in the presence of capsule by comparing unencapsulated and capsulated strains. Isogenic mutants with different capsule types similarly were less able to produce biofilms (Moscoso *et al.*, 2006). However the authors also note that presumably, given that biofilms are composed predominantly of extracellular polysaccharides, it is premature to discount a role for the pneumococcal capsule in biofilm formation.

Capsule production has been shown to be tightly regulated, and this is due to the importance in its levels in different settings of the pneumococcus. For instance, in attachment to human cells it has been shown that the capsule is down-regulated (Hammerschmidt *et al.*, 2005) however the pneumococcus must maintain the ability to produce capsule, since it is unable to further cause disease without a capsule.

A recent study has identified that some single colony variants arise in biofilms (Allegrucci and Sauer, 2007). This could be due to the conditions in biofilms such as less oxygen promoting less capsule production (Weiser *et al.*, 1994) and also more H_2O_2 promoting a higher mutation rate (Pericone *et al.*, 2000). However, it still suggests that it could be favourable, in some circumstances, to the pneumococcus for this to occur. This is interesting, since unencapsulated mutants will not be able to go on and cause invasive disease in the host. The biofilm therefore is perhaps a safe haven for the unencapsulated mutants. Since biofilms are ideal settings for the exchange of genetic material, it is also possibly that these single colony variants can reside here to act as a recipient or donor of genetic material.

Nonetheless, the role of capsule in biofilm formation, if it does indeed have a role remains to be determined. Our studies suggest that biofilm formation in TIGR4 is slightly improved in the presence of the capsular locus. This is in disagreement with a study in which the TIGR4 capsule was shown not to have a role in biofilm development (Donlan *et al.*, 2004), however this study utilised a different biofilm model. It is also in disagreement with the observation in this model that the presence of capsule limits biofilm formation (Moscoso *et al.*, 2006), however TIGR4 was not utilised in this study. This result is in agreement with the observation that capsular expression levels are similar in a biofilm

mode of growth as in liquid medium (Oggioni *et al.*, 2006), and also that some capsule proteins are upregulated in biofilms (Allegrucci *et al.*, 2006).

The observation that encapsulated TIGR4 in the absence of SP1772 is impaired in its ability to form biofilms implies a role for SP1772 in biofilm formation. This difference in phenotype shows SP1772 must be produced by the cell, although a it does not define its role in *S.pneumoniae*. This would be in agreement with observations that homologues of SP1772 have been shown to be involved in biofilm development (Froeliger and Fives-Taylor, 2001; P. S. Handley, 2005; Quin *et al.*, 2007). The size of SP1772 at 412kDa, and its putative location at the cell surface (due to the presence of the LPTXG motif and associated secretion machinery) suggest that SP1772 extends potentially beyond the capsule and could mediate contact with other cells. In its absence but in the presence of capsule, cells may be less likely to form biofilms.

Several studies have noted the expense for the cell to switch from planktonic to sessile growth. It is feasible, given the expense to the cell of synthesising SP1772 and all of the additional genes required for its full expression, that in its absence, cells have more energy to produce the components of a biofilm.

There is no *in vivo* biofilm model at present however it has also recently been noted that the gene expression profile of pneumococci in a biofilm model is very similar to the expression of pneumococci in tissue models of infection (Oggioni *et al.*, 2006).

SP1772 null mutants were also evaluated in animal models of disease. Previously, SP1772 had been identified as involved in pneumococcal pneumonia in a STM screen of TIGR4 (Hava and Camilli, 2002). SP1772 was shown in a competitive model of infection to contribute to virulence. In our studies, which were single strain infection models, SP1772 deficient mutants were no less able to cause disease than the wild type, including a pneumonia model of infection. Future work with respect to this result would be to repeat the competitive model of infection, using the same dose and breed of mouse.

Consequently, a study was published that showed that when SP1772 in TIGR4 is disrupted by a method of insertion-duplication, the mutant strain is severly impaired in its ability to cause pneumonia in a mouse model of infection

compared to the wild type strain (Obert *et al.*, 2006). The model of infection differed from that which we had used previously, in that the strain of mouse differed, as did the innoculum.

It has been noted in other studies that variation across studies is seen with the pneumococcus. For instance, in a gene expression study by Oggioni and colleagues (Oggioni *et al.*, 2006) they note that while their data fit well with those described in another study in a sepsis model (Orihuela *et al.*, 2004), the same two studies differ in their results from a meningitis model of infection. The authors attribute this to differences in host species, challenge dose and time of sampling in the two models. They further note that when compared to yet another study (LeMessurier *et al.*, 2006), the tissue infection results concur, but there is a discrepancy in the results from the sepsis model (Oggioni *et al.*, 2006).

We therefore repeated the model of infection used in the study that had identified a role for SP1772 in virulence (Obert *et al.*, 2006), however the mutant strain constructed in our laboratory was no less able to cause infection in this model than the wild type. Presumably the differences arising therefore in the behaviour of the strains is due to the differing approaches to constructing the mutants. Certainly the RNA expression patterns of the two mutants differ, in a later publication by the same group, the authors confirm that the mutant they have constructed by insertion duplication does not have polar affects on neighbouring genes in the chromosome of SP1772: SP1771 and SP1773 (Rose *et al.*, 2008). This is demonstrated by northern blot. This is in disagreement with RT-PCR studies from our laboratory which show that SP1773 RNA is not expressed in either TIGR4 Chapter 3 or TIGR4ΔSP1772 (This chapter, Figure 5.3).

There's also the possibility that mouse model isn't suitable. This gene is clearly highly variable (chapter 1, chapter 4) and that could be a host response. Therefore it may not play a role in mice, like it does in humans, or the receptor in mice is different and it would take many serial passages through mice to see this. It would be interesting to Southern blot the DNA from the resulting culture to see if there was variation in size. It would also be interesting to do the same in the strain in human blood.

Chapter 6

Results

Analysis of *SP1772* in Inv200

6 Analysis of *SP1772* in Inv200

6.1 Introduction

In chapter 3, *S.pneumoniae* strains being sequenced at the Sanger Centre were annotated for genes in RD20 where they were present (Figure 6.1). It was noted that *SP1772* in Inv200 was far smaller, at 3.3kb, than in any of the other strains. In TIGR4 *SP1772* is 14.3 kb, in 23F 13.1 kb and in Inv104b it is 13.8 kb.

There were many issues in working with a gene and predicted product of a large size, which were not seen in work describing other smaller SP1772 homologues (Bensing and Sullam, 2002; Siboo *et al.*, 2005). Studies were therefore carried out with Inv200 *SP1772*, in the hope of overcoming some of the issues relating to the size of TIGR4 *SP1772*.

Primarily, at 3.3 kb it was deemed possible to amplify *SP1772* in Inv200. Upon amplification by PCR it was noted that there were many bands of several different sizes of the gene and therefore the studies in this chapter sought to characterise the heterogenous population resulting from PCR amplification of Inv200 *SP1772*.

6.2 Results

6.2.1 Southern blotting

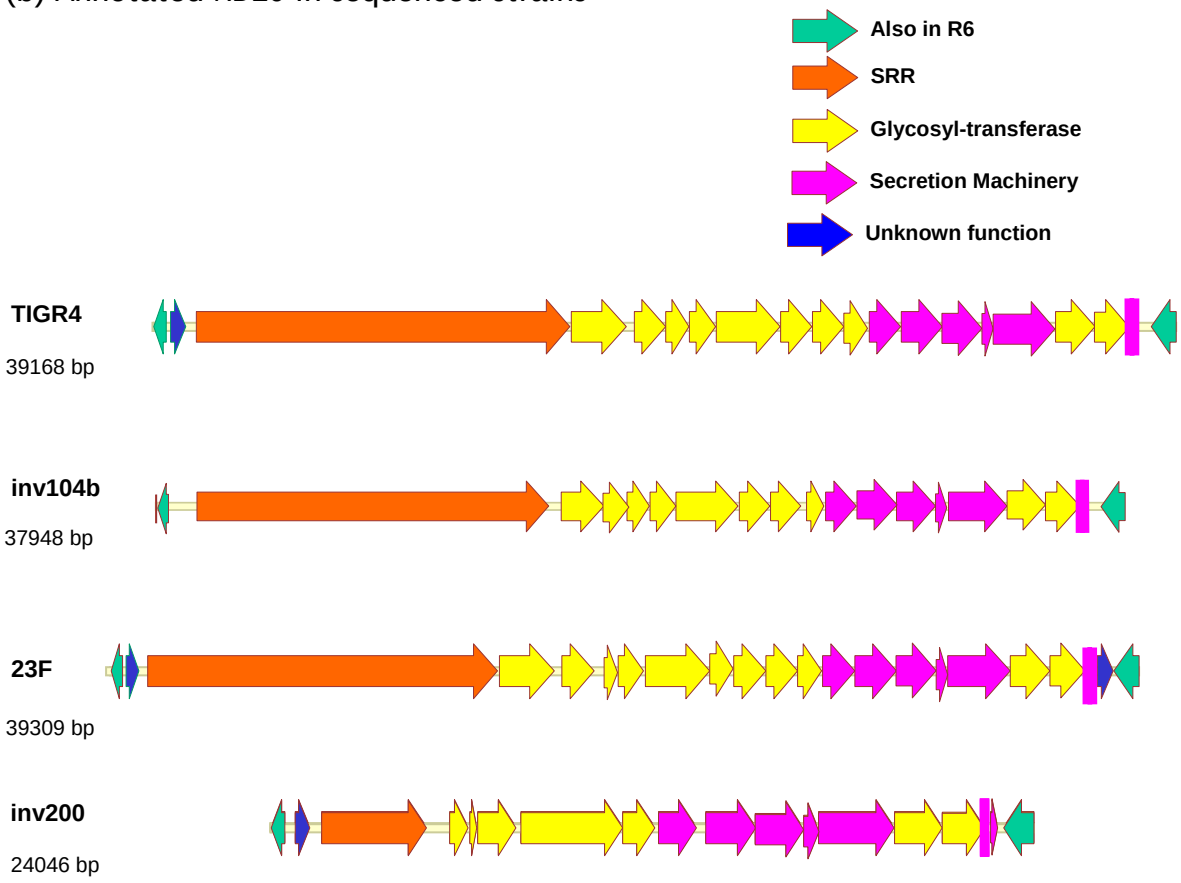
In order to confirm the size of Inv200 *SP1772* a southern blot was devised to compare *SP1772* in Inv200 with TIGR4 *SP1772*. A suitable enzyme which would digest the DNA whilst leaving *SP1772* intact was using Vector NTI (Invitrogen). The enzyme was SspI (Figure 6.2a). A probe was then designed that would incorporate a non repeat region, and also repeat region of *SP1772*. Primers were designed using the TIGR4 genome (Tettelin *et al.*, 2001). The 3' primer (56O) sequenced through a repeat region, and thus allowed a heterogenous probe containing different amounts of repeat region to maximise the chance of the probe annealing to the DNA (Figure 6.2b, 6.3a). The 5' primer (56N) sequenced through a non-repeat region upstream of the repeat

Figure 6.1 – Annotation of sequenced strains containing RD20

Strain	Glycosyl transferases	Secretion machinery	Conserved hypothetical protein
TIGR4	**10**	**7**	**1**
Inv104b	**10**	**7**	**0**
23f	**11**	**7**	**2**
Inv200	**7**	**7**	**0**

(a) Table showing functions of genes in RD20 in sequenced strains

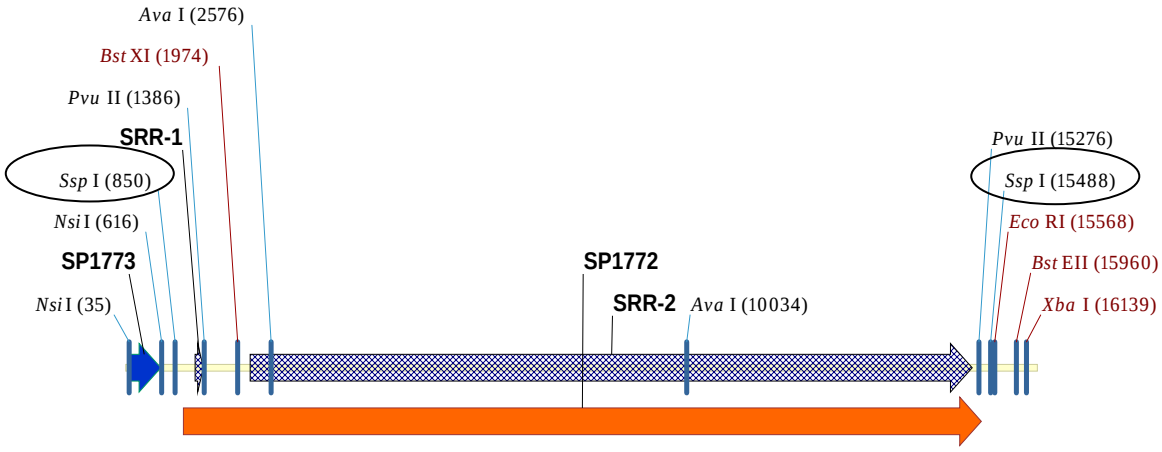
(b) Annotated RD20 in sequenced strains



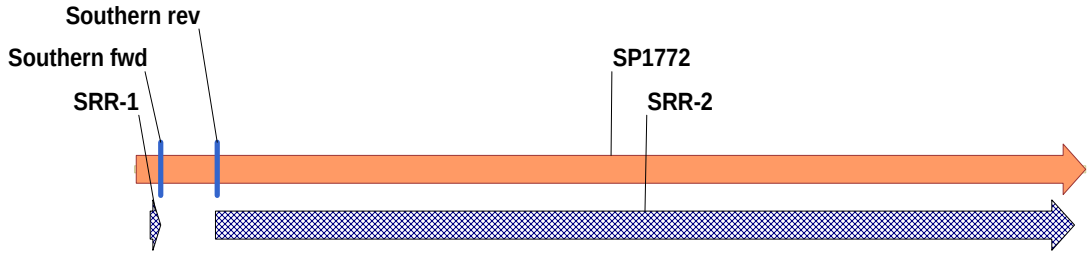
(a) Genes grouped by function in RD20 in sequenced strains. (b) RD20 in Inv104b, 23F and Inv200. Each ORF was predicted, and the translated protein product was blasted against TIGR4. Functions were assigned on the basis of homology.

Figure 6.2 – Designing a southern blot to detect *SP1772*

(a) Selecting a suitable restriction enzyme



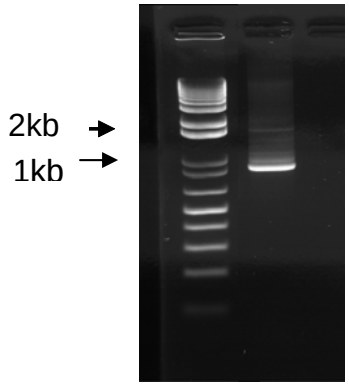
(b) Designing an appropriate probe



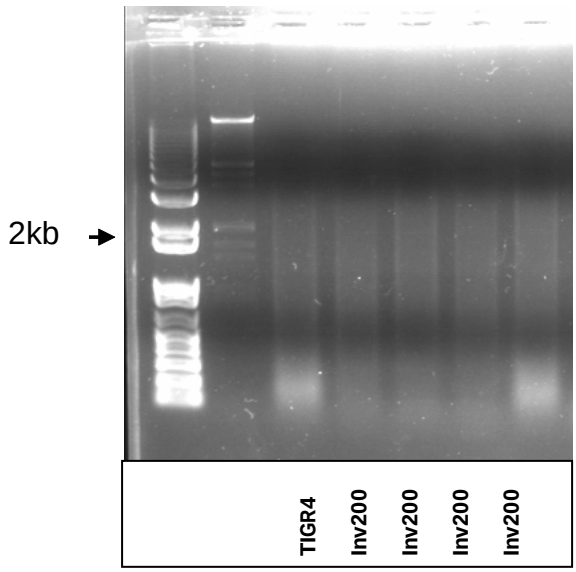
(a)A suitable restriction enzyme to digest TIGR4 DNA to get the whole *SP1772* gene was found by mapping all available restriction enzymes against the target sequence using Vector NTI. SSpI was chosen on the basis that it cut the genome at an appropriate point to obtain a suitably sized fragment containing *SP1772*. (b) A probe was designed to incorporate both repeat (SRR-1) and non-repeat region from *SP1772*. The reverse primer was designed to anneal to a repeat region, to give a probe heterogenous in size.

Figure 6.3 – Southern blotting to detect *SP1772*

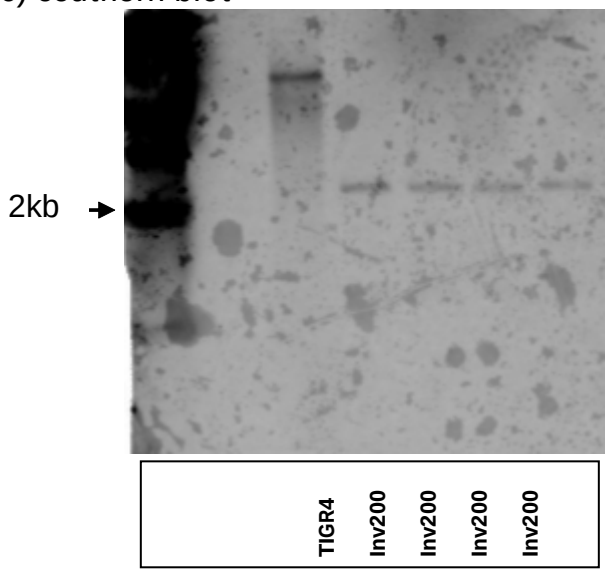
(a) Probe PCR



(b) Genomic DNA digestion for Southern blot



(c) Southern blot



(a) PCR of the *SP1772* probe. The 3' primer was designed to anneal to a repeat region, and so the probe is heterogenous in size. (b) Genomic DNA was digested using SSpI, and run on a gel to confirm digestion. Digestion of DNA by specific

enzymes generally leads to a smear down the gel since there are many different sizes of DNA present. (c) A southern blot confirming the size of *SP1772* in TIGR4 is bigger than that of Inv200.

region in *SP1772.* The southern blot confirmed that *SP1772* in TIGR4 is far larger than Inv200 *SP1772* (Figure 6.3). Interestingly, the predominant band of SP1772 in Inv200 appeared to be smaller than the 3.3kb size determined by whole genome sequencing at the Sanger Centre. Due to the confirmed smaller size of Inv200 *SP1772,* it was thought it would be possible to amplify the gene by PCR.

6.2.2 PCR optimisation

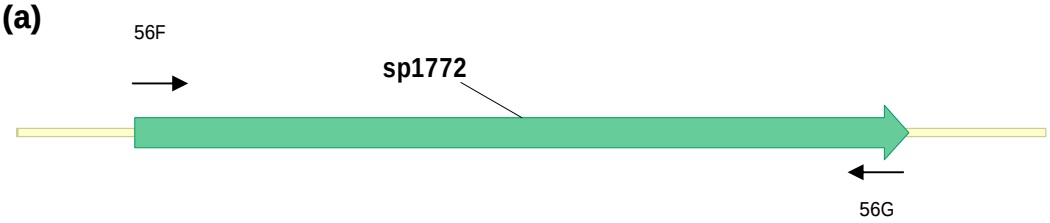
Genomic DNA was extracted from a culture grown from a single colony of *S.pneumoniae* strain Inv200. Primers 56F and 56G were then used to amplify the homologue of *SP1772* from Inv200. Primers were designed against the sequence available from ongoing sequencing projects at the Sanger Centre. A variety of enzymes were used to do this, Invitrogen Taq Hi-Fi was the enzyme that was successful. The product was predicted to be 3.3kb. The resulting PCR mixture was run on a gel and it was noted that there was a mixture of products, of varying sizes, notably none of 3.3kb (Figure 6.4). This did not match with what had been observed in the southern blot of appropriately digested TIGR4 and Inv200 gDNA, where it appeared that there may be many forms of TIGR4 *SP1772*, but Inv200 *SP1772* appeared to be only present in one form. The resulting PCR products were therefore analysed, in an attempt to determine if there were varying forms of Inv200 present in the cell, and the sequence of these different forms.

