PilB from Streptococcus sanguinis is a bimodular type IV pilin with a direct role in adhesion

C. Raynaud, D. Sheppard, J.L. Berry, I. Gurung, V. Pelicic

PilB from *Streptococcus sanguinis* is a bimodular type IV pilin with a direct role in adhesion

Claire Raynaud[a,1], Devon Sheppard[a,1], Jamie-Lee Berry[a], Ishwori Gurung[a], Vladimir Pelicic[a,b,2]

[a]MRC Centre for Molecular Bacteriology and Infection, Imperial College London, SW7 2AZ London, United Kingdom

[b]Laboratoire de Chimie Bactérienne, Aix-Marseille Université-CNRS (UMR 7283), Institut de Microbiologie de la Méditerranée, 13009 Marseille, France

[1]C.R. and D.S. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: vladimir.pelicic@inserm.fr

Abstract

Type IV pili (T4P) are functionally versatile filamentous nanomachines, nearly ubiquitous in prokaryotes. They are predominantly polymers of one major pilin, but also contain minor pilins whose functions are often poorly defined, and likely to be diverse. Here, we show that the minor pilin PilB from the T4P of *S. sanguinis* displays an unusual bimodular 3D structure, with a bulky von Willebrand factor A-like (vWA) module "grafted" onto a small pilin module via a short loop. Structural modelling suggests that PilB is only compatible with a localisation at the tip of T4P. By performing a detailed functional analysis, we found that (i) the vWA module contains a canonical metal ion-dependent adhesion site (MIDAS), preferentially binding Mg^{2+} and Mn^{2+}, (ii) abolishing metal-binding has no impact on the structure of PilB or piliation, (iii) metal-binding is important for *S. sanguinis* T4P-mediated twitching motility and adhesion to eukaryotic cells, and (iv) the vWA module shows an intrinsic binding ability to several host proteins. These findings reveal an elegant, yet simple, evolutionary tinkering strategy to increase T4P functional versatility, by grafting a functional module onto a pilin for presentation by the filaments. This strategy appears to have been extensively used by bacteria, in which modular pilins are widespread and exhibit an astonishing variety of architectures.

Significance

Type IV pili (T4P) are functionally versatile filaments widespread in prokaryotes, composed of type IV pilins and assembled by conserved multi-protein machineries. It remains unclear how such rather simple filaments can be so versatile. Our structure/function analysis of PilB, a minor pilin of *S. sanguinis* T4P, offers an elegant explanation for this paradox. We show that PilB is a modular pilin with a bulky module "grafted" onto a small pilin module, which directly mediates adhesion of *S. sanguinis* to host cells/proteins. This evolutionary tinkering strategy appears to be prevalent in bacteria, since a global analysis reveals that modular pilins are widespread and exhibit an astonishing variety of architectures.

Introduction

Type IV pili (T4P) are functionally versatile filaments widespread in prokaryotes, implicated in a variety of functions such as adhesion, twitching motility, DNA uptake *etc* (1). T4P are helical polymers consisting of type IV pilins, usually one major pilin and several minor (low abundance) ones, assembled by conserved multi-protein machineries. These defining features are shared by a superfamily of filamentous nanomachines known as type IV filaments (T4F) (1), ubiquitous in prokaryotes (2).

T4P have been intensively studied for decades in diderm bacteria because they play a central role in pathogenesis in important human pathogens (3). The following global picture of T4P biology has emerged from these studies. The pilus subunits, type IV pilins, are characterised by a short N-terminal sequence motif known as class III signal peptide, which consists of a hydrophilic leader peptide ending with a small residue (Gly or Ala), followed by a tract of 21 predominantly hydrophobic residues (4). This tract constitutes the N-terminal segment (α1N) of an α-helix (α1) of ~50 residues, which is the universally conserved structural feature in type IV pilins. Usually, the α1N helix protrudes from a globular head most often consisting of a β-sheet composed of several antiparallel β-strands, which gives pilins their characteristic "lollipop" shape (4). The hydrophilic leader peptide is then processed by a dedicated prepilin peptidase (5) after pilin translocation across the cytoplasmic membrane (CM) by the general secretory pathway (6, 7). Processed pilins remain embedded in the CM via their α1N, generating a pool of subunits ready for polymerisation. Filament assembly, which occurs from tip to base, is mediated at the CM by a complex multi-protein machinery (10-20 components) (1), centred on an integral membrane platform protein and a cytoplasmic extension ATPase (8). Recent cryo-EM structures have revealed that T4P are right-handed helical polymers where pilins are held together by extensive interactions between their α1N helices, which are partially melted and run approximately parallel to each other within the filament core (9, 10). One of the properties of T4P key for their functional versatility is their ability to retract, which has

been best characterised for T4aP (where "a" denotes the subtype). In T4aP, retraction results from rapid filament depolymerisation powered by the cytoplasmic retraction ATPase PilT (11), which generates important tensile forces (12, 13).

Studying T4P in monoderm bacteria represents a promising alternative research avenue (14). *Streptococcus sanguinis*, a commensal of the oral cavity that commonly causes life-threatening infective endocarditis (IE), has emerged as a monoderm model for deciphering T4P biology (15). Our comprehensive functional analysis of *S. sanguinis* T4P (16) revealed that they are canonical T4aP. Indeed, filaments are (i) assembled by a multi-protein machinery similar to diderm T4aP species, but simpler with only ten components, (ii) retracted by a PilT ATPase, generating tensile forces similar to diderm species, and (iii) powering intense twitching motility, leading to spreading zones around bacteria growing on plates, visible by the naked eye. Subsequently, we performed a global biochemical and structural analysis of *S. sanguinis* T4P (17), showing that (i) they are hetero-polymers composed of two major pilins, PilE1 and PilE2, rather than one as usually seen, (ii) the major pilins display classical type IV pilin 3D structure, and (iii) the filaments contain a low abundance of three minor pilins (PilA, PilB, and PilC), which are required for piliation.

The present study was prompted by a perplexing observation, *i.e.*, the minor pilin PilB harbours a protein domain that has been extensively studied in eukaryotic proteins where it mediates adhesion to a variety of protein ligands (18). This suggested that PilB might be an adhesin, promoting T4P-mediated adhesion of *S. sanguinis* to host cells and proteins. Therefore, since both the molecular mechanisms of T4P-mediated adhesion and the exact role of minor pilins in T4P biology remain incompletely understood (1), we performed a structure/function analysis of PilB, which is reported here. This uncovered a widespread strategy for minor pilins to enhance the functional properties of T4P.

Results

PilB displays a modular pilin architecture

PilB, one of the three minor pilins in *S. sanguinis* T4P (17), exhibits a canonical N-terminal class III signal peptide, the defining feature of type IV pilins (4). This sequence motif consists of a seven-residue leader peptide composed predominantly of hydrophilic and neutral amino acids (aa), ending with a Gly (Fig. 1A). This leader peptide, which is processed by the prepilin peptidase PilD (17), is followed by a stretch of 21 predominantly hydrophobic aa, except for a negatively charged Glu in position 5 (Fig. 1A). Processed PilB is unusually large for a pilin, with a predicted molecular mass of 50.5 kDa (Fig. 1B). For comparison, the two major pilins of *S. sanguinis* T4P, PilE1 and PilE2 (16), have typical pilin sizes of 14.7 and 14.1 kDa, respectively (Fig. 1B). The larger size of PilB is due to the presence of a C-terminal domain (Fig. 1B) readily detectable by bioinformatics, which belongs to the von Willebrand factor A-like domain superfamily (InterPro entry IPR036465). We will refer to this domain as vWA. The prototypical vWA domain is found in the von Willebrand factor (vWF), a human blood protein required for haemostasis (19), the physiological process that prevents/stops bleeding. vWA domains, which are found in more than 300,000 proteins in the three domains of life, have been extensively studied in eukaryotic proteins where they mediate adhesion to a variety of protein ligands (18). They have been much less studied in bacteria. Of note, the vWA domain in PilB is predicted to contain a metal coordination site known as MIDAS, for metal ion-dependent adhesion site (20) (Fig. 1A), which is important for ligand-binding in several eukaryotic vWA-containing proteins (21).

The above type IV pilin architecture is unusual for two reasons. First, in contrast to classical pilins that consist only of a pilin module (4), defined by a short N-terminal IPR012902 motif (Fig. 1B), PilB apparently has an additional module. Second, the extra C-terminal module in PilB corresponds to a well-defined functional domain not

specific to T4P biology, which has not been previously seen in pilins. Specifically, the second module – vWA – is often associated with adhesion to protein ligands (18, 21). This is what we call a modular architecture, and why we refer to PilB as a modular pilin.

Taken together, these findings suggest that PilB is a modular pilin in which a functional module has been grafted during evolution onto a pilin moiety, in order to promote T4P-mediated adhesion of *S. sanguinis* to protein ligands.

Crystal structure of PilB reveals a bimodular pilin in which a small type IV pilin module is linked to a bulky vWA module via a short loop

High-resolution structural information was required to confirm that PilB is composed of two modules, but also to understand how modular pilins are polymerised in the filaments and how they modulate T4P functionality. We therefore endeavoured to solve the 3D structure of PilB by X-ray crystallography. To facilitate protein purification, we used a synthetic gene codon-optimised for expression in *Escherichia coli*, and produced a recombinant PilB protein in which the N-terminal 35 aa (encompassing the leader peptide and hydrophobic α1N) (Fig. 1A) were replaced by a hexahistidine tag (6His) (17). This is a commonly used approach in the field to promote protein solubility since the truncation of α1N has minimal structural impact on the rest of the protein (22). The resulting 48.4 kDa 6His-PilB protein was soluble and could be purified using a combination of affinity and gel-filtration chromatography. The protein readily crystallised in multiple conditions, and after optimising the best diffracting crystals, we collected a complete dataset on crystals forming in the space group $P6_1$ (Table 1). After phase determination, done using crystals produced in the presence of seleno-methionine (SeMet), we solved a high-resolution structure (2.26 Å) of native 6His-PilB. As can be seen in Fig. 2A, this structure reveals a clear bimodular architecture with a small pilin moiety (highlighted in blue) linked, via a short nine-residue loop (grey), to a bulky vWA moiety (in red).

While bioinformatics could only predict that the extreme N-terminus of PilB corresponds to a class III signal peptide motif, our structure reveals that the first 180 residues of processed PilB clearly display a type IV pilin fold (4) and thus indeed correspond to a pilin module (Fig. 2B). The pilin module exhibits a long N-terminal α-helix packed against, not one β-sheet as usual, but two consecutive β-sheets consisting of six and three β-strands respectively, which together form the globular head of the pilin. The 432918 topology of the first β-sheet, *i.e.*, the order of the β-strands (Fig. 2B), is unusual since the β-strands are not contiguous along the protein sequence. Moreover, the last portion of this β-sheet forms a Ψ-loop (23) in which two antiparallel strands (β8 and β9) are linked via β1 in between, connected to both of them by hydrogen bonds. This motif occurs rarely in proteins (23).

As for the vWA module (Fig. 2C), the structure strengthens the bioinformatic predictions. The vWA moiety of PilB adopts a canonical vWA fold (20, 24), with a central β-sheet (composed of five parallel and one antiparallel β-strands) surrounded on both sides by a series of α-helices (Fig. 2C). Consequently, the vWA module of PilB shows high structural similarity to many vWA-containing proteins with which it shares little sequence identity. For example, the vWA module of PilB is very similar to the third vWA domain of human vWF (25) (Fig. S1), with a root mean square deviation (RMSD) of 1.72 Å when the C_α atoms of the two structures are superposed. As in eukaryotic vWA-containing proteins (20, 24), PilB exhibits a MIDAS located on top of the central β-sheet (Fig. 2C). However, in contrast to these proteins, the MIDAS in PilB is flanked by two protruding "arms", which is reminiscent of the RrgA adhesin from *Streptococcus pneumoniae* (26). While the first arm is perhaps unremarkable, the second folds into a four-stranded β-sheet (Fig. 2C). The MIDAS motif in PilB, which is formed by residues conserved in vWA-containing proteins, non-contiguous in the sequence (Fig. 1A) but in close proximity in the 3D structure (Fig. 2C), is functional since it coordinates a metal ion in the crystal. We have modelled the metal as Mg^{+2} because of its abundance in the growth medium and the high affinity of PilB for it (see below). The Ser_{206}, Ser_{208},

182 Thr_{291} and Asp_{319} residues in the MIDAS motif (20) of PilB form direct hydrogen bonds
183 with the metal through oxygen atoms (Fig. 2C), while two additional coordination sites
184 are provided by water molecules.

