# Structure and Properties of a Natural Competence-Associated Pilin Suggest a Unique Pilus Tip-Associated DNA Receptor

Mohd Zulkifli Salleh,[a] Vijaykumar Karuppiah,[a] Matthew Snee,[a] Angela Thistlethwaite,[a] Colin W. Levy,[b] David Knight,[a] Jeremy P. Derrick[a]

[a]Lydia Becker Institute of Immunology and Inflammation, School of Biological Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, The University of Manchester, Manchester, United Kingdom
[b]Manchester Institute of Biotechnology, The University of Manchester, Manchester, United Kingdom

**ABSTRACT**   Natural competence is the term used to describe the uptake of "naked" extracellular DNA by bacteria; it plays a significant role in horizontal genetic exchange. It is associated with type IV pili, and specialized competence pili mediate DNA uptake. Here, we show that the crystal structure of a competence-associated protein from *Thermus thermophilus*, ComZ, consists of a type II secretion pseudopilin-like domain, with a large $\beta$-solenoid domain inserted into the $\beta$-sheet of the pilin-like fold. ComZ binds with high affinity to another competence-associated pilin, PilA2, which lies adjacent to the *comZ* gene in the genome. The crystal structure of PilA2 revealed a similar type II secretion pseudopilin-like fold, with a small subdomain; docking simulations predicted that PilA2 binds between the pseudopilin-like and $\beta$-solenoid domains of ComZ. Electrophoretic shift analysis and DNase protection studies were used to show that ComZ alone and the ComZ/PilA2 complex are able to bind DNA. Protection against reductive dimethylation was used in combination with mass spectrometry and site-directed mutagenesis to identify two lysine residues in ComZ which are involved in DNA binding. They are located between the two domains in ComZ, on the opposite side from the predicted PilA2 binding site. These results suggest a model in which PilA2 assists ComZ in forming the competence pilus tip and DNA binds to the side of the fiber. The results demonstrate how a type IV pilin can be adapted to a specific function by domain insertion and provide the first structural insights into a tip-located competence pilin.

**IMPORTANCE**   *Thermus thermophilus* is a thermophilic bacterium which is capable of natural transformation, the uptake of external DNA with high efficiency. DNA uptake is thought to be mediated by a competence-associated pilus, which binds the DNA substrate and mediates its transfer across the outer membrane and periplasm. Here, we describe the structural and functional analysis of two pilins which are known to be essential for DNA uptake, ComZ and PilA2. ComZ adopts an unusual structure, incorporating a large $\beta$-solenoid domain into the pilin structural framework. We argue on structural grounds that this structure cannot readily be accommodated into the competence pilus fiber unless it is at the tip. We also show that ComZ binds DNA and identify two lysine residues which appear to be important for DNA binding. These results suggest a model in which ComZ and PilA2 form a tip-associated DNA receptor which mediates DNA uptake.

**KEYWORDS**   X-ray crystallography, genetic competence, natural transformation systems, pilus assembly, surface receptor

Horizontal gene transfer (HGT) is a process whereby genetic material is transferred between bacterial cells. HGT is mediated by three different mechanisms: conjugation, transduction, and natural transformation (NT); it therefore plays an important

role in bacterial adaptive evolution. NT is characterized by the ability of bacteria to take up "naked" DNA from outside the cell: one strand of the DNA molecule is degraded, and the other is transported into the cytoplasm (1–3). DNA uptake and translocation into the cytoplasm is an active process, and DNA taken up in this way can be used as food for genome repair or to generate genetic diversity. The latter property has implicated NT in mediating antimicrobial resistance in several pathogenic bacterial species, such as respiratory tract flora *Haemophilus influenzae*, *Streptococcus pneumoniae*, and *Neisseria meningitidis*. Natural competence is also found in thermophilic, as well as mesophilic, bacteria, notably *Thermus thermophilus*, where it could have a role in assisting environmental adaptation (4).

Natural competence in *T. thermophilus* is mediated by a multiprotein complex, consisting of 16 proteins which span the outer membrane, the periplasm, and the inner membrane (4, 5). Competence genes in *T. thermophilus* are upregulated under stress conditions such as starvation, nutrient deprivation, overpopulation, and even low temperature (6). *H. influenzae* and *Neisseria* species exhibit a strong preference for extracellular DNA fragments that contain specific uptake sequences (7–9). *T. thermophilus*, in contrast, can take up DNA without apparent sequence specificity; it is therefore a highly transformable and rapidly adaptable bacterial species, growing in hostile environments, with temperatures ranging from 50 to 82°C and pH values ranging from 6 to 9 (4, 10, 11). DNA uptake in *T. thermophilus* is a fast and efficient process; estimated uptake velocities of up to 40 kb s$^{-1}$ have been reported (11).

NT in Gram-negative bacteria is connected to the biogenesis of type IVa (T4a) pili (3). T4a pili are long, thin fibers which protrude from the bacterial surface; they are responsible for mediating host cell adhesion and a type of bacterial movement called twitching motility, as well as the uptake of DNA (12). Pilin subunits are exported into the periplasm, cleaved by a dedicated signal peptidase, and assembled into a pilus fiber by a complex of proteins in the cytoplasm and inner membrane. In Gram-negative organisms, this assembly platform comprises three transmembrane proteins, PilN, PilO, and PilC, a soluble protein, PilM, which binds to PilN, and a dedicated ATPase (PilF in *T. thermophilus*), which catalyzes pilus fiber assembly (6, 13–19). As well as being responsible for pilus formation, these proteins are all required for NT in *T. thermophilus* (4, 17). In addition, specialized competence-specific proteins are required; recent work in *T. thermophilus* has shown that a competence-associated protein, ComEA, is associated with the inner membrane and is responsible for binding to DNA (6). Another conserved protein, ComEC, forms a polytopic transmembrane protein and the channel for DNA passage across the inner membrane. Both proteins are highly conserved, with orthologs in Gram-positive as well as Gram-negative bacteria (1, 20).

The process of DNA uptake from the extracellular milieu and subsequent transport into the cytoplasm is divided into several discrete steps, although the details of each stage remain unclear (1, 3, 21). The initial step requires binding of DNA outside the cell, probably to a DNA-specific receptor associated with a type IV pilus. In Gram-negative bacteria, transport needs to be negotiated across the outer membrane; PilQ, a member of the secretin family, is an integral outer membrane protein which provides a channel for passage of type IV pili (22–24). The observation that PilQ is capable of binding DNA suggests a similar role for DNA transport, although details of how this occurs are obscure (25). Once inside the periplasm, the incoming DNA is bound by ComEA and one strand is degraded by an endonuclease. Transport across the inner membrane is mediated by ComEC; the DNA strand is subsequently bound by DprA (26), which recruits RecA into a complex to promote homologous recombination. Many details of this basic model remain unclear, however, notably the identity of the protein initially responsible for DNA binding and the precise mechanism by which uptake is powered. Curiously, mutations to the two retraction-specific PilT ATPases in *T. thermophilus* do not impair natural competence (19), although other ATPases could be involved in retraction of the competence pilus.

Type IV pili are made up of polymers of noncovalently linked pilin proteins; type IV pilins adopt a canonical structure, consisting of an N-terminal hydrophobic α-helix

packed against an antiparallel β-sheet (12, 27). Several structural variations on this theme have been described, including addition of entire domains (28, 29). Pili are mainly composed of a single pilin, but minor pilins, present in much lower quantities, can play a crucial role in pilus formation. For example, the minor pilins *pilHIJK* in *N. meningitidis* play a role in pilus assembly (30), possibly through modulation of pilus surface density (31, 32). Similar observations have been made in *Pseudomonas*, and minor pilins are known to adopt the type IV pilin canonical fold, with noted similarities to the type II secretion system pseudopilins (33–35). A specialized minor pilin, ComP, from *N. meningitidis* is required for DNA recognition in natural transformation but is dispensable for T4a pilin formation and other functions (36). ComP has a structure similar to that of other T4a pilins and exhibits a DNA uptake sequence (DUS) binding specificity consistent with it acting as the primary DNA receptor (37). An electropositive strip of surface residues forms a specific docking surface for the DNA ligand, and selectivity toward different DUS variants extends to ComP homologs from other *Neisseria* species (38). However, ComP is apparently confined to *Neisseria* and remains the sole example to date of a T4a pilin which specifically binds to DNA. Moreover, Hepp and Maier report that an *N. gonorrhoeae comP* mutant is not impaired in DNA uptake, suggesting that other components are involved in DNA recognition (39). The question of the identity of the primary DNA recognition pilin therefore remains open in most naturally competent Gram-negative bacteria.

Here, we describe a structural and functional analysis of ComZ, a pilin-like protein which is part of a locus in *T. thermophilus* HB27 and is known to be essential for NT but not piliation (40). We show that the ComZ structure incorporates both pilin-like and β-solenoid domains, binds specifically to an adjacent minor pilin in this specialized locus, and is able to recognize DNA. Based on these observations, we propose that ComZ functions as a tip pilin and receptor for the initial DNA binding step outside the bacterial cell.

## RESULTS

**ComZ adopts an unusual type IV pilin-like structure.** Previous work on *T. thermophilus* HB27 by Friedrich et al. identified a locus containing five open reading frames (ORFs) associated with natural competence (40). Four ORFs (*pilA1* to *pilA4* [*pilA1-4*]) were type IVa pilin-like genes, and a fifth, *comZ*, encoded an unusual and much larger protein that lacked some of the sequence features characteristic of pilin genes (12, 35). Mutation of *pilA1-3* and *comZ* leads to a loss of natural competence, although piliation is maintained (40). We set out to express and purify PilA1-3 and ComZ, in each case omitting the first 30 to 33 residues. This is a common strategy used for type IV pilin expression, because it removes the hydrophobic portion of the conserved N-terminal α-helix. ComZ was crystallized and the structure was determined to 2.72-Å resolution (see Table S1 in the supplemental material). The structure had three ComZ molecules in the asymmetric unit, but analysis by the protein interface analysis server PISA (41) did not indicate any significant crystal contacts. Size exclusion chromatography (SEC) during purification was consistent with the recombinant ComZ fragment forming a monomer in solution.

Figure 1 shows the overall structure of ComZ: it forms two clearly identifiable domains. ComZ was not identified as a type IV pilin by PilFind (42), presumably because it does not contain key characteristic sequence motifs; however, the N-terminal domain adopts a type IV pilin-like fold, suggesting that it forms a structural component of a pilus fiber. The second domain forms a complex β-solenoid structure, which is inserted between the penultimate and last β-strands in the pilin fold through two short linkers.

