The Molecular Basis of FimT-mediated DNA Uptake during Bacterial Natural Transformation

Sebastian A.G. Braus[1], Francesca L. Short[2,3], Stefanie Holz[1,#], Matthew J.M. Stedman[1,#], Alvar D. Gossert[1,4], Manuela K. Hospenthal[1,5,*]

[1] Institute of Molecular Biology and Biophysics, ETH Zürich, Otto-Stern-Weg 5, 8093 Zürich, Switzerland

[2] Monash University, Melbourne, Australia

[#] These authors contributed equally

* Correspondence: manuela.hospenthal@mol.biol.ethz.ch

[3] ORCID: 0000-0002-0025-4858
[4] ORCID: 0000-0001-7732-495X
[5] ORCID: 0000-0003-1175-6826

Abstract

Naturally competent bacteria encode sophisticated protein machineries for the uptake and translocation of exogenous DNA into the cell. If this DNA is integrated into the bacterial genome, the bacterium is said to be naturally transformed. Most competent bacterial species utilise type IV pili for the initial DNA uptake step. These proteinaceous cell-surface structures are composed of thousands of pilus subunits (pilins), designated as major or minor according to their relative abundance in the pilus. In this study, we show that the minor pilin FimT plays an important role in the natural transformation of *Legionella pneumophila*. We used NMR spectroscopy, *in vitro* DNA binding assays and *in vivo* transformation assays to understand the molecular basis of FimT's role in this process. FimT directly interacts with DNA *via* an electropositive patch, rich in arginines, several of which are well-conserved and located in FimT's conformationally flexible C-terminal tail. We also show that FimT orthologues from other γ-Proteobacteria share the ability to bind to DNA. Our functional characterisation and comprehensive bioinformatic analysis of FimT, suggest that it plays an important role for DNA uptake in a wide range of competent species.

Introduction

Competent bacteria can take up exogenous DNA, present in their environment, and integrate it into their genomes by the process of natural transformation. This is an important avenue of horizontal gene transfer (HGT), which has widespread consequences for bacterial evolution and the spread of antibiotic resistance and other pathogenicity traits. In contrast to other modes of HGT, namely transduction and conjugation, natural transformation is entirely controlled by the recipient cell that encodes all the required machinery for DNA uptake, translocation and integration[1]. More than 80 bacterial species, including Gram-negative and Gram-positive organisms, have been shown to be naturally competent[2], yet the true prevalence of this mechanism amongst bacteria likely remains underappreciated. The Gram-negative bacterium *Legionella pneumophila* is naturally competent[3], consistent with the observation that its genome bears evidence of frequent HGT and recombination events[4–6]. Although *L. pneumophila* could be described as an accidental human pathogen, it is the aetiological agent of Legionnaire's disease, a serious and life-threatening form of pneumonia, that results from an infection of alveolar macrophages by contaminated aerosols[7,8].

Legionella, like most Gram-negative bacteria, are thought to utilise type IV pili (T4P) for DNA uptake[3,9,10], defined as the movement of DNA across the outer membrane (OM) and into the periplasmic space[11]. However, the molecular mechanisms involved in this step remain poorly defined. T4P are extracellular proteinaceous appendages composed of thousands of

individual pilus subunits (pilins), designated as major or minor depending on their relative abundance in the pilus[12,13]. A prevailing model suggests that T4P can bind to DNA[9] and transport it into the cell *via* pilus retraction, which is powered by the retraction ATPase PilT[14,15]. Pilus retraction is thought to bring the DNA into proximity with the OM and be taken up across the OM-embedded secretin channel PilQ, which is the same pore traversed by the T4P themselves[16,17]. Once in the periplasm, ComEA binds to incoming DNA to prevent its back-diffusion by acting like a Brownian ratchet[18,19]. Subsequently, DNA is converted into single-stranded DNA (ssDNA) and transported across the inner membrane (IM) by a putative channel called ComEC[20]. In the cytoplasm, ssDNA is protected by single-stranded DNA binding protein (Ssb)[21] and DNA processing protein A (DprA)[22], before being integrated into the genome by homologous recombination in a RecA- and ComM-dependent manner[23,24].

In recent years, studies of several competent bacteria have shown that their T4P (or their pilins) can directly interact with DNA[15,25–29]. This function was attributed to specialised minor pilins or pilin-like proteins in *Neisseria species* (ComP)[27,28], *Vibrio cholerae* (VC0858 and VC0859)[15], and *Thermus thermophilus* (ComZ)[29], although a major pilin (PilA4) has also been suggested to contribute in the latter[30]. Of these, ComP found in *Neisseria* species, is the best-characterised DNA-binding minor pilin to date. ComP displays a sequence preference for neisserial DNA containing so-called DNA uptake sequences (DUS)[31–33] and binds to DNA through an electropositive surface patch[27,28,34]. VC0858, VC0859 and ComZ are thought to be located at the pilus tip[15,29], whereas ComP has been suggested to either be incorporated throughout the pilus fibre[28] or at the pilus tip[9]. In addition to these proteins, the minor pilin FimT has also been implicated in natural transformation, as its loss leads to a reduction in transformation efficiency in *Acinetobacter baylyi*[35]. However, this phenotype was never followed up with further DNA-binding studies.

We set out to study DNA uptake during natural transformation in *Legionella pneumophila*. It is not known whether Legionella's T4P can interact with DNA, and if so, which pilins are responsible. We tested several major and minor *Legionella* pilin candidates for their ability to bind DNA and show that FimT efficiently interacts with DNA *in vitro* and *in vivo*, and that loss of binding, just like *fimT* deletion, results in almost complete abrogation of natural transformation. We also determined the structure of FimT and show that a conserved electropositive surface patch rich in arginines is required for DNA binding. Finally, we show that FimT is not only important for natural transformation in *L. pneumophila*, but that it likely plays a role in many other bacterial species, as suggested by DNA binding studies and bioinformatic analyses. Together, our work provides the molecular basis of FimT's role in natural transformation.

Results

FimT is critical for natural transformation in *L. pneumophila* and interacts with DNA

FimT and FimU are minor type IV pilins that belong to the GspH/FimT family of proteins (Pfam: PF12019; InterPro: IPR022346), which also includes the type II secretion system (T2SS) pseudopilin GspH/XcpU. All three genes are encoded in the *L. pneumophila* genome and share an overall amino acid sequence identity of ~15–25%. *L. pneumophila* FimT ($FimT_{Lp}$) and FimU ($FimU_{Lp}$) possess all the features of typical type IV pilins, including an N-terminal signal sequence required for their targeting to the inner membrane (IM), followed by a hydrophobic transmembrane helix required for IM insertion prior to pilus assembly and proper packing into the filament structure post assembly[12,36]. First, we tested whether $FimT_{Lp}$ or $FimU_{Lp}$ are required for T4P biogenesis in *L. pneumophila*. To this end, we overexpressed a Flag-tagged version of the major pilin PilA2[10] and compared relative amounts of PilA2-Flag-containing T4P in fractions of surface appendages sheared from the cell surface of various *L. pneumophila* Lp02 strains, including *fimT* and *fimU* deletion strains (**Extended Data Fig. 1a**). These results indicate that T4P are still assembled and present on the cell surface when *fimT* or *fimU* are deleted. Next, to test whether $FimT_{Lp}$ or $FimU_{Lp}$ play a role in natural transformation in *L. pneumophila*, we performed transformation assays comparing the *fimT* and *fimU* deletion strains with the parental strain and strains harbouring deletions in genes known to be important for natural transformation (**Fig. 1a**). Deletion of *comEC*, encoding the putative IM DNA channel, *pilQ*, encoding the OM secretin, and *pilT*, encoding the retraction ATPase, resulted in undetectable levels of natural transformation in our assay. These observations are in close agreement with previous studies in *L. pneumophila*, as well as other competent Gram-negative organisms such as *V. cholerae*, where deletion of these genes resulted in severe or complete natural transformation phenotypes[10,16]. Deletion of *fimU* did not produce a phenotype, whereas natural transformation was undetectable in the *fimT* deletion strain, as observed previously in *A. baylyi*[35]. Expression of $FimT_{Lp}$ *in trans* from an IPTG-inducible promoter restored the transformation efficiency of our *L. pneumophila* strain to wild-type levels, showing that the transformation defect is specific to $FimT_{Lp}$.

We reasoned that FimT contributes to the OM DNA uptake step of natural transformation by forming a constituent part of type IV pili (T4P) able to directly bind to DNA. Therefore, we performed electrophoretic mobility shift assays (EMSA) to test whether $FimT_{Lp}$ is able to bind to DNA *in vitro* (**Fig. 1b**). In order to produce soluble protein samples, all pilins were expressed as truncations lacking the N-terminal transmembrane helix (**Extended Data Fig. 1b**). Indeed, purified $FimT_{Lp}$ interacted with all DNA probes tested (**Extended Data Table 4**), including ssDNA, dsDNA, linear and circular DNA molecules, whereas neither

142 $FimU_{Lp}$, nor the putative major pilin subunits (PilA1 and PilA2) showed any interaction
143 (**Fig. 1b** and **Extended Data Fig. 1c**). These experiments suggest that the dissociation
144 constant (K_D) of the interaction between $FimT_{Lp}$ and 30meric DNA is in the low µM range. In
145 order to determine the K_D more precisely and to learn about the binding stoichiometry of this
146 interaction, we performed isothermal titration calorimetry (ITC) utilising shorter 12meric
147 ssDNA or dsDNA fragments (**Fig. 1c**). We determined a K_D of 7.0 µM and 2.8 µM for
148 12meric ssDNA and dsDNA, respectively. Interestingly, these experiments revealed that a
149 single $FimT_{Lp}$ binds to 12meric ssDNA, whereas two molecules can bind to the dsDNA
150 ligand, suggesting two binding sites on opposite sides of the double helix.
151
152 **The solution structure of $FimT_{Lp}$**
153 We determined the solution structure of the soluble N-terminally truncated (residues 28-152,
154 mature pilin sequence numbering) $FimT_{Lp}$ by nuclear magnetic resonance (NMR)
155 spectroscopy (**Fig. 2a, Table 1**). The structure consists of an N-terminal α-helix (α1C) (the
156 transmembrane portion of this helix, α1N, has been removed in the construct), two β-sheets
157 that complete the C-terminal globular pilin domain, and a C-terminal tail, which exhibits
158 conformational flexibility. Both β-sheets are composed of antiparallel strands: β-sheet I is
159 formed by β1, β2, β3 and β5, and β-sheet II by β4, β6 and β7. The closest structural
160 homologue is FimU from *Pseudomonas aeruginosa* ($FimU_{Pa}$) (PDB ID: 4IPU, 4IPV)
161 (**Fig. 2b**). While the two structures share a common fold, there are some key differences. In
162 the $FimU_{Pa}$ structure, the loop between β2 and β3 in β-sheet I forms an additional β-hairpin
163 (β2' and β2''). It is possible, however, that this additional β-hairpin of $FimU_{Pa}$ simply
164 represents the conformation captured in the crystal structure, as the length of the β2-β3 loop
165 is similar in both proteins. In addition, the β7 strand of β-sheet II is longer in $FimU_{Pa}$ and it
166 contains an additional strand (β8)[37]. Furthermore, $FimU_{Pa}$ contains a disulphide bond
167 connecting Cys127 of β6 to the penultimate residue, Cys158, effectively stapling the
168 C-terminal tail in place on top of β-sheet II. Such a disulphide bond is found in various major
169 and minor pilins and the intervening sequence is known as the D-region[36,38]. Further
170 structures of GspH/FimT family proteins exist, including of the minor T2SS pseudopilins,
171 GspH from *Escherichia coli* (PDB ID: 2KNQ) and its orthologue EpsH from *V. cholerae* (PDB
172 ID: 2QV8[39] and 4DQ9[40]), which display similar folds (**Extended Data Fig. 2**).
173
174 The C-terminal tail (residues 140–152) of $FimT_{Lp}$ is unique amongst the currently determined
175 structures of GspH/FimT family members. Different pieces of NMR data suggest significant
176 conformational exchange, but not an entirely flexibly disordered tail. The amide resonances
177 of residues 140–149 are very weak and those of residues 142–144 are not visible at all. We
178 could not observe any intense long-range nuclear Overhauser effects (NOEs) for residues

140–152, which would be expected for a well-defined β-sheet conformation. T_2 relaxation measurements indicated conformational exchange on the millisecond timescale, as the T_2 values for backbone amide ^{1}H and ^{15}N nuclei for the C-terminal tail were approximately half the value of the structured part of the protein (**Fig. 2d, Extended Data Fig. 3**). A fully disordered C-terminal tail could however be excluded by {^{1}H}-^{15}N heteronuclear NOE measurements, as the NOE intensity for the amides in the tail was close to the theoretical value of 0.78, which is expected for amides on globular particles. Finally, the deviations of Cα chemical shifts from random coil values clearly indicated a β-strand propensity (**Fig. 2d, Extended Data Fig. 3**). The data therefore suggest that the C-terminal amino acids have a β-strand-like backbone conformation but sample different states in the micro- to millisecond timescale. These findings are further supported by low amide proton temperature coefficients[41] and increased proteolytic susceptibility of this region, compared to the rest of the structure, witnessed by disappearance of the NMR resonances of the tail after prolonged storage of samples.

$FimT_{Lp}$ interacts with DNA *via* a conserved C-terminal region rich in arginines

Next, we characterised the residues of $FimT_{Lp}$ involved in DNA binding using NMR spectroscopy (**Fig. 3a-c**). We performed binding experiments titrating increasing amounts of 12 bp dsDNA (**Extended Data Table 4**) into ^{15}N-labelled $FimT_{Lp}$ and recorded ^{15}N, ^{1}H heteronuclear single-quantum correlation (2D [^{15}N,^{1}H]-HSQC) spectra. Most $FimT_{Lp}$ resonances remained unperturbed (**Fig. 3a**), which suggests that no global conformational change occurs upon DNA binding. However, a subset of resonances exhibit marked chemical shift perturbations (CSPs) (**Fig. 3a**), indicating changes in the local chemical environment resulting from direct contact with DNA or other indirect conformational changes. A plot of CSPs against the amino acid sequence is shown in **Figure 3b**, and we mapped CSPs greater than a threshold (Δppm > 1σ) onto the $FimT_{Lp}$ surface (**Fig. 3c**, **Extended Data Fig. 4**). The largest CSPs correspond to residues located in three adjacent loop regions in the C-terminal globular domain of the protein, the β4-β5 loop (residues 103–106), the β5-β6 loop (118–126) and the C-terminal tail (140–152) (**Fig. 3b**). These shifts predominantly map to an elongated surface patch connecting the C-terminal tail with the globular C-terminal domain of $FimT_{Lp}$ (**Fig. 3c**). Most of these residues are predicted to be accessible in the context of the assembled pilus, particularly when considering the flexibility of this region (**Fig. 2d, Extended Data Fig. 3**). CSPs corresponding to residues outside this contiguous surface patch can be explained by indirect conformational changes. We attempted to further structurally characterise the DNA-bound state, with special emphasis on possible changes in the structure or dynamics of the C-terminus. However, the $FimT_{Lp}$-DNA complex was not stable long-term and NMR signals were generally strongly weakened upon

DNA binding, such that relaxation or triple resonance experiments did not yield spectra of sufficient quality. An analysis of evolutionary conservation of the $FimT_{Lp}$ surface revealed that many of the interacting residues are also well conserved (**Fig. 3c**). In particular, residues of the C-terminal tail show marked sequence conservation and include a number of positively charged arginines, which are often involved in protein-DNA contacts through binding to the negatively charged DNA backbone *via* electrostatic interactions[42].

Interface mutations inhibit DNA binding and natural transformation *in vivo*

We conducted microscale thermophoresis/temperature-related intensity change (MST/TRIC) experiments to measure the binding of labelled 12 bp dsDNA (**Extended Data Table 4**) to purified $FimT_{Lp}$ variants (**Extended Data Fig. 1b**), in order to further understand the nature of the $FimT_{Lp}$-DNA interaction and the importance of specific interface residues. First, we conducted experiments under different buffer conditions to test whether the affinity of the interaction between wild-type $FimT_{Lp}$ and DNA is dependent on ionic strength. Indeed, when we increased the NaCl concentration from 50 mM to 150 mM, thereby raising the ionic strength, the K_D increased from ~6.3 µM to ~70.1 µM (**Fig. 4a**). This is consistent with a non-sequence specific protein-DNA interaction, which is electrostatically driven. Furthermore, the K_D determined at a NaCl concentration of 50 mM agrees very well with the affinities determined from the ITC experiments (K_D of 2.8 µM) (**Fig. 1c**), as well as our NMR binding studies (K_D of ~8 µM) (**Extended Data Fig. 5**), which were all conducted in the same buffer. Next, we used MST/TRIC to test the importance of several charged residues at the DNA binding interface identified by our NMR analyses (**Fig. 4b**). We substituted arginine or lysine residues for glutamine in the three loop regions we identified to be important for binding. As expected, the loss of a single charged residue (e.g. K103 in the β4-β5 loop; R119 in the β5-β6 loop; R143, R146 or R148 in the C-terminal tail) only led to a small reduction in the affinity (~1.4–4 fold). However, the combined loss of two (R146/R148) or three (R143/R146/R148) charged residues next to each other on the $FimT_{Lp}$ surface was more detrimental to binding, resulting in a ~10 fold or ~45 fold reduction in affinity, respectively. Lastly, we tested what effect these binding mutations have on natural transformation *in vivo* (**Fig. 4c**). These data show that mutations of single charged residues reduce *Legionella's* transformability by ~30-600 fold, whereas the double and triple mutants completely abrogate DNA uptake in our assay and thus phenocopy the effect observed upon *fimT* deletion (**Fig. 1a**). These results further support a model in which $FimT_{Lp}$ contributes to natural transformation in *Legionella* by virtue of its ability to interact with DNA in the context of a DNA uptake pilus.

