bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

CryoEM structure of the outer membrane secretin channel pIV from the f1 filamentous bacteriophage

Rebecca Conners[1,2], Mathew McLaren[1,2], Urszula Łapińska[1,2], Kelly Sanders[1,2], M. Rhia L. Stone[3], Mark A. T. Blaskovich[3], Stefano Pagliara[1,2], Bertram Daum[1,2], Jasna Rakonjac[4] & Vicki A. M. Gold[1,2]*

[1] Living Systems Institute, University of Exeter, Stocker Road, Exeter EX4 4QD, UK

[2] College of Life and Environmental Sciences, Geoffrey Pope, University of Exeter, Stocker Road, Exeter EX4 4QD, UK

[3]Centre for Superbug Solutions, Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia

[4]School of Fundamental Sciences, Massey University, Palmerston North, New Zealand

*Corresponding author: v.a.m.gold@exeter.ac.uk

Abstract

The Ff family of filamentous bacteriophages infect gram-negative bacteria, but do not cause lysis of their host cell. Instead, new virions are extruded via the phage-encoded pIV protein, which has homology with bacterial secretins. Here, we have determined the structure of pIV from the f1 filamentous bacteriophage at 2.7 Å resolution by cryo-electron microscopy, the first near-atomic structure of a phage secretin. Fifteen f1 pIV monomers assemble to form a gated channel in the bacterial outer membrane, with associated soluble domains projecting into the periplasm. We model channel opening and propose a mechanism for phage-mediated gate movement. By single-cell microfluidics experiments, we demonstrate the potential for secretins such as pIV to be used as adjuvants to increase the uptake and efficacy of antibiotics

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

in bacteria. Finally, we compare the f1 pIV structure to its homologues to reveal similarities and differences between phage and bacterial secretins.

Keywords

Filamentous bacteriophage, secretin, pIV, cryo-electron microscopy, single-cell microfluidics, antimicrobial resistance

Introduction

The Ff family of filamentous bacteriophages includes f1, fd and M13, and is one of the simplest biological systems known[1]. Filamentous phages such as Ff have many uses in biotechnology and nanotechnology[2]; examples include as a cloning vector, and in phage display for screening protein-ligand and protein-protein interactions. Ff is also increasingly employed in nanotechnology, being engineered to adopt a variety of different forms such as nano-wires, nanorings and branched structures[3,4]. With the recent, and much publicised, global antimicrobial resistance crisis[5,6], there has been a renewed campaign in exploring the use of phages and their proteins to target pathogenic bacteria[7,8]. Understanding the structural and functional details of phage lifecycles is thus of broad interest for both fundamental and applied biological science research.

Phages f1, fd and M13 share 98.5% sequence identity and infect the gram-negative bacterium *Escherichia coli*. The primary binding receptor is the F-pilus - a long filamentous appendage assembled on the bacterial cell surface. Phage virions are 1 µm long and 6-7 nm wide, with a circular single-stranded DNA genome encoding for 11 viral proteins. The viral genome is encapsulated within several thousand copies of the major viral coat protein pVIII and is capped by protein complexes consisting of pIII/pVI and pVII/pIX at either end[9].

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Phages bind to the F-pilus via the pIII/pVI cap, which is followed by pilus retraction and binding of phage to the TolQRA complex in the host cell inner membrane[10]. The phage genome is released into the bacterial cytoplasm, DNA is replicated and phage proteins are expressed[11]. New virions are assembled and released from the cell via a complex of three phage-encoded proteins, pI, pXI and pIV. pI (an ATPase) and pXI span the inner membrane with pIV located in the outer membrane[12,13].

pIV is a secretin family protein, with sequence similarity to bacterial secretins at the core of the Type IV pilus assembly machinery and the Type II and Type III secretion systems[14]. A number of high-resolution structures of bacterial secretins have been determined[15-18], which have allowed detailed comparisons to be made[19,20]. All secretins are multimeric channel proteins, ranging in molecular weight from 670 kDa to >1 MD. Reports of stoichiometry range between twelve and fifteen monomers per complex, divided into a number of different domains. The C-terminal domain integrates into the bacterial outer membrane and forms a double-walled ß-barrel with a central gate forming a stricture at the centre[21]. Varying numbers of soluble N-domains connect to the ß-barrel and project into the periplasm. The N-domains are often poorly ordered in cryo-electron microscopy (cryoEM) structures, presumably due to their flexibility. Type II secretins can also differ at the externally-exposed side of the ß-barrel, with *Vibrio*-type complexes possessing a cap gate and *Klebsiella*-type lacking this feature[19].

A cryoEM structure of the f1 pIV (f1pIV) protein determined at 22 Å resolution[22] showed a barrel-like complex with a large pore running through the centre containing a pore-gate. We present here the cryoEM structure of f1pIV to 2.7 Å resolution, the first near-atomic resolution structure of a phage secretin protein. By computational modelling and correlation

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

of our structure to a large library of phenotypic mutants, we propose a mechanism for channel opening and interaction of f1pIV with phage. We also employ a mutant of f1pIV that produces a "leaky" *E. coli* phenotype and demonstrate by single-cell microfluidics that the macrolide antibiotic roxithromycin is taken up efficiently by the f1pIV-expressing bacteria. The f1pIV atomic structure is used to interpret these findings, which have important translational implications for the use of secretins as therapeutic adjuvants to increase antibiotic delivery to bacteria. Finally, we reveal the common structural motifs shared by f1pIV and bacterial secretins, and differences in their electrostatic surface charges and modular architecture.

Results

Structure of the f1pIV secretin

The 670 kDa f1pIV secretin channel was expressed recombinantly in *E. coli* and purified in the presence of CHAPS detergent by Ni^{2+} affinity and size exclusion chromatography. Pure protein was identified by SDS-PAGE, Western blot and mass spectrometry analysis (Supplementary Fig. 1). CryoEM data were collected and processed in Relion 3.1[23] (Supplementary Table 1). Averaging the protein particles together in 2D showed that f1pIV particles are composed of fifteen identical copies of the monomer (Fig. 1a); no variation in symmetry was observed. 3D reconstruction produced a 2.7 Å map with C15 symmetry applied (Fig. 1b and Supplementary Fig. 2).

The N-terminal part of f1pIV is comprised of two periplasmic N domains: the far N-terminal N0 domain (residues 1-88) and the N3 domain (108-172), which are joined together with a 19-residue linker (89-107) (Fig. 2a). The N3 domain links to the C-terminal secretin domain (175-405) via a two amino acid linker (173-174) (Fig. 2a). The quality of our map allowed

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

the N3 and secretin domains to be unambiguously modelled into the density (Fig. 1c, 2b, c), with clear side chain densities visible in most areas (Fig. 2d). The fifteen monomers pack against each other to form the secretin channel, with each monomer contributing a four-stranded ß-sheet to the multimer, culminating in a large ß-barrel composed of 60 ß-strands (Fig. 2b, c). The ß-sheets lie at an angle of 41° relative to the central axis of the pore. Within the ß-barrel is a smaller inner barrel comprised of four ß-strands from each f1pIV monomer (Fig. 2c, e-g). The inner ß-barrel is angled parallel to the pore axis and provides a scaffold for two extended loops that reach into the centre of the pore, known as Gate 1 and Gate 2[24] (Fig. 2c, g). The interface between the outer and inner ß-barrels is composed mostly of hydrophobic residues, and strengthened further with a salt bridge between Glu 185 and Arg 337 (both residues are highly conserved amongst homologues) (Fig. 2e, Supplementary Fig. 3 and 4). Extensive hydrogen bonds are formed within the inner and outer ß-barrels; within each individual monomer, and also between each monomer and its immediate neighbours (Fig. 2f). In the outer ß-barrel, monomer 1 forms 15 intra-subunit hydrogen bonds, and 12 inter-subunit hydrogen bonds (6 with each neighbouring monomer). In the inner ß-barrel, monomer 1 forms 26 intra-subunit hydrogen bonds and 14 inter-subunit ones (7 with each neighbouring monomer). The interface between neighbouring N3 domains is mostly hydrophobic, and no hydrogen bond interactions are observed between them in our structure. For reasons of flexibility (Supplementary Fig. 2), the extracellular part of the ß-barrel was missing interpretable density (36 amino acid residues, positions 199-234) as well as part of the Gate 1 loop (residues 275-289) at the very centre of the pore.

The f1pIV secretin measures 96 Å from the extracellular face of the ß-barrel to the N-terminus of the N3 domain, and 112 Å across at its widest point (the extracellular part of the ß-barrel. The diameter across the inner ß-barrel is 90 Å. The pore width is constricted further

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

in two areas; at the central gate (40 Å) and at the N3 domains (74 Å). There are however disordered loops in both the N3 domain and Gate 1, both pointing into the centre of the pore, so these distances could be smaller.

The globular N3 domain consists of 65 residues which fold to form a three-stranded ß-sheet packed against two α-helices, joined by the short two amino acid linker to the ß-barrel (Fig. 2a, c, g). The far N-terminal N0 domain and its associated 19 residue linker region (residues 1-107, Fig. 2a) could not be observed in our structure. In the 2D classes, a hazy area could be seen in the expected region, suggesting that N0 could be present but extremely flexible (Fig. 1a, red arrow). Tryptic digest mass spectrometry was used to confirm that all of the regions for which we either did not see density, or could not interpret the density, were physically present (Fig. 2a).

To provide insight into the structure of the N0 domain, we employed three complementary methods. We performed a Blast search of f1pIV N0 plus the associated linker region (residues 1-107) against all structures in the Protein Data Bank. This resulted in a single hit for the PulD secretin from *Klebsiella pneumoniae* (6HCG) with 33% sequence identity over 63 residues. We used this structure to produce a homology model of the f1pIV N0 domain in Swiss-Model[25]. The same f1pIV N0-linker domain sequence was also submitted to the I-TASSER[26] and Robetta structural prediction servers[27]. The resulting models from all methods were compared and seen to be in agreement for the folded domain (residues 2-71) (Supplementary Fig. 5). The N0 domain is thus most likely formed by two α-helices, flanked by a two-stranded β-sheet on one side and a three-stranded β-sheet on the other. We used the structure from I-TASSER to generate a composite model with our f1pIV structure (Fig. 2g). The three N0 models diverge at residues 72-107, which is in agreement with this region being

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

a flexible linker (Figure 2a, Supplementary Fig. 5). The linker region has been modelled in a fully extended conformation and the position of the N0 domain obtained by overlaying the f1pIV structure with that of the *Klebsiella* PulD secretin (6HCG[28], Supplementary Fig. 5); the distance between N0 and N3 could be closer if the linker takes on a more compact structure. The entire length of f1pIV (from the extracellular face of the β-barrel to the N-terminal periplasmic domain of N0) in our composite model is 220 Å. This agrees well with the distance needed for f1pIV to cross the *E. coli* periplasm (with a width of 210 Å)[29] and reach the pI/pXI proteins in the inner membrane.

Membrane integration of f1pIV

An additional band of density was observed in our EM map which was unaccounted for, lying in approximately the centre of the surface of the ß-barrel (Fig. 3a). This formed a ring around the outside of the secretin domain, reminiscent of detergent belts typically seen in cryoEM structures of membrane proteins[30]. We modelled the detergent CHAPS, used during purification, into this density. Two molecules of CHAPS were observed to bind to each f1pIV monomer, forming a circle of detergent around the barrel (Fig. 3b). The detergent molecules were bound to a ring of aromatic residues (Phe 184, Tyr 359 and Tyr 394), forming an aromatic girdle. An additional aromatic girdle was observed nearer the extracellular region of the ß-barrel formed by Phe 194 and Phe 196. Such clusters of aromatic residues are often observed at the lipid-water interface in membrane-spanning ß-barrels of gram-negative bacteria, and allow for protein positioning in the membrane[31]. Colouring the surface of f1pIV by hydrophobicity confirms that the two aromatic girdles lie within the two main regions of high hydrophobicity on the protein surface (Fig. 3c).

The Gate region

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The inner ß-barrel, Gate 1 (G267-G297) and Gate 2 loops (D321-Q336) form the full gate region that projects into the pore. The Gate 1 loops extend to the centre, presumably sealing the channel. Sequence alignments show high sequence conservation for Gate 1 amongst secretin homologues, while Gate 2 is less well conserved (Supplementary Fig. 3). In both cases, the areas of highest sequence conservation are found where the gates join the inner ß-barrel, with the more variable regions lying in the loop regions which face the centre of the pore. There are two highly conserved glycine residues found at the boundary of Gate 1, which are believed to form a hinge for gate opening[32] (Fig. 4a). An extensive hydrogen bonding network was observed within the gate region (Fig. 4a). Hydrogen bonds are formed within each gate (5 intra-gate bonds in Gate 1, and 2 intra-gate bonds in Gate 2), tethering each loop together tightly. As seen in all other secretin structures to date, the Gate 1 loop has an unusual twisted conformation. Inter-gate hydrogen bonds are formed between neighbouring subunits in the multimer, with Gate 1 from one chain forming two hydrogen bonds with Gate 2 from the neighbouring chain (Fig. 4a, b). A highly conserved Arg 293 in Gate 1 is observed with clear density in a buried position. It forms hydrogen bonds with the side chain of a conserved Asn 269 within the same gate, and with the backbone carbonyl oxygen of Val 332 from Gate 2 of the neighbouring subunit (Fig. 4b, Supplementary Fig. 3), adding further inter-subunit stability to the gate region.

Opening the f1pIV channel

The f1pIV pore must undergo significant structural rearrangements to allow phage to egress. To provide insight into the mechanism of channel opening, we compared our f1pIV structure to existing closed and open gate structures of a homologous protein: the Type III secretin InvG from *Salmonella typhimurium* (6PEE and 6Q15)[33]. Structural alignment of the two InvG structures highlights the main differences between the open and closed states

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(Supplementary Fig. 6). The core of the secretin domain and the conformation of the loops remain similar; differences are seen in the position of the gate loops relative to the barrel, the ß-lip and an extended loop in the N3 domain. Both gate loops move to pack against the outer ß-barrel in the open form. The Gate 1 loop moves with a hinge-like motion, while the Gate 2 loop twists. The ß-lip at the extracellular edge of the ß-barrel also moves away from the centre of the pore in the open structure, and there is movement of an extended loop (not present in f1pIV) in the N3 domain. Our closed f1pIV structure aligns well to the InvG closed state, with a Root Mean Square Deviation (RMSD) of 1.051 Å (Supplementary Fig. 6). The f1pIV closed gate loops were thus modelled into the open form based on their position in InvG (Supplementary Fig. 6 and Fig. 4c).

This model reveals how Gate 1 would likely pivot upwards and outwards on the highly conserved hinge residues Gly 267 and Gly 297, located at the top of the inner ß-barrel (Fig. 4a, c, d, Supplementary Fig. 3). The twisting motion in Gate 2 could occur around the conserved Asp 321 and Asn 335 (Fig. 4d, Supplementary Fig. 3). These movements would move both gate loops closer to the outer ß-barrel and thus result in an opening of the pore from the ~40 Å measured in our closed structure to 65 Å (Fig. 4e). This would enable the virus particle (60 Å in diameter) to pass through the pore (Fig. 4f). A fibre diffraction structure of an fd phage (2C0W) at 3.2 Å resolution[34] was coloured by electrostatic potential (Fig. 4g) and shows a clear negative charge on its outer surface. In contrast, the electrostatic surface potential within the f1pIV pore is more varied. In closed and open states, there is a band of positive charge in the area closest to the extracellular side, followed by a band of negative charge near the gate region. In the open state, the aperture revealed by gate opening is mostly positively charged (Fig. 4g).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Mapping mutations to f1pIV that cause a "leaky" *E. coli* phenotype

Wild type *E. coli* cells (strain K12) cannot transport the bile salt deoxycholate across their outer membranes, and those which are deficient in the LamB maltoporin cannot transport large maltooligosaccharide sugars. Previous studies[24,35,36] have revealed a number of mutations in the f1pIV protein that render the pore leaky to these substrates. Leaky mutations discovered by random mutagenesis (Supplementary Table 2)[24] have now been mapped on to the f1pIV structure (Fig. 5a, b). The overwhelming majority (90%) were found to lie in the gate region (the Gate 1 and Gate 2 loops or inner ß-barrel). Three mutations (A121V, D123Y, G147V) conferring leakiness to deoxycholate and maltopentaose were identified in the N3 domain, all at the interface closest to the inner ß-barrel; one mutation (I183V) was observed in the outer ß-barrel (Fig. 5a). The authors also tested susceptibility of *E. coli* expressing different f1pIV mutants to the antibiotics vancomycin and bacitracin (which are too large to cross the bacterial outer membrane). By mapping these to our f1pIV structure, we can show that all mutants which conferred antibiotic susceptibility were located in the gate region (4 with sensitivity to vancomycin alone, 12 with sensitivity to both vancomycin and bacitracin) (Fig. 5b).