The pilin fold is formed from five antiparallel β-strands packed against a single α-helix, with an extensive loop region between the end of the α-helix (Asn39) and the beginning of the first β-strand (Asn71). From a comparison of the three chains, the relative orientation of the two domains is invariant (Fig. S1). The segmental mobility between the two domains is likely to be constrained by contacts formed from the extensive loop regions, which emanate from the β-solenoid and pilin domains. Most
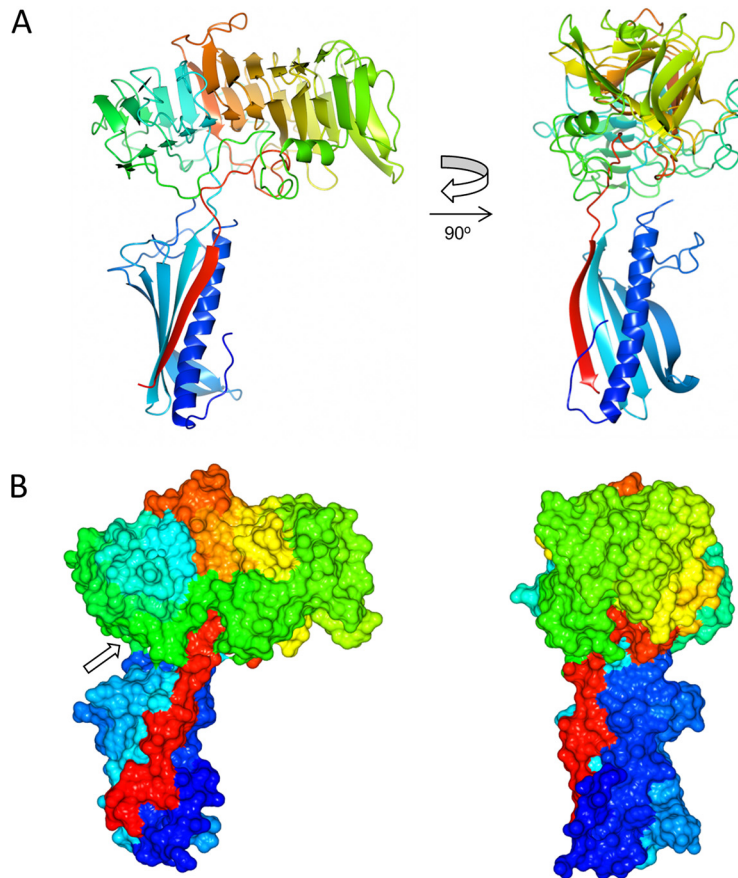
**FIG 1** Structure of *T. thermophilus* ComZ. (A) Two orthogonal ribbon plot views of ComZ, colored on a gradient from the N (blue) to the C (red) terminus. (B) Two surface plots of ComZ, with the same orientations and color scheme as panel A. The arrow indicates the long loop which connects the two ends of the β-helix.

notably, an extensive loop and short α-helix from Glu216 to Ala301 connects the two ends of the β-helix, passing between the two domains and making contact with the second of the two peptide linkers which join both domains (Fig. 1B, arrow). The loop packs against the concave face of the β-solenoid and buries hydrophobic sidechains, contributing to the rigidity of the structure. This lack of flexibility of the β-solenoid domain could play a part in the proposed role of ComZ as a tip receptor for DNA binding (discussed further below).

**Comparison of ComZ with related structures.** The Protein Data Bank now contains an extensive library of type IV pilin and pseudopilin structures (12); the ComZ pilin-like domain was therefore searched for structural homologs using the fold recognition program DALI (43). The closest match was the type II secretion system pseudopilin GspK, which forms the tip of a pseudopilus heterotrimeric complex (GspK/GspI/GspJ) (29). The fold topology of GspK has striking similarities with the ComZ pilin-like domain (Fig. 2A); the principal difference lies in the insertion point of the additional domain in each structure. In the case of ComZ, the insertion lies between the penultimate and last β-strands, whereas in GspK, insertion of an α-helical domain is between the third and fourth β-strands (Fig. 2B, topology diagrams). Our previous work has highlighted the similarity of *T. thermophilus* minor pilins to type II secretion system pseudopilins (35), but the structure of ComZ also shows how the canonical pilin domain can be adapted by various domain insertions, possibly to add additional functionality. The structure of the minor type IVb pilin CofB from enterotoxigenic *Escherichia coli* (ETEC) has parallels with ComZ, in that both are structurally related to GspK and both have additional β-rich domains outside the pilin fold (28). The CofB structure consists
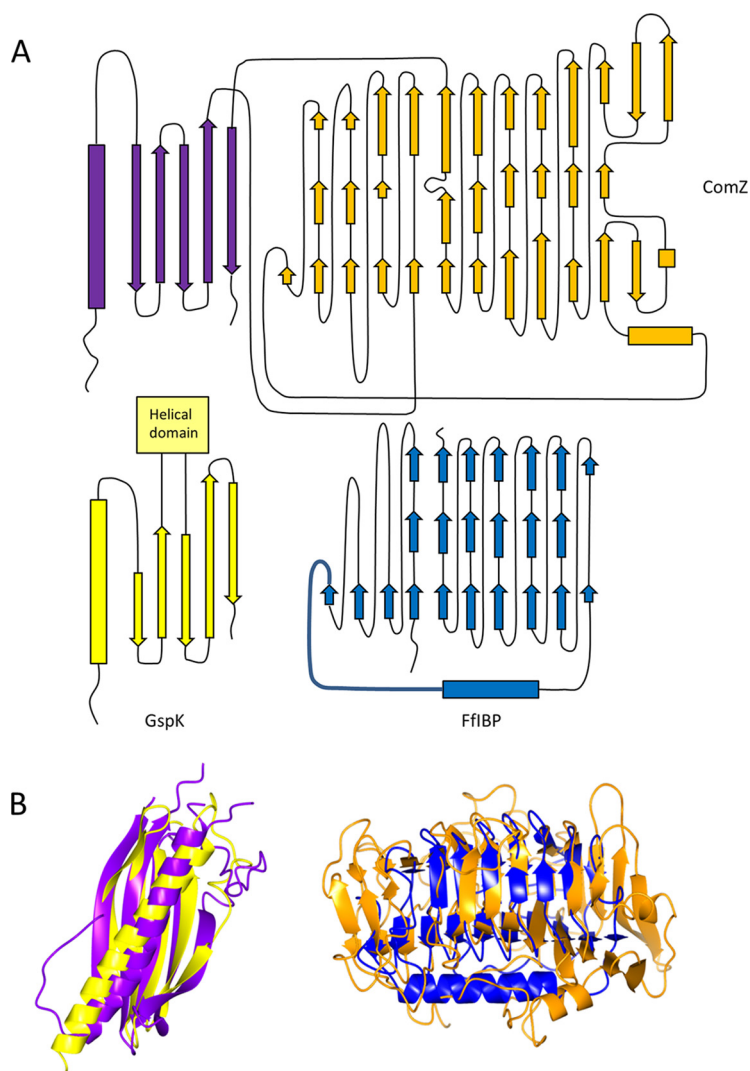
**FIG 2** Topology and comparison of ComZ domains with structurally related proteins. (A) Topology of ComZ and related folds. (Upper) Topology diagram of ComZ, with the pilin-like domain in purple and the β-solenoid domain in orange. (Lower) Folds of the related proteins GspK and FfIBP are aligned below their mapped superpositions to the equivalent ComZ domains. (B) Superpositions of related structural folds with their equivalent domains from ComZ. (Left) GspK from *E. coli* (PDB entry 3CIO; pilin domain only) is shown in yellow, and the ComZ pilin-like domain is in purple. (Right) The ice-binding protein FfIBP (4NU2) from *Flavobacterium frigoris* is in dark blue, and the β-solenoid domain is in orange. Superpositions were carried out using SSM matching, as implemented in CCP4MG (72); root mean square deviation values were 3.25 Å (GspK) and 2.33 Å (FfIBP). Colors are the same as those for panel A.

of an N-terminal pilin domain followed by a short β-repeat section and a β-sandwich domain at the C terminus. Interestingly, the C-terminal regions of CofB are thought to be necessary to initiate type IVb pilus assembly (28).

The β-solenoid domain comprises ~380 residues and constitutes the majority of the ComZ structure. It is formed from parallel β-strands arranged in a triangular β-helix, stabilized by intramolecular main-chain hydrogen bonds and a hydrophobic core characteristic of β-solenoid structures. The domain forms an elongated structure, about 60 Å in length, with a triangular cross-section (Fig. 3). There are several irregularities in the fold, notably at the end between residues 301 and 362, where the β-solenoid breaks down and is replaced by two pairs of antiparallel β-strands (Fig. 2B, far right of the topology diagram). Viewed from the side, it is apparent that the β-solenoid has a pronounced bend (Fig. 3). We labeled the faces of each side of the solenoid A, B, and C by analogy with the fold of the ice-binding protein (discussed below). Several long
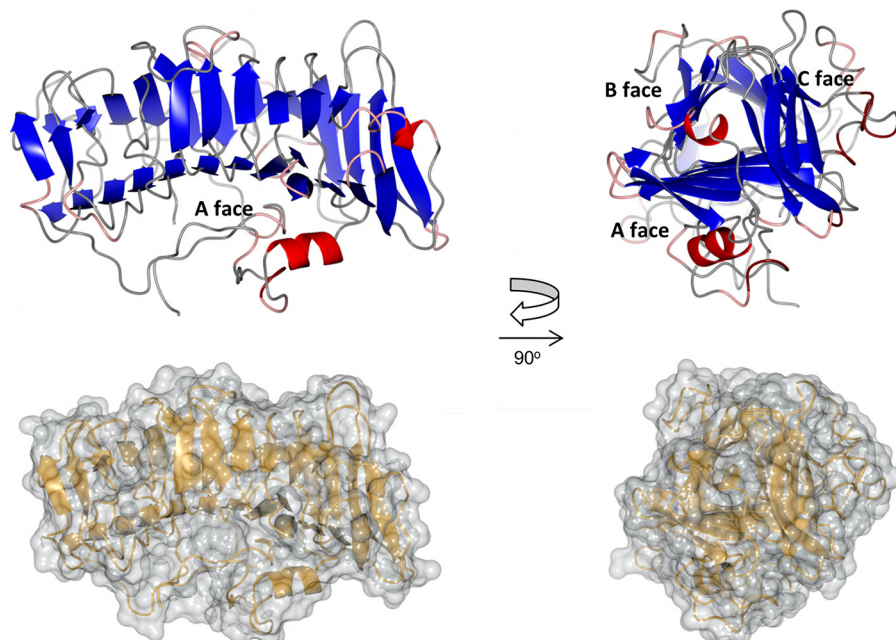
**FIG 3** Detail of the β-solenoid domain from ComZ. (Upper) Two orthogonal ribbon plot views of the β-solenoid domain from ComZ, colored by secondary structure: β-strand (blue), α-helix (red), turn (pink), and coil (gray). Faces of the solenoid are labeled on the right. (Lower) Surface models, in the same orientations as that of the upper panel, superimposed on ribbon plot.

loops emanate from the β-strands and pack against the faces of the β-solenoid; this phenomenon is most pronounced for the A face, where a network of loops and a short α-helix pack against the parallel β-sheet (Fig. 3). Another long loop from Asn182 to Ala208 packs against the C face and a shorter loop, from Thr476 to Lys489, against the B face.

The size of the β-solenoid domain, combined with the rigid orientation of the two domains discussed above, led us to examine whether it could be incorporated into a model for the assembled T4a pilus fiber. Using the structure of the pilus from *N. meningitidis* (44), we superimposed the ComZ pilin domain: the β-solenoid domain overlaps well with the fourth and fifth pilin subunits in the fiber (Fig. S2). Even with minor adjustment of the relative orientation of the β-solenoid and pilin domains in ComZ, it is unlikely that it could be stably incorporated into the fiber. In fact, the β-solenoid domain is placed in line with the central axis of the fiber, placing it in an ideal orientation to act as a tip adhesin. We acknowledge that other models for T4a pilus fiber differ in their details, but the central packing of N-terminal hydrophobic helices is a common feature; our conclusion is therefore unlikely to be altered substantially by these differences (12).

