In depth biochemical and structural analysis of a Gram-positive type IV pilus

Jamie-Lee Berry[1,†], Ishwori Gurung[1,†], Jan-Haug Anonsen[2], Ingrid Spielman[3], Elliot Harper[1], Alexander M. J. Hall[1], Vivianne J. Goosens[1], Claire Raynaud[1], Nicolas Biais[3], Stephen J. Matthews[4], Vladimir Pelicic[1,*]

[1]MRC Centre for Molecular Bacteriology and Infection, Imperial College London, London, United Kingdom

[2]Department of Biosciences, IBV Mass Spectrometry and Proteomics Unit, and Center for Integrative Microbial Evolution, University of Oslo, Oslo, Norway

[3]Department of Biology, Brooklyn College of the City University of New York, New York, USA

[4]Centre for Structural Biology, Imperial College London, London, United Kingdom

[†]These authors contributed equally

*Corresponding author

E-mail: v.pelicic@imperial.ac.uk

ABSTRACT

Type IV pili (Tfp), which belong to a large class of filamentous nanomachines called type IV filaments (Tff), are surface-exposed and functionally versatile filaments, widespread in prokaryotes. Although Tfp have been extensively studied in several Gram-negative pathogens where they are key virulence factors, many aspects of their biology remain poorly understood. Here, we have performed an in depth biochemical and structural analysis of Tfp in the opportunistic pathogen *Streptococcus sanguinis*, which has recently emerged as a Gram-positive model for the study of these filaments. In particular, we have focused on the five pilins and pilin-like proteins involved in Tfp biology in this species (PilA, PilB, PilC, PilE1 and PilE2). We found that the two major pilins (PilE1 and PilE2) (i) follow widely conserved principles for processing by the prepilin peptidase PilD and for assembly into filaments, (ii) display only one of the post-translational modifications frequently found in type IV pilins, *i.e.* a processed and methylated N-terminus, (iii) are found in the same hetero-polymeric filaments, and (iv) are not functionally equivalent with respect to twitching motility. The 3D structure of PilE1, which we solved by nuclear magnetic resonance (NMR), reveals a classical pilin fold with an uncommon highly flexible C-terminus. Intriguingly, PilE1, which fits well into available Gram-negative Tfp structures, is more structurally similar to pseudopilins forming short Tff than *bona fide* Tfp-forming major pilins, further underlining the evolutionary relatedness between different Tff. Finally, we show that *S. sanguinis* Tfp also contain a low abundance of three additional proteins processed by PilD, the minor pilins PilA, PilB and PilC. By providing the first global biochemical and structural picture of a Gram-positive Tfp, our findings have important implications for widespread filamentous nanomachines.

INTRODUCTION

Type IV pili (Tfp) are thin, long and flexible surface-exposed filaments, widespread in Bacteria and Archaea, which mediate a wide array of functions including adhesion, twitching motility, DNA uptake, electric conductance *etc.* [1, 2]. Tfp are polymers of primarily one major protein subunit - a type IV pilin with distinctive N-terminal sequence motif (class III signal peptide) and 3D structure [3] - and they are assembled by a conserved set of dedicated proteins [4]. These defining features are shared by a large class of filamentous nanomachines called type IV filaments (Tff) [1]. Tff are ubiquitous in prokaryotes since genes encoding type IV pilins and filament assembly proteins are found in virtually every bacterial and archaeal genome [1].

Most of our molecular understanding of Tff biology comes from the study, over the last three decades, of bacterial Tfp in several Gram-negative human pathogens in which they are key virulence factors [4]. The following general picture has emerged. Prepilins are translocated by the Sec machinery across the cytoplasmic membrane [5, 6], where they remain embedded via a universally conserved structural feature, *i.e.* an extended and hydrophobic N-terminal α-helix (α1N) [3]. This leaves the hydrophylic leader peptide of the class III signal peptide in the cytoplasm, which is then cleaved by a dedicated membrane-bound aspartic acid protease [7] - the prepilin peptidase - generating a pool of mature pilins in the membrane ready for polymerisation into filaments. Efficient prepilin processing by the prepilin peptidase, which does not require any other protein [8], depends on the last residue of the prepilin leader peptide (a conserved Gly) [9] and two conserved catalytic Asp residues in the prepilin peptidase [7, 10]. How filaments are polymerised remains poorly understood but is is clear that this process is mediated by a multi-protein machinery in the cytoplasmic membrane [11, 12], which transmits energy generated by a cytoplasmic hexameric assembly ATPase [13] to membrane-localised mature pilins. As a result, pilins are extruded from the membrane and polymerised into helical filaments via hydrophobic packing of their protruding α1N within the filament

core [14, 15]. Finally, once Tfp reach the outer membrane, they are extruded onto the bacterial surface through a dedicated multimeric pore - the secretin [12, 16, 17]. It is important to mention that the above picture, although already complex, is oversimplified because there are multiple additional proteins that play key roles in Tfp biology, including several proteins with class III signal peptides, named minor pilins or pilin-like proteins, whose localisation and roles remain often unclear. Moreover, Tfp are highly dynamic filaments, constantly extending and retracted. Retraction has been best characterised in a sub-class of Tfp known as Tfpa, where it results from filament depolymerisation powered by the cytoplasmic hexameric ATPase PilT [18], which generates huge tensile forces of ~100 piconewton (pN)/retracting filament [19, 20]. The second sub-class of Tfp known as Tfpb, which were long thought to be static filaments, also retract and generate smaller forces (~10 pN), by an unknown PilT-independent mechanism [21, 22].

Until recently, Tfp have not been extensively studied in Gram-positive species although it has been recognised that this represents a promising new research avenue because of these bacteria simpler surface architecture (*i.e.* absence of outer membrane) [23]. Although Gram-positive Tfp were first described in species of Clostridiales [24, 25], *Streptococcus sanguinis* has since emerged as a model because it is a workhorse for genetics [26, 27]. A comprehensive genetic analysis of *S. sanguinis* Tfp [26] has revealed that they (i) are assembled by a similar machinery with fewer components (simpler) than in Gram-negative species, (ii) are retracted by a PilT-dependent mechanism, generating tensile forces very similar to those measured in Gram-negative species, and (iii) power intense twitching motility. The main peculiarity of *S. sanguinis* filaments is that they contain two major pilins in comparable amounts, rather than one as normally seen [26].

In the present study, we have focused on the pilins and pilin-like proteins involved in Tfp biology in *S. sanguinis* and have performed an in depth biochemical and structural analysis of the filaments it produces.

RESULTS

The *pil* locus in *S. sanguinis* 2908 encodes five pilin/pilin-like proteins

All known integral components of Tfp and/or Tff - generically known as pilins or pilin-like proteins - share an N-terminal sequence motif named class III signal peptide [1, 3]. It consists of a leader peptide, composed predominantly of hydrophylic amino acids (aa) ending with a conserved Gly_{-1}, followed by a stretch of 21 predominantly hydrophobic aa (except for a negatively charged Glu_5) forming the extended α-helix (α1N) that is the main assembly interface of subunits within filaments [1, 3]. Processing by the prepilin peptidase PilD occurs after Gly_{-1}. In most Tfp and/or Tff, there are mutiple pilin and/or pilin-like proteins, whose biological roles are often poorly understood [1, 3]. Bioinformatic analysis of the proteins encoded by the *pil* locus in *S. sanguinis*, which contains all the genes involved in Tfp biology [26], predicts a total of five pilins and/or pilin-like proteins (PilA, PilB, PilC, PilE1 and PilE2) (Fig. S1A). Four proteins (PilB, PilC, PilE1 and PilE2) display a canonical N-terminal IPR012902 motif [28], which is part of the class III signal peptide (Fig. 1A). Visual inspection of the sequences of the remaining proteins reveals that the N-terminus of PilA also looks like a degenerate class III signal peptide (Fig. 1B), which is not identified by any of the bioinformatic tools available, including PilFind that is dedicated to the identification of type IV pilins [29].

PilA, PilB and PilC - all of which were previously found to be essential for piliation [26] - exhibit unique sequence features, different from PilE1 and PilE2 that are the major pilin subunits of *S. sanguinis* Tfp [26]. All three proteins have much shorter leader peptides (8 or 9 residues) than PilE1/PilE2 (18 residues) (Fig. 1B). In addition, although mature PilA has a size (17 kDa) similar to previously studied pilins and pilin-like proteins [3], its putative class III signal peptide is unique for two reasons. The first residue of the predicted mature protein is a Ser, there is an unusual Pro_4, and the highly conserved Glu_5 is missing. On the other hand, while

PilB and PilC have canonical class III signal peptides (Fig. 1B), both are much larger than classical pilin-like proteins, with mature sizes of 50 and 52.7 kDa, respectively (Fig. 1A). This is explained by a highly unusual feature, *i.e.* the presence of additional bulky protein domains at the C-terminus of PilB and PilC. Namely, PilB contains a von Willebrand factor type A motif (IPR002035) [28], while PilC contains a concanavalin A-like lectin/glucanase structural domain (SSF49899) [30] (Fig. 1A).

Taken together, these findings suggest that PilA, PilB and PilC are pilin-like proteins, which might be cleaved by PilD and polymerised into *S. sanguinis* Tfp, alongside the major pilins PilE1 and PilE2.

S. sanguinis Tfp are hetero-polymers of N-terminally methylated PilE1 and PilE2

As shown previously, purified *S. sanguinis* Tfp consist predominantly of comparable amounts of two proteins, the major pilins PilE1 and PilE2 [26] that share extensive sequence identity (Fig. S2). Tfp, sheared by vortexing, were purified by removing cells/cellular debris by centrifugation before pelleting filaments by ultra-centrifugation [26]. As assessed by Coomassie staining after SDS-PAGE (Fig. 2A), such pilus preparations contain contaminants that originate from cells/cellular debris. In order to determine the precise composition of *S. sanguinis* Tfp, we endeavoured to improve the purity of our pilus preparations. To remove more cells/cellular debris, we performed an additional centrifugation step and passed the sheared filaments through a 0.22 µm syringe filter before ultra-centrifugation. As assessed by SDS-PAGE/Coomassie (Fig. 2A), filaments prepared using this enhanced purification procedure were much purer and consisted only of two bands corresponding to PilE1 and PilE2 [26], with no visible contaminants. Intriguingly, the morphology of the purified filaments also changed. While they were previously overwhelmingly thick (12 nm) and wavy [26], filaments purified using this enhanced procedure display predominantly a classical Tfp morphology [1, 4] as assessed by transmission

electron microscopy (TEM) (Fig. 2B). Indeed, they are thin (~6 nm wide), long (several µm) and flexible.

In other piliated species, major pilins frequently undergo post-translational modifications (PTM), whose biological role often remains mysterious [3]. Their N-terminal residue is most often methylated and other PTM sometimes include the addition of a variety of glycans and phosphoforms. Methylation is catalysed by PilD that is a bi-functional enzyme in many species, *e.g.* in *Pseudomonas aeruginosa* [8, 9]. A bioinformatic analysis shows that *S. sanguinis* PilD contains the IPR010627 motif catalysing N-methylation (Fig. S1B), suggesting that the N-terminal Phe_1 residue in major pilins PilE1 and PilE2 is likely to be methylated. Since other PTM cannot be inferred bioinformatically, we used top-down and bottom-up mass spectrometry (MS) analysis to map the PTM of the two major pilins of *S. sanguinis* 2908. Top-down MS analysis of purified filaments showed mainly the presence of two major proteoforms with an approximate ratio of 4:3 (PilE1:PilE2) (Fig. 2C), with deconvoluted singly charged monoisotopic masses at *m/z* 14,745.63 and 14,087.46 Da. These masses are consistent with the predicted theoretical masses of mature PilE1 (14,731.60 Da $[M+H]^+$) and PilE2 (14,073.40 Da $[M+H]^+$) with a single N-terminal methyl group (+ 14.01 Da). Bottom-up LC-MS/MS analysis of the two bands excised separately and digested in-gel identified PilE1 and PilE2 with nearly complete sequence coverage (85% and 86%, respectively). The only peptide that was not detected, despite employing combinations of several proteolytic enzymes, was the N-terminus (^{1}FTLVELIVVIIIIAIIAAVAI^{21}) (Fig. S2). These MS results strongly suggest that *S. sanguinis* pili consist mainly of a 4:3 ratio of N-terminally methylated PilE1 and PilE2 subunits.

