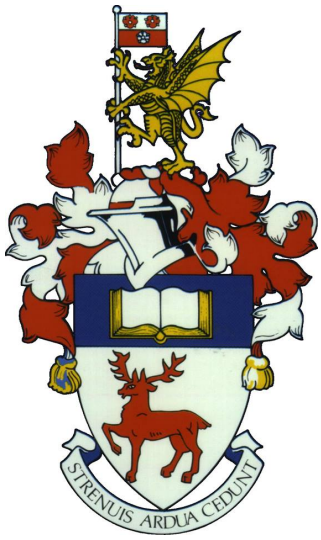
# University of Southampton

## Faculty of Natural and Environmental Sciences

## Centre for Biological Sciences


# Adaptive mutation and evolution of the RNA polymerase subunit delta during *Streptococcus pneumoniae* biofilm development


By

Robert Hull


Thesis submitted for the degree of Doctor of Philosophy

November 2018


Supervisors: Professor Jeremy Webb, Dr Stuart Clarke, Dr David Cleary and Professor Saul Faust

**UNIVERSITY OF SOUTHAMPTON**
**FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES**

**Centre for Biological Sciences**

**Doctor of Philosophy**

# Adaptive mutation and evolution during *Streptococcus pneumoniae* biofilm development

**By Robert Iain Hull**

# Abstract

*Streptococcus pneumoniae* typically resides asymptomatically within the respiratory tract, sinuses and nasopharynx of healthy carriers. However, for children, the elderly, or immunocompromised individuals, *S. pneumoniae* is the cause of considerable morbidity and mortality globally. Many of the pathologies caused by this organism are associated with the surface adhered microbial communities known as biofilms. Biofilms of the *S. pneumoniae* clinical isolate 22F ST433 have been observed to rapidly exhibit parallel evolution with mutations to the RNA polymerase (RNAP) subunit δ gene *rpoE* repeatedly occurring. All *rpoE* mutants displayed a Small Colony Variant (SCV) morphology; this phenotype is associated with carriage and disease states in *S. pneumoniae*, and therefore improved understanding of biofilm formation may be key to reducing the burden of disease of this pathogen. The *rpoE* mutations within the SCV population are heterogenous, ranging from single nucleotide polymorphisms to large-scale deletions of the C-terminal domain of RpoE. As a result of these mutation SCVs display significant altered phenotypes relative to the 22F wild-type including metabolic changes, reduced virulence and increased biofilm formation. However, these changes are likely to be due to altered RpoE function rather than a complete loss as we also observed that SCVs were phenotypically distinct from a complete *rpoE* knock-out. Alteration to, rather than loss of RpoE function within SCVs is supported by our observations that for several important characteristics, including biofilm formation, the magnitude of the phenotypic change correlates to the length of the C-terminal domain of RpoE. Phenotypic differences between the 22F wild-type and SCVs are therefore likely to be due to the C-terminal domain's roles in RNAP-δ-DNA interactions, the positioning of RNAP within gene promoter regions and open complex formation where C-terminal domain length is likely to be an important factor for RpoE function. Differences observed between SCVs and *rpoE* KO are likely due to the conserved N-terminal domain of the SCVs, which we predict to have DNA binding and enzymatic activity, and the complete loss of which causes dysregulation. These findings further highlight the diversity that can be generated within even single species biofilms and the mechanisms through which *S. pneumoniae* RpoE functions.

# Table of Contents

# Table of Figures

# Declaration of Authorship

I, Robert Iain Hull, declare that this thesis entitled "Adaptive mutation and evolution during *Streptococcus pneumoniae* biofilm development" and the work presented in it are my own and has been generated by me as a result of my own original research.

I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University;

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

- Where I have consulted the published work of others, this is always clearly attributed;

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

- I have acknowledged all main sources of help;

- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

- None of this work has been published before submission

Signed: .....

Date: .....

# Acknowledgements

I would like to thank Making People Matter Ltd for funding this project.

I would like to thank my supervisors Professor Jeremy Webb, Dr Stuart Clarke, Dr David Cleary and Professor Saul Faust who have provided guidance, and support throughout my PhD, and in applying for additional grants and awards.

I would like to express thanks to the many individuals in the Webb and Keevil groups who have contributed to this project, including but not limited to Dr Nicholas Churton for his work on *rpoE* in *S. pneumoniae* upon which this project is built; Dr Robert Howlin, and Mr Connor Frapwell for their extensive help through-out the project; Dr Ray Allen for his introduction to basic pneumococcal techniques and Dr Matthew Wand for his demonstration of *G. mellonella* methods. Particular thanks go to Dr Callum Highmore whose research discussions, humour, and great friendship have been essential part of this PhD.

The *Galleria mellonella* work was carried out in partnership with Ms Laura Pearson as part of her Masters of Biomedical Science (MBioSci) at the University of Southampton. The project design, SOP, and research was carried out by Mr Robert Hull and all text here is entirely the work of Mr Robert Hull. I would like to thank and acknowledge the μ-VIS X-ray Imaging Centre at the University of Southampton for provision of tomographic imaging facilities, supported by EPSRC grant EP-H01506X, and in particular Dr Orestis Katsamensis for his expertise, and assistance in producing the micro-CT scans of *G. mellonella*. Special thanks to Professor Vidal for his generous donation of the strain D39 and plasmid pMV158GFP which has been invaluable to this project.

However foremost I would like to thank my family, and in particular my wife Sophie and our children Lily, Alex, and Freyja. Without their support this PhD could not have been a success.

# Conferences

**Biofilms 7:** 26-28th June 2016, Porto, Portugal. "Development of the *Galleria mellonella* animal model for studies of *in vivo* biofilm formation, colonisation, virulence, and evolution in *Streptococcus pneumoniae*". Poster presentation.

**NAMRIP summer conference:** 5th June 2017, Southampton, UK. "Novel In vivo Imaging of Antimicrobial Resistant Streptococcus pneumoniae Biofilms within the *Galleria mellonella* Animal Model". Oral and poster presentations.

**Eurobiofilms2017:** 19-22nd September 2017 Amsterdam, Netherlands. "Evolution of *rpoE* During *Streptococcus pneumoniae* Biofilm Development". Poster presentation.

# Definitions and Abbreviations Used

| Abbreviation | Definition |
|---|---|
| BHI | Brain Heart Infusion |
| CBA | Columbia Blood Agar |
| CLSM | Confocal Laser Scanning Microscopy |
| COPD | Chronic Obstructive Pulmonary Disease |
| CRS | Chronic Rhinosinusitis |
| CSP | Competence Stimulating Peptide |
| CTM | Complete Transformation Medium |
| DNA | Deoxyribonucleic Acid |
| eDNA | Extracellular DNA |
| GFP | Green Fluorescent Protein |
| GTA | Gene Transfer Agent |
| $H_2O_2$ | Hydrogen Peroxide |
| HGT | Horizontal Gene Transfer |
| ICE | Integrated Conjugative Element |
| iNTPs | Initiating Nucleoside Tri-Phosphates |
| LCV | Large Colony Variant |
| LD50 | Lethal Dose for 50% of a population |
| LTSEM | Low Temperature Scanning Electron Microscopy |
| NHP | Non-Human Primate |
| NVT | Non-Vaccine Serotypes |
| $OD_{XX}$ nm | Optical Density at XX nm |
| OM | Otitis Media |
| PBS | Phosphate Buffered Saline |
| PCR | Polymerase Chain Reaction |
| PCV | Pneumococcal Conjugate Vaccines |
| Plynh | Pneumolysin - non-haemolytic |
| Plyh | Pneumolysin - haemolytic |
| RMSDa | Root mean square deviation – a measure of the distance between atoms within super imposed proteins. |
| RNA | Ribonucleic Acid |
| RNAP | Ribonucleic Acid Polymerase |
| ROS | Reactive Oxygen Species |
| RSV | Respiratory Syntical Virus |
| SCV | Small Colony Variant |
| SEM | Scanning Electron Microscopy |
| THY | Todd Hewitt broth + 0.5% Yeast |
| TSB | Tryptic Soy Broth |
| WT | Wild Type |
| μ-CT | Micro Computed X-ray Tomography |

# Chapter 1: Introduction

A typically encapsulated, Gram-positive, commensal bacteria, *Streptococcus pneumoniae* generally grows as pairs or chains of cells (Alonso De Velasco et al., 1995). Pneumococcal carriage normally occurs through asymptomatic colonisation of the nasopharynx where it forms biofilms. The transition to a disease state by *S. pneumoniae* is a rare occurrence (Weiser, 2010). However, *S. pneumoniae* has the potential to cause a wide range of pathologies, including otitis media (OM), chronic rhinosinusitis (CRS), pneumonia, bacteraemia, and meningitis (Hava et al., 2003) and is responsible for considerable morbidity and mortality.

While typically carried asymptomatically, the Gram-positive pathogen *Streptococcus pneumoniae* is responsible for approximately 300,000 deaths per year globally in children under the age of 5 (Wahl et al., 2018). *S. pneumoniae* also exerts a significant burden of morbidity and mortality upon adults, particularly the elderly (Drijkoningen and Rohde, 2014). Amongst adults, the elderly are the most impacted demographic; however those with dysfunctional spleens, cirrhosis of the liver (Bogaert et al., 2004) and compromised immune systems also exhibit increased likelihoods of pneumococcal disease (Szabo et al., 2015; WHO, 2007). Non-lethal *S. pneumoniae* infections also place a significant burden upon healthcare systems. OM is the most common cause for children to see a doctor in the USA, where the majority of children will require at least one visit to a doctor due to OM (Klein, 2000).

The burden of pneumococcal infections upon health care systems is highlighted by the economic cost of these infections. In the USA, OM is responsible for annual costs of $5 billion (Kaplan et al., 1997), while expenditure on pneumonia is measured in 10's of billions of dollars per year in the USA alone (Marrie, 1994).

The considerable burden of *S. pneumoniae* on human populations and its low infection rate are reconciled through the extremely high carriage prevalence of this organism. Carriage rates of 60% in children under 5 are recorded (Allegrucci et al., 2006). In developing nations, carriage prevalence may be as high as 95% in children under the age of 3 and up

to 40% in adults (Lloyd-Evans et al., 1996; Usuf et al., 2014). Colonised individuals are likely to carry multiple strains and serotypes of *S. pneumoniae* simultaneously (Blanchette-Cain et al., 2013).

Currently, the transition between carriage and disease states by *S. pneumoniae* is poorly understood. Both states are associated with biofilms and therefore understanding the evolutionary processes and accompanying phenotypic changes within biofilms is likely to be key to determining the factors that cause the transition from carriage to disease and therefore be targets for reducing the burden of morbidity and mortality of this organism.

*S. pneumoniae* is able to rapidly generate genetic diversity, which facilitates swift exploitation of microbial niches and adaptation to changing environmental conditions and antimicrobial measures including vaccination programs (Andam and Hanage, 2015; Boles et al., 2004; Jefferies et al.). The increasing levels of antibiotic resistance and vaccine escape observed in *S. pneumoniae* reinforce the need for greater understanding of the evolutionary mechanisms involved to ensure that effective medical interventions remain viable in the future (Lynch and Zhanel, 2005; Lynch and Zhanel, 2010; McGee et al., 2001; Picazo, 2009).

Bacteria in biofilms accumulate mutations more rapidly than planktonic cultures (Conibear et al., 2009), and it is increasingly observed that a portion of these mutations occur in a repeatable manner through parallel evolution (mutations to the same gene observed in across multiple independent experiments) (Churton et al., 2016; Churton, 2014; Cooper et al., 2014; McElroy et al., 2014). Parallel evolution within *S. pneumoniae* 22F ST433 biofilms resulting in the production of small colony variants (SCVs) has been observed over time periods as short as 3 days. All 12 of the SCVs sequenced displayed mutations to the gene *rpoE* and increased biofilm formation. Biofilms are associated with many *S. pneumoniae* pathologies and therefore the parallel evolution of *rpoE* within biofilms is likely to be relevant to the clinical outcomes of these pathologies. The mechanisms and impacts of parallel evolution of *rpoE* are not yet fully understood. In this PhD project, the aims are to explore the phenotypic changes caused by *rpoE* mutations, determine whether these changes are likely to impact on the virulence of the mutants and investigate the mechanism through which these changes occur (Churton et al., 2016).

## 1.1 Pneumococcal Conjugate Vaccines and the Relevance of 22F

Pneumococcal conjugate vaccines (PCV's) have been successful in reducing pneumococcal disease and are effective at reducing the incidence of *S. pneumoniae* biofilm-related pathologies (Ben-Shimol et al., 2014; Gisselsson-Solen et al., 2015). However, the evolutionary arms race, known as Red Queen dynamics, leads to rapid to replacement of vaccine serotypes by non-vaccine serotypes. This dynamic ensures that new serotypes of *S. pneumoniae* take the place of those counteracted by the PCV vaccines such as pneumococcal conjugate vaccine 7 (PCV7) which is active against serotypes 4, 6B, 9V, 14, 18C, 19F and 23F or the enhanced pneumococcal conjugate vaccine 13 (PCV13) which adds the additional serotypes 1, 3, 5, 6A and 19A (Chang et al., 2015; Cremers et al., 2015; Hanage et al., 2011; Jefferies et al., 2011). In the UK, following the introduction of PCV7 in 2006, increases in 22F prevalence were observed between 2006 and 2008/2009 in meningitis cases (Pichon et al., 2013) prompting investigation of this serotype (Churton et al., 2016). Similarly, carriage data observed an increase in the prevalence of 22F from 2006/2007-2009/2010, however in subsequent years 22F diminished in carriage prevalence in the Southampton, UK (Devine et al., 2017). Beyond the UK, the ongoing importance of 22F as a globally relevant serotype is underlined by the increased prevalence of *S. pneumoniae* 22F following vaccination programmes. A review of pneumococcal serotype distributions focussed on publications between March 2014 and March 2015 found the prevalence of 22F to be $\leq$ 11.3 % in children under the age of 7 and $\leq$ 6.0 % for adults over the age of 65 (Cui et al., 2017). In Canada, following the implementation of PCV13, 22F became one of the most prevalent serotypes (increasing from 6 to 11% of isolates in 2014) alongside increased antibiotic resistance profiles, with majority of isolates being ST433 (Demczuk et al., 2017; Golden et al., 2015). Increases in the prevalence of 22F have also been observed in other nations that have introduced PCV vaccinations such as PCV13 in Denmark (Slotved et al., 2016) and PCV10 in Brazil (Caierao et al., 2014). Furthermore, globally 22F is responsible for some of the highest rates of paediatric invasive pneumococcal disease for non PCV13 serotypes (Balsells et al., 2017). To the best of the authors knowledge are no direct studies of 22F virulence. This project aims to discern whether the mutations observed in SCVs impact the virulence of 22F. The trend of increased prevalence of 22F serotypes following the introduction of PCV's and the propensity of *S. pneumoniae* to mutate demonstrates the necessity of understanding the

evolutionary processes of *S. pneumoniae* 22F. This project focuses of 22F ST433 due to its local and global relevance and our observations of increased biofilm formation by small colony variants of this strain. Further phenotypic characterisation, determination of virulence and exploration of potential mechanisms will improve understanding of this serotype (Churton et al., 2016).

## 1.2 Biofilm Pathologies of *S. pneumoniae*

Biofilms are defined by Hall Stoodley et al (2009) as "surface-associated microbial communities surrounded by an extracellular polymeric substance matrix". Biofilms protect their inhabitants from host responses, antibiotic treatment and are often associated with chronic disease (Hall-Stoodley and Stoodley, 2009). These protective features ensure that biofilm-based infections are challenging to treat, often requiring intensive therapies to clear.

*S. pneumoniae* is able to colonise a broad range of tissues which is reflected in the tissues infected during biofilm pathologies. Biofilm infections include OM where biofilms form within the eustachian tube (Hoa et al., 2009), CRS which involves biofilm formation on the sinus mucosa and similar to OM is one of the most common diseases in the USA (Sanclement et al., 2005; Sanderson et al., 2006) and pneumonia during which cellular aggregates similar to biofilm have been observed within the lung (Chao et al., 2014) (Moscoso et al., 2009). Endotracheal tubes are also known to facilitate biofilm formation enabling re-emergence of infection following clinical intervention (Bauer et al., 2002). SCVs with their increased biofilm formation characteristics are likely to be relevant to these infections.

The prevalence of *S. pneumoniae* biofilm pathologies varies across demographics. OM is generally found in infants, with the greatest incidence of infections between 6 and 24 months. CRS is commonly found within adults, while the elderly are the principal risk group for pneumonia. Vaccination of children provides significant herd immunity for populations (Jose et al., 2015).

### 1.2.1 Otitis Media

OM is the best characterised biofilm related pathology of *S. pneumoniae* and is useful as a case study in pneumococcal biofilm infection. OM occurs when nasopharyngeal bacteria breach eustachian tube defences and colonise the middle ear. Usually the mucocillary and immune defences of the Eustachian tube are highly effective at preventing colonisation and infection utilising mucins, aquaporins, surfactants, and other antimicrobial measures (Giebink, 1989). In a healthy adult, these mechanisms are able to maintain a sterile environment within the Eustachian tube. However, colonisation occurs in infants or those with compromised immune systems (Giebink, 1989).

It is likely that OM infections are due to biofilm infections. Studies of OM in children have observed biofilms and cell aggregates formed of *S. pneumoniae* and *Haemophilus influenzae* within the middle ear using multiple detection methods (Bakaletz, 2007; Hall-Stoodley et al., 2006).

The pathology of OM presents in a variety of ways. However, acute OM is typically characterised by fever and inflammation, but may also be accompanied by pus formation within the middle ear. Chronic forms of OM also present with inflammation, but are more likely to include effusion and viscous fluid behind the tympanic membrane (Qureishi et al., 2014).

OM infections are often detrimental with a high potential for permanent loss of hearing, particularly where access to medical care is limited (Klein, 2000). Loss of hearing has profound effects on the development of speech, language and cognitive ability which may have ramifications for the rest of the child's life (Coticchia et al., 2013).

Numerous risk factors are known to increase the likelihood of an OM infection. Males are more likely to contract an infection, as are those who belong to certain ethnic groups. Those who contract an OM infection within the first six months of their lives are more likely to have severe and recurrent OM, as are those with respiratory mucosa allergies, or immune deficiencies such as AIDs, or co-infections with certain viruses. Congenital defects such as a cleft palate or eustachian tube dysfunction also contribute to a greater likelihood of disease; even the diet of infants affects the incidence of OM (Timby et al.,

2015). The mechanisms through which *S. pneumoniae* causes these impacts in OM and other infections are mediated through virulence factors that are detailed below.

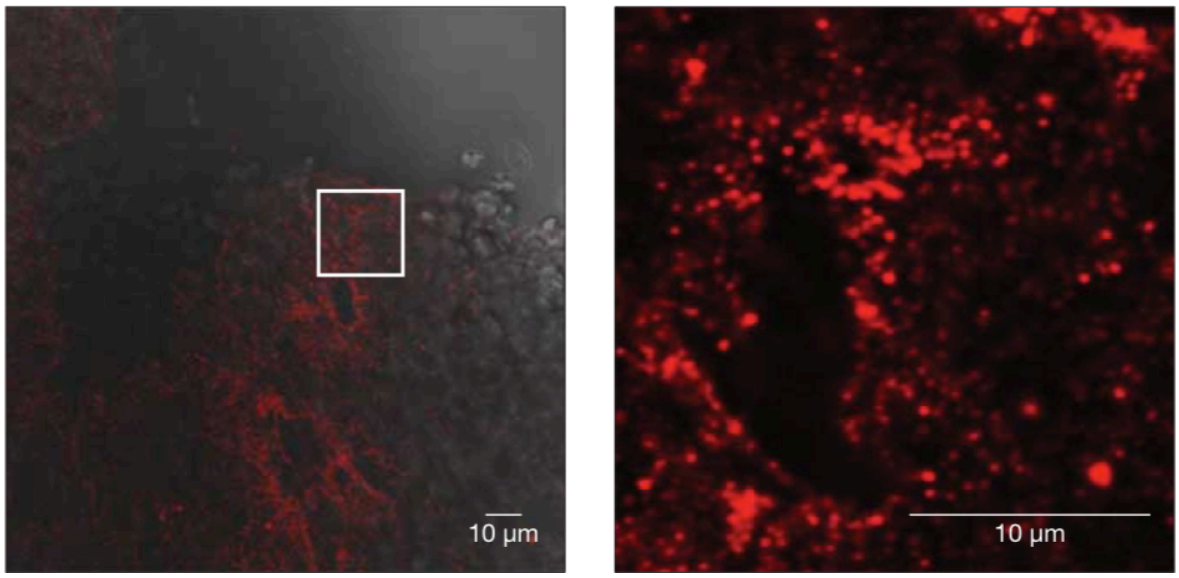



Figure 1: Confocal laser scanning microscopy (CLSM) of a biofilm from the ear biopsy of a child suffering from chronic OM. *S. pneumoniae* are stained red. Adapted from (Hall-Stoodley et al., 2006).

## 1.3 Virulence Factors

Virulence factors are the key features, and related behaviours that enable the pathologies described in 1.2. Virulence factors determine the ability of *S. pneumoniae* to colonise the host, resist or escape the immune response, extract nutrients, and facilitate the mechanisms through which *S. pneumoniae* damages its host. *S. pneumoniae* utilises a range of virulence factors to achieve successfully colonisation of the nasopharynx and succeed as an opportunistic pathogen in many tissues. Expression of virulence factors varies between host tissues, serotypes and colony variants (Hava et al., 2003; Orihuela et al., 2004). While this section focuses only on the biological factors that impact virulence, it should also be noted that environmental conditions such as high temperature and airborne dust also significantly impact *S. pneumoniae* virulence (Jusot et al., 2017).

### 1.3.1 Capsule

One of the most significant virulence factors of *S. pneumoniae* is the polysaccharide capsule which shields it from the host immune system. Encapsulated strains have been found to be more than $10^5$ fold more virulent than non-encapsulated strains and enzymatic removal of the capsule has been shown to reduce the virulence of *S. pneumoniae* (Alonso De Velasco et al., 1995; Avery and Dubos, 1931; Hyams et al., 2010; Kim and Weiser, 1998; Watson and Musher, 1990). The capsule is formed from polysaccharides, teichonic and lipoteichonic acids as shown in Figure 2 (Di Guilmi and Dessen, 2002).

Capsule production is governed by the *cps* locus. This locus is highly diverse, with at least 97 known capsule serotypes (Geno et al., 2015; Kadioglu et al., 2008). The principal genes within this locus are *cpsA-D* which are found in almost all serotypes; however, other genes and mechanisms play important capsule production roles in subsets of serotypes. For instance, serotype 18C capsule expression can be controlled through a point mutation within *cps,* while serotype 3 produces its capsule in a synthase-dependant manner (Schaffner et al., 2014).

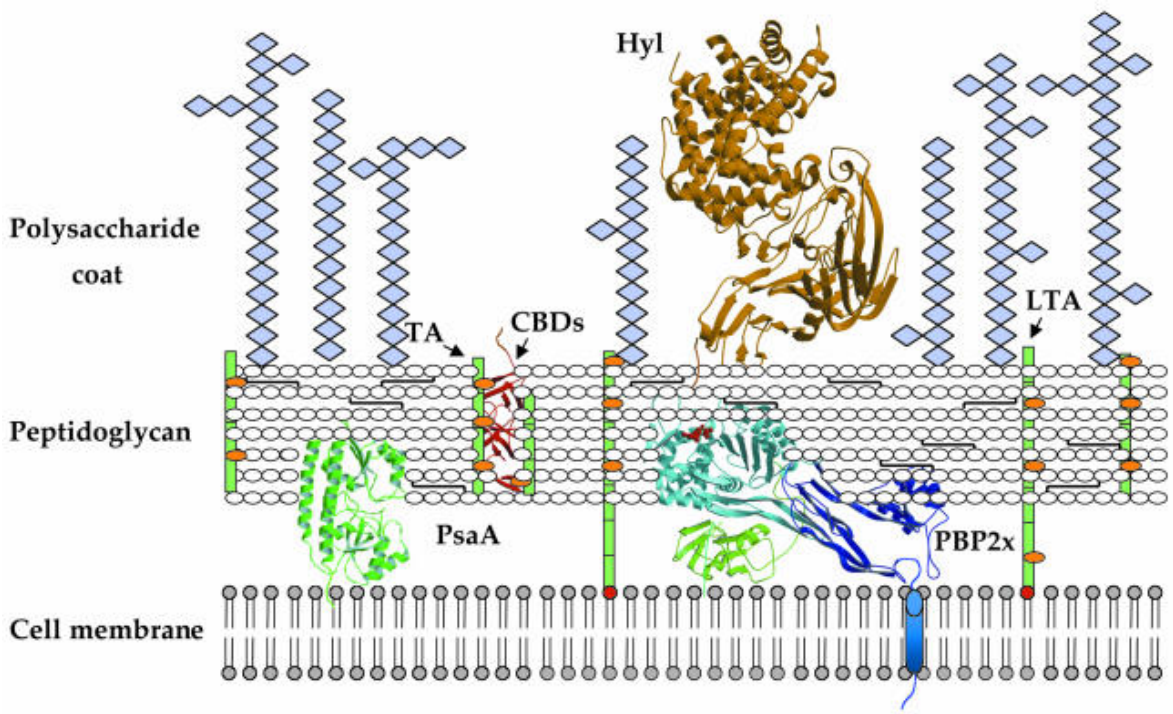



Figure 2: The structure of *S. pneumoniae* cell wall. The cell wall of *S. pneumoniae* consists of 3 layers. The base layer is formed by the cell membrane which is linked to a peptidoglycan layer through lipoteichonic acids (LTAs, displayed in green) which are themselves anchored in to the cell membrane through C-terminal fatty acyl groups (displayed as red circles). Located within the periplasmic space between the cell membrane and the peptidoglycan layer are penicillen-binding proteins which are able to interact with the peptidoglycan layer while being anchored to the cell membrane by a N-terminal transmembrane helix. Similarly the pneumococcal surface antigen (PsaA) is anchored to the cell membrane through an LXXC motif. The peptidoglycan layer contains teichonic acids (TAs), and choline-binding domains (CBD, orange circles) The CBD bind the TAs and LTAs to the peptidoglycan layer. The polysaccharide coat is formed of sugars, and hyaluronate lyase (Hyl) which is bound to the peptidoglycan layer through an LPXTG motif. Figure reproduced from (Di Guilmi and Dessen, 2002).

An intact capsule is often advantageous for *S. pneumoniae's* capability to reach the nasopharynx. Without a capsule *S. pneumoniae* is more likely to become agglutinated within the nasopharynx's luminal mucus and become unable to reach the mucosal epithelial surface of the nasopharynx where colonization and biofilm formation occur (Nelson et al., 2007). While there are many capsule variants the purpose of the capsule as a protective barrier is broadly the same. The capsule is generally 200-400 nm thick (Skov Sorensen et al., 1988), and protects the cell from harm through masking of antigens which enhances the ability of *S. pneumoniae* to evade the immune system (Kadioglu et al., 2008).

Opsonophagocytosis of *S. pneumoniae* is reduced by the capsule through its masking of immunogenic cell surface structures which may be targeted by IgG or the iC3b and other components of the complement pathway (Kadioglu et al., 2008). Additionally, the capsule is highly charged under physiological conditions which inhibits phagocytosis, with more negatively charged capsules providing greater evasion of the immune system (Lee et al., 1991; Nelson et al., 2007). The capsule may also increase antibiotic tolerance inhibiting cell lysis (Fernebro et al., 2004).

The capsule also plays a in *S. pneumoniae* survival in nutrient limited conditions. Capsule deficient mutants are observed to display greatly reduced survival under nutrient limited conditions (Hamaguchi et al., 2018). The capsule mediated increased survival under nutrient limited conditions was shown to be to be independent of the genetic background and the capsule complexity of the pneumococcal strain. During starvation the capsule is consumed as an energy source. Resistance to starvation conditions may increase the likelihood of transmission to a new host and thereby increase pneumococcal fitness (Hamaguchi et al., 2018).

While the capsule provides many protective benefits, it is also metabolically costly and inhibits many other pneumococcal activities, causing the capsule to be disadvantageous under certain circumstances. The presence of a capsule can reduce growth rate, inhibit colonisation of epithelial tissue and biofilm formation, all of which are important factors for maintaining carriage within the nasopharynx (Domenech et al., 2012; Munoz-Elias et al., 2008; Schaffner et al., 2014; Selinger and Reed, 1979). Encapsulated strains are less able to perform cell surface interactions, or to penetrate epithelial cells (Talbot et al.,

1996). Non-encapsulated strains of TIGR4 are able to reach cellular densities of 10 to 100-fold greater compared to the wild type in a mouse model (Nelson et al., 2007).

The capability of non-encapsulated *S. pneumoniae* to colonise and cause disease is highlighted by observations that 3-19 % of asymptomatic carriage isolates may be formed of non-encapsulated strains (Keller et al., 2016). Non-encapsulated strains are associated with both invasive and non-invasive disease, with particularly increased prevalence of non-invasive diseases such as OM and conjunctivitis. A potential mechanism for this may be through replacement of capsule polysaccharide genes in the cps locus with genes such as *aliC*, *aliD* and *pspK* introduced into the locus. The protein PspK is thought to increase survival through increased adhesion. A putative mechanism for PspK increased adhesion could be through support for the R1 region of the cell surface protein PspC (Andrade et al., 2010; Keller et al., 2016; Park et al., 2012). A loss of capsule can also occur through mutations to *cpsE*. In a serotype 18C clinical strain a *cpsE* mutation, believed to have occurred *in vivo* during OM infection, abolishes capsule expression and enhances colonisation (Schaffner et al., 2014).

The size of the capsule is also associated with the recombination frequency, with larger capsules displaying an increased likelihood of recombination. Increased recombination was also associated with increased carriage and diminished invasive potential (Chaguza et al., 2016). This indicates that the adaptations by *S. pneumoniae* to interventions such as vaccinations may be serotype dependant. An association between increased capsule size and increased carriage has been observed with serotype replacement favouring serotypes that utilise metabolically inexpensive polysaccharides (Weinberger et al., 2009). However, capsule expression has also been associated with shedding and transmission of *S. pneumoniae*, with reduced capsular polysaccharide expression has been associated with diminished shedding and therefore reduced transmission (Zafar et al., 2017).

The trade-offs between the protection provided by a capsule and its inhibition of cell surface interactions result in a complex and dynamic system that is constantly evolving. Pneumococcal cells adapt to the fitness costs and benefits of a capsule and its thickness depending on the prevailing environment and host factors. *S. pneumoniae* is able to adapt to the changing environment through a mechanism known as phase variation. Through

phase variation *S. pneumoniae* is able to alter its capsule between opaque and transparent morphologies. The transparent morphology is more effective at colonisation, while opaque variants are more virulent. The key changes in this dynamic are the levels of capsular polysaccharide (which are 1.2-5.6 times greater in opaque variants) and teichonic acid (which are 2.1-3.8 times higher in transparent variants) (Kim and Weiser, 1998; Li and Zhang, 2019; Weiser et al., 1994). The effects of phase variation on capsule production are accentuated under differing levels of oxygen. In in anaerobic conditions opaque variants have been observed to increase capsule production by 5.2-10.6 fold, while transparent variants are unaffected (Weiser et al., 2001). Through this reversible switching *S.pneumoniae* is able to regulate its capsule expression to gain fitness advantages in different host environments, with the aerobic nasopharynx favouring transparent variants while the reduced oxygen concentrations (such as those found within the blood) favours opaque variants (Weiser et al., 2001).

### 1.3.2 Pneumolysin

The 53 KDa thiol-activated toxin produced by the majority of *S. pneumoniae* strains, Pneumolysin (Ply), is an important virulence factor. The most significant characteristic of this toxin is its ability to cause the formation of pores in cholesterol containing cell membranes resulting in lysis of the human cell and pro-inflammatory responses. However, Ply also has extensive interactions throughout both *S. pneumoniae* and the human immune system as shown in Table 1 (Boulnois et al., 1991; Hirst et al., 2004).

| Pneumolysin Activity |
| --- |
| Inhibition of polymorphonuclear cell respiratory burst, random migration and chemotaxis |
| Inhibition of mitogen-induce proliferation and antibody production by human lymphocytes |
| Activation of both the classical and lectin complement pathways |
| Lysis of erythrocytes |
| Inhibition of ciliary beat of respiratory mucosa |
| Toxic to pulmonary alveolar epithelial cells |
| Stimulation of TNF-$\alpha$ and IL-1$\beta$ production from human monocytes |
| Activation of phospholipase $A_2$ in pulmonary endothelial cells |
| Separation of epithelial cell tight junctions |
| Initiates nitric oxide production form macrophages |
| Reduces ciliary beat frequency of cerebral ependymal cells |
| Induces production of IFN-γ in spleen cells |
| Induces synthesis and release of IL-8 from neutrophils |
| Pneumolysin may play a key role in the transmission of *S. pneumoniae* through promotion of inflammation and shedding events |
| Activation of the NLRP3 inflammasome |
| Activation of osmotic stress responses through p38-MAPK |
| Binding to the MRC-1 mannose receptor causing down regulation of inflammation in airways |

Table 1: The impacts of pneumolysin. Adapted from (Anderson and Feldman, 2017; Hirst et al., 2004; McNeela et al., 2010; Ratner et al., 2006; Subramanian et al., 2019; Zafar et al., 2017).

Ply mediated pore formation occurs in several stages. Initially, Ply monomers bind to the surface of cholesterol in cell membranes, which are most likely at lipid raft micro domains where cholesterol concentrations are high. Once a sufficient concentration (approximately 40 monomers) of Ply has accumulated, the monomers oligomerise to form a pore

precursor, which undergoes conformational change to form a large pore within the cell's membrane leading to lysis of the cell (Taylor et al., 2013; Tilley et al., 2005).

In addition to cell lysis, Ply plays an important role in transmission of *S.* pneumoniae. Through promotion of inflammation and shedding events (expelling of bacteria from the body), transmission is increased, with non-pore forming mutants displaying diminished levels of shedding events. Furthermore, increased shedding is associated with increased inflammation (Zafar et al., 2017). In the serotype 2 strain D39 Ply has been shown to play a central role in the inflammatory response to *S. pneumoniae* by the host. Immune cells may be activated in a TLR4 independent mechanism through an IL-1$\beta$ - IL-17A pathway, resulting in activation of the NLRP3 inflammasome complex (Cho et al., 2018; McNeela et al., 2010). Lysosomal damage to dendritic cells through the haemolytic properties of Ply causes activation of NLRP3 leading to increases in pro-inflammatory cytokines, particularly IL-1$\beta$ (McNeela et al., 2010). This is supported by research in serotype 3 6303 where a diminished NLRP3 response due to aged macrophages in the elderly is indicated to be a potential factor in the increased susceptibility of this demographic to *S. pneumoniae* infection (Cho et al., 2018). However, a separate study involving serotype 3 6303 found that NLRP3 may play a detrimental role in anti-pneumococcal activity, with Nlrp4 -/- mice showing strongly increased host defences (van Lieshout et al., 2018). The sources for these differences are currently unclear but may lie in the different backgrounds of the mice used, with Cho et al using a BALB/c mouse model, while van Lieshout et al used C57BI/6 mice which are known to possess different T helper cell backgrounds (Neill et al., 2012).

Beyond NLRP3 activity, Ply has been shown to promote inflammatory responses through a range of mechanisms including host complement activation and promotion of the immune response of by cells including neutrophils (Cockeran et al., 2001a; Cockeran et al., 2001b), macrophages, monocytes (Houldsworth et al., 1994) and $CD4^+$ T cells (Kadioglu et al., 2004). Furthermore, the cytolytic activities of Ply can also induce immune responses through osmotic stress sensing by epithelial cells through activation of p38 mitogen-activated protein kinase (Ratner et al., 2006).

Variations of Ply include variants without haemolytic properties particularly in serotype 1 ST306 and serotype 8 ST53 (Jefferies et al., 2007; Khan et al., 2014; Kirkham et al., 2006).

Non-haemolytic pneumolysin (Plynh) is associated with higher bacterial loads and in mouse models can result in greater mortality rates than haemolytic Ply (Khan et al., 2014). Plynh induces a reduced immune response compared to Ply, with lower levels of IL-1β, IL-6, $CD4^+$ T cell induction, dendritic cell apoptosis, and IL-17A expressing memory $CD4^+$ T cells. This may in part be because Plynh is taken up less readily by dendritic cells than Ply complemented by a reduced activation of the NLRP3 complex (Khan et al., 2014; Witzenrath et al., 2011).

The capability for Plynh to induce a weaker immune response, which allows for greater colonisation, may enable greater levels of invasive disease. This may be particularly pertinent in the case of multi-serotype infection, where colonisation by a Plynh strain established under a diminished immune response creates a suitable habitat for more pathogenic strains. Furthermore, Ply allele swapping may create issues for PCV vaccination programs with the current trend of increased serotype 3 sepsis and empyema potentially due to acquisition of Plynh (Khan et al., 2014).

Further indications that pneumolysin may also play a role in biofilm formation has been observed in Plynh serotypes. Non-haemolytic pneumolysin is commonly expressed in serotypes 1 and 8, and has been observed to promote increased biofilm colonisation densities and diminished immune response through reduced $CD4^+$ T-cell proliferation and Il-17 A activity (Khan et al., 2014; Shak et al., 2013).

Interestingly, Ply also plays a role in suppression of the immune response through binding to the mannose receptor MRC-1which is principally expressed by macrophages and dendritic cells (Subramanian et al., 2019). MRC-1 acts as a phagocytic receptor which binds to microbial antigens such as those found on the capsule of *S. pneumoniae*. Dendritic cells challenged by Ply knock outs in the serotype 4 T4R strain (T4RΔ*ply*) produced increased cytokine responses. Ply inhibits cytokine responses through binding of its $4^{th}$ domain directly to MRC-1 causing up-regulation of the cytokine repressor SOCS1. Furthermore, Ply-MRC-1 binding also facilitates non-lysosomal pneumococcal internalization (Subramanian et al., 2019).

Although Ply is an important virulence factor for *S. pneumoniae*, it does not have an N-terminal secretion signal sequence (Walker et al., 1987). This led to the theory that there was no active transport system to release Ply from *S. pneumoniae* cells, necessitating cell lysis for Ply release thought likely to be mediated through LytA. However, has been observed in the TIGR 4 strain that Ply may be exported through the SecY2A2 system, without which its contact dependant activity is decreased by an order of magnitude (Bandara et al., 2017). However, there is not yet evidence that this phenomenon occurs widely across *S. pneumoniae* strains. Bandara et al also found that SecY2A2 is also associated with biofilm formation with *secA2* mutants displaying diminished biofilm formation. Further evidence that Ply plays a role in biofilm formation comes from TIGR4 and D39 where *ply* knock out mutants have been observed to display reduced biofilm formation (Bandara et al., 2017; Shak et al., 2013).

### 1.3.3 Cell Surface Proteins PspA and PspC

The cell surface proteins PspA and PspC provide additional protection for *S. pneumoniae*. The importance of PspA for *S. pneumoniae* is demonstrated by its prevalence across serotypes, with the PspA protein or *pspA* gene found in all of over 2000 examined strains (Crain et al., 1990; Hollingshead et al., 2000; Vela Coral et al., 2001). PspA's mechanism of action is through interference of the C3 fixed complement pathway (Tu et al., 1999). Inhibition of C3b deposition onto *S. pneumoniae* results in an incomplete alternative C3 complement pathway. Mutant strains with non-functional *pspA* display reduced virulence through greater clearance from the blood of unimmunised mice when compared to wild type *pspA* strains (Tu et al., 1999).

While PspA is variable between serotypes, antibodies have been shown to be cross protective between serotypes in mouse models (Tart et al., 1996). Antibody responses against PspA enable rapid clearance of *S. pneumoniae* from blood, which is why there has been a concerted effort to develop effective vaccines against PspA. Vaccination against PspA has been shown to reduce carriage, intratracheal infections, and sepsis in mice. Other roles of PspA include inhibition of apolactoferrin and lactoferrin binding resulting in reduced bactericidal activity (Shaper et al., 2004).

PspC is a key factor in mucosal colonisation by *S. pneumoniae*, and immunisation against PspC has been shown to protect against sepsis, promote systemic mucosal antibody, and reduce nasopharyngeal carriage (Dave et al., 2004). In addition to its role in adherence to mucosal tissue, PspC may also have roles in binding to nasopharyngeal, brain, and lung tissue, and the invasion of tissue at these points (Ogunniyi et al., 2007).

Structurally similar to PspA, PspC is a choline binding surface protein. PspC is commonly found throughout different strains with ~86% of strains containing a functional copy of the gene *pspC*. PspC binds the phosphorylcholine of cell wall teichonic acids and the bacterial TAs/LTAs that facilitates the adherence to, and colonisation of mucosal epithelial, lung, and brain tissues. The mechanism of action of PspC is to bind glycoconjugates such as sialic acid and lactotetraoses (Dave et al., 2004).

PspC also inhibits immune system activity by binding to C3 on activated epithelial cells, and the secretory portion of IgA and Human Factor H (FH). Inhibition of secretory IgA function is likely to cause multiple problems for the host as secretory IgA is required to prevent microbial adherence, opsonise targets for macrophages and neutrophils, and neutralise toxins (Dave et al., 2004).

FH is a protein that is involved in the protection of cells from the activation of the alternative complement pathway by deactivating C3b bound to host cells before damage can occur. FH and secretory IgA are both able to bind to PspC non-competitively, as FH binds to the α helical domain of PspC, while IgA does not. Furthermore, it is thought that it may be that the binding of FH to *S. pneumoniae* may impart some level of protection to *S. pneumoniae* by aiding in evasion of the immune system (Dave et al., 2004).

### 1.3.4 NanA/B

NanA and NanB are neuramidases that are thought to be produced by all known clinical isolates of *S. pneumoniae* (Kelly et al., 1967; O'Toole et al., 1971). NanA plays a significant role in sepsis and colonisation through cleavage of terminal sugars on human cells, thus facilitating adherence (Kadioglu et al., 2008; Manco et al., 2006). NanB is less studied, but appears to have similar function to NanA in causing sepsis and both may be involved in scavenging carbon sources from macromolecules (Manco et al., 2006).

NanA/B are signal peptides that are exported from the cell; however, NanA also has a C-terminal domain, which is capable of anchoring it to the cell surface (Lock et al., 1988). *S. pneumoniae* strains lacking NanA are less able to colonise the nasopharynx and middle ear. Immunisation against NanA is effective at reducing colonisation by *S. pneumoniae* (Tong et al., 2000). However, NanA mutants have not been observed to be significantly diminished compared to wild-type when introduced through intravenously, intraperitoneal or intratracheal routes (Orihuela et al., 2004). An explanation for this may be due to NanA's association with biofilm formation (Parker et al., 2009) this is consistent with other studies that biofilm formation is crucial for persistence within the nasopharynx and middle ear while intravenous routes of inoculation do not favour biofilm formation, but a planktonic lifestyle within the blood.

### 1.3.5 Phase Variation

Phase variation is the process through which *S. pneumoniae* is able to alter its phenotype in a reversible manner between two states. The ability to generate diversity within a population through multiple phenotypes creates a stronger more resilient population better able to fill niches (Wisniewski-Dye and Vial, 2008). Phase variation occurs more frequently than spontaneous mutations, with phase variation occurring at frequencies greater than $10^{-5}$ per cell per generation compared to spontaneous mutation rates of $10^{-6}$ to $10^{-8}$ per cell per generation (Wisniewski-Dye and Vial, 2008). Phase generation is a widespread trait amongst bacteria, affecting traits as diverse as motility, pili synthesis, and capsule production and virulence (Manso et al., 2014; Tan et al., 2016).

In *S. pneumoniae*, phase variation is most easily observed in transitions between opaque and transparent phenotypes. The different phenotypes are important as they confer changes to virulence, and the ability of the organism to colonise. Transparent phase variants produce lower levels of capsule polysaccharides, and are able to colonise the nasopharynx more readily, while opaque phase variants show a decreased ability to adhere to and colonise the nasopharynx, but were also less easily cleared by phagocytes (Li and Zhang, 2019; Weiser et al., 1994).

Improved adaptation by *S. pneumoniae* to clinically relevant conditions has been observed in cases of acute OM, where there is generally a higher prevalence of the opaque phase variants in the middle ear fluids (Arai et al., 2011). Opaque phase variants have been observed to be more virulent highlighting the potential importance of phase variation in clinical settings (Kim and Weiser, 1998). Duplications within the *cps*3D gene have been found to increase the speed of phase switching (Waite et al., 2001) and this increased speed of switching may allow the mutants to react more rapidly to environmental changes, and to exploit them.

### 1.3.6 Carbohydrate Metabolism

*S. pneumoniae* possesses a broad carbohydrate dependant metabolism capable of exploiting a wide range of substrates. This metabolic flexibility is supported by the considerable portion of its genome committed to the use of these substrates (Tettelin et al., 2001). The importance of carbohydrates to *S. pneumoniae* is further demonstrated by the function of its membrane transporters, 30% of which are involved in carbohydrate uptake (Paulsen et al., 2000) and are capable of importing 20-32 or more carbohydrates dependant on the accessory genome present within the strain (Bidossi et al., 2012) .

Carbohydrate metabolism is controlled principally through the Catabolite Control Repression (CCR) system, which is responsible for modulating sugar utilisation and growth rate. CCR has been observed in the Gram-positive pathogen *Listeria monocytogenes* to play a role in the regulation of virulence genes (Gilbreth et al., 2004; Milenbachs et al., 1997).

In *S. pneumoniae* many of the genes involved in carbohydrate metabolism are also associated with colonisation, invasive disease and pneumonia (Buckwalter and King, 2012). This is perhaps unsurprising, as the lack of environmental sugars within the nasopharynx indicates that carbohydrates scavenged from host tissues are an important of energy source for *S. pneumoniae*. Other sources of carbohydrate for *S. pneumoniae* may include dietary sources such as plant-derived inulin (Buckwalter and King, 2012).

Alterations to the carbohydrate metabolism of *S. pneumoniae* are associated with biofilm formation. During biofilm formation, *S. pneumoniae's* metabolism is significantly altered

with down regulation of glycolytic pathways and increased expression of carbohydrate, pyruvate and arginine metabolism (Allan et al., 2014).

### 1.3.7 Hydrogen Peroxide

Pyruvate generated through carbohydrate metabolism is further catabolised through the *spxB* pathway resulting in the production of hydrogen peroxide ($H_2O_2$). *S. pneumoniae* utilizes $H_2O_2$ to inhibit the growth of other respiratory tract bacteria (Pericone et al., 2000) and damage host tissues through double strand DNA breaks (Rai et al., 2015). *S. pneumoniae* produces $H_2O_2$ through the *spxB* pathway under aerobic growth conditions. Mutants or knockouts of *spxB* are unable to kill competing organisms such as *H. influenzae*, *Moraxella catarrhalis*, or *Neisseria meningitidis* (Pericone et al., 2000). However, *S. pneumoniae* frequently reaches damaging or suicidal levels of $H_2O_2$ hindering its growth (Regev-Yochay et al., 2007).

Hydrogen peroxide production has also been linked to mutations that cause phase variation from semi-transparent to opaque (Ramos-Montanez et al., 2008) and permeant non-phase, small colony variants (Allegrucci and Sauer, 2008) derived from biofilms. All of these mutants displayed mutations to the gene *cps19F*. Quenching of $H_2O_2$ through the addition of catalase or sodium thiosulfate considerably reduced the production of SCVs by biofilms, while addition of $H_2O_2$ cause considerable increases in the production of SCVs, and biofilm formation phenotypic traits (Allegrucci and Sauer, 2008). The strong association between $H_2O_2$, virulence, SCV generation, and biofilm formation, indicate that it may play a key role in understanding the generation of colony variants and the transition from carriage to disease.

## 1.4 *S. pneumoniae* Biofilms

Bacteria have historically been studied in their planktonic form, however over the last 30 years, there has been increasing evidence to show that the majority of microbial life occurs as surface attached, mostly immobile communities of bacteria called biofilms (Costerton et al., 1995; Costerton and Stewart, 2001; Costerton et al., 1999; Hall-Stoodley et al., 2004). Recognition of the importance of biofilms was championed by Bill Costerton whose research led in the discovery that many infections are either fully or in part due to biofilm growth (Bakaletz, 2007; Hoiby et al., 2015) with estimates that 60-80% of microbial

infections are caused by biofilms (Moscoso et al., 2009; Wolcott and Ehrlich, 2008), (WHO program announcements PA-03-047 and PA-06-537). Biofilms are observed to display differences in metabolism and morphologies, while also exhibiting increased tolerances to challenges from the environment, the immune system and antimicrobial treatments.

The mechanisms through which a biofilm provide protection include higher population densities, increased coordination of activities, and the production of extracellular matrix. The extraordinary resilience exhibited by biofilms facilitates colonisation of a wide range of ecological niches, overcoming the challenges required to survive in environments as diverse as soils, and streams to hydrothermal vents (Brazelton and Baross, 2009), animals, and even to thriving on the international space station (Costerton et al., 1999; Kim et al., 2013). Fully understanding the behaviour and the adaptations made by *S. pneumoniae* when forming biofilms will be vital to generating the necessary advances required to adequately control this organism.

### 1.4.1 Biofilm Formation and Disease

There is contention as to whether biofilm formation and the biofilm lifestyle of *S. pneumoniae* contributes towards its virulence. It has been observed that nasopharyngeal biofilm resident *S. pneumoniae* are less virulent than their planktonic relatives or those that disperse from biofilms (Marks et al., 2013; Sanchez et al., 2011). However, *S. pneumoniae* biofilms are often found at sites of infections such as the Eustachian canal in OM (Bakaletz, 2007), in chronic pathologies such as chronic obstructive pulmonary disease (COPD) (Vandevelde et al., 2014) and in the lungs of patients with cystic fibrosis, and other conditions which may indicate a potential correlation between biofilm formation and chronic infections (Lizcano et al., 2010; Munoz-Elias et al., 2008). When introduced intra-cranially to a murine model biofilm derived *S. pneumoniae* have been observed to be significantly more virulent than planktonic bacteria, causing greater pneumonia and meningitis. However, biofilm cells were less virulent when introduced intravenously, and there was no significant difference observed during intra nasal challenge (Oggioni et al., 2006).

Conversely planktonic *S. pneumoniae* have also been observed to display increased virulence and are more able to cause invasive disease than cells taken directly from a biofilm (Marks et al., 2013). However, the authors have also observed that *S. pneumoniae* biofilms undergo dispersal events which release clumps of cells and planktonic bacteria into the wider nasopharynx environment. These dispersed bacteria display increased levels of several virulence factors and altered phenotypes when compared to planktonic and biofilm cells. The increased virulence resulted in greater internalization of the dispersed planktonic bacteria into the epithelial cells (Marks et al., 2013). In addition to this, dispersed planktonic bacteria displayed a decreased ability to colonise epithelial cells compared to broth planktonic cells and biofilm cells (Marks et al., 2013). This evidence indicates that typical biofilm inhabitants are less virulent, but that sub populations which are naturally dispersed from biofilms may possess increased virulence.

Phase shifting is seen during the transition from biofilm bacteria to planktonic dispersed populations. When populations of planktonic *S. pneumoniae* are grown in broth the majority of cells display the opaque phenotype, while the majority of biofilm derived cells exhibit a translucent phenotype. Planktonic dispersal variants displayed a greater proportion of opaque cells than either the biofilm or broth grown planktonic cells (Marks et al., 2013) and this may be a key factor in the increased virulence observed.

A model of diminished virulence for biofilm inhabitants is supported by data which indicates that *S. pneumoniae* derived from biofilms are unable to cause invasive disease (Sanchez et al., 2011). However, the methodology utilised in this set of experiments involved growing cells on silicon tubes under continuous flow of fluid conditions (Sanchez et al., 2011). The continuous flow of fluid would remove any virulent cells derived from biofilm dispersal events. Furthermore, the conditions under which these cells were grown are very different to those that *S. pneumoniae* would normally experience in the nasopharynx and so it is perhaps unsurprising that pneumococcal cells which are adapted to an unnatural flow cell environment are poorly suited to invading mouse epithelial tissue (Sanchez et al., 2011). The importance of the conditions under which *S. pneumoniae* are grown has been made clear by observations of major phenotypic differences between cells grown on different materials (Marks et al., 2012a).

The loss of capsule expression that many *S. pneumoniae* biofilm isolates exhibit (Moscoso et al., 2006) leaves the bacteria vulnerable to opsonophagocytosis and therefore likely to be less virulent (Hyams et al., 2010; Nelson et al., 2007). Additionally, mutants that do not produce capsule polysaccharides have been shown to be the strongest biofilm producers (Moscoso et al., 2006).

However, the increased resistance to antibiotics displayed by *S. pneumoniae* biofilms enables greater survival during antimicrobial treatments which creates a reservoir of increasingly antibiotic resistant bacteria that are able to rapidly recolonize an area (Marks et al., 2012a). This recolonizing population is then able to disperse virulent planktonic cells from the biofilm potentially leading to invasive disease.

In summary, it is likely that the biofilm lifestyle provides a reservoir of reduced virulence bacteria that generate dispersal events which release virulent planktonic *S. pneumoniae*. While some evidence may question this model, deviations may be explained through the unusual route of infection (intra-cranial) that would not normally occur, or may be potentially linked to the manipulation of competence stimulating peptide (CSP) levels at the time of infection (Oggioni et al., 2006). The biofilm lifestyle also rapidly generates mutants, some of which display increased virulence traits (Conibear et al., 2009; Syk et al., 2014). Therefore, whether biofilms can be considered pathogenic may be a matter of semantics as although the majority of a biofilm population may be considered non-virulent, the biofilm environment provided by these same cells enables the generation of the pathogenic sub populations of *S. pneumoniae* which are able to invade human tissues. Furthermore, the invasive cells generated by these mechanisms must be part of the biofilm prior to dispersal events and therefore the case can be made that *S. pneumoniae* in biofilms cannot be considered wholly non-invasive as indicated by some observations (Gilley and Orihuela, 2014), but also that the majority of cells are likely to have reduced virulence compared to planktonic broth derived cells.

The literature highlights the importance of studying biofilm variants and how a greater understanding of these variants may be key to combatting invasive pneumococcal disease through enhanced targeting of virulent sub populations of biofilm populations to remove the most virulent sub populations. Furthermore, novel *S. pneumoniae* virulence factors are

still being discovered including those that link biofilm formation and pathogenesis (Cuevas et al., 2017). This research highlights the importance of studying biofilms and their role in *S. pneumoniae* infections.

### 1.4.2 Formation and Development of Biofilms

*S. pneumoniae* biofilm formation occurs through the combination of several mechanisms. The classical model of biofilm formation involves a single cell attaching to a surface and reproducing to produce a population capable of coordinating its behaviour to build a biofilm. The transition from the planktonic lifestyle to that of a sessile biofilm community is driven by changes in gene expression, and signalling molecules, and can be broadly broken down into several stages depending on the organism (Sauer et al., 2002).

Initially a cell adheres to a surface through electrostatic forces, similarities in hydrophobicity and van der Waals forces. During this attachment stage the process is reversible, with the cell able to detach itself and regain a planktonic lifestyle (Marshall et al., 1971). Bacterial cells at this stage exhibit only very minor quantities of extracellular matrix and may be motile through pilus twitching or gliding (O'Toole and Kolter, 1998). After several hours of this stage the cell becomes irreversibly attached (Marshall et al., 1971).

Irreversible attachment occurs as cell clusters start to form, with multiple bacteria bound to both each other and the surface, while a greater volume of matrix is produced. As the population density increases, quorum sensing systems become active. Quorum sensing systems allow bacteria to sense the concentrations of various molecules related to population size and density, which will be explored in more detail.

As the biofilm matures, further layers are formed, resulting in waste and nutrient gradients. As maturation progresses, additional quorum sensing systems are activated (Sauer et al., 2002). Within this stage, formation of cell clusters known as microcolonies occurs (Klausen et al., 2003a). Biofilm structures such as water channels form and growth continues until the maximum thickness is achieved. At this growth stage the biofilm morphology may now have large tower or mushroom like structures (Sauer et al., 2002; Stoodley et al., 2002). The final stage of the biofilm phenotype is that small parts of the

biofilm now disperse either as small chunks of biofilm or as individual planktonic cells. These cells can seed biofilm formation elsewhere, and for *S. pneumoniae* may be considerably more virulent than their parents (Marks et al., 2013; Stoodley et al., 2002).

As the biofilm forms and matures, colony variants are produced creating heterogeneity even in biofilms produced from a single initial colonising cell. These sub populations of colony variants can have very different phenotypes to the wild-type parental strain. These phenotypes can play clinically important roles if they display increased biofilm formation or virulence (Churton et al., 2016; Syk et al., 2014).

As explained previously, one of the most important factors for the level of biofilm production is the presence of a capsule. The presence or absence of a capsule can mask multiple other factors that impact biofilm production, with certain receptors, sugars and proteins unable to extend past a thick capsule, and therefore unable to engage in extracellular interactions (Munoz-Elias et al., 2008). There are 97 known capsule serotypes of *S. pneumoniae* (Geno et al., 2015; Henrichsen, 1995) and each of the different serotype capsules will have different binding properties, and thicknesses. However, generally, capsule expression results in a 30-60% reduction in biofilm formation (Moscoso et al., 2006). Recognition of parallel evolution of mutants that have altered capsule expression may be important for future vaccine design, as if it is also possible to target capsule switching then it may reduce capsule-based vaccine escape.

### 1.4.3 Biofilm Structure and Physiology

The extracellular matrix of biofilm is made up of numerous components that vary on the species, nutrient availability and stresses on the population. DNA, proteins, polysaccharides, and non-cellular material, are common biofilm matrix components. *S. pneumoniae* biofilms are able to form diverse structures dependant on the strains and serotypes involved (Allegrucci et al., 2006). Structures formed vary from the construction of distinct water channels, and large microcolonies (Allegrucci et al., 2006; Marks et al., 2012b) to the formation of honeycomb like structure on polystyrene surfaces, and human epithelial cells as seen in Figures 3 and 4 (Marks et al., 2012a; Moscoso et al., 2006; Moscoso et al., 2009). Thin filaments bind cells to both intercellular matrix and to each other (Moscoso et al., 2006), and void spaces that are likely to have been filled with water

are observed through scanning electron microscopy(Marks et al., 2012a; Moscoso et al., 2006; Moscoso et al., 2009).

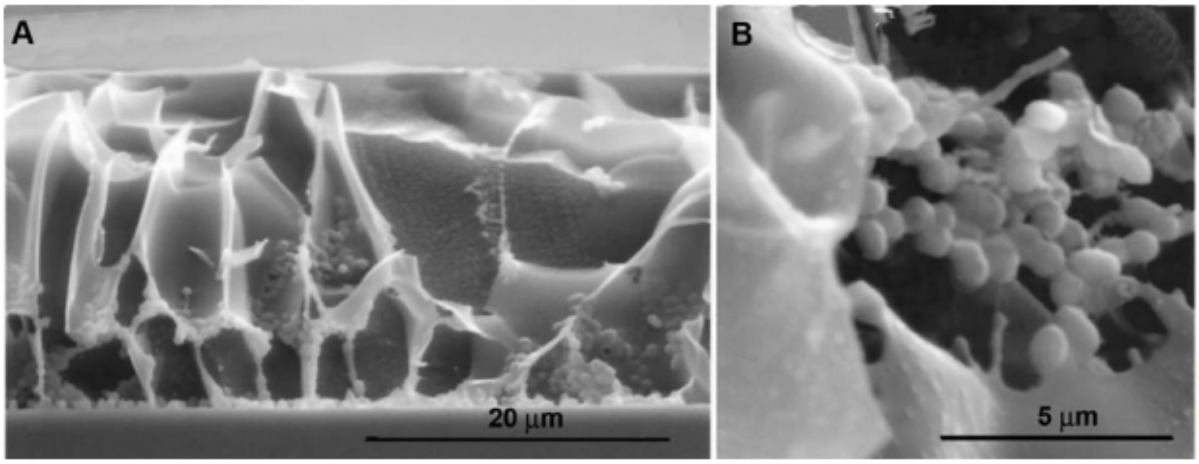



Figure 3: Low temperature scanning electron microscopy of *S. pneumoniae* biofilm structures. **A)** Honeycomb structures that *S. pneumoniae* biofilms form. **B)** Higher magnification of biofilm of cell clusters from within the honey comb structure. Adapted from (Moscoso et al., 2009).

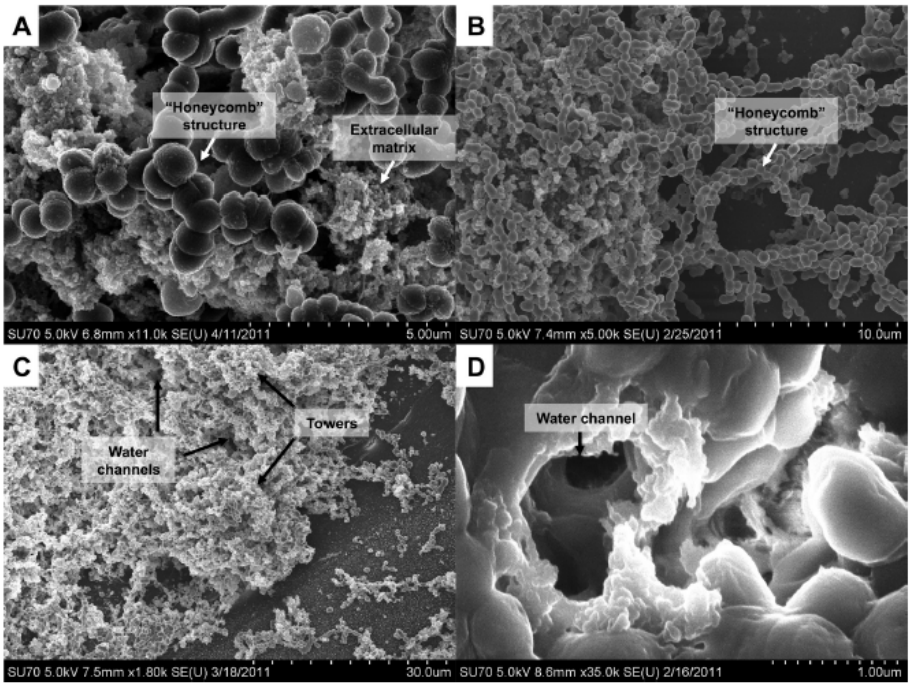



Figure 4: Scanning electron microscopy of *S. pneumoniae* D39 biofilms grown on NCI-H292 human epithelial cells. A) 48hr biofilm formation on ciliated primary bronchial epithelium with the typical honeycomb structure. B) 48hr biofilm formation on epithelial substratum, here both honeycomb structures are visible, and the left hand portion of the image is a tower structure C) 48hr biofilm formation on ciliated primary bronchial epithelium at a lower magnification shows the biofilm tower structures, and water channels D) 48hr biofilm water channel at a higher magnification. Adapted from (Marks et al., 2012a)

*S. pneumoniae* biofilm structures are dependent on the growth conditions under which they are formed. Phenotypic differences are observed between those grown *in vivo* environments such as human epithelial cells, rather than *in vitro* on a glass or polystyrene surface (Moscoso et al., 2006). Furthermore, the same cells grown on the same surfaces have been shown to display altered biofilm structures depending on the concentration of substrates present (Klausen et al., 2003b; Wimpenny and Colasanti, 1997). These potential sources for differences must be considered when comparing results from *in vitro* and *in vivo* experiments. This is pertinent to the observations of SCVs and parallel evolution of *rpoE* mutants across multiple experiments in *S. pneumoniae* 22F (Churton et al., 2016). The SCV *rpoE* mutants were generated through biofilm growth on polystyrene plates. A future avenue of research to be explored is whether the SCV mutations are *in vitro* specific or are generated in more clinically relevant *in vivo* models, as the different conditions may cause *rpoE* mutations to provide altered fitness selection.

Microcolonies are discrete growths of cells within a biofilm, often reproducing at a much greater rate than close biofilm neighbours (Klausen et al., 2003a). This cluster of cells can create conditions within itself which are very different to those experienced in the rest of the biofilm with a lack of nutrients, and severe pH and oxygen gradients being typical (de Beer et al., 1994; Sternberg et al., 1999; Stewart and Franklin, 2008).

These high stress conditions, the increased growth rate, and the greater cell density of the microcolony can cause a significantly higher level of mutant cells to be produced by the microcolony compared to the rest of the biofilm. In this way microcolonies can be seen as an engine for evolution within biofilms, with 100 fold increased mutation rates than planktonic cells (Conibear et al., 2009).

The fluid dynamics that a biofilm is exposed to also impacts on the structure of the biofilm (Domenech et al., 2012). Where there is limited flow and therefore low shear stresses, a biofilm may produce towers and mushroom shaped colonies. However, under significant shear stresses, biofilm structures become elongated with the formation of filamentous streamers with a larger ‘head’ attached to the substratum or other biofilm, and a downstream ‘tail’ that moves around freely with the flow (Stoodley et al., 1999). Streamers elongate over time, with their eventually being only single cells in chains at their tail tips.

In addition to the formation of streamers, higher shear stresses may also result in stronger, smoother, denser and more rigid biofilms being formed (Liu and Tay, 2001; Stoodley et al., 1999). Microcolony structure is maintained under shear stresses, with each microcolony attached to the substratum is able to move independently to shear forces with void spaces between streamers allowing effective nutrient access (Stoodley et al., 1998). In addition to structural changes there may also be changes to the metabolism of biofilms in response to shear forces (Liu and Tay, 2001). *S. pneumoniae* predominantly colonises the nasopharynx where any fluid flow and associated shear stresses are minor, however these stresses should be considered when interpreting the results of experiments utilising flow cell conditions.

### 1.4.4 Components of the Biofilm Matrix

The matrix of *S. pneumoniae* biofilms is formed by a broad range of materials with varying concentrations and formulations dependent on the strain, environment and the available substrates. This section provides a broad overview of the major biofilm structural macromolecules including proteins, DNA, cellular detritus, and exopolysaccharides (Flemming and Wingender, 2010). The diversity of structural material lends itself to the heterogeneity of biofilm morphologies observed for different strains of *S. pneumoniae* (Moscoso et al., 2006; Sutherland, 2001; Wimpenny and Colasanti, 1997).

Extracellular DNA is an important component of *S. pneumoniae* biofilms. A major source of DNA within pneumococcal biofilms is released by *S. pneumoniae* lysis of cells. This mechanism is mediated through LytA and LytC, with mutants for *lytA*, *lytB* and *lytC* observed to form thinner biofilms (Moscoso et al., 2006). However, it has been argued that LytA is not involved, and that stationary phase lysis is due to $H_2O_2$ mediated through *spxB* (Regev-Yochay et al., 2007). Addition of LytA has been shown to disperse biofilms by as much as 80% (Domenech et al., 2011). This may indicate that the lysis provided by LytA is necessary to provide a portion of the material required for biofilm formation, or that controlled lysis is required for the formation of the complex structures required for optimal biofilm formation, but that excessive lysis is detrimental to biofilm formation. Furthermore, addition of DNase 1 has been observed to reduce biofilm formation, supporting the case for a role for DNA in biofilm structure (Moscoso et al., 2006).

However, Moscoso *et al's* observations are challenged by Munoz-Elias *et al* who were unable to find a significant reduction in biofilm formation using a similar system (Munoz-Elias et al., 2008). A possible explanation for the inconsistencies between the two groups given is that the different media used by the groups resulted in differing quantities of DNA being released through competence mechanisms (Moscoso et al., 2009). This highlights the impact of environmental conditions on biofilm composition and is supported by observations that different growth media can result in very different biofilm structures (Klausen et al., 2003b). The importance of extracellular DNA in the biofilms has been shown in several other species, such as *P. aeruginosa* (Yang et al., 2007) and *E.coli* (Devaraj et al., 2015), and others which lends credence to the hypothesis that extracellular DNA plays a role in *S. pneumoniae* biofilm structure (Flemming and Wingender, 2010).

Proteins constitute a major fraction of the EPS (Conrad et al., 2003; Frølund et al., 1996), playing important roles in functions such as structural support, enzymatic functions, and EPS modifications (Flemming and Wingender, 2010). Enzymes within the EPS play varied roles, and some are able to break down macromolecules within the EPS, enabling a cell to utilise the products for its own use, while others may act as virulence factors against a host.

A significant portion of EPS is formed of polysaccharides (PS). PS are large macromolecules whose role in biofilm structures is to act as a binding agent, that holds the biofilm together. PS does not appear to be biofilm specific; however, the level of PS production can significantly differ from levels found in planktonic species. In addition, the concentrations of different PS vary throughout the biofilm itself (Sutherland, 2001). The quantity of PS produced by a biofilm is thought to be dependent on substrate composition, concentration and environmental conditions (Sutherland, 2001; Wrangstadh et al., 1990). PS are also thought to play a major role in maintaining biofilm structure hydration through their ability to bind high volumes of water, inhibiting desiccation (Sutherland, 2001).

PS may also play a role in reducing infection or infiltration by bacteria such as *Bdellovibrio*. If an invading cell is unable to produce appropriate enzymes to degrade the PS then it may be unable to effectively penetrate into the biofilm and prey upon its constituent cells (Sutherland, 2001). Many bacteriophage produce PS degrading enzymes, so it may also be the case that EPS provides a protective role against bacteriophage

(Hughes et al., 1998). The high level of variation in concentrations of all the constituents of biofilm matrix ensures that there is a high level physiologically heterogeneity (Stewart and Franklin, 2008) this heterogeneity provides ample opportunity for colony variants to evolve and dominate biofilm niches.