FimT of other Gram-negative bacteria also interacts with DNA

Given that FimT, and the residues involved in DNA binding identified in $FimT_{Lp}$, appear to be conserved, we wondered whether FimT orthologues from other bacteria are also capable of binding DNA. We expressed and purified FimT and FimU from the human pathogen *P. aeruginosa* and the plant pathogen *Xanthomonas campestris* (both γ-Proteobacteria) and performed EMSAs to assess DNA binding *in vitro* (**Fig. 5a**). Interestingly, FimT from both species binds to DNA and the affinity appears to be within the same order of magnitude as *L. pneumophila* FimT. On the other hand, FimU does not interact with DNA, except for the *X. campestris* homologue, which shows very weak binding at very high FimU concentrations. Since both FimT orthologues ($FimT_{Pa}$ and $FimT_{Xc}$) likely share structural similarities to $FimT_{Lp}$, we tested whether they are capable of restoring natural transformability in a *L. pneumophila fimT* deletion strain. The FimT orthologues were ectopically expressed either as wild-type full-length proteins or as chimeric proteins. The chimeric constructs replaced the flexible C-terminal tail region (lacking a disulphide bond) of $FimT_{Lp}$ with the *bona fide* D-region of the FimT orthologues (**Fig. 5b**). The expression of full-length $FimT_{Pa}$ and $FimT_{Xc}$ did not restore natural transformation. Intriguingly, when we replaced the flexible C-terminal tail of $FimT_{Lp}$ with the D-region of $FimT_{Pa}$, natural transformation levels were restored to near wild-type levels. Together, these results indicate that DNA binding by FimT is not unique to *L. pneumophila* and that FimT may be important for DNA uptake in a wide range of competent species.

We then used genomic context and sequence information from the four FimT orthologs known to either bind to DNA or contribute to competence (from *L. pneumophila*, *X. campestris*, *P. aeruginosa* and *A. baylyi*) to explore the distribution and conservation of this protein (see Methods). First, we looked at the genetic location and organisation of FimT and FimU in *Legionella* and other bacteria (**Extended Data Fig. 6**). In *L. pneumophila*, *fimU* (*lpg0632*) is encoded in a minor pilin operon upstream of *pilV (lpg0631)*, *pilW (lpg0630)*, *pilX (lpg0629)*, *pilY1 (lpg0628)* and *pilE (lpg0627).* In contrast, *fimT* (*lpg1428*) appears as an 'orphan' gene, encoded elsewhere in the genome, and seemingly distant from genes encoding other type IV pilins, components of the T4P machinery or genes with known functions in natural transformation. Interestingly, while FimT in other species could be found either as an orphan, or adjacent to other minor pilin-related genes, the location of FimU was conserved, and this pattern was seen in a broader collection of homologues as well as the functionally-characterised representatives. We then retrieved a diverse set of homologues of $FimT_{Lp}$ and classified them according to genomic location and sequence similarity to exclude sequences that were likely to be FimU proteins. We found that FimT is conserved in all sequenced *Legionella* species, and homologues are found in a wide variety of

γ-Proteobacteria from various phylogenetic orders, with representatives of the Xanthomonadales, Alteromonadales and Pseudomonadales being particularly common (**Fig. 5c**). The pairwise sequence identity was 40-50% between FimTs from *Legionella pneumophila* and other *Legionella* species, and ~25% (median) between *L. pneumophila* FimT and those from more distantly related bacteria. Around half of the FimT homologues are located in proximity (within 5 kb) to other minor pilin locus components. FimU is also present in many bacterial species, albeit not all species encode both genes. Phylogenetic analysis of FimT homologues showed that these proteins largely clustered with others from the same order and sharing the same locus type, indicating that *fimT* is likely to be vertically inherited. The best conserved regions of FimT include the N-terminal helix, important for pilus biogenesis (IM insertion, assembly and structural packing), and the C-terminal region (**Fig. 5d**). This region of conservation includes many of the residues we have identified to be important for DNA binding and thus natural transformation (**Fig. 3c, d**). Indeed, it appears as though these DNA binding residues can be identified in proteins with as little as 18% overall amino acid sequence identity with $FimT_{Lp}$. Taken together, FimT homologues share an overall fold and a conserved DNA-binding motif near the C-terminus of the protein, and can be found in diverse genomic locations within diverse proteobacterial species.

Discussion

Natural transformation is an important mode of horizontal gene transfer with widespread consequences for bacterial evolution. Furthermore, the spread of pathogenicity traits and antibiotic resistance genes leads to the emergence of increasingly virulent and difficult to treat bacterial strains. The first step of this process involves DNA uptake mediated by T4P[9], which has only been studied in a handful of competent species. The minor type IV pilin FimT, but not the closely related FimU, from *A. baylyi* was previously implicated in natural transformation[35], yet its mechanism remained obscure. Here, we characterised FimT from the naturally competent human pathogen *L. pneumophila* ($FimT_{Lp}$) and revealed the molecular mechanisms underlying its role in natural transformation.

We hypothesised that $FimT_{Lp}$ is involved in DNA uptake by binding to extracellular DNA in the context of T4P and showed that *Legionella* strains lacking *fimT* display a marked reduction in transformation efficiency (**Fig. 1a**). Indeed, purified $FimT_{Lp}$ interacted with DNA *in vitro*, regardless of the nature of DNA probe tested (**Fig. 1b**, **Extended Data Fig. 1c**). Furthermore, we determined the structure of $FimT_{Lp}$ by NMR spectroscopy (**Fig. 2**) and mapped its DNA interaction surface by chemical shift perturbation experiments (**Fig. 3**). This binding surface consists of several positively charged residues, some of which are highly conserved, located primarily in two loop regions (the β4-β5 and β5-β6 loops) and the

C-terminal tail (**Fig. 3b, c**). The importance of key residues for DNA binding and natural transformation was confirmed by *in vitro* DNA binding assays and *in vivo* transformation assays (**Fig. 4b, c**). Although our ITC experiments (**Fig. 1c**) indicate a 2:1 ($FimT_{Lp}$:DNA) binding mode, we do not think this is physiologically relevant in the context of the T4P.

Our structure of $FimT_{Lp}$ shares the same overall fold as the closely related T4P minor pilin $FimU_{Pa}$, and the T2SS minor pseudopilins $GspH_{Ec}$ and $EpsH_{Vc}$, albeit with some key differences (**Fig. 2**, **Extended Data Fig. 2**). In place of the last β-strand (β8), part of β-sheet II in all other currently determined FimT/GspH family structures, $FimT_{Lp}$ contains a conformationally flexible C-terminal tail (**Extended Data Fig. 3**). In our NMR studies, the heteronuclear $\{^1H\}$-^{15}N NOE data and Cα chemical shifts for the C-terminal residues are indicative of a β-strand conformation, while the T_2 transverse relaxation times for backbone amide ^1H and ^{15}N nuclei, increased line broadening and the absence of H-bonds indicate a less well-structured conformation. A plausible interpretation of these results is that this region can exchange between a β-strand and a less-structured conformation on a millisecond timescale. The flexibility of this region is further supported by its increased proteolytic susceptibility. $FimT_{Lp}$, as well as all FimT homologues from the order Legionellales, lack the D-region defining disulphide bond present in many major and minor pilins, including other FimT and FimU homologues (**Fig. 5d**). Therefore, it is likely that disulphide bond-containing FimT orthologues do not possess a conformationally flexible C-terminal tail. The structure of $GspH_{Ec}$ was also determined in solution by NMR spectroscopy, yet it possesses a clearly defined and complete four-stranded β-sheet II region. This suggests that this region, also shared by $FimU_{Pa}$ and $EpsH_{Vc}$, is not simply a result of crystal lattice effects and thus further highlights $FimT_{Lp}$'s unique C-terminal tail (**Fig. 2, Extended Data Fig. 2**).

FimU and GspH/EpsH have been suggested to serve as adaptors in T4P and T2SS pseudopili, respectively, linking the tip subunits to the remainder of the filament structure composed of the major pilin subunit[43–45]. Whereas minor pilins in general have been suggested to play a role in pilus priming/pilus biogenesis[37,45], the deletion of FimU, but not FimT affected pilus biogenesis in *P. aeruginosa* and *Pseudomonas syringae*[46,47]. Furthermore, FimU, but not FimT of *P. aeruginosa* has been shown to play a role in bacterial twitching motility[48]. In *A. baylyi* on the other hand, both proteins showed near wild-type levels of twitching, but FimT appeared to play a role in natural transformation[35]. Orthologues of the GspH pseudopilin are critical components of the T2SS and may play a role in binding to T2SS protein substrates[49]. To this end, the crystal structure of the *V. cholerae* orthologue EpsH revealed an extended and disordered β4-β5 loop (**Extended Data Fig. 2d**), which has

been proposed to play a role in substrate binding[40]. Interestingly, we have identified this same loop to contribute to $FimT_{Lp}$-DNA binding (**Fig. 3b**). Therefore it appears that, although sharing a common evolutionary origin[50], FimT/GspH family proteins have become functionally diverged and specialised for the binding of different macromolecular substrates[51,52]. In the case of $FimT_{Lp}$, a surface patch rich in arginines enables it to function in DNA uptake during natural transformation.

The currently best-characterised DNA binding minor pilin is ComP[27,28,34]. While ComP homologues seem to be restricted to species of the family *Neisseriaceae*[27], FimT homologues are present in diverse γ-Proteobacteria and some Hydrophilales (**Fig. 5c**). Both proteins share a conserved type IV pilin core structure, including the N-terminal helix and a four-stranded antiparallel β-sheet, but differ substantially in their C-terminal regions. In the case of ComP, this region is characterised by its so-called DD-region containing two disulphide bonds[27] (**Extended Data Fig. 7**). By contrast, FimT contains a second three-stranded antiparallel β-sheet followed by its conformationally flexible C-terminal tail and contains no disulphide bonds. In both proteins, important DNA binding residues are located near the C-terminus, which would be exposed to the solvent in the context of a fully assembled pilus[28]. Interestingly, competent *Neisseriaceae* species preferentially take up DNA sequences from related species[31–33]. This has been attributed to ComP's increased binding affinity towards DUS-sequences, which are DNA sequences that are highly enriched in their own genomes[27]. It was proposed that ComP engages DNA *via* an initial electrostatic attraction, followed by ComP's α1-β1, β1-β2, DD-region binding to successive grooves of the dsDNA to achieve specificity[28]. In contrast, no sequence selectivity has been reported for *L. pneumophila*[3], which is consistent with the electrostatic binding mode of $FimT_{Lp}$. In addition to $FimT_{Lp}$ and ComP, other type IV pilins or pilin-like proteins that contribute to T4P DNA binding include ComZ and PilA4 from *T. thermophilus*[29,30] and VC0858 and VC0859 from *V. cholerae*[15]. Once again, positively charged lysine and/or arginine residues likely contribute to DNA binding in all these proteins.

Lastly, we showed that other FimT orthologues, including FimT of the human pathogen *P. aeruginosa* and the plant pathogen *X. campestris,* are also capable of DNA binding (**Fig. 5a**). These experiments showed that FimT orthologues, whether they contain or lack the D-region defining disulphide bond, are capable of DNA binding. This was demonstrated even more strikingly by the FimT chimera, where the fusion of $FimT_{Lp}$ with $FimT_{Pa}$ introduced a non-native disulphide bond into the *Legionella* system, yet resulted in a functional protein *in vivo* capable of supporting natural transformation (**Fig. 5b**). In addition, our bioinformatic

analyses showed that FimT is present across a wide range of γ-Proteobacteria and that the DNA-binding C-terminal region is well-conserved on a sequence level (**Fig. 5d**). In particular, our alignments of high-confidence FimTs revealed a conserved GRxR motif (where x is often, but not always, a hydrophobic residue) at their C-terminus (**Fig. 5d**). In $FimT_{Lp}$ these two arginines correspond to R146 and R148, which we showed contribute to DNA binding *in vitro* and *in vivo* (**Fig. 4b, c**). This motif is less well defined or only partially present in FimU orthologues and those we tested in this study do not bind DNA *in vitro* (**Fig. 5a**). Interestingly, a similar C-terminal motif can also be found in the pilins that assemble into the Com pili of Gram-positive organisms, which have been implicated in DNA uptake during natural transformation[53–55]. It remains to be investigated, whether this motif also contributes to DNA binding and natural transformation in those proteins.

In summary, this study provides a comprehensive analysis of the molecular mechanisms underpinning FimT's interaction with DNA and demonstrated its pivotal role during natural transformation of the human pathogen *L. pneumophila*. Furthermore, we analysed FimT orthologues from other naturally competent and pathogenic γ-Proteobacteria, which together with our thorough bioinformatic analysis, suggests that FimT is a key player in the natural transformation of a wide range of bacteria.

Methods

Bacterial strains and growth conditions

L. pneumophila Lp02 (laboratory strain derived from *L. pneumophila* Philadelphia-1) was cultured in ACES [N-(2-acetamido)-2-aminoethanesulfonic acid]-buffered yeast extract (AYE) liquid medium or on ACES-buffered charcoal yeast extract (CYE) solid medium, supplemented with 100 µg/mL streptomycin and 100 µg/mL thymidine. When appropriate, chloramphenicol and kanamycin were added at 5 µg/mL and 15 µg/mL, respectively. For the construction of knockout Lp02 strains, the relevant genes and 1000 bp of upstream and downstream regions were first cloned into the pSR47S suicide plasmid (derivative of pSR47[56]). Following deletion of the gene of interest from the plasmid, the modification was introduced onto the Lp02 chromosome by triparental conjugation and subsequent selection as described previously[57,58]. All strains were verified by colony PCR and DNA sequencing (Microsynth) and are listed in **Extended Data Table 2**.

Plasmids

All protein expression constructs were generated using the pOPINS or pOPINB vectors[59,60] carrying an N-terminal His_6-SUMO or His_6 tag, respectively. Constructs for *in vivo* studies were generated using pMMB207C[61], by cloning the relevant genes downstream of the P*tac* promoter. DNA fragments were amplified from *L. pneumophila* (RefSeq NC_002942.5) genomic DNA by PCR using CloneAmp HiFi PCR premix (Takara) and the relevant primers. For FimT and FimU orthologues from *P. aeruginosa* PAO1 (RefSeq NC_002516.2) and *X. campestris pv. campestris* str. ATCC 33913 (RefSeq NC_003902.1), template DNA was first synthesised (Twist Bioscience). In-Fusion cloning and site-directed mutagenesis was carried out according to the manufacturer's guidelines (Takara). All plasmids and primers used in this study can be found in **Extended Data Table 3** and **Extended Data Table 4**, respectively. A summary of the gene locus tags of genes mentioned in this study from their respective genomes can be found in **Extended Data Table 5** and **Extended Data Table 6**.

Protein Production

Recombinant His_6-SUMO tagged proteins ($FimT_{Lp}$, $FimU_{Lp}$, $FimT_{Pa}$, $FimU_{Pa}$, $FimT_{Xc}$, $FimU_{Xc}$) and His_6-tagged proteins (PilA1, PilA2) were expressed in BL21 (DE3) or Shuffle T7 *E. coli* cells (NEB). All constructs were N-terminally truncated to remove the transmembrane helix (α1N) (**Extended Data Table 3**). Cultures were grown in Luria-Bertani (LB) media to an optical density at 600 nm (OD_{600}) of 0.6–0.8, induced using 0.5 mM IPTG and further incubated at 16°C for 12–18 h while shaking. Cells were lysed in 50 mM Tris-HCl pH 7.2, 1 M NaCl, 20 mM imidazole, 0.1 mg/mL lysozyme, 1 mg/mL DNAse and one complete mini

EDTA-free protease inhibitor cocktail tablet (Roche), by passing the sample three times through a pressurised cell disruptor (M110-L, Microfluidics) at 12000 psi. The clarified lysate was applied to a 5 mL HisTrap HP column (Cytiva) and His_6-SUMO or His_6 tagged pilins were eluted with a linear 20–500 mM imidazole gradient. The His_6-SUMO or His_6 tag was cleaved using the catalytic domain of the human SENP1 protease or PreScission protease, respectively, while the sample was dialysed against 50 mM Tris-HCl pH 7.2, 50 mM NaCl. Protein samples were further purified by cation exchange chromatography using a 5 mL HiTrap SP HP column (Cytiva), from which pilins were eluted using a linear 50–1000 mM NaCl gradient. Lastly, the pilin samples were purified by size exclusion chromatography in 50 mM Tris-HCl pH 7.2, 50 mM NaCl using a HiLoad 16/600 Superdex 75 pg column (Cytiva). Protein samples were concentrated using Amicon Ultra-15 centrifugal filters (3 kDa molecular weight cut-off, Millipore). Reducing agent (2 mM DTT) was included in the buffers for those pilins with free cysteines. All purification steps were performed at 4°C.

NMR spectroscopy

Production of isotope-labelled $FimT_{Lp}$

To produce uniformly labelled $FimT_{Lp}$, cells were grown in M9 minimal medium containing 1 g/L $^{15}NH_4Cl$ and further supplemented with 3 g/L glucose (or $^{13}C_6$-glucose for double labelled $FimT_{Lp}$), 2 mM $MgSO_4$, trace elements, vitamin mix and appropriate antibiotics for selection. Protein expression was induced at an OD_{600} of 0.6–0.8 with 0.5 mM IPTG and cells were harvested after 20 h at 16°C. $FimT_{Lp}$ was purified as described above.