6.2.3 Cloning

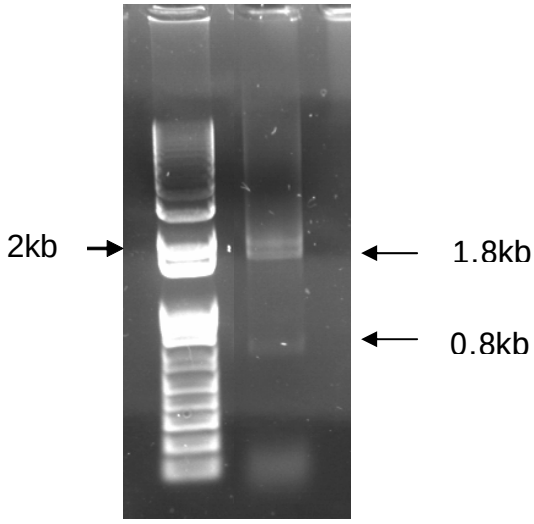
The mixture from the PCR of Inv200 SP1772 was cloned into Topo TA (Invitrogen). Successful clones were selected from a plate and colony PCR was carried out using plasmid specific primers 7F and 7G. The reactions gave bands of several different sizes, as shown in Figure 6.5a. It is worth noting that PCR of all clones resulted in only one clear band. There weren't multiple bands seen in the PCR of any clones from the original PCR. This negated the possibility that the heterogenous population of products from the original PCR of *SP1772* in Inv200 was due to enzyme slippage, and not due to the presence of many different sizes of *SP1772* in Inv200. Selected clones were re-run in approximate size order to further evaluate the differences in size present (Figure 6.5b). The clones that appeared to be representative of the different sizes present were selected for

Figure 6.4 - PCR of In200*SP1772*

(a)

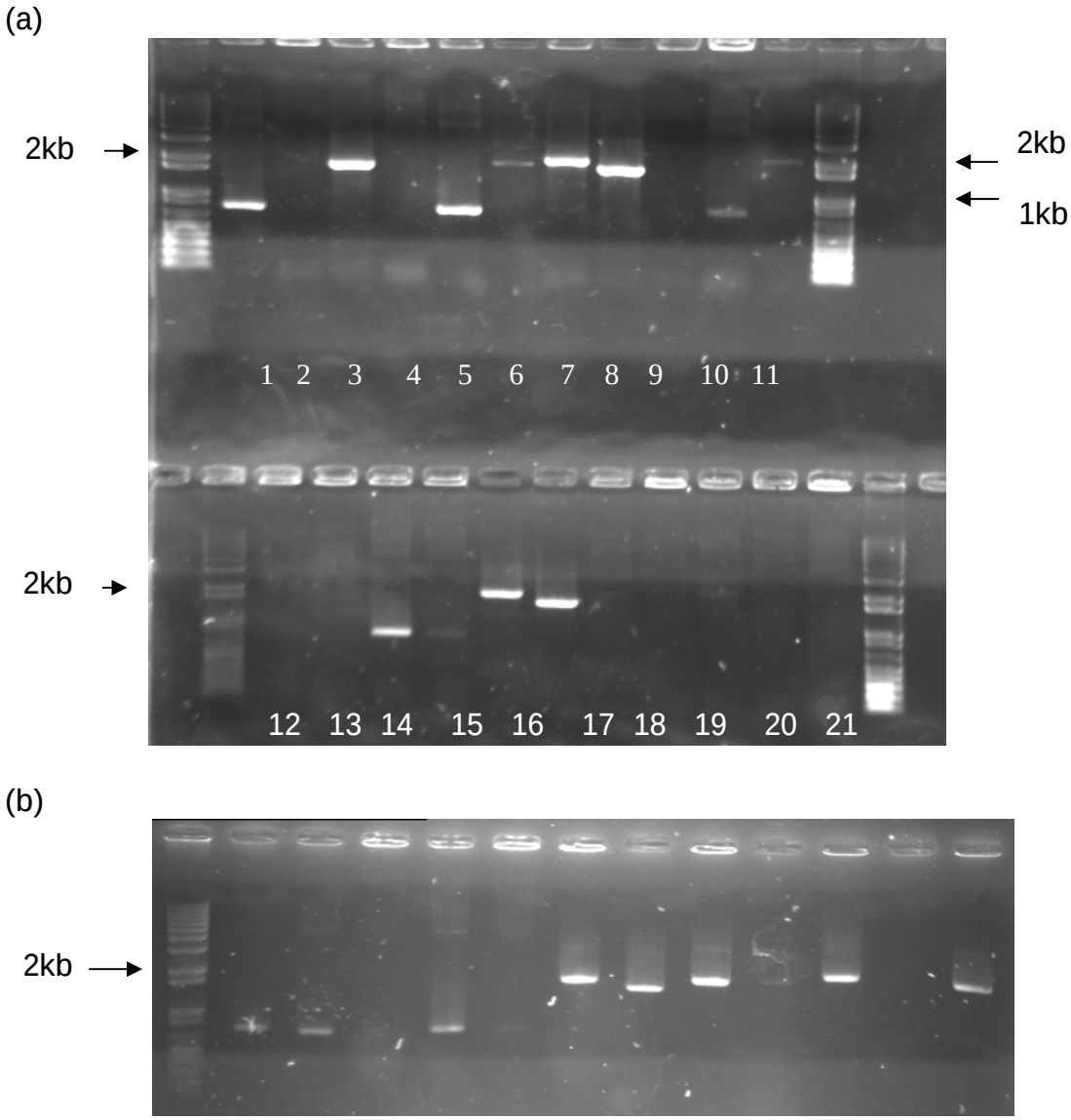


(b)



(a) Primers 56F and 56G are designed to amplify SP1772 in Inv200. A PCR using primers 56F and 56G to amplify Inv200 *SP1772*. The PCR resulted in many bands, the predominant bands being at ~ 0.8, 1.8 and 2.0 kb.

Figure 6.5 – PCR of clones resulting from PCR of Inv200*SP1772*



(a) PCR was carried out using primers T7 and 56G to determine the size of *SP1772* present in each plasmid. (b) Clones were re-ordered approximately according to size in order to select representative clones of each size for sequencing.

sequencing, to identify what the sequences were, if they were in frame and could still be functional proteins and identify sites of recombination within the gene. Since the amplification of each clone in *E.coli* resulted in single bands, even for large clones, it was deemed that the original amplification from *S.pneumoniae*, which resulted in multiple bands, was due to a heterogenous population of *SP1772* in Inv200 and not due to enzyme slippage or PCR error.

6.2.4 Sequence data of all clones

The 5 clones selected from the PCR of Inv200 *SP1772* were sequenced at the Sanger centre. The resulting sequences were aligned against the Inv200 genome and the sequences were matched to the correct place in the Inv200 genome. Data for each clone is shown in Figures 6.6-6.10, and discussed below. Data for all clones is shown in Figures 6.11 and 6.12. Figure 6.11 presents the DNA sequences of each clone mapped against the full sequence of Inv200 *SP1772*, and Figure 6.12 presents the amino acid sequences encoded by each clone mapped against the predicted amino acid sequence of full Inv200 SP1772, annotated with features.

All clones contained part of the N-terminal and part of the C-terminal. In all instances the start and stop codons of Inv200 *SP1772* were present. All clones were also found by sequencing to be in-frame.

Clone 1 is presented in Figure 6.6. It is 672bp long (additional sequence is coded in green and represents the plasmid). It contains 332bp of the N-terminal (shown in red) and 340bp of the C-terminal (shown in blue) of full length Inv200 *SP1772*. Recombination appears to have occurred between SRR-1 and SRR-2, as demonstrated by sequencing and shown in Figure 6.11 for all clones. Recombination has occurred mid-codon, and has caused GCG to change to GCA, however an alanine residue is still encoded by this codon.

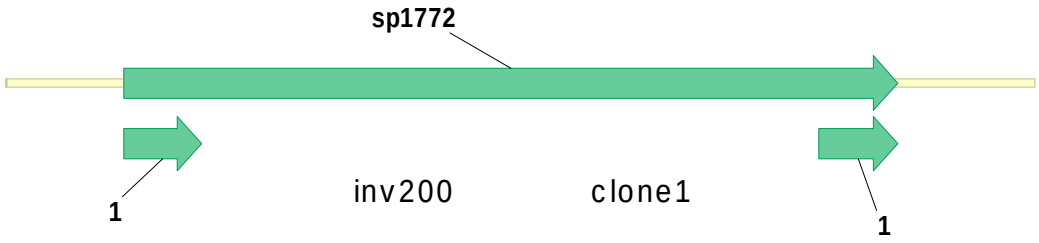
Clone 5 is presented in Figure 6.7. It is 642bp long. It contains 362bp of the N-terminal and 281bp of the C-terminal of full length Inv200 *SP1772*. Recombination appears to have occurred between SRR-1 and SRR-2. Recombination has occurred mid-codon, and has caused ACA to change to ACC, however a threonine is still encoded by this codon.

Figure 6.6 – Analysis of Inv200*SP1772* Clone 1

(a) Sequence analysis by BLAST

```
      1  AGATG CATGC TCGAG CGGCC GCCAG TGTGA TGGAT ATCTG CAGAA TTCGC
     51  CCTTG TCTTC ATCAT CTCTT TTACG ACGTT TGGCA ACCAA TCCTA AACCT
    101  GTTAA ACCAA GTATC AATCC AAGAA GTCCT GATTT CACAG ACTGT TTCTC
    151  TCCTG TATTT GGTAA TTCCT GTCTT GAATT GGTTA TTGTC TCTTG TGGTC
    201  GAGAA TTAGT TGATC CACTT GTACT TGTTG ACTCA GAGGT ACTTA CTGAC
    251  GCGCT TGTGC TTGCT GAAGC TGAGG CACTC GTTGA TGCCG ATTCA GAAGC
    301  TGACG CACTT GTCGA CGCTG AAGCC GACGC ACTGG TACTT GCTGA AGCTG
    351  ATGTA CTTGC TGAGG TCGAC GCGCT TGTTG ATGCT GATGT ACTTG CTGAG
    401  GTCGA AGCGC TGGTT GAGGC TGACT CAGAT GCAGA TGTAC TTGCC GACTC
    451  TGAAG CTGAC AAACT AGTTG AACTC GCTGA ATTAC TTGTA GATAT CGTAC
    501  CTAGA ACTAC TGTAC CATTT GTTGC CAAAG CATCC GTTTT CTCTA CAGTT
    551  TTTTC AAGTA CTGCC GACTC ATTTG TAAAT ACCTT CGTTT GAGTT GCAAC
    601  GGTTC CACTT ATGAC AGCTC CCACA GCAGC TATTC CCTTG AGGAT ATCAA
    651  GCCCA GTAAT TGAAT GACTT ACTTT ATCTT CTACC GTTTC GGTCA TGACC
    701  TGAGT AGTAT CAACA CCACC ACGCA TAAGG GCGAA TTCCA GCACA CTGGC
    751  GGCCG TTACT A
```

(b)Annotation of the sequence against the full sequence



(a) The sequence of the clone was BLASTED against the full sequence of Inv200 in parts to find the homology. The red indicates the N-Terminal portion, blue indicates the C-Terminal portion, whilst green indicates TopoTA sequences. Therefore the point between red and blue indicates the site of recombination. Clone 1 is 672bp long. It contains 332bp of the N-terminal and 340bp of the C-terminal of full length Inv200 SP1772. Recombination has occurred mid-codon, and has caused GCG to change to GCA, however an alanine residue is still encoded by this codon. (b) The sequence of the clone is indicated against the full sequence, obtained by sequencing of Inv200 at the Sanger Institute.

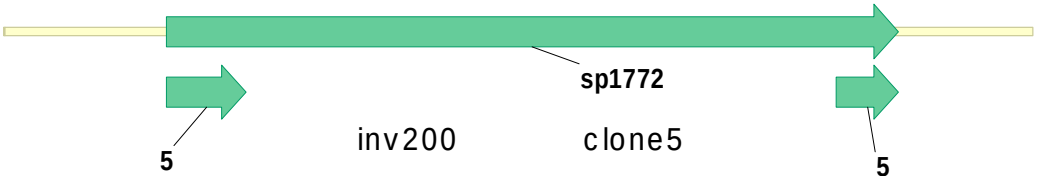
Figure 6.7 – Analysis of Inv200*SP1772* Clone 5

(a) Sequence analysis by BLAST

1	GACGG CCAGT GAATT GTAAT ACGAC TCACT ATAGG GCGAA TTGGG CCCTC
51	TAGAT GCATG CTCGA GCGGC CGCCA GTGTG ATGGA TATCT GCAGA ATTCG
101	CCCTT GTCTT CATCA TCTCT TTTAC GACGT TTGGC AACCA ATCCT AAACC
151	TGTTA AACCA AGTAT CAATC CAAGA AGTCC TGATT TCACA GACTG TTTCT
201	CTCCT GTATT TGGTA ATTCC TGTCT TGAAT TGGTT ATTGT CTCTT GTGGT
251	CGAGA ATTAG TTGAT CCACT TGTAC TTGTT GACTC AGAGG TACTT ACTGA
301	CGCGC TTGTG CTTGC TGAAG CTGAG GCACT CGTTG ATGCC GATTC AGAAG
351	CTGAC GCACT TGTCG ACGCT GAAGC CGACG CACTG GTACT TGCTG
401	GATGC ACTTG TACTC GCTGA GGTCG AAGCG CTGGT TGAGG CTGAC TCAGA
451	TGCAG ATGTA CTTGC CGACT CTGAA GCTGA CAAAC TAGTT GAACT CGCTG
501	AATTA CTTGT AGATA TCGTA CCTAG AACTA CTGTA CCATT TGTTG CCAAA
551	GCATC CGTTT TCTCT ACAGT TTTTT CAAGT ACTGC CGACT CATTT GTAAA
601	TACCT TCGTT TGAGT TGCAA CGGTT CCACT TATGA CAGCT CCCAC AGCAG
651	CTATT CCCTT GAGGA TATCA AGCCC AGTAA TTGAA CGACT TACTT TATCT
701	TCTAC CGTTT CGGTC ATGAC CTGAG TAGTA TCAAC ACCAC CACGC ATAAG
751	GGCGA ATTCC AGCAC ACTGG CGGCC GTTAC TAGTG GATCC GAGCT CGGTA
801	CCAAG CTTGG CGTAA TCATG GTCAT AGCTG TTTCC TGTGT

296 bp

(b)Annotation of the sequence against the full sequence



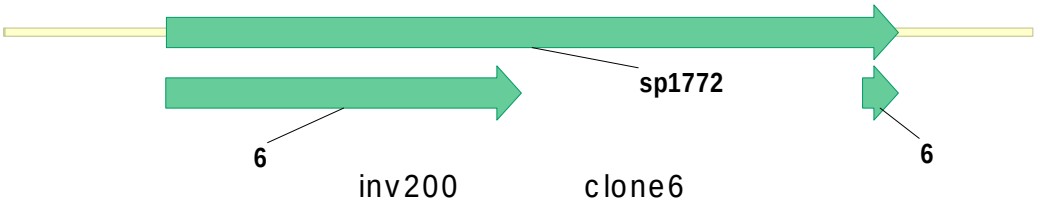
(a) The sequence of the clone was BLASTED against the full sequence of Inv200 in parts to find the homology. The red indicates the N-Terminal portion, blue indicates the C-Terminal portion, whilst green indicates TopoTA sequences. Therefore the point between red and blue indicates the site of recombination. Clone 5 is 642bp long. It contains 362bp of the N-terminal and 281bp of the C-terminal of full length Inv200 SP1772. Recombination has occurred mid-codon, and has caused ACA to change to ACC, however a threonine is still encoded by this codon. (b) The sequence of the clone is indicated against the full sequence, obtained by sequencing of Inv200 at the Sanger Institute.

Figure 6.8 – Analysis of Inv200*SP1772* Clone 6

(a) Sequence analysis by BLAST

1	CGGCC GCCAG TGTGC TGGAA TTCGC CCTTG TCTTC ATCAT CTCTT TACGA
51	CGTTT GGCAA CCAAT CCTAA ACCTG TTAAA CCAAG TATCA ATCCA AGAAG
101	TCCTG ATTTC ACAGA CTGTT TCTCT CCTGT ATTTG GTAAT TCCTG TCTTG
151	AATTG GTTAT TGTCT CTTGT GGTCG AGAAT TAGTT GATCC ACTTG CACTT
201	GTCGA TGCTG ACGTG CTTGC GCTTG TTGAT GCTGA CGTGC TTGCT GACGT
251	ACTTG CACTT GTCGA TGCTG ACGTG CTTGC CGATG TACTT GCGCT TGTTG
301	ATGCT GACGT GCTTG CTGAC GTGCT TGCGC TTGTT GATGC TGACG TGCTT
351	GCTGA CGTAC TTGCG CTTGT TGATG CCGAT GTACT TGCTG ACGTA CTTGC
401	GCTTG TTGAT GCCGA TGTAC TTGCT GAGGT CGACG CGCTT GTTGA TGCTG
451	ATGTA CTTGC TGAGG TCGAC GCGCT TGTTG ATGCC GACGT ACTTG CACTT
501	GTTGA AGCCA ATGTA CTTGC TGACG TACTT GTGCT TGCTG AAGCT GAAGC
551	GCTTG TCGAT GCTGA AGCTG AGGCG CTTTG ACTAG CACTT ACTGA TAGTG
601	ACTTA GACTG TGAAA GTGAC TGACT GGTCG TGCTA GATTC AACTA CCTTT
651	CTTGT ACCAT TGAAG TAGTT AGTTC CGATT CTATC TGTTT TAGCA GCGTA
701	AGGGG TAAAT GTAAA GGATG TATCC GTTCC AGTAA TTGGT ACAGT CCAAG
751	ATGAT GCTAA TCCAT ATCCT TTCTT AGCAA AGAAA CCATT CATTT GGGCA
801	CCATT TGCCC AAGTA TACCC ACTAC GCCAC ATCGT CATTG TAGAT GTATT
851	ATAGG GTCGA ACCTG TTTAC CATTT TTGTC AGTAA TGGAG TTCTT TACAC
901	CTGAA GGTTT ACCAA GACCA CTGCC AAGGG TTAAC ATTGT TTGTA TTGAA
951	GTTCC TGTAT TATAG ATAGA ATATC CATTA CTCAT CTTTT TTGAT ATATT
1001	ACCAA GAGTC TTTGT TTTAG GATTC ACATA CGTAA CCGTA TAGGT AAAGG
1051	TCAAT TTTGA ACCAT CATTT ATCAC TGTCA ATTGA TAGTA AATTG GAATA
1101	TTATA AAATG AATCT TTACC TGTAC CCGTG TAAAC TTTTG TCTCA CTTTT
1151	AGCAA TGTTT AGACT TGCAT TAATT GCAGG GGCGC CATTT ACAAT CGTAT
1201	TGCCA GAAAA AACAG CAGCA TTTTT TATAG AAGCC AGCAA TTGCT CGATG
1251	GAATC CACTG AACGC TTGCG TCGCT TAGCA TAAGA TTGGA GATTG ACATT
1301	TGTAA CTGAT GCTAC ATAAT CACTA GCTGG TTTCT TACGA TCTTC TTCGA
1351	CCTTC TTAGC AGTTG CTTCT GTAGC GGCAG CTGTT TGTGA ACCTA CCACA
1401	GTAGA TGATG CAGAG ATACT TGTCG AAGCC GATGT ACTTG CTGAT TCTGA
1451	TGCAC TTGTA CTCGC TGAGG TCGAA GCGCT GGTTG AGGCT GACTC AGATG
1501	CAGAT GTACT TGCCG ACTCT GAAGC TGACA AACTA GTTGA ACTCG CTGAA
1551	TTACT TGTAG ATATC GTACC TAGAA CTACT GTACC ATTTG TTGCC AAAGC
1601	ATCCG TTTTC TCTAC AGTTT TTTCA AGTAC TGCCG ACTCA TTTGT AAATA
1651	CCTTC GTTTG AGTTG CAACG GTTCC ACTTA TGACA GCTCC CACAG CAGCT
1701	ATTCC CTTGA GGATA TCAAG CCCAG TAATT GAATG ACTTA CTTTA TCTTC
1751	TACCG TTTCG GTCAT GACCT GAGTA GTATC AACAC CACCA CGCAT AAGGG
1801	CGAAT TCTGC AGATA TCCAT C

(b)Annotation of the sequence against the full sequence



(a) The sequence of the clone was BLASTED against the full sequence of Inv200 in parts to find the homology. The red indicates the N-Terminal portion, blue indicates the C-Terminal portion, whilst green indicates TopoTA sequences. Therefore the point between red and blue indicates the site of recombination. Clone 6 is 1673bp long. It contains 1378bp of the N-terminal and 295bp of the C-terminal of full length Inv200 SP1772. Recombination has occurred between two full codons. (b) The sequence of the clone is indicated against the full sequence, obtained by sequencing of Inv200 at the Sanger Institute.