185 An important biological implication of the PilB structure is that modular pilins,
186 despite their large size, are likely to be polymerised into T4P in the same way as
187 classical pilins (4), *i.e.*, via their N-terminal pilin module. We therefore tested by
188 structural modelling whether PilB could pack into filaments. First, we produced a full-
189 length 3D structural model of PilB including the missing α1N (Fig. S2), which was
190 absent in the recombinant protein that we purified. Since a portion of α1N in major
191 pilins is melted during filament assembly, as observed in several T4aP cryo-EM
192 structures (9, 10), the α1N of PilB was modelled with a melted segment. This is
193 consistent with the presence of the helix-breaking Gly residue in position 21 of α1N
194 (Fig. 1A). Then, we fitted this full-length PilB into a previously generated model of *S.*
195 *sanguinis* T4P, a right-handed helical heteropolymer where major pilins PilE1/PilE2
196 are held together by interactions between their α1N helices (Fig. 3A), which was based
197 on the cryo-EM structure of *Neisseria meningitidis* T4P (9). Despite its unusual modular
198 structure, PilB can be readily modelled into T4P, its pilin module establishing extensive
199 hydrophobic interactions via its α1N with the α1N of neighbouring major pilins (Fig.
200 3A). This suggests that PilB will assemble into filaments in the same way as classical
201 pilins (9, 10). However, PilB can only be accommodated at the tip of the filaments
202 because the bulky vWA module sits on top of the pilin module in the PilB structure, and
203 essentially prevents other pilin subunits from being modelled above it (Fig. 3B).
204 Accordingly, when PilB is modelled in the body of the filament (Fig. S3A), it exhibits
205 important steric clashes with neighbouring major pilins (Fig. S3B).

206 Together, these structural findings show that PilB is a bimodular protein
207 composed of two clearly distinct structural modules. The pilin module adopts a
208 canonical type IV pilin fold (4), explaining how modular pilins are polymerised into T4P,
209 most probably at their tip. The second module, which is linked to the end of the pilin

module via a short loop, adopts a vWA fold (20, 24) with a clearly defined MIDAS that coordinates a metal. Since the vWA motif in many eukaryotic proteins is involved in adhesion to protein ligands (18, 21), our structure strengthens our working hypothesis that PilB might be an adhesin.

Functional analysis of the MIDAS in PilB reveals that metal binding, although structurally dispensable, is important for T4P functionality

Our PilB structure revealed that Mg^{2+}, despite not being added during crystallisation, is bound by the MIDAS. In eukaryotic proteins, the MIDAS sometimes coordinates Mn^{2+} as well (20, 24). We therefore tested the metal binding specificity of the MIDAS in PilB using ThermoFluor. This fluorescent-based method, which measures changes in thermal denaturation temperature, is a commonly used approach for quantifying protein-ligand interactions (27). We determined the affinity of purified PilB for Mg^{2+}, Mn^{2+} and Ca^{2+} (Fig. 4A). While no binding was detected to Ca^{2+}, we found that PilB binds Mg^{2+} and Mn^{2+} efficiently, in the micromolar range, with estimated Kd of 70 and 54 µM, respectively. To confirm that metal binding involves the MIDAS motif, we produced $PilB_{D319A}$ in which the key MIDAS residue Asp_{319} (Fig. 2C) was changed into an Ala by site-directed mutagenesis. Binding assays performed with $PilB_{D319A}$ showed that changing this single residue abolishes the metal-binding ability of PilB for both Mg^{2+} and Mn^{2+} (Fig. 4B). These findings show that the MIDAS in PilB is functional and preferentially binds Mg^{2+} and Mn^{2+}.

Next, to determine whether metal presence/absence might impact the 3D structure of PilB, we solved the structure of $PilB_{D319A}$ by X-ray crystallography. The $PilB_{D319A}$ protein readily crystallised in the same condition as the wild-type (WT) protein. We collected a complete dataset on crystals diffracting to a resolution of 3 Å (Table 1) and solved the structure of $PilB_{D319A}$ (Fig. 5A) by molecular replacement. The structure of $PilB_{D319A}$ (Fig. 5A), solved by molecular replacement, clearly shows that no metal is occupying the mutated MIDAS pocket on top of the central β-sheet (Fig. 5B), which is

consistent with metal-binding assays. Although $PilB_{D319A}$ resolution is significantly lower than WT, it nevertheless allows for meaningful structural comparison. When the vWA modules were compared, we found that they are essentially identical, the C_α atoms superposing onto each other (Fig. 5C) with an RMSD of merely 0.45 Å, including the two arms flanking the MIDAS pocket. This shows that metal-binding by MIDAS has no detectable structural impact on PilB.

Next, we explored whether MIDAS-mediated metal-binding by PilB is important for piliation and/or T4P-powered twitching motility, both of which were previously shown to be abolished in a *ΔpilB* mutant (16). We therefore constructed an unmarked *S. sanguinis* mutant in which the endogenous *pilB* gene was altered by site-directed mutagenesis to produce $PilB_{D319A}$. We first tested whether the $pilB_{D319A}$ mutant retains the ability to assemble T4P using filament purification (16). As can be seen in Fig. 6A, in which purified T4P were separated by SDS-PAGE and stained with Coomassie blue, the $pilB_{D319A}$ mutant is piliated. This is evidenced by the presence of the two bands corresponding to major pilins PilE1 and PilE2, which are absent in a non-piliated *ΔpilD* control (Fig. 6A). Moreover, the amount of T4P that can be purified from the $pilB_{D319A}$ mutant and WT strain appear comparable. We then tested whether the pili in the $pilB_{D319A}$ mutant are able to mediate twitching motility (16). For the WT strain, twitching motility is evidenced by spreading zones around bacteria grown on agar (Fig. 6B). Spreading zones were absent for the $pilB_{D319A}$ mutant (Fig. 6B). This shows that the MIDAS-mediated metal-binding ability of PilB, while dispensable for piliation, is important for T4P-mediated twitching motility.

Together, these findings show that the MIDAS in PilB is a functional metal-binding site, dispensable for piliation and protein folding, but essential for T4P functionality.

T4P-mediated adhesion to eukaryotic cells requires PilB, which specifically binds several human proteins

Since vWA is involved in adhesion in many eukaryotic proteins (18, 21), our original hypothesis was that PilB might mediate *S. sanguinis* adhesion to host cells and/or proteins, which we aimed to test next. First, we determined whether T4P might be involved in its well-known ability of *S. sanguinis* to adhere to host cells (28). After testing a few eukaryotic cell lines, we opted for CHO cells because the WT strain adheres very efficiently to them. When CHO cells were infected by *S. sanguinis* at a multiplicity of infection (MOI) of 10, 31.6 ± 9.1 % of the bacterial inoculum adhered to the cells (Fig. 6C). In contrast, a non-piliated *ΔpilD* mutant showed a significantly reduced adhesion, with an 18-fold decrease relative to the WT (Fig. 6C). Next, we tested our original assumption that PilB might be an adhesin, by quantifying the adhesion of the $pilB_{D319A}$ mutant. As can be seen in Fig. 6C, although the $pilB_{D319A}$ mutant is piliated, its adhesion to CHO cells is dramatically impaired, with a 33-fold decrease when compared to the WT. These findings show that *S. sanguinis* T4P are multi-functional filaments important for adhesion to eukaryotic cells, and that PilB plays an important role.

Since the vWA domain in multiple eukaryotic proteins has been shown to mediate cell-extracellular matrix (ECM) interactions (21), we reasoned that PilB might recognise similar ligands because it exhibits a canonical vWA module (Fig. S1). We tested this hypothesis by performing binding assays with purified PilB using enzyme-linked immunosorbent assay (ELISA). In brief, we coated 96-well plates with selected putative ligands, added serial dilutions of purified 6His-PilB, and detected binding using an anti-6His antibody. We tested binding to fibrinogen and the ECM proteins fibronectin, elastin and laminin. While PilB exhibits no binding to BSA that was used as a negative control (Fig. 7A), we observed dose-dependent binding to fibronectin and fibrinogen, but not to the other ECM proteins that were tested (elastin and laminin). Specific binding to fibronectin and fibrinogen was in the high nanomolar range, with calculated Kd of 865 and 494 nM, respectively (Fig. 7A). Under these *in vitro* experimental conditions, metal-coordination by the MIDAS is dispensable for binding

to fibronectin and fibrinogen since $PilB_{D319A}$ binds these ligands as well as PilB (Fig. S4). Finally, to confirm the prediction that binding of PilB to the above ligands is mediated by its vWA module, we produced and purified $PilB_{vWA}$ corresponding only to the vWA module (see Fig. 1A). We found that $PilB_{vWA}$ binds to fibronectin and fibrinogen (Fig. 7B), with calculated Kd of 337 and 997 nM, respectively, which were comparable to PilB. These findings confirm that the adhesive ability of PilB is due to its vWA module.

Taken together, these findings show that *S. sanguinis* T4P are multi-functional filaments mediating adhesion to eukaryotic cells, and that PilB is a *bona fide* adhesin using its vWA module to bind several human protein ligands it shares with eukaryotic vWA-containing proteins.

Pilins with modular architectures are widespread in bacteria

PilB orthologs are ubiquitous in *S. sanguinis*, which also produces a second modular pilin PilC (17), where the extra module belongs to the concanavalin A-like lectin/glucanase domain superfamily (IPR013320) (Fig. 1B). We wondered how widespread and diverse modular pilins might be. We therefore searched the InterPro database (29) for all the proteins with an N-terminal IPR012902 domain, which also contain an extra domain not specific to T4P biology. This showed that modular pilins are (i) widespread with more than 1,200 proteins displaying such architecture (Supplementary dataset 1), (ii) present both in monoderm and diderm species, and (iii) highly diverse, with as many as 264 different architectures detected. Although a bimodular architecture is the most prevalent, there are modular pilins with multiple additional domains, the most extreme case being an 860-residue protein from *Candidatus* Falkowbacteria, with 12 copies of the IPR013211 motif of unknown function (Supplementary dataset 1). A closer inspection of the 15 most frequent modular pilin architectures offers a glimpse of their diversity (Fig. 8). While in many of these proteins the extra domain has no clear function (IPR007001, IPR011871,

IPR026906, PF05345, IPR006860, IPR003961, IPR021556), for others a function can be predicted. These functions include (i) binding to carbohydrates via PF13385 (that overlaps with the IPR013320 lectin domain superfamily), PF13620 (carbohydrate-binding-like fold), or IPR011658 (PA14 carbohydrate-binding domain), (ii) binding to proteins via IPR002035 (that overlaps with the IPR036465 vWA domain superfamily), or even (iii) peptidase activity via IPR030392. These findings suggest that the rather simple modular design strategy – where a functional module is grafted during evolution onto a pilin – appears to have been used often during evolution both by monoderm and diderm bacteria and is expected to increase the functional versatility of T4P.

Discussion

T4F are an important research topic because of their virtual ubiquity in prokaryotes and their ability to mediate several key biological processes (1). Furthermore, the molecular mechanisms of T4F-mediated functions and the exact role of minor pilins remain incompletely understood. Therefore, in this report we focused on T4aP – the prototypical T4F (1) – in the recently established monoderm model *S. sanguinis* (15), and performed a structure/function analysis of the unusual minor pilin PilB, which we predicted might play a role in T4P-mediated adhesion. This led to several notable findings discussed below, and confirmed predictions that the study of T4P in monoderms has the potential to shine new light on these filaments (14, 15).