We used the map of f1pIV mutants to understand more about the delicate balance of interactions in the gates (Supplementary Table 2). The conserved Gate 1 hinges (Gly 267 and Gly 297) pack closely together within the inner ß-barrel (Fig. 4d). Changing Gly 267 or Gly 297 to any other sidechain in the model causes steric clashes. Hinge mutations would thus disrupt the packing in the region, interfere with hinge motion and prevent proper gate closure. Asn 269, Arg 293 and Asn 295 in Gate 1 take part in a tight hydrogen bonding network in the closed structure, with the Arg 293 sidechain being buried (Fig 4b; Supplementary Fig. 7).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Disrupting the hydrogen bonding network within the Gate 1 loop and between two neighbouring gates would thus lead to a more dynamic and open gate structure.

Antibiotic uptake through an f1pIV leaky mutant

The antibiotic vancomycin is hydrophilic and bacitracin is amphipathic; both are sufficiently small to pass through the open f1pIV pore (Supplementary Fig. 8). *E. coli* expressing the previously characterised f1pIV mutant S324G are sensitive to both antibiotics, and still functional for phage assembly[24,37]. Ser 324 is located in Gate 2, where mutation to a glycine will destroy the hydrogen bond made between the Ser 324 sidechain and the backbone nitrogen of Ser 326 (Fig. 5c). Mutation to a glycine could also change the position of the peptide backbone, potentially altering the hydrogen bonds that are currently made between the Ser 324 carbonyl oxygen and the sidechain of Gln 328 from the neighbouring monomer, and those made between the backbone nitrogen of Ser 324 and the carbonyl oxygen of Ile 333 (Fig. 5c). The combined effect of these mutations likely results in increased mobilisation of the gate.

Building on this knowledge, here we aimed to understand whether the uptake and efficacy of hydrophobic antibiotics could be improved by expressing leaky f1pIV in *E. coli*. Hydrophobic macrolides are often used as first-line antibiotics, however, they are used typically to treat infections caused by gram-positive but not most gram-negative bacteria due to slow uptake across the outer membrane[38]. The macrolide antibiotic roxithromycin is widely clinically employed[39], and our modelled open gate structure reveals that it is sufficiently small to pass through the open f1pIV channel unhindered (Supplementary Fig. 8). Using single-cell analysis with the microfluidic mother machine and time-lapse microscopy[40,41], bacteria were injected into channels and the accumulation of fluorescent

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

antibiotics measured[42]. Specifically, we employed a recently introduced fluorescent derivative of roxithromycin (roxithromycin linked to nitrobenzoxadiazole -roxithromycin-NBD), which largely maintains the antibiotic potency of the parental drug[38]. *E. coli* cells expressing f1pIV^{S324G} became distinguishable from the background within 700 s after adding roxithromycin-NBD to our microfluidic device (Fig. 5d), suggesting rapid intracellular accumulation. In contrast, *E. coli* cells expressing wild type f1pIV (f1pIV^{WT}) remained largely indistinguishable from the background during exposure to roxithromycin-NBD (Fig. 5d), suggesting negligible accumulation. In agreement with these data, when we measured the minimum inhibitory concentration (MIC) for roxithromycin, there was a one-notch shift in the MIC of roxithromycin against the mutant compared to the wild-type (64 and 96 µg/ml, respectively).

Structural comparison of f1pIV with other secretins

As our structure is the first of a viral secretin at near-atomic level, we compared f1pIV to different classes of bacterial secretins[32,33,43] (Fig. 6a, b, c, Supplementary Fig. 9). All secretins share the same overall architecture of the double-walled ß-barrel and the first and last periplasmic domains (N0 and N3), but differ in the number of additional N domains (reviewed by Filloux and Voulhoux[19]). The Gate 1 and Gate 2 loops are all in similar conformations, pointing into the centre of the pore from the inner ß-barrel. The N3 domains are broadly similar, with the one notable difference being an extended loop region in the Type III InvG secretin which extends up towards the central gate area (Fig. 6c; Supplementary Fig. 9). In general, there is increasing disorder within domains closest to the N-terminus, and many structures have no clear density for their N0 domains.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Differences between secretins are seen on the extracellular face of the ß-barrel with the Klebsiella-like group of Type II secretins having a cap gate which forms a much narrower opening when compared to the other open-edged ß-barrels. From sequence analysis (Supplementary Fig. 4), we believe f1pIV will not have a cap structure at the extracellular side of the ß-barrel, but we could not see this region in our cryoEM map, likely for reasons of flexibility (Supplementary Fig. 2). We are missing density for 36 residues, which is the typical size of the ß-lip domain in phage secretins (Supplementary Fig. 4). In contrast, the bacterial Type II secretins typically have 51-60 residues in the ß-lip region, while those with a cap gate have around 80 residues here. Phage secretins are therefore expected to have a more compact ß-lip structure than their bacterial counterparts.

At the far C terminus of the protein, the Type II and Type III secretins have a helical domain (the S-domain) that packs against the outside wall of the outer ß-barrel. The S-domain binds pilotin molecules, which are essential for the assembly and/or localisation of secretins to the outer membrane. f1pIV and the Type IV pilus secretins do not have an S-domain, with their respective C-termini being located at the periplasmic side of the outer ß-barrel (Fig. 6c). Interestingly, the Type IV pilus secretins bind to their pilotin partner via a hydrophobic girdle around the ß-barrel, in a similar place to the central hydrophobic girdle that we have observed in f1pIV (Supplementary Fig.10).

We used the Consurf server[44] to perform a sequence alignment of f1pIV homologues, map them to the f1pIV structure and colour them by the degree of sequence conservation (Fig. 6d). The area of greatest conservation lies within the inner ß-barrel and the region of outer ß-barrel that packs against it, demonstrating the importance of this region for protein stability and/or function. The N3 domain is more variable, likely as a result of evolution to suit the

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

function of the individual system. We also compared the electrostatic surface potential between the different classes of secretins (Fig. 6b and Supplementary Fig. 9). The outside surfaces have a mixture of positively and negatively charged areas, with some very clear bands of hydrophobicity, as expected for outer membrane proteins. The surface charge inside the secretins is quite varied; differences presumably arose to accommodate the particular shapes and electrostatic charges of substrate that each secretin transports.

Discussion

f1pIV is the first near-atomic resolution example of a phage secretin. The protein has a similar overall architecture to bacterial secretins, but differs in its electrostatic surface potential inside the pore and in the number of N domains. f1pIV is relatively compact, with no cap gate, no C-terminal S-domain, and one N3 and one N0 domain; an arrangement which could have arisen to save valuable space in the limited capacity of the phage genome.

The phenotypes of many previously characterised mutants can be rationalised with our structure. f1pIV shows extensive hydrogen bonding within the outer and inner ß-barrel domains, as well as hydrophobic interactions and a salt bridge between the inner and outer walls. Mutation of two conserved hydrophobic residues at the interface to glycines (I312G and I314G) prevents multimer formation[24]. In addition, inclusion of a His-tag on the external loop between the inner and outer ß-barrels has a destabilising effect, but this can be rescued by introducing a hydrophobic mutation (S318I)[45], which lines the interface. The conserved hydrophobic residues between inner and outer ß-barrels are clearly important for folding and stability, and highlight the importance of the unusual double-walled architecture.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

In order for the phage to pass through f1pIV, the central gate must open. The conserved architecture of the secretin family can be used to predict the mechanism (Fig. 7). The N-terminal periplasmic N0 domain is extremely dynamic, likely due to the absence of the phage-encoded pI/pXI inner membrane interaction partners. This flexibility may allow f1pIV to sample a wider amount of periplasmic space than a rigid structure, allowing the N0 domains to locate and bind to the inner membrane assembly proteins and egressing phage. pI/pXI interaction with f1pIV is required for phage assembly[46], therefore this suggests that gate opening may be triggered by interaction of the N0 domain of f1pIV with the pI/pXI inner membrane components. It is tempting to speculate that this could result in energy transfer from the pI ATPase to help drive phage egress[13]. The egressing phage will subsequently encounter the N3 domain; the involvement of N3 in channel opening has also been predicted for Type II and Type III secretins[16,32,33,47]. Introducing mutations into the N3 domain of f1pIV causes partial gate opening[24], highlighting its role in transducing information from the periplasmic N domains into the ß-barrel.

Our model of the open state of f1pIV highlights how the central gate loops will likely move to allow phage to egress via the two conserved glycine hinge residues on the extracellular side of the inner ß-barrel[32]. A 10 residue in-frame deletion in Gate 2 results in leaky multimers that are non-functional in phage assembly and egress[24], suggesting that passage of phage through f1pIV is not a passive process. It is plausible that interaction of the f1 cap proteins pVII/pIX with the f1pIV gate will be required to fully open the pore, assisted by the force of the assembling f1 phage. The hydrogen bonding within Gate 1 and Gate 2 means that the loops do not alter their conformation significantly, rather both loops are pivoted around their hinges. The hydrogen bonds between neighbouring monomers means that all gates would move synchronously. The surface of the phage has a high negative charge, while the

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

inner pore surface of the secretin is positively charged in the periplasmic chamber. This opposing charge would presumably help to prevent friction and may drive the continued movement of phage through the pore. *E. coli* expressing the leaky mutant f1pIV^{E292K} exhibited the highest sensitivity to antibiotics and approximately a 10-fold decrease in phage assembly relative to wild-type[24]. The negatively charged sidechain of Glu 292 points into the lower cavity of the pore (Supplementary Fig. 7). Replacing this with a positively charged lysine residue would alter the electrostatic potential in this area, thus interfering with phage interaction and egress.

f1pIV shows most sequence homology to GspD from *Klebsiella pneumoniae* (Supplementary Fig. 11), and thus it is plausible that the ancestral f1 phage captured a Type II secretin which subsequently evolved to meet the needs of phage assembly and egress. Indeed, a number of filamentous phages do not encode a secretin, but rather hijack one from their bacterial host[48,49]. We predict that the outermost hydrophobic girdle in f1pIV forms the membrane embedded domain, as has been proposed for other secretin structures[17,19,20,43]. The other hydrophobic girdle (located half way down the barrel) would then lie within the periplasm, in a similar position to where pilotins are seen to bind to other secretins (Supplementary Fig. 10)[50]. The fact that this hydrophobic girdle has been evolutionarily maintained in f1pIV, even in the absence of a viral pilotin, highlights its functional importance. In support of this, the C-terminus of f1pIV is in close vicinity to the girdle, and C-terminal His-tags result in non-functional f1pIV[45]. In addition, chimeras of f1pIV with C-terminal pilotin-binding S-domains from other secretins (PulD and InvG) are only targeted to the OM and functional for phage assembly in the presence of their respective pilotins (PulS or InvH)[51,52]. The C-terminus and hydrophobic girdle are thus an important interaction surface, with one possibility being that

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

the phage system is able to hijack bacterial pilotins, or pilotin-like proteins, to aid its targeting and assembly.

Uptake of hydrophilic and amphipathic antibiotics have been tested with f1pIV previously, whereas hydrophobic drugs have not. We demonstrate that the leaky mutant f1pIV^{S324G} allows *E. coli* to accumulate the hydrophobic macrolide roxithromycin to a greater extent than its wild type counterpart. Ser 324 mutation to Gly likely destroys important hydrogen bonds and alters the position of the peptide backbone, resulting in a more open gate structure that is non-selective to the hydrophobicity of the molecule passing through. These findings, along with our f1pIV atomic model and the conserved nature of secretins, supports targeting the secretin family as an antimicrobial approach[53,54]. For example, *in silico* and *in vivo* screens could be used to identify molecules that can destabilise the gates, in effect sensitising any secretin-expressing pathogen to a wide range of antibiotics. Phage therapy is also receiving renewed attention, and filamentous phages have been engineered to carry lethal genes to bacteria[55-57]. Leaky f1pIV could be cloned into the genome of phages of interest, and on protein expression channels would open in the host bacterial outer membrane, allowing antibiotics to enter. Both of these approaches would add much needed artillery to our library of therapeutic tools needed to combat the increasing rise of antimicrobial resistance.

Methods

The system of amino acid numbering used throughout is for the mature wild-type f1pIV protein (minus the 21 signal sequence residues). Standard reagents were purchased from Sigma-Aldrich unless otherwise stated.

Bacterial strains

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

All bacterial strains used were *E. coli* K12 derivatives. TG1 electrocompetent cells were purchased from Lucigen and used for expression of f1pIV for cryoEM studies. Strain K2040 [MC4100 *ΔlamB106 degP41* (Δ Pstl-Km^R)] described in[24] was used for the microfluidics experiments.

Expression and purification of f1pIV

His-tagged f1pIV was recombinantly expressed in *E. coli* and purified in the presence of CHAPS detergent (Melford) based on previously published methods[37,58]. The wild type protein was engineered to have a 9 residue His tag (SAHHHHHHH) inserted at position 308 and a stabilizing S318I mutation[45]. Protein expression was induced overnight at 20°C and cells were lysed by sonication. The membrane fraction was solubilized in 5% (w/v) CHAPS, 50 mM Tris pH 7.6, 500 mM NaCl, 30 mM imidazole and incubated with Ni^{2+} Sepharose beads. f1pIV was eluted with a 0.03-1 M imidazole gradient, and then further purified by gel filtration chromatography on a Superose 6 Increase 10/300 GL column (GE Healthcare) in 1% (w/v) CHAPS, 25 mM Na HEPES pH 8.0, 500 mM NaCl, 0.5 mM EDTA. Purified f1pIV was separated on an Any kD mini-protean TGX gel (Biorad), electrophoretically transferred to PVDF membrane and incubated with an anti-pIV antibody.

Cryo-EM grid preparation and data collection

The f1pIV purified sample (3 µl of ~0.7mg/ml) was applied to graphene oxide-coated lacey carbon grids, 300 mesh (Agar Scientific) without any glow discharge, and frozen on a Mark IV Vitrobot (Thermo Fisher Scientific, 4°C, 100% relative humidity, blot force 0, blot time 4 sec). Micrographs were collected on a 300 kV Titan Krios microscope (Thermo Fisher Scientific) with a K3 direct electron detector (Gatan) at the Electron Bio-imaging Centre (eBIC) at Diamond Light Source, UK. Data were collected using EPU software with a pixel

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

size of 1.072 Å (0.536 Å super-resolution) and a defocus range from -2.5 µm to -1.3 µm. Further details are shown in Supplementary Table 1. A subset of data was collected in normal mode (dataset 1), and a second subset in super-resolution mode (dataset 2).

Cryo-EM image processing

Warp[59] was used for motion correction, CTF correction and particle picking. 241,591 particles were picked from 7037 micrographs for dataset 1 and 330,389 from 14,336 for dataset 2. Both sets of data were processed separately and combined in the latter stages of data processing. Several rounds of 2D classification and 3D refinement were implemented in Relion[23] followed by CTF refinement and polishing steps. The statistics of data collection and model reconstruction are shown in Supplementary Table 1, and data quality checks[60,61] are shown in Supplementary Fig. 2.

Model building and refinement

A homologous structure of the Type II secretin GspD from Enteropathogenic *Escherichia coli* (5W68) was manually placed in the cryoEM map using Chimera[62]. Subsequent building and model adjustments were performed using Coot[63]. Bulky residues, glycine residues and unique sequence patterns were used to guide sequence assignment during model building. 15-fold symmetry was applied in Chimera, and the multimeric structure was refined using Refmac from the CCPEM suite[64], with the quality of the model evaluated using the validation tools in Coot and Molprobity[65]. DeepEMhancer[66] was used for denoising and postprocessing of the maps, which were then analysed to check for additional information that could be modelled from the improved maps. A Fourier Shell Correlation (FSC) curve showing the quality of the model to map fit (calculated with Phenix)[61] is shown in Supplementary Fig. 2.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Mass spectrometry analysis

A sample of f1pIV was run on an SDS-PAGE gel (Any kD mini-protean TGX, Biorad), stained with Coomassie, and the band of interest excised with a sterile scalpel blade. The gel band was subjected to in-gel tryptic digestion using a DigestPro automated digestion unit (Intavis Ltd.) and the resulting peptides were fractionated using an Ultimate 3000 nano-LC system in line with an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific). In brief, peptides in 1% (vol/vol) formic acid were injected onto an Acclaim PepMap C18 nano-trap column (Thermo Scientific). After washing with 0.5% (vol/vol) acetonitrile 0.1% (vol/vol) formic acid peptides were resolved on a 250 mm × 75 µm Acclaim PepMap C18 reverse phase analytical column (Thermo Scientific) over a 150 min organic gradient, using 7 gradient segments (1-6% solvent B over 1 min., 6-15% B over 58 min., 15-32%B over 58 min., 32-40%B over 5 min., 40-90%B over 1 min., held at 90%B for 6min and then reduced to 1%B over 1min.) with a flow rate of 300 nl min−1. Solvent A was 0.1% formic acid and Solvent B was aqueous 80% acetonitrile in 0.1% formic acid. Peptides were ionized by nano-electrospray ionization at 2.2 kV using a stainless-steel emitter with an internal diameter of 30 µm (Thermo Scientific) and a capillary temperature of 250°C.