### 1.4.5 Biofilm-Associated Changes to Transcriptomic and Proteomic Expression

*S. pneumoniae* undergoes significant changes in gene expression when adapting to different environments. The transition between planktonic and biofilm lifestyles reflects this with major alterations to the expression of virulence factors and their regulators observed (Allan et al., 2014; Munoz-Elias et al., 2008; Oggioni et al., 2006). Over 80% of proteins studied in *S. pneumoniae* have been observed to display altered expression profiles when growing as biofilms. These changes in expression further alter during maturation of the biofilm with significant differences between 1 and 7 day old biofilms (Allan et al., 2014). Interestingly there are similarities in the expression profiles of planktonic bacteria and 7 day old biofilms, with many of the changes in expression between planktonic and biofilms at 24 hours reversing following growth for a further 6 days (Allan et al., 2014). It has also been observed that *S. pneumoniae* biofilms display increased expression of proteins related to amino acid and sugar metabolism, and this increased expression is apparent in both 1 and 7 day biofilms (Allan et al., 2014). Bacterocin transporters were up-regulated in 7 day biofilms, as was the virulence factor pyruvate oxidase. The stress protein gls24 was up-regulated in both 1 and 7 day biofilms (Allan et al., 2014). These large changes in expression profiles are supported by Allegrucci *et al* who observed that *S. pneumoniae* biofilms display a 59.4 fold greater level of the virulence gene *nanA* than a planktonic control, and a 15.8 fold increase in the virulence regulator *mgrA*. Beyond the changes in virulence factors, the transition from the planktonic to biofilm lifestyle results in the expression of over 700 proteins not observed in the planktonic state (Allegrucci et al., 2006).

Large differences in *S. pneumoniae* gene expression profiles have also been observed between tissue and blood infections in a mouse model. *S. pneumoniae* biofilm, brain, and lung tissue samples displayed broadly similar gene expression profiles, however the gene

expression observed from blood samples significantly differed, displaying characteristics more similar to those observed in planktonic culture (Oggioni et al., 2006).

The levels of change observed in genomic and proteomic expression are not dissimilar to those of other species, a review of this topic found that during biofilm formation levels of expression may be altered for 50% of the proteome when compared to planktonic bacteria as shown below (Stewart and Franklin, 2008).

| | | Fold change of pneumococcal gene expression[a] | | | | |
|---|---|---|---|---|---|---|
| Gene | | Brain | Lung | Blood | Agar | Biofilm |
| Virulence genes | | | | | | |
| *ply* | SP1923 | 0.3 (0.2) | 0.2 (0.1) | 1.3 (0.3) | 0.01 (0.01) | 0.2 (0.1) |
| *pspA* | SP0117 | 0.3 (0.1) | 0.3 (0.1) | 1.3 (0.2) | 0.4 (0.1) | 0.2 (0.1) |
| *cps4A* | SP0346 | 2.2 (0.9) | 1.6 (0.8) | 1.2 (0.2) | 7.0 (1.8) | 1.6 (0.4) |
| *pspC* | SP2190 | 1.5 (0.9) | 1.2 (0.2) | 1.8 (0.4) | 0.7 (0.1) | 0.6 (0.4) |
| *nanA* | SP1693 | 18.8 (12.0) | 16.6 (10.2) | 1.3 (0.7) | 0.3 (0.1) | 59.4 (28.0) |
| *nanB* | SP1687 | 3.8 (0.9) | 3.2 (0.7) | 2.0 (0.3) | 0.3 (0.2) | 4.8 (1.1) |
| *igA* | SP1154 | 3.3 (0.8) | 2.5 (0.7) | 1.1 (0.5) | 0.3 (0.1) | nd |
| *zmpB* | SP0664 | 2.2 (0.9) | 2.2 (1.6) | 1.2 (0.2) | 0.3 (0.1) | nd |
| *zmpC* | SP0071 | 5.2 (1.6) | 4.1 (0.9) | 1.4 (0.3) | 0.3 (0.1) | nd |
| *sodA* | SP0766 | 10.4 (3.1) | 13.9 (10.8) | 1.5 (0.6) | 1.7 (0.9) | nd |
| *nox* | SP1469 | 7.4 (4.8) | 8.7 (9.2) | 1.0 (0.3) | 1.3 (0.8) | 0.07 (0.02) |
| *lytA* | SP1937 | 2.0 (1.1) | 1.3 (0.2) | 2.3 (0.4) | 2.5 (0.2) | nd |
| *prtA* | SP0641 | 0.8 (0.5) | 0.8 (0.2) | 1.3 (0.2) | 0.2 (0.1) | 0.4 (0.3) |
| *ddlA* | SP1671 | nd | 1.8 (0.3) | 2.4 (0.8) | nd | 4.0 (0.6) |
| *pbp2x* | SP0336 | nd | 6.3 (2.2) | 1.3 (0.4) | nd | 7.7 (2.3) |
| Regulators | | | | | | |
| *comA* | SP0042 | 10.0 (1.5) | 7.1 (1.0) | 1.3 (0.4) | 2.0 (1.2) | 7.1 (0.7) |
| *comE, tcs12* | SP2235 | 6.0 (2.1) | 4.1 (1.5) | 1.3 (0.3) | 1.3 (0.5) | nd |
| *comX* | SP0014 | 12.0 (8.4) | 7.3 (4.2) | 1.6 (0.3) | 1.1 (0.1) | 5.1 (0.8) |
| *micA, tcs02* | SP1227 | 2.2 (1.2) | 1.4 (0.6) | 2.2 (0.5) | 1.2 (0.6) | 1.8 (0.6) |
| *pnpR, tcs04* | SP2082 | 3.3 (1.0) | 3.1 (0.7) | 2.0 (0.4) | 1.7 (0.9) | 4.4 (1.3) |
| *ciaR, tcs05* | SP0798 | 1.7 (0.8) | 1.3 (0.4) | 1.5 (0.3) | 1.9 (0.7) | 3.6 (0.3) |
| *blpR, tcs13* | SP0526 | 5.4 (1.0) | 4.9 (2.2) | 3.5 (0.6) | 0.4 (0.2) | 4.0 (1.9) |
| *luxS* | SP0340 | 5.0 (1.4) | 4.3 (0.6) | 2.4 (0.4) | 5.01 (0.4) | 2.8 (0.3) |
| *hrcA* | SP0515 | 0.1 (0.1) | 0.1 (0.1) | 1.2 (0.1) | 0.2 (0.1) | 0.1 (0.03) |
| *ctsR* | SP2195 | nd | 4.8 (2.8) | 1.6 (0.5) | 2.8 (1.6) | nd |
| *mgrA* | SP1800 | 31.7 (12.8) | 18.7 (9.8) | 4.9 (0.8) | 47.4 (21.9) | 15.8 (7.7) |
| *regR* | SP0330 | 4.0 (2.8) | 5.1 (2.5) | 3.4 (0.7) | 8.5 (8.0) | 3.1 (0.8) |
| *stkP* | SP1732 | 3.9 (2.1) | 5.9 (6.9) | 5.3 (3.7) | 0.4 (0.4) | 1.4 (0.1) |
| *dprA* | SP1266 | 12.2 (5.9) | 7.7 (6.9) | 0.9 (0.7) | 0.06 (0.03) | nd |
| *dtxR* | SP1638 | 10.1 (5.9) | 11.7 (9.1) | 1.0 (0.4) | 0.5 (0.3) | nd |
| *msmR* | SP1899 | 8.2 (5.6) | 5.6 (3.1) | 0.6 (0.29) | 0.03 (0.04) | 2.23 (0.6) |

**a.** Fold change in gene expression measured by quantitative real time RT-PCR was calculate using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). The internal control gene was *gyrB* and the reference condition was liquid culture in mid exponential phase. The mean and standard deviation was calculated on the data from three to five infected mice assayed individually. nd, not done.

Figure 5: Alterations in gene expression undergone by *S. pneumoniae* during tissue colonisation and biofilm growth.The fold changes in gene expression are shown for virulence genes, and regulators during growth in different tissues, and media. Adapted from Oggoni et al (Oggioni et al., 2006).

The significant alterations in gene expression in the transition from the planktonic to biofilm lifestyle is likely accompanied by a shift in the evolutionary pressures, creating a suitable environment for the evolution of colony variants. These evolutionary pressures are amplified through the heterogeneous nature of biofilms creating environmental niches in which variants can out compete the parent strain.

### 1.4.6 The Importance of Positioning Within the Biofilm

Biofilms are heterogeneous with gradients of many nutrients, wastes, and biofilm structures combined with colony variants and potentially different species competing for resources. The result of this is that many bacteria have developed methods to better position themselves within the biofilm in order to gain fitness advantages. This means that growth rate of an organism does not automatically indicate an organisms dominance within the biofilm, as a slower growing colony variant or competitor species may be able to manoeuvre more effectively, and colonise the surface or another niche of the biofilm thereby gaining greater resources (Kim et al., 2014).

Kim *et al* have observed this phenomenon in *P. fluorescens* mucoid *rsmE* mutants. The *rsmE* mutants were observed to dominate biofilms through the use of improved spatial structuring and thus were able to better access oxygen (Kim et al., 2014). The pressures of positioning ensure that parallel evolution of mutations to *rsmE* occurred across all 565 mucoid cell colony variants from multiple experiments.

In *P. aeruginosa* biofilms, sessile bacteria may form microcolonies which through rapid proliferation are able to reach towards the surface of the biofilm and potentially a more desirable habitat. Other *P. aeruginosa* are more motile and are able to exploit this motility to relocate, using twitching motions to move over microcolonies to colonize the more nutrient rich surface (Klausen et al., 2003a).

### 1.4.7 Role of Autoinducers, Quorum Sensing, Diffusion Sensing and Efficiency Sensing in *S. pneumoniae* Virulence and Biofilm Development.

Quorum sensing is the phenomenon where microbes release signalling autoinducer molecules and react to the level of this molecule in the environment using specific receptors. Once an autoinducer reaches a threshold concentration then expression of genes associated to it are induced (Parent et al., 2008; Swofford et al., 2015). This phenomenon has been observed in many species including *S. pneumoniae* where it plays a key role in competence and biofilm formation (Galante et al., 2015; Tomasz, 1965).

Biofilms are an ideal environment for the production of autoinducers, as bacterial cell densities within the matrix are likely to be higher than planktonic culture. Furthermore, within a biofilm, diffusion rates are likely to be reduced, increasing the likelihood that autoinducer threshold concentrations are reached. Clonal microcolonies also ensure that highly similar cells are spatially close, aiding in the stimulation of autoinducer related behaviours such as competence.

*S. pneumoniae* has several quorum sensing systems associated with biofilm behaviours. These include BlpABCSRH, CSP/ComABCDE, and LuxS/ Autoinducer 2 (AI-2) systems (Galante et al., 2015). Both CSP/ComABCDE, and LuxS/AI-2 are important pathways for biofilm formation with deletions to either of these systems resulting in significantly inhibited biofilm formation. The *blp* gene cluster is responsible for bacteriocin production (Vidal et al., 2011).

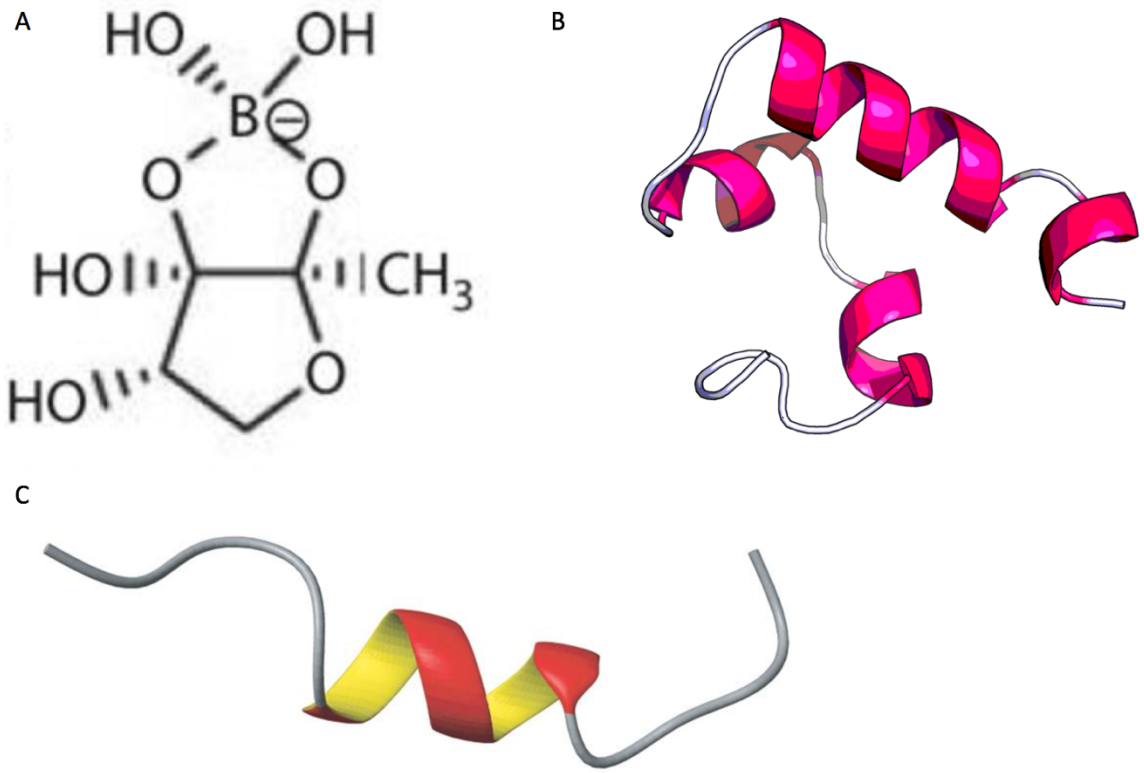



Figure 6: Structures of *S. pneumoniae* auto inducers. **A)** Autoinducer 2 structure adapted from Ferderle et al (Federle, 2009) **B)** BlpC. As far as the author is aware BlpC has yet to be structurally solved. The structure shown is a structure prediction made through the Raptor X structure prediction server (see Chapter 2 for methods) using the D39 genome CP000410.1 (Lanie et al., 2007) **C)** Competence stimulating peptide 1 structure adapted from Johnsborg et al (Johnsborg et al., 2006).

Competence stimulating peptide (CSP)/*comABCDE* quorum sensing system is *S. pneumoniae's* most studied quorum sensing system. Primarily involved in competence for transformation (Straume et al., 2015) CSP has also been shown to be an essential factor for biofilm formation with mutants unable to produce CSP also unable to produce biofilm (Oggioni et al., 2006). However, biofilm formation function may be restored through supplementation with exogenous CSP. The precursor to CSP (pre-CSP) is coded for by *comC* and has several alleles, although the majority of serotypes utilise CSP1 or CSP2. Pre-CSP is actively transported out of the cell, and it is during this transportation that pre-CSP is cleaved to CSP (Galante et al., 2015). Once CSP reaches a critical extracellular concentration, detected through membrane histidine kinase ComD, phosphorylation of ComE occurs (Galante et al., 2015; Martin et al., 2013) promoting further pre-CSP production (Weyder et al., 2018).

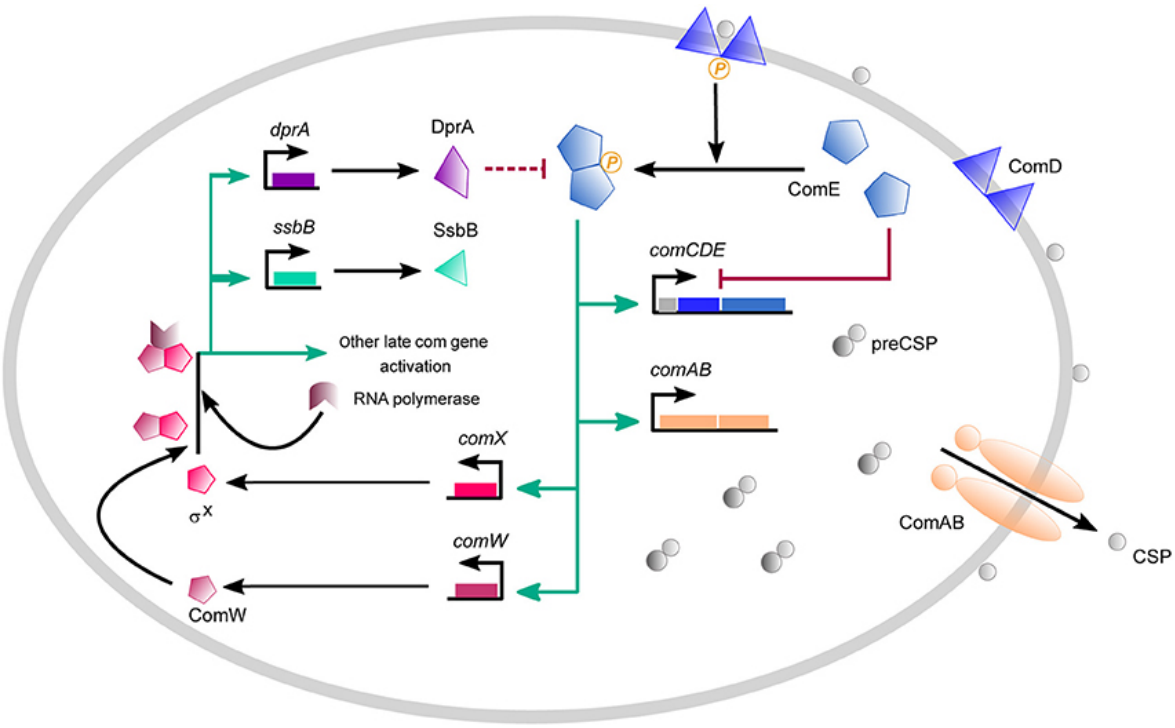



Figure 7: An overview of the competence path way mediated through ComABCDE and regulated by CSP. Competence proteins are now known to be found within the membrane. Adapted from Weyder et al 2018 (Weyder et al., 2018).

There are two known variants of CSP: CSP-1, and CSP-2, with the former being much more common than the latter (Galante et al., 2015), and these variants interact with different ComD proteins. CSP-1 is able to successfully bind to ComD1,3, and 4, while CSP-2 is thought to only able to bind to ComD2. However, work is ongoing in the investigation of the cross reactivity of CSP variants (Galante et al., 2015).

In addition to their promotion of biofilm formation, *comABCDE* genes and their products are crucial to this organism's ability to undergo genetic transformation, an integral trait of *S. pneumoniae* that will be addressed later (Knutsen et al., 2004). ComE acts as a transcription factor for not only *comABCDE,* but also *comX* and the positive feedback loop of CSP works to ensure that all cells reach a competent state simultaneously (Tovpeko and Morrison, 2014).

The Lux S/AI-2 quorum sensing system is present in numerous bacteria and is often thought to be linked to sensing population density both within a single strain, but also across different bacterial species (Pereira et al., 2013). Within *S. pneumoniae*, the Lux S/AI-2 system also plays a role in biofilm formation. Work by Vidal *et al* supports the case that AI-2 is responsible for initial attachment to surfaces, and plays a role in early biofilm formation with reductions of up to 80% in biofilm formation for *luxS* mutants (Vidal et al., 2011). Furthermore, *S. pneumoniae luxS* mutants have also been observed to display reduced virulence in a mouse model, while colonisation of the murine nasopharynx was found to be unimpaired in the *luxS* mutants, transition to the lung and blood was diminished (Stroeher et al., 2003). Interestingly, in *S. mutans* it has been shown that the addition of AI-2 induces 147-fold increases in gene expression of *rpoE* (Sztajer et al., 2008). Gene knock-outs or *rpoE* mutants may therefore be less able to respond to the stresses of high population densities and competitor species.

### 1.4.8 Biofilm Resistance to Antibiotics, Stresses and the Immune System

Bacteria that have formed biofilm communities and structures are much more tolerant of antibiotics and many other stresses than the same species in planktonic state. This increased tolerance can require concentrations of antibiotics that are hundreds to thousands of times more than would be required for eliminating the bacteria in the planktonic state

(Anderl et al., 2000; Ceri et al., 1999; Stewart and Costerton, 2001). Biofilms also provide protection for its inhabitants against the immune system, inhibiting clearance. It is when a biofilm has fully matured with associated microcolonies, structures, and potentially water channels that it is most resistant to antibiotics (Hoiby et al., 2010).

The inability of the immune system or medical intervention to eradicate biofilms enables chronic biofilm infections to occur. A chronic infection is identified by inflammation, and tissue damage at the site of infection, which differentiates it from asymptomatic colonisation which may be neutral or symbiotic.

A significant contributor towards antibiotic tolerance within biofilms is the inherent heterogeneity of cells, and microenvironments. There is a broad range of growth rates within a biofilm, with some cells, particularly those in microcolonies dividing rapidly, while other are dormant, non-dividing or very slow growing. Many antibiotics such as β lactams target growing cells, and when prescribed alone are ineffective against dormant persister cells (Wood et al., 2013), which are capable of growth after treatment has ceased. In tandem with the diversity of cellular activity, there are also gradients of nutrients and waste products within biofilms, which may interfere with drug activity.

Furthermore, biofilms promote variation through increased mutation rates within microcolonies (Conibear et al., 2009). This increased diversity within biofilm populations can protect against changing conditions through the principle of the insurance hypothesis (Boles et al., 2004). In cystic fibrosis, hypermutable phenotypes of *P. aeruginosa* regularly display mutations to DNA repair systems such as the DNA oxidative lesions repair system in genes *mutT*, *mutY*, and *mutM* or in the miss match repair system genes *mutS*, *mutL*, and *uvrD*. Furthermore, the oxidative stress within the biofilm is also thought to promote the hyper mutation phenotypes observed in cystic fibrosis patients (Hoiby et al., 2010). The increased level of mutation increases the likelihood that an antibiotic resistant strain of bacteria will emerge, and a colony variant that spontaneously develops antibiotic resistance is likely to thrive in a medical setting where that antibiotic may be a regular feature of its environment.

## 1.5 Evolution of *S. pneumoniae*

Evolution is the process through which organisms adapt to their environment through heritable changes to the genome. *S. pneumoniae* displays remarkable flexibility in its ability to evolve, and generate variant strains that exploit niches, escape the human immune system, and change capsular serotype in response to vaccines (Andam and Hanage, 2015; Jefferies et al., 2011).

This flexibility is facilitated through a large accessory genome and a pan-genome of DNA available from related strains and species (Muzzi and Donati, 2011; Straume et al., 2015). The non-essential accessory genome of *S. pneumoniae* results in differences of up to 30% between strains. Different combinations of accessory genome are likely to contribute towards the broad range of *S. pneumoniae* phenotypes observed. Furthermore *S. pneumoniae's* has the capability to undergo transformation, enabling it to acquire and integrate DNA from different strains of *S. pneumoniae* and related species. A colonised individual typically hosts multiple strains simultaneously and is likely to be carrying hundreds of strain-specific genes during carriage (Andam and Hanage, 2015; Kilian et al., 2014; Wyllie et al., 2014).

Transformation also enables rapid dissemination of genes that provide a fitness advantage throughout *S. pneumoniae* populations and from populations of related species. *S. mitis*, and *S. oralis* are thought to be responsible for the rapid generation and maintenance of beta-lactam resistance throughout *S. pneumoniae* strains (Jensen et al., 2015). These three streptococci share a common ancestor and it is thought that a shortage of hosts in ancient human communities is likely to have caused divergence resulting in differing evolutionary strategies (Kilian et al., 2008). The most striking differences, are between *S. pneumoniae,* and *S. mitis. S. mitis* has lost many of its pathogenic properties that could challenge the host, and gained greater genome stability (Kilian et al., 2014). Conversely *S. pneumoniae* maintained its pathogenic potential, and retains greater genome plasticity (Kilian et al., 2014). As a result, colonisation by *S. mitis* is less likely to be cleared by the human immune system and lifelong colonisation is common, whereas *S. pneumoniae*'s prevalence in children diminishes as the immune system develops and declines to low levels in healthy adults. Colonisation by non-pathogenic commensals provides a constant reservoir of genetic material for *S. pneumoniae* within the host that may provide genetic material that

enhances survival within the host. However, in a single species system, generation of new genes, or loss of function of current genes, can also cause these impacts and occurs through mutation.

### 1.5.1 Sources of Mutation and Heterogeneity

Mutations occur through errors made by *S. pneumoniae's* DNA replication, or failures in repairing errors caused by exogenous sources. DNA may be damaged through exposure to UV light or reactive oxygen species (ROS). Furthermore, several antibiotics are known to cause an increased rate of mutation decoding errors during translation (Stevens et al., 2011)

ROS play an important role in several *S. pneumoniae* systems such as signalling, homeostasis and control of co-colonising bacteria. Hydrogen peroxide ($H_2O_2$) is the most common ROS that *S. pneumoniae* is exposed to (*S. pneumoniae* produces near suicidal levels of $H_2O_2$ as part of its aerobic metabolism through the enzyme SpxB). *S. pneumoniae* is also deficient in many of the usual $H_2O_2$ regulation systems such as catalase, NADPH peroxidase or OxyR that other bacteria use to counteract oxidative stresses. How *S. pneumoniae* is able to withstand the oxidative stress without the protective mechanisms that are utilised by other bacteria is still unknown (Hua et al., 2014). However, in nature, *S. pneumoniae* co-colonises with numerous other species within the nasopharynx and it may be possible that other species are able to ameliorate the impacts of high levels of $H_2O_2$.

$H_2O_2$ causes damage to cells and DNA through the formation of hydroxyl radicals when reacting with ferrous ions ($Fe^{2+}$). Iron is necessary for growth of *S. pneumoniae*, requiring fine control of intracellular iron concentrations to prevent excessive damage (Imlay et al., 1988). The gene *spxB* is found in the majority of *S. pneumoniae* and codes for pyruvate oxidase, which is responsible for hydrogen peroxide produced as described in 1.3.7. During sepsis of both mice and humans, *spxB* mutants were produced which were unable to produce hydrogen peroxide (Syk et al., 2014). Work by Syk *et al* showed that these mutants grew larger colonies, were less easily cleared, and were more virulent, however they never outgrew the original strain during infection (Syk et al., 2014), which is supported by observations that *spxB* mutants are outcompeted by the wildtype strains in the nasopharynx of a rat model (Regev-Yochay et al., 2007). This indicates that *spxB* confers a

fitness advantage for *S. pneumoniae* during colonisation despite its negative impacts through high levels of $H_2O_2$. The DNA damaging effects of $H_2O_2$ are likely to be compounded by the biofilm lifestyle where microcolonies are likely to create highly mutagenic environments (Conibear et al., 2009).

### 1.5.2 Muller's Ratchet, Horizontal Gene Transfer and Transformation

Mutations are more likely to result in loss of gene function, rather than a provide a new function and so theoretically an organism's genome should degrade over time as an increasingly large portion of genes lose function, and fail to regain it; this theory is known as Muller's Ratchet (Muller, 1964). However many bacteria, including *S. pneumoniae* escape this decline in genomic function through horizontal gene transfer (HGT), involving transformation that allows escape from Muller's Ratchet if sufficient sources of extracellular DNA (eDNA) are available (Takeuchi et al., 2014). Evidence that *S. pneumoniae* utilises transformation to prevent genome degradation is given by observations that increasing the frequency of mutation through deletion of the key DNA repair genes *hexA, hexB,* and *mutB* promotes transformation competence (Gagne et al., 2013; Stevens et al., 2011). Further evidence has also shown that replication stress, such as stalled replication forks, also stimulate both CSP production and competence (Slager et al., 2014). *S. pneumoniae* is also able to acquire suitable DNA for transformation from other *S. pneumoniae* strains and related species such as *S. mitis* and *S. oralis* (Jensen et al., 2015; Muzzi and Donati, 2011).

HGT is the mechanism which enables *S. pneumoniae* to acquire DNA through non-sexual means, either from the environment or from other organisms. This facilitates the dissemination of genetic material amongst a population including facilitating the spread of antibiotic resistance. HGT is a major factor in the evolution of many organisms, but plays a particularly important role in prokaryotes, where it influences the evolution of metabolic pathways, oxygen dependant photosynthesis, and antibiotic resistance, among many other traits (Boto, 2010). *S. pneumoniae* is able to utilize HGT to such an extent that a strain may alter its genome by as much as 7.8% over the course of 7 months (Hiller et al., 2010).

There are four mechanisms through which HGT may occur in prokaryotes: these are transformation, transduction, gene transfer agents (GTA) and conjugation. Conjugation is a

process through which genetic information is passed from one bacterial cell to another directly through bridging structures such as pili. *S. pneumoniae* utilises conjugation to disseminate, and acquire mobile genetic elements known as integrative conjugative elements (ICE). The most common *S. pneumoniae* ICE groups are Tn5253, and Tn916, and have been shown by Roberts *et al* to have been important for the spread of tetracycline and erythromycin resistance throughout streptococcal species (Cochetti et al., 2007; Roberts and Mullany, 2011).

Transduction is the transfer of genetic material from one prokaryote to another by a virus. Similar to transduction is HGT through GTAs, which are thought to be relics of viral particles that have been maintained within a prokaryote's genome. GTA's encapsulate random fragments of host DNA, and promote the export of these encapsulated fragments from the prokaryote, and the uptake into other prokaryotes facilitating transfer of genetic material. There is some discussion as to whether these historic viral mechanisms have been maintained within prokaryote populations due to the advantages that HGT provides, or if they decrease fitness. *S. pneumoniae* are known to host bacteriophage, which are both a source of genetic material, and aid in the dissemination of pneumococcal genes (Croucher et al., 2009). Analysis of 10 bacteriophage that are known to infect *S. pneumoniae* has found that they hosted a variety of virulence genes which may indicate that they should be considered a further reservoir of genetic material that may be exploited by *S. pneumoniae* (Romero et al., 2009).

Transformation is the process for which *S. pneumoniae* is most renown, and is the mechanism by which a cell captures and integrates extracellular DNA (eDNA) from its environment, and integrates it into its genome. The process of transformation is mediated through competence systems which regulate the uptake of eDNA. In *S. pneumoniae* these processes are regulated by environmental conditions and the concentrations of factors such as competence stimulating peptide (CSP). The *comABCDE* genes are responsible for CSP production, quorum sensing and the induction of competence. The gene *comX* is a late stage competence gene which is responsible for the production of the RNA polymerase sigma factor $\sigma^X$ which in turn promotes the uptake and incorporation of exogenous DNA (Galante et al., 2015; Knutsen et al., 2004). $\sigma^X$ promotes the transformation of *S. pneumoniae* by altering the way in which the RNA polymerase binds. The changes allow

RNA polymerase to recognise and bind to an additional promoter sequence called the ComX box. The ComX box sits upstream of 23 genes, but enables the transcription of more than 80. These genes enable successful competence through the killing of non-competent cells, increasing DNA uptake, and promoting homologous recombination of the newly acquired DNA (Straume et al., 2015). $\sigma^X$ requires ComW to facilitate its replacement of the normal σ factor (Tovpeko and Morrison, 2014).

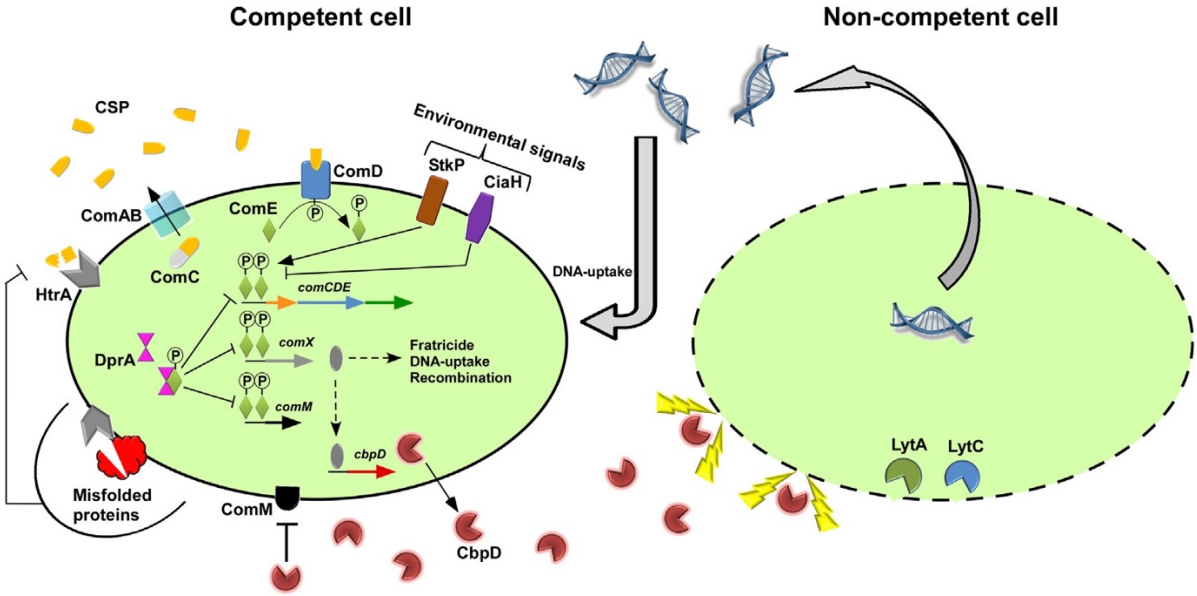



Figure 8: Regulation of competence in *S. pneumoniae* displaying the proteins involved. Adapted from Straume et al (Straume et al., 2015).

In addition to the ComABCDE competence system there are additional layers of competence regulation mediated by the proteins CiaRH, and StkP. CiaRH requires further study to elucidate its full purpose. However, it is thought to play a role in controlling ComABCDE, while StkP plays a role in competence regulation through being required to activate *comCDE* (Echenique et al., 2004). Competence is halted by the protein DprA, which is encoded by late stage competence genes, which inhibits further transformation through binding of phosphorylated ComE, halting the expression of competence genes (Mirouze et al., 2013; Weng et al., 2013).

Transformation can also be carried out artificially under experimental conditions. In this study, the aim is to utilise transformation to create a *rpoE* knock-out. Through careful

manipulation of the experimental environment and the addition of exogenous CSP, *S. pneumoniae* can be stimulated to transform artificial DNA constructs, which can be used to provide insight into the mechanisms of the gene that is altered.

It is notoriously difficult to transform encapsulated clinical isolates of *S. pneumoniae*, as some strains are not transformable through normal methodologies (Evans and Rozen, 2013; Hsieh et al., 2006; Yother et al., 1986). The majority of transformation work has been carried out in a small number of laboratory strains, the majority of which are descended from Avery's unencapsulated strain (R36A) (Avery et al., 1944), for which optimized media and techniques have been developed. The development of purified CSP has improved the breadth of strains that were transformable. One of the principal factors for the control of transformation competence is the quorum sensing pheromone competence stimulating peptide (CSP) (Havarstein et al., 1995). There are two currently known isoforms of CSP (CSP1 and CSP2), and through the use of these peptides, 50% of *S. pneumoniae* strains are transformable (Pozzi et al., 1996). However, a significant fraction of strains remain un-transformable (Evans and Rozen, 2013; Hsieh et al., 2006; Yother et al., 1986). Part of this challenge many involve the time frame for competence. For type strains, the window for transformation is limited to timeframes of less than 40 minutes. However, this may be different for clinical isolates, which fail to transform through regular methods, masking their transformability (Havarstein et al., 1995).

### 1.5.3 Parallel Evolution of *Streptococcus pneumoniae* Colony Variants

Biofilms of many species produce phenotypic variants (Boles et al., 2004; Churton et al., 2016; Cooper et al., 2014), with easily identified phenotypes such as larger or smaller colony diameters, and wrinkly or mucoid morphologies. Small colony variants are associated with poorer clinical outcomes in several species including *Pseudomonas aeruginosa* (Folkesson et al., 2012), *Staphylococcus aureus* (Hoffman et al., 2006; Wolter et al., 2013), *Burkholeria mutivorans* (Silva et al., 2011) which highlights the importance of understanding how these variants occur, and why the patient outcome is worse. The burden of SCVs may currently be underestimated as their often slower growth characteristics and atypical morphology are poorly detected during carriage studies.

Parallel evolution is the phenomenon of multiple individuals of an organism independently evolving the same mutation. Parallel evolution is promoted by positive selection for new mutants which have a fitness advantage for the current environment. There are several studies observing parallel evolution in bacterial biofilm communities. McElroy *et al* (McElroy et al., 2014) undertook deep sequencing of *P. aeruginosa*, using both the reference strain PA01, and the clinical isolate 18A, observing multiple instances of parallel evolution in both the PA01 and 18A over a 7 day experimental period. The genes involved related to electron transport, alginate synthesis, cell shape, flagella in 18A and Type IV pili in PA01, and some resulted in major fitness advantages for the organism. The mutations observed were single nucleotide variations (SNV) and deletions for both PA01 and 18A. The mutations were observed at an increased frequency compared to planktonic cultures, supporting the theory that biofilm morphology is crucial to generating the diversity necessary to adapt to environments.

Kim *et al* have shown that over short time scales colonies of *P. fluorescens* grown from a single cell generate mutants which come to dominate the colony within 5 days (Kim et al., 2014). Kim *et al*'s *rsmE* mutants are mucoid variants that are able to gain advantageous positioning compared to their parent strain, with the fitness advantage being lost if the colony structure is artificially disturbed. Kim *et al* found that all 565 mucoid variants sequenced had mutations in *rsmE* with a diverse range of loss of function mutations observed, demonstrating a clear evolutionary trajectory. Furthermore, analysis of the *rsmE* showed that this gene did not have an increased mutation rate relative to the rest of the genome, and indicating that the ET is due to strong natural selection. This demonstrates that powerful selection factors provide fertile ground for parallel evolution of mutations in single genes, and that can mutations enable dominance of a niche.

Further observations show that if an environmental niche is already filled then the likelihood of further evolution of new variants is reduced (Brockhurst et al., 2007). This may support an argument that *S. pneumoniae* colony variants that display reduced virulence, and are known to evolve repeatedly in a niche should be promoted to inhibit the evolution of more virulent niche inhabitants. Currently unpublished work by Professor R Read, at the University of Southampton has shown promising results working on a similar

principal whereby the pathogenic *N. meningitidis* is displaced by the non-pathogenic strain *N. lactamica* with the aim of reducing disease.

Parallel evolution has also been observed by in *Burkholderia cenocepacia* biofilms (Cooper et al., 2014). SCVs were observed across numerous independent experiments to have non-synonymous mutations to *B. cenocepacia wsp* operon homologue, and is likely related to the metabolism of cyclic diguanylate monophosphate (c-di-GMP), which is associated with biofilm formation in *P. aeruginosa*. *B. cenocepacia* mutants displayed increased biofilm formation, and this is hypothesised to be due to mutations causing autoinduction of the WspA promoting increased biofilm formation. Increased biofilm formation continued to be observed when these cells were grown in a planktonic state (Cooper et al., 2014). This demonstrates that bacterial parallel evolution can rapidly enhance biofilm formation, and enhance beneficial traits. Similarly, *S. pneumoniae* SCVs have been found to display increased biofilm formation (Churton et al., 2016).

*S. pneumoniae* has been observed by many groups to be capable of parallel evolution, and with similarities to the species discussed, *S. pneumoniae* isolates have been shown to repeatedly evolve variant phenotypes after short periods of growth of as little 24-72 hours (Churton et al., 2016; Churton, 2014; Morona et al., 2006; Syk et al., 2014). Variants may have mutations to one or more genes. Variants display different phenotypes beyond simple visual morphology with the clinically relevant variants displaying enhanced virulence factors that result in higher mortality (Syk et al., 2014).

These traits have been shown to evolve multiple times in parallel different individuals. There is an argument to be made that there needs to be greater recognition of *S. pneumoniae's* ability to predictably, rapidly and in parallel evolve mutants that can display altered phenotypes and virulence characteristics in response to the environment in which they find themselves. These mutations appear to occur when *S. pneumoniae* experiences certain environments and pressures. Therefore, strains observed under planktonic conditions, and labelled non-virulent may in fact have the capability to rapidly evolve virulence characteristics in a predictable manner in clinically relevant environments. This predictable parallel evolution can be described as an 'evolutionary trajectory' (ET), and understanding different strains ET may be important for controlling future virulence.

All *S. pneumoniae* 22F ST433 SCVs were found to have mutations in *rpoE* with a wide range of putative loss of function mutations, suggesting that parallel evolution is the mechanism for this phenotype, rather than horizontal transfer from another strain or clonal expansion of a single mutation event (Churton et al., 2016).

Parallel evolution by *S. pneumoniae* has not only been observed in biofilms but also in blood. *S. pneumoniae* serotype 1 ST288 colonies derived from the blood of mice post infection produced both small and large colony variants after 24-48 hours (Syk et al., 2014). This demonstrates that *S. pneumoniae* is able to rapidly produce colony variants in short time scale without requiring a biofilm to drive the adaptation. Large colony variants were observed to have loss of function mutations to *spxB* and furthermore, samples from different time points displayed different *spxB* mutations in the same mouse. This indicates parallel evolution rather than horizontal transfer of *spxB* mutations through transformation (Syk et al., 2014).

*spxB* large colony variants are more virulent in mice during both intravenous and intranasal infections, with a much higher, and more rapid mortality rate than wild type strains, and *spxB* mutants complemented with *spxB* during infection post intravenous inoculation (Syk et al., 2014). The same experiments carried out in TIGR4 *spxB* knock-out mutants showed similar although reduced results to those see for serotype 1 strain ST228, with slower clearance, and increased bacterial load in the blood (Syk et al., 2014). Related serotype 1 *spxB* mutants derived from human blood are also observed to display increased virulence, and when mixed strain challenges were undertaken *spxB* mutants rapidly out competed the parental strain by growing from 1% of the starting inoculum to 80% of the population at the point of euthanasia. This demonstrates that parallel evolution of a fitness mutation observed in one serotype may be relevant in other serotypes, and also that *S. pneumoniae* colony variants are able to rapidly dominate populations.

*spxB* mutations are thought to confer a fitness advantage through reduced early clearance of the infection by splenic marginal zone macrophages (Syk et al., 2014). However, under biofilm growth conditions, and when competing with other organisms such as

*Staphylococcus aureus*, the $H_2O_2$ production generated through *spxB* confers a fitness advantage and so is selected for.

Mucoid variants (MV) and SCVs have been repeatedly observed in the D39 *S. pneumoniae* strain, after transformations which resulted in the deletion of *cspB* (Morona et al., 2006). SCV were the primary variant observed with only 1% of the variants exhibiting a MV phenotype, and these mucoid variants were shown to have additional mutations in the *cspC* (Morona et al., 2006). However only two of each colony type were sequenced which leaves room for doubt in confirming whether this is a case of parallel evolution. Interestingly, clinical isolates of other serotypes showed identical mutations post transformation in three out of four of the mutations tested. This supports the idea that parallel evolution in one strain may be relevant to other strains.

Parallel evolution of mutations in the *cap3A/cpsA* gene has been observed in serotype 3 *S. pneumoniae* biofilms (Domenech et al., 2012). SCVs and mucoid mutants have been generated in D39 through disruption of the Cps family. Mutations to *cpsD* prevents phosphorylation of CspD through the replacement of amino acid residues in the C-terminal [YGX] repeat of CpsD which prevents its phosphorylation (Morona et al., 2004; Morona et al., 2003). Further work showed that the mucoid variants have additional mutations in CpsC when compared to the SCV. These mutations may enable *cpsC* to act independently of *cpsB*, escaping its regulation (Morona et al., 2006). SCV and mucoid variants were also observed in clinical isolates of serotypes 6A. When sequenced, 75% of mucoid variants displayed the same mutation in CspC indicating parallel evolution (Morona et al., 2006). D39 CspC SCV display a 2-fold greater adherence to human A549 cells than D39 wild-type. The SCV produced less capsule polysaccharide (CPS), although they have the same volume of CPS bound to their cell walls (Morona et al., 2006).

The usual requirement for a functional *comW* gene can be avoided by mutations in *rpoD* (Tovpeko and Morrison, 2014). When screening *comW* mutants for transformation activity, this group observed that all *comW* mutants able to undergo competence related genetic transformation displayed mutations to *rpoD*. RpoD is the main RNA polymerase sigma factor. This work may shed some light on the 22F SCV *rpoE* mutants (Churton et al., 2016), as if mutations to *rpoD* are able to stimulate competence through alternate pathways

with a potential hypothesis that mutations to *rpoE* may similarly enable alternative transformation pathways.

These examples of parallel evolution demonstrate that *S. pneumoniae* appears to be capable of rapid adaptation in response to numerous pressures. Potentially clinically relevant implications are that strains may be able to rapidly generate sub populations resistant to a medical intervention prior to the intervention, for instance a greater production of biofilm. However, if these sub populations are predictable and the evolutionary trajectory of the common virulent strains and the resulting variants are known then there is the potential for medical interventions to be tailored to take these into account. A possible mechanism for this may be through the differing metabolisms of colony variants. The metabolism of colony variants including *S. pneumoniae* 22F Δ*rpoE* SCVs have been observed to be altered when compared to the parent in several cases which may allow for more precise control options (Churton, 2014; Syk et al., 2014). It is clear that parallel evolution occurs in response to numerous changes, several of which are known to be clinically relevant. Understanding where parallel evolution is likely to occur in clinically relevant scenarios will be crucial to better understanding, and combatting *S. pneumoniae* infections, and biofilms.

## 1.6 The δ subunit RpoE of *S. pneumoniae* RNA Polymerase

The RNA polymerase (RNAP) is an enzyme that is essential to all life. RNAP enables RNA to be transcribed from DNA without which there would be no protein production. In Gram-positive bacteria, RNAP is formed of multiple subunits. The conserved core consisting of 2 α, 1 β, 1 β1 and 1 ω subunits provides the principal capabilities for transcription. However, additional subunits also play a role in this enzyme. The σ (sigma) subunit when bound to the RNAP cores enables formation of the RNAP holoenzyme. As a holoenzyme RNAP displays a significantly greater ability to recognise and initiate transcription at promoters. A further δ subunit is present in several Gram-positive bacteria including *Bacillus subtilis, Staphylococcus aureus, S. mutans* and *S. pneumoniae*.

The δ subunit was first discovered in 1975 when it co-purified with parts of the RNAP complex in *B. subtilis* (Pero et al., 1975). Pero et al observed that δ plays a role in the transcriptional specificity of phage genes. The δ subunit (RpoE) of *B. subtilis* was then

determined to be formed of two domains, a conserved structured N-terminal domain and an unstructured C-terminal domain (López de Saro et al., 1995; Motackova et al., 2010).

### 1.6.1 Localisation and Abundance of the δ Subunit

The association of the δ subunit with RNAP has been demonstrated in *B. subtilis* through fluorescent labelling which indicates that δ localises with the core subunits β1 and ω (Doherty et al., 2010). In *S. aureus*, the association between δ and the RNAP core has been observed through SDS-PAGE showing it is likely that δ directly interacts with both β and β1 subunits (Weiss et al., 2014). The interactions between the δ and RNAP core subunits are likely to be mediated through the N-terminal domain as a truncated C-terminal does not inhibit binding between δ and the RNAP core (López de Saro et al., 1995). However, the associations between the RNAP core, holoenzyme and δ are questioned by Prajapati et al who observed that while δ displays an affinity for the RNAP core it does not bind to the holoenzyme (Prajapati et al., 2016a). They argue that δ plays two roles. First it is able to binds to RNA to release it from the RNAP complex and secondly that δ is able to bind to upstream promoter sites which RNAP then interacts with to form open complexes (where the DNA strands have split apart). As these additional functions would likely still be within a close proximity to the β1 and ω this may explain why the fluorescent probes utilised by Doherty et al would observe a high level of localisation, and is consistent with observations of interactions between δ and β, and β1 subunits (Weiss et al., 2014; Weiss and Shaw, 2015).

The abundance of the δ subunit within *B. subtilis* has been measured by two different groups who have found that δ is either at a similar concentration to the RNAP core (Doherty et al., 2010) or at a 5-fold greater concentration (Lopez de Saro et al., 1999). Doherty et al observed copy numbers of the δ subunit of 23020±900 molecules per cell, similar to that observed for the β1 subunit (21700±500) (Doherty et al., 2010), while Lopez et al's observed approximately 10000 δ subunit molecules per cell. Further work is required in this area to elucidate the concentration of the δ subunit, however a higher concentration of the δ subunit relative to RNAP is supported by observations that δ may have an RNAP independent role in binding to upstream promoter elements (Prajapati et al., 2016a).

### 1.6.2 RpoE Mechanism of Action

The function of RpoE within the *S. pneumoniae* strain D39 is listed by the STRING database as:

RpoE is the "DNA-directed RNA polymerase (RNAP) subunit delta; Participates in both the initiation and recycling phases of transcription. In the presence of the delta subunit, RNAP displays an increased specificity of transcription, a decreased affinity for nucleic acids, and an increased efficiency of RNA synthesis because of enhanced recycling" (Jensen et al., 2009).

RpoE causes disruption of RNAP/promoter binding which ensures that only high specificity promoter/RNAP bindings are successful (Rabatinova et al., 2013). Therefore, mutations to *rpoE* may result in ineffective disruption of binding between weaker promoters and RNAP enabling expression of a more diverse set of promoter sites. The mechanism through which RpoE performs this role is through regulating the concentration of initiating nucleoside triphosphate (INTP) through regulation of open complex stability. If the open complex is less stable then a greater concentration of INTP is required to ensure transition into a transcriptionally competent complex, and greater specificity of transcription is achieved (Chen and Helmann, 1997; Rabatinova et al., 2013).

RpoE also plays a role in RNAP recycling. The increased RNAP recycling facilitated by RpoE enables considerably more rapid synthesis of RNA in multi-round transcription (López de Saro et al., 1995; Prajapati et al., 2016a). During transcription where only a single transcription event occurs the presence of RpoE does not increase transcript yield, however in multi-round transcription where an open complex formation may occur then the presence of RpoE facilitates several fold increases in transcription yield. This is likely to be facilitated by RpoE increasing the speed of open complex formation (Prajapati et al., 2016a). Furthermore, RpoE may additionally assist in recycling through release of RNA from the RNAP-DNA-RNA ternary complex for DNA sequences without terminators (López de Saro et al., 1995).

*S. mutans rpoE* mutants have been found to have altered gene transcription within non-coding regions with over 50% of genes in these regions displaying up regulated transcripts

and 20-30% of down regulated transcripts coming from regions considered to be noncoding (Xue et al., 2010). As it is thought that RpoE regulates the disassociation of the RNA polymerase from weakly bound promoters it is possible that in the absence of RpoE the RNA polymerase is then able to bind to the weaker promoters found in the non-coding regions (Xue et al., 2010). The products of non-coding regions may have further regulatory effects upon the rest of the genome.

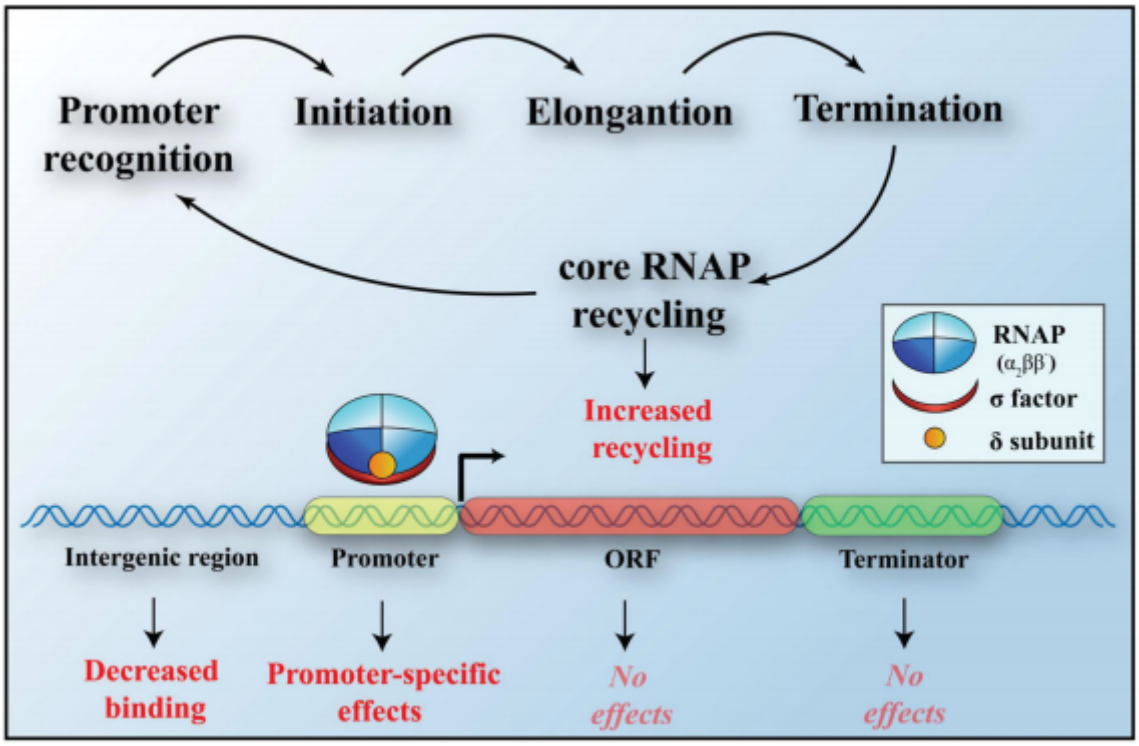



Figure 9: RNA Polymerase (RNAP) activity in gram positive bacteria and the impact that RpoE exerts upon transcription. Binding by RpoE within the promoter region alters transcrition dependent on the area of the binding site. RpoE is able to increase the rate of transcrition through increased recycling by facilitating the release of RNA from the RNAP-DNA-RNA complex. RpoE is also able to release stalled RNAP when an endogenous promoter is not present Adapted from (Weiss and Shaw, 2015).

In addition to recycling, RpoE has also been observed to play a role as a transcriptional regulator within *B. subtilis*. Prajapati et al observed that binding of RpoE to AA rich sequences at -40 is necessary for increased transcript yields at *abrB* promoters (Prajapati et al., 2016a). While decreased transcription occurred when RpoE bound to adenosine rich sequences in close proximity to the -35 region of the promoter likely due to steric clash between RpoE and the RNAP sigma factor (Prajapati et al., 2016b). Transcription may also be repressed by the binding of RpoE to AT sequences at -10 within the promoter region where it is thought that the C-terminal domain of RpoE may interact with RNAP to inhibit propagation of transcription (Aravind and Iyer, 2012; Chen and Helmann, 1997).

### 1.6.3 *rpoE* Mutant Phenotypes

Early observations of *B. subtilis* strains lacking *rpoE* found that removal of this gene did not impact viability or growth under standard laboratory conditions (Lampe et al., 1988). However, under conditions such as low pH, $H_2O_2$ stress or antibiotics, *S. mutans* mutants lacking *rpoE* displayed a reduced capability to tolerate these stresses (Xue et al., 2010). However, in the Gram-negative bacteria *Salmonella enteric* (Serovar Typhi) *rpoE* mutants exhibited increased resistance to a wide range of antimicrobial agents (Xie et al., 2016). The different impacts of *rpoE* mutations between Gram-positive, and Gram-negative bacteria antimicrobial resistance highlights the importance of the Gram type when comparing species.

*S. mutans rpoE* mutants have also displayed differences in biofilm architecture, with greater heterogeneity, and larger cell clusters observed with a greater overall density of cells within the biofilm (Xue et al., 2010). Mutations to *rpoE* also result in alterations to morphology with SCVs observed in *S. pneumoniae* (Churton et al., 2016), *S. mutans* knock-out mutants, and in *B. subtilis*.

*S. mutans rpoE* mutants were observed to grow more slowly than the wild type, with a longer lag phase, and the bacterial cells had a stronger tendency to adhere to one another and form flocs in culture medium (Xue et al., 2010). When RpoE was added to the *S. mutans* culture the majority of the wildtype functions were restored including the growth rate and a lack of floc formation. However, the extended lag phase was still observed

which indicates that the *rpoE* mutation continues to cause effects, but that there may be other contributing factors at play (Xue et al., 2010).

The extended lag phase and reduced ability to mitigate stresses indicate that a loss of RpoE results in a diminished ability to adapt to environmental changes. This is supported by observations in *B. subtilis* where *rpoE* mutants displayed reduced fitness and virulence compared to the wild-type (Rabatinova et al., 2013). Reduced virulence in *rpoE* deficient strains is supported by observations in *Streptococcus agalactiae* where *rpoE* transposon mutants display decreased survival in a neonatal rat sepsis model (Jones et al., 2000), and that increased concentrations of RpoE correlate with increased virulence (Seepersaud et al., 2006).

Virulence and infection are multi-faceted activities that require coordination of numerous genes, it may therefore be unsurprising that *rpoE* deficient mutants which are less able to adapt to a changing environment and stresses are less able to coordinate an effective virulence profile. Furthermore RNA-seq analysis of *S. aureus* reveals that numerous virulence related genes are down regulated in a *rpoE* mutant including toxins, proteases and haemolysins (Weiss et al., 2014).

## 1.7 *S. pneumoniae in vitro* Biofilm Models

*In vitro* methods for the culture of biofilms range from simple microtiter plates to more advanced flow models and bioreactors. Models may incorporate a substratum of human cells to provide a more representative environment with the human colonisation factors (Marks et al., 2012a; Shak et al., 2013).

In *S. pneumoniae* research, a range of media are utilised. Commonly used media included brain heart infusion (BHI), C medium and Todd Hewitt broth (THB), the choice of which may significantly affect biofilm formation. These media may also be supplemented with nutrients such as yeast, casein-tryptone, sugars and amino acids (Moscoso et al., 2006; Trappetti et al., 2009). Further factors to consider include pH and NaCl concentrations, both of which have been observed to cause variations in biofilm formation. Typically, models are run at 34-37°C in a 5% (v/v) $CO_2$ atmosphere to replicate the conditions within the human nasopharynx (Moscoso et al., 2006).

Microtitre plates such as 96 well plates provide a plastic surface on which a biofilm may form. The advantages of this method are the low cost of the assay and the potential for high throughput of results. Biomass from a micotitre assay can be quantified through the use of crystal violet staining. Crystal violet bound to biomass may be re-solubilised and relative biomass determined through optical densities readings (Yadav et al., 2012). Similar to microtite plates, glass bottomed plates may also be used for biofilm culture. Glass bottomed culture enables imaging of biofilms through techniques such as confocal microscopy. Software such as COMSTAT2 (http://www.comstat.dk) may then be used to quantify both live and dead biofilm biomass and structures such as microcolonies. Limitations to this model include a lack of the human host response, variability between wells and with crystal violet a lack of ability to differentiate between live and dead tissue.

Continuous flow systems and bioreactors allow for easier control of nutrient concentrations, cultivation of greater quantities of biomass, protein and polysaccharides. However, these systems can be costlier to run. Furthermore, depending on flow rate and mixing dynamics, shear forces may remove biomass or alter its structure (Allegrucci et al., 2006; Donlan et al., 2004).

Biofilms grown on human tissue provide the advantages of numerous colonisation factors and binding sites. Biofilm growth models involving a substratum of human cells many therefore be more representative of conditions *in vivo*. Limitations of models including human cell culture include the challenges in maintaining the viability of the human cells. This can result in shorter biofilm growth periods than can be achieved on plastic microtitre plates (Marks et al., 2012a; Moscoso et al., 2006).

## 1.8 *S. pneumoniae* Animal Models

When investigating pathogens such as *S. pneumoniae in vitro* methods and results provide valuable information and indications about how a bacteria's phenotype might change and be expressed. However, it is inescapable that *in vitro* models are unable to provide many of the complex signals and substrates that a pathogen takes from its host. To fully investigate the impact of mutations such as those observed in SCVs an *in vivo* model is required

particularly when investigating virulence factors and the potential impact of the mutation upon disease.

The importance of host-pathogen interactions in the validation of *S. pneumoniae* biofilm growth has been highlighted by observations that *S. pneumoniae* biofilms grown on respiratory epithelial cells produced more ordered and less antibiotic sensitive biofilms than those grown using *in vitro* plastics (Marks et al., 2012a). Furthermore, the ability of a strain to form biofilms on respiratory epithelial cells strongly correlated with the ability of strains to colonise the nasopharynx within a mouse model, and demonstrates the value of pseudo *in vivo* models. There are numerous animal models available for use with *S. pneumoniae* each with advantages and limitations, dependent on the pathology to be investigated.

### 1.8.1 Mammalian Animal Models

Mice are the most established animal model for investigating many of the pathologies of *S. pneumoniae*. Numerous inbred strains and well validated methodologies allow for tight control of a broad range experimental variables (Chiavolini et al., 2008). Mice have provided useful information on vaccination efficacy and formulation (Ogunniyi et al., 2000). However, the immunogenicity of a vaccine is dependent on the strain and method of inoculation (Caro-Aguilar et al., 2017).

A drawback of several mouse models is that they are not naturally susceptible to *S. pneumoniae* necessitating careful selection of inbred strains for pneumococcal experiments (Gingles et al., 2001; Jeong et al., 2011; Mizrachi-Nebenzahl et al., 2003). Mouse models show differential abilities to clear infections and recruit neutrophils. The principal explanation for these variation lies in the differences in host immune systems of different models (Benton et al., 1995; Benton et al., 1997a; Benton et al., 1997b). Mutations to the *xid* locus have been identified as increasing susceptibility to *S. pneumoniae* infections in mouse models through a diminished ability to produce the humoral antibody response. The *xid* deficient mice are then unable to produce antibodies against key pneumococcal antigens such as cps and phosphocholine (Kadioglu and Andrew, 2005). Differences between the commonly used mouse strains BALB/c and C57BL/6 mice have been attributed in part to the speed and intensity of the neutrophil activity following exposure.

The BALB/c mice showed higher levels of both phagocytic activity and oxidative burst, thereby reducing susceptibility. The authors attributed this to the Th2 (T helper type 2) background of BALB/c as compared to the Th1 background of C57BL/6 (Kadioglu and Andrew, 2005; Preston et al., 2004). Differences between these two mouse models is supported by research into vaccine immunogenicity. C57BL/6 mice have been observed to recruit dendritic cells and natural killer cells to lymph nodes in response to a wider range of toll like receptor (TLR) agonists than BALB/c mice. Furthermore, the two strains favoured different TLRs for stimulation of natural killer cells and CD4/CD8 T cells (Zeng et al., 2016). Differences between mouse models can be further exacerbated by pneumococcal strain choices. The neutrophil regulator IL-17 has been shown to be either advantageous (serotype 4 TIGR4) or deleterious (serotypes 3 and 6B) to survival of C57BL/6 (IL-17 knock out) mice (Ritchie et al., 2018). The breadth of variation in murine model immune system responses due to differences in genetic background is emphasised by an investigation of 39 inbred strains. This study found considerable variations in the gene expression of into the $CD4^+$ cells and neutrophils. Findings included 22% of the transcriptome varying by 2-fold or more and 119 loci with complete loss of function (Mostafavi et al., 2014).

T regulatory cells responses also vary between murine models in pneumococcal infections. BALB/c mice are more resistant to infection than CBA/Ca following infection with Serotype 2 D39 (Neill et al., 2012). This difference is thought to be due to increased levels of Foxp3+Helios+ T regulatory cells in BALB/c mice, associated with increased transforming growth factor β. These differences are diminished through transfer of T regulatory cells prior to infection (Neill et al., 2012).

The differences between murine model host responses highlight the importance of the choice of mouse model (Chiavolini et al., 2008). However, exposure of laboratory strains to feral or ‘pet shop’ mice may be a method to mitigate these differences. Exposure to the background flora of the non-laboratory is thought to facilitate development of the laboratory mouse immune system and produce immune responses which are more representative of adult humans (Sellers, 2016; Tao and Reese, 2017).

Pneumococcal biofilm formation has been principally studied within murine models. Biofilm dispersal has been studied in both chinchilla (Reid et al., 2009) and rat models (Yadav et al., 2012; Yadav et al., 2017). Gerbils are of similar size to mice and are well characterised for use with OM research, particularly in relation to determining the impact of vaccines, and drugs. Rats and rabbits allow for larger samples to be taken. However, their larger size typically leads to smaller experimental groups and loss of statistical power. Rabbits and mice are well suited to assays such as survival, CFU, histology, and pharmacokinetic assays. Chinchilla models are well suited to investigation of OM due to their large and accessible bullae. Non-human primates (NHP) have the advantage of the highest degree of biologically similarity to humans and have been used to study pneumonia.

The principal limitations of mammalian models are costs associated with housing and maintaining population, and the ethical concerns particularly for NHPs. These limitations rule out higher mammals such as NHPs from all but the most essential research.

### 1.8.2 Zebrafish (*Danio rerio*)

Zebrafish are a well-established as an animal model for the investigation of human pathogens. The tools for investigating Zebrafish are well developed, with gene editing, medium-throughput screening (Stoop et al., 2011) and pharmacological compounds testing protocols available (Bowman and Zon, 2010). The natural transparency of zebrafish allows for real-time study of host-pathogen interactions through the use of fluorescent markers (Jim et al., 2016). Both embryos and adult zebrafish are susceptible to infection by *S. pneumoniae*, and develop meningitis (Jim et al., 2016). Biofilm formation by *Lactobacillus casei* (Rieu et al., 2014) has been investigated within the zebrafish gut.

The physiology of fish is a drawback of the zebrafish model is that they have gills rather than lungs, do not possess lymph nodes, or sites of hematpoiesis. However, Zebrafish possess an innate immune systems which show a high degree of similarity to humans (Saralahti and Ramet, 2015). Furthermore, much like mammalian models, zebrafish require specialized facilities, ethical approval and are accompanied by relatively high costs compared to insect models.

### 1.8.3 *Caenorhabditis elegans*

The nematode worm *C. elegans* is a valuable model for the investigation of environmental pathogens but has only seen limited use as a model organism for the study of *S. pneumoniae* (Garsin et al., 2001) and related Streptococcal species (Jansen et al., 2002). This is largely due the aversion of this model organism to the relatively high temperatures (typically 37°C) required for the investigation of human respiratory tract pathogens. *C. elegans* is able to grow and thrive at temperatures between 12-26°C, and above this temperature, is unable to reproduce and suffers increasing levels of acute heat stress, particularly in the 35-37°C range (Zevian and Yanowitz, 2014). Furthermore, *C. elegans* small size limits the ability to inoculate a worm with a known number of bacteria.

### 1.8.1 *Galleria mellonella*

*Galleria mellonella* (greater wax moth) are increasingly being recognised as a useful animal model for the investigation of human pathogens (Loh et al., 2013; Mukherjee et al., 2013; Wand et al., 2011). Due to the combination of its size, economical costs, viable temperature range and reduced ethical concerns *G. mellonella* provides advantages over other animal models such as *C. elegans*, *D. melanogaster*, Zebrafish and mammalian models (Cook and McArthur, 2013). A full overview of this animal is detailed in Chapter 5.

## 1.9 Project Aims

This project sets out to investigate the role of *rpoE* in *Streptococcus pneumoniae* 22F ST433 and its association with the SCV phenotype. The underlying hypothesis of this research is that mutations to the RNA polymerase delta subunit RpoE are responsible for increased biofilm formation and broad phenotypic changes in *S. pneumoniae* 22F ST433. To explore the function of RpoE we aim to characterise the 22F ST433 wild-type, SCVs and a *rpoE* gene knock-out across a broad range of phenotypic traits to determine the potential importance of the *rpoE* mutations.

The first aim, detailed in Chapter 3, was to develop the methods required to produce *rpoE* knock-outs. As previously stated, all of the SCVs possess mutations to *rpoE*. However, other mutations are observed within the SCVs, albeit none that is present across all isolates. A *rpoE* gene knock-out provides validation that the observed phenotypic differences are due to mutations to *rpoE*. The objective of Chapter 4 was to characterised 22F, SCVs and *rpoE* KO across a range of phenotypic characteristics including biofilm formation, growth and attachment assays and sugar metabolism.

The *in vitro* assays of Chapter 4 provide valuable information on how the isolates behave in *in vitro* conditions. However, where possible *in vivo* models provide a more representative characterisation of the phenotypic traits. In Chapter 5 we aim to develop the *in vitro* phenotyping of Chapter 4 into an *in vivo* system through a pilot study of the animal model *Galleria mellonella*.

The objective of Chapter 6 was to investigate the potential mechanisms through which the phenotypic variations observed might occur. Protein modelling software packages are able to determine the likely structures and functions from proteins. We aim to determine the likely structures and domains RpoE and from these predict likely functions of this protein.

# Chapter 2: Materials and Methods

## 2.1 *S. pneumoniae* Strains, Culture and Phenotyping

### 2.1.1 Strain History

The carriage isolate serotype 22F ST433 isolate was obtained as part of the Southampton paediatric pneumococcal carriage study during the 2008/2009 sampling period (REC No. 06/Q1704/105, RHM MED 0704). SCVs derived from serotype 22F ST433 were received from Dr Nicholas Churton (Churton et al., 2016). The serotype 2 laboratory strain D39 was kindly donated by Professor Vidal (Emory University (Sakai et al., 2013)).

### 2.1.2 Culture Storage

*S. pneumoniae* isolates were initially grown on Columbia blood agar (Oxoid) + 5% (v/v) de-fibrinated horse blood (Oxoid) (CBA) and incubated overnight at 37°C + 5% (v/v) $CO_2$ in air. Following incubation, representative colonies were selected and inoculated into cryogenic vials containing a 1 ml solution of 75% (w/v) brain heart infusion (BHI) (Oxoid) and 25% (v/v) glycerol. These vials were then stored at -80°C.

### 2.1.3 Culture Conditions

Frozen cell stocks were streaked out onto CBA and incubated overnight at 37°C + 5% (v/v) $CO_2$ in air. Following incubation, cells were suspended in 2 ml of media (BHI, 20% BHI (80% Water, 20% standard BHI solution v/v), THY or acidified pH 6.8 THY) to the plate and the colonies suspended through the use of a cell spreader. A 1.5 ml aliquot of the suspension was transferred to a micro-centrifuge tube and centrifuged at 1300 x g to remove lysed cells. 1 ml of supernatant was then added to 9 ml of culture media and incubated at 37°C + 5% (v/v) $CO_2$ in air. For determination of the optical density vs CFU count the optical density was measured every two hours ($OD_{601}$ for BHI, $OD_{550}$ for THY), or more frequently dependant on the rate of culture growth. If there had been a change of greater than 0.050 Abs units, then an aliquot of 200 µl was taken in order to perform a colony forming unit analysis. CFU vs optical density was found to be the same as that found by Dr Nicholas Churton (Churton, 2014).

### 2.1.4 Colony Forming Unit Analysis

Colony forming units (CFU) were assessed through serial dilution of cultured bacteria in sterile culture media. 20 µl of each serial dilution was pipetted in triplicate onto CBA plates. Plates were then incubated overnight at 37°C + 5% (v/v) $CO_2$ in air. Following incubation, the colonies were then counted and converted into CFU/ml.

