# Cell attachment and tail contraction of *S. aureus* phage phi812

Ján Bíňovský[1,2]*, Marta Šiborová[1]*, Jiří Nováček[1], Pavol Bárdy[3], Roman Baška[1], Karel Škubník[1], Tibor Botka[3], Martin Benešík[3], Roman Pantůček[3], Konstantinos Tripsianes[1], Pavel Plevka[1]

[1] *Central European Institute of Technology, Kamenice 753/5, 625 00 Brno, Czech Republic*

[2] *National Centre for Biomolecular Research, Faculty of Science, Masaryk University, 625 00, Brno, Czech Republic*

[3] *Faculty of Science, Masaryk University, Kamenice 753/5, 625 00 Brno, Czech Republic*

** These authors contributed equally to this work*

# Abstract

Phages with contractile tails employ elaborate strategies to penetrate bacterial cell walls and deliver their genomes into the host cytoplasm. Most tailed phages that have been structurally characterized to date infect Gram-negative bacteria, but those targeting Gram-positive bacteria, many of which are important human pathogens, are less well understood. Here, we show that the baseplate of phage phi812, which infects Gram-negative *Staphylococcus aureus*, is formed of a core, wedge modules, and baseplate arms carrying receptor-binding proteins type 1 and 2 and tripod complexes. Upon binding to the *S. aureus* cell wall, the symmetry of the baseplate transforms from threefold to sixfold. The conformational changes propagate from the outer edge to the core of the baseplate. The receptor-binding proteins re-orient to bind to the host cell. Subsequently, central spike proteins degrade wall teichoic acid. The conformational changes to the tripod complexes trigger the release of the central spike and weld proteins from the baseplate, which enables the hub proteins to cleave cell wall peptidoglycan and facilitate penetration of the tail tube tip through the cell membrane. Changes in the positions of baseplate arms are transmitted through wedge modules to tail sheath initiator proteins. The ring formed by the tail sheath initiator proteins expands in diameter and thus triggers the contraction of the tail sheath. The tail sheath of phi812, which can bend up to 90° in its native conformation, shortens from ~200 to ~96 nm, which pushes the tip of the tail tube 10-30 nm into the cytoplasm, depending on the local thickness of the *S. aureus* cell wall. Our results provide the foundation for engineering phi812 to target specific *S. aureus* strains.

# Introduction

Tailed phages use baseplates at the end of their tails to attach to the host cell surface to initiate infection[1]. The cell binding of phages with contractile tails is coupled with a conformational change to the baseplate that triggers tail sheath contraction[2]. The tail tube penetrates the cell wall and enables the ejection of the phage genome into the bacterial cytoplasm. Most of the phages with contractile tails and contractile injection systems (CISs) structurally characterized to date attack Gram-negative bacteria[3–15]. The mechanisms of how phages with contractile tails or CISs breach the cell wall of Gram-positive bacteria are less well understood. Polyvalent phage phi812 from the genus Twort-like virus of the family *Herelleviridae* can infect more than 90 % of clinical isolates of *Staphylococcus aureus*[16,17]. This human pathogen can cause skin and epithelial infections, as well as bacteremia that can develop into life-threatening sepsis[18,19]. *S. aureus* is frequently resistant to antibiotics, which has renewed interest in phage therapy[20–22].

It has been shown that a virion of phi812 is formed by an icosahedral head with a diameter of 90 nm and a 240-nm-long tail that ends with a lentil-shaped baseplate[23]. Upon binding to the *S. aureus* cell wall, the baseplate transforms into a double-layered disc-shaped conformation. Furthermore, the structure of phage A511, which infects Gram-positive *Listeria* and is related to phi812, has also been determined[24]. However, the resolutions of the previous studies of phi812 and A511 were limited to 13 and 14 Å, respectively, which prevented the identification of some of the baseplate components and characterization of their conformational changes that enable cell wall binding, tail contraction, and genome delivery.

Here, we used a combination of cryo-EM, X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR), and AlphaFold2 predictions to build the molecular structure of the phi812 baseplate in native and cell-bound conformations. The phi812 baseplate contains more than one hundred predicted receptor-binding sites for attachment to the *S. aureus* cell wall. Based on comparing the native and contracted structures, we propose a sequence of conformational changes that propagate from the baseplate periphery to the core and trigger the tail sheath contraction. Enzymes degrading teichoic acid and peptidoglycan attached to the tail tube tip facilitate its penetration through the *S. aureus* cell wall.

# Results and Discussion

## The native structure of phi812 baseplate

The structure of the phi812 baseplate with imposed sixfold symmetry was previously determined to a resolution of 13 Å using cryo-EM[23]. However, here we show that the phi812 baseplate, similar to that of A511[24], is only approximately sixfold symmetric. Reduction of the symmetry of the reconstruction to threefold enabled the determination of the overall baseplate structure to a resolution of 7.3 Å (Fig. 1A, B, SFig. 1 and 2A, Table S1). The baseplate of phi812 is lentil-shaped with a diameter of 655 Å and height of 310 Å, and can be divided into a core, six wedge modules, and six arms decorated with three types of receptor-binding proteins (Fig. 1). A localized reconstruction of the baseplate core and wedge modules reached a resolution of 4.4 Å, which enabled the determination of the structures of the tail tube initiator, tail sheath initiator, hub, assembly, central spike, weld, wedge, and arm scaffold proteins (Fig. 1C, D, SFig. 1 and 2B, C, SMovie 1). The hexamer of tail tube initiator proteins (gp113) and trimer of hub proteins (gp110) form a channel, which extends the tail tube channel into the baseplate. The trimer of central spike proteins (gp112), stabilized by the three copies of weld protein (gp111), plugs the baseplate channel at the level of the hub proteins (Fig. 1D). The hexamer of tube initiator proteins is surrounded by six copies of each of the sheath initiator (gp115) and assembly proteins (gp114), which are placed above each other (Fig. 1C, D, and 2). The baseplate core is encircled by the six wedge modules, which enable the attachment of baseplate arms (Fig. 1C, D, and 2, SMovie 1). Each wedge module is a heterotrimer formed by two copies of wedge protein (gp116) and an N-terminal part of the arm scaffold protein (gp117) (Fig. 1C and 2, SMovie 1). One rod-shaped baseplate arm protrudes radially from each wedge module (Fig. 1A, B, SMovie 2). The baseplate arm is formed by the C-terminal extended part of the arm scaffold protein reinforced by three dimers of the arm segment protein (gp118) (Fig. 1B and 3A-C, SFig. 3). The arm scaffold protein provides attachment sites for two trimers of the tripod proteins (gp119), a trimer of receptor-binding protein 1 (RBP1, gp121), and a trimer of receptor-binding protein 2 (RBP2, gp123) (Fig. 1A, B, SMovie 2). The trimer of tripod proteins located closer to the tail axis (inner tripod) is oriented with its receptor-binding domains pointing away from the phage head (Fig. 1A, B). In contrast, the receptor-binding domains of tripod proteins attached to the baseplate arm further away from the tail axis (outer tripod) point towards the phage head (Fig. 1B). Trimers of RBP1 and RBP2 are attached to the end of the baseplate arm (Fig. 1A, B). The RBP1 reaches from one baseplate arm to the neighboring one, where it interacts with RBP2 (Fig. 1A, SFig. 4A, B). The RBP2 points towards the tail sheath (Fig. 1A, B, SFig. 4).