Next, we sought to answer the question whether *S. sanguinis* Tfp are hetero-polymers of PilE1 and PilE2 or whether two homo-polymers co-exist, each composed exclusively of one major pilin. Recently, using a markerless gene editing strategy [27], we showed that *pilE1* could be engineered *in situ* to encode a protein with a C-

terminally appended 6His tag, without affecting piliation or Tfp functionality. We therefore used this property to design an affinity co-purification procedure to answer the above question (Fig. 3A). In brief, the intention was to (i) engineer markerless *pilE1* and *pilE2* mutants encoding C-terminally 6His-tagged proteins, (ii) shear the filaments, (iii) affinity-purify (pull down) sheared filaments containing the 6His-tagged pilin using cobalt-coated beads, and (iv) assess whether the untagged pilin co-purifies, suggesting that the filaments are hetero-polymers, or does not co-purify, suggesting that two distinct homo-polymers co-exist (Fig. 3A). We therefore engineered four different unmarked mutants by either fusing a 6His tag to the C-terminus of full-length PilE1 and PilE2 ($PilE1_{6His\text{-}long}$ and $PilE2_{6His\text{-}long}$) or by replacing the last seven aa in these pilins by the tag ($PilE1_{6His\text{-}short}$ and $PilE2_{6His\text{-}short}$). We first confirmed by SDS-PAGE/Coomassie analysis of purified pilus preparations that the four unmarked mutants were all piliated (Fig. 3B). These pilus preparations contained both tagged and untagged pilins as assessed by immunoblotting using antibody specific for PilE1 and PilE2 [26], or an anti-6His tag commercial antibody (Fig. 3B), indicating that tagged pilins could be assembled into Tfp. When sheared filaments affinity-purified by pull down were analysed by immunoblotting using anti-PilE1, anti-PilE2 or anti-6His antibodies, we found that in each of the four mutants the untagged pilin co-purifies with the tagged pilin (Fig. 3B). Importantly, wild-type (WT) untagged filaments cannot be affinity-purified using this procedure, indicating that pull down is a 6His tag-specific process (Fig. 3B). These findings strongly suggest that 2908 Tfp are hetero-polymers composed of PilE1 and PilE2.

Taken together, these findings show that *S. sanguinis* Tfp have a canonical Tfp morphology and are hetero-polymers composed of a 4:3 ratio of two major pilins PilE1 and PilE2, which both harbour a single PTM, *i.e.* a methylated N-terminus.

Major pilin processing and/or assembly into filaments in Gram-positive Tfp follow widely conserved principles

Mutagenesis studies of major pilins in several Tfp and/or Tff have defined residues in class III signal peptides that are key for processing by PilD and/or assembly into filaments [9, 31-33]. Although, there are species-specific differences, a common rule has emerged concerning two of the most highly conserved residues Gly_{-1} and Glu_5 [1]. It appears that Gly_{-1} is crucial for processing by PilD, while Glu_5 is dispensable for processing but key for filament assembly because it establishes a salt bridge with the methylated N-terminal Phe_1 of the neighbouring subunit α1N [14, 15]. Since the importance of these residues has not been assessed for Gram-positive Tfp, we tested it in the context of *S. sanguinis* Tfp. Because PilE1 and PilE2 have identical N-termini (Fig. 1B and Fig. S2), we focused our efforts on PilE1 and used our gene editing strategy [27] to construct markerless mutant strains expressing PilE1 variants in which the Gly_{-1} and Glu_5 residues would be mutated. Next, we tested whether these mutants proteins were processed by PilD and assembled into filaments. Pilin processing was assessed by immunoblotting using the anti-PilE1 antibody [26] in whole-cell protein extracts and by comparison to the WT and *ΔpilD* mutant, *i.e.* processed PilE1 is 14.7 kDa while the unprocessed protein is 16.9 kDa. Assembly into Tfp was assessed by immunoblotting on purified filaments. Originally, we replaced Gly_{-1} by an Ala ($PilE1_{G-1A}$), which we found to have no effect on processing (Fig. 4A) or assembly into filaments (Fig. 4B). A similar finding was reported for *P. aeruginosa* major pilin, and was attributed to the small size of Ala since substitutions with residues with bulkier side chains abolished pilin processing [9]. Accordingly, when Gly_{-1} was replaced by a Ser ($PilE1_{G-1S}$), PilE1 could not be processed (Fig. 4A) and could not be polymerised into filaments, which consisted exclusively of PilE2 (Fig. 4B). As for the other highly conserved Glu_5 residue, when it was replaced by an Ala ($PilE1_{E5A}$), PilE1 processing was unaffected (Fig. 4A) but the protein could not be polymerised into filaments, which again consisted exclusively of PilE2 (Fig. 4B). These findings confirm that the widely conserved principles defining how major pilins are processed and assembled [1], apply to *S. sanguinis* Tfp as well despite their

peculiar hetero-polymeric structure. This further strengthens the suitability of *S. sanguinis* as a Gram-positive model species to study Tfp biology.

3D structure of PilE1 reveals a pilin fold with an uncommon highly flexible C-terminus, which more closely resembles pseudopilins

Next, to improve our structural understanding of *S. sanguinis* Tfp, we endeavoured to solve the 3D structure of its major pilins. To facilitate purification, we expressed in *Escherichia coli* the soluble portions of PilE1 (112 aa) and PilE2 (105 aa) fused to an N-terminal 6His tag (Fig. S2). These soluble portions exclude the first 27 residues of the mature proteins, which mainly form the protruding part (α1N) of the hydrophobic N-terminal α-helix in type IV pilins [1, 3]. This procedure allowed us to purify well-folded and soluble proteins using a combination of affinity and gel-filtration chromatographies. Since PilE2 shares 78% sequence identity with PilE1 (Fig. S2), we decided to focus on the longest PilE1 and determine its structure by NMR. We isotopically labelled 6His-PilE1 with ^{13}C and ^{15}N for NMR assignment (Table S1) and obtained a high-resolution NOE-derived structure in solution (Fig. 5). This structure (Fig. 5A) revealed that PilE1 adopts the classical type IV pilin fold [1, 3]. It exhibits a long N-terminal α-helix packed against a β-meander consisting of three anti-parallel β-strands, flanked by distinctive structurally variable "edges" (Fig. 5A), which usually differ between pilins [1, 3]. While the α1β1-loop connecting α1 and β1, with two short α-helices, is unexceptional except maybe for its length (50 aa), the C-terminus of the protein (after β3) is striking. Indeed, unlike in other pilins where this region is always stabilised by being "stapled" to the last β-strand of the β-meander either by a disulfide bond, a network of hydrogen bonds, or a calcium-binding site [1, 3], the 10 aa-long C-terminus in PilE1 is unstructured and highly flexible. Indeed, superposition of the structures within the NMR ensemble reveals that although these superpose well onto each other up to the last strand of the β-meander (β3), their C-termini exhibit widely different positions (Fig. 5B). Intriguingly, when compared to 3D

structures in the Protein Data Bank (PDB), PilE1 was found to be most similar to pseudopilins that form short filaments in alternative Tff rather than *bona* fide major pilus subunits of Tfp. As can be seen in Fig. 5C, the structures of PilE1 and PulG - the major pseudopilin in *Klebsiella oxytoca* type II secretion system (T2SS) - superpose [34] fairly well onto each other, with a root mean square deviation (rmsd) of 5.75 Å for all backbone atoms.

Owing to their very high sequence identity (Fig. S2), we used our PilE1 structure as a template to produce a reliable structural model of PilE2 (Fig. 6A). PilE2 structural model was found to be virtually identical to PilE1 structure except for a shorter α1β1-loop (Fig. 6B), which is explained by the fact that eight residues at the C-terminus of this loop in PilE1 are missing in PilE2 (Fig. S2). Finally, after producing full-length models of PilE1 and PilE2 using the full-length gonococcal pilin [35, 36] as a template for the missing N-terminal α1N, we were able to model packing of these pilins within recently determined structures of Tfpa [14, 15] (Fig. 6C). This revealed that PilE1and PilE2 fit readily into these filaments, which have a similar morphology to *S. sanguinis* filaments. This finding supports the notion that polymerisation of pilins into filaments in Gram-positive species also occurs via hydrophobic packing of their α1N within the filament core.

Together, these findings show that *S. sanguinis* Tfp obey structural principles common to this class of filaments, but nevertless display some intriguing peculiarities. While *S. sanguinis* major pilins adopt the canonical type IV pilin fold, their C-terminus appears to be highly flexible, which is uncommon. Moreover, while *S. sanguinis* major pilins fit readily within available Tfp structures, they are more similar structurally to pseudopilins than to *bona fide* Tfp-forming major pilins.

PilA, PilB and PilC are minor pilin components of *S. sanguinis* Tfp

In all Tff systems, there are in addition to the major pilins several proteins with class III signal peptides whose role and localisation are often unclear [3]. To start with the

experimental characterisation of PilA, PilB and PilC in *S. sanguinis*, we purified each protein and used it to immunise rabbits, to generate antisera. Immunoblotting using whole-cell protein extracts confirmed that the three proteins are expressed by *S. sanguinis* since they could be detected in WT, but not in the corresponding deletion mutants (Fig. 7A). All antibodies are thus specific for the proteins against which they were raised. We then tested whether PilA, PilB and PilC were processed by PilD, which for PilA was far from certain considering its highly degenerate class III signal peptide (Fig. 1B). Procesing by PilD, which removes the leader peptide, is expected to generate mature proteins of 17, 50.5 and 52.8 kDa for PilA, PilB and PilC, respectively, shorter than their respective 18, 51.5 and 53.9 kDa precursors. Importantly, immunoblots confirmed that all three proteins are cleaved by PilD as indicated by the detection of proteins of slightly higher molecular weight in a *ΔpilD* mutant, with masses consistent with those expected for unprocessed precursors (Fig. 7A). Next, we took advantage of our ability to prepare highly pure Tfp, to determine whether PilA, PilB and PilC could be detected in pilus preparations by immunoblotting. As can be seen in Fig. 2B, although only PilE1 and PilE2 could be detected by SDS-PAGE/Coomassie analysis of pure Tfp, PilA, PilB and PilC are all readily detected by immunoblotting in these preparations (Fig. 7B). Importantly, co-purification with the filaments was dependent upon processing by PilD since PilA, PilB and PilC were not detected in pilus preparations made from a *ΔpilD* mutant (Fig. 7B). In conclusion, findings that PilA, PilB and PilC are cleaved by PilD and co-purify with Tfp support the view that these three proteins are minor (low abundance) pilin components of *S. sanguinis* Tfp, which are likely assembled into filaments in a similar fashion to the major subunits PilE1 and PilE2.

Homo-polymeric filaments composed only of PilE1 or PilE2 are able to promote twitching motility, but at different speeds

As previously reported, single *ΔpilE1* and *ΔpilE2* mutants produce filaments

consisting of the remaining pilin, while a double *ΔpilE1ΔpilE2* mutant is non-pilated [26]. We therefore wondered whether the homo-polymeric filaments in *ΔpilE1* and *ΔpilE2* mutants are functional and, if so, whether the subtle structural differences between the two pilins (see Fig. 6B) would have a functional impact. We compared the ability of the homo-polymeric filaments produced by single *ΔpilE1* and *ΔpilE2* mutants to mediate twitching motility. We first tested whether these mutants still exhibited thin spreading zones around bacteria grown on agar plates, which was found to be the case (Fig. 8A). This confirms that Tfp consisting exclusively of PilE1 or PilE2 are functional. Motility was next assessed quantitatively at a cellular level by tracking under the microscope the movement of small chains of cells attached to glass coverslips (Fig. 8B). As previously reported for the WT [26], both mutants showed "train-like" directional motion mainly parallel to the long axis of bacterial chains. Short duration movies illustrating the movement of *ΔpilE1* and *ΔpilE2* are included as supplementary information (Movies S1 and S2). Measurement of instantaneous velocities revealed that, while the WT moved at 694 ± 4 $nm \cdot s^{-1}$ (mean ± standard error, n=22,957) consistent with previous measurements [26], the mutants moved at 462 ± 2 $nm \cdot s^{-1}$ (n=37,001) for *ΔpilE1*, and 735 ± 3 $nm \cdot s^{-1}$ (n=22,231) for *ΔpilE2* (Fig. 8B). These differences in speed, which are statistically significant, show that the two pilins are not functionally equivalent with respect to twitching motility.