### 2.1.5 Biofilm Culture - 6 Well Plate

Frozen cell stocks were plated onto CBA and incubated overnight at 37°C + 5% (v/v) $CO_2$ in air. Colonies were selected and inoculated into 20% BHI 80% Water, 20% standard BHI solution v/v) and incubated overnight at 37°C + 5% (v/v) $CO_2$ in air. The overnight culture was used to make suspensions of 1 x $10^7$ cells, which were inoculated into 6-well tissue culture plates, as previously described (Hall-Stoodley *et al*., 2008) for a total of 3 days. Biofilms were grown under static conditions at 37°C + 5% (v/v) $CO_2$ in air and harvested at 1, and 3 days post inoculation. 50% of the culture media was replaced every 24hrs with 20% BHI (80% Water, 20% standard BHI solution v/v). At each time point the supernatant was sampled, and then biofilms were washed twice with 20% BHI (80% Water, 20% standard BHI solution v/v) prewarmed to 37°C. Biomass was harvested using a cell scraper, briefly vortexed and diluted to $10^{-5}$. CFU counts were then carried out from these dilutions and 100 µL was plated onto Columbia blood agar (Oxoid) to assess for changes in colony morphology compared to the inoculum. Selection of variant colonies involved selection of large or small colonies, which were picked off the plate and stored as glycerol stocks for further analysis.

### 2.1.6 Biofilm Culture - 96 Well Plate

Frozen cell stocks were plated onto CBA and incubated overnight at 37°C + 5% $CO_2$ in air. Colonies were selected and inoculated into 20% BHI (80% Water, 20% standard BHI solution v/v) and incubated overnight at 37°C + 5% (v/v) $CO_2$ in air, this culture was used to make suspensions of 1 x $10^7$ cells, which were seeded into Costar 96 well plates, with 200 µl of inoculum per well. For assays lasting longer than 24 hrs the media in each well was refreshed through removal of 50% of culture media, and replacement with fresh 20% BHI (80% Water, 20% standard BHI solution v/v) every 24 hours.

### 2.1.7 Assessment of Colony Size

Images of colonies were taken at 100x total magnification using the Olympus BX51 epi-fluorescence microscope. Colony size was determined through comparison to a 1 mm graticule using the image manipulation software package Image J (Schneider et al., 2012).

### 2.1.8 Crystal Violet Biofilm Quantification

Biofilms were grown according to the methodology found in 2.1.6 for 24 and 72 hours. Following growth, the supernatant was removed, and the biofilm was dried at 60°C for 1 hour. The biofilms were then stained with 200 µl of 0.1% v/v crystal violet for 60 minutes, after which the crystal violet was removed, and each well washed 3 times with 210 µl of PBS. Plates were left for several days to dry at room temperature. Following drying, 200 µl of 30% (v/v) acetic acid was added to each well, and thoroughly mixed through pipetting. 100 µl from each well was then pipetted into a CoStar 96 well plate and measured using a plate reader at absorbance 560 nm.

### 2.1.9 Self Aggregation Assay

This self aggregation assay was adapted from Xue et al (Xue et al., 2011).

Frozen cell stocks of *S. pneumoniae* were cultured overnight on CBA plates at 37°C + 5% (v/v) $CO_2$ in air. Colonies were selected and inoculated into 10 ml BHI and incubated overnight at 37°C + 5% (v/v) $CO_2$ in air. Cultures were then suspended in PBS and adjusted to an $OD_{600nm}$ of 0.6 ±0.05. Cell suspensions were then incubated at 37°C in air. Prior to each time point each sample was cooled to room temperature for 5 minutes. Self aggregation was calculated as $\left(\frac{OD_{600nm} \text{ at time } 0 - OD_{600nm} \text{ at time point X}}{OD_{600nm} \text{ nm at time } 0}\right) \times 100$.

### 2.1.10 Scanning Electron Microscopy

Frozen cell stocks were plated onto CBA and incubated overnight at 37°C + 5% (v/v) $CO_2$ in air. Colonies were selected and inoculated into 20% BHI (80% Water, 20% standard BHI solution v/v) and incubated overnight at 37°C + 5% $CO_2$ in air. Following incubation, cells, were spun at 1300 x g for 3 minutes at room temperature. The supernatant from was used to make suspensions of 1 x $10^7$ cells, which were inoculated into 6-well tissue culture plates with 4 ml per well. A 13 mm circular cover glass (VWR) was placed into each well.

Biofilms were grown for 3 days replacing 2 ml of spent media with pre-warmed 2 ml of 20% BHI (80% Water, 20% standard BHI solution v/v).

On the third day all media was removed, and biofilms were gently washed with 2ml of pre-warmed 20% BHI (80% Water, 20% standard BHI solution v/v). Sterile forceps were used to remove the cover glasses and place them into a 12 well plate containing 1 ml of fixative solution per well. Fixative solution consisted of 3% glutaraldehyde (w/v), 0.1M sodium cacodylate buffer (pH 7.2) and 0.15% (w/v) Alcian blue.

Biofilms were fixed in the fixative for 3 days at 4°C. The initial fixative solution was replaced by 1 ml 0.1 M sodium cacodylate (pH 7.2) and incubated at room temperature for 1 hour. The sodium cacodylate was then removed and replaced with 1 ml of the secondary fixative solution of 0.1 M osmium tetroxide and 0.1 M sodium cacodylate (pH 7.2) and incubated for 1 hour at room temperature. The secondary fixative was removed and replaced with 1 ml of sodium cacodylate (pH 7.2) and incubated at room temperature for 1 hour.

The sodium cacodylate (pH 7.2) was removed and replaced by 1ml of ethanol solutions in water the following concentrations (v/v 20%, 50%, 70%, 95% and 100% ethanol). Each stage was incubated at room temperature for 10 minutes apart from the 100% ethanol concentration which was incubated for 20 minutes. Cover glasses were maintained in ethanol until further processing was carried out.

Cover glasses were critical point dried, mounted onto stubs using forceps and sputter coated with 7 nm platinum. Samples were then placed within the Quanta FEG scanning electron microscope and imaged.

### 2.1.11 Hydrogen Peroxide Concentration

Biofilms were grown according to the method found in 2.1.6 for 24, and 72 hours. Biofilms and supernatant were then mixed through regular pipetting into a single suspension. 50 µl was then aliquoted into an Invitrogen Amplex Red test kit well to measure the relative concentration of $H_2O_2$ production through the use of a plate reader absorbance at 584 nm.

### 2.1.12 Biolog PM1 Sugar Metabolism Assay

Biolog PM1 sugar metabolism assay was carried out as per the manufacturer's instructions with minor adjustments. Frozen cell stocks were plated onto CBA and incubated overnight at 37°C + 5% (v/v) $CO_2$ in air. PM1 inoculation fluid was made through the addition of IF-0a GN/GP 20 ml, Dye mix F (100x) 0.24 ml and PM additive 12x.

Colonies were removed from overnight plates using a sterile loop and gently suspended in IF-0a solution (Biolog) until a transmittance reading of 81% was achieved. This cell suspension was then added to PM1 inoculation fluid at a ratio of 1.76:22.24. 100 µl of this suspension was then added to each of the 96 wells in the PM1 plate. PM1 plates were then incubated for 24 hours at 37°C + 5% (v/v) $CO_2$ in air. Plates were measured using a plate reader at absorbance 570 nm. Results were normalised against the negative control well.

### 2.1.13 Extracellular Matrix (ECM) Binding

Adhesion of *S. pneumoniae* to components of the human extracellular matrix was quantified through the use of the commercially available Chemicon ECM540 Adhesion Array. This array was used in a similar manner to Xue *et al* (Xue et al., 2011) with minor modifications. The ECM540 Adhesion Array features microtitre plate wells coated in collagen I, collagen II, collagen IV, fibronectin, laminin, tenascin, vitronectin, and bovine serum albumin.

Overnight cultures of *S. pneumoniae* were grown in BHI at 37°C + 5% $CO_2$ in air. These overnight cultures were centrifuged at 1300 x g for 3 minutes. The supernatant from this centrifugation was then removed, and 1ml transferred into 10 ml of culture medium before being incubated at 37°C + 5% (v/v) $CO_2$ in air overnight. Each culture was then centrifuged at 1300 x g and adjusted to an $OD_{600}$ nm of approximately 1.0. Following adjustment, 100 µl of each culture was inoculated into each of the pre-treated wells and incubated at 37°C + 5% (v/v) $CO_2$ in air without agitation. Following a 2 hour incubation, the ECM plate was washed twice with PBS and dried for 20 minutes at 60°C. Each well was then stained with 100 µl of 0.4% (v/v) crystal violet at room temperature for 45 minutes. Each well was then washed 5 times with PBS and then dried at 60°C to remove all liquid. Crystal violet was then extracted through the addition of 100 µl of absolute

ethanol and incubated at room temperature in a shaking incubator at 50 rpm. The supernatant (50 µl) was then transferred to a CoStar 96 well plate, and absorbance was measured at 584 nm.

## 2.2 Molecular Biology

### 2.2.1 Primer Design

Primer, and plasmid sequences can be found in Appendix 8.2.1. Primers were designed using NCBI Primer-Blast (Ye et al., 2012). Nucleotide sequences of areas of interest were input into the tool with parameters selected to ensure that the regions of interest were covered. Primer selection parameters included minimum temperatures of 57.0°C, with a maximum temperature of 63.0°C and a target melting temperature of 60.0°C. Primer lengths were targeted towards 20 bp lengths with minimal cross reactivity. Suitable pairs of primers were then selected from the output primer pairings.

### 2.2.2 PCR of Colonies

Colony PCR was used to verify whether a gene was present or absent within an isolate. Representative colonies were selected for polymerase chain reaction (PCR) and the gene of interest amplified using the reagent mix shown in Table 2. PCR products were run out on agarose gels through electrophoresis with a DNA molecular weight marker ladder. Agarose gels consisted of 1% (w/v) agarose dissolved in 1x Tris-Borate-EDTA Buffer (TBE) (Fisher). During electrophoresis the gel was submerged in TBE.

| Reagent | Concentration | Volume (µl) |
|---|---|---|
| Biorad Master mix x2 | - | 25 |
| Forward Primer | 5 µM | 2.5 |
| Reverse Primer | 5 µM | 2.5 |
| DNase/RNase free H2O | - | 19 |
| DNA | - | 1 colony pick |

Table 2: PCR master mix reagents

### 2.2.3 PCR Amplification

| PCR Stage | Process | Temperature | Time<br>Minutes : Seconds |
|---|---|---|---|
| 1 | Initial Denature | 94°C | 04:00 |
| 2 | Denature | 94°C | 00:45 |
| 3 | Primer Annealing | 52°C | 00:45 |
| 4 | Elongation | 72°C | 02:30 |
| 5 | Go to Step 2 x34 | - | - |
| 6 | Hold | 4 °C | ∞ |

Table 3: PCR amplification protocol detailing the temperatures and times used.

### 2.2.4 PCR Product Purification

Linear fragments of DNA produced through PCR for transformation were purified using the QIAquick PCR purification kit.

Five volumes of buffer PB were added to 1 volume of PCR product and mixed. In order to adjust the pH to the correct level, 10 µl of 3 M sodium acetate was added to the buffered PCR solution and mixed through inverting the tube 5 times. The acidified DNA solution was then placed in a QIAquick column with a collection tube and centrifuged for 60 s at 15000 x g. The waste fluid flow through in the collection tube was then discarded. Bound DNA was then washed through addition of 0.75 ml of buffer PE to the QIAquick column. The QIAquick column was then centrifuged for 60 s at 15000 x g, flow through was discarded, and the centrifugation step was repeated. The QIAquick column was placed within a clean 1.5 ml centrifuge tube and 50 µl of molecular RNAse and DNase free water

was added. The DNA was then eluted through centrifugation at 15000 x g for 60 s. Purified DNA was then quantified using a nanodrop.

## 2.3 Plasmid Extraction

### 2.3.1 *S. pneumoniae* Plasmid Extraction

Cells containing the plasmid to be extracted were grown overnight in 24 ml of BHI supplemented with appropriate antibiotic. Cultures of cells were centrifuged 6000 x g for 15 minutes. The supernatant was removed and replaced with 1 ml lysozyme at 5 µg/ml in molecular grade $H_2O$, and vortexed for 30 seconds to suspend. The suspension was incubated at 37°C for 45-60 minutes and the pellet was washed with molecular grade water 3 times, and then suspended in 1 ml of molecular grade water.

Qiagen plasmid prep kits were then used with the following protocol. Each 1 ml tube of cells suspended in molecular grade water was centrifuged at 6000 x g for 15 minutes at 4°C. The supernatant was removed and the cells suspended in 166 µl of Qiagen Buffer P1, to which 166 µl of Qiagen Buffer P2 was added, and vigorously mixed through inverting the tube rapidly 5 times, and incubated at room temperature for 5 minutes. 166 µl of Qiagen Buffer P3 chilled on ice was added to each tube and vigorously mixed by inverting the tube rapidly 5 times and kept on ice for 5 minutes. Each suspension was then centrifuged at 16000 x g for 30 minutes at 4°C. All supernatants were then placed into a single Qiagen-tip 100 pre-treated with 4 ml of Qiagen Buffer QBT. The Qiagen-tip 100 was then washed twice with 10 ml of Qiagen Buffer QC. Plasmid DNA was then eluted through the addition of 5ml of Qiagen Buffer QF pre-warmed to 65°C. DNA was then precipitated through the addition of 3.5 ml of isopropanol at room temperature and centrifuged at 16000 x g for 30 minutes at 4°C. The supernatant was removed, and the pellet was washed 3 times with 70% ethanol and air dried for 10 minutes. The pellet was finally suspended in 100 µl of molecular grade water.

### 2.3.2 *E. coli* EPI300 Plasmid Extraction

Frozen stocks of *E. coli* were inoculated into 4 ml of LB media + antibiotics, and incubated overnight at 37°C with shaking at 200 rpm. 500 µl of the overnight culture was then inoculated into a fresh 4 ml of LB media + antibiotics, and then incubated for 30 minutes at

37°C with shaking at 200 rpm for 30 minutes. Next, 4.5 µl of 1000x induction solution was added to the 4.5 ml of *E. coli* EPI300 culture and incubated for 4 hours at 200 rpm. DNA was then extracted using the Qiagen plasmid extraction method described in 2.3.1.

## 2.4 Transformation Protocol 1

This method is based upon the complete transformation medium developed by Havarstein *et al* (Havarstein et al., 1995), which has been optimised by Vidal *et al* and has been successfully used to transform D39 with pMV158GFP (Vidal et al., 2011).

### 2.4.1 CTM Medium

Complete Transformation Medium (CTM) was produced through the combination of 50 ml THY (Todd Hewitt's + 0.5% (w/v) Yeast), 500 µl of 0.1M $CaCl_2$ and 100 mg of bovine serum albumin (BSA). This solution was sterilised by filtration through a 0.2 µm pore filter membrane.

### 2.4.2 Preparation of Competent Cells

*S. pneumoniae* strains were grown overnight using tryptic soy agar (TSA) (Oxoid) +5% (v/v) sheep blood at 37°C and 5% $CO_2$ in air. Warmed CTM medium (5 mL) was inoculated using the overnight culture from the plate, and then incubated at 37°C to $OD_{600nm}$ 0.6-0.7. This culture was diluted in CTM 1:20 and incubated at 37°C to $OD_{600m}$ 0.35-0.45 without shaking to produce competent cells. Aliquots of 270 µl of competent cells were added to 30 µl glycerol (100%), vortexed for 10 seconds and immediately frozen in liquid $N_2$. These cell stocks were stored at -80°C.

### 2.4.3 Transformation of *S. pneumoniae*

Competent cells were thawed over ice. Aliquots of 0.2 µl 100 ng/µl competence stimulating peptide 1 (CSP1) were added to 180 µl CTM and 20 µl of thawed competent cells were added to this solution followed by either 1 µg of chromosomal DNA or 5 µg of plasmid DNA (in volumes up to 15 µl). The solution was then mixed with a pipette tip, and incubated for 120 min at 37°C. Once incubated, a 100 µl of this solution was pipetted into an empty petri dish, and 15 ml of 48-50°C TSA sheep's blood agar (3% (v/v) sheep blood) with appropriate antibiotic was poured into plate, which was then mixed. Plates were

incubated at 37°C and 5% (v/v) $CO_2$ in air for 24-72 hours, with potential transformant colonies recovered using a sterile loop, and plated onto appropriate antibiotic plates.

## 2.5 Transformation Protocol 2

Due to a lack of success with Professor Vidal's protocol, the method was altered with the help of Dr Calum Johnston (Centre de Biologie Intégrative, Toulose, France), which resulted in successful transformations.

### 2.5.1 Preparation of Competent Cells

Culture of *S. pneumoniae* were streaked onto Oxoid Columbia Blood Agar, and grown at 37°C and 5% (v/v) $CO_2$ in air overnight. Complete transformation medium (CTM) was prepared through the combination of 50 mL THY (Todd Hewitt Broth (Oxoid) + 0.5% (w/v) volume yeast (Oxoid)), addition of 500 µl 0.1M CaCl2 and 100 mg BSA (Fisher). If acidified CTM (aCTM) was required, then this was done with addition of 1M HCl to adjust pH 6.8. The reagents were combined through inverting the tube by hand until all reagents had dissolved.

The CTM was then sterilized by filtration using a 0.2 or 0.22 µm syringe filter. The overnight culture was then suspended in 2 ml of filtered (aCTM), and centrifuged at 1300 x g for 3 minutes. A 1 ml aliquot of the centrifuged supernatant was added to 9 ml of acidified CTM. Cells were then gently mixed through inverting the tube containing them 5 times. This culture was then grown at 37°C with 5% (v/v) $CO_2$ to an $OD_{550nm}$ 0.10-0.15. Once grown, the culture was centrifuged at 6000 x g for 20 minutes to pellet cells, which were then suspended in $1/10^{th}$ volume of CTM (10% (v/v) glycerol) gently with a pipette and stored at -80°C.

### 2.5.2 Transformation of *S. pneumoniae*

A cryovial containing 100 µl competent cells were thawed on ice, and added to 900 ml of CTM, and gently mixed through inverting the tube 5 times. CSP1 was added to competent cells with concentrations of 100-500 ng $\mu l^{-1}$ in a 1 µl volume and incubated at 37°C for 10 minutes. 1-5 µg of plasmid DNA was added per 200 µl of cells, and incubated for 20 minutes at 30-37°C.

Serial dilutions of the competent cells incubated with the plasmid were prepared from neat to $1x10^{-5}$, and 100 µl of this was pipetted into a sterile petri dish and 10 ml of CTM + 1% (w/v) agar, and 5% (v/v) horse blood was pipetted over the competent cells. The cells were incubated at 37°C for 2-3 hours.

After incubation, a further 10ml of CTM blood agar was added over the already set agar. This agar contains twice the desired concentration of antibiotics to eliminate non-transformants and was incubated for 24-72 hours.

Plates were checked daily for transformant colonies. Putative transformant colonies were then streaked onto CBA antibiotic plates supplemented with 2% (w/v) maltose. Growth on these plates was then checked for GFP activity using an Olympus BX51 epi-fluorescence microscope, which would indicate a successful transformation as shown in Figure 14.

## 2.6 *Galleria mellonella* Methods

### 2.6.1 Selection, Inoculation, and Categorisation.

*Galleria mellonella* larvae (Livefoods Limited, Rooks Bridge) each weighing 0.29 ±0.4 g were screened for use. Healthy larvae that were non-melanised, and those that rapidly returned to their original shape following a small compression were selected.

Cultures of *S. pneumoniae* were grown in tryptic soy broth at 37°C and 5% $CO_2$ overnight. Cells were then washed twice in PBS and adjusted to suitable concentrations. Cell suspensions were then placed on ice prior to inoculation. Cell suspensions were plated on CBA plates and incubated overnight at 37°C and 5% (v/v) $CO_2$ in air and CFUs counted.

The outside of each larvae was sterilised through gentle application of 70% (v/v) ethanol to the larvae's outer surface. A syringe containing the required inoculum as a 5 µl dose was then inserted into the rear most left pro-leg and the inoculum injected into the caterpillar. Caterpillars were then incubated at 37°C and 5% (v/v) CO2 in air for the required time course. If antibiotic treatment was required then a 5 µl dose was given through the same rear left pro-leg. PBS was used as a control measure. A caterpillar was deemed to be alive

if it moved in response to minor stimuli, such as being touched by forceps, and was typically heavily melanised as shown in  Figure 10. Beyond categorisation as live or dead no further scoring system was used. A scoring system considering the level of melanisation would be considered for future work.

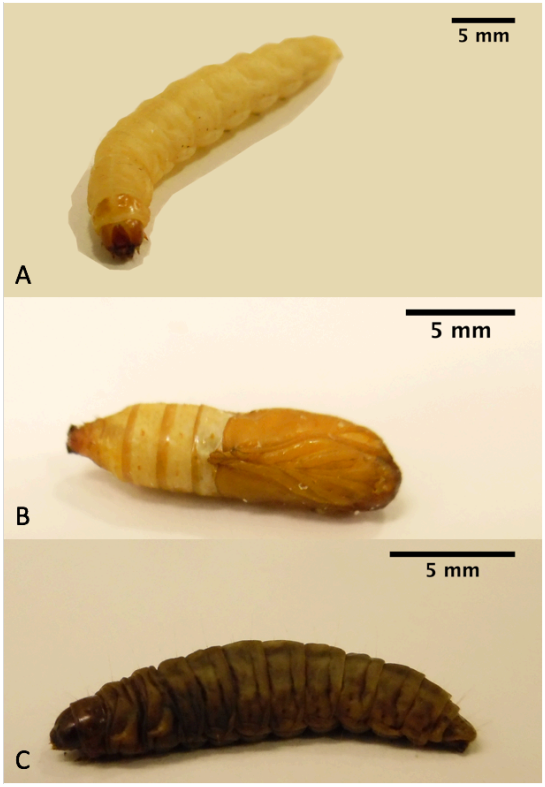



Figure 10: Images of *G. mellonella* at different life cycle stages. **A** is a healthy *G. mellonella* caterpillar without melanisation. **B** shows a healthy caterpillar that has begun to pupate; while pupating the rear end of the caterpillar is able to move and respond to stimuli. **C** shows a caterpillar that has died, the black colouration is due to melanisation which occurs as part of its stress response to infection.

### 2.6.2 *Galleria mellonella* µ-CT Contrast Agent Staining

To provide tissue contrast during micro computed tomography (µ-CT) *G. mellonella* larvae were stained with a contrasting agent. Initially the larvae were dehydrated through immersion in ethanol. Larvae were immersed in 30% (v/v) ethanol for 2 hours, and then immersed for a further 2 hours at 50% (v/v), 70% (v/v), 90% (v/v), and 100% (v/v) ethanol. Following dehydration each larva was immersed in 1% (w/v) phosphotungstic acid (PTA) in 100% ethanol for a further 18 hours.

### 2.6.3 *Galleria mellonella* Tissue Section Biofilm Labelling

Larvae sections of approximately 1 mm in depth were fixed in pH 7.4, 4% (v/v) paraformaldehyde in PBS for 10 minutes, and then washed 3 times with ice cold PBS. The sections were then permeabilised through immersion in PBS + 1% (v/v) Triton X100 for 10 minutes and washed 3 times in room temperature PBS. The sections were then incubated in blocking serum (PBS+ 0.1% (v/v) Tween 20 +1% (w/v) BSA, and 22.52 mg/ml glycine) for 30 minutes at room temperature. Sections were then incubated in diluted primary antibody (5 µg/ml of antibody in PBS+ 0.1% (v/v) Tween 20 +1% (w/v) BSA) at 4°C overnight. Each section was then washed 3 times with room temperature PBS, and then incubated in diluted secondary 10 nm gold nanoparticle conjugate antibody (5 µg/ml of antibody in PBS +1% (w/v) BSA) for 1 hour. Sections were then dehydrated in ethanol in ascending steps of 30% (v/v), 50% (v/v), 70% (v/v), 90% (v/v), and 100% (v/v) spending 2 hours at each step. Samples were then mounted in an Eppendorf tube with 100% ethanol and stored at 4°C.

### 2.6.4 *Galleria mellonella* Whole Organism Biofilm Labelling

Larvae were immersed in 400 µg/ml proteinase K (Fisher) for 30 minutes at 37°C to permeabilize the outside of the larvae facilitating penetration of chemicals in subsequent steps. Following permeabilization, larvae underwent three 5 minute washes in PBS at room temperature. Larvae were then fixed in 4% (v/v) paraformaldehyde in PBS pH 7.4, for 2 hours and were then washed 3 times for 5 minutes with ice cold PBS. The larvae were then permeabilised through immersion in 1% (v/v) Triton X100 in PBS for 10 minutes and washed 3 times in room temperature PBS. The larvae were then incubated in blocking serum (PBS+ 0.1 Tween 20 +1% (w/v) BSA, and 22.52 mg/ml glycine) at 4°C for 24

hours. The larvae were then incubated in diluted primary antibody (5 µg/ml of antibody in PBS + 0.1 Tween 20 +1% (w/v) BSA) at 4°C for 24 hours. Following incubation with the primary antibody larvae were then immersed in PBS at 4°C for 24 hours. The larvae were then incubated in diluted secondary 10 nm gold nanoparticle conjugate antibody (5 µg/ml of antibody in PBS +1% wt/volume BSA) at 4°C for 24 hours. Larvae were then dehydraten in ethanol in stages of 30% (v/v), 50% (v/v), 70% (v/v), 90% (v/v), 100% (v/v) spending 2 hours at each stage. Larvae were finally mounted within an Eppendorf tube in 100% ethanol and store at 4°C.

### 2.6.5 µ-CT Scanning

Larvae were mounted within a 1000 µl pipette tip with the bottom sealed through briefly passing through a Bunsen flame, filled with 100 % ethanol and sealed with para film (Starlab, Milton Keynes). Larvae were scanned with a Zeiss Xradia 510 Versa 3D X-ray microscope, at 80 kV, 6.5 W, 0.4x objective, Binning 2 with 1 second exposure. The resulting scanned images were then compiled through VGstudio and Avizo imaging software packages.

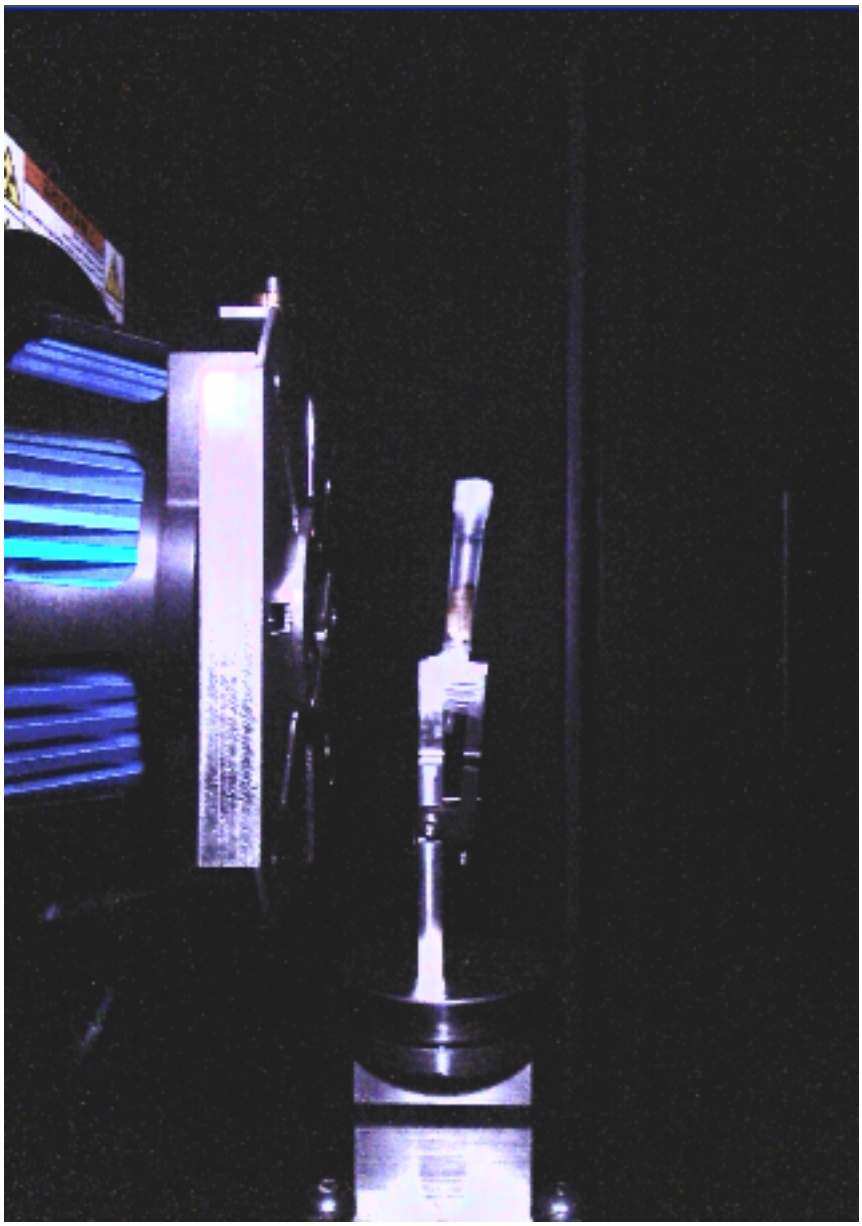


Figure 11: Whole larvae mounted within 1000 µl pipette tip within Zeiss Xradia 510 Versa 3D X-ray scanner. The pipette tip is held in place by a vice.

## 2.7 Molecular Modelling Methods

### 2.7.1 Determination of RpoE Amino Acid Sequence

The nucleotide sequences of *rpoE* determined through next generation sequencing of SCVs by Churton et al (Churton et al., 2016) were submitted to the ExPASy translate tool https://web.expasy.org/translate/. The longest reading 5' reading frame was selected and from this reading frame the first Methionine start residue was selected.

### 2.7.2 Visualisation of Protein Structures

Many programs output protein structure files in formats such as .pdb. These protein structure files were visualised within Chimera software package (Pettersen et al., 2004).

### 2.7.3 Protein Domain Prediction

Amino acid sequences were submitted in the FASTA format to the Robetta Full-chain Protein Structure Prediction Server http://robetta.bakerlab.org. This server then identified the domains within the amino acid sequence.

### 2.7.4 Protein pI Determination

Amino acid sequences were submitted in the FASTA format to ExPASy ProtParam tool. http://web.expasy.org/protparam/.

### 2.7.5 I-TASSER Protein Structure and Function Prediction Programs

Amino acid sequences were submitted in the FASTA format to the I-TASSER protein structure and function prediction server https://zhanglab.ccmb.med.umich.edu/I-TASSER/. The I-TASSER server consists of a group of programs that can be run as an automated pipeline or individually dependant on the requirements of the user.

**LOMETS:** To determine the structure and function of an amino acid sequence initially templates are identified from the PDB archive (https://www.wwpdb.org) by the program LOMETS (Wu and Zhang, 2007). LOMETS utilises 10 threading programs (MUSTER, FFAS-3D, SPARKS-X, HHSEARCH2, HHSEARCH I, Neff-PPAS, HHSEARCH, pGenTHREADER, wdPPAS and PROSPECT2) to generate alignments between the amino

acid FASTA sequence submitted. The top 10 templates for the determination of the structure of the submitted FASTA sequence are produced.

**I-TASSER:** The templates identified through LOMETS are then used by I-TASSER to generate a large number of decoys. The program SPICKER (Zhang and Skolnick, 2004) is then used to cluster decoys based on the structural similarity between decoys. The 5 largest clusters are reported as models. The quality of each model is evaluated through C-score which measures the alignment between the models produced and the target templates. C scores typically ranges between -5 and 2 where higher scores are more likely to represent a correct fold. A C-score of less than -1.5 indicates that the model is unlikely to reflect accurate folding (Yang and Zhang, 2015). I-TASSER also produces a TM-score for the highest quality model. TM-scores are used across protein prediction software and are able to indicate a likelihood of correct protein folding. A TM score of greater than 0.5 indicates that a predicted structure is likely to be correctly folded. Alignments between models or known protein structures can be measured through the program TM-align (Zhang and Skolnick, 2005).

**COACH:** COACH is a meta server that is able to determine ligand binding sites, homology to enzyme structures. Protein structures generated through I-TASSER or other programs may be submitted to the COACH as PDB files. COACH utilises a combination of three programs COFACTOR, TM-SITE and S-SITE. TM-SITE and S-SITE predict ligand binding sites.

**COFACTOR:** Structural models generated through I-TASSER or other programs can be submitted to COFACTOR as PDB files. Submitted structures are threaded against the BioLip protein function database (Yang et al., 2013). Structural homology between submitted models and database entries provides information on ligand binding sites, enzyme commission and gene ontology.

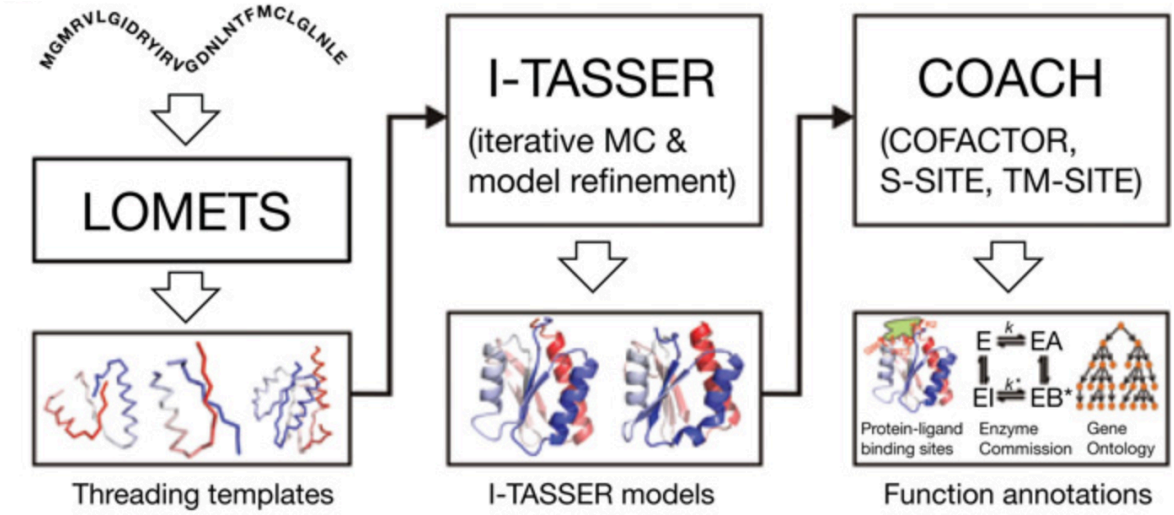



Figure 12: I-TASSER protein structure and function pipeline. Amino acid sequences submitted to LOMETS provide templates for structral models to be produced through I-TASSER which may be analysed through COACH to provide information on the structral models ligand binding affinity, structral homology to enzymes and gene ontology Adapted from Yang et al (Yang et al., 2015)

### 2.7.6 RaptorX Protein Structure and Function Prediction

Amino acid sequences were submitted in the FASTA format to RaptorX protein structure and function prediction servers http://raptorx.uchicago.edu.

Structure prediction was carried out using the online RaptorX server: http://raptorx.uchicago.edu/StructurePrediction/predict/

Property prediction was carried out using the online RaptorX server: http://raptorx.uchicago.edu/StructurePropertyPred/predict/.

Contact prediction was carried out using the online RaptorX server: http://raptorx.uchicago.edu/ContactMap/. The RaptorX contact prediction server utilises a template free (not referencing current databases of models) deep learning model to predict protein contacts. This server produces 5 models with grey scale heat maps of the potential contact sites within a protein with a darker colour denoting a higher likelihood of contact within the protein (Wang et al., 2017). This server also outputs a protein structure .pdb file which can then be viewed through the protein structure visualisation software package Chimera.

## 2.8 Statistical Analysis

All statistical analysis unless otherwise indicated was carried out using GraphPad Prism 7 (www.graphpad.com). Unless stated otherwise a one-way ANOVA was used to determine significant differences between isolates and strains.

The LD50 calculations following inoculations into *Galleria mellonella* were carried out in GraphPad Prism. In brief this was achieved through transformation of the CFU inoculations into log form, and normalisation of the *G. mellonella* responses. A non-linear regression was then carried out for each isolate. Non-linear regressions were then compared through an Extra sum-of-squares F-test. Calculated LD50s with 95% confidence intervals were then compared through a one-way ANOVA. Method details may be found in here (https://www.graphpad.com/www/graphpad/assets/File/Prism%206%20-%20Dose-response.pdf).

# Chapter 3: Transformation of *Streptococcus pneumoniae* WT 22F

## 3.1 Introduction

Transformation is the mechanism through which *S. pneumoniae* is able to incorporate DNA from its environment and related species into its genome. The capability to rapidly utilise foreign DNA facilitates rapid adaptation and uptake of genetic elements which increase fitness (Griffith, 1928; Havarstein et al., 1997). While *S. pneumoniae* is naturally capable of undergoing transformation, artificial control of this process can be used to introduce and remove genetic elements from *S. pneumoniae* genome. However, transformation is a tightly regulated process which requires precise control of experimental parameters to induce a competence state which varies between strains (Yother et al., 1986). Clinical isolates of *S. pneumoniae* such as 22F ST433 are notoriously difficult to transform (Yother et al., 1986) and often require significant method development to produce mutants.

As all SCVs displayed mutations to the gene *rpoE* we investigated whether the phenotypic changes observed are due to a loss of function in this gene using a gene knock-out from the parent strain 22F ST433 (Churton et al., 2016). In this chapter we detail the protocols necessary to transform 22F ST433 with both the plasmid pMV158gfp (Vidal et al., 2011) and a linear construct to knock-out *rpoE* and confirm that the knock-outs display the small colony morphology.

There are currently no published protocols for the transformation of 22F ST433, nor are there any pre-existing strains of *S. pneumoniae* for which *rpoE* has been knocked out. Given the predicted challenges in transforming a clinical isolate such as 22F ST433 we chose the well-defined laboratory D39 for initial method development and protocol validation. D39 was chosen as a suitable strain for this purpose as it has been used for transformation by numerous groups with several protocols available (Havarstein et al., 1995; Shak et al., 2013).

The plasmid pMV158GFP was chosen as a suitable control vector as it has been previously transformed into D39 and confers both green fluorescent protein (GFP) activity and tetracycline resistance. These two attributes allow for selection through the tetracycline resistance and rapid identification through GFP activity. The parent plasmid of pMV158GFP (pMV158) has been demonstrated to have numerous desirable characteristics that make it suitable for use in both D39 and likely to be compatible with 22F ST433. pMV158 has not been observed to significantly inhibit growth, it replicates efficiently, maintains stable inheritance and functions across multiple streptococcal species including *S. pneumoniae* with high promiscuity (Burdett, 1980; Hernandez-Arriaga et al., 2012). Furthermore, Professor Vidal kindly provided a D39 isolate carrying pMV158gfp for use as an additional transformation control. These factors indicate that pMV158GFP is likely to be functional within 22F and is therefore a suitable candidate with which to develop and test methodologies.

## 3.2 Materials and Methods

The materials and methods required to carry out the work in this chapter are detailed in Chapter 2.

## 3.3 Results

### 3.3.1 Plasmid Extraction

An initial challenge for this project was the production of sufficient plasmid for transformation protocols. Standard Qiagen plasmid DNA extraction protocols were shown to be ineffective in producing the concentrations of plasmid DNA required for transformation despite the reported high copy number of the parent plasmid pMV158 (Kramer et al., 1995; Lacks et al., 1986). Qiagen Mini kits were unable to produce a detectable level of plasmid. However, introducing a lysozyme digest increased the plasmid yield to a detectable, although very low level. The use of the Qiagen Midi kit in combination with a lysozyme digest proved successful in recovering viable concentrations of pMV158GFP for use in transformation protocols as shown in Figure 13. This is likely due to the capsule of *S. pneumoniae* inhibiting plasmid extraction during Qiagen protocols, and that its removal through a lysozyme digest improved plasmid extraction.

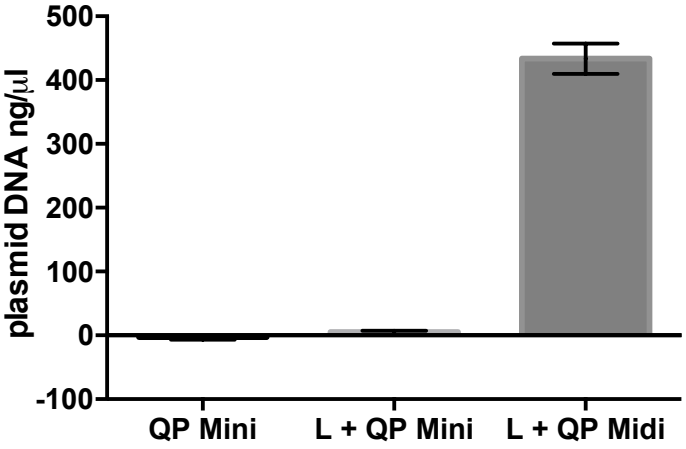



Figure 13: Plasmid pMV158GFP extraction protocol results quantified through the use of a Nannodrop spectrophotometer. QP = Qiagen Protocol, L = lysozyme pre-treatment. Use of lysozyme pre-treatment and a Qiagen Midikit provided a suitable yield that was shown by a one-way ANOVA, with Tukey's Multiple comparison test to be significantly greater than Qiagen Minikit extraction protocols with and without lysozyme treatment and that there was no significant difference between the Minikit protocols. N = 3 independent biological replicates.

### 3.3.2 D39 Generation of a Plasmid Knock-In

Although D39 is a well characterised isolate that has been successfully transformed by many groups, we were initially unable to produce a successful transformant with pMV158gfp using transformation protocol 1 (TP1). Alterations to the protocol led to the development of transformation protocol 2 (TP2) which produced a successful transformant colony verified through both tetracycline resistance selection with 1.25 μg/ml tetracycline, and through GFP activity (Figure 14).

Changes from TP1 to TP2 involved slight changes to the compositions of agar. The removal of the liquid nitrogen freezing step from TP1 and the acidification of CTM. Of these steps it is most likely that acidification of the CTM is the key alteration as this step prevents premature entry into a competent state increasing the likelihood that cells are competent during the experimental time period that knock-in DNA is added.

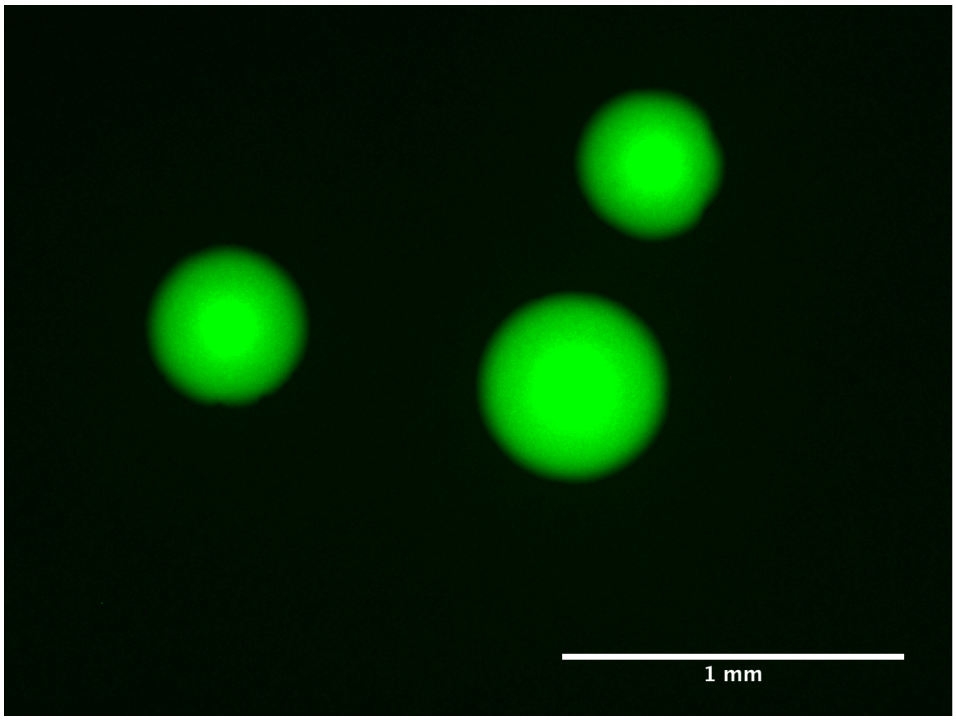



Figure 14: A representative image of colonies of *S. pneumoniae* strain D39 transformed with plasmid pMV158GFP displaying GFP activity. Transformations were carried out using transformation methodology 2 with 500 ng/ml CSP1. A colony was deemed to be transformation positive if it displayed GFP activity.

### 3.3.3 22F pMV158gfp Complementation

Having successfully used TP2 to introduce pMV158gfp to D39 we utilised the same protocol to introduce the plasmid to 22F ST433. The initial transformation protocol was unsuccessful in 22F ST433 with a CSP1 concentration of 500 ng/ml, however increasing the CSP level to 1500 ng/ml resulted in successful transformations. Not all colonies that grew on antibiotic transformation plates were subsequently able to grow as a subculture on a new antibiotic plate nor did they display GFP activity. If a colony was unable to do either of these functions it was not deemed to be a successful transformant.

Our observation that concentrations of 1500 ng/ml were necessary for transformation of 22F ST433 with pMV158gfp is contrary to previous published work (Havarstein et al.,

1995; Morrison, 1997) which observed that concentrations of greater than 100 ng/ml did not increase transformation efficiency. Putative transformants were observed at 500 ng/ml of CSP, which grew on antibiotic resistance plates with 1.25 µg/ml tetracycline, but displayed questionable GFP activity, which may have been gel artefacts combined with non-transformed cells as shown in Figure 16.

Some authors have reported that an increase in the plasmid DNA concentration resulted in increased transformation efficiency (Xu et al., 2011). However, this was not observed in the transformation of 22F with pMV158GFP where no increase in the number of successfully transformed colonies observed, with only one transformant colony observed at 5 µg/200 µl plasmid DNA in Figure 17.

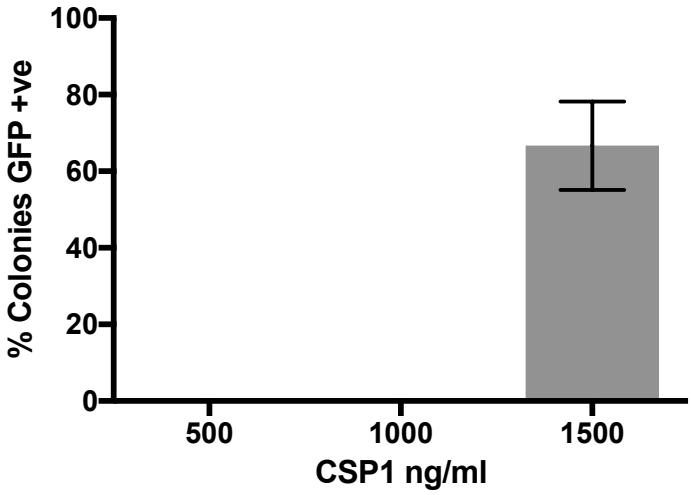



Figure 15: The impact of CSP1 concentrations on the transformation success rates for *S. pneumoniae* 22F WT. Transformations were carried out with 1 µg plasmid DNA per 200 µl of of competent cells, 1500 ng/ml of CSP and 1.25 µg/ml Tet within transformation plates. Transformation efficiency was $1.5x10^{-7}$ cells transforming per µg DNA. A colony was deemed to be transformation positive if it both grew on further antibiotic plates and displayed GFP activity. N = 3 independent biological replicates.

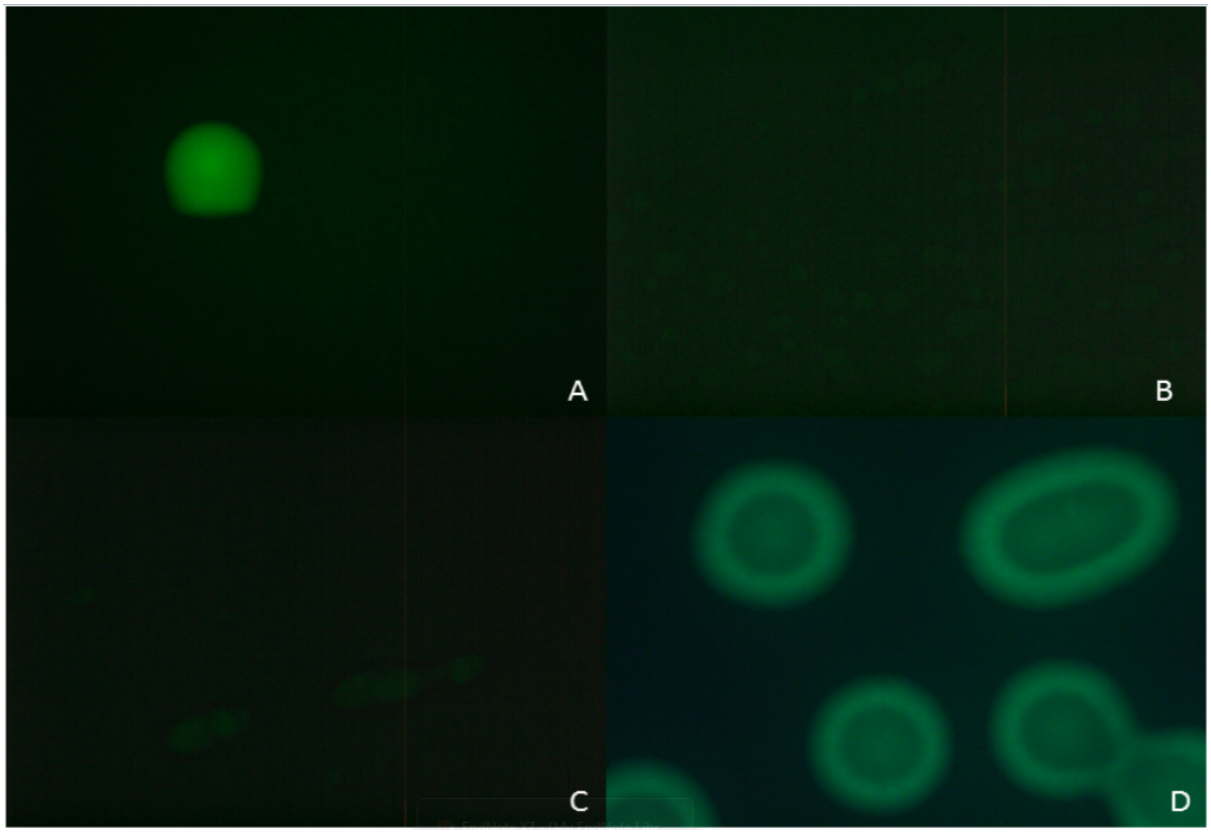



Figure 16: Representative images of GFP activity of *S. pneumoniae* strains with pMV158GFP. **A** shows the positive control D39 with pMV158GFP. **B** shows unsuccessful transformations, that are not fluorescing. **C** shows the a weak putative GFP signal, however this is likely to be a gel artefact. **D** shows the GFP positive successful 22F ST 433 transformed with pMV158GFP. Images taken with an Olympus BX51 epi-fluorescence microscope.

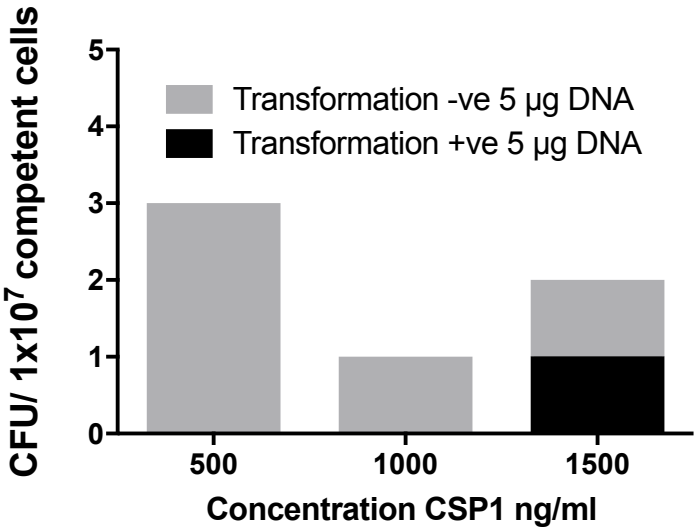



Figure 17: Impact of CSP concentrations on transformation at 5 µg plasmid DNA per 200 µl of cells. A colony was deemed to be transformation positive if it displayed GFP activity. GFP activity was only observed in 1 colony transformed at 1500 ng/ml CSP, 5 µg/200µl pMV158GFP. Transformation negative colonies grew within the initial antibiotic plate during transformation but failed to display GFP activity. N = 1 for this pilot experiment.

### 3.3.4 Generation of the *rpoE* Knock-Out

The strategy for producing the rpoE KO in 22F ST433 involved the introduction of a linear fragment of DNA into the pneumococcal genome in a similar manner to previously published methods (Lau et al., 2002). This approach takes advantage of the process of homologous recombination, through which a linear fragment of DNA with flanking regions matches those of the target gene allowing it to switch places with the wild type gene as shown in Figure 18.

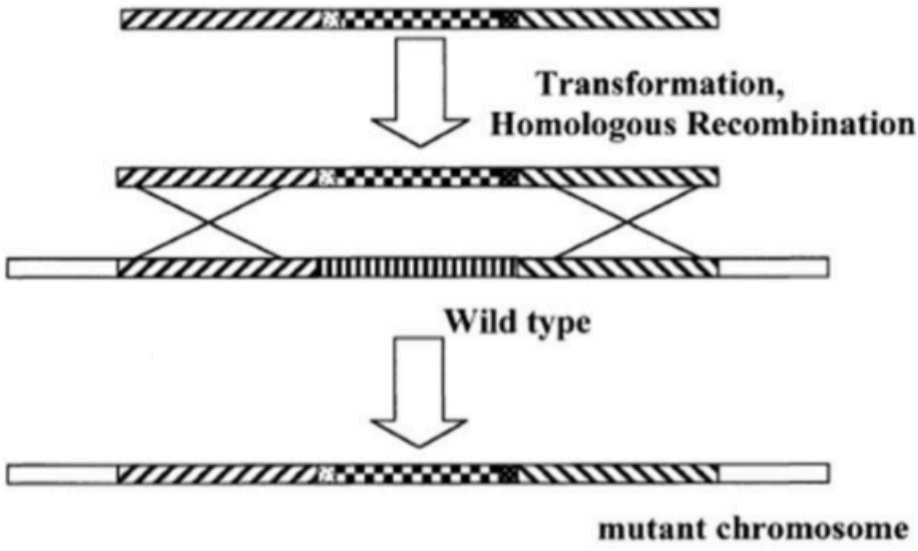



Figure 18: The process of homologous recombination. During this process a linear DNA construct flanked by regions which match the insetrion site may be introduced into the target genome Adapted from Lau et al., (2002).

We designed a linear DNA construct incorporating flanking regions matching those of *rpoE* in the 22F ST433 genome, a kanamycin resistance cassette and a constitutive promoter, as shown below. Following integration this construct would replace *rpoE* and provide the transformant strain with resistance to kanamycin enabling antibiotic selection.

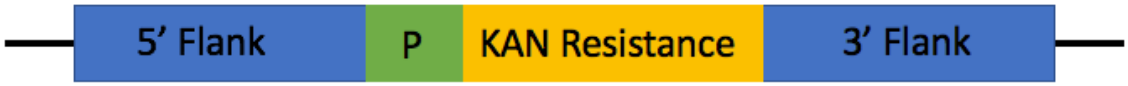



Figure 19: 22F ST433 knock-out linear construct. Flanking regions of 500 bp were placed on either side of the promoter P, and kanamycin resistance gene.

| DNA Sequence | Length (bp) | Function |
|---|---|---|
| 5' *rpoE* Flank | 500 | Natural promoter present |
| Constitutive promoter | 76 | To promote kanamycin resistance |
| Kanamycin resistance gene | 795 | Provides resistance to kanamycin |
| 3' *rpoE* Flank | 500 | Space between genetic elements |

Table 4: Components of the *rpoE* knock-out DNA construct, their lengths and their functions.

During homologous recombination experiments with *Streptococcus mutans* it has been observed that longer flanking regions produce a higher rate of transformation (Lau et al., 2002). Flanking regions of 500 nucleotides were chosen as a suitable length likely to facilitate efficient transformation. The linear fragment was synthesised within the plasmid pCC1BAC (by the company Genscript Ltd), from which it was amplified through PCR to produce sufficient quantities of DNA for transformation. Transformation under the conditions which had been previously utilised to introduce pMV158GFP into ST433 were unable to produce transformant colonies, and so further optimisation was carried out. The key factor for the production of *rpoE* knock-outs were increased incubation times and increased CSP1 concentrations.

Putative transformation colonies were verified through PCR of both the *rpoE* gene (which was absent in all the knock-outs as shown in the figures below) and the inserted kanamycin

resistance gene which was only present in the *rpoE* knock-out colonies indicating successful transformation.

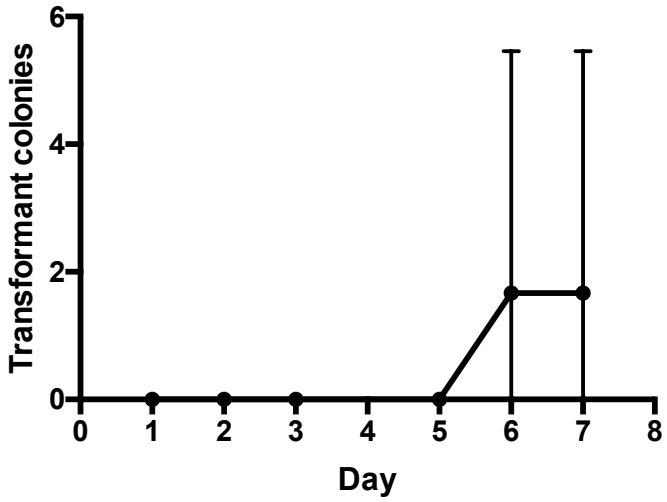



Figure 20: Transformant colonies of 22F ST433 with *rpoE* knocked out were only observed following after 6 days at 2500ng/ml of CSP1. N= 3 independent experiments.

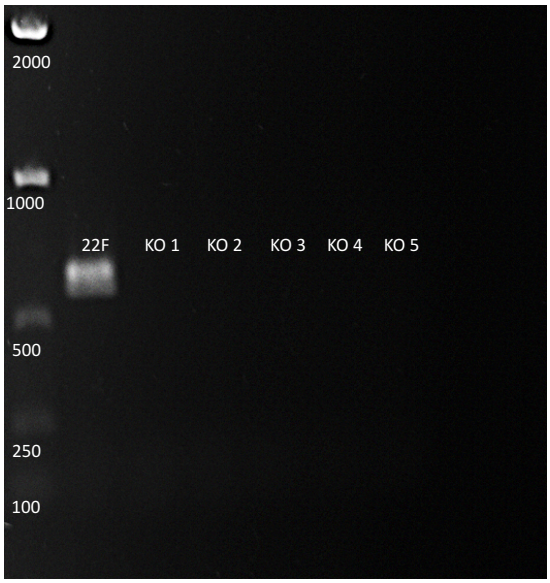



Figure 21: PCR of 22F WT and *rpoE* knock-outs using primers targetting the *rpoE* gene which is knocked out during successful transformation. In all of the knock-outs a band of the appropriate size for the *rpoE* gene is absent, while the band is observed for the 22F WT strain.

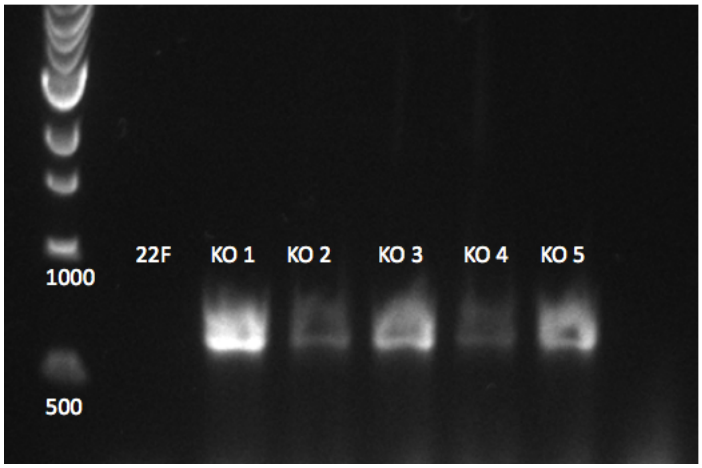



Figure 22: PCR of 22F WT and *rpoE* knock-outs using primers targetting the KAN cassette which replaces *rpoE* during the knock-out transformation. In all of the knock-outs a band of the appropriate size for the KAN cassette is present, while no such band is observed for the 22F WT strain.

The confirmed *rpoE* knock-outs were then grown on CBA to determine their colony diameters, and whether a small colony phenotype could be observed as shown in Figure 23.

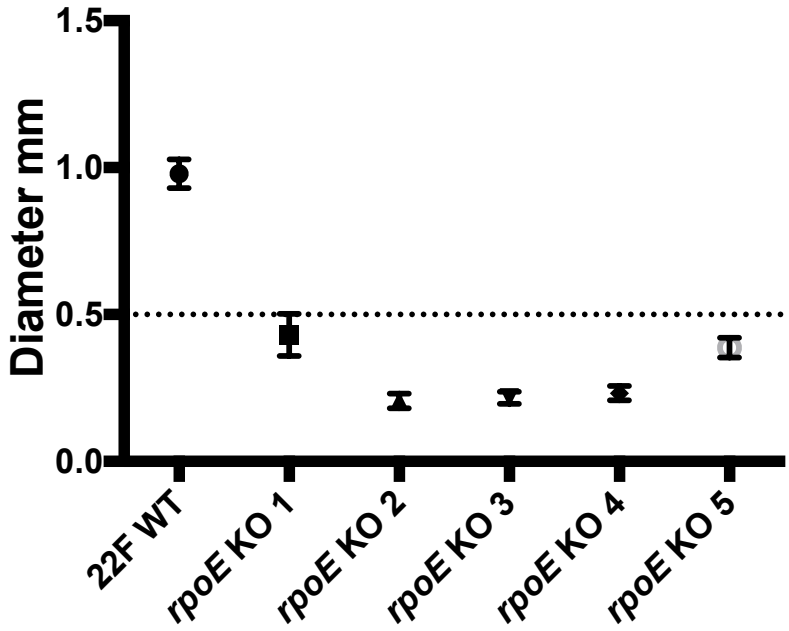



Figure 23: Colony diameter of 22F WT and *rpoE* KO mutants following 24 hours of growth on CBA. N = 5 separate colonies from the same plate.

### 3.3.5 Complementation of 22F *rpoE* KO with pCC1rpoE

To confirm that the small colony morphology observed in the knock-out mutants is due to the loss of *rpoE*, we sought to complement the knock-outs with a plasmid containing *rpoE* and a selection factor. If the phenotypic changes observed in *rpoE* KO1-5 are due to the absence of *rpoE* then complementation of the gene should reverse the phenotype. To achieve *rpoE* complementation, a plasmid was designed to incorporate *rpoE* with small flanking regions and a tetracycline resistance cassette within the plasmid pCC1. This complementation plasmid shall from this point be referred to as pCC1rpoE. The linear construct was designed with *rpoE* and *tet*(M), which was then produced by Genscript Ltd.

The plasmid pCC1 has been successfully used in transformations with D39 (Caymaris et al., 2010) and therefore thought suitable for use with 22F ST433 and *rpoE* KO. The 300 bp long 5' flank was chosen as the wildtype promoter sequence N17 was identified between bases 199, and 228 using the PePPER prokaryotic promoter element and regulon identification server (de Jong et al., 2012).


Figure 24: Diagram of the *rpoE* linear construct within the pCC1rpoE.

| DNA Sequence | Length (bp) | Function |
|---|---|---|
| 5' *rpoE* Flank | 300 | Natural promoter present |
| *rpoE* gene | 588 | Natural *rpoE* complementation |
| 3' *rpoE* Flank | 150 | Space between genetic elements |
| Constitutive promoter | 76 | To promote tetracycline resistance |
| Tetracycline resistance *tet*(M) | 1994 | Provide tetracycline resistance |

Table 5: Function and length of *rpoE* knock-in construct components.

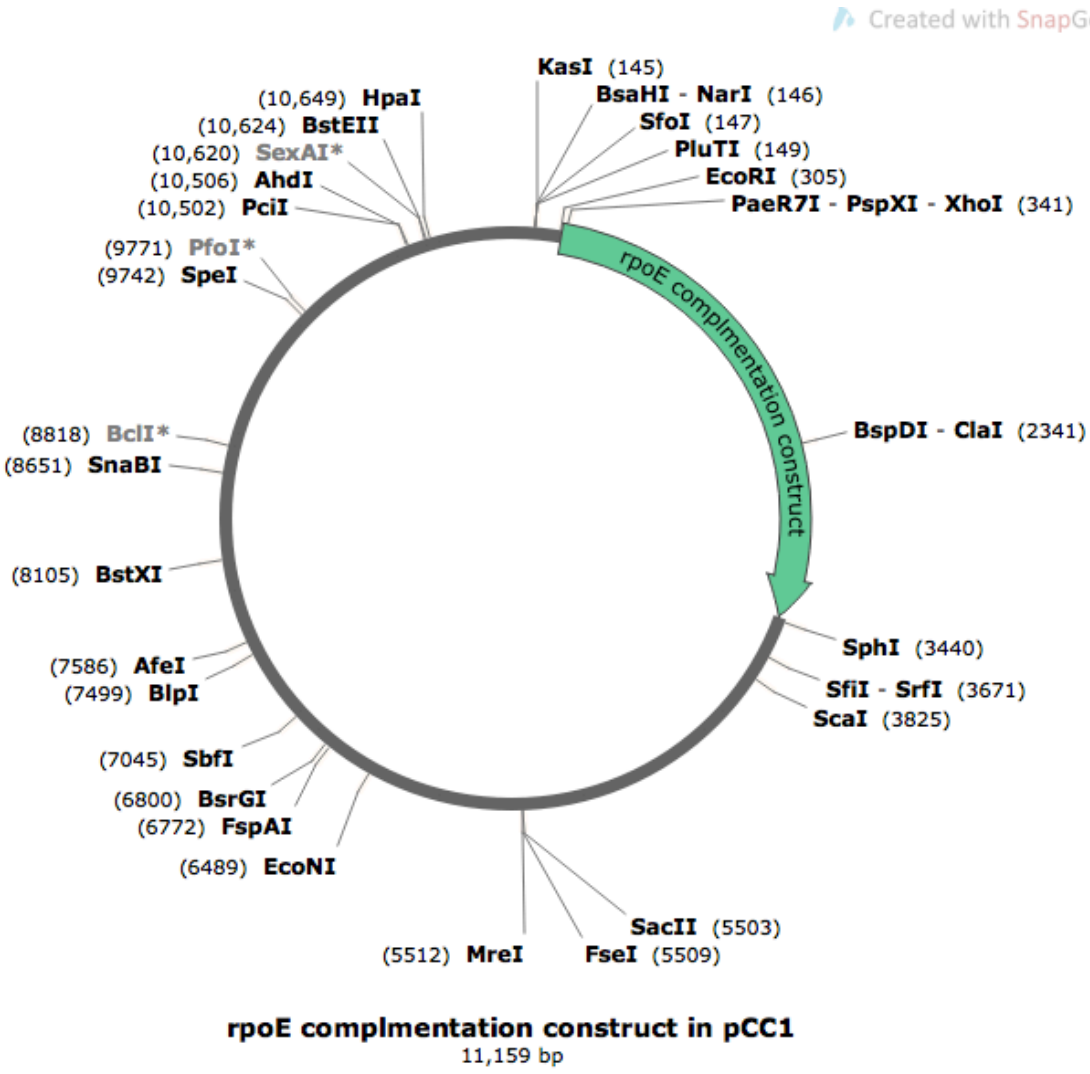



Figure 25: Map of pCC1rpoE.

## 3.4 Discussion

In order to determine whether the phenotypic changes observed in SCVs were due to mutations in *rpoE* causing a loss of function, we generated *rpoE* knock-outs. This chapter details the methods required for the transformation of *S. pneumoniae* D39 and 22F ST433 with the plasmid pMV158gfp and the generation of a *rpoE* knock-out in *S. pneumoniae* strain 22F ST433. To achieve this, we optimised the protocols necessary for the extraction of plasmid DNA from 22F ST433 and developed the methods to transform these strains with either plasmids or linear DNA constructs.

To develop the necessary protocols, we selected the commonly used *S. pneumoniae* type strain D39. D39 benefits from well characterised transformation protocols which provide a starting point from which methods could be developed to transform more difficult clinical isolates such as 22F ST433. The plasmid pMV158gfp was chosen as a suitable vector for method development due to its known compatibility with D39 and its useful selection markers in the form of a tetracycline resistance cassette and GFP production under the control of a maltose inducible promoter. These markers enable effective and rapid confirmation of transformation results (Nieto and Espinosa, 2003; Vidal et al., 2011).

An initial challenge was the production of sufficient yields of pMV158gfp for transformation. Using the standard Qiagen Mini kit protocol only negligible concentrations of plasmid were produced with the majority of samples negative. A lysozyme digest was added to the protocol which resulted in detectable although low concentrations of plasmid DNA. Lysozyme degrades the peptidoglycan cell wall which encapsulates *S. pneumoniae* (Ibrahim et al., 2001; Nash et al., 2006). Degradation of the cell wall facilitates greater release of plasmid from the cell increasing the yield however the plasmid concentrations were not sufficient. In further optimisation we increased the volume of cells for each extraction. Given the high reported copy number of pMV158 we were surprised by the low plasmid yields achieved by the Qiagen Mini kit. Through the use of Qiagen Midi kit we were able to increase plasmid yields to concentrations sufficient for transformation protocols.