The deviation of the phi812 baseplate from sixfold symmetry is most apparent in the distinct conformations of the alternating baseplate arms and the positions of the receptor-binding proteins attached to them (Fig. 1A, B, SMovie 2). The six inner tripods form the shape of an equilateral triangle when viewed along the tail axis towards the phage head (Fig. 1A). The center of mass of the inner tripod positioned at the triangle vertex is 141 Å from the tail axis, whereas that forming the edge is 87 Å from the tail axis (Fig. 1A). The differences in the positioning of the tripods are caused by the alternating arrangement of baseplate arms that point away from the tail axis at different angles (Fig. 1B). The inner tripods at the vertices are attached to upper baseplate arms that radiate from the tail axis at an angle of 72 ° (Fig. 1B). In contrast, the inner tripods forming the triangle edges are attached to the lower baseplate arms radiating at an angle of 62 ° (Fig. 1B). The distinct directions of the baseplate arms also influence the interactions of proteins at the baseplate

105 periphery. The outer tripod and RBP2 attached to the upper baseplate arm interact with a wedge
106 protein and the first tail sheath disc (Fig. 1B, SFig. 4B, D). In contrast, the RBP2 bound to the lower
107 baseplate arm only contacts a wedge protein, and the corresponding outer tripod does not interact
108 with the tail sheath (Fig. 1B, SFig. 4A, C). We speculate that the native structure of the phi812
109 baseplate with overall threefold symmetry is established during the baseplate formation and is
110 determined by the structure of the trimer of hub proteins (Fig. 1D and 2). The two wedge modules
111 that attach to each hub protein differ in their positioning relative to the tail axis (Fig. 1D, SFig. 5A-C,
112 SMovie 1). These structural differences propagate into the baseplate periphery, resulting in the
113 alternating angles of the baseplate arms relative to the tail axis (Fig. 1B, SMovie 2). The same
114 mechanism is probably employed to establish the threefold symmetry of the A511 baseplate[24]. In
115 contrast, the native baseplate of phage T4, which infects Gram-negative *E. coli*, was shown to be
116 arranged with sixfold symmetry[4].
117

118 **Central spike protein blocks the baseplate channel and degrades cell wall teichoic acid**

119 The trimer of central spike proteins of phi812 blocks the baseplate channel and protrudes ~22 nm
120 from the baseplate (Fig. 1B, D and 4A, B, SMovie 2). The central spike protein can be divided into
121 oligosaccharide-binding, β-prism, coiled-coil, knob, and petal domains (Fig. 2, SFig. 6A). The
122 oligosaccharide-binding and β-prism domains have conserved structures shared by central spike
123 proteins of other phages and CISs [7–10,12–15,25,26]. The two domains bind to the inside of the channel
124 formed by the hub proteins (Fig. 1D and 2). The coiled-coil, knob, and petal domains protrude from
125 the phi812 baseplate and are not resolved in the cryo-EM reconstruction of the baseplate, due to the
126 coiled-coil domain's flexibility (Fig. 1B and 4A, B). Structural prediction indicates that the coiled-coil
127 domains from the three central spike proteins form a 22-nm-long fiber (Fig. 2). The structures of the
128 knob and petal domains were determined by cryo-EM and single-particle reconstruction of a
129 recombinantly produced central spike protein (Fig. 1B and 4C, SFig. 1 and 2J). The knob domain has a
130 tumor necrosis factor-like fold[27], similar to that of the head domains of receptor-binding proteins
131 from lactococcal phages Tuc2009 and 1358[28–30] and a lectin from *Burkholderia cenocepacia*[31] which
132 indicates that the knob domain may bind to saccharides from the bacterial cell wall (SFig. 6A, Table
133 S2). The petal domain of the central spike protein has an α/β-barrel fold, and is structurally similar to
134 the protein GlpQ from *B. subtilis* that cleaves teichoic acid[32,33] (Fig. 4C, SFig. 6A, Table S2).
135 Phages with contractile tails and CISs attacking Gram-negative bacteria pierce the outer
136 bacterial membranes using puncturing tips formed of proteins with β-prism or β-helix domains[34–36].
137 In contrast, to infect Gram-positive *S. aureus*, phage phi812 has to penetrate a 50-70-nm-thick layer
138 of teichoic acids and peptidoglycan. Therefore, the central spike of phi812 includes the flexible
139 coiled-coil linker decorated with knob and petal domains that bind to and degrade the host teichoic
140 acids to expose the peptidoglycan layer for degradation by other proteins of the baseplate.
141

142 **Hub proteins facilitate penetration of the tail tube through the cell wall and membrane**

143 The three hub proteins form a ring, extending the tail channel into the phi812 baseplate (Fig. 1D and
144 2). Phi812 hub protein can be divided into channel-forming domains I to IV and clamp and cleaver
145 domains (SFig. 6B). The channel-forming domains I and III have beta-barrel folds and are positioned
146 next to each other so that the head-proximal interface formed by the trimer of hub proteins has

147 pseudo-sixfold symmetry for binding the hexamer of tube initiator proteins (Fig. 2, SFig. 6B). In
148 contrast, the channel-forming domains II and IV of the hub protein, positioned below domains I and
149 III respectively, are structurally distinct and form interfaces for binding the trimer of central spike
150 proteins to the inside of the hub channel (Fig. 2, SFig. 6B). The channel-forming domains I-IV are
151 structurally similar to those of the gp27 of phage T4, for which the domain naming convention was
152 established[35] (Table S2). Homologs of the channel-forming domains can also be identified among
153 phages with long non-contractile tails (Tal proteins)[3,37] and CISs (Vgrg-like proteins)[5,38,39] (Table S2).
154 The cleaver domain of the hub protein, the structure of which was determined using NMR
155 (Fig. 4D, SFig. 6B, Table S3), has the characteristic fold of the cysteine- and histidine-dependent
156 amido-hydrolase/peptidase[40], and contains the conserved catalytic triad Cys699-His759-Glu777[41]
157 (Fig. 4D). Accordingly, zymogram analysis demonstrated that a recombinant cleaver domain degrades
158 *S. aureus* cell walls (SFig. 7), which agrees with previous functional studies of hub protein homologs
159 from the related phage K[42,43]. The clamp domain of the hub protein is formed by ten short α-helices
160 and interacts with the central spike and weld proteins (Fig. 1C, D, and 4E, SFig. 6B and 8A-D). The
161 clamp domain is structurally similar to pore-forming saposin-like proteins, which employ positively
162 charged residues to interact with the negatively charged surfaces of lipid membranes (Table S2)[44,45].
163 Since the clamp domain of the phi812 hub protein contains a long positively charged patch of
164 residues, as the saposin-like proteins do (Fig. 4E), we speculate that it may bind the *S. aureus*
165 membrane to facilitate its penetration by a tail tube. In addition to well-known T4 gp27 protein
166 featuring a lysozyme domain, more hub proteins equipped with enzymatic domains have been
167 reported recently [14,46]. Therefore, we propose that the hub proteins of phi812 and related phages
168 enable genome delivery by degrading peptidoglycan and facilitating membrane penetration by the
169 tip of the tail tube.
170

171 **Tail tube initiator, weld, assembly, and tail sheath initiator proteins complete the baseplate core**

172 Three weld proteins bind to both central spike and hub proteins. Six assembly and six tail sheath
173 initiator proteins surround the hub and tail tube initiator proteins to complete the structure of the
174 baseplate core (Fig. 1C, D, and 2). The tube initiator and sheath initiator proteins are organized with
175 sixfold symmetry, whereas the central spike, weld, hub, and assembly proteins are organized with
176 threefold symmetry that determines the overall organization of the baseplate in its native
177 conformation (Fig. 1C, D, and 2).
178 The weld protein of phi812 can be divided into the N-terminal lid and C-terminal clip
179 domains, which are connected by a 14-residue-long linker (Fig. 2, SFig. 6D). The lid domain interacts
180 with an inner tripod and covers the peptidase site of the cleaver domain of the hub protein (SFig.
181 8A-D). Therefore, releasing the weld protein from the baseplate upon host cell binding enables the
182 hub protein to hydrolyze cell wall peptidoglycan. The clip domain of the weld protein is located
183 closer to the tail axis and has the CHAP protein fold similar to that of the cleaver domain of the hub
184 protein (Fig. 2 and 4F, SFig. 8). However, the clip domain lacks the conserved catalytic residues
185 Cys-His-Glu required for the peptidoglycan cleavage[41] (Fig. 4F). Thus, the clip domain is likely
186 enzymatically inactive, but may have retained peptidoglycan-binding ability, similar to the RipD
187 protein from *M. tuberculosis*[47]. There is no homolog of the weld protein in the baseplate of the
188 related phage A511[24,48]. The weld protein may be unnecessary for A511 because of the differences in
189 the cell wall composition between *Staphylococcus* and *Listeria*[49].

The assembly protein of phi812 is formed by a compact bundle of nine α-helices, and interacts with the wedge protein and C-terminal handle of the tail tube initiator protein (Fig. 1D and 2). The assembly protein of phi812 has the same position in the baseplate and compact globular structure as similar proteins from pyocin R2 and phages Pam3, E217, A-1 and Milano (Table S2). The assembly proteins stabilize the baseplate structure[7,10,12,14,15].

The hexamer of tail tube initiator proteins binds to the trimer of hub proteins and, on the other side, to a hexamer of tail tube proteins (Fig. 1C and 2). The tail tube initiator protein can be divided into the N-terminal β-barrel domain and the C-terminal α-helical handle, which stretches along the hub protein and reaches the assembly protein in the baseplate core (SFig. 6C). The tail tube initiator proteins from pyocin R2 and phages E217 and Pam3 contain extensions similar to the C-terminal handle of the phi812 tail tube initiator protein, suggesting an evolutionary link between phage and pyocin baseplates[2,3].

