Structure/Function Analysis of a Complex of Minor Pilins in the *Streptococcus sanguinis* Type IV Pili, Mediating Binding to Host Glycans

Meriam Shahin

Department of Infectious Diseases

Imperial College London

June 2022

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy and Diploma of Imperial College

Abstract

Type IV pili (T4P) are widespread multi-functional prokaryotic filaments. Typically composed of one major pilin, T4P also contain minor pilins present at a lower abundance. Minor pilins contribute to the assembly and functional versatility of the filaments but their precise role in T4P biology remains poorly understood. The Gram-positive opportunistic pathogen *Streptococcus sanguinis* provides a simple T4P machinery for studying the involvement of minor pilins in pili biogenesis. Its retractile pili comprise two major pilins and three minor pilins – PilA, PilB and PilC. While previous work has described the two major pilins and the minor pilin PilB, this study characterises PilA and PilC. PilA is a small minor pilin structurally homologous to the T2SS pseudopilin GspI. PilC is a trimodular minor pilin with a functional C-terminal lectin domain. We show that PilA and PilC form a heterodimeric complex supported by an intermolecular β-sheet between PilA and the pilin domain of PilC. The interaction allows PilA to act as a chaperone for PilC, stabilising the PilC pilin domain by β-strand complementation. Modelling suggests that the PilA-PilC complex resides at the tip of the filaments. PilA is incorporated first, followed by PilC whose large C-terminal domains cap the pili. The PilA-PilC complex thus facilitates the presentation of the PilC lectin module by the filaments. The lectin module binds to the terminal moieties of sialylated glycans with moderate affinity. The interaction is mediated by a binding pocket on the concave side of the lectin β-sandwich fold. The role that PilC plays in the *S. sanguinis* life cycle, however, remains to be established. Altogether, this study provides the first description of a minor pilin complex in Gram-positive T4P and completes the first characterisation of the role of each pilus subunit in a T4F system.

Acknowledgements

This PhD took place in the most unlikely of circumstances – a two-year-long pandemic, multiple national lockdowns and the move of my supervisor to France. I would like to express my immense gratitude to Dr Vladimir Pelicic for his support and guidance during my research. Despite the distance, he was always there for me when I needed advice or encouragement. Thank you for everything.

I would like to extend my thanks to Prof Angelika Gründling, Dr Charlotte Millership and Dr Monisha Pathania for looking after me when I was on my own. To Dr Devon for her help with the structural and computational side of this project, and to all the past members of the Pelicic lab whose work inspired and enabled my research. To Dr Marc Morgan, who got the ITC machine fixed three times in six months so that I could conduct my experiments.

Thank you to Andrea, Catrin, Celia and Claire for the many lunches, walks and hot chocolates. This PhD would have been a much less enjoyable (and flavoursome) experience without your friendship. My thanks to everyone on the third floor of the CMBI who made the lab a nice and supportive work environment.

To Dan. Thank you for all the times you made dinner when I was working late, for all the times you listened to me complain about failing experiments, for all the times you

cheered me up when I was stressed and upset, and for offering to read this thesis even though you really did not want to.

Above all, thank you to my parents. Thank you for always believing in me and my abilities, for encouraging me to pursue my goals, for supporting me throughout my undergraduate degree, and for always picking up the phone when I needed to chat. Your dedication to my education made this possible.

Statement of Originality

I certify that the research described in this thesis is my own work except where indicated by specific references and mentions. I confirm that all previous work has been referenced in the text and listed in the bibliography where possible and all sources of help have been acknowledged.

Copyright Declaration

Contents

List of Figures

List of Tables

List of Abbreviations

BACTH	Bacterial Adenylate Cyclase Two-Hybrid
CaM	Calmodulin
CAP	Catabolite Gene Activator Protein
CDM	Chemically Defined Medium
CHO	Chinese Hamster Ovaries
CRD	Carbohydrate Recognition Domain
CSP	Competence Stimulating Peptide
Da	Daltons
Dep	Depleted
dRI	Differential Refractive Index
HeLa	Henrietta Lacks
HSQC	Heteronuclear Single Quantum Correlation
IMAC	Immobilised Metal Affinity Chromatography
Inp	Input
Ig	Immunoglobulin
IPTG	Isopropyl β-D-1-Thiogalactopyranoside
ITC	Isothermal Titration Calorimetry

MALS	Multi-Angle Light Scattering
MCS	Multiple Cloning Site
NMR	Nuclear Magnetic Resonance
ONP	O-nitrophenol
ONPG	ortho-nitrophenyl-β-galactoside
PCPA	*para*-Chlorophenylalanine
PCR	Polymerase Chain Reaction
PD	Pull-Down
Pil	Pilin
RMSD	Root Mean Square Deviation
SEC	Size Exclusion Chromatography
SDM	Site-Directed Mutagenesis
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
T2SS	Type II Secretion System
T4F	Type IV Filaments
T4P	Type IV Pili
TROSY	Transverse Relaxation Optimised Spectroscopy
UV	Ultra-Violet
vWA	von Willebrand factor A-like domain
VWF	von Willebrand Factor
WCE	Whole-Cell Extract
WT	Wild-Type

Chapter 1

Introduction

1.1 *Streptococcus sanguinis*

Streptococcus sanguinis was first isolated in 1946 from patients suffering from subacute bacterial endocarditis (White and Niven, 1946). Initially referred to as *Streptococcus s.b.e.* (for subacute bacterial endocarditis), its name was changed to *Streptococcus sanguis* from the Latin word for blood, and later corrected to *Streptococcus sanguinis* for grammatical reasons (Truper and De'Clari, 1997). Even though it was recognised early on that the primary habitat of *S. sanguinis* – like other viridans streptococci – is the mouth and throat (White and Niven, 1946), it was only in the 1960s that *S. sanguinis* was successfully isolated from dental plaque samples (Carlsson, 1965). To date, *S. sanguinis* has been identified only in humans (Kreth et al, 2017).

A Gram-positive facultative anaerobe, *S. sanguinis* is a multifaceted organism. Firstly, it is a commensal of the oral cavity which antagonises caries-causing organisms and is associated with good dental health (Caufield et al, 2000). Secondly, it is an opportunistic pathogen which upon entry in the bloodstream causes bacteraemia and infective endocarditis (Zhu et al, 2018). Most recently, it has emerged as a model for type IV pili (T4P) research (Gurung et al, 2016). The simplicity of its T4P machinery, its ease of handling and natural competence make *S. sanguinis* an ideal organism for the systematic analysis of T4P biogenesis and functions. This section will expand on these three facets of *S. sanguinis*.

1.1.1 A Commensal of the Oral Cavity

1.1.1.1 Dental Plaque

The primary habitat of *S. sanguinis* is the oral cavity – home to more than 700 bacterial species which interact with each other and with the surrounding host tissues to form complex biofilms (Aas et al, 2005). The dental plaque, which accumulates on teeth, is the most common form of bacterial growth in the oral cavity and the best-characterised biofilm (Rosan and Lamont, 2000). RNA-sequencing analysis of its microbiota has revealed that 71% of its transcriptome is represented by just nine species, and 16% of all detected transcripts correspond to *S. sanguinis* (Peterson et al, 2014).

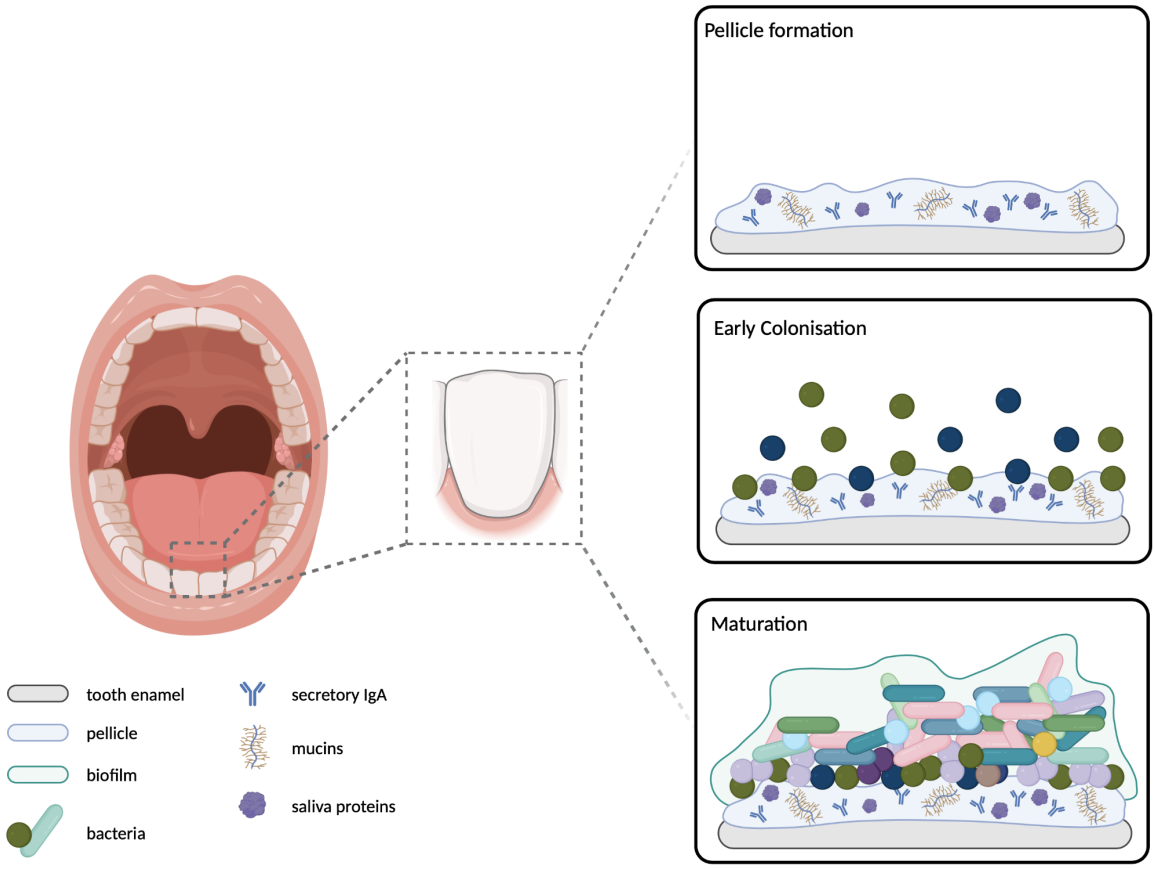


Figure 1.1: **The three phases of dental plaque development.**

The dental plaque has three distinct phases of development (Marsh, 2004). Firstly, saliva proteins – including statherin, histatin, proline-rich proteins, mucin, agglutinin,

α-amylase, glucosyltransferases and secretory IgA – are adsorbed onto all exposed tooth surfaces, forming a layer called the pellicle (Figure 1.1). Secondly, bacteria are passively transported onto the tooth surfaces where they adhere weakly through chemical and physical interactions. As the bacteria proliferate and organise into three-dimensional structures, bacterial adhesins bind to specific components of the pellicle, forming irreversible attachments to the tooth surfaces (Lafaurie et al, 2013). Adherence to the pellicle is typically performed by streptococci which constitute 60-90% of the early plaque. Thirdly, as the dominant pioneer colonisers, streptococci recruit other Gram-positive and Gram-negative bacteria, leading to the development of a mature stratified biofilm (Rosan and Lamont, 2000).

1.1.1.2 Early Colonisation

S. sanguinis participates in the early colonisation phase of dental plaque development. Colonisation by *S. sanguinis* is concurrent with the emergence of the first tooth in infants (around nine months), and the proportion of *S. sanguinis* detected in the saliva increases as new teeth erupt (Caufield et al, 2000). This suggests that *S. sanguinis* interacts with the teeth and/or the pellicle to initiate biofilm formation (Figure 1.2). Studies have shown that while *S. sanguinis* can adhere to saliva-free hydroxyapatite (the main mineral found in tooth enamel), saliva coating enhances the affinity of the interaction between the bacteria and the teeth by increasing the number of available binding sites (Tanaka et al, 1996).

The binding of *S. sanguinis* to saliva-coated hydroxyapatite was first proposed to be mediated by the bacterial adhesin SsaB (Ganeshkumar et al, 1988). SsaB is a substrate-binding protein, belonging to the LraI family of conserved metal transporters. It coordinates the

transport of the divalent metal cation Mn^{2+} across the membrane (Crump et al, 2014). As salivary proteins are mostly negatively charged, SsaB is hypothesised to engage indiscriminately with the pellicle via its bound positively charged ion (Kreth et al, 2017).

Specific components of the pellicle that support *S. sanguinis* adhesion have also been identified (Gong et al, 2000). For example, the α-amylase catalyses the hydrolysis of dietary starch and is one of the most abundant salivary proteins in humans (Peyrot des Gachons and Breslin, 2016). It has been shown to interact with the sortase-dependent pili of *S. sanguinis* (Okahashi et al, 2011a). These pili are long, rod-like structures, consisting of covalently linked pilin subunits. Each subunit carries an LPXTG amino acid motif that is recognised by a specific sortase enzyme. In *S. sanguinis*, the sortase-dependent pilus is assembled by the sortase SrtA, which catalyses the covalent attachment of the pilin subunits – FimA, FimB and FimC – to one another and to the peptidoglycan cell wall (Telford et al, 2006). While all three proteins can bind to whole human saliva, FimB and FimC interact specifically with the salivary α-amylase (Okahashi et al, 2011a).

Mucins are another major component of the saliva and the pellicle. They comprise a group of diverse glycoproteins which make up the hydrating and lubricating mucosal layer that covers epithelial cells (Kreth et al, 2017). *S. gordonii* adhesins GspB and Hsa have been shown to bind to low-molecular weight mucin MG2 and to salivary agglutinin (Takamatsu et al, 2006). The surface glycoprotein SrpA in *S. sanguinis* is a homologue of GspB and Hsa, and mutagenesis studies have confirmed its ability to bind to mucin MG2 as well (Plummer and Douglas, 2006).

The *S. gordonii* adhesin Hsa also interacts with the secretory immunoglobin A (sIgA) (Takamatsu et al, 2006). sIgA is abundant at mucosal surfaces where it plays a key role in the adaptive immune defence. sIgA, alongside α-amylase, promotes adhesion of *S.*

sanguinis to hydroxyapatite (Gong et al, 2000). This suggests that an interaction takes place between *S. sanguinis* and sIgA, and it is conceivable that the interaction is mediated by SrpA, the *S. sanguinis* homologue of Hsa.

Analysis of the *S. sanguinis* genome has revealed that it possesses an arsenal of predicted surface proteins and bacterial adhesins that may contribute to its role as a primary commensal colonising bacterium (Xu et al, 2007). *S. sanguinis* has 60 uncharacterised lipoproteins and 30 surface-exposed cell wall-anchored proteins which could be involved in attachment to the pellicle or co-adherence to other species. These numbers are significantly higher than in the sequenced strains of *S. mutans* and *S. pneumoniae*, emphasising the important role of *S. sanguinis* as an early coloniser of the oral cavity (Xu et al, 2007).

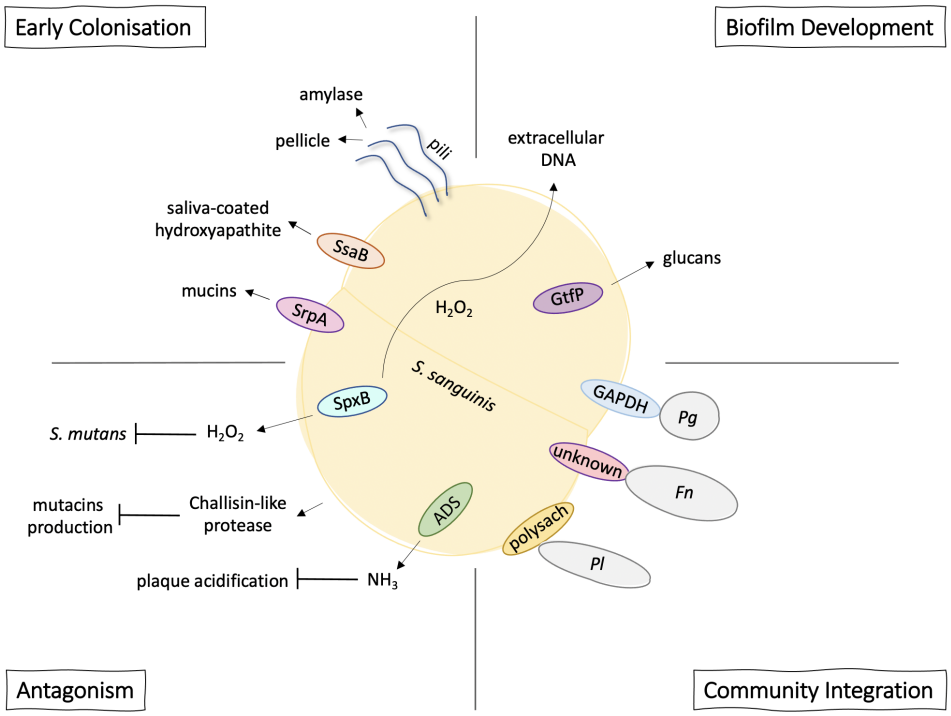


Figure 1.2: **Key factors involved in *S. sanguinis* commensalism in the oral cavity.** The schematic shows the key factors for *S. sanguinis* early colonisation, biofilm development, community integration and antagonism in dental caries. ADS = arginine deaminase system; polysach = polysaccharide, Fn = *Fusobacterium nucleatum*, Pl = *Prevotella loescheii*. Pg = *Porphyromonas gingivalis*.

1.1.1.3 Biofilm Maturation

Following attachment to the pellicle, *S. sanguinis* contributes to development of the dental plaque biofilm. A key step in biofilm maturation is the formation of the biofilm matrix which performs two key functions: 1) it provides a mesh-like structure that reinforces the integrity of the biofilm, and 2) it limits the diffusion of antimicrobial compounds. The matrix consists of carbohydrates, proteins, lipids and extracellular DNA secreted by the biofilm occupants (Kreth et al, 2017). The main biofilm polymers produced by *S. sanguinis* are glucans composed of α-1,6-linked and α-1,3-linked glucose. These glucans are synthesised from sucrose by glucosyltransferases GtfA and GtfP. Deletion of *gtfP* causes a reduction in the amount of water-soluble glucans detected in liquid culture and decreases the ability of *S. sanguinis* to form biofilms in comparison to the wild-type (WT) strain (Yoshida et al, 2014). It has also been proposed that *S. sanguinis* might engage with glucans produced by other streptococci using putative glucan-binding proteins, such as GbpB, SspC and SspD (Moraes et al, 2014).

In addition, *S. sanguinis* produces extracellular DNA. H_2O_2 generated by the pyruvate oxidase enzyme SpxB can cause the release of DNA by *S. sanguinis* without compromising cell integrity. This DNA is high-molecular weight and is derived from the bacterial chromosome (Kreth et al, 2009). Treatment of *S. sanguinis* cultures with DNase I reduces cell aggregation, confirming the contribution of extracellular DNA to cell-cell binding and biofilm maturation (Kreth et al, 2009).

A group of genes involved in nucleotide biosynthesis have also been implicated in biofilm formation and maturation by *S. sanguinis*. Nucleotides such as c-di-GMP, c-di-AMP, cAMP and (p)ppGpp play a well-defined role in quorum sensing and biofilm formation (Zhu et al, 2018). Although an analogous signalling pathway has not been reported for

S. sanguinis yet, there is speculation that *S. sanguinis* is evolving a new set of proteins for intra-species communication (Valdebenito et al, 2018).

1.1.1.4 Community Integration

As a primary coloniser, *S. sanguinis* participates in interbacterial interactions that facilitate the recruitment of later colonisers to the dental plaque. These interactions often involve multiple cell surface components engaged in reciprocal adhesin-receptor binding events (Rosan and Lamont, 2000). For instance, *Prevotella loescheii* expresses a cell surface adhesin PlaA which binds to the cell wall polysaccharides of *S. sanguinis* (London and Allen, 1990). Similarly, *Porphyromonas gingivalis* uses its pili to interact with multiple streptococcal Ssp surface proteins. In particular, the SspB surface protein is known to mediate binding between *S. gordonii* and *P. gingivalis* (Park et al, 2005). *S. sanguinis* expresses an orthologue of the SspB adhesin in which the residues involved in the interaction between *S. gordonii* and *P. gingivalis* are conserved (Xu et al, 2007). This indicates that an interaction might also take place between *S. sanguinis* and *P. gingivalis* and indeed, *P. gingivalis* exhibits a preference for colonising plaque biofilms containing higher proportions of *S. sanguinis* and other streptococci (Rosan and Lamont, 2000). A glyceraldehyde-3-phosphate dehydrogenase (GAPDH) of *S. sanguinis* has also been implicated in its interaction with *P. gingivalis* (Maeda et al, 2004).

S. sanguinis also interacts with *Fusobacterium nucleatum* (Kaplan et al, 2009). *F. nucleatum* is reported to function as a bridge organism, connecting the primary colonisers, such as *S. sanguinis*, to anaerobic secondary colonisers, such as *Aggregatibacter actinomycetemcomitans* (Brennan and Garrett, 2019). When co-cultured together, *S. sanguinis* and *F. nucleatum* assemble into highly ordered structures with one *F. nucleatum* bac-

terium bound to more than 10 *S. sanguinis* organisms (Brennan and Garrett, 2019). The interaction between the two species is mediated by the arginine-inhibitable RadD adhesin of *F. nucleatum* (Kaplan et al, 2009).

1.1.1.5 Role in Dental Caries

While *S. sanguinis* fosters interbacterial interactions leading to the development of the dental plaque, it also exerts an antagonistic effect against bacteria which cause dental caries (Zhu et al, 2018). Dental caries is a chronic disease which results in the dissolution of the tooth enamel and cavity formation. It is caused by progressive decline in the pH of the dental plaque and concurrent alterations in the dental microbiome. The presence of fermentable carbohydrates in the oral cavity lowers the environmental pH. This shifts the microbiota equilibrium towards cariogenic bacteria which in turn further acidify the plaque, and thus promote disease progression via a vicious circle (Valdebenito et al, 2018). *S. mutans* is the most common etiological agent of dental caries.

S. sanguinis and *S. mutans* are widely reported to have an antagonistic relationship (Valdebenito et al, 2018). *S. mutans* bacteria are not detectable in children who have high levels of *S. sanguinis* in their saliva (Caufield et al, 2000). The ability of *S. sanguinis* to outcompete *S. mutans* in the dental plaque is attributed to its secretion of H_2O_2. H_2O_2 is produced by the *S. sanguinis* pyruvate oxidase enzyme SpxB during the conversion of pyruvate into acetyl-phosphate (Valdebenito et al, 2018; Mu et al, 2021). H_2O_2 is a type of reactive oxygen species that can damage cellular proteins and DNA. *S. sanguinis* is thought to protect itself from oxidative damage by employing the ferritin-like iron-binding protein Dps (Xu et al, 2014), the thioredoxin reductase TrxB (Marco et al, 2013; Xu et al, 2014) and the glutathione peroxidase BasA (Valdebenito et al, 2018). *S. mutans*, however,

generates very little H_2O_2 of its own and is highly vulnerable to the antimicrobial activity of the H_2O_2 secreted by *S. sanguinis* (Yamamoto et al, 2002; Valdebenito et al, 2018). Instead, *S. mutans* produces two bacteriocins – mutacins I and IV – which kill *S. sanguinis*. In response, *S. sanguinis* has evolved a mechanism to evade the antimicrobial effect of these compounds. It expresses an extracellular challisin-like protease that degrades the *S. mutans* competence-stimulating peptide (CSP), thereby eliminating the induction signal for mutacins synthesis (Zhu et al, 2018).

S. sanguinis also moderates the plaque pH. *S. sanguinis* appears to be more sensitive to acidic conditions than *S. mutans*, suggesting that the progressive plaque acidification associated with cariogenesis might strongly disadvantage *S. sanguinis* and reduce its density. Hence, *S. sanguinis* produces ammonia from arginine via the arginine deaminase system, causing an increase in the local pH. Alkali generation is proposed to play a major role in the inhibition of dental caries (Huang et al, 2015). Bacteria isolated from the dental plaque of caries-free individuals were shown to exhibit higher activity levels of the arginine deaminase system compared to bacteria from individuals suffering from dental caries. In fact, at neutral pH, cariogenic organisms represent a very small proportion of the dental microbiome (Valdebenito et al, 2018). This suggests that *S. sanguinis* counters the development of dental caries on two fronts – 1) by directly antagonising the cariogenic *S. mutans* bacteria, and 2) by maintaining a neutral pH in the oral cavity. Thus, *S. sanguinis* is often associated with good oral health.

1.1.1.6 Role in Dental Periodontitis

S. sanguinis also plays a role in the second most common oral disease – dental periodontitis. Periodontitis is an inflammatory disease that compromises the integrity of the

teeth-supporting tissues (gums) and leads to teeth loosening and teeth loss. Several bacterial species have been implicated in the pathogenesis of dental periodontitis, including *P. gingivalis*. Even though *S. sanguinis* appears to be more abundant in healthy dental plaques compared to periodontitis-associated plaques (Zhu et al, 2018), *S. sanguinis* provides a binding site for *P. gingivalis* (Park et al, 2005). This raises the possibility that instead of promoting oral health, *S. sanguinis* merely survives better at healthy sites, making it a correlate of oral health rather than a contributor (Zhu et al, 2018).

1.1.2 An Opportunistic Pathogen

1.1.2.1 Infective Endocarditis

Long before *S. sanguinis* was recognised as a commensal of the oral cavity, it was identified as a causative agent of infective endocarditis. Indeed, *S. sanguinis* was first recovered from the bloodstream and heart valve vegetations of patients suffering from infective endocarditis (White and Niven, 1946). Infective endocarditis is an infection of the endocardial surfaces of the heart, and specifically the heart valves, the mural endocardium or a septum defect. Oral bacteria are the most frequent etiological agents of infective endocarditis. Together streptococci and staphylococci are responsible for ∼80% of all infections, while *S. sanguinis* alone causes 18-30% of endocarditis cases (Liesenborghs et al, 2020).

Overall, infective endocarditis is a rare disease, affecting between 2-12 individuals per 100,000 worldwide each year (Liesenborghs et al, 2020). Its clinical features are valvular destruction, microvascular and large vessel embolisation, and heart failure (Chambers and Bayer, 2020). At-risk groups include individuals with congenital valvular diseases or rheumatic heart disease, patients with prosthetic valves or intravenous catheters, elderly people with degenerative valve lesions and intravenous drug users (Que and Moreillon,

2011). Following treatment, estimated mortality is between 10-30%, but if left untreated, the mortality rate rises up to ~100% (Liesenborghs et al, 2020). Antibiotics are the most successful approach for treating infective endocarditis (Que and Moreillon, 2011), which makes the increasing frequency of antibiotic resistance among streptococci particularly worrying (Alves-Barroco, 2020).

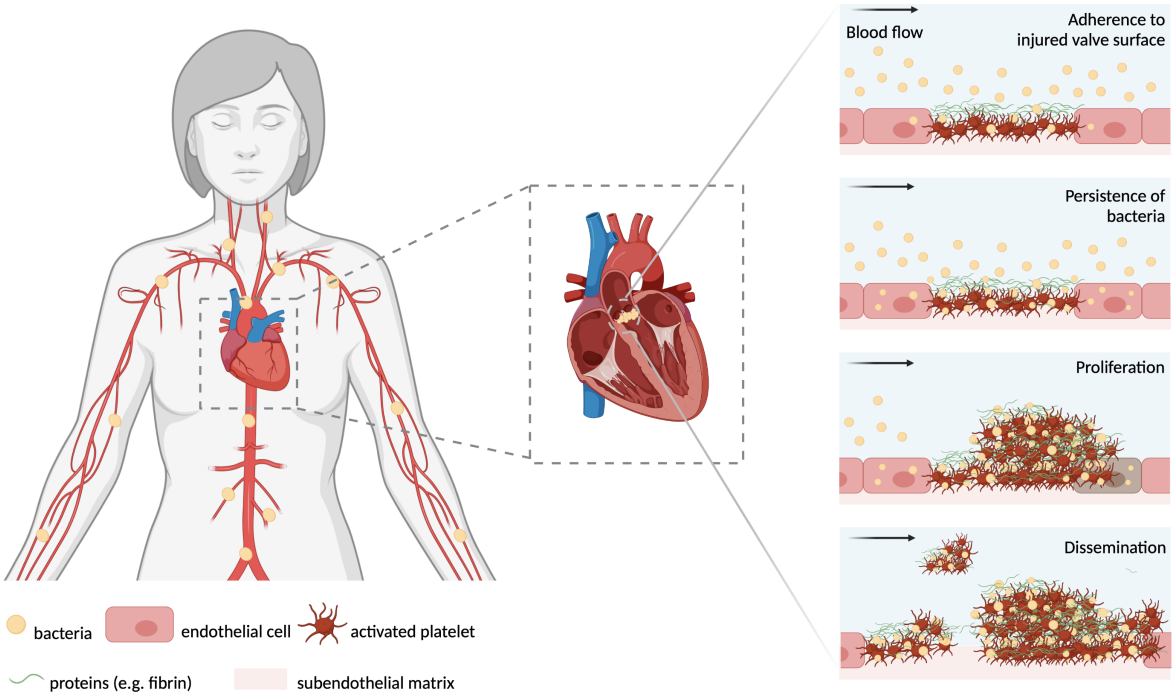


Figure 1.3: **Pathogenesis of infective endocarditis.** Infective endocarditis begins with bacteraemia and a pre-existing injury to the endothelial lining of the heart. The injury site creates a binding surface for bacteria which adhere through platelets and extracellular matrix proteins. Once attached, the bacteria persist and proliferate, forming a septic vegetation. Lastly, the septic vegetation starts to break up and the vegetation particles spread through the blood to distant organs. Adapted from Werdan et al, 2014.

1.1.2.2 Pathogenesis

There are two key factors that predispose to the development of infective endocarditis: 1) oral bacteraemia (Lafaurie et al, 2013), and 2) injury to the valvular endothelium or endocardium (Chambers and Bayer, 2020). Bacteraemia occurs when an opportunistic pathogen from the oral cavity enters the bloodstream. This typically happens after oral procedures, such as dental treatments or hygiene procedures, as well as in severe cases of periodontal disease (Lafaurie et al, 2013). The bacteraemia allows the pathogen to

be carried by the blood to the heart. Typically, the endothelial lining of the heart is resistant to bacterial and fungal infections, but the pre-existing lesion provides a milieu for bacterial colonisation (Bashore et al, 2006).

Injury to the heart endothelium leaves the underlying collagen and tissue factor exposed to the flowing blood, which facilitates the recruitment and activation of platelets (Furie and Furie, 2018). Platelets bind to the exposed collagen or to the collagen-bound von Willebrand factor via surface glycoprotein receptors. This causes the blood cells not only to adhere to the site of injury, but also induces their activation. In parallel, the exposed tissue factor initiates a proteolytic cascade that generates thrombin. Thrombin converts fibrinogen to fibrin, forming long protein threads that entangle platelets, and then activates the entangled platelets by cleaving their protease-activated platelet receptor (Furie and Furie, 2018). These two pathways are likely to act synergistically to induce the release of multiple agonists – adenosine diphosphate (ADP), serotonin and thromboxane A_2 – by the captured platelets. The secreted agonists activate other platelets and amplify the signal for thrombosis, ultimately leading to the formation of a sterile vegetation (Furie and Furie, 2018).

The pre-existing lesion creates a binding site for the opportunistic pathogens circulating in the blood by allowing them to adhere to the cellular and protein components of the sterile vegetation (Figure 1.3). As the bacteria proliferate *in situ*, they stimulate further cycles of endothelial injury, platelet recruitment and activation, as well as fibrin production, causing the now septic vegetation to expand (Werdan et al, 2014). Gradually, the pathogens become ensconced in platelet-fibrin clots, sequestering themselves away from the host immune system (Que and Moreillon, 2011). With time, the vegetation begins to fracture, and eventually small fragments detach from the endothelial lining of the heart. This

facilitates the spread of the bacterial pathogens through the blood to distant organs, including the kidneys and the brain, and can result in disease complications, such as glomerulonephritis, cerebral haemorrhage, stroke and meningitis (Chambers and Bayer, 2020).

In order to instigate infective endocarditis, *S. sanguinis* must therefore overcome three main challenges: 1) it must enter and survive in the bloodstream, 2) it must adhere to and colonise the site of valve damage, and 3) it must persist and proliferate in the septic vegetation (Figure 1.4). Unfortunately, not much is known about the virulence factors and mechanisms that allow *S. sanguinis* to overcome these challenges (Martini et al, 2020).

1.1.2.3 Survival in the Blood

Like other oral bacteria, *S. sanguinis* likely gains access to the bloodstream following dental treatments, such as hygiene procedures or surgeries (Lafaurie et al, 2013). *S. sanguinis* also induces the cell death of several epithelial cell lines by secreting H_2O_2. Since epithelial cells provide a physical barrier against infections with pathogenic microbes, *S. sanguinis* cytotoxicity could aid in overcoming host defences in the oral cavity and entering the bloodstream (Okahashi et al, 2014).

Once in the blood, the bacteria must proliferate and avoid detection by the host innate immune system in order to survive. A recent analysis of a comprehensive library of *S. sanguinis* mutants lacking non-essential genes highlighted the role of two groups of genes in the survival of *S. sanguinis* in human serum – genes involved in *de novo* purine synthesis and genes encoding proteins that belong to the energy coupling factors (ECF) family of transporters (Xu et al, 2011; Zhu et al, 2021). ECF proteins typically mediate the transport of various substrates across the cell by recruiting respective substrate-binding

proteins encoded by non-contiguous genes (Zhu et al, 2021). An ECF transporter, called SiaFGH, has been reported to act as a heme importer for group A streptococci. SiaFGH allows the bacteria to obtain nutritional iron from host haemoproteins and is crucial for bacterial colonisation and invasive infection (Chatterjee et al, 2020). It is possible that the ECF transporters of *S. sanguinis* perform a similar function, facilitating *S. sanguinis* growth and proliferation in the blood.

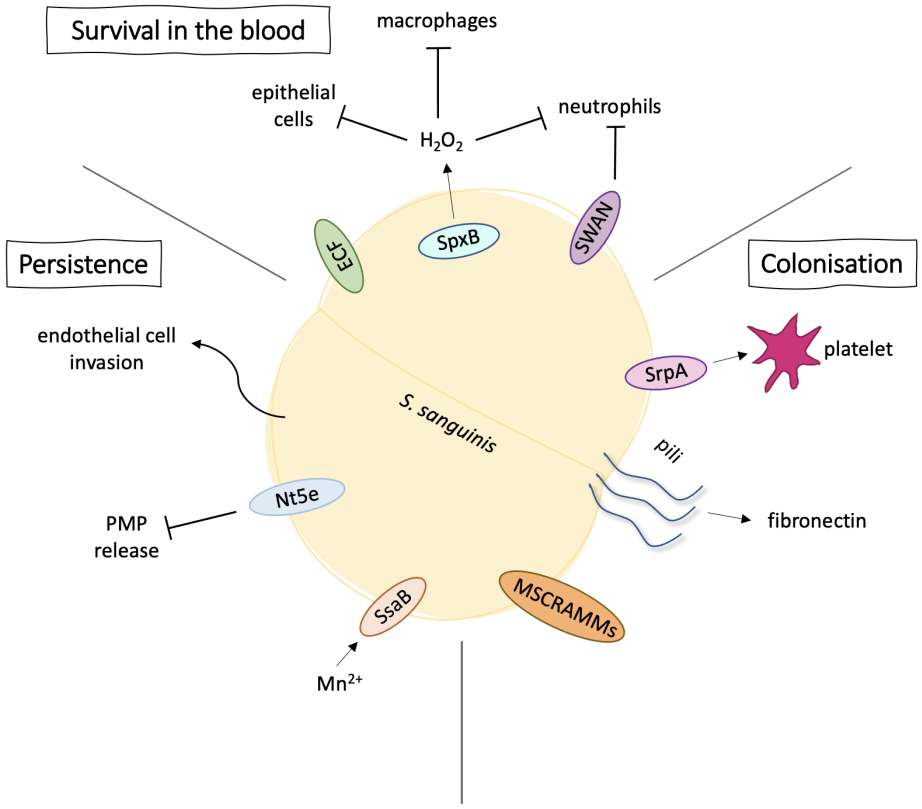


Figure 1.4: **Virulence factors involved in *S. sanguinis* pathogenesis during infective endocarditis.** The schematic shows the key virulence factors for *S. sanguinis* survival in the blood, colonisation of the heart vegetation and persistence at the site of vegetation.

Much more is known about the interaction between streptococci and the components of the host innate immune response. It is well-documented that streptococci bacteria can induce the death of both macrophages and neutrophils via H_2O_2 production. H_2O_2 secreted by *S. oralis* damages and impairs the function of macrophages, culminating in their death (Okahashi et al, 2013). H_2O_2 compromises the integrity of the macrophage organelles,

including the lysosomes which rapidly deacidify and degrade. This not only allows *S. oralis* to escape from phagocytosis, but simultaneously releases hydrolytic enzymes in the cytoplasm that inflict further damage on the macrophage cellular components, leading to its eventual death (Okahashi et al, 2016). *S. sanguinis* is also reported to induce cell death of macrophages (Okahashi et al, 2011b), but its cytotoxic mechanism has not yet been fully elucidated.

Neutrophils can also eliminate bacterial pathogens by phagocytosis, or they can use web-like chromatin structures, called neutrophil extracellular traps (NETs), to capture and kill detected microbes. Neutrophils have a critical role in eliminating streptococci from the bloodstream as evidenced by the fact that severe neutropenia is a major risk factor for viridans streptococci bacteraemia (Bochud et al, 1994). Therefore, *S. sanguinis* has evolved two separate mechanisms for evading neutrophils depending on which bactericidal activity it faces. Production of H_2O_2 induces neutrophil cell death, and disruption of the catalytic activity of the pyruvate oxidase SpxB renders *S. sanguinis* more vulnerable to phagocytosis by neutrophils (Sumioka et al, 2017). *S. sanguinis* can also free itself from NETs using a cell-surface nuclease called streptococcal wall anchored nuclease (SWAN). SWAN can digest NET DNA in the presence of Mg^{2+} and Ca^{2+} ions, and its deletion increases the susceptibility of *S. sanguinis* to NETs-mediated death (Morita et al, 2014).

The mechanisms which allow *S. sanguinis* to avoid immune detection in the bloodstream are also likely to play a key role in its persistence in vegetations. Blood monocytes are typically present in sterile and septic vegetations where they secrete cytokines and procoagulant factors and thus promote inflammation and vegetation growth. In particular, neutrophils have been suggested to contribute to infective endocarditis pathogenesis (Sumioka et al, 2017) by secreting NETs which act as a scaffold for the expansion of the

septic vegetation (Fuchs et al, 2010).

1.1.2.4 Colonisation

Once *S. sanguinis* reaches the heart, it must adhere to and colonise the sterile vegetation. The high blood flow passing through the cardiac valves creates considerable shear stress which typically impedes bacterial adhesion. Platelets overcome the shear stress by binding to the exposed collagen-bound von Willebrand factor (VWF). When VWF is subjected to shear stress, it unfolds and exposes a previously concealed domain which binds to the GpIb receptor of platelets. In turn, most endocarditis-causing pathogens bind to the platelets in the sterile vegetation and use them as an anchor in order to overcome shear stress (Liesenborghs et al, 2020).

S. sanguinis is no exception. It has been shown to interact directly with platelets and to stimulate their activation (Kerrigan et al, 2002). This interaction is mediated by the surface glycoprotein SrpA of *S. sanguinis* and the highly sialylated glycoprotein receptor GpIb of human platelets (Plummer et al, 2005). SrpA is a serine-rich repeat adhesin, part of a family of streptococcal cell wall-anchored glycoproteins that mediate adherence to host proteins and/or glycans. SrpA is a homologue of the sialoglycan binding adhesins GspB and Hsa from *S. gordonii* (Loukachevitch et al, 2016). Crystallography studies of SrpA have revealed that its sialoglycan binding region differs from those of GspB and Hsa, allowing it to accommodate either multiple or larger, more complex ligands and to bind to human platelets with a much higher affinity (Bensing et al, 2016; Loukachevitch et al, 2016). Consequently, *S. sanguinis* lacking a functional *srpA* exhibits significantly decreased binding to platelets and delayed platelet activation (Plummer et al, 2005).

Since mutating SrpA does not completely abolish the interaction between *S. sanguinis*

and platelets, it is likely that other cell-surface proteins are also involved in the binding to and activation of platelets. One candidate is the platelet-aggregation-association protein (PAAP) which has a collagen-like platelet-interactive domain (Erickson and Herzberg, 1993). PAAP promotes the accumulation of platelets in lesions and induces their activation and aggregation. The binding target of PAAP is unknown, but platelet aggregation can be inhibited with an antibody against the *S. sanguinis* PAAP protein (Erickson and Herzberg, 1993).

Apart from platelets, infective endocarditis-causing bacteria also interact with extracellular matrix proteins in the vegetation. *S. pneumonia* binds to fibronectin and collagen using its sortase-dependent pili (Hilleringmann et al, 2008), while *S. mutans* interacts with collagen via specific collagen-binding proteins (Aviles-Reyes et al, 2016). The importance of the interaction between *S. sanguinis* and fibronectin for infective endocarditis pathogenesis was discovered early on – *S. sanguinis* mutants incapable of binding to fibronectin exhibit reduced virulence in a rat disease model (Lowrance et al, 1990). The interaction between fibronectin and *S. sanguinis* is mediated by its sortase-dependent pili, and specifically the minor pilin subunit FimC (Okahashi et al, 2010). The T4P minor pilin PilB was also recently shown to bind to fibronectin and fibrinogen (Raynaud et al, 2021). To date, no binding to collagen has been reported even though genomic analysis has identified five putative cell wall-anchored proteins which possess a collagen-binding domain (Xu et al, 2007).

1.1.2.5 Persistence

Following adherence, *S. sanguinis* must persist and proliferate at the vegetation site. Many of the mechanisms that *S. sanguinis* employs to evade the host immune system in

the blood are likely to play a key role in vegetation persistence as well. Like the blood monocytes recruited to the vegetation, activated platelets also have a strong antimicrobial function (Yeaman, 2014). They secrete small cationic proteins, called platelet microbicidal proteins (PMPs), which target common bloodstream pathogens. In order to persist in the vegetation, *S. sanguinis* has evolved a way to suppress PMPs production. A cell surface protein, called ecto-5-nucleotidase (Nt5E) was shown to inhibit platelet aggregation and delay the release of PMPs without interfering with the adhesive interaction between the bacteria and the platelets. Nt5E also hydrolyses extracellular ATP into ADP, which acts as an immunosuppresant and further impairs the phagocytic and pro-inflammatory activities of the platelets and other nearby blood monocytes (Fan et al, 2012).

The facultative anaerobe *S. sanguinis* must also be able to proliferate in the presence of high levels of oxygen in the blood passing through the heart. The SsaB lipoprotein, discussed earlier, has been implicated in this survival mechanism of *S. sanguinis*. SsaB mediates the uptake of manganese into the bacterial cell through the SsaABC transport system and supplies the cations to the manganese-dependent superoxide dismutase A (SodA) enzyme. Since mutants lacking the SodA enzyme are significantly less virulent than WT, SsaB-mediated accumulation of manganese could be critical for *S. sanguinis* growth and survival during infective endocarditis (Crump et al, 2014).

Some viridans group streptococci circumvent the stringent survival requirements at the site of the infective endocarditis vegetation by invading the cardiovascular endothelial cells. For bacteria such as *S. aureus*, this is an integral event in the pathogenesis of infective endocarditis (Edwards et al, 2010). Selected strains of *S. gordonii*, *S. mutans*, *S. mitis* and *S. oralis* have also been shown to invade monolayers of human umbilical vein endothelial cells (HUVECs) leading to cytotoxicity. In *S. gordonii*, host cell invasion is

dependent on the sialic acid-binding protein Hsa, glycosyltransferases and co-aggregation proteins CshA and CshB (Stinson et al, 2003). The same study identified a few *S. sanguinis* strains that are also capable of invading HUVECs (Stinson et al, 2003). A more recent paper has shown that *S. sanguinis* strain SK36 can invade immortalised human aortic endothelial cells which are representative of the cell types that *S. sanguinis* encounters at the vegetation site of infective endocarditis (Martini et al, 2020; Martini et al, 2021). Invasion efficiency is dependent on both the sortase-dependent pili and the type IV pili of *S. sanguinis* (Okahashi et al, 2010; Martini et al, 2020; Martini et al, 2021). However, it remains unclear whether endothelial cell invasion is relevant to the *in vivo* pathogenesis of *S. sanguinis*.

1.1.3 A Model Organism for Studying Type IV Pili

S. sanguinis is a frequent source of transient bacteraemia that can lead to infective endocarditis. Antibiotic prophylaxis is often prescribed to high-risk patient groups undergoing invasive dental procedures, but recent studies have revealed that bacteraemia can occur even after routine activities such as chewing or brushing (Crump et al, 2014). As long-term antibiotic use for potential transient bloodstream infections is impractical, and infective endocarditis is a potentially lethal disease, identifying the virulence factors and mechanisms of *S. sanguinis* is of high importance. A better understanding of *S. sanguinis* pathogenesis will facilitate the development of more efficacious drugs for the prevention and treatment of endocarditis.

One promising research avenue is a prominent virulence factor that is rare among streptococci bacteria. Type IV pili (T4P) are long, thin, surface-exposed filaments of polymerised type IV pilin subunits which mediate motility, engage in host cell interactions and con-

tribute to biofilm formation in many bacterial species (Berry and Pelicic, 2015). The T4P of *S. sanguinis* mediate cell adhesion and interactions with extracellular matrix proteins fibronectin and fibrinogen (Raynaud et al, 2021). They have been shown to be critical for the development of native valve infective endocarditis in a white rabbit disease model and to contribute to pathogenesis through the development of cardiac vegetations (Martini et al, 2021). The T4P of *S. sanguinis* are also likely to play a crucial role in *S. sanguinis* colonisation of the oral cavity as a commensal organism. While the role of T4P in bacterial physiology and disease pathogenesis has been extensively reported for Gram-negative bacteria, very little was known until recently about the T4P of Gram-positive species, including *S. sanguinis* (Melville and Craig, 2013).

S. sanguinis piliation and twitching motility on agar plates were first reported 40-50 years ago. In these studies, *S. sanguinis* strains isolated from throat swabs were found to produce spreading zones around colonies on blood agar. The spreading was attributed to twitching motility, and the pili were successfully observed by electron microscopy (Henriksen and Henrichsen, 1975; Henriksen and Henrichsen, 1976). But it is only over the past decade that genetic and molecular evidence has accrued in support of *S. sanguinis* T4P expression (Gurung et al, 2016; Chen et al, 2019) and T4P involvement in infective endocarditis (Raynaud et al, 2021; Martini et al, 2021).

S. sanguinis is now recognised as a convenient model bacterial species for studying T4P due to the simplicity of its T4P machinery and its genetic tractability (Gurung et al, 2016; Pelicic, 2019). *S. sanguinis* requires only 10 proteins for T4P biogenesis (Gurung et al, 2016), and it is naturally competent (Henriksen and Eriksen, 1976; Gaustad, 1985). Competence in *S. sanguinis* is mediated by a competence pilus rather than the T4P. Mutation studies have shown that disrupting T4P functionality has no impact on trans-

formation frequencies, but deleting any of the genes associated with the competence pilus abolishes competence (Gurung et al, 2016). This facilitates the systematic analysis of the *S. sanguinis* T4P machinery through mutagenesis. Moreover, *S. sanguinis* pili can be visualised by electron microscopy and purified for biochemical analysis (Gurung et al, 2016), making it easy to assess the impact of mutations on T4P biogenesis. The next section will discuss in detail the structure, assembly and function of T4P.

1.2 Type IV Pili

1.2.1 Morphology, Functionality and Prevalence

Type IV pili (T4P) are hair-like, surface-exposed filaments of polymerised type IV pilin subunits. They belong to the type IV filament (T4F) superfamily of evolutionary related filamentous nanomachines. T4P extend up to several μm in length and 5-8 nm in width, and can interact laterally to form bundles (Berry and Pelicic, 2015). They are flexible and elastic, able to generate forces in the range of 80-100 pN (Merz et al, 2000; Maier et al, 2002).

T4P are widely spread in the prokaryotic kingdom. To date, they have been identified in more than 150 species, spanning Gram-negative, Gram-positive and archael phyla (Pelicic, 2008). Their overarching role is to mediate interactions between prokaryotes and their environment. Through cycles of extension, binding and retraction, T4P contribute to a wide array of properties, such as adhesion, motility and competence (Ellison et al, 2022). Research interest in T4P arises from their capacity as virulence factors. T4P facilitate the attachment to and colonisation of host tissues during bacterial pathogenesis, and are essential for the ability of prominent human pathogens to cause diseases (Berry and

Pelicic, 2015). This makes them potential targets in antimicrobial therapies and a pressing research topic in an age of rising antibiotic resistance (Dumenil, 2019).

This section will provide a detailed overview of the T4P assembly machinery and dynamics, their evolution and functional versatility. Specific attention is dedicated to the role of minor pilins in T4P biology as they are the main focus of this thesis.

1.2.2 Type IV Pilins

1.2.2.1 Major Pilins

T4P function through the reversible polymerisation of type IV pilins into thin helical filaments. Each filament is composed of approximately 500-1,000 copies of a single subunit, referred to as the major pilin (Giltner et al, 2012). The availability of the major pilin inside the bacterial cell dictates the number of pili expressed per cell as well as the functionality of those pili (Long et al, 2001).

Type IV pilins are small (∼7-20 kDa) structural proteins distinguished by a highly conserved N-terminal signal sequence, called the class III signal peptide. The class III signal peptide begins with a leader peptide sequence made up of 6-26 predominantly hydrophilic and neutral residues. The leader peptide ends with a conserved glycine or alanine residue at position -1 (Gly-1) and is typically followed by 21 predominantly hydrophobic residues, with the exception of a negatively charged glutamate amino acid at position 5 (Glu5) (Berry and Pelicic, 2015). Type IV pilins are expressed in a precursor form – prepilins – which possess an intact class III signal peptide sequence and are not competent for assembly. Prepilins are inserted into the cytoplasmic membrane by the SecYEG system, and their leader peptide sequence is cleaved by a dedicated aspartyl protease called prepilin peptidase. Only after prepilin processing do prepilins become mature pilins and are able

to polymerise into T4P filaments (Giltner et al, 2012). Due to its significance for membrane insertion and processing prior to assembly, the N-terminal domain of pilins is highly conserved, while the C-terminal domain presents much more variability.

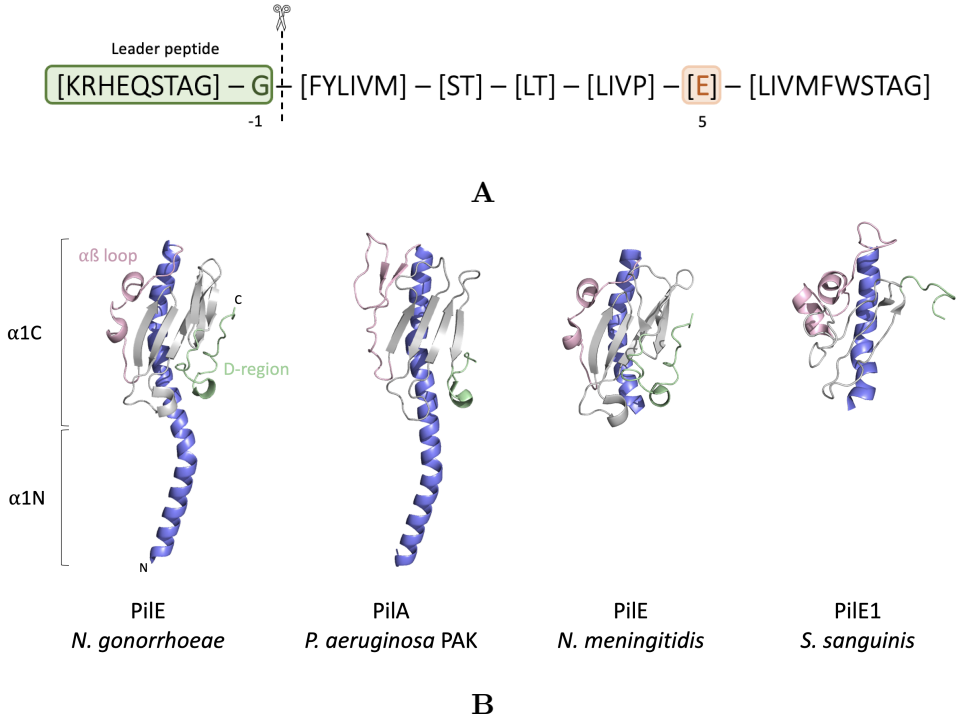


Figure 1.5: **Type IV pilins.** (A) The class III signal peptide characteristic of type IV pilins. The leader peptide, the glycine at position -1 and the glutamate at position 5 are highlighted. The prepilin peptidase cleavage site is indicated by a dotted line. The consensus sequence was obtained by Expasy ScanProsite, PS00409. (B) Structures of type IV major pilins. Type IV pilins are characterised by an N-terminal α-helix (blue), an $\alpha\beta$ loop (pink), a globular head composed of a β-sheet (silver) and a D-region (green). Full-length *N. gonorrhoaea* PilE (PDB: 1AY2); Full-length *P. aeruginosa* PAK PilA (PDB: 1OQW); Truncated *N. meningitidis* PilE (PDB: 5JW8); Truncated *S. sanguinis* PilE1 (PDB: 6I2O).

Even though there is limited sequence similarity beyond the first 25 residues of the α1N helix, type IV pilins share a common architecture according to X-ray crystallography and nuclear magnetic resonance (NMR) studies (Figure 1.5). This allows pilins from a wide range of organisms to assemble into pili using the same biogenesis principles (Craig and Li, 2008). The shape of full-length pilins resembles a lollipop, composed of an extended α-helix and a globular domain. The α-helix can be divided into two fragments – a hydrophobic α1N helix, typically spanning residues 1-28, and a soluble α1C helix, covering residues 29-52. The hydrophobic α1N helix has a triple role: 1) it acts as a

transmembrane domain, anchoring the pilin subunits to the cytoplasmic membrane before filament assembly, 2) it serves as a platform for pilin-pilin interactions during filament assembly, and 3) it supports the re-entry of disassembled type IV pilins into the cytoplasmic membrane during pilus retraction, thereby allowing the recycling of pilin subunits in subsequent rounds of pilus extension. The soluble α1C helix is amphipathic and packs against the C-terminal globular domain at a 45° angle. The globular domain contains 4 to 7 antiparallel β-strands which make up a β-meander motif (Giltner et al, 2012).

Structural studies of type IV pilins have informed our understanding of the way pilins assemble into filaments and accommodate pili functionality. For structural analysis, pilins are usually truncated by 28 residues or more in order to remove the hydrophobic α1N helix (Giltner et al, 2012). This increases the solubility of the pilin proteins and makes them easier to purify and handle. Comparisons between the truncated and full-length type IV pilin structures have shown that deletion of the α1N residues has minimal impact on the folding and structure of the α1C helix and the globular domain (Craig et al, 2003; Hartung et al, 2011; Kolappan et al, 2016). For example, the two structures of the *Pseudomonas aeruginosa* K strain major pilin PilA are virtually identical (Craig et al, 2003).

The extended α-helix of type IV pilins has two neighbouring regions whose conformations are thought to facilitate the packing of pilins into filaments and to support the inherent flexibility and elasticity of T4P. Firstly, cryo-electron microscopy (cryo-EM) reconstruction of intact purified pili from *Neisseria meningitidis* has revealed a 'melted' region between residues glycine 14 (Gly14) and proline 22 (Pro22) in the α-helix. This stretch of amino acids adopts a helical secondary structure in the structure of the *Neisseria* major pilin PilE, but it appears to have a non-helical conformation inside the pili (Kolappan et al, 2016; Wang et al, 2017). It is proposed that the 'melting' between Gly14 and Pro22

occurs upon polymerisation of the pilin subunits into filaments, thereby facilitating the packing of the hydrophobic α1N helices into the filament core. The non-helical conformation is also thought to introduce a degree of elasticity to the T4P, allowing them to resist shear stresses, such as blood flow during *Neisseria* pathogenesis (Kolappan et al, 2016; Wang et al, 2017). Secondly, the structures of full-length major pilins from *P. aeruginosa*, *Neisseria gonorrhoeae* and *Dichelobacter nodosus* all display a shallow curvature immediately after the 'melted' portion of the α1N helix. Two conserved residues – proline 22 (Pro22) and glycine/proline 42 (Gly/Pro42) – induce a kink in the extended α-helix which confers a gentle S-shaped curve to the N-termini of pilins (Parge et al, 1995; Craig et al, 2003; Hartung et al, 2011). This curvature is thought to reduce intersubunit packing and to increase the flexibility of the T4P filament, allowing them to form the hairpin loops commonly observed by electron microscopy (Burrows, 2008).

Despite the highly conserved architectural fold of type IV pilins, there are two hypervariable regions that support the structural diversity seen among pilins (Craig and Li, 2008). These regions are located on either side of the β-meander motif – the $\alpha\beta$ loop which links the α-helix to the first β-strand in the β-meander, and the D-region, which is the C-terminus stapled to the last strand of the β-meander usually by a disulphide bond (Berry and Pelicic, 2015). In some pilins, such as PilX in *Neisseria*, the D-region has a functional, not only structural role (Helaine et al, 2005; Helaine et al, 2007). The hypervariable regions often bear post-translational modifications, namely the addition of glycans, phosphorylcholine, phosphoethanolamine and phosphoglycerol. The function of these modifications, however, is unclear and tends to differ between species (Berry and Pelicic, 2015).

1.2.2.2 Classification of T4P

Historically, T4P have been grouped into two sub-classes – T4aP and T4bP – based on the length and sequence of the major pilins as well as the genetic organisation of the T4P assembly machinery (Craig et al, 2004). T4aP were the first to be discovered. They are widespread among Gram-negative and Gram-positive species and exhibit a broad host range – examples include *P. aeruginosa*, *N. gonorrhoeae*, *N. meningitidis*, *S. sanguinis* and *Clostridium difficile*. T4bP are mainly found in bacteria that colonise the human intestine, such as enteropathogenic *Escherichia coli* (EPEC), enterotoxigenic *E. coli* (ETEC), *Salmonella typhi* and *Vibrio cholerae* (Proft and Baker, 2009). Until recently, the tight-adherence (Tad) pili were also included in the T4bP group. However, phylogenetic analysis of the extension ATPases and the platform proteins of different T4F systems has demonstrated that Tad pili are clearly distinct from T4aP and T4bP, placing them in unique clade which has an archaeal origin instead (Denise et al, 2019). This suggests that Tad pili merit their own sub-class – T4cP (Pelicic, 2019).

T4a pilins have shorter leader peptide sequences than T4b pilins. Their mature pilins are also shorter – 150-175 amino acids compared to 180-200 for T4b pilins (Craig et al, 2004). The leader peptides of T4c pilins are of similar length to T4b pilins (Ellison et al, 2022), but once processed, the T4c pilins are much smaller in size (50-80 amino acids). To date, several structures of T4a and T4b pilins have been determined, revealing markedly different protein folds between the two sub-classes (Ellison et al, 2022). The pilins differ primarily in the $\alpha\beta$-loop and the D-region. The N-methylated N-terminal residue of the processed T4a pilins is usually phenylalanine but tends to vary for the T4b pilins. Overall, T4a pilins share greater N-terminal sequence homology than T4b pilins (Craig et al, 2004).

Furthermore, T4aP genes tend to be scattered across the genome in small highly conserved blocs, such as the pilus assembly sub-complex *pilMNOPQ* bloc and the core minor pilins *pilHIJK* bloc (Pelicic, 2008). In contrast, T4bP and T4cP machineries are simpler, encoding a set of 10-12 genes in a single locus (Pelicic, 2008; Tomich et al, 2017). An example is the T4bP of *V. cholerae* which is organised in one operon on the *Vibrio* pathogenicity island (Proft and Baker, 2009). Even though this should make T4bP and T4cP more easily acquirable through horizontal gene transfer, T4aP are much more widespread (Pelicic and Berry, 2015). Critically, the majority of T4bP do not have a retraction ATPase and do not mediate twitching motility or DNA uptake (Kolappan et al, 2015).

1.2.2.3 Prepilin Translocation and Processing

The translocation of prepilins across the cytoplasmic membrane and their subsequent processing by the prepilin peptidase comprise the first stages of the T4P biogenesis process (Berry and Pelicic, 2015). For both T4P and T2SS prepilins, correct insertion into the membrane occurs independently of other T4F components (Dupuy et al, 1991; Arts et al, 2007). The class III signal peptide alone promotes the translocation of prepilins across the cytoplasmic membranes and determines their membrane topology (Dupuy et al, 1991). In accordance with the positive-inside rule, prepilins are anchored in the cytoplasmic membrane by their N-terminal hydrophobic domain, the hydrophilic N-terminal leader peptide remains in the cytoplasm, and the C-terminal globular domain resides in the periplasm (Von Heijne and Gavel, 1988; Dupuy et al, 1991). The insertion of the prepilins in the membrane is carried out co-translationally by the Signal Recognition Particle (SRP) and the Sec protein-translocation pathway (Francetic et al, 2007). The SRP binds to the conserved N-terminal motif of a prepilin as it emerges from the ribosome, and delivers

the ribosome-prepilin complex to the Sec translocon. Protein translation then feeds the secretion of the nascent prepilin through the translocon which mediates the integration of the prepilin into the lipid bilayer where it adopts its typical 'lollipop' 3D structure (Green and Mecsas, 2016).

Following translocation, prepilins are processed by a dedicated prepilin peptidase (Nunn and Lory, 1991). The prepilin peptidase is a bifunctional enzyme with two distinct active sites – one cleaves the leader peptide, while the other catalyses the N-methylation of the N-terminal phenylalanine of the mature pilin (Strom et al, 1993). Prepilin peptidases form a new superfamily of polytopic membrane aspartic acid proteases which lack the canonical active site motif (D(T/S)G) and perform optimally at near neutral pH unlike other classical aspartic acid proteases. Their protease activity is dependent on two highly conserved aspartic acid residues (LaPointe and Taylor, 2000). The crystal structure of the prepilin peptidase from *Methanococcus maripaludis* shows that the enzyme must undergo conformational changes to bring the two aspartyl residues in close proximity for proteolysis to occur. These conformational changes are likely triggered by the binding of the prepilin substrate (Hu et al, 2011; Lemkul and Bevan, 2011). Indeed, prepilin processing has been shown to occur in cell-free translation systems that co-express a protein peptidase and a full-length pilin, indicating that no other T4F components are involved (Aly et al, 2013). Prepilin peptidases often exhibit substrate promiscuity and are able to process the class III signal peptides of pilins from other T4F and from other species (Berry and Pelicic, 2015).

In all cases, prepilin processing occurs after the conserved glycine residue on the cytoplasmic side of the prepilin protein (position -1), thereby removing the cytoplasmic domain of the prepilins. The full length of the leader peptide is required for efficient processing,

but deletion experiments have shown that it can be shortened to 5-6 amino acids without adverse effects (Strom and Lory, 1992; Horiuchi and Komano, 1998; Ng et al, 2009). It is the glycine residue at the end of the leader peptide (Gly-1) that is crucial for prepilin processing (Strom and Lory, 1991). In a site-directed mutagenesis study of the leader peptide of the major pilin in *P. aeruginosa*, all pilin constructs with Gly-1 mutations failed to undergo processing and led to a non-piliated phenotype. The only exception was the substitution of glycine with an alanine residue, another amino acid with a small side group, which led to partial processing of the major pilin (Strom and Lory, 1991).

Only processed pilins are competent for polymerisation into T4P. The glutamate at position 5 (Glu5) in the class III signal peptide is dispensable for prepilin processing but is essential for filament assembly. Substitutions of Glu5 prevent the pilins from assembling into homopolymeric T4P, but allow the pilins to form heteropolymers with WT subunits (Aas et al, 2007). The importance of Glu5 for assembly is attributed to the formation of a salt bridge between Glu5 of one pilin and the N-terminal amide of methylated Phe1 of the preceding pilin (Berry and Pelicic, 2015). It is thought that this electrostatic interaction facilitates the docking of incoming pilins into the nascent pilus structure and neutralises two charges in the hydrophobic environments of the inner membrane and the filament core (Kolappan et al, 2016).

1.2.2.4 Minor Pilins

While the major pilin is the main structural component of the pilus, other pilins present at a lower abundance have also been shown to contribute to pilus biogenesis and function (Giltner et al, 2012). These pilins are expressed at lower levels than the major pilin and are referred to as minor pilins. Like the major pilin, they are produced in a prepilin

form, containing a short N-terminal class III signal peptide, and are inserted into the cytoplasmic membrane with the aid of the SecYEG translocon (Jacobsen et al, 2020). Once correctly positioned in the membrane, minor pilins undergo prepilin processing by the prepilin peptidase, rendering them competent for assembly into the pilus. The incorporation of minor pilins into the T4P filaments has now been observed in multiple studies (Helaine et al, 2007; Giltner et al, 2010; Cehovin et al, 2011; Berry et al, 2019).

Minor pilins also share key structural features with the major pilin: they display the characteristic lollipop-like shape, comprised of the α1N and α1C helices and the C-terminal globular domain. The most significant differences usually occur at the two hypervariable pilin regions – the $\alpha\beta$-loop and the D-region. Minor pilins are also more varied in size than major pilins as they often encode additional functional modules (Salleh et al, 2019; Raynaud et al, 2021).

Minor pilins in T4aP are divided into two categories: core and non-core minor pilins. The core minor pilins are widespread among T4P-encoding organisms and are usually encoded in a single operon. In contrast, the non-core minor pilins are exclusive to particular species and may or may not be clustered with the core minor pilins (Giltner et al, 2012; Jacobsen et al, 2020).

Core Minor Pilins

There are typically four core minor pilins in the T4aP. They are homologous to the minor pseudopilins of the evolutionary related T2SS (Jacobsen et al, 2020) and will be referred to as PilH, PilI, PilJ and PilK for clarity (see Table 1.1). The core minor pilins are essential for T4P assembly – deletion of a core minor pilin leads to loss of piliation in both *Neisseria* and *P. aeruginosa*. Pili assembly in these mutants can be restored by a

concurrent deletion of the retraction ATPase *pilT*, resulting in piliated double mutants such as Δ*pilH*Δ*pilT* (Winther-Larsen et al, 2005; Carbonnelle et al, 2006; Giltner et al, 2010). This has led to the proposal that the core minor pilins regulate pilus homeostasis by counteracting the PilT-mediated retraction of the fibres (Winther-Larsen et al, 2005).

T4aP			T2SS		
Neisseria	*P. aeruginosa*	*M. xanthus*	*E. coli*	*P. aeruginosa*	*K. oxytoca*
PilH	FimU	FimU	GspH	XcpU	PulH
PilI	PilV	PilV	GspI	XcpV	PulI
PilJ	PilW	PilW	GspJ	XcpW	PulJ
PilK	PilX	PilX	GspK	XcpX	PulK
PilC	PilY1	PilY1	–	–	–

Table 1.1: **T4aP core minor pilins and their homologues in the T2SS.** PilC and PilY1 are T4P-associated adhesins which interact with the core minor pilins. PilK and its homologues lack the conserved glutamate residue at position 5 in the class III signal peptide.

Furthermore, PilH, PilI, PilJ and PilK are strongly dependent on one another and the T4P-associated adhesin PilC for their ability to localise to and accumulate in T4P filaments in the Δ*pilT* suppressor background (Winther-Larsen et al, 2005; Ngueyn et al, 2015a). In *P. aeruginosa*, PilIJK and PilC form a subcomplex that recruits PilH which in turn binds to the major pilin (Nguyen et al, 2015a; Nguyen et al, 2015b). Similarly, in *Myxococcus xanthus*, the four core minor pilins and the adhesin PilC have been shown to interact together by bacterial two-hydrid and pull-down assays. This multimeric complex has been visualised by cryo-electron tomography (cryo-ET) and suggested to reside at the tip of the pilus (Treuner-Lange et al, 2020).

The formation of a tip-located minor pilin complex is further substantiated by the homology between the T4P core minor pilins and the T2SS minor pseudopilins. In the enteropathogenic *E. coli* T2SS, minor pseudopilins GspI, GspJ and GspK form a het-

erotrimeric complex (Figure 1.6) whose structure suggests that it most likely resides at the tip of the pseudopilus (Korotkov and Hol, 2008). Similar interactions between T2SS minor pseudopilins have been reported in *Vibrio vulnificus* (Yanez et al, 2008a), *P. aeruginosa* (Douzi et al, 2009; Zhang et al, 2018) and *Klebsiella oxytoca* (Cisneros et al, 2012a). Specifically, in *K. oxytoca*, the interaction between minor pseudopilins PulI and PulJ leads to a significant shift of their transmembrane domains, allowing the minor pseudopilin complex to acquire a pseudopilus-like structure in the cytoplasmic membrane. It is postulated that complex formation between the minor pseudopilins shifts the equilibrium towards pilus extension by acquiring a pre-assembled conformation, thereby priming filament assembly (Cisneros et al, 2012a).

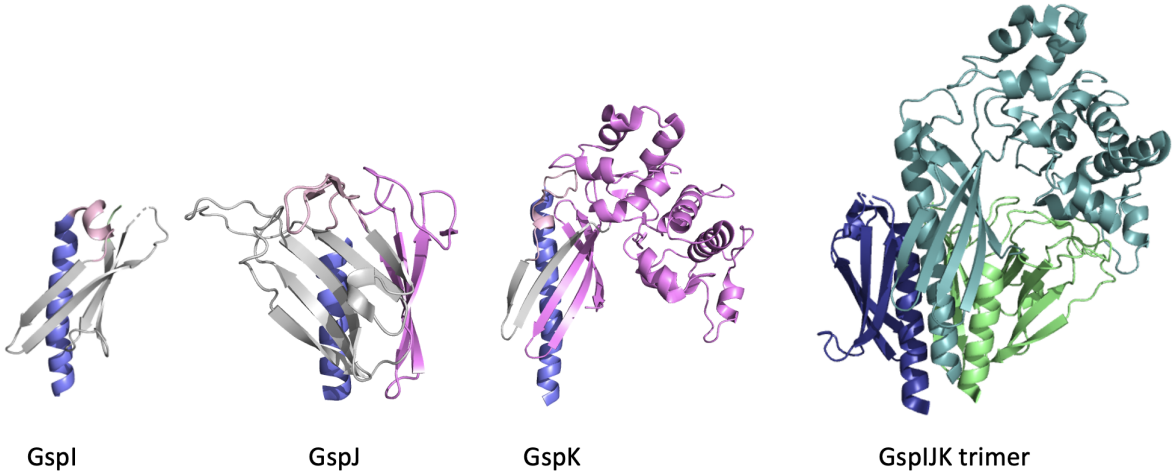


Figure 1.6: **Structures of type II secretion system minor pseudopilins.** The minor pseudopilins are coloured by their α-helix (blue), $\alpha\beta$ loop (light pink), β-sheet (gray), D-region (green) and additional modular domains (violet). GspI is shown in dark blue, GspJ in green and GspK in cyan (PDB: 3CI0).