Having developed plasmid extraction methods to produce the required volume of plasmid we attempted transformation of D39 using the Havarstein *et al* based methodologies which

have been successfully utilised to transform D39 with pMV158GFP (Vidal et al., 2011). Unfortunately, we were unable to produce successful transformants through this protocol (Transformation Protocol 1, TP1). Due to the lack of success with this method we altered our transformation method, producing a second transformation protocol (TP2) which was immediately successful in transforming pMV158GFP into D39 at efficiencies similar to other strains observed in the literature (Joloba et al., 2010). Transformation was confirmed through both antibiotic susceptibility at 1.25 μg/ml tetracycline (inhibitory to both D39 and 22F) and GFP activity.

Transformation in *S. pneumoniae* is a highly regulated process with only a limited window of competence (20-30 minutes) during which DNA may be taken up. Following competence, cells lose the ability to take up DNA and may not readily regain it in the short term. Competence is induced through quorum sensing of the CSP. Quorum sensing is affected by population density, as the larger a population is, then the more likely it is that the threshold concentration of the quorum sensing molecule is reached. To reduce the likelihood that the competence threshold was being prematurely met, TP2 utilises a lower concentration of cells (determined through optical density) than TP1. Furthermore, a pH of 6.8 was used in TP2 during the initial stages of the protocol prior to transformation. A pH of 6.8 suppresses competence during the growth of competent cells. This suppression ensures that competence occurs once the pH is increased at the point of transformant DNA addition (Havarstein et al., 1995). Rapid freezing through the use of liquid nitrogen in TP1 is thought to fracture *S. pneumoniae* cell wall facilitating the plasmids entry into the cell during transformation. As D39 was readily transformed through TP2 we did not utilise this rapid freezing further.

Transformation of 22F ST433 with pMV158GFP was not possible using the same condition as D39. This is not unexpected given the notorious difficulty of transforming encapsulated clinical isolates (Evans and Rozen, 2013; Yother et al., 1986). Protocol alterations to produce a successful transformation included alterations to incubation times, temperatures and plasmid concentrations. However, these measures did not result in successful transformation. Increasing CSP concentrations to 1500 ng/ml resulted in transformation success. This is surprising as previously no observed increases in transformation efficiency was seen for CSP concentrations between 32 and 500 ng/ml of

CSP (Håvarstein et al., 1995). Similarly, transformations of a wide variety of *S. pneumoniae* isolates has been achieved through CSP concentrations of 100 ng/ml (Evans and Rozen, 2013). However, as shown in Figure 15, 1500 ng/ml of CSP enabled repeatable transformation to occur. The level of transformation observed is at a low level, with $1.5x10^{-7}$ cells transforming per μg DNA, an order of magnitude less than previously used transformation efficiency thresholds used to determine the transformability of strains (Evans and Rozen, 2013).

Once we developed the protocol to transform pMV158GFP into 22F ST433 we moved to produce a *rpoE* knock-out (*rpoE* KO). This knock-out was designed to replace the *rpoE* gene with a kanamycin cassette through homologous recombination via 500 bp flanking regions matching those of 22F's *rpoE* either side of the cassette. Further optimisation of TP2 was required to produce a successful *rpoE* KO. The key factors were increased CSP concentrations to 2500 ng/ml and a significantly longer incubation time from the usual 24-48 hours to 144-168 hours to produce successful transformants. A potential reason for the increased incubation time required may be a combination of the reduction in protein production while under kanamycin selection, combined with the loss of RpoE which plays a role in transcription regulation. Disruption of transcriptional regulation within a *rpoE* KO may have resulted in poor expression of the kanamycin resistance cassette resulting in very slower growth characteristics, in addition to the wider global impact on gene regulation that removal of *rpoE* is likely to have caused. This is supported by observations of related bacteria. *S. mutans rpoE* knock-outs display reduced abundance of many proteins particularly those associated with carbohydrate metabolism and energy production (Xue et al., 2012). This pattern of dysregulation is also observed in *S. aureus rpoE* mutants, including the downregulation of genes associated with virulence transporters, tRNA and other regulatory genes (Weiss et al., 2014).

Five successful transformants were selected for PCR to confirm that *rpoE* was removed and that the Kanamycin cassette was present and in the correct locus. When plated onto CBA plates, all putative SCVs displayed colony diameters of less than 0.5 mm (0.163 – 0.487 mm) significantly less than that of 22F WT. To demonstrate that any colony morphology observed in the knock-outs was due to the removal of *rpoE* and are not the result of any other factor, we aimed to complement *rpoE* KO with pCC1rpoE. However,

we have been unable to produce a successful complementation of *rpoE* KO with pCC1rpoE. In a theme that is explored further throughout this thesis, it is likely that our inability to produce a knock-in is due to the loss of adaptability within *rpoE* KO, a trait shared in *S. mutans rpoE* knock-outs (Xue et al., 2012; Xue et al., 2010). As discussed in Chapter 1, competence is a tightly regulated process with time scales of as small as 5 minutes for the uptake of DNA and changes in the genes to be expressed (Peterson et al., 2004). Delays in expression due to diminished RNAP activity through complete loss of *rpoE* may render the competence system ineffective. This is further supported by evidence that competence is gained during the cell cycle during a period of growth arrest. Therefore, the *rpoE* KO with its slower growth rate on agar may require very different timings or cell concentrations in order to be transformed (Berge et al., 2017).

The cell cycle arrest caused by CSP may present an explanation for the long incubation time of 6 days required for successful transformants. The extremely high concentration of CSP required for transformation of 22F ST433 may also be inhibiting its growth and causing cell cycle arrest. It is possible that CSP is broken down over the course of several days either through environmental or *S. pneumoniae* mediated mechanisms. Once the CSP level have sufficiently declined then the cell cycle is able to resume and colonies are able to grow (Berge et al., 2017).

Another potential reason for the lack of success in complementation of *rpoE* KO with pCC1rpoE (11236 bp) may be that it is 62.8 % larger than pMV158gfp (6900 bp). Larger plasmids have been observed to have lower transformation efficiencies in *B. subtilis* and *E. coli* (Hanahan, 1983; Ohse et al., 1995).

## 3.5 Limitations and Future Work

The knock-out transformation construct utilised 500 bp flanking regions to ensure that the resistance cassette inserts into the correct location within the genome to produce the knock-out. Larger flanking regions are associated with increased transformation efficiencies (Lau et al., 2002). Increasing each flanking region to 2000 bp may have increased transformation efficiencies for *rpoE* KO and will be a factor to consider in future gene knock-out work. Conversely, the complementation assay may benefit from the use of a smaller plasmid to increase transformation efficiency. A potential candidate would be

pMV158gfp as addition of *rpoE* with flanking regions would only cause a small increase in size, it already contains selection markers and we have shown that it may be incorporated into 22F WT. However, transformation experiments with *rpoE* KO and pMV158gfp were unsuccessful suggesting that plasmid size alone is not the factor inhibiting effective plasmid transformation.

The choice of the tetracycline resistance cassette as the selection factor may also benefit from redesign in future experiments. Tetracycline inhibits protein synthesis. While the tetracycline resistance cassette reduces the impact of kanamycin on protein synthesis to non-lethal levels, transcription dysregulation in the *rpoE* KO may cause a cumulative affect resulting in a failure to grow on antibiotic plates despite the presence of antibiotic resistance genes within genetic constructs. A suitable alternative may be ampicillin. Ampicillin functions through inhibition of cell wall synthesis, and so may be less likely to have a cumulative effect with the knockout of *rpoE*. This may be supported by the long incubation time observed for transformants with the *rpoE* KO.

## 3.6 Chapter 3 Key Points

- We have developed the methods required to transform the clinical isolate 22F with both plasmid DNA and linear DNA constructs.

- Using these developed methodologies, we have produced a 22F ST433 *rpoE* KO. The *rpoE* KO contains a kanamycin resistance cassette replacing the *rpoE* gene.

- The *rpoE* knock-outs produced all exhibit a small colony variant morphology.

- We have been unable to produce a *rpoE* knock-in, this may be due to a loss of transcription regulation resulting in disruption of the finely tuned competence system.

# Chapter 4: Phenotyping of 22F WT, SCV, and *rpoE* KO Strains

## 4.1 Introduction

*Streptococcus pneumoniae* colony variants have been observed by numerous groups to display altered phenotypic characteristics. Many of the phenotypic changes have the potential to be clinically relevant such as increased virulence by *spxB* large colony variant mutants (Syk et al., 2014), increases in biofilm formation in *rpoE* SCVs (Churton et al., 2016) and alterations to capsule genes (Domenech et al., 2009). The small colony morphology which gives SCVs their name is the most readily observed phenotypic trait of these mutants, however it is the increased biofilm formation by SCVs which causes them to be of particular interest (Churton et al., 2016) due to their associations with chronic infections.

In Chapter 4 we investigated the phenotypic traits of 22F, SCVs and the *rpoE* KO produced in Chapter 3. These investigations were to determine whether *rpoE* KO is phenotypically similar to the SCVs not only in size, but also in a range of clinically relevant phenotypic characteristics including biofilm formation. Greater understanding of the phenotypic changes observed in SCVs and *rpoE* KO may provide further insight into the role of *rpoE* and the factors which exert the fitness and selection pressures required for the parallel evolution observed (Churton et al., 2016). RpoE has been demonstrated to be an important regulatory protein in *B. subtilis* (Prajapati et al., 2016a; Prajapati et al., 2016b), *S. mutans* (Xue et al., 2012; Xue et al., 2011; Xue et al., 2010) and *S. aureus* (Weiss et al., 2014) and therefore we would expect to observe phenotypic changes across a wide range of factors.

We observed that SCVs display significantly altered phenotype to that of the 22F wild type parent strain but also that the *rpoE* KO displays a range of similarities and differences to the 22F WT. These findings support the role of RpoE as a global regulator in *S. pneumoniae* and suggest that the mutations in SCVs cause changes in function distinct to the complete removal and therefore loss of function of RpoE in *rpoE* KOs.

## 4.2 Materials and Methods

The materials and methods required to carry out the work in this chapter are detailed in Chapter 2.

## 4.3 Results

### 4.3.1 Colony Size and Morphologies

22F *rpoE* KO 2 was chosen as the representative example of the *rpoE* knock-outs to be used in all further experiments and will be referred to as *rpoE* KO from this point onwards. To determine whether the difference in phenotype between the SCVs and *rpoE* KO was a transient effect, colony diameters were measured over three passages. The small colony morphology was observed in both SCVs and *rpoE* KO (Figure 26) and maintained over the three passages. The SCVs and *rpoE* KO all displayed significantly smaller colony diameters across all passages however *rpoE* KO was also significantly smaller than all SCVs (Figure 27). These results suggest that complete removal of *rpoE* may be causing a reduced growth rate on agar compared to the mutations to *rpoE* observed in the SCVs.

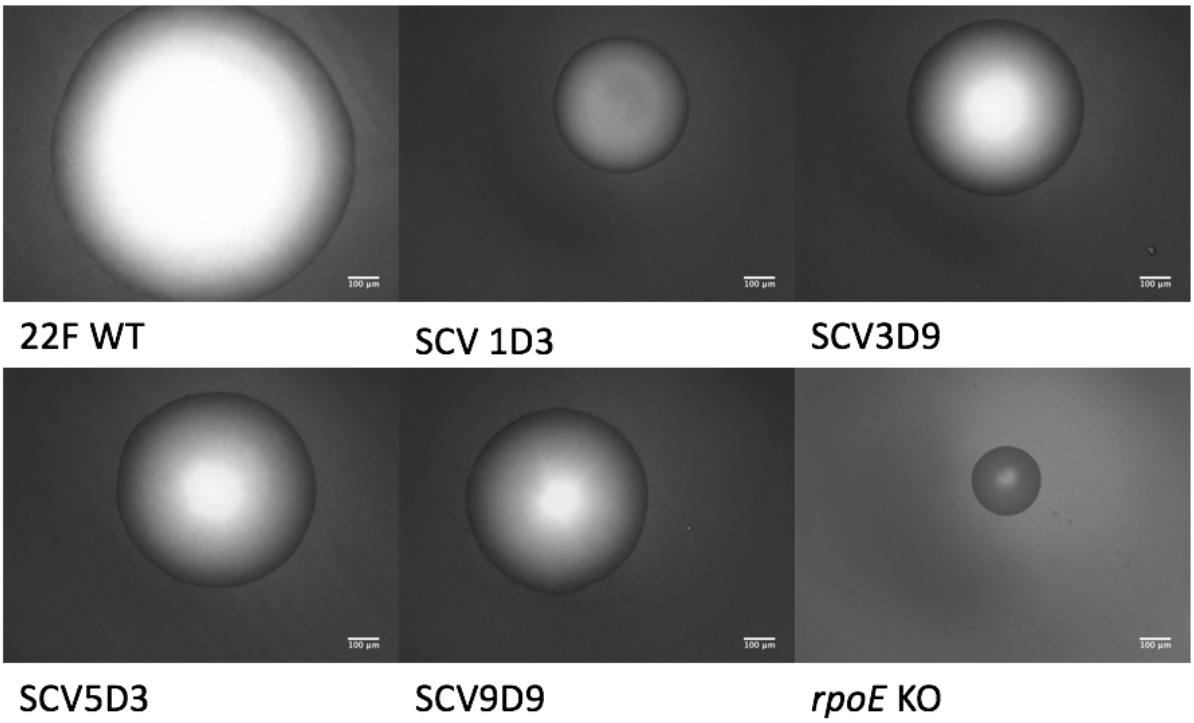



Figure 26: Representative images of the colony morphologies of 22F WT, SCVs and *rpoE* KO. Colonies were grown on CBA agar for 24 hours and imaged using a Olympus BX51 epi-fluorescence microscope microscope. The white scale bar represents 100 µm.

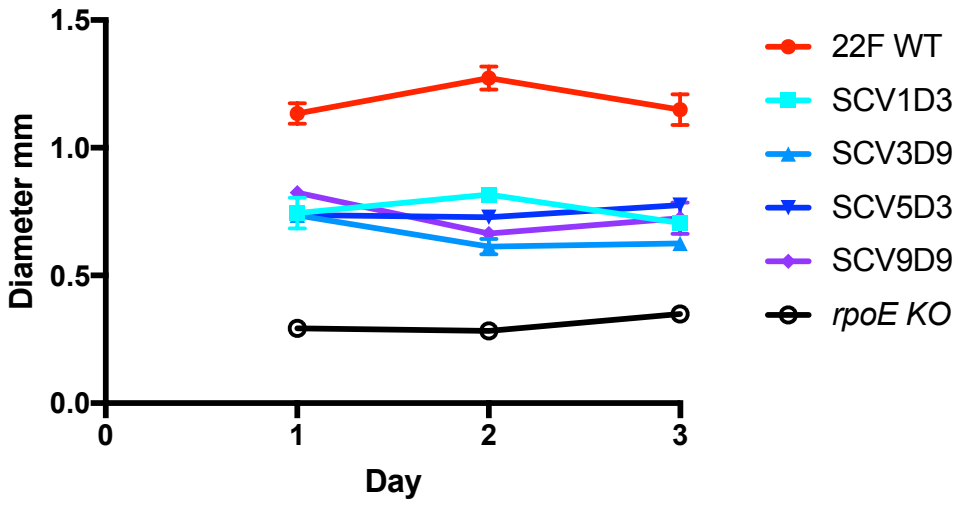



Figure 27: Colony diameters of 22F WT, SCVs and *rpoE* KO. Colonies were grown on CBA agar for 24 hours and imaged using an Olympus BX51 epi-fluorescence microscope microscope. A 1-way ANOVA of the 3rd passage revealed that all SCV colony diameters were significantly ($p < 0.0001$) smaller than 22F WT and significantly larger than *rpoE* KO.No significant differences were observed between SCVs. N= 5 colonies per symbol.

### 4.3.2 Planktonic Growth of Strains

To investigate whether the differences in growth characteristics might also be observed within planktonic growth state we investigated the growth of isolates over in liquid culture over 18 hours (Figure 28). In contrast to the phenotypic changes observed for colony diameter during planktonic growth SCVs grew to a similar or greater level than 22F WT. Following pre-treatment in BHI broth all SCVs with the exception of SCV5D3 displayed significantly greater growth than 22F WT. Pre-treatment through growth on CBA produced a slight although non-significant increase in growth for the SCVs relative to 22F WT. However, the differences in growth between CBA and BHI pre-treatments for 22F WT and each SCV were not significant.

The largest changes observed were for *rpoE* KO for which pre-treatment conditions caused significantly variations in growth. Pre-treatment of *rpoE* KO on CBA prior to inoculation into BHI broth resulted in significantly less growth than BHI broth pre-treatment. This may suggest that *rpoE* KO is less able to adapt to changes in growth media when compared to 22F WT and the SCVs which displayed broadly similar growth traits. The *rpoE* KO results also contrast to those observed over the 3-day passage on CBA where we do not observe adaptation to media to approximate the SCV phenotype as we do in BHI broth.

Growth curves carried out in BHI over a 10 hour time course showed a slightly altered trend form the 18 hour culture, although generally matching those found by Churton *et al* (Churton, 2014). 22F ST433 grew slightly faster than the SCVs and *rpoE* KO, with *rpoE* KO growing the slowest of all strains. The differences between the 10 hour time course and adaptation experiments with an 18 hour incubation indicate that the 22F WT may possess a shorter lag phase than SCVs, while the SCVs are able to grow to a higher concentration.

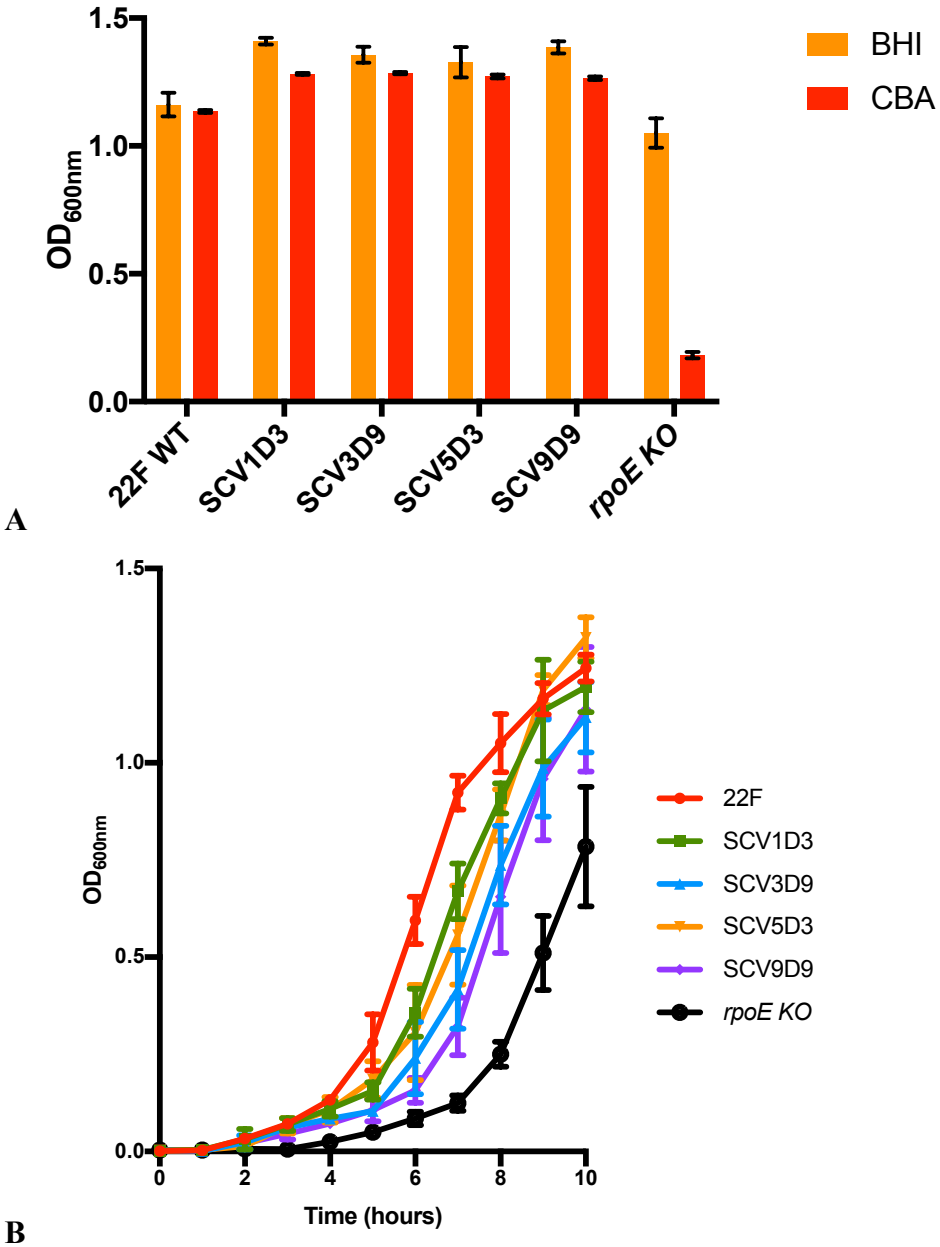



Figure 28: **A**) Assessment of growth by 22F WT, SCVs and *rpoE* KO in BHI following pre-treated in either BHI broth or on CBA plates. Strains were grown for 24 hours growth either on CBA plates or in BHI broth for 24 hours and then inoculated into BHI broth, and grown for a further 18 hours. Significance was determined through a 2-way ANOVA with Sidak's multiple comparison test. A 2-way ANOVA was chosen because it allows comparison of two independent variables, with Sidak's multiple comparison test used as the method to determine differences in growth due to media pre-treatment. The 2-way ANOVA reveals that both the strain (76.44% of total variation) and the media (12.11% of total variation), and the interactions (25.13% of total variation) all provide statistically

significant differences with ($p < 0.0001$). The *rpoE* KO grew to a higher concentration in BHI ($p < 0.0001$) when pre-treated with BHI growth conditions. For BHI pre-treatment conditions all SCVs and 22F WT produced higher optical densities than *rpoE* KO (all $p < 0.0001$). N = a minimum of 3 indpendent biological replicates. **B**) Growth curves of 22F WT, SCVs and *rpoE* KO in BHI following pre-treated in BHI broth. BHI with 1 x $10^6$ CFU / mL was incubated at 37°C for 10 hours, with hourly $OD_{600nm}$ readings. Error bars represent standard error. N = 3 independent biological replicates.

### 4.3.3 Biofilm Formation

It has been previously observed that SCVs display greater biofilm formation than 22F WT (Churton et al., 2016). To validate these results, we grew biofilms to 24 (attachment) and 72 hours (mature biofilm) and assessed the volume of biomass produced through the use of crystal violet staining assay. In the 24-hour attachment assay we observed a range of biomass although only SCV9D9 and *rpoE* KO displayed significantly increased biomass relative to 22F WT.

Interestingly for 3-day biofilm formation although all SCVs and *rpoE* KO displayed significantly greater biofilm formation than 22F WT we observe a range of biofilm formation across the SCVs with significant differences between different SCVs. Unlike the planktonic and agar assays for mature biofilm formation *rpoE* KO fell within the range of observed for SCVs. Furthermore, significant differences were also observed between the SCVs resulting in a spectrum of biofilm forming ability for 3-day biofilms.

Confirmation of the increased mature biofilm formation was quantified through CFU counts of biofilms. CFU quantification ensures that the crystal violet observations were not an artefact of the crystal violet assay and that the trends observed are due to increased live cell counts. We observed a similar trend of increased biofilm formation although only SCV5D3 and SCV9D9 displayed significantly increased CFU counts relative to 22F WT. This is likely to be explained through the lower number of repeats for the CFU counts.

A potential limitation of this assay is that the absorbance values are low. This may indicate that while statistically significant, the changes in biofilm formation may not be biologically

significant. However, the confirmation of the general trend of increased biofilm formation by SCVs with up to 2 $\log_{10}$ differences by CFU counts suggests that the trend is real.

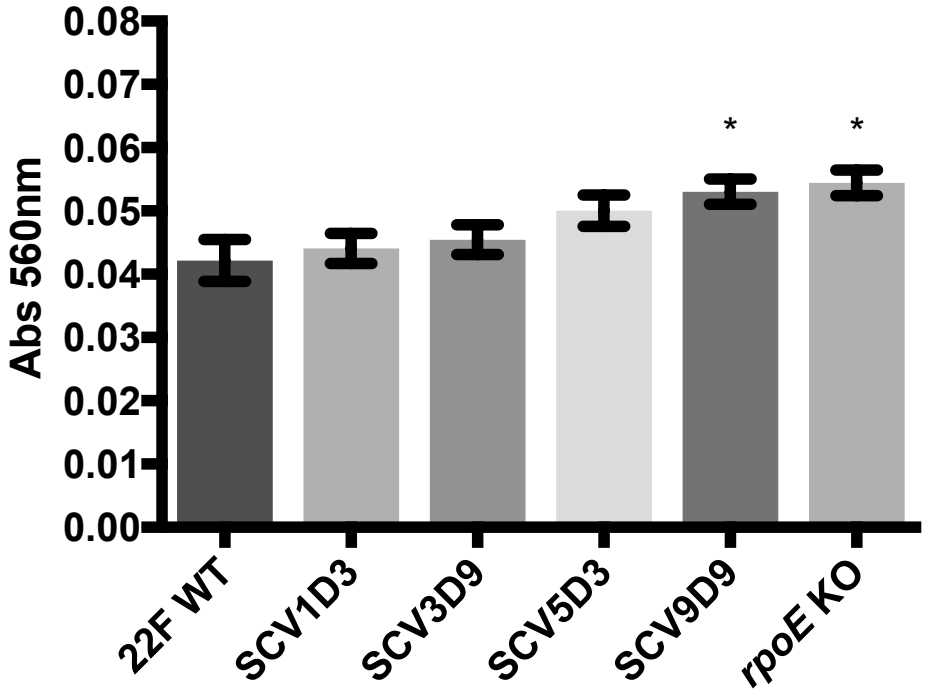



Figure 29: Crystal violet staining of 24 hour surface attachment by 22F WT, SCVs and *rpoE* KO. Strains were cultured in 20% (w/v) BHI in CoStar 96 well plate and allowed to attach for 24 hours. Crystal violet staining of attached cells was assessed through absorbance reading at 560nm on a FLOUstar Optima plate reader. A 1-way ANOVA reveals that SCV9D9 and *rpoE* KO both differ significantly from WT 22F ($p < 0.05$). The biofilm formation of each isolate was quantified through a minimum of two biologicaly independent experiments each consisting of 8 technical replicates.

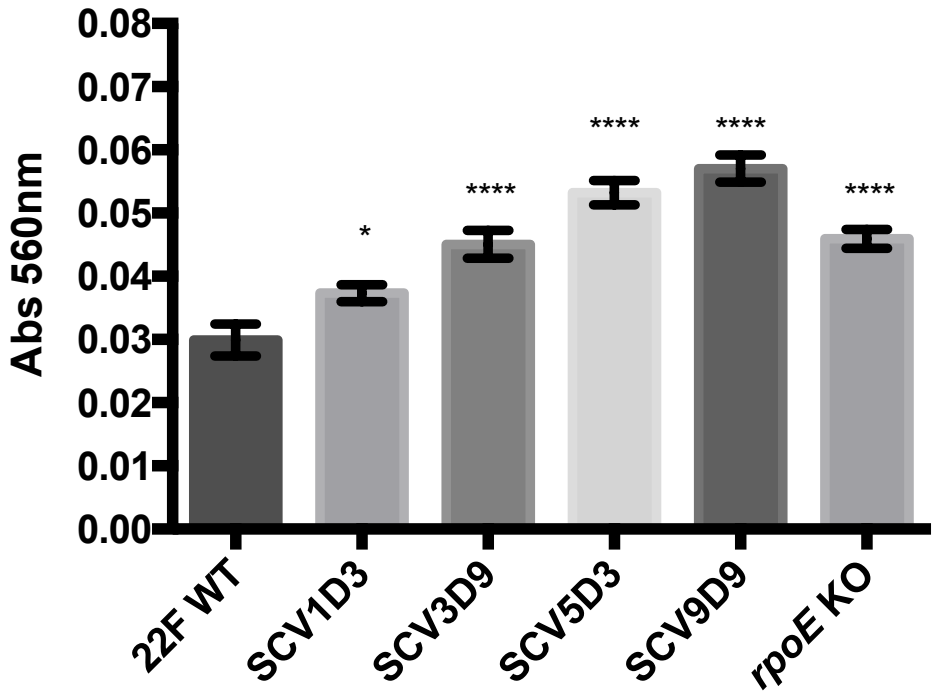



Figure 30: Crystal violet staining of 72 hour biofilm formation by 22F WT, SCVs and *rpoE* KO. Strains were cultured in 20% (w/v)BHI in CoStar 96 well plate and cultured for 72 hours with media changes every 24 hours. Crystal violet staining of the biofilm was assessed through absorbance reading at 560nm on a FLOUstar Optima plate reader. A 1-way ANOVA with Tukeys Multiple comparison test reveals a range of biofilm formation differences . All SCVs and *rpoE* KO formed significantly more biofilm than 22F as indicated by the asterixes on the graph. SCV5D3, SCV9D9 and *rpoE* all produce significantly more biofilm than SCV1D3 ($p<.0001$, $p<0.0001$ and $p<0.05$ respectively). SCV5D3 and SCV9D9 produced significantly more biofilm than SCV3D9 ($p < 0.05$ and $p < 0.001$ respectively). SCV9D9 displayed significantly greater biofilm formation than *rpoE* KO ($p < 0.01$). The biofilm formation of each isolate was quantified through formed of 3 biologically independent replicates.

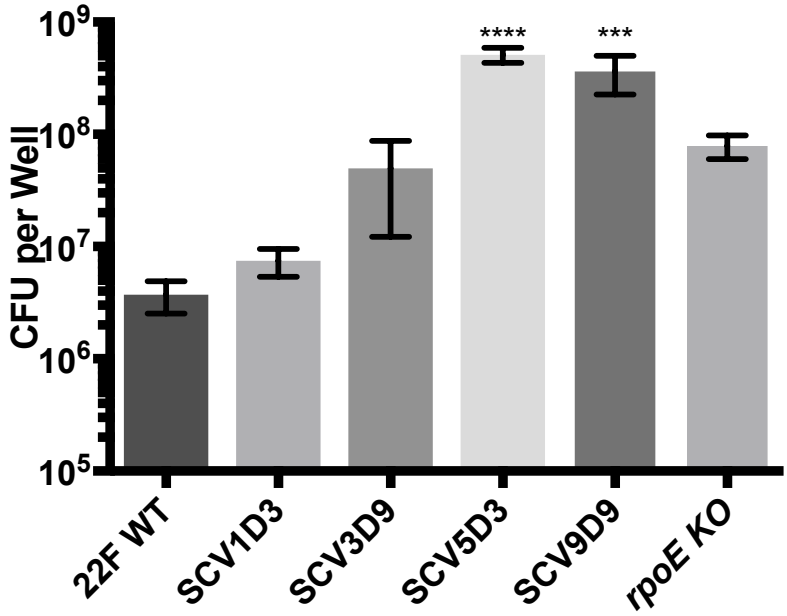



Figure 31: Colony forming unit counts from 72 hour biofilm formation by 22F WT, SCVs and *rpoE* KO. Strains were cultured in 20% (w/v)BHI in CoStar 6 well plates and cultured for 72 hours with media changes every 24 hours. A one way ANOVA revealed that SCV5D3 and SCV9D9 were both displayed significantly greater CFUs than 22F WT. N = 3 independent biologically independent replicates.

### 4.3.4 Scanning Electron Microscopy

To further investigate the mature biofilms, we utilised scanning electron microscopy (SEM) to view individual cells and biofilm structures. SEM has previously been used to investigate *S. mutans* and its *rpoE* knock-out with differences were observed between with the *rpoE* knock-out and wild-type. The *S. mutans rpoE* knock-out displayed dendrite like structures attaching to the substrate surface (Xue et al., 2011).

Representative images were taken of 22F WT, SCV5D3, SCV9D9 and *rpoE* KO. The SCVs, SCV5D3 and SCV9D9 were chosen because they displayed the greatest increases in biofilm formation and were therefore expected to be most likely isolates to display phenotypically differentiated characteristics. Biofilms of 22F WT and the SCVs appeared broadly similar. However, no dendrite like structures were observed within any of the *S. pneumoniae* biofilms.

The most striking difference observed is the diminished capacity of *rpoE* KO to form chains of cells. Chain length in *S. pneumoniae* has been linked to increased adherence and colonisation (Rodriguez et al., 2012) however, it may also increases susceptibility to human neutrophil mediated killing (Dalia and Weiser, 2011). The loss of this characteristic may indicate dysregulation within *rpoE* KO.

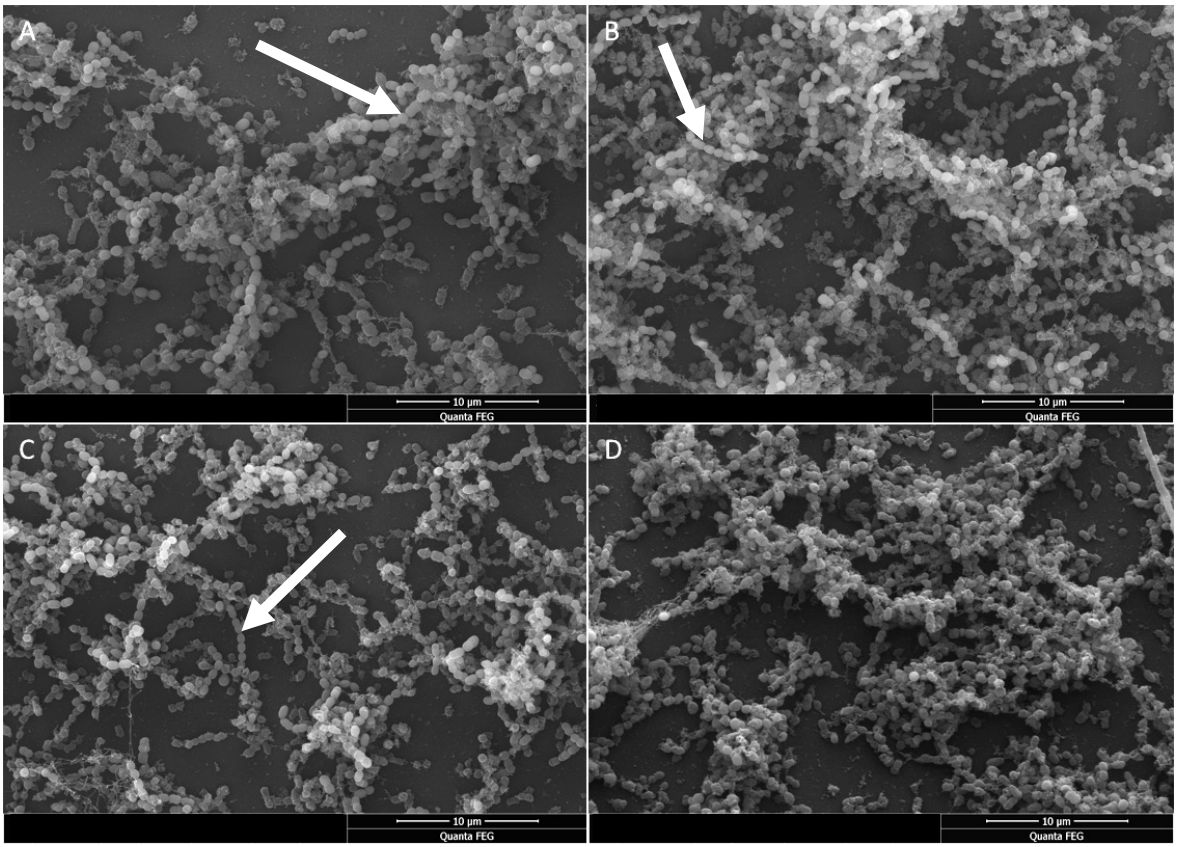



Figure 32: Scanning electron microscopy of *S. pneumoniae* 3 day biofilms **A)** 22F WT **B)** SCV5D3 **C)** SCV9D9 **D)** *rpoE* KO. *rpoE* KO appears to display reduced chain formation compared to 22F WT and SCVs. Chains indicated by white arows. Images are representative of 3 independent biological replicates.

### 4.3.5 Self-Aggregation

Due to the differences in chain formation displayed by *rpoE* KO we investigated whether the phenotypic differences observed may be due to altered self-aggregation properties. Self-aggregation is a measure of how effectively cells of a strain are able to bind to each other and may be an important factor in biofilm formation.

We investigated whether the differences in biofilm formation and chain formation observed may be due to increased self-aggregation by the SCVs and 22F WT relative to *rpoE* KO. However, we did not observe a biofilm-like phenotype continuum for self-aggregation with all strains clustering together with the exception of SCV5D3 which exhibited significant, although minor, increases in self aggregation relative to *rpoE* KO and SCV3D9.

22F WT did not display significantly altered self-aggregation properties relative to any isolate and it is therefore unlikely that this property is responsible for the differences in biofilm formation observed. This is further highlighted by SCV9D9 which displayed the greatest biofilm formation in the crystal violet assay but exhibited one of the lowest levels of self-aggregation alongside *rpoE* KO with both less than 22F WT. These findings contrast with those carried out with *S. mutans*, the *rpoE* KO displayed increased aggregation relative to the wild-type (Xue et al., 2011).

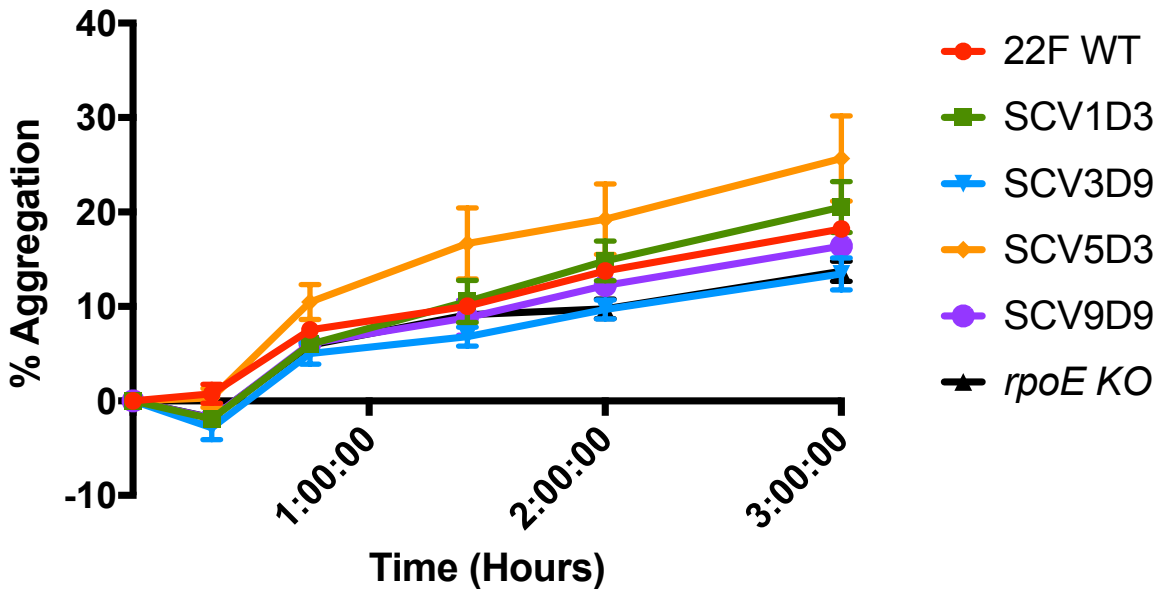



Figure 33: Self aggregation of 22F WT, SCVs and *rpoE* KO stationary phase planktonic cultures. Cultures were suspended in PBS and self aggregation measured over 3 hours. Self aggregation was calculated as $\left(\frac{\text{OD600 nm at time 0} - \text{OD600nm at time point X}}{\text{OD600 nm at time 0}}\right) \times 100$ as per (Xue et al., 2011). At 3 hours a 1-way ANOVA determines that significant differences observed are between SCV5D3 and SCV3D9 ($p < 0.05$), and SCV5D3 to *rpoE* KO ($p < 0.05$). N = 3 independent biological replicates.

### 4.3.6 Binding to Human ECM Components

Our analysis of the adherent properties of our strains was further expanded through analysis of the binding of the *S. pneumoniae* isolates to a range of human extracellular matrix (ECM) components. Increased binding to many ECM components was observed in an *S. mutans rpoE* knock-out (Xue et al., 2011). The capability of strains to bind to human extra cellular factors has important implications for colonisation and biofilm formation within the human host.

We tested the ECM binding capabilities of 22F, SCV9D9 and *rpoE* KO using the same array used with *S. mutans rpoE* knock-outs (Xue et al., 2011). We did not observe binding to any ECM component by any of our isolates. *S. pneumoniae* strains have been previously assessed for their binding capabilities for the majority of these factors including a screen of 43 strains against radiolabelled ECM components (Kostrzynska and Wadstrom, 1992). While there was variation in binding by the tested strains to several ECM components, all strains tested bound to laminin (Kostrzynska and Wadstrom, 1992). That we do not observe binding to laminin here suggests that 22F ST433 may have very different binding characteristics or that this assay is unsuitable for use with *S. pneumoniae* 22F ST433.

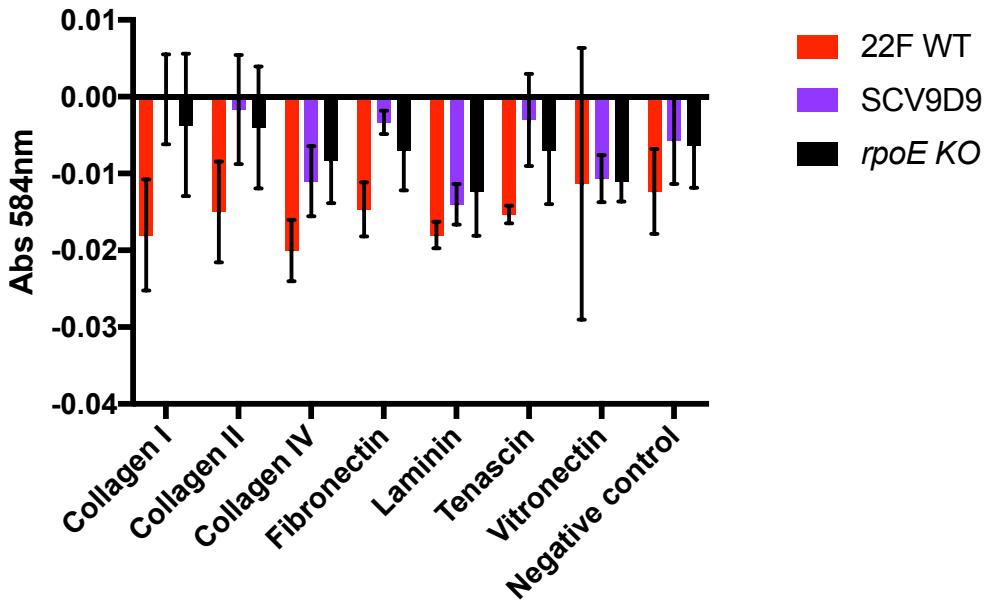



Figure 34: Binding of 22F ST433, SCV9D9 and *rpoE* KO to human extracellular matrix components over 24 hours. The absorbance results suggest that the isolates tested are not binding to the human ECM components. N = 3 independent biological replicates.

### 4.3.7 Hydrogen Peroxide Production

As detailed in Chapter 1, hydrogen peroxide is a key virulence factor of *S. pneumoniae* involved in many important processes including control of competing bacteria, host cell killing and *spxB* (the gene responsible for $H_2O_2$ production) is required for aggregate and biofilm formation (Blanchette-Cain et al., 2013). Potential associations between $H_2O_2$ and *rpoE* might be found in proteomic worked carried out previously by the Webb group (Churton, 2014). Formate acetyltransferase (pfl) and fructose-bisphosphate aldolase were both found to be down-regulated in the SCVs tested relative to the WT. Both of these proteins are associated with pyruvate metabolism which is the pathway through which $H_2O_2$ is produced (Jensen et al., 2009; Ramos-Montanez et al., 2008).

To determine whether $H_2O_2$ might be altered by *rpoE* mutations or knock-out, biofilm supernatants were analysed for $H_2O_2$ concentrations. At 24 hours several of the SCVs and *rpoE* KO displayed increased $H_2O_2$ concentrations relative to both 22F WT and SCV1D3. For the 3-day biofilms no differences were observed in $H_2O_2$ concentrations. A limitation of this assay is that it is lacking the corresponding CFU counts for 24 hour biofilms. It may be the case that each cell produces similar concentrations of $H_2O_2$ but that there is a higher concentration of cells of the *rpoE* KO SCVs 3D9, 5D3 and 9D9.

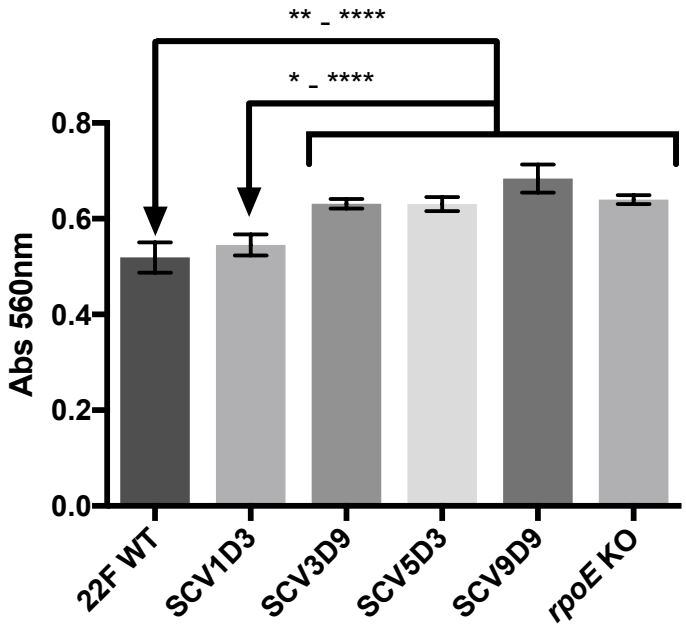



Figure 35: Hydrogen peroxide concentrations following 24 hours of biofilm growth in 20% (v/v) BHI. A 1-Way ANOVA with Tukey's multiple comparison test found 22F WT displayed signficantly reduced concentrations of $H_2O_2$ relative to SCV3D9, SCV5D3, SCV9D9, and *rpoE* KO ($p > 0.01$, $p > 0.01$, $p > 0.0001$ and $p > 0.01$ respectively), representing approximately 22-33% increases over 22F WT. Furthermore, SCV1D3 was also observed to display reduced concentrations of $H_2O_2$ relative to SCV3D9, SCV5D3, SCV9D9, and *rpoE* KO ($p > 0.05$, $p > 0.05$, $p > 0.0001$ and $p > 0.05$ respectively). N = a minimum of 16 technical replicates from 2 sets of biological replicates.

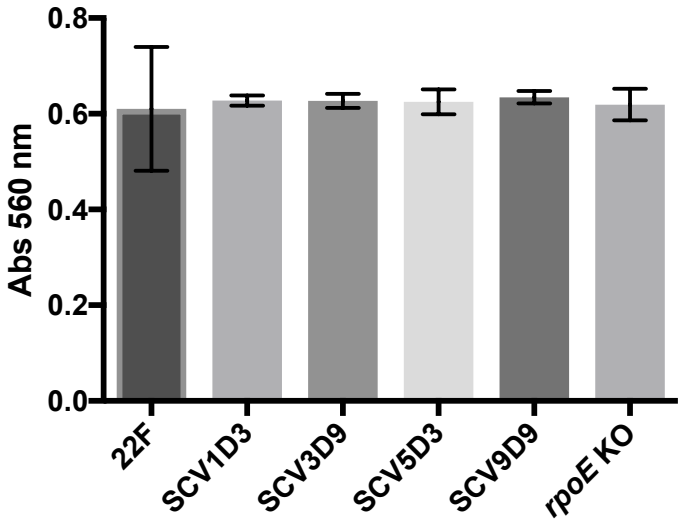



Figure 36 Hydrogen peroxide concentrations within *S. pneumoniae* 72-hour biofilms. A 1-Way ANOVA with Tukey's multiple comparison test found no significant difference between any strain ( $P > 0.05$). N = a minimum of 23 technical replicates from 3 independent biological replicates.

### 4.3.8 Carbohydrate Metabolism

To further investigate potential factors which may explain the differences in growth between planktonic and biofilms lifestyles we investigated the metabolism of a wide range of carbohydrates. *S. pneumoniae* possesses a diverse and flexible metabolism capable of exploiting a wide range of substrates and demonstrates the widest range of substrate utilization among the organisms that colonise the nasopharynx. *S. pneumoniae's* broad metabolism is facilitated through the considerable portion of its genome that is committed to carbon and nitrogen metabolism (Tettelin et al., 2001). This metabolic flexibility is further demonstrated by *S. pneumoniae's* membrane transporters of which 30% are involved in carbohydrate uptake (Paulsen et al., 2000).

Given the importance of a flexible metabolism to *S. pneumoniae*, Biolog PM1 plates were used to investigate the metabolism of a wide selection of carbohydrate sources. SCVs exhibited difference in metabolism of numerous carbohydrate sources. Across many of the sugars, SCVs (particularly SCV5D3 and SCV9D9) exhibited a trend of significantly increased sugar metabolism. SCV1D3 displays the least variation from 22F WT and for

three sugars was significantly ($P < 0.05 – 0.001$) different to at least one other SCV. This suggests that there is a spectrum of phenotypic change within SCVs for the metabolism of several sugars.

However, increases in metabolism were not observed for all sugars. SCVs and *rpoE* KO displayed significantly ($P < 0.05$-$0.01$) reduced metabolism of L-galactonic acid-γ-lactone relative to 22F WT. Several sugars including the potentially important maltotriose and sucrose displayed no significantly altered metabolic profiles, although the trends observed for these sugars mirrored those observed for the 3-day biofilm formation assays. Finally, several sugars displayed no change relative to the control such as D-glucose-1-phosphate and D-glucose-6-phosphate, despite significant ($P < 0.05$-$0.01$) differences observed in α-D-glucose.

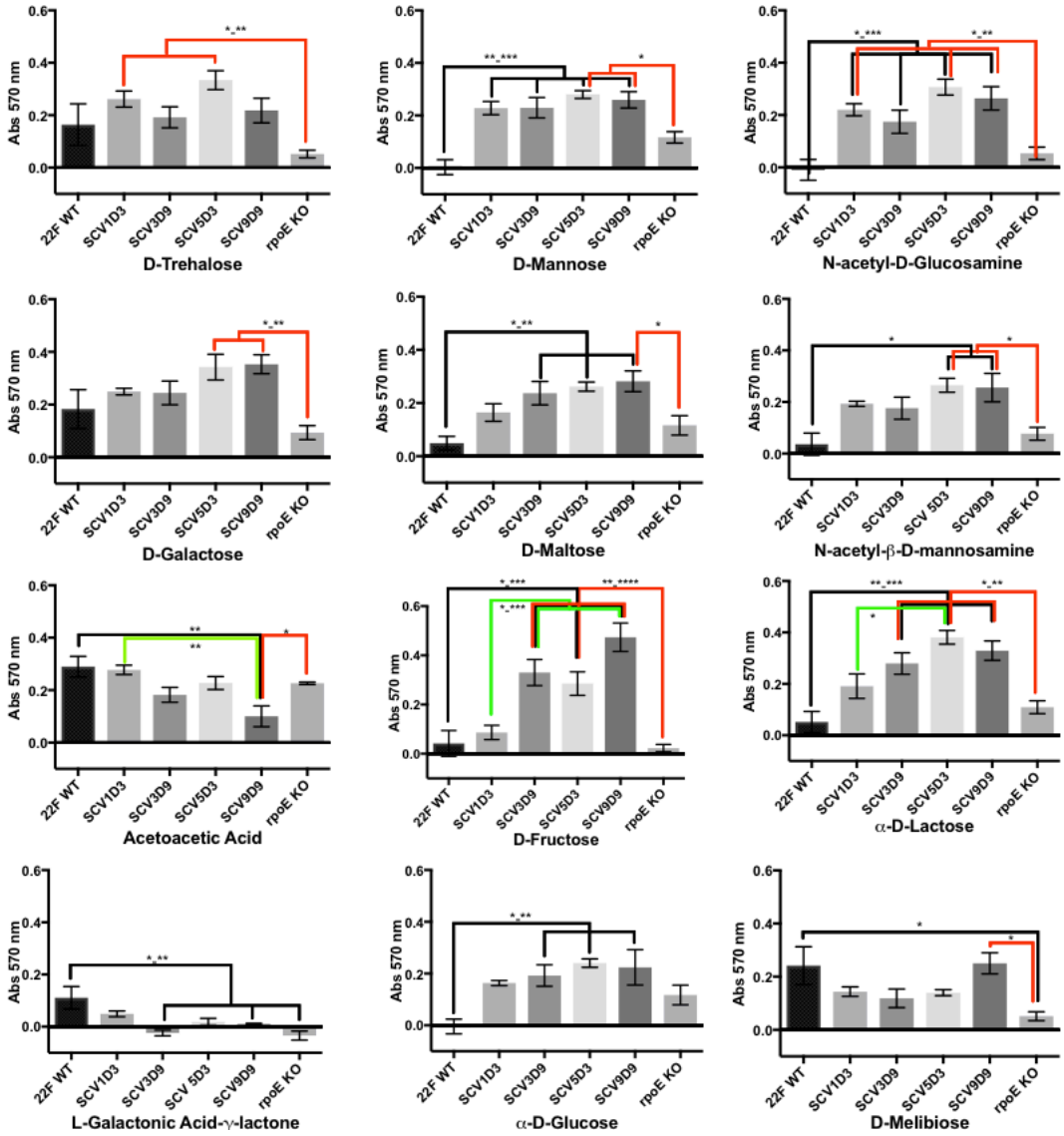



Figure 37: The change in growth following the supplementation of different carbohydrate sources through the use of Biolog PM1 plates. Significant differences are shown between different isolates. A black line indicates significant differences to 22F WT, a red line to *rpoE* KO and a green to SCV1D3. N = a minimum of 3 independent biological replicates.

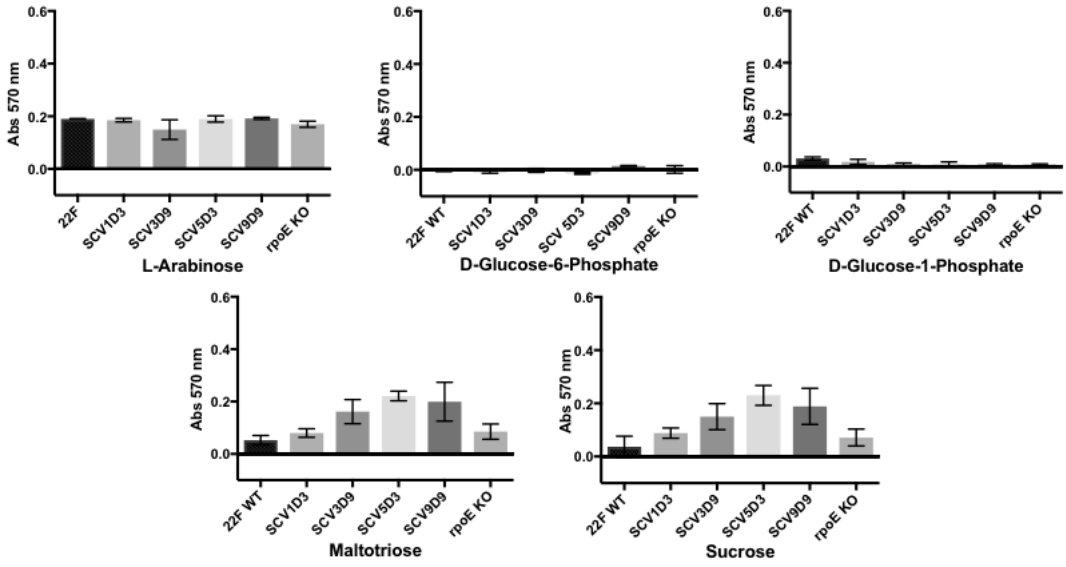



Figure 38 The change in growth due to the supplementation of carbohydrate sources in Biolog PM1 plates. For these sugars no significant differences were observed. However a continuum of phenotypes similar to that observed for 3-day biofilm formation was observed for the sugars maltotriose and sucrose. N = a minimum of 3 independent biological replicates.

## 4.4 Discussion

In Chapter 4 we carried out phenotyping of 22F WT, the *rpoE* KO generated in Chapter 3 and 4 SCVs with a range of *rpoE* mutations. Quantification of phenotypic variations between strains provides valuable indications for clinically relevant traits of these strains and potential modes of action.

In line with observations in Chapter 3, we observed significantly reduced colony diameters by the SCVs (Churton et al., 2016). Additionally, the *rpoE* KO displayed significantly smaller colony diameters than both the SCVs and the 22F WT. This indicates that a complete removal of the *rpoE* is similar to but not identical to the mutations to *rpoE* observed in SCVs.

We did not observe similar trends of phenotypic variation for planktonic growth. Cultures were grown for a 24-hour pre-treatment period either on CBA plates or within BHI broth cultures were then grown for 18 hours in BHI broth. When pre-treatment was carried out in BHI, SCV1D3, SCV3D9 and SCV9D9 all grew to a greater (although a relatively small increase) optical density than 22F WT with no significant difference observed between 22F WT and SCV5D3. However, when pre-treatment was carried out on CBA plates no significant changes in in growth were observed between the SCVs and 22F WT. The role of *rpoE* as a global regulator may provide an explanation for differences in trait variation between different environmental conditions (Weiss et al., 2014; Xue et al., 2012). Further insight into differential regulation is provided by cultures of the *rpoE* KO. When pre-treated though growth in BHI *rpoE* KO did not display significantly altered growth to 22F WT, although reduced growth compared to the SCVs was observed. The most striking differences occurred when *rpoE* KO was pre-treated on CBA plates. Following CBA pre-treatment, *rpoE* KO grew significantly less over the 18-hour time course than *rpoE* KO pre-treated in BHI broth. This suggests that *rpoE* KO may possess a reduced capability to adapt to changes in growth media compared to 22F WT and the SCVs. A diminished capacity to adapt to changing conditions has been observed in *S. mutans* (Xue et al., 2010). The growth curve of strains displayed a longer lag phase in SCVs and particularly *rpoE* KO compared to 22F WT. The shorter lag phase of 22F WT may allow it to out compete SCVs and explain why SCVs only form 1% of the biofilm populations which they are generated in (Churton et al., 2016).

The *rpoE* KO displays a further lack of adaptation on CBA plates (Figure 27). Over the course of the three passages we did not observe adaptation to CBA media by *rpoE* KO with colony diameters only slightly increasing on the third passage, and still smaller than those of SCVs and 22F WT. The different levels of growth may be linked to the environmental conditions involved. The planktonic growth in BHI and the solid medium of CBA plates may be mediated through different regulatory mechanisms and so be differentially impacted by mutations to *rpoE*.

Beyond the small colony size observed on CBA plates, one of the principal traits of *S. pneumoniae* 22F SCVs is that they have all been observed to display increased biofilm formation (Churton et al., 2016). This is perhaps unsurprising, as the SCVs were originally generated during biofilm growth, indicating that biofilm related fitness pressures select for mutations to *rpoE*. Our observations matched these findings, with significant increases in biofilm formation by both SCVs and also by *rpoE* KO, particularly for 3-day biofilms. For 24-hour biofilms (representing initial attachment), only SCV9D9 and *rpoE* KO displayed significantly greater attachment when compared to 22F WT. For 3-day biofilm formation, we observed increased differences in biofilm formation by SCVs and *rpoE* KO. This is in line with previous 22F SCV research; however, our approach differs in that we measured each SCV individually rather than as a group. While all SCVs displayed significantly greater biofilm formation than 22F WT, we did not observe a uniform increase in biofilm formation, but a spectrum. This contrasts with the agar and planktonic growth conditions, where we did not observe significant inter-SCV variability.

Furthermore, the phenotypic variation displayed by *rpoE* KO biofilms relative to 22F WT and SCVs differed to that observed in planktonic and agar-based assays. For 24-hour biofilms, *rpoE* KO produced the most biomass, significantly more than 22F WT, while in the 3-day biofilm assay *rpoE* KO biofilm formation fell within the middle of the range produced by SCVs and displayed significantly greater biofilm formation than 22F WT. When biofilm morphology was assessed through the use of SEM we observed diminished chain formation by *rpoE* KO suggesting further deviation from 22F WT and SCVs and potentially dysregulation.

*S. pneumoniae* dedicates a significant portion of its genome and membrane transporters towards carbohydrate, and nitrogen metabolism (Paulsen et al., 2000). We have observed significant changes to the metabolism of many carbohydrates by SCVs and *rpoE* KO. The trends of phenotypic variation for the sugar metabolism observed also varied. For several of the sugars, we observed a trend of phenotypic variation similar to that observed for 3-day biofilm formation; a spectrum of increased sugar metabolism was observed amongst SCVs with the most extreme phenotypic changes tending to be displayed by SCV5D3 and SCV9D9, while SCV1D3 displayed the least difference to 22F WT. Furthermore, SCVs displayed significantly greater sugar metabolism than *rpoE* KO for many of sugars. Many *S. pneumoniae* sugar metabolism genes are associated with virulence and these results may indicate increased virulence in SCVs (Buckwalter and King, 2012). For example, increased metabolism of mannose, and N-acetylglucosamine metabolism is displayed by SCVs and is associated with the gene *manLMN* which is itself linked to invasive disease. Similarly, maltose metabolism through *malT* is associated with increased pneumonia (Bidossi et al., 2012; Ogunniyi et al., 2012). Other sugars such as galactose, N-acetylmannosamine and fructose are associated with colonisation, pneumonia and invasive disease through the reading frames SP0090-2 and SP1328 highlighting the need to investigate the virulence of SCVs (Buckwalter and King, 2012; Hava and Camilli, 2002).

*rpoE* KO, with the exception of L-galactonic acid-γ-lactone and D-melibose, did not display differences to 22F WT, and was statistically significant to many of the SCVs. This further emphasises the phenotypic differences between SCVs and *rpoE* KO and is likely to be related to the slow adaptation of *rpoE* KO to changes in growth media.

Numerous sugars did not show any discernible variation in metabolism relative to the negative control, including sugars such as L-arabinose, D-glucose-1-phosphate and D-glucose-6-phosphate. This may represent sugars which are not regulated by RpoE. However, the sugars maltotriose and sucrose still displayed a similar phenotypic trend to 3-day biofilm formation despite the lack of statistical significance, additional replicates would clarify these as would RNAseq for genes involved in the metabolism of these sugars.

Hydrogen peroxide formation at 24-hours followed the trend of phenotypic expression in 3-biofilm however, no differences in hydrogen peroxide concentrations were observed at 3-days. This potentially matches proteomics carried out previous on the SCVs which observed changes in expression of *sodA* which is relevant to production of $H_2O_2$ (Churton, 2014). The gene *sodA* encodes for the superoxide dismutase SodA which under aerobic conditions catalyses the removal of superoxides produced through aerobic metabolism to oxygen, and $H_2O_2$. This gene was up-regulated 1.52 fold in SCVs (Churton, 2014).

We also investigated the self-aggregation of the strains following observations in *S. mutans* of increased self-aggregation by knock-outs (Xue et al., 2011). Our observations carried out in a similar manner found no significant difference in self aggregation between 22F WT and our *rpoE* KO. SCV5D3 displayed increased self-aggregation compared to SCV3D9 and *rpoE* KO; however, this does not fit with the trends that we observe throughout other phenotypic assays and is therefore unlikely to be a key factor.

### 4.4.1 Potential Mechanisms for Phenotypic Change

As a global regulator, RpoE is expected to have a wide range of impacts. The different *S. pneumoniae* lifestyles on agar colonies, as planktonic cultures and biofilms showed different proteomic profiles with the majority of proteins displaying altered expression profiles (Allan et al., 2014). Investigation of a *rpoE* knock-out in *S. mutans* observed changes to its proteome associated with combatting oxidative and acid stress (Xue et al., 2012). It may be the case that 22F *rpoE* KO suffers from increased stresses which in turn promote biofilm formation. This is supported by observations of proteins such as Rgg and RpoE's roles both in the response to oxidative stress from oxygen and paraquat and also in its promotion of biofilm formation (Bortoni et al., 2009).

The reduced capability for adaptation by *rpoE* KO may also play a role in its increased biofilm formation capabilities. If *rpoE* KO is slower to adapt to changing conditions then it may also have a reduced capability to adapt to the changing requirements and pressures associated with the different developmental stages of biofilm maturation. *S. pneumoniae* biofilm formation is a complex process with several developmental stages which involve large protein shifts (Allegrucci et al., 2006). In *rpoE* KO these protein shifts may be

significantly delayed, leading to increased growth, even if certain conditions which would cause a shift towards reduced growth are met.

A possible mechanism for delayed sensing of different stages of biofilm maturation may be through alterations to carbohydrate metabolism and quorum sensing. In *S. mutans*, the quorum sensing molecule AI-2 is produced as a metabolite of LuxS activities. LuxS expression is associated with many genes including those that regulate biofilm formation. The largest upregulation due to exogenous application of AI-2 was found to be with RpoE where 147-fold increases in *rpoE* regulation were observed (Sztajer et al., 2008). If AI-2 were to function in a similar manner with *S. pneumoniae*, then it may be the case that the *rpoE* KO is not responding to AI-2 mediated quorum sensing signals in a typical manner, which may result in differential biofilm formation behaviours.

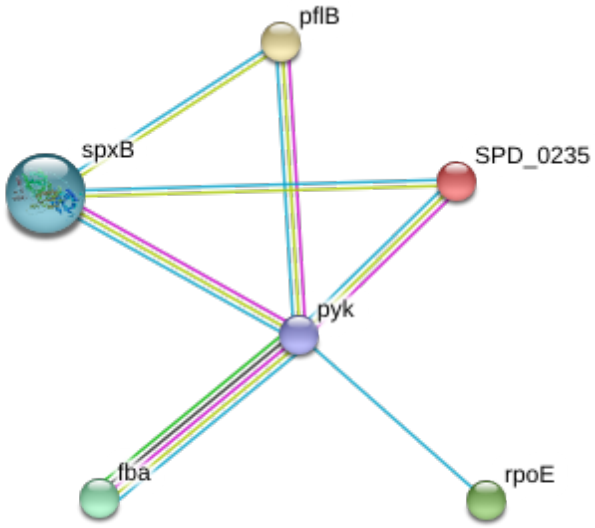
The potential for a *luxS*/AI-2 mediated mechanism may be supported by observations that *S. pneumoniae luxS* mutants display greatly reduced biofilm formation on both abiotic surfaces, *in vitro* cells (Vidal et al., 2013; Vidal et al., 2011) and in a rat model (Yadav et al., 2018). Loss of *rpoE* may cause dysregulation of key parts of the *luxS* pathways resulting in the increased biofilm formation relative to 22F WT. The potential for dysregulation of biofilm formation for *rpoE* KO is supported by the SEM results that showed that 22F WT and SCVs displayed broadly similar structures, with long chains of cells while *rpoE* KO did not display chain formation.

Exogenous AI-2 sensing by *S. pneumoniae* has been linked to the gene FruA which enables galactose metabolism (Trappetti et al., 2017). From the PM1 galactose utilisation assay, we observed a minor, although non-significant reduction, in galactose metabolism by *rpoE* KO compared to 22F WT. However, SCV5D3 and SCV9D9 do display significantly increased galactose metabolism relative to *rpoE* KO. This suggests that AI-2 activity may not be factor for SCVs in the same manner as *rpoE* KO.

The phenotypic differences observed between SCVs *rpoE* KO and 22F WT suggest that the mutations observed within SCV *rpoE* gene are not causing a loss of function, but instead indicates that SCVs are adapted to a broader metabolism which is more able to adapt to environmental sugars.

The increased metabolism of many sugars by SCVs may be the driving factor behind the increased $H_2O_2$ concentrations observed at 24-hours. Increased stress by $H_2O_2$ during attachment may promote 24-hour biofilm formation although no significant differences were observed at 72 hours. Previous work has reported that SCVs exhibit down regulation of pfl 2.27-fold (Churton, 2014). pflB is involved in the pyruvate to acetyl-CoA pathway. A reduction in this pathway may to result in increased $H_2O_2$ production conditions due to increased utilisation of the *spxB* pathway to metabolise pyruvate as shown below in Figure 39 (Ramos-Montanez et al., 2008).

**A**




**B**

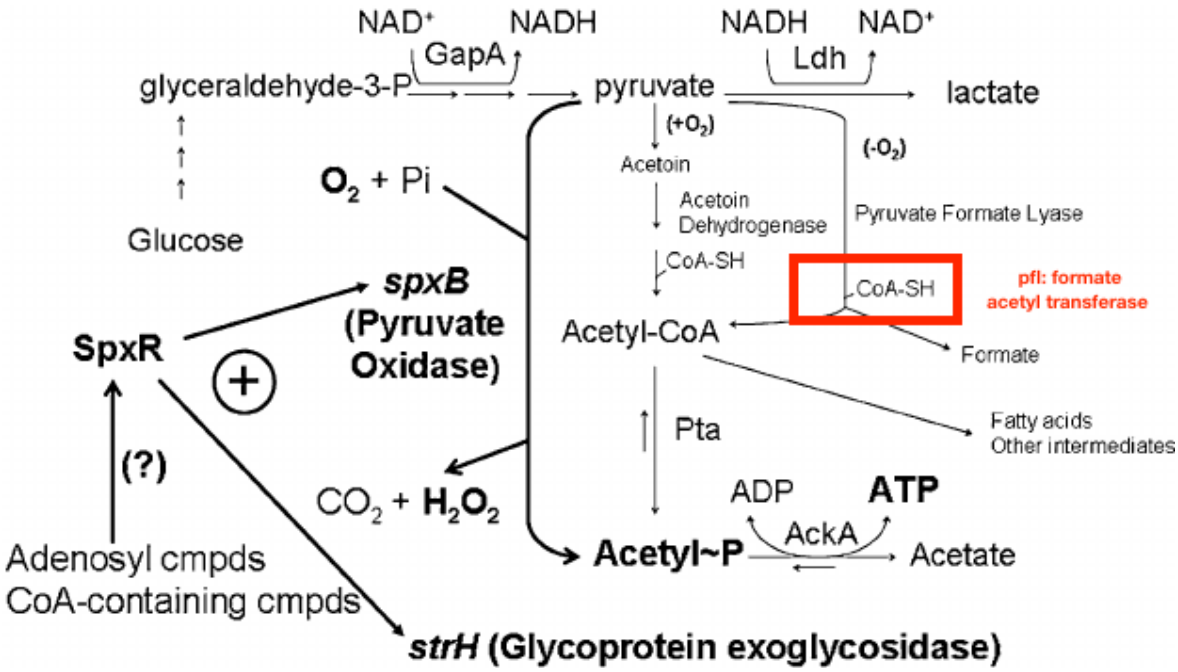



Figure 39 **A)** Map of protein interactions linking *rpoE* and *spxB*. **.B)** *S. pneumoniae* D39 spxB hydrogen peroxide pathway highlighting the pathway that is likely to be impacted by the reduction in pfl formate acetyl transferase. Adapted from Ramos-Montanez et al. (2008).