DISCUSSION

Their ubiquity in prokaryotes [1] makes Tff a hot topic for research. A better understanding of the molecular mechanisms governing Tff biology, *e.g.* how filaments are assembled and how they can mediate a vast array of seemingly unrelated properties, has potential implications for human health and nanotechnology. Perhaps one of the reasons of our limited understanding of Tff biology is that they have historically been studied in just a few Gram-negative bacterial species, all belonging to the same phylum (Proteobacteria) [4]. Therefore, it is well accepted that studying Tff in phylogenetically distant species has the potential to move the field forward, which has in recent years sparked studies in Archaea [2] and in distant phyla of Bacteria [23]. One of the most promising new Tfp models that has emerged is the Gram-positive opportunistic pathogen *S. sanguinis* [26, 27]. A recent systematic genetic analysis of Tfp biology in this species - the first to be realised in a non Proteobacterium - showed that *S. sanguinis* uses a simpler (fewer compoenents) machinery to assemble canonical Tfpa that generate high tensile forces and power twitching motility [26]. In this report, we performed an in depth biochemical and structural analysis of *S. sanguinis* filaments, which led to the notable findings discussed below.

The first important achievement in this study is the etablishment of one of the most complete biochemical pictures of a Tfp, which brings us one step closer to a complete understanding of all the integral components of these filaments and how they impact Tfp biology. Criticallly, this picture confirms the above observed trend [26] that *S. sanguinis* Tfpa are simpler filaments, whith fewer components. Indeed, in Gram-negative Tfpa models, there are in addition to the major pilin, 7-8 pilin-like proteins possessing class III signal peptides [4]. For example, in *Neisseria meningitidis*, there are four conserved pilin-like proteins required for piliation (PilH, PilI, PilJ and PilK) whose role (priming filament assembly) [37] and localisation (at the tip of the pili or distributed throughout the filaments) [38, 39] remain uncertain,

and three species-specific minor pilins (ComP, PilV and PilX) that are dispensable for piliation but modulate Tfp-associated functions [40-42]. In contrast, in *S. sanguinis*, there are besides PilE1 and PilE2 only three Pil proteins possessing class III signal peptides (PilA, PilB and PilC). As shown here, all these proteins are efficiently recognised and post-translationally modified by PilD, which processes their N-terminal leader peptides and (most likely) methylates the first residue of the resulting mature proteins, as demonstrated for PilE1 and PilE2. Processing follows conserved principles that have been originally established in Gram-negative models [1]. No other PTM, which frequently decorate major pilins in Gram-negative Tfp (*e.g.* glycans or phosphoforms), are found on *S. sanguinis* PilE1 and PilE2. Importantly, the use of highly pure pilus preparations showed that these five proteins are clearly Tfp subunits. PilE1 and PilE2 are the two major pilins, while PilA, PilB and PilC are three minor (low abundance) pilins. Although the arrangement of the major pilins in the filaments (geometric or stochastic) remains to be determined, this study makes it clear that *S. sanguinis* Tfp are hetero-polymeric structures containing comparable amounts of PilE1 and PilE2, a property previously unreported for Tff. The reason for this peculiarity remains unclear, however, since the homo-polymeric Tfp assembled by single *ΔpilE1* and *ΔpilE2* mutants are both functional as they can power efficient twitching motility. One possible explanation might be that PilE1 and PilE2, which have porbably evolved by duplication, are important for optimal stability of *S. sanguinis* Tfp as *ΔpilE1* and *ΔpilE2* mutant apparently produce less filaments than WT. On the other hand, what could be the role of the minor subunits of *S. sanguinis* Tfp (PilA, PilB and PilC)? Although they are required for piliation, these three proteins are unlikely to prime filament assembly because they are unrelated to the four conserved pilin-like proteins (PilH, PilI, PilJ and PilK) involved in this process in Gram-negative Tfp [37]. Rather, PilA, PilB and PilC might be essential for filament stability and are likely to modulate Tfp-associated functions, which is supported by the presence of additional C-terminal domains in PilB and PilC, a previously

unreported property for pilin-like proteins. Interestingly, both the von Willebrand factor type A motif (found in PilB) and the concanavalin A-like lectin/glucanase structural domain (found in PilC) are frequently involved in binding protein or carbohydrates ligands, suggesting that both PilB and PilC might be involved in adhesion, a property frequently associated with Tfp in many species [4]. Strikingly, a von Willebrand factor type A motif has been found in a minor adhesive subunit of *Streptococcus agalactiae* pili unrelated to Tfp [43], which might point to a rare case of convergent evolution between different types of pili.

The structural information generated on *S. sanguinis* Tfp is the second notable achievement in this study. Our new pilus purification strategy shows that the morphological features of *S. sanguinis* filaments are canonical of Tfpa (~6 nm width, several μm length and high flexibility) [4]. It is likely that the 12 nm thick and wavy filaments purified previously [26] were damaged during ultra-centrifugation by the presence of cells/cellular debris. Our high-resolution NMR structure of the globular domain of PilE1 confirms that Gram-positive Tfp major subunits adopt the classical type IV pilin fold, with an extended N-terminal α1 helix packed against a β-meander consisting of anti-parallel β-strands [1, 3]. The 3D structure of PilE2 is expected to be virtually identical, owing to its almost 80% sequence identity to PilE1. The full-length PilE1/PilE2 are therefore expected to adopt the canonical "lollipop" pilin structure [35, 36], since the missing hydrophobic α1N can reliably be modelled as a protruding α-helix. The α1N is likely to adopt a gentle S-shaped curve like in the full-length type IV pilin structures available, such as gonococcal major pilin [35, 36], because the helix-breaking Pro_{22} which introduces a kink in the gonococcal pilin is conserved in PilE1/PilE2. Importantly, PilE1/PilE2 fit well in recent cryoelectron microscopy reconstructions of several Gram-negative Tfpa [14, 15], which despite different helical parameters (rise and rotation) due to different arrangements of the major pilin globular domains on the filament surface, display very similar packing of the conserved α1N in the core of the filaments. Owing to the conservation of the helix-

breaking Pro_{22}, it is likely that the α1N helix will be partially melted between Ala_{14} and Pro_{22} as the pilins in the above reconstructions [14, 15], which is thought to provide flexibility and elasticity to the filaments [44]. In addition, this would allow the formation of a salt bridge between Glu_5 and the methylated Phe_1 (both conserved in PilE1 and PilE2) of neighbouring pilin in the 1-start helix. Minor subunits PilB and PilC, which have canonical class III signal peptides similar to PilE1/PilE2 are likely to assemble within filaments in a very similar fashion. It is unclear, however, whether the α1N helix will be partially melted since the helix-breaking Pro_{22} is absent in these proteins, and it remains to be seen how the bulky C-terminal domains, which appear to have been "grafted" by evolution on a pilin moiety, would be exposed on the surface of the filaments. As for PilA, its unique class III signal peptide, makes it difficult to predict how its α1N will be packed in the filament core, especially because of the unusual, potentially helix-breaking Pro_4. Critically, our high-resolution structure of PilE1 also challenges two common assumptions in the field based on previous structural findings [1, 3]. First, in contrast to all previously available pilin structures, it appears that the C-terminus in PilE1 is unstructured and highly flexible. This explains why it is a permissive insertion site and why it even can be deleted and replaced by a 6His tag [27], without interfering with the ability of PilE1/PilE2 to be polymerised into filaments. Therefore the common assumption that the C-terminus of major pilins must be stabilised (by a disulfide bond, a network of hydrogen bonds, a calcium-binding site *etc.*) to preserve pilin integrity and ability to be polymerised [3], is not always true. *S. sanguinis* might be an exception, however, since major pilins of *Clostridiuma difficile* Tfp apparently use unique strategies to stabilise their C terminus [45]. Second, PilE1 is most similar to pseudopilins that form short filaments in alternative Tff, such as PulG from *K. oxytoca* T2SS [34]. This indicates that a major pilin 3D structure cannot be used to predict whether the subunit will form pseudopili or *bona fide* Tfp, which perhaps blurs the lines between different Tff and suggests an even closer evolutionary relationship between these filamentous nanomachines.

In conclusion, by providing an uprecedented global view of a Gram-positive Tfp, this study further cements *S. sanguinis* as a model species which is fast closing the gap with historic Gram-negative Tfp models. Together with our recent reports [26, 27] and *S. sanguinis* exquisite genetic tractability, these findings pave the way for further investigations, which will undoubtedly contribute to improve our fragmentary understanding of a fascinating filamentous nano-machine almost universal in prokaryotes.

MATERIALS & METHODS

Strains and growth conditions

Strains and plasmids that were used in this study are listed in Table S2. For cloning, we used *E. coli* DH5α. *E. coli* BL21(DE3) was used for protein expression and purification. *E. coli* strains were grown in liquid or solid Lysogenic Broth (LB) (Difco) containing, when required, 100 µg/ml spectinomycin or 50 µg/ml kanamycin (both from Sigma). The WT *S. sanguinis* 2908 strain and deletion mutants were described previously [26, 27]. Bacteria were grown on plates containing Todd Hewitt (TH) broth (Difco) and 1% agar (Difco), which were incubated at 37ºC in anaerobic jars (Oxoid) under anaerobic conditions generated using Anaerogen sachets (Oxoid). Liquid cultures were grown statically under aerobic conditions in THT, *i.e.* TH broth containing 0.05% tween 80 (Merck) to limit bacterial clumping. When required, 500 µg/ml kanamycin (Km) (Sigma) was used for selection. For counterselection, we used 15 mM *p*-Cl-Phe (Sigma) [27].

Chemically competent *E. coli* cells were prepared as described [46]. DNA manipulations were done using standard molecular biology techniques [47]. All PCR were done using high-fidelity DNA polymerases from Agilent (see Table S3 for a list of primers used in this study). *S. sanguinis* genomic DNA was prepared from overnight (O/N) liquid cultures using the kit XIT Genomic DNA from Gram-Positive Bacteria (G-Biosciences). Strain 2908, which is naturally competent, was transformed as described elsewhere [26, 27]. In brief, bacteria grown O/N in THTS - THT supplemented with 2.5% heat-inactivated horse serum (Sigma) - were back-diluted in THTS and incubated at 37ºC for 1-1.5 h. When transforming DNA was added, competence was induced using a synthetic competence stimulating peptide (Peptide Protein Research). After 1 h incubation at 37ºC, transformants were selected by plating on suitable agar plates.

Unmarked *S. sanguinis* mutants in *pilE1* and *pilE2 in situ* used in this study

were constructed using a recently described two-step, cloning-independent, gene editing strategy [27]. In brief, in the first step, the target gene was cleanly replaced in the WT, by allelic exchange, by a promoterless *pheS*aphA-3* double cassette, which confers sensitivity to *p*-Cl-Phe and resistance to kanamycin. To do this, a splicing PCR (sPCR) product fusing the upstream (amplified with primers F1 and R1) and dowstream (amplified with F2 and R2) regions flanking the target gene to *pheS*aphA-3* (amplified with *pheS*-F and *aph*-R) was directly transformed into the WT, and allelic exchange mutants were selected on Km-containing plates. Allelic exchange was confirmed for a couple of transformants by PCR. In the second step, the *pheS*aphA-3* double cassette was cleanly replaced in this primary mutant, by allelic exchange, by an unmarked mutant allele of the target gene (see below). To do this, a sPCR product fusing the mutant allele to its upstream and dowstream flanking regions was directly transformed into the primary mutant and allelic exchange mutants were selected on *p*-Cl-Phe-containing plates. Markerless allelic exchange mutants, which are Km^S, were identified by re-streaking *p*-Cl-Phe^R transformants on TH plates with and without Km. The *pilE1* and *pilE2* mutant alleles encoding proteins with a C-terminal 6His tag were engineered by sPCR. We made two constructs for each gene: a LONG construct in which we fused the 6His tag to the C-terminus of the full-length protein (sPCR with F1/R3 and F3/R2), and a SHORT construct in which we replaced the last seven aa in the pilins by the tag (sPCR with F1/R4 and F4/R2). To construct the missense mutants in *pilE1*, we used as a template a pCR8/GW/TOPO plasmid in which the WT gene (amplified with F and R) was cloned (Table S2) and the Quickchange site-directed mutagenesis kit (Agilent) (with complementary primers #1 and #2). Then the sPCR product for transformation in the primary mutant was produced by fusing the mutant allele (amplified with F and R) to flanking regions upstream (amplified with F1 and R5) and dowstream (amplified with F5 and R2).