The first important finding in this study is that modular type IV pilins – T4F subunits in which an N-terminal pilin module is fused via a short linker to a functional module with a direct role in a T4F-mediated function – are widespread and extremely diverse. Modular pilins are likely to be tip-exposed in the filaments because of their peculiar architecture. However, a location in the body of the filament cannot be excluded provided that the linker is flexible enough. While previous 3D structures of a few large minor pilins suggest that they are modular pilins, their second modules do not correspond to protein domains readily identifiable by available bioinformatic tools. For example, in CofB from enterotoxigenic *E. coli* (ETEC) T4bP, there are two additional structural domains linked to the C-terminus of the pilin module by a flexible linker, a β-repeat domain followed by a β-sandwich domain (30). CofB, which forms a trimer predicted to be exposed at the tip of ETEC T4bP (31), appears to be an adapter for a secreted protein CofJ (32) that has a direct role in adhesion. TcpB, from *Vibrio cholerae* T4bP toxin-coregulated pilus, is a tip-located minor pilin forming trimers with a structure very similar to CofB (33), which probably has a similar function. Incidentally, CofB is also the receptor for the bacteriophage CTXϕ (33). In ComZ from *Thermus thermophilus* T4aP, the additional structural domain is a large β-solenoid inserted not at the end of the pilin module but into the β-sheet (34), which is thought to mediate

binding of extracellular DNA during transformation (34). This modular architecture is not restricted to T4P as it is also observed for a minor pilin from another T4F, GspK from type II secretion systems (T2SS) (35). In GspK, the additional structural domain is an α-domain of unclear function inserted into the β-sheet of the pilin module. GspK has been proposed to be at the tip of T2SS pseudopili, together with two other non-modular minor pilins (GspI and GspJ) with which it interacts to form a heterotrimer (35). These examples suggest that we have probably underestimated the global distribution of modular pilins, which are likely to be much more widespread because in many of them the additional modules are not yet defined by protein signatures in the databases. However, what is clear from our global analysis is that the functions associated with these modular pilins are potentially extremely diverse. Although a "common theme" appears to be the interaction of T4F with a variety of ligands – including proteins (via vWA in PilB, and the β-repeat/β-sandwich module in CofB), carbohydrates (via a variety of lectin domains including the concanavalin A-like lectin/glucanase domain in PilC), or DNA (the role of the β-solenoid module in ComZ) – other previously unreported T4F functions are possible. This is suggested by the modular architectures IPR012902-IPR030392 or IPR012902-IPR011493 in which the second module is a predicted peptidase belonging to S74 and M26 families, respectively.

The functional characterisation of the vWA module in PilB including its MIDAS, showing that it is a *bona fide* adhesin, is another significant achievement of this study. First, the vWA domain, which is ubiquitous in the three domains of life and has been extensively studied in eukaryotes (18, 21), has been much less studied in bacteria. Second, T4P-mediated adhesion remains among the least understood T4P functions (1). Our functional analysis of the vWA module in PilB significantly extends what was known for prokaryotic vWA-containing proteins and highlights important similarities and differences with eukaryotic vWA-containing proteins. Our 3D structure shows that the vWA module in PilB exhibits striking similarity to the vWA domain in eukaryotic proteins (20, 24), with a canonical MIDAS coordinating a metal. The main difference is

387 that the MIDAS in PilB is flanked by two protruding arms, similarly to what has been
388 described for RrgA from *S. pneumoniae* (26). Interestingly, RrgA is a subunit with
389 intrinsic adhesive properties (36) of sortase-assembled pili in monoderms (37), which
390 are unrelated to T4P. The parallel between RrgA and PilB denotes a case of
391 convergent evolution in which two unrelated types of pili have evolved a similar
392 strategy to mediate adhesion. Testing metal binding by the MIDAS in PilB, which was
393 previously done only for eukaryotic vWA-containing proteins (38), highlights important
394 similarities. MIDAS shows no significant binding to Ca^{2+}, and a slight preference for
395 Mn^{2+} over Mg^{2+}, although the difference in affinity is much smaller than in eukaryotic
396 proteins (38). Metal binding can be abolished by altering the MIDAS motif, which has
397 no impact on PilB structure (38, 39). Abolishing metal binding has no detectable effect
398 on piliation, which is analogous to what has been reported for vWA-containing
399 adhesins of sortase-assembled pili (40, 41), but it impairs T4P-mediated twitching
400 motility. It is unclear at this stage whether the lack of motility of the *pilB*$_{D319A}$ mutant is
401 due to reduced T4P-mediated adhesion to the agar, which would be consistent with
402 PilB role in adhesion, or to impaired filament retraction (11). We also provide evidence
403 that the vWA module of PilB binds several human protein ligands that it shares with
404 eukaryotic vWA-containing proteins such as integrins and/or vWF (19, 21). However,
405 unlike in these proteins where binding is often impaired when the MIDAS is inactivated
406 (20), binding to fibronectin and fibrinogen is unaffected in a PilB$_{D319A}$ mutant. This either
407 suggests that the MIDAS is not implicated in binding these specific ligands, which has
408 been described for vWF binding to collagen (25), or that our *in vitro* binding assay is
409 not sensitive enough to detect subtle but significant differences in binding.

410 The finding that PilB plays a key role in adhesion of *S. sanguinis* to host cells
411 and structures, via its vWA module, has implications for the pathogenesis of this
412 species in particular, and for our understanding of T4P-mediated adhesion in general.
413 Our findings are consistent with the possibility that PilB-mediated adhesion to host
414 proteins might play a role in IE (42), a life-threatening infection often caused by *S.*

sanguinis. Indeed, during IE, bacteria that have gained access to the bloodstream adhere to pre-existing sites of valvular damage where ECM proteins are exposed, and a blood clot is present containing large amounts of platelets, fibrinogen/fibrin and fibronectin (43). Our finding that PilB adheres directly to two of these proteins, but additional ligands cannot be excluded, suggests that PilB might be important at this early stage in IE, which could be tested in future studies. Our findings, which arguably make PilB the best characterised T4P adhesin alongside PilC/PilY1 found in diderm T4aP (44-48), have general implications for our understanding of T4P-mediated adhesion. The vWA module in PilB, which is most likely exposed at the pilus tip, is ideally placed to maximise bacterial adhesion to host protein receptors. T4P spring-like properties – gonococcal T4P can be stretched 3 times their length (49) – are expected to help bacteria that are bound via a tip-located adhesin to withstand adverse forces in their particular environment, *e.g.*, blood flow in a heart valve. This is likely to apply to other modular pilins as well, which harbour different modules predicted to function in adhesion. The parallel with the best characterised T4P adhesin PilC/PilY1 is obvious. This protein, which is not a pilin, is an adhesin that has been proposed to be presented at the T4P tip (45) via its interaction with a tip-located complex of four widely conserved minor pilins (50). All PilC/PilY1 have in common a C-terminal IPR008707 β-propeller domain while their N-termini are different (51). Since this is analogous to the situation with modular pilins, we wondered whether it could be an indication of a modular design for PilC/PilY1. This indeed seems to be the case since a search of the InterPro database (29) for all the proteins with an IPR008707 domain shows that 68 different PilC/PilY1 modular architectures are detected (Supplementary dataset 2). Strikingly, many of the N-terminal modules in PilC/PilY1 are shared with modular pilins, including vWA that was identified in PilY1 from *Pseudomonas aeruginosa* (52). These observations suggest that the same tinkering strategy has been used both by pilins and PilC/PilY1 to increase the functional versatility of T4P. In both instances, a "carrier" module for presentation at the tip of the filaments (either a

pilin, or an IPR008707 domain that interacts with a tip-located complex of minor pilins) has been fused to variety of "effector" modules, directly involved in a variety of functions.

In conclusion, by performing a detailed structure/function of the minor pilin PilB from *S. sanguinis*, we have shed light on several aspects of T4P biology. Our findings are not only of relevance for *S. sanguinis*, most notably for colonisation of its human host, they have general implications for T4F by uncovering a prevalent strategy used by these widespread filamentous nanomachines to promote their well-known exceptional functional versatility (1). The resulting conceptual framework paves the way for further investigations, which will further improve our understanding of these fascinating filaments.

Materials and methods

Strains and growth conditions

Strains and plasmids used in this study are listed in Table S1. For cloning, we used *E. coli* DH5α. For protein purification we used *E. coli* BL21(DE3) or *E. coli* BL21 B834(DE3) (Supplementary appendix). Chemically competent *E. coli* cells were prepared as described (53). DNA manipulations were done using standard molecular biology techniques (54). PCR were done using high-fidelity DNA polymerases (Agilent). Primers used in this study are listed in Table S2. The pET-28b (Novagen) derivative, pET28-$pilB_{36\text{-}461}$ for expressing 6His-$PilB_{36\text{-}461}$ was described previously (17). In this plasmid, the portion of a synthetic *pilB* gene codon-optimised for expression in *E. coli*, encoding the soluble portion of PilB, was fused to a non-cleavable N-terminal 6His tag. Similarly, we constructed pET28-$pilB_{192\text{-}461}$ for expressing $PilB_{vWA}$. To construct pET28-$pilB_{D319A}$ for expressing 6His-$PilB_{D319A}$, we introduced a missense mutation in pET28-$pilB_{36\text{-}461}$ using QuickChange site-directed mutagenesis (Agilent).

The WT *S. sanguinis* 2908 strain and deletion mutants (*ΔpilD*, *ΔpilB*) were described previously (16). *S. sanguinis* genomic DNA was prepared from overnight (O/N) liquid cultures using the XIT Genomic DNA from Gram-Positive Bacteria kit (G-Biosciences). Strain 2908, which is naturally competent, was transformed as described (16, 55). The unmarked *S. sanguinis* $pilB_{D319A}$ mutant was constructed using a previously described two-step gene editing strategy (55) (Supplementary appendix).

Protein purification

To purify native PilB, $PilB_{D319A}$ and $PilB_{vWA}$ proteins, the corresponding pET-28b derivatives were transformed in *E. coli* BL21(DE3). Expression and purification are detailed in the Supplementary appendix. To purify SeMet-labelled PilB for phasing, the corresponding pET-28b derivative was transformed in *E. coli* BL21 B834(DE3). Expression and purification are detailed in the Supplementary appendix.

Crystallisation and structure determination

Purified proteins in 50 mM HEPES pH 7.4, 200 mM NaCl were concentrated to 50 mg/ml and tested for crystallisation using sitting-drop vapor diffusion, with 100 nl drops of protein solution and mother liquor. We tested a range of commercially available kits (Molecular Dimensions, Hampton Research and Rigaku Reagents), which yielded a number of hits, mainly in high salt conditions. Crystallisation conditions were optimised to yield larger and better diffracting crystals. All data was collected and processed using the Diamond Light Source beamline i03, and integrated in $P6_1$ using the 3dii pipeline in *xia*2 (56). Initial molecular replacement was performed with Phaser (57) on the 2.26 Å resolution PilB dataset using a low-resolution partial model produced from the SeMet data using autoSHARP (58). Manual building in Coot (59) was performed on the high-resolution dataset, and the full model was then used for molecular replacement in the low-resolution datasets. All structures were produced using Coot and phenix.refine (60) and validated using MolProbity (61).

Assaying metal-binding by purified PilB

The metal binding specificity of PilB was tested using ThermoFluor, a fluorescent-based method measuring changes in thermal denaturation temperature (27). Assays were done in a 96-well plate (Applied Biosystems) format. In the wells, we added to a final volume of 40 µl (i) 0-1 mM range of concentrations of $MgCl_2$, $MnCl_2$ and $CaCl_2$, (ii) 20 µM purified PilB or $PilB_{D319A}$, and (iii) 1/5,000 dilution of SYBR Orange (Thermo Fisher Scientific). Plates were then analysed using a temperature gradient, from 25 to 99ºC, on a StepOnePlus real-time qPCR machine (Applied Biosystems). The data were exported MATLAB and analysed in GraphPad. Analyses were performed with Prism (GraphPad Software). Kd were calculated using non-linear regression fits, applying saturation binding equation (One site - Total and non-specific binding) using Ca^{2+} as non-specific binding control.

Assaying protein ligand-binding by purified PilB

Binding of PilB, $PilB_{vWA}$, $PilB_{D319A}$ to a variety of eukaryotic proteins was tested by ELISA as follows. Putative ligand proteins (elastin from human skin, fibrinogen from human plasma, laminin from human placenta, fibronectin from human plasma) (all from Sigma) were resuspended in carbonate-bicarbonate buffer (Sigma) at 5 µg/ml. Fifty µl was dispatched into the wells of MaxiSorp plates, and adsorbed O/N at 4°C. Wells were washed three times with PBS (Gibco) and blocked during 1 h with 3 % BSA (Probumin) or 1 % gelatin (Sigma) in PBS. After washing with PBST (PBS containing 0.05 % tween 20), serial twofold dilutions of PilB (from 40 to 0.625 µg/ml) were added to the wells and incubated for 2 h at 37°C. After five washes with PBST, we added 50 µl anti-6His RTM antibody (Abcam) at 1/500 dilution in PBS, and incubated for 1 h at RT. After five washes with PBST, we added 50 µl Amersham ECL anti-rabbit IgG HRP-linked whole antibody (GE Healthcare) at 1/500 dilution in PBS, and incubated for 1 h at RT. After five washes with PBST, we added 100 µl/well of TMB solution (Thermo Scientific) and incubated the plates during 20 min at RT in the dark. Finally, we stopped the reaction by adding 100 µl/well of 0.18 M sulfuric acid, before reading the plates at 450 nm using a plate reader. Analyses were performed with Prism (GraphPad Software). Kd were calculated using non-linear regression fits, applying saturation binding equation (One site - Total and non-specific binding) using BSA or gelatin as non-specific binding control.