All spectra were acquired using an Orbitrap Fusion Tribrid mass spectrometer controlled by Xcalibur 2.1 software (Thermo Scientific) and operated in data-dependent acquisition mode. FTMS1 spectra were collected at a resolution of 120 000 over a scan range (m/z) of 350-1550, with an automatic gain control (AGC) target of 400 000 and a max injection time of 100ms. Precursors were filtered according to charge state (to include charge states 2-7), with monoisotopic peak determination set to peptide and using an intensity range from 5E3 to 1E20. Previously interrogated precursors were excluded using a dynamic window (40 s +/-10 ppm). The MS2 precursors were isolated with a quadrupole mass filter set to a width of 1.6

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

m/z. ITMS2 spectra were collected with an AGC target of 5000, max injection time of 50 ms and HCD collision energy of 35%.

The raw data files were processed and quantified using Proteome Discoverer software v1.4 (Thermo Scientific) and searched against the UniProt *Escherichia coli* database (4349 sequences) plus the supplied amino acid sequence using the SEQUEST algorithm. Peptide precursor mass tolerance was set at 10 ppm, and MS/MS tolerance was set at 0.8 Da. Search criteria included carbamidomethylation of cysteine (+57.0214) as a fixed modification and oxidation of methionine (+15.9949) as a variable modification. Searches were performed with full tryptic digestion and a maximum of 1 missed cleavage was allowed. The reverse database search option was enabled and all peptide data was filtered to satisfy false discovery rate (FDR) of 5%.

Sequence alignments and homology modelling

The sequence of the N terminal region of the model that was not accounted for in our structure (residues 1-107) was subjected to a Blast search[67] against all sequences in the Protein Data Bank. A sequence alignment of the top hit was performed on f1pIV with Praline[68] and Swiss-Model[25] used to create a homology model of the f1pIV N0 domain. In addition, the sequence of the f1pIV N0 domain was submitted to the I-TASSER structural prediction server[26] (which gave 5 top models based on a number of different sequences), and also to the Robetta structural prediction server[27] (which also produced 5 top models). All models were in agreement for residues 1-71 where they started to diverge. Open and closed gate structures of InvG from *Salmonella typhimurium* were superimposed with the closed structure of f1pIV, and Gate 1 and 2 loops were modelled into an open position using Coot. 15-fold symmetry was applied to produce a model of the f1pIV open state.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The Consurf server[44] was used to perform a sequence alignment of f1pIV homologues (101 unique sequences with 35-100% homology from the Uni-Prot database were used for this analysis), map them to the f1pIV structure and colour them by the degree of sequence conservation.

Determination of minimum inhibitory concentration

E. coli cells were freshly transformed with plasmid pPMR132^{WT} or pPMR132^{S324G} and single colonies of *E. coli* f1pIV^{WT} or *E. coli* f1pIV^{S324G} were picked and cultured at 37 °C in FB media (2.5% (w/v) tryptone, 0.75% (w/v) yeast extract, 0.6% (w/v) NaCl, 0.1% (w/v) glucose, 50 mM TrisHCl (pH 7.5)). Plasmid DNA was miniprepped and sequenced for each experiment to ensure there had been no additional mutations. Protein expression was induced overnight (in FB media plus 0.5 mg/ml bovine serum albumin) with 1mM IPTG at 20 °C, then diluted 40-fold and grown to OD_{600} = 0.5. 60 µl of roxithromycin stock (640 µg/ml in DMSO) were added to the first column of a 96-well plate. 40 µl of the induction media (FB media, 1mM IPTG, 0.5 mg/ml BSA) was added to the first column, and 30 µl to all other wells. 70 µl solution was then withdrawn from the first column and serially transferred to the next column until 70 µl solution withdrawn from the last column was discharged. The mid-log phase cultures (i.e. OD_{600} = 0.5) were diluted to 10^6 colony forming units (c.f.u.)/ml and 30 µl was added to each well, to give a final concentration of 5×10^5 c.f.u/ml. Each plate contained two rows of 12 positive control experiments (i.e. bacteria growing in induction media without roxithromycin) and two rows of 12 negative control experiments (i.e. induction media only). Plates were incubated at 37 °C overnight. The minimum inhibitory concentrations (MICs) of roxithromycin against *E. coli* f1pIV^{WT} and *E. coli* f1pIV^{S324G} were determined visually, with the MIC being the lowest concentration well with no visible growth (compared to the positive control experiments).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Single-cell microfluidics

Single-cell microfluidics experiments to measure the intracellular accumulation of roxithromycin-NBD were carried out as previously reported[38,40,42]. Briefly, overnight cultures were prepared as described above and a 50 ml aliquot was centrifuged for 5 min at 4000 rpm and 37 °C and resuspended at an OD_{595} of 75. A 2 µl aliquot of this suspension was injected in the microfluidic mother machine device and incubated at 37 °C. The microfluidic device was completed by the integration of fluorinated ethylene propylene tubing (1/32" × 0.008"). The inlet tubing was connected to the inlet reservoir which was connected to a computerised pressure-based flow control system (MFCS-4C, Fluigent). The microfluidic device was mounted on an inverted microscope (IX73 Olympus, Tokyo, Japan) and the main microchamber of the mother machine was washed into the waste reservoir by flowing media at 100 µl/h for 2 h. Images were collected via a 60×, 1.2 N.A. objective (UPLSAPO60XW, Olympus) and a sCMOS camera (Zyla 4.2, Andor, Belfast, UK). After this initial 2 h growth period in growth media, the microfluidic environment was changed by flowing media containing roxithromycin-NBD at a concentration of 46 µg/ml. Upon acquiring a bright-field image the microscope was switched to fluorescent mode and FITC filter using custom built Labview software. A fluorescence image was acquired by exposing the bacteria to the blue excitation band of a broad-spectrum LED (CoolLED pE300white, Andover, UK) at 20% of its intensity. Bright-field and fluorescence imaging was carried out every 10 min. The entire assay was carried out at 37 °C in an environmental chamber (Solent Scientific, Portsmouth, UK) surrounding the microscope and microfluidics equipment. Images were processed as previously described[38,40,42].

Data availability

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Density maps of f1pIV have been deposited in the Electron Microscopy Data Bank (EMDB, https://www.ebi.ac.uk/pdbe/emdb/) with entry code EMD-12874. A model of f1pIV has been deposited in the Protein Data Bank (PDB, https://www.rcsb.org/) with accession number 7OFH.

Acknowledgements

This work was funded by a Wellcome Trust Seed Award in Science awarded to V.G. (210363/Z/18/Z), which along with the University of Exeter, supported R.C. M.M and K.S. were supported by a BBSRC responsive mode grant awarded to V.G. (BB/R008639/1). U.L. was supported through a BBSRC responsive mode grant (BB/V008021/1), an MRC Proximity to Discovery EXCITEME2 grant (MCPC17189) and an award from the Gordon and Betty Moore Foundation Marine Microbiology Initiative (GBMF5514) awarded to S.P. M.R.L.S. was supported by an Australian Postgraduate Award (APA), PhD scholarship and an IMB Research Advancement Award. We acknowledge Diamond Light Source for access and support of the cryo-EM facilities at the UK's national Electron Bio-imaging Centre (eBIC) at Diamond Light Source [under proposal BI25452], funded by the Wellcome Trust, MRC and BBRSC. We acknowledge access and support of the GW4 Facility for High-Resolution Electron Cryo-Microscopy, funded by the Wellcome Trust (202904/Z/16/Z and 206181/Z/17/Z) and BBSRC (BB/R000484/1). The deposited dataset was collected at eBIC, and the GW4 facility was used for sample screening. We are grateful to Ufuk Borucu of the GW4 Regional Facility for High-Resolution Electron Cryo-Microscopy for help with screening, Kate Heesom of University of Bristol Proteomics Facility for the mass spectrometry analysis, and to Marjorie Russel for valuable comments on the manuscript, advice, and her generous gifts of the f1pIV expression vectors (pPMR132 and pPMR132S324G) and anti-f1pIV antibody.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Author contributions

R.C. expressed and purified the protein, prepared samples for cryoEM, processed cryoEM data, built the atomic model, interpreted results and prepared figures. M.M. collected and processed cryoEM data. U.L. performed microfluidics experiments and analysed single-cell data. K.S. prepared samples for single-cell microfluidic experiments and supported R.C. S.P. designed single-cell microfluidics experiments. M.R.L.S and M.A.T.B designed and synthesised the macrolide fluorescent probes used in the microfluidics experiments. B.D. interpreted data. J.R. and V.A.M.G. conceptualised the project; V.A.M.G. designed the research, obtained funding for the project and wrote the manuscript with R.C. All authors commented on the manuscript.

Competing interests statement

The authors declare no competing interests.

References

1 Marvin, D. A., Symmons, M. F. & Straus, S. K. Structure and assembly of filamentous bacteriophages. *Prog Biophys Mol Biol* **114**, 80-122, doi:10.1016/j.pbiomolbio.2014.02.003 (2014).

2 Rakonjac, J. Filamentous Bacteriophages: Biology and Applications. *eLS*, 1-13, doi:10.1002/9780470015902.a0000777 (2012).

3 Szot-Karpińska, K. *et al.* Modified Filamentous Bacteriophage as a Scaffold for Carbon Nanofiber. *Bioconjugate Chemistry* **27**, 2900-2910, doi:10.1021/acs.bioconjchem.6b00555 (2016).

4 Mao, C. *et al.* Virus-Based Toolkit for the Directed Synthesis of Magnetic and Semiconducting Nanowires. *Science* **303**, 213-217, doi:10.1126/science.1092740 (2004).

5 Nathan, C. Resisting antimicrobial resistance. *Nature Reviews Microbiology* **18**, 259-260, doi:10.1038/s41579-020-0348-5 (2020).

6 McKenna, M. The antibiotic paradox: why companies can't afford to create life-saving drugs. *Nature* **584**, 338-341, doi:10.1038/d41586-020-02418-x (2020).

7 Brives, C. & Pourraz, J. Phage therapy as a potential solution in the fight against AMR: obstacles and possible futures. *Palgrave Communications* **6**, 1-11, doi:10.1057/s41599-020-0478-4 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

8 Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327-336, doi:10.1038/s41586-019-1894-8 (2020).

9 Rakonjac, J., Bennett, N. J., Spagnuolo, J., Gagic, D. & Russel, M. Filamentous bacteriophage: biology, phage display and nanotechnology applications. *Current issues in molecular biology* **13**, 51-76, doi:10.21775/cimb.013.051 (2011).

10 Riechmann, L. & Holliger, P. The C-terminal domain of TolA is the coreceptor for filamentous phage infection of E. coli. *Cell* **90**, 351-360, doi:10.1016/S0092-8674(00)80342-6 (1997).

11 Hay, I. D. & Lithgow, T. Filamentous phages: masters of a microbial sharing economy. *EMBO reports* **20**, doi:10.15252/embr.201847427 (2019).

12 Russel, M. & Kazmierczak, B. Analysis of the structure and subcellular location of filamentous phage pIV. *Journal of Bacteriology* **175**, 3998-4007, doi:10.1128/jb.175.13.3998-4007.1993 (1993).

13 Feng, J. n., Model, P. & Russel, M. A trans-envelope protein complex needed for filamentous phage assembly and export. *Molecular Microbiology* **34**, 745-755, doi:10.1046/j.1365-2958.1999.01636.x (1999).

14 Green, E. R. & Mecsas, J. Bacterial Secretion Systems: An Overview. *Microbiology Spectrum* **4**, 1-32, doi:10.1128/microbiolspec.vmbf-0012-2015 (2016).

15 D'Imprima, E. *et al.* Cryo-EM structure of the bifunctional secretin complex of Thermus thermophilus. *eLife* **6**, 1-23, doi:10.7554/eLife.30483 (2017).

16 Hu, J. *et al.* Cryo-EM analysis of the T3S injectisome reveals the structure of the needle and open secretin. *Nature Communications* **9**, 1-11, doi:10.1038/s41467-018-06298-8 (2018).

17 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. *Journal of Bacteriology* **200**, 1-15, doi:10.1128/JB.00521-17 (2018).

18 Yin, M., Yan, Z. & Li, X. Structural insight into the assembly of the type II secretion system pilotin-secretin complex from enterotoxigenic Escherichia coli. *Nature Microbiology* **3**, 581-587, doi:10.1038/s41564-018-0148-0 (2018).

19 Filloux, A. & Voulhoux, R. Multiple structures disclose the secretins' secrets. *Journal of Bacteriology* **200**, 1-5, doi:10.1128/JB.00702-17 (2018).

20 Majewski, D. D., Worrall, L. J. & Strynadka, N. C. J. Secretins revealed: structural insights into the giant gated outer membrane portals of bacteria. *Current Opinion in Structural Biology* **51**, 61-72, doi:10.1016/j.sbi.2018.02.008 (2018).

21 Naskar, S., Hohl, M., Tassinari, M. & Low, H. H. The structure and mechanism of the bacterial type II secretion system. *Molecular Microbiology* **115**, 412-424, doi:10.1111/mmi.14664 (2021).

22 Opalka, N. *et al.* Structure of the filamentous phage pIV multimer by cryo-electron microscopy. *Journal of Molecular Biology* **325**, 461-470, doi:10.1016/S0022-2836(02)01246-9 (2003).

23 Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, 1-22, doi:10.7554/eLife.42166 (2018).

24 Spagnuolo, J. *et al.* Identification of the gate regions in the primary structure of the secretin pIV. *Molecular Microbiology* **76**, 133-150, doi:10.1111/j.1365-2958.2010.07085.x (2010).

25 Waterhouse, A. *et al.* SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**, W296-W303, doi:10.1093/nar/gky427 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

26 Yang, J. *et al.* The I-TASSER suite: Protein structure and function prediction. *Nature Methods* **12**, 7-8, doi:10.1038/nmeth.3213 (2014).

27 Baek, M. *et al.* Accurate prediction of protein structures and interactions using a 3-track network. *bioRxiv*, 1-17, doi:10.1101/2021.06.14.448402 (2021).

28 Chernyatina, A. A. & Low, H. H. Core architecture of a bacterial type II secretion system. *Nature Communications* **10**, 1-10, doi:10.1038/s41467-019-13301-3 (2019).

29 Vollmer, W. & Seligman, S. J. Architecture of peptidoglycan: more data and more models. *Trends in Microbiology* **18**, 59-66, doi:10.1016/j.tim.2009.12.004 (2010).

30 Chaptal, V. *et al.* Quantification of Detergents Complexed with Membrane Proteins. *Scientific Reports* **7**, 1-12, doi:10.1038/srep41751 (2017).

31 Schulz, G. E. β-barrel membrane proteins. *Current Opinion in Structural Biology* **10**, 443-447, doi:10.1016/S0959-440X(00)00120-2 (2000).

32 Yan, Z., Yin, M., Xu, D., Zhu, Y. & Li, X. Structural insights into the secretin translocation channel in the type II secretion system. *Nature Structural and Molecular Biology* **24**, 177-183, doi:10.1038/nsmb.3350 (2017).

33 Hu, J. *et al.* T3S injectisome needle complex structures in four distinct states reveal the basis of membrane coupling and assembly. *Nature Microbiology* **4**, 2010-2019, doi:10.1038/s41564-019-0545-z (2019).

34 Marvin, D. A., Welsh, L. C., Symmons, M. F., Scott, W. R. P. & Straus, S. K. Molecular structure of fd (f1, M13) filamentous bacteriophage refined with respect to X-ray fibre diffraction and solid-state NMR data supports specific models of phage assembly at the bacterial membrane. *Journal of Molecular Biology* **355**, 294-309, doi:10.1016/j.jmb.2005.10.048 (2006).

35 Russel, M. Mutants at conserved positions in gene IV, a gene required for assembly and secretion of filamentous phages. *Molecular Microbiology* **14**, 357-369, doi:10.1111/j.1365-2958.1994.tb01296.x (1994).

36 Russel, M., Linderoth, N. A. & Šali, A. Filamentous phage assembly: Variation on a protein export theme. *Gene* **192**, 23-32, doi:10.1016/S0378-1119(96)00801-3 (1997).

37 Marciano, D. K., Russel, M. & Simon, S. M. An aqueous channel for filamentous phage export. *Science* **284**, 1516-1519, doi:10.1126/science.284.5419.1516 (1999).

38 Stone, M. R. L. *et al.* Fluorescent macrolide probes – synthesis and use in evaluation of bacterial resistance. *RSC Chemical Biology* **1**, 395-404, doi:10.1039/d0cb00118j (2020).

39 Bryskier, A. Roxithromycin: Review of its antimicrobial activity. *Journal of Antimicrobial Chemotherapy* **41**, 1-21, doi:10.1093/jac/41.suppl_2.1 (1998).

40 Łapińska, U., Glover, G., Capilla-Lasheras, P., Young, A. J. & Pagliara, S. Bacterial ageing in the absence of external stressors. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 1-13, doi:10.1098/rstb.2018.0442 (2019).

41 Bamford, R. A. *et al.* Investigating the physiology of viable but non-culturable bacteria by microfluidics and time-lapse microscopy. *BMC Biology* **15**, 1-12, doi:10.1186/s12915-017-0465-4 (2017).

42 Cama, J. *et al.* Single-cell microfluidics facilitates the rapid quantification of antibiotic accumulation in Gram-negative bacteria. *Lab on a Chip* **20**, 2765-2775, doi:10.1039/d0lc00242a (2020).

43 Weaver, S. J. *et al.* CryoEM structure of the type IVa pilus secretin required for natural competence in Vibrio cholerae. *Nature Communications* **11**, 1-13, doi:10.1038/s41467-020-18866-y (2020).