Clone 6 is presented in Figure 6.8. It is 1673bp long. It contains 1378bp of the N-terminal and 295bp of the C-terminal of full length Inv200 *SP1772*. Recombination appears to have occurred within SRR-2. Recombination has occurred between two full codons.

Clone 8 is presented in Figure 6.9. It is 1773bp long. It contains 1607bp of the N-terminal and 166bp of the C-terminal of full length Inv200 *SP1772*. Recombination appears to have occurred within SRR-2. Recombination has occurred mid-codon, but the new codon is still ACA, which encodes alanine.

Clone 16 is presented in Figure 6.10. It is 1749bp long. It contains 1433bp of the N-terminal and 316bp of the C-terminal of full length Inv200 *SP1772*. Recombination appears to have occurred within SRR-2. Recombination has occurred mid-codon, and has caused GCA to change to GCG, however an alanine is still encoded by this codon.

It appeared that the resulting truncated clones of SP1772 were either from recombination between SRR-1 and SRR-2 or from within SRR-2. It is interesting that in all cases, clones remain in frame, and therefore would still encode a SRR. The recombination in clone 8 resulted in two full codons from each part of *SP1772* remaining adjacent to one another. In the case of clones 1, 5, 8 and 16 recombination has occurred in the middle of a codon, and in the case of clones 1,5 and 16 a different codon is produced. However in all cases, the same amino acid is still encoded at that position.

This implies that this process is highly regulated in *S.pneumoniae*, and suggests that it may be of importance to strains to be able to produce truncated forms of SP1772.

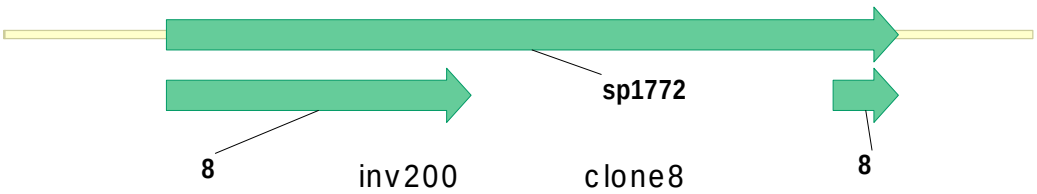
Interestingly, none of the clones sequenced contained the full predicted sequence of Inv200 *SP1772*. All of the original PCR products were examined, and no product was larger than ~2kb. The southern blot also suggests that the predominant product is smaller than 3.3kb. Inv200 *SP1772* was found to be 3.3kb in the strain sequenced at the Sanger Centre. It was possible that due to passage of the strain, the 3.3kb full *SP1772* was no longer the predominant size of

Figure 6.9 – Analysis of Inv200*SP1772* Clone 8

(a) Sequence analysis by BLAST

1	GATGG ATATC TGCAG AATTC GCCCT TGTCT TCATC ATCTC TTTTC GACGT
51	TTGGC AACCA ATCCT AAACC TGTTA AACCA AGTAT CAATC CAAGA AGTCC
101	TGATT TCACA GACTG TTTCT CTCCT GTATT TGGTA ATTCC TGTCT TGAAT
151	TGGTT ATTGT CTCTT GTGGT CGAGA ATTAG TTGAT CCACT TGTAC TTGTT
201	GACTC AGAGG TACTT ACTGA CGCGC TTGTG CTTGC TGAAG CTGAG GCACT
251	CGTTG ATGCC GATTC AGAAG CTGAC GCACT TGTCG ACGCT GAAGC CGACG
301	CACTG GTACT TGCTG AAGCT GATGT ACTTG CTGAG GTCGA CGCGC TTGTT
351	GATGC TGATG TACTT GCTGA GGTCG ACGCG CTTGT TGATG CCGAC GTACT
401	TGCAC TTGTT GAAGC CAATG TACTT GCTGA CGTAC TTGTG CTTGC TGAAG
451	CTGAA GCGCT TGTCG ATGCT GAAGC TGAGG CGCTT TGACT AGCAC TTACT
501	GATAG TGACT TAGAC TGTGA AAGTG ACTGA CTGGT CGTGC TAGAT TCAAC
551	TACCT TTCTT GTACC ATTGA AGTAG TTAGT TCCGA TTCTA TCTGT TTTAG
601	CAGCG TAAGG GGTAA ATGTA AAGGA TGTAT CCGTT CCAGT AATTG GTACA
651	GTCCA AGATG ATGTT AATCC ATATC CTTTC TTAGC AAAGA AACCA TTCAT
701	TTGGG CACCA TTTGC CCAAG TATAC CCACT ACGCC ACATC GTCAT TGTAG
751	ATGTA TTATA GGGTC GAACC TGTTT ACCAT TTTTG TCAGT AATGG AGTTC
801	TTTAC ACCTG AAGGT TTACC AAGAC CACTG CCAAG GGTTA ACATT GTTTG
851	TATTG AAGTT CCTGT ATTAT AGATA GAATA TCCAT TACTC ATCTT TTTTG
901	ATATA TTACC AAGAG TCTTT GTTTT AGGAT TCACA TACGT AACCG TATAG
951	GTAAA GGTCA ATTTT GAACC ATCAT TTATC ACTGT CAATT GATAG TAAAT
1001	TGGAA TATTA TAAAA TGAAT CTTTA CCTGT ACCCG TGTAA ACTTT TGTCT
1051	CACTT TTAGC AATGT TTAGA CTTGC ATTAA TTGCA GGGGC GCCAT TTACA
1101	ATCGT ATTGC CAGAA AAAAC AGCAG CATTT TTTAT AGAAG CCAGC AATTG
1151	CTCGA TGGAA TCCAC TGAAC GCTTG CGTCG CTTAG CATAA GATTG GAGAT
1201	TGACA TTTGT AACTG ATGCT ACATA ATCAC TAGCT GGTTT CTTAC GATCT
1251	TCTTC GACCT TCTTA GCAGT TGCTT CTGTA GCGGC AGCTG TTTGT GAACC
1301	TACCA CAGTA GATGA TGCAG AGATA CTTGT CGAAG CCGAT GTACT TGCTG
1351	ATTCT GATGC ACTTG TACTC GCTGA GGTCG AAGCG CTGGT TGAGG CTGAC
1401	TCAGA TGCAG ATGTA CTTGC CGACT CTGAA GCTGA CAAAC TAGTT GAACT
1451	CGCTG AATTA CTTGT AGATA TCGTA CCTAG AACTA CTGTA CCATT TGTTG
1501	CCAAA GCATC CGTTT TCTCT ACAGT TTTTT CAAGT ACTGC CGACT CATTT
1551	GTAAA TACCT TCGTT TGAGT TGCAA CGGTT CCACT TATGA CAGCT CCCAC
1601	AGCAG CTATT CCCTT GAGGA TATCA AGCCC AGTAA TTGAA TGACT TACTT
1651	TATCT TCTAC CGTTT CGGTC ATGAC CTGAG TAGTA TCAAC ACCAC CACGC
1701	ATAAG GGCGA ATTCC AGCAC ACTGG CGGCC G

(b)Annotation of the sequence against the full sequence



(a) The sequence of the clone was BLASTED against the full sequence of Inv200 in parts to find the homology. The red indicates the N-Terminal portion, blue indicates the C-Terminal portion, whilst green indicates TopoTA sequences. Therefore the point between red and blue indicates the site of recombination. Clone 8 is 1773bp long. It contains 1607bp of the N-terminal and 166bp of the C-terminal of full length Inv200 SP1772. Recombination has occurred mid-codon,

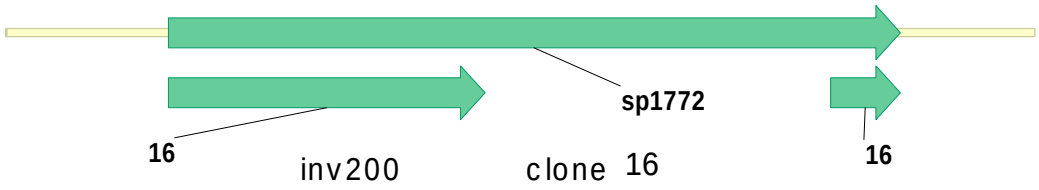
but the new codon is still ACA, which encodes alanine. (b) The sequence of the clone is indicated against the full sequence, obtained by sequencing of Inv200 at the Sanger Institute.

Figure 6.10 – Analysis of Inv200*SP1772* Clone 16

(a) Sequence analysis by BLAST

1	GATGG ATATC TGCAG AATTC GCCCT TGTCT TCATC ATCTC TTTTC GACGT
51	TTGGC AACCA ATCCT AAACC TGTTA AACCA AGTAT CAATC CAAGA AGTCC
101	TGATT TCACA GACTG TTTCT CTCCT GTATT TGGTA ATTCC TGTCT TGAAT
151	TGGTT ATTGT CTCTT GTGGT CGAGA ATTAG TTGAT CCACT TGTAC TTGTT
201	GACTC AGAGG TACTT ACTGA CGCGC TTGTG CTTGC TGAAG CTGAG GCACT
251	CGTTG ATGCC GATTC AGAAG CTGAC GCACT TGTCG ACGCT GAAGC CGACG
301	CACTG GTACT TGCTG AAGCT GATGT ACTTG CTGAG GTCGA CGCGC TTGTT
351	GATGC CGATG TACTT GCTGA CGTAC TTGCG CTTGT TGATG CCGAT GTACT
401	TGCTG AGGTC GACGC GCTTG TTGAT GCTGA TGTAC TTGCT GAGGT CGACG
451	CGCTT GTTGA TGCCG ACGTA CTTGC ACTTG TTGAA GCCAA TGTAC TTGCT
501	GACGT ACTTG TGCTT GCTGA AGCTG AAGCG CTTGT CGATG CTGAA GCTGA
551	GGCGC TTTGA CTAGC ACTTA CTGAT AGTGA CTTAG ACTGT GAAAG TGACT
601	GACTG GTCGT GCTAG ATTCA ACTAC CTTTC TTGTA CCATT GAAGT AGTTA
651	GTTCC GATTC TATCT GTTTT AGCAG CGTAA GGGGT AAATG TAAAG GATGT
701	ATCCG TTCCA GTAAT TGGTA CAGTC CAAGA TGATG TTAAT CCATA TCCTT
751	TCTTA GCAAA GAAAC CATTC ATTTG GGCAC CATTT GCCCA AGTAT ACCCA
801	CTACG CCACA TCGTC ATTGT AGATG TATTA TAGGG TCGAA CCTGT TTACC
851	ATTTT TGTCA GTAAT GGAGT TCTTT ACACC TGAAG GTTTA CCAAG ACCAC
901	TGCCA AGGGT TAACA TTGTT TGTAT TGAAG TTCCT GTATT ATAGA TAGAA
951	TATCC ATTAC TCATC TTTTT TGATA TATTA CCAAG AGTCT TTGTT TTAGG
1001	ATTCA CATAC GTAAC CGTAT AGGTA AAGGT CAATT TTGAA CCATC ATTTA
1051	TCACT GTCAA TTGAT AGTAA ATTGG AATAT TATAA AATGA ATCTT TACCT
1101	GTACC CGTGT AAACT TTTGT CTCAC TTTTA GCAAT GTTTA GACTT GCATT
1151	AATTG CAGGG GCGCC ATTTA CAATC GTATT GCCAG AAAAA ACAGC AGCAT
1201	TTTTT ATAGA AGCCA GCAAT TGCTC GATGG AATCC ACTGA ACGCT TGCGT
1251	CGCTT AGCAT AAGAT TGGAG ATTGA CATTT GTAAC TGATG CTACA TAATC
1301	ACTAG CTGGT TTCTT ACGAT CTTCT TCGAC CTTCT TAGCA GTTGC TTCTG
1351	TAGCG GCAGC TGTTT GTGAA CCTAC CACAG TAGAT GATGC AGAGA TACTT
1401	GTCGA AGCCG ATGTA CTTGC TGATT CTGAT GCACT TGTAC TCGCT GAGGT
1451	CGAAG CGCTG GTTGA GGCTG ACTCA GATGC AGATG TACTT GCCGA CTCTG
1501	AAGCT GACAA ACTAG TTGAA CTCGC TGAAT TACTT GTAGA TATCG TACCT
1551	AGAAC TACTG TACCA TTTGT TGCCA AAGCA TCCGT TTTCT CTACA GTTTT
1601	TTCAA GTACT GCCGA CTCAT TTGTA AATAC CTTCG TTTGA GTTGC AACGG
1651	TTCCA CTTAT GACAG CTCCC ACAGC AGCTA TTCCC TTGAG GATAT CAAGC
1701	CCAGT AATTG AATGA CTTAC TTTAT CTTCT ACCGT TTCGG TCATG ACCTG
1751	AGTAG TATCA ACACC ACCAC GCATA AGGGC GAATT CCAGC ACACT GGCGG
1801	CCG

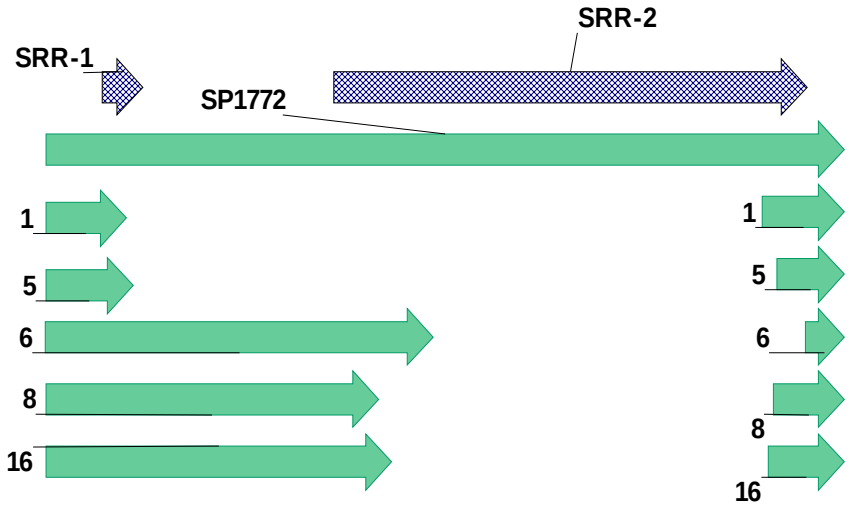
(b)Annotation of the sequence against the full sequence



(a) The sequence of the clone was BLASTED against the full sequence of Inv200 in parts to find the homology. The red indicates the N-Terminal portion, blue indicates the C-Terminal portion, whilst green indicates TopoTA sequences. Therefore the point between red and blue indicates the site of recombination. Clone 16 is 1749bp long. It contains 1433bp of the N-terminal and 316bp of the C-terminal of full length Inv200 SP1772. Recombination has occurred mid-codon, and has caused GCA to change to GCG, however an alanine is still encoded by

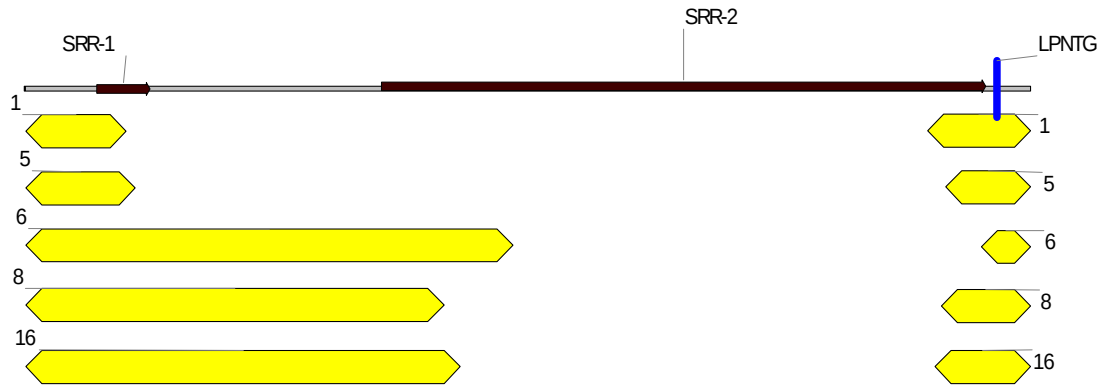
this codon. (b) The sequence of the clone is indicated against the full sequence, obtained by sequencing of Inv200 at the Sanger Institute.

Figure 6.11 – Alignment of all clones against Inv200 *SP1772* DNA



The sequence of all analysed clones is indicated against the full sequence of Inv200 *SP1772*, obtained by sequencing of Inv200 at the Sanger Institute. All of the gene products sequenced were in frame when translated to protein sequence, as shown in Figure 6.12.

Figure 6.12 – Alignment of all clones against Inv200 SP1772



The sequence of all analysed clones is indicated against the full translated protein sequence of Inv200 *SP1772*, obtained by translation of the Inv200 *SP1772* sequence in Vector NTI (Invitrogen).

SP1772 in Inv200. The corresponding area of the DNA gel to 3.3kb was cloned into Topo TA, however no clones resulted. This was done in order to establish if there was a minor quantity of full *SP1772* present, however was not successful. This could be due to full *SP1772* not being present, being present in too small a quantity or alternatively not being able to be amplified by PCR.

6.3 Discussion

SP1772, like other serine rich repeat proteins, contains 2 repeat regions: one small region at the N-terminal, and a much larger C-Terminal region. In the Inv200 genome, the first naturally truncated homologue of TIGR4 SP1772 was identified in *S.pneumoniae*. It was of a size that could be amplified by PCR. Amplification of this gene resulted in products of varying size. The results presented in this chapter show that, in Inv200, *SP1772* is heterogenous in size when grown in liquid culture. Different sizes of the gene were detected, ranging from 642bp-1773bp. Smaller clones 1 and 5 resulted from recombination between SRR-1 and SRR-2 within *SP1772*, resulting in the complete loss of the region encoded between SRR-1 and SRR-2. Larger clones 6, 8 and 16 resulted from recombination within SRR-2, resulting in clones truncated only in the size of SRR-2. All sequenced clones were found not to have altered their sequence at all, and were all in frame. Where recombination had occurred mid-codon, the resulting codon still encoded the amino acid that would have occurred at the N-terminal portion. The differing sizes of clones suggest that recombination could be successive. For instance, Clone 6, 8 and 16 could all represent an intermediate size that *SP1772* recombines to before it goes further to the size of clones 1 and 5. However, it is interesting that the site of recombination is different for all 5 clones sequenced. Due to the highly repetitive nature of the SRRs, it is not surprising that recombination occurs within these regions, nor perhaps that the site of recombination is not identical.

The presence of two different approximate sizes of *SP1772* in Inv200 suggests that these sizes of *SP1772* could be useful to the strain, or that the gene is successively getting smaller, either to then increase in size upon immune pressure (although the complete absence of the region between SRR-1 and SRR-2 would presumably the negate the ability for this to be reformed without homologous recombination) or else that the strain is attempting to rid itself of

the gene entirely. This would also be logical, since the locus is both expensive to harbour and to synthesise, and there is possibly no pressure on the strain to maintain the locus in liquid culture i.e. the locus is presumably not required for pneumococal growth given that it is absent in many pneumococcal strains. Also, although not presented here given time constraints in the ability to evaluate the mutant, a mutant of SP1754-SP1773 was constructed in the TIGR4 background and found to be viable, showing that the locus is not required for viability where it is commonly associated with a strain.

Presumably there is a pressure driving the recombination of *SP1772* in Inv200. It is not clear what this is, although as speculated above, it could be that the growth of the strain in BHI results in this. It would be interesting to analyse the behaviour of this gene in blood culture or *in vivo* models. It would also be interesting to grow cultures containing a truncated *SP1772* in blood culture or passage them to see if their size can increase.

