







CryoEM structure of the type IVa pilus secretin required for natural competence in *Vibrio cholerae*

Sara J. Weaver ^{1,5}, Davi R. Ortega ², Matthew H. Sazinsky ³, Triana N. Dalia ⁴, Ankur B. Dalia ⁴ & Grant J. Jensen ²✉

Natural transformation is the process by which bacteria take up genetic material from their environment and integrate it into their genome by homologous recombination. It represents one mode of horizontal gene transfer and contributes to the spread of traits like antibiotic resistance. In *Vibrio cholerae*, a type IVa pilus (T4aP) is thought to facilitate natural transformation by extending from the cell surface, binding to exogenous DNA, and retracting to thread this DNA through the outer membrane secretin, PilQ. Here, we use a functional tagged allele of VcPilQ purified from native *V. cholerae* cells to determine the cryoEM structure of the VcPilQ secretin in amphipol to ~ 2.7 Å. We use bioinformatics to examine the domain architecture and gene neighborhood of T4aP secretins in Proteobacteria in comparison with VcPilQ. This structure highlights differences in the architecture of the T4aP secretin from the type II and type III secretion system secretins. Based on our cryoEM structure, we design a series of mutants to reversibly regulate VcPilQ gate dynamics. These experiments support the idea of VcPilQ as a potential druggable target and provide insight into the channel that DNA likely traverses to promote the spread of antibiotic resistance via horizontal gene transfer by natural transformation.

¹Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 E. California Blvd, Pasadena, CA 91125, USA. ²Division of Biology and Biological Engineering and Howard Hughes Medical Institute, California Institute of Technology, 1200 E. California Blvd, Pasadena, CA 91125, USA.

³Department of Chemistry, Pomona College, 333N. College Way, Claremont, CA 91711, USA. ⁴Department of Biology, Indiana University, 107S. Indiana Avenue, Bloomington, IN 47405, USA. ⁵Present address: Howard Hughes Medical Institute, David Geffen School of Medicine, Departments of Biological Chemistry and Physiology, University of California Los Angeles, 615 Charles E Young Drive South, Los Angeles, CA 90095, USA. ✉email: jensen@caltech.edu

Horizontal gene transfer, or the ability of microorganisms to directly share DNA with one another, facilitates rapid evolution, can contribute to the development of antibiotic resistance, promotes the spread of virulence factors, and allows bacterial pathogens to rapidly evade host immune response¹. A clear understanding of the mechanisms of horizontal gene transfer can aid the development of tools in the fight against antibiotic resistance.

One mechanism of horizontal gene transfer is natural transformation, where a competent bacterium can take up DNA from its environment and then maintain this exogenous genetic material, either as a plasmid or by integrating it into its genome by homologous recombination². Many bacteria utilize a type IVa pilus (T4aP) nanomachine to take up genetic material^{3,4}. Intergenous transformation can result in the development of mosaic alleles that confer antibiotic resistance, and has been demonstrated in a variety of genera, including *Streptococcus*, *Neisseria*, and *Actinobacter*^{5,6}. Additionally, natural transformation of large regions of DNA can induce serotype switching in *Vibrio cholerae*^{7,8} and in *Streptococcus pneumoniae*⁹. Thus, preventing horizontal gene transfer represents a unique approach to mitigate the spread of antibiotic resistance and virulence in bacterial pathogens.

Here, we focus our attention to the structural biology of natural transformation in the gram-negative bacterium *V. cholerae*. *V. cholerae* is the causative agent of the diarrheal disease cholera¹⁰. Since 1817, cholera has spread globally in seven pandemics that each feature strains of distinct characteristics. Humans typically encounter members of the *Vibrionaceae* family through contaminated water or contaminated shellfish^{11,12}. *V. parahaemolyticus* and *V. vulnificus* are common causes of shellfish-borne illness^{13,14}. Type IV pilus (T4P) protein filaments are important in *Vibrionaceae* family pathogenicity because they mediate natural transformation, adhesion, biofilm formation, and colonization of their hosts^{15–20}. The genomes of *V. cholerae*, *V. parahaemolyticus*, and *V. vulnificus* all contain two T4aP systems: mannose-sensitive hemagglutinin (MSHA) pili and chitin-regulated T4aP used for natural competence²¹; the latter of which allows for chitin-induced natural transformation^{22–26}. Of these species, *V. cholerae* is the most genetically tractable and has emerged as a model system for studying natural transformation and bacterial pili.

Natural transformation in *V. cholerae* is tightly regulated, and is induced when these bacteria are grown on chitin, a biopolymer found in the exoskeletons of crustaceans, in their aquatic environment²³. Chitin indirectly promotes expression of the master regulator of competence TfoX^{27,28}, which, in turn, induces expression of the T4aP needed for DNA uptake^{23,29,30}. The T4aP machinery requires four elements: an inner membrane pilus assembly complex, cytoplasmic motors to extend and retract the pilus, an outer membrane pore (the secretin), and the pilin subunits that compose the pilus itself⁴. The T4aP facilitates environmental DNA uptake by extending and retracting from the cell surface through a large, outer membrane secretin pore called PilQ^{30,31}. To mediate DNA uptake, the pilus likely translocates DNA across the membrane through the PilQ secretin⁴. Thus, the PilQ secretin represents a potential target to thwart this mechanism of horizontal gene transfer.

The *V. cholerae* T4aP secretin PilQ is a member of the bacterial secretin superfamily^{32,33}. Secretins are found in the type II secretion system (T2SS), the T3SS, the T4P machine, and filamentous phage³⁴. Bacterial secretins are united by a common C-terminal secretin domain, which oligomerizes to form a large pore in the outer membrane³⁵. While the C-terminal secretin domain is remarkably similar across these secretion systems, the N-terminal region varies, which may be related to the

specialization of different secretins³⁶. The T2SS exports periplasmic folded proteins to the extracellular space. The T3SS uses a “needle and syringe” to export cytosolic effector proteins outside of the cell, or directly into another cell. The T4P have a broad range of functionality, including DNA uptake, twitching motility, biofilm formation, and adhesion^{15,16,30,37}. During natural transformation, the T4aP extends and retracts a filament to take up DNA cargo.

The structure of various T4P secretins has been examined^{38–54}, including structures of PilQ from *Neisseria meningitidis*⁴⁵, *Pseudomonas aeruginosa*⁴⁶, and *Thermus thermophilus*⁴⁷, although no work has provided high-resolution details sufficient to model the passage of DNA or to design inhibitors of this potential drug target. Thus, structural information about the T4P secretins has mainly been inferred from the related but distinct T2SS and T3SS secretins^{55–61}. While this paper was under review, a 4.3 Å structure of the T4P PilQ from *P. aeruginosa* was reported⁶².

Here, we present the structure of the *V. cholerae* T4aP secretin PilQ to ~2.7 Å using a fully functional, His-tagged allele that we express and purify from the native bacterium. We perform sequence analysis on the Proteobacteria T4aP secretins to put this structure into context. Our work highlights differences between T4aP, T2SS, and T3SS secretins, and emphasizes the need for structures of different secretin family members. In particular, we discuss differences and remaining puzzles, including how the pilus could be accommodated within VcPilQ during natural transformation and what part/s of the secretin if any penetrate the outer membrane. Finally, we report structure-inspired designs of cysteine pair mutants that allowed us to reversibly inhibit pilus assembly and natural transformation, presumably by sealing the secretin gate. These experiments support the designation of VcPilQ as a druggable target, and more broadly demonstrate how cysteine pair mutations can be employed to study the activity of bacterial secretins.

Results

Single-particle cryoEM of the T4aP secretin PilQ. To ensure properly folded and fully functional T4aP machinery, we chose to purify PilQ from *V. cholerae* rather than a recombinant system. A chromosomal mutation was made to add a deca-histidine tag to the N-terminus of PilQ. Here, expression of the T4aP system was induced via ectopic expression of TfoX using an IPTG-regulated promoter (*P_{tac-tfoX}*)^{63,64}. The bacteria expressing His-tagged VcPilQ retained wild-type levels of natural transformation (Supplementary Fig. 1a). Previous work demonstrated that similar N-terminal tags allowed for functional T4aP activity³¹.

Secretins are detergent- and heat-resistant multimers^{65,66}. Coomassie staining and western blotting of VcPilQ in amphipol run under denaturing conditions on a SDS Page gel showed monomer, the multimer, and some low molecular weight species in the purified sample (Supplementary Fig. 1b). The size difference between these contaminants (<100 kDa) and the VcPilQ multimer (~860 kDa) made it easy to distinguish VcPilQ from the milieu in electron micrographs (Supplementary Fig. 2a).

Here we report the high-resolution structure of the purified T4aP secretin VcPilQ by single-particle cryoEM (Fig. 1a and Table 1). The cryoEM data processing steps are summarized as a flow chart in Supplementary Fig. 3. The cryoEM structure (C14 symmetry, overall resolution of 2.7 Å at Fourier shell correlation (FSC) of 0.143, and 3 Å at FSC of 0.5) reached sufficient, isotropic resolution to recognize and model residues 160–571 of VcPilQ, which includes the N0, N3, and secretin domains (Fig. 1a, Supplementary Fig. 4, and Supplementary Movie 1)^{67,68}. Mass spectrometry analysis demonstrated 65% sequence coverage, with fragments identified in each domain of

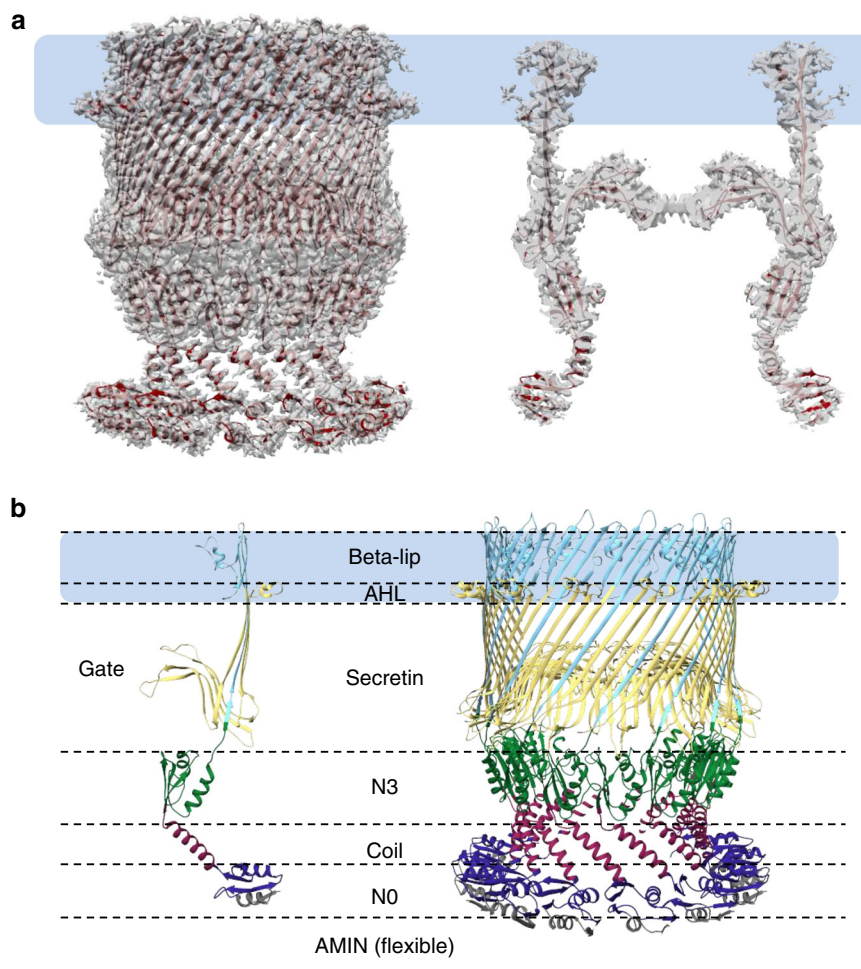


Fig. 1 CryoEM structure of the *V. cholerae* type IV competence pilus secretin PilQ. **a** Model (cartoon in dark red) and cryoEM density (dark gray transparent) of symmetrized *Vibrio cholerae* PilQ (VcPilQ) is shown from the side (left) and cut through the center (right). The putative outer membrane region is represented in a light blue rectangle. **b** The atomic model of the VcPilQ multimer is shown (right). One chain is shown by itself (left). The putative outer membrane region is represented in a light blue rectangle. Dashed lines represent the different domains of the VcPilQ structure: AMIN (not shown), N0 (gray/dark blue), coil (purple), N3 (green), secretin (yellow), beta lip (sky blue). The gate region is labeled on the monomer.