More interestingly, T2SS minor pseudopilins have been reported to prime the assembly of type IV pili. In an *E. coli* organism that expresses an indigenous Ppd T4P system and an exogenously expressed *K. oxytoca* Pul T2SS, both T4P and T2SS minor pseudopilins are functional in promoting the assembly of T4P, suggesting a similar role and mechanism of action for both sets of proteins (Cisneros et al, 2012b). This is also observed in *P. aeruginosa* which expresses one T4P and one T2SS in lab conditions. When the core

T4P minor pilins are deleted, reduced surface piliation is observed, but if both minor pilins and the minor pseudopilins are deleted, no surface pili are recovered (Nguyen et al, 2015a). In both organisms, the pili assembled in the absence of the minor pilins are primed by the minor pseudopilins. This lends further support to the role of minor pilins in initiating assembly. It is, however, unclear if the minor pilins act by triggering an elongation competent state of the T4P or by providing the scaffolding for the major pilin polymer. According to Burrows (Burrows, 2012), the two proposals are compatible.

Non-Core Minor Pilins

Non-core minor pilins are not essential for pili assembly – deletion of a non-core pilin does not affect bacterial surface piliation. However, non-core pilins confer functional properties onto the filament, such as aggregation, adhesion and natural competence (Giltner et al, 2012). *Neisseria* minor pilins ComP, PilV and PilX are among the best-characterised non-core minor pilins (Figure 1.7). All three have been shown to assemble in the T4P filaments and to be dispensable for piliation (Helaine et al, 2007; Cehovin et al, 2011; Cehovin et al, 2013).

Highly conserved throughout the *Neisseria* genomes, ComP was the first minor pilin to be identified (Figure 1.7). Mutagenesis of *comP* leads to a significant reduction in transformation efficiency, highlighting its essential role in *Neisseria* natural competence (Wolfgang et al, 1999). ComP is required for the binding of DNA by T4P (Cehovin et al, 2013). It interacts with specific conserved short DNA sequences, called DNA uptake sequences (DUS), via an electropositive stripe on its surface and mediates DNA uptake into the bacterial cell (Wolfgang et al, 1999; Cehovin et al, 2013). Structural analysis of ComP pilins from two different *Neisseria* species has revealed that the DNA receptor utilises a novel DNA-binding motif (Berry et al, 2016).

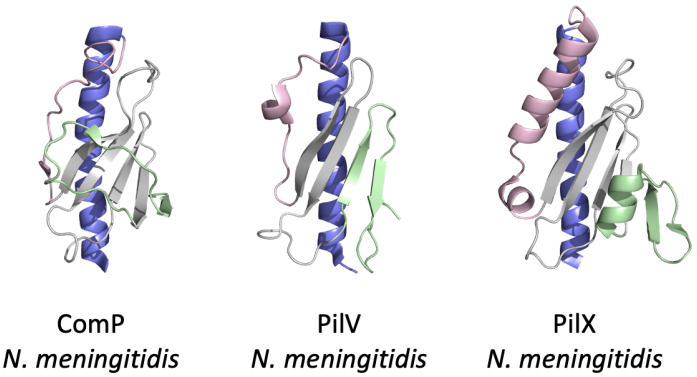


Figure 1.7: **Structures of type IV pili non-core minor pilins.** The non-core minor pilins are coloured by their α-helix (blue), $\alpha\beta$ loop (light pink), β-sheet (gray) and D-region (green). *N. meningitidis* ComP (PDB: 5HZ7); PilV (PDB: 5v0M); PilX (PDB: 2OPD).

The minor pilin PilV contributes to the adhesiveness of the T4P. Frameshift and gene disruption mutations of *pilV* lead to a ∼100-fold reduction in adhesion to human cells compared to WT (Winther-Larsen et al, 2001). The interaction between PilV and host cells is mediated by CD147, a well-established marker of brain capillaries (Bernard et al, 2014), and β2-adrenoreceptor, which is a G protein-coupled receptor expressed in brain endothelial cells (Coureuil et al, 2010). Binding to CD147 is thought to be a prerequisite for the subsequent binding to the β2-adrenoreceptor and the activation of the β2-adrenoreceptor/β-arrestin signalling pathway (Coureuil et al, 2010; Bernard et al, 2014). PilV-induced β2-adrenoreceptor/β-arrestin signalling promotes the reorganisation of the host cell surface and the formation of cellular protrusions, which together stabilise bacterial adhesion and ensure lasting microcolony adhesion against shear stresses such as blood flow (Mikaty et al, 2009). Indeed, deletion of *pilV* significantly impairs the ability of *N. meningitidis* to colonise a severe combined immunodeficiency (SCID) mouse grafted with human skin, confirming the essential role of PilV in adhesion and vascular colonisation (Barnier et al, 2021). PilV has been further implicated in the regulation of the levels of natural competence of *N. gonorrhoeae*. PilV is proposed to act as an antagonist

to ComP, opposing ComP accumulation in the T4P filaments either by obstructing its trafficking or altering its stability (Aas et al, 2002).

PilX mediates bacterial cell aggregation. Its crystal structure reveals an unusual functional D-region which forms a hook-like protrusion that is essential for PilX functionality (Helaine et al, 2005). The PilX D-region is proposed to allow PilX subunits from interacting filaments to brace against each other. This counterbalances retraction and pulls the bacteria closer together, thereby allowing aggregation to occur (Helaine et al, 2007). Furthermore, PilX is thought to be involved in T4P-mediated signalling of *N. meningitidis* colonies to endothelial cells. While PilX does not interact with the β2-adrenoreceptor, it has been shown to promote conformational changes that facilitate the elongation of pili typically associated with signalling to host cells (Brissac et al, 2012).

T4b Minor Pilins

In contrast to the multiple and functionally redundant minor pilins of the T4aP machineries, T4bP encode a single minor pilin immediately after the major pilin gene on the pilus operon. In ETEC, the minor pilin CofB is required for pilus assembly – when CofB is deleted, no major pilin is detected in the sheared pili fraction. The crystal structure of CofB shows it to be a modular protein. CofB is composed of a pilin domain with a classical T4bP fold, followed by two discrete C-terminal modules connected by flexible linkers (Kolappan et al, 2015). CofB was found to form a homotrimer in crystals and in solution. The trimer is mediated by extensive hydrophobic interactions among the two C-terminal modules. The homotrimeric structure is proposed to sit at the tip of the filament, and the two C-terminal modules that mediate oligomerisation are thought to play a role in filament assembly initiation (Kawahara et al, 2016). Similarly, the T4bP of *V. cholerae* have a single minor pilin TcpB which is required for pili assembly and functionality (Ng

et al, 2016). TcpB was also shown by crystallography to form a homotrimer, and immunogold labelling of purified pili has confirmed its position at the tip of the *V. cholerae* pili (Gutierrez-Rodarte et al, 2019). These findings resonate with the data gathered for the minor pilins of T4aP and T2SS. Together, they point to a general principle for minor pilin complex formation at the tip of pili with a critical role in filament assembly and/or functions.

1.2.3 Type IV Pili Machinery

Following membrane integration and leader peptide cleavage, pilins leave the cytoplasmic membrane and polymerise into filaments that extend beyond the cell surface (Berry and Pelicic, 2015). This process is supported and powered by the T4P machinery. The *N. meningitidis* T4P machinery consists of 23 proteins, 15 of which are essential for T4P biogenesis (Carbonnelle et al, 2006; Brown et al, 2010). Each of these proteins mediates one of the three genetically discrete steps of T4P biogenesis (Figure 1.8) – assembly of the pili, retraction antagonism and filament emergence on the cell surface (Winther-Larsen et al, 2005; Carbonnelle et al, 2006).

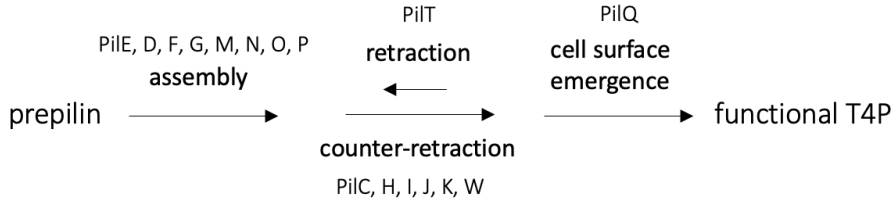


Figure 1.8: **The three genetically discrete steps of T4P biogenesis.** Filament assembly, antagonism of retraction powered by the PilT ATPase and emergence of the filaments on the cell suface. The T4P machinery proteins involved in each step are listed as reported for *N. meningitidis*. The figure is based on studies from Carbonnelle et al, 2006; Takhar et al, 2013; Goosens et al, 2017.

If a protein is required for filament assembly, its deletion leads to loss of piliation, which cannot be reversed by a concomitant deletion of the retraction ATPase PilT. Such proteins

include the major pilin PilE, the prepilin peptidase PilD, the extension ATPase PilF, the polytopic platform protein PilG and the assembly sub-complex proteins PilM, PilN, PilO and PilP (Carbonnelle et al, 2006; Takhar et al, 2013). In fact, these eight proteins are sufficient to promote filament assembly in a non-native heterologous host (Goosens et al, 2017). In contrast, the Δ*pilT* mutation can suppress piliation defects caused by the deletion of genes involved in retraction antagonism. These include the genes that encode the core minor pilin proteins in *Neisseria*. They are proposed to function as effectors of pilus homeostasis by promoting extension and countering retraction during T4P biogenesis (Winther-Larsen et al, 2005; Carbonnelle et al, 2006). Lastly, in Gram-negative bacteria, filaments need to cross the outer membrane in order to emerge on the cell surface. This is dependent on the secretin protein PilQ as the T4P filaments of the double deletion mutant Δ*pilQ*Δ*pilT* are trapped in the periplasm (Wolfgang, 2000; Carbonnelle et al, 2006).

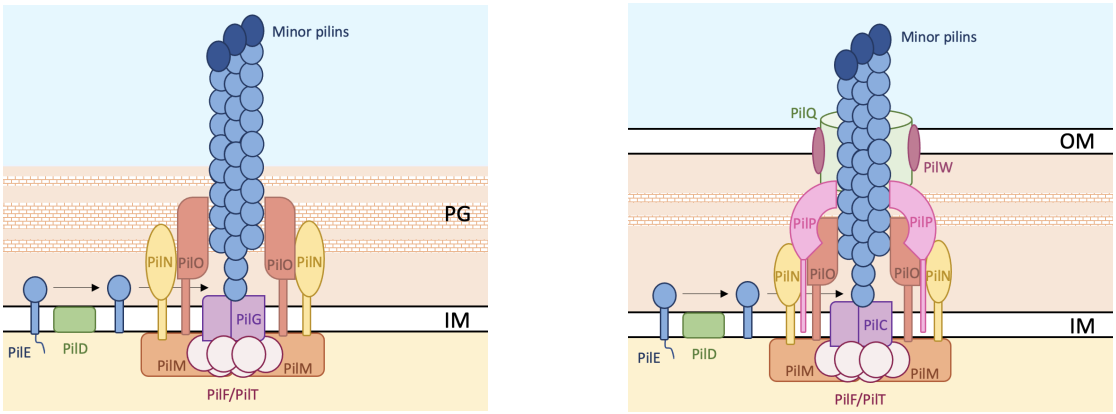


Figure 1.9: **Architecture of the T4P machinery.** (A) Gram-positive bacteria have a simpler T4P machinery due to the lack of an outer membrane. (B) Gram-negative T4P are more complex because they require the outer membrane (OM) sub-complex for filament emergence on the cell surface. The figure is based on findings by Chang et al, 2016.

While systematic mutational analysis in *Neisseria* has suggested a role for each protein in T4P biogenesis, cryo-ET studies of *M. xanthus* cell poles have provided a direct insight into the architecture of the T4P machinery (Chang et al, 2016). The T4P machinery is

a multilayered structure that spans the bacterial cell envelope (Figure 1.9). It comprises an outer membrane (OM) complex, an inner membrane (IM) assembly platform, a motor complex and a pilus stem. The pilus stem is made up of an assembly-priming complex of the minor pilins and the major pilin (Chang et al, 2016).

The OM complex is exclusive to diderm bacteria – it forms a gated pore in the outer membrane that allows the filaments to emerge on the cell surface (Wolfgang, 2000). It is composed of the secretin PilQ and a closely associated lipoprotein PilW. PilQ is a dodecamer with a ring-like structure than can accommodate the pilus fibre (Collins et al, 2001). It is composed of six stacked rings which form a central cavity surrounded by two internal gates (Collins et al, 2004; Salzer et al, 2016). In the absence of the pilus, the secretin gate prevents the escape of molecules outside the cell, while the perisplamic gate is proposed to prevent the entry of perisplasmic components into the PilQ central cavity (Salzer et al, 2016; McCallum et al, 2019). Structural studies have revealed PilQ to be a highly dynamic pore which undergoes large conformational changes in response to pilus binding (Collins et al, 2005). Shifts in its periplasmic domains and the widening of its central cavity trigger a switch from a 'closed' to an 'open' channel state and allow the assembled pilus to emerge on the cell surface (Gold and Kudryashev, 2016). The stability, multimerisation and membrane localisation of PilQ are strongly dependent on the outer membrane protein PilW which is also exclusive to Gram-negative species (Carbonnelle et al, 2004). PilQ is connected to the IM assembly platform by the inner membrane lipoprotein PilP (Balasingham et al, 2007; Chang et al, 2016).

PilP spans the periplasmic space. It interacts with PilQ via its C-terminal domain, while its N-terminal domain binds to and stabilises the PilN/PilO heterodimer (Tammam et al, 2013). PilN and PilO dimerise via their globular periplasmic domains. They are anchored

in the inner membrane by a single transmembrane helix and have a short N-terminal cytoplasmic tail. This short N-terminal segment mediates the interaction between PilN and the cytoplasmic protein PilM (Georgiadou et al, 2012). In turn, PilM interacts directly with the two motor ATPases and indirectly with the polytopic membrane platform protein PilG, thereby establishing a connection between the OM complex and the motor complex (Bischof et al, 2016). Some studies propose that the PilMNOP complex is responsible for communicating signals from the cytoplasmic to the periplasmic components of the T4P machinery regarding extension/retraction and the opening/closing of the secretin channel in Gram-negative species (Tammam et al, 2013; McCallum et al, 2019). Similarly, in Gram-positive species, where the PilQ and PilP proteins are absent due to the lack of an outer membrane (Melville and Craig, 2013; Berry and Pelicic, 2015), the PilMNO heterotrimer is proposed to play an active role in T4P assembly (Pelicic, 2019). This is consistent with findings that PilN binding to PilM favours its interaction with the extension ATPase and reduces its interactions with the retraction ATPase in *P. aeruginosa* (McCallum et al, 2016).

The two cytoplasmic ATPases power filament assembly and disassembly. As members of the superfamily of traffic ATPases, they possess an N-terminal PAS-like domain and a C-terminal RecA-like domain which contains the four signature ATPase motifs essential for their ATPase activity – Walker A and Walker B motifs, the Asp box and the His box (Berry and Pelicic, 2015). While the extension ATPase powers filament assembly, the retraction ATPase powers the depolymerisation of pilin subunits from the filament base into the cytoplasmic membrane where they await subsequent polymerisation (Morand et al, 2004). The PilT retraction ATPase is exclusive to T4aP machineries and supports their functionality (Wolfgang et al, 1998). Since retraction with low speed and low force

has been reported for *pilT* deletion mutants in T4aP systems as well as for T4bP and T4cP machineries, PilT is thought to function by enhancing the speed and force of the retraction (McCallum et al, 2019).

Biophysical experiments and structural studies have revealed that the two ATPases form elongated hexameric rings with quasi two-fold symmetry (Satyshur et al, 2007; McCallum et al, 2017). The hexamers are held together by highly conserved interactions between the N-terminal and C-terminal domains of neighbouring subunits. The interface between two neighbouring subunits forms the ATP binding site. ATP binding forces the N-terminal and C-terminal domains to come together, adopting a closed conformation (Mancl et al, 2016; Craig et al, 2019). Pairs of opposing subunits adopt distinct nucleotide-dependent conformations – a closed interface when bound to ATP or ADP, or an open interface which follows the release of ADP and has high affinity for ATP (Misic et al, 2010; McCallum et al, 2017). The ATPase contains one open interface for every two closed interfaces, conferring the hexameric structure its characteristic elongated appearance. As ATPase catalysis takes place, one open interface closes (ATP bound), a closed interface remains closed (ATP catalysed to ADP) and the other closed interface opens (ADP released), maintaining the same interface pattern (McCallum et al, 2017). These conformational changes occur in a directional manner, giving the impression that the hexamer pore is rotating. They are propagated clockwise for the extension ATPase and counter-clockwise for the retraction ATPase (McCallum et al, 2019; Ellison et al, 2022). The directionality of the conformational changes has significant implications for the mechanism of pili polymerisation and depolymerisation.

Both motor ATPases interact with the polytopic membrane protein PilG, which stimulates their ATPase activity (Bischof et al, 2016). Universally present in T4P machineries

(Berry and Pelicic, 2015), PilG has three transmembrane domains and two homologous globular cytoplasmic domains (McCallum et al, 2019) and forms a dimer in the cytoplasmic membrane (Karuppiah et al, 2010). According to cryo-ET, the T4aP machinery accommodates a single PilG dimer which is positioned on top of the cytoplasmic ATPases. The PilG N-terminal cytoplasmic domain interacts with the extension ATPase PilF, while its C-terminus interacts with the retraction ATPase PilT (Takhar et al, 2013). PilG is also proposed to bind to the major pilins in the cytoplasmic membrane via its transmembrane helices (Chang et al, 2016; Craig et al, 2019). The network of PilG interactions suggests that it plays a key role in pilin polymerisation and depolymerisation and T4P dynamics (McCallum et al, 2019).

1.2.4 Type IV Pili Dynamics

The extension/retraction dynamics of T4P are critical for their functionality. Mature pilins do not have a cytoplasmic domain, and filament polymerisation occurs on the periplasmic side of the cytoplasmic membrane. This raises the question of how the mechanical energy generated by the cytoplasmic ATPases is relayed to the pili to enable their extrusion or collapse. While it is widely accepted that the inner membrane platform protein PilG plays a central role in T4P dynamics, the precise mechanisms are still under debate.

The most recent model for T4P dynamics is based on the structural characterisation of the *Geobacter metallireducens* assembly ATPase (McCallum et al, 2017). According to McCallum et al, PilG interacts with the lumen of the ATPase hexamers via its cytoplasmic domains, while its transmembrane helices interact with the nascent T4P filament. As the extension ATPase binds and hydrolysis ATP, its hexameric pore undergoes clockwise

conformational changes which lead to the translation and rotation of the bound PilG. PilG is thrusted upwards in the inner membrane and must rotate 60° in order to settle back in the ATPase lumen. Each upwards movement is proposed to extrude a mature major pilin subunit from the membrane and integrate it in the growing filament, whereas the rotational motion positions PilG for the uptake of the next pilin subunit (McCallum et al, 2017). This would result in a right-handed helical filament with a 60° twist which is consistent with observations in the literature (Ellison et al, 2022). The retraction ATPase is proposed to operate via opposing dynamics. PilT conformational changes pull PilC downwards and induce a 60° counter-clockwise rotation, thereby facilitating the removal of major subunits one at a time from the base of the pilus fibre.

Cryo-electron tomography of the *M. xanthus* T4aP has built on this model, elaborating the role of PilG in assembly (Chang et al, 2016). It is proposed that only one PilG dimer can be accommodated in the cryo-ET structure of T4P machinery which is consistent with a one-start assembly mechanism (Chang et al, 2016). PilG is anticipated to interact with the tapered tip of the nascent filament and to anchor the pilus to the T4P machinery. This creates a complete binding pocket for the hydrophobic α1N helices of incoming major pilin subunits. ATPase-powered translation and rotation of PilG allow it to scoop a single major pilin subunits from the membrane and add it onto the base of the nascent filament while pushing the pilus upwards to free the binding pocket for the next major pilin subunit (Chang et al, 2016).

A more recent review has put forward an alternative 'compression model' in which the platform protein PilG controls the access of the pilin subunits to the filament (Craig et al, 2019). In its open conformation, the PilG dimer allows pilins to dock at the base of the pilus. The extension ATPase triggers compression of the PilG transmembrane helices, the

filament helices and the α1N helix of the incoming major pilin subunit, resulting in pilin polymerisation. The compression is subsequently relieved by the extrusion of the now elongated filament, clearing an opening for the next subunit to dock. This model does not require translational and rotational movements, but just like the other two models, it is yet to be supported by direct biochemical evidence (Craig et al, 2019).

1.2.5 Type IV Filaments

T4P belong to a superfamily of evolutionary related molecular nanomachines, called type IV filaments (T4F) (Berry and Pelicic, 2015). T4F are associated with a large diversity of functions, but their biogenesis depends on the same four conserved proteins – a type IV pilin, a prepilin peptidase, a membrane platform protein and an ATPase (Beeby, 2019). T4F are universal in prokaryotes, most likely as a result of their functional versatility and ancient nature (Berry and Pelicic, 2015). Phylogenetic analysis has revealed that T4F were present in the last universal common ancestor to bacteria and archaea which diverged more than three billion years ago (Denise et al, 2019).

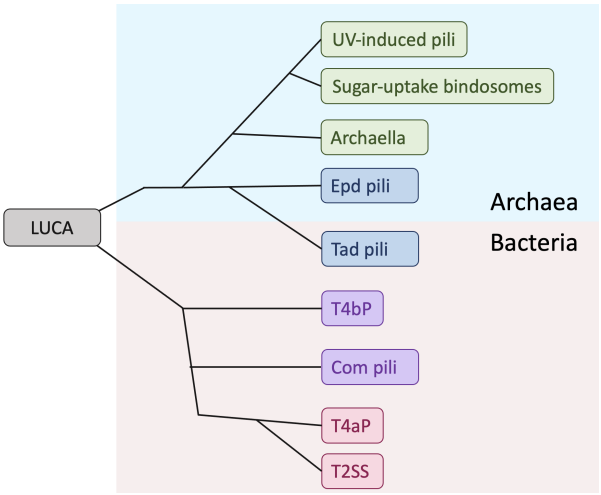


Figure 1.10: **Phylogenetic analysis of the T4F superfamily.** The upper major clade represents archaea. The lower major clade represents bacteria. Tad pili were a horizontal gene transfer from archaea to bacteria. LUCA = last universal common ancestor; T2SS = type II secretion system; Tad = tight adherence pili; T4P = type IV pili; Epd = Eppa-dependent pili; Com = competence pili (Denise et al, 2019).

Based on their origin, T4F are classed as archaeal or bacterial. The archaeal branch divides into two major clades – the first features the rotary archaella, the archaeal adhesive pili, the sugar-uptake bindosomes and the UV-inducible pili, while the second includes the Epd pili and the Tad pili which were discussed earlier. The bacterial branch is characterised by the duplication of the ATPase to produce a retraction ATPase and the acquisition of secretin. The T4bP branched first from the bacterial progenitor, confirming the validity of the T4P sub-division, and were followed by the competence (Com) pili (Figure 1.10). Then, the type II secretion system (T2SS) diverged from the T4aP. In fact, the T2SS evolved multiple times within the T4F superfamily. This suggests that T2SS are relatively modern adaptations, and their evolution is perhaps connected to the recent emergence of host-associated lifestyles among bacteria (Denise et al, 2019).

T4F perform a vast array of functions such as adhesion, motility, protein secretion and DNA uptake. Their functional versatility is likely to have contributed to their ubiquity among prokaryotes (Berry and Pelicic, 2015). Indeed, the original T4F superfamily progenitor is postulated to have been involved in DNA uptake which played a critical role during the early evolution of cellular life (Beeby, 2019). T4P were the first of the T4F to be discovered, and this section offers a brief overview of several other archaeal and bacterial T4F variants and their functions.

1.2.5.1 Archaellum

The archaellum is a cell-surface structure in archaea that powers swimming motility by rotating. Although the archaellum is a functional analogue of the bacterial flagellum, it is evolutionary related to the type IV pilus (Chaudhury et al, 2018). The archaellum is composed of 7-15 proteins encoded in one operon. Its ATPase FlaI, its polytopic platform

membrane protein FlaJ, and its prearchaellin peptidase FlaK bear strong homology to T4P components (Figure 1.11). The archaellin filament is a polymer of archaellins which are synthesised in a precursor form, bearing a class III signal peptide, and are processed by the peptidase FlaK (Ghosh and Albers, 2011; Braun et al, 2016). Unlike the flagellum, the archaellum is assembled by polymerisation of new subunits at the base, and it is not unusual for the archaellum to include more than one archaellin subunit (Ghosh and Albers, 2011). While archaellins have a typical type IV pilin structure, structural models suggest that their hydrophobic α-helices pack more tightly in the filaments and do not exhibit the partially melted N-terminal helix seen in T4P. Additionally, unlike other type IV pilins, they undergo N-linked glycosylation once embedded in the cytoplasmic membrane. The N-linked glycosylation is mediated by the AglB glycosyltransferase and is crucial for the stability of the archaellin subunits (Lassak et al, 2012). Some have suggested that the glycosylation might also facilitate species recognition and distinction between self and non-self species (Voisin et al, 2005).

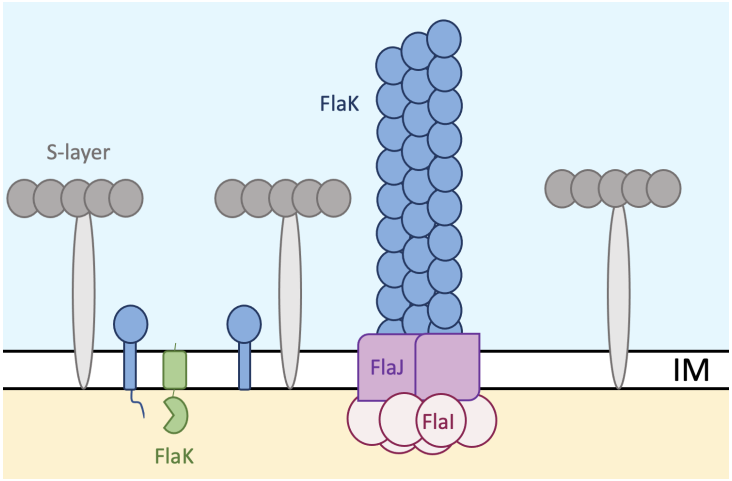


Figure 1.11: **Schematic of the archaellum assembly machinery.** FlaB is the archaellin, processed by prepilin peptidase FlaK. FlaJ is the membrane platform protein and FlaH is the hexameric ATPase. Adapted from Jarrell and Albers, 2012. IM = inner membrane.

1.2.5.2 Sugar-Uptake Bindosome

Another archaeal T4F is the sugar-uptake bindosome (Figure 1.12) which enables the thermophile and acidophile *Sulfolobus solfataricus* to grow on a broad variety of substrates (Ng et al, 2008). The bindosome filament is proposed to be a polymer of substrate-binding proteins (SBP) which feature a class III signal peptide and are processed by a prepilin peptidase PibD (Albers et al, 2003). It is assembled by the *bas* (bindosome assembly system) operon. *bas* encodes three type IV pilin-like proteins, called BasA, BasB and BasC, the BasE ATPase and the BasF integral membrane protein (Zolghadr et al, 2007). BasE and BasF form a stable protein complex that can be extracted from the *S. solfataricus* membrane, while SBP appear to form high molecular mass complexes with the main S-layer protein SlaA (Zolghadr et al, 2011). Deletion of the *basEF* genes alters cell morphology and distorts the S-layer, indicating that an interaction takes place between the S-layer and the bindosome. Our understanding of the sugar-uptake bindosome is limited by lack of visualisation, but its operon resembles a simplified ancestral form of T4F machineries (Zolghadr et al, 2011).

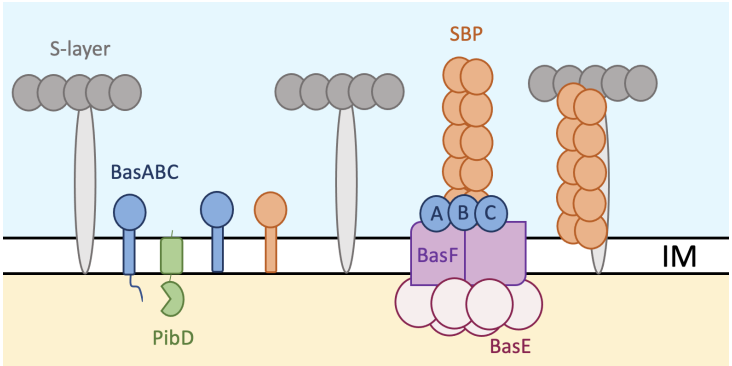


Figure 1.12: **Schematic of the bindosome assembly machinery.** The pilin-like proteins BasABC and the substrate-binding proteins (SBP) are processed by prepilin peptidase PibD. BasF is the membrane platform protein and BasE is the hexameric ATPase (Ng et al, 2008). IM = inner membrane.

1.2.5.3 Competence Pilus

In some naturally competent organisms, including *S. sanguinis*, DNA uptake is mediated by a T4P evolutionary related structure called the competence (Com) pilus. In *B. subtilis*, nearly all components of the Com pilus are encoded in a single *comG* operon. There are five pilins – the major pilin ComGC and the minor pilins ComGD, ComGE, ComGF and ComGG – which are expressed in a precursor form and processed by a dedicated prepilin peptidase ComC before they are assembled into pili (Figure 1.13). The *B. subtilis* major pilin is stabilised by an intramolecular disulphide bond introduced by two protein oxidoreductases BdbD and BdbC. The Com machinery also includes the ComGA ATPase, the membrane platform protein ComGB and a protein with no known orthologues in other systems, called ComGF (Chen and Dubnau, 2004). Although no retraction ATPase is encoded in the *comG* locus, the *B. subtilis* Com pili can generate forces in excess of 40 pN and transport DNA into the cell in linear fashion (Dubnau and Blokesch, 2019). Recent research into motor-independent retraction of T4F has revealed that the pilins of the Com pili and other motor-independent T4F machineries encode large residues within their α1C helix which introduce instability into the assembled filaments. Motor-independent retraction is proposed to rely, in part, on this instability of the filament subunits which could facilitate the spontaneous depolymerisation of stalled pili (Chlebek et al, 2021).

In *B. subtilis*, the Com pilus binds to exogenous DNA and guides it to the membrane bound receptor ComEA. Deletion of ComEA or its two helix-hairpin-helix motifs abolishes DNA uptake. ComEA is proposed to bind the DNA and deliver it to the channel formed by the polytopic membrane protein ComEC via a Brownian ratchet mechanism (Dubnau and Blokesch, 2019). The uptake process might be powered by the ComFA ATPase or by proton motive force. This remains unclear as uncoupling poisons interrupt DNA

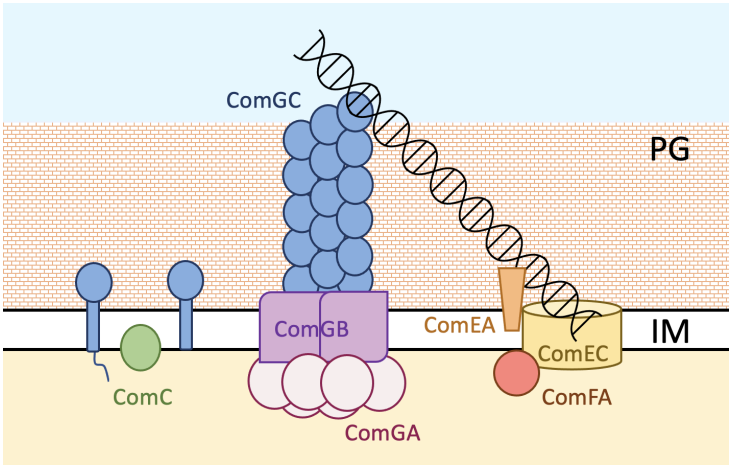


Figure 1.13: **Schematic of the competence pilus assembly machinery.** The major pilin ComGC is processed by the prepilin peptidase ComC. The platform protein ComGB and the ATPase ComGA facilitate pseudopilus assembly. The pilus facilitates access of exogenous DNA to the membrane-bound receptor ComEA which then delivers the bound DNA to the channel protein ComEC. The ATP-binding protein ComFA is postulated to be involved in DNA transport (Chen and Dubnau, 2004) IM = inner membrane; PG = peptidoglycan layer.

translocation before the ATP pool is depleted, but at the same time mutation of ComFA leads to a severe reduction in transformation frequency (Dubnau and Blokesch, 2019). During DNA uptake, only one strand of the DNA molecule enters the cytoplasm, while the other strand is degraded and released into the surrounding environment (Chen and Dubnau, 2004). In *B. subtilis*, the ComEC protein is predicted to contain both a DNA-binding domain and a β-lactamase domain which could mediate the degradation of the second strand (Draskovic and Dubnau, 2005). The incoming single-stranded DNA is then integrated into the bacterial genome in a RecA-dependent manner (Fernandez et al, 2000).

1.2.5.4 Type II Secretion System

The T2SS is highly conserved in Gram-negative bacteria with high prevalence among bacterial pathogens (Douzi et al, 2012). It transports fully folded proteins across a dedicated secretin channel in the outer membrane into the extracellular medium (Figure 1.14). The function of these effector proteins is extremely diverse and includes a large array of toxins, lipases, proteases and enzymes that break down complex carbohydrates. Overall, T2SS

effectors contribute to the environmental growth and virulence of human, animal and plant pathogens (Korotkov and Sandkvist, 2019).

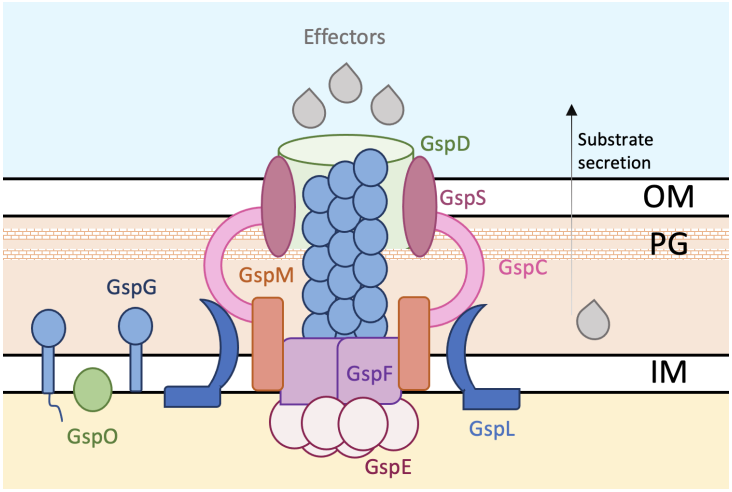


Figure 1.14: **Schematic of the T2SS assembly machinery.** The major pseudopilin GspG is processed by the prepilin peptidase GspO. The platform protein GspF, the assembly sub-complex GspCLM and the ATPase GspE facilitate pseudopilus assembly. The GspD secretin forms a pore that allows the secretion of substrates (Korotkov and Sandkvist, 2019). IM = inner membrane; OM = outer membrane; PG = peptidoglycan layer.

The T2SS components are encoded in one large operon, containing 12-16 genes. Mutation in any of these genes interferes with the secretion function and leads to accumulation of exoproteins in the periplasm. The T2SS machinery is organised in three large sub-complexes (Thomassin et al, 2017). The secretin sub-complex is located in the outer membrane and is composed of multiple copies of the secretin GspD and its lipoprotein pilotin GspS. The inner membrane sub-complex includes GspC, GspF, GspL and GspM as well as the cytoplasmic ATPase GspE which interacts with the cytoplasmic domains of GspF and GspL (Py et al, 2001). Together, these two sub-complexes support the assembly of the filamentous sub-complex which is made up of five proteins that are homologous to the type IV pilins and are called pseudopilins. GspG is the major pseudopilin, while GspH, GspI, GspJ and GspK are minor pilins (Thomassin et al, 2017). The pseudopilins are produced as precursors and processed by the prepilin peptidase GspO. As discussed above, the minor pilins GspI, GpsJ and GspK are proposed to form a trimer complex that

sits at the tip of the T2SS pseudopilus (Korotkov and Hol, 2008) and is connected to the poly-GspG filament via the minor pilin GspH (Yanez et al, 2008b). Since the T2SS only spans the periplasmic space, it has been impossible to visualise. Nevertheless, when the major pseudopilin is overexpressed, surface-exposed filaments called hyper-pseudopili can be observed (Sauvonnet et al, 2000). A recent model of a *K. oxytoca* pseudopilus based on a cryo-EM reconstruction suggests that the T2SS has a right-handed helical fibre with a 10 Å rise (Lopez-Castilla et al, 2017).

Protein secretion by the T2SS occurs in two steps. First, effector proteins are exported across the inner membrane by the Sec or Tat machineries into the periplasm where they undergo folding and/or oligomerisation. Some effector proteins require a dedicated folding chaperone, and some express a removable propeptide which functions as an intramolecular chaperone (Korotkov and Sandkvist, 2019). Second, the folded effectors engage with the T2SS pseudopilus in the periplasm and are translocated across the outer membrane in the extracellular medium. The T2SS displays high substrate-specificity – the effectors of one T2SS are not recognised by another. Since folding of the substrate is essential for secretion, the high specificity is attributed to a secretion signal formed in the folded structure. This common conformational signal is yet to be identified within the great structural variety of T2SS effectors (Douzi et al, 2012).

Currently, there are two competing models for how proteins are transported through the secretin pore by the pseudopilus. In the piston model, the pseudopilus pushes the substrates through the secretin in a linear fashion. In the Archimedes screw model, a rotary motion via interaction with the pseudopilus threads the substrates out through the pore (Korotkov and Sandkvist, 2019).

1.2.6 Type IV Pili Functions

The functional versatility of T4P is one of their defining features (Pelicic, 2019). They can perform a wide range of functions, ranging from adhesion to protein secretion. Some of these properties are dependent on the ability of T4P to retract, which makes them unique among bacterial pili (Berry and Pelicic, 2015).

1.2.6.1 Adhesion

Adhesion to a diverse range of biotic and abiotic surfaces is the most commonly reported property of T4P (Giltner et al, 2012). Attachment to and colonisation of host cells and the host extracellular matrix during bacterial pathogenesis makes T4P key virulence factors and a clinically relevant topic of research (Berry and Pelicic, 2015). Several different components of the T4P can contribute to their adhesiveness. Firstly, the major type IV pilin has been reported to mediate adhesion in several human pathogens. In *EPEC*, the major subunit, bundlin, acts as a lectin and binds to N-acetyllactosamine glycan residues on human intestinal epithelial cell receptors (Hyland et al, 2008). Similarly, the major pilin GP25 in the type IV pili of the Gram-positive *Ruminococcus albus* mediates adhesion to cellulose in the gastrointestinal tract of ruminants (Pegden et al, 1998; Rakotoarivonina et al, 2002). Secondly, the minor pilins have also been shown to possess adhesive properties. For example, in *N. meningitidis*, the non-core minor pilin PilV interacts with the sialylated N-glycans on the human endothelial cell receptor CD147 and the β2-adrenergic receptor (Bernard et al, 2014; Guennec et al, 2020; Barnier et al, 2021). Thirdly, T4P adhesion can be mediated by non-pilin proteins associated with the filaments. This has been reported for the conserved PilC/PilY1 adhesins present in many Gram-negative species. In *P. aeruginosa*, PilY1 is essential for adhesion to host epithelial cells (Heiniger et al, 2010).

It contains an arginine-glycine-aspartic acid (RGD) motif which interacts with integrin in a calcium-dependent manner (Orans et al, 2010; Johnson et al, 2011). Interestingly, T4P retraction can contribute to their adhesive function. In *N. gonorrhoeae*, attachment to epithelial cells leads to the activation of stress and cytoprotective signalling pathways in the presence of a fully functional retraction ATPase PilT (Howie et al, 2005). Likewise, *N. meningitidis* exerts retraction forces on the β2-adrenergic receptor glycans, triggering the formation of membrane protrusions around the adhering bacteria and enhancing their ability to resist the shear forces generated by blood flow (Mikaty et al, 2009; Virion et al, 2019).

1.2.6.2 Interbacterial Contacts

T4P are also involved in contacts between neighbouring bacteria via pilus-pilus contacts. T4P tend to aggregate laterally into bundles, promoting the formation of bacterial aggregates or microcolonies, which can subsequently develop into biofilms (Craig et al, 2019). Bacterial aggregation is an important part of bacterial pathogenesis and can be supported by the major pilin or minor pilins. In *V. cholerae*, mutations in the major pilin TcpA cause defects in microcolony formation and in the colonisation of an infant mouse model (Chiang et al, 1995; Kirn et al, 2000). Indeed, field-emission microscopy analysis has revealed that T4P-mediated autoaggregation of *V. cholerae* during infection leads to the formation of T4P matrices which envelop the bacteria and protect them from host antimicrobial responses (Krebs et al, 2011). In *N. meningitidis*, the minor pilin PilX is responsible for the formation of microcolonies (Helaine et al, 2005). Its D-region forms surface-exposed hooks which allow PilX pilins from interacting filaments to brace against each other to resist retraction, thereby stabilising interbacterial contacts (Helaine et al, 2007). As in adhesion, T4P retraction plays an important role in microcolony formation,

controlling the local order, viscosity and shape (Craig et al, 2019).

1.2.6.3 Twitching Motility

While other bacterial pili and fimbriae also support cellular interactions, T4P is unique in its ability to mediate twitching motility (Giltner et al, 2012). Twitching motility is powered by T4P retraction (Merz et al, 2000). For twitching motility to occur, bacteria extend their T4P, attach to a surface and then retract them, pulling themselves towards the site of attachment (Berry and Pelicic, 2015). Twitching motility is characterised by jerky movements (Eriksson et al, 2015). It consists of a linear motion, the result of the synchronous pulling of multiple T4P, and a slingshot motion, which is due to the release of a single filament while the others are still under tension (Jin et al, 2011). Twitching motility supports directed movement – chemotaxis towards phospholipids and long-chain fatty acids in *P. aeruginosa* (Miller et al, 2008) and phototaxis in the cyanobacterium *Synechocystis* (Bhaya et al, 2000). It also facilitates random surface exploration, including 2D exploration when bacteria are attached vertically to the surface by one end and walk omnidirectionally (Gibiansky et al, 2010). Furthermore, twitching motility can take place intracellularly – the pili of *P. aeruginosa* allow the bacteria to traverse human epithelial multilayers and efficiently exit invaded epithelial cells (Nieto et al, 2019). While twitching motility is considered to be an exclusive property of T4aP due to their motor-dependent retraction, a recent study has suggested that *C. crescentus* Tad pili may enable single cells to adopt an upright position and to move a very short distance away from their mother cell (Sangermani et al, 2019). This suggests that a form of T4P-mediated motility may exist outside the T4aP subclass.

1.2.6.4 DNA Uptake

Another property of T4P that is proposed to have evolved as a functional consequence of twitching motility is DNA uptake (Ellison et al, 2022). DNA uptake is the first step of natural transformation – imported DNA is typically used as a template for DNA repair, as a means of generating genetic diversity or as a food source (Berry and Pelicic, 2015). DNA uptake is not a common function of T4P – only a small subset of bacteria with retractile T4P mediate DNA uptake. The mechanism of DNA uptake was recently revealed by observing the fluorescently-labelled T4aP of *V. cholerae* during natural transformation. T4P bind to DNA via the tip and bring the DNA to the cell surface by PilT-mediated retraction, leading to DNA internalisation (Ellison et al, 2018). The interaction with the DNA is thought to be mediated by positively charged residues in two of the minor pilins of the *V. cholerae* T4aP (Ellison et al, 2018) and the minor pilin ComP in *Neisseria* (Cehovin et al, 2013). ComP is required for efficient DNA binding by T4P and interacts with specific DNA uptake sequences. It has an electropositive stripe that is exposed on the surface of the filaments and forms a new DNA-binding motif (Cehovin et al, 2013; Berry et al, 2016). Although DNA uptake in T4P is dependent on PilT-mediated retraction (Berry and Pelicic, 2015), Tad pili in *Mirococcus luteus* have been shown to be involved in genetic transformation as well (Angelov et al, 2015).

1.2.6.5 Less Common Functions

The functional versatility of T4P is emphasised by some properties that are not as widespread. For instance, initial characterisation of T4P retraction stemmed from the discovery that T4P can serve as receptors for bacterial viruses. Bacteriophages can bind to the side or the tip of the filaments and rely on pili retraction to bring them in contact

with their cognate cell-surface receptor (Bradley, 1972; Berry and Pelicic, 2015). Additionally, both retractile and non-retractile pili can act as secretion systems, extending the evolutionary homology between T4P and T2SS (discussed above). Cycles of polymerisation and depolymerisation can facilitate the recruitment of proteins from the periplasm and their delivery to the opening of the secretin channel and/or their expulsion through the channel (Craig et al, 2019). Lastly, type IV pili have also been observed to aid in the flotation of picocyanobacteria and their ability to evade predation by grazers (Aguilo-Ferretjans et al, 2021).

1.3 The Type IV Pili of *S. sanguinis*

The presence of type IV pili in *S. sanguinis* was first inferred by their functionality. Studies in the 1970s showed that *S. sanguinis* demonstrates a form of motility consistent with twitching motility. This phenotype was observed as thin spreading zones around bacteria grown on agar plates and could be easily seen by the naked eye (Henriksen and Henrichsen, 1975; Henriksen and Eriksen, 1976). In 2016, our lab obtained six of the *S. sanguinis* isolates from this study and re-tested their twitching motility ability. Several of the strains exhibited spreading zones, with 2908 showing the biggest spread (Gurung et al, 2016). Therefore, strain 2908 was selected for further T4P analysis and is the strain used in this research project as well.

1.3.1 The *pil* Locus

Whole genome sequencing of the 2908 strain revealed that all of the proteins involved in T4P biogenesis and function are clustered in a 22-kb long *pil* locus (Gurung et al, 2016). This locus (Figure 1.15) is conserved in most *S. sanguinis* strains with the exception of

SK49 and is absent in all other streptococcal species which are also incapable of twitching motility. The GC content of the *pil* locus is similar to the rest of the genome, suggesting that the cluster was not acquired by recent horizontal gene transfer (Gurung et al, 2016).

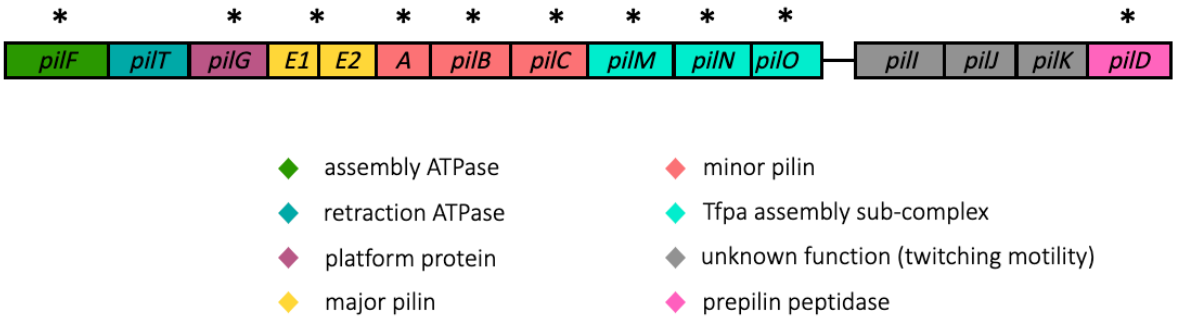


Figure 1.15: ***S. sanguinis pil* locus.** Genes encoding proteins with an essential role for T4P biogenesis are marked by an asterisk. The genes are colour-coded according to predicted function. Genes that lack recognisable motifs but are important for T4P function are depicted in grey. Six genes (SSV_2229-SSV_2294), following *pilD*, are omitted from this diagram as they appear to have no role in T4P biology (Gurung et al, 2016; Gurung et al, 2017).

The *pil* locus encodes 21 proteins, most of which are conserved in other sequenced *S. sanguinis* strains (Gurung et al, 2016). Proteins with known functions have been named after their homologues in *N. meningitidis*: PilD, PilE1, PilE2, PilF, PilG, PilM, PilN, PilO and PilT, while genes bearing no recognisable T4P sequences follow alphabetical order using the remaining letters (Gurung et al, 2016). The six genes at the end of the cluster SSV_2229-SSV_2224 appear to have no involvement in T4P biology (Gurung et al, 2017). The *pil* locus comprises the set of proteins universally conserved in all T4F: a major pilin (PilE1 and PilE2), a prepilin peptidase (PilD), a cytoplasmic traffic ATPase (PilF) to power pilus assembly and a platform polytopic membrane protein (PilG), as well as the assembly complex (PilM-PilN-PilO) which is conserved in T4aP and T2SS.

Systematic deletion of the *pil* genes has revealed that only 10 proteins are required for T4P biogenesis (Gurung et al, 2016). The 10 proteins include PilF, PilG, PilE1/E2, PilA, PilB, PilC, PilM, PilN, PilO and PilD. This makes the *S. sanguinis* T4P apparatus significantly simpler than the well-characterised Gram-negative T4P models (Gurung et

al, 2016). Like in other Gram-positive bacteria, the simplicity of the *S. sanguinis* T4P machinery can be explained by the fact that all of the components involved in the late stages of T4P biogenesis and/or localisation to the outer membrane are absent (Pelicic, 2019).

A further four proteins (PilI, PilJ, PilK and PilT) are dispensable for T4P assembly but required for twitching motility (Gurung et al, 2016). The PilT retraction ATPase powers the twitching motility of *S. sanguinis*: the Δ*pilT* mutant does not form any spreading zones on agar plates, unlike the WT strain, which exhibits train-like directional motion with occasional marked kinks. The tensile forces generated upon PilT-mediated pilus retraction are in the range of 70-100 pN, similar to those in Gram-negative T4P-expressing species (Gurung et al, 2016). Like the Δ*pilT* mutant, Δ*pilI*, Δ*pilJ* and Δ*pilK* exhibit no twitching motility on agar plates. However, Δ*pilK* is motile at a cellular level. It moves differently to the WT, in crab-like motion perpendicular to the long axis of the bacterial chain, and frequently reverses gear (Gurung et al, 2016). PilK is therefore proposed to steer the trains of cells to ensure that net macroscopic movement occurs.

1.3.2 Major Pilins

S. sanguinis pili are unusual in that they are composed of more than one major pilin – strain 2908 has two major pilins (PilE1 and PilE2), while half of the sequenced *S. sanguinis* isolates have three major pilins (Gurung et al, 2016). PilE1 and PilE2 are of similar sizes (16.9 kDa and 16.2 kDa, respectively) and sequences. The first 110 residues of the two major pilins are identical, which means that they share 78% sequence homology. PilE1 and PilE2 can be classified as type IVa prepilins even though their leader peptides (18 aa) are much longer than those of canonical T4aP pilins (Figure 1.16A).

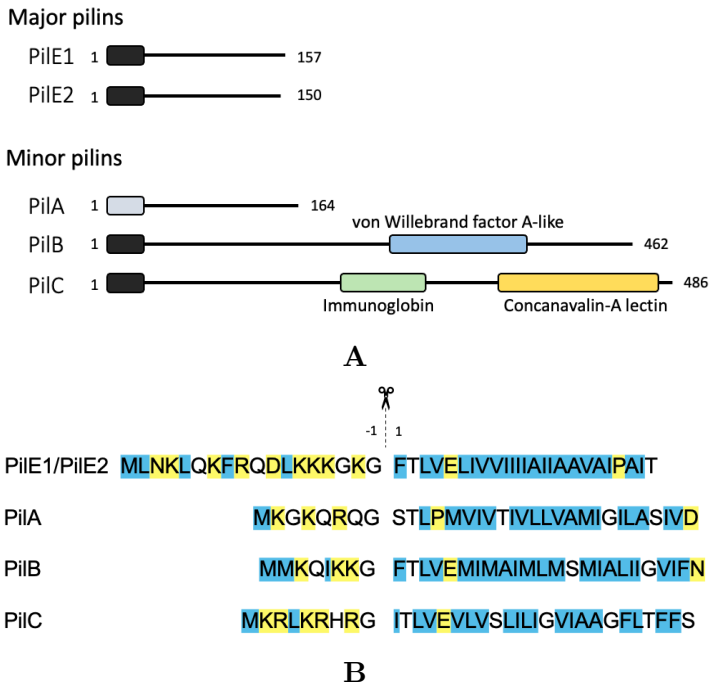


Figure 1.16: ***S. sanguinis* major and minor pilins.** (A) Schematic representation of the major and minor pilins. The class III signal peptide is presented as a square. The grey square signifies the divergent class III signal peptide of PilA. The C-terminal modular domains of PilB and PilC are highlighted in different colours. (B) Aligned sequences of the class III signal peptides of the major and minor pilins. Hydrophobic residues are highlighted in blue, and hydrophilic residues are highlighted in yellow. The dotted line marks the site of the prepilin peptidase PilD cleavage after the conserved glycine at the end of the leader peptide (Berry et al, 2019).

Both PilE1 and PilE2 possess the same class III signal peptide and are processed by the prepilin peptidase PilD (Berry et al, 2019). This results in two mature pilins of 14.7 and 14 kDa, respectively (Gurung et al, 2016). The class III signal peptide contains the conserved Gly-1 and Glu5 residues. Following widely conserved T4P principles, mutating the glycine residue into serine abolishes PilE1 prepilin processing, while substituting the glutamate residue prevents the processed PilE1 from being assembled into filaments. PilE1 and PilE2 also undergo N-terminal methylation by the PilD enzyme. Other post-translational modifications, frequently detected on Gram-negative major pilins, are absent from PilE1 and PilE2 (Berry et al, 2019).

Structural characterisation of PilE1 by NMR revealed a classical type IV pilin fold (Figure 1.17). PilE1 has a long N-terminal α-helix packed against a β-meander motif made

up of three anti-parallel β-strands. The most unusual feature of PilE1 is its flexible C-terminus. Typically, the stretch of amino acids at the C-terminus of a pilin is 'stapled' to the last β-strand by a disulphide bond, a calcium-binding site or a network of hydrogen bonds (Berry and Pelicic, 2015). But in PilE1, the 10 amino acid-long C-terminus is unstructured and highly flexible. The structure of PilE2 is expected to be identical because of the high sequence similarity to PilE1. Interestingly, PilE1 shows closest homology to the evolutionary related T2SS major pseudopilins, specifically PulG in *K. oxytoca* (Berry et al, 2019).

PilE1 and PilE2 are incorporated together into heteropolymeric pili in nearly equal ratios according to mass spectrometry (Berry et al, 2019). Neither of the major pilins is essential for piliation. Pili made up of a single major pilin, however, are less abundant than WT pilus preparations and promote reduced twitching motility velocities. The Δ*pilE1*Δ*pilE2* double mutant fails to produce pili (Gurung et al, 2016).

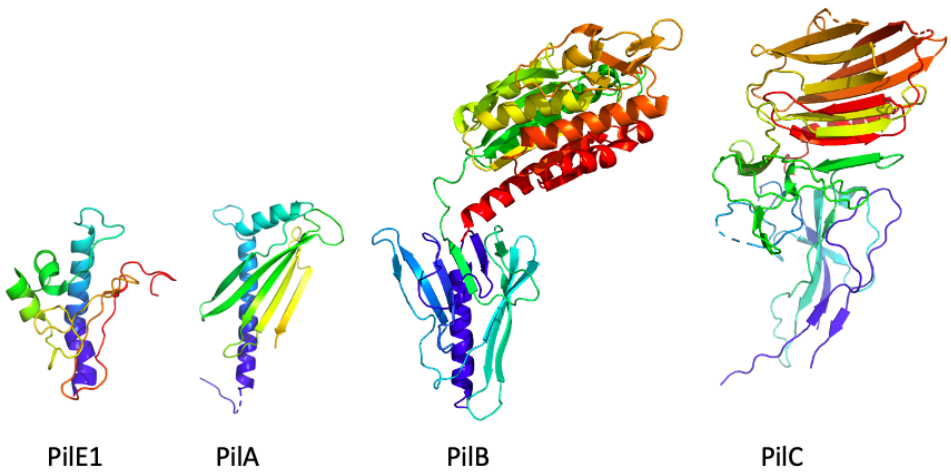


Figure 1.17: **Cartoon representations of the *S. sanguinis* major and minor pilins.** The NMR structure of the major pilin PilE1 (PDB: 6i2o). The crystal structures of the minor pilin PilA (Dr Berry, unpublished data), minor pilin PilB (PDB: 7b7p), minor pilin PilC (Dr Sheppard, unpublished data). The structure of minor pilin PilC is missing the pilin domain.

1.3.3 Minor Pilins

S. sanguinis also encodes three minor pilins – PilA, PilB and PilC (Gurung et al, 2016). All three minor pilins undergo prepilin processing by PilD and are incorporated in the filaments (Berry et al, 2019). PilA, PilB and PilC are essential for T4P biogenesis since their deletion leads to loss of piliation (Gurung et al, 2016). However, the role of the minor pilins in T4P assembly and their localisation within the filaments remains unclear. One hypothesis is that the three proteins contribute to filament stability and modulate the functionality of the T4P (Berry et al, 2019).

PilA is a small pilin protein (17 kDa for the processed pilin) with a degenerate class III signal peptide. It has no detectable sequence homologues outside of the *S. sanguinis* species and is not recognised by the bioinformatic softwares for type IV pilin identification, such as PilFind (Figure 1.16B). In its divergent class III signal peptide, the Gly-1 residue that precedes the leader peptide cleavage site is conserved, and PilA undergoes processing by the PilD prepilin peptidase (Berry et al, 2019). The first residue of the mature protein, however, is a serine, instead of a phenylalanine; there is an unusual hydrophilic proline residue at position 4; and the highly conserved Glu5, typically required for filament assembly, is absent (Figure 1.16B). The structure of the soluble domain of PilA was solved by X-ray crystallography before the beginning of this research project by Dr Berry (Figure 1.17). PilA demonstrates the classical pilin 'lollipop' shape, comprised of an α1C helix stick and a C-terminal globular domain. The globular domain contains five anti-parallel β-strands and is connected to the α1C helix through an additional short C-terminal α-helix, which is referred to as α2C. This short α-helix creates the striking 'hammer-like' shape of PilA. The structure of the minor pilin PilA will be discussed in more detail later.

In contrast, PilB and PilC are unusually large for processed pilins (50.5 kDa and 52.8 kDa, respectively) with canonical class III signal peptides (Berry et al, 2019). They are classified as modular pilins because of the bulky modules that have been 'grafted' by evolution onto their small pilin domains (Raynaud et al, 2021). Their large C-terminal modules are thought to contribute to the functionality of the *S. sanguinis* T4P and in particular, host adhesion (Figure 1.16B). The C-terminal domain of the minor pilin PilC belongs to the concanavalin A-like lectin/glucanase domain family (InterPro entry IPR013320) and is hypothesised to engage sugar residues on glycosylated host cell receptors, thereby mediating adhesion (Berry et al, 2019). The 3D structure of PilC has also been solved by X-ray crystallography by Dr Sheppard (Figure 1.17). Notably, the pilin moeity is missing from the structure of the modular minor pilin PilC, most likely due to stability issues which will be discussed in detail further below. The crystallised PilC fragment presents two modules grafted onto a pilin domain – an N-terminal and a C-terminal module, making PilC a trimodular minor pilin. The N-terminal module is composed of six β-strands in total and one very short α-helix. The C-terminal module comprises two short two-stranded β-sheets, followed by a β-sandwich, which consists of two opposed anti-parallel β-sheets each made up of six β-strands, and ends with a short α-helix. The structure of the minor pilin PilC will be discussed in more detail later.

Of the three minor pilins, only the bimodular PilB has been extensively characterised both structurally and functionally. According to bioinformatic and structural analysis, its C-terminal domain belongs to the von Willebrand factor A-like (vWA) domain superfamily (InterPro entry IPR036465) (Berry et al, 2019; Raynaud et al, 2021). The prototypical vWA domain is found in the von Willebrand Factor (VWF), a human protein required in the physiological process that prevents and impedes bleeding. X-ray crystallography has

revealed that the first 180 residues of the processed PilB display a type IV pilin fold with a long N-terminal α-helix packed against a globular pilin domain made up of two consecutive β-sheets consisting of six and three β-strands, respectively (Raynaud et al, 2021). The vWA domain, linked to the pilin module via a short, flexible loop (Figure 1.17), presents high structural similarity to other vWA-containing proteins despite sharing little sequence homology. It has a metal ion-dependent adhesion site (MIDAS) which binds to Mg^{2+} and Mn^{2+} ions. The vWA domain mediates adhesion to eukaryotic cells: the binding of the piliated MIDAS mutant to Chinese Hamster Ovary cells is severely reduced compared to WT. PilB also interacts directly with several human proteins, including fibronectin and fibrinogen in a MIDAS-independent manner (Raynaud et al, 2021). Subsequently, the minor pilin PilB from the *S. sanguinis* strain SK36 contributes to the invasion of immortalised human aortic endothelial cells and platelet-dependent biofilm formation via its MIDAS (Martini et al, 2021). These recently published data confirm the role of minor pilin PilB in adhesion and suggest a significant involvement of the *S. sanguinis* T4P in the bacterial pathogenesis of infective endocarditis.

1.4 Project Aims

The precise involvement of minor pilins in T4P pilus biogenesis and functionality has been challenging to characterise because of their multiplicity and high diversity (Giltner et al, 2012). Typically, minor pilins are divided into two categories: core and non-core minor pilins (Jacobsen et al, 2020). Non-core minor pilins are dispensable for pili assembly, but they tend to contribute to T4P functionality, while core minor pilins are highly conserved and essential for the formation of pili – their deletion impedes T4P assembly and leads to loss of piliation (Jacobsen et al, 2020).

Since the comprehensive characterisation of the structure and function of minor pilins has been hampered by their high numbers and diversity in diderm bacteria, the less complicated T4P machinery of the Gram-positive *S. sanguinis* might represent a more promising model for the study of minor pilins (Gurung et al, 2016). *S. sanguinis* encodes only three minor pilins – PilA, PilB and PilC – which is significantly fewer than in most species (Berry et al, 2019). The simplicity of the *S. sanguinis* T4P could facilitate the exhaustive characterisation of the role of each pilus subunit, which to date, has not been achieved in any T4P system. As PilB has been described elsewhere (Raynaud et al, 2021), here we focus on the two remaining minor pilins, PilA and PilC. Thus, the aims of this project are fourfold:

1. To understand how the divergent class III signal peptide of the *S. sanguinis* minor pilin PilA supports its assembly into filaments, T4P biogenesis and functionality by performing mutational analysis of conserved and divergent residues in the N-terminus of PilA (Chapter 3).

2. To determine whether the minor pilins of *S. sanguinis* form a multimeric complex by performing large-scale studies of the interactions between all pilus subunits in *S. sanguinis* (Chapter 4).

3. To characterise the putative minor pilin complex biophysically, functionally and structurally (Chapter 5).

4. To characterise the lectin module of the minor pilin PilC by establishing its affinity for a range of glycan moieties, mapping out its carbohydrate-recognition domain and determining its contribution to *S. sanguinis* eukaryotic cell adhesion (Chapter 6).

Chapter 2

Materials and Methods

2.1 Microbiological Techniques

2.1.1 Bacterial Strains and Growth Conditions

The *Streptococcus sanguinis* 2908 strains used in this study are listed in Table 2.1. *S. sanguinis* bacteria are facultative anaerobes (Gurung et al, 2016). *S. sanguinis* were grown on Todd-Hewitt (TH) broth 1% (weight/volume) agar plates at 37°C in a 3.5 L anaerobic jar with an Anaerogen sachet (Thermo Scientific). Between 10-20 ml distilled water were added to the bottom of the jar to maintain humidity. Alternatively, *S. sanguinis* were grown statically under normal atmospheric conditions at 37°C in liquid TH broth, supplemented with 0.05% Tween-80 (THT) to limit bacterial clumping. Towards the end of this project, the THT broth was further supplemented with 100 mM HEPES (THTH) to regulate the pH of overnight cultures. Antibiotics (5 μg/ml erythromycin or 500 μg/ml kanamycin) were added as required.

Chemically competent *Escherichia coli* DH5-α cells were used for cloning purposes, while BL21 (DE3) cells were used for protein expression. *E. coli* BTH101 were used in bacterial adenylate cyclase two-hybrid assays. *E. coli* were grown at 37°C either in Lysogeny Broth (LB) medium shaking at 180 revolutions per minute (rpm), or on LB 1% (w/v) agar plates. Antibiotics (100 μg/ml ampicillin, 50 μg/ml kanamycin or 100 μg/ml spectinomycin) were added as required.

Both *E. coli* and *S. sanguinis* strains were stored long-term in LB or TH, respectively, with 25% glycerol at -80°C. The optical density (OD_{600}) of liquid cultures was measured with a Biochrom Libra S11 spectrophotometer.

***S. sanguinis* 2908**	**Antibiotic Resistance**	**Reference**
WT	–	Henriksen and Henrichsen, 1975
Δ*pilA*	kanamycin	Gurung et al, 2016
Δ*pilB*	kanamycin	Gurung et al, 2016
Δ*pilC*	kanamycin	Gurung et al, 2016
Δ*pilD*	kanamycin	Gurung et al, 2016
Δ*pilT*	erythromycin	Gurung et al, 2016
Δ*pilA:pheS*-aphA3*	kanamycin	Pelicic lab
Δ*pilB:pheS*-aphA3*	kanamycin	Pelicic lab
Δ*pilC:pheS*-aphA3*	kanamycin	Pelicic lab
$pilA_{G\text{-}1A}$	–	This study
$pilA_{G\text{-}1S}$	–	This study
$pilA_{G\text{-}1R}$	–	This study
$pilA_{G\text{-}1W}$	–	This study
$pilA_{S1A}$	–	This study
$pilA_{S1F}$	–	This study
$pilA_{T2A}$	–	This study
$pilA_{L3A}$	–	This study
$pilA_{P4A}$	–	This study
$pilB_{G\text{-}1S}$	–	This study
$pilB_{G\text{-}1W}$	–	This study
$pilC_{G\text{-}1S}$	–	This study
$pilC_{K358A}$	–	This study
$pilC_{CRD7}$	–	This study
$pilC_{CRD12}$	–	This study

Table 2.1: ***S. sanguinis* strains used in this study.**

2.1.2 Preparation of *E. coli* Competent Cells

E. coli DH5-α and BL21 (DE3) competent cells were prepared using the Hanahan's method (Hanahan et al, 1991). Bacteria from frozen glycerol stocks were streaked onto a LB agar plate and incubated at 37°C. A single colony was picked from the plate to inoculate 5 ml LB and grown overnight at 37°C with shaking at 180 rpm. The following day, 1 ml of the culture was back-diluted in 50 ml LB media and grown at 37°C until the OD_{600} reached 0.6. The cells were then harvested by centrifugation at 4,000 rpm for 10 minutes at 4°C and gently resuspended in 25 ml of sterile ice-cold 0.1 M $CaCl_2$. After a 15-minute rest on ice, the cells were centrifuged again and gently resuspended in 3.3 ml

sterile ice-cold 0.1 M $CaCl_2$ buffer, containing 15% glycerol. The cell suspension was left to rest on ice for 4 hours before it was divided into 0.1 ml aliquots in pre-chilled sterile 1.5 ml Eppendorf tubes. The competent cells were stored long-term at -80°C.

2.1.3 Preparation of *E. coli* Ultracompetent Cells

E. coli BTH101 ultracompetent cells were prepared using the Inoue method (Inoue et al, 1990). Bacteria from frozen glycerol stocks were streaked onto a LB agar plate and incubated overnight at 37°C. The next day, 10-12 large colonies from the plate were used to inoculate 250 ml of SOB (Difco) and grown at 16°C while shaking. Approximately 40 hours later, when the OD_{600} reached 0.6, the culture was incubated on ice for 10 minutes in order to cool it down. The cells were then harvested by centrifugation at 2,500 rpm for 10 minutes at 4°C and gently resuspended in 80 ml of ice-cold Inoue transformation buffer (ITB) (Table 2.14). Following a 10-minute rest on ice, the cells were centrifuged once again and gently resuspended in 20 ml ice-cold ITB, containing 7% dimethyl sulfoxide (DMSO). The cell suspension was incubated on ice for 10 minutes before being divided into 0.4 ml aliquots in pre-chilled sterile 1.5 ml Eppendorf tubes. The ultracompetent cells were transferred to -80°C for long-term storage.

2.1.4 *E. coli* Transformation

Thawed chemically competent *E. coli* cells were mixed with 200 ng plasmid or 2-5 μl ligation mixture and incubated on ice for 30 minutes. The bacteria were heat-shocked for 30-45 seconds at 42°C and then rested on ice for 2 minutes. The transformed cells were grown in 600 μl LB for 1 hour before being plated on LB agar, supplemented with an appropriate antibiotic. Transformed colonies were verified by colony PCR. The transfor-

mation protocol for bacterial adenylate cyclase two-hybrid assays is described below.

2.2 DNA Recombination Techniques

2.2.1 DNA Preparation

Plasmids were extracted from 5 ml overnight *E. coli* cultures using Qiaprep spin mini-prep kit (Qiagen) according to manufacturer's instructions. Genomic DNA (gDNA) was extracted from 2 ml overnight *S. sanguinis* cultures with the XIT Genomic DNA from Gram-Positive Bacteria kit (G-Biosciences) according to manufacturer's instructions. DNA concentration and quality were assessed by NanoDrop (Thermo Fischer Scientific) and agarose gel electrophoresis, respectively. Plasmids and gDNA were stored long-term at -20°C.

2.2.2 Agarose Gel Electrophoresis

DNA quality and size were analysed by agarose gel electrophoresis. Agarose mixture of 1% (w/v) was prepared with 1x Tris-Borate-EDTA buffer (Seven Biotech Ltd) and melted by boiling in a microwave. SYBR Safe DNA stain (1:20,000) was added to the agarose mixture and set in casting trays. DNA samples were mixed with 6x loading dye ((0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF, 30% (v/v) glycerol in water) and run alongside the size marker Gene Ruler 1 kb Plus DNA ladder (Thermo Scientific) for 15-30 minutes at 100-120 V using a Mini-Sub Cell GT electrophoretic system (Biorad). DNA was visualised using a Gel Doc EZ Imager (BioRad).