A fold recognition search using DALI (43) identified an ice-binding protein, FfIBP from *Flavobacterium frigoris* (PDB accession code 4NU2), as the closest structural relative of the ComZ β-solenoid domain (45). The topology of the core fold of the two proteins is similar, covering 10 turns of the solenoid (Fig. 2B). However, there are several key differences between the two structures. The ComZ solenoid is longer, more irregular, and with many more extensive loops. FfIBP has an α-helix which packs against the A face, performing a structural role similar to that of the loops in the ComZ structure (Fig. 2A, right). FfIBP is a member of a class of ice-binding proteins, which are characterized by their β-solenoid folds (46). The ice-binding residues in FfIBP have been mapped to face B (45), but we found no obvious sequence conservation in the equivalent positions for ComZ. Given that the structural conservation of ComZ with this class of ice-binding proteins did not seem to provide any insight into its function, we sought to investigate the properties of ComZ in other ways.

**Structure of the competence-associated type IV pilin PilA2.** Three competence-associated pilin genes, *pilA1-pilA2-pilA3*, are found adjacent to *comZ* in the *T. thermophilus* strain HB27 genome (11). Unlike ComZ, each ORF was positively identified as a type IV pilin by PilFind (42). Recently we described a strategy for systematic expression and characterization of pilins from *T. thermophilus* HB8, which involved removal of the signal sequence and part of the N-terminal helix, and incorporation of a purification tag (35). We applied the same strategy to PilA1, PilA2, and PilA3 from the HB27 strain and purified each truncated pilin to homogeneity. PilA2 gave crystals with good diffraction qualities, and the structure was determined to 1.39-Å resolution; it forms a distinctly identifiable domain with a small subdomain insertion (Table S2 and Fig. S3A). PilA2 adopts a type IV pilin fold, a single $\alpha$-helix packed against four antiparallel $\beta$-strands, with a small subdomain inserted between the end of the first $\alpha$-helix and the beginning of the fourth $\beta$-strand, consisting of three short antiparallel $\beta$-strands and a short $\alpha$-helix, linked by an extensive loop region. A search with DALI (47) identified close structural homology with the type II secretion system pseudopilins EpsI from *Vibrio vulnificus* (48) and GspI from *E. coli* (29), as well as the minor type IV pilin TTHA1218 from *T. thermophilus* HB8 (35). Although all three related pilin structures share the same fold, the subdomain is consistently absent (Fig. S3B).

**ComZ binds specifically to the competence-associated pilin PilA2.** Given the structural similarity of the ComZ pilin domain to GspK, we reasoned that ComZ should bind to at least one other pilin, if it is indeed incorporated into a pilus structure. Size exclusion chromatography (SEC) profiles for ComZ and PilA2, ComZ alone, and PilA2 alone are shown in Fig. 4A. The ComZ/PilA2 complex eluted in a single peak (~80 kDa), at a higher apparent mass than ComZ alone (~60 kDa); the complex peak contained both proteins (Fig. 4A, lower). Specific binding of the two proteins was confirmed by isothermal titration calorimetry (ITC), which gave a stoichiometry of 1.1 and an equilibrium binding constant ($K_d$) of 0.69 $\mu$M (Fig. 4B).

Similar studies of binding to ComZ were carried out using the PilA1 and PilA3 pilins. PilA1 failed to show any evidence of binding to ComZ (Fig. S4A). The elution profile of PilA3 from the SEC column suggested the formation of oligomers, so it was not possible to use this method to determine binding to ComZ (Fig. S4B). Experiments were also conducted where ComZ was preincubated with PilA2 and either PilA1 or PilA3, before separation by SEC, to examine whether PilA1 or PilA3 affects the binding of PilA2 to ComZ indirectly. The results suggested that this was not the case, however, and that the presence of PilA1 or PilA3 had no discernible effect on the elution profile of the ComZ/PilA2 complex (Fig. S5A and B). We sought to confirm and extend these observations using an affinity tag assay. ComZ was incubated with PilA2 or PilA3, each of which contains a Strep-tag; the mixture was then passed through a Strep affinity column, unbound protein was eluted, the column was washed, and eluted protein was analyzed by SDS-PAGE. ComZ coeluted with PilA2 (Fig. S6A) but not with PilA3 (Fig. S6B). The reverse experiment, which used a Ni affinity column to bind ComZ, showed that PilA2, but not PilA3, coeluted with ComZ (Fig. S6C). We conclude that ComZ selectively binds to PilA2, but not PilA1 or PilA3.

**Characterization of ComZ binding to dsDNA.** It was previously reported that *comZ, pilA1, pilA2,* and *pilA3 Thermus* mutants are impaired in natural competence but retain type IV pili (3). Given that the structure of ComZ suggested it functions as a tip pilin, we examined whether it was able to bind to DNA. Studies by electrophoretic mobility shift assay (EMSA) showed that increasing quantities of ComZ reduce double-stranded DNA (dsDNA) migration (Fig. 5A). DNA binding proteins which recognize specific sites can produce smears on EMSA when tested against nonspecific DNA. In the case of ComZ, we observe discrete DNA bands. We suggest that this is explained by binding of multiple molecules of ComZ, gradually reducing the mobility of each DNA strand. This explanation is consistent with the fact that *T. thermophilus* does not exhibit any DNA sequence preference (4), unlike *Neisseria* spp. (37), for example. We sought to verify this conclusion by conducting a DNase protection experiment: increasing quan-
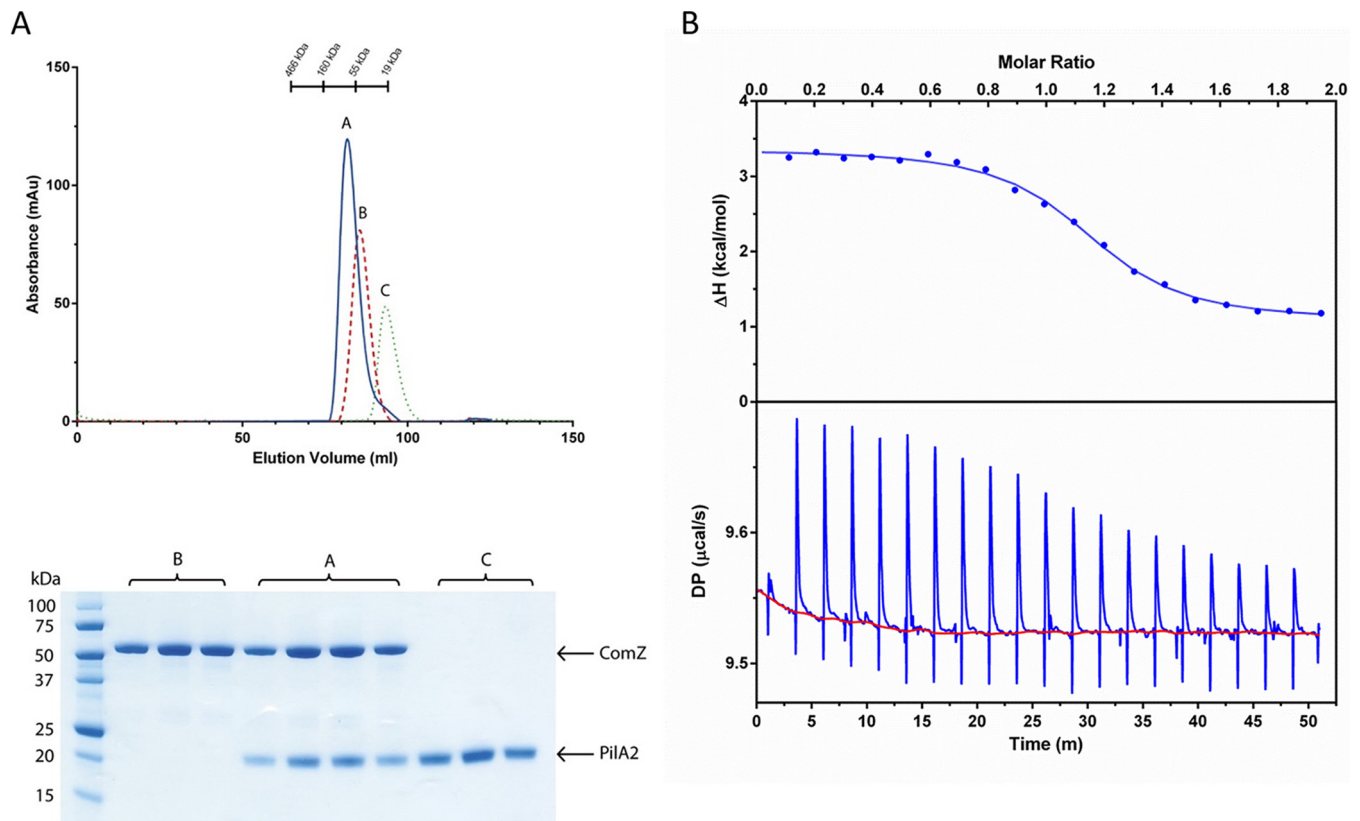
**FIG 4** Binding of PilA2 to ComZ. (A) Size exclusion chromatography analysis. (Upper) Elution profiles (absorption at 280 nm) of ComZ-PilA2 mixture (profile A), ComZ alone (profile B), and PilA2 alone (profile C). Separation was carried out on a HiLoad 16/600 Superdex 200 PG column (GE Healthcare), with a flow rate of 1 ml/min in 25 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5% glycerol. ComZ (0.036 $\mu$mol) and PilA2 (0.023 $\mu$mol) were loaded. (Lower) SDS-PAGE of eluted peaks from each SEC run. (B) Titration of PilA2 into ComZ using isothermal titration calorimetry (ITC). Two-$\mu$l aliquots of recombinant PilA2 (314 $\mu$M) were titrated into the sample cell containing 300 $\mu$l of ComZ (33 $\mu$M). (Lower) Raw data with fitted baseline. (Upper) Data fitted assuming a single binding site. Fitted parameters are the following: stoichiometry ($n$), 1.1; equilibrium binding constant ($K_d$), 0.69 $\mu$M; enthalpy change ($\Delta H$), 2.3 kcal/mol; $\Delta G = -8.4$ kcal/mol; $T\Delta S = 10.7$ kcal/mol.

tities of DNase I were added to ComZ and dsDNA (Fig. 5B). ComZ was able to inhibit nonspecific hydrolysis of DNA by DNase, effectively providing protection at concentrations where substantial degradation occurred. The results are consistent with multiple copies of ComZ binding to each dsDNA duplex in a non-sequence-dependent manner. In the presence of PilA2, the ComZ/PilA2 complex reduced DNA mobility to a greater extent than ComZ alone, indicative of the higher mass of the ComZ/PilA2 complex (Fig. 5C). Virtually no reduction in DNA mobility was recorded in the presence of PilA2 alone (Fig. 5D).