the folded protein from residues 50–567 (Supplementary Fig. 1c). Residues 50–159 were represented in the mass spectrometry analysis (Supplementary Fig. 1c), but were not clearly resolved in the cryoEM structure. A hazy density is seen in some 2D classes (Supplementary Fig. 2c, asterisk) in the region residues 50–159 would likely be present. These residues contain an AMIN Pfam domain, which is made up of two antiparallel beta sheets and is thought to bind peptidoglycan^{69,70}.

In each VcPilQ monomer, four beta strands come together to form a beta sheet (Fig. 1a). Once assembled, VcPilQ forms a 56-strand beta barrel. Inside the barrel, two further beta hairpins (beta strand-turn-beta strand each) form a gate. These regions match the topology of other secretin structures (Fig. 2a–c). In each case, the root-mean-square deviation (RMSD) calculated over the protein domains present in both structures (N3, secretin, and beta lip domains) was about 1 Å (Supplementary Fig. 5a–e). The outer membrane regions differ significantly, however, in both angle and membrane spanning distance, as discussed in more detail below.

Only one of the seven T2SS secretins and one of the T3SS secretins published structures resolve the N0 domain^{59,61}. Here we can resolve the N0 domain (residues 160–227) of VcPilQ to 4–7 Å local resolution (Supplementary Fig. 4e–h), which allowed us to build a model based on homology to previously solved

structures of N0 domains (the cross-linked *K. oxytoca* PulD⁵⁹ and *Salmonella typhimurium* InvG⁶¹), plus a crystal structure of an isolated N0 domain from *N. meningitidis* PilQ (4AR0)⁴⁵.

In VcPilQ, a 32 Å alpha helix follows the N0 domain (Fig. 1b). None of the T2SS or T3SS structures contain a helical coil to link N-terminal domains; instead their periplasmic protein domains are linked by unstructured loops (Fig. 2). Following the end of this helix, the protein chain changes direction (~104° angle) as the coil flows into the N3 domain (Fig. 1b). This dramatically reduces the channel diameter, from 90 Å at the bottom of the N0 domain to 60 Å across the N3 domain (Supplementary Fig. 7a). In the T2SS and T3SS structures, the diameter of the channel is relatively constant.

While their structures are similar, the electrostatic characteristics of the inner surfaces of the T4P, T2SS, and T3SS secretins vary (Fig. 2d). The inner beta lip, N3, and coil regions of VcPilQ (Fig. 1b) are all negatively charged (Fig. 2d). The gate region of VcPilQ is also negatively charged (Supplementary Fig. 5f, g). In contrast, in the *V. cholerae* GspD and *E. coli* K12 GspD structures, there are alternating negatively and weakly positively charged regions (Fig. 2d).

The putative outer membrane region of VcPilQ is thicker than T2SS secretins. The secretin amphipathic helix lip (AHL) is

Table 1 Summary of single-particle data collection, 3D reconstruction, and model refinement.

Imaging parameters and 3D reconstruction	
Acceleration voltage (kV)	300
Magnification (×)	81,000
Pixel size (Å)	1.104
Frame rate (s ⁻¹)	0.092
Exposure time (s)	3.7
Total exposure (e ⁻ /Å)	60
Particles	
Micrographs used for selection	2,510
Defocus range (µm)	-0.5 to -3.5
Windowed	252,319
In final 3D reconstruction	100,543
Resolution	
“Gold-standard” at FSC 0.5 (Å)	3.0 Å
“Gold-standard” at FSC 0.143 (Å)	2.7 Å
Map-sharpening B factor (Å ²)	-69
Model refinement	
Resolution in phenix.real_space_refine (Å)	3.0
Model-to-map fit (CC_mask)	0.745
Number of atoms/residues/molecules	
NCS restrained chains	14
Protein atoms, residues (per chain)	43,736, 412
Ramachandran angles (%)	
Favored	92.21
Allowed	7.79
Outliers	0
r.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	0.847
Molprobit	
Score	2.81
Clashscore	10.53
Rotamer outliers (%)	10.91
EMRinger score	3.13

thought to be a key determinant for membrane insertion and acts as the lower boundary of the outer membrane region, but the exact location of the upper boundary is less clear^{34,56,60,61}. Recently, Ghosal et al.⁷¹ solved the in situ structure of the *Legionella pneumophila* T2SS by sub-tomogram averaging and demonstrated that the putative outer membrane thickness reported in the single-particle cryoEM structures of detergent-solubilized T2SS secretins (~2.5 nm) is significantly smaller than real membranes (5–7 nm)⁷¹. In our VcPilQ atomic model, the distances between the bottom of the AHL to the top of the beta strands is about 3 nm, which is substantially taller than the same region in the previously published T2SS structures (Fig. 2c).

To investigate more than just the residue locations, we generated an inverted mask based on the atomic model density of VcPilQ and multiplied by the empirical, unsharpened cryoEM density to remove density accounted for by the atomic model (“Methods” and Supplementary Fig. 2e). The product reflects unmodeled density in the cryoEM map that is not accounted for by the atomic model (Supplementary Fig. 2e). VcPilQ was solubilized in n-Dodecyl-b-d-Maltoside (DDM) and then exchanged into amphipol, so the unaccounted for density in our structure could be some combination of DDM, amphipol, protein, lipid, and noise. Rather than select one cryoEM threshold to visualize the membrane thickness at, we chose to measure the membrane thickness as a series of thresholds so that a range (~25–50 Å) could be reported. This unmodeled density (in gray) blooms around the putative outer membrane region of the protein (Supplementary Fig. 2e). This density appears both on the outside of the beta barrel, and inside, coating the inner lip of

VcPilQ. It remains unclear precisely which residues of VcPilQ are embedded within the outer membrane in situ and whether the membrane spanning thickness is sufficient to fully penetrate the outer membrane (see “Discussion”).

Cysteine pair mutants reversibly inhibit natural transformation and surface piliation. VcPilQ is thought to mediate DNA uptake during natural transformation in *V. cholerae*³⁰. *V. cholerae* can undergo horizontal gene transfer in chitin biofilms, which can promote the spread of antibiotic resistance genes and virulence factors². Blocking DNA uptake by locking the VcPilQ gate with a small molecule could prevent this spread of genetic material. To provide a proof of concept, we designed cysteine pair mutants to reversibly lock the gate with disulfide bonds (Fig. 3a). The structure of VcPilQ was analyzed using the Disulfide by Design 2.0 web tool^{72–74} to identify residue pairs with geometries that could support a disulfide, and the top hits were analyzed in UCSF Chimera. The S448C/S453C pair is in the proximal hairpin of the gate, likely cross-linking adjacent VcPilQ monomers (Fig. 3b). The L445C/T493C pair likely cross-links the upper and lower gate hairpins of a single VcPilQ monomer (Fig. 3c). From the design, we hypothesized that the disulfide bonds would be exposed to the extracellular space.

To test if disulfide bonds could form in the gate region under normal culturing conditions, cysteine pair mutant strains were generated (Supplementary Tables 1 and 2) for natural transformation assays. Transformation efficiency was normalized by comparison to the parent strain, which expresses His-tagged VcPilQ (Fig. 3d). The dashed horizontal line marks equivalent efficiency to the parent strain. High levels of the reducing agent dithiothreitol (DTT) were toxic to wild-type *V. cholerae* cells and inhibited natural transformation, so subinhibitory concentrations of DTT were used that allow the His-tagged VcPilQ parent strain to perform natural transformation (Supplementary Fig. 6). Under oxidizing conditions, both cysteine pair mutants demonstrated lower transformation efficiency than the wild-type His-tagged PilQ parent strain (Fig. 3d). The transformation efficiency is not completely ablated in the cysteine pair mutants, which suggests that disulfide bond formation may not be 100% efficient. In the presence of reducing agent, the transformation efficiency of the cysteine pair mutants recovered transformation efficiency similar to the control parent strain (Fig. 3d). This suggests that the two cysteine pair mutants were able to assemble into functional T4aP.

To investigate if the disulfide bonds affect surface piliation, we used strains where the major pilin contains a mutation (*pilA* S67C —aka PilA-Cys) that allows for subsequent labeling with fluorescently conjugated maleimide dyes^{31,75}. The *V. cholerae* T4aP used in competence are highly dynamic (much higher than that described for many other pilus systems), such that within a snapshot, very few cells will have surface exposed pili³¹. As a result, deletion of the retraction ATPase *pilT* results in a hyperpiliated phenotype and provides a more sensitive readout for pilus assembly^{31,75}. Thus, surface piliation was qualitatively assessed in cysteine pair mutants by fluorescence microscopy under several concentrations of reducing agent using strain backgrounds containing *pilA*-Cys and Δ *pilT* (Supplementary Table 1)^{31,75,76}. The cysteine pair mutants exhibited no surface pili under oxidizing conditions (0 mM DTT, Fig. 3e) compared to the parent, which is consistent with pili not being assembled under these conditions. Conversely, piliation was comparable to the parent under reducing conditions (Fig. 3e–g). The S448C/S453C mutant is slightly more recalcitrant to DTT rescue, which may be due to the fact that these cysteines are located further down in the gate region making them less accessible to the reducing agent.