The tail sheath initiator protein of phi812 is composed of the LysM and handshake domains (Fig. 2, SFig. 6E). The handshake domain, containing two β-strands, interacts with the N- and C-termini of tail sheath proteins from the first tail sheath disc and the C-terminus of tail sheath protein from the second tail sheath disc to form a five-stranded β-sheet (SFig. 6E and 9C). The same mechanism that connects the tail sheath initiator proteins to tail sheath proteins connects successive discs of the tail sheath proteins. Therefore, the interactions of the N- and C-termini of the tail sheath proteins with the handshake domains can initiate tail sheath assembly. Proteins containing domains with structures similar to the handshake domain are integral to the baseplates of most CISs and phages with contractile tails (Table S2)[5,38,50].

**Wedge modules attach baseplate arms to the baseplate core**

Wedge modules are structural components of phage baseplates, which enable the attachment of peripheral baseplate proteins to the baseplate core[51]. Here, we show that the wedge module of phi812 is formed by two wedge proteins with different conformations, A and B, and the N-terminal 176 residues of arm scaffold protein (Fig. 1C and 2, SFig. 3B and 5, Table S2, SMovie 1). Wedge protein can be divided into N-terminal α-helices, wing, trifurcation, and dimerization domains, and the part of the arm scaffold protein belonging to the wedge module can be divided into N-terminal α-helices and a trifurcation domain (Fig. 2, SFig. 3B and 6F, G). The N-terminal helices and trifurcation domains mediate contacts between the proteins forming the wedge module (Fig. 2, SFig. 5A). The wing domain of wedge protein A binds to the hub and central spike proteins. In contrast, the wing domain of wedge protein B binds to the first disc of tail sheath proteins (Fig. 1B-D and 2). The six wedge modules form an iris around the baseplate core (Fig. 1C, D, SFig. 5A). Dimerization domains of wedge proteins mediate interactions between neighboring wedge modules (SFig. 5A). Even though the neighboring wedge modules differ in their conformations due to the overall threefold symmetry of the native baseplate, the contacts between the dimerization domains are similar (SFig. 5A-C, F).

**Baseplate arms provide attachment sites for receptor-binding proteins**

Each of the baseplate arms of phi812 is formed by linker AB, bulge, linker BC, elbow, linker Z, and connector domains of the arm scaffold protein that are reinforced by three dimers of the arm

segment proteins, named arm segments A, B, and C (Fig. 3A-E, SFig. 3B). We employed X-ray crystallography to determine the structure of the arm segment protein (Fig. 3D, Table S4), and AlphaFold2[52] to predict the structures of the arm scaffold protein domains, which could not be built into the cryo-EM map due to its limited resolution (Fig. 3E, SFig. 3A, B). Each arm segment protein is composed of an N-terminal α-helix followed by a three-layered β-sandwich (Fig. 3D). The overall structure of arm segment protein is similar to those of gp8 from phage T4 and gp105 from phage A511[24,53] (Table S2). When looking along the baseplate arm towards the baseplate core, arm segments B and C are rotated clockwise 125° and 250°, respectively, relative to arm segment A (Fig. 3B). The C-terminal beta-strands of neighboring subunits of arm segments A and B are separated by 10.2 Å (Fig. 3F). As was proposed previously[24], the gap is probably filled by a β-strand formed by the linker AB of the arm scaffold protein, creating a continuous, seven-stranded β-sheet. The linker BC of the arm scaffold protein probably connects the β-sheets of arm segments B and C using the same strand insertion mechanism.

The arm scaffold protein provides attachment sites for two trimers of tripod proteins, one trimer of RBP1, and one trimer of RBP2 (Fig. 1B). The trifurcation domain of the arm scaffold protein contains 28-residue-long docking loop 1, which forms the binding site for the inner tripod complex (Fig. 3A, E, SFig. 3B). Similarly, the bulge domain of the arm scaffold protein contains 40-residue-long docking loop 2, which enables attachment of the outer tripod complex (Fig 3A, SFig. 3B). The elbow domain with a double β-sandwich fold and C-terminal connector domain with a β-sandwich fold of the arm scaffold protein are located at the distal end of the baseplate arm, and probably provide attachment sites for RBP1 and RBP2, respectively (Fig. 1B and 3A, SFig. 3B).

**Receptor-binding proteins**

High-avidity interactions between the baseplate and the cell wall mediate the attachment of phi812 to *S. aureus*. Major components of the *S. aureus* cell wall are peptidoglycan formed by N-acetylglucosamine, N-acetylmuramic acid, and oligopeptides and teichoic acids built from poly(ribitol 5-phosphate) decorated with alpha-N-acetylglucosamine (α-GlcNAc)[49,54]. It has been shown that the backbone of the wall teichoic acid is essential for the adhesion of phi812 and related phages K, phiSA012, and phiMR003 to the host cell, whereas α-GlcNAc serves as a co-receptor[55–58]. The baseplate of phi812 contains 36 tripod proteins and 18 of each of RBP1 and RBP2 (Fig. 1A, B and 5). As discussed above, additional putative receptor-binding domains are within the hub and central spike proteins.

The structure of a trimer of tripod proteins in a native baseplate resembles a three-sided pyramid (Fig. 5A). The tripod protein can be divided into the spine, flap, fiber, base, fin, leg, anchor, and hook domains (Fig. 5C). The spine domain is positioned next to the axis of the pyramid, the flap and fiber domains form the edge, and the base domain forms the base of the pyramid. The fin, leg, anchor, and hook domains are located at the faces (Fig. 5A, C). The flap domain contains an insertion loop, three of which enable the binding of the tripod complex to the arm scaffold protein (Fig. 5C, SFig. 3B and 11A, B). The interaction between the inner tripod complex and the arm scaffold protein is well resolved in the reconstruction of a contracted baseplate (SFig. 10A-C). A three-pointed star-shaped crevice, formed by the spine domains of the tripod proteins, binds the docking loop 1 of the trifurcation domain of the arm scaffold protein (SFig. 10A-C). Since the bulge domain of the arm scaffold protein contains a docking loop 2 with hydrophobic residues in equivalent positions to those of the docking loop 1 of the trifurcation domain (SFig. 10D), the inner and outer tripod complexes

are likely attached to the arm scaffold protein using similar interactions. The docking loop 2 of the bulge domain, which binds the outer tripod complex, is twelve residues longer than the docking loop 1 of the trifurcation domain (SFig. 10D). The extra residues of the docking loop 2 probably enable the ~170° rotation of the outer tripod complexes during the baseplate transformation induced by the binding of phi812 to a host cell. The base and fin domains of tripod proteins have ten-stranded β-sandwich folds of carbohydrate-binding module families 16 and 61[59,60] (Fig. 5C, Table S2), indicating that they may bind glycans from the *S. aureus* cell wall. The leg domain includes fifteen β-strands, eight forming a β-sandwich that is structurally similar to lectin-like protein Cwp84 from *C. difficile*[61] (Fig. 5C, Table S2), suggesting another possible receptor-binding site. The anchor domain with a β-sandwich fold is followed by the hook domain with sequence similarity to LysM proteins, indicating it may bind peptidoglycan[62] (Fig. 5C, SFig. 11, Table S2). In the native phi812 baseplate, RBP1 and RBP2 are positioned between baseplate arms (Fig. 1A, B). As the cryo-EM reconstruction of the native baseplate did not resolve RBP1 and RBP2 to sufficient resolutions to enable structure determination, we employed AlphaFold2[52] prediction to obtain the structures of RBP1 and RBP2 (Fig. 5D, E, SFig. 2A, Table S4).

The predicted structure of the RBP1 trimer has the shape of a triangular bipyramid with a height of 186 Å and a width of the base of 102 Å (Fig. 5D). RBP1 consists of an antenna domain, three tower domains, and a tip domain, all positioned along the symmetry axis of the trimer, and a wing domain protruding laterally from the second tower domain (Fig. 5D, Table S2). The tower domains are structurally similar to each other and contain characteristic six-stranded β-sheets (Fig. 5D). The antenna domain mediates the binding of RBP1 to the end of the baseplate arm, probably to the connector domain of the arm scaffold protein (Fig. 1B). The tower domain III and tip domain of RBP1 interact with the tower domain 1 of the RBP2 attached to the baseplate arm located clockwise when looking at the baseplate in the direction of the phage head (Fig. 1A, B).