SDS-PAGE, antisera and immunoblotting

S. sanguinis whole-cell protein extracts were prepared using a FastPrep-24 homogeniser (MP Biomedicals) and quantified as described elsewhere [26]. Separation of the proteins by SDS-PAGE, subsequent blotting to Amersham Hybond ECL membrane (GE Healthcare) and blocking were carried out using standard molecular biology techniques [47]. To detect PilE1 and PilE2, we used previously described primary rabbit anti-peptide antibodies [26]. Antisera against PilA, PilB and PilC were produced for this study by Eurogentec by immunising rabbits with purified recombinant proteins (see below). These proteins were then used to affinity-purify the antibodies. Primary antibodies were used at between 1/2,000 and 1/5,000 dilutions, while the secondary antibody, an ECL HRP-linked anti-rabbit antibody (GE Healthcare), was used at 1/10,000 dilution. Amersham ECL Prime (GE Healthcare) was used to reveal the blots. To detect His-tagged proteins, we used a commercial HRP-linked anti-6His antibody (Sigma) at 1/10,000 dilution.

Tfp purification and visualisation

S. sanguinis 2908 Tfp were purified as described elsewhere with minor modifications [26]. Liquid cultures (10 ml), grown O/N in THT, were used the next day to re-inoculate 90 ml of THT and grown statically until the OD_{600} reached 1-1.5, at which point OD were normalised, if needed. Bacteria were pelleted at 4ºC by centrifugation for 10 min at 6,000 *g* and pellets were re-suspended in 2 ml pilus buffer (20 mM Tris pH 7.5, 50 mM NaCl). This suspension was vortexed for 2 min at full speed to shear Tfp. Bacteria were then pelleted as above, and supernatant containing the pili was transferred to a new tube. This centrifugation step was repeated, before the supernatant was passed through a 0.22 µm pore size syringe filter (Millipore) to remove residual cells and cellular debris. Pili were then pelleted by ultra-centrifugation as described [26], resuspended in pilus buffer, separated by SDS-PAGE and gels were stained using Bio-Safe Coomassie stain (Bio-Rad). Purified

filaments were visualised by TEM after negative staining as described elsewhere [26].

Proteomics and mass spectrometric analysis of purified Tfp

For the bottom-up MS analysis of purifed Tfp, we carefully excised PilE1 and PilE2 protein bands from Coomassie-stained gels and generated enzymatically derived peptides employing four separate enzymes. In brief, as previously described [48], gel pieces were destained and digested O/N with 1 µg trypsin or Lys-C in (50 mM Na_2HCO_3, pH 7.8) at 37ºC, 1 µg chymotrypsin in (100 mM Tris(hydroxymethyl)aminomethane, 10 mM $CaCl_2{\cdot}2H_2O$, pH 7.8) at 25ºC, or 0.3 µg AspN in (50 mM Tris(hydroxymethyl)aminomethane, 2.5 mM $ZnSO_4{\cdot}7H_2O$, pH 8.0) at 25ºC. Generated peptides, which were extracted as previously described [48], were vacuum-concentrated, dissolved in loading buffer (2% acetonitrile, 1% trifluoroacetic acid) and desalted using ZIP-TIP tips as instructed by the manufacturer (Millipore). The peptides were eluted with (80% acetonitrile, 1% trifluoroacetic acid) and vacuum-concentrated. Dried peptide samples were dissolved in 10 µl (2% acetonitrile, 1% formic acid) before analysis by reverse phase LC-MS/MS. Redissolved sample (2-5 µl) were injected into a Dionex Ultimate 3000 nano-UHPLC system (Sunnyvale) coupled online to a QExactive mass spectrometer (Thermo Fisher Scientific) equipped with a nano-electrospray ion source. LC separation was achieved with an Acclaim PepMap 100 column (C18, 3 µm beads, 100 Å, 75 µm inner diameter, 50 cm) and a LC-packing trap column (C18, 0.3 mm inner diameter, 300 Å). The flow rate (15 µl/min) was provided by the capillary pump. A flow rate of 300 nl/min was employed by the nano pump, establishing a solvent gradient of solvent B from 3 to 5% in 5 min and from 5 to 55% in 60 min. Solvent A was (0.1% formic acid, 2% acetonitrile), while solvent B was (0.1% formic acid, 90% acetonitrile). The mass spectrometer was operated in data-dependent mode to automatically switch between MS and MS/MS acquisition. Survey full scan MS

spectra (from *m/z* 200 to 2,000) were acquired with the resolution R = 70,000 at *m/z* 200, with an automated gain control (AGC) target of 10^6, and ion accumulation time set at 100 msec. The seven most intense ions, depending on signal intensity (intensity threshold 5.6 10^3) were considered for fragmentation using higher-energy collisional induced dissociation (HCD) at R = 17,500 and normalised collision energy (NCE)=30. Maximum ion accumulation time for MS/MS spectra was set at 180 msec. Dynamic exclusion of selected ions for MS/MS were set at 30 sec. The isolation window (without offset) was set at *m/z* 2. The lock mass option was enabled in MS mode for internal recalibration during the analysis.

To perform a complete top-down MS analysis of the PilE1 and PilE2 proteoforms present in purified *S. sanguinis* Tfp, three separate methods were employed. For the first method, sample preparation and top-down ESI-MS on a LTQ Orbitrap were performed as previously described [49], and intact protein mass spectra were acquired with a resolution of 100,000 at *m/z* 400. For the second method, after initial preliminary testing of a gradient of solvent B from 3 to 55% in 10 min, and from 55 to 85% in 12-35 min, the data was acquired on a LTQ Orbitrap operated in positive ionisation mode in the data-dependent mode, to automatically switch between MS and MS/MS acquisition. Survey full scan MS spectra (from *m/z* 200 to 2,000) were acquired with a resolution R = 100,000 at *m/z* 400, with AGC target of 10^6 and ion accumulation time set at 100 msec. The two most intense ions, depending on signal intensity (intensity threshold 5.6 10^3) were considered for fragmentation using HCD at R = 17,500 and NCE=30. The lock mass option was enabled in MS mode for internal recalibration during the analysis. For the third method, a direct injection nano-ESI top-down procedure was employed. Briefly, the Dionex Ultimate 3000 nano-UHPLC system coupled to the LTQ Orbitrap was reconfigured so that after sample loading onto the loop, valve switching allowed the nanopump to inject directly the sample from the loop into the LTQ Orbitrap MS using an isocratic gradient of (10% methanol, 10% formic acid). Data acquisition was done

manually from the LTQ Tune Plus (V2.5.5 sp2) using R=100,000 and NCE=400.

Data processing and analysis was done as follows. Bottom-up MS data were analysed using MaxQuant (v 1.5.2) and the Andromeda search engine against an in-house generated *S. sanguinis* whole proteome database, and a database containing common contaminants. Trypsin, chymotrypsin, Lys-C, AspN and no enzyme (no restriction) were selected as enzymes, allowing two missed cleavage sites. We applied a tolerance of 10 ppm for the precursor ion in the first search, 5 ppm in the second, and 0.05 Da for the MS/MS fragments. In addition to methionine oxidation, protein N-terminal methylation was allowed, in a separate search, as a variable modification. The minimum peptide length was set at 4 aa, and the maximum peptide mass at 5.5 kDa. False discovery rate was set at <0.01. Deconvolution of the PilE1 and PilE2 mass envelopes from top-down analysis was done as previously described [50]. Deconvoluted protein masses are reported as monopronated [M+H+]. Theoretical masses of PilE1 and PilE2 were determined from available sequences.

Twitching motility assays

Twitching motility was assessed macroscopically on agar plates as described elsewhere [26]. Briefly, bacteria were re-streaked as straight lines on freshly poured TH plates containing 1% Eiken agar (Eiken Chemicals), which were incubated up to several days under anaerobic conditions in a jar, in the presence of water to ensure high humidity. Motility was analysed microscopically as described elsewhere [26]. In brief, bacteria resuspended in THT were added into an open experimental chamber with a glass bottom and grown for 2 h at 37ºC in presence of 5% CO_2. The chamber was then transferred to an upright Ti Eclipse microscope (Nikon) with an environment cabinet maintaining the same growth conditions, and movies of the motion of small bacterial chains were obtained and analysed in ImageJ, as described [26]. Cell speed was measured from collected trajectories using Matlab.

Protein purification

To produce pure PilA, PilB and PilC proteins for generating antibodies, we cloned the corresponding genes in pET-28b (Novagen) (Table S2). The forward primer was designed to fuse a non-cleavable N-terminal 6His tag to the soluble portion of these proteins, *i.e.* excluding the leader peptide and the predicted hydrophobic α1N. For *pilA*, we amplified the gene from 2908 genome, while for *pilB* and *pilC*, we used synthetic genes (GeneArt), codon-optimised for expression in *E. coli*. Recombinant proteins were purified using a combination of affinity and gel-filtration chromatographies as follows. An O/N liquid culture, in selective LB, from a single colony of *E. coli* BL21 (DE3) transformed with the above expression plasmids, was back-diluted (1/500) the next day in 1 l of the same medium and grown to an OD_{600} of 0.4-0.6 on an orbital shaker. The temperature was then set to 16ºC, the culture allowed to cool for 30 min, before protein expression was induced O/N by adding 0.4 mM IPTG (Merck Chemicals). The next day, cells were harvested by centrifugation at 8,000 *g* for 20 min and subjected to one freeze/thaw cycle in binding buffer A [50 mM HEPES pH 7.4, 200 mM NaCl, 10 mM imidazole, 1 x 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 then mixed with two ml of Ni-NTA agarose resin (Qiagen), pre-washed in binding buffer A, and incubated for 2 h at 4ºC with gentle agitation. This chromatography mixture was then filtered through a Poly-Prep gravity-flow column (BioRad) and washed several times with binding buffer A, before the protein was eluted with elution buffer A (50 mM HEPES pH 7.4, 200 mM NaCl, 500 mM imidazole, 1 x SIGMAFAST EDTA-free protease inhibitor cocktail). The 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 Akta Purifier using a Superdex 75 10/300 GL column (GE

Healthcare).

For structural characterisation of PilE1, we cloned, as above, the portion of *pilE1* encoding the soluble portion of this protein in pET-28b in order to produce a protein with a non-cleavable N-terminal 6His tag (Table S2). An O/N pre-culture in LB was back-diluted 1/50 into 10 ml selective M9 minimal medium, supplemented with a mixture of vitamins and trace elements. This was grown to saturation O/N at 30ºC in an orbital shaker, then back-diluted 1/500 into 1 l of the same medium containing ^{13}C D-glucose and ^{15}N NH_4Cl for isotopic labelling. Cells were grown in an orbital shaker at 30ºC until the OD_{600} reached 0.8, then 0.4 mM IPTG was added to induce protein production O/N at 30ºC. As above, cells were then harvested and disrupted in binding buffer B (50 mM Tris-HCl pH 8.5, 200 mM NaCl, 10 mM imidazole, 1 x SIGMAFAST EDTA-free protease inhibitor cocktail). The protein was first purified by affinity chromatography and eluted in (50 mM Tris-HCl pH 8.5, 200 mM NaCl, 200 mM imidazole, 1 x SIGMAFAST EDTA-free protease inhibitor cocktail). It was then further purified and buffer-exchanged into (50 mM Na_2HPO_4/NaH_2PO_4 pH 6, 200 mM NaCl) by gel-filtration chromatography using a Superdex 75 10/300 GL column.

NMR structure determination of PilE1

Structure determination of PilE1 was done by NMR, essentially as described [51]. In brief, isotopically labelled purified 6His-PilE1 was concentrated to ~750 µM in NMR buffer (50 mM Na_2HPO_4/NaH_2PO_4 pH 6, 50 mM NaCl, 10% D_2O). A full set of triple resonance NMR spectra was recorded on a Bruker Avance III 800 MHz spectrometer equipped with triple resonance cryoprobes at 295 K, and processed with NMRPipe [52]. Backbone assignments were completed using a combination of HBHA, HNCACB, HNCO, HN(CA)CO, and CBCA(CO)NH experiments using NMRView (One Moon Scientific) [53]. Side-chain resonance assignments were obtained from a combination of CC(CO)NH, HC(C)H-TOCSY and (H)CCH-TOCSY experiments using an in-house software developed within NMRView [54]. Distance restraints were

obtained from 3D $^{1}H^{1}H^{15}N$-NOESY and $^{1}H^{1}H^{13}C$-NOESY spectra and used for structure calculations in ARIA 2.3 [55], along with dihedral angle restraints obtained from chemical shift values calculated using the TALOS+ server [56]. For each round of six calculations, 100 structures were calculated over eight iterations. In the final iteration, the 10 lowest energy structures were submitted to a water refinement stage to form the final structural ensemble.