Since the gene is cell-surface exposed, it may be a form of immune evasion that the gene can alter in its size. It may also be that the target that SP1772 interacts with in the human host can alter from host to host, and the gene adapts to this. It is interesting to note that Inv200 contains far fewer glycosyltransferases, and therefore this may be a development over time. The glycosyltransferases present in TIGR4 may exhibit some functional redundancy, and may in the evolution of the bacteria disappear. Alternatively, it could be that the size of SP1772 has been increasing, and new glycosyltransferases were acquired/copies of existing ones in the genome were copied to make the gene even more capable of immune evasion. Different glycosyltransferases will glycosylate in different ways. Different sugar residues are attached to the protein backbone and can interact with targets in different ways. The number of glycosyltransferases is reflective of the diversity where they are also diverse – and perhaps a divergence in this chromosomal region allows a diversity in host target interaction. Serotype 14 strains (such as Inv200) might have evolved to interact with a different target than Serotype 4 strains (such as TIGR4). Although both are involved in invasive disease it may be a subtle indication of the role that SP1772 is responsible for being different in these hosts, or that the capsular type is variant in its acceptance of the presence of SP1772. The locus appears to be always present in serotype 4 strains, but can differ in its presence in

serotype14 strains. Serotype 14 ST9 is commonly recovered from invasive disease as are serotype 4 strains (Enright and Spratt, 1998; Hausdorff *et al.*, 2000; McGee *et al.*, 2001) as evidenced by the fact that both serotypes have been included in the pneumococcal vaccine Prevnar (Black *et al.*, 2000).

The slight difference in the other pneumococcal sequenced strains of *SP1772* also suggests recombination within the gene can occur. The much different ASL associated with Inv200 *SP1772* is what suggests it is far more divergent than that in other sequenced pneumococcal strains. *S.pneumoniae* has been shown to acquire genetic material from other species, the most illustrative example of this is the acquisition of the penicillin binding proteins, conferring penicillin resistance on the pneumococcus, acquired from *S.mitis* (Dowson et al., 1993). *S.pneumoniae* has been shown to occupy the same niche as other bacteria and is extremely competent at acquiring DNA, not just from other strains of pneumococci, but also from other closely related species (Claverys and Havarstein, 2007; Johnsborg *et al.*, 2007; Johnsborg *et al.*, 2008). However, comparison of the Inv200 ASL with other species harbouring an ASL shows it is even more divergent from these, which, based on the evidence we have thus far, suggests that the locus was acquired into the pneumococcal population and that in Inv200 this locus has subsequently diverged in two particularly able regions: the SRR protein and the region downstream from it. The fact that this region is also divergent in other species presumably points to the fact that this region is able to diverge, as oppose to it having been acquired from elsewhere.

The variation of the locus in characterised strains of gram positive bacteria is shown in Chapter 3, Figure 3.7. *S.gordonii* and *S.aureus* both have the region SP1771-SP1764 missing and in *S.algalactiae* it is different, in that it encodes fewer glycosyl-transferases, and it encodes transposases. In *S.parasanguinis* SP1772-SP1764 and asp4 and asp5 are missing (Wu *et al.*, 2007), interestingly, while these latter two have been shown to be essential for the secretion of GspB in *S.gordonii* (Takamatsu *et al.*, 2005) they have also been shown to be missing in *S.aureus* (Siboo et al., 2005). It could be speculated that the size of SRR defines the number of accompanying genes, however, in *S.algalactiae* the SRR is much truncated from that *S.gordonii, S.aureus* or *S.parasanguinis*, and yet the ASL encodes more genes (Takamatsu *et al.*, 2004).

It is worth noting that the southern blot to identify *SP1772* in TIGR4 and Inv200 suggests that there is a heterogenous population of *SP1772* in TIGR4, and not in Inv200. TIGR4 SP1772 is too big to amplify, and simply results in a smear down the gel. This is probably because the gene is so large that the respective amounts of each size of SP1772 are not easy to distill into distinct bands. It would be interesting to clone the smear and analyse these clones. Perhaps Inv200 does not appear heterogeneously on a southern blot because the most predominant size is the one that appears on the blot. It would be informative to carry out Southern blotting on the SP1772 gene from human isolates to determine the extent of the variation in size of *SP1772* in the pneumococcal population. However, it is worth considering that a southern blot only appears to show the diversity where the gene is big enough in the first place. RT-PCR could also be informative, from liquid culture and blood culture and from *in vivo* models.

In results chapter 4, methods to detect the protein expression of SP1772 were unsuccessful, and further work in this area should certainly concentrate of finding a way to detect the SP1772 protein. Potentially, a method to detect the different sizes of the SP1772 protein would be ideal. It would also be interesting to identify the genes responsible for this recombination using knockouts of genes involved in recombination.

Chapter 7

Final Discussion

7 Final Discussion

S.pneumoniae is a highly variable bacterial species that can vary in the content of its genome by up to 10% between two strains (Hakenbeck *et al.*, 2001). This diversity helps to explain the diversity in the outcome of interactions between *S.pneumoniae* and its human host, ranging from harmless asymptomatic carriage, to more benign diseases such as otitis media, to fatal pneumonia, septicaemia and meningitis (Austrian, 1999). However, it is also true that variations in the human host will also influence the outcome of the interaction as highlighted by the increased prevalence of invasive pneumococcal disease amongst the elderly and immunocomprimised (Black *et al.*, 2000), the apparent genetic predisposition of certain populations to pneumococcal disease (Cortese *et al.*, 1992), and the different outcomes resulting from the interaction of genetically similar strains in different regions of the world. For example, serotype 1 strains have been associated with meningitis in developing countries, and pneumonia in developed countries (Sjostrom *et al.*, 2006). Nonetheless, it is the genetic diversity of the pneumococcus that has been more extensively studied, and it has been found that strains which are identical even by the sensitive technique of MLST can vary in the content of their genomes, which can help to explain the different behaviours of strains of similar genetic background (Enright and Spratt, 1998; Hanage *et al.*, 2005). Variation in the sequence of important pneumococcal virulence factors is thought to contribute to the different behaviour of strains, however the variation in whole regions of the genome also contributes. Several whole genome comparative genomic hybridisation studies have been carried out to identify regions of diversity in the pneumococcus (Hakenbeck *et al.*, 2001; Obert *et al.*, 2006; Silva *et al.*, 2006), and the work presented here focuses on one of these regions of diversity in *S.pneumoniae*, encoding a large serine rich repeat protein, SP1772, and seventeen genes putatively dedicated to the glycosylation and export of this SRR protein. *S.pneumoniae* is highly competent and can take up DNA not just from other pneumococcal strains, but also from closely related bacteria. This locus has also been described in other gram positive bacteria, including *S.gordonii* (Bensing and Sullam, 2002; Takahashi *et al.*, 2004), *S.aureus* (Siboo *et al.*, 2005) and *S.parasanguinis* (Froeliger and Fives-Taylor, 2001). It is likely that *S.pneumoniae* therefore acquired this locus by horizontal gene transfer from a closely related species, this has been documented previously in the case of

acquisition of penicillin resistance genes by *S.pneumoniae* from *S.mitis* (Dowson et al., 1993).

This genetic locus was identified originally in *S.pneumoniae* as a region of diversity (Silva *et al.*, 2006), and studies presented in chapter 3, and work by Obert *et al* have confirmed that this locus varies in its presence across the pneumococcal population (Obert *et al.*, 2006). Sequencing of pneumococcal strains has also revealed that where this locus is present it can vary in the sequence of both the SRR and accompanying strains, as presented in chapter 3. Until there are a larger number of sequences available of strains of both pneumococci and closely related species, the evolution of the locus cannot be determined. However, in spite of the sequence divergence of the ASL within the pneumococcal population, the ASL in different strains of the pneumococcus appears more similar than that in other species.

Obert and colleagues also made the notable observation that this region can be correlated with the ability of an isolate to cause invasive disease, a finding which the authors showed by creating a mutant in SP1772 by insertion-duplication and demonstrating that it was less virulent in a single strain pneumonia model of infection (Obert *et al.*, 2006). This finding may be further corroborated by the finding that SP1772 is required for the full virulence of *S.pneumoniae*, as demonstrated by a signature tagged mutagenesis screen in a competitive model of infection of pneumonia (Hava and Camilli, 2002). However, as discussed in chapter 5, a mutant of SP1772 created by clean allelic replacement of SP1772 with an erythromycin cassette, was not attenuated in its ability to cause pneumonia in a single strain model of infection. It could be speculated that this is due to the differing genomic content of the parent strain, in spite of the fact that the same strain was used in all cases. One interesting difference in the strains used is the expression of SP1773 RNA. In the TIGR4ΔSP1772 strain used by Obert and colleagues, the RNA of transposase SP1772 was found to be expressed (Rose *et al.*, 2008) whilst studies in our laboratory show that SP1773 is neither expressed in TIGR4ΔSP1772 (Chapter 5), or in the parent TIGR4 strain (Chapter 3). The differences in the behaviour of the strains could be attributed to the method of mutation of SP1772, since three different methods were used across all three studies. It would be desirable to carry out further work utilising the two different mutants and their parent

strains in the same laboratory and the same experiment to try and understand how the variability in results has arisen. It is interesting to note that studies utilising the pneumococcus frequently vary in their outcome between laboratories, one example was shown in a gene expression study by Oggioni and colleagues (Oggioni *et al.*, 2006) the authors note that while their data fit well with those described in parts of some other studies, there are also discrepancies between other parts of the study (LeMessurier *et al.*, 2006; Orihuela *et al.*, 2004). In order to examine whether the model of infection and strain of mouse used was the variable in the outcome, the conditions described in Obert and colleagues work were duplicated. This, however, did not result in the same outcome and as such it is presumably either the genetic background of the parent strain or the difference in the method of mutagenesis used.

SRRs in other bacteria have been shown to have differing roles. One SRR, fap1 in *S.parasanguinis*, has been shown to be involved in biofilm formation (Froeliger and Fives-Taylor, 2001). The ability of the pneumococcus to form biofilms has recently been studied. Studies have noted the contribution of individual genes to biofilm formation (Moscoso *et al.*, 2006), the overall increased diversity in the expressed proteome of the pneumococcus when forming biofilms (Allegrucci *et al.*, 2006) and the differential gene expression pattern of pneumococci in biofilms and in stationary growth (Oggioni *et al.*, 2006). As presented in chapter 5, in the absence of SP1772, TIGR4 was less able to form biofilms. These studies also showed that biofilm formation is slightly better in TIGR4 than in an unencapsulated derivative, and that TIGR4ΔcpsΔSP1772 was the most able of all four strains to form biofilms. The role of capsule in biofilm formation is somewhat disputed in that while it has been shown not to be required for biofilm formation (Moscoso *et al.*, 2006), it has also been shown to be expressed in biofilms where it is present (Oggioni *et al.*, 2006), and that a capsular protein is actually required for optimal biofilm formation (Allegrucci *et al.*, 2006). There remains much to be elucidated both about the role of capsule and the ASL in biofilm formation, however evidence presented to date suggests that both do play a role. It would be interesting to evaluate the mutants generated during this study in a variety of biofilm models and assess the outcomes.

Attempts to visualise SP1772 in TIGR4 by conventional methods were not successful. It was also notable that attempts to raise antibodies to a particular

region of SP1772, just upstream of SRR-2 were also unsuccessful. It is this region of Inv200 SP1772 that was found to be absent completely in some clones resulting from the recombination. It is interesting that this region is not immunogenic – not just by conventional means but also by fusing this portion of the gene to highly immunogenic ply by gateway technology, particularly as the region has two predicted regions of immunogenicity (chapter 4). Future work to determine whether SP1772 is expressed, in reality as oppose to differential judgement of mutants strains, could focus on preparing an antibody specifically against SP1772 by adsorption of sera acquired from challenge with TIGR4 against a culture of TIGR4ΔSP1772, in the manner that serum against GspB was acquired (Bensing and Sullam, 2002). Alternatively, antisera could be raised in the manner utilised by Rose and colleagues (Rose *et al.*, 2008), and similar to SP1772A as presented in chapter 4. However, there is still a limit in visualising the protein since it appears not to enter SDS-PAGE gels.

It is possible that the study of a truncated form of SP1772 would be useful. Many studies with GspB have utilised a truncated version of this SRR (Bensing *et al.*, 2004; Takamatsu *et al.*, 2004, , 2005). It would also be feasible to use naturally truncated variants of SP1772, as is present in Inv200. Studies of *SP1772* in Inv200 revealed that it is not only predicted by sequencing to be much truncated from TIGR4 *SP1772*, but also that it is present in different sizes within the strain following growth in liquid culture. This recombination has been shown to take place either between the two SRRs within the gene, or within the larger SRR (Chapter 6). The ability of the pneumococcus to harbour different sizes of *SP1772*, as demonstrated in Chapter 3, is perhaps by virtue of selective pressure for the target that the gene is interacting with which could vary between strains. It may also be a biochemical or physiochemical property of that particular strain. The reasons for a strain containing a heterogenous population of *SP1772* is less clear, although it could be speculated that growth in liquid culture has driven the strain to reduce the burden that harbouring and synthesising this large gene has placed on the cell. The mechanism by which recombinational variants of *SP1772* arise is also unclear. It could be speculated that due to the repeat regions that are present in *SP1772* it is able to recombine by regions within the gene exerting homology and upon annealing when the DNA comes into contact homologous recombination can take place. The cells in which this takes places may then be more successful in a given environment because

there is no selective pressure on the strain to harbour a large, fully functional gene product of *SP1772.* Further study is warranted to identify the behaviour of *SP1772* in blood culture, biofilm and *in vivo* models to establish whether this phenomenon is also seen here, more relevant models to pneumococcal infection, and whether this varies in different models due to differing targets in each model. Further work could also focus on more representative models of infection. A single strain model of pneumococcal behaviour is fairly definitive in defining the attributes of a single strain, however it lacks the contribution of other other species which would likely be in contact with the pneumococcus. The pneumococcus is known to occupy the human niche with other closely related strains with whom it could be constantly interacting. Particularly in the case of *SP1772* which has homologues in many closely related species, elucidation of the role of this gene in the behaviour of the pneumococcus may only be found by evaluating *SP1772* mutants and their parent strains in multi-organism models of biofilm formation and infection.

Given the difficulty in visualising SP1772, as previously discussed, attempts were made to utilise an immunogenic FLAG tag to which antibodies are commercially available. The FLAG tag was attempted to be inserted at the C-terminal of *SP1772* upstream of the codons encoding the LPNTG motif. Attempts to do this were unsuccessful, and the usefulness of this given that full TIGR4 SP1772 cannot be visualised easily is limited. However, given the later discovery of a naturally truncated form of SP1772 in Inv200, it would be desirable to both experiment with visualising this truncated form of SP1772, and also to experiment with labelling this form with an immunogenic FLAG tag.

In addition to further work to the studies already presented, it would also be interesting to look at the levels of expression of *SP1772* RNA *in vivo*, in differing disease models. However, methods to capture RNA are extremely problematic since the RNA is difficult to retrieve at varying timepoints, and is stable for a very short period of time when it is retrieved.

The work presented here shows that *SP1772* and its associated genes are present in many different serotypes and sequence types of the pneumococcus. The sequence of both the SRR protein and the related genes can vary between strains, and indeed, the sequence of the SRR has been shown to vary within a

strain. Whilst no suitable detection method for the SP1772 protein has yet been developed, an interesting role for SP1772 in biofilm formation has been elucidated in the absence of capsule.

Many additional questions regarding the distribution and behaviour of *SP1772* in pneumococcal strains have arisen from these studies. Further work could focus on several areas. It would be desirable to understand more about the prevalence of *SP1772* in the species by further CGH studies, and to understand more about the expression of *SP1772* at both the RNA and protein level under different conditions. It would also be useful to understand why mutant strains differ between different laboratories and to evaluate these mutants in different models of infection, colonisation and biofilms. Finally, it would be useful to understand the mechanisms by which SP1772 modulates its size in a single strain of the pneumococcus and to elucidate the benefit to the pneumococcus of this behaviour.

Appendix I

Buffers and Recipes

1 Appendix I (buffers and recipes)

All reagents are from Sigma-Aldrich unless otherwise stated

1.1 Media

1.1.1 Terrific broth (TB)

12g bacto-tryptone
24 g bacto-yeast extract
4 ml glycerol
In 900 ml dH_2O, autoclave
Add the following:
In 90ml H_2O
2.31g KH_2PO_4 monobasic
12.54 g K_2HPO_4 dibasic (for trihydrate 16.45 g)
Adjust volume to 100 ml with H_2O

1.2 SDS-PAGE gels

1.2.1 Stacking gel

3.21ml dH_2O
1.25ml 0.5M Tris pH 6.8
0.05ml 10% SDS
0.488ml 30% polyacrylamide
0.025ml 10% ammonium persulphate
0.005ml TEMED

1.2.2 Resolving gel

10% - nb to adjust polyacrylamide and dH_2O are adjusted)
4.05ml dH_2O
2.5ml 1.5M Tris pH 6.8
0.1ml 10% SDS
3.3ml 30% polyacrylamide

0.05ml 10% ammonium persulphate
0.005ml TEMED

1.2.3 10x Running buffer

40g SDS
144g Glycine
30g Tris base
In 1L dH_2O

1.3 Western blotting

1.3.1 Transfer buffer

3.03g Tris base
14.4g Glycine
200ml Methanol
In 1L dH_2O, keep at 4˚C

1.3.2 Tris NaCl pH7.4

1.2g Tris base
8.7g NaCl
In 1L dH_2O
Conc. HCL 800µl

1.3.3 Developer

30mg 4-chloro-1-napthol in 10ml methanol
Mix with
30µl H_2O_2 (30% w/v) in 40ml Tris NaCl pH7.4

Appendix II

Primers used in this study

2 Appendix II (Primers used in this study)

Primer name	Description	Sequence
07F	Pet T7	TAATACGACTCACTATAGGG
07G	Pet term	GCTAGTTATTGCTCAGCGGTG
19H	Srt1	GAACATCGAACGAATCGC
19I	Srt2	TTCACAAACAGGACGTGC
19W	SrtAsc1	GGCGCGCCCTGCTAGGATATTTGTCAGC
19X	SrtAsc2	GGCGCGCCCTAACAGCTTTCAATCAACC
48L	5 NT START SP1772	ATC GGG ATC CGA TGA CCG AAA CGG TAG AAG
48M	5 NT NLS SP1772	ATC GGG GAT CCG GTA CCA AAT TAC TGG AAC G
48N	3 NT NR SP1772	GAC GGA GCT CGA TTA TCC ACC ATT GAA GTA GTT AAT TC
48O	3 NT RR SP1772	GAC GGA GCT CGA TTA TGA AGC TGA GGC GCT TTG AC
48P	pET SP1772 A 513	GAATATCCAGGACGCATAC
48Q	pET SP1772 A 579	CGGTTACGTATGGTAATCC
48R	pET SP1772 A 937	CCACAGTAGATGATGCAG
48S	pET SP1772 A 1060	GAGTCGGCAAGTACATCTG
48T	pET SP1772 C 351	GATGTATCCGTTCCAGTAATTG
48U	pET SP1772 C 669	CGGTTACGTATGTGAATC
48V	pET SP1772 C 864	GCCAGCAATTGCTCGATGG
48W	pET SP1772 C 1164	GTTTGTCAGCTTCAGAGTCG
48X	pET SP1772 C 1305	CCACTTATGACAGCTCCCG
49V	SP1772 P1	AATTAGCTGATTTATACTCATTTGC
49W	SP1772 P2	CCGCCATTCTTTGCTGTTTCGGACCTGAGTAGTATCAACACCACC
49X	SP1772 P3	GGAAAGTTACACGTTACTAAAGGCACAAGTTGGAAATACTTCT