The increased $H_2O_2$ concentrations may facilitate control of competitors when establishing a biofilm. However, as *rpoE* SCVs were found to only form 1% of the population then this effect does not result in the SCVs out competing the WT(Churton, 2014).

To further explore the phenotypic traits and their potential for increased virulence an *in vivo* model is required. This is further explored in Chapter 5 through the use of the animal model *G. mellonella*.

### 4.4.2 Limitations and Future Work

We quantified biofilm formation through the use of crystal violet and CFU counts. While this provides valuable information on the quantities of cells and biomass produced, additional confirmation might be achieved through the use of confocal microscopy. Confocal microscopy would provide additional information on biofilm composition. Through the use of appropriately targeted lectins, we would quantify factors such as extracellular matrix produced and the proportion of live and dead cells. This would also provide greater information on the level of biofilm formation, and mitigate the low absorbance readings of the crystal violet readings.

We were able to discern biofilm structures through the use of SEM; however, this approach could be improved through the use of more advanced techniques. A limitation of SEM are the dehydration steps during sample preparation. These dehydration steps can produce artefacts and distortions and degrade biofilm structures. For instance, we did not observe the honeycomb sheet like structures within our biofilms which have been previously observed in *S. pneumoniae* biofilms (Figure 3) (Moscoso et al., 2006). While it is possible that 22F ST433 and the *rpoE* mutants do not produce these structures, it is also possible that the sample processing techniques by this group facilitate improved biofilm structure imaging. The main method difference between observations is the utilisation of low temperature scanning electron microscopy (LTSEM). LTSEM is able to reduce artefacts produced during the fixation and dehydration stages of processing (Bastacky et al., 1987). Utilisation of LTSEM may facilitate improved preservation of biofilm structures and enable investigation of a wider range of structures.

The Biolog PM1 assays rapidly provide information on a wide range of sugars and are therefore useful in determining potential key changes. However, a limitation to this approach is that each well is loaded with the addition of a single sugar. It may be of value in the future to determine the changes in metabolism when exposed to a more complex set of media with simultaneous increases in a range of sugars. This is addressed to an extent in our observations of adaptation to BHI from CBA and BHI pre-treatments. However, it would be of interest to look at a wider selection of complex media and sugar supplements.

RNA seq and proteomics would provide additional information on the genes which are being regulated by RpoE. This theme is explored further in Chapter 6. However,, in brief as RpoE is a global transcriptional regulator, we would expect a wide range of transcriptional changes as a result of its mutation and deletion. Quantification of the changes in expression of genes would provide valuable insights into the genes which are regulated by *rpoE* and are experiencing altered regulation by SCVs.

RNA seq would also aid in the determination of whether insensitivity to AI-2 plays a role in the phenotypic changes observed in *rpoE* KO. Exogenous application of strains including *rpoE* KO AI-2 followed by phenotypic quantification and RNA seq would enable quantification of this quorum sensing molecules effects for RpoE.

## 4.5 Chapter 4 Key Points

- 22F WT, SCVs and *rpoE* KO displayed distinct phenotypic differences.
- Both SCVs and *rpoE* KO displayed increased biofilm formation compared to 22F WT. However, *rpoE* KO biofilms did not exhibit the chain formation that is observed in 22F WT and SCV biofilms.
- SCVs displayed increased metabolism of several carbon sources relative to both 22F WT and *rpoE* KO.
- Within the SCV group, we observed a range of phenotypes with SCV5D3 and SCV9D9 exhibiting larger phenotypic changes for several traits than SCV1D3 and SCV3D9.

# Chapter 5: Pilot Investigation of *Galleria mellonella* as an Animal Model to Investigate *Streptococcus pneumoniae* Virulence, Colonisation and Biofilm Formation

## 5.1 Introduction

Biofilm formation *in vitro* does not necessarily correlate to *in vivo* observations (Blanchette-Cain et al., 2013). The host environment, epithelial cell interactions and molecular signalling are all important factors for biofilm formation that cannot be replicated within *in vitro* models (Marks et al., 2012a). In this chapter, we investigate *Galleria mellonella* as a virulence model and as the platform for the development of novel *in vivo* biofilm imaging techniques.

### 5.1.1 *Galleria mellonella* Overview

*G. mellonella* (greater wax moth) are increasingly used as an alternative animal model for human pathogens (Loh et al., 2013; Mukherjee et al., 2013; Wand et al., 2011). *G. mellonella* has several advantages over other animal models such as *C. elegans*, *D. melanogaster*, zebrafish and mammalian models due to the combination of its size, economical costs, viable temperature range and reduced ethical concerns (Cook and McArthur, 2013).

The large size of *G. mellonella* enables the precise injection of known amounts of *S. pneumoniae*, which would otherwise be impossible for smaller invertebrate models such as *D. melanogaster* and *C. elegans*. *G. mellonella* only requires minimal resource costs associated with its research when compared to more usual mammalian animal models such as mice or primates, enabling large inexpensive experimental populations (Jander et al., 2000; Olsen et al., 2011).

Furthermore, *G. mellonella* larvae are naturally suited to incubation temperatures of 34 - 37°C. This temperature range makes it suitable for the study of human pathogens such as *S. pneumoniae* and provides an advantage over the animal models *C. elegans*, *D. melanogaster*, and zebrafish whose temperature range is between 25-28°C (Glavis-Bloom et al., 2012).

Research utilising *G. mellonella* also benefits from similarities to the mammalian humoral immune system. These similarities are observed in highly conserved features such as pathogen recognition receptors, toll-like signalling pathways and the ability to phagocytose microbes in a highly similar manner to mammalian macrophages, and neutrophils (Kavanagh and Reeves, 2004; Royet et al., 2005; Tsai et al., 2016). Further similarities are observed in *G. mellonella* haemocytes which have been observed to express several proteins homologous to human phagocytic oxidase (phox) proteins, including GTP-binding protein rac 2, gp91$^{phox}$, p67 $^{phox}$ and p47 $^{phox}$. The high degree of homology between human neutrophil phox proteins and haemocytes indicates that the haemocytes of *G. mellonella* are likely through in a similar manner to human neutrophils through the use of a superoxide burst and phagocytosis (Bergin et al., 2005). During infection haemocytes are thought to aggregate with microbes attached to *G. mellonella* internal organs (Gagen and Ratcliffe, 1976) and phagocytose invading bacteria (Peleg et al., 2009). In addition to phagocytosis, *G. mellonella* possesses a wide range of antimicrobial peptides (Brown et al., 2009).

The suitability of *G. mellonella* as a model for human pathogens is further supported by observations that infections by several microorganisms in this model resemble those observed in human and murine models. *P. aeruginosa* mutant virulence in *G. mellonella* has been observed to correlate with that observed in murine models (Jander et al., 2000). Brain infections of *Listeria monocytogenes* in *G. mellonella* display similarities to the symptoms observed in humans, further indicating that the tissues of *G. mellonella* are a useful model for human infection (Mukherjee et al., 2013).

*G. mellonella* larvae have been shown to be a valuable model organism for the study of *S. pneumoniae* virulence. Differences in virulence were observed between strains of *S. pneumoniae* using *G. mellonella* with the principal factors involving the presence or

absence of a capsule and cell wall D-alanylation (Cools et al., 2018; Evans and Rozen, 2012).

However, this study has also identified several limitations for *G. mellonella* as an animal model for *S. pneumoniae* research. Reduced sensitivity by pneumolysin and *spxB* mutants has been observed with pneumolysin negative mutants displaying a 50% reduction in virulence. However, this reduction was only marginally significant and considerably less than the 10-fold virulence change observed in a mouse model (Berry et al., 1989). This indicates the *G. mellonella* may have reduced sensitivity to hydrogen peroxide or that the conditions within the larvae either mitigate its effects or supress expression of *spxB*. Further insights into infection of *G. mellonella* by *S. pneumoniae* and associations with biofilm formation come from investigation of *S. pneumoniae galU* mutants. *galU* is required for capsule production. Across multiple serotypes reduced capsule production, increased 48 hour *in vitro* biofilm formation and reduced virulence was observed (Cools et al., 2018). Although limited in scope, the literature suggests that *G. mellonella* is a suitable model for the investigation of *S. pneumoniae* strains with a recognition of the potential for a reduced response to pneumolysin and $H_2O_2$.

### 5.1.1 µ-CT for *Galleria mellonella* Biofilm Imaging

While *G. mellonella* has been utilised as a virulence model for numerous bacterial species little is known about the internal structures of *G. mellonella* larvae and which organs are colonised during infections or whether biofilm formation takes place. Current internal anatomical diagrams of *G. mellonella* physiology rely on histology with the inherent drawbacks of tissue disruption and challenges in image compilation for 3D modelling and quantification of biofilms (Perdoni et al., 2014; Rajendran et al., 2015). More recent methods have utilised GFP producing *Bacillus thurigenisis* which have been visualised within *G. mellonella*. This technique may lead to GFP based quantification of biofilm in the future. A potential drawback of such a GFP technique may be the melanisation of larvae in response to infection which is not observed within the *B. thurigenisis* model but is frequently observed in *S. pneumoniae* infections (Ramarao et al., 2012).
The size of *G. mellonella* (approximately 20 mm in length and 5 mm in diameter) also poses a challenge for the visualisation of biofilm formation. This size is too large for whole

organism visualisation with most traditional microscopy techniques without potentially disruptive sectioning.

To investigate the virulence and biofilm formation of 22F WT, SCVs and *rpoE KO* we will utilise *G. mellonella* animal model. To overcome the challenges of histology and microscopy we have utilised a micro computed X-ray tomography (µ-CT) approach. µ-CT is a non-destructive technology that is able to produce a 3D image of structures through X-ray scanning. Laval size is less challenging for µ-CT and due to X-ray detection being density-based, melanisation of the larvae does not inhibit the signal. We have developed the *G. mellonella* model through the use of micro computed X-ray tomography (µ-CT) to enable whole organism biofilm imaging.

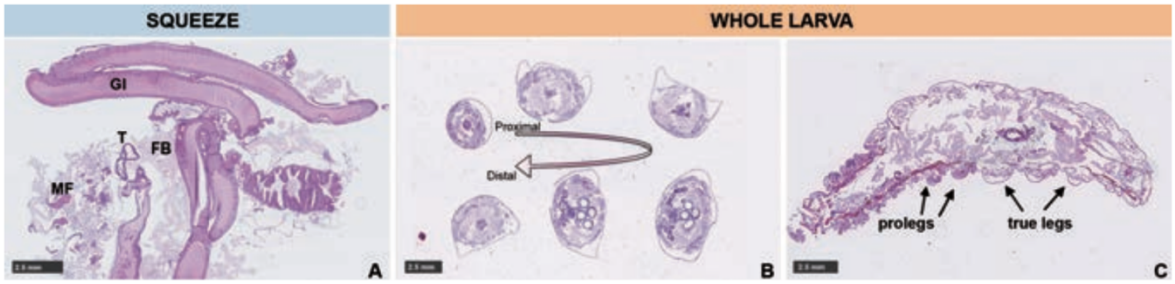



Figure 40: Histological sections of *G. mellonella* stained with haematoxylin, and eosin. **A)** Larval gut squeezed out showing the gastrointestinal tract (GI), tracheal system (T), fatty tissue (FB), and muscle fibres (MF). **B)** Proximal, middle, and distal transversal sections. **C)** Sagittal section of the whole larvae including prolegs, and true legs. The processing required to produce histological sections such as these results in distortion of tissues. Adapted from (Perdoni et al., 2014).

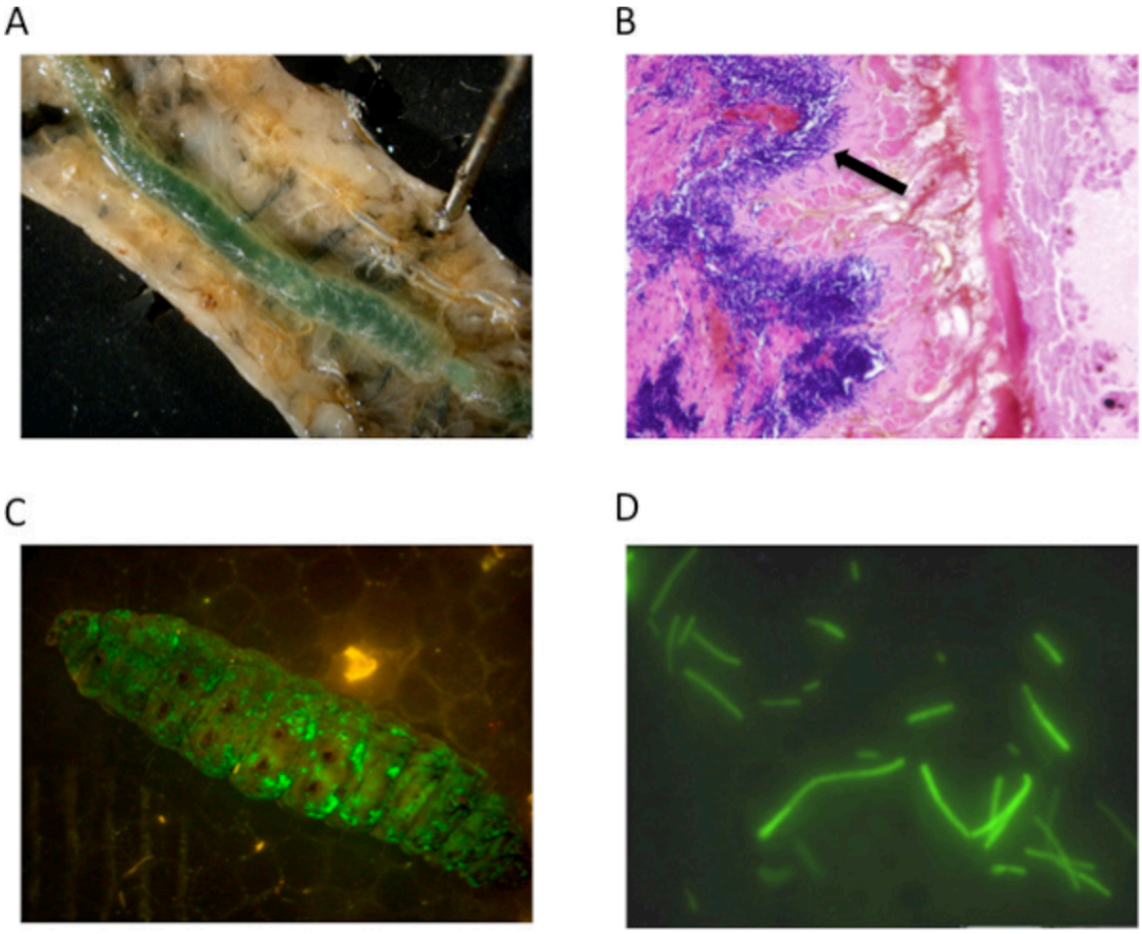



Figure 41: Techniques for the evaluation of bacterial infection within *G.* mellonella. **A)** 10 µl of blue dye was force-fed to *G. mellonella* demonstrating the capability to introduce substances into the gut. **B)** Histological section of *Bacillus thuringiensis* localised to the *G. mellonella* gut wall. **C)** Fluorescent *B. thuringiensis* that have escaped from the gut and spread throughout the hemocoel. **D)** Fluorescent *B. thuringiensis* isolated from the hemocoel following infection. Adapted from (Ramarao et al., 2012).

## 5.2 Materials and Methods

The materials and methods required to carry out the work in this chapter are detailed in Chapter 2 Materials and Methods.

## 5.3 Results

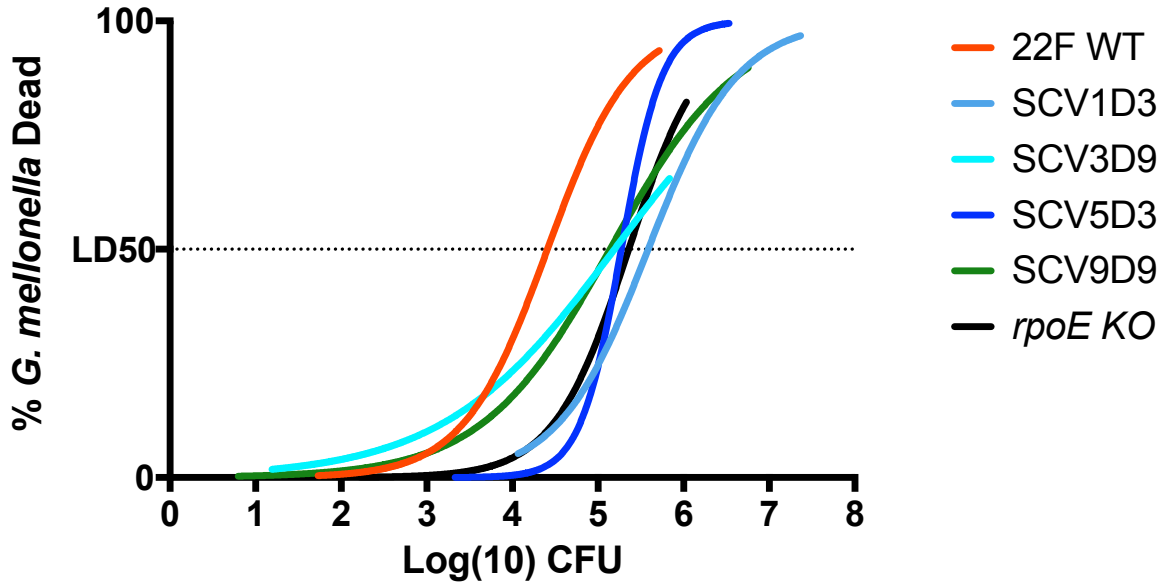
### 5.3.1 *Galleria mellonella* Virulence Model

*S. pneumoniae* isolates, were grown in suspension for inoculation of *G. mellonella* larvae. Following these inoculations, we enumerated live and dead larvae facilitating calculation of the lethal dose 50 (LD50) which is the number of CFU required to kill half the population. We observed that SCVs and *rpoE* KO displayed diminished virulence within the *G. mellonella* model following inoculation into the left rear leg. Infections by SCVs and *rpoE* KO required approximately 1 log higher concentrations for an inoculation that would kill half of infected larvae.

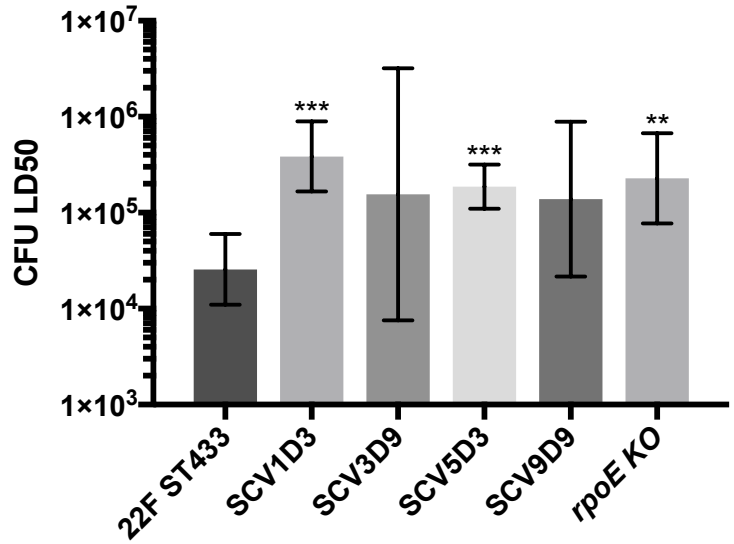
The strains SCV1D3, SCV5D3, and *rpoE* KO all displayed significantly reduced virulence compared to 22F WT. While SCV3D9 and SCV9D9 were not significantly different to the wild-type these isolates still followed the trend of reduced virulence and with an increased number of repeats we may observe significant differences. We did not observe significant differences between SCVs or *rpoE* KO.

Similar results were observed in the in the Kaplan-Meier survival assay with 22F WT displaying the highest level of virulence, while SCV9D9 and *rpoE* KO showed identical levels of virulence. These results support a model of reduced virulence for SCVs compared to 22F WT.

A


100
LD50
0
% G. mellonella Dead
0
1
2
3
4
5
6
7
8
Log(10) CFU
22F WT
SCV1D3
SCV3D9
SCV5D3
SCV9D9
rpoE KO

B


1×10^7
1×10^6
1×10^5
1×10^4
1×10^3
CFU LD50
***
***
**
22F ST433
SCV1D3
SCV3D9
SCV5D3
SCV9D9
rpoE KO

C

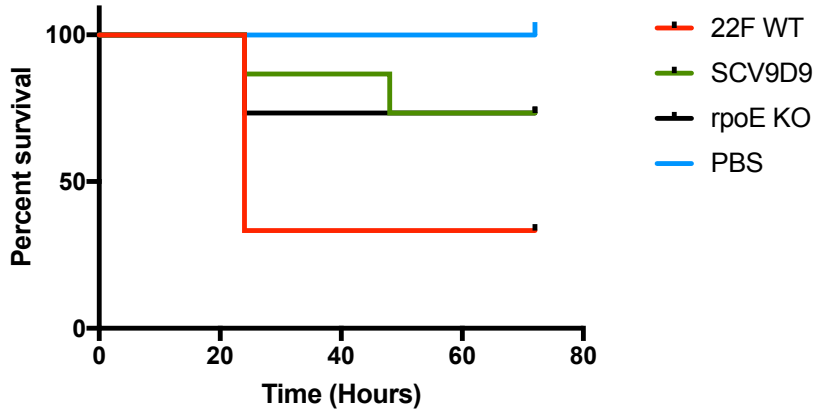

100
50
0
Percent survival
0
20
40
60
80
Time (Hours)
22F WT
SCV9D9
rpoE KO
PBS

Figure 42: Lethal dose 50 (LD50) of CFU for WT 22F, SCVs, and *rpoE* KO. The virulence required LD50 was calculated through GraphPad Prism 6.0 using a non linear fit effective concentration model. **A)** *S. pneumoniae* inoculation required to kill *G. mellonella* at 72 hours**B)**Differences compared to 22F WT. Strains displayed a variety of differences to 22F WT, SCV1D3 ($p < 0.001$) SCV3D9 ($p > 0.05$, ns), SCV5D3 ($p < 0.001$), SCV9D9 ($p > 0.05$, ns), and *rpoE* KO ($p < 0.01$). Error bars represent 95% confidence intervals. For both **A** and **B** a minimum of three independent inoculums was used. Each independent inoculum was diluted to 4 concentrations between 1 x $10^1$ and 2 x $10^7$ CFU dependent on the strain tested, e.g. more virulent strains required a lower concentration of CFUs to effectively calculate the LD50. Each of these dilutions was inoculated into 10 larvae, and the live and dead larvae were counted after 72 hours. **C**) Kaplan-Meier survival of 22F, SCV9D9, *rpoE* KO and PBS. For each isolate, 3 biologically independent suspensions of *S. pneumoniae* were made cultured, with each suspension inoculated into 10 larvae. Each larvae received 1 - 2.5 x $10^6$ CFU and wasincubated at 37°C for 72 hours. Live and dead larvae were counted every 24 hours for 72 hours. A Log-rank (Mantel-Cox) and Gehan-Breslow-Wilcoxon test both found that survival curves were significantly different with P values of <0.0001.

### 5.3.2 µ-CT of *Galleria mellonella* utilising PTA

To improve our understanding of the reductions in virulence observed we set out to develop the *G. mellonella* model to incorporate visualisation of host structures and biofilm formation, with the aim of determining whether *S. pneumoniae* infections were localising to any particular tissues within the larvae. To visualise the tissues within *G. mellonella* we utilised a combination of phosphotungstic acid and µ-CT. µ-CT uses x-rays to non-destructively image and morel 3D targets. As an x-ray system, µ-CT is dependent on density gradients in order to differentiate between tissues. In biological samples this necessitates the use of a contrast agent as biological tissues are principally water. Utilising phosphotungstic acid (PTA) we were able to visualise the internal structures of whole *G. mellonella* larvae without the need for dissection or sectioning.

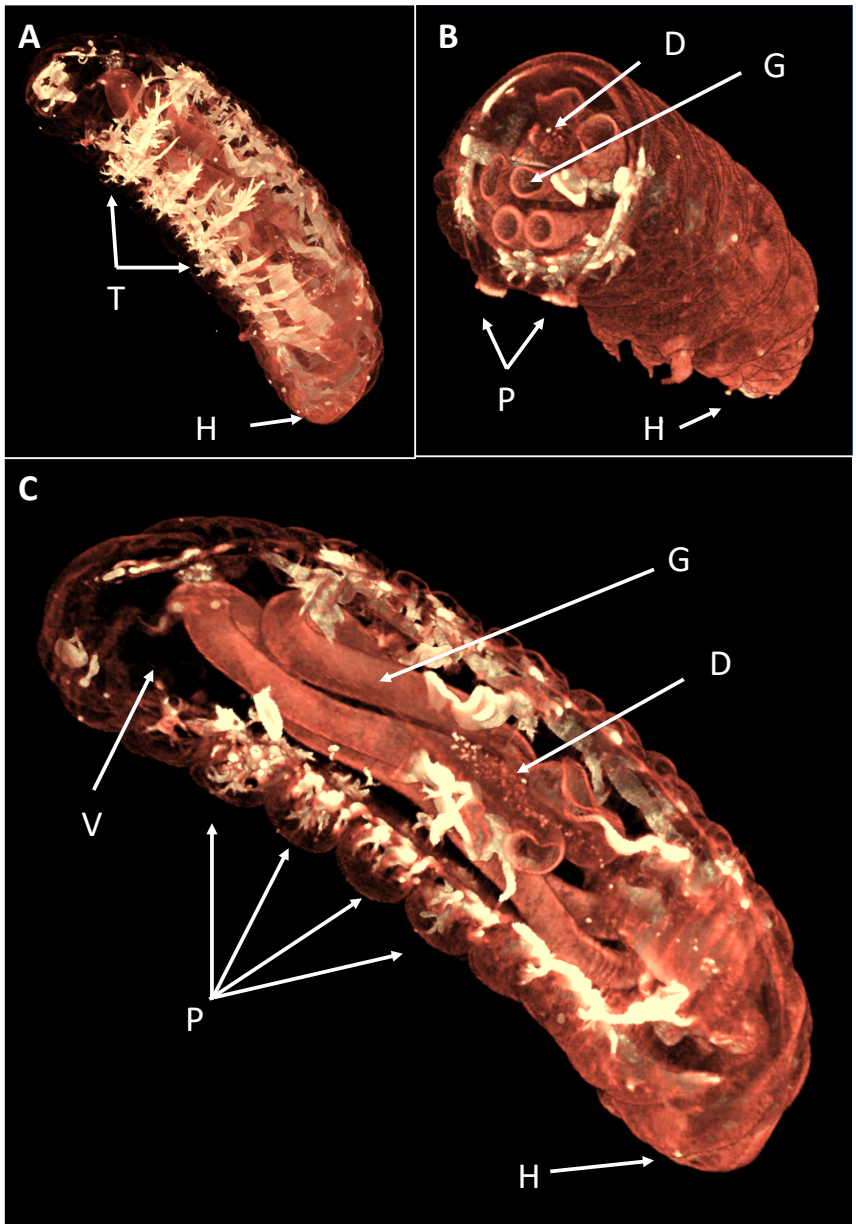



Figure 43: µ-CT imaging of pseudocoloured volume renderings of a PTA (phosphotungstic acid) – stained *G. mellonella* larvae. Internal structures are shown such as the tracheoles [T], the gut [G], the pro-legs [P], and the head of the larvae is indicated by [H]. [D] may show the dorsal vessel. A) shows a section displaying the network of tracheoles. B) shows a section through the transverse plane of the caterpillar, just after the first pro-legs. C) shows a longitudinal section through the sagittal plane. The upper left portion of C) shows the rear of the larvae and the void [V] into which inoculation takes place.

### 5.3.3 Visualisation of Biofilm Formation by D39 within *G. mellonella*

To differentiate biofilm from host tissue a *S. pneumoniae* biofilm specific marker is required. We chose a D39 specific antibody combined with a secondary conjugate antibody with 10 nm gold particles. Due to the considerabley higher density of gold than the surrounding tissues this enables effective differentiation between tissues and biofilm.

We utilised 1 mm sections of *G. mellonella* larvae (Figure 44) to validate that biofilm antibodies would bind to *S. pneumoniae* within the larvae without the potential complication of permeabalisation of the outer larvae's skin to facilatate antibody access. Unfortunately we were unable to utilise the developed antibody mediated techniques in parallel with PTA.

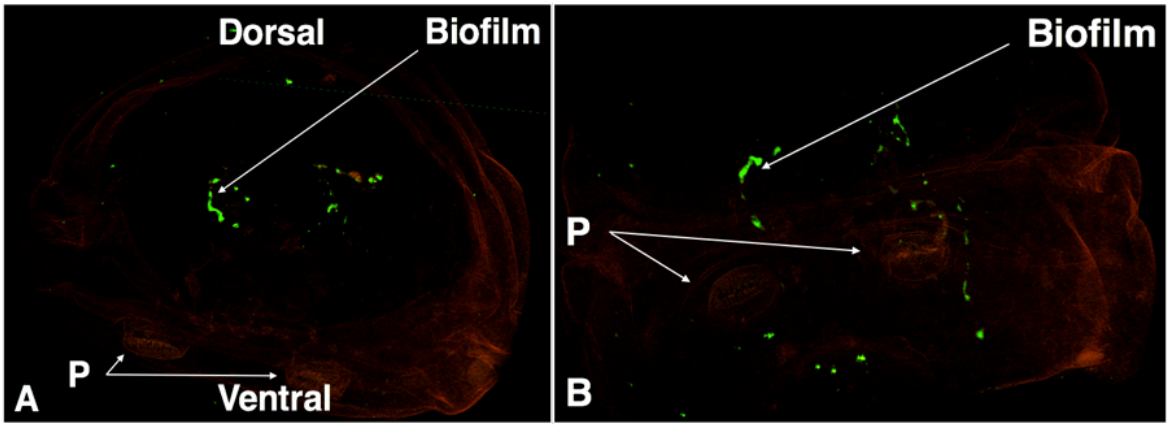



Figure 44: µ-CT scans of *S. pneumoniae* D39 biofilm growth in a 1 mm section of *G. mellonella*. Each larvae received 5 x $10^6$ CFU in a 5 µL volume in PBS and was incubated for 72 hours. Antibodies targeted to *S. pneumoniae* surface protein A are conjugated with 10 nm gold labelled secondary antibodies enabling visualisation of biofilm displayed in green. A) Shows the scan across the transverse plane. B) shows the same section viewed from the dorsal side. P = Prolegs. N = 1.

### 5.3.4 Whole Organism Biofilm Visualisation

We have determined that it is likely that *S. pneumoniae* biofilms are forming within *G. mellonella* and that it is possible to differentiate these biofilms from animal model tissue. To fully take advantage of μ-CT a dissection free approach is desirable to avoid the tissue disruption previously mentioned. To further develop the technique and enable whole organism biofilm visualisation we used a proteinase K digest to permeabilise the skin of the larvae to antibodies.

Larvae were infected with *S. pneumoniae* D39 and incubated for 24 hours to enable the infection to establish. Following this 24-hour incubation larvae were treated with tetracycline or a PBS control and incubated for a further 3 days. At the final time point 90% of larvae treated with tetracycline survived while those which received only the PBS control had all died.

Following scanning we are able to observe considerably reduced biofilm formation within the larvae which received antibiotic treatment. This indicates that the μ-CT protocols developed are sufficiently sensitive to differentiate between different levels of biofilm formation. The missing ends of the larvae on the left-hand side of the images have not been lost due to sectioning but were out of the field of view during scanning.

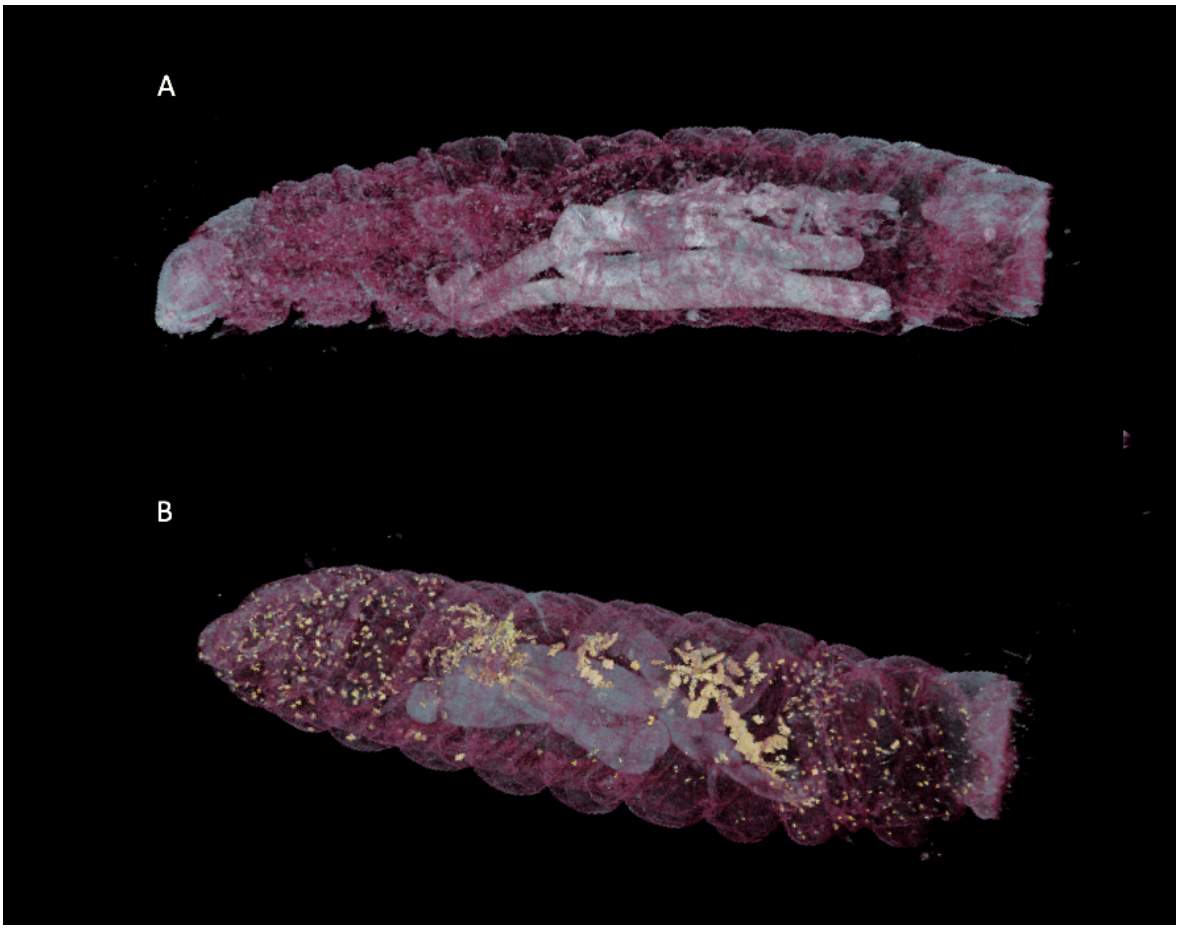



Figure 45: Pilot study of whole organism µ-CT scans of *G. mellonella* with D39 *S. pneumoniae* biofilms labelled. Each larvae received 5 x $10^3$ CFU in a 5 µL volume in PBS and was incubated for 24 hours after which each larvae received 2.5 µg/g weight of larvae dose of tetracycline in a 5 µL volume of PBS and incubated for a further 72 hours. 10 larvae were treated with the tetracycline treatment and the PBS control. 90% of larvae receiving the tetracycline treatment survived, while all larvae receiving the PBS control died. Biofilms are labelled with 10 nm gold conjugate antibodies. **A** has been treated with 1.25 µg/g tetracycline. **B** Shows the PBS control. N =1 biological replicate for pilot study scans.

### 5.3.5 Adaptation for use with 22F WT, SCV9D9 and *rpoE* KO

To investigate biofilm formation by 22F WT, SCVs and *rpoE* KO within the *G. mellonella*/μ-CT system we utilised the techniques developed in D39. We identified an anti-22F antibody as a suitable primary antibody for the investigation of biofilm formation by 22F WT, SCV9D9 and *rpoE* KO without displaying cross reactivity with the internal flora of *G. mellonella*.

For this model to be as relevant as possible for biofilm infections (the majority of which are chronic and non-lethal) we used a concentration of bacteria that was unlikely to kill all the larvae for any strain used so that only live larvae could be scanned. However, it is also necessary that a sufficient concentration of bacteria to cause a significant infection is used to ensure that the inoculation is not immediately removed by the larvae's immune system. Based on previous LD50 calculations a concentration of $1.5 \times 10^6$ CFU per larvae was chosen as providing a sufficient inoculum to produce a heavy infection while also providing sufficient surviving larvae for μ-CT analysis.

Larvae were inoculated with $1 - 2.5 \times 10^6$ CFU per larvae and incubated for 3-days to allow mature biofilms to establish which would be comparable to our *in vitro* observations. We observed melanisation within larvae infected with *S. pneumoniae*, while the PBS control displayed no such melanisation. Melanisation occurs under stress conditions and therefore the lack of melanisation in the PBS control indicates that we have not inadvertently infected this control with *S. pneumoniae* or any other environmental microbe. The greatest degree of melanisation was observed along the dorsal side of the larvae on the sagittal plane.

Through μ-CT we visualised biofilm formation within the larvae without the requirement for dissection. Biofilms were visualised in both 22F WT and *rpoE* KO. Biofilm formation was not observed by SCV9D9 with the level of gold nanoparticles observed within the larvae less than that within the PBS control. Within the PBS control we observed two aggregates. Potentially these aggregates may be antibodies bound together, or due to insufficient washing.

Considerable biofilm formation with 22F WT was observed. The biofilms did not appear to be attached to the gut but were likely to have principally localised around the dorsal vessel (the larvae analogue to the heart which circulates the haemocoel) with further minor biofilm formation observed throughout the larvae. This may be supported by the observations of melanisation to the dorsal side of the larvae indicating more intensive infection in these areas. Similar trends of biofilm formation are observed for *rpoE* KO however more diffuse and smaller aggregates are observed.

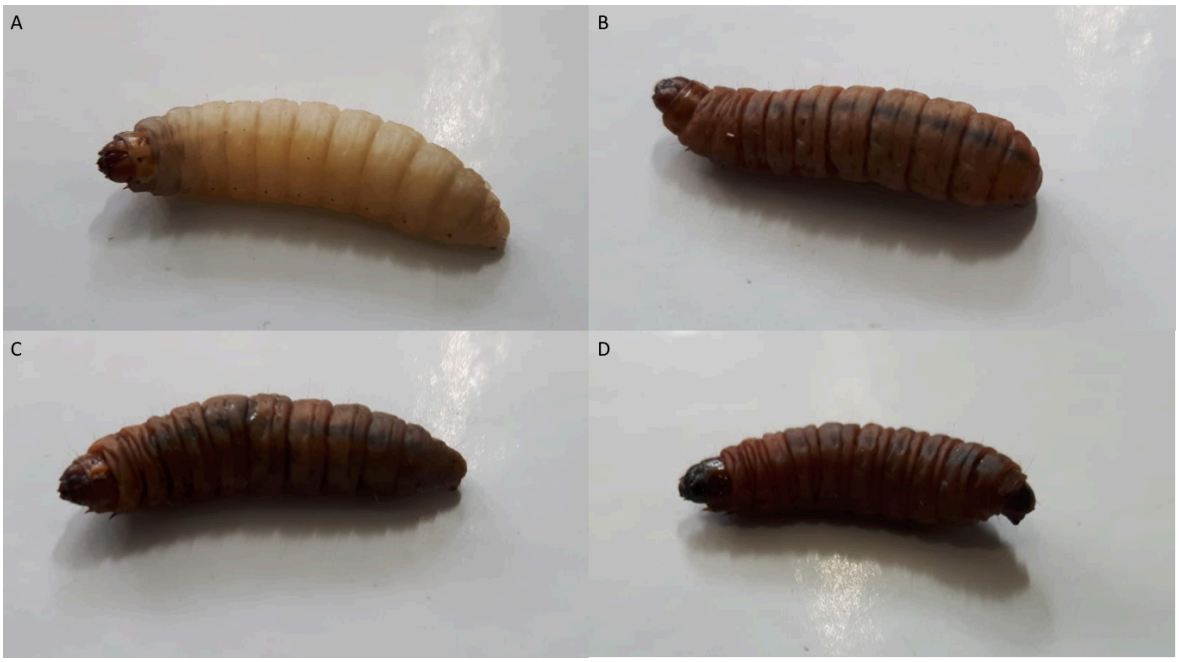



Figure 46: Representative imgages of *G. mellonella* inoculated with *S.pneomoniae* 22F isolates and incubated for 72 hours. Each larvae received 1 - 2.5 x $10^6$ CFU and were incubated for 3 days to allow for biofilm formation. A) PBS control. B) 22F C) SCV9D9 D) *rpoE* KO. All the larvae shown were alive following the 3 day time course. The larvae inocualted with *S. pneumoniae* cultures all displayed heavy melanisation as a result of the infection with the highest degree of melanisation along the dorsal side of the saggital plane. All larvae were alive at the point of processing for imaging.

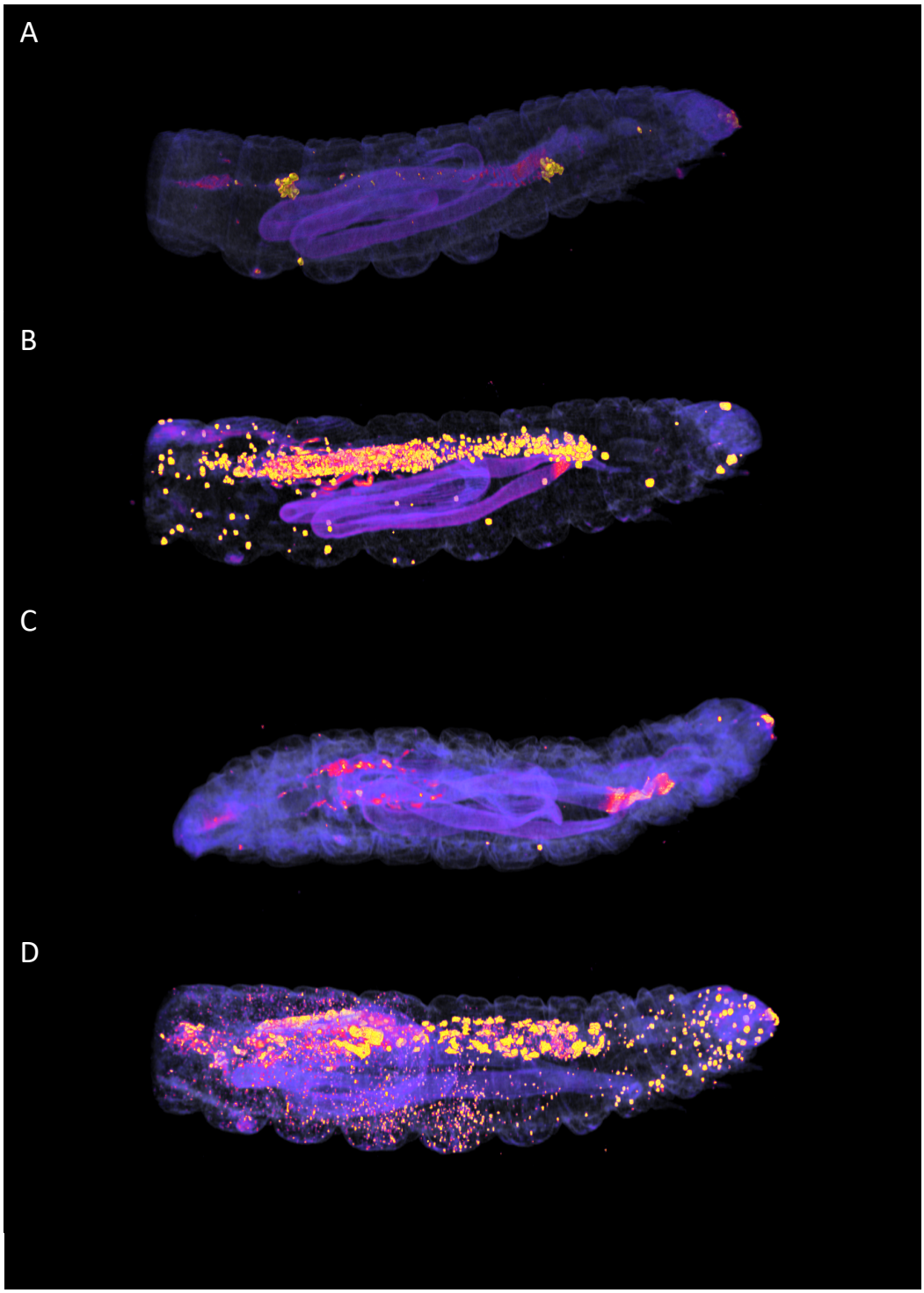



Figure 47: 3D μ-CT scans showing biofilm formation within *G. mellonella* by *S. pneumoniae* strains inoculated with 1 - 2.5 x $10^6$ CFU and incubated for 72 hours. All larvae were alive at the time of fixing. Biofilms are labelled through the use of an anti-22F antibody and a secondary antibody conjugated to a 10nm gold nanoparticle. A) PBS control. B) 22F WT. C) SCV9D9. D) *rpoE* KO. N = 1 larvae per condition.

## 5.4 Discussion

Arguably the most important phenotypic trait of a pathogen is its ability to colonise and kill its host. Phenotypic characterisation though *in vitro* assays provides valuable information for determining the potential impact that *S. pneumoniae* strains and mutants may have in human hosts. However, biofilm formation and other phenotypic traits are affected by a myriad of *in vivo* factors such as cell interactions and host signalling (Marks et al., 2012a) and therefore animal models are invaluable tools for evaluating virulence and clinically relevant factors. Insects are increasingly used as animal models including the greater wax moth *G. mellonella*. *G. mellonella* has been previously demonstrated to be a suitable model for determining the virulence of strains and serotypes of *S. pneumoniae* (Cools et al., 2018; Evans and Rozen, 2012), and that it is a suitable animal model to study this pathogen.

In the *G. mellonella* model we observed reduced virulence by the SCVs and *rpoE* KO relative to 22F WT. This is likely to be due to the lower capsule expression observed in SCVs (Churton et al., 2016) leading to faster clearance through phagocytosis. This hypothesis is supported by work across several *S. pneumoniae* serotypes which observed that reduced capsule expression correlated to diminished virulence by *galU* mutants in a *G. mellonella* model. Our observations are further supported by investigations of a *Streptococcus agalactiae rpoE* knock-out mutant which displayed diminished virulence in rat models and that *rpoE* is associated with protection from phagocytosis (Jones et al., 2003; Seepersaud et al., 2006). Additionally, reduced virulence is also observed by *S. aureus rpoE* knock-out in a mouse model and is further supported by considerable down-regulation of known virulence factors (Weiss et al., 2014). Reduced virulence has also been observed in *S. pneumoniae* SCVs with mutations to genes other than *rpoE* which were also associated with reduced capsule expression (Tikhomirova et al., 2018).

These observations contrast with research in an *S. mutans rpoE* knock-out which indicated that the loss of *rpoE* is associated with an increase in traits relevant to virulence (Xue et al., 2011). However, this investigation did not utilise an *in vivo* model to validate their predictions of increased virulence and highlights the need for such models.

The most likely mechanism of action for the reduced virulence is through a model of increased phagocytosis by haemocytes due to reduced capsule expression. The SCVs and *rpoE* KO are grouped together without significant differences between strains, we did not observe the continuum of phenotype that was seen for traits such as biofilm formation and several sugar metabolisms in Chapter 4. This suggests that virulence is impacted to the same degree by mutation and complete removal of *rpoE*. However, the confidence intervals for the virulence data are fairly large, presenting challenges for determination of more subtle phenotypic variation. Improved confidence might be achieved through increased experimental population sizes, however increased confidence may also be possible through a change in *G. mellonella* supplier. In our assays we utilised 'pet food grade' larvae, however, 'research grade' larvae are now commercially and may reduce variability between larvae.

A limitation of the *G. mellonella* model is its lack of sensitivity to pneumolysin (Ply). During infection of *G. mellonella*, Ply deficient *S. pneumoniae* strains produced only minor reductions in virulence relative to the wild-type, which is contrary to the large reductions in virulence by Ply mutants observed in mammalian infections (Evans and Rozen, 2012). However, no significant changes to Ply production were observed during proteomic analysis of 22F WT and SCVs analysed (Churton, 2014) and therefore this is unlikely to be a relevant factor for SCVs and *rpoE* KO. If hydrogen peroxide is a significant factor in infection by SCVs (which would be supported by the increased 24-hour hydrogen peroxide expression observed in Chapter 4) then it may be mitigated within this model as *spxB* mutants in *G. mellonella* have not been observed to display significantly altered virulence, while in murine models $H_2O_2$ is thought to play an important role in adhesion (Evans and Rozen, 2012).

Further investigation into the virulence of SCVs and *rpoE* KO could be improved in a number of ways. Cells for inoculation were generated through planktonic growth, while this matches the methods used in the literature (Evans and Rozen, 2012). However, as the mutations to *rpoE* in SCVs occur during biofilm growth, it may be of value to utilise biofilm cultures. This would be best carried out through two assays, one utilising only biofilm cells and another utilising the planktonic cells released from the biofilm. Such an assay would address questions surrounding virulence related to the biofilm lifestyle and

whether *S. pneumoniae* biofilms display increased virulence or alternatively that they provide a reduced virulence population from which cells with increased virulence are released (Marks et al., 2013; Sanchez et al., 2011).

Moving to a larger animal model would enable investigation of chronic infection which may be more suited to the biofilm adapted SCVs than the direct injection techniques that we have used with *G. mellonella*, which are more representative of invasive disease. The chinchilla model with their large bullae would provide an excellent platform for the investigation of a chronic biofilm infection analogous to otitis media (Reid et al., 2009).

The increased survivability of a chinchilla or other murine model would also facilitate testing of whether biofilms grown *in vivo* with all the additional host factors known to influence biofilm formation (Marks et al., 2012a) would also produce SCVs with mutations to *rpoE*. This research would provide valuable evidence as to whether *rpoE* mutations are clinically relevant and identification of whether *rpoE*/RpoE is a suitable target for future use in medical intervention design.

### 5.4.1 Development of µ-CT for the Investigation of Biofilm Formation Within *Galleria mellonella*

Internal structures were visualised through the use of the contrast agents phosphotungstic acid and iodine. Through these contrast agents we were able to investigate the internal structures of *G. mellonella* to a higher level of detail than has been previously achieved and without the disruptive effects of dissection. Gold nanoparticles have been increasingly used for µ-CT applications. However as far as the author is aware, this is the first use of gold nanoparticles conjugated with antibodies, for the investigation of bacterial biofilms within an animal model.

We have developed the methods to successfully visualise biofilm formation throughout *G. mellonella*. While these techniques are still in their infancy we have been able to see clear differences in biofilm formation following antibiotic treatment and *S. pneumoniae* strains.

While biofilm formation within *G. mellonella* was observed by 22F WT and *rpoE* KO, SCV9D9 produced no visible biofilm. This contrasts with our *in vitro* observations where SCV9D9 exhibited the greatest biofilm formation of all the SCVs in addition to exceeding that of *rpoE* KO and 22F WT.

While these observations are based on observations of only a single larva, and further work with larger sample sizes is clearly required for full validation these results further emphasise the phenotypic differences between SCVs and *rpoE* KO.

These results may suggest that SCV9D9 is being cleared from the larvae by the host immune system. However, the LD50 calculations for SCV9D9 are similar to both *rpoE* KO and 22F WT indicating that complete clearance is unlikely to be occurring. An alternative hypothesis is that within *G. mellonella* SCV9D9 is more suited to a planktonic lifestyle within the haemolymph. A potential mechanism for this lifestyle could be through the more versatile metabolism exhibited by SCV9D9. The haemolymph of *G. mellonella* is rich in the sugars glucose, fructose and sucrose which SCV9D9 may be more able to ustilise (Wyatt et al., 1956). In the Biolog assays we observed significantly increased metabolism by SCV9D9 of glucose and fructose with sucrose following this trend although not to the same degree. Increased glucose metabolism has been linked to reduced biofilm formation and biofilm dispersal in other *S. pneumoniae* strains (Chao et al., 2014; Moscoso et al., 2006). These planktonic cells may then be removed through wash steps during µ-CT antibody preparations. An alternative hypothesis is that as SCV9D9 was produced as a result of selection factors to an *in vitro* environment it is ill suited to biofilm formation under *in vivo* conditions regulated by numerous host factors (Marks et al., 2012a).

In addition to the differences in the quantity of biofilm formation observed between isolates, we also observed differential distribution. While *rpoE* KO displayed similar levels of biofilm formation to 22F WT and both appeared to be principally bound to the dorsal vessel *rpoE* KO appeared to exhibit more diffuse biofilm formation throughout the larvae. The dorsal vessel acts as the heart of the larvae pumping the haemolymph throughout its tissues. *S. pneumoniae* is known to colonise the human heart during pneumococcal endocarditis (de Egea et al., 2015; Lacalzada et al., 2016).

### 5.4.2 Limitations and Further Work

Targeting of a specific area within larger animal models may provide additional valuable information. µ-CT has already been successfully utilised to image lung and bone within mice (Ashton et al., 2015). Staining of bacterial biofilms with gold labelled antibodies may aid in the modelling of conditions such as cystic fibrosis. Quantification of biofilms without the disruption inherent in sampling would be of value in the development of biofilm treatments such as nitric oxide mediated biofilms dispersal (Barraud et al., 2009).

Further work would also benefit from moving from the qualitative assay to a quantitative approach. This would involve increased samples sizes for each inoculum combined with volume quantification of bound gold nanoparticles. Further optimisation of the techniques detailed in this chapter would incorporate contrast agents compatible with gold nanoparticles mediated visualisation. Currently we are unable to utilise both contrast agents and antibody labelling within the same tissue without the contrast agent signal overwhelming the antibody.

In order to reduce the variability present within wholesale *G. mellonella* larvae improvements may include the use of a standardised *G. mellonella* larvae such as those provided by TruLarv (https://biosystemstechnology.com). These larvae are not exposed to antibiotics or hormones during growth; a factor that is not reported by wholesale larvae suppliers, and are reported to provide a higher degree of reproducibility (Champion et al., 2016). However, these more artificial growth conditions may also impact the immune system of the larvae making them less representative.

Targeting of a specific area within larger animal models may provide additional valuable information. µ-CT has already been successfully utilised to image lung and bone within mice (Ashton et al., 2015). Staining of bacterial biofilms with gold labelled antibodies may aid in the modelling of conditions such as cystic fibrosis. Quantification of biofilms without the disruption inherent in sampling would be of value in the development of biofilm treatments such as nitric oxide mediated biofilms dispersal (Barraud et al., 2009).

## 5.5 Chapter 5 Key Findings

- Through the use of *G. mellonella* model we are able to determine that the 22F WT is more virulent in this model than the SCVs or *rpoE* KO

- We have piloted a novel method for biofilm visualisation within *G. mellonella* through the combination of immunohistochemistry and µ-CT.

- Using contrast agents, we are able to visualise the internal structures of *G. mellonella* in greater detail than has been previously achieved.

- Using µ-CT and gold labelled antibodies we are able to visualise *S. pneumoniae* within *G. mellonella* without the need for dissection or manipulation of tissues.

- This work lays the groundwork for improved visualisation of biofilms within animal models and may be applicable in larger animal models investigating targeted areas.

# Chapter 6: RpoE Structure and Function

## 6.1 Introduction

In Chapter 4 we observed both phenotypic differences and similarities between 22F WT, SCVs and *rpoE* KO. In this chapter we investigated the potential causes and mechanism for the phenotypic variations observed. Whilst the *rpoE* KO displayed some similarities to SCVs we also observed striking differences in growth rates, biofilm formation and the metabolism of several sugars. These observations suggest that the mutations to *rpoE* within SCVs are not causing complete loss of function that was observed with *rpoE* KO. Furthermore, SCVs do not display a uniform phenotype with SCV1D3 displaying less severe phenotypic differences to 22F WT than SCV9D9. These results suggest that the different mutations to *rpoE* observed within SCVs may cause a spectrum of changes to RpoE dependent upon the mutation involved; the mutations range from SNPs to large scale deletions. The mutations to *rpoE* observed in SCVs are likely to cause a range of alterations to the structure of RpoE. This spectrum of alterations may not produce identical changes to the regulatory functions of RpoE on RNAP (RNA polymerase), resulting in altered levels of expression of genes regulated by RpoE.

It is now possible to predict protein structures and functions of through computational modelling based on a protein's amino acid sequence and its homology to known protein structures. Protein structure predictions provide indications as to its function and the mechanisms through which the protein operates. Protein prediction server suites ITASSER (Yang et al., 2015), Robetta (Kim et al., 2004), and RaptorX (Kallberg et al., 2012) are able to model secondary and tertiary protein structures and predict functional properties. We explore the potential reasons for the differences observed between 22F WT, SCVs and *rpoE* KO through investigation of the RpoE isoform structures.

## 6.2 Materials and Methods

The materials and methods required to carry out the work in this chapter are detailed in Chapter 2 Materials and Methods section Molecular Modelling Methods.

## 6.3 Results

### 6.3.1 RpoE Domain Determination

The amino acid sequences of RpoE from 22F WT and SCVs were determined through the use of ExPASy Translate tool. Amino acid sequences are shown in Appendix 8.4.2 RpoE Amino Acid Sequences.

Proteins are often made up of multiple domains. A protein domain is a distinct functional unit with particular functions and interactions; as a first step in investigating RpoE, we determined its constituent domains. RpoE has been found in related species to be formed of two domains (Motackova et al., 2010; Weiss and Shaw, 2015). To confirm that 22F WT RpoE protein structure is similarly arranged, we analysed the 22F WT amino acid sequence through Ginzu prediction using the Robetta Full-chain Protein Structure Prediction Server (Kim et al., 2004). This server predicted two domains within 22F WT RpoE similar to those found within *B. subtilis* RpoE. The N-terminal domain is predicted to consist of 85 amino acids, while the C-terminal domain is made up of 95 amino acids. Sequencing of SCVs revealed a mix of SNPs and deletions (Churton, 2014); however all *rpoE* mutations observed occurred within the C-terminal domain while the N-terminal domain was conserved. The *rpoE* mutations within the SCVs resulted in truncations of the C-terminal domain producing a range of different sized isoforms as shown in Figure 48.

That all the mutations within SCVs occur within the C-terminal domain indicates that this domain must play an important functional role and is likely linked to the phenotypic range observed within SCVs. However, the striking phenotypic differences observed between SCV9D9 and *rpoE* KO are likely due to the N-terminal domain as SCV9D9 only possess a 7 amino acid C-terminal.

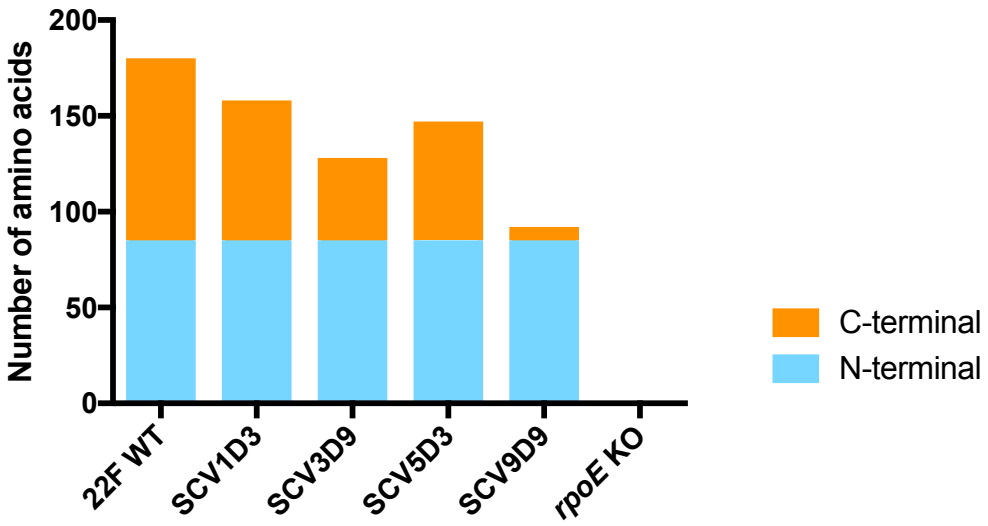



Figure 48: RpoE isoform lengths. The number of amino acids for each isoform of RpoE for 22F WT and SCVs. The N-terminal domain has been conserved during the parallel evolution of *rpoE* that produced SCVs.

### 6.3.2 Domain Order and Acidity

Analysis of the amino acid sequence of 22F WT revealed that the C-terminal domain is highly acidic, repetitive and disorganised. In contrast, the N-terminal domain is highly structured, but also contains numerous acidic residues as shown below.

The high number of acidic amino acids present within the C-terminal domain causes RpoE to be charged at physiological pH. This charge may be determined through the amino acid sequence and results in a theoretical isoelectric point (pI). The theoretical pI is the pH at which a molecule is predicted to have no net electrical charge. We observe that larger truncations of the C-terminal domain result in reduced pI and charge at a physiological pH. The extremely acidic nature of RpoE is emphasised by comparison to the other proteins within *S. pneumoniae* as shown by the protein pI distribution of several reference proteomes as shown in Figure 51.

Due to the highly acidic composition of the C-terminal domain, the truncated isoforms of the SCVs display increased pI (and therefore less charged RpoE at physiological pH). The pI of RpoE correlated strongly to the length of the truncated C-terminal.

| RpoE amino acid sequence | Number of amino acids |
|---|---|
| MIEVARAILELRGRDHEMHFSDLVNEIQNY | 30 |
| LGTSNSDIREALPLFYTELNFDGSFISLGD | 60 |
| NKWGLRSWYGVDEIDEEIIALEENDDDEVA | 90 |
| PKAKKKRVNAFMDGDSDAIDYNADDPEDED | 120 |
| AYEADPALSYDDENPDDEKNEVEAYDAEIN | 150 |
| EIAPDDLGEDVDLNEDDDEFSDDDAETSEE | 180 |

Table 6: Amino acid sequence of 22F WT RpoE. The highly structured N-terminal (underlined), while acidic (aspartic acid (D) and glutamic acid (E)) and basic residues (arginine (R), histidine (H) and lysine (K)) within the C-terminal are denoted by red and blue respectively.

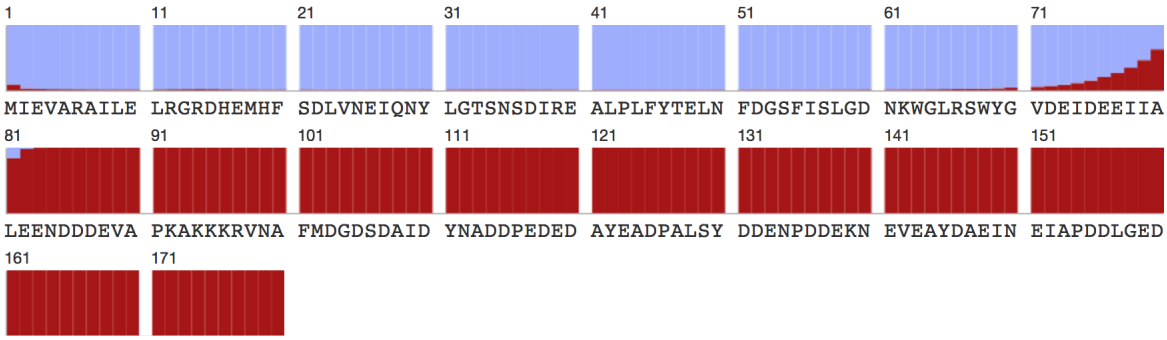



Figure 49: Order (blue) and disorder (red) within RpoE calculated through RaptorX program AUCpreD (Wang et al., 2016). The majority of the N-terminal domian is predicted to be highly ordered, while the the entire C-terminal domain is predicted to be entierly disordered.

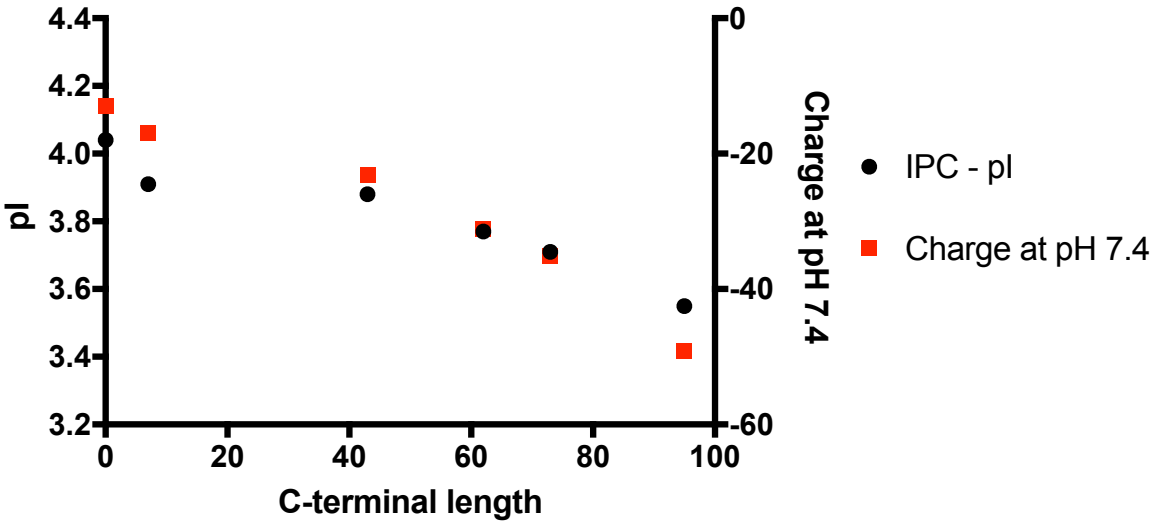



Figure 50: Predicted pI (pH at which the protein is charge neutral) and charge at ph 7.4 of 22F WT and SCV RpoE isoforms. Predictions were made using the IPC isoelectric point calculator (Kozlowski, 2016). The correlation between the length of the C-terminal domain of RpoE, and the pI of 22F WT, SCVs, and the N-terminal (C-terminal domain length = 0). More truncated isoforms of the RpoE C-terminal domain display increased pI, indicating that the protein is less acidic. We observed a high degree of correlation between the C-terminal domain and both PI, and charge (pH 7.4) with $R^2$ values for a linear regression of 0.916 and 0.943 respectively. The slopes of both pI ($p<0.027$) and charge (pH 7.4) ($p<0.0012$) were both significantly non zero.

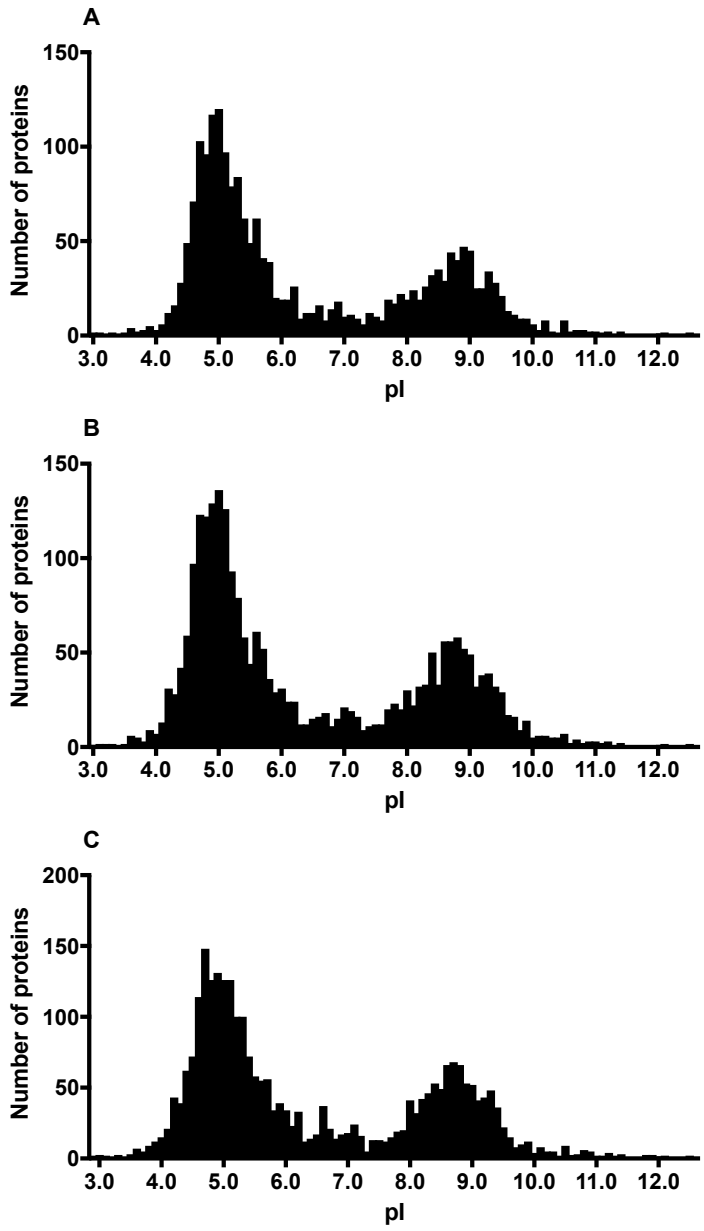



Figure 51: The pI distribution of proteins from *S. pneumoniae* reference proteomes. **A)** *S. pneumoniae* strain R6, UniProt reference proteome UP000000586, 2030 proteins **B)** *S. pneumoniae* strain GA41410, , UniProt reference proteome UP000003388 **C)** *S. pneumoniae* strain 2070335, UniProt reference proteome UP000002642, 2823 proteins. The frequency histogram was generated through IPC isoelectric point calculator (Kozlowski, 2016). For all of these proteomes RpoE is within the top 10 most acidic proteins.