Name	Sequence (5' to 3')
Sequencing Primers	
TOPO GW1	GTTGCAACAAATTGATGAGCAATGC
TOPO GW2	GTTGCAACAAATTGATGAGCAATTA
T7 promoter	TAATACGACTCACTATAGGG
T7 terminal	GCTAGTTATTGCTCAGCGG
pET Upstream	ATGCGTCCGGCGTAGA
Duet Down1	AACATGTGCCGGCGTATTAG
Duet Up2	TTGTACACGGCCGCATAATC
***S. sanguinis* Markerless Mutagenesis**	
pilA F1	ATTATCGCCATCATCGCAGC
pilA R2	TCCCTGCAAATCCCAATCCA
pilB F1	TACAACTGGACCGAAGCTGG
pilB R2	TTTGGCCTATCGTTCCCACT
pilC F1	GTGGGCTTCTCTACTTCAGC
pilC R2	CACTAGCTGCCACATCAACC
pilA-Topo F	CACAAGTGCACCAGGTTCTG
pilA-Topo R	CATACTGCATGGAGGTTCGC
pilA-G-1A F	GGGAAACAAAGACAAGCGTCTACTTTGCCCATGG
pilA-G-1A R	CCATGGGCAAAGTAGACGCTTGTCTTTGTTTCCC
pilA-G-1S F	GAAAGGGAAACAAAGACAATCGTCTACTTTGCCCATGGTTATTG
pilA-G-1S R	CAATAACCATGGGCAAAGTAGACGATTGTCTTTGTTTCCCTTTC
pilA-G-1R F	GAAAGGGAAACAAAGACAACGGTCTACTTTGCCCATGG
pilA-G-1R R	CCATGGGCAAAGTAGACCGTTGTCTTTGTTTCCCTTTC
pilA-G-1W F	GAAAGGGAAACAAAGACAATGGTCTACTTTGCCCATGG
pilA-G-1W R	CCATGGGCAAAGTAGACCATTGTCTTTGTTTCCCTTTC
pilA-S1A F	GGGAAACAAAGACAAGGGGCTACTTTGCCCATGGTTATTG
pilA-S1A R	CAATAACCATGGGCAAAGTAGCCCCTTGTCTTTGTTTTCCC
pilA-S1F F	GGGAAACAAAGACAAGGGTTTACTTTGCCCATGGTTATTG
pilA-S1F R	CAATAACCATGGGCAAAGTAAACCCTTGTCTTTGTTTTCCC
pilA-T2A F	GGGAAACAAAGACAAGGGTCTGCTTTGCCCATGGTTATTGTAAC
pilA-T2A R	GTTACAATAACCATGGGCAAAGCAGACCCTTGTCTTTGTTTCCC
pilA-L3A F	CAAAGACAAGGGTCTACTGCGCCCATGGTTATTGTAACC
pilA-L3A R	GGTTACAATAACCATGGGCGCAGTAGACCCTTGTCTTTG
pilA-P4A F	GACAAGGGTCTACTTTGGCCATGGTTATTGTAACC
pilA-P4A R	GGTTACAATAACCATGGCCAAAGTAGACCCTTGTC
pilB-G-1S F	GGAATGATGAAGCAAATAAAAAAATCGTTTACCCTAGTCGAGATGATTATG
pilB-G-1S R	CATAATCATCTCGACTAGGGTAAACGATTTTTTTATTTGCTTCATCATTCC
pilB-G-1W F	GGAATGATGAAGCAAATAAAAAAATGGTTTACCCTAGTCGAGATGATTATG
pilB-G-1W R	CATAATCATCTCGACTAGGGTAAACCATTTTTTTATTTGCTTCATCATTCC
pilC-G-1S F	GAAAAGATTAAAGAGACACAGATCGATTACACTGGTCGAAGTCTTG
pilC-G-1S R	CAAGACTTCGACCAGTGTAATCGATCTGTGTCTCTTTAATCTTTTC
pilC-K358A F	GACCGGACCGCTTCTAACCTATGCACTAGATGATACACTAACTTGG
pilC-K358A R	CCAAGTTAGTGTATCATCTAGTGCATAGGTTAGAAGCGGTCCGGTC
pilC R1	CGTACGGCCATCAAGTGCTA
pilC F2	GTGCTGCGAAGAATGGCTC
pilC-CRD F	TTATGGTGCGACTATCCCTGT
pilC-CRD R	GTTTTATTATAGACTGATTGTCC
Bacterial Adenylate Cyclase Two-Hybrid Assays	
pilA F	CGCGGATCCCATGAAAGGGAAACAAAGACAAG
pilA R	CGCGGTACCGCCTTCTGTGCCGATCTCAAATG
$pilA_{T258C}$ F	CGAGCGGGCCCATAACGTGGACCCTTCAACAGAGCCGATTAAG
$pilA_{T258C}$ R	CTTAATCGGCTCTGTTGAAGGGTCCACGTTATGGGCCCGCTCG

pilB F	CGCGGATCCCATGAAGCAAATAAAAAAAGGGTT
pilB R	CGCGGTACCGCCGGCCCTGATACAAACCACA
pilC F	CGCGGATCCCATGAAAAGATTAAAGAGACACAG
pilC R	CGCGGTACCGCTACTTATCGCGGAAGTATTCA
pilE1 F	CGCGGATCCCATGTTAAACAAATTACAAAAATTCCG
pilE1 R	CGCGGTACCGCGTTTGAGTTTACACCATTAGCA
pilE2 F	CGCGGATCCCATGTTAAACAAATTGCAAAAATTCCG
pilE2 R	CGCGGTACCGCTTTTGAATTAGCACCAGCTTC
Strep II tag	
pET28b Strep II-pilA F	GGGCCATGGATTGGAGCCACCCGCAGTTCGAAAAGGATACAGGGCAAAGCCAGAC
pET28b pilA R	CCCGTCGACTTACTTCTGTGCCGATCTCAA
pET28b Strep II-pilB F	GGGCCATGGATTGGAGCCACCCGCAGTTCGAAAAGAGCAGCCGTGAACTGATTGA
pET28b pilB R	CCCGGATCCTTACGGACCGCTAACAAACC
pET28b Strep II-pilC F	GGGCCATGGATTGGAGCCACCCGCAGTTCGAAAAGAGCGGCAGCTTTAATAACATTCTGCG
pET28b pilC R	CCCGGATCCTTAGCTTGCTTTGTATTTATCGC
pET28b Strep II-pilE1 F	GGGCCATGGATTGGAGCCACCCGCAGTTCGAAAAGCAAGATAACGCTCGTAAGAGCC
pET28b pilE1 R	CCGGATCCTTAGTTTGAGTTTACACCATTAGCAGA
pET28b Strep II-pilE2 F	GGGCCATGGATTGGAGCCACCCGCAGTTCGAAAAGCAAGATAACGCTCGTAAGAGCC
pET28b pilE2 R	CCGGATCCTTATTTTGAATTAGCACCAGCTTCG
PilC$_{254stop}$	
pET28b His-pilC F	GGGCCATGGATCATCATCATCATCATCATAGCGGCAGCTTTAATAACATTCTGCG
pet28b pilC$_{254stop}$ R	CCCGGATCCTTACAGGCCCATAATCCAAATACGATC
pETDuet1 Strep II-pilC F	GGGCATATGGATTGGAGCCACCCGCAGTTCGAAAAGAGCGGCAGCTTTAATAACATTCTGCG
pETDuet1 pilC$_{254stop}$ R	CCCCCTAGGTTACAGGCCCATAATCCAAATACGATC
PilA Interface Mutants	
pET28b His-pilA$_{A52S}$ F	GGCAATCCATCATCGGTTTATTCAGAGTTTGACGAGCGGGCCC
pET28b His-pilA$_{A52S}$ R	GGGCCCGCTCGTCAAACTCTGAATAAACCGATGATGGATTGCC
pET28b His-pilA$_{K69A}$ F	GATCCTTCAACAGAGCCGATTGCGTCAACCCATACCTTCAAAG
pET28b His-pilA$_{K69A}$ R	CTTTGAAGGTATGGGTTGACGCAATCGGCTCTGTTGAAGGATC
pET28b His-pilA$_{T71A}$ F	CAACAGAGCCGATTAAGTCAGCCCATACCTTCAAAGATGGC
pET28b His-pilA$_{T71A}$ R	GCCATCTTTGAAGGTATGGGCTGACTTAATCGGCTCTGTTG
PilC Sugar-Binding Mutants	
pET28b His-pilC$_{T356A}$ F	CAGCAGACAGGTCCGCTGCTGGCATATAAACTGGATGATACCC
pET28b His-pilC$_{T356A}$ R	GGGTATCATCCAGTTTATATGCCAGCAGCGGACCTGTCTGCTG
pET28b His-pilC$_{K358A}$ F	GGTCCGCTGCTGACATATGCACTGGATGATACCCTGAC
pET28b His-pilC$_{K358A}$ R	GTCAGGGTATCATCCAGTGCATATGTCAGCAGCGGACC
pET28b His-pilC$_{T364A}$ F	CTGGATGATACCCTGGCCTGGGGTATTAATC
pET28b His-pilC$_{T364A}$ R	GATTAATACCCCAGGCCAGGGTATCATCCAG
pET28b His-pilC$_{N368A}$ F	CCCTGACCTGGGGTATTGCTCTGGAAAATGATGGTC
pET28b His-pilC$_{N368A}$ R	GACCATCATTTTCCAGAGCAATACCCCAGGTCAGGG
pET28b His-pilC$_{E370A}$ F	CCTGGGGTATTAATCTGGCAAATGATGGTCGCATTGC
pET28b His-pilC$_{E370A}$ R	GCAATGCGACCATCATTTGCCAGATTAATACCCCAGG
pET28b His-pilC$_{R374A}$ F	GGTATTAATCTGGAAAATGATGGTGCCATTGCCATCAAAACCGTTGATAC
pET28b His-pilC$_{R374A}$ R	GTATCAACGGTTTTGATGGCAATGGCACCATCATTTTCCAGATTAATACC
pET28b His-pilC$_{G389S}$ F	CCACCACCGCAAATAATGGTTCTCAAGAATATATCCAGAACG
pET28b His-pilC$_{G389S}$ R	CGTTCTGGATATATTCTTGAGAACCATTATTTGCGGTGGTGG
pET28b His-pilC$_{Q390A}$ F	CACCGCAAATAATGGTGGTGCAGAATATATCCAGAACGTG
pET28b His-pilC$_{Q390A}$ R	CACGTTCTGGATATATTCTGCACCACCATTATTTGCGGTG
pET28b His-pilC$_{Y392A}$ F	CAAATAATGGTGGTCAAGAAGCTATCCAGAACGTGAAACTGG
pET28b His-pilC$_{Y392A}$ R	CCAGTTTCACGTTCTGGATAGCTTCTTGACCACCATTATTTG

Table 2.2: **Primers used in this study.**

2.2.3 Polymerase Chain Reaction

Polymerase chain reaction (PCR) was used to amplify target DNA regions for subsequent cloning or sequencing. The primers used in this study are shown in Table 2.2. High-fidelity PCR was performed either with Pfu Ultra II DNA polymerase or Herculase II Fusion DNA polymerase (Table 2.3 and Table 2.4) for subsequent cloning or DNA sequencing.

Taq DNA polymerase was used in analytical PCR and colony PCR to confirm successful transformation (Table 2.3 and Table 2.4). For colony PCR, single colonies of transformed *E. coli* DH5-α cells were picked from a plate with a pipette tip, re-streaked on a different plate and incubated into the Taq reaction mixture.

	Pfu (μl)	**Herculase (μl)**	**Taq (μl)**
Buffer	5	10	5
dNTPs	0.4	0.5	0.4
Forward Primer (25 μM)	0.5	0.5	0.5
Reverse Primer (25 μM)	0.5	0.5	0.5
DNA (100 ng)	1	1	1
DNA Polymerase	1	0.5	0.5
H_2O	41.6	37	42.1
Total:	50	50	50

Table 2.3: **PCR mixtures for each DNA polymerase.**

Temperature	**Pfu**	**Herculase**	**Taq**	**Cycles**
95°C	2 min	2 min	2 min	1
95°C	20 sec	20 sec	30 sec	
50°C	20 sec	20 sec	30 sec	30
72°C	15 sec/kb	30 sec/kb	1 min/kb	
72°C	10 min	10 min	5 min	1
10°C	∞	∞	∞	

Table 2.4: **PCR cycling parameters for each DNA polymerase.**

2.2.4 DNA Purification

Amplified PCR products, splicing PCR products, digested DNA fragments and DNA plasmids were purified either directly with the QIAquick PCR Purification kit (Qiagen)

or following agarose gel separation, excised from the gel and purified with QIAquick Gel Extraction kit (Qiagen).

2.2.5 Cloning

2.2.5.1 Into TOPO

Cloning into the pCR8/GW/TOPO vector (Invitrogen) generated plasmids suitable for subsequent site-directed mutagenesis and/or sub-cloning into a different vector. Cloning in TOPO occurred via the addition of 3' single-stranded adenosine ends (A-overhangs) to the insert DNA, following its PCR amplification using a high-fidelity enzyme. Addition of 3' A-overhangs was performed by incubating the reaction mixture from Table 2.5 for 10 minutes at 72°C in a thermocycler. The insert DNA and the pCR8/GW/TOPO vector were then ligated according to manufacturer's instructions. Between 2-5 μl of the ligation mixture was used to transform *E. coli* DH5-α cells, and transformed cells were selected on spectinomycin-supplemented plates. The TOPO plasmids used in this study are listed in Table 2.6.

Reaction components	**μl**
Purified PCR product	34.8
10x Polymerase Buffer	5
1 mM dATP	10
Taq DNA Polymerase	0.2

Table 2.5: **Reaction mixture for adding 3' overhangs.**

2.2.5.2 Into Other Vectors

Sub-cloning was performed in order to facilitate protein expression (pET28b, pETDuet-1, pUT18C and pKT25 vectors). Cloning into other vectors required the digestion of insert DNA from the TOPO plasmid as well as the digestion of the target vector. DNA

digestion was carried out with suitable restriction endonucleases (RE) from NEB according to manufacturer's instructions and verified by agarose gel electrophoresis. The digested DNA was purified either directly from the digestion reaction or from excised agarose gel fragments.

The insert DNA and target vector were ligated using the Ready-to-Go DNA Ligase kit (Amersham Biosciences). For cloning into the pET28b, pETDuet-1, pUT18c and pKT25 vectors, 30 fmoles of vector and insert were mixed in a total volume of 20 μl. The reaction mixture was then transferred into a tube of Ready-to-Go DNA ligase, mixed and incubated at room temperature for 5 minutes. Ligation was performed at 16°C for 30 minutes. Between 2-5 μl of the ligation mixture was used to transform *E. coli* DH5α cells.

Plasmid Name	Description of Insert DNA	Antibiotic	Reference
pCR8/GW/TOPO	Original vector	spec	Invitrogen
***S. sanguinis* Mutagenesis**			
TOPO *pilA*	*pilA*, amplified with pilA-Topo F and pilA-Topo R primers	spec	This study
TOPO *pilB*	*pilB*, amplified with pilB F1 and pilB R2 primers	spec	Pelicic lab
TOPO *pilC*	*pilC*, amplified with pilC F1 and pilC R2 primers	spec	Pelicic lab
TOPO $pilA_{G\text{-}1A}$	Mutant *pilA* allele: G-1A	spec	This study
TOPO $pilA_{G\text{-}1S}$	Mutant *pilA* allele: G-1S	spec	This study
TOPO $pilA_{G\text{-}1R}$	Mutant *pilA* allele: G-1R	spec	This study
TOPO $pilA_{G\text{-}1W}$	Mutant *pilA* allele: G-1W	spec	This study
TOPO $pilA_{S1A}$	Mutant *pilA* allele: S1A	spec	This study
TOPO $pilA_{S1F}$	Mutant *pilA* allele: S1F	spec	This study
TOPO $pilA_{T2A}$	Mutant *pilA* allele: T2A	spec	This study
TOPO $pilA_{L3A}$	Mutant *pilA* allele: L3A	spec	This study
TOPO $pilA_{P4A}$	Mutant *pilA* allele: P4A	spec	This study
TOPO $pilB_{G\text{-}1S}$	Mutant *pilB* allele: G-1S	spec	This study
TOPO $pilB_{G\text{-}1W}$	Mutant *pilB* allele: G-1W	spec	This study
TOPO $pilC_{G\text{-}1S}$	Mutant *pilC* allele: G-1S	spec	This study
TOPO $pilC_{K358A}$	Mutant *pilC* allele: K358A	spec	This study
Bacterial Adenylate Cyclase Two-Hybrid Assays			
TOPO *pilA*	Full-length WT *pilA* with 5' BamHI and 3' KpnI restriction sites	spec	This study
TOPO $pilA_{T258C}$	Full-length *pilA* missing its interior BamHI restriction site	spec	This study
TOPO *pilB*	Full-length WT *pilB* with 5' BamHI and 3' KpnI restriction sites	spec	This study
TOPO *pilC*	Full-length WT *pilC* with 5' BamHI and 3' KpnI restriction sites	spec	This study
TOPO *pilE1*	Full-length WT *pilE1* with 5' BamHI and 3' KpnI restriction sites	spec	This study
TOPO *pilE2*	Full-length WT *pilE2* with 5' BamHI and 3' KpnI restriction sites	spec	This study
Others			
TOPO Strep II-$pilC_{254stop}$	NT-Strep II-tagged codon-optimised $pilC_{254stop}$	spec	This study

Table 2.6: **TOPO plasmids used in this study.**

Plasmid Name	Description of Insert DNA	Antibiotic	Reference
pET28b	T7-based expression vector	kan	Novagen
His_6 tag			
pET28b His_6-*pilA*	NT His_6-tagged truncated *pilA* (residues 33-164)	kan	Pelicic lab
pET28b His-*pilB*	NT His_6-tagged truncated codon-optimised *pilB* (residues 36-461)	kan	Pelicic lab
pET28b His_6-*pilC*	NT His_6-tagged truncated codon-optimised *pilC* (residues 34-486)	kan	Pelicic lab
pET28b His_6-$pilC_{121\text{-}486}$	NT His_6-tagged codon-optimised *pilC* Ig and lectin domain (residues 121-486)	kan	Pelicic lab
pET28b His_6-$pilC_{254\text{-}485}$	NT His_6-tagged codon-optimised *pilC* lectin domain (residues 254-486)	kan	This study
pET28b His_6-$pilC_{254stop}$	NT His_6-tagged codon-optimised *pilC* pilin and Ig fold domain (residues 34-253)	kan	This study
pET28b His_6-*pilE1*	NT His_6-tagged truncated *pilE1* (residues 28-157)	kan	Pelicic lab
pET28b His_6-*pilE2*	NT His_6-tagged truncated *pilE2* (residues 28-150)	kan	Pelicic lab
pET28b His_6-$pilC_{T356A}$	NT His_6-tagged $pilC_{121\text{-}486}$ with a T356A mutation	kan	This study
pET28b His_6-$pilC_{K358A}$	NT His_6-tagged $pilC_{121\text{-}486}$ with a K358A mutation	kan	This study
pET28b His_6-$pilC_{T364A}$	NT His_6-tagged $pilC_{121\text{-}486}$ with a T364A mutation	kan	This study
pET28b His_6-$pilC_{N368A}$	NT His_6-tagged $pilC_{121\text{-}486}$ with a N368A mutation	kan	This study
pET28b His_6-$pilC_{E370A}$	NT His_6-tagged $pilC_{121\text{-}486}$ with a E370A mutation	kan	This study
pET28b His_6-$pilC_{R374A}$	NT His_6-tagged $pilC_{121\text{-}486}$ with a R374A mutation	kan	This study
pET28b His_6-$pilC_{G389S}$	NT His_6-tagged $pilC_{121\text{-}486}$ with a G389S mutation	kan	This study
pET28b His_6-$pilC_{Q390A}$	NT His_6-tagged $pilC_{121\text{-}486}$ with a Q390A mutation	kan	This study
pET28b His_6-$pilC_{Y392A}$	NT His_6-tagged $pilC_{121\text{-}486}$ with a Y392A mutation	kan	This study
pET28b His_6-$pilC_{CRD7}$	NT His_6-tagged $pilC_{121\text{-}486}$ with seven CRD mutations in the CRD region	kan	This study
pET28b His_6-$pilC_{CRD12}$	NT His_6-tagged $pilC_{121\text{-}486}$ with twelve mutations in the CRD region	kan	This study
Strep II tag			
pET28b Strep II-*pilA*	NT Strep II-tagged truncated *pilA* (residues 33-146)	kan	This study
pET28b Strep II-$pilA_{A52S}$	NT Strep II-tagged truncated *pilA* with a A52S mutation	kan	This study
pET28b Strep II-$pilA_{K69A}$	NT Strep II-tagged truncated *pilA* with a K69A mutation	kan	This study
pET28b Strep II-$pilA_{T71A}$	NT Strep II-tagged truncated *pilA* with a T71A mutation	kan	This study
pET28b Strep II-*pilB*	NT Strep II-tagged truncated codon-optimised *pilB* (residues 36-461)	kan	This study
pET28b Strep II-*pilC*	NT Strep II-tagged truncated codon-optimised *pilC* (residues 34-486)	kan	This study
pET28b Strep II-*pilE1*	NT Strep II-tagged truncated *pilE1* (residues 28-157)	kan	Pelicic lab
pET28b Strep II-*pilE2*	NT Strep II-tagged truncated *pilE2* (residues 28-150)	kan	Pelicic lab

Table 2.7: **pET28b plasmids used in this study.** NT stands for N-terminal.

2.2.6 Site-Directed Mutagenesis

Site-directed mutagenesis (SDM) was performed with the Pfu Ultra II DNA polymerase (Table 2.8 and Table 2.9) according to the instructions of the Quickchange site-directed mutagenesis kit (Stratagene). TOPO plasmids, containing the target gene, were used as a DNA template in varying concentrations. The desired mutations were introduced in the middle of 25-45 bp long primers (Table 2.2). Following PCR, the SDM product was incubated with 1 μl DpnI restriction enzyme (NEB) for 1 hour at 37°C according to manufacturer's instructions. DpnI targets methylated and hemi-methylated DNA, digesting the original DNA template, and thus enriches the sample for newly synthesised (mutated) DNA which is non-methylated. *E. coli* DH5-α cells were then transformed with 2-5 μl of the digestion product and selected on spectinomycin-supplemented LB agar plates. Plasmids were prepared as described above, and the successful introduction

of target mutations was verified by sequencing.

	Pfu Ultra II (μl)
Buffer	5
dNTPs	1
Forward Primer (25 μM)	0.5
Reverse Primers (25 μM)	0.5
DNA (5 ng, 25 ng, 50 ng)	1
DNA Polymerase	1
H_2O	41
Total:	50

Table 2.8: **Site-directed mutagenesis PCR mixtures.**

Temperature	**Duration**	**Cycles**
95°C	2 min	1
95°C	30 sec	
52°C	1 min	16
72°C	30 sec/kb	
72°C	10 min	1
10°C	∞	

Table 2.9: **Site-directed mutagenesis PCR cycling parameters.**

2.2.7 DNA sequencing

The sequences of amplified, cloned or mutated DNA were verified by Sanger sequencing at GATC Biotech. The sequencing samples were prepared by mixing purified PCR products (100-400 ng) or plasmids (400-500 ng) with 1 μl primer (25 μM) and distilled water in a total volume of 10 μl.

2.3 *S. sanguinis* Mutagenesis

S. sanguinis are naturally competent bacteria which possess a competence pilus independent of the type IV pilus (Gurung et al, 2016). This has facilitated systematic genetic analysis through the generation of mutants.

2.3.1 Generation of Markerless *S. sanguinis* Mutants

Primary mutants were used to generate non-polar markerless *S. sanguinis* mutants (Gurung et al, 2017). The primary mutants contain a double cassette, composed of *pheS** and *aphA3*, in place of the target gene, facilitating both negative and positive selection. The *pheS** gene encodes the point mutation A316G within the tRNA synthetase PheS, rendering the bacteria susceptible to the phenylalanine analogue *para*-chlorophenylalanine (PCPA). The A316G mutation allows the incorporation of toxic PCPA in nascent proteins, causing their misfolding and ultimately, a drastic reduction in bacterial growth. Meanwhile, *aphA3* gene confers resistance to kanamycin.

The mutated gene, alongside its upstream and downstream regions, was amplified from the SDM TOPO plasmid and PCR purified. A 2 ml overnight THT-HS culture of the primary mutant was back-diluted 1:500 into 10 ml prewarmed THT-HS and grown for 1-1.5 hours at 37°C. 330 μl of primary mutant culture were mixed with 212 ng/μl competence-stimulating peptide (CSP) and 100 ng recombinant DNA. CSP is a synthetic peptide DLRGIPNPWGWIFGR (Peptide Protein Research) which activates the transcription of competence-associated genes and stimulates the natural competence of *S. sanguinis*. After an hour-long incubation at 37°C, the bacteria were plated on TH agar, supplemented with 15 mM PCPA, and grown anaerobically for 24 hours. Single colonies were picked and re-streaked on kanamycin-supplemented TH agar plates. Transformed bacteria were resistant to PCPA and sensitive to kanamycin. Following genomic DNA extraction, the kanamycin-sensitive colonies were analysed by PCR and agarose gel electrophoresis to verify that the double cassette was replaced by the mutated target gene. The PCR products were also sent for sequencing to verify that the single amino acid mutation was introduced.

2.3.2 Splicing PCR

Splicing PCR (sPCR) was used to produce DNA constructs for allelic exchange mutagenesis in the generation of markerless *S. sanguinis* $pilC_{CRD}$ mutants. sPCR consists of two steps. In the primary PCR (Table 2.10 and 2.11), the region 5' to the CRD (including the upstream *pilC* region) and the region 3' to the CRD (including the *pilC* downstream region) were amplified using a high-fidelity DNA polymerase and two sets of primers: primers F1/R1 for the 5' region and F2/R2 for the 3' region. The upstream and downstream regions were required for homologous recombination between the *S. sanguinis* chromosome and the recombinant DNA. The inner primers – R1 and F2 – were used to delete the WT CRD and provide a stretch of complementarity between the 5' region, the 3' region and the mutant CRD. The mutant CRD were amplified from the synthetic genes using the CRD-F and CRD-R primers in a parallel reaction.

1° PCR Components	**Pfu (μl)**	**2° PCR Components**	**Pfu (μl)**
Buffer	5	Buffer	5
dNTPs	0.4	dNTPs	0.4
Outer Primers F1/R2 (25 μM)	1.6	Up region (μl)	3.3
Inner Primers R1/F2 (25 μM)	0.4	Dn region (μl)	3.3
DNA (100 ng)	1	Cassette (μl)	3.3
DNA Polymerase	1	DNA Polymerase	1
H_2O	40.6	H_2O	33.7
Total:	50	Total:	50

Table 2.10: **Primary and secondary sPCR mixtures.**

Temperature	**Duration**	**Cycles**
95°C	2 min	1
95°C	20 sec	
50°C	20 sec	25
72°C	30 sec/kb	
72°C	10 min	1
10°C	∞	

Table 2.11: **sPCR cycling parameters.**

The outer primers F1 and R2, added in excess in the primary PCR reaction, were also

used in the secondary PCR reaction (Table 2.10 and 2.11). The three PCR products from the primary PCR were spliced together via their complementary regions. The splicing reaction was verified by agarose gel electrophoresis before being transformed directly into the *S. sanguinis* Δ*pilC:phes*-aphA3* primary mutant. This generated markerless *S. sanguinis* mutants in which multiple mutations were introduced into the *pilC* carbohydrate-recognition domain.

2.4 *S. sanguinis* Phenotypic Analysis

2.4.1 Type IV Pili Purification

The piliation phenotype of *S. sanguinis* was evaluated by type IV pili purification. An overnight 10 ml THT *S. sanguinis* culture was back-diluted into 90 ml THT and grown for 3-4 hours. Once the OD_{600} reached 1-1.5, the culture was centrifuged at 6,000 *g* for 10 minutes at 4°C. The pellet was resuspended in 2 ml of ice-cold PP buffer (Table 2.14) before being transferred into a small universal tube. The sample was then vortexed at 3,000 rpm for 2 minutes in order to mechanically shear the pili from bacteria. Next, the bacteria were pelleted by two rounds of centrifugation at 6,000 *g* for 10 minutes at 4°C and finally, the supernatant was filtered through a 0.2 μm syringe filter (PALL corporation). The filtered supernatant was divided between two Eppendorf Ultracentrifuge tubes, fine-balanced and ultracentrifuged at 45,000 rpm for 1 hour at 4°C. The two resulting pellets were resuspended in 40 μl PP buffer each and combined. Aliquots were mixed with 2x Laemmli buffer (Merck) and analysed by SDS-PAGE (Table 2.12).

2.4.2 Whole-cell Protein Extracts (WCE)

Whole-cell protein extracts were used to assess whether prepilin proteins undergo processing by the prepilin peptidase. An overnight culture of 10 ml THT *S. sanguinis* was back-diluted into 90 ml THT, grown for 3-4 hours and the OD_{600} was adjusted to 1. The culture was then centrifuged at 6,000 *g* for 10 minutes at 4°C. The bacterial pellet was resuspended in 1 ml of ice-cold PP Buffer (Table 2.14) and transferred into a 2 ml tube containing lysing matrix b, 0.1 mm silica spheres (MPBio). Samples were homogenised in a Fast Prep-24 benchtop vortex (MPBio) at 6.5 speed for five cycles (1 minute on/off) to obtain whole-cell protein extracts. The lysing matrix was settled by gentle centrifugation at 1,000 *g* for 30 seconds, and the lysate was mixed with 2x Laemmli buffer. The WCE was probed by Western blotting for prepilin processing.

2.4.3 Twitching Motility Assay

S. sanguinis bacteria were streaked on TH agar, supplemented with antibiotic as required. The next day, TH plates with 1% twitching motility agar (Eiken chemical co) were prepared fresh and left to solidify on the bench. Similarly sized *S. sanguinis* colonies were picked with a pipette tip and re-streaked onto the Eiken agar TH plates as 1-2 cm long straight lines. The plates were incubated at 37°C anaerobically with 20 ml distilled H_2O at the bottom of the jar for 48-72 hours. The re-streaks were imaged with a Quantum imager.

2.5 Protein Sample Analysis

2.5.1 SDS-Polyacrylamide Gel Electrophoresis

SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) was used to separate protein samples by their size. SDS-PAGE gels were prepared as described in Table 2.12 depending on the molecular weight of the analysed proteins. Gel electrophoresis was carried out with 1x Tris/Glycine/SDS buffer (Bio-Rad) in a Mini-PROTEAN Tetra cell system for 30 min-1 hour at 200-220 V. The Precision Plus Protein All Blue Prestained Protein Standards (Bio-Rad) was loaded alongside the protein samples as a molecular weight marker.

SDS-PAGE gels were stained with Bio-Safe Coomassie stain (Bio-Rad) according to manufacturer's instructions. Following overnight destaining in H_2O, the gels were imaged with the Gel Doc EZ Imager (Bio-Rad). Alternatively, SDS-PAGE was performed as a first step in Western blotting.

Resolving Gel	**8%**	**12%**	**15%**	**Stacking Gel**	**5%**
H2O	4.6 ml	3.3 ml	2.3 ml	H2O	2.1 ml
30% acrylamide	2.7 ml	4.0 ml	5.0 ml	30% acrylamide	0.5 ml
1.5 M Tris pH 8.8	2.5 ml	2.5 ml	2.5 ml	1.0 M Tris pH 6.8	0.38 ml
10% SDS	100 μl	100 μl	100 μl	10% SDS	30 μl
10% APS	100 μl	100 μl	100 μl	10% APS	30 μl
TEMED	4 μl	4 μl	4 μl	TEMED	3 μl

Table 2.12: **Composition of SDS-PAGE gels.**

2.5.2 Western Blotting

After protein samples were separated by SDS-PAGE, they were transferred onto Amersham Hybond ECL nitrocellulose membrane (GE Healthcare). The wet transfer was carried out for 1 hour at 100 V in ice-cold Transfer Buffer (Table 2.14) using the Mini Trans Blot electrophoretic transfer cell system (Bio-Rad). After the transfer, the membrane was

blocked for 1 hour, while shaking, at room temperature in 5% (w/v) skim milk powder in PBS, supplemented with 0.1% Tween 20 (PBST). The membrane was then incubated for 1 hour with a primary antibody (Table 2.13), diluted in 5% milk PBST. Following three 10 minute washes with PBST, the membrane was incubated for 1 hour with an anti-rabbit secondary antibody, conjugated to horseradish peroxidase (HRP) (NA-934, GE Healthcare), diluted (1:10,000) in PBST. The membrane was washed again three times for 10 minutes in PBST. Finally, the membrane was developed with Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare) according to manufacturer's instructions. Protein bands were detected using the ChemiDoc Imaging System (Bio-Rad).

1° Antibody	**Peptide/Protein Used for Production**	**Dilution**	**Host**
anti-PilA	Purified recombinant His_6-tagged PilA	1:3,000	Rabbit
anti-PilB	Purified recombinant His_6-tagged PilB	1:3,000	Rabbit
anti-PilC	Purified recombinant His_6-tagged PilC	1:3,000	Rabbit
anti-PilE1	Peptide $PilE1_{105\text{-}119}$	1:2,000	Rabbit
anti-PilE2	Peptide $PilE2_{126\text{-}138}$	1:2,000	Rabbit

Table 2.13: **Primary antibodies used in this study.**

2.6 Protein Expression and Purification

2.6.1 Protein Expression

E. coli BL21 (DE3) were transformed with >200 ng pET28b derivatives (Table 2.7) as described above. pET28b derivatives encoded the soluble portion of pilins in which the hydrophobic α1N helix was replaced by a non-cleavable N-terminal His_6 tag or Strep II tag. The transformed bacteria were plated on LB agar, supplemented with 50 μg/ml kanamycin. A single colony was grown in 10 ml kanamycin-supplemented LB cultures overnight at 37°C while shaking. The next morning the culture was back-diluted into 1 l

kanamycin-supplemented LB and grown for 3-4 hours at 37°C. Once the OD_{600} reached 0.6-0.8, the culture was cooled down to 16°C and protein expression was induced with 0.1-0.4 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). Following overnight growth at 16°C, the culture was centrifuged at 6,000 g for 30 minutes at 4°C. The pellet was resuspended in 20 ml of Binding Buffer (Table 2.14), containing EDTA-free protease inhibitor mix (Sigma-Aldrich), and frozen at -80°C. A 1 ml sample was taken from the non-induced and induced cultures to verify successful induction of protein expression.

Buffer	**Components**
Inoue Transformation Buffer	55 mM $MnCl_2.4H_2O$, 15 mM $CaCl_2.2H_2O$, 250 mM KCl, 10 mM PIPES pH 6.7
Transfer Buffer	39 mM glycine, 48 mM Tris base, 0.037% SDS, 20% isopropanol
IMAC Binding Buffer	150 mM NaCl, 50 mM HEPES pH 7.5, 20 mM imidazole
IMAC Elution Buffer	150 mM NaCl, 50 mM HEPES pH 7.5, 300 mM imidazole
Strep Binding Buffer	150 mM NaCl, 50 mM HEPES pH 7.6
Strep Elution Buffer	150 mM NaCl, 50 mM HEPES pH 7.6, 2.5 mM desthiobiotin
SEC Buffer	150 mM NaCl, 50 mM HEPES pH 7.5
NMR Buffer	150 mM NaCl, 10 mM Na_2HPO_4/NaH_2PO_4, pH 7.0
PM2 Buffer	70 mM $Na_2HPO_4.12H_2O$, 30 mM $NaH_2PO_4.H_2O$, 1 mM $MgSO_4$, 0.2 mM $MnSO_4$, pH 7.0
PD Blocking Buffer	150 mM NaCl, 50 mM HEPES pH 7.6, 10 mM imidazole, 0.1% Tween-20, 5% skim milk
PD Binding Buffer	150 mM NaCl, 50 mM HEPES, pH 7.6, 10 mM imidazole, 0.1% Tween-20
PD Washing Buffer	300 mM NaCl, 50 mM HEPES, pH 7.6, 20 mM imidazole, 0.1% Tween-20
PD Elution Buffer	300 mM NaCl, 50 mM HEPES, pH 7.6, 500 mM imidazole, 0.1% Tween-20
PP Buffer	50 mM NaCl, 20 mM Tris-HCl pH 7.4
ITC P/P Buffer	150 mM NaCl, 20 mM HEPES pH 7.5
ITC P/S Buffer	50 mM NaCl, 20 mM HEPES pH 7.5

Table 2.14: **Buffers used in this study.**

2.6.2 Cell Lysis

The frozen bacterial pellet was thawed in ice-cold water and lysed by sonication with Sonics Vibra-Cell at 60% amplitude, 5 seconds on/off pulses for 5 minutes. Cell debris was then removed by centrifugation at 11,000 rpm for 30 minutes at 4°C. Samples were taken from the supernatant (soluble fraction) and from the re-suspended pellet (insoluble fraction) to test for protein solubility and stability. The proteins were purified from the cell lysate in two steps: affinity chromatography (AC), followed by size exclusion chromatography (SEC).

2.6.3 Protein Purification

2.6.3.1 Immobilised Metal Affinity Chromatography

Immobilised Metal Affinity Chromatography (IMAC) was performed on Poly-Prep or Econo-Pac Gravity-Flow Chromatography columns (Bio-Rad) loaded with 1 or 2 ml pre-equilibrated Ni-NTA agarose resin (Qiagen), respectively. The cell lysate was mixed with the resin for 30 minutes at 4°C and the flow-through (FT) collected, allowing the column to drain. Next, the nickel-charged resin was washed three times with 5-10 ml IMAC Binding Buffer (Table 2.14). The Binding Buffer contained a low concentration of imidazole to prevent non-specific interactions between proteins in the cell lysate and the column. The His_6-tagged proteins were eluted in three 2 ml fractions with IMAC Elution Buffer (Table 2.14). The eluted fractions, along with the FT and wash fractions, were analysed by SDS-PAGE as described above. The protein concentration was determined by NanoDrop (Thermo Fischer Scientific).

2.6.3.2 Strep II-Tag Affinity Chromatography

Strep II-tag affinity chromatography (AC) was performed on a 5 ml StrepTrap HP column (GE Healthcare) using an AKTA prime system (GE Healthcare). Before use, the column was washed with 5 column volumes (CV) H_2O and equilibrated with 5 CV Strep Binding Buffer (Table 2.14) at a flow rate of 5 ml/min. The cell lysate was loaded onto the column with a superloop. After the FT was collected, the column was washed with 10 CV Strep Binding Buffer. Finally, the Strep II-tagged proteins were eluted with 6 CV Strep Elution Buffer (Table 2.14) in 2 ml fractions. The Strep Elution Buffer contained desthiobiotin (iba), a derivative of biotin, which competes with the Strep II-tag for binding to the StrepTactin sepharose. The eluted fractions were analysed by SDS-PAGE alongside the

FT and wash fractions. The column was regenerated with 3 CV 0.5 M NaOH in order to remove any remaining desthiobiotin before being used for protein purification again.

2.6.3.3 Size Exclusion Chromatography

Size exclusion chromatography (SEC) was performed using an AKTA prime system (GE Healthcare) with either a Superdex S75 16/600, S200 16/600 or S200 10/300 GL column (GE Healthcare). The column, chosen based on the size of the protein to be purified, was equilibrated with 2 CV SEC Buffer (Table 2.14). The loop (100 μl, 2 ml or 5 ml) was selected in accordance with the column and the loading sample volume. The AC elution fractions were pooled together and filtered through 0.2 μm filter (PALL corporation) before being injected through the loop into the column. Fractions of 1-2 ml were collected and analysed by SDS-PAGE (Table 2.12).

2.6.4 Buffer Exchange

Buffer exchange was performed in order to decrease the salt concentration of the protein buffer or to remove interfering reagents such as imidazole and desthiobiotin. PD-10 desalting columns (GE Healthcare) were used according to manufacturer's instructions.

2.6.5 Concentrating Proteins

Protein samples were concentrated using Amicon Ultra Centrifugal Filters (Millipore) according to manufacturer's instructions. PilA and PilE1/E2 were concentrated using the 10 kDa cut-off membrane, while PilB and PilC were concentrated using the 30 kDa cut-off membrane.

2.7 Analysis of Protein-Protein Interactions

Pilin-pilin interactions were tested by bacterial adenylate cyclase two-hybrid assays, β-galactosidase assays and pull-down assays. Any complex formation was further analysed by SEC, followed by SDS-PAGE, or SEC coupled with multi-angle light scattering (SEC-MALS). The affinity of any identified interactions was determined by isothermal titration calorimetry (ITC).

2.7.1 Bacterial Adenylate Cyclase Two-Hybrid Assays

To perform bacterial adenylate cyclase two-hybrid (BACTH) assays, 100 μl of ultracompetent *E. coli* BTH101 cells were co-transformed with 20 ng (each) of recombinant pUT18C and pKT25 plasmids, encoding the full-length major and minor pilin genes of *S. sanguinis* (Table 2.15). After heat-shock treatment, the cells were rested on ice for 2 minutes and allowed to recover for 1 hour in 800 μl SOB with 20 mM glucose at 37°C while shaking at 180 rpm. Twenty μl of the transformed bacteria were plated on MacConkey agar, supplemented with 0.5 mM IPTG, 1% maltose solution (Sigma), 100 μg/ml ampicillin and 50 μg/ml kanamycin. Plates were incubated at 30°C for 48 hours. The colour of the colonies was scored after 40 hours, and the plates were imaged after 48 hours.

Plasmid Name	Description	Antibiotic	Reference
pUT18C	BACTH vector, designed to express a protein fused at its N-terminus to T18	amp	Karimova et al, 1998
pKT25	BACTH vector, designed to express a protein fused at its N-terminus to T25	kan	Karimova et al, 1998
pUT18C-zip	BACTH vector, expressing a leucine zipper domain fused at its N-terminus to T18	amp	Karimova et al, 1998
pKT25-zip	BACTH vector, expressing a leucine zipper domain fused at its N-terminus to T25	kan	Karimova et al, 1998
pUT18C-PilA	BACTH vector, expressing T18-PilA	amp	This study
pKT25-PilA	BACTH vector, expressing T25-PilA	kan	This study
pUT18C-PilB	BACTH vector, expressing T18-PilB	amp	This study
pKT25-PilB	BACTH vector, expressing T25-PilB	kan	This study
pUT18C-PilC	BACTH vector, expressing T18-PilC	amp	This study
pKT25-PilC	BACTH vector, expressing T25-PilC	kan	This study
pUT18C-PilE1	BACTH vector, expressing T18-PilE1	amp	This study
pKT25-PilE1	BACTH vector, expressing T25-PilE1	kan	This study
pUT18C-PilE2	BACTH vector, expressing T18-PilE2	amp	This study
pKT25-PilE2	BACTH vector, expressing T25-PilE2	kan	This study

Table 2.15: **BACTH plasmids used in this study.**

Positive and negative controls were performed with every BACTH assay. For the positive

control, pUT18C-zip and pKT25-zip plasmids were co-transformed into BTH101 cells, which produced deep purple colonies, indicating strong functional complementation between T18 and T25. Co-transformation with pUT18C and pKT25 empty plasmids, on the other hand, served as a negative control and generated white colonies.

2.7.2 β-galactosidase Assays

The efficiency of the functional complementation between T18 and T25 in the positive BACTH phenotypes was quantified by measuring the activity of β-galactosidase in liquid culture (Karimova et al, 1998). Single colonies of co-transformed BTH101 cells were picked from MacConkey agar 48 hours after plating and inoculated in 5 ml LB, supplemented with 0.5 mM IPTG, 100 μg/ml ampicillin and 50 μg/ml kanamycin. The bacteria were grown overnight at 30°C, while shaking at 180 rpm. The next day, the cultures were diluted 1:5 in M63 broth (15 mM $(NH_4)_2SO_4$, 100 mM KH_2PO_4, 1.8 μM $FeSO_4.7H_2O$, 3 μM vitamin B1, pH 7.0) and the OD_{600} was measured.

To permeabilise the cells, 20 μl chloroform and 20 μl 0.1% SDS were added to 1.5 ml of the diluted cultures, vortexed for 10 seconds and incubated for 40 minutes at 30°C with shaking at 250 rpm. Ten μl of the permeabilised cells were then diluted into 990 μl PM2 buffer (Table 2.14), containing 100 mM β-mercaptoethanol, and incubated at 28°C for 5 minutes. The β-galactosidase reaction was started with the addition of 250 μl 0.4% O-nitrophenol-β-galactoside (ONPG) in PM2 buffer (pre-equilibrated at 30°C). The reaction was incubated at 28°C and stopped with 500 μl 1 M Na_2CO_3. The reaction was stopped after 20 minutes for positive samples and 60 minutes for negative samples, and the OD_{420} was measured in 1.5 ml cuvettes. These time points were standardised by Dr Georgiadou (Georgiadou et al, 2012).

The enzymatic activity A (units/ml) was quantified using the following formula:

$$A \; = \; 200 \; x \; (OD_{420}/minutes \; of \; incubation) \; x \; dilution \; factor$$

where minutes of incubation refers to the length of the reaction incubation, and the dilution factor refers to the dilution of the permeabilised cells in the PM2 buffer. The results are expressed as enzymatic activity per milligram of bacterial dry weight (U/mg), so that 1 unit corresponds to 1 nmol of ONPG hydrolysed per minute at 28°C (Karimova et al, 1998). Based on the assumption that 1 ml of culture at an OD_{600} of 1 is equivalent to 300 μg dry weight bacteria, U/mg is calculated as follows:

$$U/mg \; = \; A/[(OD_{600} \; x \; 300)/1000]$$

2.7.3 Pull-Down Assays

Pull-down (PD) assays were carried out with Dynabeads His-tag Isolation and Pulldown (Invitrogen) using purified His_6-tagged and Strep II-tagged soluble pilin domains as bait and prey, respectively. The bait and prey proteins were mixed together in 1 ml of PD Binding Buffer (Table 2.14) and incubated on ice for 1 hour to allow protein-protein interactions to take place before the pull-down was performed.

For each pull-down reaction, 25 μl of magnetic beads and 1.5 nmol of protein (23 μg PilA, 71 μg PilB, 75 μg PilC, 60 μg $PilC_{121\text{-}486}$, 19 μg PilE1, 18 μg PilE2) were used, matching the maximum binding capacity of the beads. The beads were incubated for 1 hour with 700 μl PD Blocking Buffer (Table 2.14), containing 5% (w/v) skim milk powder, on a rotating wheel at 4°C in order to minimise non-specific binding between the soluble pilin

domains and the beads. The blocking reaction mixture was then placed on a magnet to catch the beads and the flow-through was discarded. The beads were rinsed twice with 700 μl PD Binding Buffer (Table 2.14) before being mixed with the 1 ml bait-prey reaction mixture. The proteins and the beads were incubated together for 20 minutes on a rotating wheel at 4°C. The bait-prey reaction mixture was then placed on a magnet to catch the beads and the flow-through was thoroughly removed. The magnetic beads were washed with 500 μl PD Washing Buffer (Table 2.14) 10 times by vortexing for 10 seconds. The PD Binding Buffer and PD Washing Buffer contained 10-20 mM imidazole to reduce non-specific binding of the prey proteins to the beads as well as 0.1% Tween-20 to prevent non-specific binding between bait and prey proteins. For each pull-down combination, a negative control reaction was included, containing only the prey protein, in order to account for the non-specific binding between the prey protein and the beads.

After the final wash, the beads were separated again by a magnet, and the flow-through was discarded. The proteins specifically bound to the magnetic beads were eluted in 100 μl PD Elution Buffer (Table 2.14), following 5 minutes of incubation on a rotating wheel at 4°C. The elution reaction was placed on a magnet, and the flow-through was carefully transferred to a fresh tube. The Input, Depletion and Elution samples were mixed with 2x Laemmli Buffer (Bio-Rad) and boiled for 5 minutes at 100°C. The Input sample was derived from the bait-prey reaction mixture, while the Depleted sample was taken from the flow-through after the incubation of the beads with the bait-prey reaction mixture. Each pull-down reaction was subsequently analysed by Western blotting.

2.7.4 Size Exclusion Chromatography Multi-Angle Light Scattering

Size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) was performed using an AKTA prime system with a S200 10/300 GL column, a 100 μl loop and a Wyatt MALS detector. The column, loop and the system were equilibrated with SEC Buffer (Table 2.14) for 48-72 hours in order to minimise and stabilise the light scattering background noise. Protein samples of 150 μl (45 μM PilA, 50 μM PilB, 50 μM PilC, 40 μM PilABC, 40 μM PilAC) were injected into the loop and loaded onto the column at 0.2 ml/min. The UV, light scattering and refractive index signals were analysed with the ASTRA chromatography software.

2.7.5 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) was performed using a MicroCal ITC200 Malvern machine at 20°C. The sample cell was loaded with 20 μM purified PilC protein. The reference cell contained filtered and degased ITC P/P (protein/protein) buffer (Table 2.14). The syringe was filled with 200 μM purified PilA protein. The experiment consisted of 20 titrations in total – the first injection was 0.4 μl, while all subsequent titrations were 1.9 μl in volume. There was a gap of 200 seconds between injections. The ITC experiments were performed three independent times. Both cells and the syringe were washed three times with H_2O and buffer between repeats. Data analysis was performed on the Origin ITC200 software in order to calculate the affinity of the interaction between the two proteins.

2.8 Analysis of Protein-Sugar Interactions

2.8.1 Isothermal Titration Calorimetry

ITC was also used to study the interactions between minor pilins and specific sugar moieties. Experiments were performed on a MicroCal ITC200 Malvern machine at 20°C in filtered and degased ITC P/S (protein/sugar) buffer (Table 2.14). The sample cell was loaded with 0.1 mM purified protein, the reference cell contained ITC P/S Buffer, and the syringe was filled with 2 mM of the sugar ligand. The experiment consisted of 30 titrations in total with a gap of 150 seconds between each titration. The first titration was 0.4 μl, while all subsequent titrations were 1.2 μl in volume. The ITC experiments were performed three independent times for each mutant-ligand pair. As above, the sample cell, the reference cell and the syringe were washed three times with H_2O and buffer between repeats. The affinity of the interaction between the protein and the sugar ligand was calculated using the Origin ITC200 software.

2.8.2 Cell Adhesion Assays

2.8.2.1 Tissue Culture Maintenance

Chinese Hamster Ovary (CHO) cells and Henrietta Lacks (HeLa) cells were acquired from Public Health England. The cells were cultured in T75 flasks (Corning) in Dulbecco's Modified Eagle Medium (DMEM) (Gibco), containing 1x minimum essential medium non-essential amino acid mix (Merck) and 5% FBS (Gibco), at 37°C and 5% CO_2. When the cells reached ∼70% confluency, they were washed twice with 10 ml PBS (Gibco) to remove excess serum and detached with 2 ml trypsin-EDTA (0.05%) at 37°C for 5 minutes. The detached cells were resuspended in 8 ml DMEM to neutralise the trypsin activity and

passaged via a 1:5 dilution.

2.8.2.2 Adhesion Assays

On day 1, eukaryotic cells were seeded at 100,000 cells/cm^2 in 24-well plates and incubated overnight at 37°C and 5% CO_2 in order to reach 90% confluency. The same day, *S. sanguinis* strains were inoculated in THH media and grown overnight. On day 2, the bacteria were back-diluted 1:10 in 10 ml THH and grown until the OD_{600} reached 0.5. If necessary, the OD_{600} was adjusted, and the bacteria were diluted in serum-free DMEM to obtain a multiplicity of infection (MOI) of 10. The eukaryotic cell monolayers were gently washed three times with 1 ml PBS in order to remove any remaining serum. Then, 1 ml of bacteria at MOI 10 was added to each well, and the plates were centrifuged at 1,000 rpm for 2 minutes. After 1 hour of incubation at 37°C and 5% CO_2, the cell monolayers were gently washed five times with 1 ml PBS before the cells and the adherent bacteria were scraped in 1 ml distilled water. Serial dilutions of the inoculums (10^{-2}, 10^{-3} and 10^{-4}) and the adherent bacteria (10^{-1}, 10^{-2}, 10^{-3}) were prepared and plated on TH agar. On day 3, the bacteria were quantified by colony-forming units (CFU) counts. The adhesion assays were performed in triplicate and repeated twice. Adhesion was calculated as percentage of the bacteria in the inoculum that have adhered to the eukaryotic cell monolayer.

2.9 Structural Studies

Structural analysis of the putative pilin complex was attempted by X-ray crystallography, while the binding interface of interacting pilins was mapped by nuclear magnetic resonance.

2.9.1 X-Ray Crystallography

For broad crystal screening, protein samples were loaded on crystallisation screen trays (Table 2.16) using the Mosquito Robot in 200 nl drops with differing ratios of protein to mother liquor (30:70, 40:60, 50:50, 60:40, 80:20). The trays were sealed, incubated at 20°C or 4°C and checked regularly.

In order to improve the quality of crystals yielded in some conditions, optimisation screens were set up using STP LabTech's Dragonfly and MRC 96-well plates (Molecular Dimensions). Optimisation screens were designed as grids with incremental changes in the concentration of precipitants along the vertical axis and buffer along the horizontal axis. This produced slight variants of the initial crystallisation condition in each well. Optimisation screens were set up as 100:100 nl drops (protein to mother liquor) and incubated at 4°C.

Plate ID	**Screen Name**	**Company**
ICL2	Wizard Screens 1 & 2	Molecular Dimensions
ICL0	Structure Screens 1 & 2	Molecular Dimensions
ICL7	PACT premier	Molecular Dimensions
ICL8	JCSG+	Molecular Dimensions
ICL12	Proplex	Molecular Dimensions
ICL13	Morpheus	Molecular Dimensions
ICL14	PGA Screen	Molecular Dimensions

Table 2.16: **Crystallisation screens used in this study.**

An alternative strategy for improving the quality of crystals was microseeding. Crystals from both commercial and optimisation screens were broken up with a Microseed Bead (Molecular Dimensions), diluted 1:100 in their original mother liquor (according to manufacturer's instruction) and used as seeds to induce the formation of larger crystals. Microseeding was tested in a range of crystallisation screen trays (Table 2.16). Using the Mosquito Robot, 300 nl drops were set up, containing 150 nl protein sample, 50 nl

microseed stock and 100 nl mother liquor. The trays were sealed, incubated at 4°C and checked regularly.

Potential crystals were sent to Diamond Light Source. Diffraction data were gathered and analysed by Dr Sheppard.

2.9.2 Nuclear Magnetic Resonance

His-PilA was isotopically labelled with ^{15}N and/or ^{13}C. ^{15}N-labelling was performed by inducing protein expression in *E. coli* BL21 (DE3) that were grown in minimal medium, supplemented with $^{15}NH_4Cl$. For double labelling (^{15}N, ^{13}C-labelling), the protein was produced by bacteria grown in minimal medium, containing $^{15}NH_4Cl$ and ^{13}C-glucose.

As described in Section 2.6.1, 10 ml LB starter cultures were grown overnight and back-diluted in 1 l kanamycin-supplemented LB. Cultures were grown at 37°C with shaking until the OD_{600} reached 1. At this point, the cells were gently harvested by centrifugation at 6,000 rpm for 10 minutes at 4°C in sterile centrifugation tubes. The pelleted cells were resuspended in 490 ml M9 salts (3.37 mM Na_2HPO_4, 2.2 mM KH_2PO_4, 0.855 mM NaCl, pH 7.2), containing 1 ml 1 M $MgSO_4$ and 250 μl 0.2 M $CaCl_2$. The cell suspension was supplemented by 5 mg vitamin B1, 50 μg/ml kanamycin, 0.5 g $^{15}NH_4Cl$ and 2 g unlabelled D-glucose (^{13}C-labelled for double isotopic labelling), all dissolved in 10 ml H_2O and filter-sterilised. The cells were then incubated for 20 minutes at 16°C with shaking before protein expression was induced with 0.5 mM IPTG. Following overnight growth at 16°C, the cells were harvested, lysed and purified via IMAC and SEC as described in Section 2.6.3. A specific NMR buffer (Table 2.14) with pH 7.0 was used during SEC purification.

The NMR samples were prepared in 200 or 400 μl volumes with 5% D_2O. All data were

collected at 25°C in a Bruker Avance III HD 800 MHz triple resonance spectrometer with a cryoprobe. NMR spectra were recorded and analysed by Dr Sheppard. In brief, the ^{15}N,^{13}C-labelled protein was used for the assignment of the backbone and side-chain ^{1}H, ^{13}C and ^{15}N atoms in TROSY (transverse relaxation optimised spectroscopy) experiments: HNCA/HNCOCA, HNCO/HNCACO and HNCACB/CHCACONH. ^{15}N-TROSY-HSQC (heteronuclear single quantum coherence) experiments were performed on the ^{15}N-labelled protein alone and the ^{15}N-labelled protein as part of the putative pilin complex. Changes in the chemical shifts upon complex formation were used to map the binding interface of the minor pilin complex.

2.9.3 Bioinformatics and Modelling

Gene sequences were analysed with SnapGene Software (Insightful Science). The prediction of protein domains was done by InterProScan, while structural homology was analysed with the DALI server (Holm, 2020). The 3D structures of proteins were prepared in PyMol (Schrödinger, LLC). The modelling of the PilA-PilC pilin complex was performed in AlphaFold (Jumper et al, 2021; Evans et al, 2022). The PISA software was used to analyse the hydrogen bonds and salt bridges maintaining the PilA homodimer and the PilAC heterodimer (Krissinel and Henrick, 2007). Modelling of the PilAC complex in the T4P was done by Dr Sheppard.

Chapter 3

Functional Analysis of Minor Pilin PilA

3.1 Introduction

PilA is one of the three minor pilins in *S. sanguinis*. It is of common size (17 kDa) and is essential for T4P biogenesis (Gurung et al, 2016). Western blot studies using a highly specific anti-PilA antibody have shown that PilA undergoes processing by the prepilin peptidase and is incorporated into the type IV pili (Berry et al, 2019). PilA has no sequence homologues outside of the *S. sanguinis* species, and its α1N helix displays a degenerate class III signal peptide (Berry et al, 2019). Although the α1N helix is dispensable for the correct folding of the pilin, it plays an important structural and functional role in T4P biology. Substitutions of the highly conserved Gly-1 residue have been extensively reported to inhibit prepilin processing (Strom and Lory, 1991), while point mutations of the Glu5 residue are known to preclude T4P assembly (Aas et al, 2007). Similarly, changes in the alanine residue at position 20 of the *M. xanthus* major pilin impact the retraction of pili. Thus, even small variations at key N-terminal residues can have major consequences for pilin and pili function (Giltner et al, 2012).

Figure 3.1: **The class III signal peptide of minor pilin PilA in *S. sanguinis* 2908.** The leader peptide is underscored. The cleavage site by the prepilin peptidase is marked by a pair of scissors and a dotted line. Conserved residues are coloured in blue and divergent residues in green.

The canonical class III signal peptide typically begins with a leader peptide sequence which is composed of 6-26 predominantly hydrophilic and neutral residues. The leader peptide ends in a highly conserved glycine residue and is followed by 21 mainly hydrophobic amino acids, with the exception of a negatively charged glutamate at position 5 (Giltner et al, 2012). The leader peptide of PilA is composed of eight residues – the majority of which are neutral in charge (Figure 3.1) – and ends in a glycine residue. The first amino acid

of the mature PilA pilin is the neutral amino acid serine. This is different from the other four *S. sanguinis* pilins, where the mature protein begins with a hydrophobic residue such as phenylalanine or isoleucine. Furthermore, the stretch of amino acids after the leader peptide is interrupted by a hydrophilic proline residue at position 4, while the highly conserved glutamate at position 5 is absent, with a hydrophobic methionine residue in its place.

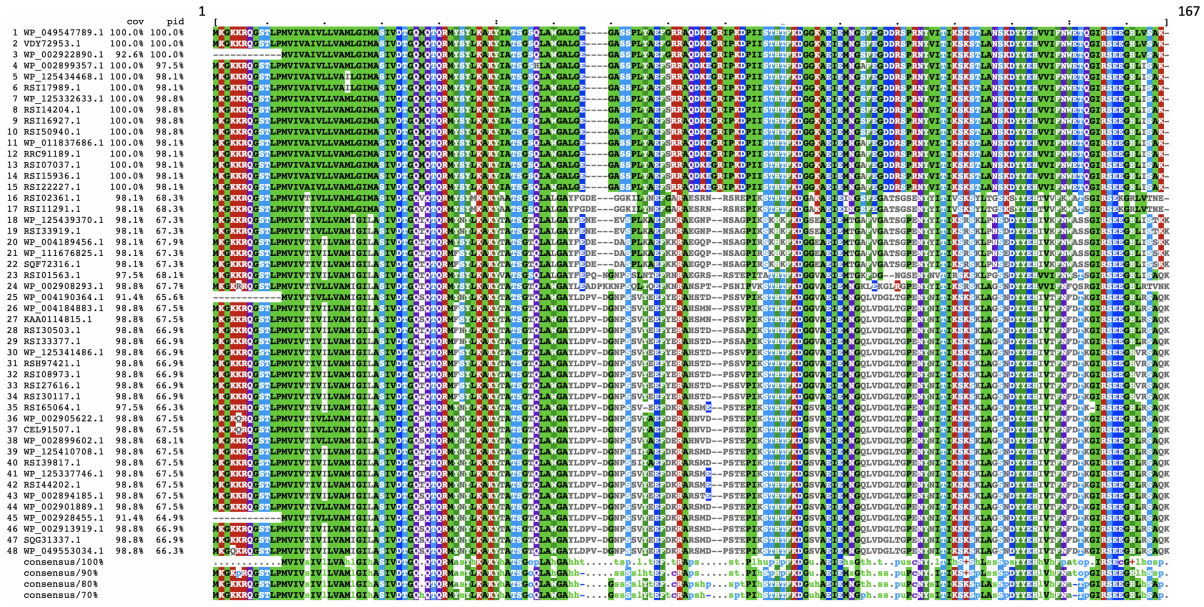


Figure 3.2: **Clustal Omega analysis of minor pilin PilA in *S. sanguinis* strains.** The analysis was performed with the 48 *S. sanguinis* strains sequenced to date. In three of the strains, the leader peptide sequence of PilA is missing. Red = small residues , Blue = acidic residues, Magenta = basic residues and histidine, Green = hydroxyl, sulfhydryl and amine residues, and glycine (Madeira et al, 2022).

Despite these unusual features, the class III signal peptide of PilA is highly conserved (Figure 3.2). Clustal Omega multiple sequence analysis (Madeira et al, 2022) of minor pilin PilA proteins from all 48 *S. sanguinis* strains sequenced to date showed a total coverage of ∼98.6% and total sequence similarity of ∼77%. The highest degree of conservation in PilA was observed at its N-terminus. The leader peptide sequence was identical in 91% of the strains: three strains bore single amino acid substitutions at position -4, and one strain carried an amino acid substitution at position -5. The longest stretch of residues conserved in all sequenced strains extended from position -3 to position 8. It included

the prepilin peptidase cleavage site as well as the N-terminal amino acids which make the class III signal peptide of PilA unusual. Such high degree of conservation suggests that these residues are central to the role of PilA in T4P biogenesis.

Importantly, the divergent class III signal peptide of minor pilin PilA supports its prepilin processing and its incorporation into the *S. sanguinis* T4P (Berry et al, 2019). This raises questions about how the residues in the class III signal peptide of PilA facilitate leader peptide cleavage and filament packing. This chapter attempts to address these questions by generating *S. sanguinis* strains with mutations in the unusual class III signal peptide of *pilA* and analysing their ability to process PilA, to incorporate the minor pilins into their pili and to support T4P functionality in the form of twitching motility.

3.2 Generating *S. sanguinis pilA* Mutants

To investigate the importance of both conserved and divergent residues in the class III signal peptide of PilA, markerless *S. sanguinis pilA* mutants were constructed. These mutants carried substitutions in the glycine residue at position -1 which is important for prepilin processing; the serine which directly follows the leader peptide cleavage site; the threonine and leucine at position 2 and 3 which are also present in minor pilins PilB and PilC; and the unusual hydrophilic proline residue at position 4 (Figure 3.3).

S. sanguinis mutagenesis was performed in two steps: side-directed mutagenesis (SDM) of *pilA*, followed by *S. sanguinis* transformation. Firstly, the *pilA* gene, alongside its upstream (0.85 kb) and downstream (0.2 kb) regions, was amplified from WT *S. sanguinis* genomic DNA (gDNA) and ligated into an empty pCR8/GW/TOPO vector. The *pilA* TOPO plasmid was used as a template for site-directed mutagenesis (SDM) facilitating

Figure 3.3: **The amino acid sequences of WT PilA and PilA mutants constructed during this project.** The leader peptide is annotated in the WT sequence. The cleavage site by the prepilin peptidase is marked by a pair of scissors and a dotted line. Conserved residues are coloured in blue, divergent in green and mutant in red.

the introduction of point mutations into the WT gene. The products from the SDM reaction were treated with the restriction enzyme DpnI, which digested any methylated or hemimethylated DNA, containing the original non-mutated template plasmid. The mutant plasmids were then transformed into *E. coli* DH5α, grown on on streptomycin-supplemented LB agar. Six individual colonies were picked for each mutant and analysed by colony PCR. Colony PCR verified the presence of the *pilA*-TOPO plasmids in the *E. coli* bacteria, while the successful introduction of the desired point mutations was verified by sequencing the purified plasmids.

Next, the mutant *pilA* constructs were amplified from the *pilA*-TOPO plasmid and transformed into the *S. sanguinis pilA:pheS*-aphA3* primary mutant in order to generate markerless *S. sanguinis* mutant strains. The primary mutant contains a double cassette, composed of *pheS** and *aphA3* instead of the WT *pilA* gene. This allows for both negative and positive selection during mutagenesis (Gurung et al, 2017). The *pheS** gene encodes a mutant tRNA synthetase PheS, which renders the bacteria susceptible to the toxic pheny-

lalanine analogue *para*-chlorophenylalanine (PCPA). PCPA causes protein misfolding and ultimately leads to a drastic reduction in bacterial growth. Meanwhile, the *aphA3* gene confers resistance to kanamycin (Gurung et al, 2017). In the successfully transformed bacteria, the *pheS*-aphA3* double cassette was replaced with a mutated *pilA* gene (Figure 3.4), rendering them resistant to PCPA and sensitive to kanamycin. The presence of the desired point mutations was verified by sequencing the *pilA* gene from the gDNA of the generated *S. sanguinis pilA* mutant strains.

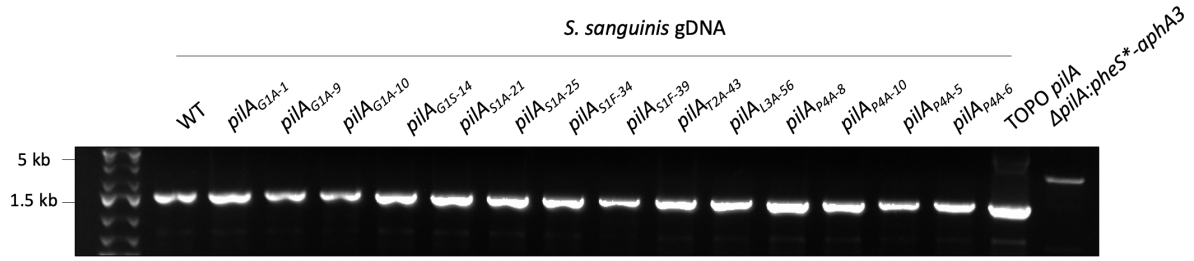


Figure 3.4: **Verification of the non-polar markerless *pilA* mutants.** In a PCR reaction, WT and mutant genomic DNA (gDNA) were amplified using specific *pilA* primers. The primary *pilA:pheS*-aphA3* mutant was included as a control. The PCR products were then analysed by agarose gel electrophoresis. The relevant DNA ladder sizes are marked in kilobases (kb). The small numbers next to the mutation refer to clone numbers.

The verified *S. sanguinis* mutants expressed a mutated PilA protein carrying a single amino acid substitution in its class III signal peptide. The mutant strains were subjected to several types of phenotypical analysis in order to determine 1) whether the mutant PilA protein underwent prepilin processing, 2) whether it supported *S. sanguinis* piliation, 3) whether it was incorporated into the resulting filaments, and 4) whether those filaments were functional.

3.3 Prepilin Processing of Mutant PilA Pilins

Firstly, we tested the ability of the mutant minor pilin PilA to undergo prepilin processing *in vivo*. Small overnight cultures of *S. sanguinis* were grown statically overnight at 37°C

before being back-diluted 1:10 into THT media (TH supplemented with 0.05% Tween-80). Once the OD_{600} reached 1, the bacteria were pelleted, resuspended in a smaller volume and homogenised to obtain whole-cell protein extracts (WCE). The whole-cell extracts were analysed for prepilin processing by Western blotting using a specific anti-PilA antibody. Wild-type *S. sanguinis* and the Δ*pilD* mutant were included as a positive and negative control, respectively. The WT whole-cell extract contained the processed PilA pilin, which is missing the class III leader peptide and is therefore smaller in size and travels further during SDS-PAGE. The Δ*pilD* mutant was deficient for prepilin processing, so PilA remained in its prepilin form, which is slightly bigger in size than WT and travels a shorter distance during SDS-PAGE.

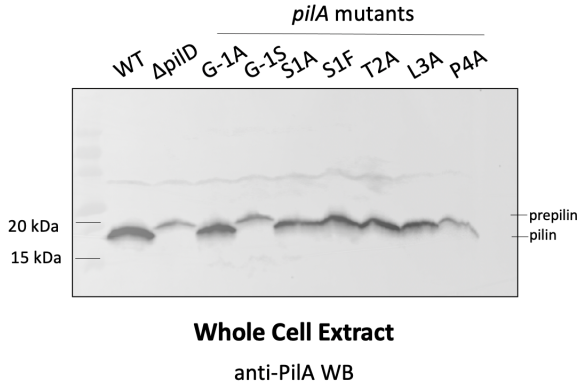


Figure 3.5: **Processing of minor pilin PilA mutants by the prepilin peptidase PilD *in vivo*.** Whole-cell extracts were separated by SDS-PAGE and probed using an anti-PilA antibody. The Western blot shows the processing of PilA mutants $PilA_{G\text{-}1A}$, $PilA_{G\text{-}1S}$, $PilA_{S1A}$,$PilA_{S1F}$, $PilA_{T2A}$, $PilA_{L3A}$, and $PilA_{P4A}$. The WT strain and the Δ*pilD* mutant were loaded as controls. Molecular weights are shown in kilo-Daltons (kDa) and the bands of prepilin (non-processed) and pilin (processed) are indicated on the right side. The same sample volume was loaded in each lane. Representative of three repeats.