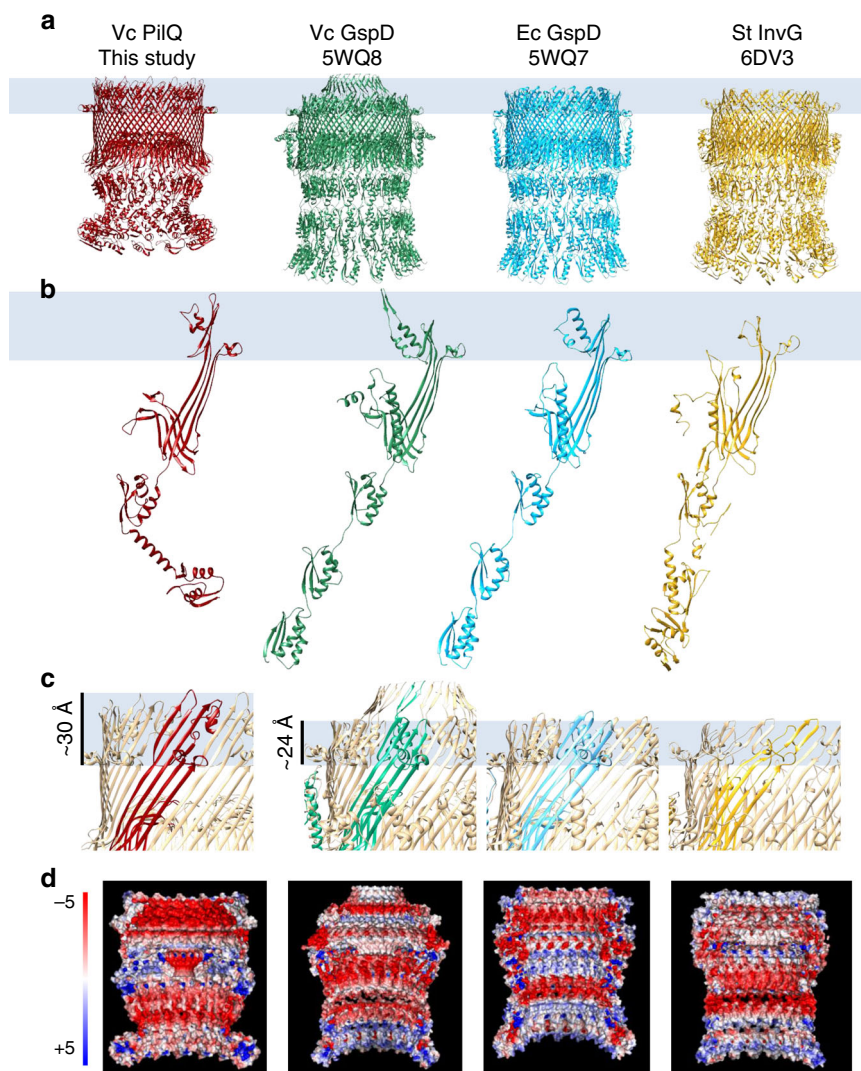


Fig. 2 Comparison of VcPilQ to T2SS and T3SS secretins. The structures of VcPilQ (this study, dark red), *V. cholerae* GspD (5WQ8) (green), *E. coli* K12 GspD (5WQ7) (blue), and *S. typhimurium* InvG (6DV3) (yellow) are compared^{55, 61}. The putative outer membrane location is depicted as a blue rectangle. The structures are shown as multimers (a) or monomers (b). c The multimer is shown in tan with one subunit colored. The putative outer membrane region is highlighted to show that VcPilQ (left, red) has a ~30 Å membrane spanning distance, while VcGspD, EcGspD, and StInvG have about a 24 Å membrane spanning distance. d Adaptive Poisson-Boltzmann solver was used to calculate the electrostatic potential calculation of each secretin in PyMol¹⁴². The inner cavity of each secretin is shown. The scale varies from -5 (red) to +5 (blue) in units of K_bT/e_c .

Diversity of T4aP secretin domain architecture in Proteobacteria. Finally, we sought to contextualize VcPilQ within T4aP secretins. T4P function in motility, communication, surface sensing, and DNA uptake¹⁶. We wondered if the functional diversity of T4aP is reflected in the architecture of secretins. We compared the domain architecture and gene neighborhood of a nonredundant set of 197 representative sequences of T4aP secretins collected from Proteobacteria. We also used phylogenetic inference of the secretin to cluster the sequences into evolutionarily close groups (Fig. 4a). Secretins do not provide a strong phylogenetic signal, which translates into a poorly resolved inference⁷⁷. Next we mapped the domain architecture of the secretins onto the phylogenetic tree (Fig. 4b). In this set of secretins, we found that the dominant domain architecture can be expressed by a variable number of AMIN domains, followed by a N0 (PFAM family STN) domain, a N3 (Secretin_N) domain, and a secretin domain. The number of AMIN domains includes none (27 sequences), one (19 sequences), two (141 sequences), or three (10 sequences) AMIN domain repeats (Supplementary Data 1).

T4aP have been implicated in natural transformation in members of the *Haemophilus*, *Moraxella*, *Neisseria*, *Pseudomonas*, and *Vibrio* genera, but many genera have not been tested⁴. In *Neisseria gonorrhoeae*, the gonococcal pilus (a T4aP) mediates adhesion, twitching motility, and competence, whereas in *P. aeruginosa*, the PAK pilus (a T4aP) is used for adhesion and motility, but not competence^{78,79}. In *V. cholerae*, the T4aP used for competence may play a role in adhesion and kin recognition, but does not promote motility^{30,37}. We mapped the *Haemophilus*, *Moraxella*, *Neisseria*, *Pseudomonas*, and *Vibrio* genera on our phylogenetic tree and observed that the ability to perform natural competence is not correlated to a particular architecture of T4aP secretin, at least in Proteobacteria (Fig. 4a and Supplementary Data 1).

Finally, we used GeneHood to identify homologs of the proteins located in the genome neighborhood of *pilQ* (Supplementary Data 1). Essentially all of the secretins of the data set were found within the complete *pilMNOPQ* cluster, with a few exceptions. Consistent with literature precedent^{80–86}, we found that the *pil* cluster is flanked by two genes involved in the Shikimate pathway

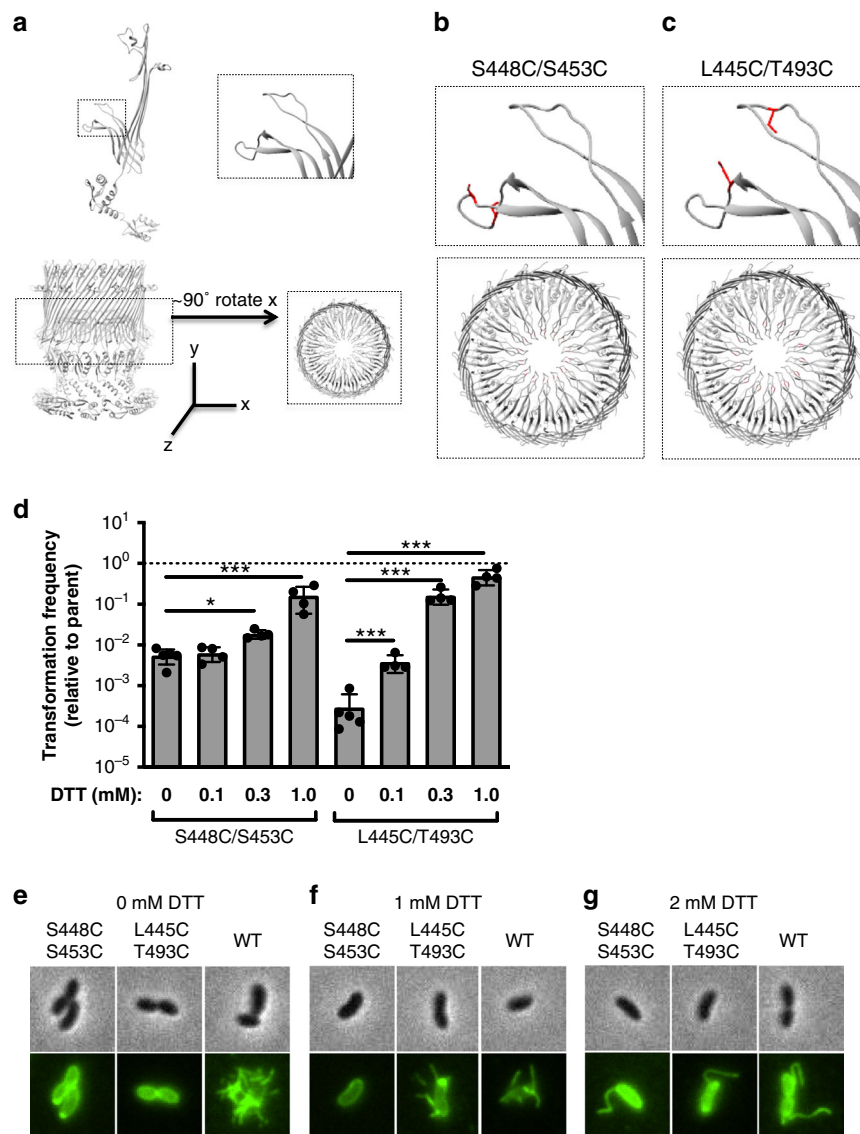


Fig. 3 Cysteine pair mutants lock the gate, reduce piliation, and reversibly inhibit transformation. **a** top: the atomic model of VcPilQ is shown as one chain. The upper and lower beta hairpins of the gate are highlighted in a dashed line box (left) and magnified (right). Bottom: the atomic model of VcPilQ is shown as a multimer and a dashed line rectangle highlights the gate region (left), which is magnified (right) after a 90° rotation about the x-axis. The coordinate system is shown. The magnified regions from **(a)** are shown for **(b)** the VcPilQ S448C/S453C mutant, which locks the lower gate and for **(c)** the VcPilQ L445C/T493C mutant, which links the upper and lower gates. In **(b, c)** the residues mutated to cysteine are shown in red. **d** The transformation frequency of two cysteine pair VcPilQ mutants (left: S448C/S453C (Strain TND2169) and right: L445C/T493C (Strain TND2170)) is plotted. Data are normalized to the parental strain that expresses wild-type VcPilQ (Strain TND2140). Natural transformation assays were performed in the presence of varying concentrations of dithiothreitol (DTT) (0–1.0 mM). Data are from at least four independent biological replicates and shown as the mean ± standard deviation. The dashed line indicates the transformation frequency expected if mutants are equivalent to the parental strain expressing wild-type VcPilQ. Statistical comparisons were made by one way ANOVA with Tukey's posttest. * $p < 0.05$, *** $p < 0.001$. The raw data are available in Supplementary Fig. 6. Representative phase contrast (top) and epifluorescence images (bottom) of *V. cholerae* in the hyperpiliated background expressing wild-type (WT) VcPilQ (Strain TND2244), or a cysteine mutant pair (S448C/S453C (Strain TND2242) or L445C/T493C (Strain TND2243)) grown in the presence of **e** 0 mM, **f** 1 mM, or **g** 2 mM dithiothreitol (DTT) prior to labeling with AlexaFluor 488-maleimide to visualize bacterial pili. More than 200 cells were imaged per condition and representative images of cells with pili are shown. In conditions where no cells analyzed exhibited an external pilus, a representative image of a non-piliated cell is shown.

(*aroK* and *aroB*) downstream, and the penicillin-binding proteins 1 A (*mrcA*, *ponA*) which is involved in cell wall formation⁸⁷ upstream of it although in a different strand.