The RBP2 trimer has the shape of a cylinder with a length of 207 Å and a diameter of 56 Å (Fig. 5E). Like RBP1, RBP2 is formed of an antenna domain, three tower domains, and the tip domain (Fig. 5E, Table S2). The antenna domain enables the binding of RBP2 to the end of the baseplate arm near arm segment C (Fig. 1B). The tip domain of RBP2 is in contact with a wedge module (SFig. 4).

**Re-arrangement of RBP and tripod complexes is required for baseplate binding to a cell wall**

Tailed phages use receptor-binding proteins from their baseplates to attach to host cells[1]. For phages with contractile tails, the receptor binding is accompanied by reorganization of the baseplates, which may include conformational changes to individual proteins[1,4,23]. It has been proposed that the reorganization starts at the baseplate periphery, propagates to the inner baseplate, and triggers the contraction of the tail sheath[2,4]. Upon binding to the *S. aureus* cell wall, the phi812 baseplate broadens to a diameter of 713 Å, and its thickness decreases to 344 Å (Fig. 6, SFig 1 and 2D-F, Table S1). The conformational change transforms the baseplate to sixfold symmetry, with all six baseplate arms protruding at 83° relative to the tail axis (Fig. 6B, C, SMovie 3). Comparing the structures of the native and contracted baseplates of phi812 enabled us to propose the sequence of steps of the baseplate rearrangement. Reorganization of the baseplate is probably initiated by the rotation of RBP1 and RBP2 trimers so that they can bind to the cell wall (Fig. 1B and 6B, C, SMovie 3). RBP1 trimers are not resolved in the reconstruction of the baseplate from phages with contracted tails, likely due to the flexible connection of RBP1 to arm scaffold protein (Fig. 6A-C, SFig. 2D). RBP2 trimers swing away from the baseplate core to orient themselves parallel with the tail axis and their

tip domains facing away from the phage head (Fig. 1B, 6B, C, SMovie 3). Subsequently, the outer tripod complexes can rotate 170° to point their receptor-binding domains away from the phage head (Fig. 1B, 6B, C, SMovie 3). The inner tripod complexes tilt by 6° and rotate 28° to assume positions with their axes parallel to the tail axis (Fig. 1B, 6B, C).

The tripod complexes not only rotate, but also undergo complex conformational changes (Fig. 5C, SMovie 4): (i) The second and third spine helices move towards the first spine helix by 10 and 16 Å, respectively. (ii) The fiber domain moves 13 Å towards the flap domain and rotates 20° counterclockwise around the tripod symmetry axis when looking from the bottom of the tripod. (iii) The base domain shifts 24 Å along the third spine helix while rotating 174° around an axis perpendicular to the tripod symmetry axis. (iv) The fin domain moves 56 Å to protrude laterally from the tripod. (v) The three leg domains detach from the flap domains and move 88 Å to interact with each other under the base domains. (vi) The anchor and hook domains were not resolved in the cryo-EM reconstruction of the contracted baseplate; however, contracted phage particles attached to the cell wall contain density under the tripod complexes, suggesting that the anchor and hook domains stretch along the tripod axis toward the cell wall (Fig. 5C, SFig. 12A, B). Similar density under the tripod complexes was observed in micrographs of contracted phage A511 attached to *Listeria* cells, and in the baseplate reconstructions of the spontaneously contracted phi812[23] and A511[24]. Because of their positioning after the tripod conformational change, we speculate that the hook loops of phi812 and related phages could bind the host cell wall.

The tripod proteins' conformational changes are required to form inter-tripod and tripod-RBP2 interactions in phages with contracted tails (Fig. 6B, C, SFig. 13A, B). In particular, the rotation and translation of the base domain expose an interface to which a fiber domain of the neighboring tripod complex binds through an area of 1,900 $Å^2$ (SFig. 13A, B). Each inner tripod complex also connects to an outer tripod complex attached to the same baseplate arm, and one fiber domain of each outer tripod complex interacts with tower domains 1 and 2 of RBP2 (Fig. 6A-C). The contacts mediated by the fiber domains stabilize the baseplate structure in phages with contracted tails.

The formation of interactions among the inner trimers of tripod proteins causes the upper and lower baseplate arms to shift 11 and 15 Å away from the tail axis, respectively, relative to their positions in the native baseplate (Fig. 1B, 6B, C). Simultaneously, all six baseplate arms are shifted to radiate from the tail axis at 83°, establishing the sixfold symmetry of the baseplate (Fig. 6A, B). Less extensive conformational changes were described for the receptor-binding proteins of other phages[10,63,64]. Overall, the recognition and adhesion of phi812 to the host cell is accompanied by re-orientation of the RBP1, RBP2, and conformational changes of tripod complexes, affecting the positions of the six arms of the contracted baseplate.

**Conformational changes to the baseplate core prompt tail sheath contraction**

The structure of baseplate core is affected by the re-arrangements of receptor-binding proteins and baseplate arms, which in turn triggers the contraction of the tail sheath. The arm scaffold proteins covalently link baseplate arms of phi812 to the wedge modules (Fig. 2 and 3A-C, SMovie 2). Therefore, the changes in the position and orientation of baseplate arms cause the iris-like structure formed by the wedge modules to expand in radius from 52 to 61 Å (SFig. 5A, D). The expansion of the iris is enabled by the movement of the dimerization and trifurcation domains relative to each other (SFig. 5B, C and E, SMovie 5). In contrast, the interactions between the dimerization domains of

wedge proteins remain the same as in the native baseplate (SFig. 5F). A similar expansion was described for gp6 and gp7 during the transition of the T4 baseplate from its native to contracted state[4]. The expansion of the radius of the hexamer of tail sheath initiator proteins of phi812, which probably occurs together with that of the wedge iris, is enabled by the partial unfolding of the linker helices between LysM and handshake domains (SFig. 9A, B, D, and E, SMovie 5). The unfolding of the linker enables the tail sheath initiator proteins to maintain the same contacts within the hexamer and with wedge modules as in the native baseplate (SFig. 9G, H). The handshake domain of the tail sheath initiator protein inclines 17° (SFig. 9B, E) and, together with the wedge iris expansion, triggers the contraction of the tail sheath, as discussed below.

**The native tail of phi812 can bend**

Contractile tails of phages differ in their length and mechanical properties. Whereas the ~100-nm-long tails of *Straboviridae* phages, represented by T4, are rigid[65]. A flexible contractile tail has been observed only in phage Milano (genus *Schmittlotzvirus*, unclassified family) with ~125-nm-long tail[15], and in the *Herelleviridae* phages phi812 and A511 with ~200-nm-long tails that can bend up to 90°[23,24]. The flexibility of the longer tails was speculated to be an adaptation to withstand mechanical forces without breaking[66]. Here, we determined the structure of the native phi812 tail to a resolution of 4.2 Å (Fig. 6D, SFig. 1, 2G, Table S1). Both tail tube and tail sheath proteins are arranged as six-start helices with a rise of 39.2 Å and twist of 21.2° (Fig. 6D). The tail tube protein of phi812 has an eight-stranded β-barrel fold with a protruding stacking loop, characteristic for tail tube proteins of long-tailed phages and CISs[66] (SFig. 6I). The inner surface of the phi812 tail tube is negatively charged to facilitate the passage of the phage DNA during ejection (SFig. 14A). The β-barrel domain of tube protein subunit contains a short loop between the β-strands 6 and 7, lack of which was suggested to allow bending of tail tube of phage Milano[15]. Since the loop is present in phi812 tail tube protein, bending of the phi812 tail is likely allowed by a different mechanism. The sheath protein of phi812 can be divided into domains I-IV, according to the convention established for the tail sheath protein (gp18) of phage T4[2] (SFig. 6H). In a native tail, the domains I and II of phi812 tail sheath proteins interact with tail tube proteins and mediate substantial contacts between tail sheath proteins. Domains III and IV are positioned successively next to domain II at the tail surface (SFig. 14B). The two-stranded β-sheet of domain I is augmented by accepting the N- and C-termini of tail sheath proteins from a disc positioned closer to the phage head (SFig. 14C). By mediating the handshake interactions, the domain I interconnects tail sheath proteins from successive discs. These interactions are conserved in tail sheath structures of numerous phages and CISs[4,7–15,25,26,67,68]. The domain II of the tail sheath protein is formed by six β-strands surrounded by eight α-helices, whereas domains III and IV each contain a beta-sandwich flanked by an alpha-helix (SFig. 6H). The domain III contacts domain IV of the tail sheath protein from the disc located closer to the baseplate (SFig. 14D). Additionally, domain IV interacts with domain II of the neighboring tail sheath protein from the same disc (SFig. 14D). The domain IV of the T4 gp18 is not involved in inter-subunit contacts[69]; however, the bulkier domain III of the T4 tail sheath protein (168 residues) interacts with another tail sheath protein from the disc closer to the phage head[69]. We speculate that interactions of domains III and IV of the phi812 tail sheath proteins allow a spring-like movement of the subunits relative to each other, which enables phi812 tail bending.