In order to map the DNA binding site on ComZ, we developed a method based on the protection of lysine residues from reductive dimethylation by bound DNA. We reasoned that lysines were likely to be involved in DNA recognition, and inspection showed that they cover both ComZ domains. Protocols for reductive dimethylation are well established and used to assist in crystallization (49). Dimethylation of ComZ impaired its ability to bind to DNA, as measured by EMSA (Fig. 6A). ComZ was exposed to the dimethylation reagent in the presence and absence of DNA and digested with protease, and peptides were identified from analysis by liquid chromatography-mass spectrometry (LC-MS) (Fig. 7A). Two peptides showed a significant reduction in intensity due to DNA protection and were identified as originating from modification of lysines 98 and 233. We noted that K98 and K233 are located between the two domains in ComZ, on the side of the L-shaped structure (Fig. 7B). To provide independent verification for this observation, both lysines were mutated to alanine, and the ability of the ComZ K98A/K233A double mutant to protect DNA from DNase digestion was
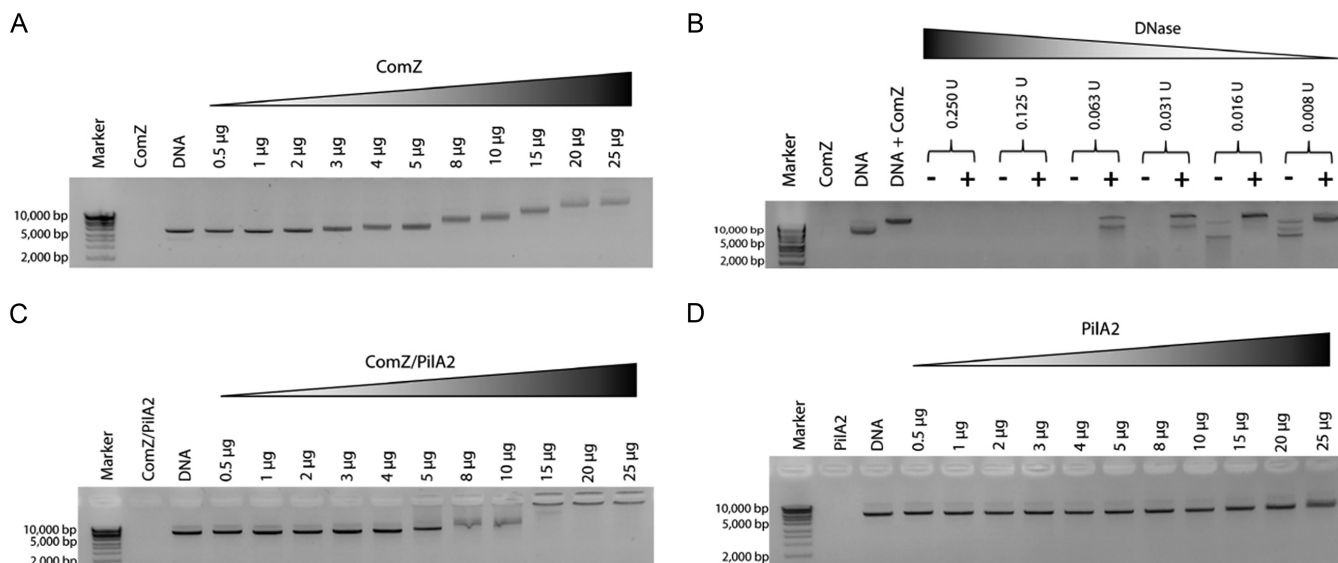
**FIG 5** Binding of ComZ to dsDNA. (A) ComZ-dsDNA interaction studied by EMSA. Increasing quantities of ComZ were added to 200 ng of linear dsDNA, as indicated. The ComZ lane is a control without DNA, and the adjacent lane, labeled DNA, is without added ComZ. (B) Protection of DNase digestion by ComZ. The digestion of linear dsDNA by various quantities of DNase was compared in the presence and absence of 8.7 $\mu$M ComZ. (C) ComZ/PilA2-dsDNA interaction studied by EMSA. Other conditions were as indicated for panel A. (D) PilA2-dsDNA interaction studied by EMSA. Other conditions were as indicated for panel A. DNA was detected using SafeView nucleic acid stain (NBS Biologicals). Marker used was HyperLadder 1 kb (Bioline).

compared to that of wild-type (WT) and dimethylated ComZ (Fig. 6B). Dimethylation had a substantial impact on the ability of ComZ to protect against degradation; the K98A/K233A mutant was also significantly impaired compared to the wild type, providing additional evidence that these lysine residues are indeed involved in DNA binding, although other lysines, and indeed other residues, are likely involved.
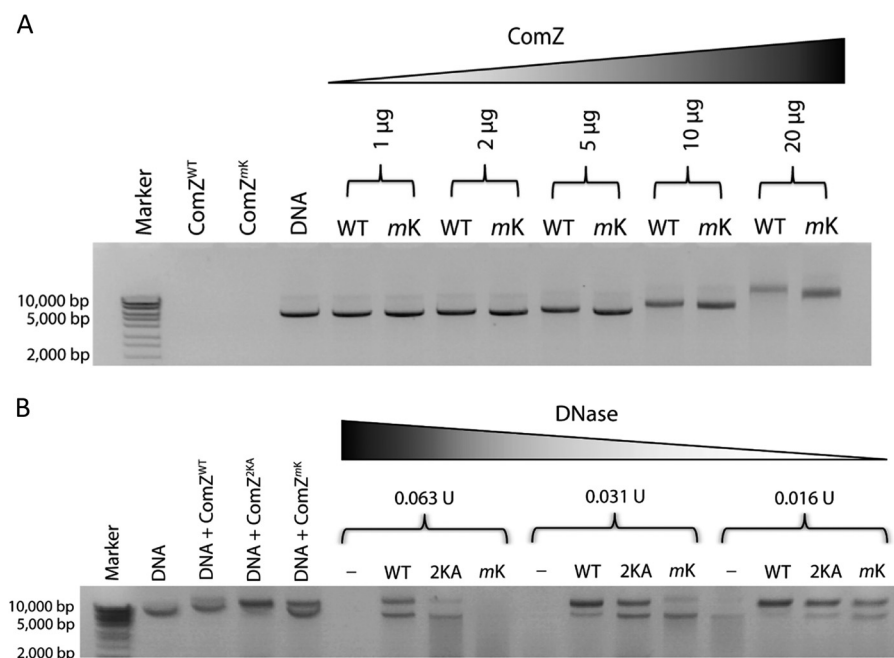


**FIG 6** Inhibition of DNA binding by dimethylated ComZ. (A) Increasing quantities of unmodified (WT) and dimethylated ComZ (mK), as indicated, were added to 200 ng of linear dsDNA. (B) DNase digestion protection was examined by comparison of the degradation of linear dsDNA in the absence of ComZ and presence of unmodified wild-type ComZ, dimethylated K98A/K233A ComZ (2KA), and dimethylated wild-type ComZ (mK). Marker used was HyperLadder 1 kb (Bioline).
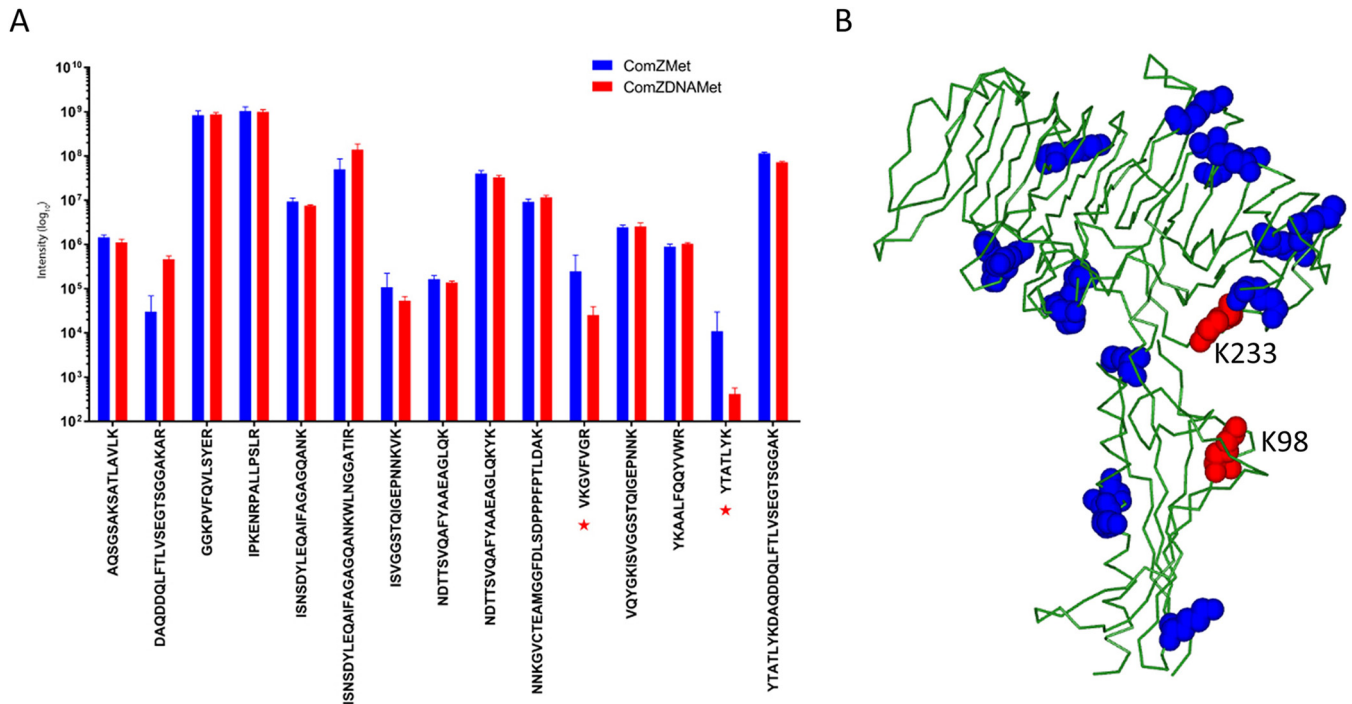
**FIG 7** LC-MS analysis of ComZ-DNA binding. (A) Intensities of dimethylated peptides in the absence (blue) and presence (red) of dsDNA ($n = 3$; values are $\pm$ standard errors from the means). (B) Location of lysine residues in ComZ. The two protected lysines, K98 and K233, are shown in red, and other lysines are in blue.

**Modelling of the ComZ-PilA2-DNA complex.** With structures of ComZ and PilA2 and some knowledge of the location of the DNA binding site on ComZ, we set out to model the tripartite complex. We reasoned that the complex between ComZ and PilA2 was likely formed by interaction between the two N-terminal helices in each structure, which is a consistent feature in the structures of assembled type IV pili. Using the complex of GspIJK as a starting model (29), the ComZ pilin domain was superposed onto GspK and PilA2 onto GspI, which is a close structural homolog. The model was adjusted to remove steric clashes, and six starting models were generated by rotation about the PilA2 helix, approximately 25° apart. Each model was sampled with 1,000 independent docking simulations implemented in ROSETTA (50), and the lowest-energy model was selected from the total of 6,000. This model places PilA2 such that its subdomain fits in the hinge region between the pilin and solenoid domains of ComZ and on the opposite side from K98 and K233 (Fig. 8). To generate a model for the ComZ-DNA complex, a 21-base B-form duplex DNA structure was generated using the make-na server (51) and docking carried out using HADDOCK2.2 (52), with the constraints that K98 and K233 were selected as active residues. The resulting model predicted contacts between the DNA and both domains in ComZ (Fig. 8). In particular, this arrangement shows DNA binding to residues 250 to 255 at the end of the $\beta$-solenoid and part of the extensive loop region, which links the pilin $\alpha$-helix with the first $\beta$-strand.