PilA was successfully processed by PilD in $pilA_{S1A}$, $pilA_{S1F}$, $pilA_{T2A}$, $pilA_{L3A}$ and $pilA_{P4A}$ mutants (Figure 3.5). This was not surprising as only the glycine at position -1 is usually required for leader peptide cleavage. In keeping with the literature, the $PilA_{G\text{-}1A}$ was also processed, while the G-1S mutation inhibited prepilin processing. In the whole-cell extract of the $pilA_{G\text{-}1S}$ mutant, only the prepilin form of PilA was detected by Western blotting.

This was consistent with previous studies which have shown that substituting the glycine with a serine residue disrupts leader peptide cleavage, while the smaller aliphatic alanine residue has no impact (Giltner et al, 2012). Indeed, this principle also applies to the major pilin PilE1 of *S. sanguinis*: $PilE1_{G\text{-}1S}$ does not undergo prepilin processing and has reduced piliation (Berry et al, 2019).

3.4 Filament Assembly in *S. sanguinis pilA* Mutants

PilA is essential for T4P biogenesis – if *pilA* is deleted, *S. sanguinis* lacks piliation. Mutagenesis of the *pilA* class III peptide could disturb its ability to pack into the T4P filaments and thereby affect filament assembly. Therefore, the generated *pilA* mutants were evaluated for their ability to support *S. sanguinis* piliation by pili purification. Overnight *S. sanguinis* cultures were back-diluted in 100 ml THT and grown until they reached the desired OD_{600} of 1.5 units. The cultures were then pelleted and resuspended in a smaller volume. Next, the pili were mechanically sheared by vortexing the bacteria for 2 minutes at maximum speed. The bacteria were pelleted again, and the supernatant containing the sheared pili was filtered and purified via ultracentrifugation for one hour. Piliation was then assessed by SDS-PAGE and Coomassie staining, while incorporation of the minor pilins into the pili was verified by Western blotting. The WT strain was included as a positive control for piliation. Following Coomassie staining, the WT sample displays two bands ∼15 kDa – these two bands represent the two major pilins PilE1 and PilE2 that make up the heteropolymeric T4P of *S. sanguinis*. If these two bands are absent in the sample of the purified pili, the corresponding *S. sanguinis* strain is not piliated.

The six mutants which underwent processing – $pilA_{G\text{-}1A}$, $pilA_{S1A}$, $pilA_{S1F}$, $pilA_{T2A}$, $pilA_{L3A}$ and $pilA_{P4A}$ – were piliated (Figure 3.6). In these strains, all three minor pilins PilA, PilB

and PilC were detected in the purified filaments at levels comparable to WT (Figure 3.5). This suggests that the conserved threonine and leucine, and the divergent serine and proline residues were all dispensable for assembly of PilA into the T4P.

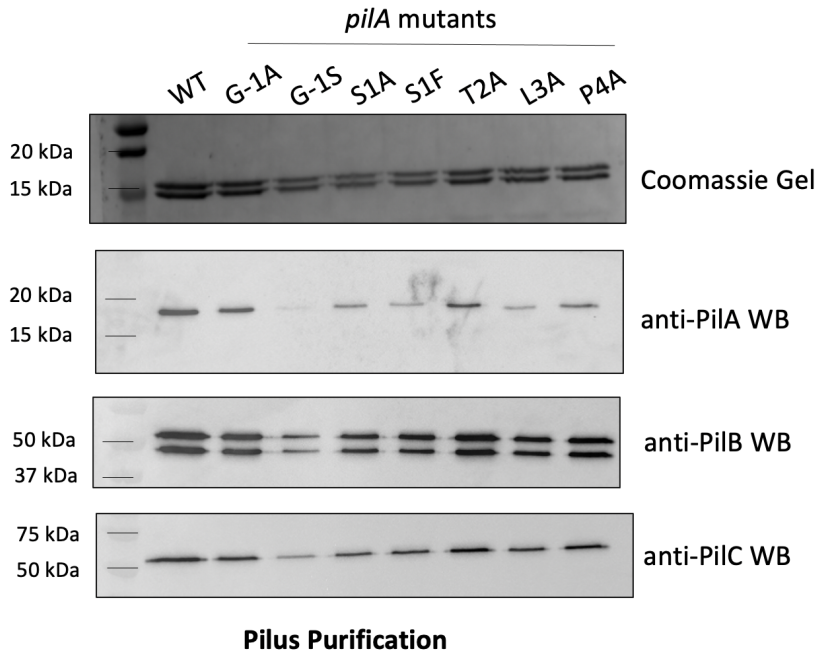


Figure 3.6: **Effect of *pilA* mutations on *S. sanguinis* piliation and minor pilin assembly into filaments.** Pili purifications were prepared for each *pilA* mutant, separated by SDS-PAGE and stained with Coomassie blue. WT *S. sanguinis* was included as a control. The purified pili were further probed by Western blotting (WB) using anti-PilA, anti-PilB and anti-PilC antibodies. The anti-PilB antibody appears to detect both PilC (higher band) and PilB (the lower band). Molecular weights are shown in kDa. The same sample volume was loaded in each lane. Representative of three repeats.

Unexpectedly, the $pilA_{G\text{-}1S}$ mutant strain was also piliated. SDS-PAGE analysis revealed that the non-processed $\mathrm{PilA_{G\text{-}1S}}$ mutant protein was able to support the assembly of T4P to WT levels (Figure 3.6). Western blot analysis of the purified pili showed a faint band of PilA assembled into the T4P with corresponding reductions in the concentrations of polymerised PilB and PilC. This was surprising because prepilin processing is essential for pilin polymerisation into filaments and indeed, the PilA protein detected in the purified pili appeared to be the same size as processed PilA (Figure 3.7). There were two possible explanations for this unusual phenotype: 1) the $pilA_{G\text{-}1S}$ mutant was contaminated with

WT *S. sanguinis*, or 2) the G-1S mutation did not abolish prepilin processing, and the residual amount of PilA that underwent leader peptide cleavage was able to support T4P assembly.

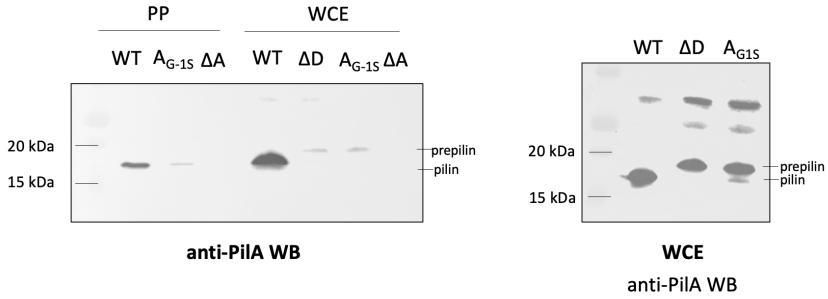


Figure 3.7: **Investigating the small fraction of PilA$_{\mathbf{G\text{-}1S}}$ which polymerises into T4P.** (Left) anti-PilA Western blot comparing the size of PilA$_{\text{G-1S}}$ in purified pili (PP) to PilA$_{\text{G-1S}}$ in whole-cell extracts (WCE). WT, Δ*pilD* and Δ*pilA* were included as controls. The same sample volume was added to each lane. (Right) anti-PilA Western blot of whole-cell extracts exposing the residual processing of the PilA$_{\text{G-1S}}$ mutant. A 4-fold dilution of the WT sample was loaded alongside a 2-fold dilution of the Δ*pilD* and PilA$_{\text{G-1S}}$. Molecular weights are shown in kDa, and the bands of prepilin (non-processed) and pilin (processed) are indicated on the right side. Representative of two repeats.

To exclude the possibility of glycerol stock contamination, we first generated a new *pilA$_{G\text{-}1S}$* mutant and repeated the same set of assays. The new *pilA$_{G\text{-}1S}$* strain yielded the same results as the original strain (data not shown), indicating that the observed phenotype was not due to the presence of WT *S. sanguinis* in the mutant cultures. Next, we modified the Western blotting step of the prepilin processing assay by loading smaller volumes of whole-cell extracts and exposing the blotted membrane for longer. This allowed us to see fainter bands that would have been obscured under normal settings. Indeed, the optimised Western blots revealed two bands in the *pilA$_{G\text{-}1S}$* sample – a strong prepilin PilA band and a much weaker band, corresponding to processed PilA (Figure 3.7). This confirmed the residual processing of the PilA$_{\text{G-1S}}$ mutant by the prepilin peptidase PilD.

3.5 Assessing the G-1S Mutation in PilB and PilC

There were two possible explanations for the residual processing of the $PilA_{G\text{-}1S}$ mutant: 1) this was a common property of *S. sanguinis* minor pilins, or 2) the unusual class III signal peptide of PilA facilitated leader peptide cleavage by PilD even when the highly conserved glycine residue was mutated. To determine which explanation was correct, the same G-1S mutation was engineered in the minor pilins PilB and PilC which have canonical class III signal peptides. The mutagenesis was performed as described above – the point mutation was introduced into TOPO plasmids carrying the WT *pilB* or *pilC* gene alongside their upstream and downstream regions. The mutated genes were then amplified by PCR and transformed into the *S. sanguinis* primary mutants *pilB: pheS*-aphA3* or *pilC: pheS*-aphA3*. Transformed bacteria were selected based on their resistance to PCPA and their sensitivity to kanamycin, and the successful introduction of the point mutation was verified by sequencing. The *S. sanguinis* $pilB_{G\text{-}1S}$ and $pilC_{G\text{-}1S}$ mutants were then tested for prepilin processing of the targeted minor pilin, piliation and incorporation of the minor pilins into the T4P filaments.

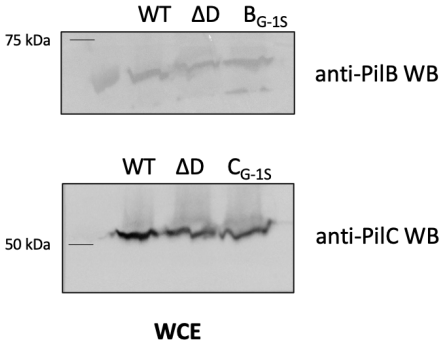


Figure 3.8: **Processing of minor pilin mutants $PilB_{G\text{-}1S}$ and $PilC_{G\text{-}1S}$ by the prepilin peptidase PilD *in vivo*.** Whole-cell extracts of the two mutants were separated by SDS-PAGE and probed using either an anti-PilB or anti-PilC antibody. WT and Δ*pilD* were loaded as controls. Molecular weights are shown in kDa. The same sample volume was loaded in each lane.

Prepilin processing of PilB and PilC was assessed by obtaining the whole-cell protein

extracts of the two mutant strains and probing them by Western blotting with anti-PilB or anti-PilC antibodies (Figure 3.8). However, due to the large size of PilB and PilC (exceeding 50 kDa in their mature form) and the short length of their leader peptides (8-9 amino acids), it was difficult to distinguish between the prepilins and the processed pilins in the Western blots. The issue could not be resolved by using a larger percentage resolving gel or running the SDS-PAGE at a lower voltage for a longer time, making it challenging to conclude whether the $PilB_{G\text{-}1S}$ and $PilC_{G\text{-}1S}$ mutations impeded leader peptide cleavage, or whether they were also susceptible to residual processing.

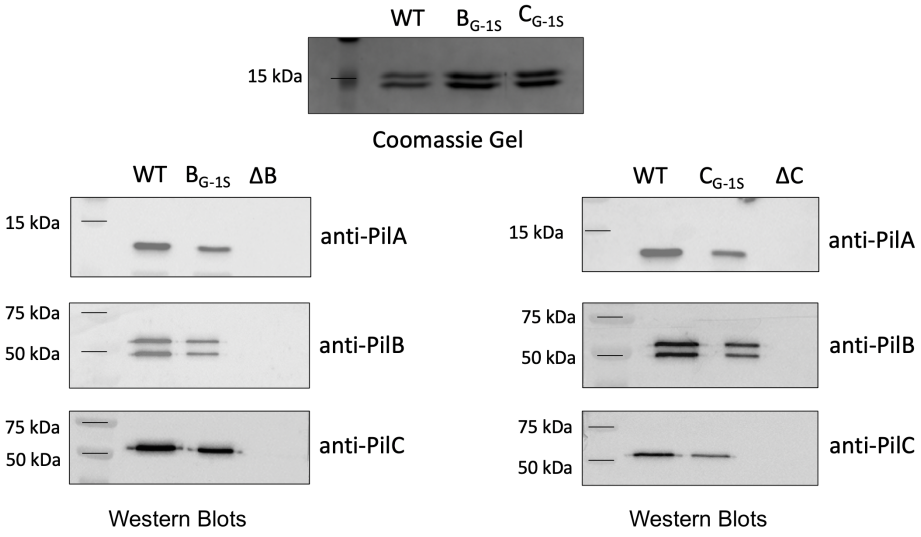


Figure 3.9: **Effect of PilB and PilC G-1S mutations on *S. sanguinis* piliation and minor pilin assembly into filaments.** Pili purifications were prepared for the $pilB_{G\text{-}1S}$ and $pilC_{G\text{-}1S}$ mutants, separated by SDS-PAGE and stained with Coomassie blue. WT was included as a control. The purified pili were further probed by Western blotting using anti-PilA, anti-PilB and anti-PilC antibodies. Deletion mutants were included as negative controls. Molecular weights are shown in kDa. Identical sample volumes were loaded in each lane. Representative of three repeats.

Nonetheless, the two mutants were analysed for piliation. If the G-1S mutation completely abolished prepilin processing, then PilB and PilC could not be incorporated into the filaments. Since both proteins are essential for T4P biogenesis, this would lead to lack of T4P in *S. sanguinis*. Piliation of $pilB_{G\text{-}1S}$ and $pilC_{G\text{-}1S}$ was tested by pili purification. Both mutants appeared to be piliated, so the purified pili were tested for minor pilin

incorporation (Figure 3.9). Similar to the phenotype observed for $pilA_{G\text{-}1S}$, the G-1S mutation led to a reduction in the amount of PilB assembled in the pili of the $pilB_{G\text{-}1S}$ strain, and PilC in the pili of the $pilC_{G\text{-}1S}$ strain, compared to the WT. Furthermore, the decreased incorporation of the mutated minor pilin correlated with a decrease in the incorporation of the other two minor pilins without affecting the overall levels of *S. sanguinis* piliation (Figure 3.9). In fact, from the Coomassie-stained gel, it seemed that the two mutants were more piliated than the WT.

It is important to note two things here. Firstly, pili purification is a qualitative method – the amount of pili detected during SDS-PAGE analysis could be artificially affected by small variations in the way the samples were treated during the performance of the experiment. Hence, pili purification only informs us as to whether the bacteria are piliated or not, but does not confer reliable information on subtle differences in quantity. Secondly, while only a trace amount of PilA was detected in the $pilA_{G\text{-}1S}$ strain, the amount of $PilB_{G\text{-}1S}$ and $PilC_{G\text{-}1S}$ in their respective mutant strains appeared to be significantly greater. This could be attributed to differences in the sensitivities of the anti-PilA, anti-PilB and anti-PilC antibodies. These differences are very pronounced in experiments performed with purified proteins which will be discussed in a different chapter.

Since the $PilB_{G\text{-}1S}$ and $PilC_{G\text{-}1S}$ mutant proteins supported piliation and were incorporated into the pili, we can conclude that the two mutant pilins underwent some degree of leader peptide cleavage. This suggests that the G-1S mutant phenotype was not specific to PilA and cannot be ascribed to its divergent class III signal peptide.

3.6 Abolishing PilA Prepilin Processing

In an attempt to completely abolish PilA prepilin processing, two further mutants were generated – $pilA_{G\text{-}1R}$ and $pilA_{G\text{-}1W}$. In both mutants, residues larger than serine were used as substitutions for the conserved Gly-1: the bulky chains of arginine and tryptophan were thought to be more likely to impact leader peptide cleavage. Mutagenesis of *S. sanguinis* was performed as described above, and the same phenotype analysis was employed, testing PilD-mediated processing and bacterial piliation.

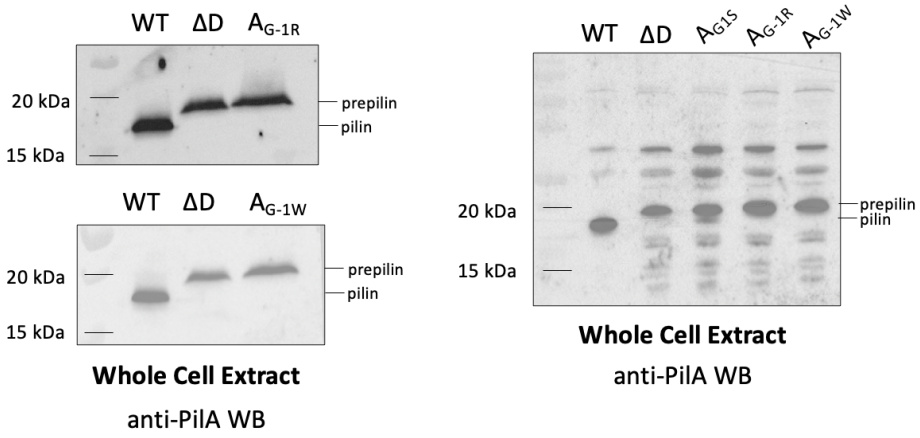


Figure 3.10: **Processing of minor pilin $PilA_{G\text{-}1R}$ and $PilA_{G\text{-}1W}$ mutants by the prepilin peptidase PilD *in vivo*.** (Left) Whole-cell extracts were separated by SDS-PAGE and probed using an anti-PilA antibody. WT strain and the Δ*pilD* mutant were loaded as controls. The same sample volume was loaded in each lane. (Right) A 4-fold dilution of the WT sample was loaded alongside 2-fold dilutions of the Δ*pilD*, $pilA_{G\text{-}1S}$, $pilA_{G\text{-}1R}$ and $pilA_{G\text{-}1W}$ mutants to visualise any residual prepilin processing. Molecular weights are shown in kDa, and the bands of prepilin (non-processed) and pilin (processed) are indicated on the right side. Representative of three repeats.

Western blot analysis of whole-cell extracts from $pilA_{G\text{-}1R}$ and $pilA_{G\text{-}1W}$ showed that the two mutations appeared to abolish PilA prepilin processing. Even the modified protocol for Western blots, which revealed the trace amount of processed pilin in the $pilA_{G\text{-}1S}$ strain, confirmed that the G-1R and G-1W mutations reduced leader peptide cleavage to the point of no detection – only the slightly longer prepilin PilA could be seen in the blots (Figure 3.10). This excluded the possibility that the mature PilA protein observed in the whole-cell extracts of the $pilA_{G\text{-}1S}$ strain was the result of translational misincorporation

of glycine at position -1 of the mutant protein. If that were the case, the $pilA_{G\text{-}1R}$ and $pilA_{G\text{-}1W}$ strains would be subject to the same frequency of misincorporation and would also demonstrate a faint band corresponding to the mature PilA in Western blots of their whole-cell extracts. We can therefore conclude that the residual processing observed in $pilA_{G\text{-}1S}$ was the result of PilD-mediated processing of the mutant leader peptide.

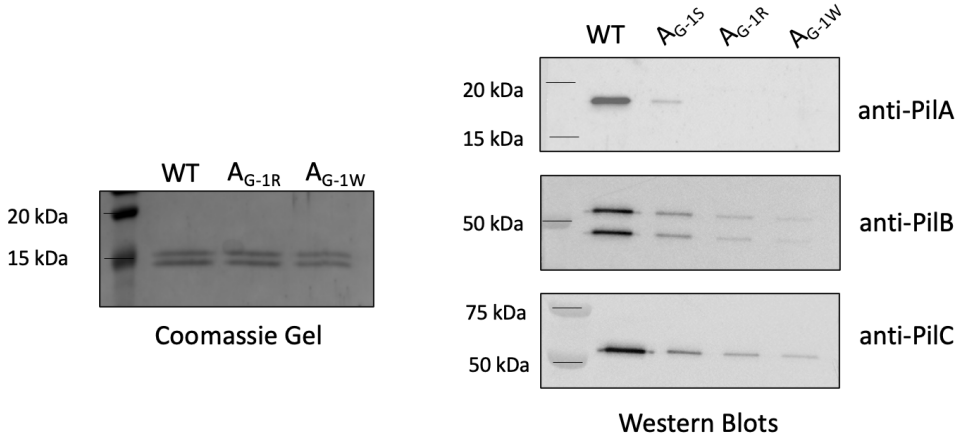


Figure 3.11: **Effect of $\mathbf{PilA_{G\text{-}1R}}$ and $\mathbf{PilA_{G\text{-}1W}}$ mutations on *S. sanguinis* piliation and minor pilin assembly into filaments.** Pili purifications were separated by SDS-PAGE and stained with Coomassie blue. WT was included as a control. The purified pili were further probed by Western blotting using anti-PilA, anti-PilB and anti-PilC antibodies. Molecular weights are shown in kDa. Identical volume of sample was loaded in each lane. Representative of three repeats.

Next, we tested the piliation phenotype of $pilA_{G\text{-}1R}$ and $pilA_{G\text{-}1W}$. Both strains were piliated at levels comparable to WT, but the minor pilin PilA was entirely absent from the purified pili, and only trace amounts of PilB and PilC could be detected (Figure 3.11). The lack of PilA in the pili was confirmed in further Western blotting experiments where the volume of the sample loaded and the exposure were varied in an attempt to expose any faint bands (data not shown). Considering that PilA is essential for *S. sanguinis* piliation (Gurung et al, 2016), it is highly unlikely that T4P assembly would occur in the absence of processed PilA. We therefore posit that the mutant PilA proteins in the $pilA_{G\text{-}1R}$ and $pilA_{G\text{-}1W}$ strains also underwent a degree of residual processing that allowed a small amount of mature PilA to be incorporated into the filaments, but the concentration

of the processed PilA was probably too low to be detected by our antibodies in Western blot analysis of the whole-cell extracts or the pili purifications.

To test this hypothesis, we re-created the G-1W mutation in the minor pilin PilB (Figure 3.12) whose antibody exhibits greater sensitivity. The $pilB_{G\text{-}1W}$ strain was piliated, and we were able to detect PilB in the purified pili. It is unlikely that non-processed PilB would be incorporated in the T4P, confirming that the G-1W mutation failed to abolish leader peptide cleavage of PilB. Furthermore, no PilA could be detected in the pili purifications of the $pilB_{G\text{-}1W}$ strain even though faint bands for both PilB and and PilC were observed. Taken together, these data indicate that the G-1R/G-1W mutations did not abolish prepilin processing but caused a stronger decrease in PilD catalytic activity than the G-1S mutation. In the $pilA_{G\text{-}1R}$ and $pilA_{G\text{-}1W}$ strains specifically, processing of the minor pilin PilA was reduced to the point of no detection due to the lower sensitivity of the anti-PilA antibody.

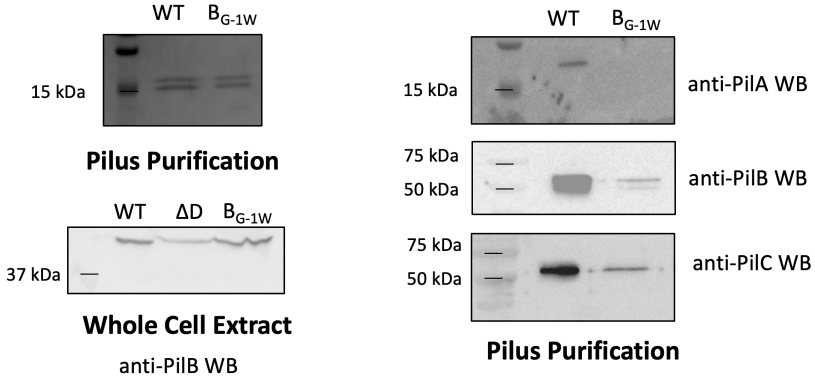


Figure 3.12: **Effect of the PilB$_{\mathbf{G\text{-}1W}}$ mutation on *S. sanguinis* piliation, minor pilin processing and assembly into filaments.** Pili purifications were prepared for the $pilB_{G\text{-}1W}$ mutant, separated by SDS-PAGE and stained with Coomassie blue. WT *S. sanguinis* was included as a control. Whole-cell extracts of WT, the $\Delta pilD$ and $pilB_{G\text{-}1W}$ mutants were probed with anti-PilB antibodies by Western blotting in order to establish whether prepilin processing occurred. The purified pili were further probed by Western blotting (WB) using anti-PilA, anti-PilB and anti-PilC antibodies. Molecular weights are shown in kDa. The same sample volume was loaded in each lane. Representative of three repeats.

3.7 T4P Functionality of *S. sanguinis* Class III Signal Peptide Mutants

The functionality of T4P in *S. sanguinis pilA*, *pilB* and *pilC* mutants was assessed by twitching motility. Twitching bacteria exhibit train-like directional motion parallel to the long axis of chains of cells, resulting in spreading zones around bacteria grown on agar plates that are visible to the naked eye. *S. sanguinis* was reported to demonstrate twitching motility long before its pili were described in molecular detail (Henriksen and Henrichsen, 1975; Gurung et al, 2016).

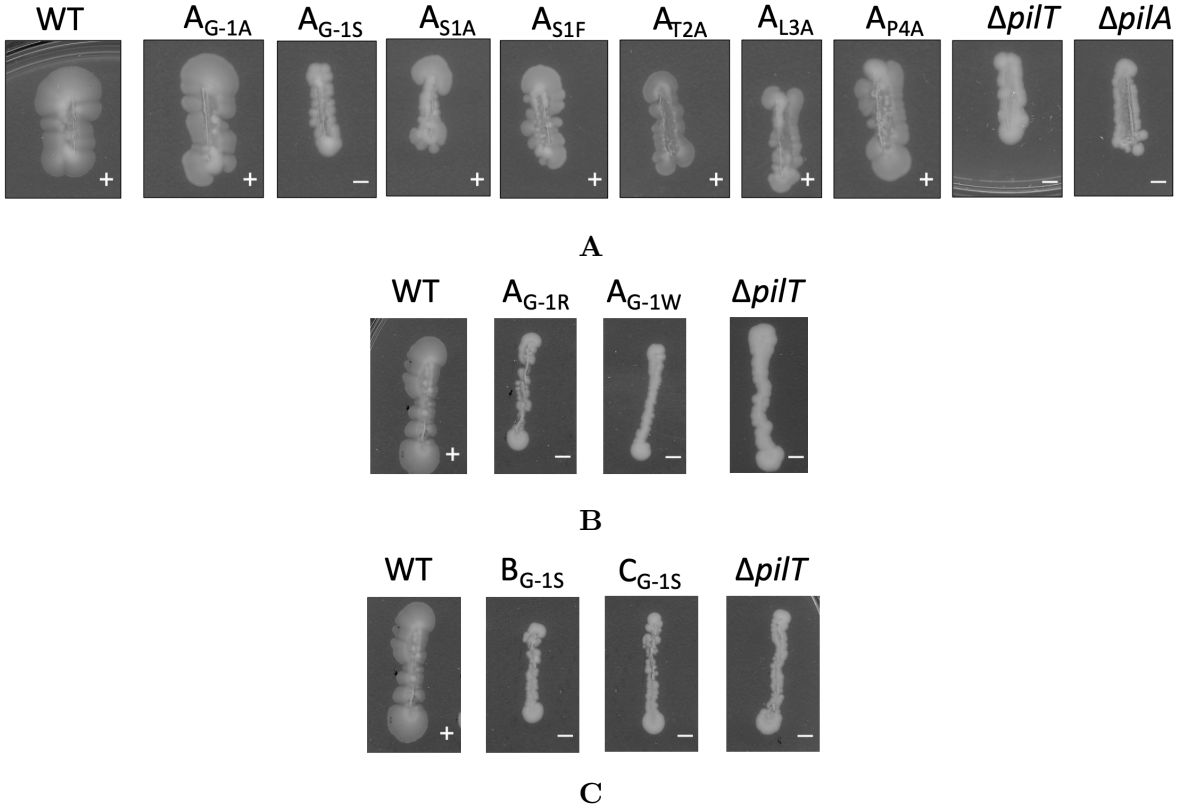


Figure 3.13: **Twitching motility assays.** All mutants were tested for their ability to produce spreading zones on agar plates by twitching motility. Each re-streak is 2 cm long. The WT strain, Δ*pilA* and Δ*pilT* mutants were included as controls. (A) The original set of *pilA* mutants. (B) $pilA_{G\text{-}1R}$ and $pilA_{G\text{-}1W}$ mutants. (C) $pilB_{G\text{-}1S}$ and $pilC_{\text{G-1S}}$ mutants. The (+) sign indicates twitching motility, while the (–) sign indicates lack of twitching motility. Representative of three repeats.

To perform the twitching motility assays, *S. sanguinis* bacteria were grown overnight on TH agar plates, supplemented with the appropriate antibiotics. The next day, single

similar-sized colonies were re-streaked on fresh Eiken agar TH plates in 2 cm-long straight lines. The plates were incubated at 37°C anaerobically with 20 ml distilled H_2O at the bottom of the jar for 72 hours. Eiken agar is thought to promote spreading by its distinct composition which confers greater 'wettability' to the agar (Clemmer et al, 2011). If the strain exhibits twitching motility, a large halo forms around the re-streaking line. If the strain is non-motile, bacteria grow densely in a straight line. The WT strain was included as a positive control, while the non-piliated Δ*pilA* and the retraction-deficient Δ*pilT* mutants were used as negative controls.

The *pilA* mutants which exhibited a WT phenotype for prepilin processing and piliation – $pilA_{G\text{-}1A}$, $pilA_{S1A}$, $pilA_{S1F}$ $pilA_{T2A}$, $pilA_{L3A}$ and $pilA_{P4A}$ – all demonstrated twitching motility (Figure 3.13). In contrast, the *pilA* mutants with disrupted prepilin processing showed no spreading zones, indicating that they are incapable of twitching motility. In $pilA_{G\text{-}1S}$, the decreased incorporation of minor pilin PilA in the T4P disrupted twitching motility, and the apparent absence of PilA polymerisation into the filaments in $pilA_{G\text{-}1W}$ and $pilA_{G\text{-}1R}$ had the same effect. Similarly, the PilB and PilC prepilin processing mutants – $pilB_{G\text{-}1S}$, $pilB_{G\text{-}1W}$ and $pilC_{G\text{-}1S}$ – did not show any twitching motility. Thus, WT levels of processed minor pilins PilA, PilB and PilC were required for T4P functionality.

3.8 Summary

The *S. sanguinis* minor pilin PilA has an unusual class III signal peptide which raises questions about how it is processed and incorporated into T4P. In this chapter, key residues within the class III signal peptide of PilA were subjected to mutagenesis in order to assess their importance for the role of PilA in T4P biogenesis.

Most mutations appeared to have no impact on PilA prepilin processing, *S. sanguinis* piliation and T4P functionality. In particular, mutating the first four residues of the mature PilA protein did not produce any distinct phenotypes. Clustal Omega multiple sequence alignment (Madeira et al, 2022) revealed that the first eight N-terminal residues of the processed PilA pilin are conserved in all 48 *S. sanguinis* species used in the alignment (Figure 3.2). Despite the high degree of conservation, these residues appeared to bear no clear significance for T4P assembly or T4P functionality. One possible explanation is that the α1N helix of PilA is highly accommodating of residue substitutions and can support PilA function regardless. Alternatively, the generated mutations might impact T4P in a way that was not detected by our limited range of experiments.

One interesting observation is that the G-1S mutation was not sufficient to abolish the prepilin processing of PilA. In the $pilA_{G\text{-}1S}$ strain, PilA underwent residual prepilin processing and was incorporated in the purified sheared pili. In contrast to the $pilA_{G\text{-}1A}$ strain, the low levels of mature PilA in the *S. sanguinis* $pilA_{G\text{-}1S}$ mutant did not support twitching motility. This unusual flexibility in the highly conserved residue preceding the leader cleavage site appears to be a common property of all three *S. sanguinis* minor pilins. The findings from this chapter indicate that a small amount of processed pilin is still available in G-1S mutants. This amount appears to be sufficient to support piliation when essential minor pilins are targeted.

Prepilin processing of the *S. sanguinis* PilA minor pilin appeared to be abolished only when the glycine at position -1 was substituted with a bulkier residue, such as tryptophan or arginine. The $pilA_{G\text{-}1R}$ and $pilA_{G\text{-}1W}$ strains were piliated to WT levels, but their pili were defective for twitching motility. Since *pilA* deletion impedes piliation, and PilA is incorporated into the filaments, it is surprising that the presence of non-processed PilA

in the membrane was sufficient to support T4P biogenesis. The most likely explanation is that the G-1W and G-1R mutations did not completely abolish prepilin processing but merely reduced it to the point of no detection by Western blotting. Even though no prepilin processing could be observed, it is feasible that sufficient levels of mature PilA were present in the cytoplasmic membrane to support T4P assembly. This is affirmed by the detection of processed $\text{PilB}_{\text{G-1W}}$ in purified pili and by the inability to detect WT PilA in the sheared pili fraction of the *pilB*$_{G\text{-}1W}$ strain. Taking together the phenotypes of the Gly-1 mutants, it appears that the prepilin peptidase active site of PilD has higher substrate flexibility than anticipated, accommodating serine, arginine and tryptophan at position -1.

The Gly-1 mutants have further demonstrated that WT levels of processed minor pilins are dispensable for the assembly of the *S. sanguinis* T4P, but may be essential for their functionality. The concentration of PilA, PilB and PilC was significantly reduced in the pili of *S. sanguinis* strains *pilA*$_{G\text{-}1S}$, *pilA*$_{G\text{-}1R}$ *pilA*$_{G\text{-}1W}$, *pilB*$_{G\text{-}1S}$ and *pilC*$_{G\text{-}1S}$. Yet, the amount of major pilins PilE1 and PilE2 detected in the pili purifications was comparable to WT. This raises the question of how WT levels of piliation can be maintained with diminished polymerisation of minor pilins, especially if the minor pilins are capping the filaments (see Chapter 5). Since the SDS-PAGE analysis of pili purifications is only indicative of the total quantity of the major pilins incorporated into the sheared pili, it is possible that the Gly-1 mutants incorporated WT levels of PilE1 and PilE2 into fewer but longer surface-expressed pili, thereby decreasing the concentration of minor pilins required to cap the filaments. This would allow the bacteria to maintain piliation while adapting to the reduced availability of processed minor pilins. The altered morphology of the prolonged pili could also explain the defects in twitching motility observed in these

mutants.

Lastly, it is worth highlighting the observed interconnection between the amount of minor pilins that are incorporated in the T4P of the $pilA_{G\text{-}1S}$, $pilA_{G\text{-}1R}$, $pilA_{G\text{-}1W}$, $pilB_{G\text{-}1S}$ and $pilC_{G\text{-}1S}$ mutants. In these mutants, the decreased polymerisation of one minor pilin led to a reduction in the polymerisation of the other two minor pilins, without affecting the overall levels of *S. sanguinis* piliation. This suggests that the three minor pilins depend on each other for stability and perhaps interact together. The next two chapters will address the correlation in minor pilin polymerisation and propose an interesting hypothesis for the role of PilA in T4P assembly.

Chapter 4

Identifying Interactions between the *S. sanguinis* Type IV Pilins

4.1 Introduction

Chapter 3 highlighted a correlation between the polymerisation of minor pilins PilA, PilB and PilC into the *S. sanguinis* type IV pili. In the $pilA_{\text{G-1S}}$, $pilA_{\text{G-1W}}$ and $pilA_{\text{G-1R}}$ strains, PilA processing was partly inhibited, leading to a significant decrease in the total amount of PilA incorporated into the type IV pili. These strains had WT levels of piliation but demonstrated a corresponding reduction in the amounts of PilB and PilC proteins incorporated into the pili. Similarly, the $pilB_{G\text{-}1S}$ and $pilC_{G\text{-}1S}$ mutants exhibited reduced levels of all three minor pilins. This interdependence between PilA, PilB and PilC suggests that an interaction might take place between them.

Recently, there has been a significant advancement in our understanding of the interactions between minor pilins in the T4P and in the evolutionary-related T2SS. In the T4P systems, interactions have been inferred from deletion studies and molecular biology assays. In *P. aeruginosa*, type IV core minor pilins PilV, PilW and PilX are proposed to form a subcomplex with adhesin PilY1. This subcomplex binds to minor pilins PilE and FimU which recruit the major pilin PilA, thus priming filament assembly (Nguyen et al, 2015b). Similarly, the *M. xanthus* adhesin PilY1 has been shown to interact with the four core minor pilins and the major pilin PilA, forming a priming complex that is essential for T4P extension (Treuner-Lange et al, 2020). In the T2SS, the interactions between the minor pseudopilins have been characterised by structural studies as well. The *E. coli* minor pseudopilin GspI interacts with minor pseudopilins GspJ and GspK in a heterotrimeric complex whose structure has been solved by X-ray crystallography (Korotkov and Hol, 2008), and similar interactions have been reported in other T2SS.

Interestingly, DALI structural comparison (Holm, 2020) revealed that GspI (13.9 kDa) is

the closest structural homologue of the *S. sanguinis* minor pilin PilA (Figure 4.1B) even though the two proteins share no detectable sequence homology. The crystal structure of the truncated (soluble) PilA protein demonstrates the typical α1C helix stick followed by a C-terminal globular domain (Figure 4.1A). The α1C helix is connected to the globular domain via an additional short C-terminal α-helix, called α2C. The α2C helix runs on top of the globular C-terminal domain and creates the 'hammer-like' shape of PilA. The globular domain contains five anti-parallel β-strands which are connected to the α2C helix by a long loop. Using AlphaFold (Jumper et al, 2021), we were able to model the structure of the full-length PilA pilin, including the hydrophobic α1N helix. The structural prediction perfectly aligns with the crystal structure of the soluble protein and reveals the characteristic 'lollipop' shape of type IV pilins (Figure 5.20).

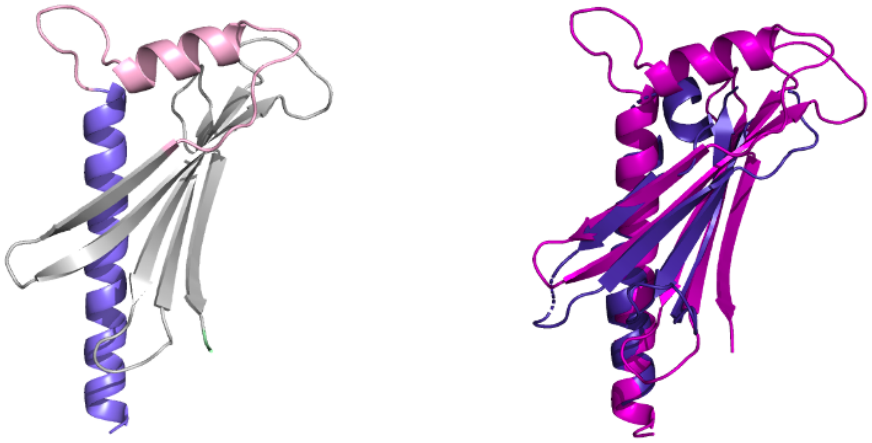


Figure 4.1: **Structural analysis of minor pilin PilA.** (A) Crystal structure of the soluble domain of PilA. The structure was solved by Dr Berry right before the start of this project. The α1-helix is shown in blue, the α2-helix and the $\alpha\beta$ loop are shown in pink, the β-sheet is shown in grey. (B) Structural superposition of *S. sanguinis* minor pilin PilA and *E. coli* T2SS minor pseudopilin GspI. The crystal structure of the truncated PilA and GspI (PDB: 3CI0) were aligned in Pymol. PilA is shown in purple, while GspI is shown in blue. The root mean square deviation (RMSD) is 1.4 Å

Superposition of PilA and GspI highlights the striking structural similarity between the two proteins. PilA and GspI (PDB: 3CI0) demonstrate near-perfect alignment between

the α1C helices and the first four β-sheets in the globular domain with a RMSD of 1.4 Å (Figure 4.1B). GspI also has a short α2C helix which connects the globular domain to the α1C helix. Although the α2C helices and connecting loops of the two minor pilins do not align, GspI has the same unusual hammer-like shape as PilA. It is worth noting that the *S. sanguinis* major pilin PilE1 also shows closest structural homology to the evolutionary-related T2SS major pilins, specifically PulG in *K. oxytoca* (Berry et al, 2019).

In the T2SS of *E. coli*, GspI interacts with minor pseudopilins GspJ and GspK in a heterotrimeric complex (Korotkov and Hol, 2008). Structural characterisation of the GspI-GspJ-GspK heterotrimer by X-ray crystallography showed that GspI is situated at the base of the complex, while the bulky GspK resides at the top. Like modular pilins PilB and PilC, GspK has a large globular C-terminal domain which would prevent further pilin assembly on top of the GspI-GspJ-GspK complex, suggesting that the complex can only be located at the tip of the pseudopilus (Korotkov and Hol, 2008). Homologues of GspI in other bacterial organisms have been shown to play a similar role in T2SS biogenesis. For example, in *P. aeruginosa*, $\mathrm{XcpV_I}$ (14.4 kDa) acts as the nucleator of a quaternary complex, involving minor pseudopilins $\mathrm{XcpX_K}$, $\mathrm{XcpW_J}$ and $\mathrm{XpcU_H}$, which is most likely situated at the pseudopilus tip (Douzi et al, 2009). Likewise, in *K. oxytoca*, PulI (13.8 kDa) is proposed to sequentially bind to PulJ and PulK, and molecular dynamics-based modelling indicates that the self-assembly of the PulI-PulJ-PulK complex promotes pseudopilus polymerisation and protein secretion (Cisneros et al, 2012a). On the other hand, EpsI (12.8 kDa) of the *V. vulnificus* T2SS has so far been found to interact only with minor pseudopilin EpsJ, with the high specificity of the heterodimer formation guided by the pilin globular domains (Yanez et al, 2008a).

The structural homology between PilA and GspI and the overall architectural resemblance between the modular *S. sanguinis* minor pilins and the bulky T2SS pseudopilins suggest that PilA might interact with PilB and/or PilC in a complex which would be presented at the tip of the type IV pilus. This chapter investigates the possible interactions between the *S. sanguinis* minor pilins, and how these minor pilins are incorporated in the pilus via interactions with the major pilins PilE1 and PilE2. Two techniques were employed to test the binding between the *S. sanguinis* full-length pilin proteins (bacterial adenylate cyclase two-hybrid (BACTH) assays) and the globular domains of the *S. sanguinis* pilins (pull-down assays).

4.2 Interactions between the Full-Length Pilin Proteins

4.2.1 Introduction to the BACTH System

The BACTH system was used to test the interactions between the *S. sanguinis* full-length major and minor pilins. This genetic test enables *in vivo* screening of functional interactions between two proteins (Figure 4.2). It is based on the interaction-mediated reconstruction of the cyclic AMP (cAMP) signal transduction pathway and the positive transcription control exerted by cAMP. The BACTH system exploits the catalytic activity of the *Bordetella pertussis* calmodulin (CaM)-dependent adenylate cyclase (CyaA), an enzyme which converts ATP to cAMP and pyrophosphate. *B. pertussis* CyaA is catalytically active when calmodulin (CaM) is available but has only low detectable activity in the absence of CaM (Ladant, 1988). Located within the first 400 residues of the 1706 amino acid-long CyaA protein, the catalytic domain can be proteolytically cleaved into

two complementary fragments T25 (comprising amino acids 1-224) and T18 (amino acids 225-399). When co-expressed as separate entities, the T25 and T18 fragments reassociate in the presence of CaM and form a fully active ternary complex capable of synthesising cAMP. However, if CaM is absent, the two fragments are physically separated and cannot reconstruct the CyaA catalytic domain, yielding only basal CaM-independent activity (Ladant et al, 1989).

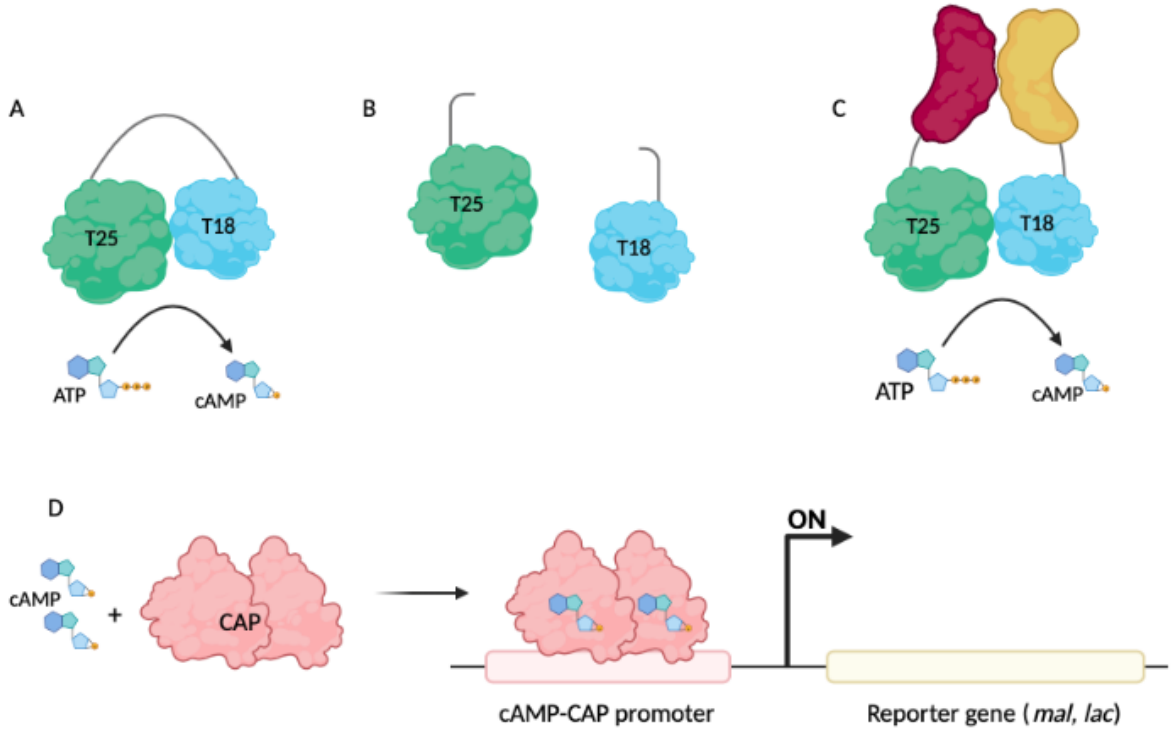


Figure 4.2: **The principle of BACTH assays.** (A) The catalytic domain of *B. pertussis* adenylate cyclase CyaA has low but detectable CaM-independent activity in *E. coli*, resulting in cAMP synthesis. (B) When co-expressed as independent polypetides in *E. coli*, the T25 and T18 fragments do not associate with each other, and no cAMP is produced. (C) T25 and T18 are fused to two interacting proteins and co-expressed in *E. coli*, bringing the fragments close together to induce functional complementation which results in cAMP synthesis. (D) In *E. coli cyaA* mutant strains, the cAMP produced by the functional complementation of T18 and T25 binds to the catabolite gene activator protein (CAP), and the cAMP-CAP complex activates the expression of *lac* and *mal* operons. The resulting fermentation of maltose and lactose can be visualised on indicator plates.

The BACTH system is based on the principle that two interacting proteins, genetically fused to the T25 and T18 fragments and co-expressed in a *cyaA*-deficient *E. coli* strain (*E. coli* lacking CaM or CaM-related proteins), will bring T25 and T18 in close proximity and induce functional complementation between the catalytic domain fragments,

resulting in cAMP synthesis (Karimova et al, 1998). cAMP is a freely diffusible secondary messenger which binds to the catabolite gene activator protein (CAP). A pleiotropic transcription regulator, the cAMP-CAP complex activates the expression of genes involved in the catabolism of carbohydrates, including lactose and maltose. Thus, T25 and T18 functional complementation restores the ability of *E. coli cyaA* mutant strains to ferment lactose and maltose which can be visualised on suitable indicator plates (Karimova et al, 1998).

Full-length pilins contain a hydrophobic α1N helix which acts as a transmembrane domain, anchoring the pilins to the cytoplasmic membrane when they are not assembled into filaments. The BACTH system was employed to study the interaction between the *S. sanguinis* major and minor pilins because it is a sensitive and simple system for the study of interactions between membrane proteins (Karimova et al, 2008). Unlike the yeast two-hybrid system, the BACTH system is not spatially constrained. The two proteins of interest do not have to interact in the vicinity of the transcription machinery – the putative interaction event can be spatially distinct from the transcription event (Karimova et al, 1998). This is because cAMP can freely diffuse from the site of interaction (and synthesis) to bind to CAP and activate transcription. The only requirement for analysing membrane protein interactions by BACTH assays is that the fused T25 and T18 fragments are located in the cytoplasm as cAMP cannot cross the lipid bilayer and therefore has to be synthesised in the cytoplasm.

The localisation of T25 and T18 fragments presents a significant limitation to the BACTH system by preventing the study of proteins that have no cytoplasmic domains. Another limitation, common to all two-hybrid systems, is potential false negative results. Spatial proximity between the CyaA fragments is essential for the BACTH assay. Any steric

constraints, causing T18 and T25 to be held too far apart or in awkward configurations, will prohibit functional complementation and yield a false negative result.

Use of the BACTH system to analyse T4P protein-protein interactions was pioneered in *N. meningitidis* to uncover an intricate interaction network between 11 Pil proteins, including inner membrane proteins PilN, PilO and PilG (Georgiadou et al, 2012). Since then, BACTH assays have been used to characterise interactions between T4P minor pilins (Nguyen et al, 2015b; Treuner-Lange et al, 2020) and T2SS minor pseudopilins (Cisneros et al, 2012a; Nivaskumar et al, 2016).

4.2.2 Construction of BACTH Recombinant Plasmids

To perform the BACTH assays for the *S. sanguinis* T4P system, the full-length pilin genes – *pilA*, *pilB*, *pilC*, *pilE1* and *pilE2* – were amplified from WT *S. sanguinis* 2908 genomic DNA, using the Pfu Ultra II DNA polymerase and primers containing BamHI or KpnI restriction sites (Table 2.2). The amplified genes were cloned into pCR8/GW/TOPO vectors and sent for sequencing to ensure that no mutations were introduced during the PCR. The full-length *pilA* gene contained an internal BamHI restriction site and so in order to facilitate sub-cloning of *pilA* into the BACTH vectors, the internal BamHI restriction site was eliminated by site-directed mutagenesis in the TOPO vector. The SDM primers were designed to alter a single nucleotide in the restriction site without modifying the amino acid sequence of the PilA protein. The point mutation was verified by sequencing.

The TOPO plasmids were then digested with BamHI and KpnI restriction enzymes. The cut-out pilin genes were separated from the TOPO vectors by agarose gel electrophoresis, extracted from the gel and sub-cloned into pUT18C and pKT25 vectors cut with the same enzymes. In the pUT18C and pKT25 vectors, the multiple cloning site (MCS) is

located at the 3' end of the T18 and T25 fragments, respectively (Karimova et al, 1998). This created full-length pilins with an N-terminal in-frame fusion to either the T18 or T25 fragment, which located the CyaA catalytic domain fragments in the cytoplasm (Table 2.15). For each pilin, two different plasmids were generated: pUT18C-pilin and pKT25-pilin (Figure 4.3). From now on, these will be referred to as T18-pilin and T25-pilin for brevity.

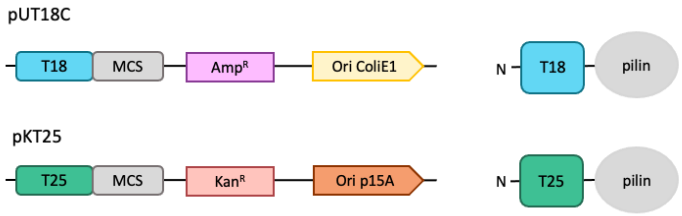


Figure 4.3: **Schematic representation of the BACTH plasmids and hybrid T18 and T25 proteins used in this study.**

The pUT18C and pKT25 plasmids have compatible origins of replication and different antibiotic genes for selection, facilitating the co-expression of the hybrid T18-pilin and T25-pilin proteins in the same *E. coli* cell. The pUT18C vector carries the ColE1 origin of replication and an ampicillin resistance selection marker, while pKT25 uses the p15A origin of replication and has a kanamycin resistance selection marker (Karimova et al, 1998). Furthermore, the pUT18C and pKT25 plasmids both contain a *lac* promoter which controls the transcription of the hybrid proteins. Hybrid protein expression therefore requires both the addition of IPTG and the presence of the cAMP/CAP complex. When functional complementation between T18 and T25 takes place, the level of cAMP rises, creating a positive feedback loop that amplifies the expression of the hybrid proteins. This minimises the possibility of false positive interactions caused by protein overexpression.

4.2.3 Identifying Protein-Protein Interactions between Pilins

The BACTH assays were performed in the *E. coli cyaA* mutant strain BTH101 on MacConkey agar indicator plates, supplemented with maltose. BTH101 cells were co-transformed with a T18-pilin and a T25-pilin plasmid and grown at 30°C. The amount of plasmid used and the volume of cells transformed and plated were standardised between assays so that a similar number of colonies were obtained on each plate. In the absence of functional complementation between T18 and T25, the co-transformed bacteria produce white colonies on MacConkey agar. When functional complementation does occur, the bacteria ferment maltose and acidify the medium, which changes the colour of the phenol red dye, yielding pink to purple colonies. Interactions were first scored after 40 hours and imaged after 48 hours. The combination of BTH101 cells and MacConkey agar was chosen over DHM1 cells and LB agar containing X-gal because previous work in the lab demonstrated that BTH101 cells had more efficient functional complementation and fewer false positive results (Georgiadou et al, 2012).

All 25 possible T18-pilin:T25-pilin combinations were tested. As a negative control, BTH101 cells were co-transformed with empty pUT18C and pKT25 vectors which yielded white colonies. The positive control was performed by co-transformation with pKT25-zip and pUT18C-zip plasmids, which contain leucine zipper domains (35 amino acids, derived from yeast protein GCN4) fused to T18 and T25. The interaction between the leucine zipper domains was very strong and produced dark purple colonies (Karimova et al, 1998). The BACTH assays were repeated twice for the plasmid combinations that gave negative results and three times for the plasmid combinations that produced coloured colonies.

Out of the 25 possible T25-pilin:T18-pilin combinations, 23 were successfully scored for colony colour (Figure 4.4). The homodimeric interactions between PilB and PilC yielded

fewer than 10 colonies for each transformation, making their scoring statistically unreliable. It is possible that the T18-PilB:T25-PilB and T18-PilC:T25-PilC plasmid combinations had very low co-transformation efficiency, or that the homodimeric expression of PilB and PilC was toxic to the cells. Furthermore, in the two combinations analysing the heterodimeric interaction between PilB and PilC (T18-PilB:T25PilC and T18-PilC:T25-PilB), a fraction of the colonies (∼7%) developed a light pink colour, while the majority remained white. Previous work in the lab has shown such interactions to be genuine, but weak and transient (Georgiadou et al, 2012).

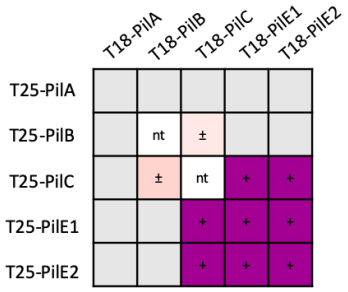


Figure 4.4: **Binary protein-protein interactions between the major and minor pilins of *S. sanguinis* using the BACTH system.** The pilin proteins were fused at the N-termini to adenylate cyclase fragments T18 and T25. All 25 possible combinations were co-transformed in *E. coli* BTH101 cells and plated on MacConkey agar, supplemented with maltose. +, coloured colonies; ±, only a fraction of the colonies were coloured; nt, combination was not tested because there were too few transformants on the plate.

Eight combinations yielded unambiguously positive results with colony colour varying from dark pink to light purple (Figure 4.5). The full-length major pilins PilE1 and PilE2 were found to form heterodimers (T18-PilE1:T25-PilE2 and T18-PilE2:T25-PilE1). This was consistent with previous data showing that the *S. sanguinis* filaments are heteropolymers of PilE1 and PilE2 (Berry et al, 2019). For the major pilins to assemble together in the T4P together, PilE1 and PilE2 must interact with one another. Furthermore, both major pilins appeared to homodimerise (T18-PilE1:T25-PilE1 and T18-PilE2:T25-PilE2). This is not surprising since the first 110 residues of PilE1 and PilE2 are identical. It is likely that the same residues involved in the PilE1-PilE2 heterodimer interaction maintain

the PilE1 and PilE2 homodimers as well.

Interestingly, the major pilins were also involved in an interaction with the full-length minor pilin PilC. This interaction was observed in all T18:T25 combinations between PilC and the major pilins (T18-PilC:T25-PilE1, T18-PilC:T25-PilE2, T18-PilE1:T25-PilC and T18-PilE2:T25-PilC), yielding dark pink to light purple colonies. No other interactions involving the minor pilins were identified in the BACTH assays.

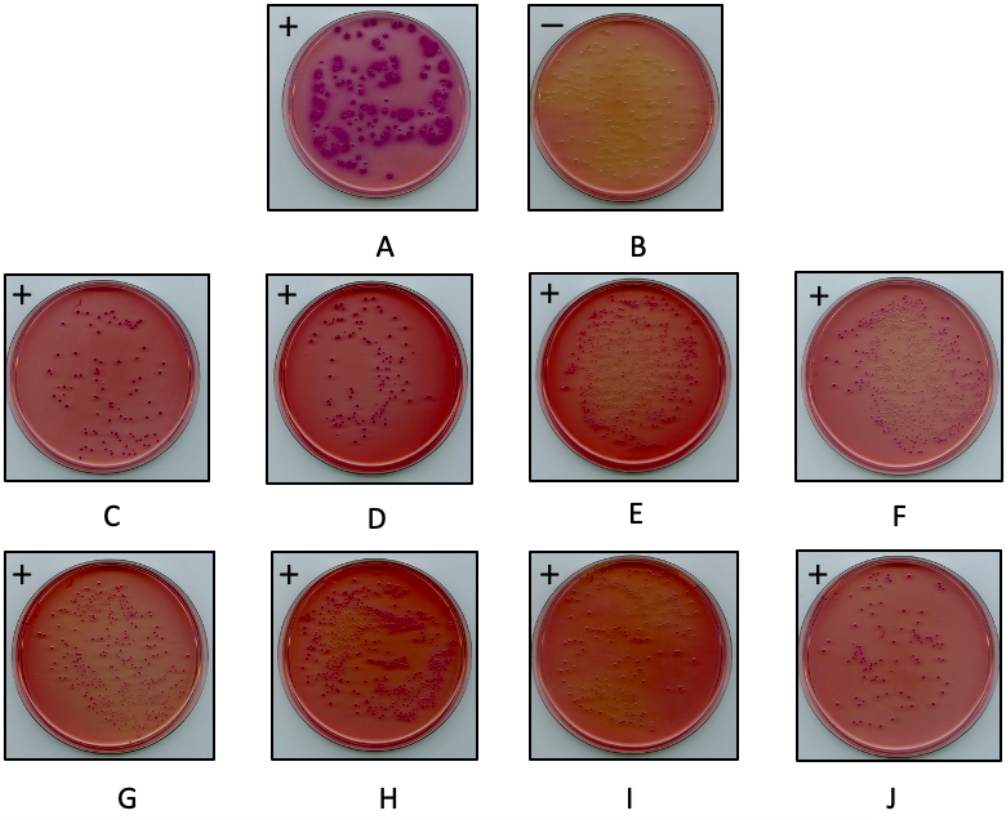


Figure 4.5: **Screening of pilin-pilin interactions with the BACTH system on MacConkey agar, supplemented with maltose.** (A) Positive control: co-transformation of BTH101 with pKT25-zip and pUT18C-zip yielded dark purple colonies. (B) Negative control: co-transformation with pKT25 and pUT18C empty vectors generated white colonies. (C) Interaction between T18-PilC and T25-PilE1, forming purple colonies (D) Interaction between T18-PilC and T25-PilE2, forming purple colonies. (E) Interaction between T18-PilE1 and T25-PilC, forming purple colonies. (F) Interaction between T18-PilE1 and T25-PilC, forming dark pink colonies. (G) Interaction between T18-PilE1 and T25-PilE2, forming dark pink colonies. (H) Interaction between T18-PilE2 and T25-PilE1, forming purple colonies.(I) Interaction between T18-PilE1 and T25-PilE1, forming dark pink colonies. (J) Interaction between T18-PilE2 and T25-PilE2, forming dark pink colonies. The (+) sign indicates the formation of coloured colonies, while the (−) sign indicates white colonies. Representative of three repeats.

4.2.4 Quantifying the Strength of the Identified Pilin-Pilin Interactions

The colouration of colonies on MacConkey agar plates supplemented with maltose is indicative of the efficiency of the complementation between the T18 and T25 hybrids, and consequently the strength of the interaction between the hybrid proteins. None of the identified pilin-pilin interactions produced as deep purple colonies as the positive control, ranging between dark pink to light purple. As discussed above, the hybrid leucine zipper domains in the positive control form an exceptionally strong interaction.

The BACTH system can be used to quantify the efficiency of the T18-T25 functional complementation by measuring the activity of the β-galactosidase enzyme in liquid culture (Karimova et al, 1998). β-galactosidase is encoded by the *lacZ* gene as part of the *lac* operon in *E. coli*, and its expression is under the control of the cAMP-CAP transcription regulator. Thus, in BTH101 cells that express interacting hybrid pilin proteins, functional complementation between the T25 and T18 fragments leads to cAMP synthesis and expression of the β-galactosidase enzyme. β-galactosidase activity therefore reflects the strength of the pilin-pilin interaction.

For the assay to be carried out, the co-transformed bacteria were permeabilised and incubated with ONPG, a colourless substrate of β-galactosidase. Upon hydrolysis, ONPG turns into O-nitrophenol (ONP) which possesses a yellow colour. ONPG hydrolysis was stopped after 20 minutes for positive interactions and 60 minutes for negative reactions (Georgiadou et al, 2012). ONP formation was measured with a spectrophotometer at OD_{420} and used to calculate the β-galactosidase activity. The amount of ONP formed in the liquid culture is directly proportional to the amount of β-galactosidase expressed in

the cells and to the strength of the interaction between the hybrid pilins.

β-galactosidase activity was calculated as described in Section 2.7.2 and expressed in U/mg, where 1 unit corresponds to 1nmol of ONPG hydrolysed per minute at 28°C (Karimova et al, 1998). The β-galactosidase assay was performed three times for each plasmid combination that produced coloured colonies.

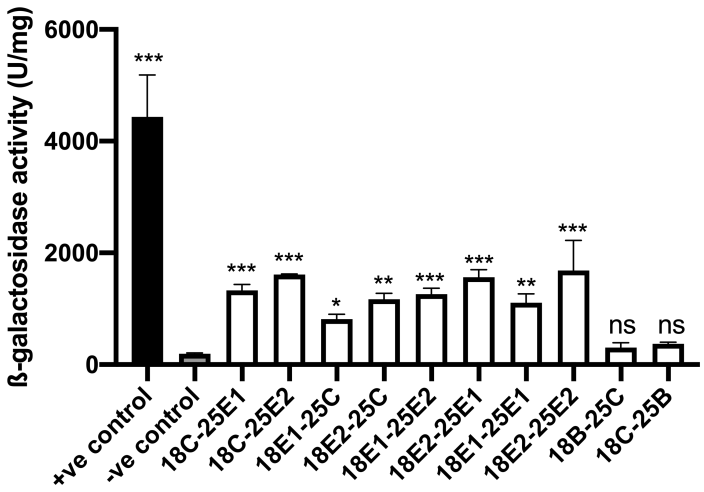


Figure 4.6: **Quantification of the pilin-pilin interactions identified by BACTH.** β-galactosidase assays were performed to quantify the efficiency of functional complementation between T18 and T25, fused to interacting pilin proteins. The results are expressed as units of β-galactosidase activity per milligram of dry-weight bacteria. The graph shows the mean and standard deviation of three independent repeats. The statistical analysis included a one-way ANOVA test, followed by Dunnett's multiple comparison test (ns = non-significant; * $p = 0.033$, ** $p = 0.002$, *** $p < 0.001$).

The negative control defined the background activity of β-galactosidase as 196 $\pm$ 11 U/mg (Figure 4.6). For the positive control, the β-galactosidase activity was measured as 4,433 $\pm$ 749 U/mg. The identified pilin-pilin interactions registered medium β-galactosidase activity, all above the background level. For the T18-PilB:T25-PilC and T18-PilC:T25-PilB combinations, which yielded only a small fraction of coloured colonies, the β-galactosidase activities (303 $\pm$ 90 U/mg and 373 $\pm$ 28 U/mg, respectively) were very close to the background level and statistically not significant, suggesting that the PilB-PilC interactions are quite weak.

The highest β-galactosidase activity was measured for the PilE2 dimer at 1,685 $\pm$ 538

U/mg, followed by the T18-PilC:T25-PilE2 interaction at 1,615 ± 9 U/mg. The lowest activity was recorded for the T18-PilE1:T25-PilC interaction, but it was still eight times higher than the background level. The majority of the identified *S. sanguinis* pilin interactions had β-galactosidase activities above 1,000 U/mg, indicating interactions of medium strength.

4.3 Interactions between the Globular Domains of the Pilin Proteins

4.3.1 Introduction to Pull-Down Assays

While the BACTH assays investigated the interactions between the full-length *S. sanguinis* pilins, pull-down assays were used to test the interactions between the globular domains of the major and minor pilin proteins. Pull-down assays are an *in vitro* screening method for identifying physical interactions between two or more proteins. They exploit the principles of affinity purification and bear strong resemblance to immunoprecipitation, except that a 'bait' protein is used instead of an antibody. The bait protein is tagged and captured by an immobilised affinity ligand specific for that tag. This creates a secondary affinity support, enabling the co-purification of any proteins that interact with the bait (Einarson et al, 2002). Unlike in the BACTH system, pull-down assays allow the direct visualisation of any protein interactions by SDS-PAGE or Western blotting.

In this project, the affinity ligands were immobilised on magnetic beads. Magnetic beads allow quick and easy bead separation through the application of magnetic force. The magnetic stabilisation of the bead pellet facilitates the complete removal of the flow-through, and the non-porous nature of the magnetic beads prevents the trapping of any

contaminants (Ruffert, 2016). This leads to reduction in non-specific binding and overall higher purity of the pull-downs.

4.3.2 Preparation of the Pilin Proteins Used in the Pull-Down Assays

Pull-down assays were performed with truncated *S. sanguinis* pilins, carrying a His_6 (HHHHHH) or a Strep II tag (WSHPQFEK) at their N-termini. In the truncated pilins, the leader peptide and the hydrophobic α1N helix are removed and replaced with an N-terminal affinity tag. The deletion of the α1N helix results in a soluble globular domain whose structure and functional properties are unaltered (Craig et al, 2004). The soluble domains allow for straightforward protein purification and easier protein handling.

The His_6-tagged pilins were cloned into expression vector pET28b by Dr Gurung. In the pET vector system, the expression of target genes is controlled by the bacteriophage T7 RNA polymerase. The *E. coli* BL21 (DE3) competent cells contain a chromosomal copy of the T7 RNA polymerase gene under the control of the *lacUV5* promoter (Dubendorf and Studier, 1991). T7 RNA polymerase expression can be induced by the addition of IPTG (a non-degradable molecular mimic of allolactose) to the bacterial culture. Upon T7 RNA polymerase induction, almost all of the bacterial resources are directed towards the expression of the target gene. In the uninduced state, however, the target gene is transcriptionally silent (Studier and Moffat, 1986).

The Strep II-tagged pilins were generated by amplifying the truncated pilin genes directly from the His_6-tagged pET28b vectors. The primers used in the amplification were designed to replace the His_6 tag with a Strep II tag. The amplified PCR products were purified and sub-cloned into empty pET28b vectors. Colony PCR was used to confirm the successful

insertion of the truncated pilin genes in the plasmids. The plasmids were also sent for sequencing to ensure that no errors were introduced during cloning and to verify the presence of the Strep II tag.

The soluble His_6-tagged and Strep II-tagged pilins were transformed into *E. coli* BL21 (DE3) cells plated on LB agar, supplemented with 50 μg/ml kanamycin. A single colony was used to start an overnight 5 ml LB culture, which was back-diluted into 250 ml kanamycin-supplemented LB the next morning. When the culture reached the desired OD_{600}, protein expression was induced with 0.5 mM IPTG and carried out overnight at 16°C. Expression of the Strep II-tagged PilC protein was performed at 37°C for 3 hours as the protein appeared to be unstable overnight (Figure 4.7).

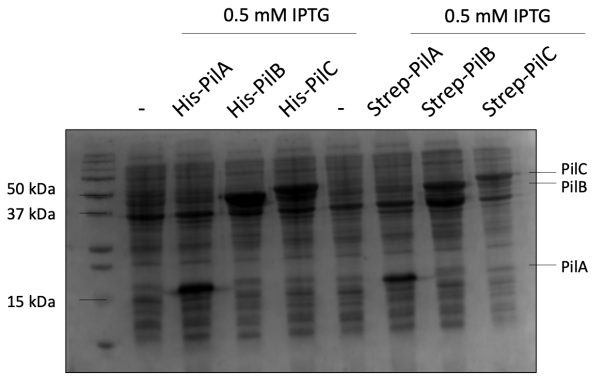


Figure 4.7: **Expression of pilin constructs used in the pull-down assays.** His-PilA, His-PilB, His-PilC, Strep-PilA and Strep-PilB were expressed in BL21 (DE3) *E. coli* cells with 0.5 mM IPTG at 16°C overnight. Strep-PilC was expressed with 0.5 mM IPTG at 37° for 4 hours. One ml samples were taken from each culture before induction (-) and after induction with IPTG and the OD_{600} was recorded. The samples were centrifuged and the pelleted cells were resuspended in 10 μl 1x Laemmli buffer for each 0.1 unit of OD_{600}. The protein expression was analysed by SDS-PAGE, followed by Coomassie staining.

The induced cultures were pelleted, and the cells were lysed by sonication. The expressed proteins were purified by affinity chromatography from the cell lysate. His_6-tagged pilins were purified by immobilised metal affinity chromatography using Ni-NTA agarose resin and gravity-flow chromatography columns, while Strep II tag affinity chromatography was performed on a 5 ml StrepTrap HP column (Section 2.6.3). The elution samples with

the highest protein concentration were pooled together and buffer-exchanged into the PD Binding Buffer (Table 2.14). The imidazole and desthiobiotin from the IMAC Elution Buffer and the Strep Elution Buffer were removed so as not to interfere with the binding between the bait proteins and the beads during the pull-down assays.