44 Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research* **44**, W344-W350, doi:10.1093/nar/gkw408 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

45 Linderoth, N. A., Model, P. & Russel, M. Essential role of a sodium dodecyl sulfate-resistant protein IV multimer in assembly-export of filamentous phage. *Journal of Bacteriology* **178**, 1962-1970, doi:10.1128/jb.178.7.1962-1970.1996 (1996).

46 Russel, M. Protein-protein interactions during filamentous phage assembly. *Journal of Molecular Biology* **231**, 689-697, doi:10.1006/jmbi.1993.1320 (1993).

47 Worrall, L. J. *et al.* Near-atomic-resolution cryo-EM analysis of the Salmonella T3S injectisome basal body. *Nature* **540**, 597-601, doi:10.1038/nature20576 (2016).

48 Davis, B. M. *et al.* Convergence of the secretory pathways for cholera toxin and the filamentous phage, CTXφ. *Science* **288**, 333-335, doi:10.1126/science.288.5464.333 (2000).

49 Bille, E. *et al.* A chromosomally integrated bacteriophage in invasive meningococci. *Journal of Experimental Medicine* **201**, 1905-1913, doi:10.1084/jem.20050112 (2005).

50 McCallum, M., Tammam, S., Rubinstein, J. L., Burrows, L. L. & Howell, P. L. CryoEM map of Pseudomonas aeruginosa PilQ enables structural characterization of TsaP. *Structure* **29**, 457-466, doi:10.1016/j.str.2020.11.019 (2021).

51 Daefler, S., Guilvout, I., Hardie, K. R., Pugsley, A. P. & Russel, M. The C-terminal domain of the secretin PulD contains the binding site for its cognate chaperone, PulS, and confers PulS dependence on pIV(f1) function. *Molecular Microbiology* **24**, 465-475, doi:10.1046/j.1365-2958.1997.3531727.x (1997).

52 Daefler, S. & Russel, M. The Salmonella typhimurium InvH protein is an outer membrane lipoprotein required for the proper localization of InvG. *Molecular Microbiology* **28**, 1367-1380, doi:10.1046/j.1365-2958.1998.00908.x (1998).

53 Baron, C. & Coombes, B. Targeting Bacterial Secretion Systems: Benefits of Disarmament in the Microcosm. *Infectious Disorders - Drug Targets* **7**, 19-27, doi:10.2174/187152607780090685 (2008).

54 McShan, A. C. & De Guzman, R. N. The Bacterial Type III Secretion System as a Target for Developing New Antibiotics. *Chemical Biology and Drug Design* **85**, 30-42, doi:10.1111/cbdd.12422 (2015).

55 Hagens, S. & Bläsi, U. Genetically modified filamentous phage as bactericidal agents: A pilot study. *Letters in Applied Microbiology* **37**, 318-323, doi:10.1046/j.1472-765X.2003.01400.x (2003).

56 Moradpour, Z. *et al.* Genetically engineered phage harbouring the lethal catabolite gene activator protein gene with an inducer-independent promoter for biocontrol of Escherichia coli. *FEMS Microbiology Letters* **296**, 67-71, doi:10.1111/j.1574-6968.2009.01620.x (2009).

57 Westwater, C. *et al.* Use of Genetically Engineered Phage To Deliver Antimicrobial Agents to Bacteria: an Alternative Therapy for Treatment of Bacterial Infections. *Antimicrobial Agents and Chemotherapy* **47**, 1301-1307, doi:10.1128/AAC.47.4.1301-1307.2003 (2003).

58 Linderoth, N. A., Simon, M. N. & Russel, M. The filamentous phage pIV multimer visualized by scanning transmission electron microscopy. *Science* **278**, 1635-1638, doi:10.1126/science.278.5343.1635 (1997).

59 Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nature Methods* **16**, 1146-1152, doi:10.1038/s41592-019-0580-y (2019).

60 Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature Methods* **11**, 63-65, doi:10.1038/nmeth.2727 (2014).

61 Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology* **75**, 861-877, doi:10.1107/S2059798319011471 (2019).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.20.453082; this version posted July 21, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

62 Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

63 Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).

64 Wood, C. *et al.* Collaborative computational project for electron cryo-microscopy. *Acta Crystallographica Section D: Biological Crystallography* **71**, 123-126, doi:10.1107/S1399004714018070 (2015).

65 Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science* **27**, 293-315, doi:10.1002/pro.3330 (2018).

66 Sanchez-Garcia, R. *et al.* DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *bioRxiv*, doi:10.1101/2020.06.12.148296 (2020).

67 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).

68 Simossis, V. A. & Heringa, J. PRALINE: A multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Research* **33**, W289-W294, doi:10.1093/nar/gki390 (2005).

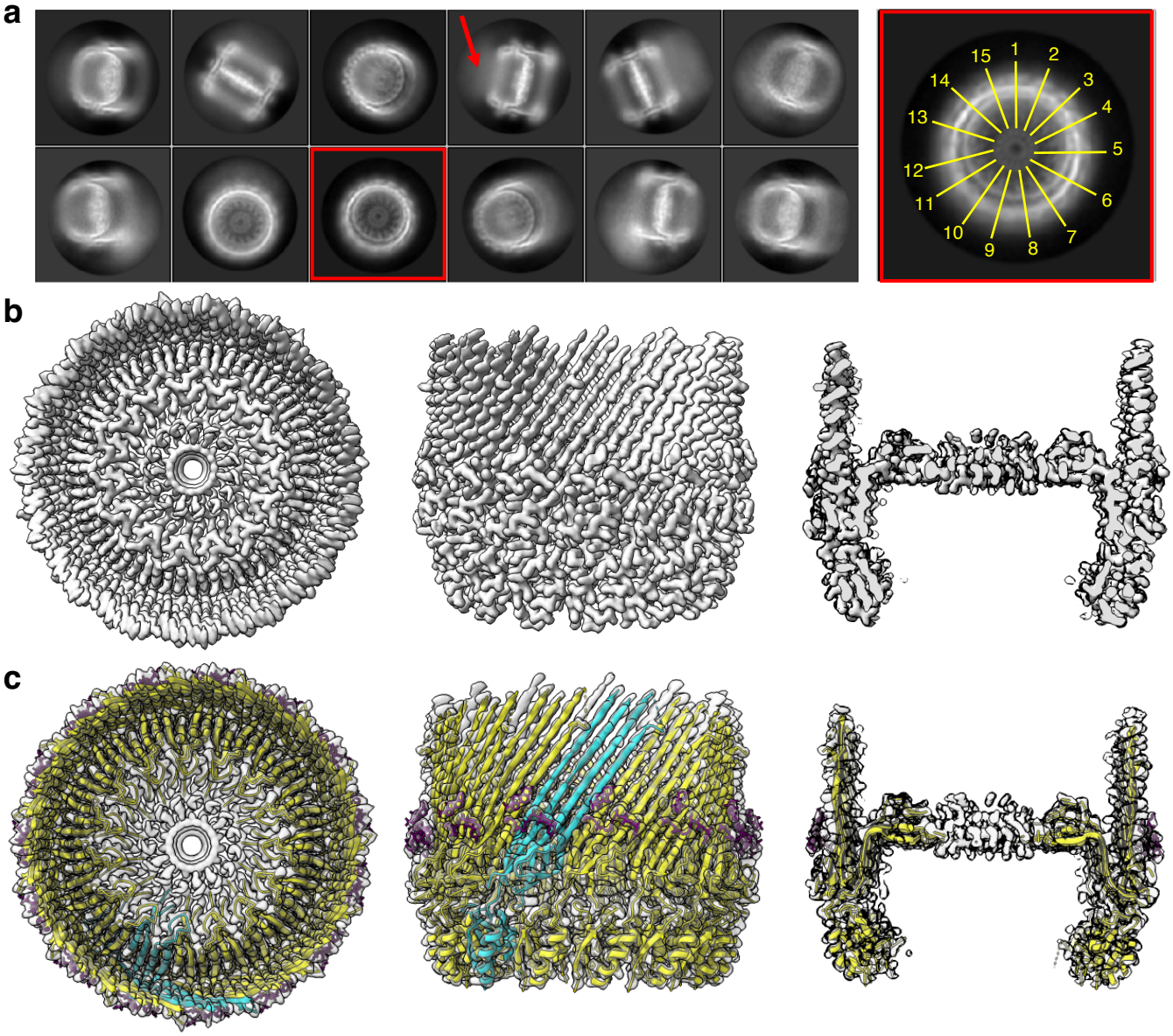


Figure 1. f1pIV cryoEM data and atomic model. a) Left, 2D class averages from Relion. A red arrow highlights a hazy area, likely resulting from protein flexibility. The red box highlights an enlarged top view (right) depicting observed 15-fold symmetry (indicated with yellow lines). **b)** Final 3D reconstruction shown from left to right as top view, side view and side view sliced through the centre. **c)** f1pIV protein structure modelled into the experimental density map shown as a top view, side view and side view sliced through the centre. A single f1pIV monomer is shown as a cyan cartoon with the remaining 14 monomers in yellow; the EM map is shown in light grey. A band of CHAPS detergent molecules are shown as purple sticks. The structure has been deposited with the PDB accession number 7OFH.

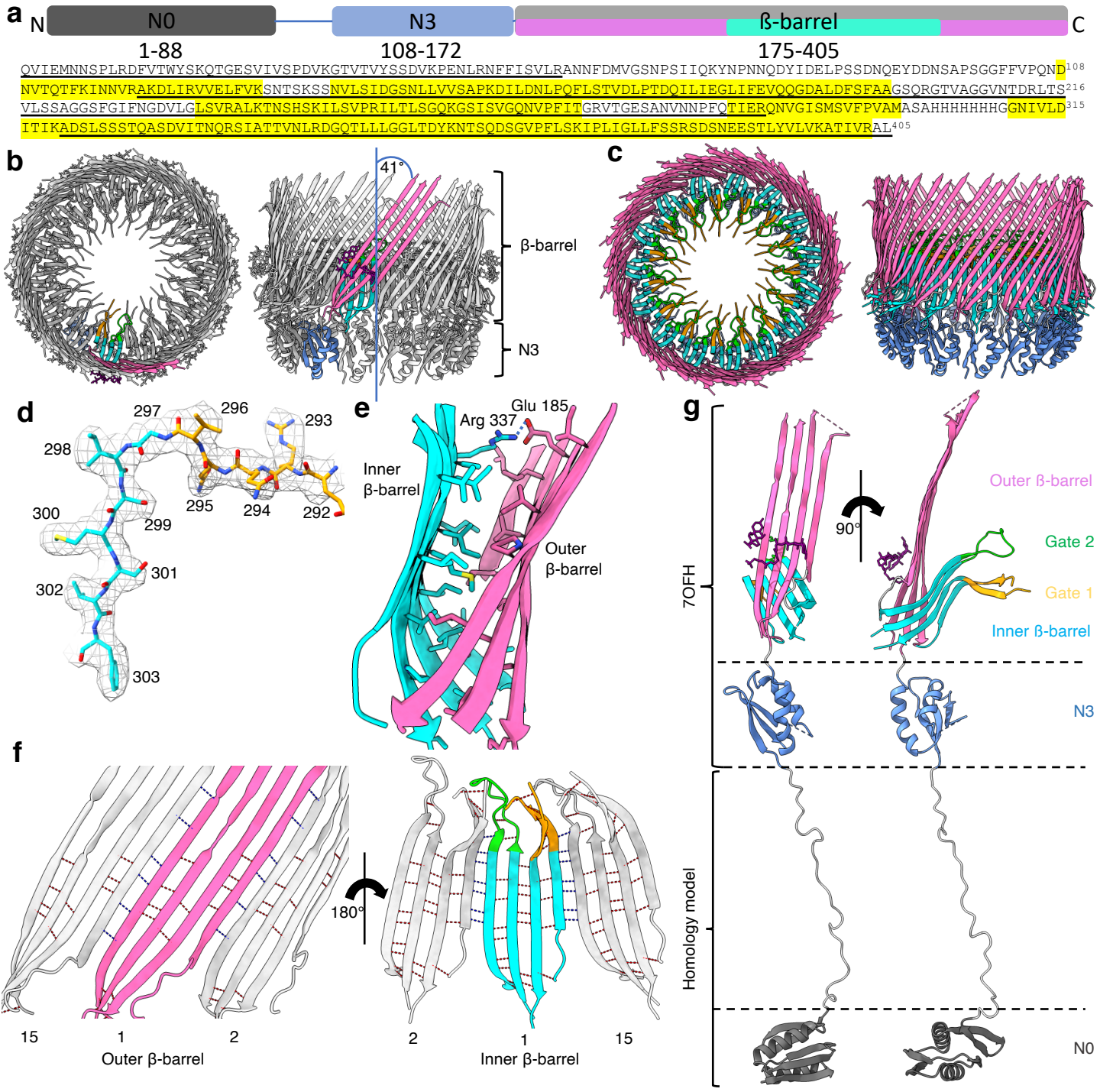


Figure 2. Structural details of f1pIV. a) Domain structure and amino acid sequence of f1pIV, showing the N0 domain (dark grey), the N3 domain (blue), and the secretin domain (light grey). The regions forming the outer β-barrel are shown in pink, and those forming the inner β-barrel in cyan. Residues highlighted in yellow in the amino acid sequence were observed in the cryoEM map and have been modelled into the density. Underlined residues were confirmed to be present in the cryoEM sample by mass spectrometry analysis. **b)** Cartoon representation of the f1pIV multimer (top and side views) with a single molecule coloured as follows: N3 domain in blue, outer β-barrel in pink, inner β-barrel in cyan, Gate 1 in orange, Gate 2 in green, and CHAPS molecules as purple sticks. **c)** Cartoon representation of the f1pIV multimer with all chains coloured as in b). CHAPS molecules have been removed for clarity. **d)** Residues 292-303 (incorporating part of the Gate 1 loop and the inner β-barrel) showing representative density. **e)** The interface between the inner and outer β-barrels is lined mostly with hydrophobic residues which are shown as sticks, and coloured by atom (N in blue, O in red, S in yellow). A salt bridge is formed at the edge of this interface between conserved residues Glu 185 and Arg 337 (dashed blue line). **f)** Inter- (blue dashed lines) and intra- (red dashed lines) subunit hydrogen bonds are formed connecting the inner and outer β-barrels. One subunit (position 1) is coloured as in b), and the neighbouring subunits (in positions 2 and 15) are shown in grey. Images have been clipped for clarity. **g)** Composite model of the f1pIV monomer shown as front and side views. The structure built into the EM map is coloured as in b), with a homology model of the N0 domain in dark grey.

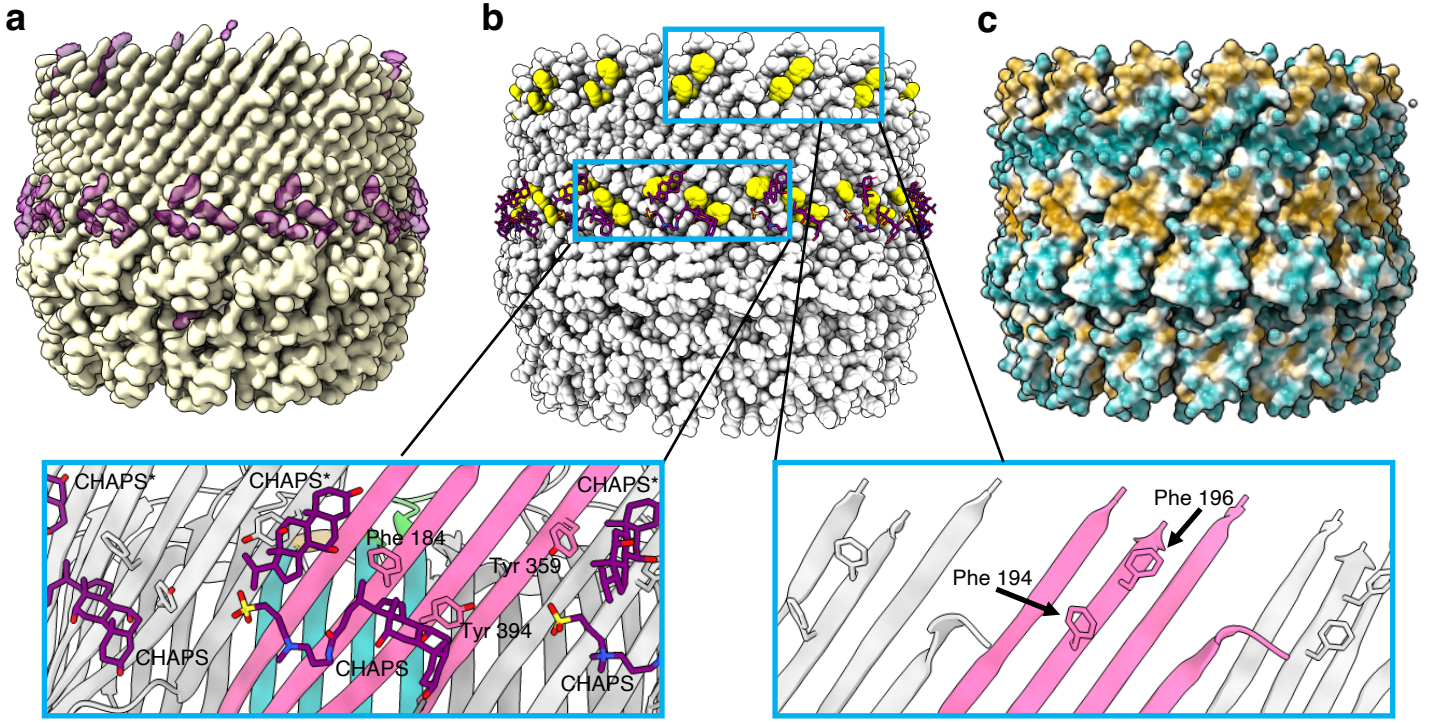


Figure 3. Positioning of the f1pIV multimer in the *E. coli* outer membrane. **a)** The f1pIV cryoEM map in side view (yellow) shows density that was unaccounted for after the protein had been modelled (purple). **b)** The f1pIV atomic model is shown as space-filled atoms in light grey with the CHAPS shown as purple sticks. There are two bands of aromatic residues shown in yellow (aromatic girdles). The enlarged panels (blue boxes) show the two girdles in greater detail with the aromatic residues labelled. One full molecule of CHAPS, and one consisting only of the aromatic rings (CHAPS*) were modelled in to the ring of density. CHAPS is shown as purple sticks, and the f1pIV monomers are coloured as in Fig. 2f. **c)** f1pIV atomic model shown as surface representation is coloured by hydrophobicity, ranging from the most hydrophilic areas (dark cyan), through white, to the most hydrophobic areas (orange).