Discussion

Here we present the high-resolution structure of a bacterial T4aP secretin. This protein complex facilitates DNA uptake into diverse

bacterial species to aid in their evolution. The *V. cholerae* T4aP is a model system to study natural transformation in bacteria. We observed key differences in the outer membrane region and the periplasmic region among the different members of the secretin family. These differences emphasize the weakness in relying on homology models of evolutionarily related secretins, like the T2SS secretin GspD, to understand VcPilQ. We also discuss the domain architecture of VcPilQ in the context of related T4aP

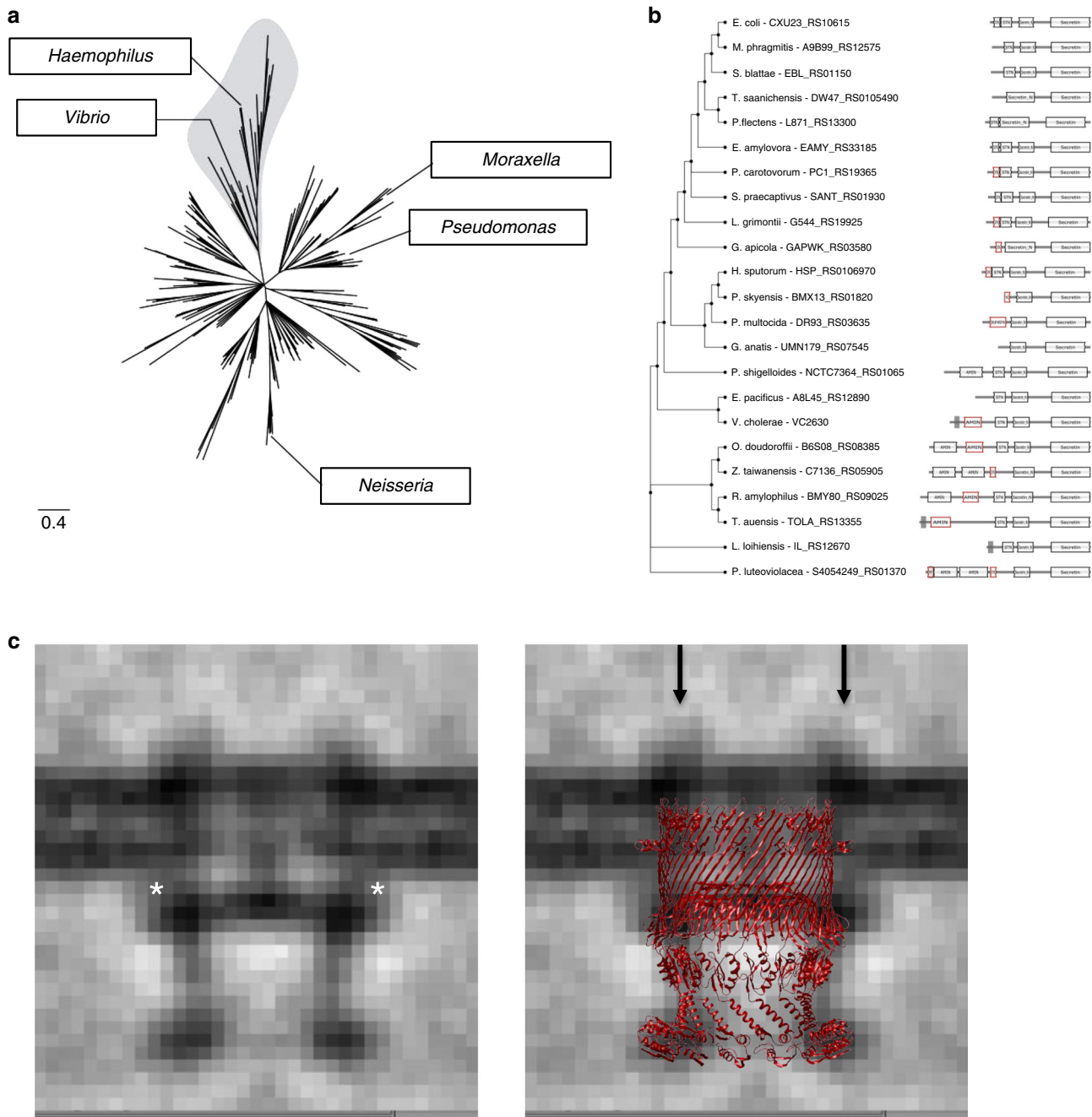


Fig. 4 The domain architecture diversity of T4aP secretins and a comparison of VcPilQ with the sub-tomogram average of *M. xanthus* type IVa pilus machinery. **a** Phylogenetic inference of the secretin domain in a representative set of 197 secretins in Proteobacteria. The shaded region is a fully resolved part of the tree that includes the VcPilQ and is displayed in details in panel **(b)**. The locations of representatives of the *Haemophilus*, *Moraxella*, *Myxococcus*, *Neisseria*, *Pseudomonas*, and *Vibrio* genera are indicated. **b** Detail of the branch that includes the *V. cholerae* (Vc) secretin with domain architecture for each representative sequence. The domains were identified using Comprehensive Domain Visualization Tool (CD-VIST) pipeline by first executing a HMMER search (a hidden Markov model package) (black) and subsequently by a HHblits (HMM-HMM-based lightning-fast iterative sequence search) search (red), while transmembrane regions were predicted using the TMHMM (transmembrane hidden Markov model) package (gray region)¹⁴¹. The tree and the gene neighborhood information is available in Supplementary Data 1–3. **c** VcPilQ is compared to the non-piliated sub-tomogram average of the type IVa pilus machinery in the Δ pilP *Myxococcus xanthus* strain (EMD-3249)⁴⁹. In the Δ pilP strain, only PilQ and TsaP are present in the complex. The putative TsaP location is marked with an asterisk in the left image. In the right image, the structure of VcPilQ (dark red) is aligned to the sub-tomogram average. Arrows indicate unidentified extracellular densities.

secretins. We designed cysteine pair mutants to reversibly seal the VcPilQ gate and inhibit natural transformation, which can be used as a tool to further investigate the function of the T4aP machinery in situ. We also compare our structure to previous T4aP sub-tomogram averaging results in *Myxococcus xanthus*

which call in question which parts if any of PilQ extend across the outer membrane. Together, these results elucidate the structure of PilQ and provide a foundation for future mechanistic studies.

In some previous secretin structures, the trans-outer-membrane region appears to be only 2–3 nm thick (Fig. 2c),

which is much thinner than a typical membrane^{55,61,71}. In *Thermus thermophilus*, the cryoEM structure of the secretin (TtPilQ) contained a “crown domain” that was unaccounted for by the sequence of TtPilQ⁴⁷. This has left it unclear which regions of secretin molecules are actually embedded in the outer membrane, and whether any residues are exposed to the extracellular surface. Based on the atomic model, we estimate the putative outer membrane region in VcPilQ is at least 3 nm thick and we wondered how this compares to real membranes. The in situ structure of VcPilQ is not available, but an in situ structure of the related *M. xanthus* T4aP (MxT4aP) has been solved by sub-tomogram averaging⁴⁹. We docked VcPilQ into a sub-tomogram average of the MxT4aP secretin (EMD-3249) (Fig. 4c). The gate region of VcPilQ nicely superimposes onto the gate of the *M. xanthus* secretin, which suggests the VcPilQ structure was positioned correctly. However, the putative outer membrane region in VcPilQ does not extend across the entire outer membrane of the MxT4aP sub-tomogram averages. The shortness of the other predicted transmembrane domains of secretins (Fig. 2c) and the two bumps outside the outer membrane in the sub-tomogram average directly above the secretin barrel (arrows in Fig. 4c) call into question whether and how parts of PilQ may penetrate the outer membrane, and if any other accessory proteins are also present in the membrane in situ.

The T4aP machinery extends and retracts the T4aP through the PilQ outer membrane pore^{30,88–90}. It remains controversial if the DNA would associate only with the tip of the pilus^{31,91} or along the length of the pilus^{92,93}. The DNA may accompany the T4P through the secretin pore to enter the periplasm, or an alternative mechanism is possible^{4,30,38,94,95}. Several T4aP structures have been solved by cryoEM and reveal pili diameters of 50–80 Å^{93,96–98}. In contrast, the inner chamber of VcPilQ varies in diameter from 25 to 108 Å, and docking the *E. coli* T4aP cryoEM structure⁹⁸ into VcPilQ reveals steric clashes at the gate (Supplementary Fig. 7). In Yan et al.⁵⁵, a glycine in the VcGspD gate (G453) was identified as a putative hinge point to facilitate gate opening, and showed that the G453A mutant trapped the T2SS secretin in a partially open state. In our VcPilQ structure, the inner channel distance between the corresponding glycine (G439) alpha carbons is about 8 nm (Fig. 4a). Thus, we hypothesize that a gate hinge mechanism could accommodate pili up to 7 nm in diameter. In an open state the gate loops could flip up toward the extracellular space to accommodate the pilus. Additionally, the inner surface of VcPilQ is negatively charged (Fig. 2d and Supplementary Fig. 5f, g). In contrast, the inner surface of the T2SS shows alternating negative and positive electrostatic character (Fig. 2d). These differences may relate to the function of VcPilQ in natural competence: the DNA cargo of the T4aP is also negatively charged, so it is possible that this electrostatic repulsion will help the cargo pass through the cavity, rather than getting stuck. By comparison, for T2SS secretins, the charge alternates in the inner cavity, which could reflect its broad scope of cargo.

Our analysis of putative T4aP secretins in Proteobacteria shows a relatively conserved domain architecture (Supplementary Data 1) that is consistent with literature precedent^{35,70,99,100}. The model used to identify T4aP secretin sequences was previously validated by the Rocha lab⁷⁷, but it could be biased by our current understanding of T4aP gene cluster organization. Perhaps the conservation of the *pilMNOPQ* cluster is linked to a specific domain architecture of secretin. Notably, while the majority of the T4aP secretin sequences contained the N0 (STN), N3 (Secretin_N), and secretin domains, we observed variation in the presence and abundance of AMIN domain repeats (Supplementary Data 1). Looking specifically in the well-resolved branch containing the VcPilQ, we note the gradual loss of AMIN domain over the evolution of secretin (Fig. 4b). Several peptidoglycan-

binding domains are represented in T4P machinery proteins, including the AMIN domain⁷⁰ and the LysM domain^{101,102}, and may play roles in protein localization and stability¹⁰³. The loss of the AMIN domain in some secretins merits further investigation. Thus, we conclude that the variations we observed in domain architecture, in particular the number of AMIN domains in the N-terminus of the secretin, do not determine functional variability, which is expected. Alternatively, if there is a specialization of the secretin sequences toward function it might occur at the amino-acid level scale.