## Contraction of the tail sheath

In phages with contractile tails and CISs, the rearrangement of the baseplate induced by binding to a target cell wall triggers the contraction of the tail sheath, which transforms from the native state to a lower energy contracted state[70]. The released energy of the tail sheath contraction is used to drive the tail tube through the host cell wall[7,10,14,23,24]. Upon contraction, the tail sheath of phi812 broadens from 219 to 276 Å in diameter, which is coupled to a change in the rise and twist of the six-start helix formed by the tail sheath proteins to 18.8 Å and 30.6°, respectively (Fig. 6E, Sfig. 1 and 2H, I, Table S1). The successive discs of the contracted tail sheath are brought closer to each other due to the tilting of domain I 47° compared to its native conformation while maintaining the inter-subunit handshake interactions among domains I from successive discs (Fig. 6F, SFig. 14E, F). The buried surface area of intersubunit contacts per tail sheath monomer increases from ~5,500 $Å^2$ in the native tail sheath to ~9,500 $Å^2$ in the contracted tail sheath. The enlarged contacts stabilize the straight structure of the contracted tail sheath, which is less flexible than the native one[23]. During contraction, the individual domains of the tail sheath protein of phi812, unlike those of T4 and some CISs[8,69,71], move relative to each other, which enables the formation of the more extensive contacts between the subunits (SFig. 14G).

It has been shown for phages T4 and A511 and the R-type diffocin that the contraction of the tail sheath is gradual and occurs as a wave along the tail axis[24,46,70,72]. In agreement with the previous *in vitro* observations, we captured particles of phi812 with a partially contracted tail sheath attached to the *S. aureus* cell wall (SFig. 12A, B). Furthermore, the cryo-EM reconstructions of baseplate-proximal parts of the tails in the native and contracted state provide structures of discrete intermediates of contraction (SFig. 2B, E, and 16A-D, SMovie 5). The first three baseplate-proximal tail sheath discs of the native phi812 tail sheath differ in conformation from those in the central part of the native tail sheath (SFig. 15A, B). The deviation occurs gradually, with the first tail sheath disc being the most affected (SFig. 15B). Similar deviations of baseplate-proximal tail sheath proteins were observed in the R2 pyocin and T6SS[7,73]. The first three sheath discs of phi812 have their organization shifted toward that of the contracted tail sheath, which may reduce the activation energy required to initiate the tail sheath contraction[74].

In the contracted state of phi812, the structures of the first six baseplate-proximal tail sheath discs deviate from those from the central part of the contracted tail sheath (SFig. 15C, D). Similar to the native tail sheath, the first contracted sheath disc is the most affected and the sixth one the least affected (SFig. 15D). The first two baseplate-proximal sheath discs have the rise increased by 5.5 Å relative to the tail sheath disc from the center of the contracted tail sheath (SFig. 15D). The absence of a twist difference may be due to the more extensive contacts between consecutive tail sheath discs than in the native tail, which is probably incompatible with deviations in the angle of the helical twist. Interactions with the sheath initiator and wedge module proteins likely cause the altered structure of the baseplate-proximal tail sheath discs and directly couple the conformational changes of baseplate proteins with the sheath contraction.

## Changes in the baseplate structure required for cell wall penetration

Phages with contractile tails infecting Gram-positive bacteria employ enzymes that degrade teichoic acid and peptidoglycan to facilitate tail tube penetration through the cell wall[54,75]. Phage phi812

449 probably uses the petal domain of the central spike protein to cleave teichoic acid (Fig. 4B, C and 7A,
450 B); however, the central spike protein needs to be released from the tail channel to enable the
451 genome ejection. We speculate that the conformational change to inner tripod trimers, which
452 disrupts their interaction with weld proteins, triggers the release of the weld and central spike
453 proteins from the baseplate (Fig. 7B, C, SFig. 8A-D). Dissociation of the weld and central spike
454 proteins is required to enable the cleaver domains of the hub protein to interact with and degrade
455 the cell wall peptidoglycan (Fig. 4D, E, SFig. 8D). This indicates that the weld and central spike
456 proteins of phi812 are released before the cell wall penetration by the tail tube (Fig. 7D). In contrast,
457 the central spike proteins of phages T4 and E217, which infect Gram-negative bacteria, are released
458 from baseplates after puncturing the host outer membrane[10,76,77].
459 As the phi812 tail sheath contracts, the tip of the tail tube is pushed through the degraded
460 teichoic acid and peptidoglycan layers while rotating about its axis by 475° (Fig. 7D). The rotation of
461 the hub proteins at the tip of the tail tube can enhance their peptidoglycan degradation activity by
462 bringing them into contact with more substrate. Once the peptidoglycan layer is breached, the clamp
463 domains of the hub proteins can interact with the *S. aureus* cytoplasmic membrane. It was shown
464 that after the penetration of the outer membrane and peptidoglycan layer, gp27 from T4 (a homolog
465 of the hub protein from phi812) remains associated with the tip of the tail tube and probably
466 interacts with ejected tape measure protein to form a pore through the inner membrane of *E. coli*[78].
467 However, the tail tube of phi812 penetrates 10-30 nm into the *S. aureus* cytoplasm, depending on
468 the thickness of the cell wall, which varies between 50-70 nm[23] (Fig. 7E). Therefore, unlike in T4, the
469 hub proteins of phi812 are probably not required to form a pore in the *S. aureus* membrane, and
470 their role in genome delivery may be limited to facilitation of the penetration of the cell wall and
471 cytoplasmic membrane by the tail tube tip.
472
473 **Conclusions**

474 In this study, we employed a combination of cryo-EM, X-ray crystallography, NMR, and structure
475 prediction methods to elucidate the intricate mechanisms employed by phage phi812 to penetrate
476 the *S. aureus* cell wall in preparation for transferring its genetic material into the host cytoplasm.
477 While most of the previously structurally characterized phages with contractile tails target
478 Gram-negative bacteria, our focus has been on phi812 attacking *S. aureus*, a representative of
479 Gram-positive bacteria, which are comparatively less explored. We show that the baseplate of the
480 phage phi812 comprises a core, wedge modules, and baseplate arms that carry RBP1, RBP2, and
481 tripod complexes (Fig. 1A-C, SMovies 1, 2). Upon binding to the *S. aureus* cell wall, the baseplate
482 symmetry transitions from threefold to sixfold, which is enabled by a series of conformational
483 changes that propagate from the baseplate periphery to the core (Fig 6A, B, SMovies 3, 4). These
484 changes encompass a reorientation of receptor-binding proteins and conformational changes to
485 tripod complexes to enable host cell surface binding (Fig. 7A, B, , SMovie 3). During the attachment,
486 the central spike protein degrades wall teichoic acid (Fig. 7A, B). Conformational changes to the
487 tripod complexes trigger the release of the central spike and weld proteins from the baseplate, which
488 enables the hub proteins to cleave peptidoglycan (Fig. 7C, D) and facilitate penetration of the tail
489 tube through the cell membrane (Fig. 7D, E). Changes in the baseplate arms' positions relayed
490 through wedge modules to the hexamer of tail sheath initiator proteins cause an expansion of its
491 diameter, triggering a tail sheath contraction (Fig. 7C-E, SMovie 5). The baseplate-proximal tail sheath
492 discs of the native phage particle have an altered structure, which may lower the energy barrier

493 required to initiate the tail sheath contraction. The tail sheath shortens from 200 to 96 nm upon
494 contraction[23], pushing the tail tube's tip 10-30 nm into the cytoplasm, depending on the local
495 thickness of the *S. aureus* cell wall. The molecular characterization of the phi812 baseplate structure
496 and the mechanism of cell wall binding open up the possibility of a rational design of phi812 to target
497 specific strains of *S. aureus*.