Data acquisition and structure determination

For resonance assignments and structure determination the following spectra were recorded on a 580 µM sample of (u-^{13}C,^{15}N)-labeled FimT 28–152 in 25 mM NaP_i pH 7.2, 150 mM NaCl and 10% D_2O at 298 K in a 3 mm diameter NMR tube: 3D HNCACB and 3D CBCACONH spectra[62] were recorded on a 700 MHz AVIIIHD spectrometer equipped with a TCI cryoprobe (Bruker). The spectra consisted of 2048×50×100 complex points in the ^{1}H, ^{15}N and ^{13}C dimensions with respective spectral widths of 16, 34 and 64 ppm, and were recorded with 8 scans per increment resulting in 2 and 1.5 days of measurement time, respectively. A 3D HcC(aliaro)H-TOCSY[63] was recorded on a 600 MHz AVIIIHD spectrometer equipped with a TCI cryoprobe (Bruker). The spectrum consisted of 1536×100×150 complex points in the ^{1}H, ^{1}H and ^{13}C dimensions with respective spectral widths of 16, 12 and 140 ppm and was recorded with 2 scans per increment in 3 days using a recycle delay of 2 s. A time shared 3D [^{13}C/^{15}N,^{1}H]-HSQC NOESY (modified from[64]) was recorded on a 900 MHz AVIIIHD spectrometer equipped with a TCI cryoprobe (Bruker). The spectrum consisted of 1536×100×256 complex points in the ^{1}H, ^{1}H and ^{13}C/^{15}N dimensions

with respective spectral widths of 16, 12 and 140/58 ppm and was recorded with 2 scans per increment in 3 days.

Resonance assignments were determined with the program cara (www.cara.nmr.ch, Keller R (2005), ETH Zürich) to 98.2% completeness. Signals in the NOESY spectra were subsequently automatically picked in the program analysis of the ccpnmr 2.5.1 software package[65]. Peaklists and assignments were used as input for a structure calculation with cyana[66] where angle constraints were automatically generated from Cα chemical shifts. Manual inspection of the automatically picked peak lists resulted in a set of 4595 picked NOE peaks of which 4220 were assigned in the final cyana calculation which yielded an average target function value of 0.21. The structures were finally energy minimized in the program amber20[67]. Statistics for the resulting bundle of 20 conformers can be found in **Table 1**. Additional analysis of the structural bundle after the cyana calculation revealed 42 hydrogen bonds (each present in more than 6 structures) and the following Ramachandran statistics: 72.2%, 27.4% and 0.4% of residues in favoured, allowed and additionally allowed regions, respectively. All structural figures were generated using PyMOL (https://www.pymol.org).

DNA binding studies by NMR

To map the surface patch of $FimT_{Lp}$ involved in DNA binding, chemical shift perturbation experiments were performed using 12 bp or 36 bp dsDNA fragments (**Extended Data Table 4**). [^{15}N,^{1}H]-HSQC experiments of 80 µM ^{15}N-labelled $FimT_{Lp}$ at saturating concentrations of DNA were recorded. In order to use the same conditions as other assays, all protein and DNA samples for NMR binding studies were dialysed into 50 mM Tris-HCl pH 7.2, 50 mM NaCl buffer. Weighted chemical shift perturbations (CSPs), defined as $((\Delta^{1}H^{2})^{0.5}+((\Delta^{15}N/5)^{2})^{0.5}$ (ppm), were measured by comparing spectra of unbound and bound states. The standard deviation (σ) of the chemical shift range was calculated, CSP maps were plotted in GraphPad Prism v9 and residues for which the shift change was greater than σ were mapped onto the $FimT_{Lp}$ surface. To estimate the equilibrium dissociation constant (K_D) of this interaction, [^{15}N,^{1}H]-HSQC experiments of 40 µM ^{15}N-labelled $FimT_{Lp}$ at different concentrations (0–600 µM) of DNA were recorded. For selected residues undergoing large CSPs, binding curves were plotted and fitted to a model assuming one set of binding sites using the software fitKD (four representative curves are shown in (**Extended data Fig. 5**). The spectra were recorded on a 700 MHz AV-NEO spectrometer equipped with a TCI cryoprobe (Bruker) and consisted of 2048×128 complex points using 32 scans per increment resulting in an experiment time of 2 h.

Electrophoretic mobility shift assay

Various DNA probes were tested for interaction with purified pilin samples using an agarose gel-based electrophoretic mobility shift assay (EMSA). Short 30 bp dsDNA fragments were generated by annealing complementary strands of the appropriate length. To generate fluorescently labelled dsDNA probes, one of the two annealing strands was labelled at the 5' end with fluorescein (FAM). All oligonucleotides were obtained from Microsynth and are listed in **Extended Data Table 4**. The pTRC99A-*lpg2953-2958::Kan* (9074 bp) plasmid, left intact or linearised by a single-cutter restriction enzyme (ClaI), was used for the comparison between circular and linear dsDNA probes, respectively. All DNA probes were resuspended in or dialysed into the same buffer as the protein samples prior to the assay. DNA samples (1 µM of 30-meric ssDNA and dsDNA; 20 ng/µl for longer DNA fragments) were incubated with increasing concentrations (0-100 µM) of pilins in 50 mM Tris-HCl pH 7.2, 50 mM NaCl, 15% (v/v) glycerol in a final volume of 20 µL. These samples were incubated at 25°C for 30 min and subsequently separated by gel electrophoresis at 10 V/cm for 30 min using 0.9-2.5% (w/v) agarose gels containing SYBR Safe DNA stain (Invitrogen). DNA was visualised using UV illumination in a gel imaging system (Carestream).

Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) experiments were carried out in duplicate on a VP-ITC microcalorimeter (MicroCal). All measurements were performed in 50 mM Tris-HCl pH 7.2, 50 mM NaCl buffer at 30°C. Following a pre-injection of 1 µL, titrations consisted of 19 consecutive 15 µL injections of 320 µM 12meric dsDNA or 350 µM ssDNA (syringe) into 30 µM $FimT_{Lp}$ (cell) performed at 180 s or 240 s intervals, respectively. The heat of ligand dilution, obtained by injecting DNA into buffer, was subtracted from the reaction heat, and curve fitting was performed in Origin (OriginLab) using a model assuming two binding sites of equal affinity or "one set" of binding sites.

Microscale thermophoresis/temperature-related intensity change measurements

Microscale thermophoresis (MST) experiments were conducted measuring the temperature-related intensity change (TRIC) of the fluorescence signal[68]. A 12 bp fluorescently labelled dsDNA probe was generated by annealing a 5' FAM-labelled and an unlabelled strand (Microsynth; **Extended Data Table 4**) and used in all MST/TRIC experiments. Equilibrium binding assays were performed in 50 mM Tris-HCl pH 7.2, 50-150 mM NaCl, 0.05% (v/v) Tween-20. Increasing concentrations of purified wild-type or mutant $FimT_{Lp}$ samples were incubated with 100 nM of FAM-labelled 12 bp dsDNA probe at 25°C for 30 min prior to measurement. MST/TRIC measurements were performed at 20°C using a Monolith NT.115 instrument (NanoTemper) at 25% LED power and 20% MST laser power. Curve fitting was

performed with data derived from the TRIC effect. For the experiment conducted with wild-type $FimT_{Lp}$ measured at 50 mM NaCl, the data appeared slightly biphasic in nature. This suggested the presence of two binding sites with similar, yet non-identical binding affinities. When these data were fitted with a binding model assuming two non-identical binding sites, $K_D(1)$ was indeed very similar to that obtained when fit according to two identical sites (~2.9 *vs* 6.3 µM). All other binding experiments using other methods (ITC and NMR), as well as MST/TRIC experiments conducted with $FimT_{Lp}$ mutants, did not reveal an obvious biphasic binding signature, which could be explained by insufficient resolution. Therefore, we chose to fit all data in the same manner, assuming two identical binding sites, to allow for their comparison. All MST/TRIC measurements were performed at least three times. In addition, all samples were measured twice, 30 min apart, resulting in very similar binding curves and derived dissociation constants, indicating that the binding equilibrium had been attained at the time of measurement.

Transformation assay

All transformation assays were performed with the *L. pneumophila* Lp02 strain in liquid medium at 30°C, similar to transformation assays performed by others[10,69]. Strains were streaked onto CYE solid medium from frozen stocks and incubated at 37°C for 3-4 days. From this plate, bacteria were resuspended in a liquid starter culture (5 mL of AYE medium) and incubated at 37°C overnight while shaking at 200 rpm. The starter culture was diluted into a fresh 10 mL AYE culture (starting OD_{600} of 0.02) and cultured at 30°C while shaking. Once the culture reached an OD_{600} of 0.3, 1 mL was transferred into a new tube and incubated with 1 µg of transforming DNA at 30°C for a further 24 h. The transforming DNA consisted of a 4906 bp PCR product encompassing the *L. pneumophila* genomic region spanning *lpg2953-2958*, where the *hipB* gene (*lpg2955*) is interrupted by a kanamycin resistance cassette (based on[70]). This provides 2000 bp regions of homology up- and downstream of the resistance cassette. Tenfold serial dilutions of the culture were plated on selective (supplemented with 15 µg/mL kanamycin) and non-selective CYE media. The plates were incubated at 37°C for 4-5 days and colony forming units (CFUs) were counted. The transformation efficiency corresponds to the ratio of the number of CFUs obtained on selective medium divided by the number of CFUs counted on non-selective medium. The minimum counting threshold was set at 10 colonies per plate. Transformation assays to test complementation of knockout strains with protein ectopically expressed from the pMMB207C plasmid were performed in the same manner, except for the addition of 0.5 mM IPTG during the incubation step of the bacteria with transforming DNA. Transformation assays requiring direct comparison between strains or complemented strains were performed in parallel.

Western blot detection of pilin in sheared surface fractions

Lp02 strains (parental, Δ*fimT* and Δ*fimU*) harbouring pMMB207C-*pilA2-flag* were cultured at 37°C for 24 h on CYE media, additionally supplemented with 0.5 mM IPTG. Cells were resuspended in AYE media containing a complete mini EDTA-free protease inhibitor cocktail tablet (Roche) and adjusted to an OD_{600} of 20. To shear appendages from the cell surface, the resuspended cells were vortexed at maximal speed for 30 s. Subsequently, the depiliated cells were pelleted by two rounds of centrifugation at 20'000 *g* for 20 min at 4°C. The supernatants containing surface appendages, including T4P, were transferred to new tubes and the pellets were washed twice by resuspension in 1 mL AYE followed by centrifugation at 20'000 *g* for 20 min at 4°C. Both pellets and supernatants were separated by SDS-PAGE. Proteins were transferred to polyvinylidene fluoride (PVDF) membranes (Amersham) and PilA2-Flag was detected using a horse radish peroxidase (HRP)-coupled primary anti-Flag antibody at a 1:2000 dilution (Sigma, cat. no. SAB4200119). Enhanced chemiluminescence (ECL) (Cytiva) was used for the detection of the protein signal in a Amercham Imager 600. PVDF membranes were stained with Ponceau S to verity even loading across all lanes.

Bioinformatic analyses

Collection of putative FimT and FimU sequences

Three sets of FimT or FimU sequences were collected as follows: 1) a FimT set was retrieved by BlastP against $FimT_{Lp}$, $FimT_{Pa}$, $FimT_{Ab}$ and $FimT_{Xc}$ with a 95% query coverage cutoff, 2) a FimU set was retrieved by BlastP against $FimU_{Lp}$, $FimU_{Pa}$, $FimU_{Ab}$ and $FimU_{Xc}$ with a 95% (Pa, Ab, Xc) or 80% (Lp) query coverage cutoff, 3) a diverse FimT/U set was retrieved by a PSI-blast[71] search against $FimT_{Lp}$, with >95% query coverage and e-value >0.005 cutoffs applied at each iteration, and the search continued for 8 iterations. To limit redundancy in the results all searches were conducted against the refseq_select protein database which, for prokaryotes, contains only sequences from representative and reference genomes. The FimT and FimU sets were used for initial gene neighbourhood analyses beyond the four functionally characterised representatives (**Extended Data Fig. 6**), while the diverse set was used for phylogenetic analysis and to define conserved regions.

Gene neighbourhood analysis

The gene neighbourhood of each *fimT* and *fimU* was examined using custom Biopython[72] scripts and NCBI resources as follows 1) source genome(s) for each protein entry were identified from the Identical Protein Groups (IPG) resource (this was necessary because many of the blast results were non-redundant entries comprising multiple identical proteins), 2) the genome region corresponding to the gene of interest and 5000 bp up- and

downstream was downloaded from the nucleotide database for one representative of each IPG (if <5000 bp flanking up- and downstream sequence was available the entry was excluded from further analysis), and 3) coding sequences in the neighbouring region were extracted as multifasta and searched against the Pfam[73] database of domain profiles using HMMER[74] (hmm scan function, e-value threshold 0.0001). *fimT* or *fimU* genes were classified as orphans or minor pilin locus components based on the presence of one or more of the Pfam domains PilC, PilX, PilX_N and PilW in the flanking region. The presence of just one of these domains was defined as indicating a minor pilin locus, to account for the possibility that proteins only weakly matching the relevant Pfam domain would be missed, or that relevant proteins may be found >5000 bp away. NCBI scripts used in this study are available at https://github.com/francesca-short/NCBI_scripts.

Generation of high-confidence FimT set and phylogenetic analysis

Because FimT is a GspH-domain protein and shares overall structural similarity with the type IV minor pilin FimU and the T2SS protein GspH, putative homologues from the diverse FimT/U set were filtered based on their gene neighbourhood to exclude likely *fimU* genes and generate a subset of high-confidence putative *fimT* genes for further analyses. As 100% of genes in the FimU set were located in minor pilin operons, orphan genes within the diverse FimT/U set were presumed to encode genuine FimT proteins, and these sequences were aligned along with $FimT_{Lp}$, $FimT_{Pa}$, $FimT_{Ab}$ and $FimT_{Xc}$ and used to generate a FimT HMM profile using HMMER[74] (hmmbuild function). A FimU HMM profile was generated from sequence set 2 (FimU homologues), following alignment with MUSCLE and removal of entries showing >80% amino acid identity to another entry. Each sequence from the diverse FimT/U set was scanned against both the FimU and FimT sequence HMMs and reported as a likely FimT if its match score to the FimT profile was >20 points greater than its match to the FimU profile. In this way, a set of 196 putative FimT protein sequences was obtained. FimT protein sequences were aligned using MUSCLE[75] with default (high-accuracy) settings, and the alignment was visualised and manually improved using JalView[76]. The FimT alignment was processed using TrimAL[77] to remove low-quality positions and uninformative sequences (parameters: -strictplus -resoverlap 0.8 -seqoverlap 75). A maximum-likelihood phylogenetic tree of the FimT homologues was constructed using IQtree[78] with the substitution model LG+F+R5[79] and ultrafast bootstrapping[80]. The phylogenetic tree and associated metadata was viewed using iTol[81]. The tree was midpoint-rooted and branches with less than 50% bootstrap support removed. Gene neighbourhood diagrams for selected FimT homologues were generated using Clinker[82]. The FimT motif diagram was generated using WebLogo[83].

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. NMR spectra and corresponding model coordinates have been deposited in the BioMag Resonance Data Bank (BMRB: XXX) and Protein Data Bank (PDB ID: XXXX), respectively.

References

1. Johnsborg, O., Eldholm, V. & Håvarstein, L. S. Natural genetic transformation: prevalence, mechanisms and function. *Research in Microbiology* **158**, 767–778 (2007).

2. Johnston, C., Martin, B., Fichant, G., Polard, P. & Claverys, J.-P. Bacterial transformation: distribution, shared mechanisms and divergent control. *Nature reviews. Microbiology* **12**, 181–196 (2014).

3. Stone, B. J. & Kwaik, Y. A. Natural Competence for DNA Transformation by *Legionella pneumophila* and Its Association with Expression of Type IV Pili. *Journal of bacteriology* **181**, 1395–1402 (1999).

4. Gomez-Valero, L. *et al.* Extensive recombination events and horizontal gene transfer shaped the *Legionella pneumophila* genomes. *BMC Genomics* **12**, 536 (2011).

5. Sánchez-Busó, L., Comas, I., Jorques, G. & González-Candelas, F. Recombination drives genome evolution in outbreak-related *Legionella pneumophila* isolates. *Nat Genet* **46**, 1205–1211 (2014).

6. David, S. *et al.* Multiple major disease-associated clones of *Legionella pneumophila* have emerged recently and independently. *Genome Res* **26**, 1555–1564 (2016).

7. Newton, H. J., Ang, D. K. Y., Driel, I. R. van & Hartland, E. L. Molecular Pathogenesis of Infections Caused by *Legionella pneumophila*. *Clin Microbiol Rev* **23**, 274–298 (2010).

8. Cunha, P. B. A., Burillo, A. & Bouza, P. E. Legionnaires' disease. *The Lancet* **387**, 376–385 (2016).

9. Piepenbrink, K. H. DNA Uptake by Type IV Filaments. *Frontiers in Molecular Biosciences* **6**, 1441–13 (2019).

10. Hardy, L., Juan, P.-A., Coupat-Goutaland, B. & Charpentier, X. Transposon Insertion Sequencing in a Clinical Isolate of *Legionella pneumophila* Identifies Essential Genes and Determinants of Natural Transformation. *J Bacteriol* **203**, e00548-20 (2021).

11. Dubnau, D. & Blokesch, M. Mechanisms of DNA Uptake by Naturally Competent Bacteria. *Annual Review of Genetics* **53**, 217–237 (2019).