## DISCUSSION

The initial encounter of DNA by a specific receptor outside the cell is an essential first step in the DNA uptake by NT in Gram-negative bacteria. Although this outline model is consistent across different naturally competent Gram-negative organisms, details vary. For example, some species have specificity for certain DNA uptake sequences (DUS); this is the case for *H. influenzae* and *Neisseria* spp. This would require a receptor that is specific for each DUS, likely to be confined to a limited range of organisms, as is the case for neisserial ComP (38). In many cases, current knowledge of
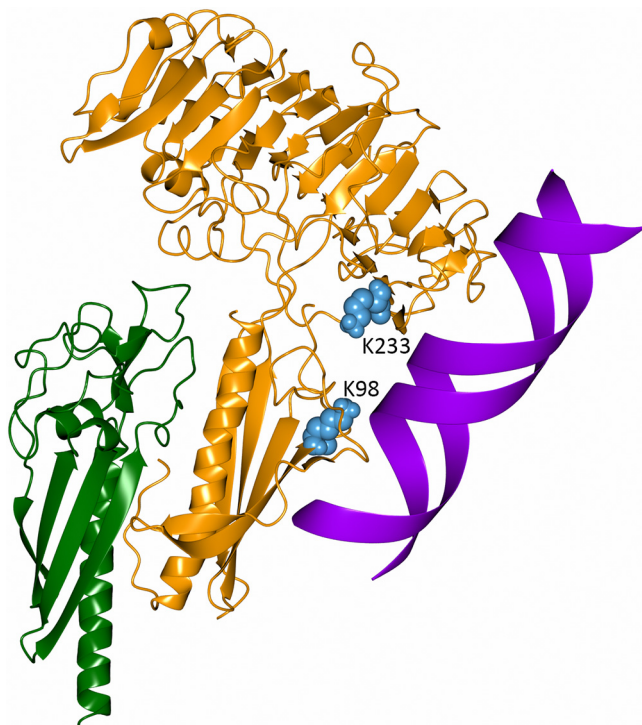
**FIG 8** Model for the ComZ-PilA2-DNA complex. PilA2 is shown in green, ComZ in orange, and DNA in purple. See the text for further details. The figure was generated using CCP4MG (72).

the atomic details of each step in the uptake process is scant. Recent developments in cryoelectron tomography have provided valuable insights into the assembly of the type IV pilus biogenesis system *in vivo*, revealing a complex with components in the inner and outer membranes, connected by a channel which spans the periplasm (22, 53). These reconstructions are, however, at comparatively low resolution; a complete description of the DNA uptake machine will require atomic-level detail, which can only be achieved by higher-resolution structure determination of the component proteins and their complexes.

The current model for DNA uptake therefore requires a receptor which is able to provide the first encounter with the DNA substrate. It is necessary to invoke the existence of such a receptor, as it is the most plausible way by which DNA is guided into and through the PilQ secretin channel, which is thought to be the conduit for DNA passage across the outer membrane (25). Direct evidence for this hypothesis was recently published by Ellison et al., who demonstrated, in *Vibrio cholerae*, the binding of type IV competence pili to extracellular DNA and pilus retraction transporting the bound substrate to the cell surface (54). It is reasonable to infer, based on current evidence for the requirement for competence pilins, that such specialized receptors are associated with the competence pilus fiber, most likely present as a specialized pilin. The structure of ComZ reveals a larger and more complex competence-associated pilin than those studied to date. ComP from *N. meningitidis* adopts a type IV pilin structure but has no additional domains (37). Interaction with DNA is through positively charged residues on the surface, and ComP binds DNA with a specificity which reflects the DUS specificity in *Neisseria* spp. (37, 38). It is unclear, however, if this model for a DUS-specific pilin applies outside the *Neisseriae*. In addition, the observation that uptake is not impaired in a *comP* mutant suggests that other components are involved in the initial process of DNA recognition (39). Competence pili have been directly observed in *S. pneumoniae* (55), and the solution structure of the component pilin, ComGC, was determined by nuclear magnetic resonance (56). ComGC is the major component of the *S. pneumoniae* competence-associated pilus (57). It has some characteristics in common

with T4a pilins, in that it has a hydrophobic N-terminal α-helix and a Glu at position 5 and is processed by the PilD peptidase. In contrast to ComP, DNA does not bind to the ComGC monomer, although it does do so to the assembled, mature competence pilus (57). There are therefore few points of similarity between *N. meningitidis* ComP and *S. pneumoniae* ComGC.

It is in this context that we examined the structures and functions of the competence-associated pilins previously identified in *T. thermophilus* HB27 (35). Mutation of these pilins gives a noncompetent but piliated phenotype, suggesting that PilA1-3 and ComZ combine to form a competence-specific pilus fiber. Of these four genes, ComZ stood out as the largest and had atypical features for a type IVa pilin, with a Gly at position 5, for example. ComZ is not predicted to be a substrate for the PilD peptidase by PilFind (42), although it should be noted that such a prediction is not infallible, and independent experimental evidence would be required to confirm whether or not ComZ is processed in this way. The structure is indicative of a type IV pilin fold, but with closest similarities to the GspK pseudopilin, which forms a heterotrimer proposed to form the tip of a pseudopilus which drives substrate secretion in the type II secretion system (T2SS) (29). This reflects our earlier observation that other *T. thermophilus* minor pilins are more closely related to T2SS pseudopilins than T4a pilin structures (35). We also show that ComZ binds to another pilin in the locus, PilA2, but not PilA1 or PilA3, suggesting some specificity of interaction, which would be expected if the pilins were incorporated into the fiber in a specific order. We note that ComZ contains a predicted hydrophobic α-helix, running from residues 5 to 27, which could form the basis of interaction with PilA2 in a fashion similar to that of the assembly of the GspK/GspI/GspJ heterotrimer. The dominant structural feature of ComZ, however, is the large β-solenoid domain, which appears to have been inserted into the pilin fold. The result is a large macromolecule, measuring ~60 Å across, similar to the diameter of type IV pilus fibers from Gram-negative organisms (27). The latter measurement is significant because a competence pilus fiber would need to navigate through the PilQ secretin channel. Recent structures of the T2SS GspD secretin, which is closely related to PilQ, confirm earlier observations that the central secretin pore is approximately 60 Å across, with the potential to widen to accommodate the emerging pilus fiber and therefore, by implication, a tip-located ComZ (58). In contrast, it is difficult to reconcile incorporation of ComZ into the central body of a type IVa pilus fiber. Atomic models for type IV pilus fibers vary in detail—although none are available for *T. thermophilus* pilus fibers at present—but are based on the association of hydrophobic N-terminal α-helices at the center of the fiber. We find that the size of the β-solenoid domain, and its limited flexibility with respect to the pilin domain, argues against its incorporation into the main competence pilus fiber; steric constraints make this unlikely. We are left with the more plausible option, which is that ComZ is a tip pilin, as is the case for GspK. This suggests that ComZ is incorporated first into the nascent competence fiber as it is being assembled, followed by PilA2. Our model for the ComZ/PilA2 complex underscores this point by locating PilA2 such that its smaller subdomain sits between the ComZ pilin and β-solenoid domains, suggesting that ComZ could effectively cap the end of the competence pilus fiber (Fig. 8).

DNA binding is an unusual function for a β-solenoid structure, usually associated with hydrolytic enzyme activity (e.g., K5 lyase tailspike protein [59]). An examination of surface electrostatics of the β-solenoid domain did not suggest any characteristic positive charge patches, which are often associated with DNA binding (Fig. S7), but this does not necessarily preclude a DNA binding function. We obtained direct evidence for DNA binding to ComZ by developing a method based on protection of lysine residues against reductive dimethylation. Our resulting structural model suggests a side-on association of DNA with the end of the pilus fiber (Fig. 8). *In vivo*, association of DNA with the competence pilus could well involve other competence pilins and will require further study of the assembly of the fiber.

Figure 9 summarizes our proposed model for the role of ComZ in the DNA uptake into *T. thermophilus* by NT. ComZ is located at the tip of a competence pilus, supported
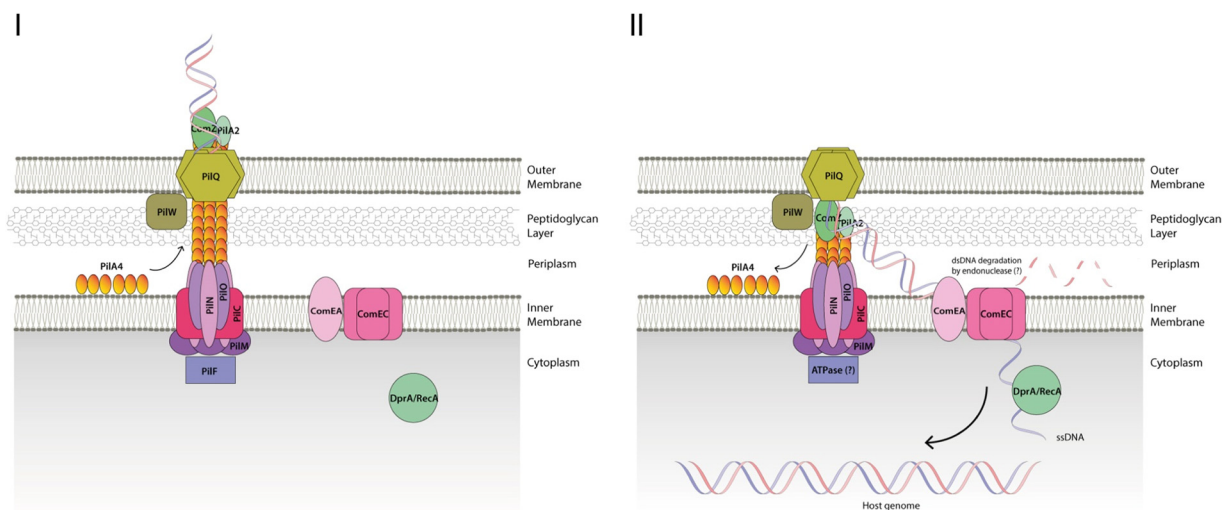
**FIG 9** Schematic model for DNA uptake in *T. thermophilus*. In stage I, the pilin PilA4 (17) is assembled into a competence pilus fiber by the inner membrane complex PilMNO and PilC (14, 73). Elongation of the pilus is powered by the hexameric AAA-ATPase PilF, which provides energy for the assembly via ATP hydrolysis (13, 74). In stage II, double-stranded DNA is bound by the pilus tip-associated DNA receptor ComZ/PilA2 and transported into the periplasm through the secretin pore PilQ (24). A competence-associated retraction ATPase has not been identified and may not, in any case, be required. Once in the periplasm, double-stranded DNA binds to ComEA (6), one strand is degraded by an unidentified endonuclease, and the remaining DNA strand is translocated across the inner membrane through ComEC (20). In the cytoplasm, DNA is used as a source for intracellular metabolism or recombined with the host chromosomal genome (75).

in position by the minor pilin PilA2. It acts as the initial receptor for DNA binding outside the cell; the assembled pilus is retracted to bring the DNA substrate into the periplasm and into close proximity with ComEA and ComEC. DNA is then transferred to ComEA, which is inner membrane associated, and probably works in concert with the ComEC channel. The role of DNA binding to ComEA has been highlighted by the recent observations of Hepp and Maier in *Neisseria* (39). Their model suggests that ComEA binding, in a ratchet-type model, drives DNA uptake. This hypothesis requires that ComEA binds with higher affinity than any DNA receptors or components, which function earlier in the process. This could include a protein with a function similar to that of ComZ and perhaps also the PilQ secretin; indeed, there is evidence that *N. meningitidis* PilQ binds DNA (25).