Methods

Bacterial strains and culture conditions. All *V. cholerae* strains were derived from the El Tor strain E7946¹⁰⁴. Strains were routinely grown in LB Miller broth and agar. Media were supplemented with kanamycin (50 µg/mL), spectinomycin (200 µg/mL), and/or erythromycin (10 µg/mL) as appropriate. All strains were generated by natural transformation as described below in “Natural transformation assays”^{64,105}. Mutant constructs were generated by splicing-by-overlap extension PCR to stitch the upstream region of homology (UP arm; amplified with F1/R1), the mutational cargo (MIDDLE arm), and the downstream region of homology (DOWN arm; amplified with F2/R2) together. The full descriptions of the strains are available in Supplementary Table 1 and the primers used are in Supplementary Table 2. The following systems were knocked out of all *V. cholerae* strains used in this study: the MSHA pilus (ΔMSHA::CarbR), the toxin co-regulated pilus (TCP) (ΔTCP::ZeoR), and the cholera toxin (ΔCTX::KanR). Also, all strains used in this study contained mutations to constitutively activate natural competence³¹; the quorum sensing regulatory protein LuxO was knocked out (ΔluxO) and expression of the master regulator of competence, TfoX, was placed under control of the IPTG-inducible Ptac promoter (Ptac-tfoX)^{64,104–106}. A deca-histidine tag was added to the N-terminus of PilQ at the native locus to facilitate protein purification. The full genotype of this strain (TND1751) is 10xHis-PilQ, ΔVC1807::SpecR, lacZ::lacIq, comEA-mCherry, ΔluxO, Ptac-tfoX, ΔTCP::ZeoR, ΔMSHA::CarbR, ΔCTX::KanR.

For purification of PilQ, *V. cholerae* expressing His-tagged PilQ (Strain TND1751) were streaked on Luria Broth agar plates and grown overnight at 30 °C. Small cultures were seeded (5 mL) and grown overnight at 30 °C. The next day, 500 mL cultures were seeded with the 5 mL culture. LB broth was supplemented with 20 mM MgCl₂, 10 mM CaCl₂, and 100 µM IPTG to induce expression of TfoX and therefore the T4aP system. The large cultures were grown overnight at 30 °C in beveled flasks. The following day, cultures were spun down (4790 × g, 4 °C, 20 min) and the cell paste was weighed, aliquoted, and stored at –80 °C.

For testing the impact of cysteine pair mutants on natural transformation, *V. cholerae* strains were generated where the native copy of PilQ was deleted and the corresponding PilQ allele was expressed at a chromosomally integrated ectopic site under the control of an arabinose-inducible Pbad promoter¹⁰⁷. The full genotype of the parent strain (TND2140) was ΔlacZ::Pbad-10XHis-PilQ CmR, ΔpilQ::TetR, ΔCTX::KanR, ΔMSHA::CarbR, ΔluxO, ΔTCP::ZeoR, comEA-mCherry, Ptac-tfoX. The cysteine pair mutants were isogenic other than the cysteine mutations introduced into the Pbad-10XHis-PilQ construct, which were PilQ S448C S453C (TND2169) and PilQ L445C T493C (TND2170).

To test the impact of cysteine pair mutants on pilus biogenesis, *V. cholerae* strains were generated akin to those described above, with the exception that the retraction ATPase PilT was deleted^{23,30} and the strains contained a cysteine substitution mutation in the major pilin (pilA S67C) that allows for pilus labeling with AlexaFluor 488-maleimide dye³¹. The full genotype of the parent strain (TND2244) was ΔlacZ::Pbad-10XHis-PilQ CmR, ΔpilT::TmR, ΔpilQ::TetR, ΔCTX::KanR, ΔMSHA::CarbR, ΔluxO, ΔTCP::ZeoR, pilA S67C, comEA-mCherry, Ptac-tfoX. The cysteine pair mutants were isogenic other than the cysteine mutations introduced into the Pbad-10XHis-PilQ construct, which were PilQ S448C S453C (TND2242) and PilQ L445C T493C (TND2243).

Natural transformation assays. Chitin-independent natural transformation assays were performed³¹. Briefly, competence was induced in late-log phase *V. cholerae* cells as described above. The cells were resuspended in Instant Ocean medium (7 g l⁻¹; Aquarium Systems) and incubated with or without DNA (~500 ng) at 30 °C for 5 h. After the incubation period, cells were shaken with additional LB (1 mL) at 37 °C for 2 h and plated in the presence or absence of antibiotics, and the number of colonies was assessed the following day to calculate transformation frequency. For reactions where strains harbored Pbad-10XHis-PilQ constructs, arabinose was added to a final concentration of 0.2%. Where indicated, DTT was added at the indicated concentrations throughout the assay.

Competence pilus labeling and microscopy. Cells were labeled with AlexaFluor 488-maleimide dye and imaged to visualize competence pili^{31,75}. Competence was induced in late-log phase *V. cholerae* cells as described earlier in methods. Cells were labeled with AlexaFluor 488-maleimide dye (25 µg/mL) in Instant Ocean with 20 mM MgCl₂ and 10 mM CaCl₂, and then imaged using a Nikon Ti-2 microscope

at $\times 60$ magnification. All strains were grown with arabinose added to a final concentration of 0.2% to induce expression of the Pbad-10XHis-PilQ construct. Where indicated, cells were grown in the presence of the indicated concentration of DTT prior to labeling. More than 200 cells were imaged per condition, and representative images are reported.

Purification. Cell pellet (15 g) was resuspended in lysis buffer (50 mM Tris HCl, pH 8, 300 mM NaCl, 1% DDM, 20 mM imidazole) supplemented with lysozyme (40 mg/mL in 50% glycerol/water), DNase I (4 mg/mL in 50% glycerol/water), and EDTA-free Protease Inhibitor tablet (Roche, 11697498001). Lysis proceeded with stirring at 4 °C for 20 h. Lysate was clarified by ultracentrifugation (Beckman L8-M ultracentrifuge, Rotor Type 45 Ti, 50,000 \times g, 1 h). The supernatant was mixed with Ni NTA agarose beads (Anatrace, SUPER-NINTA25) and incubated with stirring (4 °C, 8 h). In a gravity column at 4 °C, proteins conjugated to Ni NTA agarose beads were washed (50 mM Tris HCl, pH 8, 300 mM NaCl, 0.05% DDM, 70 mM imidazole), (50 mM Tris HCl, pH 8, 300 mM NaCl, 0.05% DDM, 300 mM imidazole), and eluted (50 mM Tris HCl, pH 8, 300 mM NaCl, 0.05% DDM, 1 M imidazole). Eluant was concentrated to ~ 1 mg/L (EMD Millipore Amicon Ultra-15, 30 kDa cutoff, UFC903024). Concentrated PilQ (150 μ L of ~ 1 mg/mL protein) was exchanged into Amphipol A8-35 (0.585 mg for a 3:1 ratio, Anatrace, A835) and allowed to incubate at 4 °C for 1 h. Excess DDM was removed using Bio-Beads SM2 (Bio-Rad, 1523920) by incubating overnight at 4 °C. The protein was concentrated. Protein was analyzed on Bio-Rad Any kD™ Mini-PROTEAN+™ TGX Stain-Free™ Protein Gels (Bio-Rad, 4568126) by stain-free exposure, Coomassie staining, or western blot with 6x-His Tag Monoclonal Antibody (HIS.H8), HRP (Invitrogen, MA1-21315-HRP).

Electron microscopy. For cryoEM, Quantifoil R2/2 300 Mesh grids (EMS, Q33100CR2) were glow discharged (Pelco EasiGlow, 20 mA, 60 s). PilQ in amphipol (3 μ L of ~ 0.8 mg/L) was frozen on a Mark IV Vitrobot (FEI, 20 °C, 100% relative humidity, blot force -6 , blot time 4 s). Micrographs were collected on a 300 kV Titan Krios microscope (FEI) with energy filter (Gatan) and equipped with a K3 direct electron detector (Gatan). Data were collected using Serial EM software with a pixel size of 1.104 Å ($\times 81,000$ magnification) and a defocus range from -1.0 to -3.0 μ m¹⁰⁸. A fluence of 19.8 electrons/pixel/second was used with a 3.7 s exposure time to collect 60 e⁻/Å².

Image processing. The cryoEM image processing workflow is summarized in Supplementary Fig. 3. MotionCor2 was used for motion correction and dose weighting of 3808 movies¹⁰⁹. Contrast transfer function (CTF) correction was used to evaluate micrograph quality¹¹⁰. CryoSPARC blob picking on 2510 micrographs yielded 3,100,353 potential particles¹¹¹. After inspection, the 252,319 particles were analyzed by several rounds of 2D classification and 3D classification to yield 100,543 particles. These particles were moved to Relion using the UCSF PyEM package script (<https://github.com/asarnow/pyem/>)¹¹². In Relion, several rounds of 3D refinement, polishing, and CTF refinement were used^{113–115}. ResMap was used to calculate local resolution⁶⁸.

Model building and refinement. The initial model (residues 230–571) was auto-built using Buccaneer¹¹⁶. Subsequent building and model adjustments were performed by hand using COOT¹¹⁷. A homology model of the N0 domain (residues 160–229) was created using I-TASSER and manually docked using COOT^{118–120}. Coulombic potential density for residues 1–159 was not observed. The model was refined in PHENIX version 1.16-dev3549 using phenix.real_space_refine with the resolution set to 3 Å¹²¹. NCS constraints were applied for the 14 subunits and were automatically detected and refined. Automatically determined secondary structure restraints, rotamer restraints, and Ramachandran restraints were applied as well. The quality of the model was evaluated using EMRinger¹²² and Molprobit¹²³ (Table 1).

CryoEM structure analysis. The structure of VcPilQ was compared to the *V. cholerae* T2SS secretin GspD (PDB 5WQ8), the *E. coli* T2SS GspD (PDB 5WQ7), and the *S. typhimurium* T3SS InvG (PDB 6DV3)^{55,124}. UCSF Chimera Match-Maker was used to calculate the RMSD between the C-alpha carbons of each pair^{125,126}.

The Positioning of Proteins in Membrane web server was used to predict the location of the transmembrane region of VcPilQ¹²⁷. Using UCSF Chimera, the structure of VcPilQ was docked into sub-tomogram averages of the non-piliated MxT4aP in *M. xanthus* Δ pilP (EMD-3249), the non-piliated MxT4aP in *M. xanthus* Δ pilB (EMD-3260), and the pilated MxT4aP in wild-type *M. xanthus* (EMD-3247)^{49,125}. At a high cryoEM density threshold, the lipid bilayer is clearly observed as two leaflets in the MxT4aP sub-tomogram averages, so there was little doubt where the hydrophobic region of VcPilQ should be placed in the sub-tomogram average. The bilayer produces stronger features in the MxT4aP sub-tomogram averages than the protein, so a lower cryoEM density threshold is required to visualize the full MxT4aP machinery. Here we focus on the Δ pilP MxT4aP sub-tomogram average (EMD-3249) because in this mutant only PilQ and TsaP localize correctly, and are therefore the only proteins likely to be present in the sub-tomogram average (Fig. 4c)⁴⁹. The putative location of TsaP suggested in

Chang et al.⁴⁹ is marked with an asterisk in Fig. 4c. In *M. xanthus*, TsaP is a peptidoglycan-binding protein¹⁰². In *V. cholerae*, the TsaP homolog LysM has not been implicated in T4aP function.

Unmodeled density analysis. The UCSF Chimera command molmap was used to generate a synthetic three-angstrom resolution cryoEM density map based on the VcPilQ atomic model (VcPilQ.pdb) (command: molmap <VcPilQ.pdb model number> 3)¹²⁵. The resulting density map (molmap_VcPilQ.mrc) was resampled to match the pixel size and box size of the VcPilQ cryoEM density map (the output from Relion Refine3D, cryoEM_PilQ.mrc)¹¹⁴. Resampling was performed in UCSF Chimera (command: vop resample <molmap_VcPilQ.mrc model number> onGrid <cryoEM_PilQ.mrc model number>), which yielded a synthetic density map based only on the atoms in the atomic model (molmap_VcPilQ_resampled.mrc).