12. Jacobsen, T., Bardiaux, B., Francetic, O., Izadi-Pruneyre, N. & Nilges, M. Structure and function of minor pilins of type IV pili. *Medical Microbiology and Immunology* **209**, 301–308 (2020).

731 13. Berry, J.-L. & Pelicic, V. Exceptionally widespread nanomachines composed of type IV
732 pilins: the prokaryotic Swiss Army knives. *FEMS Microbiology Reviews* **39**, 134–154 (2015).

733 14. Wolfgang, M. *et al.* PilT mutations lead to simultaneous defects in competence for natural
734 transformation and twitching motility in piliated *Neisseria gonorrhoeae*. *Molecular*
735 *Microbiology* **29**, 321–330 (1998).

736 15. Ellison, C. K. *et al.* Retraction of DNA-bound type IV competence pili initiates DNA
737 uptake during natural transformation in *Vibrio cholerae*. *Nature Microbiology* **3**, 773–780
738 (2018).

739 16. Seitz, P. & Blokesch, M. DNA-uptake machinery of naturally competent *Vibrio cholerae*.
740 *Proceedings of the National Academy of Sciences* **110**, 17987–17992 (2013).

741 17. Weaver, S. J. *et al.* CryoEM structure of the type IVa pilus secretin required for natural
742 competence in *Vibrio cholerae*. *Nat Commun* **11**, 5080 (2020).

743 18. Seitz, P. *et al.* ComEA Is Essential for the Transfer of External DNA into the Periplasm
744 in Naturally Transformable *Vibrio cholerae* Cells. *PLoS Genet* **10**, e1004066-15 (2014).

745 19. Hepp, C. & Maier, B. Kinetics of DNA uptake during transformation provide evidence
746 for a translocation ratchet mechanism. *Proceedings of the National Academy of Sciences* **113**,
747 12467–12472 (2016).

748 20. Draskovic, I. & Dubnau, D. Biogenesis of a putative channel protein, ComEC, required
749 for DNA uptake: membrane topology, oligomerization and formation of disulphide bonds.
750 *Molecular microbiology* **55**, 881–896 (2004).

751 21. Attaiech, L. *et al.* Role of the Single-Stranded DNA–Binding Protein SsbB in
752 Pneumococcal Transformation: Maintenance of a Reservoir for Genetic Plasticity. *Plos Genet*
753 **7**, e1002156 (2011).

754 22. Bergé, M., Mortier-Barrière, I., Martin, B. & Claverys, J.-P. Transformation of
755 *Streptococcus pneumoniae* relies on DprA- and RecA-dependent protection of incoming
756 DNA single strands. *Molecular microbiology* **50**, 527–536 (2003).

757 23. Mortier-Barrière, I. *et al.* A Key Presynaptic Role in Transformation for a Widespread
758 Bacterial Protein: DprA Conveys Incoming ssDNA to RecA. *Cell* **130**, 824–836 (2007).

759 24. Nero, T. M. *et al.* ComM is a hexameric helicase that promotes branch migration during
760 natural transformation in diverse Gram-negative species. *Nucleic Acids Research* **46**, 6099–
761 6111 (2018).

762 25. Schaik, E. J. van *et al.* DNA Binding: a Novel Function of *Pseudomonas aeruginosa*
763 Type IV Pili. *Journal of bacteriology* **187**, 1455–1464 (2005).

764 26. Laurenceau, R. *et al.* A Type IV Pilus Mediates DNA Binding during Natural
765 Transformation in *Streptococcus pneumoniae*. *PLoS pathogens* **9**, e1003473-12 (2013).

766 27. Cehovin, A. *et al.* Specific DNA recognition mediated by a type IV pilin. *Proceedings of*
767 *the National Academy of Sciences* **110**, 3065–3070 (2013).

768 28. Berry, J.-L. *et al.* A Comparative Structure/Function Analysis of Two Type IV Pilin DNA
769 Receptors Defines a Novel Mode of DNA Binding. *Structure* **24**, 926–934 (2016).

770 29. Salleh, M. Z. *et al.* Structure and Properties of a Natural Competence-Associated Pilin
771 Suggest a Unique Pilus Tip-Associated DNA Receptor. *mBio* **10**, e00614-19 (2019).

772 30. Neuhaus, A. *et al.* Cryo-electron microscopy reveals two distinct type IV pili assembled
773 by the same bacterium. *Nature Communications* 1–13 (2020) doi:10.1038/s41467-020-
774 15650-w.

775 31. Goodman, S. D. & Scocca, J. J. Identification and arrangement of the DNA sequence
776 recognized in specific transformation of *Neisseria gonorrhoeae*. *Proc National Acad Sci* **85**,
777 6982–6986 (1988).

778 32. Graves, J. F., Biswas, G. D. & Sparling, P. F. Sequence-specific DNA uptake in
779 transformation of *Neisseria gonorrhoeae*. *J Bacteriol* **152**, 1071–1077 (1982).

780 33. Mell, J. C. & Redfield, R. J. Natural competence and the evolution of DNA uptake
781 specificity. *Journal of bacteriology* **196**, 1471–1483 (2014).

782 34. Berry, J.-L., Cehovin, A., McDowell, M. A., Lea, S. M. & Pelicic, V. Functional Analysis
783 of the Interdependence between DNA Uptake Sequence and Its Cognate ComP Receptor
784 during Natural Transformation in *Neisseria* Species. *Plos Genet* **9**, e1004014 (2013).

785 35. Leong, C. G. *et al.* The role of core and accessory type IV pilus genes in natural
786 transformation and twitching motility in the bacterium *Acinetobacter baylyi*. *PloS one* **12**,
787 e0182139-25 (2017).

788 36. Giltner, C. L., Nguyen, Y. & Burrows, L. L. Type IV pilin proteins: versatile molecular
789 modules. *Microbiology and molecular biology reviews : MMBR* **76**, 740–772 (2012).

790 37. Nguyen, Y. *et al. Pseudomonas aeruginosa* Minor Pilins Prime Type IVa Pilus Assembly
791 and Promote Surface Display of the PilY1 Adhesin. *J Biol Chem* **290**, 601–611 (2015).

792 38. Craig, L., Pique, M. E. & Tainer, J. A. Type IV pilus structure and bacterial
793 pathogenicity. *Nature reviews. Microbiology* **2**, 363–378 (2004).

794 39. Yanez, M. E., Korotkov, K. K., Abendroth, J. & Hol, W. G. J. Structure of the Minor
795 Pseudopilin EpsH from the Type 2 Secretion System of *Vibrio cholerae*. *Journal of*
796 *Molecular Biology* **377**, 91–103 (2008).

797 40. Raghunathan, K. *et al.* The 1.59 Å resolution structure of the minor pseudopilin EpsH of
798 *Vibrio cholerae* reveals a long flexible loop. *Biochimica et Biophysica Acta (BBA) - Proteins*
799 *and Proteomics* **1844**, 406–415 (2014).

800 41. Cierpicki, T. & Otlewski, J. Amide proton temperature coefficients as hydrogen bond
801 indicators in proteins. *J Biomol Nmr* **21**, 249–261 (2001).

802 42. Corona, R. I. & Guo, J. Statistical analysis of structural determinants for protein-DNA-
803 binding specificity. *Proteins* **84**, 1147–1161 (2016).

804 43. Douzi, B. *et al.* The XcpV/GspI Pseudopilin Has a Central Role in the Assembly of a
805 Quaternary Complex within the T2SS Pseudopilus. *J Biol Chem* **284**, 34580–34589 (2009).

806 44. Korotkov, K. V. & Sandkvist, M. Protein Secretion in Bacteria. *Ecosal Plus* **8**, 227–244
807 (2019).

808 45. Treuner-Lange, A. *et al.* PilY1 and minor pilins form a complex priming the type IVa
809 pilus in *Myxococcus xanthus*. *Nat Commun* **11**, 5054 (2020).

810 46. Alm, R. A. & Mattick, J. S. Identification of two genes with prepilin-like leader
811 sequences involved in type 4 fimbrial biogenesis in *Pseudomonas aeruginosa*. *J Bacteriol*
812 **178**, 3809–3817 (1996).

813 47. Taguchi, F. & Ichinose, Y. Role of Type IV Pili in Virulence of *Pseudomonas syringae*
814 pv. tabaci 6605: Correlation of Motility, Multidrug Resistance, and HR-Inducing Activity on
815 a Nonhost Plant. *Mol Plant-microbe Interactions* **24**, 1001–1011 (2011).

816 48. Belete, B., Lu, H. & Wozniak, D. J. *Pseudomonas aeruginosa* AlgR Regulates Type IV
817 Pilus Biosynthesis by Activating Transcription of the fimU-pilVWXY1Y2E Operon. *J*
818 *Bacteriol* **190**, 2023–2030 (2008).

819 49. Douzi, B., Ball, G., Cambillau, C., Tegoni, M. & Voulhoux, R. Deciphering the Xcp
820 *Pseudomonas aeruginosa* Type II Secretion Machinery through Multiple Interactions with
821 Substrates. *J Biol Chem* **286**, 40792–40801 (2011).

822 50. Denise, R., Abby, S. S. & Rocha, E. P. C. The Evolution of Protein Secretion Systems by
823 Co-option and Tinkering of Cellular Machineries. *Trends Microbiol* **28**, 372–386 (2020).

824 51. Korotkov, K. V. & Sandkvist, M. Architecture, Function, and Substrates of the Type II
825 Secretion System. *Ecosal Plus* **8**, (2019).

826 52. DebRoy, S., Dao, J., Söderberg, M., Rossier, O. & Cianciotto, N. P. *Legionella*
827 *pneumophila* type II secretome reveals unique exoproteins and a chitinase that promotes
828 bacterial persistence in the lung. *Proc National Acad Sci* **103**, 19146–19151 (2006).

829 53. Lam, T. *et al.* Competence pili in *Streptococcus pneumoniae* are highly dynamic
830 structures that retract to promote DNA uptake. *Mol Microbiol* 00: 1-16 (2021)
831 doi:10.1111/mmi.14718.

832 54. Chung, Y. S. & Dubnau, D. All Seven comG Open Reading Frames Are Required for
833 DNA Binding during Transformation of Competent *Bacillus subtilis*. *Journal of bacteriology*
834 **180**, 41–45 (1998).

835 55. Laurenceau, R. *et al.* A Type IV Pilus Mediates DNA Binding during Natural
836 Transformation in *Streptococcus pneumoniae*. *PLoS pathogens* **9**, e1003473-12 (2013).

837 56. Merriam, J. J., Mathur, R., Maxfield-Boumil, R. & Isberg, R. R. Analysis of the
838 *Legionella pneumophila* fliI Gene: Intracellular Growth of a Defined Mutant Defective for
839 Flagellum Biosyntheis. *Infection and immunity* **65**, 2497–2501 (1997).

840 57. Zuckman, D. M., Hung, J. B. & Roy, C. R. Pore-forming activity is not sufficient for
841 *Legionella pneumophila* phagosome trafficking and intracellular growth. *Molecular*
842 *microbiology* **32**, 990–1001 (1999).

843 58. Roy, C. R. & Isberg, R. R. Topology of *Legionella pneumophila* DotA: an Inner
844 Membrane Protein Required for Replication in Macrophages. *Infection and immunity* **65**,
845 571–578 (1997).

846 59. Berrow, N. S. *et al.* A versatile ligation-independent cloning method suitable for high-
847 throughput expression screening applications. *Nucleic Acids Research* **35**, e45–e45 (2007).

848 60. Assenberg, R. *et al.* Expression, purification and crystallization of a lyssavirus matrix (M)
849 protein. *Acta Crystallogr Sect F Struct Biology Cryst Commun* **64**, 258–262 (2008).

850 61. Chen, J. *et al. Legionella* Effectors That Promote Nonlytic Release from Protozoa.
851 *Science* **303**, 1358–1361 (2004).

852 62. Muhandiram, D. R. & Kay, L. E. Gradient-Enhanced Triple-Resonance Three-
853 Dimensional NMR Experiments with Improved Sensitivity. *J Magnetic Reson Ser B* **103**,
854 203–216 (1994).

855 63. Kovacs, H. & Gossert, A. Improved NMR experiments with 13C-isotropic mixing for
856 assignment of aromatic and aliphatic side chains in labeled proteins. *J Biomol Nmr* **58**, 101–
857 112 (2014).

858 64. Frueh, D. P. *et al.* Time-shared HSQC-NOESY for accurate distance constraints
859 measured at high-field in 15N-13C-ILV methyl labeled proteins. *J Biomol Nmr* **45**, 311
860 (2009).

861 65. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: Development of a
862 software pipeline. *Proteins Struct Funct Bioinform* **59**, 687–696 (2005).

863 66. Güntert, P. & Buchner, L. Combined automated NOE assignment and structure
864 calculation with CYANA. *J Biomol Nmr* **62**, 453–471 (2015).

865 67. Case, D. A. *et al. Amber 2021*. (2021), Amber 2021, University of California, San
866 Francisco.

867 68. López-Méndez, B. *et al.* Reproducibility and accuracy of microscale thermophoresis in
868 the NanoTemper Monolith: a multi laboratory benchmark study. *Eur Biophys J* **50**, 411–427
869 (2021).

870 69. Sexton, J. A. & Vogel, J. P. Regulation of hypercompetence in *Legionella pneumophila*.
871 *Journal of bacteriology* **186**, 3814–3825 (2004).

872 70. Charpentier, X., Kay, E., Schneider, D. & Shuman, H. A. Antibiotics and UV radiation
873 induce competence for natural transformation in *Legionella pneumophila*. *Journal of*
874 *bacteriology* **193**, 1114–1121 (2011).

875 71. Schäffer, A. A. *et al.* Improving the accuracy of PSI-BLAST protein database searches
876 with composition-based statistics and other refinements. *Nucleic Acids Res* **29**, 2994–3005
877 (2001).

878 72. Cock, P. J. A. *et al.* Biopython: freely available Python tools for computational molecular
879 biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).

880 73. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412-
881 D419 (2020).

882 74. Eddy, S. R. Accelerated Profile HMM Searches. *Plos Comput Biol* **7**, e1002195 (2011).

883 75. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and
884 space complexity. *Bmc Bioinformatics* **5**, 113 (2004).

885 76. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview
886 Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**,
887 1189–1191 (2009).

888 77. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated
889 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973
890 (2009).

891 78. Nguyen, L.-T., Schmidt, H. A., Haeseler, A. von & Minh, B. Q. IQ-TREE: A Fast and
892 Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol*
893 *Evol* **32**, 268–274 (2015).

894 79. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Haeseler, A. von & Jermiin, L. S.
895 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**, 587–
896 589 (2017).

897 80. Hoang, D. T., Chernomor, O., Haeseler, A. von, Minh, B. Q. & Vinh, L. S. UFBoot2:
898 Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* **35**, 518–522 (2017).

899 81. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic
900 tree display and annotation. *Nucleic Acids Res* **49**, W293-W296 (2021)
901 doi:10.1093/nar/gkab301.

902 82. Gilchrist, C. L. M. & Chooi, Y.-H. clinker & clustermap.js: automatic generation of gene
903 cluster comparison figures. *Bioinformatics* (2021) doi:10.1093/bioinformatics/btab007.

904 83. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo
905 Generator. *Genome Res* **14**, 1188–1190 (2004).

906 84. Keene, O. N. The log transformation is special. *Stat Med* **14**, 811–819 (1995).

907 85. Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores of
908 residues on protein structures. *Nucleic Acids Res* **33**, W299–W302 (2005).

909 **Acknowledgements**

910 This work was funded by an SNSF PRIMA grant PR00P3_179728 to MKH. FLS is supported
911 by an Australian Research Council Discovery Early Career Research Award DE200101524.
912 We would like to thank G. Waksman and A. Meir for the Lp02, CR019 and DH5α λpir strains,
913 and the pSR47S plasmid. We would also like to thank H. Hilbi for the pMMB207C plasmid.
914 We are grateful to J. Scheuermann for the use of the VP-ITC instrument.

915

916 **Author Contributions**

917 SAGB cloned constructs, created *Legionella* strains, purified proteins, performed DNA
918 binding studies, transformation assays, Western blots and analysed results. FLS designed
919 and performed all bioinformatic analyses. SH constructed FimT chimera constructs and
920 performed the corresponding transformation assays. MJMS purified proteins and performed
921 ITC experiments. ADG performed and analysed all NMR-related experiments with help from
922 SAGB. MKH designed and supervised the study, made figures and wrote the manuscript
923 with help from all authors.