Bioinformatics

All the sequences were from the genome of *S. anguinis* 2908 [26]. Protein alignments were done using the Clustal Omega server at EMBL-EBI, with default parameters. Pretty-printing and shading of alignment files was done using BOXSHADE server at ExPASy. Prediction of functional domains was done by scanning the protein sequence either against InterPro protein signatures [28], or against the SUPERFAMILY database of structural protein domains [30]. Both analyses were done using default parameters. MODELLER was used for modelling protein 3D structures [57]. In brief, a homology model for full-length PilE1 was produced using MODELLER in multiple template mode, with the N-terminal 25 residues of PilE from the gonococcal Tfp [15] and a representative of the PilE1 NMR ensemble determined in this study. This PilE1 model was then used as a template to produce a PilE2 model using MODELLER in single template mode. These two molecules were then superimposed onto a single chain each of the gonococcal filament cryo-EM reconstruction [15], using COOT [58] using the SSM superpose function. Sections of the polypeptide were then refined into the electron densities using the real space refine function in COOT to adjust fitting of PilE1 /PilE2 monomers into the helical filament. Multiple copies of these models were then used to produce the final representation. Structural homologs of PilE1 were identified by scanning the Protein Data Bank (PDB) using the Dali server [59].

ACKNOWLEDGMENTS

This work was supported by a grant from the Medical Research Council (MRC) to VP (MR/P022197/1). IG was a recipient of a PhD studentship from the MRC Centre of Molecular Bacteriology and Infection. We are grateful to Michael Koomey (University of Oslo) for useful discussions. We would like to express our gratitude to Christian Köhler (University of Oslo) for his valuable input and assistance in re-configuring the LC.

REFERENCES

1. Berry JL, Pelicic V. Exceptionally widespread nano-machines composed of type IV pilins: the prokaryotic Swiss Army knives. FEMS Microbiol Rev. 2015;39:134-54.

2. Chaudhury P, Quax TEF, Albers SV. Versatile cell surface structures of archaea. Mol Microbiol. 2018;107:298-311.

3. Giltner CL, Nguyen Y, Burrows LL. Type IV pilin proteins: versatile molecular modules. Microbiol Mol Biol Rev. 2012;76:740-72.

4. Pelicic V. Type IV pili: *e pluribus unum*? Mol Microbiol. 2008;68:827-37.

5. Francetic O, Buddelmeijer N, Lewenza S, Kumamoto CA, Pugsley AP. Signal recognition particle-dependent inner membrane targeting of the PulG pseudopilin component of a type II secretion system. J Bacteriol. 2007;189:1783-93.

6. Arts J, van Boxtel R, Filloux A, Tommassen J, Koster M. Export of the pseudopilin XcpT of the *Pseudomonas aeruginosa* type II secretion system via the signal recognition particle-Sec pathway. J Bacteriol. 2007;189:2069-76.

7. LaPointe CF, Taylor RK. The type 4 prepilin peptidases comprise a novel family of aspartic acid proteases. J Biol Chem. 2000;275:1502-10.

8. Aly KA, Beebe ET, Chan CH, Goren MA, Sepulveda C, Makino S, et al. Cell-free production of integral membrane aspartic acid proteases reveals zinc-dependent methyltransferase activity of the *Pseudomonas aeruginosa* prepilin peptidase PilD. MicrobiologyOpen. 2013;2:94-104.

9. Strom MS, Lory S. Amino acid substitutions in pilin of *Pseudomonas aeruginosa*. Effect on leader peptide cleavage, amino-terminal methylation, and pilus assembly. J Biol Chem. 1991;266:1656-64.

10. Hu J, Xue Y, Lee S, Ha Y. The crystal structure of GXGD membrane protease FlaK. Nature. 2011;475:528-31.

11. Goosens VJ, Busch A, Georgiadou M, Castagnini M, Forest KT, Waksman G, et al. Reconstitution of a minimal machinery capable of assembling periplasmic type IV pili. Proc Natl Acad Sci USA. 2017;114:E4978-E86.

12. Chang YW, Rettberg LA, Treuner-Lange A, Iwasa J, Sogaard-Andersen L, Jensen GJ. Architecture of the type IVa pilus machine. Science. 2016;351:aad2001.

13. Mancl JM, Black WP, Robinson H, Yang Z, Schubot FD. Crystal structure of a type IV pilus assembly ATPase: insights into the molecular mechanism of PilB from *Thermus thermophilus*. Structure. 2016;24:1886-97.

14. Kolappan S, Coureuil M, Yu X, Nassif X, Egelman EH, Craig L. Structure of the *Neisseria meningitidis* type IV pilus. Nat Commun. 2016;7:13015.

15. Wang F, Coureuil M, Osinski T, Orlova A, Altindal T, Gesbert G, et al. Cryoelectron microscopy reconstructions of the *Pseudomonas aeruginosa* and *Neisseria gonorrhoeae* type IV pili at sub-nanometer resolution. Structure. 2017;25:1423-35.

16. Koo J, Lamers RP, Rubinstein JL, Burrows LL, Howell PL. Structure of the *Pseudomonas aeruginosa* type IVa pilus secretin at 7.4 A. Structure. 2016;24:1778-87.

17. Gold VA, Salzer R, Averhoff B, Kuhlbrandt W. Structure of a type IV pilus machinery in the open and closed state. Elife. 2015;4:e07380.

18. Satyshur KA, Worzalla GA, Meyer LS, Heiniger EK, Aukema KG, Misic AM, et al. Crystal structures of the pilus retraction motor PilT suggest large domain movements and subunit cooperation drive motility. Structure. 2007;15:363-76.

19. Maier B, Koomey M, Sheetz MP. A force-dependent switch reverses type IV pilus retraction. Proc Natl Acad Sci USA. 2004;101:10961-6.

20. Clausen M, Jakovljevic V, Sogaard-Andersen L, Maier B. High-force generation is a conserved property of type IV pilus systems. J Bacteriol. 2009;191:4633-8.

21. Ng D, Harn T, Altindal T, Kolappan S, Marles JM, Lala R, et al. The *Vibrio cholerae* minor pilin TcpB initiates assembly and retraction of the toxin-coregulated pilus. PLoS Pathog. 2016;12:e1006109.

22. Ellison CK, Kan J, Dillard RS, Kysela DT, Ducret A, Berne C, et al. Obstruction of pilus retraction stimulates bacterial surface sensing. Science. 2017;358:535-8.

23. Melville S, Craig L. Type IV pili in Gram-positive bacteria. Microbiol Mol Biol Rev. 2013;77:323-41.

24. Rakotoarivonina H, Jubelin G, Hebraud M, Gaillard-Martinie B, Forano E, Mosoni P. Adhesion to cellulose of the Gram-positive bacterium *Ruminococcus albus* involves type IV pili. Microbiology. 2002;148:1871-80.

25. Varga JJ, Nguyen V, O'Brien D K, Rodgers K, Walker RA, Melville SB. Type IV pili-dependent gliding motility in the Gram-positive pathogen *Clostridium perfringens* and other *Clostridia*. Mol Microbiol. 2006;62:680-94.

26. Gurung I, Spielman I, Davies MR, Lala R, Gaustad P, Biais N, et al. Functional analysis of an unusual type IV pilus in the Gram-positive *Streptococcus sanguinis*. Mol Microbiol. 2016;99:380-92.

27. Gurung I, Berry JL, Hall AMJ, Pelicic V. Cloning-independent markerless gene editing in *Streptococcus sanguinis*: novel insights in type IV pilus biology. Nuc Acids Res. 2017;45:e40.

28. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236-40.

29. Imam S, Chen Z, Roos DS, Pohlschröder M. Identification of surprisingly diverse type IV pili, across a broad range of Gram-positive bacteria. PLoS One. 2011;6:e28919.

30. Wilson D, Pethica R, Zhou Y, Talbot C, Vogel C, Madera M, et al. SUPERFAMILY-sophisticated comparative genomics, data mining, visualization and phylogeny. Nuc Acids Res. 2009;37:D380-6.

31. Pugsley AP. Processing and methylation of PuIG, a pilin-like component of the general secretory pathway of *Klebsiella oxytoca*. Mol Microbiol. 1993;9:295-308.

32. Horiuchi T, Komano T. Mutational analysis of plasmid R64 thin pilus prepilin: the entire prepilin sequence is required for processing by type IV prepilin peptidase. J Bacteriol. 1998;180:4613-20.

33. Thomas NA, Chao ED, Jarrell KF. Identification of amino acids in the leader peptide of *Methanococcus voltae* preflagellin that are important in posttranslational processing. Arch Microbiol. 2001;175:263-9.

34. Lopez-Castilla A, Thomassin JL, Bardiaux B, Zheng W, Nivaskumar M, Yu X, et al. Structure of the calcium-dependent type 2 secretion pseudopilus. Nat Microbiol. 2017;2:1686-95.

35. Parge HE, Forest KT, Hickey MJ, Christensen DA, Getzoff ED, Tainer JA. Structure of the fibre-forming protein pilin at 2.6 Å resolution. Nature. 1995;378:32-8.

36. Craig L, Taylor RK, Pique ME, Adair BD, Arvai AS, Singh M, et al. Type IV pilin structure and assembly. X-ray and EM analyses of *Vibrio cholerae* toxin-coregulated pilus and *Pseudomonas aeruginosa* PAK pilin. Mol Cell. 2003;11:1139-50.

37. Cisneros DA, Pehau-Arnaudet G, Francetic O. Heterologous assembly of type IV pili by a type II secretion system reveals the role of minor pilins in assembly initiation. Mol Microbiol. 2012;86:805-18.

38. Giltner CL, Habash M, Burrows LL. *Pseudomonas aeruginosa* minor pilins are incorporated into type IV pili. J Mol Biol. 2010;398:444-61.

39. Korotkov KV, Hol WG. Structure of the GspK-GspI-GspJ complex from the enterotoxigenic *Escherichia coli* type 2 secretion system. Nat Struct Mol Biol. 2008;15:462-8.

40. Helaine S, Dyer DH, Nassif X, Pelicic V, Forest KT. 3D structure/function analysis of PilX reveals how minor pilins can modulate the virulence properties of type IV pili. Proc Natl Acad Sci USA. 2007;104:15888-93.

41. Cehovin A, Simpson PJ, McDowell MA, Brown DR, Noschese R, Pallett M, et al. Specific DNA recognition mediated by a type IV pilin. Proc Natl Acad Sci USA. 2013;110:3065-70.

42. Bernard SC, Simpson N, Join-Lambert O, Federici C, Laran-Chich MP, Maissa N, et al. Pathogenic *Neisseria meningitidis* utilizes CD147 for vascular colonization. Nat Med. 2014;20:725-31.

43. Konto-Ghiorghi Y, Mairey E, Mallet A, Dumenil G, Caliot E, Trieu-Cuot P, et al. Dual role for pilus in adherence to epithelial cells and biofilm formation in *Streptococcus agalactiae*. PLoS Pathog. 2009;5:e1000422.

44. Biais N, Higashi DL, Brujic J, So M, Sheetz MP. Force-dependent polymorphism in type IV pili reveals hidden epitopes. Proc Natl Acad Sci USA. 2010;107:11358-63.

45. Piepenbrink KH, Maldarelli GA, Martinez de la Pena CF, Dingle TC, Mulvey GL, Lee A, et al. Structural and evolutionary analyses show unique stabilization strategies in the type IV pili of *Clostridium difficile*. Structure. 2015;23:385-96.

46. Inoue H, Nojima H, Okayama H. High efficiency transformation of *Escherichia coli* with plasmids. Gene. 1990;96:23-8.

47. Sambrook J, Russell DW. Molecular cloning. A laboratory manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 2001.

48. Anonsen JH, Vik A, Egge-Jacobsen W, Koomey M. An extended spectrum of target proteins and modification sites in the general O-linked protein glycosylation system in *Neisseria gonorrhoeae*. J Proteome Res. 2012;11:5781-93.