### 6.3.3 Biofilm Formation Correlates with C-terminal Domain Length

The trend of increased pI in relation to shorter C-terminal domains is a trend that we have observed in numerous phenotypic traits. Biofilm formation in particular displays a strong correlation to the length of the C-terminal with greater truncated correlating with increased biofilm formation as shown in Figure 52. This may provide an indication towards the mechanism of action through which the variation for several traits in SCVs may occur.

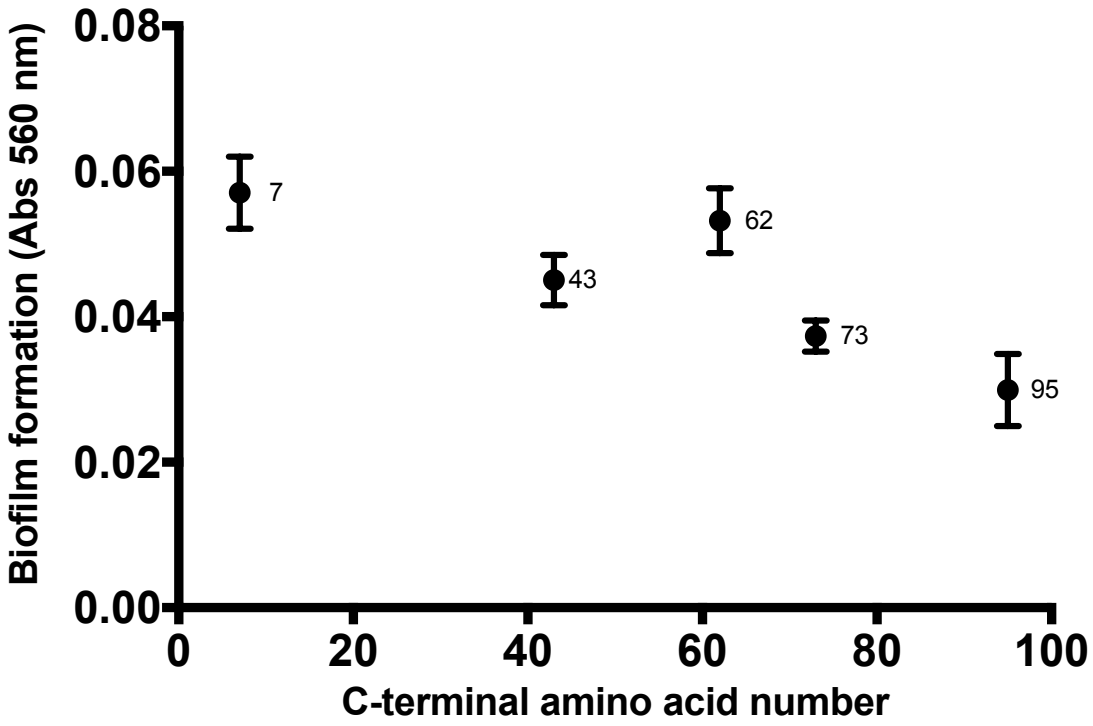



Figure 52: The correlation between the length of the C-terminal domain of RpoE, and mean biofilm crystal violet formation. More truncated isoforms of the RpoE C-terminal display a trend of increased biofilm formation.

### 6.3.4 RpoE C-terminal Domain Structure

In order to determine the potential functional changes caused by the truncations to the RpoE C-terminal domain we utilised I-TASSER protein structure modelling server to analyse the C-terminal domain for structural properties. I-TASSER utilises a mixture of homology modelling to known structures and *ab initio* structural predictions of regions without alignment which are produced through Monte Carlo simulations through the program LOMETS (Yang et al., 2015). The end result of this process is a set of models with C-scores. For model predictions, C-scores range between -5 and 2, with higher scores indicating a higher confidence of accurate protein folding with a C-score of -1.5 considered the lower cut off for accurate folding (Yang and Zhang, 2015).

We were unable to predict a C-terminal domain model for the 22F WT with a sufficiently high C-score to be considered a likely structure with the highest C-score of -1.87. Furthermore, we were unable to determine models of sufficient similarity to provide homology modelling of sufficient quality. The high degree of disorder within 22F WT RpoE C-terminal domain mirrors that was observed in *B. subtilis* and is probably responsible for the difficulties in structure prediction (Papouskova et al., 2013). An unstructured C-terminal would support its role in displacing RNA from the DNA-RNAP-RNA complex, which is likely to require flexibility (de Jong et al., 2017).

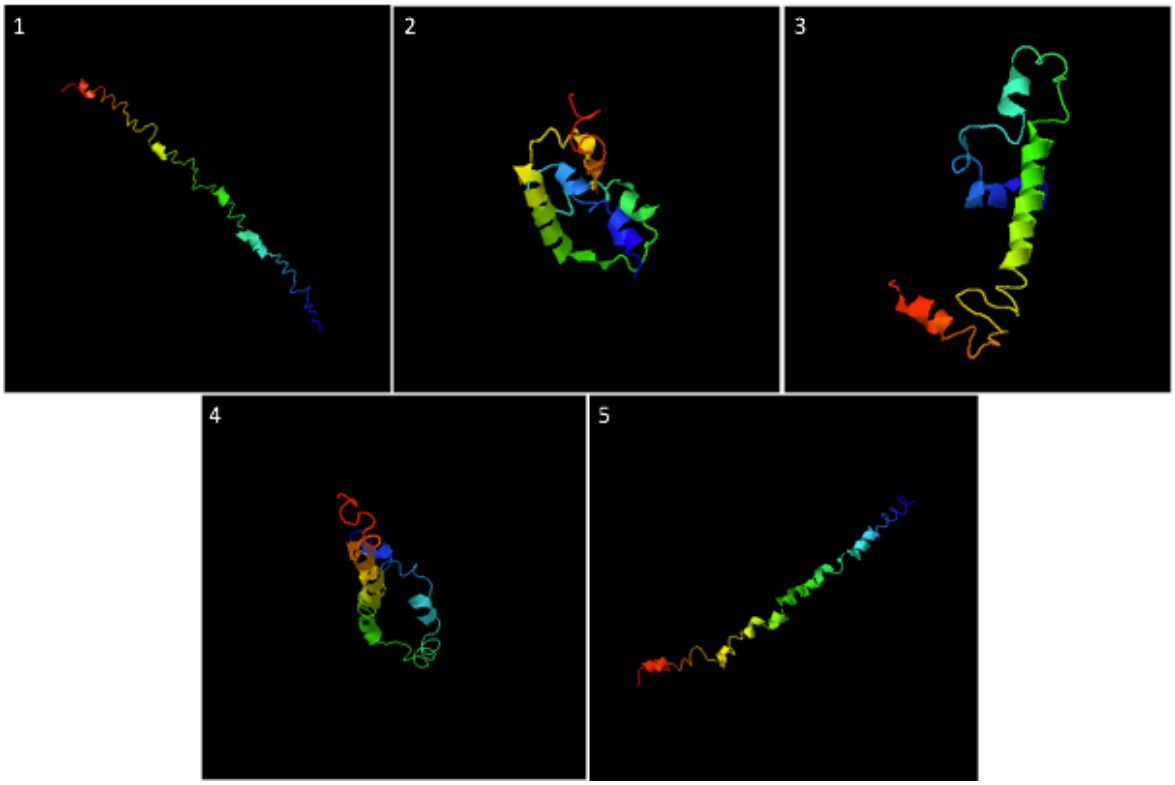



Figure 53: Potential structures for 22F WT RpoE C-terminal determined through I-TASSER structural prediction. Model C-scores **1)** -1.87. **2)** -2.93. **3)** -2.67. **4)** -5. **5)** -3.16. All structure predictions produced C-scores below the threshold of -1.5 indicating that the models produced are unlikely to reflect accurate folding (Yang and Zhang, 2015). This is further emphasised by the high degree of variation seen in the highest ranked I-TASSER protein structure models for the C-terminal.

### 6.3.5 RpoE N-terminal Domain Structure

We have observed that the *rpoE* KO differs phenotypically to both 22F WT, and the SCVs including SCV9D9 which possess a C-terminal domain of only 7 amino acids in length indicating that the N-terminal of RpoE also plays an important role. To determine the structure of RpoE we used the I-TASSER protein prediction server.

Initially the program LOMETS identifies likely homology through an array of 10 threading programs as shown in Figure 54. All of the threading programs revealed homology between 22F RpoE N-terminal structures and those found in a few closely related *B. subtillis* RpoE structures: 2krcA (Motackova et al., 2010), 2m4kA (Papouskova et al., 2013) and 4nc7A (Demo et al., 2014). The highest degree of similarity was to 2krcA. The N-terminal domain structure of 2krcA RpoE has been previously determined through NMR in *B. subtilis*.

The conserved N-terminal of 22F ST433 is predicted to display a high degree of similarity to *B. subtilis* N-terminal of RpoE PDB 2KRC. From the model produced we were able to predict functional properties for the 22F N-terminal domain, which may indicate its role and the differences observed between SCV9D9 and *rpoE* KO. We also observed from the structure produced, that a winged helix-turn-helix (wHTH) motif is likely to be present within the N-terminal domain. HTH motifs play important roles in DNA binding within many proteins and this may provide an indication to the role that RpoE plays within *S. pneumoniae* (Harrison, 1991).

| Rank / Threading Program | PDB ID | Identity 1 | Identity 2 | Coverage | Normalised Z-score |
|---|---|---|---|---|---|
| 1: MUSTER | 2krcA | 0.37 | 0.40 | 0.98 | 2.81 |
| 2: FFAS-3D | 2krcA | 0.40 | 0.38 | 0.92 | 1.74 |
| 3: SPARKS-X | 2krcA | 0.37 | 0.40 | 0.98 | 2.63 |
| 4: HHSEARCH2 | 2m4kA | 0.40 | 0.33 | 0.79 | 4.54 |
| 5: HHSEARCH1 | 2m4kA | 0.40 | 0.33 | 0.79 | 3.72 |
| 6: Neff-PPAS | 2krcA | 0.37 | 0.40 | 0.98 | 3.17 |
| 7: HHSEARCH | 2m4kA | 0.40 | 0.33 | 0.79 | 5.91 |
| 8: pGenThreader | 4nc7A | 0.40 | 0.33 | 0.79 | 1.72 |
| 9: wdPPAS | 2krcA | 0.37 | 0.40 | 0.98 | 2.71 |
| 10: PROSPECT2 | 2krcA | 0.37 | 0.40 | 0.98 | 1.49 |

Figure 54: Threading alignments of the 22F WT N-terminal identified through LOMETS, identifying the top 10 alignments generated by the indicated threading programs. All of the PDB ID refer to *B. subtillis* solved RpoE N-terminal structures 2krcA (Motackova et al., 2010), 2m4kA (Papouskova et al., 2013) and 4nc7A (Demo et al., 2014). Identity 1 is the percentage of sequence identity in the regions aligned between the query sequence and aligned PDB template. Identity 2 is the percentage of sequence identity of the whole template chains wih query sequence. Coverage represents the number of aligned residues divided by the query sequence length.The normalised Z-score is a measurement of how good alignments are, Z-scores of greater than 1 are considered good alignments.

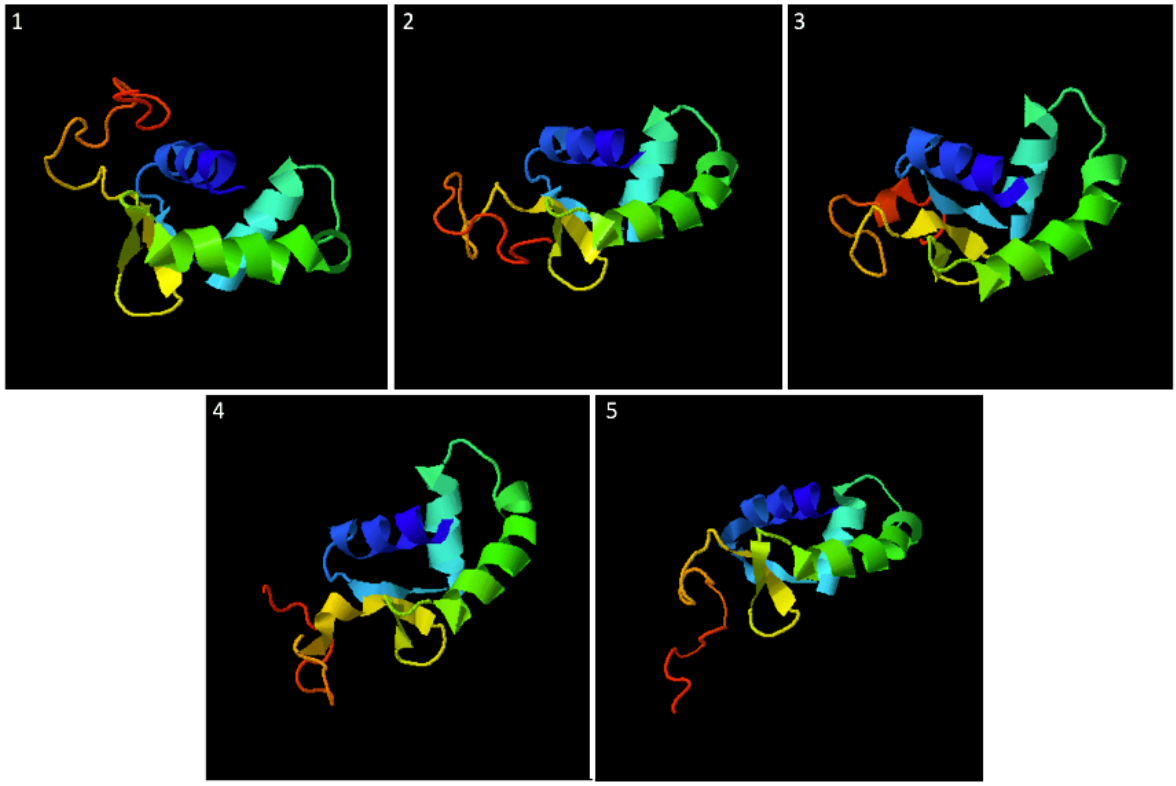



Figure 55: Potential structures for 22F WT RpoE N-terminal determined through I-TASSER structural prediction server. C-terminal structure predictions through I-TASSER all produced C-scores below the threshold of -1.5 indicating that these models are unlikely to reflect accurate folding (Yang and Zhang, 2015). This is further emphasised by the high degree of variation observed in the highest ranked I-TASSER protein structure models for the C-terminal C-scores are **1)** 1.02 **2)** -4.67 **3)** -4.92 **4)** -5 **5)** -5. Model 1 was chosen as the highest quality structure due to its high C-score of 1.02 and the low C-score of all other models generated indicating that it is likely to represent an accurate protein folding.

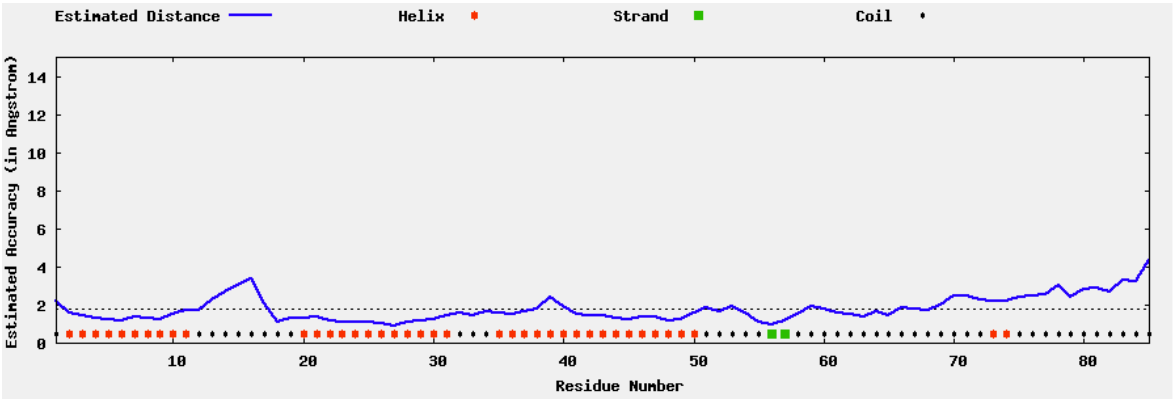



Figure 56: Local structure accuracy modelling of 22F WT RpoE N-terminal model 1. Modelling is achieved through ResQ as part of the I-TASSER protein prediction server suite. The estimated root mean square of the deviations = 1.8±1.5Å. The TM score for model 1 = 0.85±0.08 indicating that this model is likely to be correctly folded. We observe 3 large helix regions, 1 minor helix motif formed (helix regions denoted in red) of two amino acid residues and 1 small strand region.

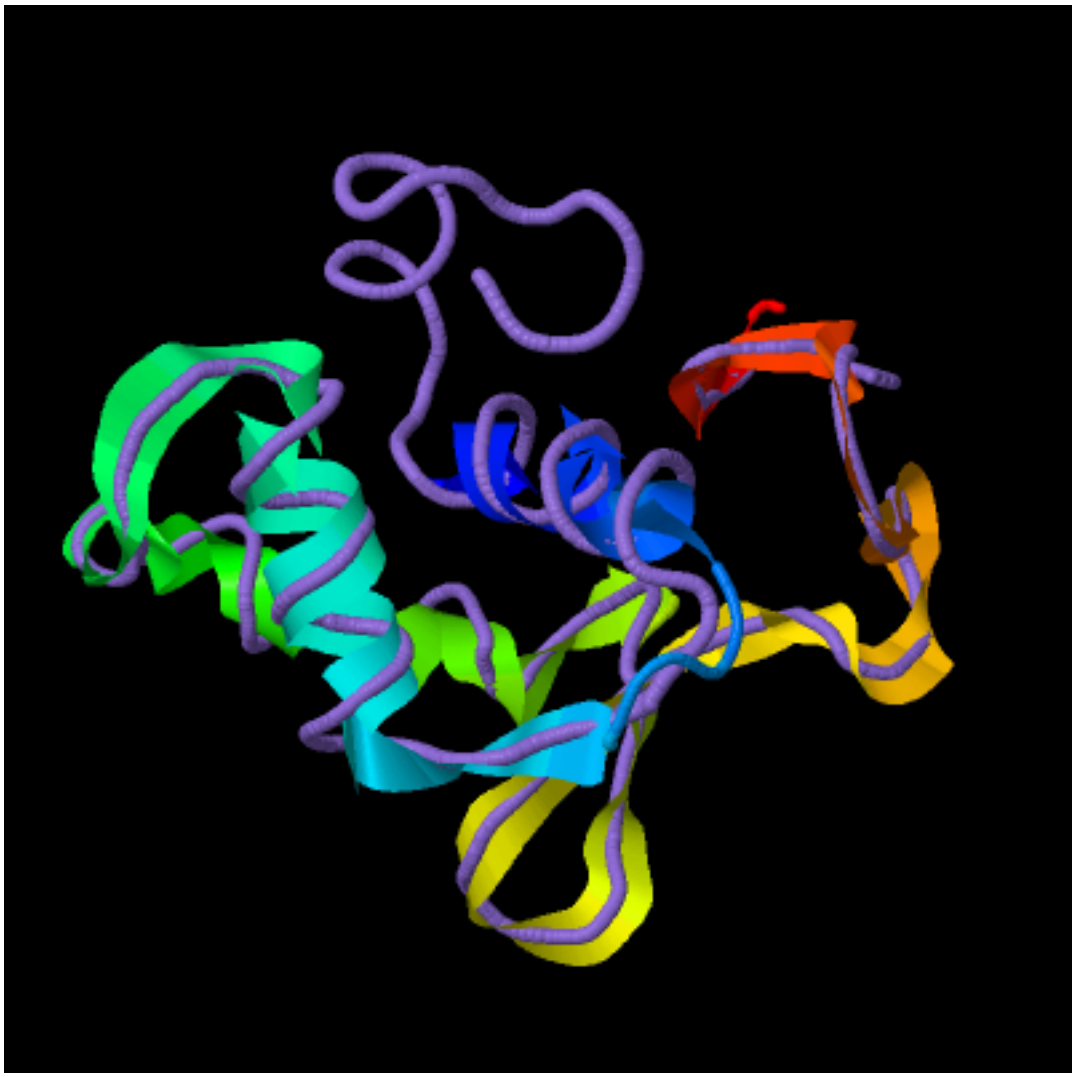


Figure 57: Homology modelling of the RpoE N-terminal domain protein structure of 22F WT and *B. subtilis* (PDB id: 2KRC). This domain is conserved across SCV, and 22F WT. The multi-coloured heat map represents the 22F WT N-terminal domain o RpoE, while the purple backbone represents the structure of RpoE determined through nuclear magnetic resonance (NMR) (Motackova et al., 2010). The 22F WT RpoE N-terminal displays a high degree of structural similarity to 2KRC. The TM score for this alignment = 0.953. RMSD[a] = 0.56. Proportion of structurally aligned residues = 0.977.

### 6.3.6 N-terminal Domain Predicted Functions and Binding Sites

In order to better determine the function of the conserved N-terminal domain we utilised I-TASSER and RaptorX servers to predict the potential binding sites and functions of the N-terminal domain model.

For domain functions C-score ranges from 0-1 where a higher score indicates a more reliable prediction of correct binding pocket folding. We predicted only a relatively low C-score values through I-TASSER for the DNA ligand binding prediction and so we further validated potential binding properties through use of the RaptorX Model-assisted protein binding site prediction server. The RaptorX server predictions support a DNA binding hypothesis for the N-terminal domain, with ligand binding predicted between the N-terminal and the nucleotides Adenosine, Guanine and Thymine; Cytosine was not predicted to be likely to bind to RpoE.

In addition to investigating potential binding sites we also investigated the possibility that the N-terminal domain of RpoE may exhibit enzymatic activity based on its structure. We observed that the N-terminal domain showed a reasonable degree of homology to multiple DNA isomerase, transferase and hydrolase enzymes, further supporting a DNA binding function (Table 7).

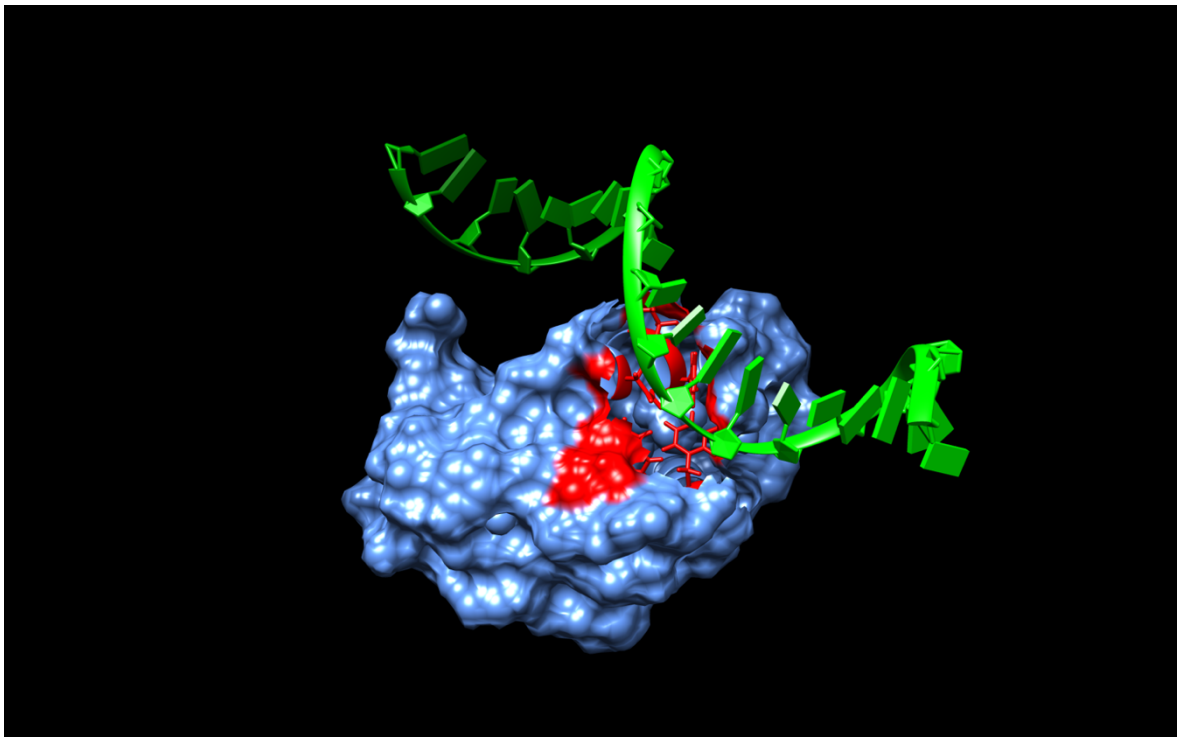


Figure 58: Prediction of N-terminal function through the use of COFACTOR and COACH applications applied to the 22F WT RpoE N-terminal domain structure produced through the I-TASSER pipeline. 22F WT RpoE N-terminal (blue) shows similarity to the protein Foxp2, with single stranded DNA binding domain (red) mediated through residues F20, Y46, T47, N50, S57, and W63 with a C-score of 0.33. (C-score ranges from 0-1 where a higher score indicates a more reliable prediction).

| Pocket | Multiplicity | Ligands | Binding residues |
|---|---|---|---|
| 1 | 99 | DA, DG, DT | M1 E40 A41 P43 L44 T47 E48 |
| 2 | 50 | DA, DG, DT | M1 L44 E48 E83 N84 D85 |
| 3 | 50 | DG, DT | F20 S21 Y46 T47 N50 D60 N61 W63 |
| 4 | 35 | DG, DT | T47 F51 |
| 5 | 20 | DT, EDO, MIX | N25 S36 R39 E40 L42 P43 |
| 6 | 20 | DG, DT, DC | T33 I38 R39 E40 A41 |

Table 7: 22F N-terminal ligand binding sites calculated through RaptorX Model-assisted protein binding site prediction server. Ligands abbreviations: DA = 2'-Deoxyadenosine-5'-monophosphate, DT = Thymine-5'-monophosphate, DG = Deoxyguanosine-5'-monophosphate, DC = Deoxycytidine-5'-monophosphate, EDO = 1,2-Ethanediol, MIX= Mitoxantrone. Multiplicity indicates the quality of the ligand binding pocket. A multiplicity of greater than 40 indicates that the folding, and structure of the pocket is likely to be correct. The P-value for this structure = $4.1 \times 10^{-4}$ indicating that the folding and structure for the N-terminal is highly likely to be correct. A high model quality was also confirmed through uGDT and GDT scores of 62, and 79 where scores of greater than 50 indicate a high-quality model. The results indicate that DA, DG and DT are likely to be bound at several sites on the N-terminal while it is unlikely that EDO, MIX or DC are bound due to the high likelihood of an incorrectly folded pocket.

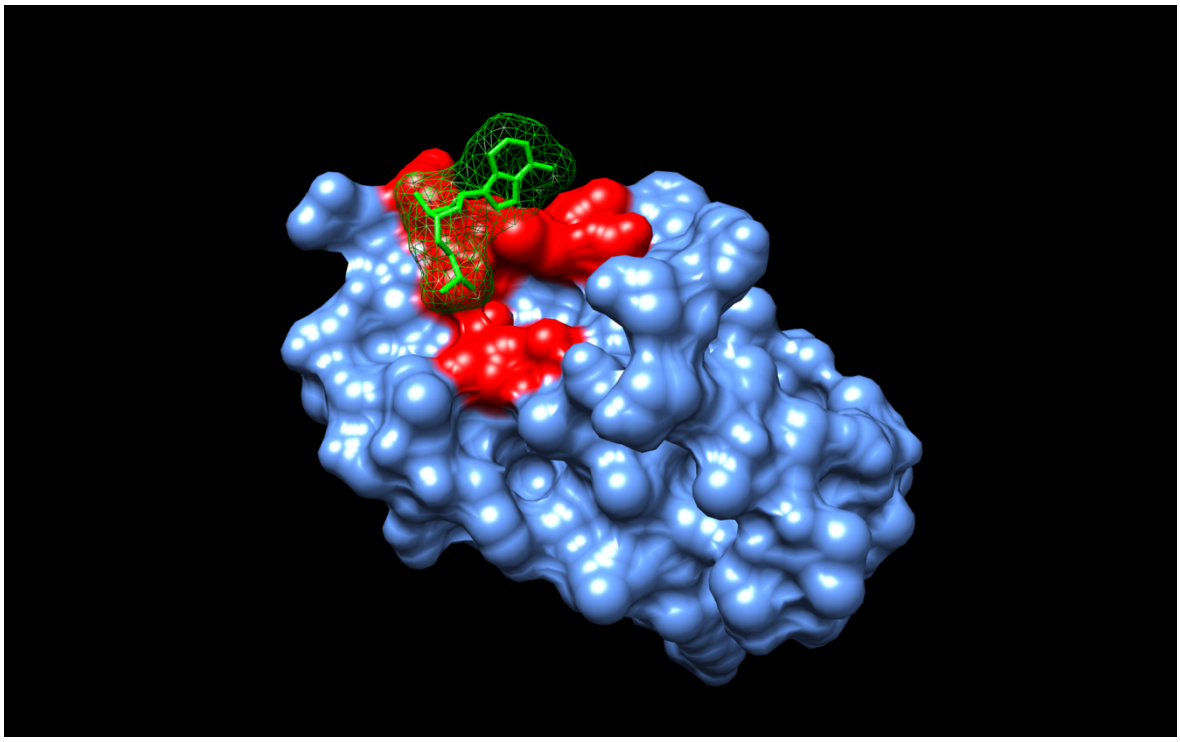


Figure 59: 2'-Deoxyadenosine-5'-monophosphate binding pocket prediction through the RaptorX-Binding Server. Binding between DA and the N-terminal domain is mediated through the residues: M1 E40 A41 P43 L44 T47 E48 with a multiplicity of 100.

| **Rank** | **TM score** | **PDB Hit** | **Coverage** | **Species** | **Enzyme / Activity classification** | **Reference** |
|---|---|---|---|---|---|---|
| 1 | 0.581 | 1fp1D | 0.894 | *Medicago sativa* | Chalcone O-methyltransferase / Transferase | (Zubieta, He, Dixon, & Noel, 2001) |
| 2 | 0.585 | 2wwyA | 0.824 | *Homo sapiens* | Hydrolase / DNA | (Pike et al., 2015) |
| 3 | 0.583 | 1fokA | 0.906 | *Planomicrobium okeanokoites* | Topoisomerase I / Isomerase | (Wah et al., 1997) |
| 4 | 0.569 | 3fofA | 0.918 | *Streptococcus pneumoniae* | Isomerase / DNA | (Laponogov et al., 2009) |
| 5 | 0.574 | 3k9fA | 0.894 | *Streptococcus pneumoniae* | Isomerase / DNA | (Laponogov et al., 2010) |

Table 8: N-terminal domain model 1 structral similarities to enzyme crystal structures determined through I-TASSER. The TM score represents the structral similarity between

the 22F WT N-terminal and the PDB protein. A TM score of greater than 0.5 indicates a liklihood that the structal homology is correct.

### 6.3.7 Inter-Domain Interactions

Due to our inability to predict a structure and therefore potential functions for the C-terminal domain, we sought to determine whether its predicted functions may be altered through interactions between the terminal domains. Interactions between the N-terminal and C-terminal domains were predicted through the RaptorX Contact prediction server which utilises a template free methodology to predict interactions within proteins (Wang et al., 2017). As a template free method, RaptorX is not biased by currently determined structures.

Contact maps and protein structures with potential contacts are visualised in Figure 60 and Figure 61 respectively. Predicted structures for SCV3D9, SCV5D3 and SCV9D9 indicate that it is unlikely that inter-domain interactions are occurring. However, 22F WT and SCV1D3 may have some level of inter-domain interactions. This suggests that inter-domain interactions are not responsible for the gradient of biofilm formation phenotypes observed.

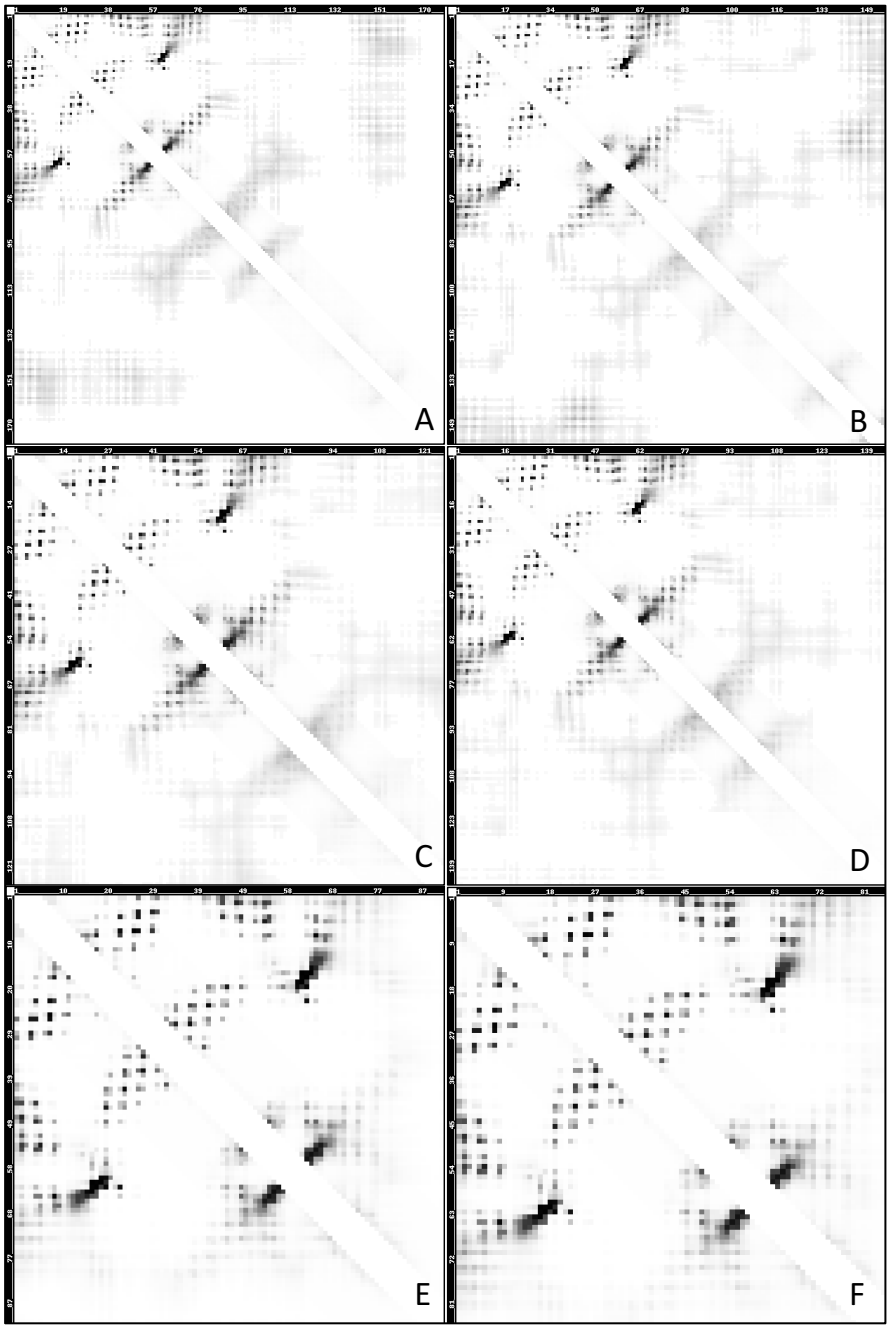



Figure 60: Contact maps for RpoE generated through RaptorX. **A)** 22F WT **B)** SCV1D3 **C)** SCV3D9 **D)** SCV5D3 **E)** SCV9D9 **F)** N-terminal domain. Darker spots indicate a higher liklihood of contact between residues. A high degree of contact is observed within the N-terminal domain. In 22F WT and SCV1D3 there is a low probablility observed for contact between the N-terminal and C-terminal domains indicated by the light grey areas in the top right, and bottom left portions of the contact map.

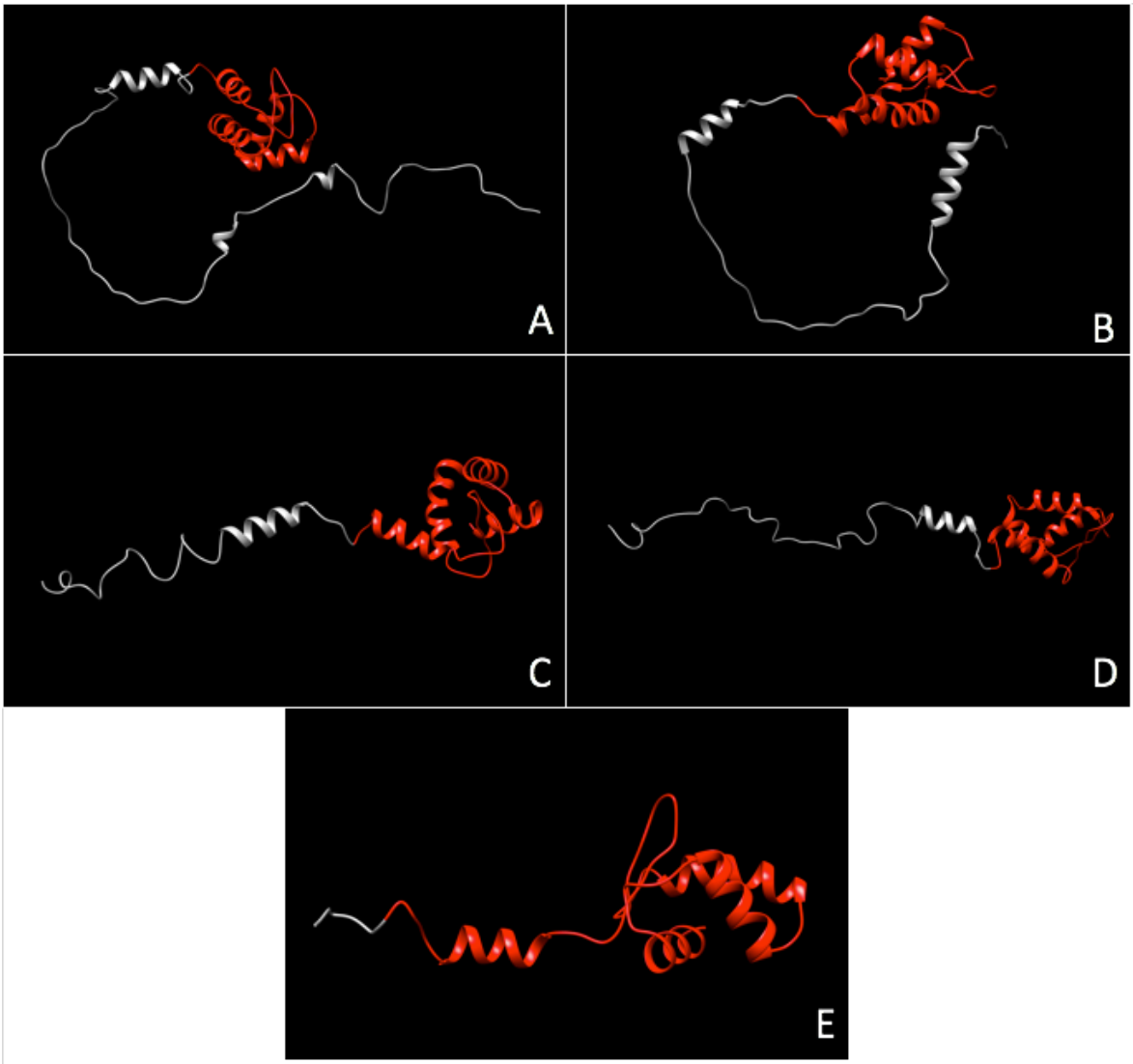



Figure 61: Protein structures for RpoE generated through RaptorX contact server. The N-terminal domain is coloured red while the C-terminal domain is coloured grey. **A)** 22F WT **B)** SCV1D3 **C)** SCV3D9 **D)** SCV5D3 **E)** SCV9D9. The areas of close proximity indicating contact between terminals is only observed for 22F WT and SCV1D3. Structures are determined *de novo* based only on the binding properties of the amino acids without the use of structral templates (Wang et al., 2017).

### 6.3.8 Homology Between Predicted Structures

The template free methodology of the RaptorX contact server utilises an ultra-deep learning model to predict protein structures without relying on sequence homology to currently know structures. In addition to providing information on intra-protein contacts this technique also predicts protein structures. To confirm that the protein structural models generated through the RaptorX contact prediction server and I-TASSER protein structure prediction servers were not predicting radically different structures, we calculated the homology between the models produced through the I-TASSERTM align server (Zhang and Skolnick, 2005).

We found a high degree of similarity between the models as shown in Figure 62 and Figure 63 indicating that the models produced are comparable. This is perhaps unsurprising as the structure of the N-terminal observed is a wHTH which is well defined and conserved structure. However, the orientation of the N-terminal is reversed although the helixes display a high degree of alignment.

Additionally, we found a lower degree of homology between the I-TASSER 22F N-terminal model and the RaptorX full RpoE model supporting the hypothesis that the C-terminal domain may be impacting the structure and potentially therefore the function of the N-terminal domain.

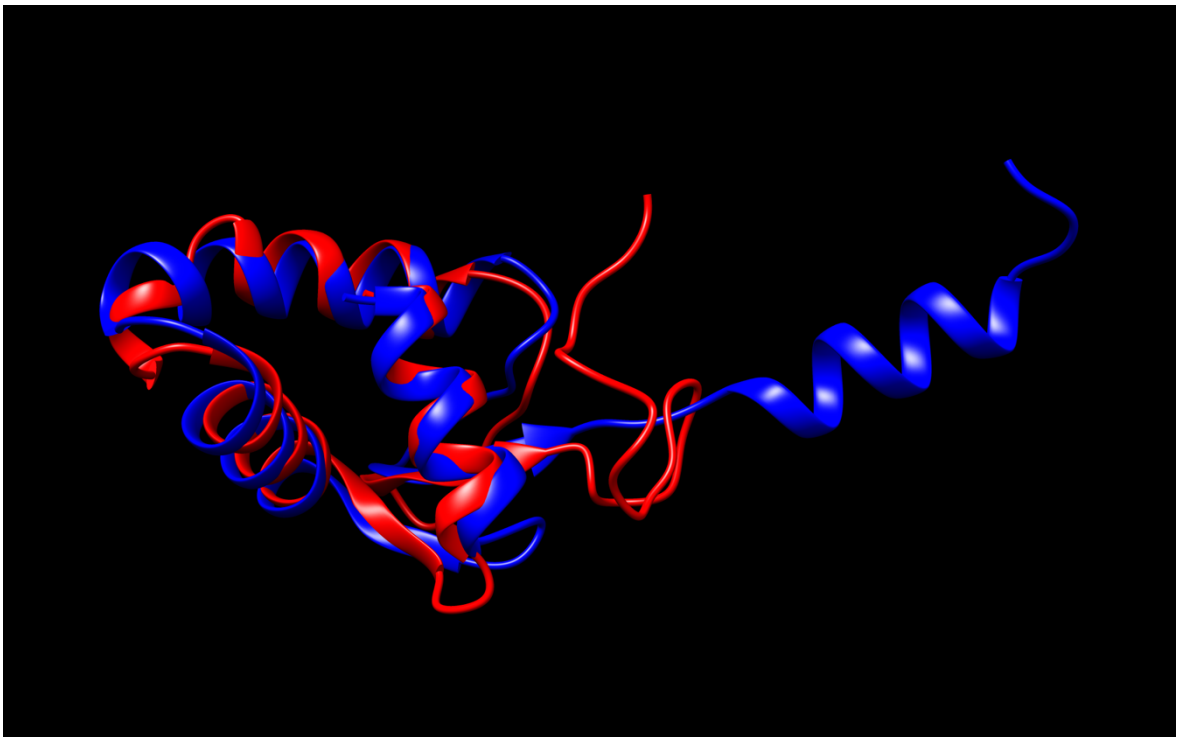


Figure 62: Structural homology between RaptorX Contact Server structure prediction for RpoE N-terminal domain structure prediction (Blue) and I-TASSER 22F N-terminal domain structure prediction (Red). The TM score between the two structures was calculated using the I-TASSER TM-align server. The TM-score for the alignment of the two models is 0.622 indicating that these structures are similar and that the different modelling techniques have not produced overly dissimilar structures with the main differences appearing within the tail of the domain.

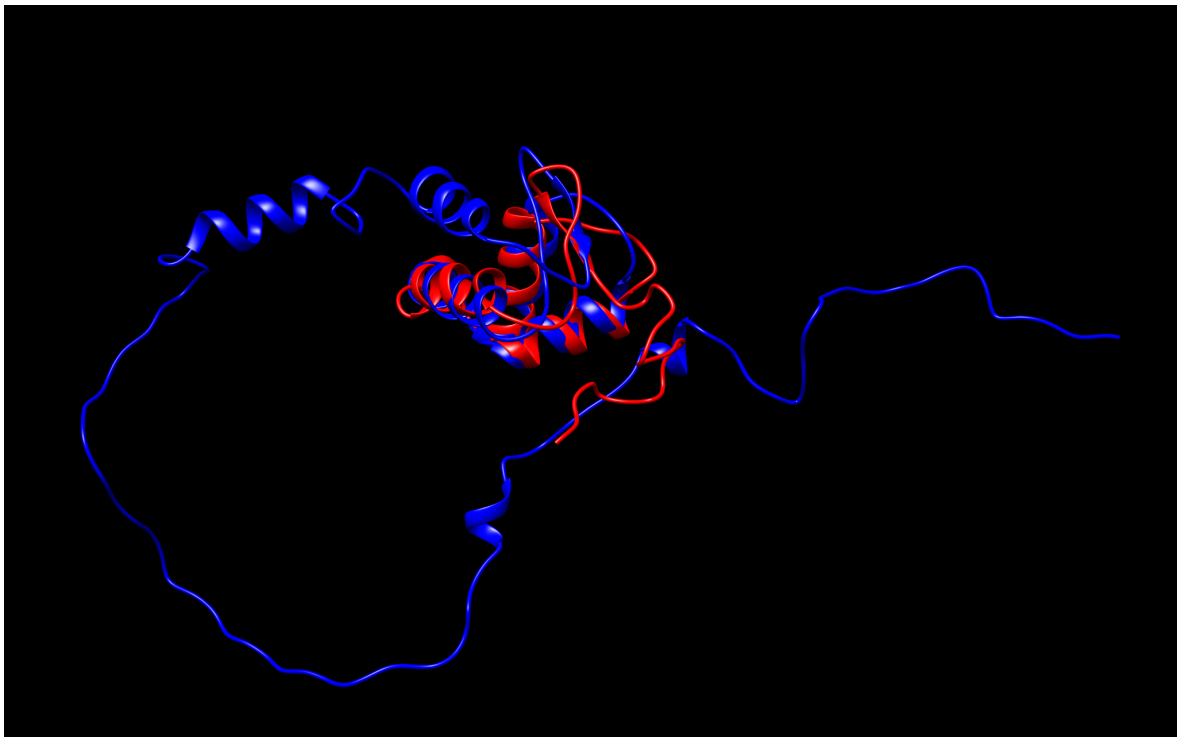


Figure 63: Structural homology between RaptorX Contact Server structure prediction for the full sequence length of RpoE (Blue) predictions and I-TASSER 22F N-terminal domain structure prediction (Red). The TM score between the two structures was calculated using the I-TASSER TM-align server. The TM-score for the two models normalised against the I-TASSER 22F N-terminal structure between the two models is 0.505 indicating that these structures are similar and that the different modelling techniques have not produced overly dissimilar structures. In the alignment, the helix turn helix motif is conserved for both structures, however they display opposite orientations of the C-terminal (Zhang and Skolnick, 2005).

## 6.4 Discussion

### 6.4.1 RpoE Structure

We have identified that 22F WT *S. pneumoniae* RpoE is formed of two domains. This two-domain structure is in line with that observed in other species such as *B. subtilis*. The 22F WT RpoE N-terminal also displays a high degree of homology to the NMR solved *B. subtilis* RpoE N-terminal structure (Motackova et al., 2010). The high C score for our N-terminal model 1 suggests with a high likelihood that the predicted structure is correctly folded.

We were unable to determine a structure for the C-terminal domain of 22F WT RpoE. This is unsurprising as in *B. subtilis* the C-terminal domain has been identified as being unstructured and flexible, matching our predictions that the C-terminal domain is highly disordered. We have also identified that both terminals are highly acidic, which matched observations in *B. subtilis* (López de Saro et al., 1995).

### 6.4.2 Shorter RpoE C-Terminals Are Less Charged, and Are Correlated with Increased Biofilm Formation

The C-terminal domain of 22F WT RpoE is predicted to be highly charged and repetitive. These predictions are in line with RpoE in *B. subtilis* (Papouskova et al., 2013). As the C-terminal of the SCVs become more truncated the pI increases, indicating that RpoE is becoming less charged as the acidic amino acids are lost.

While the N-terminal domain displays a high degree of conservation across multiple species, the length of the C-terminal domain is highly variable (Weiss and Shaw, 2015). A potential mechanism of action for the C-terminal domain length dependant activity may lie in its ability to displace RNA and DNA from RNAP (López de Saro et al., 1995; Prajapati et al., 2016a). It has been observed that the N-terminal domain alone does not have an impact upon RNA displacement, however the C-terminal domain is 100-fold more efficient in displacing RNA when bound to the N-terminal domain, indicating that the N-terminal domain may help to orientate the C-terminal for RNA displacement. The C-terminal domain alone was as efficient as the whole RpoE protein in displacing DNA, suggesting that this activity does not require orientation or the binding properties exhibited by the N-

terminal domain(López de Saro et al., 1995). Positioning of the C-terminal domain by the N-terminal domain to facilitate C-terminal domain interactions in the DNA binding cleft of RNAP is supported by protein cross linking studies carried out in *B. subtillis* (de Jong et al., 2017).

Loss of RNA and DNA displacement activity from RNAP has the potential to cause genes that do not have an endogenous promoter to have a reduced expression through reduced RNAP recycling. Therefore, SCVs with more truncated C-terminal domains may have reduced transcriptional cycling for terminator-less genes, resulting in differential expression dependant on C-terminal domain length.

The DNA displacement activity of the C-terminal domain might alter transcription in two ways. First, RNAP may stall in response to stresses during transcription or DNA template damage (Edenberg et al., 2014). SCVs with increasingly truncated C-terminal domains may be less able to displace a stalled RNAP from its DNA template. Alternatively, the C-terminal domain may compete with the target DNA template for RNAP binding. More truncated and less charged C-terminal domains may be less effective in competing for RNAP binding, resulting in the gradient of phenotypes observed. A potential indicator that this is not the case may be that DNA displacement is unlikely to be affected by the orientation of the C-terminal domain, which it might be assumed is important for RNAP binding (López de Saro et al., 1995). In favour of this hypothesis is the evidence that RpoE acts as a transcriptional regulator by binding to upstream sites (Prajapati et al., 2016a; Prajapati et al., 2016b).

Inter-domain interactions may provide an alternative putative mechanism for the gradient of phenotype observed across the 22F WT and SCVs. Shorter C-terminal domains may be less able to bind to distant sites on the N-terminal domain, and therefore would be less likely to cause any conformational or folding changes. However, for the gradient of phenotypic change to occur there would have to be a gradient of binding sites distributed across the N-terminal domain. We have only observed potential interactions between the two domains of RpoE for 22F WT and SCV1D3, which would not provide a gradient of effects that would fully explain the gradient of biofilm formation observed, although it may still it may be a contributory factor.

A further hypothesis may lie in a function of the C-terminal domain identified in *S. aureus* that linked the length of this domain to the stability of RpoE. A truncated C-terminal domain missing half of its length was observed at a lower concentration than the full RpoE and a N-terminal alone mutant was not detectable, suggesting that the C-terminal domain may play a role in stability of the N-terminal (Weiss et al., 2014). However, we observed that SCV9D9 also displays considerable differences in numerous phenotypic traits to the *rpoE* KO, despite possessing a C-terminal domain that is only 7 amino acid residues long, indicating that the N-terminal domain possess its own function.

### 6.4.3 RpoE N-terminal Domain Displays DNA Binding Homology

We predicted that the N-terminal domain I-TASSER model displays a high level of structural homology to DNA binding structure. The structure and DNA binding properties are in line with those observed through His-tagged analysis of RpoE within *B. subtilis*, which revealed that N-terminal forms a "DNA/RNA-binding three-helical bundle fold" (Papouskova et al., 2013). In our structural predictions, we observe a structure formed principally of three helixes with DNA binding homology. The three-helix structure observed is likely to be a HARE-HTH domain. The HARE-HTH domain is defined by its winged helix-turn-helix structure (wHTH) and was initially discovered within **H**B1, **A**SXL and **R**estriction **E**ndonuclease proteins (HARE). The HTH structure is prevalent across multiple kingdoms of life where it plays numerous roles including in DNA binding activities (Aravind and Iyer, 2012).

It has been argued that the N-terminal domain only utilises DNA/RNA binding fold as part of its biological activity and that it is not directly associated with nucleic acids (Motackova et al., 2010). However, we have observed that the N-terminal domain displays homology to a number of DNA processing enzymes (Table 8) and is predicted to have a nucleic acid binding site and in particular to bind A, G, and T. This may suggest that the N-terminal domain may play a role in nucleic acid binding in the RNAP complex. Furthermore, the bases identified as having a high likelihood of binding exclude cytosine. This is in line with the binding behaviour of RpoE observed in *B. subtilis* where binding was observed only to adenosine and thymine rich sequences, and binding was abolished when these

sequences were changed to cytosine or cytosine and guanine rich sequences (Prajapati et al., 2016a).

It is likely that the predicted DNA/nucleotide binding activity is involved in the role of RpoE as a transcriptional regulator. *B. subtilis* RpoE has been demonstrated to display a strong affinity for promoters at upstream AA and AT sequences. Binding of RpoE at these sites has been demonstrated to alter transcription yields of genes at these promoters, with binding at sites of -40 increasing transcription (Prajapati et al., 2016a) while binding to sites between -35 and -10 represses transcription (Chen and Helmann, 1997; Prajapati et al., 2016b). Binding to AA and AT sequences is in line with our predictions of ligand binding by RpoE with further confirmation through observations that replacement of AA sequences with CC at -40 abolished RpoE mediated transcription increases (Prajapati et al., 2016a). These observations match our RaptorX ligand binding predictions that binding pockets are likely only for A, G and T nucleotides and therefore the N-terminal is likely to be unable to bind to CC sequences and accordingly unable to act as a transcriptional regulator at such sites. The *rpoE* KO does not possess the N-terminal domain and therefore the mutant will have lost the regulatory properties of N-terminal-promoter binding. This dysregulation is likely to contribute towards the phenotypic differences that we have observed between SCVs and *rpoE* KO including the loss of chain formation.

In addition to its DNA binding role, the N-terminal domain is also predicted to have a high degree of homology to several enzymatic functions. The loss of the N-terminal DNA binding site, and putative transferase/topoisomerase groups may lead to less effective processing of DNA into the RNAP complex. Transferases, and topoisomerases play an important role in unwinding of DNA.

Such an enzymatic function mechanism may impact open complex formation. In *B. subtilis*, RpoE has been shown to affect the binding of RNAP to some rRNA promoters through destabilisation of open complexes. RpoE regulates RNAP sensitivity through interference between RNAP, and initiating nucleotide triphosphates (iNTP), which in Gram-positive bacteria this is thought to be principally GTP (Rabatinova et al., 2013).

However, not all promoters are sensitive to iNTP. iNTP-insensitive promoters are thought to be key during rapid changes in gene expression, such as during a change in nutrient availability (Krasny and Gourse, 2004; Lopez et al., 1981; Murray et al., 2003; Ochi et al., 1982). If this were the key area of functionality for RpoE in *S. pneumoniae* then this contrasts with our observations of a loss of adaptability in *rpoE* KO.

### 6.4.4 Phenotype Variation Within SCVs and *rpoE* KO Are Likely To Be Mediated Through At Least Two Independent, Domain Specific Mechanisms

We have observed that greater truncations of the C-terminal domain correlate with increased biofilm formation. This suggests that the C-terminal domain provides a continuum of regulatory effects upon this phenotype dependant on C-terminal domain length. We also observed that the C-terminal domain is highly disordered, in line with other groups that were unable to determine NMR structures for this domain (Motackova et al., 2010).

The mechanism for a continuum of phenotypes based upon the C-terminal domain is likely to be involved in its role in releasing RNA from the DNA-RNAP-RNA complex, thereby aiding RNAP recycling (López de Saro et al., 1995). Our observation that this terminal is highly charged supports a hypothesis of the C-terminal domain acting as a nucleic acid mimic (López de Saro et al., 1995; Papouskova et al., 2013); shorter C-terminals may have a diminished capacity to carry out RNA displacement. This hypothesis is supported by the observations in *B. subtilis* that the RpoE N-terminal may bind above the RNAP β' subunit in a manner that would orientate the C-terminal domain into the DNA binding channel (de Jong et al., 2017). From this position, we expect that the C-terminal domain plays a role in releasing RNA from the RNAP-DNA-RNA complex, thus increasing the rate of transcription for genes without terminators. In this model, shorter C-terminal domains would be less able to interfere with binding within the RNAP DNA binding channel and thereby transcription is reduced for terminator less genes in correlation with the length of the C-terminal domain producing the gradient of biofilm phenotypes.

However, when *rpoE* is knocked-out, we did not observe increased or similar biofilm formation to SCV9D9, as might be expected if the C-terminal domain mediated nucleic

acid mimicry were the sole function of RpoE. Compared to SCV9D9, we observed diminished biofilm formation by *rpoE* KO accompanied by diminished chain formation within biofilm and a loss of adaptability. SCV9D9 differs from *rpoE* KO as it possesses a C-terminal of only 7 amino acids and an N-terminal domain. Given the small C-terminal domain differences and the dysregulation observed in not only *rpoE* KO, but also in *S. mutans* (Xue et al., 2010), this suggests that the N-terminal comain possesses a function independent of the C-terminal domain.

Through molecular modelling with I-TASSER and RaptorX, we predicted that the N-terminal domain was likely to possess single stranded DNA (ssDNA) binding characteristics in particular to the nucleotides adenosine, thymine and guanine although not cytosine. The ssDNA binding is unsurprising due to the winged helix turn helix motif observed and is in line with observations in *B. subtilis* and *S. mutans*. The low likelihood of cytosine binding is supported by observations that replacement of AA regions with CC or GC in upstream promoter regions inhibited RpoE function in *B. subtilis* (Prajapati et al., 2016b). Binding to the upstream promoter sites may be linked to open complex formation inhibition by RpoE through steric clash with the σ subunit of RNAP. The loss of the N-terminal domain in the *rpoE* KO removes this transcriptional regulation and is likely responsible for the dysregulation that we and other groups have observed for *rpoE* KO.

These observations may also suggest that RpoE acts in a similar manner to MerR-like proteins which require both terminals to function. Similar to RpoE, MerR proteins are transcriptional regulators formed of a conserved HTH N-terminal domain attached to C-terminal effectors (Brown et al., 2003). Further similarities between RpoE and MerR proteins are observed in MerR like protein's capabilities in binding to -35 and -10 promoter regions, although this binding is mediated through cysteine residues rather than the adenosine that we observe for RpoE. MerR proteins facilitate Promoter-RNAP binding at these sites through crosslinking and binding of coactivator molecules (Heldwein and Brennan, 2001). There is limited evidence through crystallisation of RpoE that it may form dimers in this manner, with nickel ions acting as the co-factor (Demo et al., 2014). However, the nickel binding and dimer formation observed may be due to the experimental parameters used and has not been observed in NMR data. While currently inconclusive,

future work would benefit from investigation of potential MerR-like protein activities for RpoE which may further elucidate the function of this protein.

Typically, investigations into *rpoE* and other genes are based upon observations of functional changes following a full knock-out of the gene. In clinical samples and other 'real world' settings, full removal of a gene is unlikely to occur. This work highlights that even mutations which result in large deletions cannot be assumed to cause a loss of function.

### 6.4.5 Limitations and Future Work

While our modelling and predictions indicate a high likelihood that the N-terminal has a highly similar structure to that of *B. subtilis* it would be beneficial to confirm the structure through NMR based methods (Motackova et al., 2010). This further confirmation, should it complement our predictions, would lend additional confidence to our arguments for the role of the N-terminal domain within *S. pneumoniae*.

In order to further explore the genes and promoters that are impacted by mutations to, and knock-outs of, *rpoE* we could utilise RNA-Seq. RNA-Seq would allow us to evaluate the differences in transcription between the different isoforms of RpoE in SCVs and *rpoE* KO. Should differential production of RNA occur between RpoE isoforms and *rpoE* KO then we may be able to identify regions upstream of genes where RpoE binding may occur. Mutations to these sites which change double adenosine sites to cytosine may then allow for confirmation that RpoE is binding to these sites in *S. pneumoniae* (Prajapati et al., 2016a; Prajapati et al., 2016b).

The C-terminal domain's RNAP recycling function may be validated by investigating expression of genes that requires a high degree of recycling. Differential expression of such a gene by different SCV isoforms would suggest that the C-terminals domai length impacts its capability to release RNA from the DNA-RNAP-RNA complex for terminator-less genes.

Identification of terminator-less genes may be achieved through adaptation of current terminator identification algorithms. Methods developed for *E. coli* and *B. subtilis* may be

suitable as a starting point and have been found to have been found to predict Rho-independent terminators with a 95% success rate in *S. pneumoniae* strain R6 (de Hoon et al., 2005). Correlation between transcription, proteomics and identified terminator-less genes would further validate the mechanism of action which we propose for *rpoE*.

## 6.5 Chapter 6 Key Findings

- All mutations to *rpoE* within SCVs occur within the C-terminal domain of RpoE.
- The mutations observed occur at different sites within RpoE resulting in a range of RpoE isoforms between SCVs
- Larger truncations of the C-terminal are associated in greater phenotypic changes in SCVs from 22F WT. Increased biofilm formation is associated with shorter C-terminal domains.
- N-terminal domain is highly likely to have DNA binding properties and to regulate transcription by binding to upstream promoter sites through Adenosine, Thymine and Guanine residues.
- The lack of an N-terminal is likely to be responsible for the differences observed between *rpoE* KO and SCV9D9.

# Chapter 7: Conclusion

*S. pneumoniae* is a globally important pathogen that exerts a considerable burden of disease on human populations and is responsible for hundreds of thousands of deaths each year, particularly of young children (Wahl et al., 2018) and the elderly (Drijkoningen and Rohde, 2014).

*S. pneumoniae* biofilms are associated with a range of pathologies including otitis media, chronic rhinosinusitis and pneumonia (Bakaletz, 2007; Chao et al., 2014; Sanderson et al., 2006; Yadav et al., 2012). Biofilms are surface-adhered communities of microbes which provide protection against environmental conditions, the human immune system and display increased tolerance to antibiotic treatment (Fux et al., 2005; Koo et al., 2017; Stewart and Costerton, 2001). *S. pneumoniae* 22F has been observed to rapidly produce SCVs during biofilm growth and that these SCVs display greater biofilm formation than the wild type. In a display of parallel evolution, all of the SCVs investigated possessed mutations to the RNA polymerase δ subunit gene *rpoE* (Churton et al., 2016).

This study set out to further investigate the *S. pneumoniae* isolate 22F ST433 and the SCVs produced by this strain during biofilm formation (Churton et al., 2016). All of the SCVs investigated contained mutations to the gene *rpoE*; mutations to *rpoE* ranged from single nucleotide polymorphisms to large scale deletions removing 264 base pair from the gene. Through investigation of 22F, its SCVs and *rpoE*, we aimed to advance our understanding of parallel evolution during biofilm formation by this pathogen and to explore the mechanisms through which these mutations exert their effects.

To achieve these aims, we produced a *rpoE* gene knock-outs from 22F ST433. The knock-outs displayed the characteristic small colony phenotype observed in the SCVs supporting the hypothesis that the SCV phenotype is due to the mutations observed within *rpoE*. However, further phenotyping of 22F WT, SCVs and *rpoE* KO across a broader range of phenotypic characteristics revealed a range of similarities and differences between each of the groups.

We observed phenotypic differences across planktonic growth, biofilm formation, sugar metabolism and virulence. For biofilm formation we observed that SCVs are not a homogenous population, but display a continuum of biofilm forming abilities; although all displayed greater biofilm formation than the 22F WT. This continuum of phenotypes was also observed in the metabolism of many sugars. However, for other traits such as planktonic growth, virulence and colony diameter size, we did not observe a divergence of phenotypes between SCVs.

Similarly, alterations to the phenotype of *rpoE* KO were not displayed in a uniform manner. During assays such as biofilm formation, colony diameter and virulence *rpoE* KO, displayed a phenotype that was in-line with that of the SCVs. However, for phenotypic traits such as planktonic growth, biofilm morphology and colony diameter, we observed that *rpoE* KO displays phenotypic characteristics significantly different to both SCVs and 22F WT.

Through protein structure modelling and predictions, we predicted that it is likely the continuum of phenotypic variations observed are a result of the differing RpoE structures within SCVs and the complete absence of RpoE in *rpoE* KO. Structural analysis of RpoE amino acid sequences revealed that *S. pneumoniae* 22F ST433 RpoE was formed of two domains, a N-terminal domain and a C-terminal domain. All the mutations occurred within the C-terminal domain causing truncations of this domain and a different isoform of RpoE to be produced within each SCV investigated within this study.

The length of the C-terminal domain is correlates with the level of biofilm formation of each of the isolates. This observation is supported by observations of *B. subtilis* that RpoE is likely to bind to the β subunit of RNAP in such a way that RpoE is orientated to enable the C-terminal domain to interact with the RNAP-DNA binding cleft (de Jong et al., 2017). From this position the C-terminal domain is able to facilitate the release of RNA from the DNA-RNAP-RNA complex enabling increased transcription.

The differences between SCVs and *rpoE* KO are likely to be due the activity of the N-terminal domain which is conserved within SCVs but absent in *rpoE* KO. This is supported by the observation that SCV9D9 possesses a C-terminal domain of only 7 amino acids in

length and is therefore unlikely to be responsible itself for the striking differences we observe between these two strains. Through the use of multiple protein structure prediction methods, we predicted that the N-terminal domain forms a wHTH motif. wHTH are known to act as a DNA-protein binding domain often involved in transcription (Teichmann et al., 2012). The role of the N-terminal domain as a DNA binding factor is supported by our predictions of homology between the N-terminal domain and other proteins with DNA binding properties. Furthermore, our protein binding predictions support this function with observations of a likely single strand DNA binding site, which is likely to be capable of binding the adenosine, thymine and guanine bases but not cytosine. These observations fit with the additional functionality of RpoE as a transcription inhibitor (Prajapati et al., 2016a; Prajapati et al., 2016b). In the proposed model, RpoE binds to regions 10-40 bases upstream of genes and thereby prevents RNAP binding through steric clash. Our predictions match the published literature in that replacement of adenosine and thymine by cytosine at RpoE binding sites in these regions diminishes its effects (Prajapati et al., 2016a; Prajapati et al., 2016b).

From these findings we conclude that RpoE possesses at least two independent domain-specific functions. The dual functions of this protein explain the phenotypic variation observed between SCVs which are mediated through the C-terminal domain and the large phenotypic changes of the *rpoE* KO mediated through the N-terminal domain functions.

Our observations suggest that evolutionary pressure on *rpoE* produces the SCV mutants that are able to better metabolise a wider range sugars, display increased biofilm formation and grow to higher optical densities than wild-type 22F WT. The increased fitness in these key traits provides the strong selection factors required to facilitate the parallel evolution observed. However, it has also been observed that the SCVs only form approximately 1% of biofilm populations following 9 days of growth (Churton et al., 2016). If SCV populations possess increased fitness in these key areas then this raises the question as to why they do not dominate the biofilm, and it is therefore likely that SCVs may be adapted to a niche biofilm environment when in competition with 22F WT.

This work highlights that RpoE is an important protein with a global impact within *S. pneumoniae*. While questions remain as to the genes regulated by RpoE it is clear that

numerous clinically relevant traits are affected. Further investigation into the *in vivo* role of this protein will be required to validate its potential role as a target for medical intervention, however the striking parallel evolution of increased biofilm formation characteristics by SCVs combined with the alterations to virulence and *in vivo* phenotypes indicate that RpoE might provide a valuable target for the future interventions to combat *S. pneumoniae*.

## 7.1 Limitations and Future Work

### 7.1.1 Relevance of *rpoE* mutations to other serotypes

Two key unresolved questions for this project are whether the mutations to *rpoE* might cause similar effects in other strains and serotypes and the extent to which *rpoE* mutations are an artefact of the artificial growth conditions under which they occurred. It may be the case that the *rpoE* mutations observed are a unique response of 22F ST433 to the artificial environment and growth conditions rather than biofilm associated evolution with clinical relevance.

To address broader strain and serotype relevance, next steps would include creation of *rpoE* mutants in well characterised *S. pneumoniae* laboratory strains such as D39 and TIGR4. These mutations would cause truncations of the *rpoE* C-terminal domain. This would allow for validation of multi-strain relevance of the *rpoE* mutations in readily transformable strains. Should this be achieved further work would include investigation of whether *rpoE* mutants could be produced in relevant clinical strains.