924

925 **Competing Interests Statement**

926 The authors declare no competing interests.

927

Figures

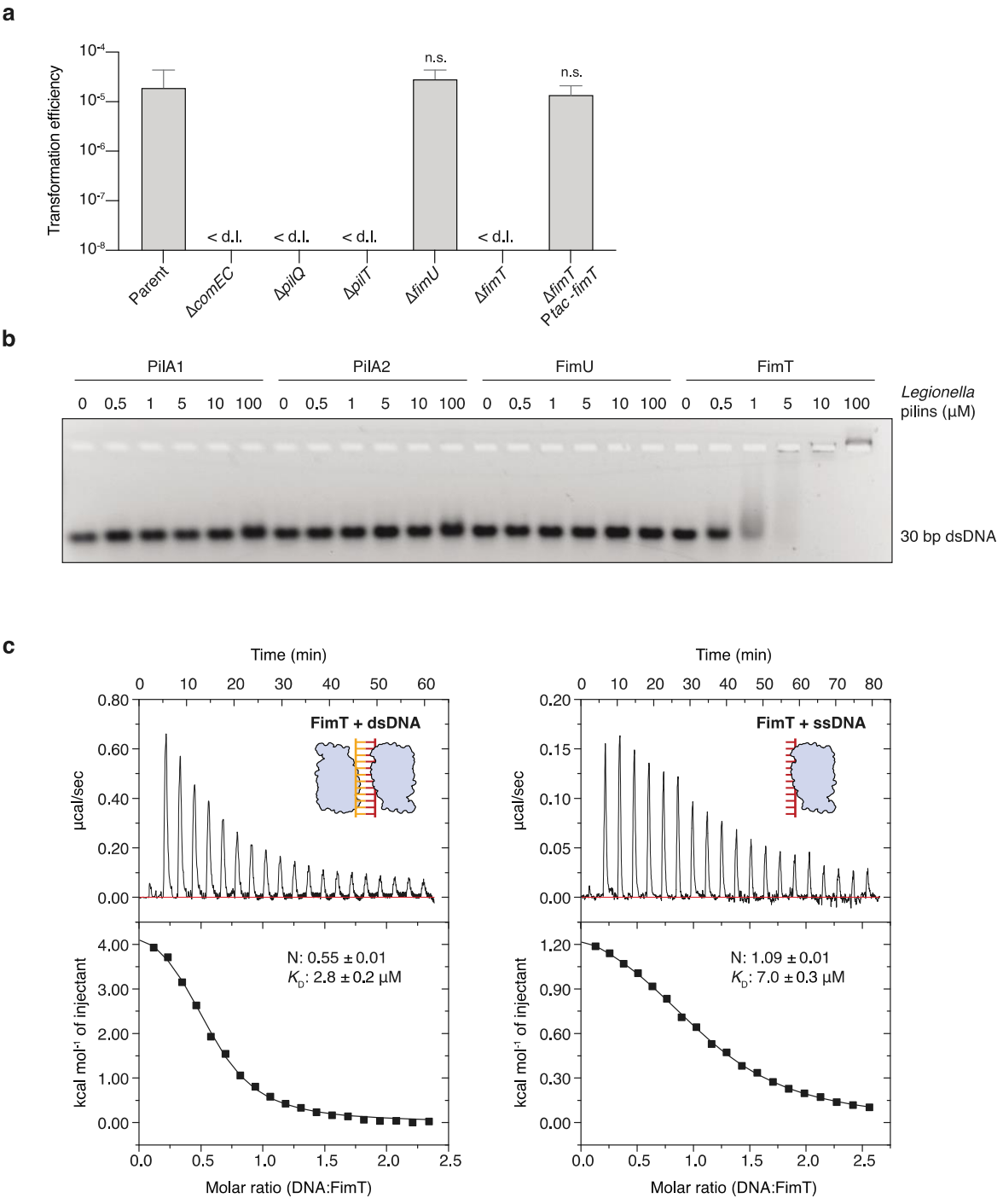


Figure 1: FimT is critical for the transformation of *L. pneumophila* and binds to DNA

a, Natural transformation efficiencies of the parental *L. pneumophila* Lp02 strain and Lp02 strains harbouring deletions of genes known to play a role in transformation compared to the *fimU* and *fimT* deletion strains. The Δ*fimT* strain was complemented by ectopic expression of wild-type FimT, under the control of an IPTG-inducible promoter. The mean transformation efficiencies of three independent biological replicates is shown (error bars represent

938 standard deviation [SD]). <d.l., below detection limit (d.l.) (average d.l. = 2.0 x 10^{-8} ± 8.2
939 x 10^{-9}). Statistical significances of transformation differences were determined on log-
940 transformed[84] data using an unpaired t-test with Welch's correction. n.s., not statistically
941 significant (p>0.05). **b**, *In vitro* DNA binding of purified *L. pneumophila* PilA1, PilA2, FimU
942 and FimT assessed by an EMSA. A 30 bp dsDNA fragment (1 µM) was incubated with
943 increasing concentrations of purified pilins (0–100 µM) and resolved by agarose gel
944 electrophoresis. **c**, ITC binding studies of wild-type FimT binding to 12meric dsDNA (right)
945 and ssDNA (left). In both cases, DNA (syringe) was injected into FimT (cell). Data were fitted
946 using the "one set" of sites model, assuming that both binding sites on the dsDNA are of
947 equal affinity.
948

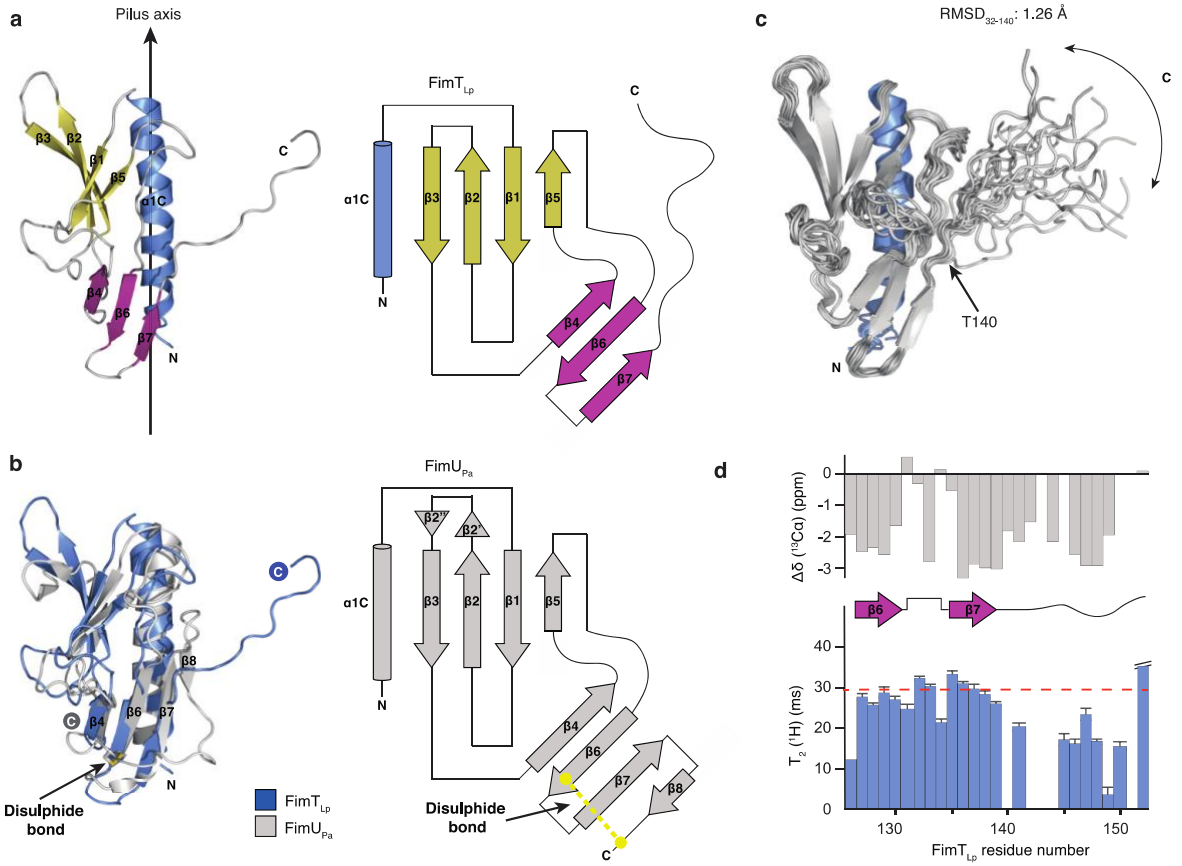


Figure 2: The structure of FimT$_{Lp}$

a, The solution structure of FimT$_{Lp}$ 28-152 (state 18) in ribbon representation (left) and the corresponding topology diagram (right). Secondary structure elements are indicated: truncated N-terminal α-helix (α1C) (blue), β-sheet I formed by β1, β2, β3 and β5 (yellow), and β-sheet II formed by β4, β6 and β7 (magenta). A vertical arrow indicates the pilus axis from the cell surface towards the pilus tip. **b**, Structure alignment of FimT$_{Lp}$ (blue) and FimU$_{Pa}$ (grey; PDB ID: 4IPV) (left) and the topology diagram of FimU$_{Pa}$ (right). The disulphide bond of FimU$_{Pa}$ is indicated in stick representation with sulphur atoms in yellow. **c**, Superimposed 20 lowest energy structures calculated by NMR spectroscopy. An arrow indicates the conformational flexibility of the C-terminal tail (140-152). The pairwise backbone root-mean-square deviation (RMSD) for the structured region (residues 32 to 140) is 1.26 Å. N- and C-termini are indicated in each panel. **d**, Cα chemical shift values (top) and T_2(^{1}H) transverse relaxation data (bottom), encompassing the last 27 residues of FimT$_{Lp}$. Secondary structural elements are indicated and error bars represent the fitting errors of the respective exponential decay curves.

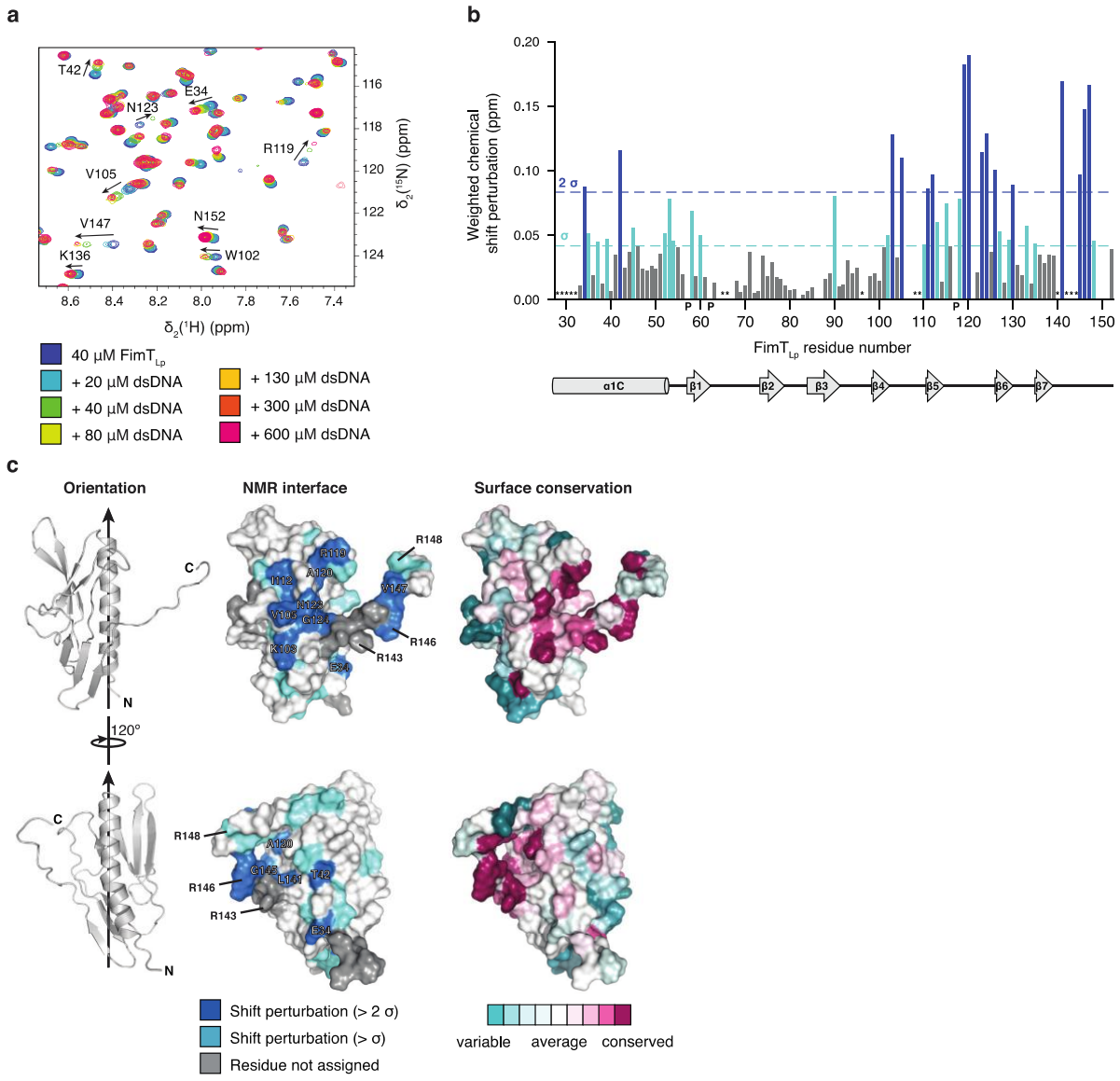


Figure 3: Identification of the DNA interaction surface of $FimT_{Lp}$

a, Selected region of ^{1}H, ^{15}N-HSQC spectra showing ^{15}N-labeled $FimT_{Lp}$ alone and in presence of increasing concentrations of 12 bp dsDNA. Full spectra are in **Source Data**. **b**, Weighted CSP map generated from a. Residues experiencing CSPs (Δppm > 1 σ), light blue; residues experiencing CSPs (Δppm > 2 σ), dark blue; P, prolines; *, residues not assigned. **c**, Left, $FimT_{Lp}$ is shown in two orientations rotated by 120° in ribbon representation. Arrows indicate the pilus axis as in Fig. 2a. Middle, CSPs are mapped onto the surface of $FimT_{Lp}$ and coloured as in b. Residues producing large shifts are labelled on the molecular surface. Right, Surface residues of $FimT_{Lp}$ are coloured according to conservation (full multisequence alignment in **Source Data**). This image was generated using the ConSurf server[85].

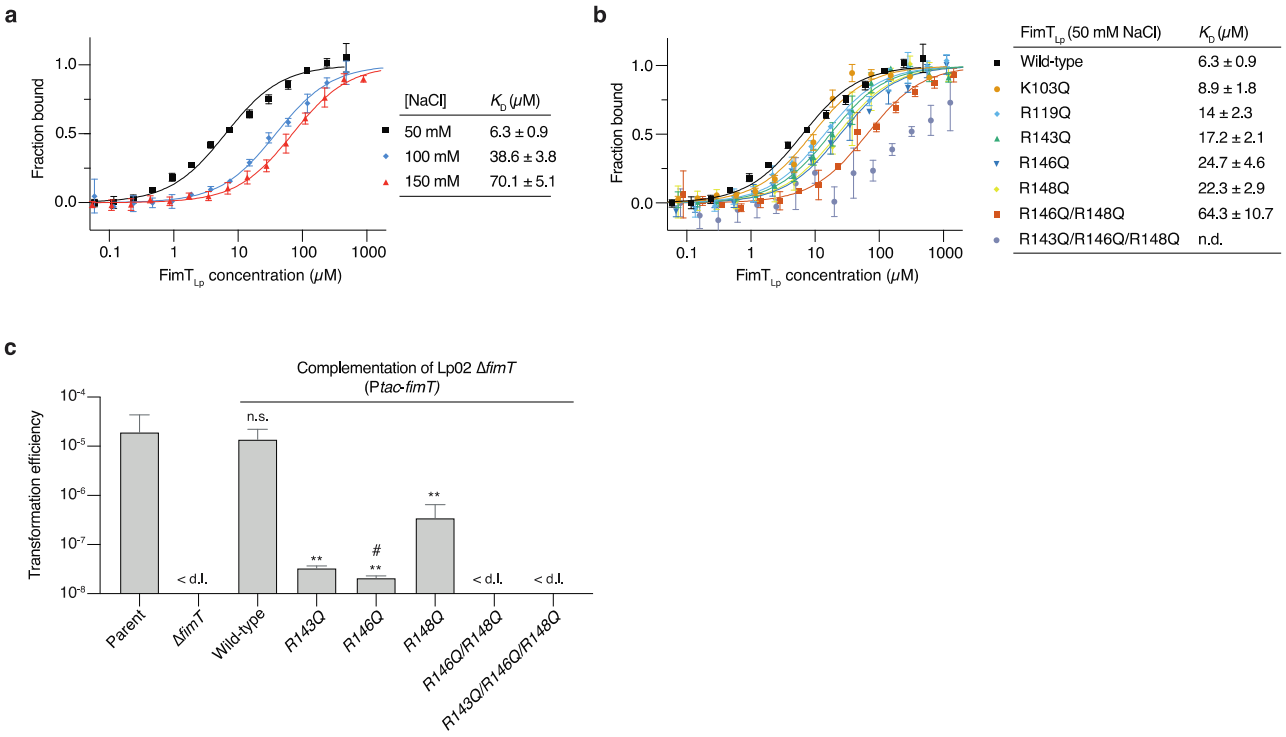


Figure 4: Characterisation of FimT$_{Lp}$ binding to DNA *in vitro* and *in vivo*

MST/TRIC binding assay of 12 bp FAM-labelled dsDNA with **a,** wild-type FimT$_{Lp}$ performed at increasing NaCl concentrations (ionic strength) and **b**, wild-type FimT$_{Lp}$ compared to FimT mutants. n.d., not determined. The MST/TRIC data were fitted according to two binding sites with equal affinity. Error bars represent the mean ± SD. **c,** Natural transformation efficiencies of parental Lp02, Lp02 Δ*fimT*, and the Lp02 Δ*fimT* strain complemented by ectopic expression of wild-type and FimT$_{Lp}$ mutants. The mean transformation efficiencies of three independent biological replicates are plotted with error bars representing the SD. <d.l., below d.l. (average d.l. = 2.0 x 10^{-8} ± 8.2 x 10^{-9}); #, below d.l. in at least one replicate (average d.l. used to calculate the mean transformation efficiency). These assays were performed in parallel to those displayed in Fig. 1a, and statistical differences were determined on log-transformed data using an unpaired t-test with Welch's correction. **, p<0.01; n.s., not statistically significant (p>0.05).