4.3.3 Optimisation of the Pull-Down Assays

All five *S. sanguinis* pilins were expressed with a His_6 tag and a Strep II tag. This meant that the pull-downs could be performed using either His-tag magnetic beads or Strep-Tactin magnetic beads. Firstly, the specific binding of bait proteins to both types of magnetic beads was evaluated. The His_6-tagged pilins and the Strep II-tagged pilins were mixed with the His-tag beads and the Strep-Tactin beads, respectively, in 10 separate pull-down reactions. The bait proteins and the beads were incubated for 20 minutes on a rotating wheel at 4°C. The reactions were placed on a magnet to catch the beads and the flow-through was discarded. The bait proteins bound to the beads were eluted by boiling the beads in 1x Laemmli buffer at 100°C for 10 minutes and visualised by Western blotting. The His-tag beads showed higher binding capacity for the His_6-tagged pilins than the Strep-Tactin beads for the Strep II-tagged pilins (data not shown). Therefore, the pull-down assays were performed with His-tag beads using His_6-tagged proteins as bait and Strep II-tagged proteins as prey.

Next, the non-specific binding between the Strep II-tagged pilins and the His-tag beads was tested. In accordance with manufacturer's instructions, the prey proteins were mixed with the beads for 20 minutes on a rotating wheel at 4°C in 50 mM HEPES, 150 mM NaCl, pH 7.6, 0.01% Tween-20. The five reactions were placed on a magnet to catch the beads and the flow-through was discarded. The beads were washed four times with

the same buffer before being boiled. Western blots revealed high-intensity bands of Strep II-tagged proteins bound to His-tag (Figure 4.8). The non-specific binding was especially high for minor pilins PilB and PilC which have large functional adhesive domains (Berry et al, 2019).

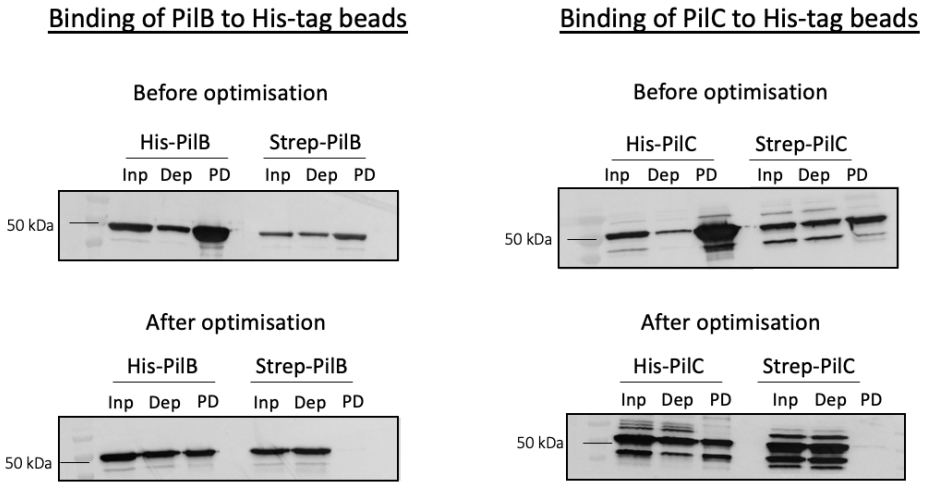


Figure 4.8: **Optimisation of the pull-down assays.** Western blots showing specific and non-specific binding of PilB (left) and PilC (right) to the His-tag beads before and after protocol optimisation. Non-specific binding refers to the interactions between the His-tag beads and prey proteins (Strep-PilB and Strep-PilC), while specific binding occurs between the His-tag beads and the bait proteins (His-PilB and His-PilC). Before optimisation, the proteins were mixed with the beads in 50 mM HEPES, 150 mM NaCl, pH 7.6, 0.01% Tween-20, and the beads were washed four times with the same buffer. The proteins bound to the beads were eluted by boiling in 1x Laemmli buffer at 100°C for 10 minutes. After optimisation, the beads were first blocked with PD Blocking Buffer, containing 5% (w/v) skim milk powder, for 1 hour. The proteins and the beads mixed together for 20 minutes in PD Binding Buffer, containing 0.1% Tween-20 and 10 mM imidazole. The beads were washed 10 times with PD Washing Buffer which includes 300 mM NaCl and 20 mM imidazole. The proteins bound to the beads were eluted in PD Elution Buffer, which contains 500 mM imidazole. Inp, input sample was derived from the bait-prey reaction mixture. Dep, depleted sample was taken from the flow-through after the incubation of the beads with the bait-prey reaction mixture. PD, pull-down sample was the flow-through collected after the elution step.

The non-specific binding was minimised by small consecutive improvements to the protocol. The washing steps were increased from four to ten, and the NaCl concentration in the PD Washing Buffer was doubled to 300 mM. Imidazole, which competes for the binding to the His-tag beads, was added to the PD Binding Buffer (10 mM) and the PD Washing Buffer (20 mM). The concentration of Tween-20 was increased 10-fold to 0.1% in order to prevent not only non-specific binding between the prey proteins and the beads but also non-specific binding between bait and prey proteins.

Two more steps were also added to the protocol. Elution of the proteins bound to the beads was found to minimise the amount of non-specific binding detected by Western blots, compared to boiling of the beads. Elution was performed in 100 μl PD Elution Buffer, containing 500 mM imidazole, for 5 minutes on a rotating wheel at 4°C. The elution reaction was placed on a magnet and the flow-through was collected.

Finally, a pre-blocking step was also included. Before the proteins were mixed with the beads, the beads were incubated for 1 hour in PD Blocking Buffer, containing 5% (w/v) skim milk powder, on a rotating wheel at 4°C. The milk saturated excess protein binding sites and minimised non-specific interactions with the beads. As shown in Figure 4.8, the optimised pull-down assay protocol significantly decreased the binding between Strep II-tagged proteins and the beads without substantially affecting the interaction between the His_6-tagged proteins and the His-tag beads.

While the BACTH system uses maltose fermentation or ONPG hydrolysis as a proxy for protein-protein interactions, the pull-down assays allow the direct analysis of interactions by SDS-PAGE or Western blotting. Western blotting was chosen in this study for its higher sensitivity and hence, its ability to detect even low-affinity interactions. Western blots were performed with antibodies against the individual pilin proteins (anti-PilA, anti-PilB, anti-PilC, anti-PilE1 and anti-PilE2). Each pull-down reaction could thus be probed for the presence of one protein at a time, allowing us to distinguish between proteins with very similar molecular weights, such as PilB (48.5 kDa) and PilC (51.3 kDa). This distinction would have been challenging if SDS-PAGE analysis was performed instead. The higher sensitivity of Western blots is likely to have contributed to the amount of non-specific binding detected initially. It is possible that the non-specific binding would have been negligible if SDS-PAGE analysis was performed instead.

4.3.4 Identifying Dimeric Pilin Interactions

In order to identify dimeric pilin interactions, pull-down assays were performed with alternating pairs of truncated pilins, i.e. His-PilA:Strep-PilB and His-PilB:Strep-PilA. All possible pilin-pilin combinations were tested. The pull-down assays were carried out three times, using different protein preparations. If two pilins interact, both should be detected in the pull-down sample. A negative control was included for each pull-down reaction to account for any non-specific binding between the prey proteins and the beads. In the negative control, the prey protein was incubated with the beads in the absence of the bait protein.

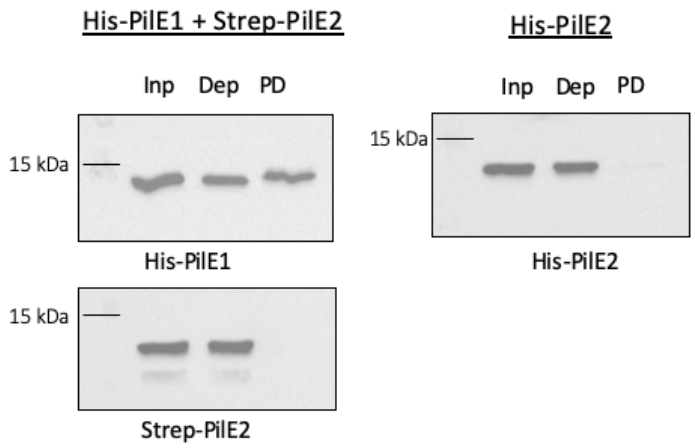


Figure 4.9: **Pull-down assays testing dimeric interactions between *S. sanguinis* major pilins.** (Left) PilE1-PilE2 pull-down: His-PilE1 as bait and Strep-PilE2 as prey. (Right) His-PilE2 binding to His-beads. The His-PilE2 protein failed to bind to the His-tag beads, so it could not be used as bait in pull-down assays. The pull-down assays were performed with magnetic His-tag beads. The Western blots were performed with anti-PilE1 and anti-PilE2 antibodies. Inp, input sample was derived from the bait-prey reaction mixture. Dep, depleted sample was taken from the flow-through after the incubation of the beads with the bait-prey reaction mixture. PD, pull-down sample was the flow-through collected after the elution step.

The pull-downs did not show any co-purification between the *S. sanguinis* major pilins, indicating that the soluble domains of PilE1 and PilE2 did not interact (Figure 4.9). The His-PilE2 protein failed to bind to the His-tag beads (the reason for this is unknown), so the assays could only be performed with His-PilE1 acting as bait and Strep-PilE2 as prey. Similarly, interactions between PilE2 and the minor pilins were also tested with

Strep-PilE2 as the prey protein. PilE1 and PilE2 did not bind to any of the minor pilins either. Notably, the interactions between the full-length PilC and major pilins observed in the BACTH assay were not confirmed by the pull-downs. The soluble PilC pilin failed to co-purify with the soluble major pilin proteins (Figure 4.10).

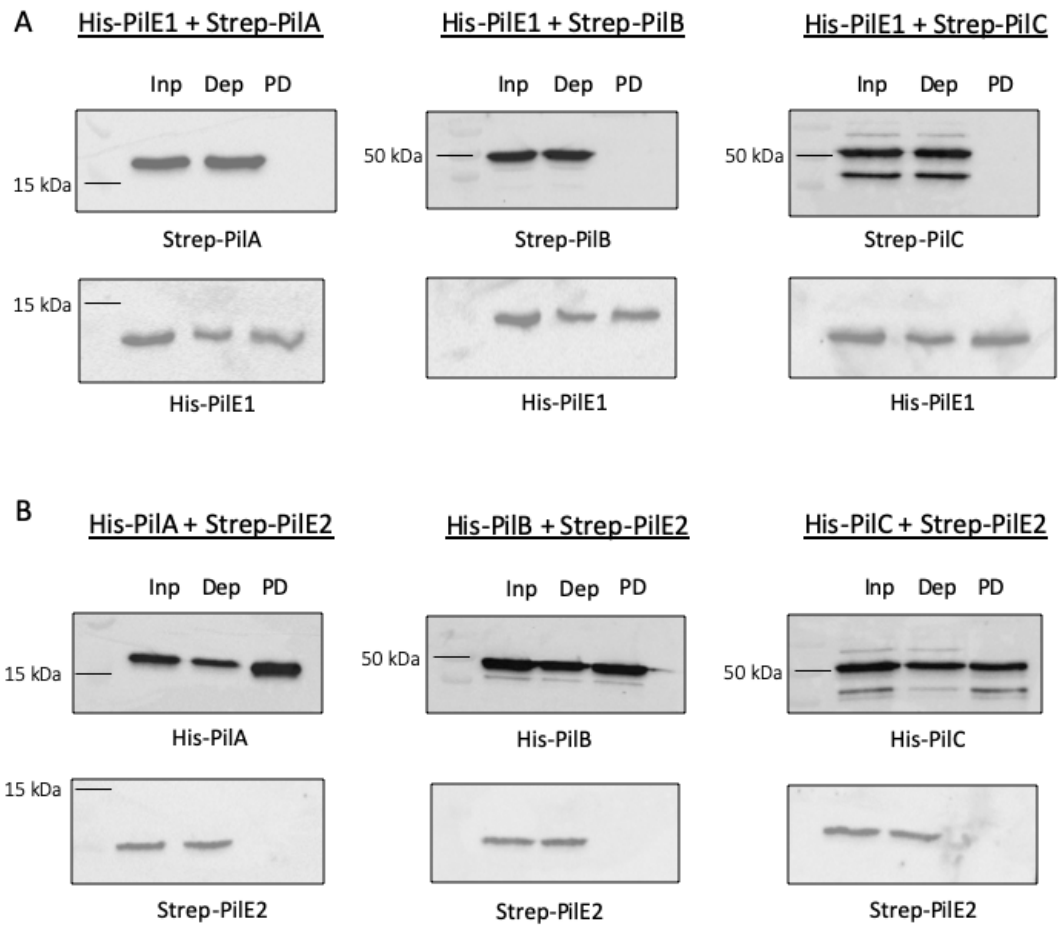


Figure 4.10: **Pull-down assays testing dimeric interactions between *S. sanguinis* major and minor pilins.** (A) Pull-downs between His-PilE1 and the minor pilins. (B) Pull-downs between the minor pilins and Strep-PilE2. The pull-down assays were performed using magnetic His-tag beads. The Western blots were performed with anti-PilA, anti-PilB, anti-PilC, anti-PilE1 and anti-PilE2 antibodies. Inp, input sample was derived from the bait-prey reaction mixture. Dep, depleted sample was taken from the flow-through after the incubation of the beads with the bait-prey reaction mixture. PD, pull-down sample was the flow-through collected after the elution step.

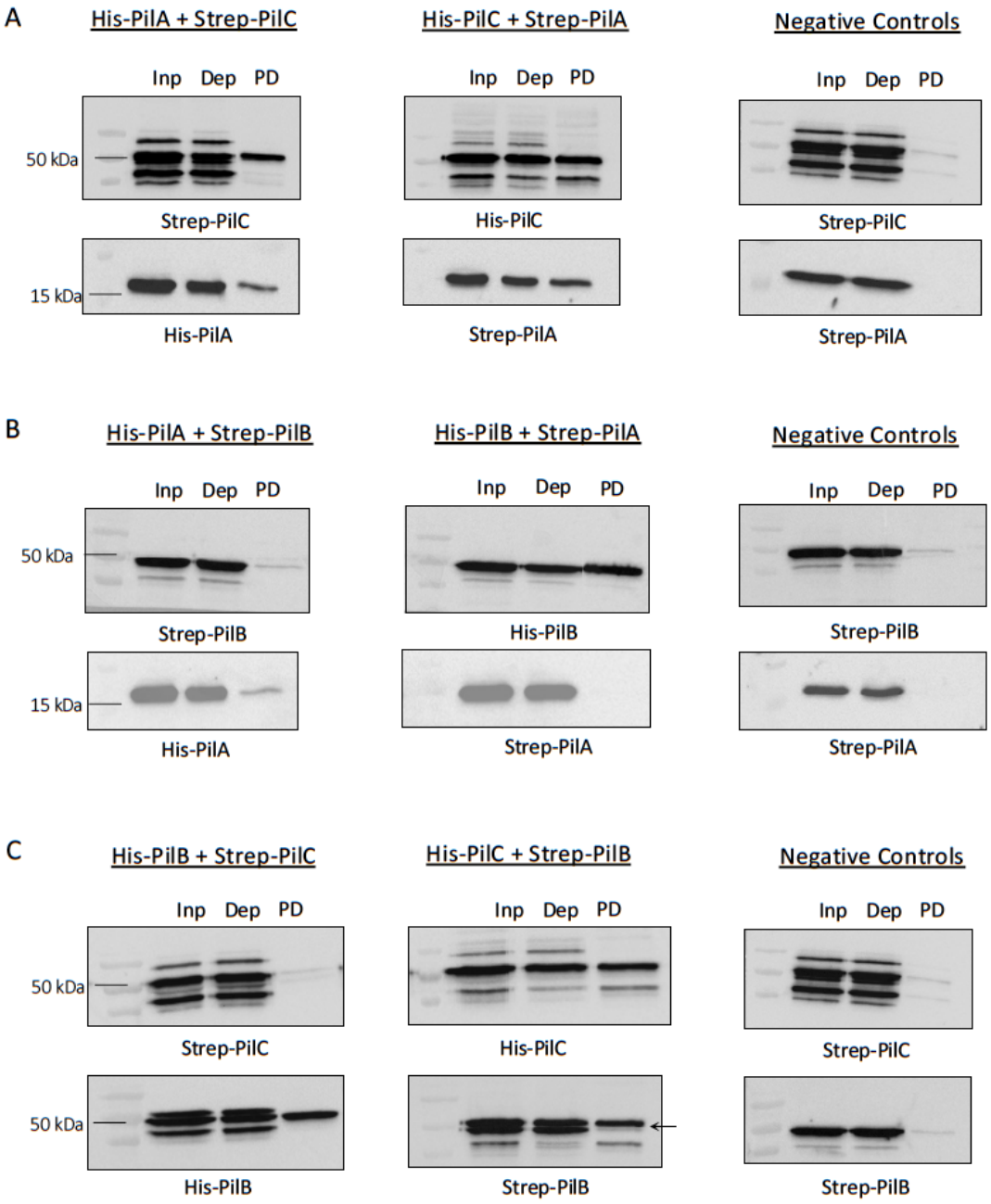


Figure 4.11: **Pull-down assays testing dimeric interactions between *S. sanguinis* minor pilins.** (A) PilA-PilC pull-down: His-PilA as bait (left), His-PilC as bait (medium), negative controls (right). (B) PilA-PilB pull-down: His-PilA as bait (left), His-PilB as bait (medium), negative controls (right). (C) PilB-PilC pull-down: His-PilB as bait (left), His-PilC as bait (medium), negative controls (right). The arrow indicates the band which corresponds to the PilB protein. The pull-down assays were performed using magnetic His-tag beads. The Western blots were performed with anti-PilA, anti-PilB and anti-PilC antibodies. Inp, input sample was derived from the bait-prey reaction mixture. Dep, depleted sample was taken from the flow-through after the incubation of the beads with the bait-prey reaction mixture. PD, pull-down sample was the flow-through collected after the elution step. In the negative control, the prey protein was incubated with the beads in the absence of the bait protein.

Minor pilin PilB also failed to interact with any of the pilins (Figure 4.11). In some of the Western blots, faint bands of the prey proteins were detected in the pull-down (PD) samples, which comprised the flow-through collected after elution. For example, the pull-downs between PilB and PilC showed traces of both Strep-PilB and Strep-PilC in the PD sample (Figure 4.11C), possibly suggesting a weak interaction between the two pilins. The matter was complicated by the cross-reactivity of the anti-PilB antibody which appeared to bind to PilC. The molecular weights of the two proteins indicated that the top band in the anti-PilB Western blots of the PilB-PilC pull-downs corresponded to PilC, whereas the band immediately below it represented PilB (see arrow in Figure 4.11C). Upon close inspection of the respective negative controls, the low-level co-purification of the prey proteins with the bait proteins was attributed to non-specific binding between the prey protein and the magnetic beads.

The pull-down assays revealed an interaction between minor pilins PilA and PilC (Figure 4.11A). The interaction was detected in both pull-down combinations (His-PilA:Strep-PilC and Strep-PilA:His-PilC). Western blots showed two strong bands in the PD samples, corresponding to the PilA and PilC proteins, indicating that the two proteins co-purified. The intensity of the bands was significantly greater than that of the negative controls and was suggestive of a strong interaction.

4.3.5 Testing Oligomeric Pilin Interactions

A significant advantage of the pull-down assays over the BACTH system is the ability to test interactions between more than two proteins. This can be achieved by combining one bait protein with multiple prey proteins and following the same protocol for binding the bait-prey reaction mixture to the beads, washing and elution. Oligomeric interactions

were of interest to us because previous studies have reported pilin complex formation through sequential binding. For instance, in the *P. aeruginosa* T2SS, minor pseudopilin $XcpV_I$ acts as a nucleator, binding $XcpX_K$ and $XcpW_J$ at different sites before this ternary complex recruits $XcpU_H$ (Douzi et al, 2009). Similarly, in the *P. aeruginosa* T4P, minor pilins quaternary complex PilVXWY1 interacts with minor pilins PilE and FimU which mediate the coupling of the complex to the major subunit (Nguyen et al, 2015b).

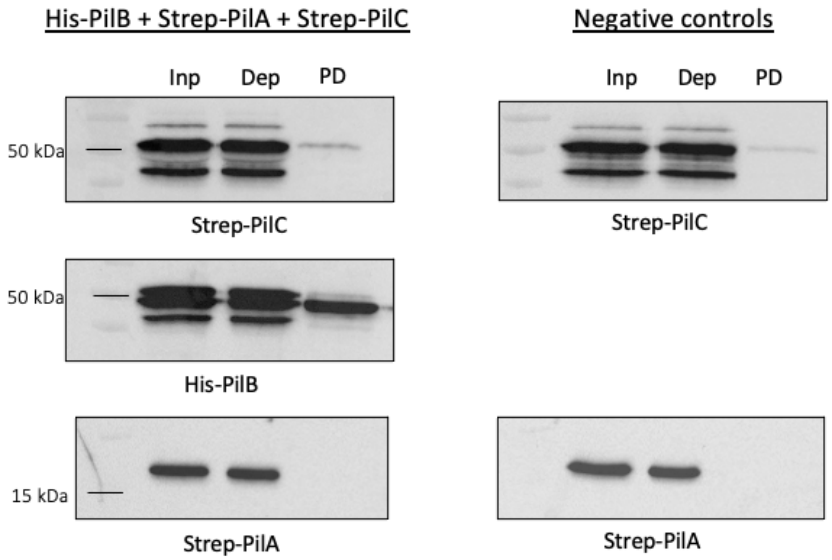


Figure 4.12: **Pull-down assays testing the interactions between all *S. sanguinis* minor pilins.** His-PilB was used as bait protein, while Strep-PilA and Strep-PilC were prey proteins. The pull-down assays were performed using magnetic His-tag beads. The Western blots were performed with anti-PilA, anti-PilB and anti-PilC antibodies. Inp, input sample was derived from the bait-prey reaction mixture. Dep, depleted sample was taken from the flow-through after the incubation of the beads with the bait-prey reaction mixture. PD, pull-down sample was the flow-through collected after the elution step. In the negative control, the prey protein was incubated with the beads in the absence of the bait protein.

It is possible that even though PilB did not interact with either PilA or PilC, the PilAC complex might create a binding interface for PilB. To test this hypothesis, His-PilB was used as bait and mixed together with prey proteins Strep-PilA and Strep-PilC (Figure 4.12). No PilA was detected in the PD sample by Western blotting. A trace amount of PilC seemed to co-purify with PilB, but the intensity of the band was too close to the negative control, and so was most likely the result of non-specific binding. This suggests that the PilAC complex did not interact with minor pilin PilB.

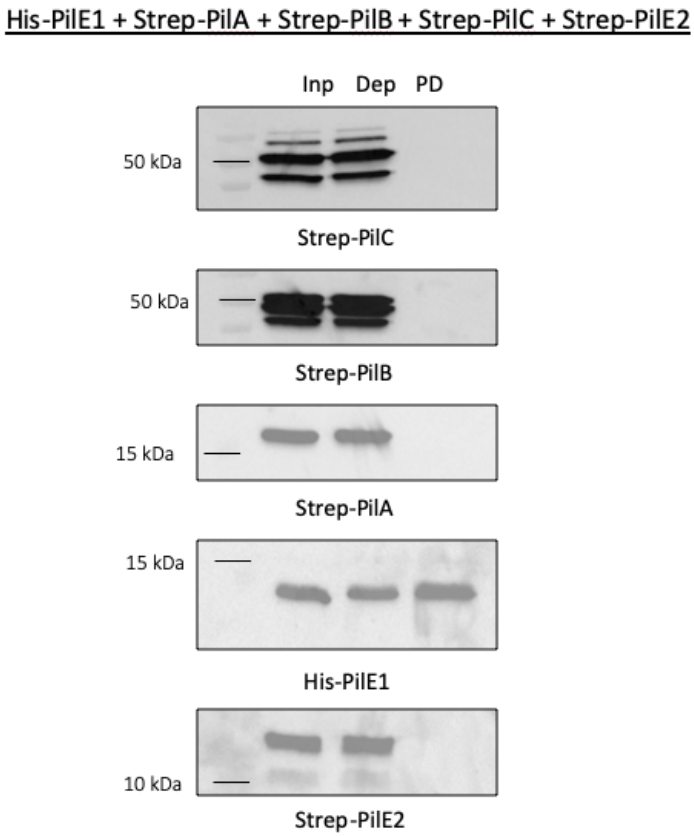


Figure 4.13: **Pull-down assays testing the interactions between all *S. sanguinis* pilins.** His-PilE1 was used as bait protein, while Strep-PilA, Strep-PilB, Strep-PilC and Strep-PilE2 were prey proteins. The pull-down assays were performed using magnetic His-tag beads. The Western blots were performed with anti-PilA, anti-PilB, anti-PilC, anti-PilE1 and anti-PilE2 antibodies. Inp, input sample was derived from the bait-prey reaction mixture. Dep, depleted sample was taken from the flow-through after the incubation of the beads with the bait-prey reaction mixture. PD, pull-down sample was the flow-through collected after the elution step.

Lastly, protein-protein interactions were tested between all five pilins simultaneously. The pull-down assay was performed with His-PilE1 as bait and Strep-PilA, Strep-PilB, Strep-PilC and Strep-PilE2 as prey proteins. The four pilin proteins did not co-purify with the major pilin PilE1 (Figure 4.13). This suggests that no sequential interactions took place between the soluble domains of the *S. sanguinis* major and minor pilins.

4.4 Summary

In this chapter, two screening techniques were employed to test the interactions between the major and minor pilins of the *S. sanguinis* T4P. The BACTH system was used to

interrogate the binding between full-length pilin proteins and to quantify the strength of these interactions, while the pull-down assays focused on the interactions between the soluble globular domains of the pilin proteins.

The full-length major pilins PilE1 and PilE2 were found to both homodimerise and heterodimerise. The heterodimeric interaction was consistent with previous studies showing that the *S. sanguinis* T4P are heteropolymers and that the two major pilins co-purify in sheared filaments (Berry et al, 2019). Interestingly, the soluble domains of PilE1 and PilE2 did not interact, suggesting that the interaction between the two major pilins is primarily mediated by the hydrophobic N-terminal portion of the proteins. This is consistent with the fact that T4P polymerisation is mediated by hydrophobic interactions (Berry and Pelicic, 2015). Moreover, since the major pilins were involved in both homo- and heterodimers, and the first 110 residues of PilE1 and PilE2 are identical, it is highly likely that the major pilin binding interface resides within those N-terminal residues.

Three new pilin-pilin interactions were identified in this chapter. Minor pilin PilC was found to bind to PilA, PilE1 and PilE2. PilC binding to PilA took place via their soluble domains. This was consistent with our hypothesis that PilA is part of a *S. sanguinis* minor pilin complex, presented at the tip of T4P. Indeed, minor pilin complex formation in both T2SS and T4P has been shown to occur through the soluble domains of the comprising proteins (Korotkov and Hol, 2008; Yanez et al, 2008a; Douzi et al, 2009; Cisneros et al, 2012a).

The PilAC interaction was not detected in the BACTH assays, but as discussed above, the BACTH system is prone to potential false negative results. One feasible explanation for the failure to detect the PilAC interaction is that the T18 and T25 domains were not brought together in an optimal configuration, prohibiting the restoration of adenylate

cyclase catalytic activity. It is also possible that the PilAC interaction requires partial membrane extraction of one or both pilins. This has been reported for the T2SS minor pilins in *K. oxytoca* – molecular dynamics has revealed that PulI and PulJ form a staggered complex in the plasma membrane, and binding of PulK to this dimer leads to its partial extraction from the membrane (Cisneros et al, 2012). Similarly, a degree of flexibility within the membrane might be required to accommodate the PilA-PilC interaction. In the BACTH assays, however, the two pilins still possessed their cytoplasmic leader peptides and were further attached to the T18 and T25 fragments, making them tightly anchored to the cytoplasmic membrane.

PilC was also found to bind to the full-length PilE1 and PilE2. The interaction between PilC and the major pilins was observed in the BACTH system, but not in the pull-down assays. This suggests that the hydrophobic α1N helix mediates the binding between PilC-PilE1 and PilC-PilE2. PilC therefore might act as a linker between the minor pilin PilAC complex and the nascent pilus.

While no interaction was detected between PilA and PilB, some binding appeared to take place between PilB and PilC. The interaction between the full-length pilins PilB and PilC resulted in partial and faint functional complementation in the BACTH assays, characterised by β-galactosidase activity that was only slightly higher than background levels. This suggested that the PilB-PilC interaction was weak and transient. In line with this, no binding was observed between the soluble domains of PilB and PilC in the pull-down assays. If an interaction does occur between PilB and PilC, our methods and conditions were not suitable for its detection.

Chapter 5

Characterising the PilA-PilC Interaction

5.1 Introduction

In the previous chapter, two screening techniques were employed to dissect the binding events between the *S. sanguinis* pilin proteins. The pull-down assays revealed a strong interaction between the soluble domains of minor pilins PilA and PilC. The PilAC interaction corroborates our hypothesis that the *S. sanguinis* minor pilins form a tip-located complex as previously reported for the T2SS minor pseudopilins. To our knowledge, this is the first minor type IV pilin complex reported in Gram-positive bacteria.

This chapter focuses on elucidating the PilAC interaction in greater detail, employing the same range of techniques as those used in the study of the interactions between T2SS minor pilins. Firstly, SEC-MALS (size exclusion chromatography multi-angle light scattering) was used to determine the stoichiometry of the PilAC complex. Next, the binding affinity of the minor pilin interaction was obtained by ITC (isothermal titration calorimetry). Structural characterisation of the PilAC complex was also attempted. The complex, unfortunately, proved to be unsuitable for X-ray crystallography studies. Instead, the PilA interface in the minor pilin complex was mapped out by NMR (nuclear magnetic resonance), while the PilC interface was deduced by mutagenesis studies. Protein stability assays were employed to ascertain the function of the PilA-PilC interaction. Lastly, a model was generated for the PilAC complex using AlphaFold, and this model was successfully fitted into the *S. sanguinis* pili.

5.2 Determining the Stoichiometry of the PilAC Interaction

5.2.1 Purification of the PilAC Complex

Determining the stoichiometry of the PilAC interaction by SEC-MALS required that the proteins remain associated during size exclusion chromatography (SEC). SEC is a well-established biochemical technique for separating proteins and protein complexes in solution based on their size and shape. To test whether the PilAC interaction withstands SEC purification, His_6-tagged PilA and PilC soluble proteins were expressed. The same His_6-tagged minor pilin pET28b expression plasmids were used as in the pull-down assays.

The His_6-tagged pilins expression plasmids were transformed into *E. coli* BL21 (DE3) cells and plated on LB agar, supplemented with 50 μg/ml kanamycin. A single colony was used to inoculate a 10 ml overnight LB culture, which was back-diluted into 1 l kanamycin-supplemented LB the next morning. The culture was grown with shaking at 37°C until the OD_{600} reached 0.6-0.8. Protein expression was then induced with 0.5 mM IPTG and performed overnight at 16°C.

The bacterial cultures were pelleted, lysed by sonication, and the cell debris was removed from the cell lysate by centrifugation. Protein purification was performed by immobilised metal affinity chromatography (IMAC), followed by SEC (Section 2.6.3). The eluted proteins from each step were analysed by SDS-PAGE. The purification of His_6-tagged PilA and His_6-tagged PilC is shown in the Appendix. SEC peak fractions, containing the highest protein concentration and the fewest impurities, were pooled together and concentrated appropriately. The purified minor pilins were then mixed together in 1:1

molar ratios and incubated overnight at 4°C in order to allow protein-protein interactions to take place.

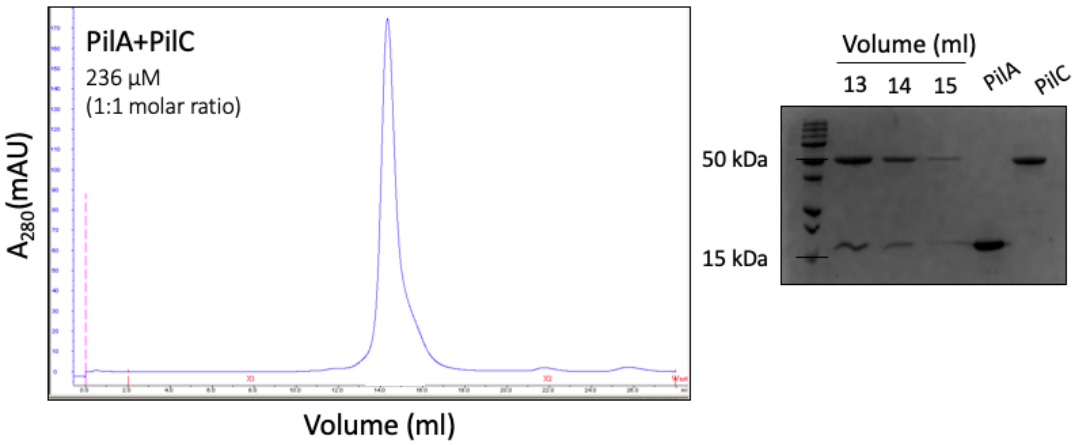


Figure 5.1: **Purification of the PilA-PilC complex.** Complex formation between PilA and PilC mixed in 1:1 molar ratio at 236 μM concentration. SEC was performed using an AKTA Prime system, a S200 10/300 GL column and a 100 μl loop. SDS-PAGE was performed on a 15% gel at 200-220 V for 1 hour and followed by Coomassie Blue staining. Fractions from the peak were analysed by SDS-PAGE alongside individual purified PilA and PilC minor pilins.

The PilAC complex was then run through SEC with a S200 10/300 GL column, which was to be used later in the SEC-MALS experiments. PilA and PilC eluted together in one peak (Figure 5.1). The collected fractions were analysed by SDS-PAGE. Coomassie staining showed that the peak contained both PilA and PilC, and the intensities of the two protein bands remained proportional throughout the peak. Therefore, PilA and PilC co-purified together as a complex. Furthermore, the interactions between PilA and PilC appeared to be strong – it was detected even at low μM concentrations – suggesting that the PilAC interaction was amenable to SEC-MALS analysis.

5.2.2 Introduction to SEC-MALS

SEC is a liquid chromatography method that separates macromolecules of different sizes based on the partial exclusion of these molecules from the pores of the stationary phase. Small molecules diffuse into the pores and their flow through the column is slowed down

according to their size. Large molecules, on the other hand, do not enter the pores and are eluted quicker. Thus, the molecules in the sample are separated based on their size as they pass through the column and are eluted in order of decreasing molecular weight. SEC relies on the elution volume of the macromolecule to estimate its molecular weight. This estimation can often be inaccurate since the elution volume does not depend solely on molecular weight, but can be affected by the hydrodynamic radius of the macromolecule and any non-ideal interactions between the macromolecule and the stationary phase of the column (Some et al, 2019).

Unlike SEC, multi-angle light scattering (MALS) is an 'absolute' technique that determines the molecular weight of a macromolecule in solution based on its light scattering properties. Coupling of SEC to MALS provides a versatile, reliable tool for characterizing solutions of one or more protein species including monomers, native oligomers or aggregates, and heterocomplexes. In SEC-MALS, the column is coupled in-line with a ultraviolet (UV) detector, a MALS detector and a refractomeric detector. The UV signal facilitates the detection of a macromolecule as it elutes from the column; the light scattering (LS) signal is a measure of the proportion of light scattered by the macromolecule into multiple angles relative to the incident laser beam; the differential refractive index (dRI) signal determines concentration based on the change in the solution refractive index due to the presence of the macromolecule. Protein detection by the system results in simultaneous peaks from all three parameters (Some et al, 2019).

SEC-MALS has a few notable limitations. Firstly, the technique separates proteins only by size, so molecules with the same size cannot be separated and properly analysed. Only well-resolved peaks can generate reliable information about molecular weight and oligomeric state. High protein concentration may be required in order to obtain enough

LS signal for MALS analysis, especially for small macromolecules. Lastly, SEC-MALS requires extensive equilibration for achieving a clean baseline signal. This is because SEC columns often release particles from the stationary phase that interfere with the light scattering measurements (Some et al, 2019). For the experiments described below, the equilibration took place over 48-72 hours.

5.2.3 SEC-MALS of the Individual Pilin Proteins

SEC-MALS analysis was first performed on the individual minor pilin proteins. This established the elution profile of each pilin protein and confirmed that the molecular weight estimations were accurate. PilB (48.5 kDa) and PilC (51.3 kDa) behaved as monomers, and their molecular weights were accurately estimated (Figure 5.2B and C). It is worth highlighting that PilB eluted at a higher elution volume than expected, underscoring the unreliability of molecular weight estimates based solely on SEC. The retention time of PilB is likely to have been affected by the small degree of flexibility between its pilin domain and its C-terminal globular domain. It is also possible that PilB interacted with the column due to the adhesive properties of its vWA domain (Raynaud et al, 2021).

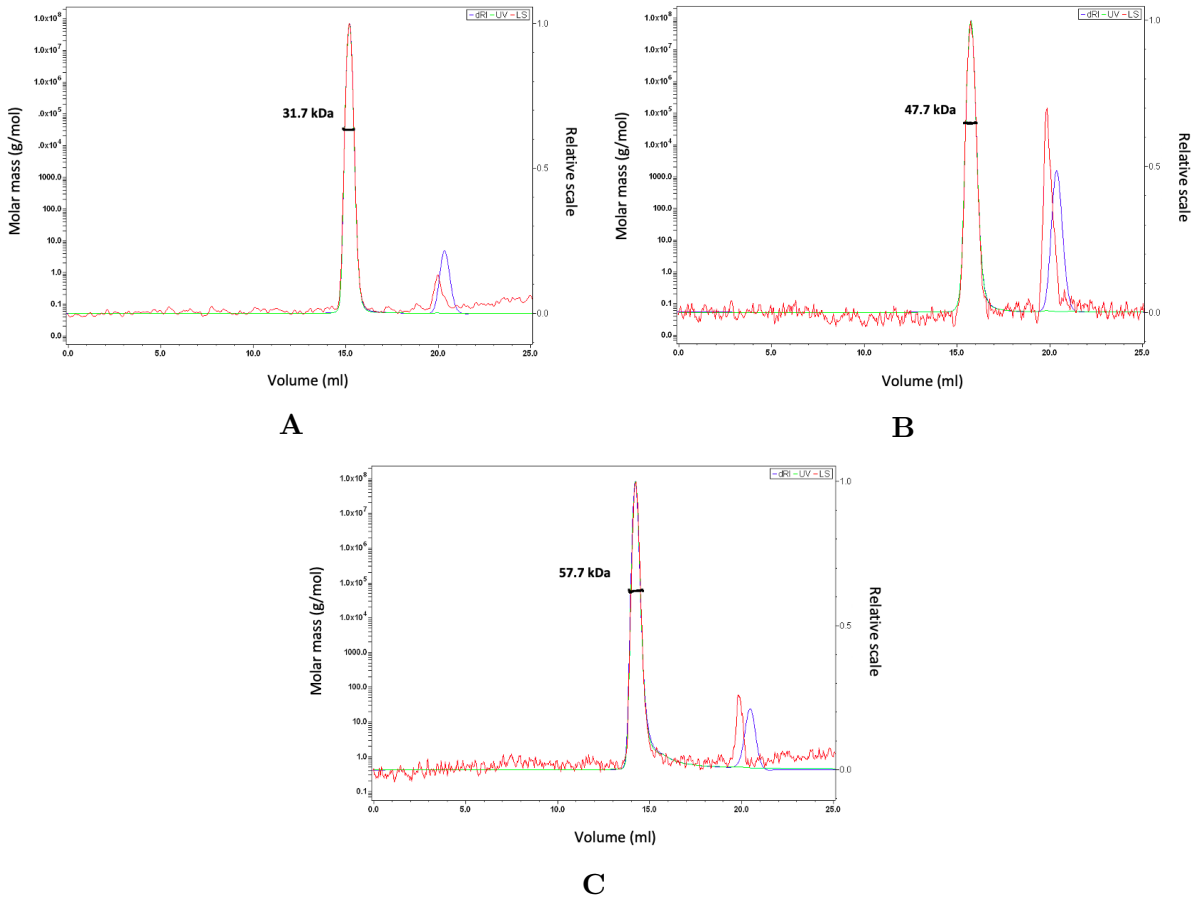


Figure 5.2: **SEC-MALS analysis of individual minor pilin proteins.** (A) PilA (15.7 kDa) (B) PilB (48.5 kDa). (C) PilC (51.3 kDa). The individual proteins were purified by IMAC and SEC and concentrated appropriately. SEC-MALS was performed using an AKTA Prime system with a S200 10/300 GL column, a 100 μl loop and a Wyatt MALS detector. Light scattering (LS) signal is shown in red, ultaviolet light signal (UV) is shown in green and differential refractive index signal (dRI) is shown in blue. Protein detection by the system results in simultaneous peaks from all three parameters. The molecular weight estimated by the ASTRA Chromatography software is included on the left side of the peak.

The molecular weight of PilA (15.7 kDa) was estimated to be ~31.7 kDa, indicating that PilA formed homodimers in solution (Figure 5.2A). Close inspection of the crystal packing in the PilA structure solved by Dr Berry also showed PilA dimerisation with the constitutive PilA monomers oriented in head-to-toe fashion (Figure 5.3). PISA analysis of the dimer revealed that the PilA-PilA interaction was maintained by hydrogen bonds between the residues in the α1C helix and the last ß-strand of the globular domain. Due to the opposing orientations of the monomers, the PilA dimer cannot be incorporated into the T4P filaments and is therefore biologically irrelevant. It would also be impossible for

two PilA molecules to dimerise in *cis* in the cell membrane, which explains why the PilA homodimer was not detected in the BACTH assays. The oligomeric interaction, however, is likely to be important for stabilising PilA in solution.

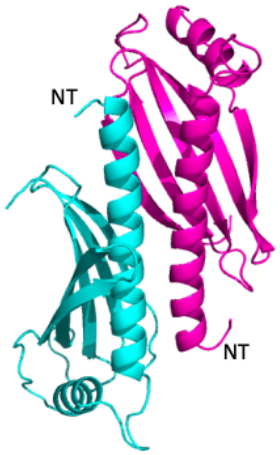


Figure 5.3: **Homodimer in the crystal structure of PilA.**

5.2.4 PilA and PilC Form a Heterodimeric Complex

Next, SEC-MALS analysis was performed on the three minor pilins together. Even though complex formation was detected only between PilA and PilC, PilB was included in the experiment as an internal control. The PilB peak served as a benchmark for the accuracy of the molecular weight estimation by the ASTRA software. Purified PilA, PilB and PilC were mixed together in a 40 μM sample with a final concentration of 4.7 mg/ml.

The SEC-MALS chromatogram revealed two defined, well-resolved peaks (Figure 5.4A). The first peak corresponded to a molecular weight of ~67.8 kDa ± 1.7%; the second peak corresponded to ~48 kDa ± 2.8%. Based on the molecular weights of the minor pilin monomers, the first peak was most likely a heterodimer between PilA and PilC (67 kDa) with 1:1 stoichiometry, while the second peak was monomeric PilB (48.5 kDa). Indeed, overlap of the chromatograms for the individual PilB and PilC proteins with the

chromatogram for the mixed PilA, PilB and PilC sample demonstrated a shift in the PilC peak to the left, representing its interaction with PilA (Figure 5.4B). The second peak, on the other hand, overlapped with the peak of the monomeric PilB, and no shift was detected. This indicated that PilA and PilC formed a heterodimeric complex without the involvement of PilB.

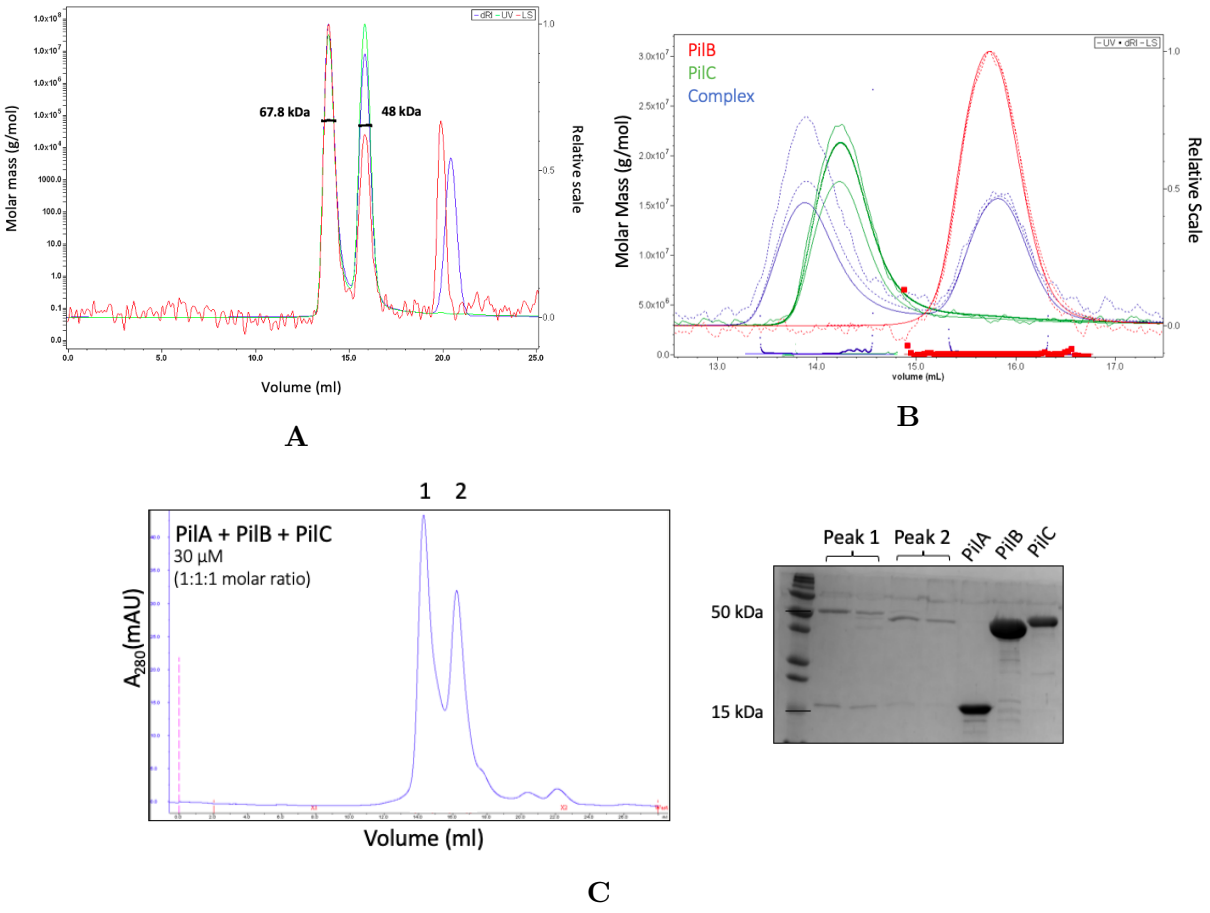


Figure 5.4: **SEC-MALS analysis of complex formation between PilA, PilB and PilC.** (A) Complex formation between PilA, PilB and PilC mixed in 1:1:1 molar ratio at 40 μM concentration. SEC-MALS was performed as described in Figure 5.2. (B) Chromatogram overlap of individual PilB (red), PilC (green) and the mixed PilA, PilB and PilC sample (blue). Chromatogram analysis was performed using the ASTRA Chromatography software. (C) Fraction analysis of complex formation between PilA, PilB and PilC mixed in 1:1:1 molar ratio at 40 μM concentration. SEC was performed using an AKTA Prime system, a S200 10/300 GL column and a 100 μl loop. SDS-PAGE was performed on a 15% gel at 200-220 V for 1 hour and followed by Coomassie Blue staining. Fractions from Peak 1 and Peak 2 were loaded on the gel, along with the individual purified pilin proteins.

SEC-MALS does not facilitate fraction collection, so to verify that the first peak corresponded to a PilAC heterodimer, the PilA-PilB-PilC sample was also run through SEC. The fractions for the two peaks were collected and analysed by SDS-PAGE (Figure 5.4C).

Coomassie staining confirmed that Peak 1 contained both PilA and PilC, while Peak 2 comprised only of PilB. Lastly, SEC-MALS was performed on PilAC in the absence of PilB. Purified PilA and PilC were mixed together at ∼40 μM (Figure 5.5). The chromatogram confirmed the 1:1 stoichiometry of the PilA-PilC interaction with an estimated molecular weight of ∼70.8 kDa ± 0.1% for the complex peak. The slight discrepancy in the molecular weights (∼1 kDa) for the PilAC heterodimer in the PilA-PilB-PilC sample and the PilA-PilC sample might have been caused by the release of small particles from the stationary phase of the column (Some et al, 2019).

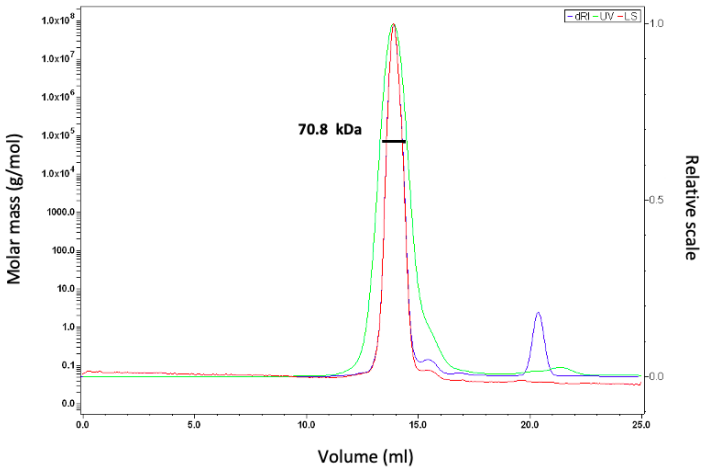


Figure 5.5: **SEC-MALS analysis of PilAC complex formation.** Complex formation between PilA and PilC mixed in 1:1 molar ratio at 40 μM concentration. SEC-MALS was performed as described in Figure 5.2. Protein detection by the system results in simultaneous peaks from all three parameters. The molecular weight was estimated by the ASTRA Chromatography software.

Both the SEC and SEC-MALS chromatograms showed a single peak for the PilAC complex and no peaks for the individual PilA and PilC proteins, indicating that the majority of the available PilA and PilC were incorporated into the complex (Figure 5.1 and Figure 5.5). This suggests that PilA preferred binding to PilC over its homodimeric interaction. In fact, the 1:1 stoichiometry of the PilAC complex indicated that simultaneous binding of PilA and PilC to the monomeric PilA was incompatible. PilC displaced the PilA homodimer, hinting at a partial overlap between the binding interfaces of the PilAC

heterodimer and the PilA homodimer.

5.3 Determining the Affinity of the PilAC Interaction

Analysis of the PilAC complex by SEC-based methods demonstrated that the interaction between PilA and PilC was strong enough to withstand purification at low μM concentrations. Furthermore, PilC binding was found to displace the stable PilA homodimer by engaging with the same binding interface, with higher affinity. To determine the precise binding affinity of PilA and PilC, isothermal titration calorimetry (ITC) was employed. ITC is a well-established technique for analysing the thermodynamic parameters of a binding interaction in a single experiment, including the binding affinity, enthalpy, entropy and stoichiometry. To perform ITC, one reactant is titrated in small volumes into a second reactant under isothermal conditions (Duff et al, 2011). When a titration takes place, a certain amount of complex is formed between the two reactants, causing the release or absorption of heat. ITC records the amount of power (μcal/sec) required to maintain the constant temperature between the sample cell, where the reaction is taking place, and the reference cell, which contains only buffer and serves as a temperature reference (Velazquez et al, 2014). As the reactant in the sample cell becomes saturated, the heat signal diminishes until only background heat is observed, thereby producing a binding curve.

The binding between PilA and PilC was tested using His_6-tagged PilC and His_6-tagged PilA. The two proteins were expressed in *E. coli* BL21 (DE3) cells and purified by affinity chromatography, followed by size exclusion chromatography. PilC was diluted to 20 μM

and placed in the sample cells, while PilA was concentrated to 200 μM and loaded into the syringe. The reference cell was filled with buffer. The ITC experiment revealed an exothermic binding event taking place between PilA and PilC (Figure 5.6). Integration of the peaks generated a dissociation constant (K_D) of 24.4 ± 6.6 nM, indicating a high binding affinity between the two minor pilins.

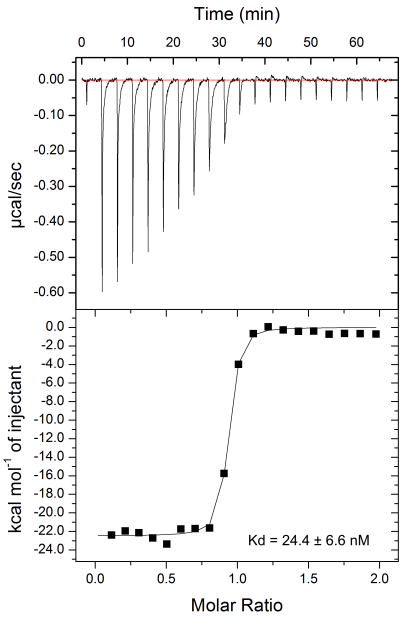


Figure 5.6: **ITC analysis of the PilA-PilC interaction.** ITC was performed as described in Section 2.7.5 using 20 μM PilC and 200 μM PilA. The $K_{(D)}$ value is shown in the bottom right corner. Representative of three repeats.

5.4 Determining the PilAC Interaction Interface

5.4.1 Attempting Crystallography of the PilAC Complex

The high affinity of the PilAC interaction and its ability to withstand SEC purification suggested that the complex would be well-suited to crystallography studies. Solving the structure of the PilAC complex would provide further insight into how the two pilins are incorporated into the filaments and would allow us to generate mutants that disrupt the

PilAC interaction.

In order to establish conditions suitable for the crystallisation of the PilAC complex, His-PilA and His-PilC were purified by affinity chromatography and SEC and mixed together (1:1) at a range of different concentrations. The PilAC complex was then loaded on crystallisation screen trays (Table 2.16) in 200 nl drops with differing ratios of protein to mother liquor (30:70, 40:60, 50:50, 60:40, 80:20). The trays were incubated at 20°C or 4°C to allow for crystals to grow.

Only one crystallisation condition yielded protein crystals – Morpheus F11, which comprised of 0.1 M Tris-BICINE pH 8.5, 60% precipitant mix (40% v/v glycerol, 20% w/v PEG 4000) and 0.12 M monosaccharides (D-glucose, D-Mannose, D-galactose, L-fucose, D-xylose, N-Acetyl-D-glucosamine). The PilAC complex was concentrated to 350 μM (24 mg/ml) and added to the mother liquor at a ratio of 40:60 or 50:50. Following two weeks of incubation at 4°C, the two drops yielded one crystal each. Both crystals appeared to be multilayered, resembling a collection of very thin crystals layered on top of each other (Figure 5.7). Nonetheless, they were sent to the Diamond Light Source and shown to diffract at a resolution of ~4 Å.

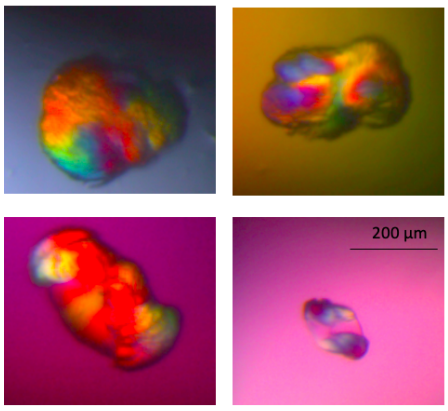


Figure 5.7: **Examples of PilAC crystal morphology.** PilAC crystals obtained from the optimisation screen.

A couple of strategies were pursued in an attempt to improve the quality of the crystals. Firstly, optimisation screens were set up using the Morpheus F11 condition as a starting point and slightly varying the concentration of precipitants along the vertical axis and buffer along the horizontal axis. Optimisation screens were set up as 100:100 nl drops (protein to mother liquor) and incubated at 4°C. The optimised conditions yielded eight new crystals, but their multiplicity was not much improved (Figure 5.7). These crystals were also shown to diffract – one of the collected data sets had a resolution of 2.5 Å and seemed to occupy space group P222. However, none of the crystals could be used to generate structural data due to their multiplicity.

Next, we attempted to use microseed matrix screening to optimise the crystal morphology. In microseed matrix screening, crystalline material (microseeds) is transferred from an initial crystallisation experiment into a range of different and unrelated crystallisation conditions. The microseeds act as a nucleation site for crystal growth, bypassing the *de novo* crystallisation step. The reasoning behind this approach is that the optimal conditions for crystal nucleation might differ from the optimal conditions for crystal growth. Microseed matrix screening has been widely reported to improve crystal quality and diffraction properties, producing high-quality crystals from poor starting points when classical optimisation strategies have failed (D'Arcy et al, 2014).

Thus, PilAC crystals from both the commercial and optimisation screens were broken up and used as microseeds in a range of crystallisation screen conditions (Table 2.16). Each drop contained 150 nl PilAC complex at 350 μM, 50 nl microseed stock and 100 nl mother liquor. After two weeks of incubation at 4°C, twelve new crystals appeared. One of these crystals grew in the original Morpheus F11 condition, while the other eleven crystals developed in unrelated conditions in commercial screens such as JCSG, Wizard,

ProPlex and PACT. These crystals had the same multilayered morphology as the crystals used for microseeding. This time the crystals were broken apart and crystal fragments were sent to the Diamond Light Source to try to obtain data from isolated regions of the crystals. However, once again the crystals appeared to be multiplet and had lower resolution (∼4 Å) than previously observed. Therefore, we decided to pursue a different avenue for studying the PilAC interaction interface.

5.4.2 Mapping the PilA Interaction Interface by NMR

Since the attempts to obtain well-diffracting crystals for the PilAC complex were unsuccessful, nuclear magnetic resonance (NMR) was employed to map the interaction interface of the complex. The smaller size of PilA (as a monomer and a homodimer) made it suitable for NMR analysis. His-PilA was therefore isotopically labelled with ^{15}N as follows. The expression plasmid encoding His-PilA was transformed into *E. coli* BL21 (DE3) and plated on kanamycin-supplemented LB agar. A single colony was used to innoculate a 10 ml LB overnight culture, which was back-diluted into 1 l kanamycin-supplemented LB the next day. This culture was grown at 37°C until the OD_{600} reached 1. The *E. coli* cells were then pelleted and resuspended into 500 ml chemically defined medium (CDM), containing $^{15}NH_4Cl$ and kanamycin. After a 20-minute rest at 16°C, protein expression was induced with 0.5 mM IPTG. The pilin protein was then purified as described above, except SEC was performed in a buffer (Table 2.14), containing 10 mM Na_2HPO_4/NaH_2PO_4 as the HEPES buffer can interefere with the NMR spectrum. 2D-NMR was performed on ^{15}N-PilA on its own at 1.5 mM concentration and on ^{15}N-PilA at 1 mM concentration in the presence of 0.5 mM unlabelled PilC. The NMR experiments and the 2D heteronuclear single quantum correlation (HSQC) spectra analysis were performed by Dr Sheppard.

His-PilA was also labelled with ^{13}C and ^{15}N by expressing the protein in kanamycin-supplemented CDM, containing $^{15}NH_4Cl$ and ^{13}C-D-glucose. The double labelling enabled the assignment of the backbone and side-chain ^{1}H, ^{13}C and ^{15}N atoms in transverse relaxation-optimised spectroscopy (TROSY) triple-resonance experiments and thus the identification of the PilA amino acids in the chemical shift spectrum. The NMR experiments and the chemical shift assignment were performed by Dr Sheppard. Due to the large size of the PilA dimer, the complete assignment of the chemical shift peaks could not be achieved (Figure 5.8A). In total, 71 chemical shift peaks were assigned, amounting to 53% of the PilA residues. There is a probability of misassignment due to the incompleteness of the data, but any increase in certainty or assignment expansion would have required the use of deuterated samples. This was not deemed to be a priority as the crystal structure of PilA was already available.

The overlay between the HSQC spectrum of the ^{15}N-PilA sample and the ^{15}N-PilA-PilC sample allowed the identification of chemical shifts perturbations that took place when PilA bound to PilC. The spectral overlay contained two events: the falling apart of the PilA dimer and the formation of the PilAC heterodimer. The peaks that underwent chemical shifts perturbations could have corresponded to residues involved in either event, so a strategy was required to distinguish between the amino acids important for homodimerisation and PilC binding.

To perform this distinction, Dr Sheppard measured the inter-chain distance between amide moieties in the PilA crystal structure to identify the key residues for homodimer formation. The identified residues were located in the α1C helix of PilA and the last β-sheet of its globular domain (Figure 5.8B). As PilC binding was incompatible with the PilA homodimer form, it was highly likely that residues involved in the PilA dimer overlapped

with the residues in the PilAC binding interface. Indeed, the α1C helix typically maintains pilin-pilin interactions in the type IV pili (Giltner et al, 2012) and can therefore be expected to be important for the PilAC interaction as well. Identifying the residues that were involved in the homodimerisation of PilA, however, allowed us to focus on the residues which were unique for the binding of PilA to PilC.

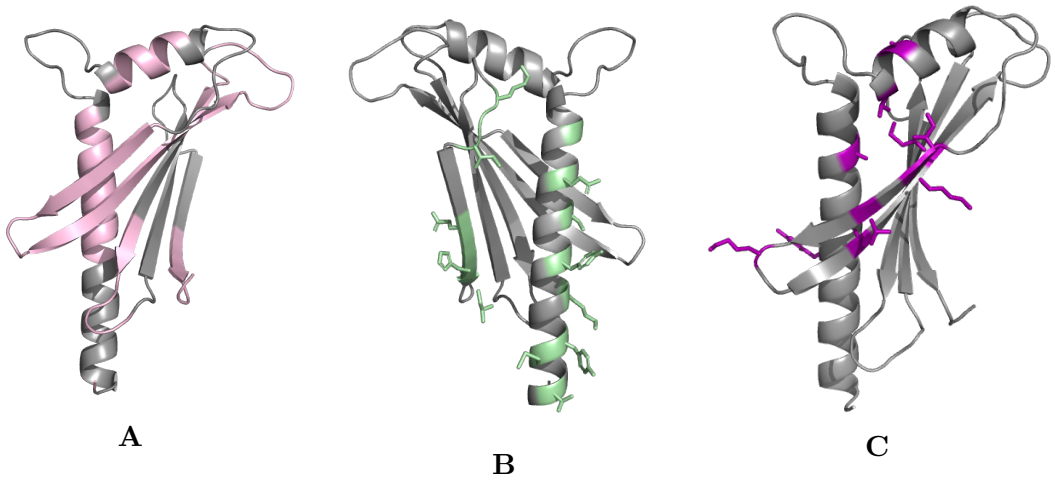


Figure 5.8: **Mapping the PilA binding interface in the PilAC complex.** (A) Degree of chemical shift assignment of PilA. PilA residues assigned for the double-labelled PilA sample by NMR are highlighted in pink. (B) PilA residues involved in the PilA homodimeric interaction are highlighted in green. (C) PilA residues involved in the PilAC interaction. Residues that underwent chemical shifts perturbations of at least one standard deviation over the average are highlighted in purple.

Overall, 14 peaks were observed to undergo chemical shifts changes in the spectral overlay. Of these 14 peaks, 10 underwent chemical shifts perturbations of at least one standard deviation above the average and were confidently assigned to PilA residues. These 10 peaks were identified as key residues in the PilAC interaction (Figure 5.8C): tyrosine 26, alanine 35, valine 50, alanine 52, isoleucine 68, lysine 69, threonine 71, lysine 75, alanine 80 and methionine 84. They were located in the α1C helix, the α2C helix, β-strands 1 and 2 and the loop connecting them. Tyrosine 26, located in the middle of the α1C helix, was marked as a key residue for both PilA dimer formation and PilAC complex formation. The side chains of alanine 35, valine 50, isoleucine 68, alanine 80 and methionine 84 were pointed towards the interior of the PilA structure, suggesting that their chemical shifts

perturbations might be due to conformational changes upon PilAC binding, rather than direct interaction with PilC residues. The remaining four residues – alanine 51, located in the α2C helix; lysine 69 and threonine 71, located on β-strand 1; and lysine 75, located on the loop connecting the first two β-strands – were suitably positioned for engaging with PilC residues upon PilAC complex formation. As the chemical shift assignment was incomplete (only 53%), more residues are likely to be involved in the PilAC interaction than just the ones identified in the NMR experiments.

Nonetheless, we can conclude that the interaction interface in the PilAC complex involved the α1 and α2 helices and the first two β-strands (β1 and β2). Apart from the involvement of the α1 helix, the PilAC interaction interface was completely different from the interaction interface in the PilA dimer. In fact, the heterodimer and the homodimer engaged two opposite sides of PilA.

5.4.3 Disrupting the PilAC Interaction

Next, we attempted to disrupt the interaction between PilA and PilC by mutating the residues involved in the PilA binding site. NMR revealed four PilA residues likely to be involved in the PilAC interaction: alanine 51, located in the α2C helix; lysine 69 and threonine 71, located on β-strand 1; and lysine 75, located on the loop connecting the first two β-strands. In addition, tyrosine 27, found in the middle of the α1C helix, appeared to be a key residue for both PilA dimer formation and PilAC complex formation (Figure 5.9). These five residues were targeted for mutagenesis – bulky and/or charged residues were substituted with alanine, while alanine was substituted with the slightly bigger and charged serine amino acid.

The pET28b plasmid encoding Strep II-tagged PilA was used as a template during site-

directed mutagenesis. The SDM PCR products were incubated with DpnI restriction enzyme for one hour at 37°C to facilitate the digestion of the non-mutated original plasmid before being transformed into *E. coli* DH5α cells and plated on LB agar, supplemented with kanamycin. The pET28b plasmids were then isolated from individual colonies and sequenced to check that the mutation was introduced without any incidental errors. Of the five mutants, three were successfully generated: Strep-PilA$_{A52S}$, Strep-PilA$_{K69A}$ and Strep-PilA$_{T71A}$. PCR amplification with the SDM primers for the other two mutants was ineffective and could not be optimised.

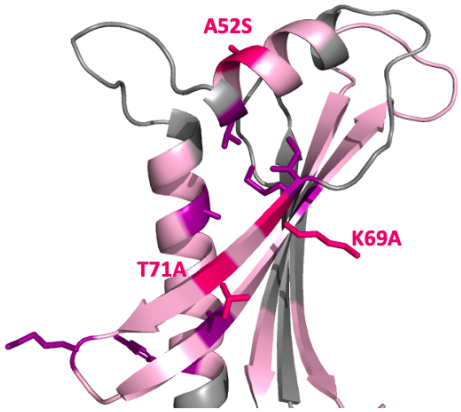


Figure 5.9: **PilA mutations generated to disrupt the PilAC interaction.** The assigned residues are shown in light pink. Residues that underwent chemical shifts perturbations of at least one standard deviation over the average are highlighted in purple. The three residues that were successfully mutated – A52, K69 and T71 – are shown in hot pink and labelled with the corresponding mutation.

The mutant pET28b Strep-PilA plasmids were next transformed into *E. coli* BL21 (DE3) cells. Individual colonies were picked and grown overnight at 37°C in 5 ml kanamycin-supplemented LB. The next day, the cultures were back-diluted into 50 ml LB and when the OD_{600} reached 0.6-0.8, protein expression was induced with 0.5 mM IPTG. Protein expression was tested both at 37°C and 16°C after 3 hours and after 18 hours. The three mutant proteins were produced both at 37°C and after overnight growth at 16°C (Figure 5.10). Strep-PilA$_{K69A}$ appeared to run at a slightly higher molecular weight to the other two mutants and to WT Strep-PilA. The protein was stable and its plasmid

sequence showed no additional mutations. The difference in its motility in the SDS-PAGE could be explained by the substitution of a lysine residue which could serve as a binding site for SDS. The abolishing of an SDS-binding site would decrease the overall negative charge of the protein in Laemmli buffer and consequently decrease its motility during gel electrophoresis.

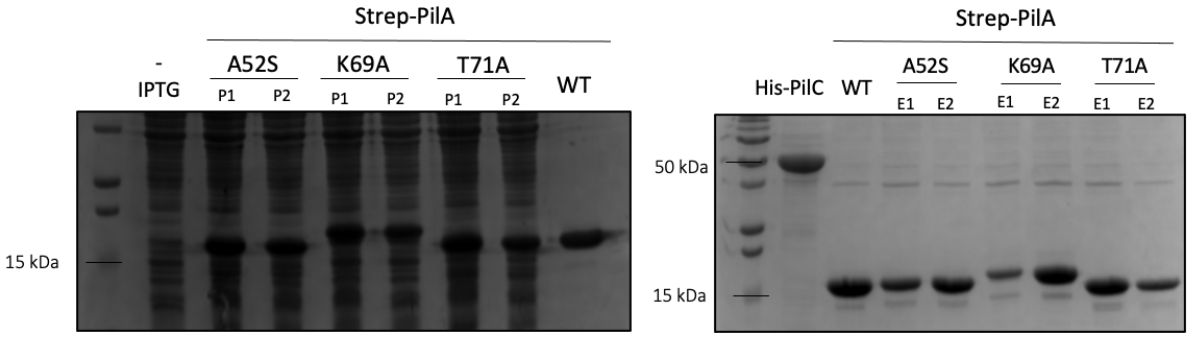


Figure 5.10: **Expression and purification of Strep II-tagged PilA mutants.** (Left) Test expression of mutants Strep-PilA$_{A52S}$, Strep-PilA$_{K69A}$ and Strep-PilA$_{T71A}$. P1 and P2 refer to the two pET28b plasmids tested for each Strep-PilA mutant. Samples were taken before induction (- IPTG) and 18 hours after induction at 16°C. (Right) Affinity chromatography purification of Strep-PilA WT and mutants and His-PilC. E1 and E2 refer to the elution fractions of each protein. SDS-PAGE analysis was performed on 15% gels at 200-220 V for 40 minutes, followed by Coomassie staining.