The general applicability of our conclusions on the function of ComZ lie in the role of a tip-located pilin with a specific DNA-binding function. This is a feature which could be replicated in other species, even if the structural details are different. Within Gram-negative organisms, NT uptake systems are diverse. Although they have some common constituents, such as the ComA and ComE proteins, other components vary. One reason may be that some organisms exhibit DNA sequence specificity through DUSs, whereas others do not. In addition, regulation differs between different organisms (21). Nevertheless, the structure and properties of ComZ point to the involvement of a specialized, tip-associated T4a pilin for initial DNA recognition, which could have general applicability to other NT systems.

## MATERIALS AND METHODS

**Cloning, expression, and purification of ComZ and PilA1-3.** The *comZ* gene (omitting the region coding for the signal peptidase sequence and hydrophobic part of the N-terminal helix) was amplified by PCR using primers CTTCACCATGGCCATAGAGCTCTGGACCACCCGCAACGAC and CGGTGTGACTCGAG GCGGCGCTCATAGGAGAGCACCTG and *T. thermophilus* HB27 genomic DNA as the template. The amplified gene and the pET-22b vector (Novagen) were digested with restriction enzymes NcoI and XhoI, purified, and ligated. The *comZ*-22b construct coded for the ComZ protein (residues 31 to 554) with the PelB leader sequence at the N terminus and hexahistidine tag at the C terminus. Synthetic constructs corresponding to the soluble domains PilA1 (residues 30 to 156), PilA2 (residues 33 to 193), and PilA3 (residues 33 to 233) were designed, optimized for expression in *E. coli*, and synthesized (GeneArt) with a Strep-tag II prior to subcloning into pET-22b vector (BamHI and XhoI sites), which includes the PelB leader at the N terminus. Recombinant plasmids were transformed into T7 Express cells (New England Biolabs [NEB]) for ComZ expression or Lemo21(DE3) cells (NEB) for minor prepilin expression and grown

on Luria-Bertani (LB) plates with 100 $\mu$g/ml ampicillin (T7 Express) or 100 $\mu$g/ml ampicillin and 30 $\mu$g/ml chloramphenicol (Lemo21 cells) at 37°C for 14 h. Several colonies were inoculated into 50 ml of starter culture (LB medium with antibiotics) and grown until the optical density at 600 nm ($OD_{600}$) reached 0.6 to 0.8. Cells were subsequently grown in Terrific broth (TB) medium supplemented with antibiotics and were induced with 0.4 mM isopropyl-$\beta$-D-thiogalactopyranoside (IPTG) when the $OD_{600}$ reached 0.8. Cells were left shaking at 16°C for 16 h to allow protein expression before being harvested by centrifugation at 8,000 rpm (SLA3000; Sorvall) for 30 min at 4°C. Cell pellets were resuspended in buffer A (25 mM Tris-HCl, pH 8.5, 100 mM NaCl for ComZ and 25 mM Tris-HCl, pH 8.0, 200 mM NaCl for prepilins) supplemented with protease inhibitor cocktail (Roche) and DNase I (5 $\mu$g/ml). Cells were disrupted by sonication (Sonopuls; Bandelin) for 7 min at 30% power. Debris and unbroken cells were removed by centrifugation at 18,000 rpm (F21; Sorvail) for 45 min at 4°C. For ComZ, purification was carried out by using a HisTrap HP column (GE Healthcare), followed by an elution step using buffer A plus increasing concentrations of imidazole (up to 500 mM). The purification of Strep-tag II-tagged pilins was carried out on a StrepTrap HP column (GE Healthcare). The streptavidin column was first equilibrated with 5 column volumes (CV) of water and 5 CV of buffer A. Crude cell lysates were applied to the column, which was washed with 6 CV of buffer A, followed by elution with 6 CV of buffer A plus 5% glycerol and 2.5 mM desthiobiotin. Collected fractions from cell lysates, wash, and elution buffer were analyzed by SDS-PAGE. A final purification step by SEC was carried out using a HiLoad 16/600 Superdex 75 PG column (GE Healthcare) in 25 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5% glycerol at 25°C.

**Protein crystallization, data collection, and structural analysis.** Purified ComZ (8 to 10 mg/ml) was crystallized by sitting drop vapor diffusion. Equal volumes (200 nl) of protein and a reservoir solution containing 0.2 M potassium thiocyanate, 0.1 M Bis-Tris propane, pH 6.5, and 20% (wt/vol) polyethylene glycol (PEG) 3350 were mixed and incubated at 293K. For phasing, crystals were soaked with reservoir solution supplemented with 10 mM potassium tetrachloroplatinate(II) ($K_2PtCl_4$) for 2 to 3 h. Crystals were cryoprotected with reservoir solution supplemented with 20% glycerol and flash-cooled in liquid nitrogen. The native ComZ data set was processed using the xia2 (60) automated pipeline implementing XDS (61), XSCALE, and AIMLESS (62) to a resolution of 2.72 Å. For the Pt derivative, as the anomalous signal was weak, the XDS integrated data from four ComZ crystals (treated with $K_2PtCl_4$) were manually scaled and merged using AIMLESS, as implemented in the CCP4 suite (63), to a resolution of 3.5 Å. This gave an anomalous multiplicity of 33 and midslope anomalous normal probability of 1.19. Automated substructure identification, calculation of phases, density modification, and low-resolution preliminary model building were carried out using the AutoSol wizard in Phenix (64). Thirteen Pt sites were identified and produced an interpretable electron density map. At this stage, it was clear that there are three molecules in the asymmetric unit and the presence of the $\beta$-solenoid domain. Manual model building was carried out to assign as many residues as possible to regions of interpretable electron density for one of the chains. This chain was then used as a search model for molecular replacement to extend the phases to the native data set using Phaser (65). The model was built using the AutoBuild wizard in Phenix (64) and completed using iterative rounds of manual model building using Coot (66) and refinement using phenix.refine (64). The structure was analyzed using the PDB_REDO server (67) and validated using MolProbity (68). X-ray data collection and refinement statistics are presented in Table S1 in the supplemental material.

Crystals of PilA2 were obtained using MRC 2-well plates containing 200 nl of protein (13 mg/ml) and 200 nl of 0.2 M ammonium nitrate, 0.1 M Bis-Tris propane, pH 8.5, and 18% (vol/vol) PEG Smear High using a Mosquito robot (TTP Labtech). For phasing, crystals were soaked with the reservoir solution, supplemented with solid potassium iodide (KI) and 15% PEG 200 for 5 min. Crystals were cryoprotected with the reservoir solution, supplemented with 20% glycerol, and then flash-cooled in liquid nitrogen. The native data set was processed using xia2 (69), the automated pipeline implementing XDS (61), fast DP, and autoPROC 1.0.5. For the iodide derivative, the XDS integrated data from PilA2 crystals (treated with KI) were manually scaled and merged using AIMLESS (62), as implemented in the CCP4 suite (63), to a resolution of 2.81 Å. This gave an anomalous multiplicity of 6.3 and a midslope anomalous normal probability of 1.103. The model was built using the AutoBuild wizard in Phenix (64) and completed using iterative rounds of the manual model building using Coot (66) and refinement using refmac (70). X-ray data collection and refinement statistics are presented in Table S2.

**Biophysical binding measurements.** For studies of the interaction of ComZ with PilA1-3, analysis was carried out on a HiLoad 16/600 Superdex 200 PG column (GE Healthcare). Prior to the chromatography, the target proteins were incubated together in 25 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5% glycerol at 4°C for 30 min. Affinity tag binding assays used the principle that both ComZ and pilin proteins were expressed with different tags. The ComZ reading frame encoded a C-terminal hexahistidine tag, whereas PilA1, PilA2, and PilA3 incorporated a C-terminal Strep-tag II (WSHPQFEK). A HisTrap HP column was used to study the ComZ-PilA2-PilA3 interaction, and a StrepTrap HP column was used for interaction analysis of ComZ-PilA2 and ComZ-PilA3. Three mg of ComZ and 2 mg of each pilin were loaded in each case. Washing and elution from each affinity column was carried out under the same conditions as those used for affinity chromatography purification in the protein purification protocols described above.

ITC was performed on a MicroCal PEAQ-ITC instrument (Malvern) by titrating 20 2-$\mu$l aliquots of PilA2 (314 $\mu$M) into the sample cell containing 300 $\mu$l of ComZ (33 $\mu$M), with rapid stirring at 25°C. Thermodynamic parameters, i.e., binding constant ($K_d$), stoichiometry ($n$), enthalpy ($\Delta H$), and entropy ($\Delta S$), were calculated using the manufacturer's software by measuring heat absorbed or released (71).

Electrophoretic mobility shift assays (EMSA) were carried out by mixing ComZ and linearized DNA plasmid (pET-22b predigested with BamHI), incubation at 4°C for 1 h, and separation by electrophoresis on a 1.5% agarose gel, supplemented with SafeView nucleic acid stain (NBS Biologicals). Increasing

concentrations of ComZ (0 $\mu$M to 44 $\mu$M) were added to a fixed quantity of linearized pET-22b (200 ng) in 25 mM Tris-HCl, pH 8.5, 100 mM NaCl in a final volume of 10 $\mu$l. After electrophoresis at a constant voltage (110 V) for 30 min, the gel was visualized and photographed under UV light using a UVpro transilluminator (UVItec). For the DNase protection assay, linearized pET-22b (200 ng) was added to 8.7 $\mu$M ComZ in 25 mM Tris-HCl, pH 8.5, 100 mM NaCl in a final volume of 10 $\mu$l and incubated at 55°C for 15 min. Samples were cooled to 37°C, DNase I was added to the specified concentration, and samples were incubated for a further 30 min at 37°C prior to separation by electrophoresis.

**Reductive dimethylation protection.** Protection from reductive dimethylation was used to map surface-exposed lysine residues on ComZ, which bind DNA. The protocol was adapted from a method described by Rayment (49). Purified ComZ was exchanged into 50 mM HEPES, pH 7.5, using a Pierce polyacrylamide spin desalting column (7-kDa molecular weight cutoff; Thermo Fisher). Reductive dimethylation was carried out by addition of 20 mM borane-dimethylamine (BDC; Sigma-Aldrich) and 38 mM formaldehyde (final concentrations) in 500 $\mu$l of 70 $\mu$M ComZ in the presence or absence of 1 $\mu$g pET22b linearized by BamHI digestion. The solution was incubated at 0°C for 2 h before addition of a further 20 mM BDC–38 mM formaldehyde and incubation at the same temperature for a further 2 h. Following a third addition of 10 mM BDC, the solution was incubated for a further 14 h at 0°C. The reaction was quenched by addition of glycine to a final concentration of 124 mM, incubated for 1 h at 0°C, and dialyzed into 50 mM HEPES, pH 7.5, for 4 h at 4°C. A further 14-h dialysis was performed against 50 mM HEPES, pH 7.5, 2 mM dithiothreitol (DTT) at the same temperature to reverse any modifications of cysteine and methionine residues.