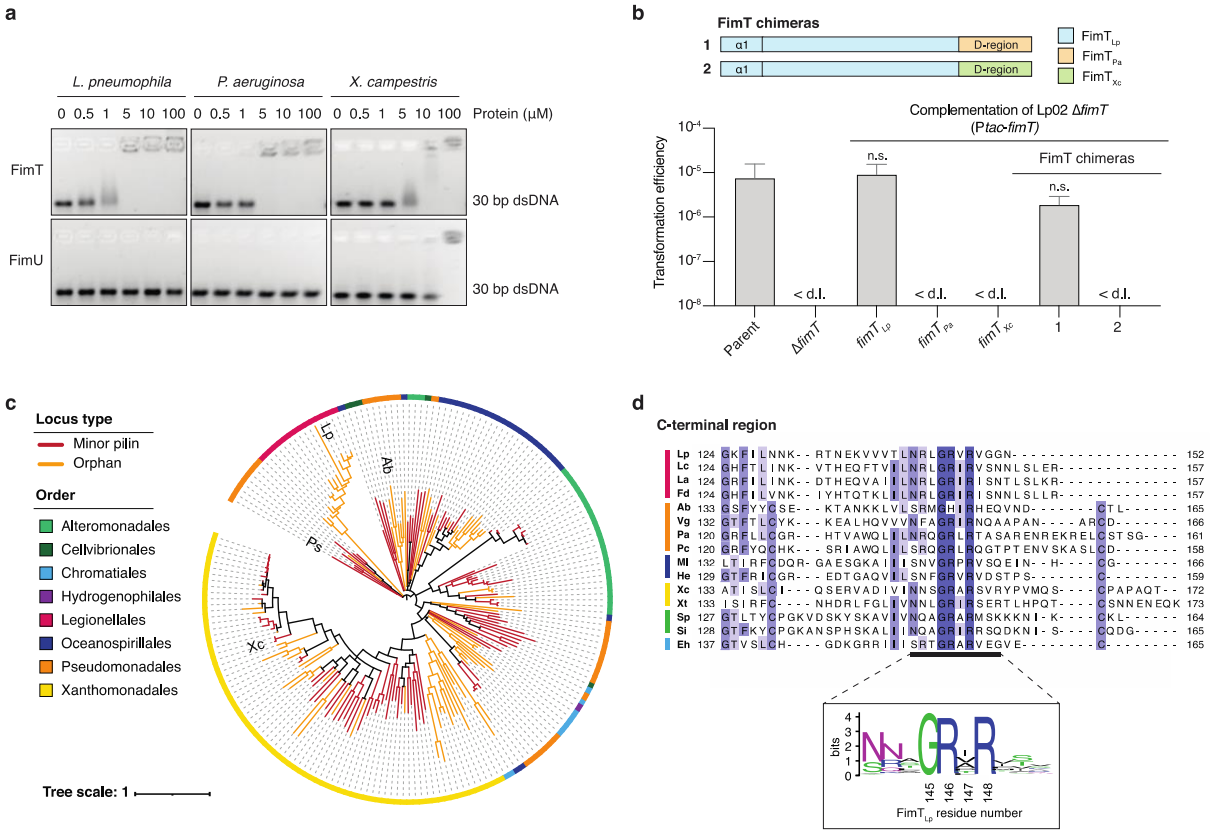
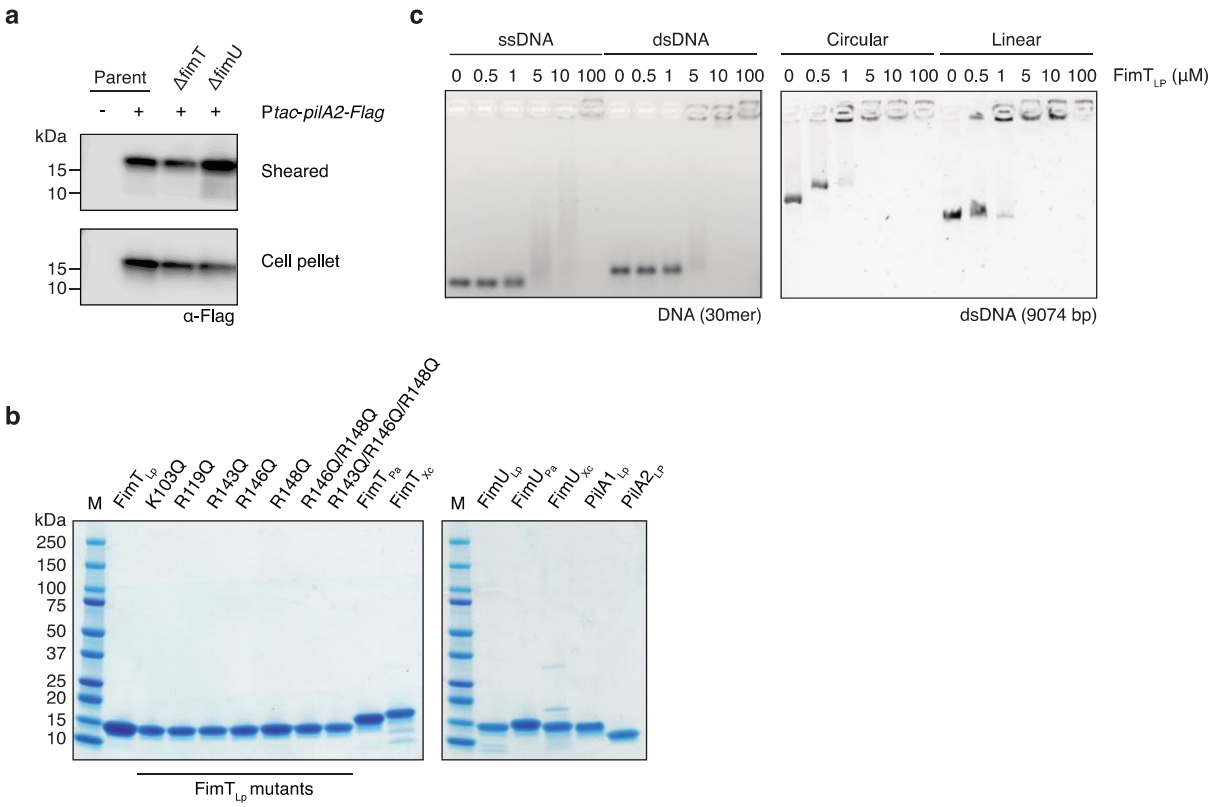


Figure 5: Bioinformatic and functional analysis of FimT orthologues

a, EMSA showing *in vitro* DNA binding of purified FimT and FimU orthologues from *L. pneumophila*, *P. aeruginosa* and *X. campestris*. A 30 bp dsDNA fragment (1 μM) was incubated with increasing concentrations of purified pilins (0-100 μM) and resolved by agarose gel electrophoresis. **b**, A comparison of natural transformation efficiencies of the Lp02 Δ*fimT* strain complemented by ectopic expression of $FimT_{Lp}$, FimT orthologues from *P. aeruginosa* ($FimT_{Pa}$) and *X. campestris* ($FimT_{Xc}$), or chimeric FimT mutants (1-2). The corresponding composition of these FimT chimeras (1-2) is explained by a schematic drawing (top). The mean transformation frequencies of three independent biological replicates are shown with error bars representing the SD. <d.l., below d.l. (average d.l. = 4.8 x 10^{-8} ± 2.1 x 10^{-8}). An unpaired t-test with Welch's correction, using log-transformed data, was used to analyse statistical significance. n.s., not statistically significant (p>0.05). **c**, Phylogenetic tree of FimT homologues, comprising eight orders of γ-proteobacteria illustrated by the coloured circumferential ring. Branches coloured in orange represent FimTs encoded as orphan genes, whereas those coloured red represent FimTs encoded within minor pilin operons. The positions of the four functionally characterised FimT orthologues in the tree are indicated (Lp, *L. pneumophila*; Ab, *A. baylyi*; Pa, *P. aeruginosa*; and Xc,

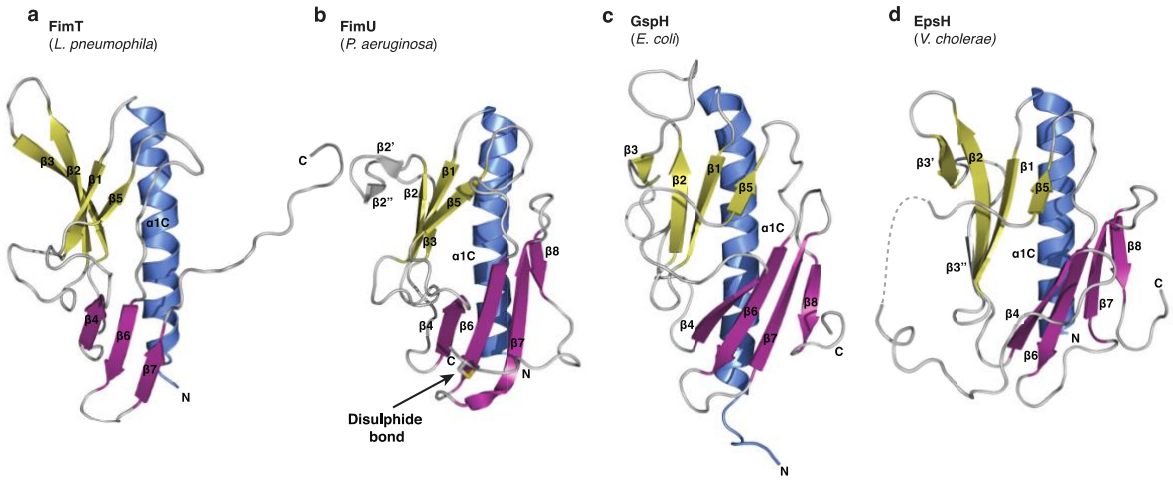
X. campestris). The scale bar indicates the average number of substitutions per site. **d**, Top, multisequence alignment of representative FimT orthologues across six orders (indicated by a coloured line as in c) focusing on their C-terminal region (Lc, *Legionella cherrii*; La, *Legionella anisa*; Fd, *Fluoribacter dumoffii*; Vg, *Ventosimonas gracilis*; Pc, *Pseudomonas chloritidismutans*; Ml, *Marinicella litoralis*; He, *Halomonas endophytica*; Xt, *Xylella taiwanensis*; Sp, *Shewanella polaris*; Si, *Shewanella indica*; Eh, *Ectothiorhodospira haloalkaliphile*). Residues are coloured according to sequence identity. Bottom, sequence logo generated from the full multisequence alignment of 196 high-confidence FimTs (**Source Data**).

Extended Data



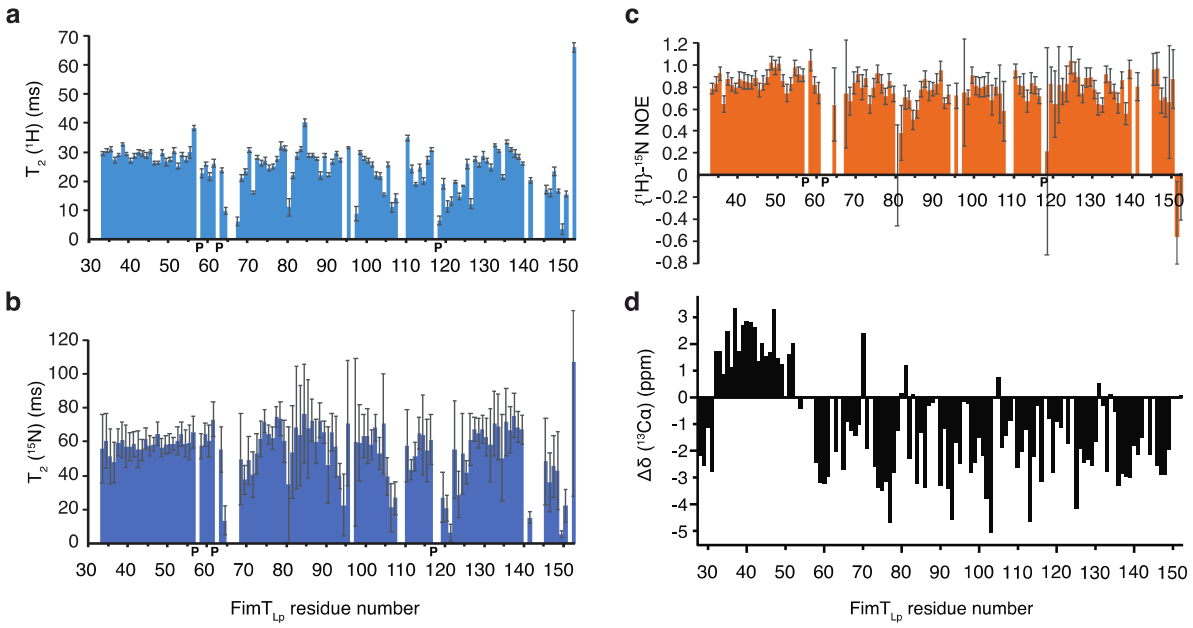
Extended Data Figure 1: Cell surface expression of PilA2-Flag, *in vitro* DNA binding of $FimT_{Lp}$ and purified proteins utilised in this study

a, Immunodetection of ectopically expressed PilA2-Flag in various Lp02 strains using anti-Flag antibodies (**Source Data**). Sheared pili were detected in supernatants (sheared) and the whole cell lysates of depiliated cells (cell pellet). **b**, All purified N-terminally truncated pilins (construct boundaries can be found in **Extended Data Table 3**), utilised in this study, resolved by SDS-PAGE. M, marker; Lp, *L. pneumophila*; Pa, *P. aeruginosa*; Xc, *X. campestris*. **c**, EMSAs showing *in vitro* DNA binding of $FimT_{Lp}$ to ssDNA *vs* dsDNA (left) and linear *vs* circular DNA (right). DNA probes were incubated with increasing concentrations of $FimT_{Lp}$ and resolved by agarose gel electrophoresis.



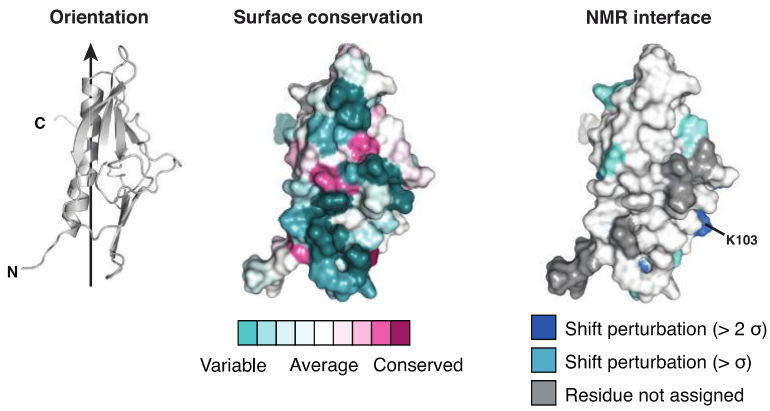
Extended Data Figure 2: Structures of GspH/FimT family members

a, The structure of FimT from *L. pneumophila* (state 18, this study); **b**, FimU from *P. aeruginosa* (PDB ID: 4IPV); **c**, GspH from *E. coli* (state 1, PDB ID: 2KNQ); and **d**, EpsH from *V. cholerae* (PDB ID: 2QV8). The $FimT_{Lp}$ and $GspH_{Ec}$ structures were determined using NMR spectroscopy, while those of $FimU_{Pa}$ and $EpsH_{Vc}$ are crystal structures. The disulphide bond of FimU is shown in stick representation (sulphur atoms in yellow), indicated by an arrow. The previously named β3 and β4-strands of the EpsH structure[1] have been labelled as β3' and β3'' for consistency of strand nomenclature across all depicted structures. All structures are shown in ribbon representation with their N-and C-termini indicated and secondary structural elements are coloured and labelled as in Fig. 2a.