The effect of the PilA mutations on binding to PilC was tested by pull-down assays using the Strep II-tagged PilA proteins as prey and His-PilC as bait. In fact, the Strep-II tagged PilA construct was chosen as a template for the site-directed mutagenesis because the pull-down assays in Chapter 4 showed no detectable non-specific binding between Strep-PilA and the beads. The pull-down assays were performed as described in Section 2.7.3 and analysed by Western blotting. The pull-down reaction between WT Strep-PilA and His-PilC was included as a positive control. Strep-PilA$_{A52S}$, Strep-PilA$_{K69A}$ and Strep-PilA$_{T71A}$ all exhibited binding to His-PilC comparable to WT Strep-PilA (Figure 5.11). The binding between minor pilins PilA and PilC could not be abolished by a single mutation.The NMR data indicated that the PilA-PilC binding was extensive, suggesting that exhaustive disruption of the interaction interface might be required to abolish the

formation of the PilAC complex.

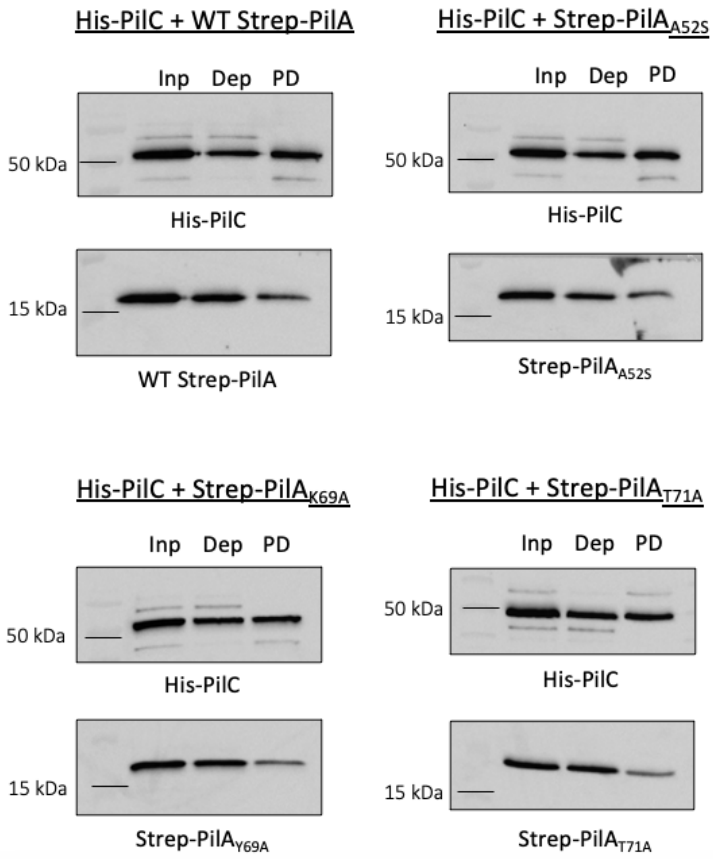


Figure 5.11: **Pull-down assays testing interactions between WT His-PilC, WT PilA and PilA mutants.** The pull-down assays were performed using magnetic His-tag beads.The Western blots were performed with anti-PilA and anti-PilC antibodies. Inp, input sample was derived from the bait-prey reaction mixture. Dep, depleted sample was taken from the flow-through after the incubation of the beads with the bait-prey reaction mixture. PD, pull-down sample was the flow-through collected after the elution step.

Even though the individual mutations failed to disrupt the PilA-PilC interaction, they could still have a negative impact on the binding affinity between the two proteins. We tested this hypothesis by ITC. Since the K_D for the PilA-PilC complex was originally obtained using His_6-tagged PilA, and the mutants are Strep II-tagged, ITC was first performed with the WT Strep II-tagged PilA to confirm that the N-terminal tags do not impact the interaction between the two minor pilins. Indeed, Strep-PilA and His-PilA showed the same binding affinity (Table 5.1). Two of the mutants – Strep-$PilA_{A52S}$ and Strep-$PilA_{K69A}$ – exhibited WT-like binding affinity to His-PilC. In contrast, Strep-$PilA_{T71A}$ had a K_D of 69.2 $\pm$ 10.6 nM, showing a nearly three-fold decrease in binding

affinity compared to WT. This result validated the PilA-PilC interface mapped out by the multidimensional NMR experiments. It also indicated that the threonine 71 residue located at the end of the β1 strand of PilA played a significant role in the interaction between PilA and PilC. As threonine contains a hydroxyl group in its side chain and is thus a polar non-charged amino acid, it is likely that threonine 71 participates in a key hydrogen bond between PilA and PilC.

Cell	**Syringe**	**K_D (nM)**	**ΔH (kcal/mol)**
His-PilC	WT His-PilA	24.4 ± 6.6	– 21.7 ± 0.9
His-PilC	WT Strep-PilA	26.1 ± 6.4	– 18.3 ± 0.7
His-PilC	Strep-PilA$_{A52S}$	30.6 ± 6.7	– 19.3 ± 0.6
His-PilC	Strep-PilA$_{K69A}$	26.8 ± 4.4	– 18.9 ± 1.3
His-PilC	Strep-PilA$_{T71A}$	69.2 ± 10.6	– 19.7 ± 0.3

Table 5.1: **Thermodynamic parameters of the PilA-PilC interaction.**

5.4.4 Pinpointing the Interaction Interface on PilC

The PilC interaction interface could not be analysed by NMR as its molecular size (51.3 kDa) exceeded the current limits of the technique. The modular structure of PilC, however, allowed its binding interface to be mapped by mutagenesis studies. As described in Chapter 1, the soluble domain of PilC consists of a pilin domain, followed by an immunoglobin (Ig)-like fold domain and a concanavalin A-like lectin domain at the C-terminus. PilC deletion mutants were generated so as to either express the C-terminal lectin domain without the pilin domain (His-PilC$_{121\text{-}486}$), or to express the N-terminal pilin domain without the lectin domain (His-PilC$_{34\text{-}253}$). These mutants were then tested for binding to PilA in pull-down assays in order to determine which part of the PilC modular structure was essential for the PilAC interaction.

Pull-Down Assays with the Lectin and Pilin Domains of PilC

The His_6-tagged PilC construct, comprising residues 121 to 486 of the mature pilin, was cloned into pET28b by Dr Raynaud. This construct encompassed both the lectin domain and the Ig-like fold domain. His-$PilC_{121\text{-}486}$ was used as bait in pull-down assays with the Strep-PilA protein acting as prey. Pull-down assays were performed as described in Section 2.7.3. The two proteins failed to co-purify, indicating that no interaction took place between PilA and the PilC lectin domain (Figure 5.12).

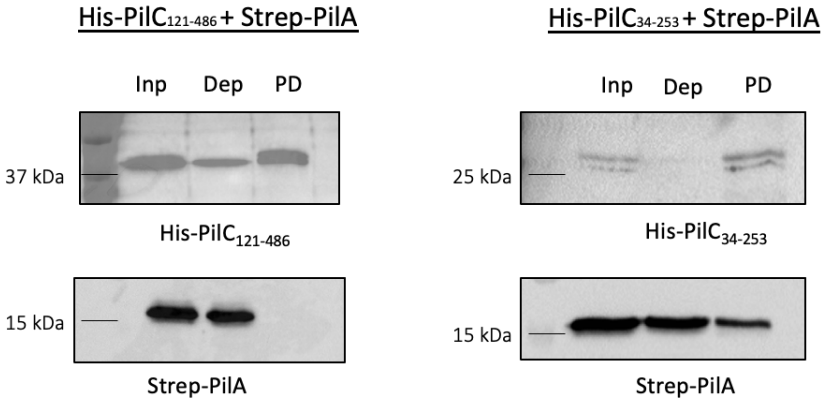
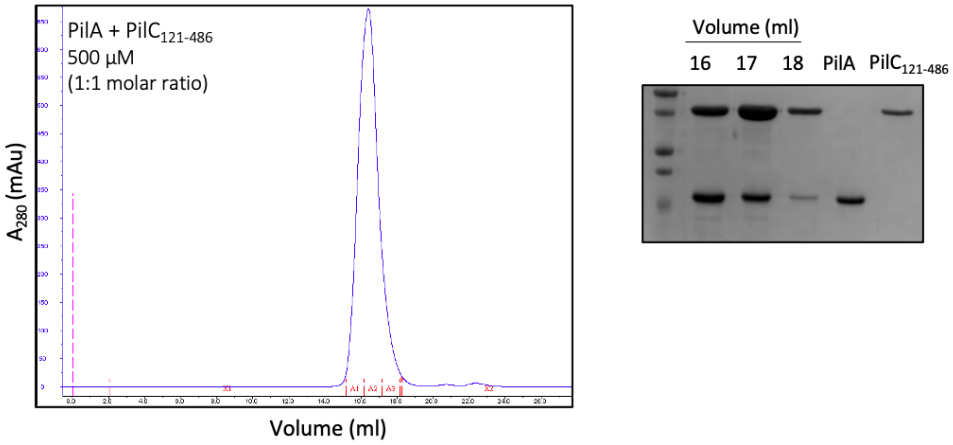


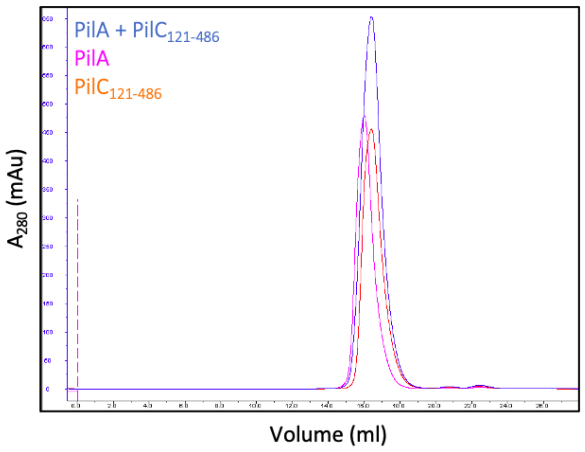
Figure 5.12: **Pull-down assays with the lectin and pilin domains of PilC.** (Left) Pull-down assay between His-$PilC_{121\text{-}486}$ and Strep-PilA. (Right) Pull-down assay between His-$PilC_{34\text{-}253}$ and Strep-PilA. Pull-down assays were performed as described in Section 2.7.3. Western blot analysis was performed with anti-His_6 primary antibody for the detection of the His_6-tagged PilC domains and the anti-PilA primary antibody for the detection of Strep II-tagged PilA. Inp, input sample was derived from the bait-prey reaction mixture. Dep, depleted sample was taken from the flow-through after the incubation of the beads with the bait-prey reaction mixture. PD, pull-down sample was the flow-through collected after the elution step. Representative of three repeats.

The absence of binding was also examined by SEC. Samples of purified His-PilA and His-$PilC_{121\text{-}486}$ were mixed at 500 μM concentration overnight and analysed with the S200 10/300 column. SEC revealed one peak for both proteins, but SDS-PAGE analysis of the peak fractions showed no correlation in the elution profiles of PilA and $PilC_{121\text{-}486}$ (Figure 5.13A). PilA appeared to elute first, closely followed by $PilC_{121\text{-}486}$. This was consistent with the elution volumes of the individual PilA and $PilC_{121\text{-}486}$ proteins. As Figure 5.13B demonstrates, the elution volumes of PilA and $PilC_{121\text{-}486}$ were so close that

the peaks of the two proteins overlapped, but PilA eluted slightly earlier than $PilC_{121\text{-}486}$. PilA failed to interact with $PilC_{121\text{-}486}$, suggesting that the PilAC interaction was likely maintained by the pilin domain, not the lectin domain.



A



B

Figure 5.13: **SEC Analysis of the PilA-$PilC_{121\text{-}486}$ Interaction.** (A) His-PilA and His-$PilC_{121\text{-}486}$ were mixed in 1:1 molar ratio at 500 μM concentration. SEC was performed using an AKTA Prime system, a S200 10/300 GL column and a 100 μl loop. SDS-PAGE was performed on a 15% gel at 200-220 V for 1 hour and followed by Coomassie Blue staining. Fraction peaks were analysed by SDS-PAGE alongside individual PilA (15.7 kDa) and $PilC_{121\text{-}486}$ (41.6 kDa) proteins. (B) Chromatogram overlap of individual PilA (pink), $PilC_{121\text{-}486}$ (orange) proteins and the mixed PilA-$PilC_{121\text{-}486}$ sample (blue). Representative of three repeats.

To test this hypothesis, a His_6-tagged PilC construct (His-$PilC_{34\text{-}253}$), expressing the pilin domain and the Ig-like fold domain, was cloned into pET28b. This construct consisted of residues 34-253 of the mature pilin protein with a predicted molecular weight of 25.3 kDa. Previous work in the lab had shown that the PilC pilin domain was too unstable

to be expressed on its own (Dr Raynaud, unpublished data). His-PilC$_{34\text{-}253}$ therefore was designed to include the Ig-like fold domain in an attempt to stabilise the pilin domain. Expression tests showed that His-PilC$_{34\text{-}253}$ was strongly expressed by *E. coli* BL21 (DE3), so IMAC purification was attempted. Following cell lysis, the majority of His-PilC$_{34\text{-}253}$ was found to locate to the insoluble fraction of the cell lysate (Figure 5.14). This indicated that the His-PilC$_{34\text{-}253}$ protein was also unstable and accumulated in inclusion bodies.

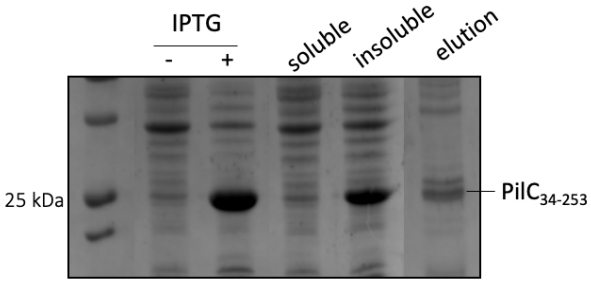


Figure 5.14: **Expression and purification of His-PilC$_{34\text{-}253}$.** SDS-PAGE analysis of His-PilC$_{34\text{-}253}$ expression performed on 15% gels at 200-220 V for 40 minutes, followed by Coomassie staining. Samples were taken from before induction (-IPTG) and after 3h induction at 37°C (+IPTG). After cell lysis, samples were also taken from the soluble and insoluble fractions. Lastly, a sample was taken from the elution following IMAC purification.

Nonetheless, a small amount of His-PilC$_{34\text{-}253}$ was available in the soluble fraction of the cell lysate and could be purified by IMAC (Figure 5.14). The purification yields were very low but sufficient to perform pull-down assays with Strep-PilA. Both the PilC$_{34\text{-}253}$ domain and PilA were detected in the PD samples by Western Blots, indicating that PilA bound to the His-PilC$_{34\text{-}253}$ (Figure 5.12). This suggested that the PilAC interaction was maintained by the pilin domain of PilC.

Constructing a pETDuet-1 Vector, Expressing PilC$_{34\text{-}253}$ and PilA

In order to verify the results from the PilA-PilC$_{34\text{-}253}$ pull-down assays, another attempt was made to stabilise the PilC pilin domain. This time the PilC$_{34\text{-}253}$ construct was cloned into a pETDuet-1 vector alongside PilA. pETDuet vectors are designed for the

co-expression of multiple target genes in *E. coli*, allowing the study of protein complexes. Co-expression has been shown to lead to improvements in yield, solubility, and the protection of proteins or protein subunits from degradation (Tolia and Joshua-Tor, 2006). The pETDuet-1 vector facilitates the co-expression of two genes by encoding two MCS. Each MCS has its own T7 promoter, *lac* operator and ribosome binding site. The pETDuet-1 vector also carries the ColE1 origin of replication, the *lacI* gene and an ampicillin resistance marker (Tolia and Joshua-Tor, 2006). By cloning PilA and the $PilC_{34\text{-}253}$ construct into a pETDuet-1 vector, it was anticipated that PilA might stabilise the pilin domain of PilC and increase its solubility, allowing for more rigorous analysis of the PilA-$PilC_{34\text{-}253}$ interaction.

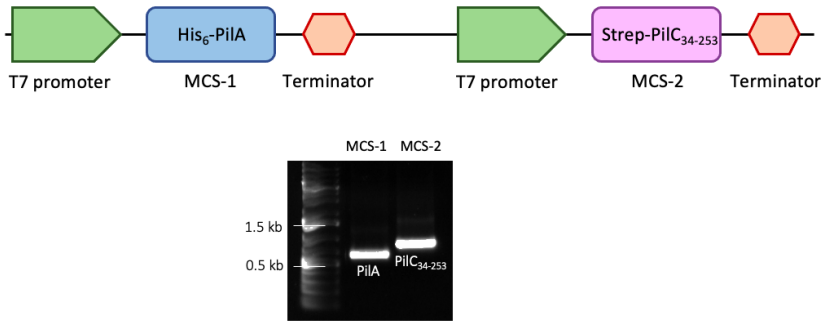


Figure 5.15: **Cloning PilA and $PilC_{34\text{-}253}$ into the pETDuet-1 vector.** Schematic representation of the pETDuet-1 vector, containing His-PilA at MCS-1 and Strep II-tagged $PilC_{34\text{-}253}$ at MCS-2 (top). DNA agarose gel of MCS-1 and MCS-2 amplified by PCR from the final pETDuet-1 plasmid (bottom).

The two proteins were differently tagged in the pETDuet-1 vector: His-PilA and Strep-$PilC_{34\text{-}253}$. Strep-$PilC_{34\text{-}253}$ was cloned first into MCS-2 and sequenced to ensure no mutations had been introduced. Then, His-PilA was cloned into MCS-1 and sent for sequencing to verify the absence of any errors. Lastly, the presence of both PilA and $PilC_{34\text{-}253}$ in the same pETDuet-1 vector was confirmed by PCR and agarose gel electrophoresis (Figure 5.15). The pETDuet-1 plasmid was transformed into *E. coli* BL21 (DE3), and expression tests were performed under a variety of conditions: at 37°C, 22°C and 16°C

with 0.1- 0.5 mM IPTG induction for 3 hours or overnight. At 37°C and 22°C, PilA and the $PilC_{34\text{-}253}$ were successfully produced after IPTG induction for 3 hours and overnight. These cultures were pelleted and lysed by sonication. The insoluble and soluble fractions of the cell lysate were then analysed by SDS-PAGE analysis (Figure 5.16). Both PilA and the $PilC_{34\text{-}253}$ protein were found to locate to the insoluble fraction of the lysate, suggesting that the $PilC_{34\text{-}253}$ bound to PilA and pulled it into the inclusion bodies. Neither protein could be purified from the soluble fraction. Although the stabilisation of the $PilC_{34\text{-}253}$ construct was unsuccessful, the results of the expression tests also substantiated an interaction taking place between PilA and the pilin domain of PilC.

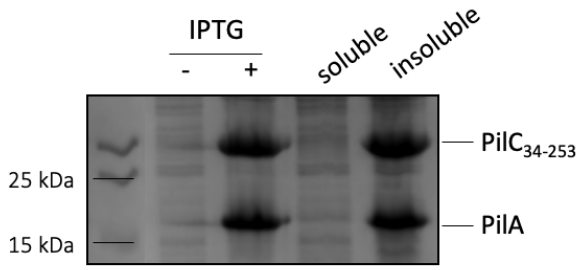


Figure 5.16: **Expression of His-PilA and Strep-$PilC_{34\text{-}253}$ from the pETDuet-1 plasmid.** SDS-PAGE analysis of pETDuet-1 plasmid expression was performed on 15% gels at 200-220 V for 40 minutes, followed by Coomassie staining. Samples were taken from before induction (-IPTG) and after 3h induction at 37°C (+IPTG). Samples were also taken from the supernatant (soluble fraction) and the pellet (insoluble fractions) of the centrifuged cell lysate.

5.5 Determining the Function of the PilAC Interaction

So far, PilA and PilC have been shown to form a heterodimeric complex of high affinity, but the function of this complex remains unclear. Working with the truncated PilC construct led to the observation that PilC was an unstable protein. PilA, in comparison, was a very stable protein, it was well-expressed in *E. coli* cells and had high purification

yields. Since our data indicated an interaction between PilA and the PilC pilin domain, we hypothesised that PilA binding to PilC might have a stabilising effect on PilC.

5.5.1 PilAC Long-Term Stability Tests

To test this hypothesis, long-term stability tests were performed on PilA, PilC and the PilAC complex. Firstly, His-PilA and His-PilC were expressed and purified by IMAC and SEC. The two minor pilin proteins were then mixed together at a concentration of 400 μM, and the PilAC complex was isolated via SEC. PilAC was purified to ensure that all the proteins in the sample were bound in the complex, and no free PilA or PilC was present. The PilAC complex was kept at 4°C for a month alongside aliquots of purified PilA and PilC proteins. Samples were taken once a week and analysed by SDS-PAGE to monitor the stability of the pilins. Degradation of the purified PilA and PilC was compared with the degradation of PilA and PilC bound together in the PilAC complex (Figure 5.17). This experiment was performed in triplicate.

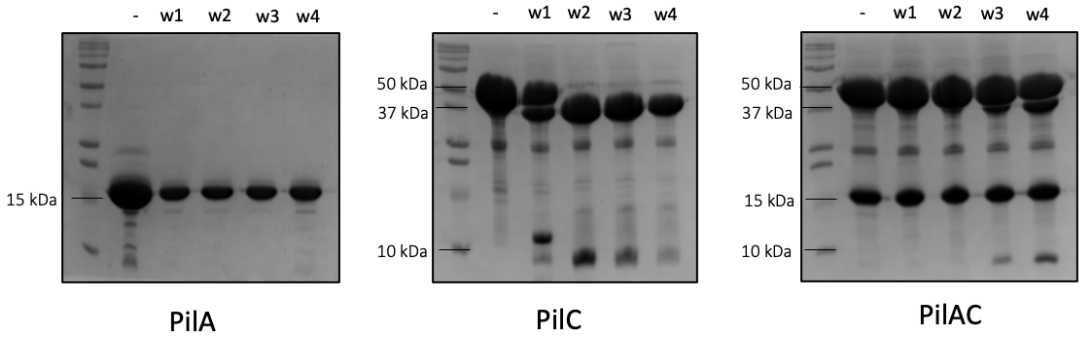


Figure 5.17: **Protein stability of PilA, PilC and PilAC.** SDS-PAGE analysis of PilA, PilC and the PilAC complex. The proteins were stored at 4°C for one month with samples taken once a week.

The purified PilA remained unchanged over the period of five weeks, confirming its stability. The purified PilC began degrading at week one and and was completely degraded by week two. PilC degradation resulted in two fragments with molecular weights of ∼40 kDa and ∼10 kDa. The bigger fragment was sent for mass spectrometry analysis, and

Edman degradation sequencing revealed its first six residues to be NAIETP (Figure 5.18). This placed the site of proteolysis at residue 111 of the full-length PilC protein, which is located right after the pilin domain and before the Ig fold domain. Thus, the bigger fragment of PilC degradation corresponded to the Ig fold and lectin modules, while the smaller fragment was the N-terminal pilin domain.

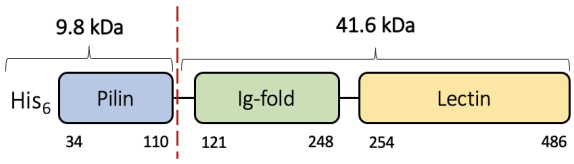


Figure 5.18: **The proteolysis site of PilC in long-term stability assays.** The schematic representation of His-PilC (lacking the hydrophobic α1N residues) shows its three modules: the pilin module, the Ig fold module and the lectin module. The proteolysis site is indicated by a red line. The estimated molecular weights of the two fragments are included above.

In contrast, the PilC in the PilAC complex remained stable for longer and the degradation process was more gradual. During the first two weeks, PilC was still intact; degradation began only at week three, yielding the same two fragments as the purified PilC, but even at week four the majority of PilC was still full-length. PilA binding, thus, delayed and slowed down PilC degradation. PilA stabilised PilC, and specifically its pilin domain.

5.5.2 PilAC Trypsin Sensitivity Assays

The stabilising effect of PilA on PilC was also studied by trypsin sensitivity assays. Trypsin is a pancreatic serine protease which digests proteins by cleaving peptide bonds on the C-terminal side of the amino acid residues lysine and arginine (Baird and Craik, 2013). Trypsin's catalytic activity can be used to evaluate the folding state and stability of proteins. In the trypsin sensitivity assays, the trypsin digestion of PilA and PilC on their own was compared to the digestion of the two proteins bound together in the PilAC complex.

His-PilA and His-PilC were purified by affinity chromatography and SEC. One hundred μl aliquots of PilA, PilC and the PilAC complex were prepared at 80 μM concentration and incubated at 4°C overnight. The protein concentration of each aliquot was measured by Nanodrop (mg/ml), and the total protein concentration was calculated (mg in 100 μl). The PilA, PilC and PilAC aliquots were incubated with the trypsin enzyme for 60 minutes on ice with samples taken at 1, 5, 10, 20, 40 and 60 minutes. The samples were immediately mixed with 2x Laemmli Buffer and boiled at 100°C for 10 minutes in order to terminate the digestion reaction. Initially, a range of trypsin dilutions was tested (1:500, 1:1,000 and 1:2,000), and the 1:1,000 trypsin dilution was deemed to be optimal. The sensitivity assays were therefore performed three times with 1:1,000 trypsin dilution, meaning that for 1 mg total protein concentration in the aliquot, 1 μg of trypsin was added. The digestion samples from each aliquot were analysed by SDS-PAGE, followed by Coomassie staining.

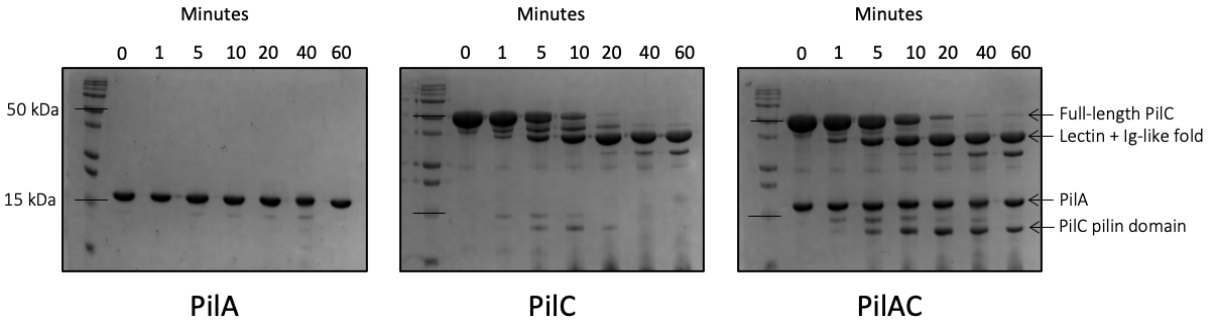


Figure 5.19: **Trypsin sensitivity assay of the PilAC complex.** PilA, PilC and PilAC aliquots were incubated with trypsin (1:1,000 dilution) for 60 minutes on ice. Samples were taken before trypsin was added (0) and at 1, 5, 10, 20, 40, 60 minutes post-trypsin addition. The samples were analysed by SDS-PAGE, followed by Coomassie staining. The digestion fragments are labelled on the right. Representative of three repeats.

The trypsin sensitivity experiments confirmed our previous observations about the stability of the PilA and PilC proteins (Figure 5.19). Following one-hour long incubation with trypsin, PilA was completely intact. PilC digestion, on the other hand, began within the first minute of the enzyme reaction. By 20 minutes, PilC was digested into a ∼40 kDa

fragment, most likely corresponding to the PilC lectin and Ig-like fold domain, and a ∼10 kDa fragment, most likely consisting of the PilC pilin domain. After 20 minutes, the pilin domain was no longer detectable by Coomassie staining, suggesting that it was quickly digested by trypsin (Figure 5.19).

The trypsin sensitivity assays also confirmed that PilA had a stablising effect on PilC. Firstly, PilA slowed down the complete digestion of the full-length PilC into two fragments (Figure 5.19). In the PilAC aliquot, full-length PilC was still present after 20 minutes of trypsin digestion, compared to the PilC aliquot in which full-length PilC was completely digested after 10 minutes. Secondly, the PilAC complex stabilised the smaller digestion fragment of PilC, corresponding to the pilin domain. A strong band for the PilC pilin domain was observed five minutes after the trypsin digestion reaction began and was present until the end of the experiment. In contrast, the pilin domain band in the PilC aliquot was fainter from the beginning and was completely digested after 20 minutes (Figure 5.19). PilA delayed the overall digestion of PilC and stabilised its pilin domain.

5.6 Modelling of the PilAC Complex

Since the PilAC complex was not amenable to structural analysis, we visualised the interaction between the two minor pilins by AlphaFold modelling (Jumper et al, 2021). AlphaFold is an artificial intelligence program developed by Google's Deepmind. It uses the amino acid sequence of a protein as well as information about the physical and geometric constraints of protein folding to predict the protein's 3D structure with atomic accuracy (Jumper et al, 2021). In particular, AlphaFold relies on multiple sequence alignments to identify pairs of residues that have been conserved during evolution and residues that have co-evolved together in order to model intra- and inter-chain contacts. The

most recent version of AlphaFold (AlphaFold-Multimer) uses this principle to predict the interaction interface of protein complexes (Evans et al, 2022).

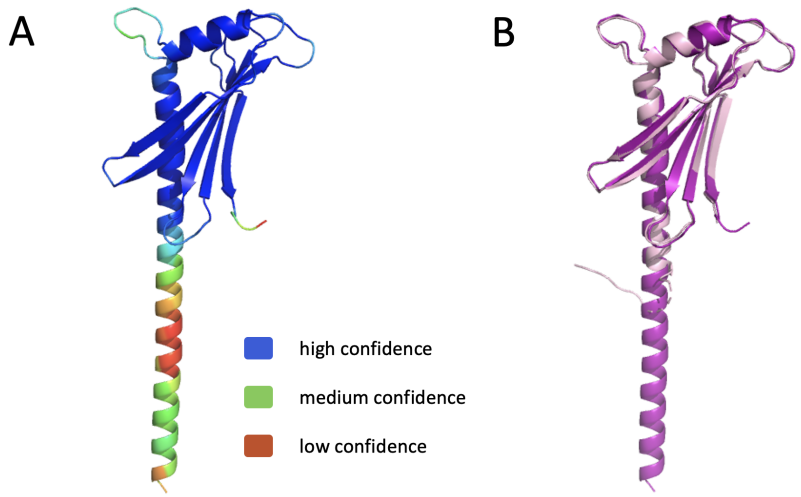


Figure 5.20: **AlphaFold model of minor pilin PilA.** (A) Model confidence. Regions of high confidence (90-100%) are shown in dark blue, regions of low confidence (0-50%) are shown in red (Jumper et al, 2021). (B) Structural superposition of the AlphaFold PilA model and the crystal structure of PilA.

We first modelled the PilA and PilC proteins separately and compared the predicted structures to the crystal structures. While the AlphaFold model of the mature PilA pilin (Figure 5.20) showed near-perfect alignment with the crystal structure of the truncated His-PilA protein (RMSD = 0.52 Å), the prediction for the mature PilC protein was suboptimal. The lectin module and the mature pilin module were modelled with high confidence, but the Ig fold module was largely disordered (Figure 5.21). The AlphaFold model for the lectin domain and the lectin domain from the His-PilC$_{121\text{-}486}$ crystal structure aligned with a RMSD of 0.55 Å (Figure 5.21A). The pilin module, whose structure could not be experimentally determined, was predicted to have a long N-terminal α-helix connected to a highly unusual C-terminal β-sheet via a short loop (Figure 5.21D). The β-sheet was made up of three anti-parallel β-strands: β1 and β2 were positioned perpendicularly to each other and linked by β2 which was kinked in a way that allowed it to run anti-parallel to both β1 and β3. This atypical globular head might explain the inherent instability

of the PilC pilin domain. The Ig-module, however, yielded a low-confidence AlphaFold model which was largely disordered and could not be aligned with our structural data.

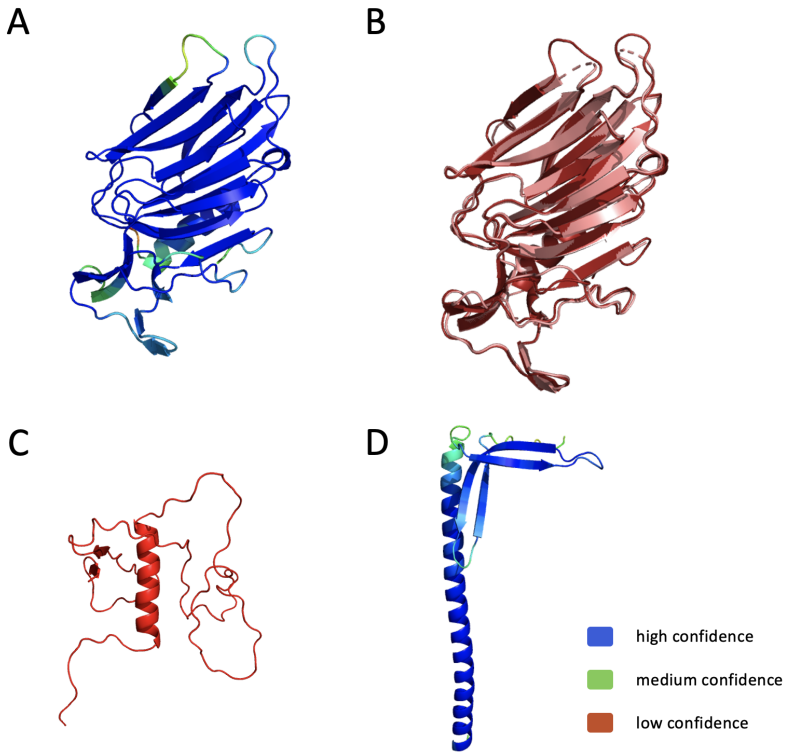


Figure 5.21: **AlphaFold model of minor pilin PilC.** (A) Model confidence for the PilC lectin domain. (B) Structural superposition of the AlphaFold model of the PilC lectin domain and the crystal structure of the PilC lectin domain. (C) AlphaFold model of the PilC Ig fold domain coloured according to confidence. (D) AlphaFold model of the PilC pilin domain coloured according to confidence. Regions of high confidence (90-100%) are shown in dark blue, regions of low confidence (0-50%) are shown in red (Jumper et al, 2021).

We therefore modelled the interaction between the pilin domains of PilA and PilC. AlphaFold generated a high-confidence model (Figure 5.22) which complemented our experimental findings for the PilAC complex. According to PISA analysis (Krissinel and Henrick, 2007), the PilAC binding interface comprised the α1 helices of the two proteins, the PilA α2C helix, β1 and β5 sheets, and the PilC β1 and β3 sheets (Figure 5.23). It was maintained by 21 hydrogen bonds and four salt bridges which supported the high binding affinity recorded for the PilA-PilC interaction (Figure 5.6). In contrast, the PilA homodimer was maintained by only 12 hydrogen bonds and no salt bridges, suggesting that the PilA-PilA interaction was significantly weaker than the PilA-PilC interaction. This

offered a potential explanation for the preferential binding of PilA to PilC observed in the SEC and SEC-MALS experiments (Figure 5.1 and 5.5). The stronger binding between PilC and PilA and the partially overlapping binding interface between the heterodimer and the homodimer allowed PilC to displace the PilA homodimer.

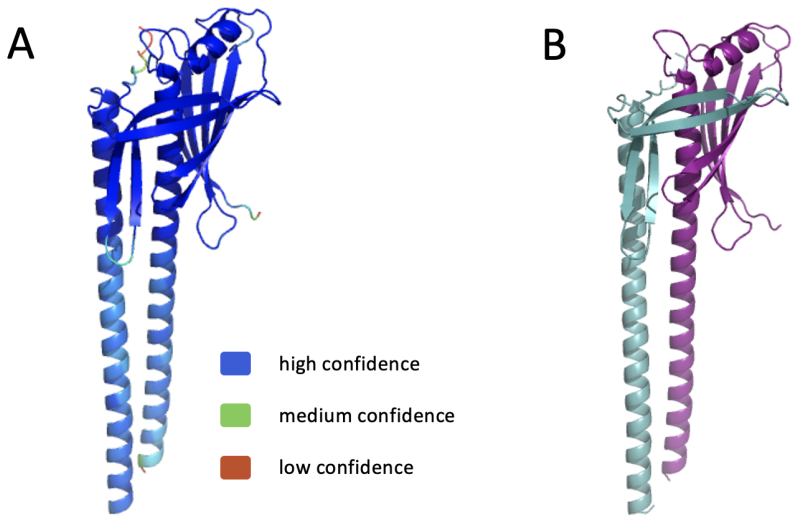


Figure 5.22: **AlphaFold model of the PilA-PilC complex.** (A) Model confidence of the PilAC complex. Regions of high confidence (90-100%) are shown in dark blue, regions of low confidence (0-50%) are shown in red (Jumper et al, 2021; Ellison et al, 2022). (B) AlphaFold model of the interaction between the mature minor pilin PilA protein and the processed pilin domain of minor pilin PilC. PilA is coloured in purple, PilC is coloured in blue.

Furthermore, the model provided a structural basis for the stabilising effect of PilA on PilC. The modelled complex revealed the formation of non-covalent bonds between the PilA α2 helix and β1 strand and the PilC β1 strand, as well as the PilA α1 helix and the PilC β3 strand. The PilA β1 strand and the PilC β1 strand were anti-parallel to each other, and their interaction resulted in an intermolecular β-sheet. Thus, PilA appeared to complement the highly unusual β-sheet of PilC. This β-strand complementation is likely to explain the ability of PilA to stabilise the PilC pilin domain and protect it against degradation (Figure 5.17 and 5.19).

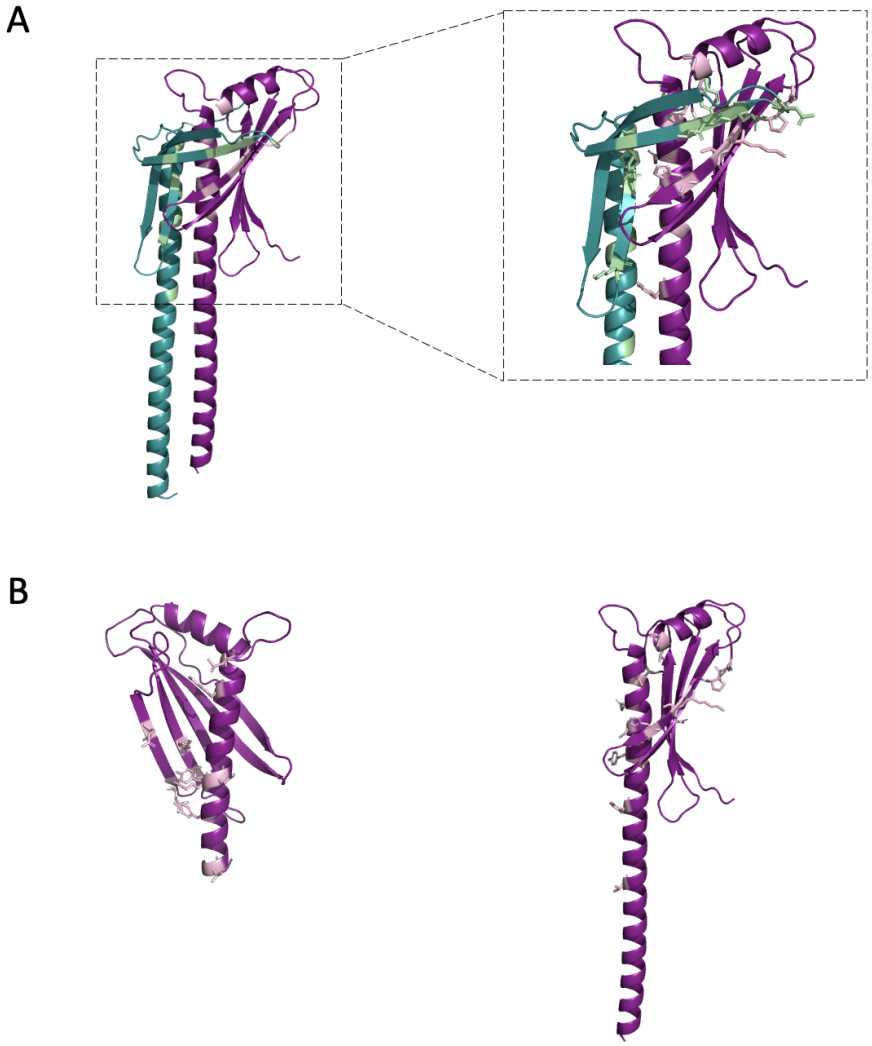


Figure 5.23: **PISA analysis of the PilA heterodimeric and homodimeric interfaces.** (A) PISA analysis of the PilA interface in the AlphaFold model of the PilA-PilC interaction. PilA is shown in purple, PilC is shown in blue. The interacting residues of PilA are highlighted in light pink and PilC in light green. On the right, the interaction between PilA and the unusual β-sheet of the PilC pilin domain is shown in more detail. (B) Comparison of the PilA-PilA interface (left) and the PilA-PilC interface (right). PilA is shown in purple and the interacting residues are shown in light pink. The list of hydrogen bonds and salt bridges is available in the Appendix.

Critically, the PilAC model was in strong agreement with our NMR findings. The regions in PilA that experienced chemical shifts perturbations in the presence of PilC (located in the α1 and α2 helices and the first β-strand of PilA) were also involved in the interaction with PilC in the AlphaFold model. Furthermore, both the AlphaFold model and the NMR analysis indicated that the PilA-PilA and the PilA-PilC binding interfaces engaged opposite sides of the C-terminal globular head of PilA. The α1C helix was involved in

both homodimer and heterodimer formation. But the PilAC interaction engaged the N-terminal portion of the PilA globular head (the α2 helix and the first two β-strands), and the PilA homodimer was maintained by the C-terminal portion (the last two β-strands). This not only provided experimental validation for the AlphaFold model of the PilAC complex, but also suggested the order in which PilA and PilC might be assembled into the filaments. Based on the deduced PilAC binding interface and the axial rise observed between the α1 helices of PilA and PilC in the AlphaFold model, it is highly likely that PilA is incorporated first, followed by PilC.

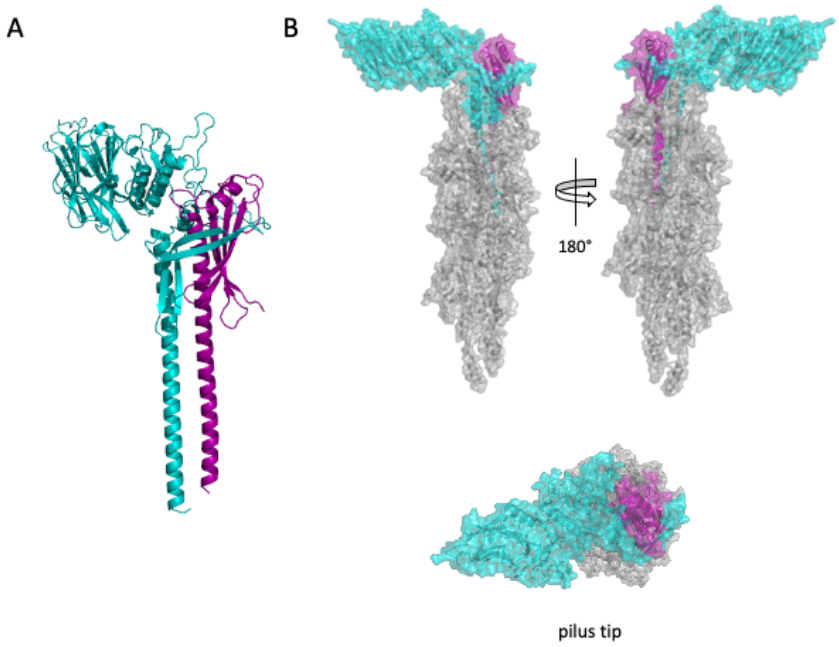


Figure 5.24: **3D model of the PilAC complex in *S. sanguinis* T4P.** (A) Model of the interaction between PilA and the full-length PilC generated by pasting the predicted pilin module onto the crystal structure of the His-PilC$_{121\text{-}486}$ construct (B) Packing of the PilAC complex into the *S. sanguinis* T4P. The heteropolymeric pilus is shown in grey, PilA is purple and PilC is blue. The T4P model was adapted from Berry et al, 2019.

To test this hypothesis, Dr Sheppard generated a model of the PilAC complex in the T4P filaments of *S. sanguinis*. The full-length PilC structure was obtained by pasting the predicted pilin module onto the crystal structure of the His-PilC$_{121\text{-}486}$ (Figure 5.24A). This enabled the PilA-PilC complex to be fitted into a previously described model of the

S. sanguinis T4P (Berry et al, 2019). The complex was readily modelled into the T4P (Figure 5.24B). The α1 helices of the two minor pilins established extensive hydrophobic interactions with the α1 helices of the neighbouring two major pilins. The PilAC complex could only be accommodated at the tip of the filaments in order to avoid steric clashes between the bulky C-terminal modules of PilC and the T4P core. As suggested by the interaction interface, PilA was incorporated first into the pili, followed by PilC whose large C-terminal modules tilted to one side and capped the filaments.

5.7 Summary

In this chapter, we described in detail the interaction between the *S. sanguinis* minor pilins PilA and PilC that was identified in the pull-down assays in Chapter 4. The interaction was amenable to purification which allowed us to determine its stoichiometry. PilA and PilC were found to form a heterodimeric complex with a molecular weight of ∼70 kDa. The interaction between the two minor pilins was confirmed to take place in real time with a K_D of ∼24 nM.

Since attempts to solve the structure of the PilAC complex through X-ray crystallography proved unsuccessful, NMR sprectroscopy was used to determine the binding interface of PilA. Due to the partial chemical shift assignment of PilA, only 10 residues were identified as key for the PilAC interaction. Nonetheless, the NMR data provided a general insight into the PilAC interface. PilC appeared to engage in the usual pilin-pilin interaction with the α1C helix of PilA, came in contact with the α2C helix, which creates the hammer-like shape of PilA, and the first two β-strands in the globular domain of PilA. The NMR analysis was validated by mutagenesis as mutating one of the residues identified as key for the PilAC interaction (threonine 71) caused a nearly three-fold reduction in binding

affinity.

As PilC was too large to be studied by NMR, its PilAC binding interface was instead mapped out by systematic deletion of its modules, followed by pull-down assays with PilA. PilC was found to bind to PilA via its pilin domain; no interaction was detected between PilA and the PilC lectin and Ig fold modules. Indeed, this was confirmed by modelling the PilA-PilC interaction using AlphaFold. The pilin domain of PilC and the mature PilA pilin protein interacted via 21 hydrogen bonds and four salt bridges. The model also provided an explanation of how PilA stabilised PilC. In long-term stability assays and trypsin sensitivity assays, the PilAC interaction protected the PilC pilin domain from degradation. This stabilising effect was achieved by the formation of an intermolecular β-sheet between the two minor pilins. It has led us to propose that PilA acts as a chaperone for PilC in the type IV pili of *S. sanguinis*.

Finally, according to the NMR data and the AlphaFold model, the PilAC interaction involved the N-terminal portion of the PilA globular domain – mainly the α2C helix and the first two β-strands. This suggested that PilA was incorporated into the pili first, followed by PilC. Modelling the PilAC complex in the T4P filaments revealed that it could be fitted only at the tip of the filaments. PilA was the first pilin to be assembled, but PilC still capped the filaments. Its large lectin module loomed over PilA and prevented other pilin subunits from being modelled above it. This likely facilitates the role of PilC as a filament adhesin which will be investigated in the next chapter.

Chapter 6

Functional Analysis of Minor Pilin PilC

6.1 Introduction

While PilA is a small pilin protein of typical size, PilC is unusually large for a processed pilin (52.8 kDa). A recently published study in our lab proposed an elegant evolutionary strategy for increasing the versatility of T4P by grafting functional modules onto pilin domains for presentation by the filaments (Raynaud et al, 2021). The so-called modular pilins are widespread among bacteria and extremely diverse – examples include the minor pilin PilB in *S. sanguinis* (Raynaud et al, 2021), the minor pilin CofB in *E. coli* T4bP (Kolappan et al, 2015), the competence-associated pilin ComZ in *Thermus thermophilus* (Salleh et al, 2019) as well as the minor pseudopilin GspK in the ETEC T2SS (Korotkov and Hol, 2008). Bioinformatic and structural analyses of the *S. sanguinis* minor pilin PilC indicated that it is also a modular pilin.

Figure 6.1: **Structural analysis of minor pilin PilC.** Crystal structure of the His-PilC$_{121\text{-}486}$ construct. The crystal structure was solved by Dr Raynaud before the beginning of this project. The discrete modules of PilC are shown in different colours: the N-terminal Ig module is coloured in light pink, while the C-terminal lectin module is in raspberry. The pilin domain is missing from the structure.

The crystal structure of the PilC$_{121\text{-}486}$ construct, in which the entire pilin domain was replaced by a His$_6$ tag, revealed two distinct modules – an N-terminal immunoglobin

(Ig)-like fold and a C-terminal lectin module, classifying PilC as a trimodular protein (Figure 6.1). The Ig module consists of two β-sheets packed against each other. This type of fold has been found in increasing numbers of bacterial proteins with no sequence homology and diverse functions (Bodelon et al, 2013). The Ig module of PilC shows significant structural homology to the Ig fold of the N-acetylglucosamine-binding protein GbpA from *V. cholerae* (PDB: 2XWX) despite sharing no sequence identity (Wong et al, 2012). Commonly, the Ig fold domains mediate protein-protein interactions and are found in extracellular and cell surface proteins that function in adhesion and host cell invasion (Bodelon et al, 2013). Ig fold domains have been characterised in the *E. coli* fimbrial adhesins FimH and PapG (De Greve et al, 2007), the outer membrane protein intimin (Kelly et al, 1999; Luo et al, 2000) and its *Yersinia pseudotuberculosis* homologue invasin (Hamburger et al, 1999). Curiously, all these proteins encode both an Ig-like fold domain and a lectin domain. FimH and PapG have an N-terminal lectin and a conserved C-terminal pilin domain, both of which display Ig folds (De Greve et al, 2007).

The C-terminal domain of PilC was predicted to belong to the concanavalin A-like lectin/glucanase domain family (InterPro entry IPR013320) (Berry et al, 2019). Lectins bind to carbohydrates specifically and reversibly without structurally altering them (Cavada et al, 2018). In pathogenic organisms, lectins interact with host cell surface carbohydrate ligands and mediate host recognition and tissue adhesion (Vasta, 2009). The lectin module of PilC presents a β-sandwich fold which consists of two opposed anti-parallel β-sheets, each made up of six β-strands. Previous work in the lab has confirmed that the PilC lectin domain is a functional adhesin and determined its glycan specificity through glycan microarrays (Dr Raynaud, unpublished data). PilC was found to bind almost exclusively to two types of glycans: sialylated glycans and glycosaminoglycans

(GAGs). The sialylated ligands are terminated with 3'-sialyllactose (3'-SL) or 3'-sialyl-N-acetyllactosamine (3'-SLN), which are similar trisaccharides (α-NeuNAc-[2-3]-β-D-Gal-[1-4]-D-Glc vs α-NeuNAc-[2-3]-β-D-Gal-[1-4]-D-GlcNAc). The GAG-related ligands of PilC are linear polysaccharides of various lengths consisting of repeating disaccharide units, related to heparin and chondroitin sulfate. The high prevalence of the PilC ligands in the human glycome (Cummings, 2009), suggests a role for this minor pilin in host-pathogen interactions. This chapter focuses on further elucidating the carbohydrate-binding function of PilC and its role in *S. sanguinis* pathogenesis.

6.2 Determining the Affinity of PilC for Sialylated Glycans

The glycan microarrays performed by Dr Raynaud provided an insight into the glycan specificity of the minor pilin PilC (unpublished data). To determine the binding affinity of PilC for the identified sialylated ligands, ITC was employed. ITC is ideally suited for quantifying protein-carbohydrate interactions as the large number of hydrogen bonds involved in the protein-sugar binding produce strongly exothermic interactions which are readily recorded by ITC (Audfray et al, 2013).

The glycan arrays had demonstrated that the pilin domain is dispensable for the binding of PilC to its ligands. Therefore, the ITC experiments were performed with the His-PilC$_{121\text{-}486}$ construct, which encoded the Ig module and the lectin module, but lacked the pilin domain. This construct was selected over His-PilC due to its greater stability and higher purification yields. His-PilC$_{121\text{-}486}$ was purified by affinity chromatography and size exclusion chromatography, buffer-exchanged into ITC P/S Buffer (Table 2.14) and

concentrated to 100 μM. The chemically synthesised sugar ligands were obtained from Dextra Laboratories as lyophilised powder and re-suspended in the same buffer as the PilC construct to obtain a final concentration of 2 mM. The sample cell was loaded with His-PilC$_{121\text{-}486}$, while the syringe was filled with the sugar ligand. The ITC experiments consisted of 30 titrations in total with a gap of 150 seconds between each titration. The affinity of PilC for 3'-SL, 3'-SLN and their constituent subunits (sialic acid and lactose) was tested in three independent experiments.

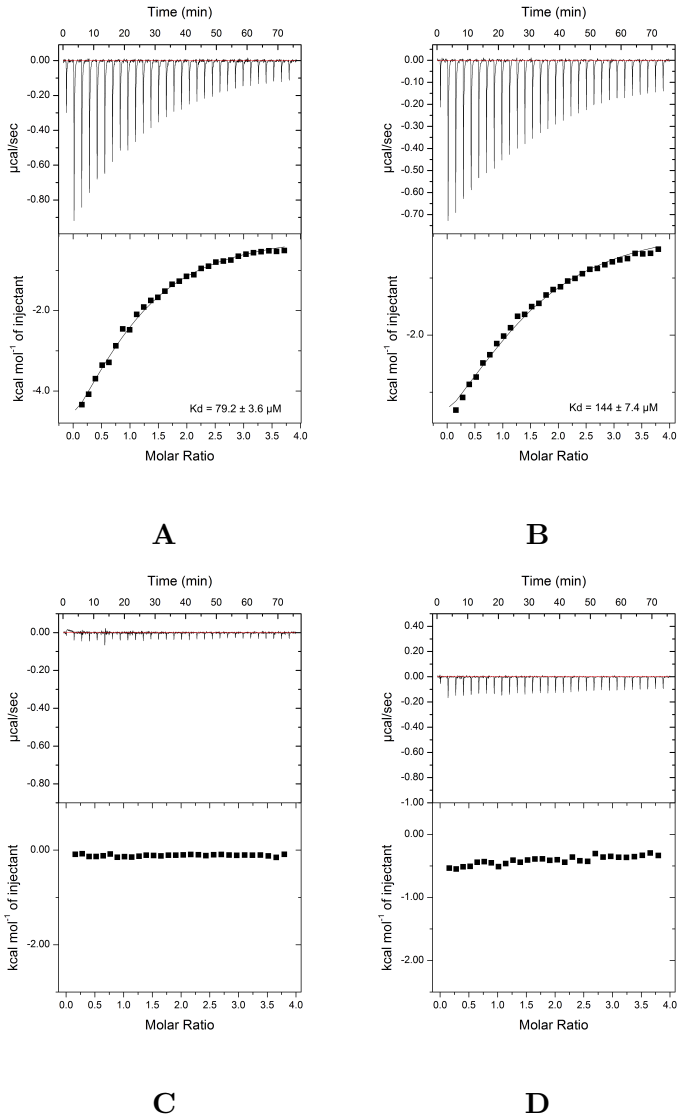


Figure 6.2: **ITC analysis of the PilC-glycan interactions.** (A) His-PilC$_{121\text{-}486}$ binding to 3'-SL. (B) His-PilC$_{121\text{-}486}$ binding to 3'-SLN. (C) His-PilC$_{121\text{-}486}$ binding to lactose. (D) His-PilC$_{121\text{-}486}$ binding to sialic acid.

No binding was observed between His-PilC$_{121\text{-}486}$ and lactose or sialic acid, which are the terminal monosaccharides in 3'-SL and 3'-SLN, but PilC demonstrated clear binding to both 3'-SL and 3'-SLN (Figure 6.2). This was in line with the findings from the glycan microarrays and confirmed that the pilin domain was not involved in the pilin-sugar interaction. The affinity of PilC for 3'-SL appeared to be slightly higher (K_D = 79.2 ± 0.4 μM) compared to 3'-SLN (K_D = 144 ± 7.4 μM). The interaction between PilC and the two ligands occurred in the absence of Ca^{2+}, suggesting that carbohydrate binding is not metal-ion dependent, setting PilC apart from the C-type lectins.

We hypothesised that the interaction between PilC and its sugar ligands was mediated by the C-terminal lectin module. To confirm this, a PilC$_{254\text{-}486}$ construct was designed in which the pilin module and the Ig fold were replaced by a His_6-tag. His-PilC$_{254\text{-}486}$ was expressed in *E. coli* BL21(DE3) and purified by affinity chromatography and size exclusion chromatography. While the protein was well-expressed and appeared to be stable, its chromatogram showed that it was prone to oligomerisation (Figure 6.3A). Oligomerisation is a common property of lectins and is proposed to enhance their glycan-binding valence (Laaf et al, 2019). The SEC peak fractions whose elution volume was consistent with the size of the His-PilC$_{254\text{-}486}$ monomer (Peak 2) were pooled together, concentrated to 100 μM and tested for binding to 3'-SLN by ITC. As above, His-PilC$_{254\text{-}486}$ was loaded in the sample cell, and the sugar ligand was titrated via 30 injections.

A weak binding event took place between His-PilC$_{254\text{-}486}$ and 3'-SLN (Figure 6.3B). The strength of the μcal/sec signal was significantly weaker than for His-PilC$_{121\text{-}486}$, and even after 30 titrations, the reaction did not appear to reach full saturation. The K_D was estimated to be 550 mM, but its accuracy is debatable due to the lack of saturation. The binding affinity of the PilC lectin domain alone was significantly worse than the binding

affinity of the PilC lectin domain in the presence of the Ig module. This could be due to the propensity of His-PilC$_{254\text{-}486}$ to oligomerise which could result in obstruction of its binding site or due to intrinsic instability. The Ig-like fold appeared to be important for the PilC-sugar interaction – it might prevent lectin oligomerisation, stabilising the monomeric lectin module, or it could play a supporting role in ligand binding.

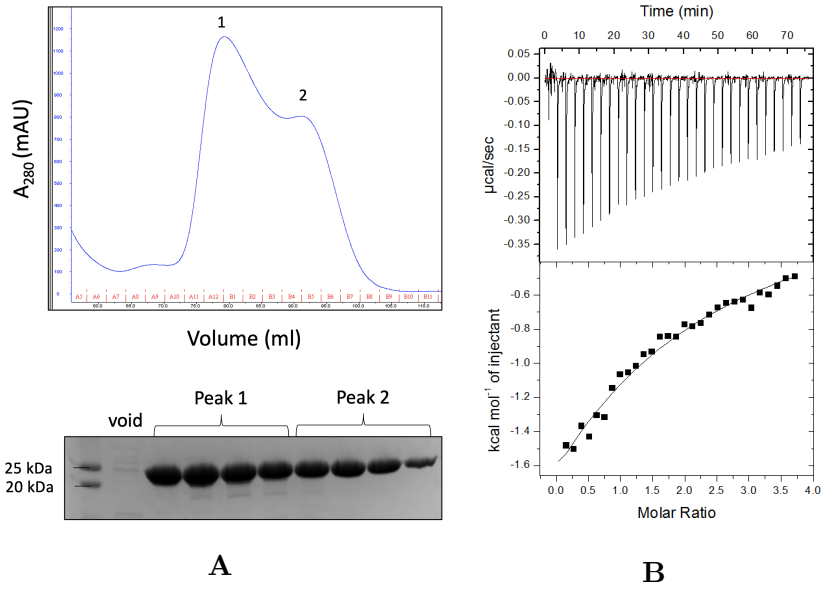


Figure 6.3: **Analysis of His-PilC$_{254\text{-}486}$** (A) SEC purification of His-PilC$_{254\text{-}486}$ (26 kDa) and Coomassie gel of the peak fractions reveal that His-PilC$_{254\text{-}486}$ forms oligomers. (B) ITC analysis of the interaction between His-PilC$_{254\text{-}486}$ and 3'-SLN.

It is also worth noting that the K$_D$s reported in this chapter may not faithfully represent the strength of the *in vivo* interaction between PilC and its ligands. Firstly, glycan ligands are often immobilised on the cell surface, while both protein and ligands were free in solution in the ITC experiments described here. Techniques, such as surface plasmon resonance, in which one reactant is immobilised, are reported to yield lower apparent K$_D$ values than those obtained by ITC and are thought to be closer in value to the real K$_D$ (Laaf et al, 2019). Secondly, the ITC experiments tested the affinity of PilC for chemically synthesised trisaccharides. It is likely that PilC establishes multiple contacts with its glycan ligands, beyond a single trisaccharide subunit as well as with the protein

receptors that host the glycan ligands. This would significantly contribute to both the affinity and the avidity of the host-pathogen interaction. Hence, it is possible that the affinity of the PilC-glycan interaction is much stronger than what was observed in this study.

6.3 Determining the Carbohydrate-Recognition Domain of PilC

6.3.1 Bioinfomatic Analysis of the PilC Lectin Module

Having confirmed the carbohydrate-binding activity of the PilC lectin module and established an assay for testing its interaction with oligosaccharides, we next wanted to determine its carbohydrate-recognition domain (CRD). This would allows us to perform structure/function analysis of the PilC minor pilin. However, the affinity of PilC for the 3'-SL and 3'-SLN ligands was too low to allow us to crystallise the protein-sugar complex, and the His-PilC$_{121\text{-}486}$ protein was too large (41.6 kDa) for the protein-sugar interactions to be studied by NMR. Furthermore, the lack of sequence homology among bacterial lectins made it difficult to determine the CRD based on conserved sequence motives, such as those found in eukaryotic lectins (Hirabayashi and Kasai, 1993).

Instead, we used structural homology to make an educated prediction about the location of the PilC CRD. According to the crystal structure of the His-PilC$_{121\text{-}486}$ protein, the lectin module is a β-sandwich characterised by two opposing anti-parallel β-sheets which consist of six strands each. Close inspection of the two β-sheets revealed that they are oriented in such a way as to create a concave side and a convex side (Figure 6.4A). The concave side could provide a binding site for the glycan ligands of PilC. Indeed, other lectins in the

PDB database which bear structural homology to the PilC lectin module display a similar concave surface formed by the β-sandwich, and both structural and mutagenesis studies have confirmed that these surfaces are involved in carbohydrate binding (Hirabayashi and Kasai, 1993; Henrick et al, 1998; Salomonsson et al, 2010).

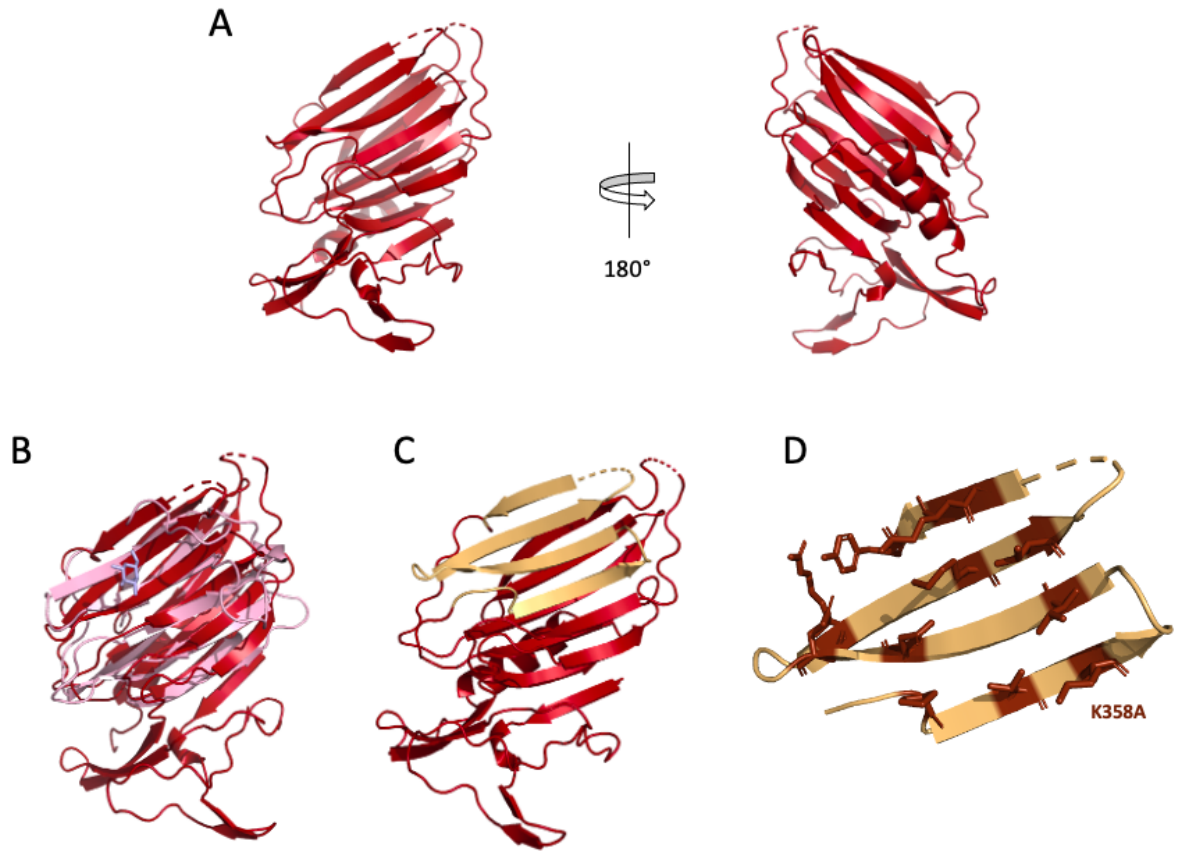


Figure 6.4: **Determining the CRD of the PilC lectin module.** (A) The concave (left) and convex (right) side of the lectin module. (B) Structural superposition of the *S. sanguinis* PilC lectin module and human Gal-7 (PDB: 2GAL). The lectin module is shown in red, Gal-7 is in light pink. The galactose sugar is shown in blue. (C) The predicted location of the PilC lectin CRD based on this structural analysis. (D) The PilC CRD residues targeted in mutagenesis. The mutated residues are shown in brown and K358 is specifically labelled.

Critical insight into the PilC CRD was gained by DALI structural analysis (Holm, 2020) which highlighted the human galectin Gal-7 (PDB: 2GAL) as a close structural homologue of the PilC lectin module (Figure 6.4B). The two proteins bear no sequence homology and have different ligand preferences (Leonidas et al, 1998), but their 3D structures aligned with a RMSD of 1.95 Å. This homology was greatly advantageous as the ligand-binding ability of galectins is well-characterised. In fact, the structure of Gal-7 was solved in

complex with galactose, revealing that its CRD resembles a pocket on the concave side of the protein (Leonidas et al, 1998). The good superposition between this region in the PilC lectin module and Gal-7 crystal structure led us to speculate that the PilC lectin CRD was located between residues 352 and 392 (Figure 6.4C).

6.3.2 Mutational Studies of the Predicted CRD

Narrowing the CRD to a small region of the PilC lectin module allowed us to perform mutagenesis studies in order to identify residues that are important for the PilC-glycan interaction. We generated nine individual mutations in the predicted PilC lectin CRD. The mutations targeted polar and charged amino acids, located between residues 352 and 392, whose side chains were externally oriented and capable of engaging in hydrogen bonds with glycan ligands (Figure 6.4D). The pET28b plasmid encoding His-PilC$_{121\text{-}486}$ was used as a template for the site-directed mutagenesis. The SDM products were incubated with DpnI restriction enzyme for one hour at 37°C and transformed into *E. coli* DH5α cells. The pET28b plasmids were then isolated from individual colonies grown on LB-kanamycin agar and sequenced to check that the mutation was introduced without any incidental errors.

Next, the His-PilC$_{121\text{-}486}$ mutant plasmids were transformed into BL21 (DE3) to test protein expression and solubility. All nine mutant proteins were well-expressed and stable. They were easily purified by affinity chromatography and size exclusion chromatography, providing similar yields to the WT His-PilC$_{121\text{-}486}$ construct. The affinity of each mutant for both 3'-SL and 3'-SLN oligosaccharides was then tested by ITC in order to determine whether any of the targeted residues are directly involved in the PilC-glycan interaction.

PilC$_{121\text{-}486}$	**3'-SL**	**3'-SLN**
WT	79.2 ± 0.4 μM	144 ± 7.4 μM
His-PilC$_{T356A}$	98.9 ± 10 μM	135 ± 11 μM
His-PilC$_{K358A}$	201.4 ± 7.6 μM	290 ± 22 μM
His-PilC$_{T364A}$	76.9 ± 6.9 μM	147 ± 8.1 μM
His-PilC$_{N368A}$	92.9 ± 6.6 μM	120 ± 4.3 μM
His-PilC$_{E370A}$	68.2 ± 0.7 μM	172 ± 4.1 μM
His-PilC$_{R374A}$	68.5 ± 0.4 μM	142 ± 7.3 μM
His-PilC$_{G389S}$	69.8 ± 3.5 μM	153 ± 8.0 μM
His-PilC$_{Q390A}$	92.8 ± 8.1 μM	140 ± 29 μM
His-PilC$_{Y392A}$	75.9 ± 3.8 μM	181 ± 14 μM

Table 6.1: **Binding affinities of the PilC mutants for 3'-SL and 3'SLN.**

As shown in Table 6.1, the majority of the mutations had no significant impact on the interaction between PilC and its oligosaccharide ligands. Only the His-PilC$_{K358A}$ displayed reduced binding affinity – it showed 39.3 ± 1.5% of the binding of WT PilC for 3'-SL and 49.8 ± 3.6 % of the binding of WT PilC for 3'-SLN. This finding validated our prediction for the location of the PilC CRD and highlighted a key residue for the PilC-glycan interaction.

6.4 Determining the Role of PilC in *S. sanguinis* T4P Functionality

As PilC is essential for T4P biogenesis, its deletion mutant Δ*pilC* is non-piliated. This has hindered the investigation of the role PilC plays in T4P functionality. To determine its contribution to host-pathogen interactions, we required a *pilC* mutant which could support filament assembly, but was defective for glycan binding. The K358A mutation decreased the affinity of PilC for its ligands almost three-fold. It was located in the lectin module of PilC and did not impact the stability of the recombinant protein, strongly suggesting that it would not interfere with PilC polymerisation into T4P. We therefore

re-produced the K358A mutation in *S. sanguinis* and tested its impact on *S. sanguinis* T4P functionality.