Assaying piliation of *S. sanguinis*

S. sanguinis T4P were purified as described (16, 17) (Supplementary appendix). Pili were resuspended in pilus buffer, separated by SDS-PAGE, before gels were stained with Bio-Safe Coomassie (Bio-Rad).

Assaying twitching motility of *S. sanguinis*

Twitching motility was assessed on agar plates as described (16) (Supplementary appendix). Plates were photographed using an Epson Perfection V700 photo scanner.

Assaying adhesion of *S. sanguinis* to eukaryotic cells

We tested adhesion of *S. sanguinis* to CHO cells (Public Health England) as follows. Cells were replicated in flasks in DMEM medium (Gibco) containing 1 x MEM non-essential amino acid mix (Gibco) and 5 % fetal bovine serum (Gibco) and seeded at 100,000 cells/cm^2 in 24-well plates, which were incubated O/N at 37°C in the presence of 5 % CO_2. The next day, cell monolayers were gently rinsed with PBS, and infected at an MOI of 10 with bacteria grown in TH. In brief, bacteria were grown for a few hours to OD_{600} 0.5 units, adjusted at the same OD, pelleted by centrifugation at 1,100 *g* during 10 min, and resuspended in PBS. Bacteria in the inoculum were quantified by performing CFU counts on TH plates. After 1h of infection at 37ºC, cell monolayers were gently rinsed four times with PBS, before cells with adherent bacteria were scraped in distilled water. Adherent bacteria were then quantified by performing CFU counts. Statistical analyses were performed with Prism. Comparisons were done by one-way ANOVA, followed by Dunnett's multiple comparison tests. An adjusted *P* value < 0.05 was considered significant (* $P<0.05$, ** $P<0.01$, *** $P<0.001$, **** $P<0.0001$).

Bioinformatics

Protein sequences were routinely analysed using DNA Strider (62). Prediction of protein domains, their global distribution and associated architectures was done by using InterProScan (29) to interrogate the InterPro database. This database was also used to download all the protein entries discussed in this paper. Molecular visualisation of protein 3D structures was done using PyMOL (Schrödinger). The PDBsum Generate (63) server was used to provide at-a-glance overviews – secondary structure, topology diagram, protein motifs, schematic diagram of metal-protein interactions – of the 3D

structures determined during this work. The DALI (64) server was used for comparing protein structures in 3D. Protein 3D structures were downloaded from the RCSB PDB server. The 3d-SS (65) server was used to superpose 3D protein structures with the STAMP algorithm (66).

The cryo-EM structure of *N. meningitidis* T4P (PDB 5KUA) (9) was used to model, using SWISS-MODEL (67), the N-terminal helices of PilE1, PilE2 and PilB within the filaments. Coot and PyMOL were then used to place the full-length structures within the T4P model.

Data availability

3D structures have been deposited in the PDB, under accession codes 7B7P (PilB), and 7BA2 ($PilB_{D319A}$). All the datasets generated during this study are included in this paper and its Supplementary appendix.

References

1. J. L. Berry, V. Pelicic, Exceptionally widespread nano-machines composed of type IV pilins: the prokaryotic Swiss Army knives. *FEMS Microbiol. Rev.* **39**, 134-154 (2015).

2. R. Denise, S. S. Abby, E. P. C. Rocha, Diversification of the type IV filament superfamily into machines for adhesion, protein secretion, DNA uptake, and motility. *PLoS Biol.* **17**, e3000390 (2019).

3. V. Pelicic, Type IV pili: *e pluribus unum*? *Mol. Microbiol.* **68**, 827-837 (2008).

4. C. L. Giltner, Y. Nguyen, L. L. Burrows, Type IV pilin proteins: versatile molecular modules. *Microbiol. Mol. Biol. Rev.* **76**, 740-772 (2012).

5. C. F. LaPointe, R. K. Taylor, The type 4 prepilin peptidases comprise a novel family of aspartic acid proteases. *J. Biol. Chem.* **275**, 1502-1510 (2000).

6. O. Francetic, N. Buddelmeijer, S. Lewenza, C. A. Kumamoto, A. P. Pugsley, Signal recognition particle-dependent inner membrane targeting of the PulG pseudopilin component of a type II secretion system. *J. Bacteriol.* **189**, 1783-1793 (2007).

7. J. Arts, R. van Boxtel, A. Filloux, J. Tommassen, M. Koster, 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 (2007).

8. Y. W. Chang *et al.*, Architecture of the type IVa pilus machine. *Science* **351**, aad2001 (2016).

9. S. Kolappan *et al.*, Structure of the *Neisseria meningitidis* type IV pilus. *Nat. Commun.* **7**, 13015 (2016).

10. F. Wang *et al.*, Cryoelectron microscopy reconstructions of the *Pseudomonas aeruginosa* and *Neisseria gonorrhoeae* type IV pili at sub-nanometer resolution. *Structure* **25**, 1423-1435 (2017).

11. A. J. Merz, M. So, M. P. Sheetz, Pilus retraction powers bacterial twitching motility. *Nature* **407**, 98-102 (2000).

12. B. Maier *et al.*, Single pilus motor forces exceed 100 pN. *Proc. Natl. Acad. Sci. USA* **99**, 16012-16017 (2002).

606 13. N. Biais, B. Ladoux, D. Higashi, M. So, M. Sheetz, Cooperative retraction of
607 bundled type IV pili enables nanonewton force generation. *PLoS Biol.* **6**, e87 (2008).
608 14. S. Melville, L. Craig, Type IV pili in Gram-positive bacteria. *Microbiol. Mol. Biol.*
609 *Rev.* **77**, 323-341 (2013).
610 15. V. Pelicic, Monoderm bacteria: the new frontier for type IV pilus biology. *Mol.*
611 *Microbiol.* **112**, 1674-1683 (2019).
612 16. I. Gurung *et al.*, Functional analysis of an unusual type IV pilus in the Gram-positive
613 *Streptococcus sanguinis*. *Mol. Microbiol.* **99**, 380-392 (2016).
614 17. J. L. Berry *et al.*, Global biochemical and structural analysis of the type IV pilus
615 from the Gram-positive bacterium *Streptococcus sanguinis*. *J. Biol. Chem.* **294**, 6796-
616 6808 (2019).
617 18. C. A. Whittaker, R. O. Hynes, Distribution and evolution of von Willebrand/integrin
618 A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol.*
619 *Biol. Cell* **13**, 3369-3387 (2002).
620 19. J. E. Sadler, Biochemistry and genetics of von Willebrand factor. *Annu. Rev.*
621 *Biochem.* **67**, 395-424 (1998).
622 20. J. O. Lee, P. Rieu, M. A. Arnaout, R. Liddington, Crystal structure of the A domain
623 from the alpha subunit of integrin CR3 (CD11b/CD18). *Cell* **80**, 631-638 (1995).
624 21. S. K. Dickeson, S. A. Santoro, Ligand recognition by the I domain-containing
625 integrins. *Cell. Mol. Life Sci.* **54**, 556-566 (1998).
626 22. L. Craig *et al.*, Type IV pilin structure and assembly. X-ray and EM analyses of
627 *Vibrio cholerae* toxin-coregulated pilus and *Pseudomonas aeruginosa* PAK pilin. *Mol.*
628 *Cell* **11**, 1139-1150 (2003).
629 23. E. G. Hutchinson, J. M. Thornton, HERA – a program to draw schematic diagrams
630 of protein secondary structures. *Proteins* **8**, 203-212 (1990).
631 24. A. Qu, D. J. Leahy, Crystal structure of the I-domain from the CD11a/CD18 (LFA-
632 1, αLβ2) integrin. *Proc. Natl. Acad. Sci. USA* **92**, 10277-10281 (1995).

633 25. R. A. Romijn *et al.*, Identification of the collagen-binding site of the von Willebrand
634 factor A3-domain. *J. Biol. Chem.* **276**, 9985-9991 (2001).
635 26. T. Izoré *et al.*, Structural basis of host cell recognition by the pilus adhesin from
636 *Streptococcus pneumoniae*. *Structure* **18**, 106-115 (2010).
637 27. F. H. Niesen, H. Berglund, M. Vedadi, The use of differential scanning fluorimetry
638 to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212-2221
639 (2007).
640 28. J. Kreth, R. A. Giacaman, R. Raghavan, J. Merritt, The road less traveled – defining
641 molecular commensalism with *Streptococcus sanguinis*. *Mol. Oral. Microbiol.* **32**, 181-
642 196 (2017).
643 29. P. Jones *et al.*, InterProScan 5: genome-scale protein function classification.
644 *Bioinformatics* **30**, 1236-1240 (2014).
645 30. S. Kolappan, D. Ng, G. Yang, T. Harn, L. Craig, Crystal structure of the minor pilin
646 CofB, the initiator of CFA/III pilus assembly in enterotoxigenic *Escherichia coli*. *J. Biol.*
647 *Chem.* **290**, 25805-25818 (2015).
648 31. K. Kawahara *et al.*, Homo-trimeric structure of the type IVb minor pilin CofB
649 suggests mechanism of CFA/III pilus assembly in human enterotoxigenic *Escherichia*
650 *coli*. *J. Mol. Biol.* **428**, 1209-1226 (2016).
651 32. H. Oki *et al.*, Interplay of a secreted protein with type IVb pilus for efficient
652 enterotoxigenic *Escherichia coli* colonization. *Proc. Natl. Acad. Sci. USA* **115**, 7422-
653 7427 (2018).
654 33. M. Gutierrez-Rodarte, S. Kolappan, B. A. Burrell, L. Craig, The *Vibrio cholerae*
655 minor pilin TcpB mediates uptake of the cholera toxin phage CTXϕ. *J. Biol. Chem.* **294**,
656 15698-15710 (2019).
657 34. M. Z. Salleh *et al.*, Structure and properties of a natural competence-associated
658 pilin suggest a unique pilus tip-associated DNA receptor. *mBio* **10**, e00614-00619
659 (2019).

660 35. K. V. Korotkov, W. G. Hol, Structure of the GspK-GspI-GspJ complex from the
661 enterotoxigenic *Escherichia coli* type 2 secretion system. *Nat. Struct. Mol. Biol.* **15**,
662 462-468 (2008).
663 36. M. Hilleringmann *et al.*, Pneumococcal pili are composed of protofilaments
664 exposing adhesive clusters of RrgA. *PLoS Pathog.* **4**, e1000026 (2008).
665 37. J. L. Telford, M. A. Barocchi, I. Margarit, R. Rappuoli, G. Grandi, Pili in Gram-
666 positive pathogens. *Nat. Rev. Microbiol.* **4**, 509-519 (2006).
667 38. E. T. Baldwin *et al.*, Cation binding to the integrin CD11b I domain and activation
668 model assessment. *Structure* **6**, 923-935 (1998).
669 39. A. Qu, D. J. Leahy, The role of the divalent cation in the structure of the I domain
670 from the CD11a/CD18 integrin. *Structure* **4**, 931-942 (1996).
671 40. Y. Konto-Ghiorghi *et al.*, Dual role for pilus in adherence to epithelial cells and
672 biofilm formation in *Streptococcus agalactiae*. *PLoS Pathog.* **5**, e1000422 (2009).
673 41. H. V. Nielsen *et al.*, The metal ion-dependent adhesion site motif of the
674 *Enterococcus faecalis* EbpA pilin mediates pilus function in catheter-associated urinary
675 tract infection. *mBio* **3**, e00177-00112 (2012).
676 42. Y. A. Que, P. Moreillon, Infective endocarditis. *Nat. Rev. Cardiol.* **8**, 322-336
677 (2011).
678 43. K. Werdan *et al.*, Mechanisms of infective endocarditis: pathogen-host interaction
679 and risk states. *Nat. Rev. Cardiol.* **11**, 35-50 (2014).
680 44. X. Nassif *et al.*, Roles of pilin and PilC in adhesion of *Neisseria meningitidis* to
681 human epithelial and endothelial cells. *Proc. Natl. Acad. Sci. USA* **91**, 3769-3773
682 (1994).
683 45. T. Rudel, I. Scheuerpflug, T. F. Meyer, *Neisseria* PilC protein identified as a type-
684 4 pilus-tip located adhesin. *Nature* **373**, 357-359 (1995).
685 46. R. W. Heiniger, H. C. Winther-Larsen, R. J. Pickles, M. Koomey, M. C. Wolfgang,
686 Infection of human mucosal tissue by *Pseudomonas aeruginosa* requires sequential