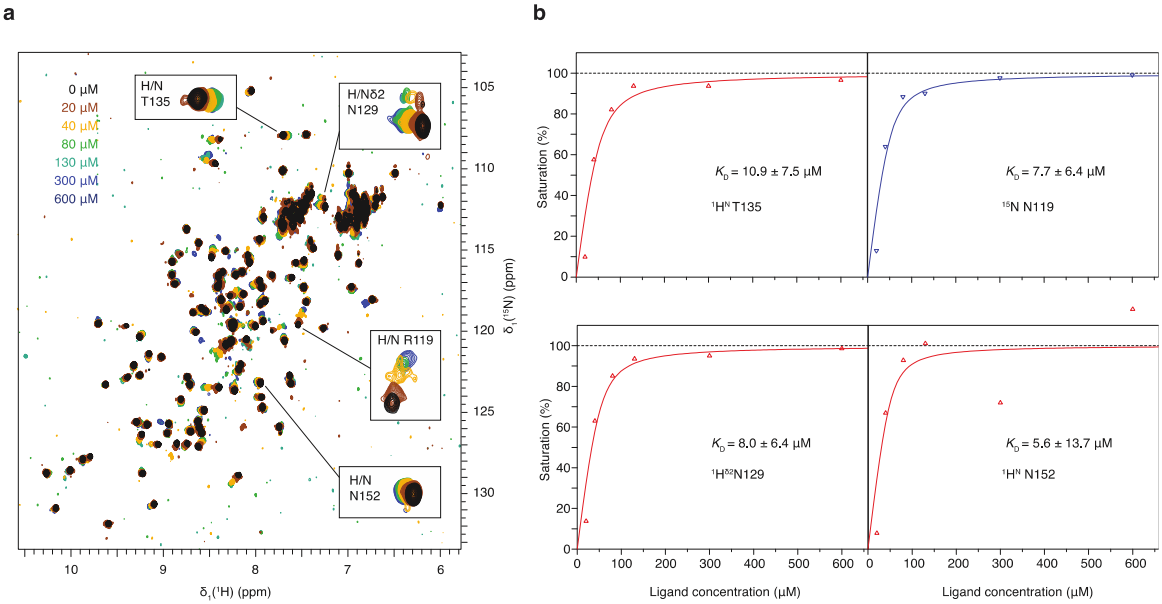
Extended Data Figure 3: Relaxation data for FimT$_{Lp}$ indicate dynamics of the C-terminal residues 140–150 on the millisecond timescale

a, **b**, Backbone amide T2 transverse relaxation data of FimT$_{Lp}$ for ^{1}H (**a**) and ^{15}N (**b**) nuclei, where amide groups of the loops and the C-terminus show significantly decreased T_2 values compared to the folded part of the domain. The low T_2 values for the C-terminal tail (signals of amides of residues 140 and 142–144 were too weak to be analysed), indicate dynamics of the C-terminal residues (140–150) on the microsecond to millisecond timescale. Proline residues are indicated with a bold letter P. Error bars represent the fitting errors of the respective exponential decay curves. **c**, Heteronuclear {^{1}H}-^{15}N NOE data show that only the last two residues (151 and 152) exhibit fast dynamics on the nanosecond timescale, typical for flexibly disordered termini. Error bars reflect the error from the signal-to-noise ratio of the individual signals used for the analysis. **d**, Cα chemical shift deviation from random coil values (Δδ(^{13}Cα)) indicate predominantly β-strand secondary structure for the C-terminal residues. Significant (>0.5 ppm) positive and negative deviations of ^{13}Cα chemical shifts from random coil values indicate α-helical and β-strand conformations of the backbone, respectively. ^{13}Cα chemical shifts are shown without smoothing, representing the raw data after calibration of the ^{13}C chemical shift to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS).



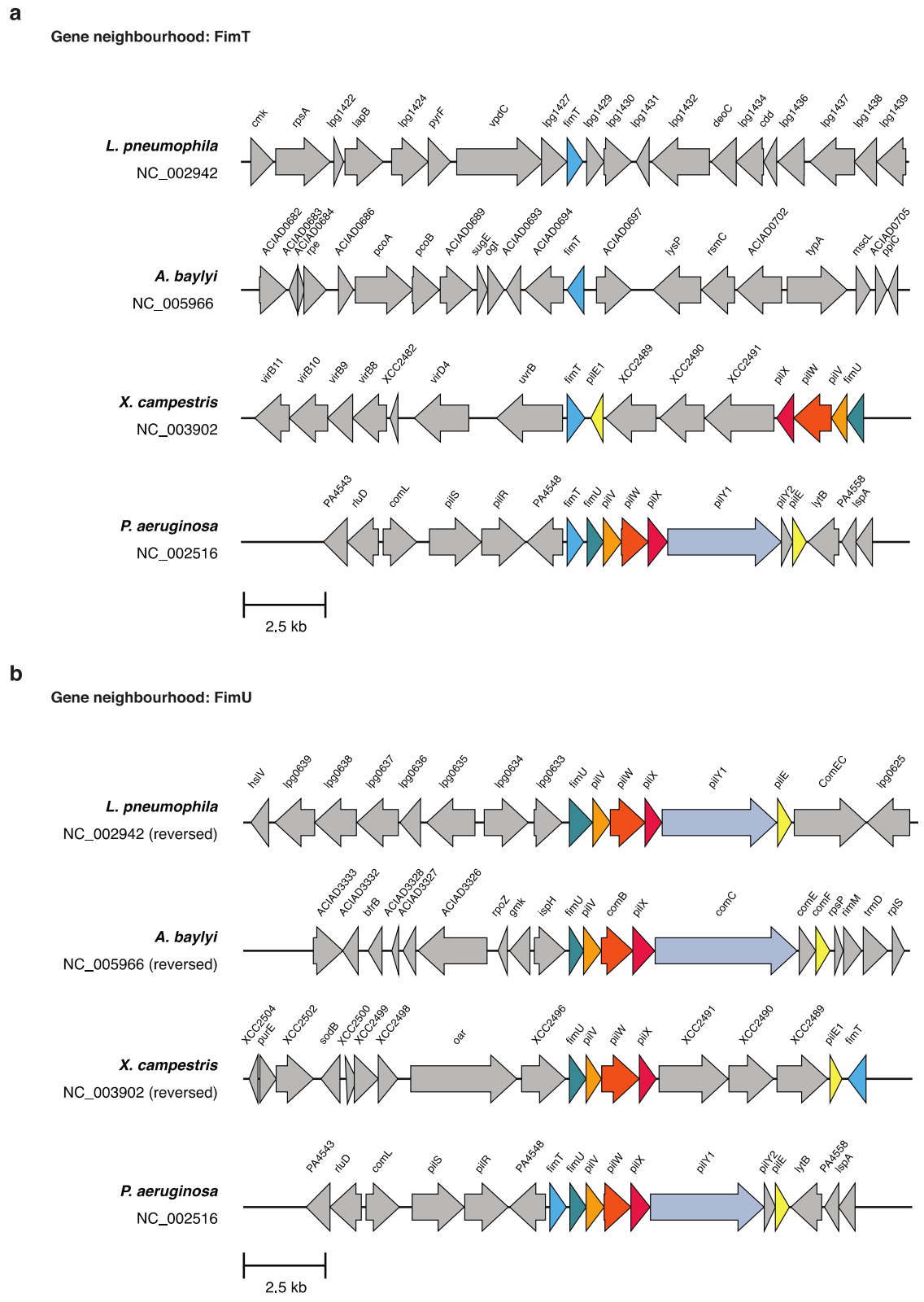
Extended Data Figure 4: NMR binding studies of $FimT_{Lp}$ to DNA

Left, $FimT_{Lp}$ is shown in ribbon representation rotated a further 120° with respect to the orientations displayed in Fig. 3c. Middle, residues experiencing chemical shift perturbations due to DNA binding are mapped onto the surface of $FimT_{Lp}$. Right, surface residues of $FimT_{Lp}$ are coloured according to conservation.



Extended Data Figure 5: Affinity determination of FimT to 12 bp dsDNA by NMR

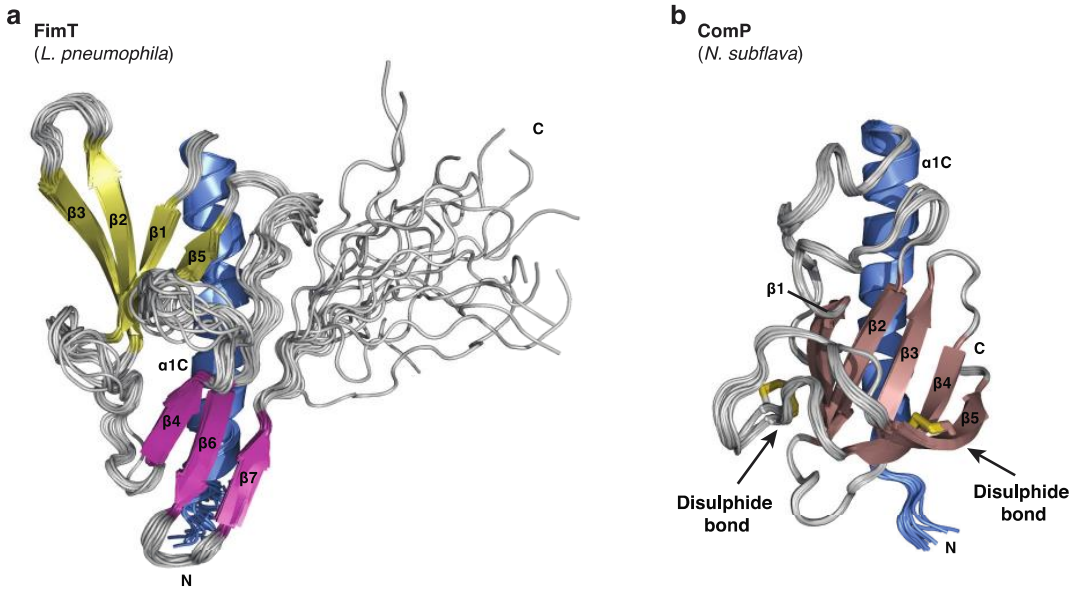
a, DNA binding studies of $FimT_{Lp}$ performed by NMR spectroscopy. Increasing concentrations of 12 bp dsDNA (see colour code on top left in spectra overlay) were added to 40 µM of ^{15}N-labelled $FimT_{LP}$ and the CSPs of four peaks were plotted against the ligand (12 bp DNA) concentration. **b**, For the four signals indicated in the spectra overlay, the binding curves are shown on the right-hand side, for ^{1}H and ^{15}N nuclei in red and blue triangles, respectively. The data were fitted assuming two identical binding sites (solid lines) and averaged to estimate a K_D of ~8 µM of the interaction.



Extended Data Figure 6: Gene neighbourhoods of FimT and FimU

Genomic regions around FimT (**a**) and FimU (**b**) in *L. pneumophila*, *A. baylyi*, *X. campestris* and *P. aeruginosa*. Each gene is labelled with its name or locus tag (if unannotated). Genes coding for T4P homologues are colour-coded identically across the different bacterial

1100 species. Among FimT and FimU homologues collected by BlastP using the four
1101 representative sequences, 25% of FimT sequences were located close to other minor pilin
1102 operon components, while 100% of FimU sequences were located in minor pilin operons
1103 (see **Source Data**).
1104



Extended Data Figure 7: Comparison of the NMR structures of FimT and ComP

a, b, Superimposed 20 lowest energy structures calculated by NMR spectroscopy of FimT from *L. pneumophila* (**a**) and ComP from *Neisseria subflava* (PDB ID: 2NBA[3]) (**b**). The DD-region defining disulphide bonds of ComP are shown in stick representation (sulphur atoms in yellow) and are indicated by arrows. Both structures are shown in ribbon representation with their N-and C-termini indicated.

Extended Data Table 1 NMR and refinement statistics for $FimT_{Lp}$

	$FimT_{Lp}$
NMR distance and dihedral constraints	
Distance constraints	
Total NOE	2311
Intra-residue	635
Inter-residue	1676
Sequential ($\|i - j\| = 1$)	522
Medium-range ($\|i - j\| < 4$)	344
Long-range ($\|i - j\| > 5$)	810
Hydrogen bonds	-
Total dihedral angle restraints*	
Backbone	666
Other	558
Structure statistics	
Average Cyana target function	0.21 ± 0.02
Violations (mean and s.d.)**	
Distance constraints (Å)	0
Max. dihedral angle violation (º)	123.98 ± 28.54
Max. distance constraint violation (Å)	0.57 ± 0.19
Deviations from idealized geometry	
Bond lengths (Å)	0.0035 ± 0.0012
Bond angles (º)	1.377 ± 0.459
Average pairwise r.m.s. deviation*** (Å)	
Heavy	1.13 ± 0.13
Backbone	0.56 ± 0.16

* Dihedral angle restraints were derived from Cα chemical shifts using TALOS+ as implemented in cyana 3.98

** Restraints violated in 6 or more structures

*** Pairwise r.m.s. deviation for structured regions (res. 32–62, 70–139) was calculated among 20 refined structures.

Extended Data Table 2: Strains used in this study

Name	Relevant genotype/description	Source/Reference
Escherichia coli		
BL21 (DE3)	*E. coli expression strain*	NEB (cat. no. C2527H/I)
Shuffle T7	*E. coli expression strain*	NEB (cat. no. C3026J)
Stellar (HST08 strain)	*E. coli* cloning strain	Takara (cat. no. 636763/636766)
DH5α λpir	*E. coli cloning strain:* Encodes π protein for the replication of the *pir*-dependent origin of replication - *oriR*(*R6K)*	4,5
CR019	*E. coli mobilizing strain:* MT607 *E. coli* containing pRK600 plasmid [*oriR*(ColE1) *oriT*(RK2); CmR]	6
Legionella pneumophila		
Lp02 WT	Philadelphia-1 *rpsL hsdR thyA*; SmR	7
Lp02 Δ*fimT*	Lp02 Δ*fimT (lpg1428)*	This study
Lp02 Δ*fimU*	Lp02 Δ*fimU (lpg0632)*	This study
Lp02 Δ*pilQ*	Lp02 Δ*pilQ (lpg0931)*	This study
Lp02 Δ*pilT*	Lp02 Δ*pilT (lpg2013)*	This study
Lp02 Δ*comEC*	Lp02 Δ*comEC (lpg0626)*	This study

Extended Data Table 3: Plasmids used in this study

Name	Relevant genotype/description	Source/Reference
pMMB207C	*Legionella* expression vector derived from RSF1010: IncQ lacI^q P*tac oriT* Δ*mobA;* CmR	8
pMMB207C-*fimT*$_{Lp}$	*L. pneumophila* wild-type *fimT*	This study
pMMB207C-*fimT*$_{Lp}$ *R143Q*	pMMB207C-*fimT*$_{Lp}$, *with fimT R143Q mutation*	This study
pMMB207C-*fimT*$_{Lp}$ *R146Q*	pMMB207C-*fimT*$_{Lp}$, *with fimT R146Q mutation*	This study
pMMB207C-*fimT*$_{Lp}$ *R148Q*	pMMB207C-*fimT*$_{Lp}$, *with fimT R148Q mutation*	This study
pMMB207C-*fimT*$_{Lp}$ *R146Q, R148Q*	pMMB207C-*fimT*$_{Lp}$, with *fimT R146Q, R148Q mutations*	This study
pMMB207C-*fimT*$_{Lp}$ *R143Q, R146Q, R148Q*	pMMB207C-*fimT*$_{Lp}$, *with fimT R143Q, R146Q, R148Q mutations*	This study
pMMB207C-*pilA2-flag*	*L. pneumophila pilA2 (lpg1915)-flag*	This study
pMMB207C-*fimT*$_{Pa}$	*Pseudomonas aeruginosa PAO1 fimT (PA4549)*	This study
pMMB207C-*fimT*$_{\text{chimera 1}}$	pMMB207C-*fimT*$_{Lp}$ residues 1-128, fused to *fimT*$_{Pa}$ residues 125-161	This study
pMMB207C-*fimT*$_{Xc}$	*Xanthomonas campestris ATCC 33913 fimT (XCC2486)*	This study
pMMB207C-*fimT*$_{\text{chimera 2}}$	pMMB207C-*fimT*$_{Lp}$ residues 1-128, fused to *fimT*$_{Xc}$ residues 138-172	This study
pSR47S	Suicide plasmid: *oriR*(*R6K*) *oriT*(RP4) *sacB*; KanR	Vogel, J. P., et al. (unpublished data) 9
pSR47S-*fimT*	*L. pneumophila fimT* gene with 1000 bp up- and downstream sequence (homology regions)	This study
pSR47S-*fimU*	*L. pneumophila fimU* gene with 1000 bp up- and downstream sequence (homology regions)	This study
pSR47S-*pilQ*	*L. pneumophila pilQ* gene with 1000 bp up- and downstream sequence (homology regions)	This study
pSR47S-*pilT*	*L. pneumophila pilT* gene with 1000 bp up- and downstream sequence (homology regions)	This study
pSR47S-*comEC*	*L. pneumophila comEC* gene with 1000 bp up- and downstream sequence (homology regions)	This study
pSR47S-Δ*fimT*	pSR47S-*fimT*, with *fimT* deletion	This study
pSR47S-Δ*fimU*	pSR47S-*fimU*, with *fimU* deletion (52 nt left intact at 5' end of gene)	This study
pSR47S-Δ*pilQ*	pSR47S-*pilQ*, with *pilQ* deletion	This study
pSR47S-Δ*pilT*	pSR47S-*pilT*, with *pilT* deletion	This study
pSR47S-Δ*comEC*	pSR47S-*comEC*, with *comEC* deletion	This study
pOPINS	*E. coli* expression vector: N-terminal His_6-SUMO tag, T7 promoter; KanR	10
pOPINS-*fimT*$_{Lp}$	*L. pneumophila wild-type fimT, residues 28-152*	This study
pOPINS-*fimT*$_{Lp}$ *K103Q*	pOPINS-*fimT*$_{Lp}$, *with fimT K103Q mutation*	This study
pOPINS-*fimT*$_{Lp}$ *R119Q*	pOPINS-*fimT*$_{Lp}$, *with fimT R119Q mutation*	This study
pOPINS-*fimT*$_{Lp}$ *R143Q*	pOPINS-*fimT*$_{Lp}$, *with fimT R143Q mutation*	This study
pOPINS-*fimT*$_{Lp}$ *R146Q*	pOPINS-*fimT*$_{Lp}$, *with fimT R146Q mutation*	This study
pOPINS-*fimT*$_{Lp}$ *R148Q*	pOPINS-*fimT*$_{Lp}$, *with fimT R148Q mutation*	This study
pOPINS-*fimT*$_{Lp}$ *R146Q, R148Q*	pOPINS-*fimT*$_{Lp}$, with *fimT R146Q, R148Q mutations*	This study
pOPINS-*fimT*$_{Lp}$ *R143Q, R146Q, R148Q*	pOPINS-*fimT*$_{Lp}$, *with fimT R143Q, R146Q, R148Q mutations*	This study
pOPINS-*fimU*$_{Lp}$	*L. pneumophila fimU, residues 28-167*	This study
pOPINS-*fimT*$_{Pa}$	*Pseudomonas aeruginosa PAO1 fimT (PA4549),* residues 28-161	This study
pOPINS-*fimU*$_{Pa}$	*P. aeruginosa PAO1 fimU (PA4550),* residues 28-159	This study

pOPINS-*fimT*$_{Xc}$	*Xanthomonas campestris ATCC 33913 fimT (XCC2486),* residues 28-172	This study
pOPINS-*fimU*$_{Xc}$	*X. campestris ATCC 33913 fimU (XCC2495),* residues 28-163	This study
pOPINB	*E. coli* expression vector: N-terminal His_6-tag, T7 promoter; KanR	11
pOPINB-*pilA1*$_{Lp}$	*L. pneumophila pilA1 (lpg1914), residues 25-132*	This study
pOPINB-*pilA2*$_{Lp}$	*L. pneumophila pilA2,* residues 25-131	This study
pTRC99A	P*trc oriR*(pBR322); AmpR	12
pTRC99A-*lpg2953-2958::Kan*	*L. pneumophila* genomic region spanning *lpg2953-2958*. The *hipB* gene (*lpg2955*) is interrupted by kanamycin cassette, KanR	This study