6.4.1 Generating the K358A mutation in *S. sanguinis*

The generation of a non-polar markerless *S. sanguinis pilC* mutant was a two-step process. Firstly, we used a PCR8/GW/TOPO vector containing the WT *S. sanguinis pilC* gene, alongside its upstream (1 kb) and downstream regions (0.69 kb), as a template for site-directed mutagenesis. Following SDM PCR, the plasmids were treated with DpnI and transformed into *E. coli* DH5α selectively grown on streptomycin-LB agar. Once the introduction of the desired point mutation was verified by sequencing, the $pilC_{K358A}$ gene was amplified from its TOPO plasmid and transformed into the *S. sanguinis pilC:pheS*-aphA3* primary mutant. The primary mutant's double cassette was replaced with the $pilC_{K358A}$ gene by allelic exchange. The successfully transformed bacteria were identified by their resistance to PCPA and sensitivity to kanamycin. The presence of the *pilC* gene was then confirmed by PCR, and the isolated *pilC* gene was sent for sequencing to check for the K358A mutation.

Next, the *S. sanguinis* $pilC_{K358A}$ mutant underwent phenotypical analysis to assert 1) that the mutation did not interfere with *S. sanguinis* piliation, 2) that the mutant PilC protein was stable and not prone to proteolysis or degradation, and 3) whether the mutant PilC protein supported T4P functionality in the form of twitching motility (Figure 6.5). The prepilin processing of $pilC_{K358A}$ was not tested because the mutation is located outside of the class III signal peptide and thus cannot impact leader peptide cleavage. Pili purification showed that $PilC_{K358A}$ supported piliation. The stability of the $PilC_{K358A}$ protein was assessed by Western blotting of whole-cell protein extracts using the WT *S.*

sanguinis strain as a control. As predicted, the full-length mutant protein was detected in the whole-cell extracts in amounts comparable to the WT. The K358A mutation did not appear to interfere with T4P-mediated twitching motility either. This confirmed the successful generation of a piliated *S. sanguinis* mutant which could be used to test the role of PilC-mediated glycan binding in *S. sanguinis* adhesion.

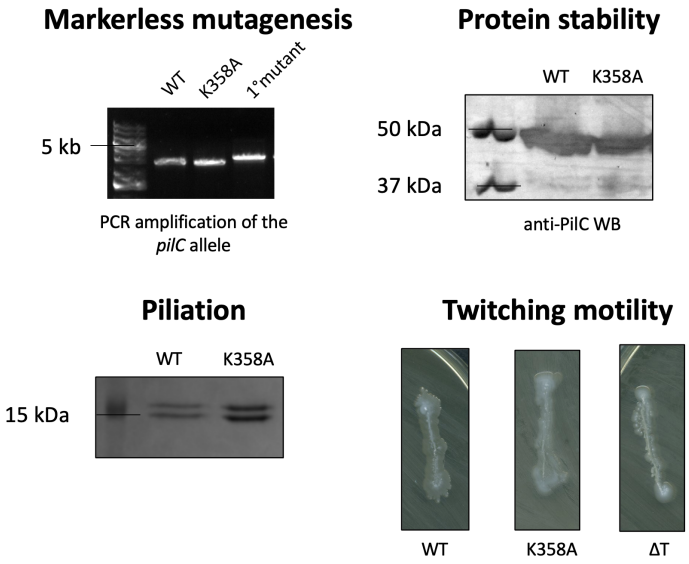


Figure 6.5: **Phenotypical analysis of *S. sanguinis pilC*$_{K358A}$ mutant.** Markerless mutagenesis was performed as described in Chapter 2. In a PCR reaction, gDNA was amplified using specific *pilC* primers. The WT and primary *pilC:pheS*-aphA3* mutant were included as controls. The PCR products were then analysed by agarose gel electrophoresis. The stability of the protein was assessed by its quantity in whole-cell extracts compared to WT. Whole-cell extracts were obtained by homogenisation of 100 ml THTH cultures of WT and mutant bacteria and analysed by Western blotting using an anti-PilC antibody. Piliation of the mutant was determined by SDS-PAGE analysis of the sheared and purified pili. Twitching motility was assessed by re-streaking *S. sanguinis* colonies in 2 cm-long lines on TH 1% Eiken agar plates. The WT and non-retractile Δ*pilT* mutant were included as controls.

6.4.2 Assessing *S. sanguinis* Adhesion to Eukaryotic Cells

The ability of the PilC lectin module to bind to a range of glycans which are highly prevalent in the human glycome suggests that the minor pilin PilC could mediate interactions between *S. sanguinis* and glycosylated host proteins and/or glycosylated host cell receptors. To test this hypothesis, we compared the ability of WT *S. sanguinis*, the *S. sanguinis pilC*$_{K358A}$ mutant and the non-piliated Δ*pilC* mutant to adhere to monolayers

of cultured eukaryotic cells.

We first tested the binding of *S. sanguinis* to Chinese Hamster Ovary (CHO) cells. CHO cells have been reported to produce heparan sulfate and chondroitin-sulfate proteoglycans and α2,3-sialylated terminal residues (Patnaik et al, 2006), matching the glycan specificity of the minor pilin PilC. CHO cells were seeded at a density of 10,000 cells/cm^2 in a 24-well plate and infected with 1 ml of *S. sanguinis* bacteria at MOI of 10. Following an hour-long incubation, the cell monolayers were gently washed and the adherent bacteria were collected. The degree of adhesion was quantified by CFU counts and calculated as percentage of the bacteria in the inoculum that have adhered to the eukaryotic cell monolayer (Figure 6.6A). For the WT *S. sanguinis*, ∼79% of the bacteria in the inoculum adhered to the CHO cells. The non-piliated Δ*pilC* mutant suffered a 10-fold reduction in adhesion compared to WT. This was consistent with previously published data (Raynaud et al, 2021) and re-affirmed the central role of T4P in host-pathogen interactions. The *S. sanguinis* $pilC_{K358A}$ mutant, however, displayed WT-like levels of adhesion, with ∼80% of bacteria in the inoculum adhering to the monolayers.

Similar results were obtained for the binding of *S. sanguinis* to HeLa (Henrietta Lacks) cells (Figure 6.6B). WT *S. sanguinis* demonstrated ∼70% adhesion to the cell monolayers relative to the inoculum. The non-piliated Δ*pilC* mutant suffered a three-fold decrease in adhesion compared to the WT. This decrease was still statistically significant but suggested that the pili are less important for the interaction of *S. sanguinis* with HeLa cells than with CHO cells. As above, the *S. sanguinis* $pilC_{K358A}$ mutant displayed WT-like levels of adhesion, with ∼72% of bacteria in the inoculum adhering to the HeLa monolayers. Thus, even though the K358A mutation significantly decreased the affinity of PilC for its sugar ligands *in vitro*, it did not affect *in vivo* PilC interactions.

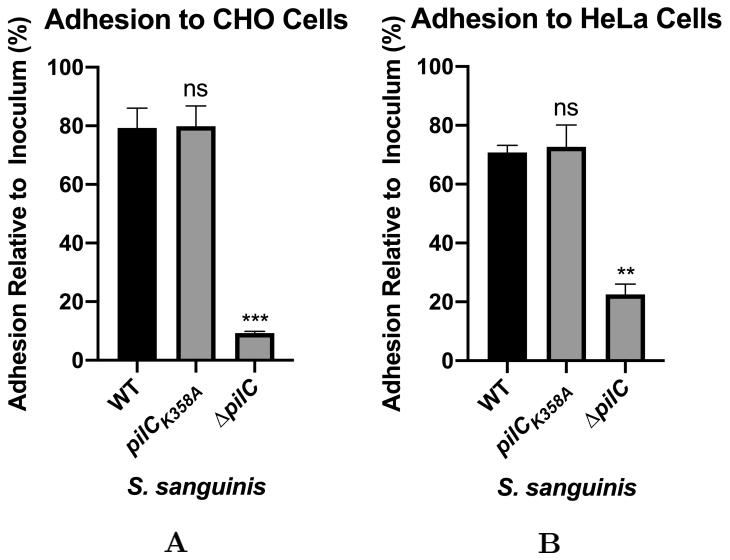


Figure 6.6: **Cell adhesion of *S. sanguinis* to eukaryotic cells.** (A) Adhesion to CHO cells. (B) Adhesion to HeLa cells. Non-adherent bacterial cells were removed by five consecutive washing steps, and the adherent bacteria were plated on TH agar. Results are expressed as the percentage of adhered bacteria relative to the bacteria in the inoculum. The statistical analysis included a one-way ANOVA test, followed by Dunnett's multiple comparison test (ns = non-significant, *** $p < 0.001$).

6.4.3 Further Mutagenesis

Since we were unable to disrupt *S. sanguinis* adhesion to eukaryotic cells with a single mutation, we performed more extensive mutagenesis of the PilC CRD (Figure 6.7). Two further mutants were generated – one carried seven amino acid substitutions in the predicted sugar binding site ($PilC_{CRD7}$), while the other carried 12 ($PilC_{CRD12}$). Because of the low GC content of the CRD region of the *pilC* gene and the close proximity of the targeted CRD residues, the mutants were obtained by gene synthesis instead of site-directed mutagenesis. One pair of genes was designed to encode the codon-optimised His-$PilC_{121\text{-}486}$ construct bearing the CRD7 or CRD12 mutations. These genes were subcloned into pET28b expression vectors and used to assess the stability of the mutant proteins and their binding affinity for the PilC ligands. The other set of synthetic genes encoded the CRD region of the *S. sanguinis pilC* gene harbouring either the seven or 12 mutations. The mutant CRD regions were introduced into the *S. sanguinis* genome, and the mutant strains were phenotypically analysed.

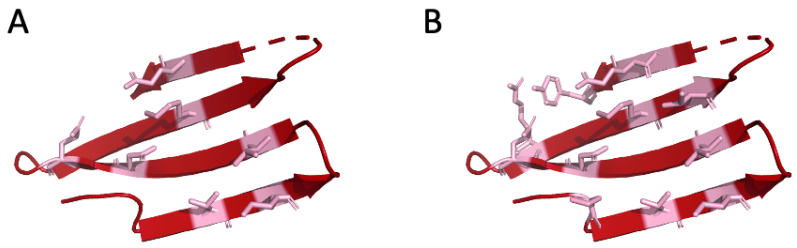


Figure 6.7: **Further mutagenesis of the PilC CRD.** (A) The residues mutated in $PilC_{CRD7}$. (B) The residues mutated in the $PilC_{CRD12}$. Both mutants include the K358A mutation.

In the codon-optimised synthetic genes, the mutant constructs were flanked by a 5' NcoI restriction site and a 3' BamHI restriction site. This facilitated their excision from the synthetic gene plasmid and subsequent sub-cloning into the pET28b vector. In brief, the synthetic gene plasmids and the pET28b vector were digested with the NcoI and BamHI restriction enzymes. The digestion reaction products were separated by agarose gel electrophoresis, extracted from the gel and purified. The digested mutant constructs and the linearised pET28b vector were then ligated and transformed into *E. coli* DH5α bacteria. pET28b-His-$PilC_{CRD7}$ and pET28b-His-$PilC_{CRD12}$ were isolated from single DH5α colonies and transformed into *E. coli* BL21-DE3 for recombinant protein expression.

Expression of the two polymutant proteins was tested as described previously (Section 2.6). A single colony of the transformed *E. coli* cells was used to inoculate a 10 ml kanamycin-supplemented LB culture overnight. The next day, the cultures were back-diluted into 1 l of kanamycin-supplemented LB and grown at 37°C. Once the OD_{600} reached 0.8 units, protein expression was induced with 0.5 mM IPTG, and the cultures either remained at 37°C or were transferred to 16°C. Samples from each culture were taken after four or 18 hours and tested for protein expression, solubility and purification yields.

Both His-PilC$_{CRD7}$ and His-PilC$_{CRD12}$ were strongly expressed in all tested conditions. Following cell harvesting, however, the majority of His-PilC$_{CRD12}$ was detected in the insoluble fraction of the cell lysate, and very little protein could be purified by affinity chromatography (Figure 6.8). This indicated that the PilC$_{CRD12}$ mutant was unfolded and accumulated in inclusion bodies. While PilC$_{CRD7}$ was also insoluble when expressed at 37°C, about half of the protein was found to locate in the soluble fraction of the cell lysate when expressed at 16°C overnight. The soluble PilC$_{CRD7}$ fraction was successfully purified by affinity chromatography and SEC. The protein appeared to be stable and generated sufficient yields for further analysis.

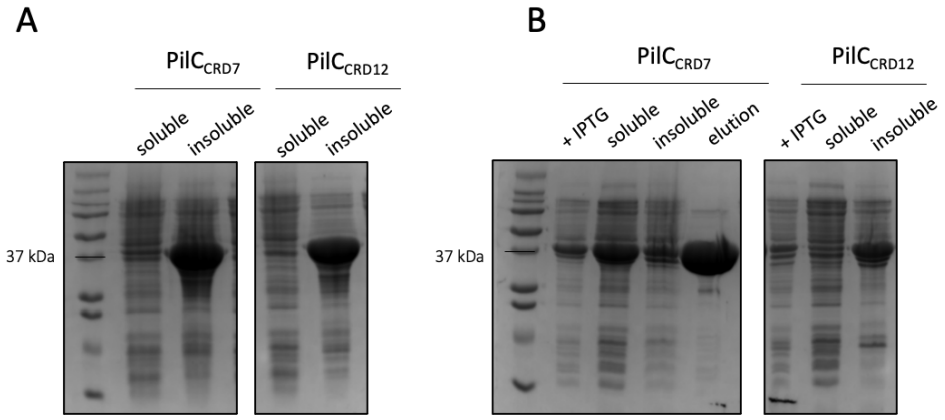


Figure 6.8: **Protein expression of His-PilC$_{CRD7}$ and His-PilC$_{CRD12}$.** (A) Protein expression at 37°C. (B) Protein expression at 16°C. The Coomassie-stained gels show the protein expression, the soluble and insoluble fractions of the cell lysates and the first elution fraction of the affinity chromatography.

The binding affinity of the PilC$_{CRD7}$ mutant for the trisaccharide ligands was tested by ITC. PilC$_{CRD7}$ showed reduced binding affinity compared to WT PilC. It exhibited 31.3 ± 1.1% of the binding of WT for 3'-SL and 38.4 ± 1.7% for 3'-SLN (Figure 6.9). Compared to the previously tested PilC$_{K358A}$ mutant, PilC$_{CRD7}$ showed a small decrease in ligand binding affinity. The interaction between both sugars and PilC$_{CRD7}$ was ∼1.3-fold weaker than the interaction with the previously tested PilC$_{K358A}$. This further substantiated our prediction for the sugar-binding site of the PilC lectin module and provided us with

another mutant that could have an impact on *in vivo* PilC-mediated interactions.

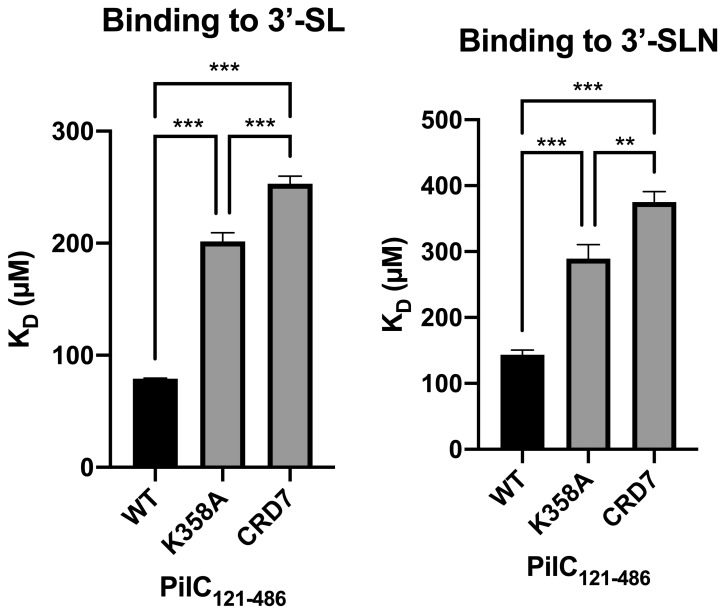


Figure 6.9: **Binding affinity of the PilC mutants to trisaccharides.** The K_Ds were obtained by ITC experiments performed at least three times. The sugar ligands were loaded in the syringe, and the $PilC_{121\text{-}486}$ protein in the cell as described above. The statistical analysis included a one-way ANOVA test, followed by Tukey's multiple comparison test (** $p = 0.02$, *** $p < 0.001$).

The CRD7 mutation was therefore re-produced in *S. sanguinis*. The purchased synthetic gene encoded only the mutant CRD region of the *S. sanguinis pilC* gene. This region was amplified from the synthetic plasmid and joined together with the rest of the *pilC* gene in splicing PCR (sPCR). As described in Section 2.3.1, sPCR consists of two steps. In the primary PCR, the *pilC* sequence 5' to the CRD (including the *pilC* upstream region) and the sequence 3' to the CRD (including the *pilC* downstream region) were amplified from the TOPO plasmid containing the WT *S. sanguinis pilC* gene using a high-fidelity DNA polymerase. The primers in the primary PCR were designed to create a 40-nucleotide-long region of complementarity between the 5' fragment and the CRD, and the CRD and the 3' fragment. This allowed the three PCR products from the primary PCR to be spliced together in the secondary PCR. The splicing was verified by agarose gel electrophoresis and the sPCR product was transformed directly into *S. sanguinis pilC:pheS*-aphA3* primary

mutant. Successfully transformed bacteria were once again identified by their resistance to PCPA and their sensitivity to kanamycin. The substitution of the double cassette with the $pilC_{CRD7}$ gene was then confirmed by PCR.

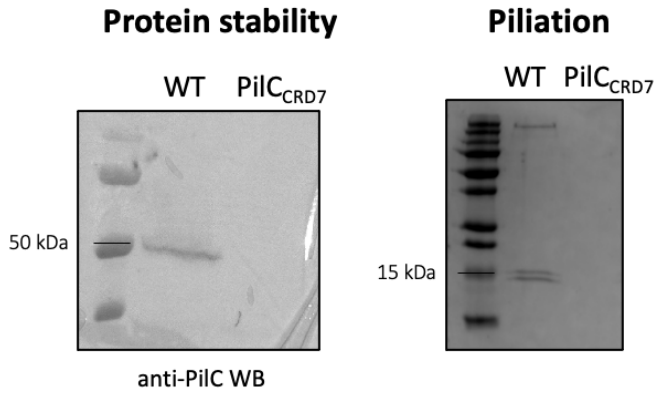


Figure 6.10: **Phenotypical analysis of the *S. sanguinis* $pilC_{CRD7}$ mutant.** Protein stability was assessed by Western blotting whole-cell extracts using an anti-PilC antibody. Piliation was determined by SDS-PAGE analysis of the sheared and purified pili. The WT *S. sanguinis* strain was included as a positive control in both experiments.

The next step was to characterise the phenotype of the $pilC_{CRD7}$ mutant. Because of the insolubility issues observed for the recombinant His-PilC$_{\mathrm{CRD7}}$ protein, the *S. sanguinis* $pilC_{CRD7}$ mutant was first tested for PilC protein stability. Whole-cell protein extracts were obtained from 100 ml of bacterial cultures and analysed by Western blotting to determine whether the introduced mutations affected the production of the PilC pilin in *S. sanguinis*. The WT strain was included as a positive control. No PilC protein was detected in the whole-cell extracts of the $pilC_{CRD7}$ mutant (Figure 6.10). The polymutant appeared to be unstable and vulnerable to degradation and/or proteolysis. Incidentally, the same phenotype was observed for the unmarked *S. sanguinis* $pilC_{CRD12}$ mutant (data not shown). Both observations were consistent with the insolubility of the recombinant polymutant proteins when expressed at 37°C.

Since PilC is essential for piliation, its absence in the $pilC_{CRD7}$ mutant was likely to have a severe impact on the production of T4P. Indeed, no major pilins were detected in the

sheared pili fraction of the *S. sanguinis* $pilC_{CRD7}$ mutant (Figure 6.10). This means that the $pilC_{CRD7}$ bacteria were double-defective: defective for PilC-mediated glycan binding and defective for piliation. So even though the $pilC_{CRD7}$ mutant exhibited only 28% of the adhesion of WT *S. sanguinis* to CHO cells (data not shown), the reduction in eukaryotic cell adhesion was most likely caused by the loss of piliation.

6.4.4 Competition Assays

As mutagenesis proved unsuitable for determining the role of PilC in *S. sanguinis* T4P functionality, we attempted to determine the overall contribution of the PilC lectin module to the binding of *S. sanguinis* to eukaryotic cells by performing competition assays. These experiments assessed the ability of the recombinant His-$PilC_{121\text{-}486}$ protein to disrupt WT *S. sanguinis* adhesion to CHO cells. The cell adhesion assays were performed as described above with the exception that prior to infection, the CHO monolayers were washed three times with PBS and incubated with 500 μg/ml of the purified protein for 30 minutes at 37°C. The purified protein was then gently removed and the cells were immediately infected with 1 ml of bacteria at MOI of 10. These assays were performed in triplicate and repeated three independent times.

In the absence of competition, the WT strain demonstrated a consistent efficiency of adhesion – ∼75% of the bacteria in the inoculum adhered to the CHO cells (Figure 6.11C). When the cells were incubated with His-$PilC_{121\text{-}486}$ protein prior to infection, the adhesion decreased to ∼56%. In relative terms, competition with the lectin module reduced WT *S. sanguinis* adhesion by a quarter. The effect of the recombinant protein on cell adhesion was statistically significant but not as pronounced as what has been reported for other minor pilins involved in adhesion (Barnier et al, 2021). This suggests an overall modest

contribution of the PilC lectin domain to the interaction between *S. sanguinis* and CHO cells.

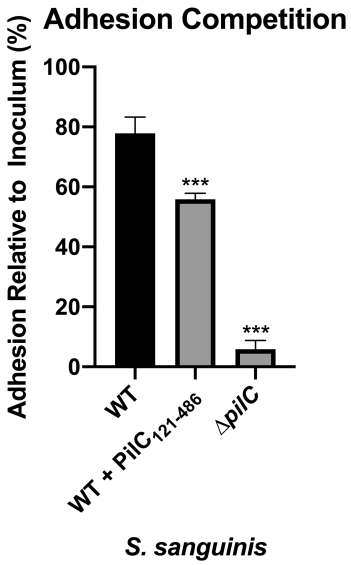


Figure 6.11: **Cell adhesion competition assays.** The cell adhesion competition assays were performed on CHO cells. The CHO cell monolayers were washed three times with PBS and incubated with 500 μg/ml purified recombinant His-PilC$_{121\text{-}486}$ for 30 minutes at 37°C prior to infection. The recombinant protein was removed and 1 ml of bacteria at MOI of 10 was added. The rest of the cell adhesion assay was then conducted as described above. Results are expressed as adhesion relative to the inoculum in percentages. The statistical analysis included a one-way ANOVA test, followed by Dunnett's multiple comparison test (*** $p < 0.001$).

6.5 Summary

The *S. sanguinis* minor pilin PilC is a trimodular protein, displaying an N-terminal pilin domain, followed by an Ig fold domain and a C-terminal lectin module. The lectin module was predicted to belong to the concanavalin A-like lectin/glucanase domain family (InterPro entry IPR013320) (Berry et al, 2019), and previous work in the lab has shown that it acts as a functional adhesin. Taken together with the findings from the previous chapter that PilC is presented at the tip of the T4P filaments in a complex with minor pilin PilA, this strongly suggests that PilC might be involved in host-pathogen interactions by binding to glycosylated host proteins and/or cell receptors. In this chapter, we further characterised the carbohydrate-binding function of PilC and attempted to determine its

role in *S. sanguinis* pathogenesis.

Two trisaccharides (3'-SL and 3'-SLN) were identified as ligands of PilC in glycan microarrays (Dr Raynaud, unpublished data). The interaction between the minor pilin and these two sugars was observed in real-time by ITC experiments. PilC demonstrated a binding affinity of 79.2 μM for 3'-SL and 144 μM for 3'-SLN. The binding occurred independently of the PilC pilin domain and Ca^{2+} or Mn^{2+} ions. This distinguishes PilC from other members of the concanavalin A-like lectin family in which carbohydrate binding is metal-ion dependent. As predicted, the interaction between PilC and its sugar ligands occurred via the lectin module, but the Ig fold domain was required for optimum affinity. The Ig fold domain appeared to contribute towards the stability and the monomeric state of the lectin module. These observations reinforce the role of the minor pilin PilC as a T4P adhesin.

The carbohydrate-binding site of the PilC adhesin was predicted by structural comparison between PilC and its close homologue Gal-7. The CRD was speculated to reside on the concave side of the lectin module and to extend between residues 352-392. The location of the CRD was verified by extensive mutagenesis, and a residue involved in the PilC-trisaccharide interaction was identified. The K358A mutation caused a nearly 2.5-fold decrease in the binding affinity of PilC for both 3'-SL and 3'-SLN, while the CRD7 mutations reduced the affinity nearly 3-fold.

As lectins in pathogenic organisms are often involved in host recognition and tissue adhesion, PilC was also likely to play a key role in T4P functionality. Moreover, PilC was found to interact with a range of glycans that are widespread in the human glycome, suggesting that the PilC-sugar interaction might be important for host interactions during *S. sanguinis* pathogenesis. Since the lab had an established protocol for cell adhesion assays,

we tested the contribution of PilC to bacterial binding to eukaryotic cells. Decreases in the binding affinity of PilC for its ligands *in vitro* did not lead to a corresponding decrease in *S. sanguinis* adhesion to either CHO or HeLa cells *in vivo* as evidenced by the $PilC_{K358A}$ mutation. Similarly, purified PilC was found to interfere modestly with *S. sanguinis* binding to CHO cells in competition assays. In contrast, the *S. sanguinis* minor pilin PilB was recently found to be indispensable for cell adhesion, with a PilB MIDAS mutant showing a 33-fold decrease in adhesion to CHO cells compared to the WT (Raynaud et al, 2021). From the data presented in this study we cannot determine the role PilC plays in the interaction between *S. sanguinis* T4P and eukaryotic cells.

Chapter 7

Discussion and Future Perspectives

The ubiquity and functional versatility of type IV pili make them an interesting and pressing topic of research. Found in nearly all phyla of the prokaryotic kingdom, T4P facilitate key bacterial functions, such as motility, natural competence and adherence, and are essential for the ability of several prominent human pathogens to colonise their hosts and cause diseases (Berry and Pelicic, 2015). Over the last three decades, in-depth structure/function studies have provided key insights into the mechanisms of T4P biogenesis and functionality. However, there remain important gaps in our understanding of T4P biology. One such gap is the role minor pilins play in T4P assembly and T4P-associated functions.

The high number and great diversity of minor pilins found in Gram-negative T4P systems have made their structural and functional characterisation challenging. The Gram-positive opportunistic pathogen *S. sanguinis* provides a simpler context for studying the role of minor pilins in T4P biology. Due to the absence of an outer membrane, the T4P machinery of *S. sanguinis* is significantly less complex than the machinery of Gram-negative bacteria: *S. sanguinis* requires only 10 proteins for T4P biogenesis, compared to the 15 proteins essential for T4P biogenesis in *Neisseria* (Carbonnelle et al, 2005; Gurung et al, 2016). More importantly, *S. sanguinis* encodes only three minor pilins – PilA, PilB and PilC – whose deletion leads to loss of piliation, in contrast to the multiple core and non-core minor pilins identified in *Neisseria* (Carbonnelle et al, 2005; Gurung et al, 2016).

The simplicity of the *S. sanguinis* T4P machinery and its genetic tractability provide us with an unprecedented opportunity to define the role of each pilus subunit in a single T4P system. Since the two *S. sanguinis* major pilins PilE1 and PilE2 and the minor pilin PilB have already been described by our lab (Berry et al, 2019; Raynaud et al, 2021), this thesis focused on the two remaining minor pilins – PilA and PilC. This chapter positions

our findings within the field of T4P research.

7.1 Functional Analysis of Minor Pilin PilA

Type IV pilins are distinguished by the highly distinctive N-terminal class III signal peptide. The class III signal peptide facilitates the incorporation of prepilins into the cytoplasmic membrane and their processing into mature pilins which lack a cytoplasmic domain (Dupuy et al, 1991; Strom and Lorry, 1992). The canonical class III signal peptide is made up of a leader peptide, containing 6-26 predominantly hydrophilic and neutral amino acids, followed by 21 predominantly hydrophobic amino acids. Mutational analysis of major pilins has demonstrated the importance of the class III signal peptide for pilin and pili functionality. Even small variations at key N-terminal residues can have major consequences for T4P biogenesis (Giltner et al, 2012). In particular, the glycine at position -1 is critical for leader peptide cleavage (Strom and Lory, 1991), and the glutamate at position 5 is essential for pilin polymerisation (Aas et al, 2007).

This made the *S. sanguinis* minor pilin PilA an interesting case study in the contribution of class III signal peptides to minor pilin functionality. Despite its highly divergent class III signal peptide, PilA undergoes leader peptide cleavage, polymerises into T4P filaments and is essential for T4P biogenesis (Gurung et al, 2016; Berry et al, 2019). Our systematic mutational analysis further revealed that the α1N helix of the mature PilA protein is highly accommodating of residue substitutions. Mutagenesis of the first eight processed pilin residues, which are conserved in all *S. sanguinis* strains sequenced to date, had no impact on PilA stability, T4P assembly or T4P-associated functions. This suggests that a canonical class III signal peptide sequence is not required for the functionality of this minor pilin. Indeed, while extensive sequence homology has been noted for the first

25 residues of the T4a pilins (Craig and Li, 2008), T4b pilins, T4c pilins and archaeal pilins exhibit a much greater diversity in their α1N sequences (Craig et al, 2004). We can therefore conclude that the unusual class III signal peptide of PilA bears no specific significance for its role in T4P biogenesis. In other words, it remains to be established why PilA evolved an N-terminal motif that is divergent not only from the other four pilins in *S. sanguinis*, but also from pilins in other T4F systems.

The glycine at position -1 was the only class III signal peptide residue whose mutation impacted PilA functionality. The role of Gly-1 in prepilin processing was first reported for the *P. aeruginosa* major pilin – mutating the glycine to alanine led to partial prepilin processing, while a serine substitution completely abolished leader peptide cleavage and resulted in the loss of bacterial piliation (Strom and Lorry, 1991). Since then, G-1S mutations have been shown to successfully abolish major pilin processing and its subsequent assembly into filaments in the Gram-positive *S. sanguinis* and the Gram-negative *N. meningitidis* and *N. gonorrhoaea* (Berry and Pelicic, 2015; Berry et al, 2019). However, the impact of Gly-1 disruptions on minor pilins has not been studied as extensively.

In our functional analysis of the minor pilin PilA class III signal peptide, we generated unmarked *S. sanguinis* mutants carrying a G-1S, G-1R or G-1W substitution in the *pilA* gene. While such mutations in the major pilins have been consistently reported to hinder T4P biogenesis, in the minor pilin PilA they had no impact on bacterial piliation. In the $pilA_{G\text{-}1S}$ strain, PilA underwent residual prepilin processing which could be detected by Western blotting. The residually processed PilA protein was incorporated into the T4P and supported WT-levels of piliation, but the pili were incapable of twitching motility. Similarly, the $pilA_{G\text{-}1R}$ and $pilA_{G\text{-}1W}$ strains presented non-functional pili on their surface. In these two strains, however, the levels of mature PilA protein were reduced to the

point of no detection by Western blotting, in both the whole-cell protein extracts and the sheared pili fractions. Since T4P biogenesis does not occur in the absence of PilA, and only processed pilins can polymerise into filaments, the most likely explanation for the $pilA_{G\text{-}1R}$ and $pilA_{G\text{-}1W}$ phenotypes is that very low levels of mature PilA are still available in those strains to support piliation. Having ruled out the possibility of glycerol stock contamination and translational misincorporation of glycine at position -1, we postulate that the low levels of mature PilA protein are the result of PilD-mediated cleavage of the mutant leader peptides.

These findings lead to three interesting conclusions. Firstly, glycine substitutions with residues carrying a polar and/or bulky side group do not completely abolish prepilin processing. This property does not appear to be exclusive either to PilA – $pilB_{G\text{-}1S}$ and $pilB_{G\text{-}1W}$ mutants displayed the same phenotypes as the respective *pilA* mutants – or to *S. sanguinis*. In fact, similar observations have been made in *N. gonorrhoeae*. Mutants carrying a G-1S mutation in the non-core minor pilin ComP appeared to retain partial transformability and DNA uptake proficiency. The authors speculate that the high sensitivity of the transformation assays allowed them to uncover the residual activity of PilD for the missense mutant minor pilin (Wolfgang et al, 1999). While they could not detect ComP in the whole-cell protein extracts due to difficulties generating anti-ComP antibodies, the phenotypical analysis of $pilA_{G\text{-}1S}$ supports their hypothesis. It is already accepted that PilD can process prepilins from non-related species and different T4F systems (Berry and Pelicic, 2015). Our study has further shown that PilD is capable of accommodating different residues at position -1 although its enzymatic efficiency drastically decreases with increases in the size of the amino acid side chain at position -1. Indeed, substituting glycine with another small amino acid, such as alanine, has no impact on prepilin pro-

cessing; but a serine residue leads to a significant reduction in the amount of detectable mature pilin, and bulkier residues such as arginine and tryptophan hinder detection of any cleaved protein.

This raises the question whether major pilins carrying G-1 mutations are also susceptible to residual processing. Major pilins are expressed in much higher abundance than the minor pilins. Therefore, the trace amount of mature major pilin protein generated by residual processing could be easily obscured in Western blots of whole-cell protein extracts by the strong intensity of the band representing the non-processed pilin. Indeed, the residual processing of $PilA_{G\text{-}1S}$ was nearly overlooked in this study due to the tendency of strong bands to be overexposed and faint bands to fade into the background in chemiluminescent Western blot detection. While low quantities of processed minor pilins can support T4P assembly, the G-1 mutations in major pilins have catastrophic consequences for T4P biogenesis. Since the availability of major pilin in the cell dictates the number of filaments expressed per cell (Long et al, 2001), it is likely that residual major pilin processing is insufficient to enable polymerisation of pili that can be easily detected by purification or by functionality assays. This highlights the importance of expanding the study of key T4P biogenesis principles beyond the major pilin.

Secondly, while low levels of processed *S. sanguinis* pilins appear to be sufficient to support T4P assembly and cell surface expression, WT-like levels are required for T4P functionality, or at least T4P-mediated twitching motility. This is in disagreement with the observations of the Wolfang et al, 1999 report, where presumably low levels of the processed $ComP_{G\text{-}1S}$ mutant were sufficient to promote residual ComP-mediated T4P functions such as transformability and DNA uptake proficiency. It would be interesting to test the effect of the PilA, PilB and PilC G-1 substitutions on a different form of T4P functionality, such

as the ability of *S. sanguinis* to adhere to eukaryotic cells. It would also be interesting to interrogate the ability of a drastically decreased population of processed minor pilins to support WT levels of major pilin incorporation into pili. If, as our data indicate, the minor pilins are located at the tip of the T4P filaments, how can the sheared pili of the Gly-1 mutants contain the same amount of major pilins as the WT but significantly reduced levels of minor pilins? One possible explanation is that the processed major pilins available in the cytoplasmic membrane are incorporated into fewer but longer fibres. This hypothesis could be tested by electron microscopy studies of the surface piliation of the Gly-1 mutants and could shed light on the mechanism by which the length and quantity of assembled pili are controlled in *S. sanguinis*, as well as the relationship between T4P morphology and functionality.

Thirdly, the *S. sanguinis* $pilA_{G\text{-}1S}$, $pilA_{G\text{-}1R}$ and $pilA_{G\text{-}1W}$ mutants revealed an interdependence between the three minor pilins PilA, PilB and PilC with regards to incorporation into the pili. The decreased polymerisation of minor pilin PilA in these three strains led to corresponding reductions in the amount of PilB and PilC detected in the pili without impacting overall levels of *S. sanguinis* piliation. Similarly, the purified pili fractions of the $pilB_{G\text{-}1S}$ and $pilB_{G\text{-}1W}$ strains contained dramatically less PilA and PilC than WT, while the $pilC_{G\text{-}1S}$ filaments had reduced levels of PilA and PilB.

Interdependence between minor pilins that are essential for piliation has been reported in both *Neisseria* and *P. aeruginosa*. The core minor pilins in these two species are dependent on each other and on a T4P-associated adhesin PilC/PilY1 for pili polymerisation (Winther-Larsen et al, 2005; Nguyen et al, 2015a). For example, in *P. aeruginosa*, PilV, PilW and PilX require each other for incorporation into filaments, and PilY1 is only detected in the sheared pili fractions when PilV, PilW and PilX are all also present (Nguyen

et al, 2015a). The minor pilins and the adhesin are proposed to form a multimeric complex which supports T4P assembly (Carbonnelle et al, 2006; Giltner et al, 2010). Altogether, this suggests that a high-order interaction might also take place between the *S. sanguinis* minor pilins.

7.2 Minor Pilin Complex Formation in *S. sanguinis*

A key indication that the *S. sanguinis* minor pilins form a multimeric complex came from the close structural homology between PilA and the *E. coli* T2SS minor pseudopilin GspI. The homology between the two proteins could be revealed only by structural biology since they share no sequence identity. GspI interacts with minor pseudopilins GspJ and GspK in a heterotrimeric complex at the tip of the pseudopilus. Crystallography studies have shown that GspI is situated at the base of the complex, while the modular GspK pseudopilin resides at the top, preventing further pilin assembly (Korotkov and Hol, 2008). Similar minor pilin interactions involving GspI homologues have been described in other T2SS: in *P. aeruginosa*, $XcpV_I$ acts as the nucleator of a complex, involving the minor pseudopilins $XcpX_K$ and $XcpW_J$ (Douzi et al, 2009), while in *K. oxytoca*, PulI binds to PulJ and PulK, thereby promoting pseudopilus polymerisation and protein secretion (Cisneros et al, 2012a). These studies reveal a conserved pattern in T2SS of a small minor pseudopilin (12-14 kDa) interacting with two bigger minor pseudopilins, often containing additional C-terminal modules, to form a multimeric complex at the tip of the pseudopilus.

Based on 1) the striking structural homology between PilA and the GspI-like T2SS pseudopilins, 2) the modular architectures of PilB, PilC and the GspJ, GspK pseudopilins, and 3) the close similarities between the major pilin PilE1 and the T2SS major pseudopilins (Berry et al, 2019), we hypothesised that this pattern extends to the *S. sanguinis*

T4P machinery. Indeed, pull-down assays revealed a clear and exclusive interaction between two of the *S. sanguinis* minor pilins, PilA and PilC. PilA and PilC engaged in a high-affinity heterodimeric complex, with PilC acting as the preferred binding partner to PilA. To our knowledge, this is the first time the formation of a minor pilin complex has been reported in Gram-positive T4P, and the first time a minor pilin complex has been characterised to a level of detail rivalling the reports on the T2SS minor pseudopilins.

The binding interface of the PilAC dimer was determined by multidimensional NMR studies, pull-down assays with bimodular configurations of the PilC protein and computational modelling using Google's Deepmind AlphaFold program (Jumper et al, 2021; Evans et al, 2022). The PilAC structural model indicated that PilA was incorporated first into the filaments, followed by PilC whose large C-terminal modules loomed over PilA. In fact, due to the large size of PilC, the PilAC dimer could only be accommodated at the tip of a computational model of the *S. sanguinis* T4P heteropolymer. A key feature of the divergent class III signal peptide of PilA is the lack of the highly conserved glutamate residue at position 5 deemed essential for filament assembly (Berry et al, 2019). Glu5 forms a salt bridge with the N-terminal amide of Phe1 of the preceding pilin subunit, thereby facilitating the docking of incoming pilins into the nascent pilus structure (Kolappan et al, 2016). Both T2S and T4P systems have been noted to possess one unique minor pilin which lacks the highly conserved Glu5 residue – PilX in *P. aeruginosa*, PilK in *Neisseria* and GspK in the *E. coli* T2SS (Giltner et al, 2012). It is thought that the lack of a charged side chain improves the ability of these unique pilins to leave the membrane during assembly initiation (Giltner et al, 2012). Indeed, the crystal structure of the *E. coli* T2SS minor pseudopilin complex shows GspK residing at the top and GspI at the base of the complex (Korotkov and Hol, 2008). Since these pilins are located at the

apex of the filament, they have no preceding subunit with which to establish a salt bridge (Kolappan et al, 2016) and consequently do not require the Glu5 residue. Accordingly, PilA follows this principle and assumes the forefront position in the *S. sanguinis* minor pilin complex. Even though the position of PilA in the minor pilin complex differs from that of its structural homologue GspI, its function is analogous: in binding to PilC, PilA acts as the nucleator in the formation of the minor pilin complex.

The AlphaFold PilAC model also provided structural insight into the stabilising effect of PilA on PilC. PilA was found to protect the pilin domain of PilC from degradation in both long-term stability assays and trypsin sensitivity assays. According to the AlphaFold structural model, PilA extends the highly unusual β-sheet of the PilC pilin domain, thereby forming an intermolecular β-sheet. This stabilising effect is reminiscent of the donor-strand complementation characteristic of the chaperone-usher pili, which are not related to T4P. Chaperone-usher pili are important virulence factors which play a key role in the pathogenesis of urinary tract infections caused by uropathogenic *E. coli* (Hospenthal et al, 2017). The chaperone-usher pilins display an incomplete Ig-like fold domain which lacks the seventh C-terminal β-strand usually found in Ig-like folds. The lack of this strand results in a hydrophobic groove that runs along the length of the pilin and makes the individual pilins inherently unstable. Once transported across the inner membrane, the pilins need to be stabilised by a periplasmic chaperone which complements the hydrophobic groove by donating one of its own β-strands. Pilins then polymerise into filaments by donor strand exchange, in which the donor strand of the chaperone is replaced with a donor strand from another pilus subunit (Hospenthal et al, 2017; Werneburg and Thanassi, 2018). This secures the stability of the pilins and provides a binding interface for incoming pilus subunits, thereby supporting pilus assembly. Similarly, in *S. sanguinis*,

PilA donates a single β-strand to complement the three-stranded β-sheet of the PilC pilin domain. Thus, PilA acts simultaneously as a nucleator for complex formation by providing a binding interface for minor pilin PilC, and as a chaperone for minor pilin PilC by protecting it from degradation. It would be interesting to test the effect of disrupting the chaperone function of PilA by mutagenesis on PilC stability and T4P biogenesis *in vivo*.

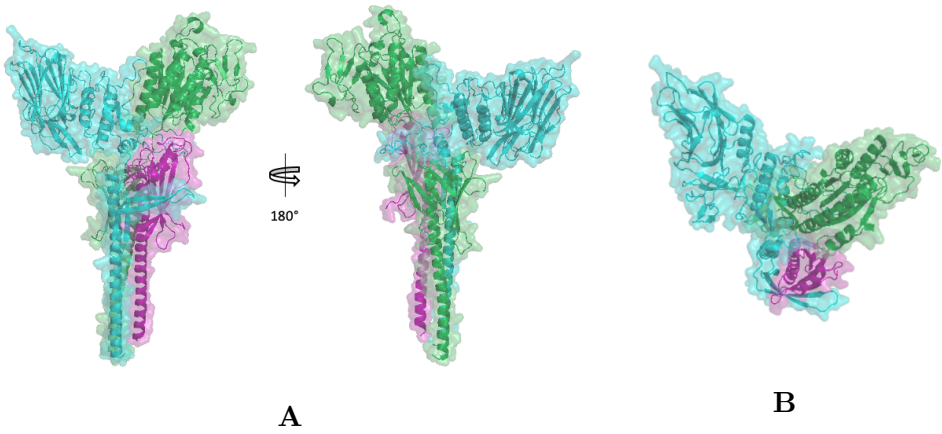


Figure 7.1: **AlphaFold model of the PilA-PilB-PilC complex.** (A) 3D structural model of the PilABC interaction. PilA is shown in purple, PilB in green and PilC in blue. (B) The winged shape of the PilABC complex viewed from the top.

While we have characterised the PilA-PilC interaction in great detail, one question that remains unsolved is the involvement of the minor pilin PilB in the *S. sanguinis* minor pilin complex. Trimeric minor pilin interactions are a hallmark of the T2SS – the GspIJK complex in *E. coli* (Korotkov and Hol, 2008), the $XcpV_IW_JX_K$ complex in *P. aeruginosa* (Douzi et al, 2009) and the PulIJK complex in *K. oxytoca* (Cisneros et al, 2012). They are also consistent with the predicted 3.6 units per turn of the helical T4P filaments (Giltner et al, 2012). The dependence of the three minor pilins on one another for pili polymerisation (Chapter 3) suggests that PilA, PilB and PilC all interact together, but our experimental findings have been ambiguous. While we were unable to detect an interaction between PilB, PilA and/or PilC in the pull-down assays, a weak and transient interaction between the full-length PilB and PilC was observed in the BACTH assays. AlphaFold modelling

of the PilABC complex supports the occurrence of a PilBC interaction – PilB is predicted to form the base of the minor pilin complex, binding to the PilAC heterodimer via a limited number of contacts between the PilB vWA domain and the PilC Ig fold domain (Figure 7.1). Ig-like folds typically mediate protein-protein interactions (Bodelon et al, 2013), which makes it plausible that the function of this middle PilC module is to engage PilB in the minor pilin complex. Weaker interactions between a minor pilin dimer and a third minor pilin are not uncommon in T2SS. In the pivotal study of the *E. coli* T2SS minor pilin complex, GspK was found to interact with a GspI-GspJ dimer (Korotkov and Hol, 2008). Similarly, the dimeric interaction between the *K. oxytoca* PulI and PulJ pseudopilins was readily detected by BACTH assays, whereas binding of PulK to the PulIJ dimer could only be observed following cysteine crosslinking (Cisneros et al, 2012a). And in the *V. vulnificus* T2SS, only an interaction between EpsI and EpsJ has been reported so far (Yanez et al, 2008a). Thus, our inability to detect an interaction between PilB and the PilAC heterodimer does not exclude the possibility that such an interaction takes place in nature.

Indeed, computational modelling done by Dr Sheppard has demonstrated that it is possible to fit either the PilABC trimer or the PilAC dimer at the tip of the pili (Figure 7.2). The AlphaFold model of the PilAC dimer indicates that PilA is incorporated into the pili first, followed by PilC. This is in keeping with our earlier findings that PilC might act as a linker between the tip-located complex and the nascent heteropolymeric pilus. The interaction between PilC and the two major pilins, PilE1 and PilE2, detected in the BACTH assays, would enable the localisation of the PilAC heterodimer at the tip of the filament. Conversely, in the PilABC heterotrimer, PilB is located at the base of the tip-located complex. However, this configuration fails to explain the strong interactions detected between

PilC and the major pilins and is not supported by evidence of interactions between PilB, PilE1 and PilE2. While we posit that the formation of a PilAC dimer at the tip of the pili is more likely based on our results, we are unable to conclusively rule out that a PilABC complex is formed or that it is located at the tip of the *S. sanguinis* T4P. This raises the possibility that *S. sanguinis* might express two distinct kinds of T4P capped by either PilB or PilAC, or only one kind of T4P capped by the PilABC complex. Consequently, it would be useful to perform immunogold electron microscopy on T4P (either attached to or sheared from the bacterial cells) in order to pinpoint the localisation of the three minor pilins. This would not only confirm that the minor pilin complex resides at the tip of the filaments, but would also indicate whether the minor pilin complex is the dimeric PilAC or the trimeric PilABC. Alternatively, cryo-ET of the *S. sanguinis* T4P machinery with and without an extended pilus could provide an electron density map of the minor pilin tip complex in which we could try to fit the PilAC or PilABC AlphaFold models. Similar studies were recently performed in *M. xanthus* (Treuner-Lange et al, 2020).

There are two possible, mutually compatible roles that minor pilin complex formation could play in T4P biogenesis: 1) it could support T4P assembly and 2) it could facilitate T4P functionality. Evidence for the former derives mainly from the homologous T2SS. In *P. aeruginosa*, $XcpV_I$ which acts as the nucleator of minor pseudopilin complex formation is essential for pseudopilus assembly (Durand et al, 2005). Similarly, in *K. oxytoca*, the minor pseudopilin complex was found to acquire a pseudopilus-like structure in the cytoplasmic membrane, thereby priming filament assembly (Cisneros et al, 2012a). Two independent observations that minor pseudopilins can support T4P assembly in the absence of minor pilins have led to the theory that the T2SS and T4P minor pilins play an interchangeable role in filament assembly, implicating both sets of proteins in the

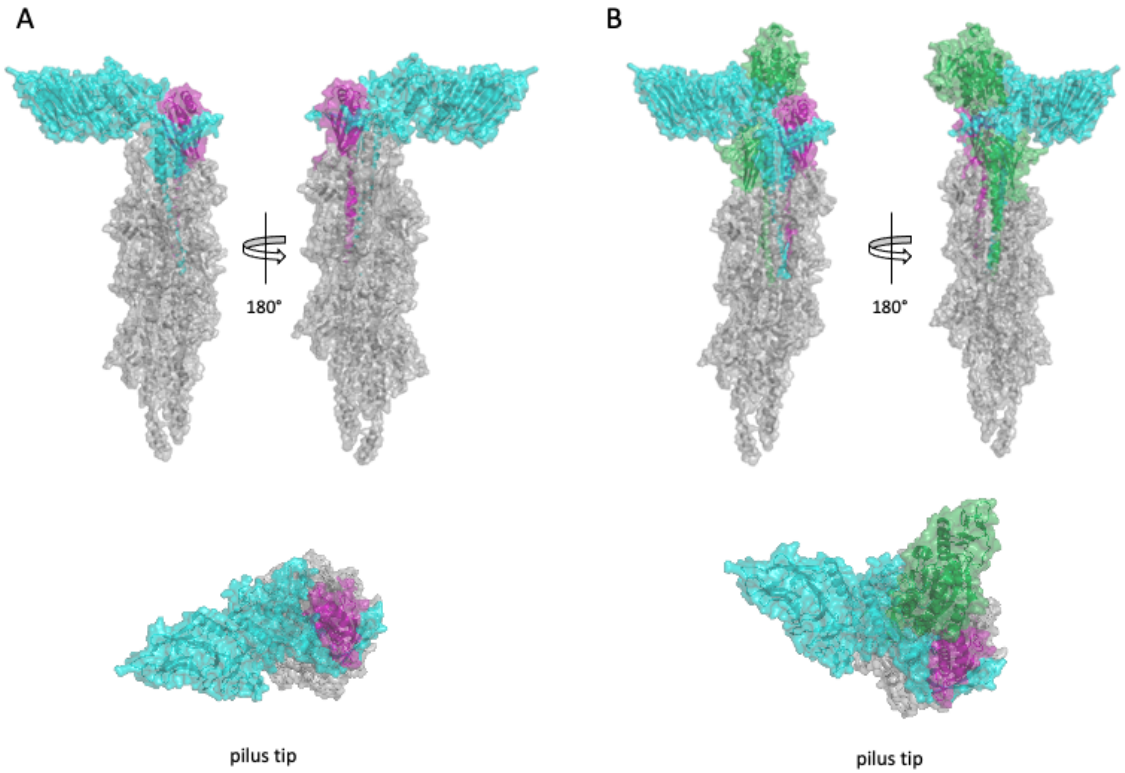


Figure 7.2: **3D model of the minor pilin complex in *S. sanguinis* T4P.** (A) Packing of the PilAC complex into *S. sanguinis* T4P. (B) Packing of the PilABC complex into *S. sanguinis* T4P. The heteropolymeric pilus is shown in grey, PilA is purple, PilB is green and PilC is blue.

mechanisms of pilus initiation (Cisneros et al, 2012b; Nguyen et al, 2015a). While it is unclear how minor pilin complexes trigger filament assembly – whether by acquiring a pre-assembled conformation or by transducing their binding energy to the T4P assembly platform – the critical role of minor pilins in T4P biogenesis has been documented in multiple organisms (Jacobsen et al, 2020), including *S. sanguinis* (Gurung et al, 2016). There are several arguments in favour of a role for the *S. sanguinis* minor pilin complex in filament assembly: 1) PilA, PilB and PilC are essential for *S. sanguinis* T4P biogenesis (Gurung et al, 2016); 2) deletion of one or all three minor pilins in a Δ*pilT* background does not restore piliation (unpublished data), indicating that the *S. sanguinis* minor pilins are not involved in counteracting PilT-mediated retraction like the core minor pilins of *Neisseria* (Winther-Larsen et al, 2005; Carbonnelle et al, 2006); 3) the close structural homology between the *S. sanguinis* pilins and the T2SS pilins (Berry et al, 2019); 4) the

anticipated location of the minor pilin complex at the tip of the pili; and 5) the axial rise between the pilin transmembrane domains in the AlphaFold models of the PilAC dimer and the PilABC trimer which suggests a pre-assembled conformation. Thus, it is possible that the formation of the *S. sanguinis* minor pilin complex plays a role in initiating filament assembly.

Our lab recently proposed that minor pilin complex formation allows the presentation of functional modules at the tip of T4P, thereby increasing the functional versatility of the filaments (Raynaud et al, 2021). An excellent example of this is the presentation of the pilus-associated adhesin PilC/PilY1 in Gram-negative T4P systems via interactions with the four widely conserved core minor pilins (Rudel et al, 1995). In both the PilAC and the PilABC pilus tip configurations, it is clear that the minor pilin complex facilitates the presentation of the modular pilins' C-terminal domains on the surface of the pili. Even though PilA resides at the top of the complex, its small size allows the Ig fold and the lectin domains of PilC to tower over it and cap the filaments (Figure 7.2A). This results in a pilus tip that is tilted to one side and not in line with the pilus body, similar to the tip structures observed in the cryo-ET studies of the *M. xanthus* T4aP (Treuner-Lange et al, 2020). Further binding of PilB to the PilAC heterodimer creates a wing-like shape of the minor pilin complex in which the vWA domain of PilB opens up to the right of the pilus axis and the lectin domain of PilC occupies the left (Figure 7.2B). The winged structure thus accommodates the simultaneous presentation of two different functional domains on the tip of the *S. sanguinis* T4P – the vWA domain of PilB and the lectin domain of PilC – both of which are implicated in host adhesion. The PilB vWA domain enables *S. sanguinis* T4P-mediated binding to eukaryotic cells and to host extracellular matrix proteins, including fibronectin and fibrinogen (Raynaud et al, 2021), and is proposed to

contribute to infective endocarditis pathogenesis in *S. sanguinis*. The PilC lectin domain is predicted to bind to host glycosylated proteins, suggesting a possible role for PilC in the commensal or pathogenic lifestyle of *S. sanguinis*.

Thus, the small minor pilin PilA plays a central role in T4P biology: PilA initiates the formation of the minor pilin complex, it acts as chaperone for PilC, stabilising its pilin domain, and it facilitates the presentation of functional adhesins at the tip of the *S. sanguinis* filaments. We propose that the role of PilA could be extrapolated to the small core minor pilins of other T4F systems. The core minor pilins PilI are homologues of the T2SS minor pseudopilins GspI (Table 1.1). They are distinguished by their small size and are proposed to form a tip-located minor pilin complex with the bigger, modular minor pilins PilJ and PilK. They are also essential for T4P biogenesis – in their absence, no or suboptimal T4P assembly is observed (Carbonnelle et al, 2006; Giltner et al, 2010; Nguyen et al, 2015a). These features apply to the *S. sanguinis* minor pilin PilA as well – its deletion leads to loss of piliation (Gurung et al, 2016), it is significantly smaller than *S. sanguinis* minor pilins PilB and PilC (Berry et al, 2019), and we have demonstrated that it forms a minor pilin complex. Furthermore, PilA exhibits striking structural homology to the T2SS GspI proteins, specifically GspI from *E. coli*. These observations point to a functional homology between the *S. sanguinis* PilA and the PilI minor pilins in Gram-negative species. Considering the essential role of PilA in the formation and stabilisation of the *S. sanguinis* minor pilin complex, this functional homology could establish an overarching principle for core minor pilin complex formation in T4P biogenesis.

Lastly, it is worth noting that PilI is the only one of the four core T4aP minor pilins to be found in *S. sanguinis*. PilB and PilC bear no structural or sequence homology to either PilH, PilJ or PilK. This not only reaffirms the central role of the PilI minor

pilins in T4P biogenesis, but also suggests that the *S. sanguinis* minor pilin complex is a rudimentary version of the Gram-negative T4P and the T2SS tip-located complexes. The simplicity of the *S. sanguinis* minor pilin complex is further underscored by the lack of tip-associated adhesins or effectors in the *S. sanguinis* T4P machinery. Instead, the functional modules of the *S. sanguinis* T4P have been grafted directly onto its minor pilins (Raynaud et al, 2021). It is tempting to propose *S. sanguinis* as a prototype of the tip-located minor pilin complexes found in T4F, but more structure/function and phylogenetic studies are required to trace the evolution of these essential lower-abundance pilins and their interactions.

7.3 Functional Analysis of Minor Pilin PilC

The *S. sanguinis* minor pilin PilC is one of the first functional T4P-associated lectins to be studied in detail. While a database search (conducted in May 2022) for type IV pilins whose architecture comprises an IPR012902 pilin motif and a PF13385 lectin motif revealed 259 proteins, the large majority were uncharacterised. Known examples of type IV pilin lectins include the major subunit BfpA of the bundle-forming T4P of EPEC and the minor pilin CofB of the T4bP of ETEC. BfpA demonstrates lectin-like binding activity towards N-acetyllactosamine glycan residues on human intestinal epithelial cells (Hyland et al, 2008) but does not possess a structural lectin motif. CofB has a C-terminal H-type lectin domain but exhibits no carbohydrate-binding activity (Kawahara et al, 2016). H-type lectins typically bind N-acetylgalactosamine (GalNac) at their trimeric interface, but no interactions have been detected between GalNac and CofB to date. Instead, the CofB trimer uses its GalNac-binding site to anchor the secreted protein CofJ to the pilus (Oki et al, 2018).

In contrast, the *S. sanguinis* minor pilin PilC displays both the structure and functionality of a lectin in its C-terminal domain. On the structural side, the PilC lectin module belongs to the concanavalin A-like (ConA-like) lectin/glucanase domain family (InterPro entry IPR013320) (Berry et al, 2019). It shows close homology to members of this family, specifically human Gal-7, despite sharing no sequence identity. Structural similarity without sequence homology is often reported for bacterial lectins which are proposed to have arisen by convergent evolution as mimicry of eukaryotic proteins (Zelensky and Gready, 2005). On the functional side, PilC was found to bind to two types of glycans – sialylated glycans and GAGs (Dr Raynaud, unpublished data). The interactions between PilC and the 3'-SL/3'-SLN termini of the identified sialylated ligands were characterised by low μM affinity, similar to the affinities reported for galectins and other bacterial lectins (Topin et al, 2013; Patnaik et al, 2006). Binding was found to occur on the concave side of the lectin module where the top four β-sheets of the β-sandwich form a shallow binding pocket – mutating a key residue in this region caused a significant decrease in the binding affinity of PilC for both synthetic trisaccharides. Importantly, in both models of the tip-located minor pilin complex (PilAC or PilABC), the carbohydrate binding site was surface-exposed and unobstructed, allowing the lectin module to freely engage with its glycan ligands. This substantiates the role of PilC as a functional T4P adhesin in *S. sanguinis*.

Even though the structure of the PilC lectin domain places it in the ConA-like lectin family, its functionality is disparate. While ConA-like lectins present preferential binding to mannose, glucose and their derivatives without binding to sialic acid (Cavada et al, 2018), PilC binds almost exclusively to sialylated glycans. A large subgroup of ConA-like lectins display metal-dependent carbohydrate binding (Zelensky and Gready, 2005), but

Ca^{2+} or Mn^{2+} ions were not detected in the structure of PilC (Dr Raynaud, unpublished data), and PilC-ligand binding took place in their absence. Furthermore, the PilC lectin module binds to carbohydrates in a monomeric state, which is maintained by the N-terminal Ig fold module. Deletion of the Ig fold module led to lectin oligomerisation which appeared to interfere with ligand binding. In contrast, most ConA-like lectins are organised as homo- or heterooligomers which enhances their glycan-binding valence (Vasta, 2009; Laaf et al, 2019). Lastly, although the carbohydrate binding sites of PilC and the ConA-like lectins are closely aligned structurally, the lack of sequence conservation means that the lectin-glycan interactions are coordinated by different residues (Zelensky and Gready, 2005; Modenutti et al, 2019).

Lectins in bacterial pathogens are often involved in host recognition and tissue adhesion. Host tissues display a large array of free or covalently attached glycans that create a binding milieu for colonising microorganisms (Audfray et al, 2013). Specifically, sialic acid residues often make up the termini of glycan chains of cell surface glycoproteins and glycolipids and so are a major target for numerous pathogens (Imberty and Varrot, 2008). For instance, both the *S. aureus* enterotoxin SEC2 and the *Clostridium botulinum* neurotoxin B bind to 3'-SL (Swaminathan et al, 1995; Swaminathan and Eswaramoorthy, 2000). Our findings that PilC binds to sialylated glycans, and that its lectin domain is presented at the tip of the T4P suggest a possible role for the *S. sanguinis* minor pilin in tissue adhesion. However, PilC appeared to play a minimal role in T4P-mediated cell adhesion. PilC mutants defective for carbohydrate-binding showed similar levels of adhesion to CHO and HeLa cells as the WT. Meanwhile, mutating a single MIDAS residue in the minor pilin PilB led to a 33-fold decrease in *S. sanguinis* cell adhesion compared to WT (Raynaud et al, 2021). *S. sanguinis* binding to CHO cells was also not impeded by the

addition of WT recombinant PilC. In contrast, addition of recombinant *Neisseria* minor pilin PilV to endothelial cells interfered with meningococcal binding to the sialylated N-glycans on the CD147 receptor and resulted in ∼70% decrease in *N. meningitidis* adhesion (Barnier et al, 2021).

There are several feasible explanations for why we could not detect a role for PilC in eukaryotic cell adhesion. Firstly, our prediction for the lectin CRD could be inaccurate. The prediction was based on the structural homology between PilC and Gal-7 which have different sugar binding preferences (Leonidas et al, 1998). Furthermore, there was only a slight difference between the ligand binding affinities of the $PilC_{K358A}$ and the $PilC_{CRD7}$ mutants. If the CRD prediction was correct, mutating multiple residues across those four β-strands should have resulted in a much more pronounced carbohydrate binding defect than the single missense mutation. Secondly, the two cells types we tested (CHO and HeLa cells) might not have been ideal for PilC adhesion assays. Perhaps more specialised cell types, such as immortalised human aortic endothelial cells (Martini et al, 2021) or immortalised cell lines of the oral mucosa (Bierbaumer et al, 2018), that are relevant to the lifestyle of *S. sanguinis* as an opportunistic pathogen or as a commensal might be more suitable for assaying the role of PilC in cell adhesion. Thirdly, PilC might not be involved in binding to eukaryotic cells at all. *S. sanguinis* T4P are already equipped with a minor pilin (PilB) that mediates host cell adhesion (Raynaud et al, 2021). If, as our lab has proposed, modular pilins evolved as a means of increasing the functional versatility of T4P (Raynaud et al, 2021), it seems unlikely that one T4P system would evolve two large modular pilins for the same function, making them redundant. Furthermore, the ability of *S. sanguinis* to proliferate in two very different host niches – the human oral cavity and the heart – suggests that its virulence factors must be highly versatile and adaptable.

Since PilB plays a key role in the pathogenesis of infective endocarditis (Raynaud et al, 2021), it is possible that PilC is instead involved in the commensal lifestyle of *S. sanguinis*. The primary habitat of *S. sanguinis* is the oral cavity where it actively contributes to the development of the dental plaque. *S. sanguinis* makes up 16% of all dental plaque microbiota species (Peterson et al, 2014) and as the dominant pioneer coloniser, it has been shown to initiate adhesion to the pellicle, to promote biofilm maturation and to recruit Gram-positive and Gram-negative bacteria to the mature dental plaque (Rosan and Lamont, 2000). The huge influence that *S. sanguinis* exerts in the colonisation of the mouth is made possible by its multitudinous reciprocal adhesin-receptor interactions. As a surface-exposed carbohydrate-binding adhesin, the minor pilin PilC is likely to play a central role in the attachment of *S. sanguinis* to the pellicle and the dental plaque biofilm matrix as well as in the maintenance of interbacterial contacts with other dental plaque colonisers.

In fact, the binding specificity of PilC for sialylated glycans makes it ideally suited to act as an adhesin in the oral cavity. Sialylated glycans are prominently displayed on salivary proteins and thus often serve as binding sites for commensal bacteria in the mouth (Cross and Ruhl, 2018). For instance, the *S. sanguinis* serine-rich repeat protein SrpA facilitates attachment to the pellicle by binding to the terminal sialic acid moiety of the low molecular weight mucin MG2 and to salivary agglutinin (Takamatsu et al, 2006; Plummer and Douglas, 2006). Both mucins and salivary agglutinins feature 3'-SLN trisaccharides at the termini of their respective O- and N-linked glycans (Cross and Ruhl, 2018). These terminal 3'-SLN moieties could serve as a binding site for the PilC lectin domain, allowing the minor pilin to attach to mucins, agglutinins and other salivary glycoproteins and thereby facilitate T4P-mediated adherence to the pellicle during the

early colonisation of *S. sanguinis*.

Interestingly, it has been suggested that the ability of *S. sanguinis* to interact efficiently with sialic-acid-containing structures on mucins in the oral cavity has also enabled it to bind to platelets during infective endocarditis (Plummer and Douglas, 2006). In keeping with this, SrpA and its *S. gordonii* homologues GspB and Hsa have been reported to bind to the sialic acid moieties on the platelet cell surface receptor GpIb (Plummer et al, 2005). Interactions with GpIb are proposed to help *S. sanguinis* and other pathogenic bacteria resist the shear forces of the blood and adhere to sterile vegetations in the heart. The N-terminal ligand-binding domain of the platelet receptor GpIb features 3'-SLN trisaccharides at the termini of its N-linked glycans (Li et al, 2016). This creates a potential binding site for the PilC lectin domain on the surface of platelets and raises the possibility that the minor pilin PilC acts as an accessory to PilB during the early stages of infective endocarditis. Consequently, it would be interesting to test whether PilC does interact with a fully glycosylated recombinant GpIb construct and whether PilC can support adhesion to blood platelets using a blood flow system that mimics *in vivo* blood flow. Identifying the binding partners of the minor pilin PilC in the oral cavity and/or the heart will not only improve our understanding of T4P-mediated adhesion, but will also shed much needed light on the role T4P play in *S. sanguinis* commensalism and pathogenicity.

7.4 Final Conclusion and Perspectives

The simplicity of the *S. sanguinis* T4P machinery and its genetic tractability provided us with the unprecedented opportunity to determine the role of each type IV pilin in a T4F system. Previous work described the two major pilins PilE1 and PilE2 (Berry et al, 2019) as well as the minor pilin PilB (Raynaud et al, 2021). We have now completed the study

of the *S. sanguinis* pilus subunits by characterising the remaining two minor pilins – PilA and PilC – in this thesis. We report that PilA and PilC form a heterodimeric complex which, according to modelling, can only be accommodated at the tip of the filaments. The high-affinity interaction between the two minor pilins fulfils three main functions: 1) it allows PilA to act as a chaperone for PilC, stabilising its unusual pilin domain, 2) it positions PilC at the pilus tip where its C-terminal lectin domain can interact with a range of host glycans that are widely distributed in the human glycome, 3) it plays a key role in filament assembly.

Together with our previous publications, these results suggest the following model for the *S. sanguinis* T4P. All five pilin proteins of *S. sanguinis* undergo prepilin processing by the dedicated prepilin peptidase PilD and are incorporated into the filaments (Gurung et al, 2016; Berry et al, 2019). The two major pilins PilE1 and PilE2 are incorporated together in nearly equal ratios into heteropolymeric pili (Berry et al, 2019). While neither major pilin alone is essential for piliation, homopolymeric pili are less abundant than WT and display reduced twitching motility velocities (Gurung et al, 2016). Indeed, recent studies of archaeal T4P have suggested that heteropolymeric filaments increase cellular motility and filament stability, conferring an evolutionary advantage over homopolymeric pili (Gambelli et al, 2022). Work is already under way in our lab to obtain the structure of the *S. sanguinis* pilus by cryo-EM and to determine the polymerisation pattern of the two major pilins.

The heteropolymeric filaments are capped by the three minor pilins – PilA, PilB and PilC. *S. sanguinis* is the paragon of the evolutionary tinkering strategy for increasing functional T4P versatility, having evolved two modular minor pilins. First characterised, PilB is a bimodular minor pilin. Its C-terminal von Willebrand factor A-like (vWA)

domain contributes to *S. sanguinis* twitching motility, mediates eukaryotic cell adhesion and binds to several human proteins, including fibrinogen and fibronectin (Raynaud et al, 2021). The wide range of functions that PilB confers onto the *S. sanguinis* T4P suggests that it plays a central role in the pathogenesis of *S. sanguinis* with a recent study showing impaired platelet-dependent biofilm formation in *pilB* mutants (Martini et al, 2021).

The trimodular minor pilin PilC consists of a pilin domain, an Ig-like fold domain and a concanavalin A-like lectin domain. The C-terminal lectin domain shows specific and exclusive binding to an array of monosialylated glycans and glycosaminoglycans which are highly prevalent in the human glycome. The protein-ligand interactions are mediated by a carbohydrate-recognition domain that is structurally homologous to that of other lectin proteins. Although we were unable to define at which stage of the *S. sanguinis* life cycle the PilC lectin domain plays a role, we propose that its affinity towards sialylated glycans makes it ideally suited to function as a T4P adhesin in the oral cavity and the early stages of infective endocarditis. The role of the middle Ig fold domain is less well-understood. Its main function appears to be the stabilisation of the lectin domain, but it might also be involved in protein-protein interactions. It is tempting to speculate that the Ig fold domain binds to the proteins whose N-terminal glycans are recognised by the PilC lectin domain, thereby increasing the affinity and specificity of the adhesin-receptor interaction. AlphaFold modelling has also suggested that the Ig fold interacts with the vWA domain of PilB, recruiting the bimodular pilin to the tip-located PilAC minor pilin complex.

PilA is the nucleator of the *S. sanguinis* minor pilin complex. It binds to minor pilin PilC and acts as its chaperone, stabilising its pilin domain by β-strand complementation. The divergent class III signal peptide of PilA, which lacks the highly conserved Glu5 residue,

allows it to enter the filaments first, followed by its high-affinity binding partner PilC. The PilAC complex most likely caps the *S. sanguinis* filaments, thereby positioning the PilC lectin domain optimally for binding to its glycan ligands.