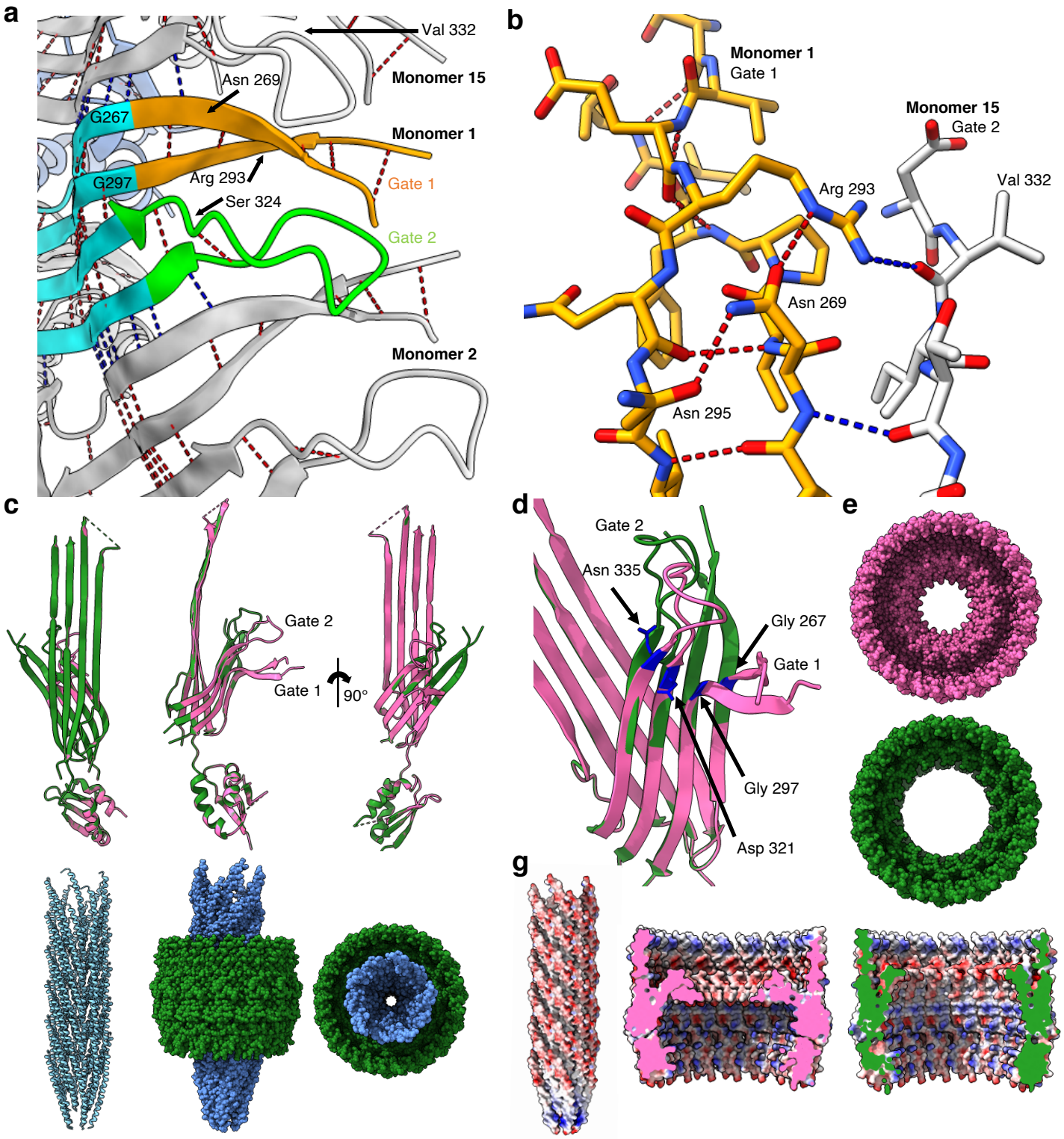


Figure 4. The Gate region, channel opening, and interaction with phage. a) Extensive intra- (red dashed lines) and inter- (blue dashed lines) gate hydrogen bonding is seen in Gate 1 and Gate 2, and between neighbouring subunits, colouring as in Fig. 2f. Glycine 267 (G267) and Glycine 297 (G297) hinge residues are shown, along with key residues Asn 269 and Arg 293 in Gate 1, and Ser 324 and Val 332 in Gate 2. **b)** Close up to show the hydrogen bonding interactions of the conserved and buried Arg 293 in Gate 1. Arg 293 forms hydrogen bonds within Gate 1 with a conserved Asn 269, and with Val 332 in Gate 2 of the neighbouring monomer. Gate 1 from monomer 1 is shown in orange and Gate 2 from monomer 15 in grey, and coloured by atom (N in blue, O in red). **c)** Structural superimposition of the f1pIV monomer closed state structure (pink) with f1pIV modelled in the open gate position (green), shown in front, side and back views. **d)** Close up showing the Gate 1 (Gly 267, Gly 297) and Gate 2 (Asp 321, Asn 335) hinge residues in blue. **e)** The closed (pink) and open (green) multimers shown as space-fill representation in top view. **f)** Left, structure of an fd virion (2C0W) in cartoon form. Centre and right, manual docking of the fd virion (blue space-filled representation) into the open f1pIV secretin pore (green surface representation) in side and top views respectively. **g)** Left, electrostatic surface potential of the fd virion coloured red for negative and blue for positive. Centre and right, electrostatic surface potential of the closed (pink) and open (green) states of the f1pIV pore, coloured as per the fd virion.

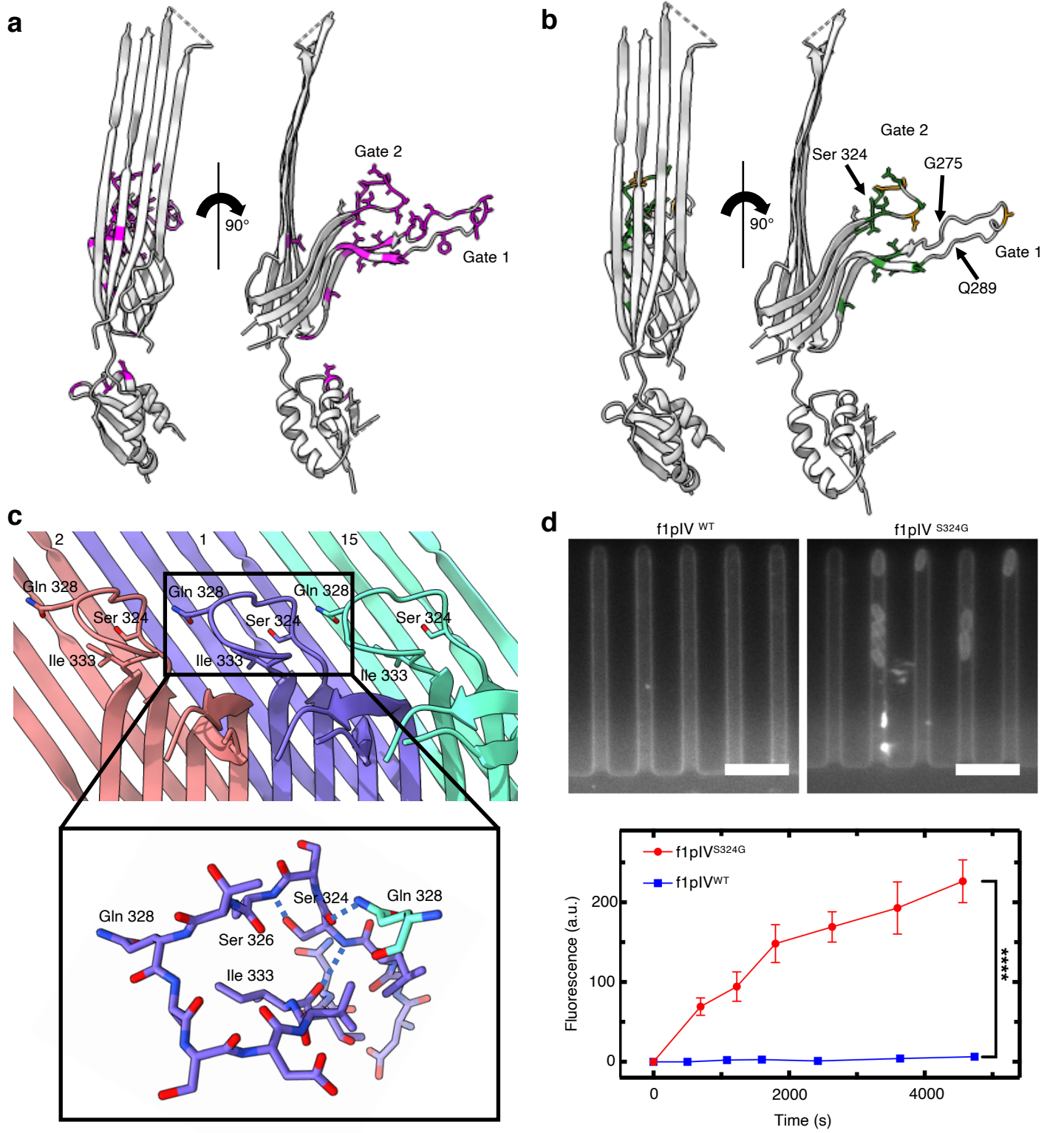


Figure 5. Antibiotic susceptibility in *E. coli* expressing leaky mutants of f1pIV. a) Front and side views showing all mutants[24] that were leaky to both maltopentaose and deoxycholate as magenta sticks. **b)** Front and side views showing the mutations which caused sensitivity to the antibiotics vancomycin and bacitracin as green sticks. Mutations which caused sensitivity to vancomycin only are shown as orange sticks. In parts a) and b) the 15-residue loop from residues G275-Q289 (labelled), which is disordered in our structure, has been modelled in for representation purposes. The position of the Ser 324 mutation to Gly (S324G), analysed in part d), is also shown. **c)** Gate 2 loop showing the key hydrogen bond interactions made by Ser 324. Monomer 1 is shown in purple, with neighbouring monomers 2 and 15 in red and green respectively. The close up shows that Ser 324 forms hydrogen bonds (blue dashed lines) with Ile 333 and Ser 326 within the same monomer, and with Gln 328 from a neighbouring monomer (N in blue, O in red). **d)** Top, individual *E. coli* cells expressing f1pIV^{S324G} take up fluorescently labelled roxithromycin. No fluorescence change was observed for the cells expressing wild type f1pIV (f1pIV^{WT}). Scale bar, 5 µm. Bottom, scatter graph depicting the fluorescence change for 350 cells for f1pIV^{WT} and f1pIV^{S324G} samples, confirming roxithromycin uptake in the S324G mutant. Error bars represent standard error of the mean.

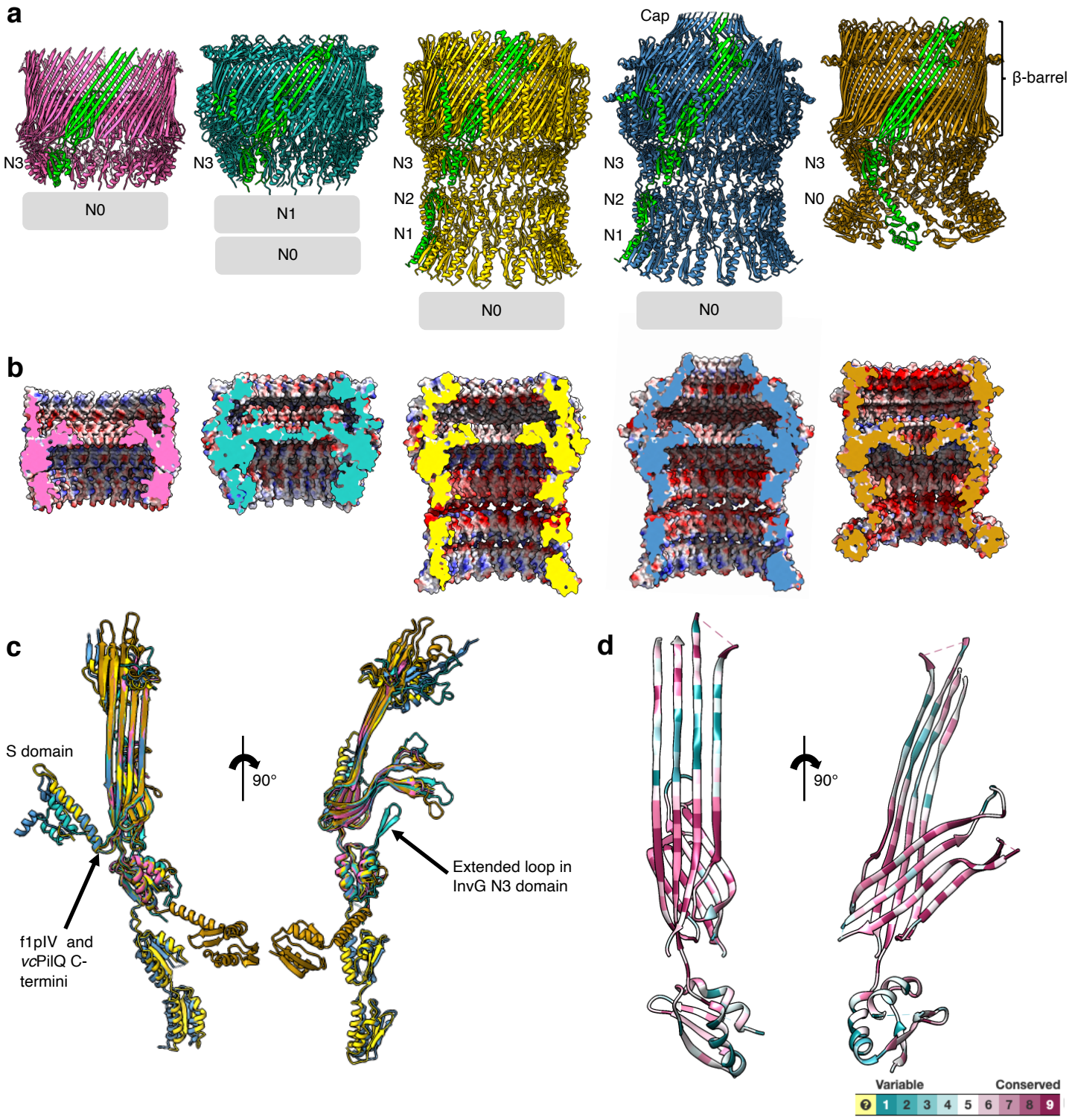


Figure 6. Structural comparison of f1pIV with a selection of different classes of bacterial secretins. Secretins are aligned at the β-barrel domain. f1pIV is shown in pink (this study:7OFH), the Type III secretin InvG from *Salmonella typhimurium* in sea green (6PEE), the Type II secretin GspD from *E. coli* K12 in gold (5WQ7), the Type II secretin GspD from *Vibrio cholerae* in blue (5WQ8) and the Type IV pilus secretin PilQ from *Vibrio cholerae* in orange (6W6M). The secretins are shown in **a)** as multimers (with one monomer coloured lime green). Periplasmic N domains not observed in the maps are shown as grey boxes. **b)** Electrostatic surface potential comparison of multimers cut-through to show the inner surface inside the pore. **c)** Structural superimposition of the selected secretin monomers (shown individually in Supplementary Fig. 8) shown in front (left) and side (right) views. **d)** Sequence conservation amongst the secretin family plotted on to the f1pIV structure, calculated using Consurf and based on 100 unique homologues from the Uni-Prot database sharing 35-100% identity**.**