Relevance of *rpoE* mutations across strains and serotypes may also be achieved through investigation of carriage study data sets through bioinformatic approaches. We are currently working with collaborators at the University Hospital Southampton to review their *S. pneumoniae* carriage data for *rpoE* mutations. Similarly, we are also working to analyse the Maela refugee camp carriage data sets, where we understand that there may be mutations to *rpoE* in 22F and potentially other strains. (Turner et al., 2012). Should we confirm the presence of these mutations, this may indicate greater relevance of *rpoE* mutations beyond 22F. Next steps would be to carry out phenotypic analysis on available isolates to investigate associations between phenotypic characteristics and *rpoE* mutations.

### 7.1.2 Improved *In Vitro* and *In Vivo* Biofilm Models

The environment within microtitre plates, such as those used for the *in vitro* biofilm assays, is considerably different to that of the human nasopharynx and is therefore likely to impact the phenotypic assays carried out, including biofilm formation. Future work would involve investigation of more representative *in vitro* models for biofilm formation. Use of respiratory cells as the substrate upon which biofilms form would be beneficial due to the numerous host factors provided by human respiratory tissue which are known to influence biofilm formation (Marks et al., 2012a; Moscoso et al., 2006). A potential model would be the use fixed H292 cells. This model has been observed to support *S. pneumoniae* biofilms which display a high degree of similarity to *in vivo* biofilms (Marks et al., 2012a). Biofilm quantification could also be improved through the use of confocal laser scanning microscopy (CLSM) rather than the crystal violet biomass quantification. Through the use of appropriate stains, we would be able to quantify the live and dead cells in biofilms and determine where their distribution. CLSM would also enable investigation of biofilm topology through software packages such as COMSTAT2 (http://www.comstat.dk) and IMARIS (https://imaris.oxinst.com).

In addition to improving *in vitro* models, this project would benefit from further use of *in vivo* models. Initially, this would include greater use of the *G. mellonella* animal model with the μ-CT methods that we have developed. The striking differences between the biofilm formation of strains observed *in vivo* requires replication of results to ensure that our observations are statistically significant. This model has the potential to facilitate improved understanding of *S. pneumoniae* biofilm formation and virulence.

However, murine models would provide a more representative model of *S. pneumoniae* infection than *G. mellonella*. Virulence could be modelled in animal models such as chinchilla (Reid et al., 2009), rat (Yadav et al., 2012; Yadav et al., 2017) or mouse (Chiavolini et al., 2008) models which are more suited to longer term carriage investigations. Due to the variations in immune response use in mouse models, ideally we would use both of the well characterised inbred C57BL/6 and BALB/c mice. The combination of these two mouse models ensures that both Th1 and Th2 immune backgrounds are covered (Kadioglu and Andrew, 2005; Preston et al., 2004).

### 7.1.1 Proteomic Analysis of Strains

The phenotypic analysis we have carried out would benefit from further investigation through proteomic approaches. Proteomic analysis of the strains investigated in this PhD would enable quantification of the changes in expression of proteins due to *rpoE* mutations. These approaches would both provide greater understanding of the mechanisms through which the phenotypic changes observed occur and may also provide further indications upon how *rpoE* regulates expression in *S.pneumoniae*.

Suitable methods would include iTRAQ proteomic profiling which has successfully been used to discern phenotypic changes during biofilm formation by *S. pneumoniae* serotype 14 (Allan et al., 2014). Through this approach we may be able to determine the metabolic proteins involved in the increased sugar metabolism that we observed. Furthermore, the iTRAQ approach may be used to identify changes in immunogenic protein expression by SCVs. Should further investigation determine that *rpoE* mutations are clinically relevant across multiple strains this may facilitate vaccine development against *rpoE* mutant SCVs (Allan et al., 2014).

## 7.2 Discussion Key Points

- In this PhD, a clinical isolate of *S. pneumoniae* 22F and its SCVs generated through biofilm growth were studied.

- All SCVs present with mutations to the gene *rpoE* and increased biofilm formation.

- To investigate these potentially clinically relevant changes we produced a *rpoE* knock-out in strain 22F.

- Phenotypic differences were observed between 22F WT, SCVs and the *rpoE* knock-out. We also observed a gradient of phenotypes for SCVs, which may be linked to the length of the C-terminal domain.

- The differences observed between SCVs and *rpoE* KO are likely due to the loss of the N-terminal domain of RpoE in *rpoE* KO, which was conserved in SCVs.

- This work highlights the need for phenotypic evaluation of isolates as even a loss of almost half of *rpoE* such as in SCV9D9 does not result in a loss of function for *rpoE*.

- The striking phenotypic changes observed both *in vitro* and *in vivo* indicate that RpoE may be a promising target for future drug or vaccine development.

# References


Allan, R.N., P. Skipp, J. Jefferies, S.C. Clarke, S.N. Faust, L. Hall-Stoodley, and J. Webb. 2014. Pronounced metabolic changes in adaptation to biofilm growth by *Streptococcus pneumoniae*. *PLoS One* 9:e107015.

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

Allegrucci, M., and K. Sauer. 2008. Formation of *Streptococcus pneumoniae* non-phase-variable colony variants is due to increased mutation frequency present under biofilm growth conditions. *J. Bacteriol.* 190:6330-6339.

Alonso De Velasco, E., A.F. Verheul, J. Verhoef, and H. Snippe. 1995. *Streptococcus pneumoniae*: virulence factors, pathogenesis, and vaccines. *Microbiological Reviews* 59:591-603.

Andam, C.P., and W.P. Hanage. 2015. Mechanisms of genome evolution of *Streptococcus*. *Infect Genet Evol* 33:334-342.

Anderl, J.N., M.J. Franklin, and P.S. Stewart. 2000. Role of antibiotic penetration limitation in *Klebsiella pneumoniae* biofilm resistance to ampicillin and ciprofloxacin. *Antimicrob. Agents Chemother.* 44:1818-1824.

Anderson, R., and C. Feldman. 2017. Pneumolysin as a potential therapeutic target in severe pneumococcal disease. *J. Infect.* 74:527-544.

Andrade, A.L., C.M. Franco, J. Lamaro-Cardoso, M.C. Andre, L.L. Oliveira, A. Kipnis, C.G. Rocha, J.G. Andrade, S.L. Alves, I.H. Park, M.H. Nahm, S.G. Almeida, and M.C. Brandileone. 2010. Non-typeable *Streptococcus pneumoniae* carriage isolates genetically similar to invasive and carriage isolates expressing capsular type 14 in Brazilian infants. *J. Infect.* 61:314-322.

Arai, J., M. Hotomi, S.K. Hollingshead, Y. Ueno, D.E. Briles, and N. Yamanaka. 2011. *Streptococcus pneumoniae* isolates from middle ear fluid and nasopharynx of children with acute otitis media exhibit phase variation. *J. Clin. Microbiol.* 49:1646-1649.

Aravind, L., and L.M. Iyer. 2012. The HARE-HTH and associated domains: novel modules in the coordination of epigenetic DNA and protein modifications. *Cell Cycle* 11:119-131.

Ashton, J.R., J.L. West, and C.T. Badea. 2015. *In vivo* small animal micro-CT using nanoparticle contrast agents. *Front Pharmacol* 6:256.

Avery, O.T., and R. Dubos. 1931. The protective action of a specific enzyme against type Iii pneumococcus infection in mice. *J. Exp. Med.* 54:73-89.

Avery, O.T., C.M. Macleod, and M. McCarty. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus Type III. *J. Exp. Med.* 79:137-158.

Bakaletz, L.O. 2007. Bacterial biofilms in otitis media: evidence and relevance. *Pediatr. Infect. Dis. J.* 26:S17-19.

Balsells, E., L. Guillot, H. Nair, and M.H. Kyaw. 2017. Serotype distribution of *Streptococcus pneumoniae* causing invasive disease in children in the post-PCV era: A systematic review and meta-analysis. *PLoS One* 12:e0177113.

Bandara, M., J.M. Skehel, A. Kadioglu, I. Collinson, A.H. Nobbs, A.J. Blocker, and H.F. Jenkinson. 2017. The accessory Sec system (SecY2A2) in *Streptococcus pneumoniae* is involved in export of pneumolysin toxin, adhesion and biofilm formation. *Microbes Infect* 19:402-412.

Barraud, N., D. Schleheck, J. Klebensberger, J.S. Webb, D.J. Hassett, S.A. Rice, and S. Kjelleberg. 2009. Nitric oxide signaling in *Pseudomonas aeruginosa* biofilms mediates phosphodiesterase activity, decreased cyclic di-GMP levels, and enhanced dispersal. *J. Bacteriol.* 191:7333-7342.

Bastacky, J., G.R. Hook, G.L. Finch, J. Goerke, and T.L. Hayes. 1987. Low-temperature scanning electron-microscopy of frozen hydrated mouse lung. *Scanning* 9:57-70.

Bauer, T., A. Torres, R. Ferrer, C. Heyer, C. Schultze-Werninghaus, and K. Rasche. 2002. Biofilm formation in endotracheal tubes. Association between pneumonia and the persistence of pathogens. *Monaldi Archives for Chest Disease* 57:84-87.

Ben-Shimol, S., N. Givon-Lavi, E. Leibovitz, S. Raiz, D. Greenberg, and R. Dagan. 2014. Near-elimination of otitis media caused by 13-valent pneumococcal conjugate vaccine (PCV) serotypes in southern Israel shortly after sequential introduction of 7-valent/13-valent PCV. *Clin. Infect. Dis.* 59:1724-1732.

Benton, K.A., M.P. Everson, and D.E. Briles. 1995. A pneumolysin-negative mutant of *Streptococcus pneumoniae* causes chronic bacteremia rather than acute sepsis in mice. *Infect. Immun.* 63:448-455.

Benton, K.A., J.C. Paton, and D.E. Briles. 1997a. Differences in virulence for mice among *Streptococcus pneumoniae* strains of capsular types 2, 3, 4, 5, and 6 are not attributable to differences in pneumolysin production. *Infect. Immun.* 65:1237-1244.

Benton, K.A., J.C. Paton, and D.E. Briles. 1997b. The hemolytic and complement-activating properties of pneumolysin do not contribute individually to virulence in a pneumococcal bacteremia model. *Microb. Pathog.* 23:201-209.

Berge, M.J., C. Mercy, I. Mortier-Barriere, M.S. VanNieuwenhze, Y.V. Brun, C. Grangeasse, P. Polard, and N. Campo. 2017. A programmed cell division delay preserves genome integrity during natural genetic transformation in *Streptococcus pneumoniae*. *Nat Commun* 8:1621.

Bergin, D., E.P. Reeves, J. Renwick, F.B. Wientjes, and K. Kavanagh. 2005. Superoxide production in *Galleria mellonella* hemocytes: identification of proteins homologous to the NADPH oxidase complex of human neutrophils. *Infect. Immun.* 73:4161-4170.

Berry, A.M., J. Yother, D.E. Briles, D. Hansman, and J.C. Paton. 1989. Reduced virulence of a defined pneumolysin-negative mutant of *Streptococcus pneumoniae*. *Infect. Immun.* 57:2037-2042.

Bidossi, A., L. Mulas, F. Decorosi, L. Colomba, S. Ricci, G. Pozzi, J. Deutscher, C. Viti, and M.R. Oggioni. 2012. A functional genomics approach to establish the complement of carbohydrate transporters in *Streptococcus pneumoniae*. *PLoS One* 7:e33320.

Blanchette-Cain, K., C.A. Hinojosa, R. Akula Suresh Babu, A. Lizcano, N. Gonzalez-Juarbe, C. Munoz-Almagro, C.J. Sanchez, M.A. Bergman, and C.J. Orihuela. 2013. *Streptococcus pneumoniae* biofilm formation is strain dependent, multifactorial, and associated with reduced invasiveness and immunoreactivity during colonization. *MBio* 4:e00745-00713.

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

Boles, B.R., M. Thoendel, and P.K. Singh. 2004. Self-generated diversity produces "insurance effects" in biofilm communities. *Proc Natl Acad Sci U S A* 101:16630-16635.
Bortoni, M.E., V.S. Terra, J. Hinds, P.W. Andrew, and H. Yesilkaya. 2009. The pneumococcal response to oxidative stress includes a role for Rgg. *Microbiology* 155:4123-4134.
Boto, L. 2010. Horizontal gene transfer in evolution: facts and challenges. *Proc Biol Sci* 277:819-827.
Boulnois, G.J., J.C. Paton, T.J. Mitchell, and P.W. Andrew. 1991. Structure and function of pneumolysin, the multifunctional, thiol-activated toxin of *Streptococcus pneumoniae*. *Mol. Microbiol.* 5:2611-2616.
Bowman, T.V., and L.I. Zon. 2010. Swimming into the future of drug discovery: in vivo chemical screens in zebrafish. *ACS Chem Biol* 5:159-161.
Brazelton, W.J., and J.A. Baross. 2009. Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. *ISME J* 3:1420-1424.
Brockhurst, M.A., N. Colegrave, D.J. Hodgson, and A. Buckling. 2007. Niche occupation limits adaptive radiation in experimental microcosms. *PLoS One* 2:e193.
Brown, N.L., J.V. Stoyanov, S.P. Kidd, and J.L. Hobman. 2003. The MerR family of transcriptional regulators. *FEMS Microbiol. Rev.* 27:145-163.
Brown, S.E., A. Howard, A.B. Kasprzak, K.H. Gordon, and P.D. East. 2009. A peptidomics study reveals the impressive antimicrobial peptide arsenal of the wax moth *Galleria mellonella*. *Insect Biochem. Mol. Biol.* 39:792-800.
Buckwalter, C.M., and S.J. King. 2012. Pneumococcal carbohydrate transport: food for thought. *Trends Microbiol.* 20:517-522.
Burdett, V. 1980. Identification of tetracycline-resistant R-plasmids in *Streptococcus agalactiae* (group B). *Antimicrob. Agents Chemother.* 18:753-760.
Caierao, J., P. Hawkins, F.H. Sant'anna, G.R. da Cunha, P.A. d'Azevedo, L. McGee, and C. Dias. 2014. Serotypes and genotypes of invasive *Streptococcus pneumoniae* before and after PCV10 implementation in southern Brazil. *PLoS One* 9:e111129.
Caillaud, F., P. Trieu-Cuot, C. Carlier, and P. Courvalin. 1987. Nucleotide sequence of the kanamycin resistance determinant of the pneumococcal transposon Tn1545: evolutionary relationships and transcriptional analysis of aphA-3 genes. *Mol. Gen. Genet.* 207:509-513.
Caro-Aguilar, I., L. Indrawati, R.M. Kaufhold, C. Gaunt, Y. Zhang, D.K. Nawrocki, C. Giovarelli, M.A. Winters, W.J. Smith, J. Heinrichs, and J.M. Skinner. 2017. Immunogenicity differences of a 15-valent pneumococcal polysaccharide conjugate vaccine (PCV15) based on vaccine dose, route of immunization and mouse strain. *Vaccine* 35:865-872.
Caymaris, S., H.J. Bootsma, B. Martin, P.W. Hermans, M. Prudhomme, and J.P. Claverys. 2010. The global nutritional regulator CodY is an essential protein in the human pathogen *Streptococcus pneumoniae*. *Mol. Microbiol.* 78:344-360.
Ceri, H., M.E. Olson, C. Stremick, R.R. Read, D. Morck, and A. Buret. 1999. The Calgary Biofilm Device: new technology for rapid determination of antibiotic susceptibilities of bacterial biofilms. *J. Clin. Microbiol.* 37:1771-1776.
Chaguza, C., C.P. Andam, S.R. Harris, J.E. Cornick, M. Yang, L. Bricio-Moreno, A.W. Kamng'ona, J. Parkhill, N. French, R.S. Heyderman, A. Kadioglu, D.B. Everett, S.D. Bentley, and W.P. Hanage. 2016. Recombination in *Streptococcus pneumoniae* lineages increase with carriage duration and size of the polysaccharide capsule. *MBio* 7:

Chang, Q., A.E. Stevenson, N.J. Croucher, G.M. Lee, S.I. Pelton, M. Lipsitch, J.A. Finkelstein, and W.P. Hanage. 2015. Stability of the pneumococcal population structure in Massachusetts as PCV13 was introduced. *BMC Infect. Dis.* 15:68.
Chao, Y., L.R. Marks, M.M. Pettigrew, and A.P. Hakansson. 2014. *Streptococcus pneumoniae* biofilm formation and dispersion during colonization and disease. *Front Cell Infect Microbiol* 4:194.
Chen, Y.F., and J.D. Helmann. 1997. DNA-melting at the *Bacillus subtilis* flagellin promoter nucleates near -10 and expands unidirectionally. *J. Mol. Biol.* 267:47-59.
Chiavolini, D., G. Pozzi, and S. Ricci. 2008. Animal models of *Streptococcus pneumoniae* disease. *Clin. Microbiol. Rev.* 21:666-685.
Cho, S.J., K. Rooney, A.M.K. Choi, and H.W. Stout-Delgado. 2018. NLRP3 inflammasome activation in aged macrophages is diminished during *Streptococcus pneumoniae* infection. *Am J Physiol Lung Cell Mol Physiol* 314:L372-L387.
Churton, N.W., R.V. Misra, R.P. Howlin, R.N. Allan, J. Jefferies, S.N. Faust, S.E. Gharbia, R.J. Edwards, S.C. Clarke, and J.S. Webb. 2016. Parallel evolution in *Streptococcus pneumoniae* biofilms. *Genome Biol Evol* 8:1316-1326.
Churton, N.W.V. 2014. Genetic and phenotypic diversification within biofilms formed by clinically relevant strains of *Streptococcus pneumoniae*. In Faculty of Natural and Enviromental Sciences. University of Southampton, University of Southampton. 260.
Cochetti, I., E. Tili, M. Vecchi, A. Manzin, M. Mingoia, P.E. Varaldo, and M.P. Montanari. 2007. New Tn916-related elements causing erm(B)-mediated erythromycin resistance in tetracycline-susceptible pneumococci. *J. Antimicrob. Chemother.* 60:127-131.
Cockeran, R., H.C. Steel, T.J. Mitchell, C. Feldman, and R. Anderson. 2001a. Pneumolysin potentiates production of prostaglandin E(2) and leukotriene B(4) by human neutrophils. *Infect. Immun.* 69:3494.
Cockeran, R., A.J. Theron, H.C. Steel, N.M. Matlola, T.J. Mitchell, C. Feldman, and R. Anderson. 2001b. Proinflammatory interactions of pneumolysin with human neutrophils. *The Journal of Infectious Diseases* 183:604-611.
Conibear, T.C., S.L. Collins, and J.S. Webb. 2009. Role of mutation in *Pseudomonas aeruginosa* biofilm development. *PLoS One* 4:e6289.
Conrad, A., M.K. Suutari, M.M. Keinanen, A. Cadoret, P. Faure, L. Mansuy-Huault, and J.C. Block. 2003. Fatty acids of lipid fractions in extracellular polymeric substances of activated sludge flocs. *Lipids* 38:1093-1105.
Cook, S.M., and J.D. McArthur. 2013. Developing *Galleria mellonella* as a model host for human pathogens. *Virulence* 4:350-353.
Cools, F., E. Torfs, B. Vanhoutte, M.B. de Macedo, L. Bonofiglio, M. Mollerach, L. Maes, G. Caljon, P. Delputte, D. Cappoen, and P. Cos. 2018. *Streptococcus pneumoniae* galU gene mutation has a direct effect on biofilm growth, adherence and phagocytosis *in vitro* and pathogenicity *in vivo*. *Pathog Dis* 76:
Cooper, V.S., R.K. Staples, C.C. Traverse, and C.N. Ellis. 2014. Parallel evolution of small colony variants in *Burkholderia cenocepacia* biofilms. *Genomics* 104:447-452.
Costerton, J.W., Z. Lewandowski, D.E. Caldwell, D.R. Korber, and H.M. Lappin-Scott. 1995. Microbial biofilms. *Annu. Rev. Microbiol.* 49:711-745.
Costerton, J.W., and P.S. Stewart. 2001. Battling biofilms. *Sci. Am.* 285:74-81.
Costerton, J.W., P.S. Stewart, and E.P. Greenberg. 1999. Bacterial biofilms: a common cause of persistent infections. *Science* 284:1318-1322.
Coticchia, J.M., M. Chen, L. Sachdeva, and S. Mutchnick. 2013. New paradigms in the pathogenesis of otitis media in children. *Front Pediatr* 1:52.

Crain, M.J., W.D. Waltman, 2nd, J.S. Turner, J. Yother, D.F. Talkington, L.S. McDaniel, B.M. Gray, and D.E. Briles. 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.

Cremers, A.J., F.M. Mobegi, M.I. de Jonge, S.A. van Hijum, J.F. Meis, P.W. Hermans, G. Ferwerda, S.D. Bentley, and A.L. Zomer. 2015. The post-vaccine microevolution of invasive *Streptococcus pneumoniae*. *Sci Rep* 5:14952.

Croucher, N.J., D. Walker, P. Romero, N. Lennard, G.K. Paterson, N.C. Bason, A.M. Mitchell, M.A. Quail, P.W. Andrew, J. Parkhill, S.D. Bentley, and T.J. Mitchell. 2009. Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone *Streptococcus pneumoniae* Spain 23F ST81. *J. Bacteriol.* 191:1480-1489.

Cuevas, R.A., R. Eutsey, A. Kadam, J.A. West-Roberts, C.A. Woolford, A.P. Mitchell, K.M. Mason, and N.L. Hiller. 2017. A novel streptococcal cell-cell communication peptide promotes pneumococcal virulence and biofilm formation. *Mol. Microbiol.* 105:554-571.

Cui, Y.A., H. Patel, W.M. O'Neil, S. Li, and P. Saddier. 2017. Pneumococcal serotype distribution: A snapshot of recent data in pediatric and adult populations around the world. *Hum Vaccin Immunother* 13:1-13.

Dalia, A.B., and J.N. Weiser. 2011. Minimization of bacterial size allows for complement evasion and is overcome by the agglutinating effect of antibody. *Cell Host Microbe* 10:486-496.

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

de Beer, D., P. Stoodley, F. Roe, and Z. Lewandowski. 1994. Effects of biofilm structures on oxygen distribution and mass transport. *Biotechnol. Bioeng.* 43:1131-1138.

de Hoon, M.J., Y. Makita, K. Nakai, and S. Miyano. 2005. Prediction of transcriptional terminators in *Bacillus subtilis* and related species. *PLoS Comput Biol* 1:e25.

de Jong, A., H. Pietersma, M. Cordes, O.P. Kuipers, and J. Kok. 2012. PePPER: a webserver for prediction of prokaryote promoter elements and regulons. *BMC Genomics* 13:299.

de Jong, L., E.A. de Koning, W. Roseboom, H. Buncherd, M.J. Wanner, I. Dapic, P.J. Jansen, J.H. van Maarseveen, G.L. Corthals, P.J. Lewis, L.W. Hamoen, and C.G. de Koster. 2017. In-culture cross-linking of bacterial cells reveals large-scale dynamic protein-protein interactions at the peptide level. *J Proteome Res* 16:2457-2471.

Del Grosso, M., A. Scotto d'Abusco, F. Iannelli, G. Pozzi, and A. Pantosti. 2004. Tn2009, a Tn916-like element containing *mef*(E) in *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* 48:2037-2042.

Demczuk, W.H.B., I. Martin, L. Hoang, P. Van Caeseele, B. Lefebvre, G. Horsman, D. Haldane, J. Gubbay, S. Ratnam, G. German, J. Daley Bernier, L. Strudwick, A. McGeer, G.G. Zhanel, G. Van Domselaar, M. Graham, and M.R. Mulvey. 2017. Phylogenetic analysis of emergent *Streptococcus pneumoniae* serotype 22F causing invasive pneumococcal disease using whole genome sequencing. *PLoS One* 12:e0178040.

Demo, G., V. Papouskova, J. Komarek, P. Kaderavek, O. Otrusinova, P. Srb, A. Rabatinova, L. Krasny, L. Zidek, V. Sklenar, and M. Wimmerova. 2014. X-ray vs. NMR structure of N-terminal domain of delta-subunit of RNA polymerase. *J Struct Biol* 187:174-186.

Devaraj, A., S.S. Justice, L.O. Bakaletz, and S.D. Goodman. 2015. DNABII proteins play a central role in UPEC biofilm structure. *Mol. Microbiol.* 96:1119-1135.
Devine, V.T., D.W. Cleary, J.M. Jefferies, R. Anderson, D.E. Morris, A.C. Tuck, R.A. Gladstone, G. O'Doherty, P. Kuruparan, S.D. Bentley, S.N. Faust, and S.C. Clarke. 2017. The rise and fall of pneumococcal serotypes carried in the PCV era. *Vaccine* 35:1293-1298.
Di Guilmi, A.M., and A. Dessen. 2002. New approaches towards the identification of antibiotic and vaccine targets in *Streptococcus pneumoniae*. *EMBO Rep* 3:728-734.
Doherty, G.P., M.J. Fogg, A.J. Wilkinson, and P.J. Lewis. 2010. Small subunits of RNA polymerase: localization, levels and implications for core enzyme composition. *Microbiology* 156:3532-3543.
Domenech, M., E. Garcia, and M. Moscoso. 2009. Versatility of the capsular genes during biofilm formation by *Streptococcus pneumoniae*. *Environ. Microbiol.* 11:2542-2555.
Domenech, M., E. Garcia, and M. Moscoso. 2011. In vitro destruction of *Streptococcus pneumoniae* biofilms with bacterial and phage peptidoglycan hydrolases. *Antimicrob. Agents Chemother.* 55:4144-4148.
Domenech, M., E. Garcia, and M. Moscoso. 2012. Biofilm formation in *Streptococcus pneumoniae*. *Microb Biotechnol* 5:455-465.
Donlan, R.M., J.A. Piede, C.D. Heyes, L. Sanii, R. Murga, P. Edmonds, I. El-Sayed, and M.A. El-Sayed. 2004. Model system for growing and quantifying *Streptococcus pneumoniae* biofilms in situ and in real time. *Appl. Environ. Microbiol.* 70:4980-4988.
Drijkoningen, J.J., and G.G. Rohde. 2014. Pneumococcal infection in adults: burden of disease. *Clin. Microbiol. Infect.* 20 Suppl 5:45-51.
Echenique, J., A. Kadioglu, S. Romao, P.W. Andrew, and M.C. Trombe. 2004. Protein serine/threonine kinase StkP positively controls virulence and competence in *Streptococcus pneumoniae*. *Infect. Immun.* 72:2434-2437.
Edenberg, E.R., M. Downey, and D. Toczyski. 2014. Polymerase stalling during replication, transcription and translation. *Curr. Biol.* 24:R445-452.
Evans, B.A., and D.E. Rozen. 2012. A *Streptococcus pneumoniae* infection model in larvae of the wax moth *Galleria mellonella*. *Eur. J. Clin. Microbiol. Infect. Dis.* 31:2653-2660.
Evans, B.A., and D.E. Rozen. 2013. Significant variation in transformation frequency in *Streptococcus pneumoniae*. *ISME J* 7:791-799.
Federle, M.J. 2009. Autoinducer-2-based chemical communication in bacteria: complexities of interspecies signaling. *Contrib Microbiol* 16:18-32.
Fernebro, J., I. Andersson, J. Sublett, E. Morfeldt, R. Novak, E. Tuomanen, S. Normark, and B.H. Normark. 2004. Capsular expression in *Streptococcus pneumoniae* negatively affects spontaneous and antibiotic-induced lysis and contributes to antibiotic tolerance. *J. Infect. Dis.* 189:328-338.
Flemming, H.C., and J. Wingender. 2010. The biofilm matrix. *Nat. Rev. Microbiol.* 8:623-633.
Folkesson, A., L. Jelsbak, L. Yang, H.K. Johansen, O. Ciofu, N. Hoiby, and S. Molin. 2012. Adaptation of *Pseudomonas aeruginosa* to the cystic fibrosis airway: an evolutionary perspective. *Nat. Rev. Microbiol.* 10:841-851.
Frølund, B., R. Palmgren, K. Keiding, and P.H. Nielsen. 1996. Extraction of extracellular polymers from activated sludge using a cation exchange resin. *Water Res.* 30:1749-1758.

Fux, C.A., J.W. Costerton, P.S. Stewart, and P. Stoodley. 2005. Survival strategies of infectious biofilms. *Trends Microbiol.* 13:34-40.
Gagen, S.J., and N.A. Ratcliffe. 1976. Studies on the *in vivo* cellular reactions and fate of injected bacteria in *Galleria mellonella* and *Pieris brassicae* larvae. *J. Invertebr. Pathol.* 28:17-24.
Gagne, A.L., K.E. Stevens, M. Cassone, A. Pujari, O.E. Abiola, D.J. Chang, and M.E. Sebert. 2013. Competence in *Streptococcus pneumoniae* is a response to an increasing mutational burden. *PLoS One* 8:e72613.
Galante, J., A.C. Ho, S. Tingey, and B.M. Charalambous. 2015. Quorum sensing and biofilms in the pathogen, *Streptococcus pneumoniae*. *Curr. Pharm. Des.* 21:25-30.
Garsin, D.A., C.D. Sifri, E. Mylonakis, X. Qin, K.V. Singh, B.E. Murray, S.B. Calderwood, and F.M. Ausubel. 2001. A simple model host for identifying Gram-positive virulence factors. *Proc Natl Acad Sci U S A* 98:10892-10897.
Geno, K.A., G.L. Gilbert, J.Y. Song, I.C. Skovsted, K.P. Klugman, C. Jones, H.B. Konradsen, and M.H. Nahm. 2015. Pneumococcal capsules and their types: past, present, and future. *Clin. Microbiol. Rev.* 28:871-899.
Giebink, G.S. 1989. The microbiology of otitis media. *Pediatr. Infect. Dis. J.* 8:S18-20.
Gilbreth, S.E., A.K. Benson, and R.W. Hutkins. 2004. Catabolite repression and virulence gene expression in *Listeria monocytogenes*. *Curr. Microbiol.* 49:95-98.
Gilley, R.P., and C.J. Orihuela. 2014. Pneumococci in biofilms are non-invasive: implications on nasopharyngeal colonization. *Front Cell Infect Microbiol* 4:163.
Gingles, N.A., J.E. Alexander, A. Kadioglu, P.W. Andrew, A. Kerr, T.J. Mitchell, E. Hopes, P. Denny, S. Brown, H.B. Jones, S. Little, G.C. Booth, and W.L. McPheat. 2001. Role of genetic resistance in invasive pneumococcal infection: identification and study of susceptibility and resistance in inbred mouse strains. *Infect. Immun.* 69:426-434.
Gisselsson-Solen, M., G. Henriksson, A. Hermansson, and A. Melhus. 2015. Effect of pneumococcal conjugate vaccination on nasopharyngeal carriage in children with early onset of acute otitis media - a randomized controlled trial. *Acta Otolaryngol* 135:7-13.
Glavis-Bloom, J., M. Muhammed, and E. Mylonakis. 2012. Of model hosts and man: using *Caenorhabditis elegans, Drosophila melanogaster* and *Galleria mellonella* as model hosts for infectious disease research. *Adv. Exp. Med. Biol.* 710:11-17.
Golden, A.R., H.J. Adam, M.W. Gilmour, M.R. Baxter, I. Martin, K.A. Nichol, W.H. Demczuk, D.J. Hoban, and G.G. Zhanel. 2015. Assessment of multidrug resistance, clonality and virulence in non-PCV-13 *Streptococcus pneumoniae* serotypes in Canada, 2011-13. *J. Antimicrob. Chemother.* 70:1960-1964.
Griffith, F. 1928. The significance of pneumococcal types. *J Hyg (Lond)* 27:113-159.
Hall-Stoodley, L., J.W. Costerton, and P. Stoodley. 2004. Bacterial biofilms: from the natural environment to infectious diseases. *Nat. Rev. Microbiol.* 2:95-108.
Hall-Stoodley, L., F.Z. Hu, A. Gieseke, L. Nistico, D. Nguyen, J. Hayes, M. Forbes, D.P. Greenberg, B. Dice, A. Burrows, P.A. Wackym, P. Stoodley, J.C. Post, G.D. Ehrlich, and J.E. Kerschner. 2006. Direct detection of bacterial biofilms on the middle-ear mucosa of children with chronic otitis media. *JAMA* 296:202-211.
Hall-Stoodley, L., and P. Stoodley. 2009. Evolving concepts in biofilm infections. *Cell. Microbiol.* 11:1034-1043.
Hamaguchi, S., M.A. Zafar, M. Cammer, and J.N. Weiser. 2018. Capsule prolongs survival of *Streptococcus pneumoniae* during starvation. *Infect. Immun.* 86:e00802-00817.

Hanage, W.P., C.J. Bishop, S.S. Huang, A.E. Stevenson, S.I. Pelton, M. Lipsitch, and J.A. Finkelstein. 2011. Carried pneumococci in Massachusetts children: the contribution of clonal expansion and serotype switching. *Pediatr. Infect. Dis. J.* 30:302-308.
Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166:557-580.
Harrison, S.C. 1991. A structural taxonomy of DNA-binding domains. *Nature* 353:715-719.
Hava, D.L., and A. Camilli. 2002. Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors. *Mol. Microbiol.* 45:1389-1406.
Hava, D.L., J. LeMieux, and A. Camilli. 2003. From nose to lung: the regulation behind *Streptococcus pneumoniae* virulence factors. *Mol. Microbiol.* 50:1103-1110.
Havarstein, L.S., G. Coomaraswamy, and D.A. Morrison. 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., R. Hakenbeck, and P. Gaustad. 1997. Natural competence in the genus *Streptococcus*: evidence that streptococci can change pherotype by interspecies recombinational exchanges. *J. Bacteriol.* 179:6589-6594.
Heldwein, E.E., and R.G. Brennan. 2001. Crystal structure of the transcription activator BmrR bound to DNA and a drug. *Nature* 409:378-382.
Henrichsen, J. 1995. Six newly recognized types of *Streptococcus pneumoniae*. *J. Clin. Microbiol.* 33:2759-2762.
Hernandez-Arriaga, A.M., M. Espinosa, and G. del Solar. 2012. Fitness of the pMV158 replicon in *Streptococcus pneumoniae*. *Plasmid* 67:162-166.
Hiller, N.L., A. Ahmed, E. Powell, D.P. Martin, R. Eutsey, J. Earl, B. Janto, R.J. Boissy, J. Hogg, K. Barbadora, R. Sampath, S. Lonergan, J.C. Post, F.Z. Hu, and G.D. Ehrlich. 2010. Generation of genic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal pediatric infection. *PLoS Pathog* 6:e1001108.
Hirst, R.A., A. Kadioglu, C. O'Callaghan, and P.W. Andrew. 2004. The role of pneumolysin in pneumococcal pneumonia and meningitis. *Clin. Exp. Immunol.* 138:195-201.
Hoa, M., M. Syamal, L. Sachdeva, R. Berk, and J. Coticchia. 2009. Demonstration of nasopharyngeal and middle ear mucosal biofilms in an animal model of acute otitis media. *Ann Otol Rhinol Laryngol* 118:292-298.
Hoffman, L.R., E. Deziel, D.A. D'Argenio, F. Lepine, J. Emerson, S. McNamara, R.L. Gibson, B.W. Ramsey, and S.I. Miller. 2006. Selection for *Staphylococcus aureus* small-colony variants due to growth in the presence of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* 103:19890-19895.
Hoiby, N., T. Bjarnsholt, M. Givskov, S. Molin, and O. Ciofu. 2010. Antibiotic resistance of bacterial biofilms. *Int. J. Antimicrob. Agents* 35:322-332.
Hoiby, N., T. Bjarnsholt, C. Moser, G.L. Bassi, T. Coenye, G. Donelli, L. Hall-Stoodley, V. Hola, C. Imbert, K. Kirketerp-Moller, D. Lebeaux, A. Oliver, A.J. Ullmann, C. Williams, E.S.G.f. Biofilms, and Z. Consulting External Expert Werner. 2015. ESCMID guideline for the diagnosis and treatment of biofilm infections 2014. *Clin. Microbiol. Infect.* 21 Suppl 1:S1-25.
Hollingshead, S.K., R. Becker, and D.E. Briles. 2000. Diversity of PspA: mosaic genes and evidence for past recombination in *Streptococcus pneumoniae*. *Infect. Immun.* 68:5889-5900.

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

Hsieh, Y.C., J.T. Wang, W.S. Lee, P.R. Hsueh, P.L. Shao, L.Y. Chang, C.Y. Lu, C.Y. Lee, F.Y. Huang, and L.M. Huang. 2006. Serotype competence and penicillin resistance in *Streptococcus pneumoniae*. *Emerg Infect Dis* 12:1709-1714.

Hua, C.Z., A. Howard, R. Malley, and Y.J. Lu. 2014. Effect of nonheme iron-containing ferritin Dpr in the stress response and virulence of pneumococci. *Infect. Immun.* 82:3939-3947.

Hughes, K.A., I.W. Sutherland, and M.V. Jones. 1998. Biofilm susceptibility to bacteriophage attack: the role of phage-borne polysaccharide depolymerase. *Microbiology* 144 ( Pt 11):3039-3047.

Hyams, C., E. Camberlein, J.M. Cohen, K. Bax, and J.S. Brown. 2010. The *Streptococcus pneumoniae* capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. *Infect. Immun.* 78:704-715.

Ibrahim, H.R., T. Matsuzaki, and T. Aoki. 2001. Genetic evidence that antibacterial activity of lysozyme is independent of its catalytic function. *FEBS Lett.* 506:27-32.

Imlay, J.A., S.M. Chin, and S. Linn. 1988. Toxic DNA damage by hydrogen peroxide through the Fenton reaction *in vivo* and *in vitro*. *Science* 240:640-642.

Jander, G., L.G. Rahme, and F.M. Ausubel. 2000. Positive correlation between virulence of *Pseudomonas aeruginosa* mutants in mice and insects. *J. Bacteriol.* 182:3843-3845.

Jansen, W.T., M. Bolm, R. Balling, G.S. Chhatwal, and R. Schnabel. 2002. Hydrogen peroxide-mediated killing of *Caenorhabditis elegans* by *Streptococcus pyogenes*. *Infect. Immun.* 70:5202-5207.

Jefferies, J.M., S.C. Clarke, J.S. Webb, and A.R. Kraaijeveld. 2011. Risk of red queen dynamics in pneumococcal vaccine strategy. *Trends Microbiol.* 19:377-381.

Jefferies, J.M.C., C.H.G. Johnston, L.-A.S. Kirkham, G.J.M. Cowan, K.S. Ross, A. Smith, S.C. Clarke, A.B. Brueggemann, R.C. George, B. Pichon, G. Pluschke, V. Pflüger, and T.J. Mitchell. 2007. Presence of nonhemolytic pneumolysin in serotypes of *Streptococcus pneumoniae* associated with disease outbreaks. *The Journal of Infectious Diseases* 196:936-944.

Jensen, A., O. Valdorsson, N. Frimodt-Moller, S. Hollingshead, and M. Kilian. 2015. Commensal streptococci serve as a reservoir for beta-lactam resistance genes in *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* 59:3529-3540.

Jensen, L.J., M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, T. Doerks, P. Julien, A. Roth, M. Simonovic, P. Bork, and C. von Mering. 2009. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* 37:D412-416.

Jeong, D.G., E.S. Jeong, J.H. Seo, S.H. Heo, and Y.K. Choi. 2011. Difference in resistance to *Streptococcus pneumoniae* infection in mice. *Lab Anim Res* 27:91-98.

Jim, K.K., J. Engelen-Lee, A.M. van der Sar, W. Bitter, M.C. Brouwer, A. van der Ende, J.W. Veening, D. van de Beek, and C.M. Vandenbroucke-Grauls. 2016. Infection of zebrafish embryos with live fluorescent *Streptococcus pneumoniae* as a real-time pneumococcal meningitis model. *J Neuroinflammation* 13:188.

Johnsborg, O., P.E. Kristiansen, T. Blomqvist, and L.S. Havarstein. 2006. A hydrophobic patch in the competence-stimulating Peptide, a pneumococcal competence pheromone, is essential for specificity and biological activity. *J. Bacteriol.* 188:1744-1749.

Joloba, M.L., B.R. Kidenya, D.P. Kateete, F.A. Katabazi, J.K. Muwanguzi, B.B. Asiimwe, S.P. Alarakol, J.L. Nakavuma, S. Bajaksouzian, A. Windau, and M.R. Jacobs. 2010. Comparison of transformation frequencies among selected *Streptococcus pneumoniae* serotypes. *Int. J. Antimicrob. Agents* 36:124-128.

Jones, A.L., K.M. Knoll, and C.E. Rubens. 2000. Identification of *Streptococcus agalactiae* virulence genes in the neonatal rat sepsis model using signature-tagged mutagenesis. *Mol. Microbiol.* 37:1444-1455.

Jones, A.L., R.H.V. Needham, and C.E. Rubens. 2003. The delta subunit of RNA polymerase is required for virulence of S*treptococcus agalactiae*. *Infect. Immun.* 71:4011-4017.

Jose, R.J., J.N. Periselneris, and J.S. Brown. 2015. Community-acquired pneumonia. *Curr Opin Pulm Med* 21:212-218.

Jusot, J.F., D.R. Neill, E.M. Waters, M. Bangert, M. Collins, L. Bricio Moreno, K.G. Lawan, M.M. Moussa, E. Dearing, D.B. Everett, J.M. Collard, and A. Kadioglu. 2017. Airborne dust and high temperatures are risk factors for invasive bacterial disease. *J. Allergy Clin. Immunol.* 139:977-986 e972.

Kadioglu, A., and P.W. Andrew. 2005. Susceptibility and resistance to pneumococcal disease in mice. *Brief Funct Genomic Proteomic* 4:241-247.

Kadioglu, A., W. Coward, M.J. Colston, C.R.A. Hewitt, and P.W. Andrew. 2004. CD4-T-lymphocyte interactions with pneumolysin and pneumococci suggest a crucial protective role in the host response to pneumococcal infection. *Infect. Immun.* 72:2689-2697.

Kadioglu, A., J.N. Weiser, J.C. Paton, and P.W. Andrew. 2008. The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nat. Rev. Microbiol.* 6:288-301.

Kallberg, M., H. Wang, S. Wang, J. Peng, Z. Wang, H. Lu, and J. Xu. 2012. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* 7:1511-1522.

Kaplan, B., T.L. Wandstrat, and J.R. Cunningham. 1997. Overall cost in the treatment of otitis media. *Pediatr. Infect. Dis. J.* 16:S9-S11.

Kavanagh, K., and E.P. Reeves. 2004. Exploiting the potential of insects for in vivo pathogenicity testing of microbial pathogens. *FEMS Microbiol. Rev.* 28:101-112.

Keller, L.E., D.A. Robinson, and L.S. McDaniel. 2016. Nonencapsulated *Streptococcus pneumoniae*: emergence and pathogenesis. *mBio* 7:e01792-01715.

Kelly, R.T., S. Farmer, and D. Greiff. 1967. Neuraminidase activities of clinical isolates of *Diplococcus pneumoniae*. *J. Bacteriol.* 94:272-273.

Khan, M.N., J.R. Coleman, J. Vernatter, A.K. Varshney, C. Dufaud, and L.A. Pirofski. 2014. An ahemolytic pneumolysin of *Streptococcus pneumoniae* manipulates human innate and CD4(+) T-cell responses and reduces resistance to colonization in mice in a serotype-independent manner. *J. Infect. Dis.* 210:1658-1669.

Kilian, M., K. Poulsen, T. Blomqvist, L.S. Havarstein, M. Bek-Thomsen, H. Tettelin, and U.B. Sorensen. 2008. Evolution of *Streptococcus pneumoniae* and its close commensal relatives. *PLoS One* 3:e2683.

Kilian, M., D.R. Riley, A. Jensen, H. Bruggemann, and H. Tettelin. 2014. Parallel evolution of *Streptococcus pneumoniae* and *Streptococcus mitis* to pathogenic and mutualistic lifestyles. *MBio* 5:e01490-01414.

Kim, D.E., D. Chivian, and D. Baker. 2004. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* 32:W526-531.

Kim, J.O., and J.N. Weiser. 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.
Kim, W., F. Racimo, J. Schluter, S.B. Levy, and K.R. Foster. 2014. Importance of positioning for microbial evolution. *Proc Natl Acad Sci U S A* 111:E1639-1647.
Kim, W., F.K. Tengra, Z. Young, J. Shong, N. Marchand, H.K. Chan, R.C. Pangule, M. Parra, J.S. Dordick, J.L. Plawsky, and C.H. Collins. 2013. Spaceflight promotes biofilm formation by *Pseudomonas aeruginosa*. *PLoS One* 8:e62437.
Kirkham, L.-A.S., J.M.C. Jefferies, A.R. Kerr, Y. Jing, S.C. Clarke, A. Smith, and T.J. Mitchell. 2006. Identification of invasive serotype 1 pneumococcal isolates that express nonhemolytic pneumolysin. *J. Clin. Microbiol.* 44:151.
Klausen, M., A. Aaes-Jorgensen, S. Molin, and T. Tolker-Nielsen. 2003a. Involvement of bacterial migration in the development of complex multicellular structures in *Pseudomonas aeruginosa* biofilms. *Mol. Microbiol.* 50:61-68.
Klausen, M., A. Heydorn, P. Ragas, L. Lambertsen, A. Aaes-Jorgensen, S. Molin, and T. Tolker-Nielsen. 2003b. Biofilm formation by *Pseudomonas aeruginosa* wild type, flagella and type IV pili mutants. *Mol. Microbiol.* 48:1511-1524.
Klein, J.O. 2000. The burden of otitis media. *Vaccine* 19 Suppl 1:S2-8.
Knutsen, E., O. Ween, and L.S. Havarstein. 2004. Two separate quorum-sensing systems upregulate transcription of the same ABC transporter in *Streptococcus pneumoniae*. *J. Bacteriol.* 186:3078-3085.
Koo, H., R.N. Allan, R.P. Howlin, P. Stoodley, and L. Hall-Stoodley. 2017. Targeting microbial biofilms: current and prospective therapeutic strategies. *Nat. Rev. Microbiol.* 15:740-755.
Kostrzynska, M., and T. Wadstrom. 1992. Binding of laminin, type IV collagen, and vitronectin by *Streptococcus pneumoniae*. *Zentralbl. Bakteriol.* 277:80-83.
Kozlowski, L.P. 2016. IPC - Isoelectric Point Calculator. *Biol Direct* 11:55.
Kramer, M.G., G. del Solar, and M. Espinosa. 1995. Lagging-strand origins of the promiscuous plasmid pMV158: physical and functional characterization. *Microbiology* 141 ( Pt 3):655-662.
Krasny, L., and R.L. Gourse. 2004. An alternative strategy for bacterial ribosome synthesis: *Bacillus subtilis* rRNA transcription regulation. *EMBO J.* 23:4473-4483.
Lacks, S.A., P. Lopez, B. Greenberg, and M. Espinosa. 1986. Identification and analysis of genes for tetracycline resistance and replication functions in the broad-host-range plasmid pLS1. *J. Mol. Biol.* 192:753-765.
Lampe, M., C. Binnie, R. Schmidt, and R. Losick. 1988. Cloned gene encoding the delta subunit of *Bacillus subtilis* RNA polymerase. *Gene* 67:13-19.
Lanie, J.A., W.L. Ng, K.M. Kazmierczak, T.M. Andrzejewski, T.M. Davidsen, K.J. Wayne, H. Tettelin, J.I. Glass, and M.E. Winkler. 2007. Genome sequence of Avery's virulent serotype 2 strain D39 of *Streptococcus pneumoniae* and comparison with that of unencapsulated laboratory strain R6. *J. Bacteriol.* 189:38-51.
Laponogov, I., X.S. Pan, D.A. Veselkov, K.E. McAuley, L.M. Fisher, and M.R. Sanderson. 2010. Structural basis of gate-DNA breakage and resealing by type II topoisomerases. *PLoS One* 5:e11338.
Laponogov, I., M.K. Sohi, D.A. Veselkov, X.S. Pan, R. Sawhney, A.W. Thompson, K.E. McAuley, L.M. Fisher, and M.R. Sanderson. 2009. Structural insight into the quinolone-DNA cleavage complex of type IIA topoisomerases. *Nat. Struct. Mol. Biol.* 16:667-669.

Lau, P.C., C.K. Sung, J.H. Lee, D.A. Morrison, and D.G. Cvitkovitch. 2002. PCR ligation mutagenesis in transformable streptococci: application and efficiency. *J. Microbiol. Methods* 49:193-205.
Lee, C.J., S.D. Banks, and J.P. Li. 1991. Virulence, immunity, and vaccine related to *Streptococcus pneumoniae*. *Crit. Rev. Microbiol.* 18:89-114.
Li, J., and J.R. Zhang. 2019. Phase variation of *Streptococcus pneumoniae*. *Microbiol Spectr* 7:
Liu, Y., and J.H. Tay. 2001. Metabolic response of biofilm to shear stress in fixed-film culture. *J. Appl. Microbiol.* 90:337-342.
Lizcano, A., T. Chin, K. Sauer, E.I. Tuomanen, and C.J. Orihuela. 2010. Early biofilm formation on microtiter plates is not correlated with the invasive disease potential of *Streptococcus pneumoniae*. *Microb. Pathog.* 48:124-130.
Lloyd-Evans, N., T.J.D. O'Dempsey, I. BALDEH, O. Secka, E. Demba, J.E. Todd, T.F. Mcardle, W.S. Banya, and B.M. Greenwood. 1996. Nasopharyngeal carriage of pneumococci in Gambian children and in their families. *The Pediatric Infectious Disease Journal* 15:866-871.
Lock, R.A., J.C. Paton, and D. Hansman. 1988. Comparative efficacy of pneumococcal neuraminidase and pneumolysin as immunogens protective against *Streptococcus pneumoniae*. *Microb. Pathog.* 5:461-467.
Loh, J.M., N. Adenwalla, S. Wiles, and T. Proft. 2013. *Galleria mellonella* larvae as an infection model for group A streptococcus. *Virulence* 4:419-428.
López de Saro, F.J., A.Y. Moon Woody, and J.D. Helmann. 1995. Structural analysis of the *Bacillus subtilis* δ factor: A protein polyanion which displaces RNA from RNA Polymerase. *J. Mol. Biol.* 252:189-202.
Lopez de Saro, F.J., N. Yoshikawa, and J.D. Helmann. 1999. Expression, abundance, and RNA polymerase binding properties of the delta factor of *Bacillus subtilis*. *J. Biol. Chem.* 274:15953-15958.
Lopez, J.M., A. Dromerick, and E. Freese. 1981. Response of guanosine 5'-triphosphate concentration to nutritional changes and its significance for *Bacillus subtilis* sporulation. *J. Bacteriol.* 146:605-613.
Lynch, J.P., 3rd, and G.G. Zhanel. 2005. Escalation of antimicrobial resistance among *Streptococcus pneumoniae*: implications for therapy. *Semin Respir Crit Care Med* 26:575-616.
Lynch, J.P., 3rd, and G.G. Zhanel. 2010. *Streptococcus pneumoniae*: epidemiology and risk factors, evolution of antimicrobial resistance, and impact of vaccines. *Curr Opin Pulm Med* 16:217-225.
Manco, S., F. Hernon, H. Yesilkaya, J.C. Paton, P.W. Andrew, and A. Kadioglu. 2006. Pneumococcal neuraminidases A and B both have essential roles during infection of the respiratory tract and sepsis. *Infect. Immun.* 74:4014-4020.
Manso, A.S., M.H. Chai, J.M. Atack, L. Furi, M. De Ste Croix, R. Haigh, C. Trappetti, A.D. Ogunniyi, L.K. Shewell, M. Boitano, T.A. Clark, J. Korlach, M. Blades, E. Mirkes, A.N. Gorban, J.C. Paton, M.P. Jennings, and M.R. Oggioni. 2014. A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. *Nat Commun* 5:5055.
Marks, L.R., B.A. Davidson, P.R. Knight, and A.P. Hakansson. 2013. Interkingdom signaling induces *Streptococcus pneumoniae* biofilm dispersion and transition from asymptomatic colonization to disease. *MBio* 4:13.
Marks, L.R., G.I. Parameswaran, and A.P. Hakansson. 2012a. Pneumococcal interactions with epithelial cells are crucial for optimal biofilm formation and colonization in vitro and in vivo. *Infect. Immun.* 80:2744-2760.

Marks, L.R., R.M. Reddinger, and A.P. Hakansson. 2012b. High levels of genetic recombination during nasopharyngeal carriage and biofilm formation in *Streptococcus pneumoniae*. *MBio* 3:

Marrie, T.J. 1994. Community-acquired pneumonia. *Clin. Infect. Dis.* 18:501-513; quiz 514-505.

Marshall, K.C., R. Stout, and R. Mitchell. 1971. Mechanism of the initial events in the sorption of marine bacteria to surfaces. *Journal of General Microbiology* 68:337-348.

Martin, B., A.L. Soulet, N. Mirouze, M. Prudhomme, I. Mortier-Barriere, C. Granadel, M.F. Noirot-Gros, P. Noirot, P. Polard, and J.P. Claverys. 2013. ComE/ComE~P interplay dictates activation or extinction status of pneumococcal X-state (competence). *Mol. Microbiol.* 87:394-411.

McElroy, K.E., J.G. Hui, J.K. Woo, A.W. Luk, J.S. Webb, S. Kjelleberg, S.A. Rice, and T. Thomas. 2014. Strain-specific parallel evolution drives short-term diversification during *Pseudomonas aeruginosa* biofilm formation. *Proc Natl Acad Sci U S A* 111:E1419-1427.

McGee, L., L. McDougal, J. Zhou, B.G. Spratt, F.C. Tenover, R. George, R. Hakenbeck, W. Hryniewicz, J.C. Lefevre, A. Tomasz, and K.P. Klugman. 2001. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the pneumococcal molecular epidemiology network. *J. Clin. Microbiol.* 39:2565-2571.

McNeela, E.A., Á. Burke, D.R. Neill, C. Baxter, V.E. Fernandes, D. Ferreira, S. Smeaton, R. El-Rachkidy, R.M. McLoughlin, A. Mori, B. Moran, K.A. Fitzgerald, J. Tschopp, V. Pétrilli, P.W. Andrew, A. Kadioglu, and E.C. Lavelle. 2010. Pneumolysin activates the NLRP3 inflammasome and promotes proinflammatory cytokines independently of TLR4. *PLoS Path.* 6:e1001191.

Milenbachs, A.A., D.P. Brown, M. Moors, and P. Youngman. 1997. Carbon-source regulation of virulence gene expression in *Listeria monocytogenes*. *Mol. Microbiol.* 23:1075-1085.

Mirouze, N., M.A. Berge, A.L. Soulet, I. Mortier-Barriere, Y. Quentin, G. Fichant, C. Granadel, M.F. Noirot-Gros, P. Noirot, P. Polard, B. Martin, and J.P. Claverys. 2013. Direct involvement of DprA, the transformation-dedicated RecA loader, in the shut-off of pneumococcal competence. *Proc Natl Acad Sci U S A* 110:E1035-1044.

Mizrachi-Nebenzahl, Y., S. Lifshitz, R. Teitelbaum, S. Novick, A. Levi, D. Benharroch, E. Ling, and R. Dagan. 2003. Differential activation of the immune system by virulent *Streptococcus pneumoniae* strains determines recovery or death of the host. *Clin. Exp. Immunol.* 134:23-31.

Morona, J.K., D.C. Miller, R. Morona, and J.C. Paton. 2004. The effect that mutations in the conserved capsular polysaccharide biosynthesis genes *cpsA*, *cpsB*, and *cpsD* have on virulence of *Streptococcus pneumoniae*. *J. Infect. Dis.* 189:1905-1913.

Morona, J.K., R. Morona, D.C. Miller, and J.C. Paton. 2003. Mutational analysis of the carboxy-terminal (YGX)4 repeat domain of CpsD, an autophosphorylating tyrosine kinase required for capsule biosynthesis in *Streptococcus pneumoniae*. *J. Bacteriol.* 185:3009-3019.

Morona, J.K., R. Morona, and J.C. Paton. 2006. Attachment of capsular polysaccharide to the cell wall of *Streptococcus pneumoniae* type 2 is required for invasive disease. *Proc Natl Acad Sci U S A* 103:8505-8510.

Morrison, D.A. 1997. Streptococcal competence for genetic transformation: regulation by peptide pheromones. *Microb. Drug Resist.* 3:27-37.

Moscoso, M., E. Garcia, and R. Lopez. 2006. Biofilm formation by *Streptococcus pneumoniae*: role of choline, extracellular DNA, and capsular polysaccharide in microbial accretion. *J. Bacteriol.* 188:7785-7795.
Moscoso, M., E. Garcia, and R. Lopez. 2009. Pneumococcal biofilms. *Int. Microbiol.* 12:77-85.
Mostafavi, S., A. Ortiz-Lopez, M.A. Bogue, K. Hattori, C. Pop, D. Koller, D. Mathis, C. Benoist, and C. Immunological Genome. 2014. Variation and genetic control of gene expression in primary immunocytes across inbred mouse strains. *Journal of immunology (Baltimore, Md. : 1950)* 193:4485-4496.
Motackova, V., H. Sanderova, L. Zidek, J. Novacek, P. Padrta, A. Svenkova, J. Korelusova, J. Jonak, L. Krasny, and V. Sklenar. 2010. Solution structure of the N-terminal domain of *Bacillus subtilis* delta subunit of RNA polymerase and its classification based on structural homologs. *Proteins* 78:1807-1810.
Mukherjee, K., T. Hain, R. Fischer, T. Chakraborty, and A. Vilcinskas. 2013. Brain infection and activation of neuronal repair mechanisms by the human pathogen *Listeria monocytogenes* in the lepidopteran model host *Galleria mellonella*. *Virulence* 4:324-332.
Muller, H.J. 1964. The relation of recombination to mutational advance. *Mutat. Res.* 106:2-9.
Munoz-Elias, E.J., J. Marcano, and A. Camilli. 2008. Isolation of *Streptococcus pneumoniae* biofilm mutants and their characterization during nasopharyngeal colonization. *Infect. Immun.* 76:5049-5061.
Murray, H.D., D.A. Schneider, and R.L. Gourse. 2003. Control of rRNA expression by small molecules is dynamic and nonredundant. *Mol. Cell* 12:125-134.
Muzzi, A., and C. Donati. 2011. Population genetics and evolution of the pan-genome of *Streptococcus pneumoniae*. *Int. J. Med. Microbiol.* 301:619-622.
Nash, J.A., T.N.S. Ballard, T.E. Weaver, and H.T. Akinbi. 2006. The peptidoglycan-degrading property of lysozyme Is not required for bactericidal activity *in vivo*. *The Journal of Immunology* 177:519.
Neill, D.R., V.E. Fernandes, L. Wisby, A.R. Haynes, D.M. Ferreira, A. Laher, N. Strickland, S.B. Gordon, P. Denny, A. Kadioglu, and P.W. Andrew. 2012. T regulatory cells control susceptibility to invasive pneumococcal pneumonia in mice. *PLoS Path.* 8:e1002660.
Nelson, A.L., A.M. Roche, J.M. Gould, K. Chim, A.J. Ratner, and J.N. Weiser. 2007. Capsule enhances pneumococcal colonization by limiting mucus-mediated clearance. *Infect. Immun.* 75:83-90.
Nieto, C., and M. Espinosa. 2003. Construction of the mobilizable plasmid pMV158GFP, a derivative of pMV158 that carries the gene encoding the green fluorescent protein. *Plasmid* 49:281-285.
O'Toole, G.A., and R. Kolter. 1998. Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. *Mol. Microbiol.* 30:295-304.
O'Toole, R.D., L. Goode, and C. Howe. 1971. Neuraminidase activity in bacterial meningitis. *J. Clin. Invest.* 50:979-985.
Ochi, K., J. Kandala, and E. Freese. 1982. Evidence that *Bacillus subtilis* sporulation induced by the stringent response is caused by the decrease in GTP or GDP. *J. Bacteriol.* 151:1062-1065.
Oggioni, M.R., C. Trappetti, A. Kadioglu, M. Cassone, F. Iannelli, S. Ricci, P.W. Andrew, and G. Pozzi. 2006. Switch from planktonic to sessile life: a major event in pneumococcal pathogenesis. *Mol. Microbiol.* 61:1196-1210.

Ogunniyi, A.D., R.L. Folland, D.E. Briles, S.K. Hollingshead, and J.C. Paton. 2000. Immunization of mice with combinations of pneumococcal virulence proteins elicits enhanced protection against challenge with *Streptococcus pneumoniae*. *Infect. Immun.* 68:3028-3033.

Ogunniyi, A.D., K.S. LeMessurier, R.M. Graham, J.M. Watt, D.E. Briles, U.H. Stroeher, and J.C. Paton. 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.

Ogunniyi, A.D., L.K. Mahdi, C. Trappetti, N. Verhoeven, D. Mermans, M.B. Van der Hoek, C.D. Plumptre, and J.C. Paton. 2012. Identification of genes that contribute to the pathogenesis of invasive pneumococcal disease by *in vivo* transcriptomic analysis. *Infect. Immun.* 80:3268-3278.

Ohse, M., K. Takahashi, Y. Kadowaki, and H. Kusaoke. 1995. Effects of plasmid DNA sizes and several other factors on transformation of *Bacillus subtilis* ISW1214 with plasmid DNA by electroporation. *Biosci Biotechnol Biochem* 59:1433-1437.

Olsen, R.J., M.E. Watkins, C.C. Cantu, S.B. Beres, and J.M. Musser. 2011. Virulence of serotype M3 Group A Streptococcus strains in wax worms (*Galleria mellonella* larvae). *Virulence* 2:111-119.

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

Papouskova, V., P. Kaderavek, O. Otrusinova, A. Rabatinova, S.S. H, J. Novacek, L. Krasny, V. Sklenar, and L. Zidek. 2013. Structural study of the partially disordered full-length delta subunit of RNA polymerase from *Bacillus subtilis*. *ChemBioChem* 14:1772-1779.

Parent, M.E., C.E. Snyder, N.D. Kopp, and D. Velegol. 2008. Localized quorum sensing in *Vibrio fischeri*. *Colloids Surf. B. Biointerfaces* 62:180-187.

Park, I.H., K.-H. Kim, A.L. Andrade, D.E. Briles, L.S. McDaniel, and M.H. Nahm. 2012. Nontypeable pneumococci can be divided into multiple cps types, including one type expressing the novel gene *pspK*. *mBio* 3:e00035-00012.

Parker, D., G. Soong, P. Planet, J. Brower, A.J. Ratner, and A. Prince. 2009. The NanA neuraminidase of *Streptococcus pneumoniae* is involved in biofilm formation. *Infect. Immun.* 77:3722-3730.

Paulsen, I.T., L. Nguyen, M.K. Sliwinski, R. Rabus, and M.H. Saier, Jr. 2000. Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. *J. Mol. Biol.* 301:75-100.

Peleg, A.Y., S. Jara, D. Monga, G.M. Eliopoulos, R.C. Moellering, Jr., and E. Mylonakis. 2009. *Galleria mellonella* as a model system to study *Acinetobacter baumannii* pathogenesis and therapeutics. *Antimicrob. Agents Chemother.* 53:2605-2609.

Perdoni, F., M. Falleni, D. Tosi, D. Cirasola, S. Romagnoli, P. Braidotti, E. Clementi, G. Bulfamante, and E. Borghi. 2014. A histological procedure to study fungal infection in the wax moth *Galleria mellonella*. *Eur J Histochem* 58:2428.

Pereira, C.S., J.A. Thompson, and K.B. Xavier. 2013. AI-2-mediated signalling in bacteria. *FEMS Microbiol. Rev.* 37:156-181.

Pericone, C.D., K. Overweg, P.W. Hermans, and J.N. Weiser. 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.

Pero, J., J. Nelson, and T.D. Fox. 1975. Highly asymmetric transcription by RNA polymerase containing phage-SP01-induced polypeptides and a new host protein. *Proc Natl Acad Sci U S A* 72:1589-1593.

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

Pettersen, E.F., T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, and T.E. Ferrin. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605-1612.

Picazo, J.J. 2009. Management of antibiotic-resistant *Streptococcus pneumoniae* infections and the use of pneumococcal conjugate vaccines. *Clin. Microbiol. Infect.* 15 Suppl 3:4-6.

Pichon, B., S.N. Ladhani, M.P. Slack, A. Segonds-Pichon, N.J. Andrews, P.A. Waight, E. Miller, and R. George. 2013. Changes in molecular epidemiology of *Streptococcus pneumoniae* causing meningitis following introduction of pneumococcal conjugate vaccination in England and Wales. *J. Clin. Microbiol.* 51:820-827.

Pike, A.C., S. Gomathinayagam, P. Swuec, M. Berti, Y. Zhang, C. Schnecke, F. Marino, F. von Delft, L. Renault, A. Costa, O. Gileadi, and A. Vindigni. 2015. Human RECQ1 helicase-driven DNA unwinding, annealing, and branch migration: insights from DNA complex structures. *Proc Natl Acad Sci U S A* 112:4286-4291.

Pozzi, G., L. Masala, F. Iannelli, R. Manganelli, L.S. Havarstein, L. Piccoli, D. Simon, and D.A. Morrison. 1996. Competence for genetic transformation in encapsulated strains of *Streptococcus pneumoniae*: two allelic variants of the peptide pheromone. *J. Bacteriol.* 178:6087-6090.

Prajapati, R.K., S. Sengupta, P. Rudra, and J. Mukhopadhyay. 2016a. *Bacillus subtilis* delta Factor Functions as a Transcriptional Regulator by Facilitating the Open Complex Formation. *J. Biol. Chem.* 291:1064-1075.

Prajapati, R.K., R. Sur, and J. Mukhopadhyay. 2016b. A Novel Function of delta Factor from *Bacillus subtilis* as a Transcriptional Repressor. *J. Biol. Chem.* 291:24029-24035.

Preston, J.A., K.W. Beagley, P.G. Gibson, and P.M. Hansbro. 2004. Genetic background affects susceptibility in nonfatal pneumococcal bronchopneumonia. *Eur. Respir. J.* 23:224-231.

Qureishi, A., Y. Lee, K. Belfield, J.P. Birchall, and M. Daniel. 2014. Update on otitis media - prevention and treatment. *Infect Drug Resist* 7:15-24.

Rabatinova, A., H. Sanderova, J. Jirat Matejckova, J. Korelusova, L. Sojka, I. Barvik, V. Papouskova, V. Sklenar, L. Zidek, and L. Krasny. 2013. The delta subunit of RNA polymerase is required for rapid changes in gene expression and competitive fitness of the cell. *J. Bacteriol.* 195:2603-2611.

Rai, P., M. Parrish, I.J. Tay, N. Li, S. Ackerman, F. He, J. Kwang, V.T. Chow, and B.P. Engelward. 2015. *Streptococcus pneumoniae* secretes hydrogen peroxide leading to DNA damage and apoptosis in lung cells. *Proc Natl Acad Sci U S A* 112:E3421-3430.

Rajendran, R., E. Borghi, M. Falleni, F. Perdoni, D. Tosi, D.F. Lappin, L. O'Donnell, D. Greetham, G. Ramage, and C. Nile. 2015. Acetylcholine protects against *Candida albicans* infection by inhibiting biofilm formation and promoting hemocyte function in a *Galleria mellonella* infection model. *Eukaryot. Cell* 14:834-844.

Ramarao, N., C. Nielsen-Leroux, and D. Lereclus. 2012. The insect *Galleria mellonella* as a powerful infection model to investigate bacterial pathogenesis. *J Vis Exp* e4392.

Ramos-Montanez, S., H.C. Tsui, K.J. Wayne, J.L. Morris, L.E. Peters, F. Zhang, K.M. Kazmierczak, L.T. Sham, and M.E. Winkler. 2008. Polymorphism and regulation

of the *spxB* (pyruvate oxidase) virulence factor gene by a CBS-HotDog domain protein (SpxR) in serotype 2 *Streptococcus pneumoniae*. *Mol. Microbiol.* 67:729-746.

Ratner, A.J., K.R. Hippe, J.L. Aguilar, M.H. Bender, A.L. Nelson, and J.N. Weiser. 2006. Epithelial cells are sensitive detectors of bacterial pore-forming toxins. *J. Biol. Chem.* 281:12994-12998.

Regev-Yochay, G., K. Trzcinski, C.M. Thompson, M. Lipsitch, and R. Malley. 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.

Reid, S.D., W. Hong, K.E. Dew, D.R. Winn, B. Pang, J. Watt, D.T. Glover, S.K. Hollingshead, and W.E. Swords. 2009. *Streptococcus pneumoniae* forms surface-attached communities in the middle ear of experimentally infected chinchillas. *J. Infect. Dis.* 199:786-794.

Rieu, A., N. Aoudia, G. Jego, J. Chluba, N. Yousfi, R. Briandet, J. Deschamps, B. Gasquet, V. Monedero, C. Garrido, and J. Guzzo. 2014. The biofilm mode of life boosts the anti-inflammatory properties of *Lactobacillus*. *Cell. Microbiol.* 16:1836-1853.

Ritchie, N.D., R. Ritchie, H.K. Bayes, T.J. Mitchell, and T.J. Evans. 2018. IL-17 can be protective or deleterious in murine pneumococcal pneumonia. *PLoS Path.* 14:e1007099.

Roberts, A.P., and P. Mullany. 2011. Tn916-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. *FEMS Microbiol. Rev.* 35:856-871.

Rodriguez, J.L., A.B. Dalia, and J.N. Weiser. 2012. Increased chain length promotes pneumococcal adherence and colonization. *Infect. Immun.* 80:3454-3459.

Romero, P., N.J. Croucher, N.L. Hiller, F.Z. Hu, G.D. Ehrlich, S.D. Bentley, E. Garcia, and T.J. Mitchell. 2009. Comparative genomic analysis of ten *Streptococcus pneumoniae* temperate bacteriophages. *J. Bacteriol.* 191:4854-4862.

Royet, J., J.M. Reichhart, and J.A. Hoffmann. 2005. Sensing and signaling during infection in *Drosophila*. *Curr. Opin. Immunol.* 17:11-17.

Sakai, F., S.J. Talekar, C.F. Lanata, C.G. Grijalva, K.P. Klugman, J.E. Vidal, R.P. Group, and G. Investigators. 2013. Expression of *Streptococcus pneumoniae* virulence-related genes in the Nasopharynx of healthy children. *PLoS One* 8:e67147.