**Mass spectrometry.** For trypsin digestion, 500 $\mu$l of 10.5 $\mu$M purified ComZ was incubated at 90°C for 1 h in 20 mM ammonium bicarbonate and 10 mM DTT and cooled to 22°C, and iodoacetamide was added to a final concentration of 10 mM. The reaction mix was incubated in the dark at 22°C for 45 min before addition of trypsin at a mass ratio of 1:75 and incubated for 14 h at 37°C. Samples were desalted by reverse-phase, solid-phase extraction using Oligo R3 beads (Thermo Scientific) with elution in 40% acetonitrile in 0.1% formic acid, followed by evaporation via vacuum centrifugation and subsequent resuspension in 10 $\mu$l of 5% acetonitrile in 0.1% formic acid. Samples were analyzed by LC-tandem MS (LC-MS/MS) using an UltiMate 3000 rapid separation LC (RSLC; Dionex Corporation, Sunnyvale, CA) coupled to an Orbitrap Elite (Thermo Fisher Scientific, Waltham, MA) mass spectrometer. Peptide mixtures were separated using a gradient from 92% A (0.1% FA in water) and 8% B (0.1% FA in acetonitrile) to 33% B in 44 min at 300 nl min$^{-1}$, using a 75-mm by 250-$\mu$m-inner-diameter 1.7 $\mu$M CSH C$_{18}$ analytical column (Waters). Peptides were selected for fragmentation automatically by data-dependent analysis. The data produced were then analyzed using Progenesis QI for Proteomics software (Waters) to align and quantify the peptide signals, which were then identified using Mascot software (Matrix Sciences) from the Swiss-Prot database of protein sequences (European Bioinformatics Institute) from January 2018, using oxidation on methionines and dimethylation on lysines as variable modifications. The signal intensities of peptides that were identified from ComZ were compared to determine the level of protection from dimethylation.

**Data availability.** Structure factors and atomic coordinates for ComZ and PilA2 have been deposited in the Protein Data Bank with accession codes 6QVI and 6QVF, respectively.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio .00614-19.

**FIG S1**, DOCX file, 1 MB.
**FIG S2**, DOCX file, 2 MB.
**FIG S3**, DOCX file, 0.4 MB.
**FIG S4**, DOCX file, 0.6 MB.
**FIG S5**, DOCX file, 0.5 MB.
**FIG S6**, DOCX file, 0.8 MB.
**FIG S7**, DOCX file, 0.9 MB.
**TABLE S1**, DOCX file, 0.01 MB.
**TABLE S2**, DOCX file, 0.01 MB.

## ACKNOWLEDGMENTS

the structure of ComZ; M.Z.S., J.P.D., and C.W.L. crystallized and determined the structure of PilA2; M.Z.S. carried out protein-protein and protein-DNA interaction studies; M.Z.S. and D.K. conducted the mass spectrometry mapping experiments; M.Z.S., J.P.D., and V.K. prepared the figures and wrote the manuscript. We have no competing financial interests to declare.

## REFERENCES

1. Chen I, Dubnau D. 2004. DNA uptake during bacterial transformation. Nat Rev Microbiol 2:241–249. https://doi.org/10.1038/nrmicro844.
2. Averhoff B, Friedrich A. 2003. Type IV pili-related natural transformation systems: DNA transport in mesophilic and thermophilic bacteria. Arch Microbiol 180:385–393. https://doi.org/10.1007/s00203-003-0616-6.
3. Muschiol S, Balaban M, Normark S, Henriques-Normark B. 2015. Uptake of extracellular DNA: competence induced pili in natural transformation of Streptococcus pneumoniae. Bioessays 37:426–435. https://doi.org/10.1002/bies.201400125.
4. Averhoff B. 2009. Shuffling genes around in hot environments: the unique DNA transporter of Thermus thermophilus. FEMS Microbiol Rev 33:611–626. https://doi.org/10.1111/j.1574-6976.2008.00160.x.
5. Friedrich A, Hartsch T, Averhoff B. 2001. Natural transformation in mesophilic and thermophilic bacteria: identification and characterization of novel, closely related competence genes in Acinetobacter sp. strain BD413 and Thermus thermophilus HB27. Appl Environ Microbiol 67:3140–3148. https://doi.org/10.1128/AEM.67.7.3140-3148.2001.
6. Salzer R, Kern T, Joos F, Averhoff B. 2016. The Thermus thermophilus comEA/comEC operon is associated with DNA binding and regulation of the DNA translocator and type IV pili. Environ Microbiol 18:65–74. https://doi.org/10.1111/1462-2920.12820.
7. Maughan H, Wilson LA, Redfield RJ. 2010. Bacterial DNA uptake sequences can accumulate by molecular drive alone. Genetics 186:613–627. https://doi.org/10.1534/genetics.110.119438.
8. Mell JC, Hall IM, Redfield RJ. 2012. Defining the DNA uptake specificity of naturally competent Haemophilus influenzae cells. Nucleic Acids Res 40:8536–8549. https://doi.org/10.1093/nar/gks640.
9. Treangen TJ, Ambur OH, Tonjum T, Rocha EP. 2008. The impact of the neisserial DNA uptake sequences on genome evolution and stability. Genome Biol 9:R60. https://doi.org/10.1186/gb-2008-9-3-r60.
10. Koyama Y, Hoshino T, Tomizuka N, Furukawa K. 1986. Genetic transformation of the extreme thermophile Thermus thermophilus and of other Thermus spp. J Bacteriol 166:338–340. https://doi.org/10.1128/jb.166.1.338-340.1986.
11. Schwarzenlander C, Averhoff B. 2006. Characterization of DNA transport in the thermophilic bacterium Thermus thermophilus HB27. FEBS J 273:4210–4218. https://doi.org/10.1111/j.1742-4658.2006.05416.x.
12. Berry JL, Pelicic V. 2015. Exceptionally widespread nanomachines composed of type IV pilins: the prokaryotic Swiss Army knives. FEMS Microbiol Rev 39:134–154. https://doi.org/10.1093/femsre/fuu001.
13. Collins RF, Hassan D, Karuppiah V, Thistlethwaite A, Derrick JP. 2013. Structure and mechanism of the PilF DNA transformation ATPase from Thermus thermophilus. Biochem J 450:417–425. https://doi.org/10.1042/BJ20121599.
14. Karuppiah V, Collins RF, Thistlethwaite A, Gao Y, Derrick JP. 2013. Structure and assembly of an inner membrane platform for initiation of type IV pilus biogenesis. Proc Natl Acad Sci U S A 110:E4638–E4647. https://doi.org/10.1073/pnas.1312313110.
15. Karuppiah V, Hassan D, Saleem M, Derrick JP. 2010. Structure and oligomerization of the PilC type IV pilus biogenesis protein from Thermus thermophilus. Proteins 78:2049–2057.
16. Karuppiah V, Derrick JP. 2011. Structure of the PilM-PilN inner membrane type IV pilus biogenesis complex from Thermus thermophilus. J Biol Chem 286:24434–24442. https://doi.org/10.1074/jbc.M111.243535.
17. Rumszauer J, Schwarzenlander C, Averhoff B. 2006. Identification, subcellular localization and functional interactions of PilMNOWQ and PilA4 involved in transformation competency and pilus biogenesis in the thermophilic bacterium Thermus thermophilus HB27. FEBS J 273:3261–3272. https://doi.org/10.1111/j.1742-4658.2006.05335.x.
18. Salzer R, Herzberg M, Nies DH, Joos F, Rathmann B, Thielmann Y, Averhoff B. 2014. Zinc and ATP binding of the hexameric AAA-ATPase PilF from Thermus thermophilus: role in complex stability, piliation, adhesion, twitching motility, and natural transformation. J Biol Chem 289:30343–30354. https://doi.org/10.1074/jbc.M114.598656.
19. Salzer R, Joos F, Averhoff B. 2014. Type IV pilus biogenesis, twitching motility, and DNA uptake in Thermus thermophilus: discrete roles of antagonistic ATPases PilF, PilT1, and PilT2. Appl Environ Microbiol 80:644–652. https://doi.org/10.1128/AEM.03218-13.
20. Draskovic I, Dubnau D. 2005. Biogenesis of a putative channel protein, ComEC, required for DNA uptake: membrane topology, oligomerization and formation of disulphide bonds. Mol Microbiol 55:881–896. https://doi.org/10.1111/j.1365-2958.2004.04430.x.
21. Seitz P, Blokesch M. 2013. Cues and regulatory pathways involved in natural competence and transformation in pathogenic and environmental Gram-negative bacteria. FEMS Microbiol Rev 37:336–363. https://doi.org/10.1111/j.1574-6976.2012.00353.x.
22. Gold VAM, Salzer R, Averhoff B, Kuhlbrandt W. 2015. Structure of a type IV pilus machinery in the open and closed state. Elife 4:eLife.07380. https://doi.org/10.7554/eLife.07380.
23. Berry J-L, Phelan MM, Collins RF, Adomavicius T, Tonjum T, Frye SA, Bird L, Owens R, Ford RC, Lian L-Y, Derrick JP. 2012. Structure and assembly of a trans-periplasmic channel for type IV pili in Neisseria meningitidis. PLoS Pathog 8:e1002923. https://doi.org/10.1371/journal.ppat.1002923.
24. Burkhardt J, Vonck J, Averhoff B. 2011. Structure and function of PilQ, a secretin of the DNA transporter from the thermophilic bacterium Thermus thermophilus HB27. J Biol Chem 286:9977–9984. https://doi.org/10.1074/jbc.M110.212688.
25. Assalkhou R, Balasingham S, Collins RF, Frye SA, Davidsen T, Benam AV, Bjoras M, Derrick JP, Tonjum T. 2007. The outer membrane secretin PilQ from Neisseria meningitidis binds DNA. Microbiology 153:1593–1603. https://doi.org/10.1099/mic.0.2006/004200-0.
26. Hovland E, Beyene GT, Frye SA, Homberset H, Balasingham SV, Gómez-Muñoz M, Derrick JP, Tønjum T, Ambur OH. 2017. DprA from Neisseria meningitidis: properties and role in natural competence for transformation. Microbiology 163:1016–1029. https://doi.org/10.1099/mic.0.000489.
27. Craig L, Li J. 2008. Type IV pilli: paradoxes in form and function. Curr Opin Struct Biol 18:267–277. https://doi.org/10.1016/j.sbi.2007.12.009.
28. Kolappan S, Ng D, Yang GX, Harn T, Craig L. 2015. Crystal structure of the minor pilin CofB, the initiator of CFA/III pilus assembly in enterotoxigenic Escherichia coli. J Biol Chem 290:25805–25818. https://doi.org/10.1074/jbc.M115.676106.
29. Korotkov KV, Hol W. 2008. Structure of the GspK-GspI-GspJ complex from the enterotoxigenic Escherichia coli type 2 secretion system. Nat Struct Mol Biol 15:462–468. https://doi.org/10.1038/nsmb.1426.
30. Carbonnelle E, Helaine S, Nassif X, Pelicic V. 2006. A systematic genetic analysis in Neisseria meningitidis defines the Pil proteins required for assembly, functionality, stabilization and export of type IV pili. Mol Microbiol 61:1510–1522. https://doi.org/10.1111/j.1365-2958.2006.05341.x.
31. Imhaus AF, Dumenil G. 2014. The number of Neisseria meningitidis type IV pili determines host cell interaction. EMBO J 33:1767–1783. https://doi.org/10.15252/embj.201488031.
32. Karuppiah V, Derrick JP. 2014. Type IV pili-a numbers game. EMBO J 33:1732–1734. https://doi.org/10.15252/embj.201489096.
33. Nguyen Y, Sugiman-Marangos S, Harvey H, Bell SD, Charlton CL, Junop MS, Burrows LL. 2015. Pseudomonas aeruginosa minor pilins prime type IVa pilus assembly and promote surface display of the PilY1 adhesin. J Biol Chem 290:601–611. https://doi.org/10.1074/jbc.M114.616904.
34. Nguyen Y, Harvey H, Sugiman-Marangos S, Bell SD, Buensuceso RN, Junop MS, Burrows LL. 2015. Structural and functional studies of the Pseudomonas aeruginosa minor pilin, PilE. J Biol Chem 290:26856–26865. https://doi.org/10.1074/jbc.M115.683334.
35. Karuppiah V, Thistlethwaite A, Derrick JP. 2016. Structures of type IV pilins from Thermus thermophilus demonstrate similarities with type II