687 and mutually dependent virulence factors and a novel pilus-associated adhesin. *Cell.*
688 *Microbiol.* **12**, 1158-1173 (2010).
689 47. M. D. Johnson *et al.*, *Pseudomonas aeruginosa* PilY1 binds integrin in an RGD-
690 and calcium-dependent manner. *PLoS One* **6**, e29629 (2011).
691 48. E. A. Porsch *et al.*, Calcium binding properties of the *Kingella kingae* PilC1 and
692 PilC2 proteins have differential effects on type IV pilus-mediated adherence and
693 twitching motility. *J. Bacteriol.* **195**, 886-895 (2013).
694 49. N. Biais, D. L. Higashi, J. Brujic, M. So, M. P. Sheetz, Force-dependent
695 polymorphism in type IV pili reveals hidden epitopes. *Proc. Natl. Acad. Sci. USA* **107**,
696 11358-11363 (2010).
697 50. A. Treuner-Lange *et al.*, PilY1 and minor pilins form a complex priming the type
698 IVa pilus in *Myxococcus xanthus*. *Nat. Commun.* **11**, 5054 (2020).
699 51. J. Orans *et al.*, Crystal structure analysis reveals *Pseudomonas* PilY1 as an
700 essential calcium-dependent regulator of bacterial surface motility. *Proc. Natl. Acad.*
701 *Sci. USA* **107**, 1065-1070 (2010).
702 52. S. L. Kuchma *et al.*, Cyclic-di-GMP-mediated repression of swarming motility by
703 *Pseudomonas aeruginosa*: the *pilY1* gene and its impact on surface-associated
704 behaviors. *J. Bacteriol.* **192**, 2950-2964 (2010).
705 53. H. Inoue, H. Nojima, H. Okayama, High efficiency transformation of *Escherichia*
706 *coli* with plasmids. *Gene* **96**, 23-28 (1990).
707 54. J. Sambrook, D. W. Russell, *Molecular cloning. A laboratory manual* (Cold Spring
708 Harbor Laboratory Press, Cold Spring Harbor, New York, 2001).
709 55. I. Gurung, J. L. Berry, A. M. J. Hall, V. Pelicic, Cloning-independent markerless
710 gene editing in *Streptococcus sanguinis*: novel insights in type IV pilus biology. *Nucleic*
711 *Acids Res.* **45**, e40 (2017).
712 56. G. Winter, C. M. Lobley, S. M. Prince, Decision making in *xia*2. *Acta Crystallogr.*
713 *Sect. D: Biol. Crystallogr.* **69**, 1260-1273 (2013).

714 57. A. J. McCoy *et al.*, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-
715 674 (2007).
716 58. C. Vonrhein, E. Blanc, P. Roversi, G. Bricogne, Automated structure solution with
717 autoSHARP. *Meth. Mol. Biol.* **364**, 215-230 (2007).
718 59. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of
719 Coot. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **66**, 486-501 (2010).
720 60. P. V. Afonine *et al.*, Towards automated crystallographic structure refinement with
721 phenix.refine. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **68**, 352-367 (2012).
722 61. V. B. Chen *et al.*, MolProbity: all-atom structure validation for macromolecular
723 crystallography. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **66**, 12-21 (2010).
724 62. C. Marck, 'DNA Strider': a 'C' program for the fast analysis of DNA and protein
725 sequences on the Apple Macintosh family of computers. *Nucleic Acids Res.* **16**, 1829-
726 1836 (1988).
727 63. R. A. Laskowski, J. Jablonska, L. Pravda, R. S. Varekova, J. M. Thornton,
728 PDBsum: structural summaries of PDB entries. *Protein Sci.* **27**, 129-134 (2018).
729 64. L. Holm, L. M. Laakso, Dali server update. *Nucleic Acids Res.* **44**, W351-355
730 (2016).
731 65. K. Sumathi, P. Ananthalakshmi, M. N. Roshan, K. Sekar, 3dSS: 3D structural
732 superposition. *Nucleic Acids Res.* **34**, W128-W132 (2006).
733 66. R. B. Russell, G. J. Barton, Multiple protein sequence alignment from tertiary
734 structure comparison: assignment of global and residue confidence levels. *Proteins*
735 **14**, 309-323 (1992).
736 67. A. Waterhouse *et al.*, SWISS-MODEL: homology modelling of protein structures
737 and complexes. *Nucleic Acids Res.* **46**, W296-W303 (2018).

Acknowledgements

This work was funded by the Medical Research Council (MR/P022197/1). We acknowledge the use of the crystallisation facility at Imperial College London, which is supported by the Biotechnology and Biological Sciences Research Council (BB/D524840/1) and Wellcome Trust (202926/Z/16/Z). We thank Claire Poyart (Institut Cochin) for advice on adhesion experiments. We thank Angelika Gründling (Imperial College London), Sophie Helaine (Harvard Medical School), Christoph Tang (University of Oxford), and Romé Voulhoux (Laboratoire de Chimie Bactérienne) for critical reading of the manuscript.

Figure legends

Fig. 1. Bioinformatic analysis of PilB. A) Relevant features of PilB from *S. sanguinis* 2908. The N-terminal class III signal peptide is boxed. The 7-aa long leader peptide contains mostly hydrophilic (shaded in orange) and neutral (no shading) residues, and it ends with a conserved Gly. This leader peptide is processed by the prepilin peptidase PilD, which is indicated by the vertical arrow, generating a protein of 454 residues (50.5 kDa). The processed protein starts with a tract of 21 predominantly hydrophobic residues (shaded in blue), which invariably form an extended α-helix that is the main assembly interface within filaments. The C-terminal vWA module (IPR036465) is boxed, with the conserved residues forming the MIDAS highlighted in yellow. Arrowheads indicate the proteins that were produced and purified in this study, consisting of either two modules (black arrowhead) or just the vWA module (red arrowhead). **B**) Modular architectures of PilB and PilC minor pilins compared to the major pilins PilE1/PilE2. The proteins, from *S. sanguinis* 2908, have been drawn to scale. The black rounded rectangles correspond to the IPR012902 motif that is part of the class III signal peptide. The C-terminal domains in PilB and PilC are highlighted by coloured rounded rectangles, vWA domain (red) and lectin domain (yellow).

Fig. 2. Crystal structure of PilB. A) Orthogonal cartoon views of the PilB structure in which the two distinct modules have been highlighted in blue (pilin module) and red (vWA module), while the short loop connecting them is in grey. The orange sphere represents a magnesium ion. **B**) Left, close-up cartoon view of the pilin module coloured in rainbow spectrum from blue (N-terminus) to red (C-terminus). Right, topology diagram of the pilin module structure. **C**) Left, close-up cartoon view of the vWA module in which the β-strands composing the central β-sheet are highlighted in red, while the surrounding α-helices are highlighted in yellow. The connecting loops

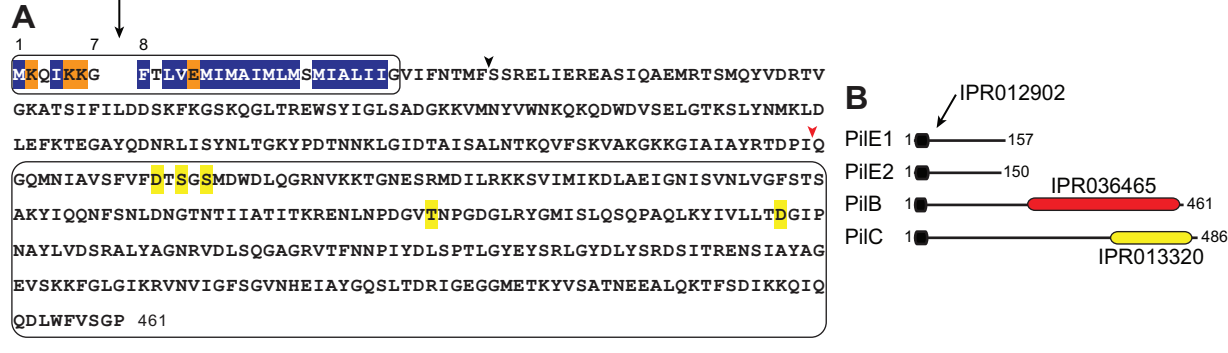
774 are in grey, except for the two "arms" on top of the structure (coloured in orange and
775 blue), which surround the MIDAS. Right, diagram of the magnesium coordination by
776 the conserved MIDAS residues. Coordinating oxygen atoms are shown, with dashed
777 lines corresponding to hydrogen bonds.
778
779 **Fig. 3. 3D model of PilB in *S. sanguinis* T4P. A**) Packing of PilB (red) into *S.*
780 *sanguinis* T4P, which is a right-handed helical heteropolymer of two major pilins PilE1
781 (blue) and PilE2 (grey). **B**) View of the T4P tip capped by PilB, or not.
782
783 **Fig. 4. Metal binding by PilB.** Purified PilB was incubated with increasing
784 concentrations of divalent ions (Ca^{2+}, Mg^{2+}, Mn^{2+}) and binding was quantified by
785 ThermoFluor. **A**) Metal binding by PilB. **B**) Metal binding by $PilB_{D319A}$, with an inactive
786 MIDAS. Results are the average ± standard deviations from 3 independent
787 experiments.
788
789 **Fig. 5. 3D crystal structure of $PilB_{D319A}$. A**) Close-up cartoon view of the vWA module
790 in $PilB_{D319A}$. **B**) Comparison of electron density maps in the MIDAS pocket for the
791 $PilB_{D319A}$ (upper panel) and PilB (lower panel) structures. **C**) Superposition of the vWA
792 modules of PilB with bound Mg^{2+} (grey) and $PilB_{D319A}$ (orange). The two structures
793 superpose with a RMSD of 0.45 Å.
794
795 **Fig. 6. Phenotypic characterisation of a *S. sanguinis* mutant expressing $PilB_{D319A}$**
796 **with an inactive MIDAS. A**) Piliation was quantified by purifying T4P from cultures
797 adjusted to the same OD_{600}. Purified T4P (identical volumes were loaded in each lane)
798 were separated by SDS-PAGE and stained with Coomassie blue. A molecular weight
799 marker (MW) was run in the first lane. Molecular weights are indicated in kDa. **B**)
800 Twitching motility was assessed by a macroscopic motility assay. Bacteria were

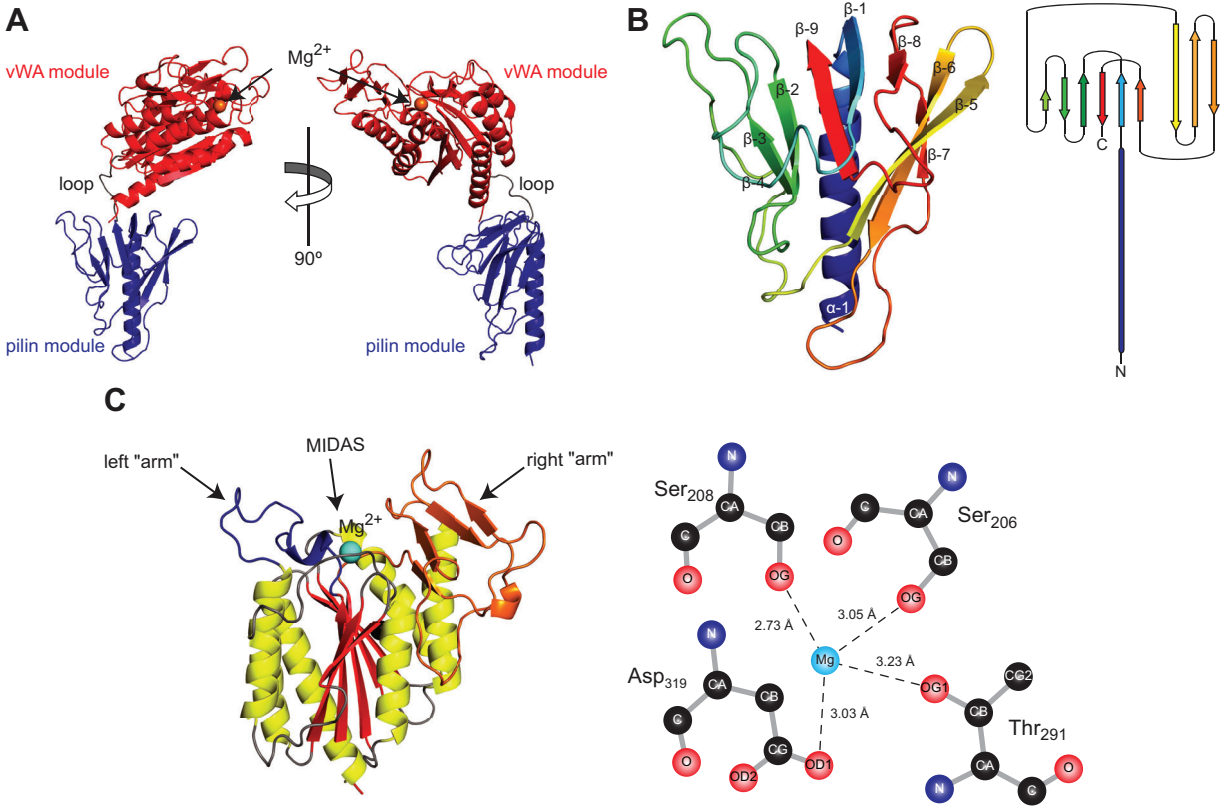
streaked on plates, which were incubated several days at 37ºC in a humid atmosphere and then photographed. Twitching motility is characterised by spreading zones around colonies. **C**) Adhesion of *S. sanguinis* to eukaryotic cells was quantified by incubating bacteria (MOI 10) with CHO cells for 1 h. After removing non-adherent bacteria by several washes, bacteria adhering to cells were enumerated by performing CFU counts. Results are expressed as adhesion relative to WT (set to 1) and are the average ± standard deviations from five independent experiments. For statistical analysis, one-way ANOVA followed by Dunnett's multiple comparison tests were performed (****$P < 0.0001$).