Sanchez, C.J., N. Kumar, A. Lizcano, P. Shivshankar, J.C.D. Hotopp, J.H. Jorgensen, H. Tettelin, and C.J. Orihuela. 2011. *Streptococcus pneumoniae* in biofilms are unable to cause invasive disease due to altered virulence determinant production. *Plos One* 6:13.

Sanclement, J.A., P. Webster, J. Thomas, and H.H. Ramadan. 2005. Bacterial biofilms in surgical specimens of patients with chronic rhinosinusitis. *Laryngoscope* 115:578-582.

Sanderson, A.R., J.G. Leid, and D. Hunsaker. 2006. Bacterial biofilms on the sinus mucosa of human subjects with chronic rhinosinusitis. *Laryngoscope* 116:1121-1126.

Saralahti, A., and M. Ramet. 2015. Zebrafish and Streptococcal Infections. *Scand. J. Immunol.* 82:174-183.

Sauer, K., A.K. Camper, G.D. Ehrlich, J.W. Costerton, and D.G. Davies. 2002. *Pseudomonas aeruginosa* displays multiple phenotypes during development as a biofilm. *J. Bacteriol.* 184:1140-1154.

Schaffner, T.O., J. Hinds, K.A. Gould, D. Wuthrich, R. Bruggmann, M. Kuffer, K. Muhlemann, M. Hilty, and L.J. Hathaway. 2014. A point mutation in *cpsE* renders *Streptococcus pneumoniae* nonencapsulated and enhances its growth, adherence and competence. *BMC Microbiol.* 14:210.

Schneider, C.A., W.S. Rasband, and K.W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9:671-675.
Seepersaud, R., R.H. Needham, C.S. Kim, and A.L. Jones. 2006. Abundance of the delta subunit of RNA polymerase is linked to the virulence of *Streptococcus agalactiae*. *J. Bacteriol.* 188:2096-2105.
Selinger, D.S., and W.P. Reed. 1979. Pneumococcal adherence to human epithelial cells. *Infect. Immun.* 23:545-548.
Sellers, R.S. 2016. Translating mouse models: Immune variation and efficacy testing. *Toxicol. Pathol.* 45:134-145.
Shak, J.R., H.P. Ludewick, K.E. Howery, F. Sakai, H. Yi, R.M. Harvey, J.C. Paton, K.P. Klugman, and J.E. Vidal. 2013. Novel role for the *Streptococcus pneumoniae* toxin pneumolysin in the assembly of biofilms. *MBio* 4:e00655-00613.
Shaper, M., S.K. Hollingshead, W.H. Benjamin, Jr., and D.E. Briles. 2004. PspA protects *Streptococcus pneumoniae* from killing by apolactoferrin, and antibody to PspA enhances killing of pneumococci by apolactoferrin *Infect. Immun.* 72:5031-5040.
Silva, I.N., A.S. Ferreira, J.D. Becker, J.E. Zlosnik, D.P. Speert, J. He, D. Mil-Homens, and L.M. Moreira. 2011. Mucoid morphotype variation of *Burkholderia multivorans* during chronic cystic fibrosis lung infection is correlated with changes in metabolism, motility, biofilm formation and virulence. *Microbiology* 157:3124-3137.
Skov Sorensen, U.B., J. Blom, A. Birch-Andersen, and J. Henrichsen. 1988. Ultrastructural localization of capsules, cell wall polysaccharide, cell wall proteins, and F antigen in pneumococci. *Infect. Immun.* 56:1890-1896.
Slager, J., M. Kjos, L. Attaiech, and J.W. Veening. 2014. Antibiotic-induced replication stress triggers bacterial competence by increasing gene dosage near the origin. *Cell* 157:395-406.
Slotved, H.C., T. Dalby, and S. Hoffmann. 2016. The effect of pneumococcal conjugate vaccines on the incidence of invasive pneumococcal disease caused by ten non-vaccine serotypes in Denmark. *Vaccine* 34:769-774.
Sternberg, C., B.B. Christensen, T. Johansen, A. Toftgaard Nielsen, J.B. Andersen, M. Givskov, and S. Molin. 1999. Distribution of bacterial growth activity in flow-chamber biofilms. *Appl. Environ. Microbiol.* 65:4108-4117.
Stevens, K.E., D. Chang, E.E. Zwack, and M.E. Sebert. 2011. Competence in *Streptococcus pneumoniae* is regulated by the rate of ribosomal decoding errors. *MBio* 2:
Stewart, P.S., and J.W. Costerton. 2001. Antibiotic resistance of bacteria in biofilms. *Lancet* 358:135-138.
Stewart, P.S., and M.J. Franklin. 2008. Physiological heterogeneity in biofilms. *Nat. Rev. Microbiol.* 6:199-210.
Stoodley, P., Z. Lewandowski, J.D. Boyle, and H.M. Lappin-Scott. 1998. Oscillation characteristics of biofilm streamers in turbulent flowing water as related to drag and pressure drop. *Biotechnol. Bioeng.* 57:536-544.
Stoodley, P., Z. Lewandowski, J.D. Boyle, and H.M. Lappin-Scott. 1999. Structural deformation of bacterial biofilms caused by short-term fluctuations in fluid shear: an in situ investigation of biofilm rheology. *Biotechnol. Bioeng.* 65:83-92.
Stoodley, P., K. Sauer, D.G. Davies, and J.W. Costerton. 2002. Biofilms as complex differentiated communities. *Annu. Rev. Microbiol.* 56:187-209.
Stoop, E.J., T. Schipper, S.K. Rosendahl Huber, A.E. Nezhinsky, F.J. Verbeek, S.S. Gurcha, G.S. Besra, C.M. Vandenbroucke-Grauls, W. Bitter, and A.M. van der Sar. 2011. Zebrafish embryo screen for mycobacterial genes involved in the initiation of

granuloma formation reveals a newly identified ESX-1 component. *Dis Model Mech* 4:526-536.
Straume, D., G.A. Stamsas, and L.S. Havarstein. 2015. Natural transformation and genome evolution in *Streptococcus pneumoniae*. *Infect Genet Evol* 33:371-380.
Stroeher, U.H., A.W. Paton, A.D. Ogunniyi, and J.C. Paton. 2003. Mutation of *luxS* of *Streptococcus pneumoniae* affects virulence in a mouse model. *Infect. Immun.* 71:3206-3212.
Subramanian, K., D.R. Neill, H.A. Malak, L. Spelmink, S. Khandaker, G. Dalla Libera Marchiori, E. Dearing, A. Kirby, M. Yang, A. Achour, J. Nilvebrant, P.-Å. Nygren, L. Plant, A. Kadioglu, and B. Henriques-Normark. 2019. Pneumolysin binds to the mannose receptor C type 1 (MRC-1) leading to anti-inflammatory responses and enhanced pneumococcal survival. *Nature microbiology* 4:62-70.
Sutherland, I. 2001. Biofilm exopolysaccharides: a strong and sticky framework. *Microbiology* 147:3-9.
Swofford, C.A., N. Van Dessel, and N.S. Forbes. 2015. Quorum-sensing *Salmonella* selectively trigger protein expression within tumors. *Proc Natl Acad Sci U S A* 112:3457-3462.
Syk, A., M. Norman, J. Fernebro, M. Gallotta, S. Farmand, A. Sandgren, S. Normark, and B. Henriques-Normark. 2014. Emergence of hypervirulent mutants resistant to early clearance during systemic serotype 1 pneumococcal infection in mice and humans. *J. Infect. Dis.* 210:4-13.
Szabo, B.G., K.S. Lenart, B. Kadar, A. Gombos, B. Dezsenyi, J. Szanka, I. Bobek, and G. Prinz. 2015. Thousand faces of *Streptococcus pneumoniae* (pneumococcus) infections. *Orv Hetil* 156:1769-1777.
Sztajer, H., A. Lemme, R. Vilchez, S. Schulz, R. Geffers, C.Y. Yip, C.M. Levesque, D.G. Cvitkovitch, and I. Wagner-Dobler. 2008. Autoinducer-2-regulated genes in *Streptococcus mutans* UA159 and global metabolic effect of the *luxS* mutation. *J. Bacteriol.* 190:401-415.
Takeuchi, N., K. Kaneko, and E.V. Koonin. 2014. Horizontal gene transfer can rescue prokaryotes from muller's ratchet: benefit of DNA from dead cells and population subdivision. *G3-Genes Genomes Genetics* 4:325-339.
Talbot, U.M., A.W. Paton, and J.C. Paton. 1996. Uptake of *Streptococcus pneumoniae* by respiratory epithelial cells. *Infect. Immun.* 64:3772-3777.
Tan, A., J.M. Atack, M.P. Jennings, and K.L. Seib. 2016. The capricious nature of bacterial pathogens: phasevarions and vaccine development. *Front Immunol* 7:586.
Tao, L., and T.A. Reese. 2017. Making mouse models that reflect human immune responses. *Trends Immunol.* 38:181-193.
Tart, R.C., L.S. McDaniel, B.A. Ralph, and D.E. Briles. 1996. Truncated *Streptococcus pneumoniae* PspA molecules elicit cross-protective immunity against pneumococcal challenge in mice. *J. Infect. Dis.* 173:380-386.
Taylor, S.D., M.E. Sanders, N.A. Tullos, S.J. Stray, E.W. Norcross, L.S. McDaniel, and M.E. Marquart. 2013. The cholesterol-dependent cytolysin pneumolysin from *Streptococcus pneumoniae* binds to lipid raft microdomains in human corneal epithelial cells. *PLoS One* 8:e61300.
Teichmann, M., H. Dumay-Odelot, and S. Fribourg. 2012. Structural and functional aspects of winged-helix domains at the core of transcription initiation complexes. *Transcription* 3:2-7.
Tettelin, H., K.E. Nelson, I.T. Paulsen, J.A. Eisen, T.D. Read, S. Peterson, J. Heidelberg, R.T. DeBoy, D.H. Haft, R.J. Dodson, A.S. Durkin, M. Gwinn, J.F. Kolonay, W.C. Nelson, J.D. Peterson, L.A. Umayam, O. White, S.L. Salzberg, M.R. Lewis, D.

Radune, E. Holtzapple, H. Khouri, A.M. Wolf, T.R. Utterback, C.L. Hansen, L.A. McDonald, T.V. Feldblyum, S. Angiuoli, T. Dickinson, E.K. Hickey, I.E. Holt, B.J. Loftus, F. Yang, H.O. Smith, J.C. Venter, B.A. Dougherty, D.A. Morrison, S.K. Hollingshead, and C.M. Fraser. 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* 293:498-506.

Tikhomirova, A., C. Trappetti, A.J. Standish, Y. Zhou, J. Breen, S. Pederson, P.S. Zilm, J.C. Paton, and S.P. Kidd. 2018. Specific growth conditions induce a *Streptococcus pneumoniae* non-mucoidal, small colony variant and determine the outcome of its co-culture with *Haemophilus influenzae*. *Pathog Dis* 76:fty074-fty074.

Tilley, S.J., E.V. Orlova, R.J. Gilbert, P.W. Andrew, and H.R. Saibil. 2005. Structural basis of pore formation by the bacterial toxin pneumolysin. *Cell* 121:247-256.

Timby, N., O. Hernell, O. Vaarala, M. Melin, B. Lonnerdal, and M. Domellof. 2015. Infections in infants fed formula supplemented with bovine milk fat globule membranes. *J Pediatr Gastroenterol Nutr* 60:384-389.

Tomasz, A. 1965. Control of the competent state in Pneumococcus by a hormone-like cell product: an example for a new type of regulatory mechanism in bacteria. *Nature* 208:155-159.

Tong, H.H., L.E. Blue, M.A. James, and T.F. DeMaria. 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.

Tovpeko, Y., and D.A. Morrison. 2014. Competence for genetic transformation in *Streptococcus pneumoniae*: mutations in sigmaA bypass the *comW* requirement. *J. Bacteriol.* 196:3724-3734.

Trappetti, C., A. Kadioglu, M. Carter, J. Hayre, F. Iannelli, G. Pozzi, P.W. Andrew, and M.R. Oggioni. 2009. Sialic acid: a preventable signal for pneumococcal biofilm formation, colonization, and invasion of the host. *J. Infect. Dis.* 199:1497-1505.

Trappetti, C., L.J. McAllister, A. Chen, H. Wang, A.W. Paton, M.R. Oggioni, C.A. McDevitt, and J.C. Paton. 2017. Autoinducer 2 signaling via the phosphotransferase FruA drives galactose utilization by *Streptococcus pneumoniae*, resulting in hypervirulence. *MBio* 8:

Tsai, C.J., J.M. Loh, and T. Proft. 2016. *Galleria mellonella* infection models for the study of bacterial diseases and for antimicrobial drug testing. *Virulence* 7:214-229.

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

Turner, P., C. Turner, A. Jankhot, N. Helen, S.J. Lee, N.P. Day, N.J. White, F. Nosten, and D. Goldblatt. 2012. A longitudinal study of *Streptococcus pneumoniae* carriage in a cohort of infants and their mothers on the Thailand-Myanmar border. *PLoS One* 7:e38271.

Usuf, E., C. Bottomley, R.A. Adegbola, and A. Hall. 2014. Pneumococcal carriage in sub-Saharan Africa--a systematic review. *PLoS One* 9:e85001.

van Lieshout, M.H.P., A.F. de Vos, M.C. Dessing, A.P.N.A. de Porto, O.J. de Boer, R. de Beer, S. Terpstra, S. Florquin, C. van't Veer, and T. van der Poll. 2018. ASC and NLRP3 impair host defense during lethal pneumonia caused by serotype 3 *Streptococcus pneumoniae* in mice. *Eur. J. Immunol.* 48:66-79.

Vandevelde, N.M., P.M. Tulkens, Y. Diaz Iglesias, J. Verhaegen, H. Rodriguez-Villalobos, I. Philippart, J. Cadrobbi, N. Coppens, A. Boel, K. Van Vaerenbergh, H. Francart, R. Vanhoof, G. Liistro, P. Jordens, J.P. d'Odemont, Y. Valcke, F. Verschuren, and F. Van Bambeke. 2014. Characterisation of a collection of *Streptococcus*

*pneumoniae* isolates from patients suffering from acute exacerbations of chronic bronchitis: *in vitro* susceptibility to antibiotics and biofilm formation in relation to antibiotic efflux and serotypes/serogroups. *Int. J. Antimicrob. Agents* 44:209-217.

Vela Coral, M.C., N. Fonseca, E. Castaneda, J.L. Di Fabio, S.K. Hollingshead, and D.E. Briles. 2001. Pneumococcal surface protein A of invasive *Streptococcus pneumoniae* isolates from Colombian children. *Emerg Infect Dis* 7:832-836.

Vidal, J.E., K.E. Howery, H.P. Ludewick, P. Nava, and K.P. Klugman. 2013. Quorum-sensing systems LuxS/autoinducer 2 and Com regulate *Streptococcus pneumoniae* biofilms in a bioreactor with living cultures of human respiratory cells. *Infect. Immun.* 81:1341-1353.

Vidal, J.E., H.P. Ludewick, R.M. Kunkel, D. Zahner, and K.P. Klugman. 2011. The LuxS-dependent quorum-sensing system regulates early biofilm formation by *Streptococcus pneumoniae* strain D39. *Infect. Immun.* 79:4050-4060.

Wah, D.A., J.A. Hirsch, L.F. Dorner, I. Schildkraut, and A.K. Aggarwal. 1997. Structure of the multimodular endonuclease FokI bound to DNA. *Nature* 388:97-100.

Wahl, B., K.L. O'Brien, A. Greenbaum, A. Majumder, L. Liu, Y. Chu, I. Luksic, H. Nair, D.A. McAllister, H. Campbell, I. Rudan, R. Black, and M.D. Knoll. 2018. Burden of *Streptococcus pneumoniae* and Haemophilus influenzae type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000-15. *Lancet Glob Health* 6:e744-e757.

Waite, R.D., J.K. Struthers, and C.G. Dowson. 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.

Walker, J.A., R.L. Allen, P. Falmagne, M.K. Johnson, and G.J. Boulnois. 1987. Molecular cloning, characterization, and complete nucleotide sequence of the gene for pneumolysin, the sulfhydryl-activated toxin of *Streptococcus pneumoniae*. *Infect. Immun.* 55:1184-1189.

Wand, M.E., C.M. Muller, R.W. Titball, and S.L. Michell. 2011. Macrophage and *Galleria mellonella* infection models reflect the virulence of naturally occurring isolates of *B. pseudomallei, B. thailandensis* and *B. oklahomensis*. *BMC Microbiol.* 11:11.

Wang, S., J. Ma, and J. Xu. 2016. AUCpreD: proteome-level protein disorder prediction by AUC-maximized deep convolutional neural fields. *Bioinformatics* 32:i672-i679.

Wang, S., S. Sun, Z. Li, R. Zhang, and J. Xu. 2017. Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Comput Biol* 13:e1005324.

Watson, D.A., and D.M. Musher. 1990. Interruption of capsule production in *Streptococcus pneumonia* serotype 3 by insertion of transposon Tn916. *Infect. Immun.* 58:3135-3138.

Weinberger, D.M., K. Trzcinski, Y.J. Lu, D. Bogaert, A. Brandes, J. Galagan, P.W. Anderson, R. Malley, and M. Lipsitch. 2009. Pneumococcal capsular polysaccharide structure predicts serotype prevalence. *PLoS Pathog* 5:e1000476.

Weiser, J.N. 2010. The pneumococcus: why a commensal misbehaves. *J Mol Med (Berl)* 88:97-102.

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

Weiser, J.N., D. Bae, H. Epino, S.B. Gordon, M. Kapoor, L.A. Zenewicz, and M. Shchepetov. 2001. Changes in availability of oxygen accentuate differences in capsular polysaccharide expression by phenotypic variants and clinical isolates of *Streptococcus pneumoniae*. *Infect. Immun.* 69:5430-5439.

Weiss, A., J.A. Ibarra, J. Paoletti, R.K. Carroll, and L.N. Shaw. 2014. The delta subunit of RNA polymerase guides promoter selectivity and virulence in *Staphylococcus aureus*. *Infect. Immun.* 82:1424-1435.
Weiss, A., and L.N. Shaw. 2015. Small things considered: the small accessory subunits of RNA polymerase in Gram-positive bacteria. *FEMS Microbiol. Rev.* 39:541-554.
Weng, L., A. Piotrowski, and D.A. Morrison. 2013. Exit from competence for genetic transformation in *Streptococcus pneumoniae* is regulated at multiple levels. *PLoS One* 8:e64197.
Weyder, M., M. Prudhomme, M. Berge, P. Polard, and G. Fichant. 2018. Dynamic modeling of *Streptococcus pneumoniae* competence provides regulatory mechanistic insights into its tight temporal regulation. *Front Microbiol* 9:1637.
WHO. 2007. Pneumococcal conjugate vaccine for childhood immunization--WHO position paper. *Wkly Epidemiol Rec* 82:93-104.
Wimpenny, J.W.T., and R. Colasanti. 1997. A unifying hypothesis for the structure of microbial biofilms based on cellular automaton models. *FEMS Microbiol. Ecol.* 22:1-16.
Wisniewski-Dye, F., and L. Vial. 2008. Phase and antigenic variation mediated by genome modifications. *Antonie Van Leeuwenhoek* 94:493-515.
Witzenrath, M., F. Pache, D. Lorenz, U. Koppe, B. Gutbier, C. Tabeling, K. Reppe, K. Meixenberger, A. Dorhoi, J. Ma, A. Holmes, G. Trendelenburg, M.M. Heimesaat, S. Bereswill, M. van der Linden, J. Tschopp, T.J. Mitchell, N. Suttorp, and B. Opitz. 2011. The NLRP3 inflammasome is differentially activated by pneumolysin variants and contributes to host defense in pneumococcal pneumonia. *J. Immunol.* 187:434-440.
Wolcott, R.D., and G.D. Ehrlich. 2008. Biofilms and chronic infections. *JAMA* 299:2682-2684.
Wolter, D.J., J.C. Emerson, S. McNamara, A.M. Buccat, X. Qin, E. Cochrane, L.S. Houston, G.B. Rogers, P. Marsh, K. Prehar, C.E. Pope, M. Blackledge, E. Deziel, K.D. Bruce, B.W. Ramsey, R.L. Gibson, J.L. Burns, and L.R. Hoffman. 2013. *Staphylococcus aureus* small-colony variants are independently associated with worse lung disease in children with cystic fibrosis. *Clin. Infect. Dis.* 57:384-391.
Wood, T.K., S.J. Knabel, and B.W. Kwan. 2013. Bacterial persister cell formation and dormancy. *Appl. Environ. Microbiol.* 79:7116-7121.
Wrangstadh, M., U. Szewzyk, J. Ostling, and S. Kjelleberg. 1990. Starvation-specific formation of a peripheral exopolysaccharide by a marine *Pseudomonas sp*., strain S9. *Appl. Environ. Microbiol.* 56:2065-2072.
Wu, S., and Y. Zhang. 2007. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res.* 35:3375-3382.
Wyllie, A.L., M.L. Chu, M.H. Schellens, J. van Engelsdorp Gastelaars, M.D. Jansen, A. van der Ende, D. Bogaert, E.A. Sanders, and K. Trzcinski. 2014. *Streptococcus pneumoniae* in saliva of Dutch primary school children. *PLoS One* 9:e102045.
Xie, X., H. Zhang, Y. Zheng, A. Li, M. Wang, H. Zhou, X. Zhu, Z. Schneider, L. Chen, B.N. Kreiswirth, and H. Du. 2016. RpoE is a putative antibiotic resistance regulator of *Salmonella enteric* serovar Typhi. *Curr. Microbiol.* 72:457-464.
Xu, P., X. Ge, L. Chen, X. Wang, Y. Dou, J.Z. Xu, J.R. Patel, V. Stone, M. Trinh, K. Evans, T. Kitten, D. Bonchev, and G.A. Buck. 2011. Genome-wide essential gene identification in *Streptococcus sanguinis*. *Sci Rep* 1:125.
Xue, X., J. Li, W. Wang, H. Sztajer, and I. Wagner-Dobler. 2012. The global impact of the delta subunit RpoE of the RNA polymerase on the proteome of *Streptococcus mutans*. *Microbiology* 158:191-206.

Xue, X., H. Sztajer, N. Buddruhs, J. Petersen, M. Rohde, S.R. Talay, and I. Wagner-Dobler. 2011. Lack of the delta subunit of RNA polymerase increases virulence related traits of *Streptococcus mutans*. *PLoS One* 6:e20075.
Xue, X., J. Tomasch, H. Sztajer, and I. Wagner-Dobler. 2010. The delta subunit of RNA polymerase, RpoE, is a global modulator of *Streptococcus mutans* environmental adaptation. *J. Bacteriol.* 192:5081-5092.
Yadav, M.K., S.W. Chae, and J.J. Song. 2012. *In vitro Streptococcus pneumoniae* biofilm formation and *in vivo* middle ear mucosal biofilm in a rat model of acute otitis induced by *S. pneumoniae*. *Clin Exp Otorhinolaryngol* 5:139-144.
Yadav, M.K., Y.Y. Go, S.H. Kim, S.W. Chae, and J.J. Song. 2017. Antimicrobial and antibiofilm effects of human amniotic/chorionic membrane extract on *Streptococcus pneumoniae*. *Front Microbiol* 8:1948.
Yadav, M.K., J.E. Vidal, Y.Y. Go, S.H. Kim, S.W. Chae, and J.J. Song. 2018. The LuxS/AI-2 quorum-sensing system of *Streptococcus pneumoniae* is required to cause disease, and to regulate virulence- and metabolism-related genes in a rat model of middle ear infection. *Front Cell Infect Microbiol* 8:138.
Yang, J., A. Roy, and Y. Zhang. 2013. BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Res.* 41:D1096-1103.
Yang, J., R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang. 2015. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* 12:7-8.
Yang, J., and Y. Zhang. 2015. Protein structure and function prediction using I-TASSER. *Curr Protoc Bioinformatics* 52:5 8 1-15.
Yang, L., K.B. Barken, M.E. Skindersoe, A.B. Christensen, M. Givskov, and T. Tolker-Nielsen. 2007. Effects of iron on DNA release and biofilm development by *Pseudomonas aeruginosa*. *Microbiology* 153:1318-1328.
Ye, J., G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, and T.L. Madden. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13:134.
Yother, J., L.S. McDaniel, and D.E. Briles. 1986. Transformation of encapsulated *Streptococcus pneumoniae*. *J. Bacteriol.* 168:1463-1465.
Zafar, M.A., Y. Wang, S. Hamaguchi, and J.N. Weiser. 2017. Host-to-host transmission of *Streptococcus pneumoniae* is driven by its inflammatory toxin, pneumolysin. *Cell Host & Microbe* 21:73-83.
Zeng, M., E. Nourishirazi, E. Guinet, and M. Nouri-Shirazi. 2016. The genetic background influences the cellular and humoral immune responses to vaccines. *Clin. Exp. Immunol.* 186:190-204.
Zevian, S.C., and J.L. Yanowitz. 2014. Methodological considerations for heat shock of the nematode *Caenorhabditis elegans*. *Methods* 68:450-457.
Zhang, Y., and J. Skolnick. 2004. SPICKER: a clustering approach to identify near-native protein folds. *J Comput Chem* 25:865-871.
Zhang, Y., and J. Skolnick. 2005. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 33:2302-2309.

# Chapter 8: Appendices

### 8.1.1 Bacterial CFU to Optical Density Calculations.

Bacterial growth curves were generated in order to enable quantification of bacterial cultures through optical density readings.

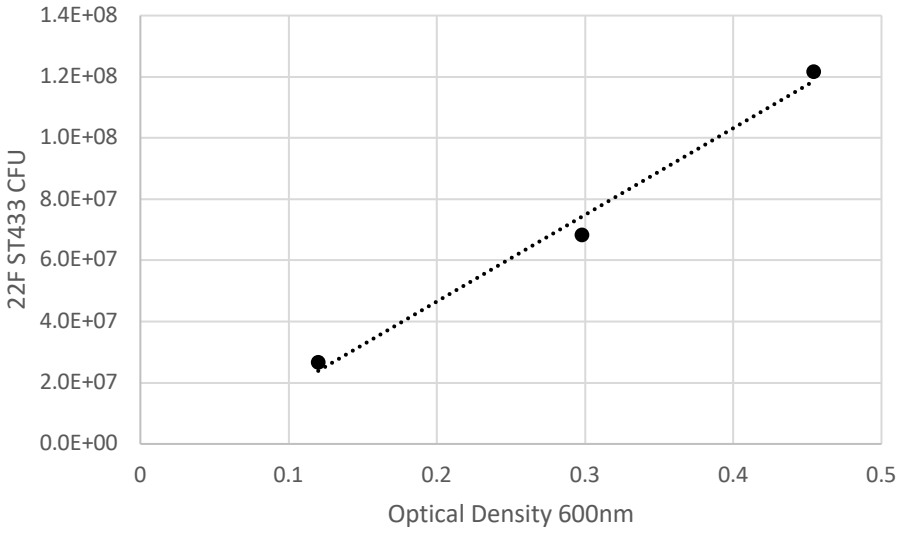



Figure 64: Colony forming unit enumeration and regression analysis of *S. pneumoniae* 22F WT at optical density 600 nm. N = 1.

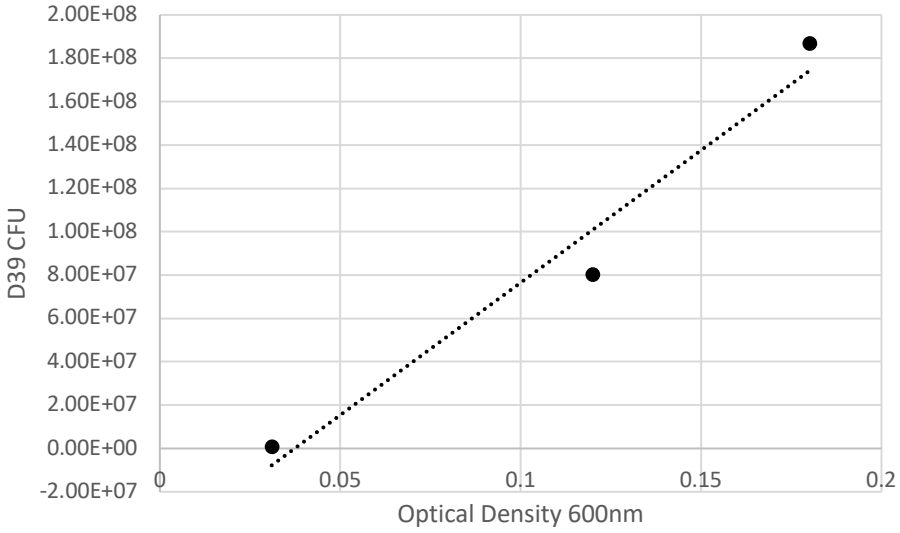



Figure 65: Colony forming unit enumeration and regression analysis of *S. pneumoniae* D39 at optical density 600 nm N = 1.

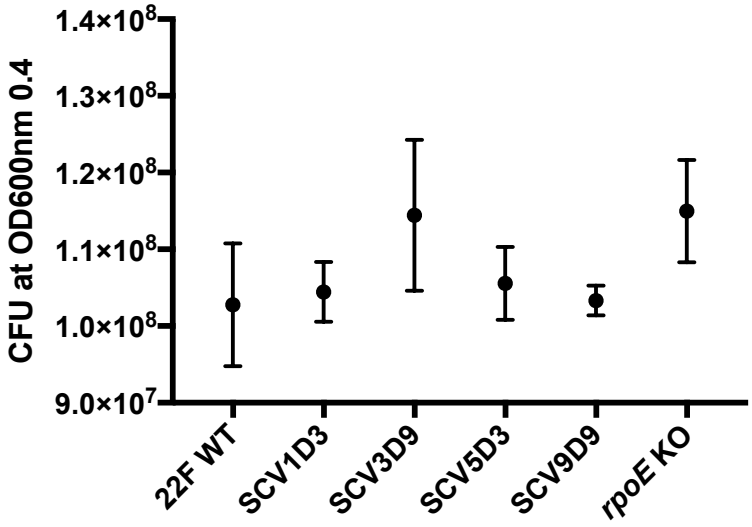



Figure 66: CFU enumeration for *S. pneumoniae* isolates at optical densities of 0.4 read at 600 nm. No significant differences were observed between isolates. Protocol carried out in triplicate with 3 independent biological replicates.

## 8.2 DNA Constructs

### 8.2.1 Primers

| Primer | Nucleotide Sequence |
|---|---|
| *rpoE* Forward | GAG-GAG-AAA-CGC-TTT-GGA-ATT-AGA-AG |
| *rpoE* Reverse | GCT-AAC-TCT-TAT-TCC-TCG-CTG-GTT-TC |
| D39 *rpoE* knock-out construct Forward | GCG-ACC-TTG-GAA-GCC-ATG-TA |
| D39 *rpoE* knock-out construct Reverse | CCT-AGA-CTC-GCT-GCC-ACA-AT |
| 22F *rpoE* knock-out construct Forward | CAG-TCA-AGC-CGA-CCT-TGG-AA |
| KAN cassette Forward primer | GAT-ACT-GAG-GGT-ATG-TCG-CCC |
| KAN cassette Reverse primer | AAG-TCG-AAA-AAG-AGC-TCG-ACG |
| KAN cassette with 3' flank Forward | ACG-TCG-AGC-TCT-TTT-TCG-ACT |
| KAN cassette with 3' flanks Reverse | CAA-AAA-CTT-CCC-CGT-GCT-GG |

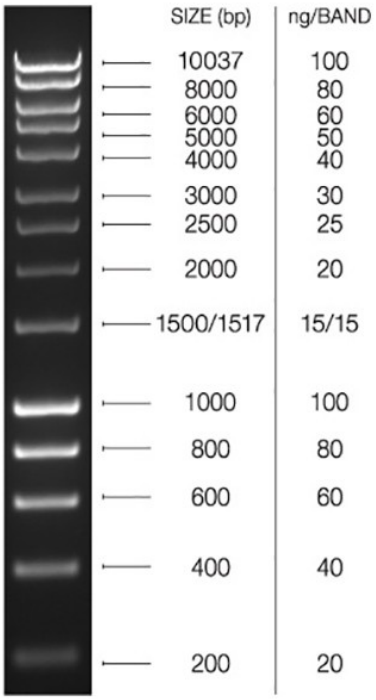



Table 9: Bioline Hyperladder 1 molecular weight ladder. PCR products are compared to this ladder to provide an indication on the weight

### 8.2.2 D39 rpoE knock-out construct

| D39 *rpoE* Knock-out Left Flank-500bp |
| --- |
| CCTCGTAGATGGGGTGAAGCTTCCTTACAGTCAAGCGACCTTGGAAGCCATG<br>TATTCCTATTTGGATCGGCATGGTGAGCTTTTTTGGATTGAAGTCAAGGAGA<br>AGGGAGAATGGTTTCCAATTGGGGATATTACACTATCTCAGGATAATCTCCC<br>CATTGTGATTGGGAATTCCGCTTACCAACATCGAGGACTTGGAAAAAAGATT<br>CTAAGTGCTTTGATTGAATTGGCTCGAGTAAAAGGATGGAAAGAATTGAGA<br>GTCAAGGAAATCTACACCTACAATCATGCTTCTAGGAGGTGTTTCAAGTCGC<br>TTGGATTTGTGGAAAATGGAGCAACAGAAAAAGGAAGGAGTTTTATATTGG<br>AATTAGTCTAATCCTGCCACTTTCTCCCCCTTAACCTTTGACTATTCCGCAAA<br>ATTATCGTAAAATAAAGAGTAAATGATAAAATGAGGTCAGAGTCTGTTCGCT<br>CTGGCGATAGTAGTATAAATGAGGAGAAACGCT |

| *rpoE* allele 1 – As per D39 Genome gb\|CP000410.1 |
| --- |
| TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAACTATCTATGATTG<br>AGGTAGCGCGTGCTATCTTGGAACTTCGTGGTCGCGATCATGAGATGCATTT<br>TAGCGATCTTGTAAACGAAATTCAAAACTACCTTGGAACATCAAACAGCGAT<br>ATCCGCGAAGCTTTGCCTTTGTTCTACACAGAGTTGAACTTTGACGGTAGCTT<br>CATCTCACTTGGGGACAACAAATGGGGTCTTCGTTCATGGTATGGTGTGGAC<br>GAAATCGACGAAGAAATCATCGCTCTTGAAGAAAATGACGACGATGAAGTA<br>GCACCAAAAGCTAAGAAAAAACGTGTCAATGCCTTTATGGATGGTGATTCA<br>GATGCCATTGACTACAATGCAGATGATCCAGAAGACGAAGATGCATACGAA<br>GCAGATCCAGCTCTTTCATACGATGATGAAAATCCAGATGATGAAAAAAAT<br>GAAGTGGAAGCTTATGATGCAGAAATCAACGAAATCGCTCCAGATGACTTG<br>GGAGAAGATGTGGATCTCAACGAAGACGACGACGAGTTTTCAGATGACGAC<br>GCTGAAACCAGCGAGGAATAA |

| Kanamycin Cassette – This kanamycin resistance gene sequence was kindly provided by Dr Ray Allan, and is based on the apha3 resistance gene (Caillaud et al., 1987). |
| --- |
| ATGGCCAAAATGCGCATTAGTCCGGAACTGAAAAAATTGATTGAAAAGTAC<br>CGCTGTGTCAAAGATACTGAGGGTATGTCGCCCGCCAAAGTCTACAAGCTGG<br>TCGGCGAAAATGAAAACCTGTACTTGAAAATGACAGATAGTCGCTACAAAG |

GCACCACCTACGACGTAGAGCGCGAGAAAGATATGATGTTATGGCTGGAAG
GTAAACTGCCTGTTCCGAAAGTTCTGCATTTCGAACGCCACGACGGTTGGAG
CAACCTGCTGATGTCGGAAGCAGATGGCGTATTGTGTAGCGAAGAATACGA
AGACGAACAATCGccgGAGAAAATCATCGAATTGTACGCGGAATGCATCcgcT
TGTTTCACAGCATCGACATCAGTGATTGCCCTTACACCAACTCCTTAGATAG
CCGCCTGGCTGAACTTGATTATTTGTTGAATAACGACTTGGCTGATGTAGAC
TGCGAAAACTGGGAGGAAGATACACCCTTCAAGGACCCGCGCGAGCTCTAC
GACTTTCTGAAAACTGAAAAACCGGAAGAAGAGCTGGTTTTCTCCCACGGCG
ATCTGGGCGACTCGAATATTTTCGTAAAAGATGGCAAAGTTTCCGGCTTTAT
CGACTTGGGCCGCAGCGGGCGCGCAGATAAATGGTACGACATTGCGTTCTGC
GTCCGCAGCATCCGTGAAGATATCGGTGAAGAACAATACGTCGAGCTCTTTT
TCGACTTGCTGGGTATTAAACCGGATTGGGAAAAAATCAAATATTATATCCT
GTTGGATGAATTATTCTAG

| Right Flank-500bp |
| --- |
| GAGTTAGCTATTGACAATTATCCTATTTTTAGGTATCATATTGTTTGGGCACC |
| TCTTTTAGAGGTCGGGGCTCCCTAGTTCTTAGGGAGCTATTTTTGTTTTTTCA |
| AGAAGTTATCTTCTTGTATTTTATACTCAATGAAAATCAAAGTGCAAGCTAG |
| GAAACTAGCCGTAGGCTGCTCAAAACACTGTTTTGAGGTTGTAGATAAGACT |
| GACAAAGTCAGGAACACATATCTACGGCAAGGCGACGTTGACGCGGTTTGA |
| AGAGATTTTCGAAGAGTATTAGTTGTGAATCTGGTGCAGTCGTCCCAGATTA |
| TTCTTATTAGTAGGGTCTTGTTTTCTATATCCCCTCGTAGTTAACAAGACCTT |
| GAGCATTTTAGAAAGAGGAATCTATGTCTACGAAATATATTTTTGTAACTGG |
| TGGTGTGGTATCGTCTATTGGGAAAGGGATTGTGGCAGCGAGTCTAGGCCGT |
| CTCTTGAAAAATCGTGGTCTCAAAGTAACC |

| D39 Complete Construct |
| --- |
| CCTCGTAGATGGGGTGAAGCTTCCTTACAGTCAAGCGACCTTGGAAGCCATG |
| TATTCCTATTTGGATCGGCATGGTGAGCTTTTTTGGATTGAAGTCAAGGAGA |
| AGGGAGAATGGTTTCCAATTGGGGATATTACACTATCTCAGGATAATCTCCC |
| CATTGTGATTGGGAATTCCGCTTACCAACATCGAGGACTTGGAAAAAAGATT |
| CTAAGTGCTTTGATTGAATTGGCTCGAGTAAAAGGATGGAAAGAATTGAGA |
| GTCAAGGAAATCTACACCTACAATCATGCTTCTAGGAGGTGTTTCAAGTCGC |

TTGGATTTGTGGAAAATGGAGCAACAGAAAAAGGAAGGAGTTTTATATTGG
AATTAGTCTAATCCTGCCACTTTCTCCCCCTTAACCTTTGACTATTCCGCAAA
ATTATCGTAAAATAAAGAGTAAATGATAAAATGAGGTCAGAGTCTGTTCGCT
CTGGCGATAGTAGTATAAATGAGGAGAAACGCTGGGCCCAAAATTTGTTTG
ATTTGTATCTTAAAATTTTGTATAATAGGAATTGAAGTTATCTAGAAAGGGG
AAAATTATGGCCAAAATGCGCATTAGTCCGGAACTGAAAAAATTGATTGAA
AAGTACCGCTGTGTCAAAGATACTGAGGGTATGTCGCCCGCCAAAGTCTACA
AGCTGGTCGGCGAAAATGAAAACCTGTACTTGAAAATGACAGATAGTCGCT
ACAAAGGCACCACCTACGACGTAGAGCGCGAGAAAGATATGATGTTATGGC
TGGAAGGTAAACTGCCTGTTCCGAAAGTTCTGCATTTCGAACGCCACGACGG
TTGGAGCAACCTGCTGATGTCGGAAGCAGATGGCGTATTGTGTAGCGAAGA
ATACGAAGACGAACAATCGCCGGAGAAAATCATCGAATTGTACGCGGAATG
CATCCGCTTGTTTCACAGCATCGACATCAGTGATTGCCCTTACACCAACTCCT
TAGATAGCCGCCTGGCTGAACTTGATTATTTGTTGAATAACGACTTGGCTGA
TGTAGACTGCGAAAACTGGGAGGAAGATACACCCTTCAAGGACCCGCGCGA
GCTCTACGACTTTCTGAAAACTGAAAAACCGGAAGAAGAGCTGGTTTTCTCC
CACGGCGATCTGGGCGACTCGAATATTTTCGTAAAAGATGGCAAAGTTTCCG
GCTTTATCGACTTGGGCCGCAGCGGGCGCGCAGATAAATGGTACGACATTGC
GTTCTGCGTCCGCAGCATCCGTGAAGATATCGGTGAAGAACAATACGTCGAG
CTCTTTTTCGACTTGCTGGGTATTAAACCGGATTGGGAAAAAATCAAATATT
ATATCCTGTTGGATGAATTATTCTAGGAGTTAGCTATTGACAATTATCCTATT
TTTAGGTATCATATTGTTTGGGCACCTCTTTTAGAGGTCGGGGCTCCCTAGTT
CTTAGGGAGCTATTTTTGTTTTTTCAAGAAGTTATCTTCTTGTATTTTATACTC
AATGAAAATCAAAGTGCAAGCTAGGAAACTAGCCGTAGGCTGCTCAAAACA
CTGTTTTGAGGTTGTAGATAAGACTGACAAAGTCAGGAACACATATCTACGG
CAAGGCGACGTTGACGCGGTTTGAAGAGATTTTCGAAGAGTATTAGTTGTGA
ATCTGGTGCAGTCGTCCCAGATTATTCTTATTAGTAGGGTCTTGTTTTCTATA
TCCCCTCGTAGTTAACAAGACCTTGAGCATTTTAGAAAGAGGAATCTATGTC
TACGAAATATATTTTTGTAACTGGTGGTGTGGTATCGTCTATTGGGAAAGGG
ATTGTGGCAGCGAGTCTAGGCCGTCTCTTGAAAAATCGTGGTCTCAAAGTAA
CC

### 8.2.3 22F ST433 3298 rpoE knock-out construct

| 5’ *rpoE* Flanking 500 bp |
|---|
| CTCGTAGATGGGGTGAAGCTTCCTTACAGTCAAGCCGACCTTGGAAGCCATG<br>TATTCCTATTTGGATCGGCATGGTGAGCTTTTTTGGATTGAAGTCAAGGAGA<br>AGGGAGAATGGTTTCCAATTGGGGATATTACACTATCTCAGGATAATCTCCC<br>CATTGTGATTGGGAATTCCGCTTACCAACATCGAGGACTTGGAAAAAAGATT<br>CTAAGTGCTTTGATTGAATTGGCTCGAGTAAAAGGATGGAAAGAATTGAGA<br>GTCAAGAAAATCTACACCTACAATCATGCTTCTAGGAGGTGTTTCAAGTCGC<br>TTGGATTTGTGGAAAATGGAGCAACAGAAAAAGGAATGAGTTTTATATTGG<br>AATTAGTCTAATCCTGCCACTTTCTCCCCCTTAACCTTTGACTATTCCGCAAA<br>ATTATCGTAAAATAAAGAGTAAATGATAAAATGAGGTCAGAGTCTGTTCGCT<br>CTGGCGATAGTAGTATAAATGAGGAGAAACGCT |

| 3’ *rpoE* Flanking 500 bp |
|---|
| GAGTTAGCTATTGACAATTATCCTATTTTTAGGTATCATATTGTTTGGGCACC<br>TCTTTTAGAGGTCGGGGCTCCCTAGTTCTTAGGGAGCTATTTTTGTTTTTTCA<br>AGAAGTTATCTTCTTGTATTTTATACTCAATGAAAATCAAAGAGCAAGCTAG<br>GAAACTAGCCGTAGGCTGCTCAAAACACTGTTTTGAGGTTGTAGATAAGACT<br>GACAAAGTCAGGAACACATATCTACGGCAAGGCGACGTTGACGCGGTTTGA<br>AGAGATTTTCGAAGAGTATTAGTTGTGAATCTGGTGCAGTCGTCCCAGATTA<br>TTCTTATTAGTAGGGTCTTGTTTTCTATATCCCCTCGTAGTTAACAAGACCTT<br>GAGCATTTTAGAAAGAGGAATCTATGTCTACGAAATATATTTTTGTAACTGG<br>TGGTGTGGTATCGTCCATTGGGAAAGGGATTGTGGCAGCGAGTCTAGGCCGT<br>CTCTTGAAAAATCGTGGTCTCAAAGTAACC |

| Kanamycin Cassette 795 bp |
|---|
| ATGGCCAAAATGCGCATTAGTCCGGAACTGAAAAAATTGATTGAAAAGTAC<br>CGCTGTGTCAAAGATACTGAGGGTATGTCGCCCGCCAAAGTCTACAAGCTGG<br>TCGGCGAAAATGAAAACCTGTACTTGAAAATGACAGATAGTCGCTACAAAG<br>GCACCACCTACGACGTAGAGCGCGAGAAAGATATGATGTTATGGCTGGAAG<br>GTAAACTGCCTGTTCCGAAAGTTCTGCATTTCGAACGCCACGACGGTTGGAG<br>CAACCTGCTGATGTCGGAAGCAGATGGCGTATTGTGTAGCGAAGAATACGA<br>AGACGAACAATCGCCGGAGAAAATCATCGAATTGTACGCGGAATGCATCCG |

CTTGTTTCACAGCATCGACATCAGTGATTGCCCTTACACCAACTCCTTAGATA
GCCGCCTGGCTGAACTTGATTATTTGTTGAATAACGACTTGGCTGATGTAGA
CTGCGAAAACTGGGAGGAAGATACACCCTTCAAGGACCCGCGCGAGCTCTA
CGACTTTCTGAAAACTGAAAAACCGGAAGAAGAGCTGGTTTTCTCCCACGGC
GATCTGGGCGACTCGAATATTTTCGTAAAAGATGGCAAAGTTTCCGGCTTTA
TCGACTTGGGCCGCAGCGGGCGCGCAGATAAATGGTACGACATTGCGTTCTG
CGTCCGCAGCATCCGTGAAGATATCGGTGAAGAACAATACGTCGAGCTCTTT
TTCGACTTGCTGGGTATTAAACCGGATTGGGAAAAAATCAAATATTATATCC
TGTTGGATGAATTATTCTAG

| Constitutive Promoter |
| --- |
| GGGCCCAAAATTTGTTTGATTTGTATCTTAAAATTTTGTATAATAGGAATTGA<br>AGTTATCTAGAAAGGGGAAAATT |

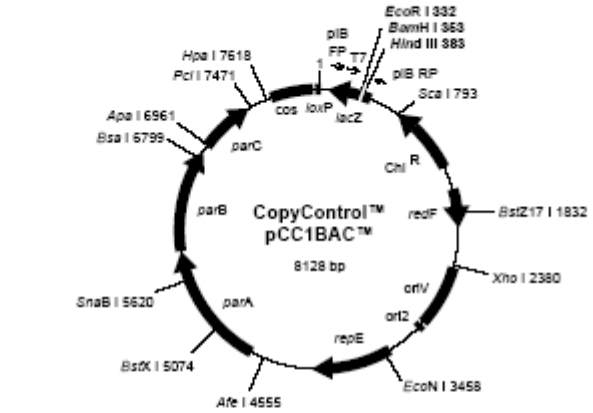
| 22F *rpoE* knock-out complete construct |
| --- |
| CTCGTAGATGGGGTGAAGCTTCCTTACAGTCAAGCCGACCTTGGAAGCCATG<br>TATTCCTATTTGGATCGGCATGGTGAGCTTTTTTGGATTGAAGTCAAGGAGA<br>AGGGAGAATGGTTTCCAATTGGGGATATTACACTATCTCAGGATAATCTCCC<br>CATTGTGATTGGGAATTCCGCTTACCAACATCGAGGACTTGGAAAAAAGATT<br>CTAAGTGCTTTGATTGAATTGGCTCGAGTAAAAGGATGGAAAGAATTGAGA<br>GTCAAGAAAATCTACACCTACAATCATGCTTCTAGGAGGTGTTTCAAGTCGC<br>TTGGATTTGTGGAAAATGGAGCAACAGAAAAAGGAATGAGTTTTATATTGG<br>AATTAGTCTAATCCTGCCACTTTCTCCCCCTTAACCTTTGACTATTCCGCAAA<br>ATTATCGTAAAATAAAGAGTAAATGATAAAATGAGGTCAGAGTCTGTTCGCT<br>CTGGCGATAGTAGTATAAATGAGGAGAAACGCTGGGCCCAAAATTTGTTTG<br>ATTTGTATCTTAAAATTTTGTATAATAGGAATTGAAGTTATCTAGAAAGGGG<br>AAAATTATGGCCAAAATGCGCATTAGTCCGGAACTGAAAAAATTGATTGAA<br>AAGTACCGCTGTGTCAAAGATACTGAGGGTATGTCGCCCGCCAAAGTCTACA<br>AGCTGGTCGGCGAAAATGAAAACCTGTACTTGAAAATGACAGATAGTCGCT<br>ACAAAGGCACCACCTACGACGTAGAGCGCGAGAAAGATATGATGTTATGGC<br>TGGAAGGTAAACTGCCTGTTCCGAAAGTTCTGCATTTCGAACGCCACGACGG<br>TTGGAGCAACCTGCTGATGTCGGAAGCAGATGGCGTATTGTGTAGCGAAGA<br>ATACGAAGACGAACAATCGCCGGAGAAAATCATCGAATTGTACGCGGAATG |

CATCCGCTTGTTTCACAGCATCGACATCAGTGATTGCCCTTACACCAACTCCT
TAGATAGCCGCCTGGCTGAACTTGATTATTTGTTGAATAACGACTTGGCTGA
TGTAGACTGCGAAAACTGGGAGGAAGATACACCCTTCAAGGACCCGCGCGA
GCTCTACGACTTTCTGAAAACTGAAAAACCGGAAGAAGAGCTGGTTTTCTCC
CACGGCGATCTGGGCGACTCGAATATTTTCGTAAAAGATGGCAAAGTTTCCG
GCTTTATCGACTTGGGCCGCAGCGGGCGCGCAGATAAATGGTACGACATTGC
GTTCTGCGTCCGCAGCATCCGTGAAGATATCGGTGAAGAACAATACGTCGAG
CTCTTTTTCGACTTGCTGGGTATTAAACCGGATTGGGAAAAAATCAAATATT
ATATCCTGTTGGATGAATTATTCTAGGAGTTAGCTATTGACAATTATCCTATT
TTTAGGTATCATATTGTTTGGGCACCTCTTTTAGAGGTCGGGGCTCCCTAGTT
CTTAGGGAGCTATTTTTGTTTTTTCAAGAAGTTATCTTCTTGTATTTTATACTC
AATGAAAATCAAAGAGCAAGCTAGGAAACTAGCCGTAGGCTGCTCAAAACA
CTGTTTTGAGGTTGTAGATAAGACTGACAAAGTCAGGAACACATATCTACGG
CAAGGCGACGTTGACGCGGTTTGAAGAGATTTTCGAAGAGTATTAGTTGTGA
ATCTGGTGCAGTCGTCCCAGATTATTCTTATTAGTAGGGTCTTGTTTTCTATA
TCCCCTCGTAGTTAACAAGACCTTGAGCATTTTAGAAAGAGGAATCTATGTC
TACGAAATATATTTTTGTAACTGGTGGTGTGGTATCGTCCATTGGGAAAGGG
ATTGTGGCAGCGAGTCTAGGCCGTCTCTTGAAAAATCGTGGTCTCAAAGTAA
CC

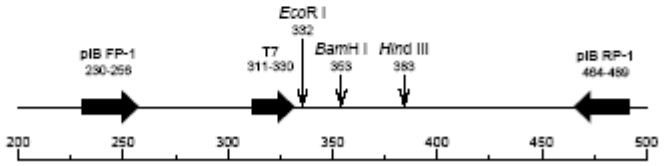
### 8.2.4 pCC1BAC Vector Map

DNA constructs will be inserted into pCC1BAC and transformed into *E. coli*.

**pCC1BAC™ Vector Map.**




Note: Not all restriction enzymes that cut only once are indicated above.
See Appendix E (pg. 15) for complete restriction information.
Primers are not drawn to scale.



### 8.2.5 rpoE complementation plasmid

| Tetracycline Resistance gene *tet*(M) 1994 bp – GenBank: AY466395.1 (Del Grosso et al., 2004) |
|---|
| CACATGAAAATTATTAATATTGGAGTTTTAGCTCATGTTGATGCGGGAAAAA |
| CTACCTTAACAGAAAGCTTATTATATAACAGTGGAGCGATTACAGAATTAGG |
| AAGCGTGGACAGAGGTACAACGAAAACGGATAATACGCTTTTAGAACGTCA |
| GAGAGGAATTACAATTCAGACAGGAATAACCTCTTTTCAGTGGGAAAATAC |
| GAAGGTGAACATCATAGACACGCCAGGACATATGGATTTCTTAGCAGAAGT |
| ATATCGTTCATTATCAGTTTTAGATGGGGCAATTCTACTGATTTCTGCAAAAG |
| ATGGCGTACAAGCACAAACTCGTATATTGTTTCATGCACTTAGGAAAATAGG |
| TATTCCCACAATCTTTTTTATCAATAAGATTGACCAAAATGGAATTGATTTAT |
| CAACGGTTTATCAGGATATTAAAGAGAAACTTTCTGCCGAAATTGTAATCAA |
| ACAGAAGGTAGAACTGTATCCTAATATGTGTGTGACGAACTTTACCGAATCT |
| GAACAATGGGATACGGTAATAGAGGGAAACGATGACCTTTTAGAGAAATAT |
| ATGTCCGGTAAATCATTAGAAGCATTGGAACTCGAACAAGAGGAAAGCATA |
| AGATTTCAGAATTGTTCCCTGTTCCCTGTTTATCACGGAAGTGCAAAAAACA |
| ATATAGGGATTGATAACCTTATAGAAGTGATTACGAATAAATTTTATTCATC |
| AACACATCGAGGTCAGTCTGAACTTTGCGGAAAAGTTTTCAAAATTGAGTAT |
| TCGGAAAAAAGACAGCGTCTTGCATATATACGTCTTTATAGTGGCGTACTGC |
| ATTTGCGAGATTCGGTTAGAATATCGGAAAAGGAAAAAATAAAAATTACAG |
| AAATGTATACTTCAATAAATGGTGAATTATGTAAAATCGATAAGGCTTATTC |
| CGGGGAAATTGTTATTTTGCAAAATGAGTTTTTGAAGTTAAATAGTGTTCTTG |
| GAGATACAAAACTATTGCCACAGAGAAAAAAGATTGAAAATCCGCACCCTC |
| TACTACAAACAACTGTTGAACCGAGTAAACCTGAACAGAGAGAAATGTTGC |
| TTGATGCCCTTTTGGAAATCTCAGATAGTGATCCGCTTCTACGATATTACGTG |
| GATTCTACGACACATGAAATTATACTTTCTTTCTTAGGGAAAGTACAAATGG |
| AAGTGATTAGTGCACTGTTGCAAGAAAAGTATCATGTGGAGATAGAACTAA |
| AAGAGCCTACAGTCATTTATATGGAGAGACCGTTAAAAAATGCAGAATATA |
| CCATTCACATCGAAGTGCCGCCAAATCCTTTCTGGGCTTCCATTGGTTTATCT |
| GTATCACCGCTTCCGTTGGGAAGTGGAATGCAGTATGAGAGCTCGGTTTCTC |
| TTGGATACTTAAATCAATCGTTTCAAAATGCAGTTATGGAGGGGATACGCTA |
| TGGCTGTGAACAAGGATTGTATGGTTGGAATGTGACGGACTGTAAAATCTGT |

TTTAAGTATGGCTTATACTATAGCCCTGTTAGTACCCCAGCAGATTTTCGGAT
GCTTGCTCCTATTGTATTGGAACAAGTCTTAAAAAAAGCTGGAACAGAATTG
TTAGAGCCATATCTTAGTTTTAAAATTTATGCGCCACAGGAATATCTTTCACG
AGCATACAACGATGCTCCTAAATATTGTGCGAACATCGTAGACACTCAATTG
AAAAATAATGAGGTCATCTTAGTGGAGAAATCCCTGCTCGGTGTATTCAAGA
ATATCGTAGTGATTTAACTTTCTTTACAAATGGACGTAGTGTTTGTTTAACAG
AGTTAAAAGGGTACCATGTTACTACCGGTGAACCTGTTTGCCAGCCCCGTCG
TCCAAATAGTCGGATAGATAAAGTACGATATATGTTCAATAAAATAACTTAG
TGTATTTTATGTTGTTATATAAATATGGTTTCTTGTTAAATAAGATGAAATAT
TTTTTAATAAAGATTTGA

## 8.3 Biolog Arrays

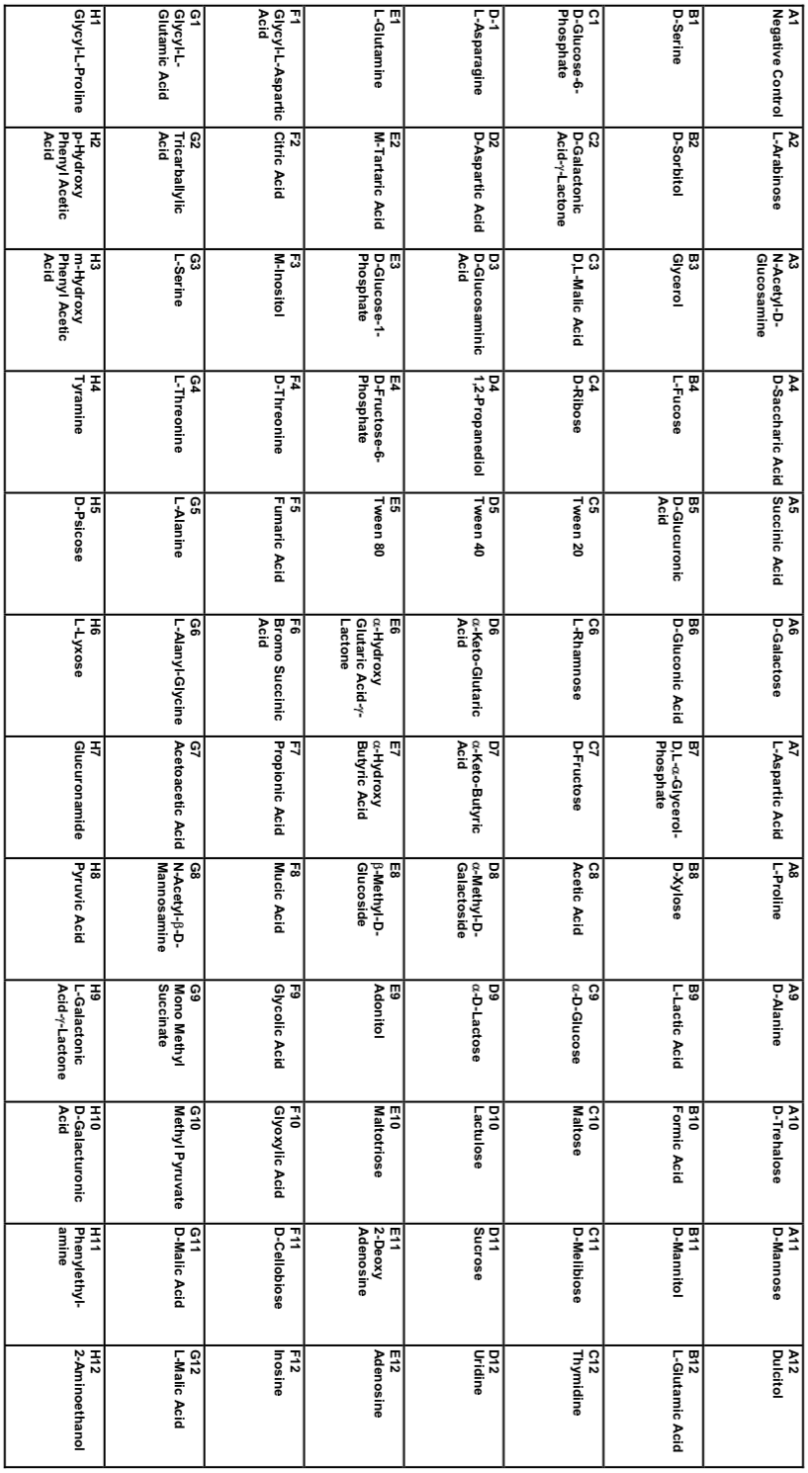



Figure 67: Biolog PM1 array full list of sugars.

## 8.4 Molecular Modelling Supplemental Data

### 8.4.1 *rpoE* Allele Nucleotide Sequences

| Strain | Nucleotide Sequence |
|---|---|
| 22F | TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAA<br>CTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTC<br>GTGGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAA<br>CGAAATTCAAAACTACCTTGGAACATCAAACAGCGATAT<br>CCGCGAAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTT<br>GACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTC<br>TTCGTTCATGGTATGGTGTGGACGAAATCGACGAAGAAAT<br>CATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACC<br>AAAAGCTAAGAAAAAACGTGTCAATGCCTTTATGGATGG<br>TGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAA<br>GACGAAGATGCATACGAAGCAGATCCAGCTCTTTCATAC<br>GATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAA<br>GCTTATGATGCAGAAATCAACGAAATCGCTCCAGATGACT<br>TGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGT<br>TTTCAGATGACGACGCTGAAACCAGCGAGGAATAA |
| SCV1D3 | TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAA<br>CTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTC<br>GTGGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAA<br>CGAAATTCAAAACTACCTTGGAACATCAAACAGCGATAT<br>CCGCGAAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTT<br>GACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTC<br>TTCGTTCATGGTATGGTGTGGACGAAATCGACGAAGAAAT<br>CATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACC<br>AAAAGCTAAGAAAAAACGTGTCAATGCCTTTATGGATGG<br>TGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAA<br>GACGAAGATGCATACGAAGCAGATCCAGCTCTTTCATAC<br>GATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAA<br>GCTTATGATGCAGAAATCAACGAAATCGCTCCAGATGACT<br>TGGGATAAGATGTGGATCTCAACGAAGACGACGACGAGT<br>TTTCAGATGACGACGCTGAAACCAGCGAGGAATAA |
| SCV3D9 | TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAA<br>CTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTC<br>GTGGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAA<br>CGAAATTCAAAACTACCTTGGAACATCAAACAGCGATAT<br>CCGCGAAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTT<br>GACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTC<br>TTCGTTCATGGTATGGTGTGGACGAAATCGACGAAGAAAT<br>CATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACC<br>AAAAGCTAAGAAAAAACGTGTCAATGCCTTTATGGATGG<br>TGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAA<br>GACGAAGATGCATACGAAGCAGATCCAGCTCTTTAATAC<br>GATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAA<br>GCTTATGATGCAGAAATCAACGAAATCGCTCCAGATGACT |

| | |
|---|---|
| | TGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGT<br>TTTCAGATGACGACGCTGAAACCAGCGAGGAATAA |
| SCV5D3 | TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAA<br>CTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTC<br>GTGGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAA<br>CGAAATTCAAAACTACCTTGGAACATCAAACAGCGATAT<br>CCGCGAAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTT<br>GACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTC<br>TTCGTTCATGGTATGGTGTGGACGAAATCGACGAAGAAAT<br>CATCGCTCTTGAAGAAAATGACGACGATGAAGTAGCACC<br>AAAAGCTAAGAAAAAACGTGTCAATGCCTTTATGGATGG<br>TGATTCAGATGCCATTGACTACAATGCAGATGATCCAGAA<br>GACGAAGATGCATACGAAGCAGATCCAGCTCTTTCATAC<br>GATGATGAAAATCCAGATGATGAAAAAAATGAAGTGGAA<br>GCTTATGATGCATAAATCAACGAAATCGCTCCAGATGACT<br>TGGGAGAAGATGTGGATCTCAACGAAGACGACGACGAGT<br>TTTCAGATGACGACGCTGAAACCAGCGAGGAATAA |
| SCV9D9 | TTGGAATTAGAAGTATTTGCTGGGCAAGAAAAAAGTGAA<br>CTATCTATGATTGAGGTAGCGCGTGCTATATTAGAACTTC<br>GTGGTCGCGATCATGAGATGCATTTTAGCGATCTTGTAAA<br>CGAAATTCAAAACTACCTTGGAACATCAAACAGCGATAT<br>CCGCGAAGCTTTGCCTCTGTTCTACACAGAGTTGAACTTT<br>GACGGTAGCTTCATCTCACTTGGGGACAACAAATGGGGTC<br>TTCGTTCATGGTATGGTGTGGACGAAATCGACGAAGAAAT<br>CATCGCTCTTGAAGAAAATGACGACGCTGAAACCAGCGA<br>GGAATAA |

### 8.4.2 RpoE Amino Acid Sequences

| Strain | Amino Acid Sequence |
|---|---|
| 22F | MIEVARAILELRGRDHEMHFSDLVNEIQNYLGTSNSDIREAL<br>PLFYTELNFDGSFISLGDNKWGLRSWYGVDEIDEEIIALEEND<br>DDEVAPKAKKKRVNAFMDGDSDAIDYNADDPEDEDAYEA<br>DPALSYDDENPDDEKNEVEAYDAEINEIAPDDLGEDVDLNE<br>DDDEFSDDDAETSEE |
| SCV1D3 | MIEVARAILELRGRDHEMHFSDLVNEIQNYLGTSNSDIREAL<br>PLFYTELNFDGSFISLGDNKWGLRSWYGVDEIDEEIIALEEND<br>DDEVAPKAKKKRVNAFMDGDSDAIDYNADDPEDEDAYEA<br>DPALSYDDENPDDEKNEVEAYDAEINEIAPDDLG |
| SCV3D9 | MIEVARAILELRGRDHEMHFSDLVNEIQNYLGTSNSDIREAL<br>PLFYTELNFDGSFISLGDNKWGLRSWYGVDEIDEEIIALEEND<br>DDEVAPKAKKKRVNAFMDGDSDAIDYNADDPEDEDAYEA<br>DPAL |
| SCV5D3 | MIEVARAILE LRGRDHEMHFSDLVNEIQNYLGTSNSDIRE<br>ALPLFYTELN FDGSFISLGDNKWGLRSWYG VDEIDEEIIA<br>LEENDDDEVAPKAKKKRVNAFMDGDSDAIDYNADDPEDED<br>AYEADPALSYDDENPDDEKN EVEAYDA |

| | |
|---|---|
| SCV9D9 | MIEVARAILELRGRDHEMHFSDLVNEIQNYLGTSNSDIREAL PLFYTELNFDGSFISLGDNKWGLRSWYGVDEIDEEIIALEEND DAETSEE |

### 8.4.3 22F WT RpoE N-terminal Structural Alignments with Enzymes

| Rank | PDB Hit | TM-score | RMSD[a] | IDEN[a] | Cov |
|---|---|---|---|---|---|
| 1 | 2krcA | 0.953 | 0.56 | 0.373 | 0.977 |
| 2 | 2m4kA | 0.702 | 1.23 | 0.403 | 0.788 |
| 3 | 1vtnC | 0.679 | 2.2 | 0.122 | 0.871 |
| 4 | 2c6yB | 0.674 | 2.69 | 0.115 | 0.894 |
| 5 | 1w5tC | 0.659 | 2.54 | 0.05 | 0.871 |
| 6 | 3zcoA | 0.651 | 2.23 | 0.092 | 0.847 |
| 7 | 5ocnA | 0.651 | 2.49 | 0.118 | 0.871 |
| 8 | 2as5F | 0.646 | 2.35 | 0.083 | 0.847 |
| 9 | 5u8jA | 0.644 | 3.06 | 0.063 | 0.941 |
| 10 | 2a3sA | 0.644 | 2.95 | 0.09 | 0.918 |

Figure 68: Ranking of proteins is based on TM-score of the structural alignment between the query structure and known structures in the PDB library. RMSDa is the root mean square deviation which measures the distances between atoms of superimposed proteins. between residues that are structurally aligned by TM-align. IDENa is the percentage sequence identity in the structurally aligned region. Cov represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

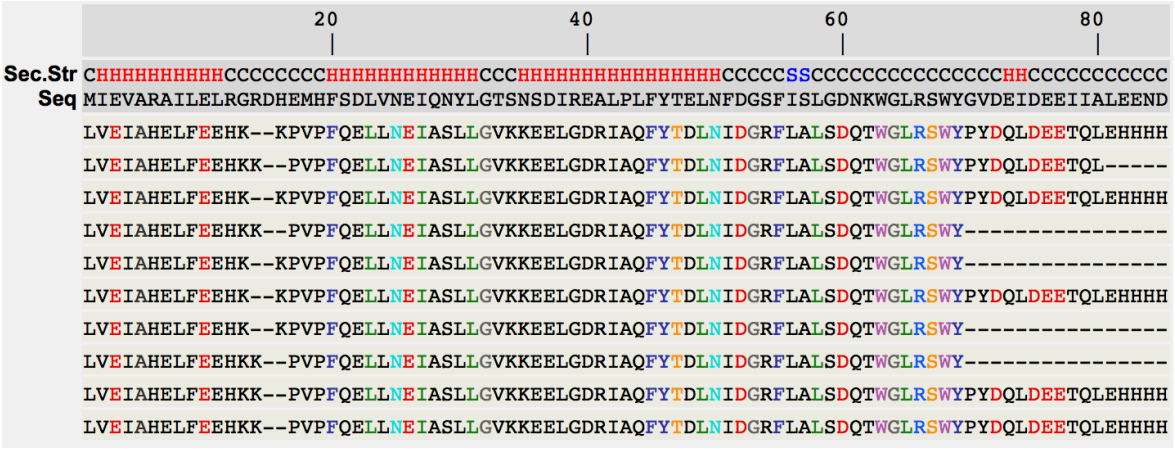



Figure 69: Threading of amino acids between the 22F N-terminal sequence and the top 10 sequence alignments. Residues are coloured in black; however, those residues in template which are identical to the residue in the query sequence are highlighted in color.