49. Anonsen JH, Borud B, Vik A, Viburiene R, Koomey M. Structural and genetic analyses of glycan O-acetylation in a bacterial protein glycosylation system: evidence for differential effects on glycan chain length. Glycobiology. 2017;27:888-99.

50. Anonsen JH, Vik A, Borud B, Viburiene R, Aas FE, Kidd SW, et al. Characterization of a unique tetrasaccharide and distinct glycoproteome in the O-Linked protein glycosylation system of *Neisseria elongata* subsp. *glycolytica*. J Bacteriol. 2016;198:256-67.

51. Berry JL, Xu Y, Ward PN, Lea SM, Matthews SJ, Pelicic V. A comparative structure/function analysis of two type IV pilin DNA receptors defines a novel mode of DNA binding. Structure. 2016;24:926-34.

52. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995;6:277-93.

53. Johnson BA, Blevins RA. NMR View: A computer program for the visualization and analysis of NMR data. J Biomol NMR. 1994;4:603-14.

54. Marchant J, Sawmynaden K, Saouros S, Simpson P, Matthews S. Complete resonance assignment of the first and second apple domains of MIC4 from *Toxoplasma gondii*, using a new NMRView-based assignment aid. Biomol NMR Assign. 2008;2:119-21.

55. Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, Nilges M. ARIA2: automated NOE assignment and data integration in NMR structure calculation. Bioinformatics. 2007;23:381-2.

56. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR. 2009;44:213-23.

57. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinformatics. 2014;47:5.6.1-5.6.32.

58. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr Sect D: Biol Crystallogr. 2010;66:486-501.

59. Holm L, Laakso LM. Dali server update. Nuc Acids Res. 2016;44:W351-5.

LEGENDS TO FIGURES

Figure 1. Bioinformatic analysis of the two major pilins (PilE1 and PilE2) and three pilin-like proteins (PilA, PilB and PilC) encoded by the *pil* locus in *S. sanguinis* 2908. (**A**) Protein architecture of the Pil proteins harbouring the N-terminal IPR012902 motif that is part of the class III signal peptide defining type IV pilins (black rounded rectangle). In PilA, this motif is degenerate and could be detected only by visual inspection and is therefore represented by a grey rounded rectangle. PilB and PilC contain additional C-terminal domains (blue and orange rounded rectangles). PilB contains a von Willebrand factor type A motif (IPR002035), while PilC contains a concanavalin A-like lectin/glucanase structural domain (SSF49899). Proteins have been drawn to scale and the subscript numbers indicate their lengths. (**B**) N-terminal sequence alignement of the putative class III signal peptides of the above five Pil proteins. The 8-18 aa-long leader peptides, which contain a majority of hydrophilic (shaded in orange) or neutral (no shading) residues, end with a conserved Gly_{-1}. Processing by PilD (indicated by the vertical arrow) is expected to occur after Gly_{-1}. The mature proteins start with a tract of 21 predominantly hydrophobic residues (shaded in blue), which usually forms an extended α-helix that is the main assembly interface within filaments.

Figure 2. *S. sanguinis* Tfp exhibit a characteristic Tfp morphology and are composed primarily of N-terminally methylated PilE1 and PilE2 subunits. (**A**) SDS-PAGE/Coomassie analysis of *S. sanguinis* Tfp purified using previously described (lane 1) and enhanced (lane 2) purification procedures. Samples were prepared from cultures at similar OD_{600}, separated by SDS-PAGE and stained with Coomassie blue. Identical volumes were loaded in each lane. MW indicates the molecular weight marker lane, with weights in kDa. (**B**) WT filament morphology in pilus preparations prepared using the enhanced purification procedure as assessed

by TEM after negative staining. Scale bar represents 200 nm. (**C**) Top-down mass profiling of purified Tfp from *S. sanguinis*. Shown are deconvoluted molecular mass spectra over the range of 14,000 to 15,300 *m/z*. Masses are presented as monoisotopic $[M+H]^+$.

Figure 3. *S. sanguinis* Tfp are hetero-polymers composed of PilE1 and PilE2. (**A**) Schematics of the affinity-purification strategy. In brief, one of the genes encoding major pilin PilE1 (red) or PilE2 (black) is engineered to produce a protein fused to an affinity 6His tag, *e.g* PilE1-6His (small red sphere). Sheared pili are mixed with cobalt-coated beads (large black sphere) and purified by pull down, which is expected to yield filaments containing both pilins (tagged and untagged) if *S. sanguinis* Tfp are hetero-polymers. Conversely, if the filaments are distinct homo-polymers, only the tagged filaments will be purified. (**B**) Analysis of filaments purified by shearing/ultra-centrifugation from four unmarked mutants harbouring a 6His tag fused either to the C-terminus of full-length PilE1 and PilE2 ($PilE1_{6His\text{-}long}$ and $PilE2_{6His\text{-}long}$) or replacing the last seven aa in these pilins ($PilE1_{6His\text{-}short}$ and $PilE2_{6His\text{-}short}$). The WT strain was included as a control. Samples were prepared from cultures at similar OD_{600}, separated by SDS-PAGE and either stained with Coomassie blue (upper panel) or analysed by immunoblot using anti-PiE1, anti-PilE2 or anti-6His antibodies (bottom three panels). MW indicates the molecular weight marker lane, with weights in kDa. (**C**) Immunoblot analysis using anti-PiE1, anti-PilE2 or anti-6His antibodies of sheared filaments that were affinity purified by pull down using cobalt-coated beads. The WT strain was included as a control. Sheared filaments were prepared from cultures adjusted to the same OD_{600}, affinity purified, eluted in the same final volume and identical volumes were loaded in each lane. MW indicates the molecular weight marker lane, with weights in kDa.

Figure 4. Processing of *S. sanguinis* major pilin PilE1 by PilD and its assembly

in filaments follow widely conserved general principles. (**A**) Immunoblot analysis of PilE1 expression and processing by PilD in strains expressing PilE1$_{G-1A}$, PilE1$_{G-1S}$ and PilE1$_{E5A}$ mutants. The WT strain *ΔpilD* mutant have been included as controls. Whole-cell protein extracts were separated by SDS-PAGE and probed using anti-PilE1 antibody or anti-PilE2 antibody as a control. Protein extracts were quantified, equalised and identical volumes were loaded in each lane. Molecular weights are indicated in kDa. (**B**) Immunoblot analysis of PilE1 assembly in filaments in strains expressing PilE1$_{G-1A}$, PilE1$_{G-1S}$ and PilE1$_{E5A}$ mutants. The WT strain *ΔpilD* mutant have been included as controls. Purified Tfp were separated by SDS-PAGE and probed using anti-PilE1 antibody or anti-PilE2 antibody as a control. Samples were prepared from cultures adjusted to the same OD_{600}, and identical volumes were loaded in each lane but to improve detection of faint bands, we loaded 20-fold dilutions of WT and PilE1$_{G-1A}$ samples. Molecular weights are indicated in kDa.

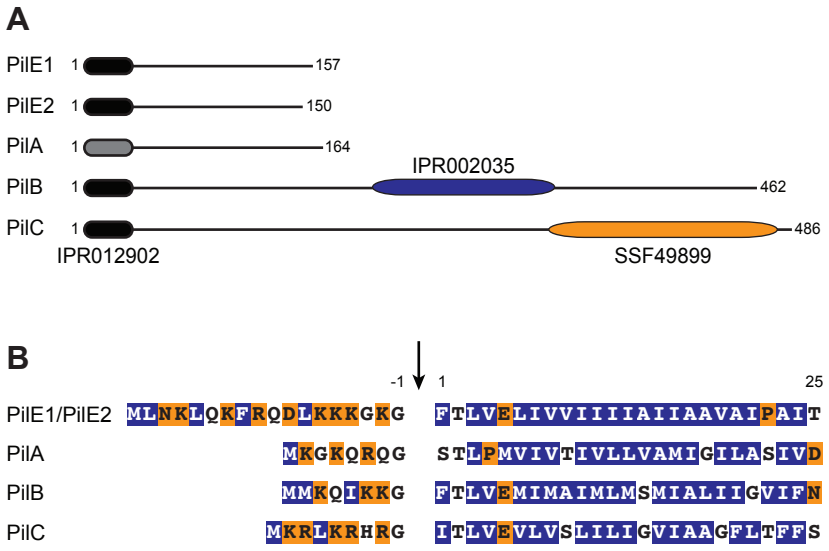
Figure 5. High-resolution 3D structure of PilE1 reveals a canonical type IV pilin fold with an uncommon highly flexible C-terminus, which more closely resembles pseudopilins. (**A**) Cartoon representation of the NMR structure of globular domain of PilE1. The conserved core in type IV pilins (the N-terminal α-helix and 3-stranded antiparallel β-meander) is depicted in grey. Distinctive/key structural features flanking the β-meander are highlighted in color, the α1β1-loop in red and the unstructured C-terminus in green. (**B**) Ribbon representation of the superposition of the ensemble of 10 most favourable PilE1 structures determined by NMR. (**C**) Cartoon representation of the superposition of structures globular domains of PilE1 (magenta) and *K. oxytoca* pseudopilin PulG (grey). The two structures superpose with a rmsd of 5.75 Å over their entire length.

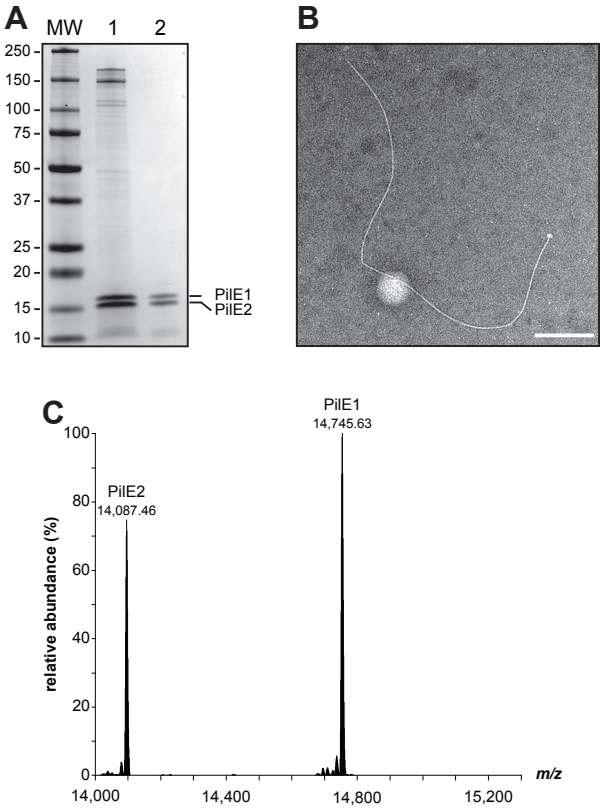
Figure 6. Modelling shows that full-length PilE1/PilE2 fit readily into available Tfp structures. (**A**) Cartoon representation of the homology structural model of PilE2