1129 **Extended Data Table 4:** Oligonucleotides used in this study

Name	Sequence (5' to 3')	Construct
Cloning		
pMMB207_lin_F	aattcgagctcggtacccgg	pMMB207C
pMMB207_lin _R	ctgtttcctgtgtgaaattgttatccgc	
fimT$_{Lp}$_pMMB207C_F	tcacacaggaaacagatgcggcttcaattgatgaaaataacaggatttac	pMMB207C-*fimT$_{Lp}$*
fimT$_{Lp}$_pMMB207C_R	taccgagctcgaattttaattacccctaccctaaccctgcc	
fimT$_{Lp}$R143Q_ pMMB207C_F	ccctaaccctgccaagctgatttaaagtaaccacaac	pMMB207C-*fimT$_{Lp}$ R143Q*
fimT$_{Lp}$R143Q_ pMMB207C_R	ggttactttaaatcagcttggcagggttagggtag	
fimT$_{Lp}$R146Q_ pMMB207C_F	cttggccaggttagggtagggggtaattaaaattcg	pMMB207C-*fimT$_{Lp}$ R146Q*
fimT$_{Lp}$R146Q_ pMMB207C_R	taccctaacctggccaagcctatttaaagtaaccacaac	
fimT$_{Lp}$R148Q_ pMMB207C_F	cagggttcaggtaggggggtaattaaaattcgagctc	pMMB207C-*fimT$_{Lp}$ R148Q*
fimT$_{Lp}$R148Q_ pMMB207C_R	cccctacctgaaccctgccaagcctatttaaag	
fimT$_{Lp}$R146QR148Q_ pMMB207C_F	cttggccaggttcaggtaggggggtaattaaaattcg	pMMB207C-*fimT$_{Lp}$ R146Q, R148Q*
fimT$_{Lp}$R146QR148Q_ pMMB207C_R	cctacctgaacctggccaagcctatttaaagtaac	
fimT$_{Lp}$ R143QR146QR148Q_ pMMB207C_F	gccaggttcaggtaggggggtaattaaaattcgagctc	pMMB207C-*fimT$_{Lp}$ R143Q, R146Q, R148Q*
fimT$_{Lp}$ R143QR146QR148Q_ pMMB207C_R	ctacctgaacctggccaagctgatttaaagtaaccacaacttttttcattg	
pilA2$_{Lp}$-flag_ pMMB207C_F	tcacacaggaaacagatggagatggtcatgagacaaaagggttttac	pMMB207C-*pilA2-flag*
pilA2$_{Lp}$-flag_ pMMB207C_R	atcgtctttgtagtctggtctgcaactggcaggtc	
fimT$_{Pa}$ _pMMB207C_F	tcacacaggaaacagatggtcgaaaggtcgcagagagc	pMMB207C-*fimT$_{Pa}$*
fimT$_{Pa}$ _pMMB207C_R	taccgagctcgaatttcatccggaagtgctgcatagctc	
fimT$_{Pa}$_125_pMMB207C_F	ggtaaatttatttgtgcggaaggcataccgttgc	pMMB207C-*fimT*$_{chimera\ 1}$
fimT$_{Lp}$_138_pMMB207C_R	caaaataaatttaccattactcatcgcacgattcg	
fimT$_{Xc}$ _pMMB207C_F	tcacacaggaaacagatgcagacaggacctcagtcacc	pMMB207C-*fimT$_{Xc}$*
fimT$_{Xc}$ _pMMB207C_R	taccgagctcgaattttatgtctgcgcaggtgcc	
*fimT$_{Xc}$*_138_pMMB207C_F	ggtaaatttatttgtgcatccagtcagagcgagtgg	pMMB207C-*fimT*$_{chimera\ 2}$
fimT$_{Lp}$_138_pMMB207C_R	caaaataaatttaccattactcatcgcacgattcg	
pSR47S_lin_F	ggatcccccgggctgcaggaattcg	pSR47S
pSR47S_lin_R	ccactagttctagagcggccgcc	
fimT_HR_pSR47S_F	ggccgctctagaactagtggtggcaaatgggatttaggtctccctcaatg	pSR47S-*fimT*
fimT_HR_pSR47S_R	cctgcagcccgggggatccataaatgcctcagacaagctgacctctcc	
fimU_HR_pSR47S_F	ggccgctctagaactagtggccaacacatcactaccttgttgagcattgcc	pSR47S-*fimU*
fimU_HR_pSR47S_R	tcctgcagcccgggggatcccaatcactattgatgatttgccctttgttggtg	
pilQ_HR_pSR47S_F	ggccgctctagaactagtggttgaaaaaaagcaacatcaggcagc	pSR47S-*pilQ*
pilQ_HR_pSR47S_R	tcctgcagcccgggggatccatcgaaacatcaacctcggcataaag	
pilT_HR_pSR47S_F	ggccgctctagaactagtggtatcgtaatgagtgcgaatattttctttactaatgc	pSR47S-*pilT*

pilT_HR_pSR47S_R	tcctgcagcccgggggatccccgttacaataacacgtaattttaccaattatgc	
comEC_HR_pSR47S_F	ggccgctctagaactagtggggtttatccacaaacattatcactgccactg	pSR47S-*comEC*
comEC_HR_pSR47S_R	tcctgcagcccgggggatccactctgcttgaaaggtatcccagg	
ΔfimT_HR_pSR47S_F	tcttaaattataagcaatggttgttcataaagagg	pSR47S-Δ*fimT*
ΔfimT_HR_pSR47S_R	ccattgcttataatttaagacatctacaaaattttatgatgaagataagatgcg	
ΔfimU_HR_pSR47S_F	agcattatccctattgtttgatcgaacccac	pSR47S-Δ*fimU*
ΔfimU_HR_pSR47S_R	caaacaatagggataatgctaacaacacccggccaagcagtc	
ΔpilQ_HR_pSR47S_F	tcaagattggactaattttatctcattaataaagataaaaaacattaatttaatagc	pSR47S-Δ*pilQ*
ΔpilQ_HR_pSR47S_R	ttagtccaatcttgagcctcactcctgc	
ΔpilT_HR_pSR47S_F	atacacatgacttgtgaaaaagacccaaggtc	pSR47S-Δ*pilT*
ΔpilT_HR_pSR47S_R	acaagtcatgtgtatactctataattcccgcc	
ΔcomEC_HR_pSR47S_F	atggattggctgacccatgttatatctaagc	pSR47S-Δ*comEC*
ΔcomEC_HR_pSR47S_R	ggtcagccaatccatttcaaattaagttggactttcc	
pOPINS_lin_F	taaagctttctagaccatttaaacaccaccac	pOPINS
pOPINS_lin_R	accaccgatctgttcgcgat	
$fimT_{Lp}$_28_pOPINS_F	atcgcgaacagatcggtggtatacaaaataatgagagagaaacattagttaatagtataaaaacagccattc	pOPINS-*$fimT_{Lp}$*
$fimT_{Lp}$_152_pOPINS_R	aaatggtctagaaagctttattaattacccccctaccctaaccctgcc	
$fimT_{Lp}$K103Q_pOPINS_F	tggaatattaattggcagggcgtagattcaaaccatag	pOPINS-*$fimT_{Lp}$ K103Q*
$fimT_{Lp}$K103Q_pOPINS_R	tacgccctgccaattaatattccaggaattagaactcc	
$fimT_{Lp}$R119Q_pOPINS_F	ccaatattccgaatcaggcgatgagtaatggtaaatttattttg	pOPINS-*$fimT_{Lp}$ R119Q*
$fimT_{Lp}$R119Q_pOPINS_R	catcgcctgattcggaatattggatataataattctatggtttgaatc	
$fimT_{Lp}$R143Q_pOPINS_F	ggttactttaaatcagcttggcagggttagggtag	pOPINS-*$fimT_{Lp}$ R143Q*
$fimT_{Lp}$R143Q_pOPINS_R	ccctaaccctgccaagctgatttaaagtaaccacaac	
$fimT_{Lp}$R146Q_pOPINS_F	gcttggccaggttagggtagggggtaattaataaag	pOPINS-*$fimT_{Lp}$ R146Q*
$fimT_{Lp}$R146Q_pOPINS_R	cctaacctggccaagcctatttaaagtaaccacaac	
$fimT_{Lp}$R148Q_pOPINS_F	cagggttcaggtagggggtaattaataaagctttctagac	pOPINS-*$fimT_{Lp}$ R148Q*
$fimT_{Lp}$R148Q_pOPINS_R	cccctacctgaaccctgccaagcctatttaaagtaac	
$fimT_{Lp}$R146QR148Q_pOPINS_F	cttggccaggttcaggtagggggtaattaataaagc	pOPINS-*$fimT_{Lp}$ R146Q, R148Q*
$fimT_{Lp}$R146QR148Q_pOPINS_R	ccctacctgaacctggccaagcctatttaaagtaac	
$fimT_{Lp}$ R143QR146QR148Q_pOPINS_F	ggccaggttcaggtagggggtaattaataaagctttctag	pOPINS-*$fimT_{Lp}$ R143Q, R146Q, R148Q*
$fimT_{Lp}$ R143QR146QR148Q_pOPINS_R	tacctgaacctggccaagctgatttaaagtaaccac	
$fimU_{Lp}$_28_pOPINS_F	atcgcgaacagatcggtggtattttgaatagccgtttgacttcaaacattgac	pOPINS-*$fimU_{Lp}$*
$fimU_{Lp}$_167_pOPINS_R	atggtctagaaagctttattaagggcagttcaaagctccattattcc	
$fimT_{Pa}$_28_pOPINS_F	atcgcgaacagatcggtggtctggacggcaatcgcgagc	pOPINS-*$fimT_{Pa}$*
$fimT_{Pa}$_161_pOPINS_R	aaatggtctagaaagctttatcatccggaagtgctgcatagctc	
$fimU_{Pa}$_28_pOPINS_F	atcgcgaacagatcggtggtctgacagaacgcaacgaactgcag	pOPINS-*$fimU_{Pa}$*
$fimU_{Pa}$_159_pOPINS_R	aaatggtctagaaagctttatcaatagcatgactggggcgc	
$fimT_{Xc}$_28_pOPINS_F	atcgcgaacagatcggtggtatcgagcggcagcggttg	pOPINS-*$fimT_{Xc}$*
$fimT_{Xc}$_172_pOPINS_R	aaatggtctagaaagctttattatgtctgcgcaggtgccgg	
$fimU_{Xc}$_28_pOPINS_F	atcgcgaacagatcggtggtattcggtcgaatcgcgctgttac	pOPINS-*$fimU_{Xc}$*
$fimU_{Xc}$_163_pOPINS_R	aaatggtctagaaagctttatcattgacagttatcctttctacttctgacttgc	
pOPINB_lin_F	agcagcggtctggaagttctgtttcag	pOPINB
pOPINB_lin_R	atggtctagaaagcttta	
$pilA1_{Lp}$_25_pOPINB_F	aagttctgtttcagggcccgcaggactataccatcagagcac	pOPINB-*$pilA1_{Lp}$*
$pilA1_{Lp}$_132_pOPINB_R	atggtctagaaagctttattaagggcggcagtagg	

pilA2$_{Lp}$_28_pOPINB_F	aagttctgtttcagggcccgcaagattacacaatacgagctcg	pOPINB-*pilA2$_{Lp}$*
pilA2$_{Lp}$_131_pOPINB_R	atggtctagaaagctttattatggtctgcaactggcag	
pTRC99A_lin_F	gtgtctagagtcgacctgcaggcat	pTRC99A
pTRC99A_lin_R	gaacacaccagagatatctggcagaattc	
Lpg2953_F	atctctggtgtgttcggatagattatgcgagaggtctatttgaagattctctg actatg	pTRC99A-*lpg2953-2958::Kan* Amplification of transforming DNA
Lpg2958_R	gtcgactctagacacagacatggcctggaaacgttggtggg	
KanR_lin_F	cattcaaatatgtatccgctcatga	pTRC99A-*lpg2953-2958::Kan*
KanR_lin_R	cggggtctgacgctcagt	
pTRC99A_lpg2953_F	atacatatttgaatgcacgaatttctattctttggcc	pTRC99A-*lpg2953-2958::Kan*
pTRC99A_lpg2958_R	gagcgtcagaccccggctttggcagtttttctcttca	
*DNA-binding assays**		
FAM-12mer	# gttcgcaacgaa	MST/TRIC
12mer	gttcgcaacgaa	NMR titrations/ITC
FAM-30mer	# ttaaataggcttggcagggttagggtaggg	EMSA
30mer	ttaaataggcttggcagggttagggtaggg	EMSA

* The complementary strand for dsDNA probes is not shown. Only one of the two strands is fluorescein (FAM)-labelled.

# Indicates the position of the FAM label.

Extended Data Table 5: Gene locus tags of *fimT* and *fimU* genes from previous and recently updated genomes

	L. pneumophila* Philadelphia 1 (old)	***L. pneumophila* Philadelphia 1 (new)**	***X. campestris* ATCC 33913 (old)***	***X. campestris* ATCC 33913 (new)**	***A. baylyi* ADP1 (old)**	***A. baylyi ADP1 (new)***
RefSeq	NC_002942.5	NC_002942	NC_003902.1	NC_003902	NC_005966.1	NC_005966
Release date	2014	2021	2014	2021	2015	2020
fimT	*lpg1428*	LPG_RS07155	XCC2486	XCC_RS12930	ACIAD0695	ACIAD_RS03200
fimU	*lpg0632*	LPG_RS03130	XCC2495	XCC_RS12975	ACIAD3321	ACIAD_RS15030

* In this study we have referred to the old locus tags throughout.

Extended Data Table 6: Gene locus tags of selected genes from this study from previous and recently updated genomes

	L. pneumophila* Philadelphia 1 (old)	***L. pneumophila* Philadelphia 1 (new)**
RefSeq	NC_002942.5	NC_002942
Release date	2014	2021
pilQ	lpg0931	LPG_RS04620
pilT	lpg2013	LPG_RS10105
comEC	lpg0626	LPG_RS03100
pilA1	lpg1914	LPG_RS09600
pilA2	lpg1915	LPG_RS09605
hipB	lpg2955	LPG_RS14950
pilV	lpg0631	LPG_RS03125
pilW	lpg0630	LPG_RS03120
pilX	lpg0629	LPG_RS03115
pilY1	lpg0628	LPG_RS03110
pilE	lpg0627	LPG_RS03105

* In this study we have referred to the old locus tags throughout.

References

1. Yanez, M. E., Korotkov, K. K., Abendroth, J. & Hol, W. G. J. Structure of the Minor Pseudopilin EpsH from the Type 2 Secretion System of *Vibrio cholerae*. *Journal of Molecular Biology* **377**, 91–103 (2008).

2. Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* **33**, W299–W302 (2005).

3. Berry, J.-L. *et al.* A Comparative Structure/Function Analysis of Two Type IV Pilin DNA Receptors Defines a Novel Mode of DNA Binding. *Structure* **24**, 926–934 (2016).

4. Zuckman, D. M., Hung, J. B. & Roy, C. R. Pore-forming activity is not sufficient for *Legionella pneumophila* phagosome trafficking and intracellular growth. *Molecular microbiology* **32**, 990–1001 (1999).

5. Kolter, R., M, I. & R, H. D. Trans-Complementation-Dependent Replication of a Low Molecular Weight Origin Fragment from Plasmid R6K. *Cell* **15**, 1199–1208 (1978).

6. Finan, T. M., Kunkel, B., Vos, G. F. D. & Signer, E. R. Second Symbiotic Megaplasmid in *Rhizobium meliloti* Carrying Exopolysaccharide and Thiamine Synthesis Genes. *Journal of bacteriology* **167**, 66–72 (1986).

7. Berger, K. H. & Isberg, R. R. Two distinct defects in intracellular growth complemented by a single genetic locus in Legionella pneumophila. *Molecular microbiology* **7**, 7–19 (1993).

8. Chen, J. *et al. Legionella* Effectors That Promote Nonlytic Release from Protozoa. *Science* **303**, 1358–1361 (2004).

9. Merriam, J. J., Mathur, R., Maxfield-Boumil, R. & Isberg, R. R. Analysis of the *Legionella pneumophila* fliI Gene: Intracellular Growth of a Defined Mutant Defective for Flagellum Biosyntheis. *Infection and immunity* **65**, 2497–2501 (1997).

10. Assenberg, R. *et al.* Expression, purification and crystallization of a lyssavirus matrix (M) protein. *Acta Crystallogr Sect F Struct Biology Cryst Commun* **64**, 258–262 (2008).

11. Berrow, N. S. *et al.* A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Research* **35**, e45–e45 (2007).

12. Amann, E., Ochs, B. & Abel, K.-J. Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene* **69**, 301–315 (1988).