The Relion command relion_mask_create was used to generate an inverted mask based on the synthetic density map (command: relion_mask_create --i molmap_VcPilQ_resampled.mrc --o mask_molmap_VcPilQ_resampled.mrc --ini_threshold 0.013 --invert). The resulting mask (mask_molmap_VcPilQ_resampled.mrc) has zeros where atoms are modeled and ones everywhere else (unmodeled space). To remove the parts of the empirical cryoEM map where atoms are modeled, the inverted mask is multiplied by the empirical cryoEM map (command: relion_image_handler --i cryoEM_PilQ.mrc --o cryoEM_PilQ_unmodeled_density.mrc --multiply mask_molmap_VcPilQ_resampled.mrc). The result is a cryoEM density volume representing unmodeled density (cryoEM_PilQ_unmodeled_density.mrc).

Mass spectrometry. After running a Bio-Rad Stain-Free gel and performing a Coomassie staining, the band of interest was excised with a clean razor blade. The gel piece was destained with ammonium bicarbonate and reduced with DTT (50 °C, 30 min). Next the sample was alkylated with iodoacetamide (room temperature, dark, 20 min). The gel pieces were then dehydrated. Trypsin was used to digest the protein in the gel (37 °C, overnight). Peptides were extracted from the gel matrix, dried, and desalted with a zip tip.

The in-gel-digested samples were subjected to LC-MS/MS analysis on a nanoflow LC system, EASY-nLC 1200 (Thermo Fisher Scientific), coupled to a QExactive HF Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a Nanospray Flex ion source.

Samples were directly loaded onto a C18 Aurora series column (Ion Opticks, Parkville, Australia). The 25 cm \times 50 μ m ID column (1.6 μ m) was heated to 45 °C. The peptides were separated with a 60 min gradient at a flow rate of 350 nL/min. The gradient was as follows: 2–6% Solvent B (3.5 min), 6–25% B (42.5 min), and 25–40% B (14.5 min), to 100% B (1 min) and 100% B (12 min). Solvent A consisted of 97.8% water, 2% acetonitrile, and 0.2% formic acid and solvent B consisted of 19.8% water, 80% acetonitrile, and 0.2% formic acid.

The QExactive HF Orbitrap was operated in data-dependent mode. Spray voltage was set to 1.8 kV, S-lens RF level at 20, and heated capillary at 275 °C. Full scan resolution was set to 60,000 at m/z 200. Full scan target was 3×10^6 with a maximum injection time of 15 ms (profile mode). Mass range was set to 300–1650 m/z . For data-dependent MS2 scans the loop count was 12, target value was set at 1×10^5 , and intensity threshold was kept at 1×10^5 . Isolation width was set at 1.2 m/z and a fixed first mass of 100 was used. Normalized collision energy was set at 28. Peptide match was set to off, and isotope exclusion was on. Ms2 data were collected in centroid mode.

Raw data were analyzed using MaxQuant (v. 1.6.5.0)^{128,129}. Spectra were searched against UniProt *V. cholerae* entries (3784 sequences) and a contaminant protein database (246 sequences). Trypsin was specified as the digestion enzyme and up to two missed cleavages were allowed. Precursor mass tolerance was 4.5 ppm after recalibration and fragment mass tolerance was 20 ppm. Variable modifications included oxidation of methionine and protein N-terminal acetylation. Carbamidomethylation of cysteine was specified as a fixed modification. A decoy database was used to set score thresholds to ensure a 1% false discovery rate at the protein and peptide level. Protein abundances were estimated using iBAQ and the fractional abundance was calculated as the protein abundance divided by the sum of all non-contaminant protein abundances¹³⁰.

Bioinformatics resources and software. Sequences were selected in the MiST3 database¹³¹ as of June 2020. The secretin protein domain model was taken from the PFAM database¹³² and specific models for individual type four filamentous family from the Rocha Lab⁷⁷. Matches to the protein domains were found using HMMER 3.1b2¹³³. Data manipulation was executed by custom scripts written in Typescript and available at <https://gitlab.com/jensenlab/seccomp>. To decrease the redundancy of the unaligned secretin sequences we used CD-HIT v4.6¹³⁴. Similarity scores were calculated with BLAST v2.7.1+¹³⁵. To further filter out divergent sequences, we used dyno cluster (<https://gitlab.com/jensenlab/dyno-cluster>) to parse the blastp results and generate connected graphs. Sequences were aligned using L-INS-I from the software package MAFFT v.7305b¹³⁶. We used Jalview¹³⁷ to manually inspect the multiple sequence alignment. We used the TREND¹³⁸ domain pipeline for initial explorations of the data set. Phylogenetic inferences were built using RAxML v8.2.1059¹³⁹. To collapse branches with low support in phylogenetic trees, we used TreeCollapseCL4 v3.0¹⁴⁰. We used CD-VIST for detailed domain architecture

identification and visualization¹⁴¹. The gene neighborhood data set and images were generated by GeneHood command line application v0.2.8-1 (<https://npmjs.org/package/genehood-cli>) and visualization together with homolog assignment using GeneHood viewer v0.16.0 (<https://genehood.io>).

Secretin sequence selection and analysis. We selected all 198,900 genes from the MiST3 database to which the gene product contained at least a single match to the Pfam domain model of secretin from Proteobacteria genomes. MiST3 contains pre-computed domain architecture prediction against Pfam 31 and taxonomy ranks from the National Center for Biotechnology Information. We used hhpres from HMMER to generate a hidden Markov model database from the secretin models defined by the Rocha Lab⁷⁷. We selected 56,941 sequences had the highest score against the T4aP model and above $1E-40$ e value. Next, we trimmed the region of the sequences matching the boundaries of the secretin domain model from Pfam 33. We used CD-HIT to decrease the redundancy of the data set at 65% identity to a total of 386 representative trimmed sequences and recollect the full-length sequences of these representatives. We used dyno cluster with $1E-110$ threshold to select over 50% of the sequences in the largest connected subgraph. We aligned the 203 full-length sequences using L-INS-I. We eliminated six sequences that open large gaps in the alignment. With the final data set of 197 sequences, we generated a phylogeny inference using RAxML with “-m PROTGAMMAILG -p 1234555 -x 9876545 -f a -N 200” parameters. We used TreeCollapse to collapse nodes with <50% bootstrap support. The alignment and the phylogenetic tree can be found in Supplementary Data 2. Also, we submitted the sequences to CD-VIST to detail domain architecture analysis with default parameters but skipping RPSBLAST step. Finally, we ran genehood-cli to fetch six genes up and downstream of each gene identifier in the alignment. We use GeneHood viewer to visualize the gene neighborhoods mapped to the phylogeny. GeneHood uses BLAST all vs. all similarity scores as edges and each gene of the data set as vertices in graphs to search for homologs among the displayed genes. The software filters edges according to a selected threshold for each gene and performs a breadth-first search to find all vertices that belong to the connected subgraph of the selected gene. All vertices of the selected subgraph are marked with the same color as the selected gene. Using this feature, we marked the homologs of each *Vibrio* gene in the data set. The thresholds used can be accessed by loading the Supplementary Data 3 to genehood.io.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data supporting the findings of this paper are available from the corresponding author upon reasonable request. The cryoEM reconstruction and model have been deposited in the Electron Microscopy Data Bank (<https://www.ebi.ac.uk/pdbe/entry/emdb/EMD-21559>) and the Protein Data Bank (PDB 6W6M).

Code availability

The scripts underlying the bioinformatics analysis are available at <https://gitlab.com/jensenlab/seccomp>. The alignment and the phylogenetic tree are available in Supplementary Data 2, while the thresholds used in the genehood.io analysis are in Supplementary Data 3. Other data are available from the corresponding author upon reasonable request.

Received: 6 March 2020; Accepted: 15 September 2020;