		GG
49Y	SP1772 P4	AATTTCCGGACATATTATTCAGACC
49Z	Erm1	CGAAACAGCAAAGAATGGCGG
50A	Erm2	CCCTTTAGTAACGTGTAACTTTCC
50J	1772P1A	CATTAAAGGAGGGCACAGACATGTCG
50K	1772P4A	AGATAAACCACTAGCTTGACCGTTGG
50V	SP1774e	CACAAGATGATTGCGCATG
50W	SP1773s	GCATAGCTTTATTAGACC
50X	SP1773e	GATTCACAAATTAGAGAATG
50Y	SP1772s	CTTCTACCGTTTCGGTC
50Z	SP1772e	GTGTTACTTGGAGTTCTAGC
51A	SP1771s	CCTGAAGCATCCGTAGAACC
51B	SP1771e	GAGAGAAATAAGCAAATG
51C	SP1770s	CTATAGATTTTGTATTC
51D	SP1770e	GTTACGATTTATCCTTG
51E	SP1769s	CTAATTAGTTCGTCTGCC
51F	SP1769e	GATTACAAAGAGTGATAGC
51G	SP1768s	CCCAGTTGACTAGCTATC
51H	SP1768e	CTGGAAGAGGTTCATGAG
51I	SP1767s	GTTGTTTCTAACTGCCTGGTG
51J	SP1767e	CTGCAGAACATAATTTCG
51K	SP1766s	GACTATTGTGTCTACAGAG
51L	SP1766e	CTTACCAATGTTATTGGAAC
51M	SP1765s	CTAATACTATTGATTTTCTC
51N	SP1765e	CATAGTAGAATTGATTGGC
51O	SP1764s	CAGAATATGCTCGACACAC
51P	SP1764e	CCGTTCTGCGTATTCGATTTG
51Q	SP1763s	GACGAATAAATCTTAGTC
51R	SP1763e	CGATTGGCAATGATTCCAGG
51S	SP1762s	CCATGGAGTGATATCTGC
51T	SP1762e	GCAGATGTTACAGAATTTTC
51U	SP1761s	CTCTGGAATACTGACCATC
51V	SP1761e	GCCTATGAACAATTAGTG
51W	SP1760s	GTAATAAGATACAGTTGTTC
51X	SP1760e	CTTATCCAGAAGATAGTTATAC
51Y	SP1759s	GCTACTAATTCTTGATCCG
51Z	SP1759e	CAGAAAAATCCAATCGTAG
52A	SP1758s	CACCAGCACGATAGGCTTG
52B	SP1758e	CTGACCATGTAGAAGACC
52C	SP1757s	CGTAACCTAGATAATAGG
52D	SP1757e	CAGACGGTGCACAATAGAC
52E	SP1756s	GTTTCAACTCTTCCATGCG
52F	SP1756e	GATGATTCTGATTGGTTTGC
52G	SP1755s	GTTTACCAATATTACTAGTG
52H	SP1755e	CTGGCAGAGCAACACCTTGG
52I	SP1754s	GAAGCCAAAACCTTCATCC
52W	SP1774 RT f	CCAAGCGAGCCACCATCCTA
52X	SP1774 RT r	GACCTGCCAGATAAGCTTC
52Y	SP1773 RT f	GATGTCGGCGTCAGAACTCT

52Z	SP1773 RTr	TCCGCTCTAAGTTCCCTTG
53A	SP1772 RTf	AATGAGTCAGCAGTACTTG
53B	SP1772 RTr	TCGCTGAATTACTTGTAG
53C	SP1771 RTf	CTATCTGAGGAAGTGCCTA
53D	SP1771 RTr	CCATCATTAACGACAACA
53E	SP1770 RTf	CTAGATGGTTATGCCTTGG
53F	SP1770 RTr	GGTAGAAGGAACATCACC
53G	SP1769 RTf	GTGTTTGGATAGCATTCAGA
53H	SP1769 RTr	CTCTACAAATTCTTCACAT
53I	SP1768 RTf	GAAATGAATAAGCGTCTGGA
53J	SP1768 RTr	GGTTGGAGACTGAAAGACT
53K	SP1767 RTf	CAATTACACCAGGCAGTTA
53L	SP1767 RTr	TCTGGCATGATATCTTGA
53M	SP1766 RTf	ACATTTGCACGCTATTTTA
53N	SP1766 RTr	CAAATCACCAGTCACAATCA
53O	SP1765 RTf	GCCTATCTTATTCCTTTA
53P	SP1765 RTr	CCATTCAGGAGCTATATCA
53Q	SP1764 RTf	GTGGAGTGTGTCGAGCATA
53R	SP1764 RTr	GAACCGTCATCAACTAAA
53S	SP1763 RTf	GTGTTACTTGCTTTATTA
53T	SP1763 RTr	CAGGAACCAAATAGCGAGCT
53U	SP1762 RTf	GGTGTGTTAGAAATGGATACT
53V	SP1762 RTr	CTTGGGTATGGAGATTATGA
53W	SP1761 RTf	GTCCCCAACAACTAATCAA
53X	SP1761 RTr	CTTGTATCGTCGGAGGAA
53Y	SP1760 RTf	GTTCGTCTGTATAATCCTCTA
53Z	SP1760 RT	CTGGTAATTGACACTAGA

	r	
54A	SP1759 RT f	CCATGCCTGTCTATTTGAA
54B	SP1759 RT r	CATTTCCTCGGCATCACGCT
54C	SP1758 RT f	GAACTTTCGCAGATTTATAG
54D	SP1758 RT r	CTGAACCAATAGCTTCCATCAA
54E	SP1757 RT f	GCAGATAGCGAGCCGATTATGA
54F	SP1757 RT r	CAATAAGATATCACCCGTCACA
54G	SP1756 RT f	GTCAGAGGAAGATTTATT
54H	SP1756 RT r	CAACTCTTCCATGCGGCCTT
54I	SP1755 RT f	GATTGAAATACTAATTGTTT
54J	SP1755 RT r	GACGGGGTTGTATAGTTA
54K	SP1754 RT f	GGTTTCAAATTCTTCTCTT
54L	SP1754 RT r	GTCTTCCGGAACCGCAG
56F	INV200 SP1772 fwd	ATGCGTGGTGGTGTTGATAC
56G	INV200 SP1772 rev	GTCTTCATCATCTCTTTTAC
56L	SP1772D-Ply Gateway fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGTACCAATTACTGGAACG
56M	SP1772D-Ply Gateway rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCTGAAGCTGAGGCGCTTTGAC
56N	SP1772 Southern fwd	GTGGTAGGTTCACAAACAGCTGCC
56O	SP1772 Southern rev	CTGAAGCCGAGGCACTGGTTGATG
57C	inv200 sp1772 1 rev	CTCCTGTATTTGGTAATTCC
57D	inv200 sp1772 1 fwd	GTAGAGAAAACGGATGCTTTG
57E	Inv200 SP1772 6	GTATGTGAATCCTAAAACAAAG

	fwd	
57F	Inv200 SP1772 6 rev	GAGAATTAGTTGATCCACTTG
57G	Inv200 SP1772 7 fwd	CTGGCAATACGATTGTAAATG
57H	Inv200 SP1772 7 rev	GCTGAGGCACTCGTTGATGCCG
57I	Inv200 SP1772 8 rev	GCCGACGCACTGGTACTTG
57J	Inv200 SP1772 8 fwd	CAAATGTCAATCTCCAATC
57K	Inv200 SP1772 16 fwd	CAAAATTGACCTTTACCTATAC
57L	Inv200 SP1772 16 rev	CTGGTACTTGCTGAAGCTG
57M	inv200 sp1772 pET33b-F	ATCGGGATCCGATGCGTGGTGGTGTTGATAC
57N	inv200 sp1772 pET33b-F	GACGGAGCTCGATTATCATCTCTTTTACGACGTTT

References

References

Aanensen, D.M., Mavroidi, A., Bentley, S.D., Reeves, P.R., and Spratt, B.G. (2007) Predicted Functions and Linkage Specificities of the Products of the Streptococcus pneumoniae Capsular Biosynthetic Loci. *J. Bacteriol.* **189**: 7856-7876.

Alanee, S.R., McGee, L., Jackson, D., Chiou, C.C., Feldman, C., Morris, A.J., Ortqvist, A., Rello, J., Luna, C.M., Baddour, L.M., Ip, M., Yu, V.L., and Klugman, K.P. (2007) Association of serotypes of Streptococcus pneumoniae with disease severity and outcome in adults: an international study. *Clin Infect Dis* **45**: 46-51.

Allegrucci, M., Hu, F.Z., Shen, K., Hayes, J., Ehrlich, G.D., Post, J.C., and Sauer, K. (2006) Phenotypic characterization of Streptococcus pneumoniae biofilm development. *J Bacteriol* **188**: 2325-2335.

Allegrucci, M., and Sauer, K. (2007) Characterization of colony morphology variants isolated from Streptococcus pneumoniae biofilms. *J Bacteriol* **189**: 2030-2038.

Allegrucci, M., and Sauer, K. (2008) The formation of Streptococcus pneumoniae non-phase variable colony variants is due to increased mutation frequency present during biofilm growth conditions. *J Bacteriol*.

AlonsoDeVelasco, E., Verheul, A.F., Verhoef, J., and Snippe, H. (1995) Streptococcus pneumoniae: virulence factors, pathogenesis, and vaccines. *Microbiol. Rev.* **59**: 591-603.

Amdahl, B.M., Rubins, J.B., Daley, C.L., Gilks, C.F., Hopewell, P.C., and Janoff, E.N. (1995) Impaired natural immunity to pneumolysin during human immunodeficiency virus infection in the United States and Africa. *Am J Respir Crit Care Med* **152**: 2000-2004.

Andersson, H., and von Heijne, G. (1991) A 30-residue-long "export initiation domain" adjacent to the signal sequence is critical for protein translocation across the inner membrane of Escherichia coli. *Proc Natl Acad Sci U S A* **88**: 9751-9754.

Aniansson, G., Alm, B., Andersson, B., Larsson, P., Nylen, O., Peterson, H., Rigner, P., Svanborg, M., and Svanborg, C. (1992) Nasopharyngeal colonization during the first year of life. *J Infect Dis* **165 Suppl 1**: S38-42.

Appelbaum, P.C. (2002) Resistance among Streptococcus pneumoniae: Implications for drug selection. *Clin Infect Dis* **34**: 1613-1620.

Arrecubieta, C., Lopez, R., and Garcia, E. (1994) Molecular characterization of cap3A, a gene from the operon required for the synthesis of the capsule of Streptococcus pneumoniae type 3: sequencing of mutations responsible for the unencapsulated phenotype and localization of the capsular cluster on the pneumococcal chromosome. *J Bacteriol* **176**: 6375-6383.

Austrian, R. (1981) Some observations on the pneumococcus and on the current status of pneumococcal disease and its prevention. *Rev Infect Dis* **3 Suppl**: S1-17.

Austrian, R. (1999) The pneumococcus at the millennium: not down, not out. *J Infect Dis* **179 Suppl 2**: S338-341.

Avery, O., MacLeod, C., and McCarty, M. (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III. *J Exp Med* **79**: 137-157.

Baddour, L.M. (1994) Virulence factors among gram-positive bacteria in experimental endocarditis. *Infect Immun* **62**: 2143-2148.

Bagnoli, F., Moschioni, M., Donati, C., Dimitrovska, V., Ferlenghi, I., Facciotti, C., Muzzi, A., Giusti, F., Emolo, C., Sinisi, A., Hilleringmann, M., Pansegrau, W., Censini, S., Rappuoli, R., Covacci, A., Masignani, V., and Barocchi, M.A. (2008) A second pilus type in Streptococcus pneumoniae is prevalent in emerging serotypes and mediates adhesion to host cells. *J Bacteriol* **190**: 5480-5492.

Barocchi, M.A., Ries, J., Zogaj, X., Hemsley, C., Albiger, B., Kanth, A., Dahlberg, S., Fernebro, J., Moschioni, M., Masignani, V., Hultenby, K., Taddei, A.R., Beiter, K., Wartha, F., von Euler, A., Covacci, A., Holden, D.W., Normark, S., Rappuoli, R., and Henriques-Normark, B. (2006) A pneumococcal pilus influences virulence and host inflammatory responses. *Proc Natl Acad Sci U S A* **103**: 2857-2862.

Bayer, A.S., McNamara, P., Yeaman, M.R., Lucindo, N., Jones, T., Cheung, A.L., Sahl, H.G., and Proctor, R.A. (2006) Transposon disruption of the complex I NADH oxidoreductase gene (snoD) in Staphylococcus aureus is associated with reduced susceptibility to the microbicidal activity of thrombin-induced platelet microbicidal protein 1. *J Bacteriol* **188**: 211-222.

Bender, M.H., and Weiser, J.N. (2006) The atypical amino-terminal LPNTG-containing domain of the pneumococcal human IgA1-specific protease is required for proper enzyme localization and function. *Mol Microbiol* **61**: 526-543.

Bensing, B.A., and Sullam, P.M. (2002) An accessory sec locus of Streptococcus gordonii is required for export of the surface protein GspB and for normal levels of binding to human platelets. *Mol Microbiol* **44**: 1081-1094.

Bensing, B.A., Gibson, B.W., and Sullam, P.M. (2004a) The Streptococcus gordonii platelet binding protein GspB undergoes glycosylation independently of export. *J Bacteriol* **186**: 638-645.

Bensing, B.A., Lopez, J.A., and Sullam, P.M. (2004b) The Streptococcus gordonii surface proteins GspB and Hsa mediate binding to sialylated carbohydrate epitopes on the platelet membrane glycoprotein Ibalpha. *Infect Immun* **72**: 6528-6537.

Bensing, B.A., Takamatsu, D., and Sullam, P.M. (2005) Determinants of the streptococcal surface glycoprotein GspB that facilitate export by the accessory Sec system. *Mol Microbiol* **58**: 1468-1481.

Bensing, B.A., Siboo, I.R., and Sullam, P.M. (2007) Glycine residues in the hydrophobic core of the GspB signal sequence route export toward the accessory Sec pathway. *J Bacteriol* **189**: 3846-3854.

Bentley, S.D., Aanensen, D.M., Mavroidi, A., Saunders, D., Rabbinowitsch, E., Collins, M., Donohoe, K., Harris, D., Murphy, L., Quail, M.A., Samuel, G., Skovsted, I.C., Kaltoft, M.S., Barrell, B., Reeves, P.R., Parkhill, J., and Spratt, B.G. (2006) Genetic Analysis of the Capsular Biosynthetic Locus from All 90 Pneumococcal Serotypes. *PLoS Genetics* **2**: e31.

Bernard, P., Kezdy, K.E., Van Melderen, L., Steyaert, J., Wyns, L., Pato, M.L., Higgins, P.N., and Couturier, M. (1993) The F plasmid CcdB protein induces efficient ATP-dependent DNA cleavage by gyrase. *J Mol Biol* **234**: 534-541.

Berry, A.M., and Paton, J.C. (2000) Additive attenuation of virulence of Streptococcus pneumoniae by mutation of the genes encoding pneumolysin and other putative pneumococcal virulence proteins. *Infect Immun* **68**: 133-140.

Biswas, I., Gruss, A., Ehrlich, S.D., and Maguin, E. (1993) High-efficiency gene inactivation and replacement system for gram-positive bacteria. *J Bacteriol* **175**: 3628-3635.

Bjarnsholt, T., Jensen, P.O., Burmolle, M., Hentzer, M., Haagensen, J.A., Hougen, H.P., Calum, H., Madsen, K.G., Moser, C., Molin, S., Hoiby, N., and Givskov, M. (2005) Pseudomonas aeruginosa tolerance to tobramycin, hydrogen peroxide and polymorphonuclear leukocytes is quorum-sensing dependent. *Microbiology* **151**: 373-383.

Black, S., Shinefield, H., Fireman, B., Lewis, E., Ray, P., Hansen, J.R., Elvin, L., Ensor, K.M., Hackell, J., Siber, G., Malinoski, F., Madore, D., Chang, I., Kohberger, R., Watson, W., Austrian, R., and Edwards, K. (2000) Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Northern California Kaiser Permanente Vaccine Study Center Group. *Pediatr Infect Dis J* **19**: 187-195.

Blue, C.E., Paterson, G.K., Kerr, A.R., Berge, M., Claverys, J.P., and Mitchell, T.J. (2003) ZmpB, a novel virulence factor of Streptococcus pneumoniae that induces tumor necrosis factor alpha production in the respiratory tract. *Infect Immun* **71**: 4925-4935.

Bogaert, D., De Groot, R., and Hermans, P.W. (2004a) Streptococcus pneumoniae colonisation: the key to pneumococcal disease. *Lancet Infect Dis* **4**: 144-154.

Bogaert, D., van Belkum, A., Sluijter, M., Luijendijk, A., de Groot, R., Rumke, H.C., Verbrugh, H.A., and Hermans, P.W. (2004b) Colonisation by Streptococcus pneumoniae and Staphylococcus aureus in healthy children. *Lancet* **363**: 1871-1872.

Bohr, V., Rasmussen, N., Hansen, B., Gade, A., Kjersem, H., Johnsen, N., and Paulson, O. (1985) Pneumococcal meningitis: an evaluation of prognostic

factors in 164 cases based on mortality and on a study of lasting sequelae. *J Infect* **10**: 143-157.
Boulnois, G.J. (1992) Pneumococcal proteins and the pathogenesis of disease caused by Streptococcus pneumoniae. *J Gen Microbiol* **138**: 249-259.
Braunstein, M., Brown, A.M., Kurtz, S., and Jacobs, W.R., Jr. (2001) Two Nonredundant SecA Homologues Function in Mycobacteria. *J. Bacteriol.* **183**: 6979-6990.
Braunstein, M., Espinosa, B.J., Chan, J., Belisle, J.T., and Jacobs, W.R., Jr. (2003) SecA2 functions in the secretion of superoxide dismutase A and in the virulence of Mycobacterium tuberculosis. *Mol Microbiol* **48**: 453-464.
Briles, D.E., Yother, J., and McDaniel, L.S. (1988) Role of pneumococcal surface protein A in the virulence of Streptococcus pneumoniae. *Rev Infect Dis* **10 Suppl 2**: S372-374.
Brooks-Walter, A., Briles, D.E., and Hollingshead, S.K. (1999) The pspC gene of Streptococcus pneumoniae encodes a polymorphic protein, PspC, which elicits cross-reactive antibodies to PspA and provides immunity to pneumococcal bacteremia. *Infect Immun* **67**: 6533-6542.
Brueggemann, A.B., Griffiths, D.T., Meats, E., Peto, T., Crook, D.W., and Spratt, B.G. (2003) Clonal relationships between invasive and carriage Streptococcus pneumoniae and serotype- and clone-specific differences in invasive disease potential. *J Infect Dis* **187**: 1424-1432.
Brueggemann, A.B., Pai, R., Crook, D.W., and Beall, B. (2007) Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS Pathog* **3**: e168.
Bu, S., Li, Y., Zhou, M., Azadin, P., Zeng, M., Fives-Taylor, P., and Wu, H. (2008) Interaction between two putative glycosyltransferases is required for glycosylation of a serine-rich streptococcal adhesin. *J Bacteriol* **190**: 1256-1266.
Burnette-Curley, D., Wells, V., Viscount, H., Munro, C.L., Fenno, J.C., Fives-Taylor, P., and Macrina, F.L. (1995) FimA, a major virulence factor associated with Streptococcus parasanguis endocarditis. *Infect Immun* **63**: 4669-4674.
Cartee, R.T., Forsee, W.T., Jensen, J.W., and Yother, J. (2001) Expression of the Streptococcus pneumoniae type 3 synthase in Escherichia coli. Assembly of type 3 polysaccharide on a lipid primer. *J Biol Chem* **276**: 48831-48839.
Cartwright, K. (2002) Pneumococcal disease in western Europe: burden of disease, antibiotic resistance and management. *Eur J Pediatr* **161**: 188-195.
Caspers, M., and Freudl, R. (2008) Corynebacterium glutamicum possesses two secA homologous genes that are essential for viability. *Arch Microbiol* **189**: 605-610.
Chen, H., Ma, Y., Yang, J., O'Brien, C.J., Lee, S.L., Mazurkiewicz, J.E., Haataja, S., Yan, J.H., Gao, G.F., and Zhang, J.R. (2007a) Genetic requirement for pneumococcal ear infection. *PLoS ONE* **3**: e2950.

Chen, Q., Wu, H., and Fives-Taylor, P.M. (2002) Construction of a novel transposon mutagenesis system useful in the isolation of Streptococcus parasanguis mutants defective in Fap1 glycosylation. *Infect Immun* **70**: 6534-6540.

Chen, Q., Wu, H., and Fives-Taylor, P.M. (2004) Investigating the role of secA2 in secretion and glycosylation of a fimbrial adhesin in Streptococcus parasanguis FW213. *Mol Microbiol* **53**: 843-856.

Chen, Q., Wu, H., Kumar, R., Peng, Z., and Fives-Taylor, P.M. (2006) SecA2 is distinct from SecA in immunogenic specificity, subcellular distribution and requirement for membrane anchoring in Streptococcus parasanguis. *FEMS Microbiol Lett* **264**: 174-181.

Chen, Q., Sun, B., Wu, H., Peng, Z., and Fives-Taylor, P.M. (2007b) Differential roles of individual domains in selection of secretion route of a Streptococcus parasanguinis serine-rich adhesin, Fap1. *J Bacteriol* **189**: 7610-7617.

Chi, F., Nolte, O., Bergmann, C., Ip, M., and Hakenbeck, R. (2007) Crossing the barrier: evolution and spread of a major class of mosaic pbp2x in Streptococcus pneumoniae, S. mitis and S. oralis. *Int J Med Microbiol* **297**: 503-512.

Claverys, J.P., and Havarstein, L.S. (2002) Extracellular-peptide control of competence for genetic transformation in Streptococcus pneumoniae. *Front Biosci* **7**: d1798-1814.