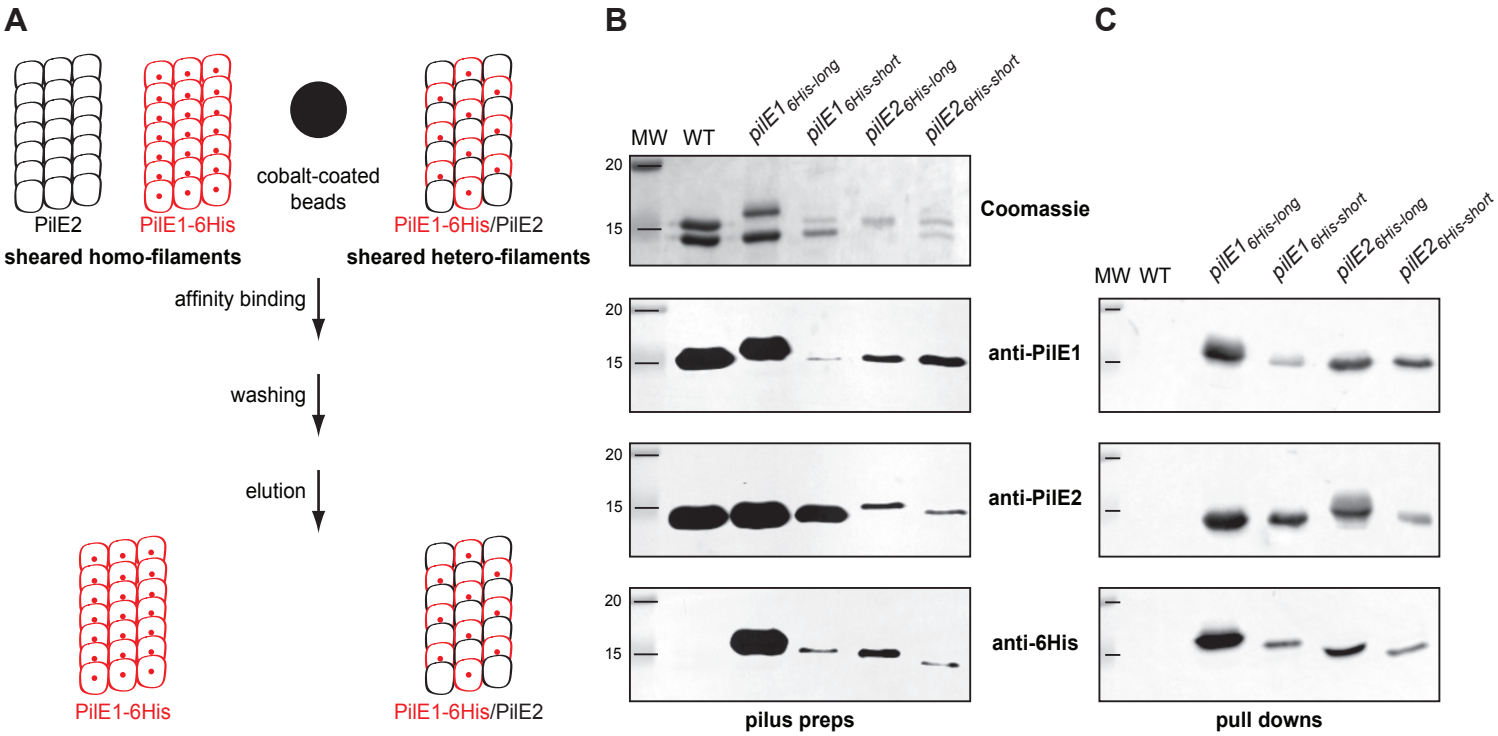
based on PilE1 structure. Same coulour codes have been used than in Fig. 5A. (**B**) Superposition of globular domains of PilE1 and PilE2 (right) with a rmsd of 0.61 Å. (**C**) Packing of full-length models of PilE1 and PilE2 into the recently determined structure of gonoccoccal Tfp. Half of the gonoccoccal subunits in the structure were replaced by PilE1 and the other half with PilE2.

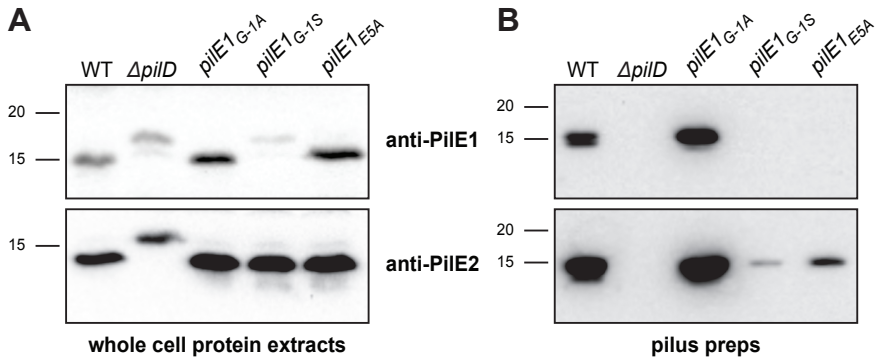
Figure 7. Pilin-like proteins PilA, PilB and PilC are processed by PilD and co-purify with Tfp. (**A**) Immunoblot analysis of PiA, PilB and PilC expression and processing by PilD. Whole-cell protein extracts were probed using specific anti-peptide antibodies, which were generated for this study. Protein extracts were quantified, equalised and equivalent amounts of total proteins were loaded in each lane. Molecular weights are indicated in kDa. (**B**) Immunoblot analysis of pilus preparations using anti-PiA, anti-PilB and anti-PilC antibodies. Samples were prepared from cultures adjusted to the same OD_{600} and identical volumes were loaded in each lane. Molecular weights are indicated in kDa.

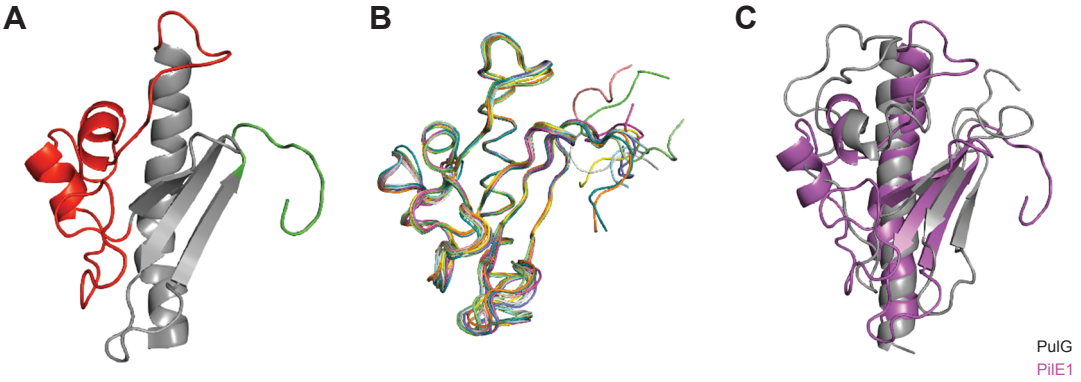
Figure 8. *ΔpilE1* and *ΔpilE2* mutants produce Tfp capable of powering motility, albeit at different speeds. (**A**) Macroscopic motility assay. Spreading zones, or lack thereof, around single and double *ΔpilE1* and *ΔpilE2* mutants. The WT strain and *ΔpilT* mutant have been included as positive and negative controls, respectively. (**B**) Microscopic motility assay. The histograms represents the distribution curve of velocities (in 100 $nm \cdot s^{-1}$ intervals) measured for *ΔpilE1* and *ΔpilE2* mutants. Insets are representative 30 sec trajectories of movement of small chains of cells. Scale bar represents 5 µm. Corresponding movies are available as supplementary information (Movies S1 and S2).

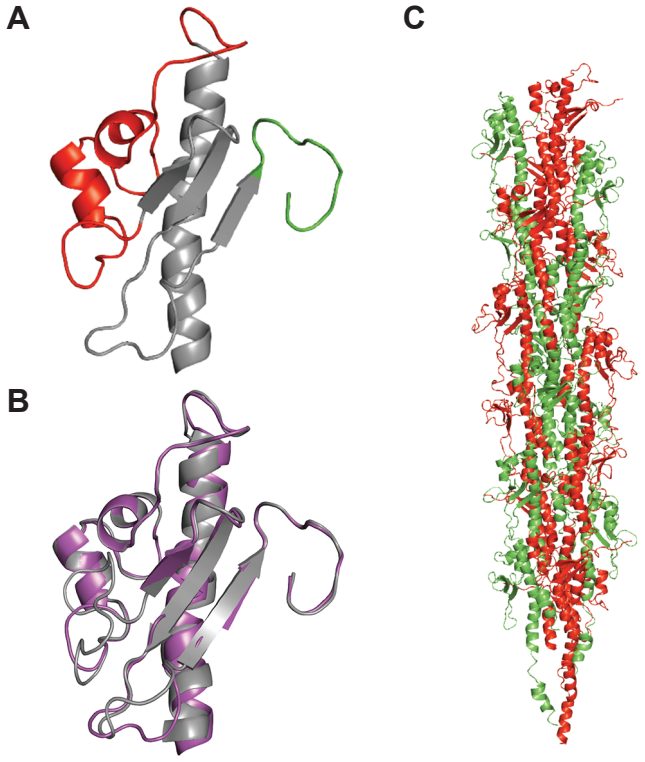

A
PilE1
1
157
PilE2
1
150
PilA
1
164
IPR002035
PilB
1
462
PilC
1
486
IPR012902
SSF49899
B
-1
1
25
PilE1/PilE2
MLNKLQKFRQDLKKKGKG
FTLVELIVVIIIIAIIAAVAIPAIT
PilA
MKGKQRQG
STLPMVIVTIVLLVAMIGILASIVD
PilB
MMKQIKKG
FTLVEMIMAIMLMSMIALIIGVIFN
PilC
MKRLKRHRG
ITLVEVLVSLILIGVIAAGFLTFFS

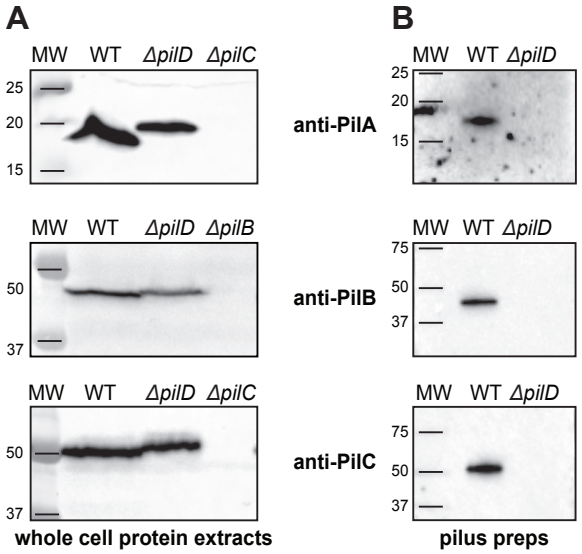

A
MW
1
2
250
150
100
75
50
37
25
20
15
10
PilE1
PilE2
B
C
PilE1
14,745.63
PilE2
14,087.46
relative abundance (%)
100
80
60
40
20
0
m/z
14,000
14,400
14,800
15,200

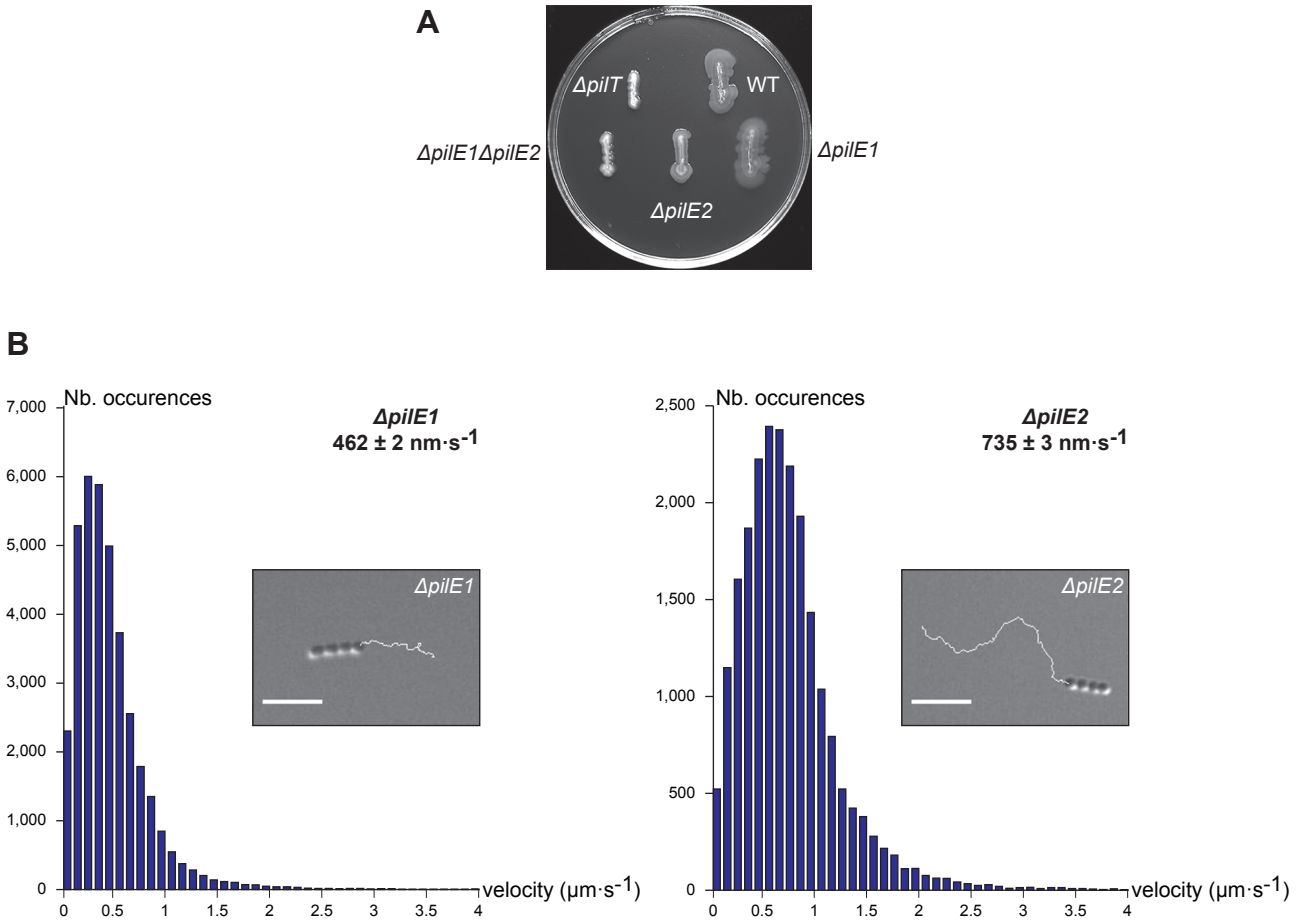

A
PilE2
PilE1-6His
cobalt-coated beads
PilE1-6His/PilE2
sheared homo-filaments
sheared hetero-filaments
affinity binding
washing
elution
PilE1-6His
PilE1-6His/PilE2
B
MW
WT
pilE1 6His-long
pilE1 6His-short
pilE2 6His-long
pilE2 6His-short
20
15
Coomassie
anti-PilE1
anti-PilE2
anti-6His
pilus preps
C
MW
WT
pilE1 6His-long
pilE1 6His-short
pilE2 6His-long
pilE2 6His-short
pull downs