Published online: 08 October 2020

References

- von Wintersdorff, C. J. H. et al. Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. *Front. Microbiol.* **7**, 173 (2016).
- Blokesch, M. In and out—contribution of natural transformation to the shuffling of large genomic regions. *Curr. Opin. Microbiol.* **38**, 22–29 (2017).
- Berry, J.-L. & Pelicic, V. Exceptionally widespread nanomachines composed of type IV pili: the prokaryotic Swiss Army knives. *FEMS Microbiol. Rev.* **39**, 134–154 (2015).
- Piepenbrink, K. H. DNA uptake by type IV filaments. *Front. Mol. Biosci.* **6**, 1 (2019).
- Maiden, M. C. J. Horizontal genetic exchange, evolution, and spread of antibiotic resistance in bacteria. *Clin. Infect. Dis.* **27**, S12–S20 (1998).
- Domingues, S. et al. Natural transformation facilitates transfer of transposons, integrons and gene cassettes between bacterial species. *PLoS Pathog.* **8**, e1002837 (2012).
- Blokesch, M. & Schoolnik, G. K. Serogroup conversion of *Vibrio cholerae* in aquatic reservoirs. *PLoS Pathog.* **3**, e81 (2007).
- Miller, M. C., Keymer, D. P., Avelar, A., Boehm, A. B. & Schoolnik, G. K. Detection and transformation of genome segments that differ within a coastal population of *Vibrio cholerae* strains. *Appl. Environ. Microbiol.* **73**, 3695–3704 (2007).
- Croucher, N. J. et al. Rapid pneumococcal evolution in response to clinical interventions. *Science* **331**, 430–434 (2011).
- Hu, D. et al. Origins of the current seventh cholera pandemic. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E7730–E7739 (2016).
- Sun, Y., Bernardy, E. E., Hammer, B. K. & Miyashiro, T. Competence and natural transformation in vibrios. *Mol. Microbiol.* **89**, 583–595 (2013).
- Takemura, A. F., Chien, D. M. & Polz, M. F. Associations and dynamics of Vibrionaceae in the environment, from the genus to the population level. *Front. Microbiol.* **5**, 38 (2014).
- Wittman, R. J. & Flick, G. J. Microbial contamination of shellfish: prevalence, risk to human health, and control strategies. *Annu. Rev. Public Health* **16**, 123–140 (1995).
- Butt, A. A., Aldridge, K. E. & Sanders, C. V. Infections related to the ingestion of seafood Part I: viral and bacterial infections. *Lancet Infect. Dis.* **4**, 201–212 (2004).
- Pelicic, V. Type IV pili: e pluribus unum? *Mol. Microbiol.* **68**, 827–837 (2008).
- Craig, L., Forest, K. T. & Maier, B. Type IV pili: dynamics, biophysics and functional consequences. *Nat. Rev. Microbiol.* **17**, 429–440 (2019).
- Burrows, L. L. *Pseudomonas aeruginosa* twitching motility: Type IV Pili in action. *Annu. Rev. Microbiol.* **66**, 493–520 (2012).
- Hospenthal, M. K., Costa, T. R. & Waksman, G. A comprehensive guide to pilus biogenesis in gram-negative bacteria. *Nat. Rev. Microbiol.* **15**, 365–379 (2017).
- Maier, B. & Wong, G. C. L. How bacteria use type IV pili machinery on surfaces. *Trends Microbiol.* **23**, 775–788 (2015).
- Proft, T. & Baker, E. Pili in gram-negative and gram-positive bacteria—structure, assembly and their role in disease. *Cell. Mol. Life Sci.* **66**, 613–635 (2009).
- Aagesen, A. M. & Häse, C. C. Sequence analyses of type IV pili from *Vibrio cholerae*, *Vibrio parahaemolyticus*, and *Vibrio vulnificus*. *Microb. Ecol.* **64**, 509–524 (2012).
- Simpson, C. A., Podicheti, R., Rusch, D. B., Dalia, A. B. & van Kessel, J. C. Diversity in natural transformation frequencies and regulation across *Vibrio* species. *mBio* **10**, e02788-19 (2019).
- Meibom, K. L., Blokesch, M., Dolganov, N. A., Wu, C. -Y. & Schoolnik, G. K. Chitin induces natural competence in *Vibrio cholerae*. *Science* **310**, 1824–1827 (2005).
- Gulig, P. A., Tucker, M. S., Thiaville, P. C., Joseph, J. L. & Brown, R. N. USER friendly cloning coupled with chitin-based natural transformation enables rapid mutagenesis of *Vibrio vulnificus*. *AEM* **75**, 4936–4949 (2009).
- Chen, Y., Dai, J., Morris, J. G. & Johnson, J. A. Genetic analysis of the capsule polysaccharide (K antigen) and exopolysaccharide genes in pandemic *Vibrio parahaemolyticus* O3:K6. *BMC Microbiol.* **10**, 274 (2010).
- Neiman, J., Guo, Y. & Rowe-Magnus, D. A. Chitin-induced carbotype conversion in *Vibrio vulnificus*. *Infect. Immun.* **79**, 3195–3203 (2011).
- Yamamoto, S. et al. Regulation of natural competence by the orphan two-component system sensor kinase ChiS involves a non-canonical transmembrane regulator in *Vibrio cholerae*. *Mol. Microbiol.* **91**, 326–347 (2014).
- Dalia, A. B., Lazinski, D. W. & Camilli, A. Identification of a membrane-bound transcriptional regulator that links chitin and natural competence in *Vibrio cholerae*. *mBio* **5**, e01028-13 (2014).
- Meibom, K. L. et al. The *Vibrio cholerae* chitin utilization program. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2524–2529 (2004).
- Seitz, P. & Blokesch, M. DNA-uptake machinery of naturally competent *Vibrio cholerae*. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17987–17992 (2013).
- Ellison, C. K. et al. Retraction of DNA-bound type IV competence pili initiates DNA uptake during natural transformation in *Vibrio cholerae*. *Nat. Microbiol.* **3**, 773–780 (2018).
- Sandkvist, M. Biology of type II secretion. *Mol. Microbiol.* **40**, 271–283 (2001).
- Korotkov, K. V., Gonen, T. & Hol, W. G. Secretins: dynamic channels for protein transport across membranes. *Trends Biochem. Sci.* **36**, 433–443 (2011).
- Majewski, D. D., Worrall, L. J. & Strynadka, N. C. Secretins revealed: structural insights into the giant gated outer membrane portals of bacteria. *Curr. Opin. Struct. Biol.* **51**, 61–72 (2018).
- Genin, S. & Boucher, C. A. A superfamily of proteins involved in different secretion pathways in gram-negative bacteria: modular structure and specificity of the N-terminal domain. *Molec. Gen. Genet.* **243**, 112–118 (1994).
- Green, E. R. & Mecsas, J. Bacterial secretion systems—an overview. *Microbiol. Spectr.* **4**, 215–239 (2016).
- Adams, D. W., Stutzmann, S., Stoudmann, C. & Blokesch, M. DNA-uptake pili of *Vibrio cholerae* are required for chitin colonization and capable of kin

- recognition via sequence-specific self-interaction. *Nat. Microbiol.* **4**, 1545–1557 (2019).
38. Burkhardt, J., Vonck, J., Langer, J. D., Salzer, R. & Averhoff, B. Unusual N-terminal $\alpha\alpha\beta\beta\beta\alpha$ fold of PilQ from *Thermus thermophilus* mediates ring formation and is essential for piliation. *J. Biol. Chem.* **287**, 8484–8494 (2012).
 39. Collins, R. F., Davidsen, L., Derrick, J. P., Ford, R. C. & Tønnum, T. Analysis of the PilQ Secretin from *Neisseria meningitidis* by transmission electron microscopy reveals a dodecameric quaternary structure. *J. Bacteriol.* **183**, 3825–3832 (2001).
 40. Collins, R. F. et al. Three-dimensional structure of the *Neisseria meningitidis* secretin PilQ determined from negative-stain transmission electron microscopy. *J. Bacteriol.* **185**, 2611–2617 (2003).
 41. Collins, R. F. et al. Structure of the *Neisseria meningitidis* outer membrane PilQ secretin complex at 12 Å resolution. *J. Biol. Chem.* **279**, 39750–39756 (2004).
 42. Collins, R. F. et al. Interaction with type IV Pili induces structural changes in the bacterial outer membrane secretin PilQ. *J. Biol. Chem.* **280**, 18923–18930 (2005).
 43. Burkhardt, J., Vonck, J. & Averhoff, B. Structure and function of PilQ, a secretin of the DNA transporter from the thermophilic bacterium *Thermus thermophilus* HB27. *J. Biol. Chem.* **286**, 9977–9984 (2011).
 44. Balasingham, S. V. et al. Interactions between the lipoprotein PilP and the Secretin PilQ in *Neisseria meningitidis*. *J. Bacteriol.* **189**, 5716–5727 (2007).
 45. Berry, J.-L. et al. Structure and assembly of a trans-periplasmic channel for type IV Pili in *Neisseria meningitidis*. *PLOS Pathog.* **8**, e1002923 (2012).
 46. Koo, J., Lamers, R. P., Rubinstein, J. L., Burrows, L. L. & Howell, L. P. Structure of the *Pseudomonas aeruginosa* type IVa pilus secretin at 7.4 Å. *Structure* **24**, 1778–1787 (2016).
 47. D’Imprima, E. et al. Cryo-EM structure of the bifunctional secretin complex of *Thermus thermophilus*. *eLife* **6**, e30483 (2017).
 48. Bitter, W., Koster, M., Latijnhouwers, M., Cock, H. & Tommassen, J. Formation of oligomeric rings by XcpQ and PilQ, which are involved in protein transport across the outer membrane of *Pseudomonas aeruginosa*. *Mol. Microbiol.* **27**, 209–219 (1998).
 49. Chang, Y. -W. et al. Architecture of the type IVa pilus machine. *Science* **351**, aad2001 (2016).
 50. Chang, Y. -W. et al. Architecture of the *Vibrio cholerae* toxin-coregulated pilus machine revealed by electron cryotomography. *Nat. Microbiol.* **2**, 16269 (2017).
 51. Gold, V. A., Salzer, R., Averhoff, B. & Kühlbrandt, W. Structure of a type IV pilus machinery in the open and closed state. *eLife* **4**, e07380 (2015).
 52. Jain, S. et al. Structural characterization of outer membrane components of the type IV pili system in pathogenic *Neisseria*. *PLoS ONE* **6**, e16624 (2011).
 53. Tarry, M. et al. The extra-membranous domains of the competence protein HofQ show dna binding, flexibility and a shared fold with type I KH domains. *J. Mol. Biol.* **409**, 642–653 (2011).
 54. Lieberman, J. A. et al. Outer membrane targeting, ultrastructure, and single molecule localization of the enteropathogenic *Escherichia coli* type IV pilus secretin BfpB. *J. Bacteriol.* **194**, 1646–1658 (2012).
 55. Yan, Z., Yin, M., Xu, D., Zhu, Y. & Li, X. Structural insights into the secretin translocation channel in the type II secretion system. *Nat. Struct. Mol. Biol.* **24**, 177–183 (2017).
 56. Hay, I. D., Belousoff, M. J., Dunstan, R. A., Bamert, R. S. & Lithgow, T. Structure and membrane topography of the vibrio-type secretin complex from the type 2 secretion system of enteropathogenic *Escherichia coli*. *J. Bacteriol.* **200**, e00521–17 (2018).
 57. Yin, M., Yan, Z. & Li, X. Structural insight into the assembly of the type II secretion system pilotin–secretin complex from enterotoxigenic *Escherichia coli*. *Nat. Microbiol.* **3**, 581–587 (2018).
 58. Hay, I. D., Belousoff, M. J. & Lithgow, T. Structural basis of type 2 secretion system engagement between the inner and outer bacterial membranes. *mBio* **8**, e01344–17 (2017).
 59. Chernyatina, A. A. & Low, H. H. Core architecture of a bacterial type II secretion system. *Nat. Commun.* **10**, 1–10 (2019).
 60. Howard, P. S. et al. Structure and assembly of pilotin-dependent and -independent secretins of the type II secretion system. *PLOS Pathog.* **15**, e1007731 (2019).
 61. Worrall, L. et al. Near-atomic-resolution cryo-EM analysis of the Salmonella T3S injectisome basal body. *Nature* **540**, 597 (2016).
 62. McCallum, M., Tammam, S., Rubinstein, J. L., Burrows, L. L. & Howell, P. L. TsaP binds PilQ and is involved in cyclic di-GMP signaling. Preprint at <https://www.biorxiv.org/content/10.1101/2020.05.29.123786v1> (2020).
 63. Lo Scudato, M. & Blokesch, M. A transcriptional regulator linking quorum sensing and chitin induction to render *Vibrio cholerae* naturally transformable. *Nucleic Acids Res.* **41**, 3644–3658 (2013).
 64. Dalia, A. B., McDonough, E. & Camilli, A. Multiplex genome editing by natural transformation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8937–8942 (2014).
 65. Chami, M. et al. Structural insights into the secretin PulD and its trypsin-resistant core. *J. Biol. Chem.* **280**, 37732–37741 (2005).
 66. Guilvout, I., Nickerson, N. N., Chami, M. & Pugsley, A. P. Multimerization-defective variants of dodecameric secretin PulD. *Res. Microbiol.* **162**, 180–190 (2011).
 67. Tan, Y. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).
 68. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
 69. Rocaboy, M. et al. The crystal structure of the cell division amidase AmiC reveals the fold of the AMIN domain, a new peptidoglycan binding domain. *Mol. Microbiol.* **90**, 267–277 (2013).
 70. de Souza, R. F., Anantharaman, V., de Souza, S. J., Aravind, L. & Gueiros-Filho, F. J. AMIN domains have a predicted role in localization of diverse periplasmic protein complexes. *Bioinformatics* **24**, 2423–2426 (2008).
 71. Ghosal, D. et al. In vivo structure of the Legionella type II secretion system by electron cryotomography. *Nat. Microbiol.* **4**, 2101–2108 (2019).
 72. Dombkowski, A. A. Disulfide by design: a computational method for the rational design of disulfide bonds in proteins. *Bioinformatics* **19**, 1852–1853 (2003).
 73. Craig, D. B. & Dombkowski, A. A. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinform.* **14**, 346 (2013).
 74. Dombkowski, A. A., Sultana, K. Z. & Craig, D. B. Protein disulfide engineering. *FEBS Lett.* **588**, 206–212 (2014).
 75. Ellison, C. K., Dalia, T. N., Dalia, A. B. & Brun, Y. V. Real-time microscopy and physical perturbation of bacterial pili using maleimide-conjugated molecules. *Nat. Protoc.* **14**, 1–17 (2019).
 76. Ellison, C. K. et al. Obstruction of pilus retraction stimulates bacterial surface sensing. *Science* **358**, 535–538 (2017).
 77. Denise, R., Abby, S. S. & Rocha, E. P. C. Diversification of the type IV filament superfamily into machines for adhesion, protein secretion, DNA uptake, and motility. *PLOS Biol.* **17**, e3000390 (2019).
 78. Craig, L., Pique, M. E. & Tainer, J. A. Type IV pilus structure and bacterial pathogenicity. *Nat. Rev. Microbiol.* **2**, 363 (2004).
 79. Plant, L. J. & Jonsson, A.-B. Type IV pili of *Neisseria gonorrhoeae* influence the activation of human CD4+ T cells. *Infect. Immun.* **74**, 442–448 (2006).
 80. Friedrich, A., Prust, C., Hartsch, T., Henne, A. & Averhoff, B. Molecular analyses of the natural transformation machinery and identification of pilus structures in the extremely thermophilic bacterium *Thermus thermophilus* strain HB27. *Appl. Environ. Microbiol.* **68**, 745–755 (2002).
 81. Martin, P. R., Hobbs, M., Free, P. D., Jeske, Y. & Mattick, J. S. Characterization of pilQ, a new gene required for the biogenesis of type 4 fimbriae in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **9**, 857–868 (1993).
 82. Parkhill, J. et al. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* **404**, 502–506 (2000).
 83. Mattick, J. S., Whitchurch, C. B. & Alm, R. A. The molecular genetics of type-4 fimbriae in *Pseudomonas aeruginosa*—a review. *Gene* **179**, 147–155 (1996).
 84. Ropp, P. A. & Nicholas, R. A. Cloning and characterization of the ponA gene encoding penicillin-binding protein 1 from *Neisseria gonorrhoeae* and *Neisseria meningitidis*. *J. Bacteriol.* **179**, 2783–2787 (1997).
 85. Wall, D., Kolenbrander, P. E. & Kaiser, D. The *Myxococcus xanthus* pilQ (sglA) gene encodes a secretin homolog required for type IV pilus biogenesis, social motility, and development. *J. Bacteriol.* **181**, 24–33 (1999).
 86. Palchevskiy, V. & Finkel, S. E. *Escherichia coli* competence gene homologs are essential for competitive fitness and the use of DNA as a nutrient. *J. Bacteriol.* **188**, 3902–3910 (2006).
 87. Sauvage, E., Kerff, F., Terrak, M., Ayala, J. A. & Charlier, P. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiol. Rev.* **32**, 234–258 (2008).
 88. Merz, A. J., So, M. & Sheetz, M. P. Pilus retraction powers bacterial twitching motility. *Nature* **407**, 98–102 (2000).
 89. Skerker, J. M. & Berg, H. C. Direct observation of extension and retraction of type IV pili. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6901–6904 (2001).
 90. Wolfgang, M. et al. PilT mutations lead to simultaneous defects in competence for natural transformation and twitching motility in piliated *Neisseria gonorrhoeae*. *Mol. Microbiol.* **29**, 321–330 (1998).
 91. Salleh, M. Z. et al. Structure and properties of a natural competence-associated pilin suggest a unique pilus tip-associated DNA receptor. *mBio* **10**, e00614–e00619 (2019).
 92. Craig, L. et al. Type IV pilus structure by cryo-electron microscopy and crystallography: implications for pilus assembly and functions. *Mol. Cell* **23**, 651–662 (2006).
 93. Neuhaus, A. et al. Cryo-electron microscopy reveals two distinct type IV pili assembled by the same bacterium. *Nat. Commun.* **11**, 2231 (2020).
 94. Kruse, K., Salzer, R., Joos, F. & Averhoff, B. Functional dissection of the three N-terminal general secretory pathway domains and the Walker motifs of the traffic ATPase PilF from *Thermus thermophilus*. *Extremophiles* **22**, 461–471 (2018).