# Figures

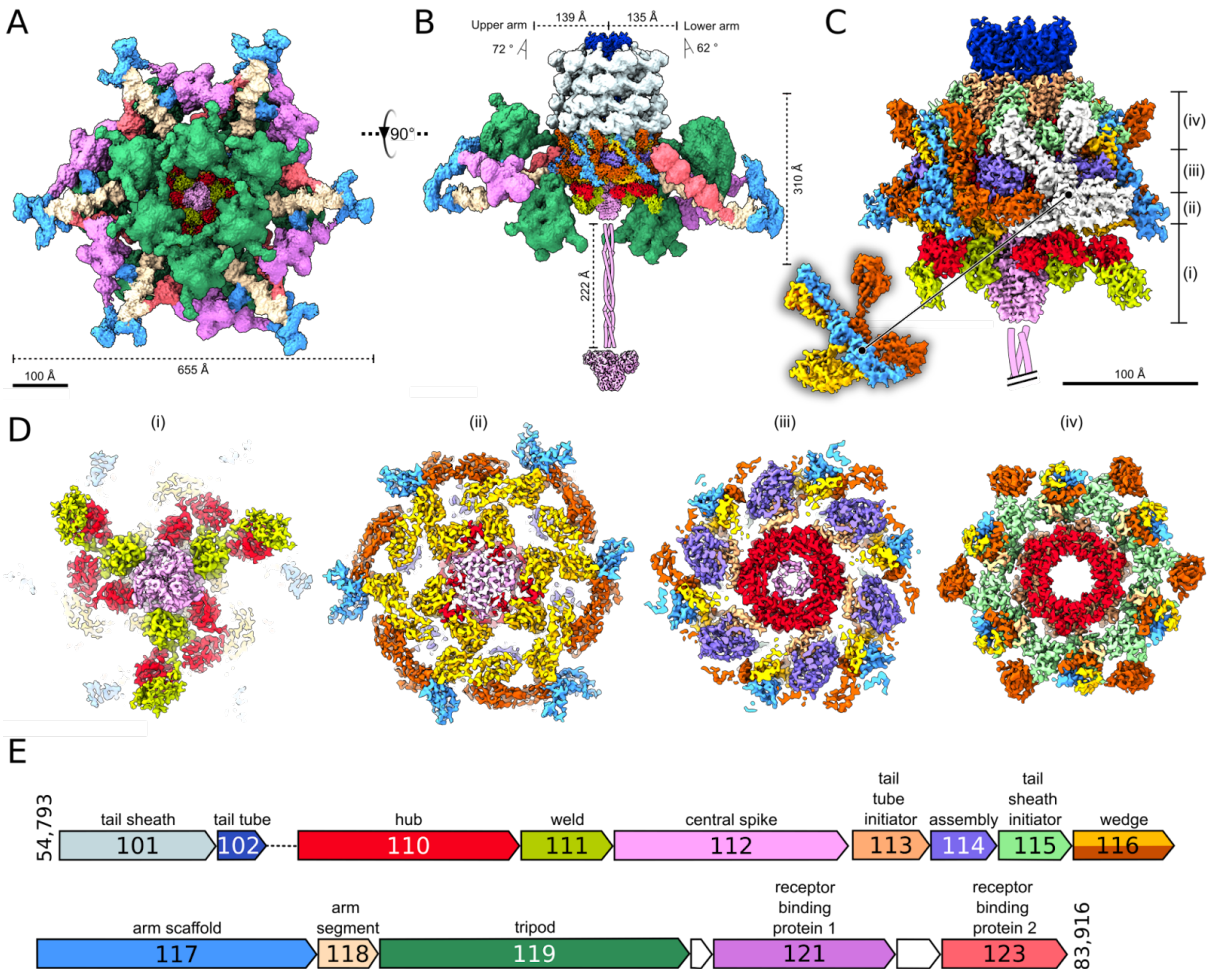
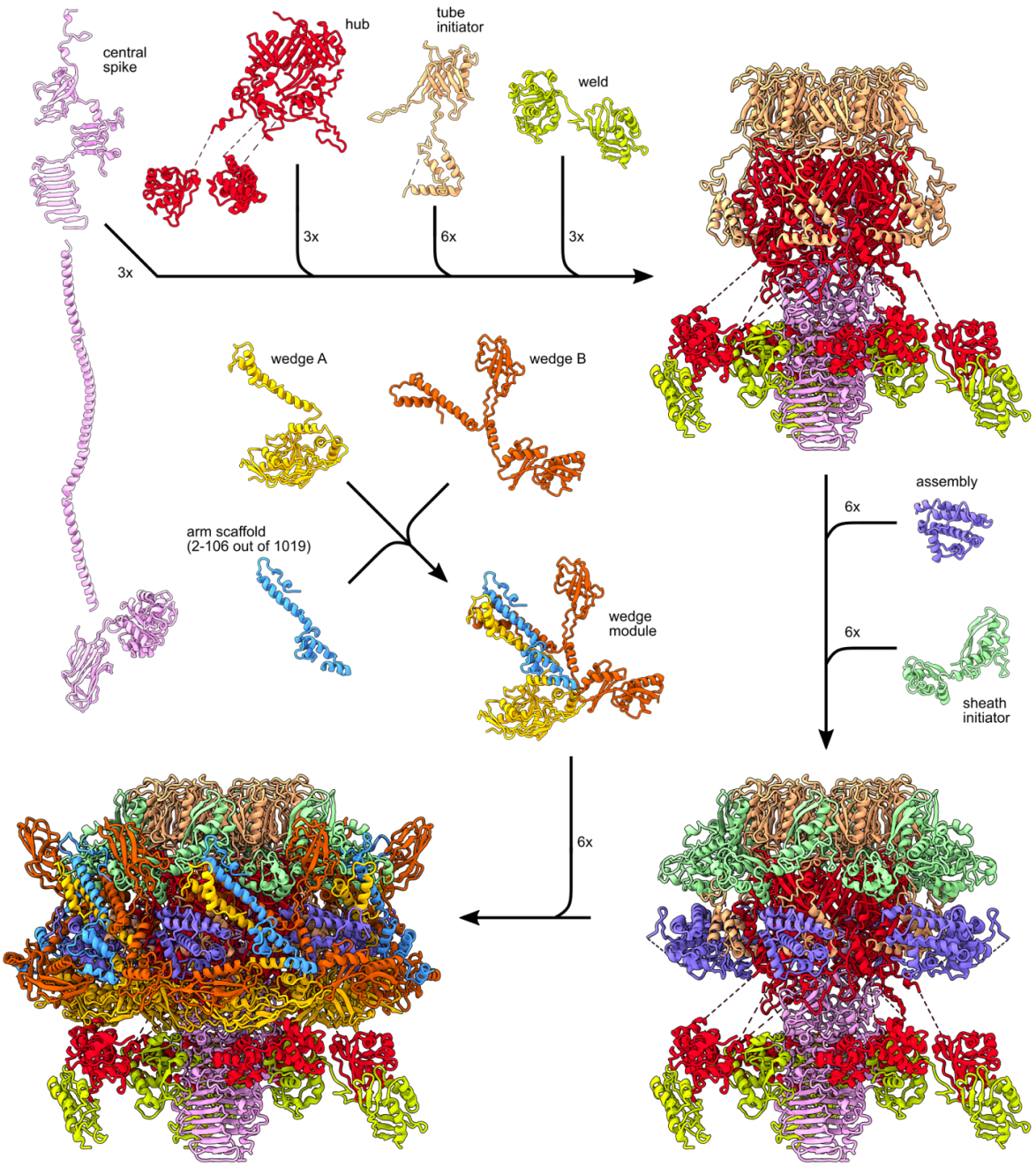



**Fig. 1. Structure of native phi812 baseplate. (A-D)** The surface representation of the composite cryo-EM map of the phi812 baseplate is viewed along (A, D) and perpendicular to (B, C) the tail axis. Protein components of the baseplate are colored according to the scheme depicted in panel E. The part of the central spike protein protruding from the baseplate is not shown in (A). **(B)** Side view of phi812 baseplate. Only two opposing baseplate arms are shown to display the structure of the baseplate core. The height of the baseplate, angles of baseplate arms relative to the tail axis, distances of baseplate arms from the tail axis, and length of the coiled-coil domain of the central spike protein are indicated. **(C)** A complex of tail tube disc, baseplate core proteins, and wedge modules. A single wedge module shown in white in the complex is displayed at the bottom left to highlight interactions among the arm scaffold and two wedge proteins. The positions of sections shown in panel (D) are indicated on the right. **(D)** Sections through the baseplate core proteins and wedge modules at positions indicated in panel C. **(E)** Scheme of part of the phi812 genome encoding tail and baseplate genes.




514

515

516 **Fig. 2. Components of phi812 baseplate core and wedge modules.** The baseplate core and wedge
517 module proteins are shown in cartoon representation and colored according to Fig. 1E. Protein
518 names and the numbers of subunits of each protein forming the baseplate core and wedge modules
519 are indicated.