secretion system pseudopilins. J Struct Biol 196:375–384. https://doi.org/10.1016/j.jsb.2016.08.006.

36. Wolfgang M, van Putten JP, Hayes SF, Koomey M. 1999. The *comP* locus of *Neisseria gonorrhoeae* encodes a type IV prepilin that is dispensable for pilus biogenesis but essential for natural transformation. Mol Microbiol 31:1345–1357. https://doi.org/10.1046/j.1365-2958.1999.01269.x.

37. Cehovin A, Simpson PJ, McDowell MA, Brown DR, Noschese R, Pallett M, Brady J, Baldwin GS, Lea SM, Matthews SJ, Pelicic V. 2013. Specific DNA recognition mediated by a type IV pilin. Proc Natl Acad Sci U S A 110:3065–3070. https://doi.org/10.1073/pnas.1218832110.

38. Berry JL, Cehovin A, McDowell MA, Lea SM, Pelicic V. 2013. Functional analysis of the interdependence between DNA uptake sequence and its cognate ComP receptor during natural transformation in *w* species. PLoS Genet 9:e1004014. https://doi.org/10.1371/journal.pgen.1004014.

39. Hepp C, Maier B. 2016. Kinetics of DNA uptake during transformation provide evidence for a translocation ratchet mechanism. Proc Natl Acad Sci U S A 113:12467–12472. https://doi.org/10.1073/pnas.1608110113.

40. Friedrich A, Rumszauer J, Henne A, Averhoff B. 2003. Pilin-like proteins in the extremely thermophilic bacterium *Thermus thermophilus* HB27: implication in competence for natural transformation and links to type IV pilus biogenesis. Appl Environ Microbiol 69:3695–3700. https://doi.org/10.1128/aem.69.7.3695-3700.2003.

41. Krissinel E, Henrick K. 2007. Inference of macromolecular assemblies from crystalline state. J Mol Biol 372:774–797. https://doi.org/10.1016/j.jmb.2007.05.022.

42. Imam S, Chen Z, Roos DS, Pohlschröder M. 2011. Identification of surprisingly diverse type IV Pili, across a broad range of Gram-positive bacteria. PLoS One 6:e28919. https://doi.org/10.1371/journal.pone.0028919.

43. Holm L, Rosenstrom P. 2010. Dali server: conservation mapping in 3D. Nucleic Acids Res 38:W545–W549. https://doi.org/10.1093/nar/gkq366.

44. Kolappan S, Coureuil M, Yu X, Nassif X, Egelman EH, Craig L. 2016. Structure of the *Neisseria meningitidis* Type IV pilus. Nat Commun 7:13015. https://doi.org/10.1038/ncomms13015.

45. Do H, Kim SJ, Kim HJ, Lee JH. 2014. Structure-based characterization and antifreeze properties of a hyperactive ice-binding protein from the Antarctic bacterium *Flavobacterium frigoris* PS1. Acta Crystallogr D Biol Crystallogr 70:1061–1073. https://doi.org/10.1107/S1399004714000996.

46. Bar Dolev M, Braslavsky I, Davies PL. 2016. Ice-binding proteins and their function. Annu Rev Biochem 85:515–542. https://doi.org/10.1146/annurev-biochem-060815-014546.

47. Holm L, Kääriäinen S, Rosenström P, Schenkel A. 2008. Searching protein structure databases with DaliLite v.3. Bioinformatics 24:2780–2781. https://doi.org/10.1093/bioinformatics/btn507.

48. Lam AY, Pardon E, Korotkov KV, Hol WGJ, Steyaert J. 2009. Nanobody-aided structure determination of the EpsI:EpsJ pseudopilin heterodimer from Vibrio vulnificus. J Struct Biol 166:8–15. https://doi.org/10.1016/j.jsb.2008.11.008.

49. Rayment I. 1997. Reductive alkylation of lysine residues to alter crystallization properties of proteins. Methods Enzymol 276:171–179. https://doi.org/10.1016/S0076-6879(97)76058-0.

50. Chaudhury S, Berrondo M, Weitzner BD, Muthu P, Bergman H, Gray JJ. 2011. Benchmarking and analysis of protein docking performance in Rosetta v3.2. PLoS One 6:13.

51. Macke TJ, Case DA. 1998. Modeling unusual nucleic acid structures. ACS Symp Ser 682:379–393. https://doi.org/10.1021/bk-1998-0682.ch024.

52. van Zundert GCP, Rodrigues J, Trellet M, Schmitz C, Kastritis PL, Karaca E, Melquiond ASJ, van Dijk M, de Vries SJ, Bonvin A. 2016. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. J Mol Biol 428:720–725. https://doi.org/10.1016/j.jmb.2015.09.014.

53. Chang Y-W, Rettberg LA, Treuner-Lange A, Iwasa J, Søgaard-Andersen L, Jensen GJ. 2016. Architecture of the type IVa pilus machine. Science 351:aad2001. https://doi.org/10.1126/science.aad2001.

54. Ellison CK, Dalia TN, Vidal Ceballos A, Wang J-Y, Biais N, Brun YV, Dalia AB. 2018. Retraction of DNA-bound type IV competence pili initiates DNA uptake during natural transformation in Vibrio cholerae. Nat Microbiol 3:773–780. https://doi.org/10.1038/s41564-018-0174-y.

55. Laurenceau R, Péhau-Arnaudet G, Baconnais S, Gault J, Malosse C, Dujeancourt A, Campo N, Chamot-Rooke J, Le Cam E, Claverys J-P, Fronzes R. 2013. A type IV pilus mediates DNA binding during natural transformation in *Streptococcus pneumoniae*. PLoS Pathog 9:e1003473. https://doi.org/10.1371/journal.ppat.1003473.

56. Muschiol S, Erlendsson S, Aschtgen MS, Oliveira V, Schmieder P, de Lichtenberg C, Teilum K, Boesen T, Akbey U, Henriques-Normark B. 2017.

Structure of the competence pilus major pilin ComGC in *Streptococcus pneumoniae*. J Biol Chem 292:14134–14146. https://doi.org/10.1074/jbc.M117.787671.

57. Balaban M, Bättig P, Muschiol S, Tirier SM, Wartha F, Normark S, Henriques-Normark B. 2014. Secretion of a pneumococcal type II secretion system pilus correlates with DNA uptake during transformation. Proc Natl Acad Sci U S A 111:E758–E765. https://doi.org/10.1073/pnas.1313860111.

58. Yan ZF, Yin M, Xu DD, Zhu YQ, Li XM. 2017. Structural insights into the secretin translocation channel in the type II secretion system. Nat Struct Mol Biol 24:177–183. https://doi.org/10.1038/nsmb.3350.

59. Thompson JE, Pourhossein M, Waterhouse A, Hudson T, Goldrick M, Derrick JP, Roberts IS. 2010. The K5 lyase KflA combines a viral tail spike structure with a bacterial polysaccharide lyase mechanism. J Biol Chem 285:23963–23969. https://doi.org/10.1074/jbc.M110.127571.

60. Winter G, Lobley CM, Prince SM. 2013. Decision making in xia2. Acta Crystallogr D Biol Crystallogr 69:1260–1273. https://doi.org/10.1107/S0907444913015308.

61. Kabsch W. 2010. XDS. Acta Crystallogr D Biol Crystallogr 66:125–132. https://doi.org/10.1107/S0907444909047337.

62. Evans PR, Murshudov GN. 2013. How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr 69:1204–1214. https://doi.org/10.1107/S0907444913000061.

63. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. 2011. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 67:235–242. https://doi.org/10.1107/S0907444910045749.

64. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66:213–221. https://doi.org/10.1107/S0907444909052925.

65. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. J Appl Crystallogr 40:658–674. https://doi.org/10.1107/S0021889807021206.

66. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66:486–501. https://doi.org/10.1107/S0907444910007493.

67. Joosten RP, Long F, Murshudov GN, Perrakis A. 2014. The PDB_REDO server for macromolecular structure model optimization. IUCrJ 1:213–220. https://doi.org/10.1107/S2052252514009324.

68. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66:12–21. https://doi.org/10.1107/S0907444909042073.

69. Winter G. 2010. xia2: an expert system for macromolecular crystallography data reduction. J Appl Crystallogr 43:186–190. https://doi.org/10.1107/S0021889809045701.

70. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. 2011. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 67:355–367. https://doi.org/10.1107/S0907444911001314.

71. Freire E, Mayorga OL, Straume M. 1990. Isothermal titration calorimetry. Anal Chem 62:950A–959A. https://doi.org/10.1021/ac00217a002.

72. Potterton L, McNicholas S, Krissinel E, Gruber J, Cowtan K, Emsley P, Murshudov GN, Cohen S, Perrakis A, Noble M. 2004. Developments in the CCP4 molecular-graphics project. Acta Crystallogr D Biol Crystallogr 60:2288–2294. https://doi.org/10.1107/S0907444904023716.

73. Collins RF, Saleem M, Derrick JP. 2007. Purification and three-dimensional electron microscopy structure of the Neisseria meningitidis type IV pilus biogenesis protein PilG. J Bacteriol 189:6389–6396. https://doi.org/10.1128/JB.00648-07.

74. Collins R, Karuppiah V, Siebert CA, Dajani R, Thistlethwaite A, Derrick JP. 2018. Structural cycle of the Thermus thermophilus PilF ATPase: the powering of type IVa pilus assembly. Sci Rep 8:14022. https://doi.org/10.1038/s41598-018-32218-3.

75. Mell JC, Redfield RJ. 2014. Natural competence and the evolution of DNA uptake specificity. J Bacteriol 196:1471–1483. https://doi.org/10.1128/JB.01293-13.