The ambiguity surrounding the involvement of PilB in the minor pilin complex gives rise to two possible models for the tip of the *S. sanguinis* minor pilins. In the first model, the T4P of *S. sanguinis* are specialised. *S. sanguinis* is able to thrive in two distinct niches in the host – in the oral cavity as a commensal and in the valves of the heart as an opportunistic pathogen (Kreth et al, 2017). We propose that *S. sanguinis* expresses two distinct kinds of pili, capped by either PilB or the dimeric PilAC complex, in order to best adapt to these two environments. PilB-presenting pili might be best-suited to mediate the adherence and proliferation of opportunistic *S. sanguinis* in sterile vegetations in the heart, while the PilAC dimer might play a central role in the attachment of commensal *S. sanguinis* to the tooth pellicle and subsequent colonisation of the dental plaque. This proposal is based on our inability to detect any interactions between PilB and the other two minor pilins as well as computational modelling showing that PilB (Raynaud et al, 2021) and the PilAC dimer can be incorporated at the filament tips on their own.

Alternatively, *S. sanguinis* pili could present both modular pilins at all times as part of a heterotrimeric PilABC complex. This model is supported by the dependence of the three minor pilins on one another for pili polymerisation in both mutational and deletion studies. Furthermore, AlphaFold modelling has generated a high-confidence prediction for the PilABC complex in which PilB is recruited to the PilAC dimer via the PilC Ig fold domain. The PilABC complex is not only readily accommodated at the tip of the *S. sanguinis* T4P, but also adopts a winged shape which allows the adhesin modules of PilB and PilC to bind to their respective ligands simultaneously. It is worth noting that

the two proposed models are not incompatible. *S. sanguinis* could reformulate its T4P tip composition to best suit its survival requirements. Distinguishing between these three possibilities will improve our understanding of the role T4P play in the lifestyle of the multifaceted *S. sanguinis* species.

Apart from completing the study of all five *S. sanguinis* pilus subunits, we have also provided the first detailed characterisation of a T4P minor pilin complex. We hope that the minor pilin interactions uncovered in this thesis and their role in T4P biology will inform future studies of T4P systems, bringing us closer to understanding these complex and versatile machineries.

References

Aas FE, Lovold C and Koomey M (2002) An inhibitor of DNA binding and uptake events dictates the proficiency of genetic transformation in *Neisseria gonorrhoeae*: mechanism of action and links to Type IV pilus expression. *Mol. Microbiol.* **46**: 1441–1450.

Aas JA, Paster BJ, Stokes LN, Olsen I and Dewhirst FE (2005) Defining the Normal Bacterial Flora of the Oral Cavity. *J. Clin. Microbiol.* **43**: 5721–5732.

Aas FE, Winther-Larsen HC, Wolfgang M, Frye S, Lovold C, Roos N, van Putten JPM and Koomey M (2007) Substitutions in the N-terminal alpha helical spine of *Neisseria gonorrhoeae* pilin affect Type IV pilus assembly, dynamics and associated functions. *Mol. Microbiol.* **63**: 69–85.

Aguilo-Ferretjans M del M, Bosch R, Puxty RJ, Latva M, Zadjelovic V, Chhun A, Sousoni D, Polin M, Scanlan DJ and Christie-Oleza JA (2021) Pili allow dominant marine cyanobacteria to avoid sinking and evade predation. *Nat. Commun.* **12**: 1857.

Albers S-V, Szabo Z and Driessen AJM (2003) Archaeal Homolog of Bacterial Type IV Prepilin Signal Peptidases with Broad Substrate Specificity. *J. Bacteriol.* **185**: 3918–3925.

Alves-Barroco C, Rivas-Garcia L, Fernandes AR and Baptista PV (2020) Tackling Mul-

tidrug Resistance in Streptococci. From Novel Biotherapeutic Strategies to Nanomedicines. *Front. Microbiol.* **11**: 1–21.

Aly KA, Beebe ET, Chan CH, Goren MA, Sepulveda C, Makino SI, Fox BG and Forest KT (2013) Cell-free production of integral membrane aspartic acid proteases reveals zinc-dependent methyltransferase activity of the *Pseudomonas aeruginosa* prepilin peptidase PilD. *Microbiologyopen* **2**: 94–104.

Angelov A, Bergen P, Nadler F, Hornburg P, Lichev A, Aoebelacker M, Pachl F, Kuster B and Liebl W (2015) Novel Flp pilus biogenesis-dependent natural transformation. *Front. Microbiol.* **6**: 1–11.

Arts J, Van Boxtel R, Filloux A, Tommassen J and Koster M (2007) Export of the pseudopilin XcpT of the *Pseudomonas aeruginosa* type II secretion system via the signal recognition particle-Sec pathway. *J. Bacteriol.* **189**: 2069–2076.

Audfray A, Varrot A and Imberty A (2013) Bacteria love our sugars: Interaction between soluble lectins and human fucosylated glycans, structures, thermodynamics and design of competing glycocompounds. *Comptes Rendus Chim.* **16**: 482–490.

Aviles-Reyes A, Miller JH, Lemos JA and Abranches J (2017) Collagen-binding proteins of *Streptococcus mutans* and related streptococci. *Mol. Oral Microbiol.* **32**: 89–106.

Baird TT and Craik CS (2013) Chapter 575 –Trypsin. in Rawlings, NS and Salvesen G *Handb. Proteolytic Enzym.* Third Edition. Academic Press. 2594-2600.

Balasingham SV, Collins RF, Assalkhou R, Homberset H, Frye SA, Derrick JP, Tonjum T (2007) Interactions between the lipoprotein PilP and the secretin PilQ in *Neisseria meningitidis J. Bacteriol.* **189**: 5716-5727.

Barnier J-P, Meyer J, Kolappan S, Bouzinba-Segard H, Gesbert G, Jamet A, Frapy E, Schonherr-Hellec S, Capel E, Virion Z, Dupuis M, Bille E, Morand P, Schmitt T, Bourdoulous S, Nassif X, Craig L and Coureuil M (2021) The minor pilin PilV provides a conserved adhesion site throughout the antigenically variable meningococcal type IV pilus. *Proc. Natl. Acad. Sci.* **118**: e2109364118.

Bashore TM, Cabell C and Fowler, Jr V (2006) Update on Infective Endocarditis. *Curr. Probl. Cardiol.* **31**: 274–352.

Beeby M (2019) Evolution of a family of molecular Rube Goldberg contraptions. *PLoS Biol.* **17**: 1–9.

Bensing BA, Loukachevitch L V., Mcculloch KM, Yu H, Vann KR, Wawrzak Z, Anderson S, Chen X, Sullam PM and Iverson TM (2016) Structural basis for sialoglycan binding by the *Streptococcus sanguinis* SrpA adhesin. *J. Biol. Chem.* **291**: 7230–7240.

Bernard SC, Simpson N, Join-Lambert O, Federici C, Laran-Chich M-P, Maïssa N, Bouzinba-Segard H, Morand PC, Chretien F, Taouji S, Chevet E, Janel S, Lafont F, Coureuil M, Segura A, Niedergang F, Marullo S, Couraud P-O, Nassif X and Bourdoulous S (2014) Pathogenic *Neisseria meningitidis* utilizes CD147 for vascular colonization. *Nat. Med.* **20**: 725–731.

Berry JL and Pelicic V (2015) Exceptionally widespread nanomachines composed of type IV pilins: The prokaryotic Swiss Army knives. *FEMS Microbiol. Rev.* **39**: 134–154.

Berry J-L, Xu Y, Ward PN, Lea SM, Matthews SJ and Pelicic V (2016) A Comparative Structure/Function Analysis of Two Type IV Pilin DNA Receptors Defines a Novel Mode of DNA Binding. *Structure* **24**: 926–934.

Berry J-L, Gurung I, Anonsen JH, Spielman I, Harper E, Hall AMJ, Goosens VJ, Ray-

naud C, Koomey M, Biais N, Matthews S and Pelicic V (2019) Global biochemical and structural analysis of the type IV pilus from the Gram-positive bacterium *Streptococcus sanguinis*. *J. Biol. Chem.* **294**: 6796–6808.

Bhaya D, Bianco NR, Bryant D and Grossman A (2000) Type IV pilus biogenesis and motility in the cyanobacterium *Synechocystis sp.* PCC6803. *Mol. Microbiol.* **37**: 941–951.

Bierbaumer L, Schwarze UY, Gruber R and Neuhaus W (2018) Cell culture models of oral mucosal barriers: A review with a focus on applications, culture conditions and barrier properties. *Tissue Barriers.* **6**: 1–42.

Bischof LF, Friedrich C, Harms A, Sogaard-Andersen L and van der Does C (2016) The Type IV Pilus Assembly ATPase PilB of *Myxococcus xanthus* Interacts with the Inner Membrane Platform Protein PilC and the Nucleotide-binding Protein PilM. *J. Biol. Chem.* **291**: 6946–6957.

Bochud PY, Calandra T, Francioli P (1994) Bacteremia due to viridans streptococci in neutropenic patients: a review. *Am J Med.* **97**: 256–64.

Bodelon G, Palomino C and Fernandez LA (2013) Immunoglobulin domains in *Escherichia coli* and other enterobacteria: From pathogenesis to applications in antibody technologies. *FEMS Microbiol. Rev.* **37**: 204–250.

Bradley DE (1972) Shortening of *Pseudomonas aeruginosa* pili after RNA-phage adsorption. *J Gen Microbiol* **72**:303–19.

Braun T, Vos MR, Kalisman N, Sherman NE, Rachel R, Wirth R, Schroder GF and Egelman EH (2016) Archaeal flagellin combines a bacterial type IV pilin domain with an Ig-like domain. *Proc. Natl. Acad. Sci. U. S. A.* **113**: 10352–10357.

Brennan CA and Garrett WS (2019) *Fusobacterium nucleatum* — symbiont, opportunist and oncobacterium. *Nat. Rev. Microbiol.* **17**: 156–166.

Brissac T, Mikaty G, Dumenil G, Coureuil M and Nassif X (2012) The Meningococcal Minor Pilin PilX Is Responsible for Type IV Pilus Conformational Changes Associated with Signaling to Endothelial Cells. *Infect. Immun.* **80**: 3297–3306.

Brown DR, Helaine S, Carbonnelle E and Pelicic V (2010) Systematic functional analysis reveals that a set of seven genes is involved in fine-tuning of the multiple functions mediated by type IV pili in Neisseria meningitidis. *Infect. Immun.* **78**: 3053–3063.

Burrows LL (2008) A Nice Return on the 'Stalk' Exchange. *Structure* **16**: 19–20.

Burrows LL (2012) *Pseudomonas aeruginosa* twitching motility: Type IV pili in action. *Annu. Rev. Microbiol.* **66**: 493–520.

Carbonnelle E, Helaine S, Prouvensier L, Nassif X and Pelicic V (2004) Type IV pilus biogenesis in *Neisseria meningitidis*: PilW is involved in a step occurring after pilus assembly, essential for fibre stability and function. *Mol. Microbiol.* **55**: 54–64.

Carbonnelle E, Helaine S, Nassif X and Pelicic V (2006) A systematic genetic analysis in *Neisseria meningitidis* defines the Pil proteins required for assembly, functionality, stabilization and export of type IV pili. *Mol. Microbiol.* **61**: 1510–1522.

Carlsson J (1965) Zooglea-forming streptococci, resembling *Streptococcus sanguis*, isolated from dental plaque in man. *Odontol. Revy* **16**: 348–358.

Caufield PW, Dasanayake AP, Li Y, Pan Y, Hsu JAY and Hardin JM (2000) Natural History of *Streptococcus sanguinis* in the Oral Cavity of Infants: Evidence for a Discrete Window of Infectivity. *Infect. Immun.* **68**: 4018–4023.

Cavada B, Pinto-Junior V, Osterne V and Nascimento K (2018) ConA-Like Lectins: High Similarity Proteins as Models to Study Structure/Biological Activities Relationships. *Int. J. Mol. Sci.* **20**: 30.

Cehovin A, Kroll JS and Pelicic V (2011) Testing the vaccine potential of PilV, PilX and ComP, minor subunits of *Neisseria meningitidis* type IV pili. *Vaccine* **29**: 6858–6865.

Cehovin A, Simpson PJ, McDowell MA, Brown DR, Noschese R, Pallett M, Brady J, Baldwin GS, Lea SM, Matthews SJ and Pelicic V (2013) Specific DNA recognition mediated by a type IV pilin. *Proc. Natl. Acad. Sci.* **110**: 3065–3070.

Chambers HF and Bayer AS (2020) Native-Valve Infective Endocarditis. *N. Engl. J. Med.* **383**: 567–576.

Chang Y-W, Rettberg LA, Treuner-Lange A, Iwasa J, Sogaard-Andersen L and Jensen GJ (2016) Architecture of the type IVa pilus machine. *Science.* **351**: 1–18.

Chatterjee N, Cook LCC, Lyles K V., Nguyen HAT, Devlin DJ, Thomas LS and Eichenbaum Z (2020) A novel heme transporter from the energy coupling factor family is vital for group a streptococcus colonization and infections. *J. Bacteriol.* **202**.

Chaudhury P, Quax TEF and Albers SV (2018) Versatile cell surface structures of archaea. *Mol. Microbiol.* **107**: 298–311.

Chen I and Dubnau D (2004) DNA uptake during bacterial transformation. *Nat. Rev. Microbiol.* **2**: 241–249.

Chen YYM, Chiang YC, Tseng TY, Wu HY, Chen YY, Wu CH and Chiu CH (2019) Molecular and functional analysis of the type IV pilus gene cluster in *Streptococcus sanguinis* SK36. *Appl. Environ. Microbiol.* **85**: 1–15.

Chiang SL, Taylor RK, Koomey M and Mekalanos JJ (1995). Single amino acid substitutions in the N-terminus of *Vibrio cholerae* TcpA affect colonization, autoagglutination, and serum resistance. *Mol. Microbiol.* **17**: 1133–1142.

Chlebek JL, Denise R, Craig L and Dalia AB (2021) Motor-independent retraction of type IV pili is governed by an inherent property of the pilus filament. *Proc. Natl. Acad. Sci.* **118**: e2102780118.

Cisneros DA, Bond PJ, Pugsley AP, Campos M and Francetic O (2012a) Minor pseudopilin self-assembly primes type II secretion pseudopilus elongation. *EMBO J.* **31**: 1041–1053.

Cisneros DA, Pehau-Arnaudet G and Francetic O (2012b) Heterologous assembly of type IV pili by a type II secretion system reveals the role of minor pilins in assembly initiation. *Mol. Microbiol.* **86**: 805–818.

Clemmer KM, Bonomo RA and Rather PN (2011) Genetic analysis of surface motility in *Acinetobacter baumannii*. *Microbiology* **157**: 2534–2544.

Collins RF, Davidsen L, Derrick JP, Ford RC and Tonjum T (2001) Analysis of the PilQ secretin from *Neisseria meningitidis* by transmission electron microscopy reveals a dodecameric quaternary structure. *J. Bacteriol.* **183**: 3825-3832.

Collins RF, Frye SA, Kitmitto A, Ford RC, Tonjum T and Derrick JP (2004) Structure of the *Neisseria meningitidis* outer membrane PilQ complex at 12 A resolution. *J. Biol. Chem.* **279**: 39750-39756.

Collins RF, Frye SA, Balasingham S, Ford RC, Tonjum T and Derrick JP (2005) Interaction with type IV pili induces structural changes in the bacterial outer membrane secretin PilQ. *J. Biol. Chem.* **280**: 18923–18930.

Coureuil M, Lecuyer H, Scott MGH, Boularan C, Enslen H, Soyer M, Mikaty G, Bourdoulous S, Nassif X and Marullo S (2010) Meningococcus Hijacks a β2-Adrenoceptor/β-Arrestin Pathway to Cross Brain Microvasculature Endothelium. *Cell* **143**: 1149–1160.

Craig L and Li J (2008) Type IV pili: paradoxes in form and function. *Curr. Opin. Struct. Biol.* **18**: 267–277.

Craig L, Taylor RK, Pique ME, Adair BD, Arvai AS, Singh M, Lloyd SJ, Shin DS, Getzoff ED, Yeager M, Forest KT and Tainer JA (2003) Type IV Pilin Structure and Assembly. *Mol. Cell* **11**: 1139–1150.

Craig L, Pique ME and Tainer JA (2004) Type IV pilus structure and bacterial pathogenicity. *Nat. Rev. Microbiol.* **2**: 363–378.

Craig L, Forest KT and Maier B (2019) Type IV pili: dynamics, biophysics and functional consequences. *Nat. Rev. Microbiol.* **17**: 429–440.

Cross BW and Ruhl S (2018) Glycan recognition at the saliva – oral microbiome interface. *Cell. Immunol.* **333**: 19–33.

Crump KE, Bainbridge B, Brusko S, Turner LS, Ge X, Stone V, Xu P and Kitten T (2014) The relationship of the lipoprotein SsaB, manganese and superoxide dismutase in *Streptococcus sanguinis* virulence for endocarditis. *Mol. Microbiol.* **92**: 1243–1259.

Cummings RD (2009) The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.* **5**: 1087–1104.

D'Arcy A, Bergfors T, Cowan-Jacob SW and Marsh M (2014) Microseed matrix screening for optimization in protein crystallization: What have we learned? *Acta Crystallogr. Sect. Struct. Biol. Commun.* **70**: 1117–1126.

De Greve H, Wyns L and Bouckaert J (2007) Combining sites of bacterial fimbriae. *Curr Opin Struct Biol* **17**: 506–512.

Denise R, Abby SS and Rocha EPC (2019) Diversification of the type IV filament superfamily into machines for adhesion, protein secretion, DNA uptake, and motility. *PLOS Biol.* **17**: e3000390.

Douzi B, Durand E, Bernard C, Alphonse S, Cambillau C, Filloux A, Tegoni M and Voulhoux R (2009) The XcpV/GspI Pseudopilin Has a Central Role in the Assembly of a Quaternary Complex within the T2SS Pseudopilus. *J. Biol. Chem.* **284**: 34580–34589.

Douzi B, Filloux A and Voulhoux R (2012) On the path to uncover the bacterial type II secretion system. *Philos. Trans. R. Soc. B Biol. Sci.* **367**: 1059–1072.

Draskovic I and Dubnau D (2004) Biogenesis of a putative channel protein, ComEC, required for DNA uptake: membrane topology, oligomerization and formation of disulphide bonds. *Mol. Microbiol.* **55**: 881–896.

Dubendorf JW and Studier FW (1991) Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* **219**: 45–59.

Dubnau D and Blokesch M (2019) Mechanisms of DNA Uptake by Naturally Competent Bacteria. *Annu. Rev. Genet.* **53**: 217–237.

Duff MR, Grubbs J and Howell EE (2011) Isothermal titration calorimetry for measuring macromolecule-ligand affinity. *J. Vis. Exp.* doi:10.3791/2796.

Dumenil G (2019) Type IV Pili as a Therapeutic Target. *Trends Microbiol.* **27**: 658–661.

Dupuy B, Taha MK, Pugsley AP and Marchal C (1991) *Neisseria gonorrhoeae* prepilin

export studied in *Escherichia coli*. *J. Bacteriol.* **173**: 7589–7598.

Durand E, Michel G, Voulhoux R, Kürner J, Bernadac A and Filloux A (2005) XcpX Controls Biogenesis of the *Pseudomonas aeruginosa* XcpT-containing Pseudopilus. *J. Biol. Chem.* **280**: 31378–31389.

Edwards AM, Potts JR, Josefsson E and Massey RC (2010) *Staphylococcus aureus* host cell invasion and virulence in sepsis is facilitated by the multiple repeats within FnBPA. *PLoS Pathog.* **6**: e1000964.

Einarson, M.B. and Orlinick, J.R. (2002). Identification of Protein-Protein Interactions with Glutathione S-Transferase Fusion Proteins. *In Protein-Protein Interactions: A Molecular Cloning Manual* 37-57.

Ellison CK, Dalia TN, Vidal Ceballos A, Wang JCY, Biais N, Brun Y V. and Dalia AB (2018) Retraction of DNA-bound type IV competence pili initiates DNA uptake during natural transformation in *Vibrio cholerae*. *Nat. Microbiol.* **3**: 773–780.

Ellison CK, Whitfield GB and Brun Y V (2022) Type IV Pili: dynamic bacterial nanomachines. *FEMS Microbiol. Rev.* **46**: 1–14.

Erickson PR and Herzberg MC (1993) The *Streptococcus sanguis* platelet aggregation-associated protein. Identification and characterization of the minimal platelet-interactive domain. *J. Biol. Chem.* **268**: 1646–1649.

Eriksson J, Eriksson OS, Maudsdotter L, Palm O, Engman J, Sarkissian T, Aro H, Wallin M and Jonsson A-B (2015) Characterization of motility and piliation in pathogenic *Neisseria*. *BMC Microbiol.* **15**: 92.

Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Židek A, Bates R,

Blackwell S, Yim J, Ronneberger O, Bodenstein S, Zielinski M, Bridgland A, Potapenko A, Cowie A, Tunyasuvunakool K, Jain R, Clancy E, Kohli P, Jumper J and Hassabis D (2022) Protein complex prediction with AlphaFold-Multimer. *bioRxiv* doi: 2021.10.04.463034.

Fan J, Zhang Y, Chuang-Smith ON, Frank KL, Guenther BD, Kern M, Schlievert PM and Herzberg MC (2012) Ecto-5'-Nucleotidase: A Candidate Virulence Factor in *Streptococcus sanguinis* Experimental Endocarditis. *PLoS One.* **7**: e38059.

Fernandez S, Ayora S, Alonson JC (2000) *Bacillus subtilis* homologous recombination: genes and products *Res. Microb.* **151**: 481-486

Francetic O, Buddelmeijer N, Lewenza S, Kumamoto CA and Pugsley AP (2007) Signal Recognition Particle-Dependent Inner Membrane Targeting of the PulG Pseudopilin Component of a Type II Secretion System. *J. Bacteriol.* **189**: 1783–1793.

Fuchs TA, Brill A, Duerschmied D, Schatzberg D, Monestier M, Myers DD, Wrobleski SK, Wakefield TW, Hartwig JH and Wagner DD (2010) Extracellular DNA traps promote thrombosis. *Proc. Natl. Acad. Sci.* **107**: 15880–15885.

Furie B and Furie BC (2008) Mechanisms of thrombus formation. *N. Engl. J. Med.* **359**: 938–49.

Gambelli L, Isupov MN, Conners R, McLaren M, Bellack A, Gold V, Rachel R and Daum B (2022) An archaellum filament composed of two alternating subunits. *Nat. Commun.* **13**: 1–11.

Ganeshkumar N, Song M and McBride BC (1988) Cloning of a *Streptococcus sanguis* adhesin which mediates binding to saliva-coated hydroxyapatite. *Infect. Immun.* **56**: 1150–1157.

Gaustad P (1985) Genetic Transformation in *Streptococcus sanguinis*. *Acta Path Microbiol Immunol Scand* **93**: 277–282.

Georgiadou M, Castagnini M, Karimova G, Ladant D and Pelicic V (2012) Large-scale study of the interactions between proteins involved in type IV pilus biology in *Neisseria meningitidis*: characterization of a subcomplex involved in pilus assembly. *Mol. Microbiol.* **84**: 857–873.

Ghosh A and Albers SV (2011) Assembly and function of the archaeal flagellum. *Biochem. Soc. Trans.* **39**: 64–69.

Gibiansky ML, Conrad JC, Jin F, Gordon VD, Motto DA, Mathewson MA, Stopka WG, Zelasko DC, Shrout JD and Wong GCL (2010) Bacteria use type IV pili to walk upright and detach from surfaces. *Science* **330**: 8–10.

Giltner CL, Habash M and Burrows LL (2010) *Pseudomonas aeruginosa* minor pilins are incorporated into type IV Pili. *J. Mol. Biol.* **398**: 444–461.

Giltner CL, Nguyen Y and Burrows LL (2012) Type IV Pilin Proteins: Versatile Molecular Modules. *Microbiol. Mol. Biol. Rev.* **76**: 740–772.

Gold V and Kudryashev M (2016) Recent progress in structure and dynamics of dual-membrane-spanning bacterial nanomachines. *Curr. Opin. Struct. Biol.* **39**: 1-7.

Gong K, Mailloux L and Herzberg MC (2000) Salivary Film Expresses a Complex, Macromolecular Binding Site for *Streptococcus sanguis*. *J. Biol. Chem.* **275**: 8970–8974.

Goosens VJ, Busch A, Georgiadou M, Castagnini M, Forest KT, Waksman G and Pelicic V (2017) Reconstitution of a minimal machinery capable of assembling periplasmic type IV pili. *Proc. Natl. Acad. Sci.* **114**: e4978–E4986.

Green ER and Mecsas J (2016) Bacterial Secretion Systems: An Overview. *Microbiol. Spectr.* **4**: 215–239.

Guennec L, Virion Z, Bouzinba-Segard H, Robbe-Masselot C, Leonard R, Nassif X, Bourdoulous S and Coureuil M (2020) Receptor recognition by meningococcal type IV pili relies on a specific complex N-glycan. *Proc. Natl. Acad. Sci. U. S. A.* **117**: 2606–2612.

Gurung I, Spielman I, Davies MR, Lala R, Gaustad P, Biais N and Pelicic V (2016) Functional analysis of an unusual type IV pilus in the Gram-positive *Streptococcus sanguinis*. *Mol. Microbiol.* **99**: 380–392.

Gurung I, Berry J-L, Hall AMJ and Pelicic V (2017) Cloning-independent markerless gene editing in *Streptococcus sanguinis*: novel insights in type IV pilus biology. *Nucleic Acids Res.* **5**: e40.

Gutierrez-Rodarte M, Kolappan S, Burrell BA and Craig L (2019) The *Vibrio cholerae* minor pilin TcpB mediates uptake of the cholera toxin phage CTXϕ. *J. Biol. Chem.* **294**: 15698–15710.

Hamburger ZA, Brown MS, Isberg RR and Bjorkman PJ (1999) Crystal structure of invasin: A bacterial integrin-binding protein. *Science* **286**: 291–295.

Hanahan D, Jessee J and Bloom FR (1991) Plasmid transformation of *Escherichia coli* and other bacteria. *Methods Enzymol.* **204**: 63–113.

Hartung S, Arvai AS, Wood T, Kolappan S, Shin DS, Craig L and Tainer JA (2011) Ultrahigh resolution and full-length pilin structures with insights for filament assembly, pathogenic functions, and vaccine potential. *J. Biol. Chem.* **286**: 44254–44265.

Heiniger RW, Winther-Larsen HC, Pickles RJ, Koomey M and Wolfgang MC (2010) In-

fection of human mucosal tissue by *Pseudomonas aeruginosa* requires sequential and mutually dependent virulence factors and a novel pilus-associated adhesin. *Cell. Microbiol.* **12**: 1158–1173.

Helaine S, Carbonnelle E, Prouvensier L, Beretti JL, Nassif X and Pelicic V (2005) PilX, a pilus-associated protein essential for bacterial aggregation, is a key to pilus-facilitated attachment of *Neisseria meningitidis* to human cells. *Mol. Microbiol.* **55**: 65–77.

Helaine S, Dyer DH, Nassif X, Pelicic V and Forest KT (2007) 3D structure/function analysis of PilX reveals how minor pilins can modulate the virulence properties of type IV pili. *Proc. Natl. Acad. Sci.* **104**: 15888–15893.

Henrick K, Bawumia S, Barboni EAM, Mehul B and Hughes RC (1998) Evidence for subsites in the galectins involved in sugar binding at the nonreducing end of the central galactose of oligosaccharide ligands: Sequence analysis, homology modeling and mutagenesis studies of hamster galectin-3. *Glycobiology* **8**: 45–57.

Henriksen SD and Eriksen J (1976) Transformation of twitching strains of *Streptococcus sanguinis*. *Acta Pathol Microbiol Scand* **84**: 433–436.

Henriksen S and Henrichsen J (1975) Twitching motility and possession of polar fimbriae in spreading *Streptococcus sanguinis* isolates from the human throat. *Acta Pathol Microbiol Scand* **83**: 133–140.

Henriksen S and Henrichsen J (1976) Further studies of twitching *Streptococcus sanguinis* isolated from the human throat. Isolation of strains with a new antigen. *Acta Pathol Microbiol Scand* **84**: 428–432.

Hilleringmann M, Giusti F, Baudner BC, Masignani V, Covacci A, Rappuoli R, Barocchi MA and Ferlenghi I (2008) Pneumococcal Pili Are Composed of Protofilaments Exposing

Adhesive Clusters of RrgA. *PLoS Pathog.* **4**: e1000026.

Hirabayashi J and Kasai KI (1993) The family of metazoan metal-independent β-galactoside-binding lectins: Structure, function and molecular evolution. *Glycobiology* **3**: 297–304.

Holm L (2020) Using Dali for protein structure comparison. *Methods Mol. Biol.* **2112**: 29-42.

Horiuchi T and Komano T (1998) Mutational analysis of plasmid R64 thin pilus prepilin: The entire prepilin sequence is required for processing by type IV prepilin peptidase. *J. Bacteriol.* **180**: 4613–4620.

Hospenthal MK, Costa TRD and Waksman G (2017) A comprehensive guide to pilus biogenesis in Gram-negative bacteria. *Nat. Rev. Microbiol.* **15**: 365–379.

Howie HL, Glogauer M and So M (2005) The *N. gonorrhoeae* Type IV Pilus Stimulates Mechanosensitive Pathways and Cytoprotection through a pilT-Dependent Mechanism. *PLoS Biol.* **3**: e100.

Hu J, Xue Y, Lee S and Ha Y (2011) The crystal structure of GXGD membrane protease FlaK. *Nature* **475**: 528–531.

Huang X, Schulte RM, Burne RA and Nascimento MM (2015) Characterization of the arginolytic microflora provides insights into pH homeostasis in human oral biofilms. *Caries Res.* **49**: 165–176.

Hyland RM, Sun J, Griener TP, Mulvey GL, Klassen JS, Donnenberg MS and Armstrong GD (2008) The bundlin pilin protein of enteropathogenic *Escherichia coli* is an N-acetyllactosamine-specific lectin. *Cell. Microbiol.* **10**: doi:10.1111/j.1462-5822.2007.01028.x.

Imberty A and Varrot A (2008) Microbial recognition of human cell surface glycoconju-

gates. *Curr. Opin. Struct. Biol.* **18**: 567–576.

Inoue H, Nojima H and Okayama H (1990) High efficiency transformation of *E. coli* with plasmids. *Gene* **96**: 23–28.

Jacobsen T, Bardiaux B, Francetic O, Izadi-Pruneyre N and Nilges M (2020) Structure and function of minor pilins of type IV pili. *Med. Microbiol. Immunol.* **209**: 301–308.

Jarrell KF and Albers SV (2012) The archaellum: An old motility structure with a new name. *Trends Microbiol.* **20**: 307–312.

Jin F, Conrad JC, Gibiansky ML and Wong GCL (2011) Bacteria use type-IV pili to slingshot on surfaces. *Proc. Natl. Acad. Sci.* **108**: 12617–12622.

Johnson MDL, Garrett CK, Bond JE, Coggan KA, Wolfgang MC and Redinbo MR (2011) *Pseudomonas aeruginosa* PilY1 binds integrin in an RGD- and calcium-dependent manner. *PLoS One* **6**: 1–8.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P and Hassabis D (2021) Highly accurate protein structure prediction with AlphaFold. *Nature.* **596**: 583–589.

Kaplan CW, Lux R, Haake SK and Shi W (2009) The *Fusobacterium nucleatum* outer membrane protein RadD is an arginine-inhibitable adhesin required for inter-species adherence and the structured architecture of multispecies biofilm. *Mol. Microbiol.* **71**: 35–47.

Karimova G, Pidoux J, Ullmann A and Ladant D (1998) A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl. Acad. Sci.* **95**: 5752–5756.

Karimova G, Dautin N and Ladant D (2008) Interaction Network among *Escherichia coli* Membrane Proteins Involved in Cell Division as Revealed by Bacterial Two-Hybrid Analysis. *J. Bacteriol.* **190**: 8248–8248.

Karuppiah V, Hassan D, Saleem M and Derrick JP (2010) Structure and oligomerization of the PilC type IV pilus biogenesis protein from *Thermus thermophilus*. *Proteins Struct. Funct. Bioinforma.* **78**: 2049–2057.

Kawahara K, Oki H, Fukakusa S, Yoshida T, Imai T, Maruno T, Kobayashi Y, Motooka D, Iida T, Ohkubo T and Nakamura S (2016) Homo-trimeric Structure of the Type IVb Minor Pilin CofB Suggests Mechanism of CFA/III Pilus Assembly in Human Enterotoxigenic *Escherichia coli*. *J. Mol. Biol.* **428**: 1209–1226.

Kelly G, Prasannan S, Daniell S, Fleming K, Frankel G, Dougan G, Connerton L and Matthews S (1999) Structure of the cell-adhesion fragment of intimin from enteropathogenic *E. coli*. *Nat. Struct. Biol.* **6**: 313–318.

Kerrigan SW, Douglas I, Wray A, Heath J, Byrne MF, Fitzgerald D and Cox D (2002) A role for glycoprotein Ib in *Streptococcus sanguis*-induced platelet aggregation. *Blood* **100**: 509–516.

Kirn TJ, Lafferty MJ, Sandoe CMP and Taylor RK (2000) Delineation of pilin domains required for bacterial association into microcolonies and intestinal colonization by *Vibrio cholerae*. *Mol. Microbiol.* **35**: 896–910.

Kolappan S, Ng D, Yang G, Harn T and Craig L (2015) Crystal Structure of the Minor

Pilin CofB, the Initiator of CFA/III Pilus Assembly in Enterotoxigenic *Escherichia coli*. *J. Biol. Chem.* **290**: 25805–25818.

Kolappan S, Coureuil M, Yu X, Nassif X, Egelman EH and Craig L (2016) Structure of the *Neisseria meningitidis* type IV pilus. *Nat. Commun.* **7**: 13015.

Korotkov K V and Hol WGJ (2008) Structure of the GspK–GspI–GspJ complex from the enterotoxigenic *Escherichia coli* type 2 secretion system. *Nat. Struct. Mol. Biol.* **15**: 462–468.

Korotkov K V and Sandkvist M (2019) Architecture, Function, and Substrates of the Type II Secretion System. *EcoSal Plus.* **8**. doi: 10.1128/ecosalplus.esp-0034-2018

Krebs SJ and Taylor RK (2011) Protection and attachment of *Vibrio cholerae* mediated by the toxin-coregulated pilus in the infant mouse model. *J. Bacteriol.* **193**: 5260–5270.

Kreth J, Vu H, Zhang Y and Herzberg MC (2009) Characterization of hydrogen peroxide-induced DNA release by *Streptococcus sanguinis* and *Streptococcus gordonii*. *J. Bacteriol.* **191**: 6281–6291.

Kreth J, Giacaman RA, Raghavan R and Merritt J (2017) The road less traveled - defining molecular commensalism with *Streptococcus sanguinis*. *Mol. Oral Microbiol.* **32**: 181–196.

Krissinel E and Henrick K (2007). Inference of macromolecular assemblies from crystalline state.*J. Mol. Biol.* **372**: 774-797.

Laaf D, Bojarova P, Elling L and Křen V (2019) Galectin–Carbohydrate Interactions in Biomedicine and Biotechnology. *Trends Biotechnol.* **37**: 402–415.

Ladant D, Michelson S, Sarfati R, Gilles AM, Predeleanu R and Barzu O (1989) Charac-

terization of the calmodulin-binding and of the catalytic domains of *Bordetella pertussis* adenylate cyclase. *J. Biol. Chem.* **264**: 4015–4020.

Lafaurie G, Castillo D, Romero-Sanchez M and Millan L (2013) Infective Endocarditis Associated to Oral Bacteremia. in *Horizons World Cardiovasc. Res.* 1318–1330.

LaPointe CF and Taylor RK (2000) The Type 4 Prepilin Peptidases Comprise a Novel Family of Aspartic Acid Proteases. *J. Biol. Chem.* **275**: 1502–1510.

Lassak K, Ghosh A and Albers SV (2012) Diversity, assembly and regulation of archaeal type IV pili-like and non-type-IV pili-like surface structures. *Res. Microbiol.* **163**: 630–644.

Lemkul JA and Bevan DR (2011) Characterization of interactions between PilA from *Pseudomonas aeruginosa* strain K and a model membrane. *J. Phys. Chem.* **115**: 8004–8008.

Leonidas DD, Vatzaki EH, Vorum H, Celis JE, Madsen P, Acharya KR (1998) Structural basis for the recognition of carbohydrates by human galectin-7 *Biochem.* **37**: 13930-13940.

Li R, Hoffmeister KM and Falet H (2016) Glycans and the platelet life cycle. *Platelets* **27**: 505–511.

Liesenborghs L, Meyers S, Vanassche T and Verhamme P (2020) Coagulation: At the heart of infective endocarditis. *J. Thromb. Haemost.* **18**: 995–1008.

London J, Allen J. (1990) Purification and characterization of a Bacteroides loeschei adhesin that interacts with procaryotic and eucaryotic cells. *J Bacteriol.* **172**: 2527–2534.

Long CD, Hayes SF, Van Putten JPM, Harvey HA, Apicella MA and Seifert HS (2001) Modulation of gonococcal piliation by regulatable transcription of pilE. *J. Bacteriol.* **183**:

1600–1609.

López-Castilla A, Thomassin J-L, Bardiaux B, Zheng W, Nivaskumar M, Yu X, Nilges M, Egelman EH, Izadi-Pruneyre N and Francetic O (2017) Structure of the calcium-dependent type 2 secretion pseudopilus. *Nat. Microbiol.* **2**: 1686–1695.

Loukachevitch L V., Bensing BA, Yu H, Zeng J, Chen X, Sullam PM and Iverson TM (2016) Structures of the *Streptococcus sanguinis* SrpA Binding Region with Human Sialoglycans Suggest Features of the Physiological Ligand. *Biochemistry* **55**: 5927–5937.

Lowrance JH, Baddour LM and Simpson WA (1990) The role of fibronectin binding in the rat model of experimental endocarditis caused by *Streptococcus sanguis*. *J. Clin. Invest.* **86**: 7–13.

Luo Y, Frey EA, Pfuetzner RA, Creaght AL, Knoechel DG, Haynes CA, Finlay BB and Strynadka NCJ (2000) Crystal structure of enteropathogenic *Escherichia coli* intimin-receptor complex. *Nature* **405**: 1073–1077.

Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A and Lopez R (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* gkac240.

Maeda K, Nagata H, Nonaka A, Kataoka K, Tanaka M and Shizukuishi S (2004) Oral streptococcal glyceraldehyde-3-phosphate dehydrogenase mediates interaction with fimbriae. *Microbes Infect.* **6**: 1163–1170.

Maier B, Potter L, So M, Seifert HS and Sheetz MP (2002) Single pilus motor forces exceed 100 pN. *Proc. Natl. Acad. Sci.* **99**: 16012–16017.

Mancl JM, Black WP, Robinson H, Yang Z and Schubot FD (2016) Crystal Structure of

a Type IV Pilus Assembly ATPase: Insights into the Molecular Mechanism of PilB from *Thermus thermophilus. Structure.* **24**: 1886–1897.

Marco S, Rullo R, Albino A, Masullo M, De Vendittis E and Amato M (2013) The thioredoxin system in the dental caries pathogen *Streptococcus mutans* and the food-industry bacterium *Streptococcus thermophilus. Biochimie.* **95**: 2145–2156.

Marsh PD (2004) Dental Plaque as a Microbial Biofilm. *Caries Res.* **38**: 204–211.

Martini AM, Moricz BS, Ripperger AK, Tran PM, Sharp ME, Forsythe AN, Kulhankova K, Salgado-Pabón W and Jones BD (2020) Association of Novel *Streptococcus sanguinis* Virulence Factors With Pathogenesis in a Native Valve Infective Endocarditis Model. *Front. Microbiol.* **11**: 1–15.

Martini AM, Moricz BS, Woods LJ and Jones BD (2021) Type IV Pili of *Streptococcus sanguinis* Contribute to Pathogenesis in Experimental Infective Endocarditis. *Microbiol. Spectr.* **9**: 1–18.

McCallum M, Tammam S, Little DJ, Robinson H, Koo J, Shah M, Calmettes C, Moraes TF, Burrows LL, Howell PL (2016) PilN binding modulates the structure and binding partners of the *Pseudomonas aeruginosa* type IVa pilus protein PilM. *J. Biol. Chem.* **291**: 11003–11015.

McCallum M, Tammam S, Khan A, Burrows LL and Howell PL (2017) The molecular mechanism of the type IVa pilus motors. *Nat. Commun.* **8**: 15091.

McCallum M, Burrows LL and Howell PL (2019) The Dynamic Structures of the Type IV Pilus. *Microbiol. Spectr.* **7**: 1–12.

Melville S and Craig L (2013) Type IV Pili in Gram-Positive Bacteria. *Microbiol. Mol.*

Biol. Rev. **77**: 323–341.

Merz AJ, So M and Sheetz MP (2000) Pilus retraction powers bacterial twitching motility. *Nature* **407**: 98–102.

Mikaty G, Soyer M, Mairey E, Henry N, Dyer D, Forest KT, Morand P, Guadagnini S, Prevost MC, Nassif X and Dumenil G (2009) Extracellular bacterial pathogen induces host cell surface reorganization to resist shear stress. *PLoS Pathog.* **5**: e1000314.

Miller RM, Tomaras AP, Barker AP, Voelker DR, Chan ED, Vasil AI and Vasil ML (2008) *Pseudomonas aeruginosa* twitching motility-mediated chemotaxis towards phospholipids and fatty acids: Specificity and metabolic requirements. *J. Bacteriol.* **190**: 4038–4049.

Misic AM, Satyshur KA and Forest KT (2010) *P. aeruginosa* PilT Structures with and without Nucleotide Reveal a Dynamic Type IV Pilus Retraction Motor. *J. Mol. Biol.* **400**: 1011–1021.

Modenutti CP, Capurro JIB, Di Lella S and Marti MA (2019) The Structural Biology of Galectin-Ligand Recognition: Current Advances in Modeling Tools, Protein Engineering, and Inhibitor Design. *Front. Chem.* **7** doi: 10.3389/fchem.2019.00823.

Moraes JJ, Stipp RN, Harth-Chu EN, Camargo TM, Hofling JF and Mattos-Graner RO (2014) Two- component system VicRK regulates functions associ- ated with establishment of *Streptococcus sanguinis* in biofilms. *Infect. Immun.* **82**: 4941–4951.

Morand PC, Bille E, Morelle S, Eugène E, Beretti JL, Wolfgang M, Meyer TF, Koome M and Nassif X (2004) Type IV pilus retraction in pathogenic *Neisseria* is regulated by the PilC proteins. *EMBO J.* **23**: 2009–2017.

Morita C, Sumioka R, Nakata M, Okahashi N, Wada S, Yamashiro T, Hayashi M, Hamada

S, Sumitomo T and Kawabata S (2014) Cell Wall-Anchored Nuclease of *Streptococcus sanguinis* Contributes to Escape from Neutrophil Extracellular Trap-Mediated Bacteriocidal Activity. *PLoS One.* **9**: e103125.

Mu R, Anderson D, Merritt J, Wu H and Kreth J (2021) Post-translational modification of *Streptococcus sanguinis* SpxB influences protein solubility and H_2O_2 production. *Mol. Oral Microbiol.* **36**: 267–277.

Ng SYM, Zolghadr B, Driessen AJM, Albers S-V and Jarrell KF (2008) Cell Surface Structures of Archaea. *J. Bacteriol.* **190**: 6039–6047.

Ng SYM, VanDyke DJ, Chaban B, Wu J, Nosaka Y, Aizawa S-I and Jarrell KF (2009) Different Minimal Signal Peptide Lengths Recognized by the Archaeal Prepilin-Like Peptidases FlaK and PibD. *J. Bacteriol.* **191**: 6732–6740.

Ng D, Harn T, Altindal T, Kolappan S, Marles JM, Lala R, Spielman I, Gao Y, Hauke CA, Kovacikova G, Verjee Z, Taylor RK, Biais N and Craig L (2016) The *Vibrio cholerae* Minor Pilin TcpB Initiates Assembly and Retraction of the Toxin-Coregulated Pilus. *PLoS Pathog.* **12**: e1006109.

Nguyen Y, Sugiman-Marangos S, Harvey H, Bell SD, Charlton CL, Junop MS and Burrows LL (2015a) *Pseudomonas aeruginosa* minor pilins prime type IVa pilus assembly and promote surface display of the PilY1 adhesin. *J. Biol. Chem.* **290**: 601–611.

Nguyen Y, Harvey H, Sugiman-Marangos S, Bell SD, Buensuceso RNC, Junop MS and Burrows LL (2015b) Structural and Functional Studies of the *Pseudomonas aeruginosa* Minor Pilin, PilE. *J. Biol. Chem.* **290**: 26856–26865.

Nieto V, Kroken AR, Grosser MR, Smith BE, Metruccio MME, Hagan P, Hallsten ME, Evans DJ and Fleiszig SMJ (2019) Type IV Pili Can Mediate Bacterial Motility within

Epithelial Cells. *MBio* **10**: 1–9.

Nivaskumar M, Santos-Moreno J, Malosse C, Nadeau N, Chamot-Rooke J, Tran Van Nhieu G and Francetic O (2016) Pseudopilin residue E5 is essential for recruitment by the type 2 secretion system assembly platform. *Mol. Microbiol.* **101**: 924–941.

Nunn DN and Lory S (1991) Product of the *Pseudomonas aeruginosa* gene pilD is a prepilin leader peptidase. *Proc. Natl. Acad. Sci.* **88**: 3281–3285.

Okahashi N, Nakata M, Sakurai A, Terao Y, Hoshino T, Yamaguchi M, Isoda R, Sumitomo T, Nakano K, Kawabata S and Ooshima T (2010) Pili of oral *Streptococcus sanguinis* bind to fibronectin and contribute to cell adhesion. *Biochem. Biophys. Res. Commun.* **391**: 1192–1196.

Okahashi N, Nakata M, Terao Y, Isoda R, Sakurai A, Sumitomo T, Yamaguchi M, Kimura RK, Oiki E, Kawabata S and Ooshima T (2011a) Pili of oral *Streptococcus sanguinis* bind to salivary amylase and promote the biofilm formation. *Microb. Pathog.* **50**: 148–154.

Okahashi N, Okinaga T, Sakurai A, Terao Y, Nakata M, Nakashima K, Shintani S, Kawabata S, Ooshima T and Nishihara T (2011b) *Streptococcus sanguinis* induces foam cell formation and cell death of macrophages in association with production of reactive oxygen species. *FEMS Microbiol. Lett.* **323**: 164–170.

Okahashi N, Nakata M, Sumitomo T, Terao Y and Kawabata S (2013) Hydrogen Peroxide Produced by Oral Streptococci Induces Macrophage Cell Death. *PLoS One* **8**: 1–6.

Okahashi N, Sumitomo T, Nakata M, Sakurai A, Kuwata H and Kawabata S (2014) Hydrogen Peroxide Contributes to the Epithelial Cell Death Induced by the Oral Mitis Group of Streptococci. *PLoS One.* **9**: e88136.

Okahashi N, Nakata M, Kuwata H and Kawabata S (2016) *Streptococcus oralis* Induces Lysosomal Impairment of Macrophages via Bacterial Hydrogen Peroxide. *Infect. Immun.* **84**: 2042–2050.

Oki H, Kawahara K, Maruno T, Imai T, Muroga Y, Fukakusa S, Iwashita T, Kobayashi Y, Matsuda S, Kodama T, Iida T, Yoshida T, Ohkubo T and Nakamura S (2018) Interplay of a secreted protein with type IVb pilus for efficient enterotoxigenic *Escherichia coli* colonization. *Proc. Natl. Acad. Sci. U. S. A.* **115**: 7422–7427.

Orans J, Johnson MDL, Coggan KA, Sperlazza JR, Heiniger RW, Wolfgang MC and Redinbo MR (2010) Crystal structure analysis reveals Pseudomonas PilY1 as an essential calcium-dependent regulator of bacterial surface motility. *Proc. Natl. Acad. Sci. U. S. A.* **107**: 1065–1070.

Parge HE, Forest KT, Hickey MJ, Christensen DA, Getzoff ED, and Tainer JA (1995) Structure of the fibre-forming protein pilin at 2.6 A resolution. *Nature* **378**: 32–38.

Park Y, Simionato MR, Sekiya K, Murakami Y, James D, Chen W, Hackett M, Yoshimura F, Demuth DR and Lamont RJ (2005) Short fimbriae of *Porphyromonas gingivalis* and their role in coadhesion with *Streptococcus gordonii*. *Infect. Immun.* **73**: 3983–3989.

Patnaik SK, Potvin B, Carlsson S, Sturm D, Leffler H and Stanley P (2006) Complex N-glycans are the major ligands for galectin-1, -3, and -8 on Chinese hamster ovary cells. *Glycobiology* **16**: 305–317.

Pegden RS, Larson MA, Grant RJ and Morrison M (1998) Adherence of the gram-positive bacterium *Ruminococcus albus* to cellulose and identification of a novel form of cellulose-binding protein which belongs to the Pil family of proteins. *J. Bacteriol.* **180**: 5921–5927.

Pelicic V (2008) Type IV pili: e pluribus unum? *Mol. Microbiol.* **68**: 827–837.

Pelicic V (2019) Monoderm bacteria: the new frontier for type IV pilus biology. *Mol. Microbiol.* **112**: 1674–1683.

Peterson SN, Meissner T, Su AI, Snesrud E, Ong AC, Schork NJ and Bretz WA (2014) Functional expression of dental plaque microbiota. *Front. Cell. Infect. Microbiol.* **4**: 1–13.

Peyrot des Gachons C and Breslin PAS (2016) Salivary Amylase: Digestion and Metabolic Syndrome. *Curr. Diab. Rep. 16*: 102

Plummer C and Douglas CWI (2006) Relationship between the ability of oral streptococci to interact with platelet glycoprotein Ibα and with the salivary low-molecular-weight mucin, MG2. *FEMS Immunol. Med. Microbiol.* **48**: 390–399.

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

Proft T and Baker EN (2009) Pili in Gram-negative and Gram-positive bacteria - Structure, assembly and their role in disease. *Cell. Mol. Life Sci.* **66**: 613–635.

Py B, Loiseau F and Barras F (2001) An inner membrane platform in the type II secretion machinery of Gram-negative bacteria. *EMBO Rep* **2**: 244–248

Que Y-A and Moreillon P (2011) Infective endocarditis. *Nat. Rev. Cardiol.* **8**: 322–336.

Rakotoarivonina H, Jubelin G, Hebraud M, Gaillard-Martinie B, Forano E and Mosoni P (2002) Adhesion to cellulose of the Gram-positive bacterium *Ruminococcus albus* involves type IV pili. *Microbiology* **148**: 1871–1880.

Raynaud C, Sheppard D, Berry J-L, Gurung I and Pelicic V (2021) PilB from *Streptococcus*

sanguinis is a bimodular type IV pilin with a direct role in adhesion. *Proc. Natl. Acad. Sci.* **118**: e2102092118.

Rosan B and Lamont RJ (2000) Dental plaque formation. *Microbes Infect.* **2**: 1599–1607.

Rudel T, Schleurerpflug I and Meyer TF (1995) Neisseria PilC protein identified as type-4 pilus tip-located adhesin. *Nature* **373**: 357-359

Ruffert C (2016) Magnetic Bead—Magic Bullet. *Micromachines* **7** doi:10.3390/mi7020021.

Salleh MZ, Karuppiah V, Snee M, Thistlethwaite A, Levy CW, Knight D and Derrick JP (2019) Structure and properties of a natural competence-associated pilin suggest a unique pilus Tip-associated DNA receptor. *Mbio* **10**.

Salomonsson E, Carlsson MC, Osla V, Hendus-Altenburger R, Kahl-Knutson B, Oberg CT, Sundin A, Nilsson R, Nordberg-Karlsson E, Nilsson UJ, Karlsson A, Rini JM and Leffler H (2010) Mutational tuning of galectin-3 specificity and biological function. *J. Biol. Chem.* **285**: 35079–35091.

Salzer R, D'Imprima E, Gold VAM, Rose I, Drechsler M, Vonck J and Averhoff B (2016) Topology and structure/function correlation of ring-and gate-forming domains in the dynamic secretin complex of *Thermus thermophilus*. *J. Biol. Chem.* **291**: 14448–14456.

Sangermani M, Hug I, Sauter N, Pfohl T and Jenal U (2019) Tad Pili Play a Dynamic Role in Caulobacter crescentus Surface Colonization. *Mbio* **10**: e01237-19.

Satyshur KA, Worzalla GA, Meyer LS, Heiniger EK, Aukema KG, Misic AM and Forest KT (2007) Crystal Structures of the Pilus Retraction Motor PilT Suggest Large Domain Movements and Subunit Cooperation Drive Motility. *Structure* **15**: 363–376.

Sauvonnet N, Vignon G, Pugsley AP and Gounon P (2000) Pilus formation and protein

secretion by the same machinery in *Escherichia coli*. *EMBO J.* **19**: 2221–2228.

Some D, Amartely H, Tsadok A and Lebendiker M (2019) Characterization of Proteins by Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering (SEC-MALS). *J Vis Exp.* **20** doi: 10.3791/59615.

Stinson MW, Alder S and Kumar S (2003) Invasion and Killing of Human Endothelial Cells by Viridans Group Streptococci. *Infect. Immun.* **71**: 2365–2372.

Strom MS and Lory S (1991) Amino acid substitutions in pilin of *Pseudomonas aeruginosa*. *J. Biol. Chem.* **266**: 1656–1664.

Strom MS and Lory S (1992) Kinetics and sequence specificity of processing of prepilin by PilD, the type IV leader peptidase of *Pseudomonas aeruginosa*. *J. Bacteriol.* **174**: 7345–7351.

Strom MS, Nunn DN and Lory S (1993) A single bifunctional enzyme, PilD, catalyzes cleavage and N-methylation of proteins belonging to the type IV pilin family. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 2404–2408.

Studier FW and Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol.* **189**: 113-30.

Sumioka R, Nakata M, Okahashi N, Li Y, Wada S, Yamaguchi M, Sumitomo T, Hayashi M and Kawabata S (2017) *Streptococcus sanguinis* induces neutrophil cell death by production of hydrogen peroxide. *PLoS One* **12**: 1–19.

Swaminathan S, Furey W, Pletcher J and Sax M (1995) Residues defining V beta specificity in staphylococcal enterotoxins. *Nat Struct Biol* **2**: 680-686.

Swaminathan S and Eswaramoorthy S (2000) Structural analysis of the catalytic and

binding sites of *Clostridium botulinum* neurotoxin B. *Nat Struct Biol* **7**: 693-699.

Takamatsu D, Bensing BA, Prakobphol A, Fisher SJ and Sullam PM (2006) Binding of the streptococcal surface glycoproteins GspB and Hsa to human salivary proteins. *Infect. Immun.* **74**: 1933–1940.

Takhar HK, Kemp K, Kim M, Howell PL and Burrows LL (2013) The Platform Protein Is Essential for Type IV Pilus Biogenesis. *J. Biol. Chem.* **288**: 9721–9728.

Tammam S, Sampaleanu LM, Koo J, Manoharan K, Daubaras M, Burrows LL and Howell PL (2013) PilMNOPQ from the *Pseudomonas aeruginosa* type IV pilus system form a transenvelope protein interaction network that interacts with PilA. *J. Bacteriol.* **195**: 2126–2135.

Tanaka H, Ebara S, Otsuka K and Hayashi K (1996) Adsorption of saliva-coated and plain streptococcal cells to the surfaces of hydroxyapatite beads. *Arch. Oral Biol.* **41**: 505–508.

Telford JL, Barocchi MA, Margarit I, Rappuoli R and Grandi G (2006) Pili in Gram-positive pathogens. *Nat. Rev. Microbiol.* **4**: 509–519.

Thomassin JL, Santos Moreno J, Guilvout I, Tran Van Nhieu G and Francetic O (2017) The trans-envelope architecture and function of the type 2 secretion system: new insights raising new questions. *Mol. Microbiol.* **105**: 211–226.

Tomich M, Planet PJ and Figurski DH (2007) The tad locus: Postcards from the widespread colonization island. *Nat. Rev. Microbiol.* **5**: 363–375.

Toria NH and Joshua-Tor L (2006) Strategies for protein coexpression in *Escherichia coli*. *Nat Methods.* **3**: 55-64.

Topin J, Arnaud J, Sarkar A, Audfray A, Gillon E, Perez S, Jamet H, Varrot A, Imberty A and Thomas A (2013) Deciphering the glycan preference of bacterial lectins by glycan array and molecular docking with validation by microcalorimetry and crystallography. *PLoS One* **8**: e71149.

Treuner-Lange A, Chang Y-W, Glatter T, Herfurth M, Lindow S, Chreifi G, Jensen GJ and Sogaard-Andersen L (2020) PilY1 and minor pilins form a complex priming the type IVa pilus in *Myxococcus xanthus*. *Nat. Commun.* **11**: 5054.

Truper H and Clari LD (1997) Taxonomic note: necessary corrections of specific epithets formed as substantives (nouns) "in apposition". *Int. J. Syst. Bacteriol.* **47**: 908–909.

Valdebenito B, Tullume-Vergara PO, Gonzalez W, Kreth J and Giacaman RA (2018) In silico analysis of the competition between *Streptococcus sanguinis* and *Streptococcus mutans* in the dental biofilm. *Mol. Oral Microbiol.* **33**: 168–180.

Vasta GR (2009) Roles of galectins in infection. *Nat. Rev. Microbiol.* **7**: 424–438.

Velazquez-Campoy A, Ohtaka H, Nezami A, Muzammil S and Freire E (2004) Isothermal Titration Calorimetry. *Curr. Protoc. Cell Biol.* **23**: 1–24.

Virion Z, Doly S, Saha K, Lambert M, Guillonneau F, Bied C, Duke RM, Rudd PM, Robbe-Masselot C, Nassif X, Coureuil M and Marullo S (2019) Sialic acid mediated mechanical activation of $\beta 2$ adrenergic receptors by bacterial pili. *Nat. Commun.* **10**: 4752.

Voisin S, Houliston RS, Kelly J, Brisson JR, Watson D, Bardy SL, Jarrell KF and Logan SM (2005) Identification and characterization of the unique N-linked glycan common to the flagellins and S-layer glycoprotein of *Methanococcus voltae*. *J. Biol. Chem.* **280**: 16586–16593.

von Heijne G and Gavel Y (1988) Topogenic signals in integral membrane proteins. *Eur. J. Biochem.* **174**: 671–678.

Wang F, Coureuil M, Osinski T, Orlova A, Altindal T, Gesbert G, Nassif X, Egelman EH and Craig L (2017) Cryoelectron Microscopy Reconstructions of the *Pseudomonas aeruginosa* and *Neisseria gonorrhoeae* Type IV Pili at Sub-nanometer Resolution. *Structure* **25**: 1423-1435.e4.

Werdan K, Dietz S, Loffler B, Niemann S, Bushnaq H, Silber RE, Peters G and Müller-Werdan U (2014) Mechanisms of infective endocarditis: Pathogen-host interaction and risk states. *Nat. Rev. Cardiol.* **11**: 35–50.

Werneburg GT and Thanassi DG (2018) Pili Assembled by the Chaperone/Usher Pathway in *Escherichia coli* and *Salmonella*. *EcoSal Plus.* **8**: 139–148.

White JC and Niven CF (1946) Streptococcus S.B.E.: A Streptococcus Associated with Subacute Bacterial Endocarditis. *J. Bacteriol.* **51**: 717–722.

Winther-Larsen HC, Hegge FT, Wolfgang M, Hayes SF, Van Putten JPM and Koomey M (2001) *Neisseria gonorrhoeae* PilV, a type IV pilus-associated protein essential to human epithelial cell adherence. *Proc. Natl. Acad. Sci. U. S. A.* **98**: 15276–15281.

Winther-Larsen HC, Wolfgang M, Dunham S, Van Putten JPMM, Dorward D, Lovold C, Aas FE and Koomey M (2005) A conserved set of pilin-like molecules controls type IV pilus dynamics and organelle-associated functions in *Neisseria gonorrhoeae*. *Mol. Microbiol.* **56**: 903–917.

Wolfgang M, Park HS, Hayes SF, van Putten JP and Koomey M (1998) Suppression of an absolute defect in type IV pilus biogenesis by loss-of-function mutations in pilT, a twitching motility gene in *Neisseria gonorrhoeae*. *Proc. Natl. Acad. Sci. U. S. A.* **95**:

14973–8.

Wolfgang M, Van Putten JPM, Hayes SF and Koomey M (1999) The comP locus of *Neisseria gonorrhoeae* encodes a type IV prepilin that is dispensable for pilus biogenesis but essential for natural transformation. *Mol. Microbiol.* **31**: 1345–1357.

Wolfgang M (2000) Components and dynamics of fiber formation define a ubiquitous biogenesis pathway for bacterial pili. *EMBO J.* **19**: 6408–6418.

Wong E, Vaaje-Kolstad G, Ghosh A, Hurtado-Guerrero R, Konarev P V., Ibrahim AFM, Svergun DI, Eijsink VGH, Chatterjee NS and van Aalten DMF (2012) The *Vibrio cholerae* colonization factor GbpA possesses a modular structure that governs binding to different host surfaces. *PLoS Pathog.* **8**: 1–12.

Xu P, Alves JM, Kitten T, Brown A, Chen Z, Ozaki LS, Manque P, Ge X, Serrano MG, Puiu D, Hendricks S, Wang Y, Chaplin MD, Akan D, Paik S, Peterson DL, Macrina FL and Buck GA (2007) Genome of the Opportunistic Pathogen *Streptococcus sanguinis*. *J. Bacteriol.* **189**: 3166–3175.

Xu P, Ge X, Chen L, Wang X, Dou Y, Xu JZ, Patel JR, Stone V, Trinh M, Evans K, Kitten T, Bonchev D and Buck GA (2011) Genome-wide essential gene identification in *Streptococcus sanguinis*. *Sci. Rep.* **1**: srep00125.

Xu Y, Itzek A and Kreth J (2014) Comparison of genes required for H_2O_2 resistance in *Streptococcus gordonii* and *Streptococcus sanguinis*. *Microbiology* **160**: 2627–2638.

Yamamoto Y, Poole LB, Hantgan RR and Kamio Y (2002) An iron-binding protein, Dpr, from *Streptococcus mutans* prevents iron-dependent hydroxyl radical formation *in vitro*. *J. Bacteriol.* **184**: 2931–2939.

Yanez ME, Korotkov KV, Abendroth J and Hol WGJ (2008a) The Crystal Structure of a Binary Complex of two Pseudopilins: EpsI and EpsJ from the Type 2 Secretion System of *Vibrio vulnificus*. *J. Mol. Biol.* **375**: 471–486.

Yanez ME, Korotkov KV, Abendroth J and Hol WGJ (2008b) Structure of the minor pseudopilin EpsH from the type 2 secretion system of *Vibrio cholerae*. *J. Mol. Biol.* **377**: 91-103.

Yeaman MR (2014) Platelets: At the nexus of antimicrobial defence. *Nat. Rev. Microbiol.* **12**: 426–437.

Yoshida Y, Konno H, Nagano K, Abiko Y, Nakamura Y, Tanaka Y and Yoshimura F (2014) The influence of a glucosyltransferase, encoded by gtfP , on biofilm formation by *Streptococcus sanguinis* in a dual-species model. *APMIS* **122**: 951–960.

Zelensky AN and Gready JE (2005) The C-type lectin-like domain superfamily. *FEBS J.* **272**: 6179–6217.

Zhang Y, Faucher F, Zhang W, Wang S, Neville N, Poole K, Zheng J and Jia Z (2018) Structure-guided disruption of the pseudopilus tip complex inhibits the Type II secretion in *Pseudomonas aeruginosa*. *PLOS Pathog.* **14**: e1007343.

Zhu B, Macleod LC, Kitten T and Xu P (2018) *Streptococcus sanguinis* biofilm formation and interaction with oral pathogens. *Future Microbiol.* **13**: 915–932.

Zhu B, Green SP, Ge X, Puccio T, Nadhem H, Ge H, Bao L, Kitten T and Xu P (2021) Genome-wide identification of *Streptococcus sanguinis* fitness genes in human serum and discovery of potential selective drug targets. *Mol. Microbiol.* **115**: 658–671.

Zolghadr B, Weber S, Szabó Z, Driessen AJM and Albers SV (2007) Identification of a

system required for the functional surface localization of sugar binding proteins with class III signal peptides in *Sulfolobus solfataricus. Mol. Microbiol.* **64**: 795–806.

Zolghadr B, Klingl A, Rachel R, Driessen AJM and Albers SV (2011) The bindosome is a structural component of the Sulfolobus solfataricus cell envelope. *Extremophiles* **15**: 235–244.

Appendix

Protein Purification

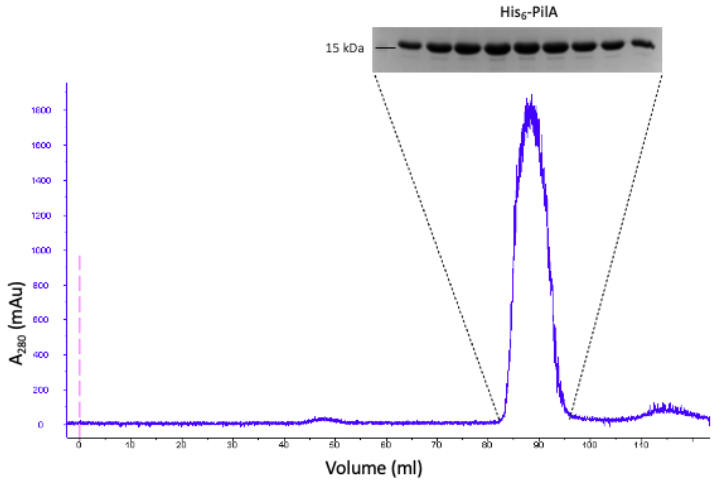


Figure A.1: Protein purification of His-PilA. SEC and SDS-PAGE analysis of His-PilA peak fractions.

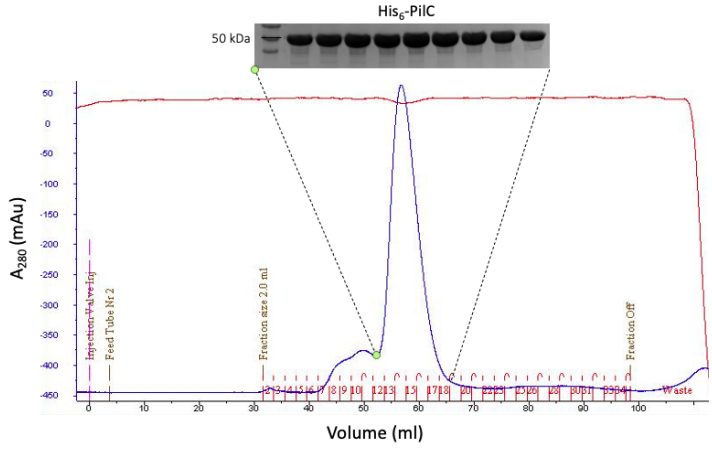


Figure A.2: Protein purification of His-PilC. SEC and SDS-PAGE analysis of His-PilC peak fractions.

PISA Analysis

Hydrogen bonds XML

No disulfide bonds found
No covalent bonds found
No salt bridges found

##	Structure 1	Dist. [Å]	Structure 2
1	A:TRP 36[NE1]	3.53	D:THR 15[O]
2	A:HIS 117[NE2]	2.70	D:TYR 21[OH]
3	A:TYR 26[OH]	3.59	D:THR 29[O]
4	A:TYR 26[OH]	3.50	D:SER 30[OG]
5	A:TYR 26[OH]	2.85	D:GLN 33[OE1]
6	A:LYS 25[NZ]	2.65	D:TYR 115[OH]
7	A:TYR 21[OH]	2.63	D:GLU 131[OE1]
8	A:LEU 135[N]	2.87	D:GLU 131[OE2]
9	A:THR 15[O]	3.62	D:TRP 36[NE1]
10	A:TYR 21[OH]	2.80	D:HIS 117[NE2]
11	A:LEU 40[O]	2.88	D:HIS 7[NE2]
12	A:TYR 115[OH]	2.80	D:LYS 25[NZ]
13	A:GLU 131[OE1]	2.69	D:LEU 135[N]
14	A:GLU 131[OE2]	2.66	D:TYR 21[OH]

Figure A.3: PISA analysis of the PilA homodimer interface. Structure 1 and Structure 2 refer to the two PilA monomers which make up the PilA homodimer. The PISA analysis was performed on the crystal structure of His-PilA solved by Dr Berry.

Hydrogen bonds XML

##	Structure 1	Dist. [Å]	Structure 2
1	B:GLN 34[NE2]	3.02	C:VAL 24[O]
2	B:LYS 56[NZ]	2.64	C:GLU 147[OE1]
3	B:TYR 70[OH]	2.41	C:SER 65[OG]
4	B:GLN 71[N]	2.86	C:LYS 85[O]
5	B:GLY 73[N]	3.65	C:PRO 83[O]
6	B:ASN 74[N]	3.54	C:GLU 82[OE1]
7	B:LYS 95[N]	2.93	C:TYR 35[OH]
8	B:PHE 99[N]	2.65	C:SER 46[OG]
9	B:ALA 101[N]	3.53	C:GLN 49[OE1]
10	B:ILE 44[O]	3.75	C:TYR 42[OH]
11	B:GLN 45[OE1]	3.69	C:LYS 41[NZ]
12	B:GLU 49[OE1]	3.12	C:LYS 41[NZ]
13	B:GLU 66[OE2]	2.91	C:HIS 88[NE2]
14	B:THR 69[O]	2.45	C:SER 86[OG]
15	B:TYR 70[OH]	3.66	C:VAL 66[N]
16	B:GLN 71[O]	2.87	C:LYS 85[N]
17	B:LYS 95[O]	2.94	C:TYR 35[OH]
18	B:ILE 104[O]	3.61	C:TRP 52[NE1]
19	B:GLU 105[OE1]	3.87	C:GLN 49[NE2]
20	B:THR 106[O]	3.06	C:ILE 144[N]

Salt bridges XML

No disulfide bonds found
No covalent bonds found

##	Structure 1	Dist. [Å]	Structure 2
1	B:LYS 56[NZ]	2.64	C:GLU 147[OE1]
2	B:LYS 56[NZ]	3.00	C:GLU 147[OE2]
3	B:GLU 49[OE1]	3.12	C:LYS 41[NZ]
4	B:GLU 66[OE2]	2.91	C:HIS 88[NE2]

Figure A.4: PISA analysis of the PilAC heterodimer interface. Structure 1 refers to PilC, and Structure 2 refers to PilA. The PISA analysis was performed on the AlphaFold model of the interaction between the PilA and PilC pilin domains.