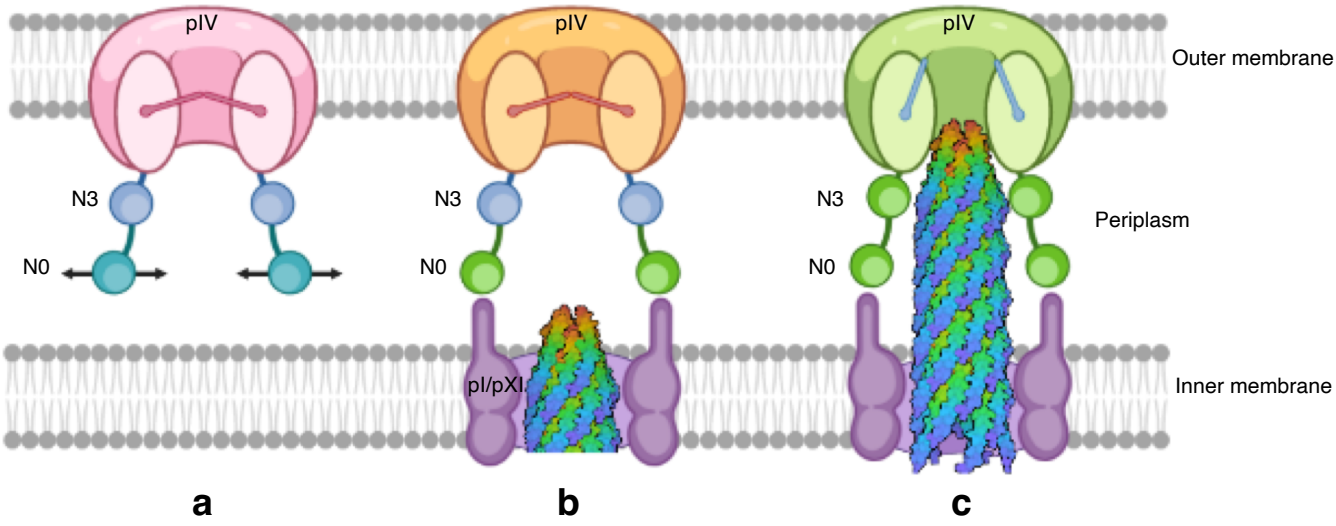
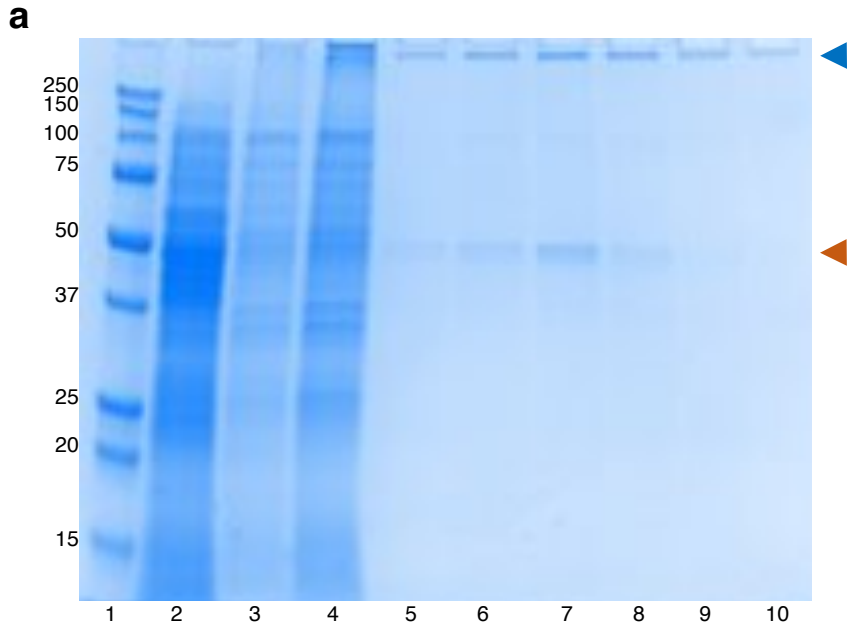
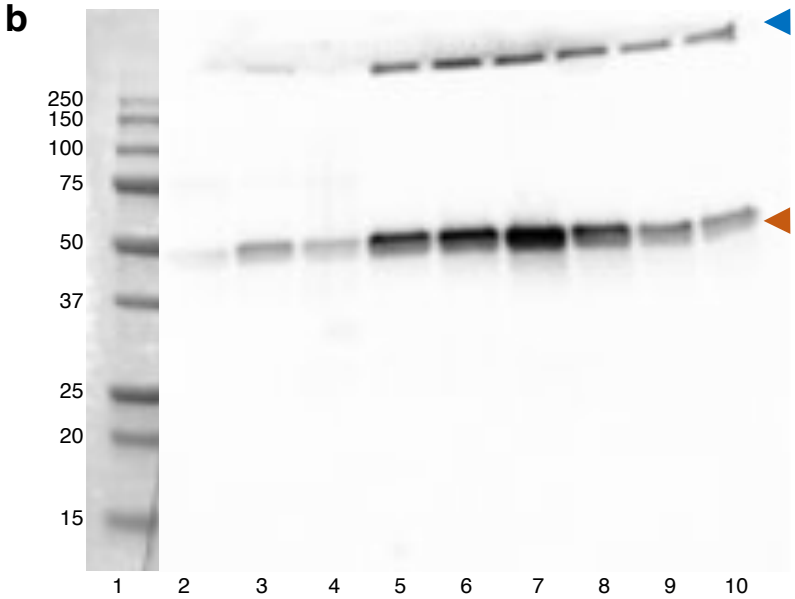


Figure 7. Model for extrusion of filamentous phage. a) f1pIV with the gate in the closed state (pink). The N3 (blue) and N0 (teal) domains project into the periplasm. N0 is attached to N3 via a long linker and is flexible (indicated by black arrows). **b)** N0 (now green, activated) binds to the inner membrane complex of pI/pXI (purple) and the pIV gate (orange) is primed to open. **c)** Egressing phage (blue and green; cap proteins pVII/pIX red and orange) interacts with the N3 domain (now green, activated), triggering opening of pIV (now green, activated) by rearrangement of the gate (blue). Created with BioRender.com; the phage was drawn using 2C0W.

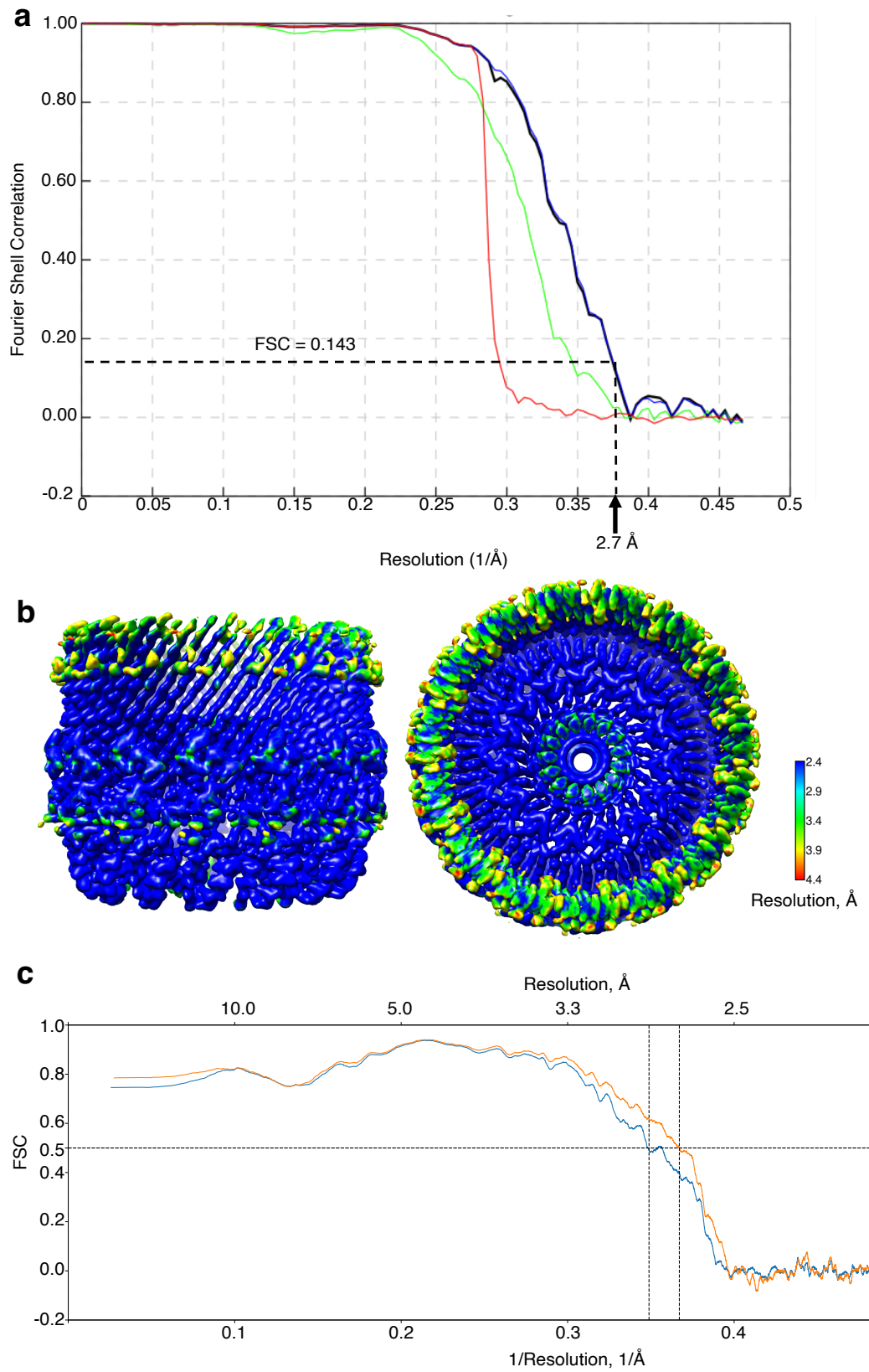




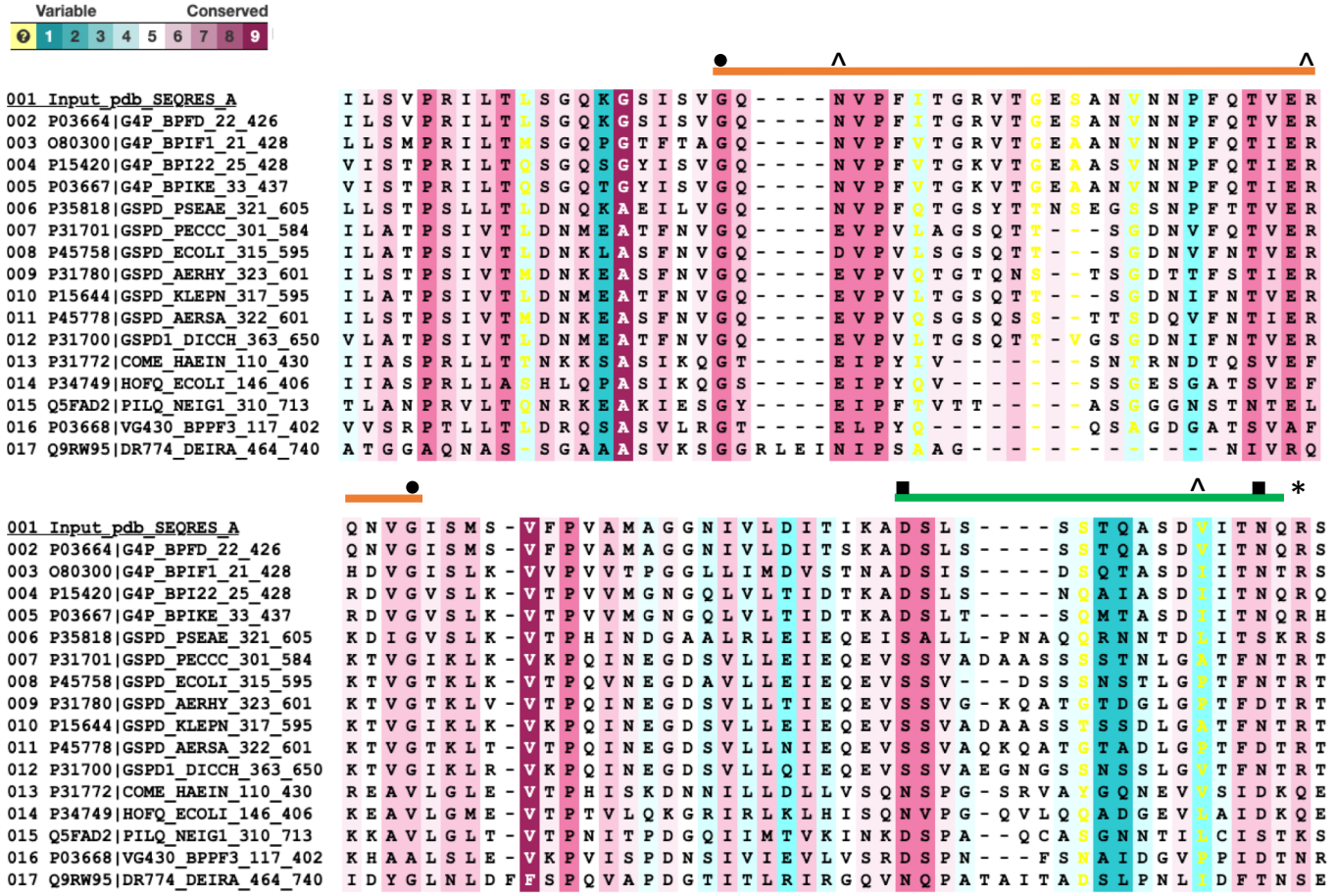
c

Accession	Description	Score	Coverage	# proteins	# unique peptides	# peptides	# PSMs	Area	# AAs	MW (kDa)	Calc. pI
f1pIV	f1pIV sequence	2641.11	77.78	1	40	40	973	3.630E11	405	43.4	5.06
P0A853	Tryptophanase OS=Escherichia coli (strain K12)	248.61	79.83	1	39	39	91	4.152E9	471	52.7	6.23
P0ABB4	ATP synthase subunit beta OS=Escherichia coli (strain K12)	180.28	77.61	1	28	28	56	7.750E8	460	50.3	5.01
P0ABB0	ATP synthase subunit alpha OS=Escherichia coli (strain K12)	120.71	49.32	1	21	22	36	2.840E8	513	55.2	6.13
P0A6U8	Glycogen synthase OS=Escherichia coli (strain K12)	118.33	41.93	1	19	19	38	9.668E8	477	52.8	6.62
P13035	Aerobic glycerol-3-phosphate dehydrogenase OS=Escherichia coli (strain K12)	94.42	56.09	1	23	23	32	1.994E8	501	56.7	7.44
P0A6F3	Glycerol kinase OS=Escherichia coli (strain K12)	93.32	54.78	1	25	25	32	3.567E8	502	56.2	5.50
P77804	Protein YdgA OS=Escherichia coli (strain K12)	91.30	58.17	1	23	23	29	9.258E7	502	54.7	5.17
P0CE47	Elongation factor Tu 1 OS=Escherichia coli (strain K12)	71.19	73.60	2	21	21	27	1.377E8	394	43.3	5.45
P0AFG6	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex OS=E. coli (strain K12)	63.11	34.57	1	15	15	20	1.909E8	405	44.0	5.81