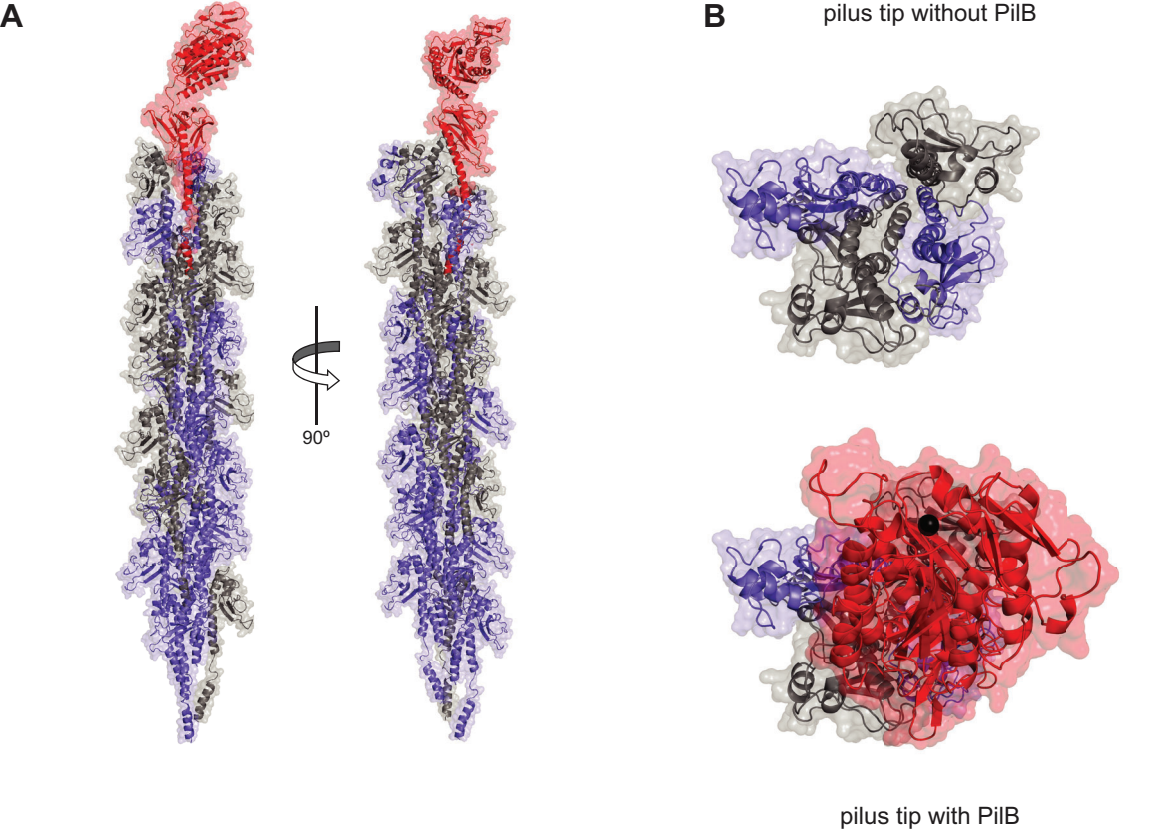
Fig. 7. Dose-dependent binding of PilB to various protein ligands. Increasing concentrations of purified PilB was added to constant concentrations of immobilised putative ligands, and binding was quantified by ELISA. BSA served as negative control. Results are the average ± standard deviations from at least three independent experiments. **A**) Binding of PilB to fibrinogen, fibronectin, elastin and laminin. **B**) Binding of $PilB_{vWA}$, consisting only of the vWA module, to fibrinogen and fibronectin.

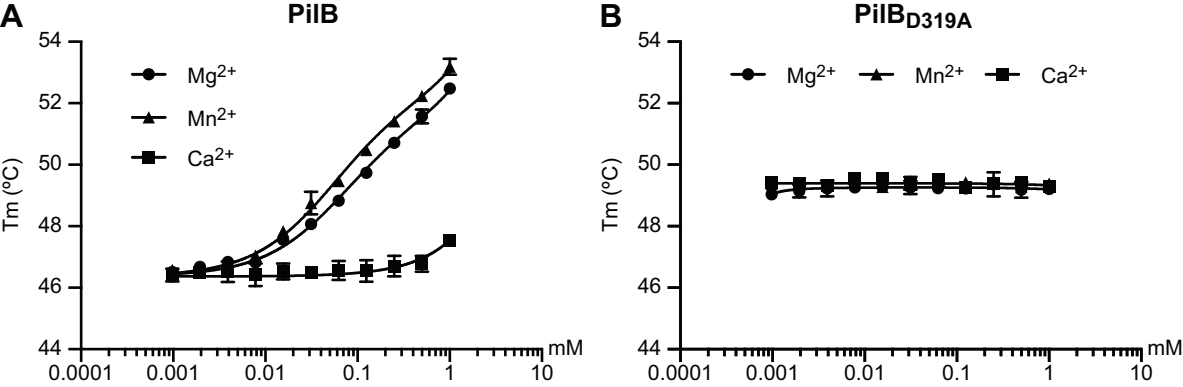
Fig. 8. Global distribution of modular pilins. Fifteen most widespread modular pilin architectures in the InterPro database are presented. The numbers in parenthesis represent the number of proteins displaying that architecture. The representative proteins depicted, drawn to scale, are from the following species. 1, *Candidatus* Magasanikbacteria (UniProtKB/TrEMBL protein A0A0G0IU57). 2, *Photorhabdus luminescens* (A0A022PI42). 3, *Desulfuribacillus stibiiarsenatis* (A0A1E5L295). 4, *Candidatus* Falkowbacteria (A0A1J4TDE2). 5, *Candidatus* Wolfebacteria (A0A0G1WFE5). 6, Clostridiales bacterium (A0A101H8M7). 7, *Corynebacterium glutamicum* (A0A1Q6BQB1). 8, *Thermosulfidibacter takaii* (A0A0S3QUH2). 9, *Candidatus* Gracilibacteria (A0A1J5F7A7). 10, *Candidatus* Saccharibacteria (A0A1Q3NLQ9). 11, Planctomycetes bacterium (A0A1G2ZHU9). 12, *Actinoplanes*

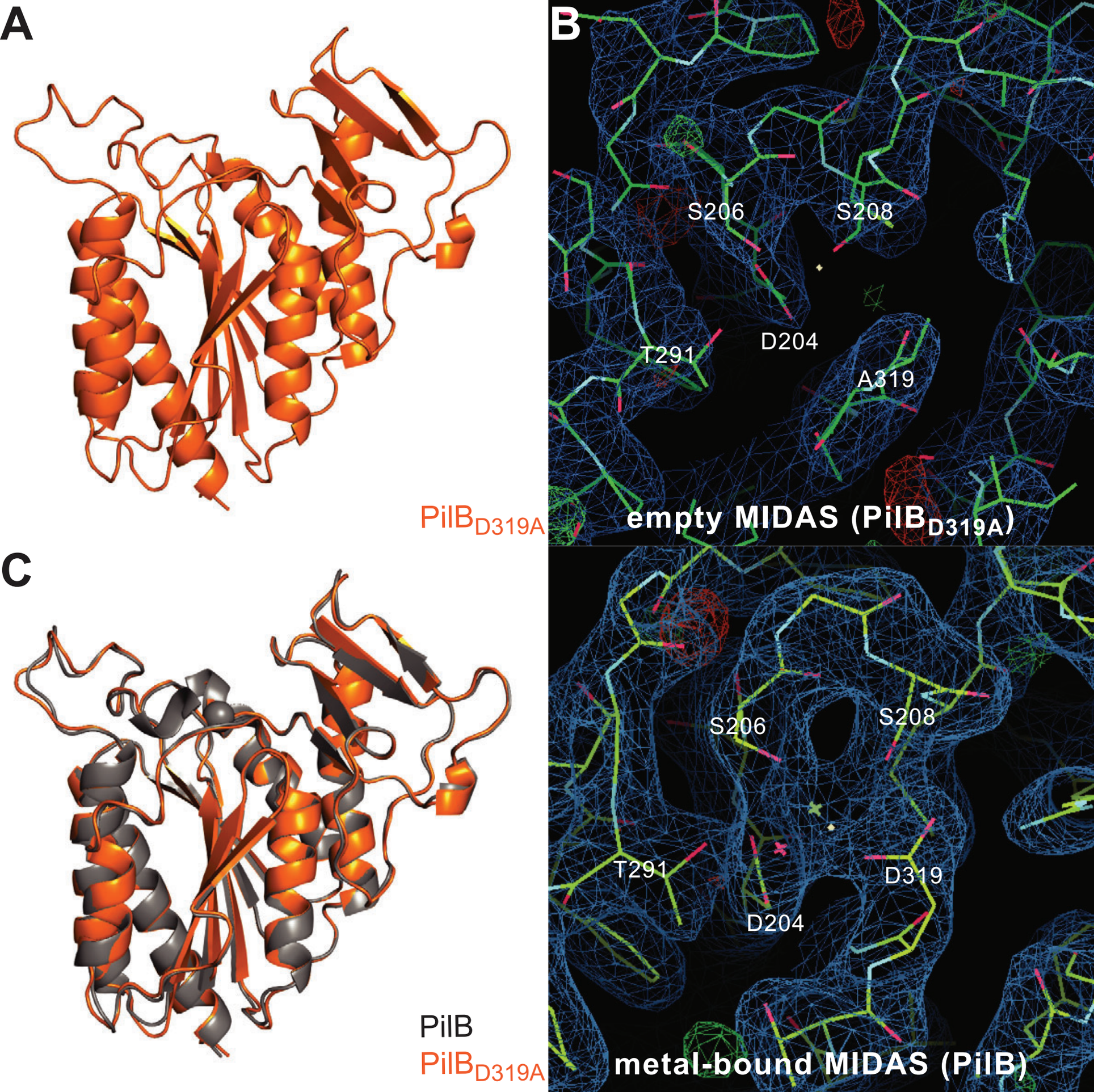
awajinensis subsp. *mycoplanecinus* (A0A101J7V4). 13, *Actinoplanes derwentensis* (A0A1H2D7E9). 14, Desulfobacterales bacterium (A0A1V1WSE4). 15, Parcubacteria group bacterium (A0A2D6FLV5).