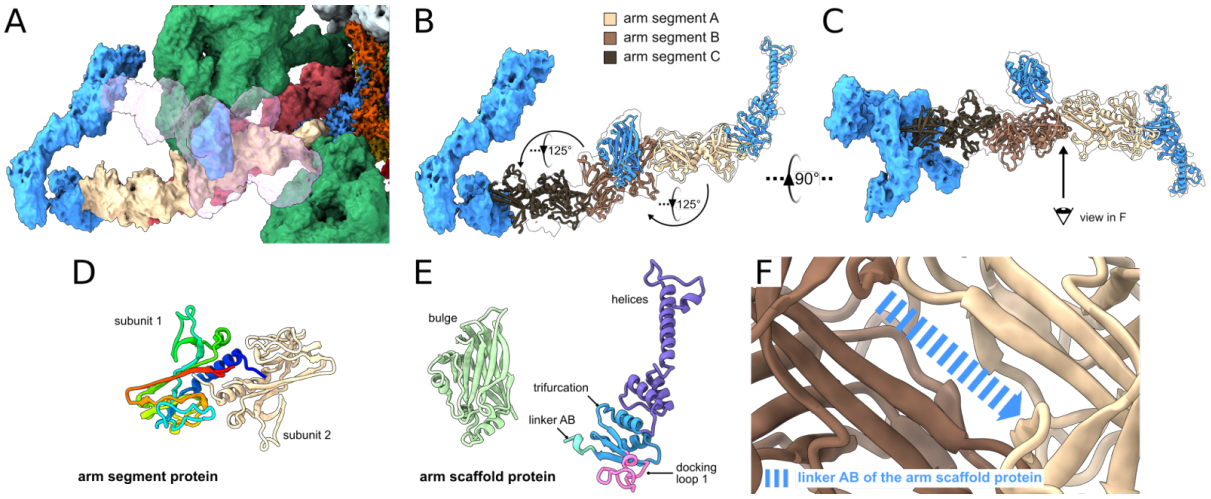



**Fig. 3. Structure of phi812 baseplate arm. (A)** Surface representation of cryo-EM reconstruction of phi812 baseplate in native conformation, showing one baseplate arm with attached receptor-binding proteins. The densities corresponding to the baseplate arm, wedge module, and tail sheath proteins were segmented and are shown in color according to Fig. 1E, except for RBP1, which is shown as a semi-transparent surface. **(B, C)** The N-terminal helices, trifurcation, and bulge domains of the arm scaffold protein and arm segment proteins are shown in cartoon representation. The corresponding cryo-EM map of a native baseplate is depicted as a semi-transparent white surface with a black outline. Cryo-EM density of the distal part of the baseplate arm, probably corresponding to elbow and connector domains, is shown as a blue surface. The arm segment protein dimers A, B, and C are colored according to the legend. **(C)** Same as B, but rotated 90°. The direction in which the interface between arm segment protein dimers A and B is viewed in panel F is indicated. **(D)** Cartoon representation of the arm segment protein dimer, with one of the subunits rainbow colored from the N-terminus in blue to the C-terminus in red. **(E)** Cartoon representations of the bulge, helices, and trifurcation domains of the arm scaffold protein predicted using AlphaFold2. **(F)** Cartoon representation of the interface between the C-terminal β-strands of arm segment dimer A (beige) and dimer B (light brown) shown in cartoon representation. The dashed blue arrow indicates a speculative position of a beta-strand formed by the linker AB of the arm scaffold protein.

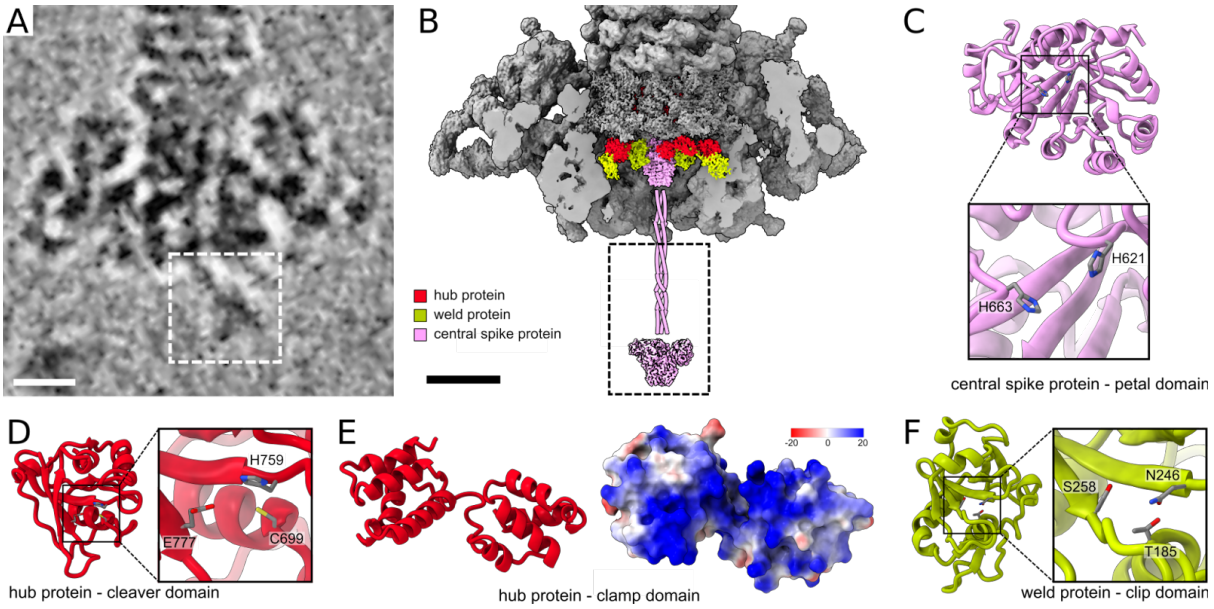



**Fig. 4. Central spike, hub, and weld proteins. (A)** Cryo-electron micrograph of the phi812 baseplate. The white dashed rectangle contains the coiled-coil, knob, and petal domains of the trimer of central spike proteins. Scale bar 100 Å. **(B)** The surface representation of the cryo-EM reconstruction of the native phi812 baseplate is colored in gray except for central spike, hub, and weld proteins. The front part of the baseplate has been removed. Scale bar 100 Å. **(C)** Cartoon representation of the petal domain of the central spike protein. Side chains of residues forming a putative hydrolytic site are shown in stick representation. The inset shows details of the putative hydrolytic site. **(D)** The cleaver domain of the hub protein is shown in cartoon representation with side chains of residues forming a putative hydrolytic site shown in stick representation. **(E)** Clamp domain of the hub protein shown in cartoon representation (left) and the molecular surface colored according to Coulombic electrostatic potential (right) in the range from -20 to 20 kcal/(mol·*e*) at 298 K. **(F)** Cartoon representation of the clip domain of the weld protein with side chains of residues forming a putative receptor-binding site shown in stick representation.