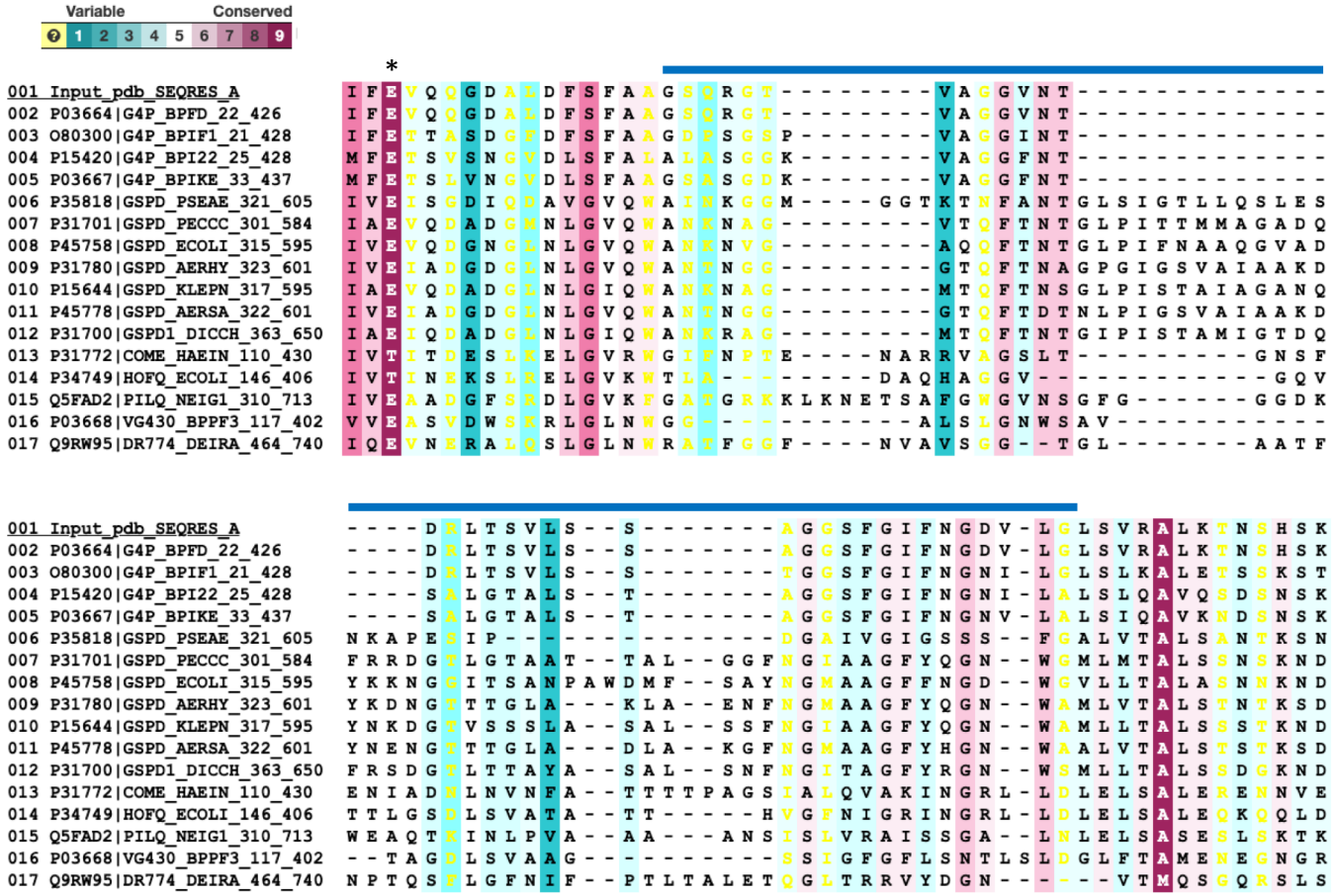
Supplementary Figure 1. Purification of f1pIV verified by **a)** SDS-PAGE and **b)** Western blot. Lane 1, Biorad Precision Plus All Blue Marker; lane 2, total soluble protein; lane 3, total membrane protein; lane 4, solubilized membrane protein; lanes 5-10, fractions from size exclusion chromatography. The f1pIV monomer is indicated with an orange arrowhead and the multimer by a blue arrowhead. Lane 1 (protein markers) was cut away and visualized with white light, hence the difference in colouring compared to the rest of the blot. **c)** Mass spectrometry analysis. The top ten proteins identified in the mass spectrometry analysis are shown, with f1pIV clearly being the most abundant protein present.



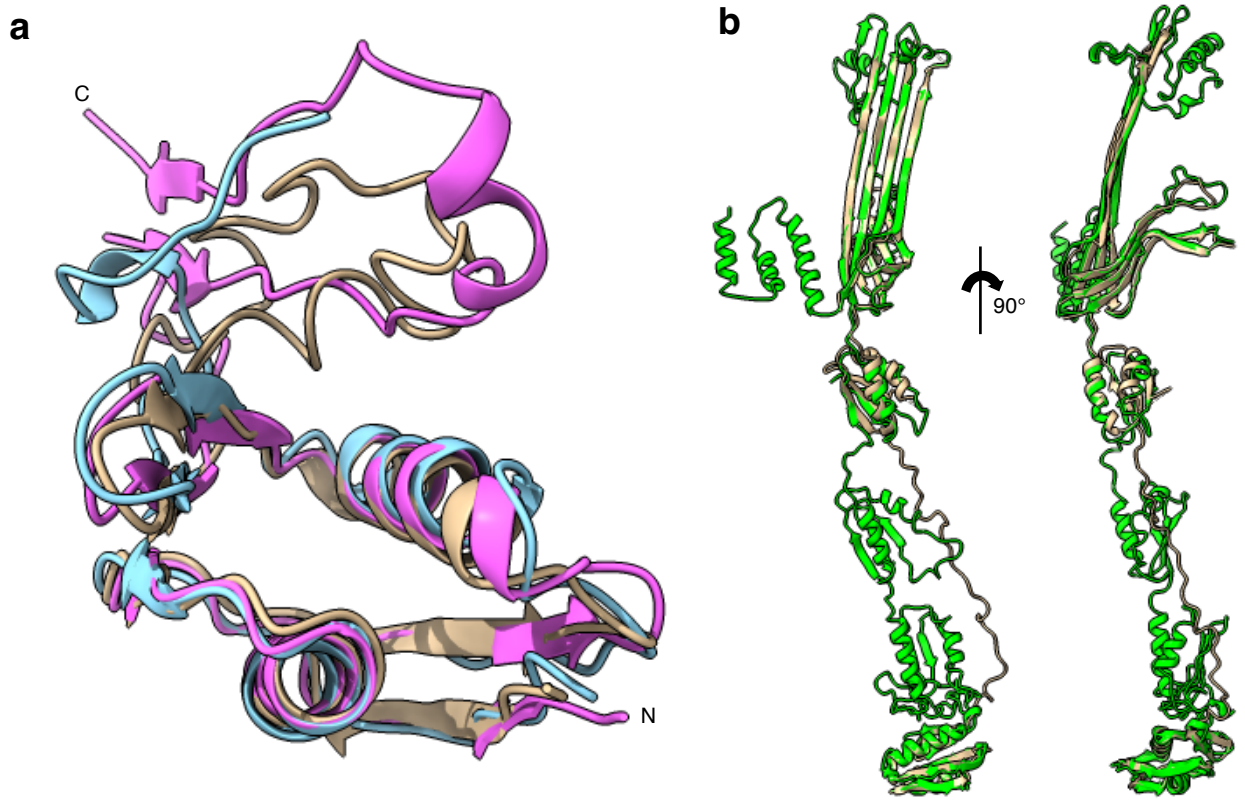
Supplementary Figure 2. f1pIV data quality. a) FSC curve obtained from Relion. The calculated final map was determined at 2.7 Å resolution using Fourier shell correlation at the 0.143 cut-off.
Red curve = rlnCorrectedFourierShellCorrelationPhaseRandomizedMaskedMaps.
Green curve = rlnFourierShellCorrelationUnmaskedMaps.
Blue curve = rlnFourierShellCorrelationMaskedMaps.
Black curve = rlnFourierShellCorrelationCorrected.
b) f1pIV cryoEM map coloured by local resolution (calculated with Resmap)[60].
c) FSC curve showing the quality of the model to map fit (calculated with Phenix)[61].



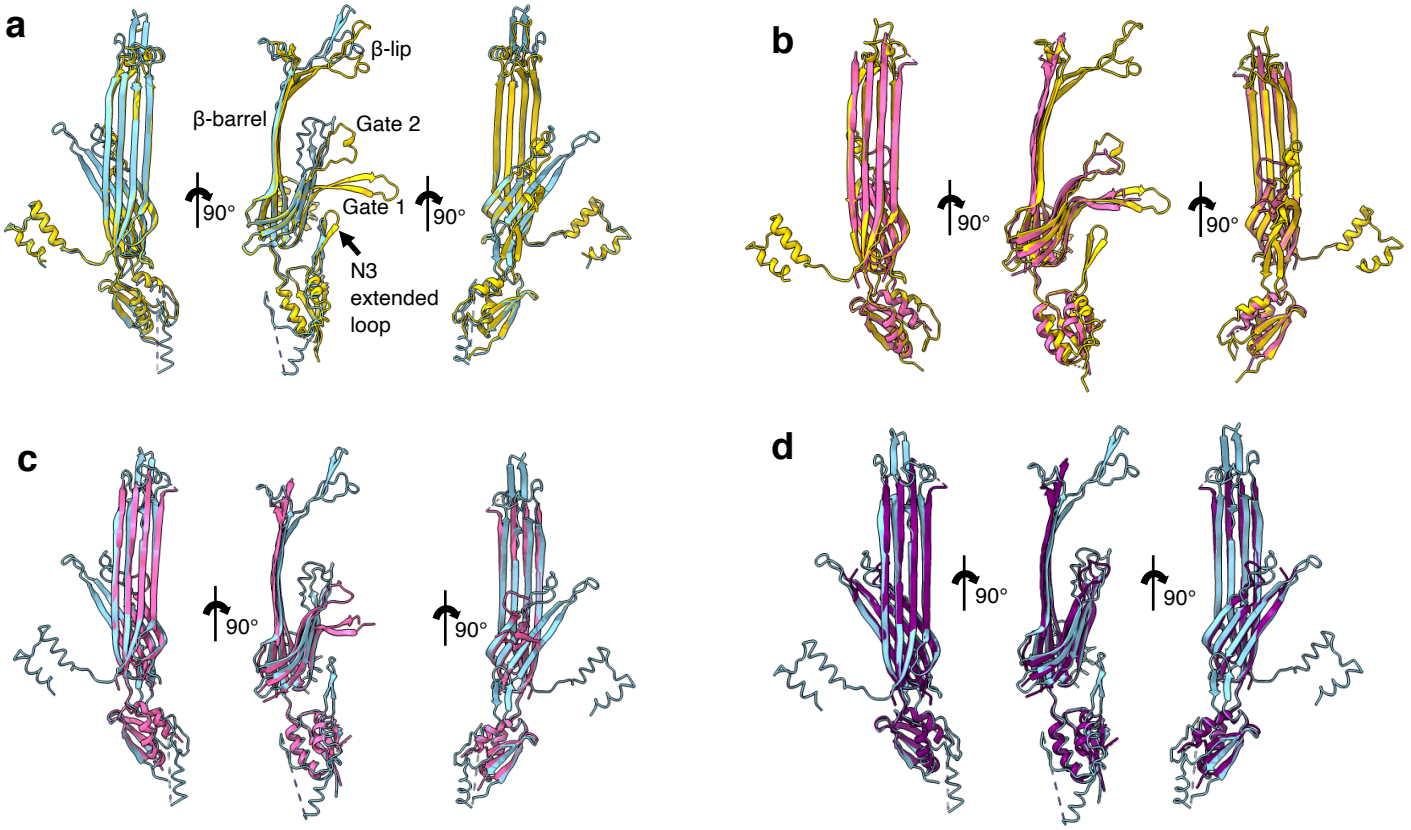
Supplementary Figure 3. Sequence alignment of the Gate region of 16 homologues from Swiss-Prot coloured by conservation. Gate 1 is indicated with an orange line and Gate 2 with a green line. ● Hinge residues Gly 267 and Gly 297 in the Gate 1 loop and ■ hinge residues Asp 321 and Asn 335 in the Gate 2 loop. ^ highly conserved Asn 269 and Arg 293 involved in intra-gate hydrogen bonding, and non-conserved Val 332 involved in inter-gate hydrogen bonding via its main chain carbonyl oxygen. * highly conserved Arg 337 found in salt bridge in the interface between the inner and outer β-barrels.



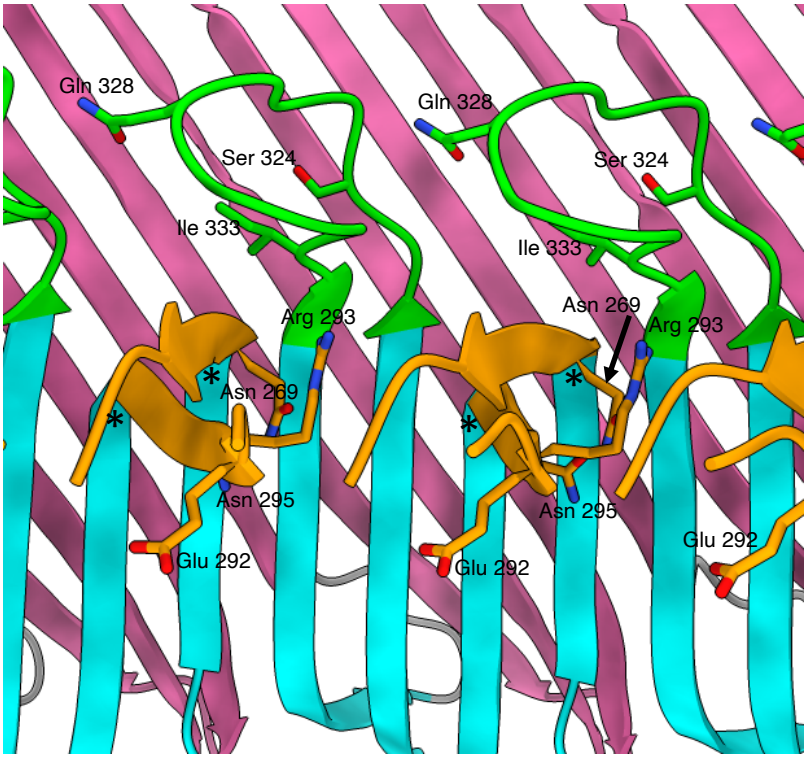
Supplementary Figure 4. Sequence alignment of the ß-barrel lip region of 16 homologues from Swiss-Prot coloured by conservation. The cap region is indicated with a blue line. * highly conserved Glu 185 found at top of the interface between the inner and outer β-barrels. Sequences 001-005 and 016 are phage sequences and have 36-37 residues in the ß-lip region. The remaining sequences are from bacterial Type II secretion systems and have 51-59 residues in this region.



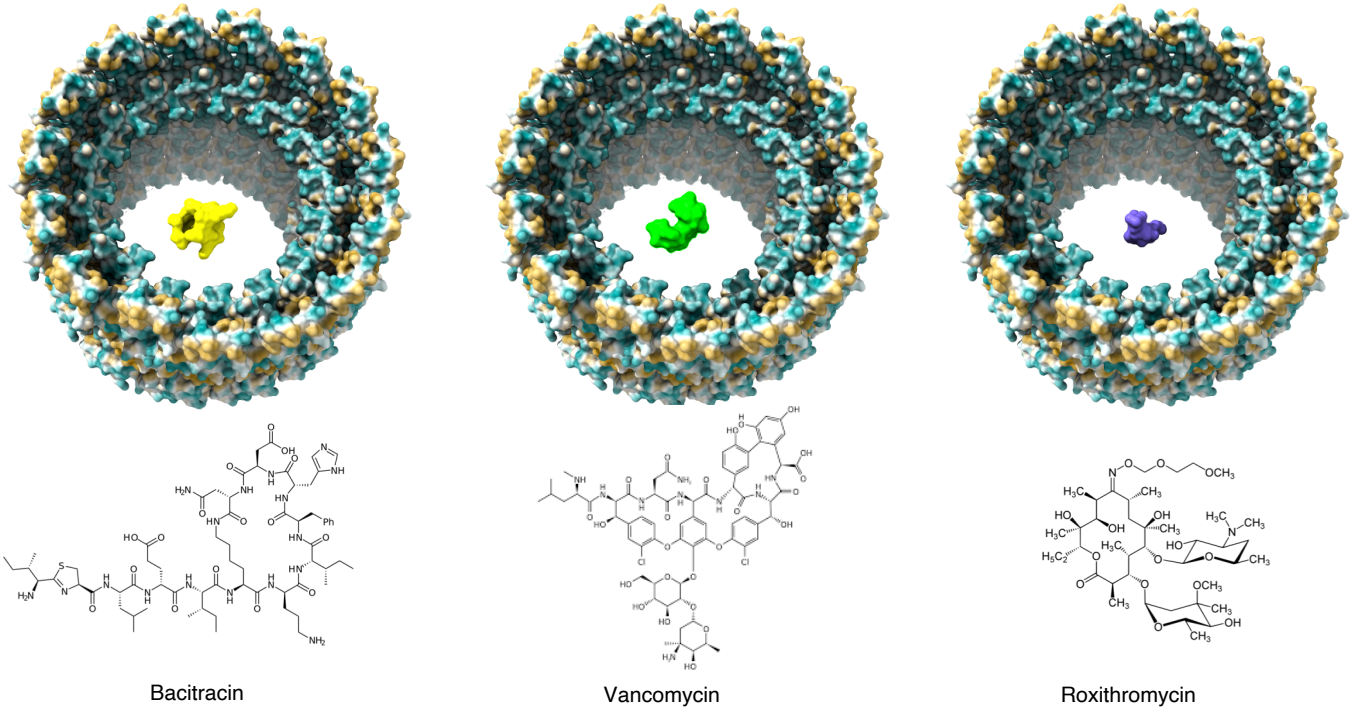
Supplementary Figure 5. Modelling the N0 domain. a) Structural superimposition of N0 homology models produced by Swiss-model (blue), I-TASSER (brown) and Robetta (RoseTTAfold algorithm, magenta). The models are in agreement for residues 2-71 (the folded domain) and differ from residues 72-107 (the flexible linker). **b)** Structural superimposition of the PulD secretin from *Klebsiella pneumoniae* (6HCG, green) with f1pIV composite model (brown).