Claverys, J.P., and Havarstein, L.S. (2007) Cannibalism and fratricide: mechanisms and raisons d'etre. *Nat Rev Microbiol* **5**: 219-229.

Coffey, T.J., Dowson, C.G., Daniels, M., Zhou, J., Martin, C., Spratt, B.G., and Musser, J.M. (1991) Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthetic genes, in natural populations of Streptococcus pneumoniae. *Mol Microbiol* **5**: 2255-2260.

Coffey, T.J., Enright, M.C., Daniels, M., Morona, J.K., Morona, R., Hryniewicz, W., Paton, J.C., and Spratt, B.G. (1998) Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of Streptococcus pneumoniae. *Mol Microbiol* **27**: 73-83.

Collier, D.N., Strobel, S.M., and Bassford, P.J., Jr. (1990) SecB-independent export of Escherichia coli ribose-binding protein (RBP): some comparisons with export of maltose-binding protein (MBP) and studies with RBP-MBP hybrid proteins. *J Bacteriol* **172**: 6875-6884.

Cortese, M.M., Wolff, M., Almeido-Hill, J., Reid, R., Ketcham, J., and Santosham, M. (1992) High incidence rates of invasive pneumococcal disease in the White Mountain Apache population. *Arch Intern Med* **152**: 2277-2282.

Cossart, P. (1998) Interactions of the bacterial pathogen Listeria monocytogenes with mammalian cells: bacterial factors, cellular ligands, and signaling. *Folia Microbiol (Praha)* **43**: 291-303.

Costerton, J.W. (1995) Overview of microbial biofilms. *J Ind Microbiol* **15**: 137-140.

Costerton, J.W., Stewart, P.S., and Greenberg, E.P. (1999) Bacterial biofilms: a common cause of persistent infections. *Science* **284**: 1318-1322.

Crain, M.J., Waltman, W.D., 2nd, Turner, J.S., Yother, J., Talkington, D.F., McDaniel, L.S., Gray, B.M., and Briles, D.E. (1990) Pneumococcal surface protein A (PspA) is serologically highly variable and is expressed by all clinically important capsular serotypes of Streptococcus pneumoniae. *Infect Immun* **58**: 3293-3299.

Cundell, D.R., and Tuomanen, E.I. (1994) Receptor specificity of adherence of Streptococcus pneumoniae to human type-II pneumocytes and vascular endothelial cells in vitro. *Microb Pathog* **17**: 361-374.

Dagkessamanskaia, A., Moscoso, M., Henard, V., Guiral, S., Overweg, K., Reuter, M., Martin, B., Wells, J., and Claverys, J.P. (2004) Interconnection of competence, stress and CiaR regulons in Streptococcus pneumoniae: competence triggers stationary phase autolysis of ciaR mutant cells. *Mol Microbiol* **51**: 1071-1086.

Danese, P.N., and Silhavy, T.J. (1998) Targeting and assembly of periplasmic and outer-membrane proteins in Escherichia coli. *Annu Rev Genet* **32**: 59-94.

Dave, S., Carmicle, S., Hammerschmidt, S., Pangburn, M.K., and McDaniel, L.S. (2004) Dual roles of PspC, a surface protein of Streptococcus pneumoniae, in binding human secretory IgA and factor H. *J Immunol* **173**: 471-477.

Daws, M.R., Sullam, P.M., Niemi, E.C., Chen, T.T., Tchao, N.K., and Seaman, W.E. (2003) Pattern recognition by TREM-2: binding of anionic ligands. *J Immunol* **171**: 594-599.

Dillard, J.P., Vandersea, M.W., and Yother, J. (1995) Characterization of the cassette containing genes for type 3 capsular polysaccharide biosynthesis in Streptococcus pneumoniae. *J Exp Med* **181**: 973-983.

Dintilhac, A., Alloing, G., Granadel, C., and Claverys, J.P. (1997) Competence and virulence of Streptococcus pneumoniae: Adc and PsaA mutants exhibit a requirement for Zn and Mn resulting from inactivation of putative ABC metal permeases. *Mol Microbiol* **25**: 727-739.

Donlan, R.M., and Costerton, J.W. (2002) Biofilms: Survival Mechanisms of Clinically Relevant Microorganisms. *Clin. Microbiol. Rev.* **15**: 167-193.

Donlan, R.M., Piede, J.A., Heyes, C.D., Sanii, L., Murga, R., Edmonds, P., El-Sayed, I., and El-Sayed, M.A. (2004) Model system for growing and quantifying Streptococcus pneumoniae biofilms in situ and in real time. *Appl Environ Microbiol* **70**: 4980-4988.

Douglas, C.W., Heath, J., Hampton, K.K., and Preston, F.E. (1993) Identity of viridans streptococci isolated from cases of infective endocarditis. *J Med Microbiol* **39**: 179-182.

Dowson, C.G., Coffey, T.J., Kell, C., and Whiley, R.A. (1993) Evolution of penicillin resistance in Streptococcus pneumoniae; the role of Streptococcus mitis in the formation of a low affinity PBP2B in S. pneumoniae. *Mol Microbiol* **9**: 635-643.

Dowson, C.G., Coffey, T.J., and Spratt, B.G. (1994) Origin and molecular epidemiology of penicillin-binding-protein-mediated resistance to beta-lactam antibiotics. *Trends Microbiol* **2**: 361-366.

Duong, F., and Wickner, W. (1997) The SecDFyajC domain of preprotein translocase controls preprotein movement by regulating SecA membrane cycling. *Embo J* **16**: 4871-4879.

Durack, D.T. (1975) Experimental bacterial endocarditis. IV. Structure and evolution of very early lesions. *J Pathol* **115**: 81-89.

Durack, D.T. (1995) Prevention of infective endocarditis. *N Engl J Med* **332**: 38-44.

Edwards, A.M., Grossman, T.J., and Rudney, J.D. (2006) Fusobacterium nucleatum transports noninvasive Streptococcus cristatus into human epithelial cells. *Infect Immun* **74**: 654-662.

Eichler, J., Brunner, J., and Wickner, W. (1997) The protease-protected 30 kDa domain of SecA is largely inaccessible to the membrane lipid phase. *Embo J* **16**: 2188-2196.

Engelhard, D., Pomeranz, S., Gallily, R., Strauss, N., and Tuomanen, E. (1997) Serotype-related differences in inflammatory response to Streptococcus pneumoniae in experimental meningitis. *J Infect Dis* **175**: 979-982.

Enright, M.C., and Spratt, B.G. (1998) A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. *Microbiology* **144 (Pt 11)**: 3049-3060.

Fekkes, P., van der Does, C., and Driessen, A.J. (1997) The molecular chaperone SecB is released from the carboxy-terminus of SecA during initiation of precursor protein translocation. *Embo J* **16**: 6105-6113.

Feldman, C., Mitchell, T.J., Andrew, P.W., Boulnois, G.J., Read, R.C., Todd, H.C., Cole, P.J., and Wilson, R. (1990) The effect of Streptococcus pneumoniae pneumolysin on human respiratory epithelium in vitro. *Microb Pathog* **9**: 275-284.

Feldman, C., Cockeran, R., Jedrzejas, M.J., Mitchell, T.J., and Anderson, R. (2007) Hyaluronidase augments pneumolysin-mediated injury to human ciliated epithelium. *Int J Infect Dis* **11**: 11-15.

Fine, D.P. (1975) Pneumococcal type-associated variability in alternate complement pathway activation. *Infect Immun* **12**: 772-778.

Finlay, B.B., and Falkow, S. (1989) Common themes in microbial pathogenicity. *Microbiol Rev* **53**: 210-230.

Finlay, B.B., and Falkow, S. (1997) Common themes in microbial pathogenicity revisited. *Microbiol Mol Biol Rev* **61**: 136-169.

Froeliger, E.H., and Fives-Taylor, P. (2001) Streptococcus parasanguis Fimbria-Associated Adhesin Fap1 Is Required for Biofilm Formation. *Infect. Immun.* **69**: 2512-2519.

Garau, J., and Calbo, E. (2007) Capsular types and predicting patient outcomes in pneumococcal bacteremia. *Clin Infect Dis* **45**: 52-54.

Garcia-Patrone, M., and Tandecarz, J.S. (1995) A glycoprotein multimer from Bacillus thuringiensis sporangia: dissociation into subunits and sugar composition. *Mol Cell Biochem* **145**: 29-37.

Garcia, E., Llull, D., Munoz, R., Mollerach, M., and Lopez, R. (2000) Current trends in capsular polysaccharide biosynthesis of Streptococcus pneumoniae. *Res Microbiol* **151**: 429-435.

Gibbons, R.J., and Houte, J.V. (1975) Bacterial adherence in oral microbial ecology. *Annu Rev Microbiol* **29**: 19-44.

Gillespie, S.H. (1989) Aspects of pneumococcal infection including bacterial virulence, host response and vaccination. *J Med Microbiol* **28**: 237-248.

Goebel, W., and Adams, M. (1943) The immunological properties of the heterophile antigen and somatic polysaccharide of pneumococcus. *J Exp Med* **77**: 435-449.

Gosink, K.K., Mann, E.R., Guglielmo, C., Tuomanen, E.I., and Masure, H.R. (2000) Role of novel choline binding proteins in virulence of Streptococcus pneumoniae. *Infect Immun* **68**: 5690-5695.

Gray, B.M., Converse, G.M., 3rd, Huhta, N., Johnston, R.B., Jr., Pichichero, M.E., Schiffman, G., and Dillon, H.C., Jr. (1981) Epidemiologic studies of Streptococcus pneumoniae in infants: antibody response to nasopharyngeal carriage of types 3, 19, and 23. *J Infect Dis* **144**: 312-318.

Gray, B.M., Turner, M.E., and Dillon, H.C., Jr. (1982) Epidemiologic studies of Streptococcus pneumoniae in infants. The effects of season and age on pneumococcal acquisition and carriage in the first 24 months of life. *Am J Epidemiol* **116**: 692-703.

Griffith, F. (1928) The significance of pneumococcal types. *Journal of Hygene* **27**: 113-159.

Guckian, J.C., Christensen, G.D., and Fine, D.P. (1980) The role of opsonins in recovery from experimental pneumococcal pneumonia. *J Infect Dis* **142**: 175-190.

Guiral, S., Mitchell, T.J., Martin, B., and Claverys, J.P. (2005) Competence-programmed predation of noncompetent cells in the human pathogen Streptococcus pneumoniae: genetic requirements. *Proc Natl Acad Sci U S A* **102**: 8710-8715.

Habash, M., and Reid, G. (1999) Microbial biofilms: their development and significance for medical device-related infections. *J Clin Pharmacol* **39**: 887-898.

Hakenbeck, R., Balmelle, N., Weber, B., Gardes, C., Keck, W., and de Saizieu, A. (2001) Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of Streptococcus pneumoniae. *Infect Immun* **69**: 2477-2486.

Hall-Stoodley, L., and Stoodley, P. (2005) Biofilm formation and dispersal and the transmission of human pathogens. *Trends Microbiol* **13**: 7-10.

Hammerschmidt, S., Muller, A., Sillmann, H., Muhlenhoff, M., Borrow, R., Fox, A., van Putten, J., Zollinger, W.D., Gerardy-Schahn, R., and Frosch, M. (1996) Capsule phase variation in Neisseria meningitidis serogroup B by slipped-strand mispairing in the polysialyltransferase gene (siaD): correlation with bacterial invasion and the outbreak of meningococcal disease. *Mol Microbiol* **20**: 1211-1220.

Hammerschmidt, S., Talay, S.R., Brandtzaeg, P., and Chhatwal, G.S. (1997) SpsA, a novel pneumococcal surface protein with specific binding to secretory immunoglobulin A and secretory component. *Mol Microbiol* **25**: 1113-1124.

Hammerschmidt, S., Bethe, G., Remane, P.H., and Chhatwal, G.S. (1999) Identification of pneumococcal surface protein A as a lactoferrin-binding protein of Streptococcus pneumoniae. *Infect Immun* **67**: 1683-1687.

Hammerschmidt, S., Wolff, S., Hocke, A., Rosseau, S., Muller, E., and Rohde, M. (2005) Illustration of Pneumococcal Polysaccharide Capsule during Adherence and Invasion of Epithelial Cells. *Infect. Immun.* **73**: 4653-4667.

Hanage, W.P., Auranen, K., Syrjanen, R., Herva, E., Makela, P.H., Kilpi, T., and Spratt, B.G. (2004) Ability of pneumococcal serotypes and clones to cause acute otitis media: implications for the prevention of otitis media by conjugate vaccines. *Infect Immun* **72**: 76-81.

Hanage, W.P., Kaijalainen, T., Herva, E., Saukkoriipi, A., Syrjanen, R., and Spratt, B.G. (2005) Using multilocus sequence data to define the pneumococcus. *J Bacteriol* **187**: 6223-6230.

Handley, P., Coykendall, A., Beighton, D., Hardie, J.M., and Whiley, R.A. (1991) Streptococcus crista sp. nov., a viridans streptococcus with tufted fibrils, isolated from the human oral cavity and throat. *Int J Syst Bacteriol* **41**: 543-547.

Handley, P.S., Correia, F.F., Russell, K., Rosan, B., and DiRienzo, J.M. (2005) Association of a novel high molecular weight, serine-rich protein (SrpA) with fibril-mediated adhesion of the oral biofilm bacterium Streptococcus cristatus. *Oral Microbiol Immunol* **20**: 131-140.

Hanein, D., Matlack, K.E., Jungnickel, B., Plath, K., Kalies, K.U., Miller, K.R., Rapoport, T.A., and Akey, C.W. (1996) Oligomeric rings of the Sec61p complex induced by ligands required for protein translocation. *Cell* **87**: 721-732.

Hardy, G.G., Magee, A.D., Ventura, C.L., Caimano, M.J., and Yother, J. (2001) Essential role for cellular phosphoglucomutase in virulence of type 3 Streptococcus pneumoniae. *Infect Immun* **69**: 2309-2317.

Hartley, J.L., Temple, G.F., and Brasch, M.A. (2000) DNA cloning using in vitro site-specific recombination. *Genome Res* **10**: 1788-1795.

Hassett, D.J., Ma, J.F., Elkins, J.G., McDermott, T.R., Ochsner, U.A., West, S.E., Huang, C.T., Fredericks, J., Burnett, S., Stewart, P.S., McFeters, G., Passador, L., and Iglewski, B.H. (1999) Quorum sensing in Pseudomonas aeruginosa controls expression of catalase and superoxide dismutase genes and mediates biofilm susceptibility to hydrogen peroxide. *Mol Microbiol* **34**: 1082-1093.

Hausdorff, W.P., Bryant, J., Kloek, C., Paradiso, P.R., and Siber, G.R. (2000a) The Contribution of Specific Pneumococcal Serogroups to Different Disease Manifestations: Implications for Conjugate Vaccine Formulation and Use, Part II. *Clinical Infectious Diseases* **30**: 122-140.

Hausdorff, W.P., Bryant, J., Paradiso, P.R., and Siber, G.R. (2000b) Which Pneumococcal Serogroups Cause the Most Invasive Disease: Implications for Conjugate Vaccine Formulation and Use, Part I. *Clinical Infectious Diseases* **30**: 100-121.

Hausdorff, W.P., Feikin, D.R., and Klugman, K.P. (2005) Epidemiological differences among pneumococcal serotypes. *The Lancet Infectious Diseases* **5**: 83-93.

Hava, D.L., and Camilli, A. (2002) Large-scale identification of serotype 4 Streptococcus pneumoniae virulence factors. *Mol Microbiol* **45**: 1389-1406.

Havarstein, L.S., Coomaraswamy, G., and Morrison, D.A. (1995) An unmodified heptadecapeptide pheromone induces competence for genetic transformation in Streptococcus pneumoniae. *Proc Natl Acad Sci U S A* **92**: 11140-11144.

Havarstein, L.S., Gaustad, P., Nes, I.F., and Morrison, D.A. (1996) Identification of the streptococcal competence-pheromone receptor. *Mol Microbiol* **21**: 863-869.

Havarstein, L.S., Martin, B., Johnsborg, O., Granadel, C., and Claverys, J.P. (2006) New insights into the pneumococcal fratricide: relationship to clumping and identification of a novel immunity factor. *Mol Microbiol* **59**: 1297-1307.

Heilmann, C., Niemann, S., Sinha, B., Herrmann, M., Kehrel, B.E., and Peters, G. (2004) Staphylococcus aureus fibronectin-binding protein (FnBP)-mediated adherence to platelets, and aggregation of platelets induced by FnBPA but not by FnBPB. *J Infect Dis* **190**: 321-329.

Henrichsen, J. (1995) Six newly recognized types of Streptococcus pneumoniae. *J Clin Microbiol* **33**: 2759-2762.

Hiller, N.L., Janto, B., Hogg, J.S., Boissy, R., Yu, S., Powell, E., Keefe, R., Ehrlich, N.E., Shen, K., Hayes, J., Barbadora, K., Klimke, W., Dernovoy, D., Tatusova, T., Parkhill, J., Bentley, S.D., Post, J.C., Ehrlich, G.D., and Hu, F.Z. (2007) Comparative Genomic Analyses of Seventeen Streptococcus pneumoniae Strains: Insights into the Pneumococcal Supragenome. *J. Bacteriol.* **189**: 8186-8195.

Hirst, R.A., Kadioglu, A., O'Callaghan, C., and Andrew, P.W. (2004) The role of pneumolysin in pneumococcal pneumonia and meningitis. *Clin Exp Immunol* **138**: 195-201.

Hirst, R.A., Gosai, B., Rutman, A., Guerin, C.J., Nicotera, P., Andrew, P.W., and O'Callaghan, C. (2008) Streptococcus pneumoniae deficient in pneumolysin or autolysin has reduced virulence in meningitis. *J Infect Dis* **197**: 744-751.

Hoberman, A., Greenberg, D.P., Paradise, J.L., Rockette, H.E., Lave, J.R., Kearney, D.H., Colborn, D.K., Kurs-Lasky, M., Haralam, M.A., Byers, C.J., Zoffel, L.M., Fabian, I.A., Bernard, B.S., and Kerr, J.D. (2003) Effectiveness of Inactivated Influenza Vaccine in Preventing Acute Otitis Media in Young Children: A Randomized Controlled Trial. *JAMA* **290**: 1608-1616.

Horn, C., Namane, A., Pescher, P., Riviere, M., Romain, F., Puzo, G., Barzu, O., and Marchal, G. (1999) Decreased Capacity of Recombinant 45/47-kDa Molecules (Apa) of Mycobacterium tuberculosis to Stimulate T Lymphocyte Responses Related to Changes in Their Mannosylation Pattern. *J. Biol. Chem.* **274**: 32023-32030.

Hoskins, J., Alborn, W.E., Jr., Arnold, J., Blaszczak, L.C., Burgett, S., DeHoff, B.S., Estrem, S.T., Fritz, L., Fu, D.J., Fuller, W., Geringer, C., Gilmour, R., Glass, J.S., Khoja, H., Kraft, A.R., Lagace, R.E., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S.M., McHenney, M., McLeaster, K., Mundy, C.W., Nicas, T.I., Norris, F.H., O'Gara, M., Peery, R.B., Robertson, G.T., Rockey, P., Sun, P.M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C.A., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R., Jr., Skatrud, P.L., and Glass, J.I. (2001) Genome of the bacterium Streptococcus pneumoniae strain R6. *J Bacteriol* **183**: 5709-5717.

Hostetter, M.K. (1986) Serotypic variations among virulent pneumococci in deposition and degradation of covalently bound C3b: implications for phagocytosis and antibody production. *J Infect Dis* **153**: 682-693.

Houldsworth, S., Andrew, P.W., and Mitchell, T.J. (1994) Pneumolysin stimulates production of tumor necrosis factor alpha and interleukin-1 beta by human mononuclear phagocytes. *Infect Immun* **62**: 1501-1503.

Hui, F.M., and Morrison, D.A. (1991) Genetic transformation in Streptococcus pneumoniae: nucleotide sequence analysis shows comA, a gene required for competence induction, to be a member of the bacterial ATP-dependent transport protein family. *J Bacteriol* **173**: 372-381.

Ibrahim, Y.M., Kerr, A.R., Silva, N.A., and Mitchell, T.J. (2005) Contribution of the ATP-dependent protease ClpCP to the autolysis and virulence of Streptococcus pneumoniae. *Infect Immun* **73**: 730-740.