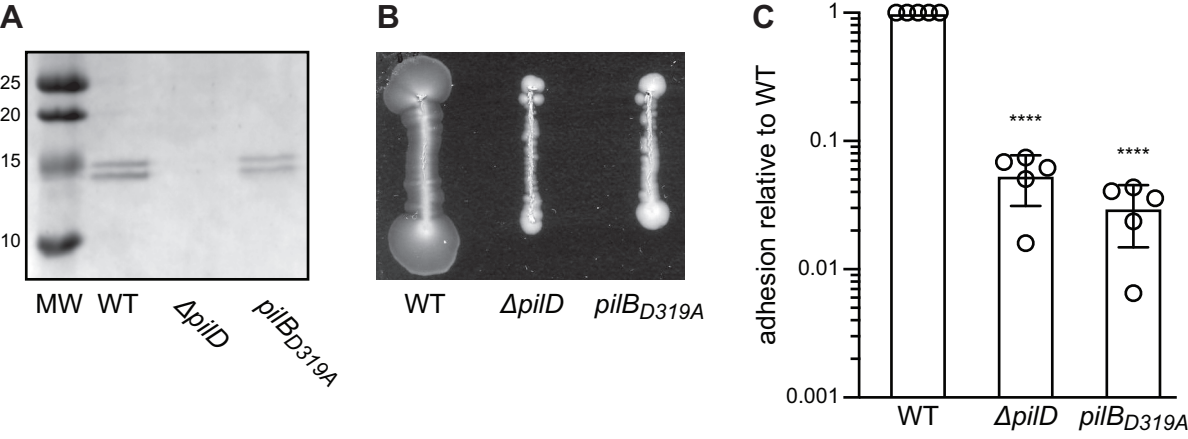

A
1
7
8
MKQIKKG FTLVEMIMAIMLMSMIALIIGVIFNTMFSSRELIEREASIQAEMRTSMQYVDRTV
GKATSIFILDDSKFKGSKQGLTREWSYIGLSADGKKVMNYVWNKQKQDWDVSELGTKSLYNMKLD
LEFKTEGAYQDNRLISYNLTGKYPDTNNKLGIDTAISALNTKQVFSKVAKGKKGIAIAYRTDPIQ
GQMNIAVSFVFDTSGSMDWDLQGRNVKKTGNESRMDILRKKSVIMIKDLAEIGNISVNLVGFSTS
AKYIQQNFSNLDNGTNTIIATITKRENLNPDGVTNPGDGLRYGMISLQSQPAQLKYIVLLTDGIP
NAYLVDSRALYAGNRVDLSQGAGRVTFNNPIYDLSPTLGYEYSRLGYDLYSRDSITRENSIAYAG
EVSKKFGLGIKRVNVIGFSGVNHEIAYGQSLTDRIGEGGMETKYVSATNEEALQKTFSDIKKQIQ
QDLWFVSGP 461
B
IPR012902
PilE1 1 157
PilE2 1 150
PilB 1 461
PilC 1 486
IPR036465
IPR013320

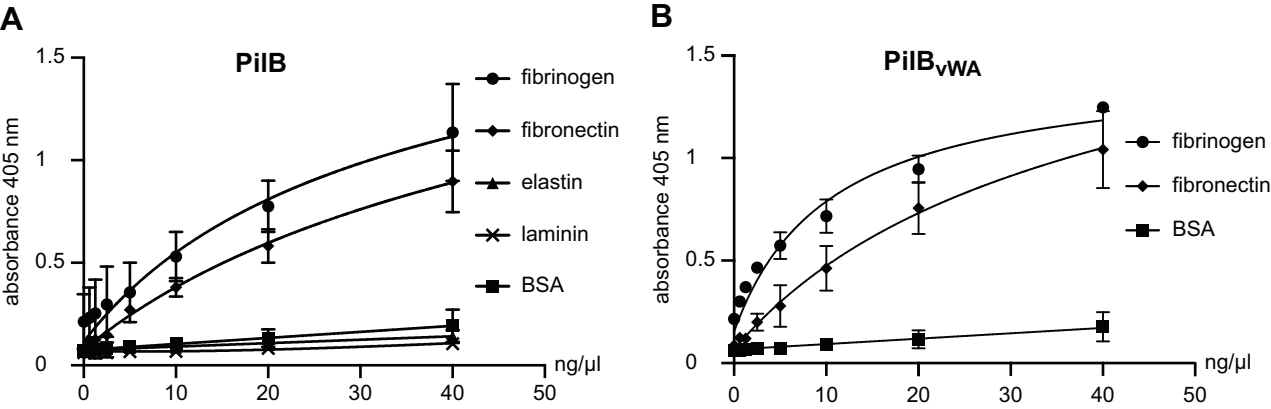

A
vWA module
Mg2+
vWA module
loop
loop
90°
pilin module
pilin module
B
β-1
β-9
β-8
β-6
β-2
β-5
β-3
β-7
β-4
α-1
C
N
C
left "arm"
MIDAS
right "arm"
Mg2+
Ser208
Ser206
2.73 Å
3.05 Å
3.23 Å
Asp319
3.03 Å
Thr291

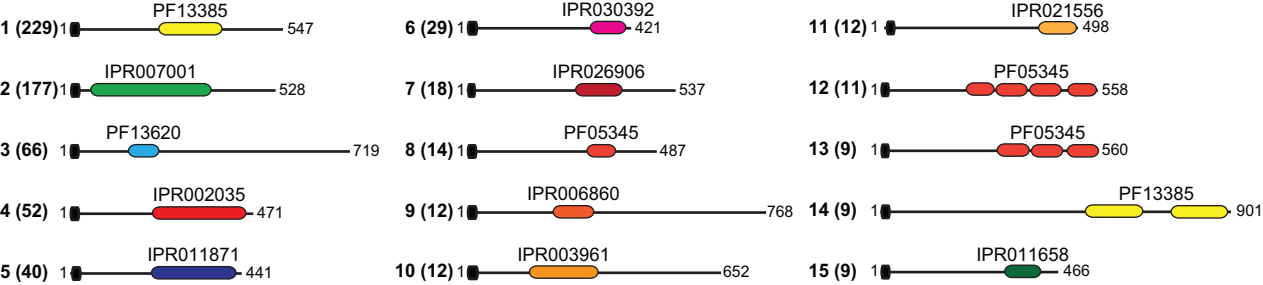

A
90°
B
pilus tip without PilB
pilus tip with PilB


A
PilB
Mg2+
Mn2+
Ca2+
Tm (ºC)
54
52
50
48
46
44
0.0001
0.001
0.01
0.1
1
10
mM
B
PilBD319A
Mg2+
Mn2+
Ca2+
Tm (ºC)
54
52
50
48
46
44
0.0001
0.001
0.01
0.1
1
10
mM


A
PilB$_{D319A}$
B
S206
S208
D204
T291
A319
empty MIDAS (PilB$_{D319A}$)
C
PilB
PilB$_{D319A}$
S206
S208
T291
D319
D204
metal-bound MIDAS (PilB)


A
25
20
15
10
MW
WT
ΔpilD
pilB$_{D319A}$
B
WT
ΔpilD
pilB$_{D319A}$
C
adhesion relative to WT
1
0.1
0.01
0.001
****
****
WT
ΔpilD
pilB$_{D319A}$


A
PilB
absorbance 405 nm
1.5
1
0.5
0
0
10
20
30
40
50
ng/µl
fibrinogen
fibronectin
elastin
laminin
BSA
B
PilB vWA
absorbance 405 nm
1.5
1
0.5
0
0
10
20
30
40
50
ng/µl
fibrinogen
fibronectin
BSA


1 (229)
PF13385
1
547
2 (177)
IPR007001
1
528
3 (66)
PF13620
1
719
4 (52)
IPR002035
1
471
5 (40)
IPR011871
1
441
6 (29)
IPR030392
1
421
7 (18)
IPR026906
1
537
8 (14)
PF05345
1
487
9 (12)
IPR006860
1
768
10 (12)
IPR003961
1
652
11 (12)
IPR021556
1
498
12 (11)
PF05345
1
558
13 (9)
PF05345
1
560
14 (9)
PF13385
1
901
15 (9)
IPR011658
1
466

PilB from *Streptococcus sanguinis* is a bimodular type IV pilin with a direct role in adhesion

Claire Raynaud[a,1], Devon Sheppard[a,1], Jamie-Lee Berry[a], Ishwori Gurung[a], Vladimir Pelicic[a,b,2]

[a]MRC Centre for Molecular Bacteriology and Infection, Imperial College London, London SW7 2AZ, United Kingdom

[b]Laboratoire de Chimie Bactérienne, Aix-Marseille Université-CNRS (UMR7283), Institut de Microbiologie de la Méditerranée, 13009 Marseille, France

[1]C.R. and D.S. contributed equally to this work

[2]To whom correspondence may be addressed. Email: vladimir.pelicic@inserm.fr

Materials and methods

Strains and growth conditions

E. coli was grown in liquid or solid Lysogenic Broth (LB) (Difco) containing, when required, 100 µg/ml spectinomycin or 50 µg/ml kanamycin (both from Sigma). For purification of protein labelled with seleno-methionine (SeMet), bacteria were grown in chemically defined medium (CDM) supplemented with 20 mg/ml SeMet (Sigma).

S. sanguinis was grown on plates containing Todd Hewitt (TH) broth (Difco) and 1 % agar (Difco), incubated at 37ºC in anaerobic jars (Oxoid) under anaerobic conditions generated using Anaerogen sachets (Oxoid) (1). Liquid cultures were grown statically under aerobic conditions in THT, *i.e.*, TH broth containing 0.05 % tween 80 (Merck) to limit bacterial clumping(1). When required, 500 µg/ml kanamycin was used for selection and 15 mM *p*-Cl-Phe (Sigma) for counterselection (2). To construct the unmarked *S. sanguinis* $pilB_{D319A}$ mutant, we replaced the gene in the WT by a promoterless *pheS*aphA-3* double cassette, which confers sensitivity to *p*-Cl-Phe and resistance to kanamycin (2). To do this, we fused by splicing PCR the upstream and downstream regions flanking *pilB* to *pheS*aphA-3*, directly transformed the PCR product into the WT, and selected allelic exchange mutants on kanamycin plates. Allelic exchange was confirmed by PCR. In the second step, we replaced the *pheS*aphA-3* double cassette in this primary mutant by allelic exchange, with an unmarked $pilB_{D319A}$ mutation. To do this, we first constructed the missense mutation by site-directed mutagenesis, using as a template a pCR8/GW/TOPO (Invitrogen) derivative in which the WT gene was cloned. Then, the PCR product was directly transformed into the primary mutant, with plating on *p*-Cl-Phe-containing plates. Markerless allelic exchange mutants, which are the only one sensitive to kanamycin, were identified by re-streaking colonies on TH plates with and without antibiotic.

Protein purification

To purify native PilB, $PilB_{D319A}$ and $PilB_{vWA}$ proteins, the corresponding pET-28b derivatives were transformed in *E. coli* BL21(DE3). Transformants were grown O/N at 37ºC in liquid LB with kanamycin. The next day, this culture was diluted (1/500) in 1 l of the same medium and grown at 37ºC to an OD_{600} of 0.4-0.6. Proteins were purified as described (3).The temperature was then set to 16ºC, the culture allowed to cool for 30 min, before protein expression was induced O/N by adding 0.5 mM IPTG (Merck). The next day, cells were harvested by centrifugation at 8,000 *g* for 20 min and subjected to one cycle of -80ºC freeze/thaw in binding buffer (50 mM HEPES pH 7.4, 200 mM NaCl, 15 mM imidazole), to which we added SIGMAFAST EDTA-free protease inhibitor cocktail (Sigma). Cells were disrupted by repeated cycles of sonication, *i.e.*, pulses of 5 sec on and 5 sec off during 3-5 min, until the cell suspension was visibly less viscous. The cell lysate was then centrifuged for 30 min at 17,000 *g* to remove cell debris. The clarified lysate was first affinity-purified on an ÄKTA Purifier using His-Trap HP columns (GE Healthcare) and eluted with elution buffer (50 mM HEPES pH 7.4, 200 mM NaCl, 300 mM imidazole). Affinity-purified proteins were further purified, and simultaneously buffer-exchanged into (50 mM HEPES pH 7.4, 200 mM NaCl), by gel-filtration chromatography on an ÄKTA Purifier using a Superdex 75 10/300 GL column (GE Healthcare). Protein concentration was quantified spectrophotometrically using a NanoDrop Lite (Thermo Fisher Scientific).

To purify SeMet-labelled PilB for phasing, the corresponding pET-28b derivative was transformed in *E. coli* BL21 B834(DE3). Transformants were grown at 37ºC in liquid LB with kanamycin, until OD_{600} reached 0.6-0.7. Next, the cells were pelleted at 8,000 *g* for 5 min, and washed twice with 2 ml of CDM, which contains no Met. The pellet was then washed with 2 ml of CDM supplemented with 20 mg/ml L-Met (Sigma) and used to inoculate, at 1/200 dilution, 20 ml of CDM supplemented with 20 mg/ml Met. This culture was grown at 37°C for 16-18 h. Cells were pelleted and washed three times with CDM. Then, the pellet was re-suspended in 20 ml of

CDM supplemented with 20 mg/ml SeMet, which was used to inoculate 1 l of CDM supplemented with SeMet. Cells were grown at 37°C until OD_{600} reached 0.5-0.7. The temperature was then set to 16ºC, the culture allowed to cool for 30 min, before protein expression was induced O/N by adding 1 mM IPTG (Merck) and 4 ml of 36 % glucose (w/v). Two and half hours later, we again added 4 ml of 36 % glucose to the culture. The next day, SeMet-labelled PilB was purified as above.