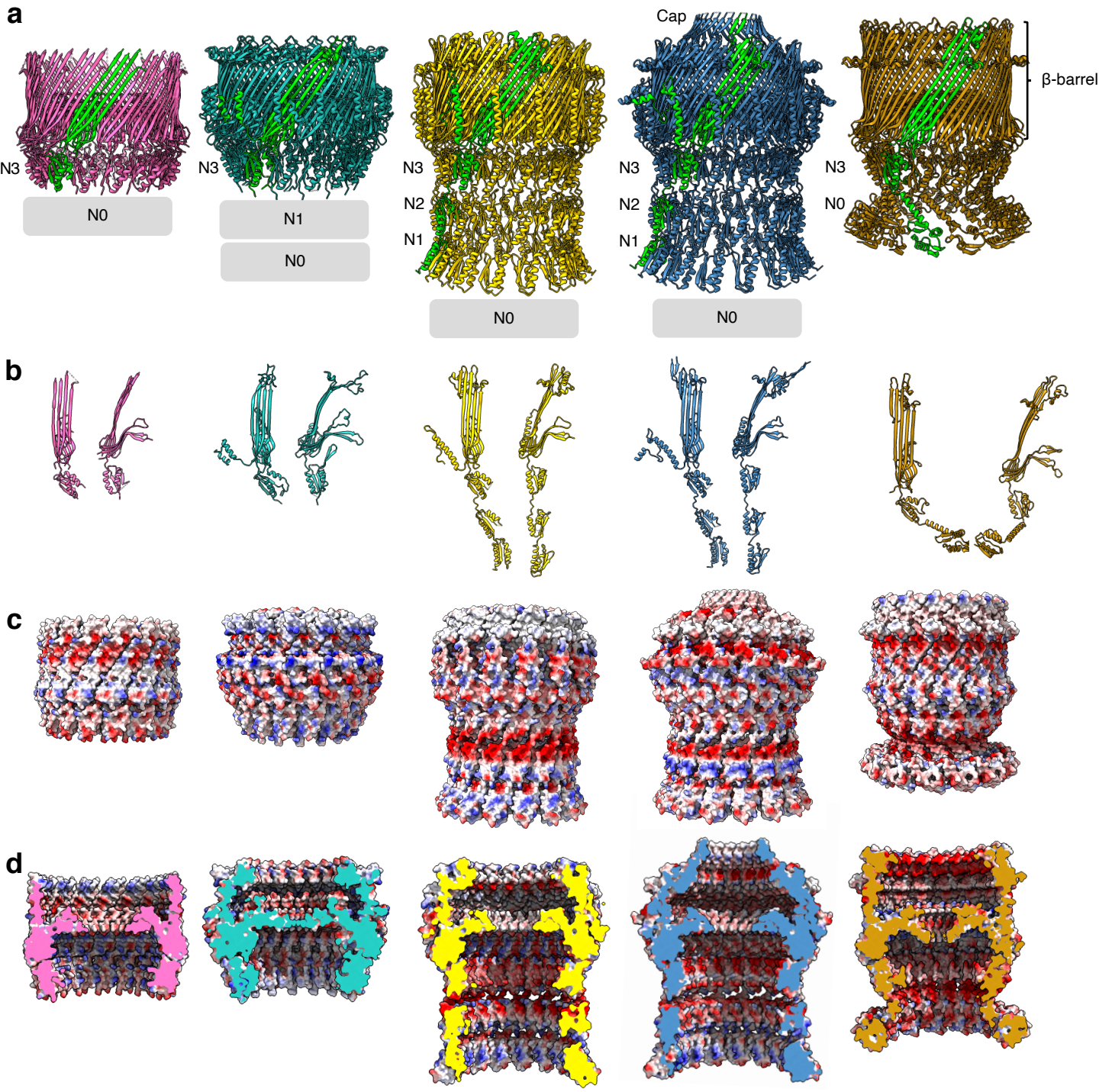
Supplementary Figure 6. Modelling f1pIV in the open state. a) Structural alignment of open (blue; 6Q15[33]) and closed (yellow; 6PEE[33]) states of InvG from *Salmonella typhimurium* in front, side and back views from left to right. **b)** Structural alignment of the closed state of InvG (yellow; 6PEE) and f1pIV (this study: 7OFH; pink) in front, side and back views from left to right. The average distance between the atoms of the two superimposed proteins was measured using the Root Mean Square Deviation (RMSD), giving a value of 1.051 Å between 94 pruned atom pairs. **c)** Structural alignment of the open state of InvG (blue; 6Q15) and f1pIV (this study: 7OFH; pink) in front, side and back views from left to right. **d)** Structural alignment of the open state of InvG (blue, 6Q15) and f1pIV modelled in an open form (purple) in front, side and back views from left to right.



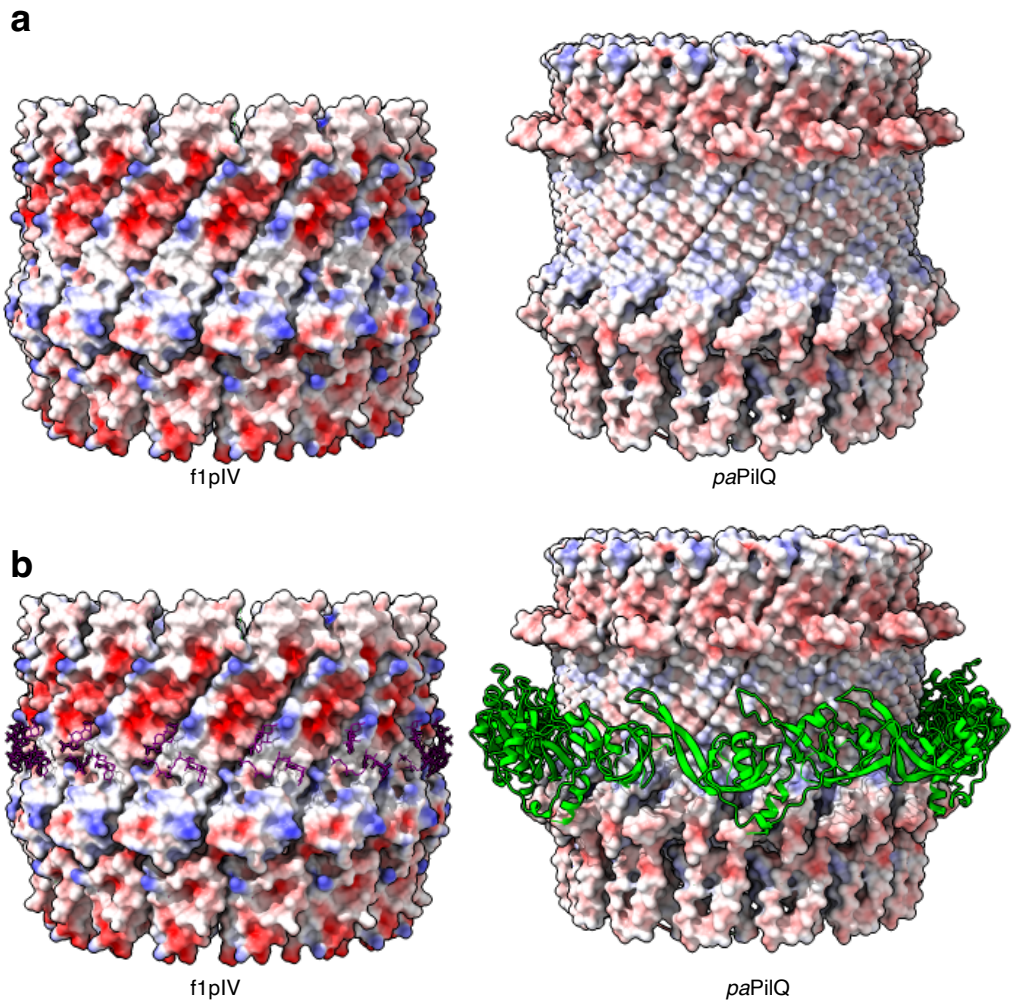
Supplementary Figure 7. f1pIV leaky mutants. An f1pIV 15-mer is shown as a cut-through, viewed from the centre of the pore. Residues Asn 269, Glu 292, Arg 293 and Asn 295 are shown in stick form in the Gate 1 loop (orange), residues Ser 324, Ile 333 and Gln 328 are shown in stick form in the Gate 2 loop (green). Hinge residues Gly 267 and Gly 297 are shown with *. Colouring as per Fig. 2b.



Supplementary Figure 8. Antibiotic uptake through open f1pIV. The open f1pIV model (coloured by hydrophobicity with the most hydrophobic areas coloured orange, through white, to the most hydrophilic in dark cyan) with bacitracin (yellow, amphipathic), vancomycin (lime green, hydrophilic) and roxithromycin (purple, hydrophobic) shown as surfaces. The chemical structures of each antibiotic are shown below.



Supplementary Figure 9. Structural comparison of f1pIV with a selection of different classes of bacterial secretins. Secretins are aligned at the β-barrel domain. f1pIV is shown in pink (this study: 7OFH), the Type III secretin InvG from *Salmonella typhimurium* in sea green (6PEE), the Type II secretin GspD from *E. coli* K12 in gold (5WQ7), the Type II secretin GspD from *Vibrio cholerae* in blue (5WQ8) and the Type IV pilus secretin PilQ from *Vibrio cholerae* in orange (6W6M). The secretins are shown in **a)** as multimers (with one monomer coloured lime green). Periplasmic N domains not observed in the map are shown as grey boxes. The corresponding monomers are shown in **b)** in front and side view. **c)** Electrostatic surface potential comparison of multimers and in **d)** cut-through to show the inner surface inside the pore.



Supplementary Figure 10. Comparison of electrostatic surfaces and pilotin binding. Secretins are aligned at their β-barrel domains. **a)** Electrostatic surface charge representation of f1pIV (this study: 7OFH) and PilQ from *Pseudomonas aeruginosa* (*pa*PilQ) with pilotin removed (6VE2)[50]. **b)** Electrostatic surface charge representation of f1pIV showing CHAPS molecules bound (purple sticks) and *p*aPilQ showing the pilotin TsaP bound (cartoon representation in green).

Results colour-coded for amino acid conservation

The current colourscheme of the alignment is for **amino acid conservation**.

The conservation scoring is performed by PRALINE. The scoring scheme works from 0 for the least conserved alignment position, up to 10 for the most conserved alignment position.
The colour assignments are:

Unconserved 0 1 2 3 4 5 6 7 8 9 10 **Conserved**

```
           .........10 .........20 .........30 .........40 .........50
f1pIV      ---------- --------QV IEMNNSPLRD FVTWYSKQTG ESVIVSPDVK
KpGspD     MKKLLASILL FFTFTVYAIP VELNNVPLRE FVSWYSKVSG QPVIVSPDVK
Consistency0000000000 00000000012 8*7**2***6 **5****25* 63********

           .........60 .........70 .........80 .........90 .........100
f1pIV      GTVTVYSSDV KPE---NLRN FFISVLRANN FDMVGSNPSI IQKYNPNNQD
KpGspD     GEVTVYSADV KREELSQFFV S---VLRANG YDLSNGNPGV VQKY----SS
Consistency*3*****6** *2*00044 11 2000*****4 6*7244**48 8***0000 44

           .........110 .........120 .........130 .........140 .........150
f1pIV      YIDELPSSDN QEYDDNSAPS GGF------- -FVPQNDNVT QTFKINNVRA
KpGspD     RNYEYADSFS DDVDYVPQDS SSAPAGDFFN PAKTQASLIT QTYPVNNVRA
Consistency210*334*15 463*01332* 4420000000 0223*2418* **638*****

           .........160 .........170 .........180 .........190 .........200
f1pIV      KDLIRVVELF VKSNTSKSSN VLSIDGSNLL VVSAPKDILD NLPQFLSTVD
KpGspD     KDLAPVIDVF LKGDNIAGTK VFPFDGANIL AVTASASQHK QLKEFFPSVD
Consistency***32*866* 6*45423454 *434**6*7* 5*5*334113 4*36*435**

           .........210 .........220 .........230 .........240 .........250
f1pIV      LPTDQILIEG LIFEVQQGDA LDFSFAAGSQ RGT-VAGGVN TDRLTSVLSS
KpGspD     VPRIQVLVES VIFETSASDG FDFSFAAGDP SGHPIAGGVN TDRLGKVLSS
Consistency6*31*8*8*4 6***4434*4 4*******43 3*108***** ****24****

           .........260 .........270 .........280 .........290 .........300
f1pIV      AGGSFGIFNG DVLGLSVRAL KTNSHSKILS VPRILTLSGQ KGSISVGQNV
KpGspD     SGGSFGIFNG NVLALSLKAL ETSNKSTLLS MPRILTMSGQ PGTFTAGQNV
Consistency6********* 5**4**66** 5*552*37** 5*****7*** 3*5455****

           .........310 .........320 .........330 .........340 .........350
f1pIV      PFITGRVTGE SANVNNPFQT IERQNVGISM SVFPVAMAGG NIVLDITSKA
KpGspD     PFVTGRITGE AASVNNPFQT IERRDVGISL QVVPVVTPAG LLIMDVKTNA
Consistency**8***8*** 6*5******* ***55****7 4*3**5334* 1787*8354*

           .........360 .........370 .........380 .........390 .........400
f1pIV      DSLSSSTQAS DVITNQRSIA TTVNLRDGQT LLLGGLTDYK NTSQDSGVPF
KpGspD     DSISDSQAAS DIITNTRSIS TTVQLKSGQT VLLGGMVDNR DSEADSSVPW
Consistency**7*4*33** *8***3***6 ***4*64*** 6****74*16 5543**4**4

           .........410 .........420 .........430.....
f1pIV      LSKIPLIGLL FSSRSDSNEE STLYVLVKAT IVRAL
KpGspD     VSKIPLIGAL FTSKSTNASK RTLYMLIRAR VVNPL
Consistency6*******3* *5*6*35245 3***5*86*3 8*43*
```

Supplementary Figure 11. Sequence alignment performed with Praline. The closest homologue of f1pIV based on sequence identity is GspD from *Klebsiella pneumoniae.*

Data collection and processing	Dataset 1	Dataset 2
Acceleration voltage (kV)	300	300
Nominal magnification (x)	81k	81k
Pixel size (Å)	1.072	0.536 (super-res)
Frame rate (s^{-1})	11.93	11.93
Exposure time (s)	3.52	3.52
Total exposure (e^-/Å)	42.059	42.059
Particles		
Micrographs used for selection	7037	14,336
Defocus range (μm)	-2.5 to -1.3 (0.2 μm steps)	-2.5 to -1.3 (0.2 μm steps)
Total number of particles	241,591	330,389
In final 3D reconstruction	16104	95575
	Dataset 1 and Dataset 2 combined	
Resolution		
"Gold-standard" at FSC 0.143 (Å)	2.7	
Map-sharpening B factor ($Å^2$)	0	
Model refinement		
Model composition	15 identical subunits	
Non-hydrogen atoms, per chain	1742	
Protein residues, per chain	231	
Ligands, per chain	2	
R.M.S. Z scores		
Bond lengths (Å)	0.48	
Bond angles (°)	0.72	
R.M.S.D		
Bond lengths (Å)	0.031	
Bond angles (°)	1.903	
Validation		
Ramachandran angles (%)		
Favoured	88	
Allowed	12	
Outliers	0	
Clashscore	14	
Rotamer outliers (%)	2	
Resolution estimates (Å)		
Model resolution (0/0.143/0.5) - masked	2.5/2.6/2.7	
Model resolution (0/0.143/0.5) - unmasked	2.4/2.5/2.9	

Supplementary Table 1. The statistics of data collection, model reconstruction, refinement and validation.

f1pIV leaky mutation	Location in structure	Leaky to maltopentaose sugars /deoxycholate	Leaky (sensitive) to vancomycin	Leaky (sensitive) to bacitracin
A121V	N3 domain	✓	×	×
D123Y	N3 domain	✓	×	×
G147V	N3 domain	✓	×	×
I183V	Secretin domain	✓	×	×
G259S/D	Gate region	✓	×	×
S263F	Gate region	✓	✓	✓
G267D	Gate region	✓	×	×
N269D	Gate region	✓	✓	✓
V270I	Gate region	✓	✓	✓
P271S	Gate region	✓	×	×
I273V	Gate region	✓	×	×
G275D	Gate region	✓	×	×
R276C	Gate region	✓	×	×
V277A	Gate region	✓	×	×
G279D	Gate region	✓	×	×
S281P	Gate region	✓	×	×
A282G	Gate region	✓	✓	×
N283K	Gate region	✓	×	×
V284A	Gate region	✓	×	×
F288L	Gate region	✓	×	×
E292K	Gate region	✓	✓	✓
R293C	Gate region	✓	×	×
N295S	Gate region	✓	✓	✓
V296I	Gate region	✓	×	×
G297V	Gate region	✓	×	×
S322N	Gate region	✓	✓	✓
S324G	Gate region	✓	✓	✓
S325P	Gate region	✓	✓	✓
S326F	Gate region	✓	✓	×
T327A	Gate region	✓	✓	×
Q328G	Gate region	✓	✓	✓
A329T/V	Gate region	✓	×	×
S330N	Gate region	✓	×	×
D331N	Gate region	✓	✓	✓
I333V	Gate region	✓	✓	×
T334A	Gate region	✓	✓	✓

Supplementary Table 2. Leaky f1pIV mutants. The leaky mutants from Spagnuolo et al[24] were plotted onto the f1pIV structure shown in Fig. 5a and b.