A
WT
ΔpilD
pilE1 G-1A
pilE1 G-1S
pilE1 E5A
20
15
anti-PilE1
15
anti-PilE2
whole cell protein extracts
B
WT
ΔpilD
pilE1 G-1A
pilE1 G-1S
pilE1 E5A
20
15
20
15
pilus preps


A
B
C
PulG
PilE1


A
B
C


A
B
MW WT ΔpilD ΔpilC
25
20
15
anti-PilA
MW WT ΔpilD ΔpilB
50
37
anti-PilB
MW WT ΔpilD ΔpilC
50
37
anti-PilC
whole cell protein extracts
MW WT ΔpilD
25
20
15
MW WT ΔpilD
75
50
37
MW WT ΔpilD
75
50
37
pilus preps


A
ΔpilT
WT
ΔpilE1ΔpilE2
ΔpilE1
ΔpilE2
B
Nb. occurences
ΔpilE1
462 ± 2 nm·s-1
ΔpilE1
velocity (µm·s-1)
Nb. occurences
ΔpilE2
735 ± 3 nm·s-1
ΔpilE2
velocity (µm·s-1)

SUPPLEMENTARY INFORMATION

Figure S1. Bioinformatic analysis of the genes in the *pil* locus in *S. sanguinis* 2908 and of its prepilin peptidase PilD. (**A**) Gene organisation in the *pil* locus in *S. sanguinis* 2908. All the genes are drawn to scale and the scale bar represents 1 kb. Genes essential for Tfp biogenesis are boxed by a thick line. (**B**) Protein architecture of the prepilin peptidase PilD in *S. sanguinis* and its ortholog in *N. meningitidis*. The N-terminal IPR010627 motif (orange rounded rectangle) catalyses N-methylation, while the C-terminal IPR000045 motif (blue rounded rectangle) catalyses proteolytic processing of pilins. Proteins have been drawn to scale and the subscript numbers indicate protein length.

Figure S2. Sequence alignement of major pilins PilE1 and PilE2 in *S. sanguinis* 2908. Residues are shaded in dark blue (identical), light blue (conserved) or unshaded (different). Structural features - α-helices and β-strands - are indicated below the sequences. The vertical arrow indicates the N-terminal portion of PilE1 and PilE2 that was truncated to facilitate protein purification.

Movies S1. Cellular motility of a *ΔpilE1* mutant. A small chain of cells attached to a coverslip was imaged for 30 sec. The scale bar represents 5 µm.

Movies S2. Cellular motility of a *ΔpilE2* mutant. A small chain of cells attached to a coverslip was imaged for 30 sec. The scale bar represents 5 µm.

Table S1. NMR structural statistics.

Table S2. Strains and plasmids used in this study.

Table S3. Primers used in this study.

Table S1

6His-PilE1	
Number of distance restraints	1,750
intra-residual	712
sequential	416
medium range	287
long range	335
NOE violations >0.5 Å (%)	0.85
Dihedral violations >5º (%)	0
Ramachandran favoured (%)	84.4
Ramachandran allowed (%)	13.5
Ramachandran generously allowed (%)	0.7
Ramachandran disallowed (%)	1.3

Table S2

Name	Details	Source
***E. coli* strains**		
DH5α	used for cloning	
BL21(DE3)	used for protein expression and purification	
***S. sanguinis* strains**		
2908	sequenced WT isolate	[26]
ΔpilA	*ΔpilA::aphA-3* deletion mutant	[26]
ΔpilB	*ΔpilB::aphA-3* deletion mutant	[26]
ΔpilC	*ΔpilC::aphA-3* deletion mutant	[26]
ΔpilD	*ΔpilD::aphA-3* deletion mutant	[26]
ΔpilE1	*ΔpilE1::aphA-3* deletion mutant	[26]
ΔpilE1 primary mutant	*ΔpilE1::pheS*aphA-3* deletion mutant	[27]
pilE1$_{G-1A}$	*pilE1* point mutant expressing PilE1$_{G-1A}$	this study
pilE1$_{G-1S}$	*pilE1* point mutant expressing PilE1$_{G-1A}$	this study
pilE1$_{E5A}$	*pilE1* point mutant expressing PilE1$_{G-1A}$	this study
pilE1-6His$_{SHORT}$	*pilE1* point mutant expressing 6His-tagged PilE1	this study
pilE1-6His$_{LONG}$	*pilE1* point mutant expressing 6His-tagged PilE1	[27]
ΔpilE2	*ΔpilE2::aphA-3* deletion mutant	[26]
ΔpilE2 primary mutant	*ΔpilE2::pheS*aphA-3* deletion mutant	this study
pilE2-6His$_{SHORT}$	*pilE2* point mutant expressing 6His-tagged PilE2	this study
pilE2-6His$_{LONG}$	*pilE2* point mutant expressing 6His-tagged PilE2	this study
ΔpilE1ΔpilE2	*ΔpilE1pilE2::aphA-3* double deletion mutant	[26]
ΔpilT	*ΔpilT::aphA-3* deletion mutant	[26]
Plasmids		
pCR8/GW/TOPO	TA cloning vector	Invitrogen
TOPO-*pheS*aphA-3*	*pheS*aphA-3* double cassette in pCR8/GW/TOPO	[27]
TOPO-*pilE1*	full-length *pilE1* in pCR8/GW/TOPO	this study
TOPO-*pilE1*$_{G-1A}$	*pilE1*$_{G-1A}$ in pCR8/GW/TOPO	this study
TOPO-*pilE1*$_{G-1S}$	*pilE1*$_{G-1S}$ in pCR8/GW/TOPO	this study
TOPO-*pilE1*$_{E5A}$	*pilE1*$_{E5A}$ in pCR8/GW/TOPO	this study
pMK-*pilB*	codon-optimised *pilB* in pMK	GeneArt
pMK-RQ-*pilC*	codon-optimised *pilC* in pMK-RQ	GeneArt
pET-28b	T7-based expression vector	Novagen
pET28-*pilA*	pET-28b derivative for expressing 6His-PilA$_{33-164}$	this study
pET28-*pilB*	pET-28b derivative for expressing 6His-PilB$_{37-462}$	this study
pET28-*pilC*	pET-28b derivative for expressing 6His-PilC$_{34-486}$	this study
pET28-*pilE1*	pET-28b derivative for expressing 6His-PilE1$_{46-157}$	this study
pET28-*pilE2*	pET-28b derivative for expressing 6His-PilE2$_{46-150}$	this study

Table S3

Name	Sequence
Cloning in pET-28b	
pilA-pETF	gggccatggatcatcatcatcatcatcatGATACAGGGCAAAGCCAGAC
pilA-pETR	cccgtcgacTTACTTCTGTGCCGATCTCAA
pilB-pETF	gggccatggatcatcatcatcatcatcatAGCAGCCGTGAACTGATTGA
pilB-pETR	cccggatccTTACGGACCGCTAACAAACC
pilC-pETF	gggccatggatcatcatcatcatcatcatAATAACATTCTGCGTCAGCGTAGCCA
pilC-pETR	cccggatccTTAGCTTGCTTTGTATTTATCGC
pilE1-pETF	ggccatggatcatcatcatcatcatcatCAAGATAACGCTCGTAAGAGCC
pilE1-pETR	ccggatccTTAGTTTGAGTTTACACCATTAGCAGA
pilE2-pETF	ggccatggatcatcatcatcatcatcatCAAGATAACGCTCGTAAGAGCC
pilE2-pETR	ccggatccTTATTTTGAATTAGCACCAGCTTCG
Cloning in pCR8/GW/TOPO	
pilE1-F	AGGAACAAAACAAATGCCCCT
pilE1-R	TCTCAAATGCAGGGTTTTACTACA
Site-directed mutagenesis	
pilE1$_{G-1A}$#1	GACTTGAAGAAAAAAGGTAAAG**C**TTTTACCTTGGTTGAGTTGATC
pilE1$_{G-1A}$#2	GATCAACTCAACCAAGGTAAAA**G**CTTTACCTTTTTTCTTCAAGTC
pilE1$_{G-1S}$#1	GACTTGAAGAAAAAAGGTAAA**A**GTTTTACCTTGGTTGAGTTGATC
pilE1$_{G-1S}$#2	GATCAACTCAACCAAGGTAAAAC**T**TTTACCTTTTTTCTTCAAGTC
pilE1$_{E5A}$#1	GTAAAGGTTTTACCTTGGTTG**C**GTTGATCGTGGTAATTATC
pilE1$_{E5A}$#2	GATAATTACCACGATCAAC**G**CAACCAAGGTAAAACCTTTAC
Engineering *N. meningitidis* mutants	
pheS-F	ATGACGAAAACGATTGAAGAAC
aph-R	CTAAAACAATTCATCCAGTAAAA
pilE1-F1	CAGGCCGGTGAAAAGACTG
pilE1-R1	GTTCTTCAATCGTTTTCGTCATTTTGAATAGATCTCCTGTTTTT
pilE1-F2	TTTTACTGGATGAATTGTTTTAGCGACTGGTCTGCTAATGGTG
pilE1-R2	GCTCTGTTGAAGGATCCACG
pilE1-R3	TTAgtgatggtgatggtgatgGTTTGAGTTTACACCATTAGCAG
pilE1-F3	catcaccatcaccatcacTAATTGTCAAATCATCTAAATAAGATGTA
pilE1-R4	TTAgtgatggtgatggtgatgAGACCAGTCGTAGGTCAAAAC
pilE1-F4	catcaccatcaccatcacTAATTGTCAAATCATCTAAATAAGATGTA
pilE1-R5	AGGGGCATTTGTTTTGTTCCT
pilE1-F5	TGTAGTAAAACCCTGCATTTGAGA
pilE2-F1	GGAACGTCTGACAGGGATGA
pilE2-R1	GTTCTTCAATCGTTTTCGTCATATGTATTTTCTCCTAATGTTTTTATG
pilE2-F2	TTTTACTGGATGAATTGTTTTAGACCGAAGCTGGTGCTAATTC
pilE2-R2	TACCATCCGCAGAAAGACCA
pilE2-R3	TTAgtgatggtgatggtgatgTTTTGAATTAGCACCAGCTTC
pilE2-F3	catcaccatcaccatcacTAATAACTTGAATTAATTTGAGTTATTCAT
pilE2-R4	TTAgtgatggtgatggtgatgGGTCCAGTTGTATGTTAAAACG
pilE2-F4	catcaccatcaccatcacTAATAACTTGAATTAATTTGAGTTATTCAT

Overhangs are in lower case, with restriction sites underlined. Mismatched bases generating point mutations are in bold upper case.