Inga Benz, M.A.S. (2002) Never say never again: protein glycosylation in pathogenic bacteria. *Molecular Microbiology* **45**: 267-276.

Ito, K., Wittekind, M., Nomura, M., Shiba, K., Yura, T., Miura, A., and Nashimoto, H. (1983) A temperature-sensitive mutant of E. coli exhibiting slow processing of exported proteins. *Cell* **32**: 789-797.

Jakubovics, N.S., Kerrigan, S.W., Nobbs, A.H., Stromberg, N., van Dolleweerd, C.J., Cox, D.M., Kelly, C.G., and Jenkinson, H.F. (2005) Functions of cell surface-anchored antigen I/II family and Hsa polypeptides in interactions of Streptococcus gordonii with host receptors. *Infect Immun* **73**: 6629-6638.

Jedrzejas, M.J. (2001) Pneumococcal Virulence Factors: Structure and Function. *Microbiol. Mol. Biol. Rev.* **65**: 187-207.

Jenkins, S.G., Brown, S.D., and Farrell, D.J. (2008) Trends in antibacterial resistance among Streptococcus pneumoniae isolated in the USA: update from PROTEKT US Years 1-4. *Ann Clin Microbiol Antimicrob* **7**: 1.

Johnsborg, O., Eldholm, V., and Havarstein, L.S. (2007) Natural genetic transformation: prevalence, mechanisms and function. *Res Microbiol* **158**: 767-778.

Johnsborg, O., Eldholm, V., Bjornstad, M.L., and Havarstein, L.S. (2008) A predatory mechanism dramatically increases the efficiency of lateral gene transfer in Streptococcus pneumoniae and related commensal species. *Mol Microbiol* **69**: 245-253.

Johnson, A.P., Warner, M., Broughton, K., James, D., Efsratiou, A., George, R.C., and Livermore, D.M. (2001) Antibiotic susceptibility of streptococci and related genera causing endocarditis: analysis of UK reference laboratory referrals, January 1996 to March 2000. *Bmj* **322**: 395-396.

Johnston, R.B., Jr. (1991) Pathogenesis of pneumococcal pneumonia. *Rev Infect Dis* **13 Suppl 6**: S509-517.

Joly, J.C., and Wickner, W. (1993) The SecA and SecY subunits of translocase are the nearest neighbors of a translocating preprotein, shielding it from phospholipids. *Embo J* **12**: 255-263.

Kalin, M. (1998) Pneumococcal serotypes and their clinical relevance. *Thorax* **53**: 159-162.

Kausmally, L., Johnsborg, O., Lunde, M., Knutsen, E., and Havarstein, L.S. (2005) Choline-binding protein D (CbpD) in Streptococcus pneumoniae is essential for competence-induced cell lysis. *J Bacteriol* **187**: 4338-4345.

Kawamura, T., and Shockman, G.D. (1983) Purification and some properties of the endogenous, autolytic N-acetylmuramoylhydrolase of Streptococcus faecium, a bacterial glycoenzyme. *J Biol Chem* **258**: 9514-9521.

Kerrigan, S.W., Jakubovics, N.S., Keane, C., Maguire, P., Wynne, K., Jenkinson, H.F., and Cox, D. (2007) Role of Streptococcus gordonii surface proteins SspA/SspB and Hsa in platelet function. *Infect Immun* **75**: 5740-5747.

Kihara, A., Akiyama, Y., and Ito, K. (1995) FtsH is required for proteolytic elimination of uncomplexed forms of SecY, an essential protein translocase subunit. *Proc Natl Acad Sci U S A* **92**: 4532-4536.

Kim, J.O., and Weiser, J.N. (1998) Association of intrastrain phase variation in quantity of capsular polysaccharide and teichoic acid with the virulence of Streptococcus pneumoniae. *J Infect Dis* **177**: 368-377.

Kolenbrander, P.E., and London, J. (1993) Adhere today, here tomorrow: oral bacterial adherence. *J Bacteriol* **175**: 3247-3252.

Kolenbrander, P.E. (2000) Oral microbial communities: biofilms, interactions, and genetic systems. *Annu Rev Microbiol* **54**: 413-437.

Kolkman, M.A., van der Zeijst, B.A., and Nuijten, P.J. (1998) Diversity of capsular polysaccharide synthesis gene clusters in Streptococcus pneumoniae. *J Biochem* **123**: 937-945.

Kolter, R. (2005) Surfacing views of biofilm biology. *Trends Microbiol* **13**: 1-2.

Kuboniwa, M., Tribble, G.D., James, C.E., Kilic, A.O., Tao, L., Herzberg, M.C., Shizukuishi, S., and Lamont, R.J. (2006) Streptococcus gordonii utilizes several distinct gene functions to recruit Porphyromonas gingivalis into a mixed community. *Mol Microbiol* **60**: 121-139.

Kumamoto, C.A., and Beckwith, J. (1985) Evidence for specificity at an early step in protein export in Escherichia coli. *J Bacteriol* **163**: 267-274.

Kuo, C., Takahashi, N., Swanson, A.F., Ozeki, Y., and Hakomori, S. (1996) An N-linked high-mannose type oligosaccharide, expressed at the major outer membrane protein of Chlamydia trachomatis, mediates attachment and infectivity of the microorganism to HeLa cells. *J Clin Invest* **98**: 2813-2818.

Lancy, P., Jr., Appelbaum, B., Holt, S.C., and Rosan, B. (1980) Quantitative in vitro assay for "corncob" formation. *Infect Immun* **29**: 663-670.

Lancy, P., Jr., Dirienzo, J.M., Appelbaum, B., Rosan, B., and Holt, S.C. (1983) Corncob formation between Fusobacterium nucleatum and Streptococcus sanguis. *Infect Immun* **40**: 303-309.

Lau, G.W., Haataja, S., Lonetto, M., Kensit, S.E., Marra, A., Bryant, A.P., McDevitt, D., Morrison, D.A., and Holden, D.W. (2001) A functional genomic analysis of type 3 Streptococcus pneumoniae virulence. *Mol Microbiol* **40**: 555-571.

Lee, M.S., and Morrison, D.A. (1999) Identification of a new regulator in Streptococcus pneumoniae linking quorum sensing to competence for genetic transformation. *J Bacteriol* **181**: 5004-5016.

Lee, V.T., and Schneewind, O. (2001) Protein secretion and the pathogenesis of bacterial infections. *Genes Dev* **15**: 1725-1752.

Leimkugel, J., Adams Forgor, A., Gagneux, S., Pfluger, V., Flierl, C., Awine, E., Naegeli, M., Dangy, J.P., Smith, T., Hodgson, A., and Pluschke, G. (2005) An outbreak of serotype 1 Streptococcus pneumoniae meningitis in northern Ghana with features that are characteristic of Neisseria meningitidis meningitis epidemics. *J Infect Dis* **192**: 192-199.

LeMessurier, K.S., Ogunniyi, A.D., and Paton, J.C. (2006) Differential expression of key pneumococcal virulence genes in vivo. *Microbiology* **152**: 305-311.

LeMieux, J., Hava, D.L., Basset, A., and Camilli, A. (2006) RrgA and RrgB are components of a multisubunit pilus encoded by the Streptococcus pneumoniae rlrA pathogenicity islet. *Infect Immun* **74**: 2453-2456.

Lenz, L.L., and Portnoy, D.A. (2002) Identification of a second Listeria secA gene associated with protein secretion and the rough phenotype. *Mol Microbiol* **45**: 1043-1056.

Li, Y.H., Tang, N., Aspiras, M.B., Lau, P.C., Lee, J.H., Ellen, R.P., and Cvitkovitch, D.G. (2002) A quorum-sensing signaling system essential for genetic competence in Streptococcus mutans is involved in biofilm formation. *J Bacteriol* **184**: 2699-2708.

Lloyd-Evans, N., O'Dempsey, T.J., Baldeh, I., Secka, O., Demba, E., Todd, J.E., McArdle, T.F., Banya, W.S., and Greenwood, B.M. (1996) Nasopharyngeal carriage of pneumococci in Gambian children and in their families. *Pediatr Infect Dis J* **15**: 866-871.

Llull, D., Munoz, R., Lopez, R., and Garcia, E. (1999) A single gene (tts) located outside the cap locus directs the formation of Streptococcus pneumoniae type 37 capsular polysaccharide. Type 37 pneumococci are natural, genetically binary strains. *J Exp Med* **190**: 241-251.

Loo, C.Y., Corliss, D.A., and Ganeshkumar, N. (2000) Streptococcus gordonii biofilm formation: identification of genes that code for biofilm phenotypes. *J Bacteriol* **182**: 1374-1382.

Lopez, R., Garcia, E., Garcia, P., and Garcia, J.L. (1997) The pneumococcal cell wall degrading enzymes: a modular design to create new lysins? *Microb Drug Resist* **3**: 199-211.

Magee, A.D., and Yother, J. (2001) Requirement for capsule in colonization by Streptococcus pneumoniae. *Infect Immun* **69**: 3755-3761.

Manco, S., Hernon, F., Yesilkaya, H., Paton, J.C., Andrew, P.W., and Kadioglu, A. (2006) Pneumococcal neuraminidases A and B both have essential roles during infection of the respiratory tract and sepsis. *Infect Immun* **74**: 4014-4020.

Mann, B., Orihuela, C., Antikainen, J., Gao, G., Sublett, J., Korhonen, T.K., and Tuomanen, E. (2006) Multifunctional role of choline binding protein G in pneumococcal pathogenesis. *Infect Immun* **74**: 821-829.

McCullers, J.A., and Tuomanen, E.I. (2001) Molecular pathogenesis of pneumococcal pneumonia. *Front Biosci* **6**: D877-889.

McDougal, L.K., Facklam, R., Reeves, M., Hunter, S., Swenson, J.M., Hill, B.C., and Tenover, F.C. (1992) Analysis of multiply antimicrobial-resistant isolates of Streptococcus pneumoniae from the United States. *Antimicrob Agents Chemother* **36**: 2176-2184.

McGee, L., McDougal, L., Zhou, J., Spratt, B.G., Tenover, F.C., George, R., Hakenbeck, R., Hryniewicz, W., Lefevre, J.C., Tomasz, A., and Klugman, K.P. (2001) Nomenclature of Major Antimicrobial-Resistant Clones of Streptococcus pneumoniae Defined by the Pneumococcal Molecular Epidemiology Network. *J. Clin. Microbiol.* **39**: 2565-2571.

McNab, R., and Jenkinson, H.F. (1992) Gene disruption identifies a 290 kDa cell-surface polypeptide conferring hydrophobicity and coaggregation properties in Streptococcus gordonii. *Mol Microbiol* **6**: 2939-2949.

Mescher, M.F., and Strominger, J.L. (1976) Structural (shape-maintaining) role of the cell surface glycoprotein of Halobacterium salinarium. *Proc Natl Acad Sci U S A* **73**: 2687-2691.

Messner, P. (2004) Prokaryotic Glycoproteins: Unexplored but Important. *J. Bacteriol.* **186**: 2517-2519.

Mitchell, C., and Oliver, D. (1993) Two distinct ATP-binding domains are needed to promote protein export by Escherichia coli SecA ATPase. *Mol Microbiol* **10**: 483-497.

Mitchell, J., Siboo, I.R., Takamatsu, D., Chambers, H.F., and Sullam, P.M. (2007) Mechanism of cell surface expression of the Streptococcus mitis platelet binding proteins PblA and PblB. *Mol Microbiol* **64**: 844-857.

Mizrachi Nebenzahl, Y., Porat, N., Lifshitz, S., Novick, S., Levi, A., Ling, E., Liron, O., Mordechai, S., Sahu, R.K., and Dagan, R. (2004) Virulence of Streptococcus pneumoniae may be determined independently of capsular polysaccharide. *FEMS Microbiol Lett* **233**: 147-152.

Morona, J.K., Morona, R., and Paton, J.C. (1999) Comparative Genetics of Capsular Polysaccharide Biosynthesis in Streptococcus pneumoniae Types Belonging to Serogroup 19. *J. Bacteriol.* **181**: 5355-5364.

Mortier-Barriere, I., de Saizieu, A., Claverys, J.P., and Martin, B. (1998) Competence-specific induction of recA is required for full recombination proficiency during transformation in Streptococcus pneumoniae. *Mol Microbiol* **27**: 159-170.

Moscoso, M., and Claverys, J.P. (2004) Release of DNA into the medium by competent Streptococcus pneumoniae: kinetics, mechanism and stability of the liberated DNA. *Mol Microbiol* **54**: 783-794.

Moscoso, M., Garcia, E., and Lopez, R. (2006) Biofilm formation by Streptococcus pneumoniae: role of choline, extracellular DNA, and capsular polysaccharide in microbial accretion. *J Bacteriol* **188**: 7785-7795.

Mouton, C., Reynolds, H.S., and Genco, R.J. (1980) Characterization of tufted streptococci isolated from the "corn cob" configuration of human dental plaque. *Infect Immun* **27**: 235-245.

Muller, M. (2005) Twin-arginine-specific protein export in Escherichia coli. *Res Microbiol* **156**: 131-136.

Munoz, R., Coffey, T.J., Daniels, M., Dowson, C.G., Laible, G., Casal, J., Hakenbeck, R., Jacobs, M., Musser, J.M., Spratt, B.G., and et al. (1991) Intercontinental spread of a multiresistant clone of serotype 23F Streptococcus pneumoniae. *J Infect Dis* **164**: 302-306.

Myers, C., and Gervaix, A. (2007) Streptococcus pneumoniae bacteraemia in children. *Int J Antimicrob Agents* **30 Suppl 1**: S24-28.

Nguyen, T., Ghebrehiwet, B., and Peerschke, E.I. (2000) Staphylococcus aureus protein A recognizes platelet gC1qR/p33: a novel mechanism for staphylococcal interactions with platelets. *Infect Immun* **68**: 2061-2068.

Nielsen, S.V., and Henrichsen, J. (1992) Capsular types of Streptococcus pneumoniae isolated from blood and CSF during 1982-1987. *Clin Infect Dis* **15**: 794-798.

Nishiyama, K., Hanada, M., and Tokuda, H. (1994) Disruption of the gene encoding p12 (SecG) reveals the direct involvement and important function of SecG in the protein translocation of Escherichia coli at low temperature. *Embo J* **13**: 3272-3277.

Nobbs, A.H., Zhang, Y., Khammanivong, A., and Herzberg, M.C. (2007) Streptococcus gordonii Hsa environmentally constrains competitive binding by Streptococcus sanguinis to saliva-coated hydroxyapatite. *J Bacteriol* **189**: 3106-3114.

O'Brien, L., Kerrigan, S.W., Kaw, G., Hogan, M., Penades, J., Litt, D., Fitzgerald, D.J., Foster, T.J., and Cox, D. (2002) Multiple mechanisms for the activation of human platelet aggregation by Staphylococcus aureus: roles for the clumping factors ClfA and ClfB, the serine-aspartate repeat protein SdrE and protein A. *Mol Microbiol* **44**: 1033-1044.

O'Toole, G.A., and Kolter, R. (1998) Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development. *Mol Microbiol* **30**: 295-304.

O'Toole, G.A. (2004) Microbiology: Jekyll or hide? *Nature* **432**: 680-681.

Obert, C., Sublett, J., Kaushal, D., Hinojosa, E., Barton, T., Tuomanen, E.I., and Orihuela, C.J. (2006) Identification of a Candidate Streptococcus pneumoniae core genome and regions of diversity correlated with invasive pneumococcal disease. *Infect Immun* **74**: 4766-4777.

Obregon, V., Garcia, J.L., Garcia, E., Lopez, R., and Garcia, P. (2003) Genome Organization and Molecular Analysis of the Temperate Bacteriophage MM1 of Streptococcus pneumoniae. *J. Bacteriol.* **185**: 2362-2368.

Oggioni, M.R., Trappetti, C., Kadioglu, A., Cassone, M., Iannelli, F., Ricci, S., Andrew, P.W., and Pozzi, G. (2006) Switch from planktonic to sessile life: a major event in pneumococcal pathogenesis. *Mol Microbiol* **61**: 1196-1210.

Ogunniyi, A.D., LeMessurier, K.S., Graham, R.M., Watt, J.M., Briles, D.E., Stroeher, U.H., and Paton, J.C. (2007) Contributions of pneumolysin, pneumococcal surface protein A (PspA), and PspC to pathogenicity of Streptococcus pneumoniae D39 in a mouse model. *Infect Immun* **75**: 1843-1851.

Oliver, D.B. (1993) SecA protein: autoregulated ATPase catalysing preprotein insertion and translocation across the Escherichia coli inner membrane. *Mol Microbiol* **7**: 159-165.

Orihuela, C.J., Gao, G., McGee, M., Yu, J., Francis, K.P., and Tuomanen, E. (2003) Organ-specific models of Streptococcus pneumoniae disease. *Scand J Infect Dis* **35**: 647-652.

Orihuela, C.J., Gao, G., Francis, K.P., Yu, J., and Tuomanen, E.I. (2004a) Tissue-specific contributions of pneumococcal virulence factors to pathogenesis. *J Infect Dis* **190**: 1661-1669.

Orihuela, C.J., Radin, J.N., Sublett, J.E., Gao, G., Kaushal, D., and Tuomanen, E.I. (2004b) Microarray analysis of pneumococcal gene expression during invasive disease. *Infect Immun* **72**: 5582-5596.

Osborne, R.S., and Silhavy, T.J. (1993) PrlA suppressor mutations cluster in regions corresponding to three distinct topological domains. *Embo J* **12**: 3391-3398.

Overweg, K., Pericone, C.D., Verhoef, G.G., Weiser, J.N., Meiring, H.D., De Jong, A.P., De Groot, R., and Hermans, P.W. (2000) Differential protein expression in phenotypic variants of Streptococcus pneumoniae. *Infect Immun* **68**: 4604-4610.

Parsek, M.R., and Singh, P.K. (2003) Bacterial biofilms: an emerging link to disease pathogenesis. *Annu Rev Microbiol* **57**: 677-701.

Paton, J.C., Andrew, P.W., Boulnois, G.J., and Mitchell, T.J. (1993) Molecular analysis of the pathogenicity of Streptococcus pneumoniae: the role of pneumococcal proteins. *Annu Rev Microbiol* **47**: 89-115.

Pearce, B.J., Iannelli, F., and Pozzi, G. (2002) Construction of new unencapsulated (rough) strains of Streptococcus pneumoniae. *Res Microbiol* **153**: 243-247.

Peng, Z., Fives-Taylor, P., Ruiz, T., Zhou, M., Sun, B., Chen, Q., and Wu, H. (2008a) Identification of critical residues in Gap3 of Streptococcus parasanguinis involved in Fap1 glycosylation, fimbrial formation and in vitro adhesion. *BMC Microbiol* **8**: 52.

Peng, Z., Wu, H., Ruiz, T., Chen, Q., Zhou, M., Sun, B., and Fives-Taylor, P. (2008b) Role of gap3 in Fap1 glycosylation, stability, in vitro adhesion, and fimbrial and biofilm formation of Streptococcus parasanguinis. *Oral Microbiol Immunol* **23**: 70-78.

Pericone, C.D., Overweg, K., Hermans, P.W., and Weiser, J.N. (2000) Inhibitory and bactericidal effects of hydrogen peroxide production by Streptococcus pneumoniae on other inhabitants of the upper respiratory tract. *Infect Immun* **68**: 3990-3997.

Pericone, C.D., Park, S., Imlay, J.A., and Weiser, J.N. (2003) Factors contributing to hydrogen peroxide resistance in Streptococcus pneumoniae include pyruvate oxidase (SpxB) and avoidance of the toxic effects of the fenton reaction. *J Bacteriol* **185**: 6815-6825.

Pestova, E.V., Havarstein, L.S., and Morrison, D.A. (1996) Regulation of competence for genetic transformation in Streptococcus pneumoniae by an auto-induced peptide pheromone and a two-component regulatory system. *Mol Microbiol* **21**: 853-862.

Petersen, F.C., Pecharki, D., and Scheie, A.A. (2004) Biofilm mode of growth of Streptococcus intermedius favored by a competence-stimulating signaling peptide. *J Bacteriol* **186**: 6327-6331.

Peterson, S.N., Sung, C.K., Cline, R., Desai, B.V., Snesrud, E.C., Luo, P., Walling, J., Li, H., Mintz, M., Tsegaye, G., Burr, P.C., Do, Y., Ahn, S., Gilbert, J., Fleischmann, R.D., and Morrison, D.A. (2004) Identification of competence pheromone responsive genes in Streptococcus pneumoniae by use of DNA microarrays. *Mol Microbiol* **51**: 1051-1070.