95. Seitz, P. et al. ComEA is essential for the transfer of external DNA into the periplasm in naturally transformable *Vibrio cholerae* cells. *PLOS Genet.* **10**, e1004066 (2014).
96. Kolappan, S. et al. Structure of the *Neisseria meningitidis* type IV pilus. *Nat. Commun.* **7**, 13015 (2016).
97. Wang, F. et al. Cryoelectron microscopy reconstructions of the *Pseudomonas aeruginosa* and *Neisseria gonorrhoeae* type IV pili at sub-nanometer resolution. *Structure* **25**, 1423–1435.e4 (2017).
98. Bardiaux, B. et al. Structure and assembly of the enterohemorrhagic *Escherichia coli* type 4 pilus. *Structure* **27**, 1082–1093.e5 (2019).
99. Pugsley, A. P. The complete general secretory pathway in gram-negative bacteria. *Microbiol. Rev.* **57**, 50–108 (1993).
100. Russel, M. Macromolecular assembly and secretion across the bacterial cell envelope: type II protein secretion systems. *J. Mol. Biol.* **279**, 485–499 (1998).
101. Wehbi, H. et al. The peptidoglycan-binding protein FimV promotes assembly of the *Pseudomonas aeruginosa* type IV pilus secretin. *J. Bacteriol.* **193**, 540–550 (2011).
102. Siewering, K. et al. Peptidoglycan-binding protein TsaP functions in surface assembly of type IV pili. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E953–E961 (2014).
103. Carter, T. et al. The type IVa pilus machinery is recruited to sites of future cell division. *mBio* **8**, e02103–e02116 (2017).
104. Miller, V. L., DiRita, V. J. & Mekalanos, J. J. Identification of toxS, a regulatory gene whose product enhances toxR-mediated activation of the cholera toxin promoter. *J. Bacteriol.* **171**, 1288–1293 (1989).
105. Dalia, A. B. Natural cotransformation and multiplex genome editing by natural transformation (MuGENT) of *Vibrio cholerae*. In *Vibrio Cholerae: Methods and Protocols* (ed. Sikora, A. E.) 53–64 (Springer, 2018).
106. Zhu, J. et al. Quorum-sensing regulators control virulence gene expression in *Vibrio cholerae*. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3129–3134 (2002).
107. Dalia, T. N., Chlebek, J. L. & Dalia, A. B. A modular chromosomally integrated toolkit for ectopic gene expression in *Vibrio cholerae*. *Sci. Rep.* **10**, 15398 (2020).
108. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
109. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
110. Rohou, A. & Grigorieff, N. CTFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
111. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
112. D. Asarnow, E. Palovcak & Y. Cheng. asarnow/pyem: UCSF pyem v0.5. Zenodo. <https://doi.org/10.5281/zenodo.3576630> (2019).
113. Kimanius, D., Forsberg, B. O. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, e18722 (2016).
114. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
115. Zivanov, J., Nakane, T. & Scheres, S. H. W. Estimation of high-order aberrations and anisotropic magnification from cryo-EM datasets in RELION-3.1. *IUCr* **7**, 253–267 (2020).
116. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Cryst. D* **62**, 1002–1011 (2006).
117. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of coot. *Acta Cryst. D* **66**, 486–501 (2010).
118. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).
119. Yang, J. et al. The I-TASSER suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
120. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinform.* **9**, 40 (2008).
121. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D* **66**, 213–221 (2010).
122. Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. <https://doi.org/10.1038/nmeth.3541> (2015).
123. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst. D* **66**, 12–21 (2010).
124. Hu, J. et al. Cryo-EM analysis of the T3S injectisome reveals the structure of the needle and open secretin. *Nat. Commun.* **9**, 1–11 (2018).
125. Pettersen, E. F. et al. UCSF chimera—a visualization system for exploratory research and analysis. *J. Computational Chem.* **25**, 1605–1612 (2004).
126. Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C. & Ferrin, T. E. Tools for integrated sequence-structure analysis with UCSF chimera. *BMC Bioinform.* **7**, 339 (2006).
127. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).
128. Cox, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
129. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
130. Schwanhäusser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
131. Gumerov, V. M., Ortega, D. R., Adebali, O., Ulrich, L. E. & Zhulin, I. B. MiST 3.0: an updated microbial signal transduction database with an emphasis on chemosensory systems. *Nucleic Acids Res.* **48**, D459–D464 (2020).
132. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
133. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
134. Li, W., Jaroszewski, L. & Godzik, A. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **17**, 282–283 (2001).
135. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 1–9 (2009).
136. Katoh, K. & Toh, H. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* **9**, 286–298 (2008).
137. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
138. Gumerov, V. M. & Zhulin, I. B. TREND: a platform for exploring protein function in prokaryotes based on phylogenetic, domain architecture and gene neighborhood analyses. *Nucleic Acids Res.* **48**, W72–W76 (2020).
139. Stamatakis, A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
140. Hodcroft, E. TreeCollapserCL4 <http://emmahodcroft.com/TreeCollapseCL.html> (2015).
141. Adebali, O., Ortega, D. R. & Zhulin, I. B. CDvist: a webserver for identification and visualization of conserved domains in protein sequences. *Bioinformatics* **31**, 1475–1477 (2015).
142. Jurrus, E. et al. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 (2018).

Acknowledgements

Cryo Electron microscopy was performed in the Beckman Institute Resource Center for Transmission Electron Microscopy at Caltech. Dr. Songye Chen and Dr. Andrey Mal-yutin assisted with data collection. Dr. Spiros D. Garbis, Dr. Annie Moradian, Dr. Michael Sweredoski, and Dr. Brett Lomenick at the Caltech Proteome Exploration Laboratory (PEL) performed and analyzed mass spectrometry results. Dr. Naima Sharaf, Jeffery Lai, and Prof. Doug Rees provided invaluable advice on membrane protein biochemistry and instrumentation. Jane Ding and Welison Floriano provided computational support. Dr. Debnath Ghosal, Dr. Mohammed Kaplan, Dr. Catherine Oikonomou, Dr. Lauren Ann Metskas, Dr. Christopher Barnes, Claudia Jette, and Andrew Schacht provided feedback and advice. This work was supported in part by National Institutes of Health grant AI127401 to G.J.J. and National Institutes of Health grant R35GM128674 to A.B.D.

Author contributions

S.J.W. conceptualized the project, expressed and purified the protein, prepared samples for cryoEM, collected cryoEM data, processed cryoEM data, assisted in atomic model building, interpreted results, designed figures, and wrote the paper. D.R.O. designed research, performed sequence analyses, wrote software, collected sequence data, interpreted results, designed figures, and contributed text to the paper. M.H.S. purified protein, assisted with cryoEM sample prep and data collection, built the atomic model, interpreted results, and provided feedback on the paper. T.N.D. engineered the *V. cholerae* constructs, performed microbial assays, and interpreted results. A.B.D. conceptualized the project, obtained funding, engineered the *V. cholerae* constructs, performed microbial assays, interpreted results, designed figures, and provided feedback on the paper. G.J.J. conceptualized the project, obtained funding, interpreted results, and provided feedback on the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-18866-y>.

Correspondence and requests for materials should be addressed to G.J.J.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contributions to the peer review of this work. Peer review reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020