Crystallisation and structure determination

The PilB crystals used for the high-resolution structure determination were obtained when the purified protein was mixed 1:1 with crystallisation liquor (0.1 M Bis-tris propane pH 7, 3 M NaCl). The lower resolution $PilB_{D319A}$ crystals were obtained with crystallisation liquor (0.1 M Bis-tris pH 6.5, 3 M NaCl). Crystals were cryoprotected with 30 % glycerol in crystallisation liquor, and flash-frozen in liquid nitrogen.

Assaying piliation of *S. sanguinis*

To purify *S. sanguinis* T4P (1, 3), bacteria grown in 100 ml THT until the OD_{600} reached 1-1.5, at which point OD were normalised, bacteria pelleted and re-suspended in pilus buffer (20 mM Tris, pH 7.5, 50 mM NaCl). This suspension was vortexed for 2 min at full speed to shear T4P. After removing bacterial cells by two centrifugation steps and filtration through a 0.22 µm pore size syringe filter (Millipore), pili were pelleted by ultracentrifugation and resuspended in pilus buffer.

Assaying twitching motility of *S. sanguinis*

To assess twitching motility on agar plates (1), bacteria grown O/N were streaked as straight lines on freshly poured TH plates containing 1 % Eiken agar (Eiken Chemicals). Plates were grown for several days at 37ºC in anaerobic condition under high humidity, which is necessary for twitching.

References

1. I. Gurung *et al.*, Functional analysis of an unusual type IV pilus in the Gram-positive *Streptococcus sanguinis*. *Mol. Microbiol.* **99**, 380-392 (2016).

2. I. Gurung, J. L. Berry, A. M. J. Hall, V. Pelicic, Cloning-independent markerless gene editing in *Streptococcus sanguinis*: novel insights in type IV pilus biology. *Nucleic Acids Res.* **45**, e40 (2017).

3. J. L. Berry *et al.*, Global biochemical and structural analysis of the type IV pilus from the Gram-positive bacterium *Streptococcus sanguinis*. *J. Biol. Chem.* **294**, 6796-6808 (2019).

4. R. A. Romijn *et al.*, Identification of the collagen-binding site of the von Willebrand factor A3-domain. *J. Biol. Chem.* **276**, 9985-9991 (2001).

5. S. Kolappan *et al.*, Structure of the *Neisseria meningitidis* type IV pilus. *Nat. Commun.* **7**, 13015 (2016).

6. P. Jones *et al.*, InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240 (2014).

113 **Table S1. Strains and plasmids used in this study.**
114

Name	Details	Source
***E. coli* strains**		
DH5α	used for cloning	
BL21(DE3)	used for protein expression/purification	
BL21 B834(DE3)	used for SeMet protein expression/purification	
***S. sanguinis* strains**		
2908	sequenced WT isolate	(1)
ΔpilB	*ΔpilB::aphA-3* deletion mutant	(1)
ΔpilD	*ΔpilD::aphA-3* deletion mutant	(1)
ΔpilB primary mutant	*ΔpilB::pheS*aphA-3* deletion mutant	this study
$pilB_{D319A}$	*pilB* point mutant expressing $PilB_{D319A}$	this study
Plasmids		
pCR8/GW/TOPO	TA cloning vector	Invitrogen
TOPO-*pheS*aphA-3*	*pheS*aphA-3* double cassette in TOPO	(2)
TOPO-*pilB*	*pilB* in TOPO	this study
TOPO-$pilB_{D319A}$	$pilB_{D319A}$ in TOPO	this study
pET-28b	T7-based expression vector	Novagen
pMK-*pilB*	codon-optimised synthetic *pilB* in pMK	(3)
pET28-*pilB*	pET-28 derivative for expressing 6His-$PilB_{36-461}$	(3)
pET28-$pilB_{D319A}$	pET-28 derivative for expressing 6His-$PilB_{D319A}$	this study
pET28-$pilB_{vWA}$	pET-2b derivative for expressing 6His-$PilB_{192-461}$	this study

115
116 *pilB*, codon-optimised synthetic gene. *pheS**, point mutant encoding $PheS_{A316G}$.

Table S2. Primers used in this study.

Name	Sequence
pilB$_{D319A}$ #1	GAAATATATCGTTCTGCTGACCGcTGGTATTCCGAATGCATATCTGG
pilB$_{D319A}$ #2	CCAGATATGCATTCGGAATACCAgCGGTCAGCAGAACGATATATTTC
pilB$_{vWA}$-pET-F	ggg**ccatgg**atcatcatcatcatcatcatCAGGGCCAGATGAATATTGC
pilB$_{vWA}$-pET-R	ccc**ggatcc**TTACGGACCGCTAACAAACC
pilB-F	TACAACTGGACCGAAGCTGG
pilB-R	TTTGGCCTATCGTTCCCACT
pilB-F1	TACAACTGGACCGAAGCTGG
pilB-R1	gttcttcaatcgttttcgtcatcaTTCCTACCTATTTATTTTTACTTCTG
pilB-F2	ttttactggatgaattgttttagAGGATTTGTGGTTTGTATCAGGG
pilB-R2	TTTGGCCTATCGTTCCCACT
pheS-F*	ATGACGAAAACGATTGAAGAAC
aph-R	CTAAAACAATTCATCCAGTAAAA
pilB$_{D319A}$ *#1*	GCTTAAATATATAGTTCTATTGACAGcTGGCATACCTAATGCTTATTTAGTAG
pilB$_{D319A}$ *#2*	CTACTAAATAAGCATTAGGTATGCCAgCTGTCAATAGAACTATATATTTAAGC

pilB, codon-optimised synthetic gene. Overhangs are in lower case. Restriction sites are in bold. Mismatches are in red.

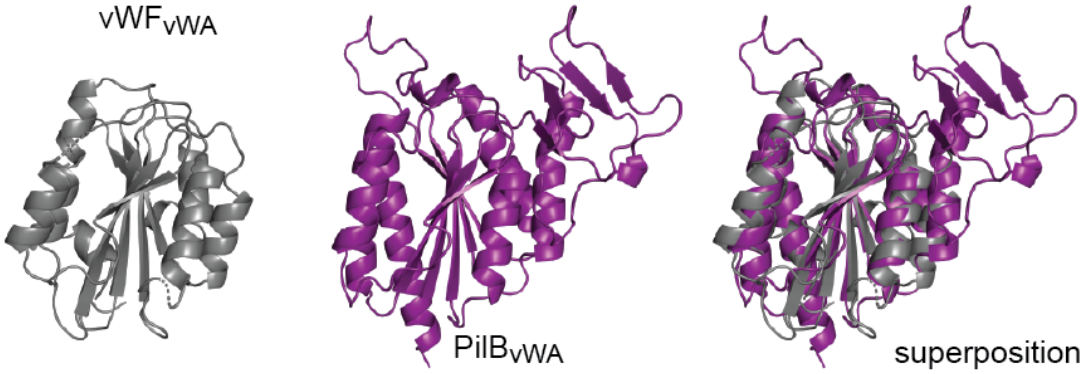


Fig. S1. Structural similarity of vWA domains in PilB and human vWF. Left, A3 vWA domain in vWF (from PDB 1FE8) (grey) (4). Center, vWA module of PilB (purple). Right, superposition of the two structures. While the two sequences share only 15.6 % sequence identity, the structures superpose with an RMSD of 1.72 Å.

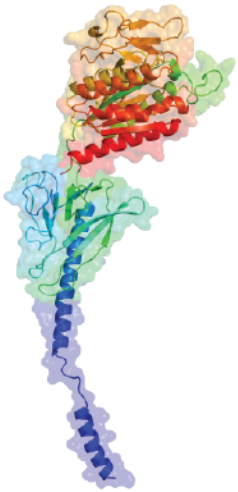


Fig. S2. 3D model of full-length PilB with a melted α1N segment. The cryo-EM structure of the *N. meningitidis* T4P (PDB 5KUA) (5) has been used as a template.

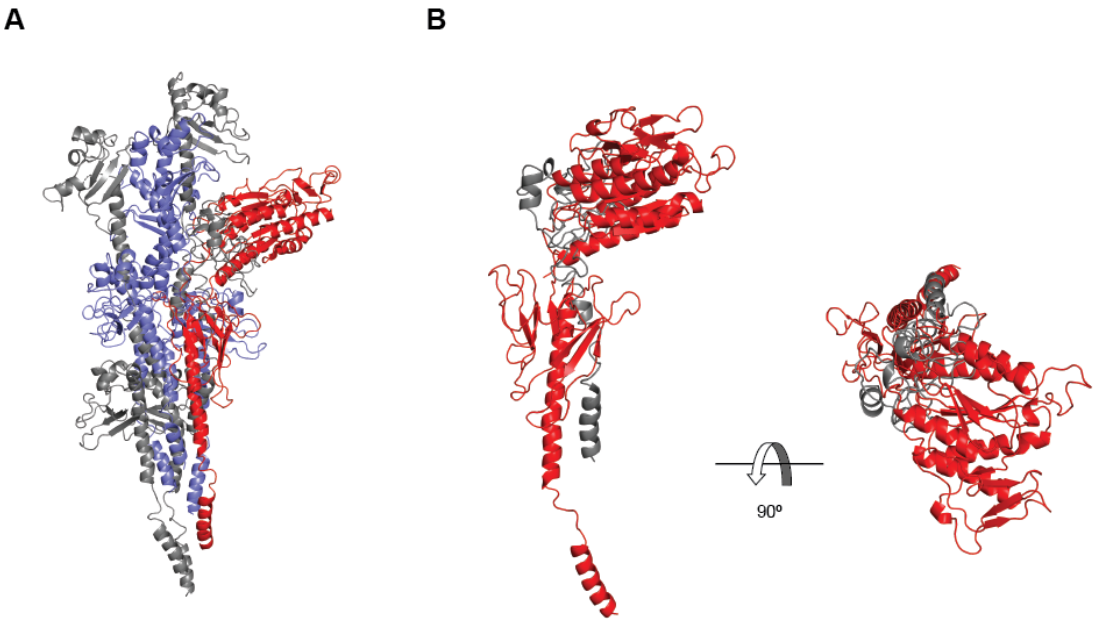


Fig. S3. Modelling of PilB in the body of a T4P leads to important steric clashes. A) Packing of PilB (red) into *S. sanguinis* T4P composed of PilE1 (blue) and PilE2 (grey). **B**) Close-up view of the important steric clashes between PilB and PilE subunit above in the filament.

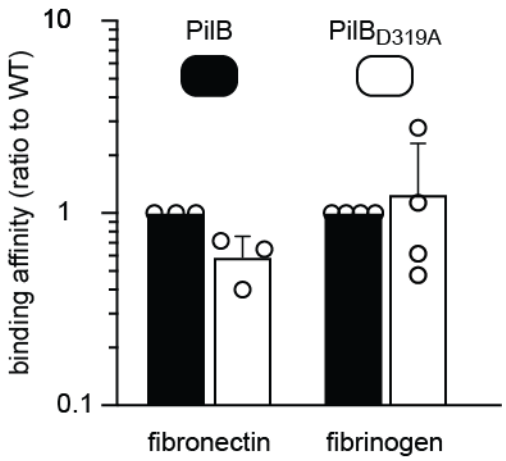


Fig. S4. Binding of PilB$_{D319A}$ to fibrinogen and fibronectin. Increasing concentrations of purified PilB were added to constant concentration of immobilised ligands, and binding was quantified by ELISA. Results are represented as Kd relative to WT, which is set to 1. Results are the average ± standard deviations from 3-4 independent experiments.

Supplementary datasets legends

Supplementary dataset 1. List of all the pilin architectures in the InterPro database. This list was generated by searching the database (November 2020) for entries displaying an IPR012902 pilin motif (6). Modular pilins display an N-terminal pilin motif together with additional module(s) not specific to T4P biology. Proteins in which the pilin motif is not N-terminal are listed as unclear.

Supplementary dataset 2. List of all the PilC/PilY1 architectures in the InterPro database. This list was generated by searching the database (November 2020) for entries displaying the IPR008707 PilC/PilY1 β-propeller domain (6). Modular PilC/PilY1 display a C-terminal IPR008707 motif together with additional module(s) not specific to T4P biology. Proteins in which the IPR008707 motif is not C-terminal are listed as unclear.