Pettigrew, M.M., Fennie, K.P., York, M.P., Daniels, J., and Ghaffar, F. (2006) Variation in the presence of neuraminidase genes among Streptococcus pneumoniae isolates with identical sequence types. *Infect Immun* **74**: 3360-3365.

Plummer, C., Wu, H., Kerrigan, S.W., Meade, G., Cox, D., and Ian Douglas, C.W. (2005) A serine-rich glycoprotein of Streptococcus sanguis mediates adhesion to platelets via GPIb. *Br J Haematol* **129**: 101-109.

Polissi, A., Pontiggia, A., Feger, G., Altieri, M., Mottl, H., Ferrari, L., and Simon, D. (1998) Large-scale identification of virulence genes from Streptococcus pneumoniae. *Infect Immun* **66**: 5620-5629.

Potera, C. (1999) Forging a link between biofilms and disease. *Science* **283**: 1837, 1839.

Pozzi, G., Masala, L., Iannelli, F., Manganelli, R., Havarstein, L.S., Piccoli, L., Simon, D., and Morrison, D.A. (1996) Competence for genetic transformation in encapsulated strains of Streptococcus pneumoniae:

two allelic variants of the peptide pheromone. *J Bacteriol* **178**: 6087-6090.

Prakobphol, A., Xu, F., Hoang, V.M., Larsson, T., Bergstrom, J., Johansson, I., Frangsmyr, L., Holmskov, U., Leffler, H., Nilsson, C., Boren, T., Wright, J.R., Stromberg, N., and Fisher, S.J. (2000) Salivary agglutinin, which binds Streptococcus mutans and Helicobacter pylori, is the lung scavenger receptor cysteine-rich protein gp-340. *J Biol Chem* **275**: 39860-39866.

Pratt, L.A., and Kolter, R. (1998) Genetic analysis of Escherichia coli biofilm formation: roles of flagella, motility, chemotaxis and type I pili. *Mol Microbiol* **30**: 285-293.

Prellner, K., Hermansson, A., White, P., Melhus, A., and Briles, D. (1999) Immunization and protection in pneumococcal otitis media studied in a rat model. *Microb Drug Resist* **5**: 73-82.

Pugsley, A.P. (1993) The complete general secretory pathway in gram-negative bacteria. *Microbiol Rev* **57**: 50-108.

Quin, L.R., Carmicle, S., Dave, S., Pangburn, M.K., Evenhuis, J.P., and McDaniel, L.S. (2005) In vivo binding of complement regulator factor H by Streptococcus pneumoniae. *J Infect Dis* **192**: 1996-2003.

Quin, L.R., Onwubiko, C., Moore, Q.C., Mills, M.F., McDaniel, L.S., and Carmicle, S. (2007) Factor H binding to PspC of Streptococcus pneumoniae increases adherence to human cell lines in vitro and enhances invasion of mouse lungs in vivo. *Infect Immun* **75**: 4082-4087.

Raj K. Upreti, M.K.V.S. (2003) Bacterial glycoproteins: Functions, biosynthesis and applications. *PROTEOMICS* **3**: 363-379.

Rajam, G., Phillips, D.J., White, E., Anderton, J., Hooper, C.W., Sampson, J.S., Carlone, G.M., Ades, E.W., and Romero-Steiner, S. (2008) A functional epitope of the pneumococcal surface adhesin A activates nasopharyngeal cells and increases bacterial internalization. *Microbial Pathogenesis* **44**: 186-196.

Rapoport, T.A. (2007) Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**: 663-669.

Regev-Yochay, G., Trzcinski, K., Thompson, C.M., Lipsitch, M., and Malley, R. (2007) SpxB is a suicide gene of Streptococcus pneumoniae and confers a selective advantage in an in vivo competitive colonization model. *J Bacteriol* **189**: 6532-6539.

Reinert, R.R. (2004) Pneumococcal conjugate vaccines--a European perspective. *Int J Med Microbiol* **294**: 277-294.

Ren, B., McCrory, M.A., Pass, C., Bullard, D.C., Ballantyne, C.M., Xu, Y., Briles, D.E., and Szalai, A.J. (2004) The virulence function of Streptococcus pneumoniae surface protein A involves inhibition of complement activation and impairment of complement receptor-mediated protection. *J Immunol* **173**: 7506-7512.

Ribes, S., Taberner, F., Cabellos, C., Tubau, F., Ardanuy, C., Gerber, J., Linares, J., Nau, R., and Gudiol, F. (2008) Contribution of capsular and

clonal types and beta-lactam resistance to the severity of experimental pneumococcal meningitis. *Microbes Infect* **10**: 129-134.
Rigel, N.W., and Braunstein, M. (2008) A new twist on an old pathway - accessory secretion systems. *Mol Microbiol* **69**: 291-302.
Ring, A., Weiser, J.N., and Tuomanen, E.I. (1998) Pneumococcal trafficking across the blood-brain barrier. Molecular analysis of a novel bidirectional pathway. *J Clin Invest* **102**: 347-360.
Robinson, C., and Bolhuis, A. (2004) Tat-dependent protein targeting in prokaryotes and chloroplasts. *Biochim Biophys Acta* **1694**: 135-147.
Robson, A., and Collinson, I. (2006) The structure of the Sec complex and the problem of protein translocation. *EMBO Rep* **7**: 1099-1103.
Roche, F.M., Massey, R., Peacock, S.J., Day, N.P., Visai, L., Speziale, P., Lam, A., Pallen, M., and Foster, T.J. (2003) Characterization of novel LPXTG-containing proteins of Staphylococcus aureus identified from genome sequences. *Microbiology* **149**: 643-654.
Romain, F., Horn, C., Pescher, P., Namane, A., Riviere, M., Puzo, G., Barzu, O., and Marchal, G. (1999) Deglycosylation of the 45/47-Kilodalton Antigen Complex of Mycobacterium tuberculosis Decreases Its Capacity To Elicit In Vivo or In Vitro Cellular Immune Responses. *Infect. Immun.* **67**: 5567-5572.
Rosan, B., and Lamont, R.J. (2000) Dental plaque formation. *Microbes Infect* **2**: 1599-1607.
Rose, L., Shivshankar, P., Hinojosa, E., Rodriguez, A., Sanchez, C.J., and Orihuela, C.J. (2008) Antibodies against PsrP, a novel Streptococcus pneumoniae adhesin, block adhesion and protect mice against pneumococcal challenge. *J Infect Dis* **198**: 375-383.
Rosenow, C., Ryan, P., Weiser, J.N., Johnson, S., Fontan, P., Ortqvist, A., and Masure, H.R. (1997) Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of Streptococcus pneumoniae. *Mol Microbiol* **25**: 819-829.
Rouphael, N.G., Atwell-Melnick, N., Longo, D., Whaley, M., Carlone, G.M., Sampson, J.S., and Ades, E.W. A real-time polymerase chain reaction for the detection of Streptococcus pneumoniae in blood using a mouse model: a potential new "gold standard". *Diagnostic Microbiology and Infectious Disease* **In Press, Corrected Proof**.
Sandgren, A., Sjostrom, K., Olsson-Liljequist, B., Christensson, B., Samuelsson, A., Kronvall, G., and Henriques Normark, B. (2004) Effect of clonal and serotype-specific properties on the invasive capacity of Streptococcus pneumoniae. *J Infect Dis* **189**: 785-796.
Sandgren, A., Albiger, B., Orihuela, C.J., Tuomanen, E., Normark, S., and Henriques-Normark, B. (2005) Virulence in mice of pneumococcal clonal types with known invasive disease potential in humans. *J Infect Dis* **192**: 791-800.
Schaffer, C., Graninger, M., and Messner, P. (2001) Prokaryotic glycosylation. *Proteomics* **1**: 248-261.

Schmidt, M.A., Riley, L.W., and Benz, I. (2003) Sweet new world: glycoproteins in bacterial pathogens. *Trends in Microbiology* **11**: 554-561.

Schneider, O., Michel, U., Zysk, G., Dubuis, O., and Nau, R. (1999) Clinical outcome in pneumococcal meningitis correlates with CSF lipoteichoic acid concentrations. *Neurology* **53**: 1584-1587.

Scott, J.A., Hall, A.J., Dagan, R., Dixon, J.M., Eykyn, S.J., Fenoll, A., Hortal, M., Jette, L.P., Jorgensen, J.H., Lamothe, F., Latorre, C., Macfarlane, J.T., Shlaes, D.M., Smart, L.E., and Taunay, A. (1996) Serogroup-specific epidemiology of Streptococcus pneumoniae: associations with age, sex, and geography in 7,000 episodes of invasive disease. *Clin Infect Dis* **22**: 973-981.

Shah, P., and Swiatlo, E. (2006) Immunization with polyamine transport protein PotD protects mice against systemic infection with Streptococcus pneumoniae. *Infect Immun* **74**: 5888-5892.

Shaper, M., Hollingshead, S.K., Benjamin, W.H., Jr., and Briles, D.E. (2004) PspA protects Streptococcus pneumoniae from killing by apolactoferrin, and antibody to PspA enhances killing of pneumococci by apolactoferrin [corrected]. *Infect Immun* **72**: 5031-5040.

Siboo, I.R., Bensing, B.A., and Sullam, P.M. (2003) Genomic organization and molecular characterization of SM1, a temperate bacteriophage of Streptococcus mitis. *J Bacteriol* **185**: 6968-6975.

Siboo, I.R., Chambers, H.F., and Sullam, P.M. (2005) Role of SraP, a Serine-Rich Surface Protein of Staphylococcus aureus, in binding to human platelets. *Infect Immun* **73**: 2273-2280.

Siboo, I.R., Chaffin, D.O., Rubens, C.E., and Sullam, P.M. (2008) Characterization of the Accessory Sec System of Staphylococcus aureus. *J Bacteriol*.

Silva, N.A., McCluskey, J., Jefferies, J.M., Hinds, J., Smith, A., Clarke, S.C., Mitchell, T.J., and Paterson, G.K. (2006) Genomic diversity between strains of the same serotype and multilocus sequence type among pneumococcal clinical isolates. *Infect Immun* **74**: 3513-3518.

Sjostrom, K., Spindler, C., Ortqvist, A., Kalin, M., Sandgren, A., Kuhlmann-Berenzon, S., and Henriques-Normark, B. (2006) Clonal and capsular types decide whether pneumococci will act as a primary or opportunistic pathogen. *Clin Infect Dis* **42**: 451-459.

Steinmoen, H., Knutsen, E., and Havarstein, L.S. (2002) Induction of natural competence in Streptococcus pneumoniae triggers lysis and DNA release from a subfraction of the cell population. *Proc Natl Acad Sci U S A* **99**: 7681-7686.

Steinmoen, H., Teigen, A., and Havarstein, L.S. (2003) Competence-induced cells of Streptococcus pneumoniae lyse competence-deficient cells of the same strain during cocultivation. *J Bacteriol* **185**: 7176-7183.

Stephenson, A.E., Wu, H., Novak, J., Tomana, M., Mintz, K., and Fives-Taylor, P. (2002) The Fap1 fimbrial adhesin is a glycoprotein: antibodies specific for the glycan moiety block the adhesion of Streptococcus parasanguis in an in vitro tooth model. *Mol Microbiol* **43**: 147-157.

Stickler, D. (1999) Biofilms. *Curr Opin Microbiol* **2**: 270-275.
Sullam, P.M., Drake, T.A., and Sande, M.A. (1985) Pathogenesis of endocarditis. *Am J Med* **78**: 110-115.
Szymanski, C.M., Logan, S.M., Linton, D., and Wren, B.W. (2003) Campylobacter - a tale of two protein glycosylation systems. *Trends in Microbiology* **11**: 233-238.
Szymanski, C.M., and Wren, B.W. (2005) Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Micro* **3**: 225-237.
Takahashi, Y., Yajima, A., Cisar, J.O., and Konishi, K. (2004) Functional analysis of the Streptococcus gordonii DL1 sialic acid-binding adhesin and its essential role in bacterial binding to platelets. *Infect Immun* **72**: 3876-3882.
Takahashi, Y., Takashima, E., Shimazu, K., Yagishita, H., Aoba, T., and Konishi, K. (2006) Contribution of sialic acid-binding adhesin to pathogenesis of experimental endocarditis caused by Streptococcus gordonii DL1. *Infect Immun* **74**: 740-743.
Takamatsu, D., Bensing, B.A., and Sullam, P.M. (2004a) Genes in the accessory sec locus of Streptococcus gordonii have three functionally distinct effects on the expression of the platelet-binding protein GspB. *Mol Microbiol* **52**: 189-203.
Takamatsu, D., Bensing, B.A., and Sullam, P.M. (2004b) Four proteins encoded in the gspB-secY2A2 operon of Streptococcus gordonii mediate the intracellular glycosylation of the platelet-binding protein GspB. *J Bacteriol* **186**: 7100-7111.
Takamatsu, D., Bensing, B.A., Cheng, H., Jarvis, G.A., Siboo, I.R., Lopez, J.A., Griffiss, J.M., and Sullam, P.M. (2005a) Binding of the Streptococcus gordonii surface glycoproteins GspB and Hsa to specific carbohydrate structures on platelet membrane glycoprotein Ibalpha. *Mol Microbiol* **58**: 380-392.
Takamatsu, D., Bensing, B.A., and Sullam, P.M. (2005b) Two additional components of the accessory sec system mediating export of the Streptococcus gordonii platelet-binding protein GspB. *J Bacteriol* **187**: 3878-3883.
Takamatsu, D., Bensing, B.A., Prakobphol, A., Fisher, S.J., and Sullam, P.M. (2006) Binding of the streptococcal surface glycoproteins GspB and Hsa to human salivary proteins. *Infect Immun* **74**: 1933-1940.
Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K., and Fraser, C.M. (2001) Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. *Science* **293**: 498-506.

Throup, J.P., Koretke, K.K., Bryant, A.P., Ingraham, K.A., Chalker, A.F., Ge, Y., Marra, A., Wallis, N.G., Brown, J.R., Holmes, D.J., Rosenberg, M., and Burnham, M.K. (2000) A genomic analysis of two-component signal transduction in Streptococcus pneumoniae. *Mol Microbiol* **35**: 566-576.

Toledo-Arana, A., Merino, N., Vergara-Irigaray, M., Debarbouille, M., Penades, J.R., and Lasa, I. (2005) Staphylococcus aureus develops an alternative, ica-independent biofilm in the absence of the arlRS two-component system. *J Bacteriol* **187**: 5318-5329.

Tong, H.H., Blue, L.E., James, M.A., and DeMaria, T.F. (2000) Evaluation of the virulence of a Streptococcus pneumoniae neuraminidase-deficient mutant in nasopharyngeal colonization and development of otitis media in the chinchilla model. *Infect Immun* **68**: 921-924.

Tseng, H.J., McEwan, A.G., Paton, J.C., and Jennings, M.P. (2002) Virulence of Streptococcus pneumoniae: PsaA mutants are hypersensitive to oxidative stress. *Infect Immun* **70**: 1635-1639.

Tu, A.H., Fulgham, R.L., McCrory, M.A., Briles, D.E., and Szalai, A.J. (1999) Pneumococcal surface protein A inhibits complement activation by Streptococcus pneumoniae. *Infect Immun* **67**: 4720-4724.

Tuomanen, E., Liu, H., Hengstler, B., Zak, O., and Tomasz, A. (1985) The induction of meningeal inflammation by components of the pneumococcal cell wall. *J Infect Dis* **151**: 859-868.

Verhaegen, J., Glupczynski, Y., Verbist, L., Blogie, M., Verbiest, N., Vandeven, J., and Yourassowsky, E. (1995) Capsular types and antibiotic susceptibility of pneumococci isolated from patients in Belgium with serious infections, 1980-1993. *Clin Infect Dis* **20**: 1339-1345.

von Heijne, G., Liljestrom, P., Mikus, P., Andersson, H., and Ny, T. (1991) The efficiency of the uncleaved secretion signal in the plasminogen activator inhibitor type 2 protein can be enhanced by point mutations that increase its hydrophobicity. *J Biol Chem* **266**: 15240-15243.

Waite, R.D., Struthers, J.K., and Dowson, C.G. (2001) Spontaneous sequence duplication within an open reading frame of the pneumococcal type 3 capsule locus causes high-frequency phase variation. *Mol Microbiol* **42**: 1223-1232.

Waite, R.D., Penfold, D.W., Struthers, J.K., and Dowson, C.G. (2003) Spontaneous sequence duplications within capsule genes cap8E and tts control phase variation in Streptococcus pneumoniae serotypes 8 and 37. *Microbiology* **149**: 497-504.

Ware, D., Jiang, Y., Lin, W., and Swiatlo, E. (2006) Involvement of potD in Streptococcus pneumoniae polyamine transport and pathogenesis. *Infect Immun* **74**: 352-361.

Watnick, P.I., and Kolter, R. (1999) Steps in the development of a Vibrio cholerae El Tor biofilm. *Mol Microbiol* **34**: 586-595.

Weiser, J.N. (1993) Relationship between colony morphology and the life cycle of Haemophilus influenzae: the contribution of lipopolysaccharide phase variation to pathogenesis. *J Infect Dis* **168**: 672-680.

Weiser, J.N., Austrian, R., Sreenivasan, P.K., and Masure, H.R. (1994) Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. *Infect. Immun.* **62**: 2582-2589.

Weiser, J.N., and Kapoor, M. (1999) Effect of intrastrain variation in the amount of capsular polysaccharide on genetic transformation of Streptococcus pneumoniae: implications for virulence studies of encapsulated strains. *Infect Immun* **67**: 3690-3692.

Winkelstein, J.A. (1984) Complement and the host's defense against the pneumococcus. *Crit Rev Microbiol* **11**: 187-208.

Wu, H., Mintz, K.P., Ladha, M., and Fives-Taylor, P.M. (1998) Isolation and characterization of Fap1, a fimbriae-associated adhesin of Streptococcus parasanguis FW213. *Mol Microbiol* **28**: 487-500.

Wu, H., and Fives-Taylor, P.M. (1999) Identification of dipeptide repeats and a cell wall sorting signal in the fimbriae-associated adhesin, Fap1, of Streptococcus parasanguis. *Mol Microbiol* **34**: 1070-1081.

Wu, H., and Fives-Taylor, P.M. (2001) Molecular strategies for fimbrial expression and assembly. *Crit Rev Oral Biol Med* **12**: 101-115.

Wu, H., Bu, S., Newell, P., Chen, Q., and Fives-Taylor, P. (2007a) Two gene determinants are differentially involved in the biogenesis of Fap1 precursors in Streptococcus parasanguis. *J Bacteriol* **189**: 1390-1398.

Wu, H., Zeng, M., and Fives-Taylor, P. (2007b) The glycan moieties and the N-terminal polypeptide backbone of a fimbria-associated adhesin, Fap1, play distinct roles in the biofilm development of Streptococcus parasanguinis. *Infect Immun* **75**: 2181-2188.

Xiong, Y.Q., Bensing, B.A., Bayer, A.S., Chambers, H.F., and Sullam, P.M. (2008) Role of the serine-rich surface glycoprotein GspB of Streptococcus gordonii in the pathogenesis of infective endocarditis. *Microb Pathog*.

Yoon, S.S., Hennigan, R.F., Hilliard, G.M., Ochsner, U.A., Parvatiyar, K., Kamani, M.C., Allen, H.L., DeKievit, T.R., Gardner, P.R., Schwab, U., Rowe, J.J., Iglewski, B.H., McDermott, T.R., Mason, R.P., Wozniak, D.J., Hancock, R.E., Parsek, M.R., Noah, T.L., Boucher, R.C., and Hassett, D.J. (2002) Pseudomonas aeruginosa anaerobic respiration in biofilms: relationships to cystic fibrosis pathogenesis. *Dev Cell* **3**: 593-603.

Young, N.M., Brisson, J.R., Kelly, J., Watson, D.C., Tessier, L., Lanthier, P.H., Jarrell, H.C., Cadotte, N., St Michael, F., Aberg, E., and Szymanski, C.M. (2002) Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, Campylobacter jejuni. *J Biol Chem* **277**: 42530-42539.

Zhang, J.-R., Mostov, K.E., Lamm, M.E., Nanno, M., Shimida, S.-i., Ohwaki, M., and Tuomanen, E. (2000) The Polymeric Immunoglobulin Receptor Translocates Pneumococci across Human Nasopharyngeal Epithelial Cells. *Cell* **102**: 827-837.