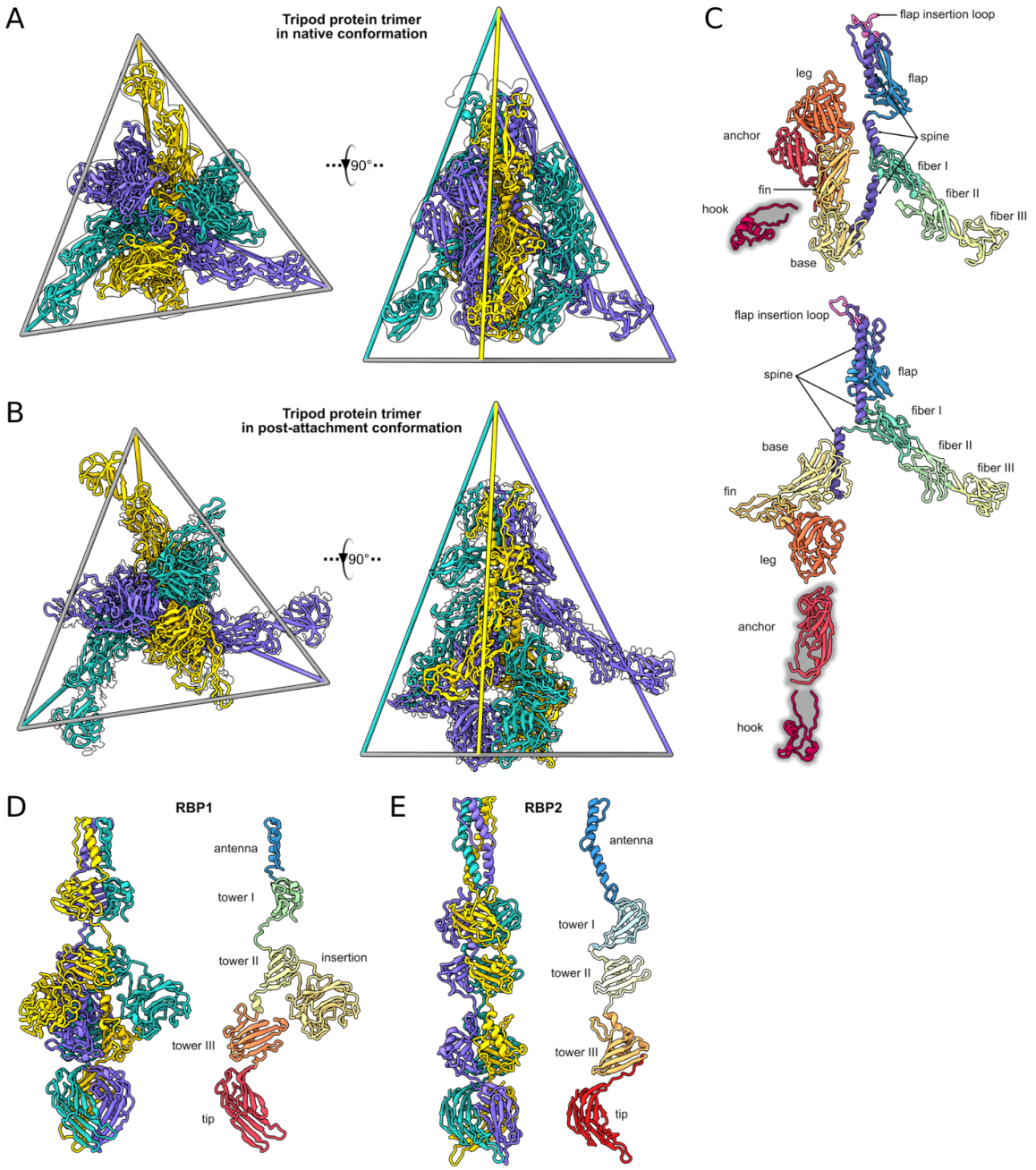



**Fig. 5. Tripod and receptor-binding proteins. (A)** Cartoon representation of an AlphaFold2-predicted structure of a trimer of tripod proteins in native conformation. Subunits are distinguished by color. The structure is fitted into the cryo-EM density of a native baseplate, which is shown as a black outline. Only the domains of tripod proteins for which the density was resolved are shown. **(B)** The trimer structure of tripod proteins in the post-attachment conformation was determined using cryo-EM reconstruction of the contracted phi812 baseplate. Only the domains of tripod proteins for which the density was resolved are shown. Triangular cages in (A) and (B) highlight the shape of the tripod protein complexes. **(C)** Cartoon representations of tripod proteins in native (top) and post-attachment (bottom) phi812 baseplates. The structures are colored according to domain. Domains with gray backgrounds were not resolved in the corresponding reconstructions, and their positions are speculative. **(D, E)** RBP1 (D) and RBP2 (E) structures predicted using AlphaFold2 are shown in cartoon representations colored according to subunit (left). A single subunit of each protein

569 colored according to domain is shown (right).

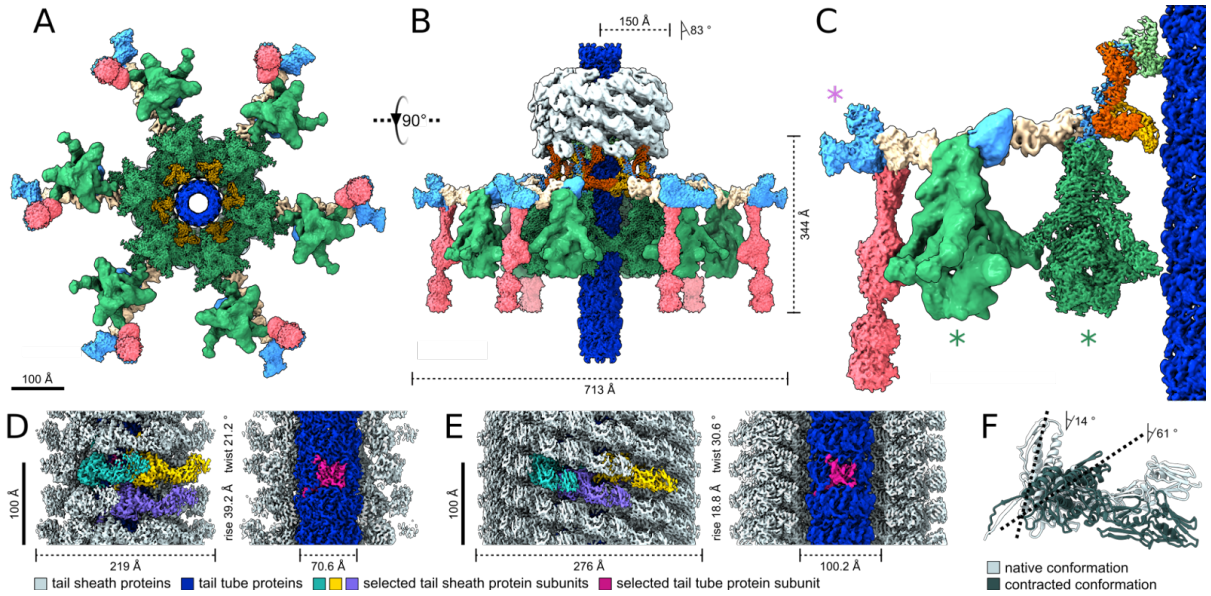



**Fig. 6. Structure of baseplate from phi812 particle with contracted tail and conformational changes of phi812 tail sheath induced by contraction. (A-B)** The surface representation of the composite cryo-EM map of the phi812 baseplate is viewed along (A) and perpendicular to (B, C) the tail axis. Protein components are colored according to the scheme depicted in Fig. 1E. (B) The width and height of the baseplate are indicated at the bottom and on the right, respectively. The angle of the baseplate arm and its distance from the tail axis are displayed at the top. **(C)** One arm of phi812 baseplate. A pink asterisk indicates the putative attachment site of RBP1, which was not resolved in the reconstruction. Green asterisks indicate volumes occupied by anchor and hook domains of the tripod complexes, which were not resolved in the reconstruction. **(D, E)** Surface representation of cryo-EM map of the native (D) and contracted (E) tail of phi812. For each panel, complete maps are shown on the left, and maps with the front half of the tail sheath removed on the right. Protein components are colored according to the legend below the panel. **(F)** Comparison of the structures of tail sheath proteins in native and contracted conformations. The structures are shown in cartoon representation, and their domains II were superimposed. Angles of the long axes of domain I relative to the tail axis are indicated for each conformation.

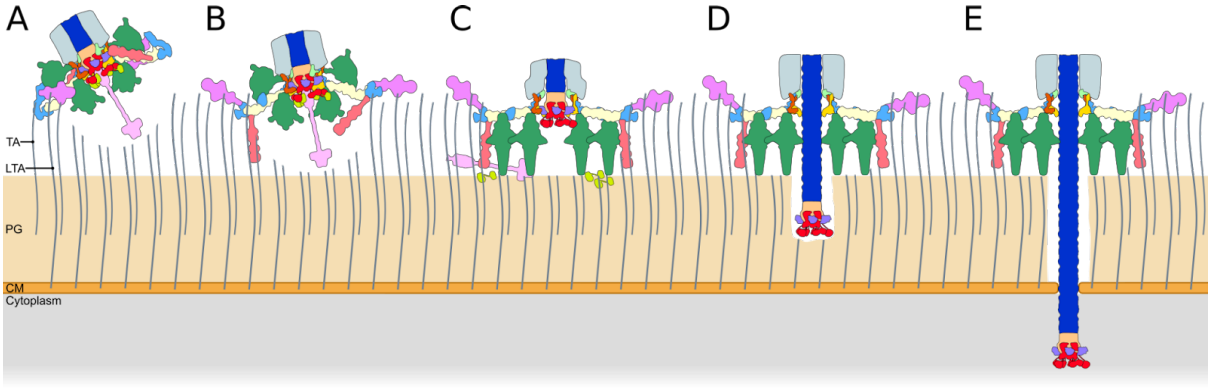
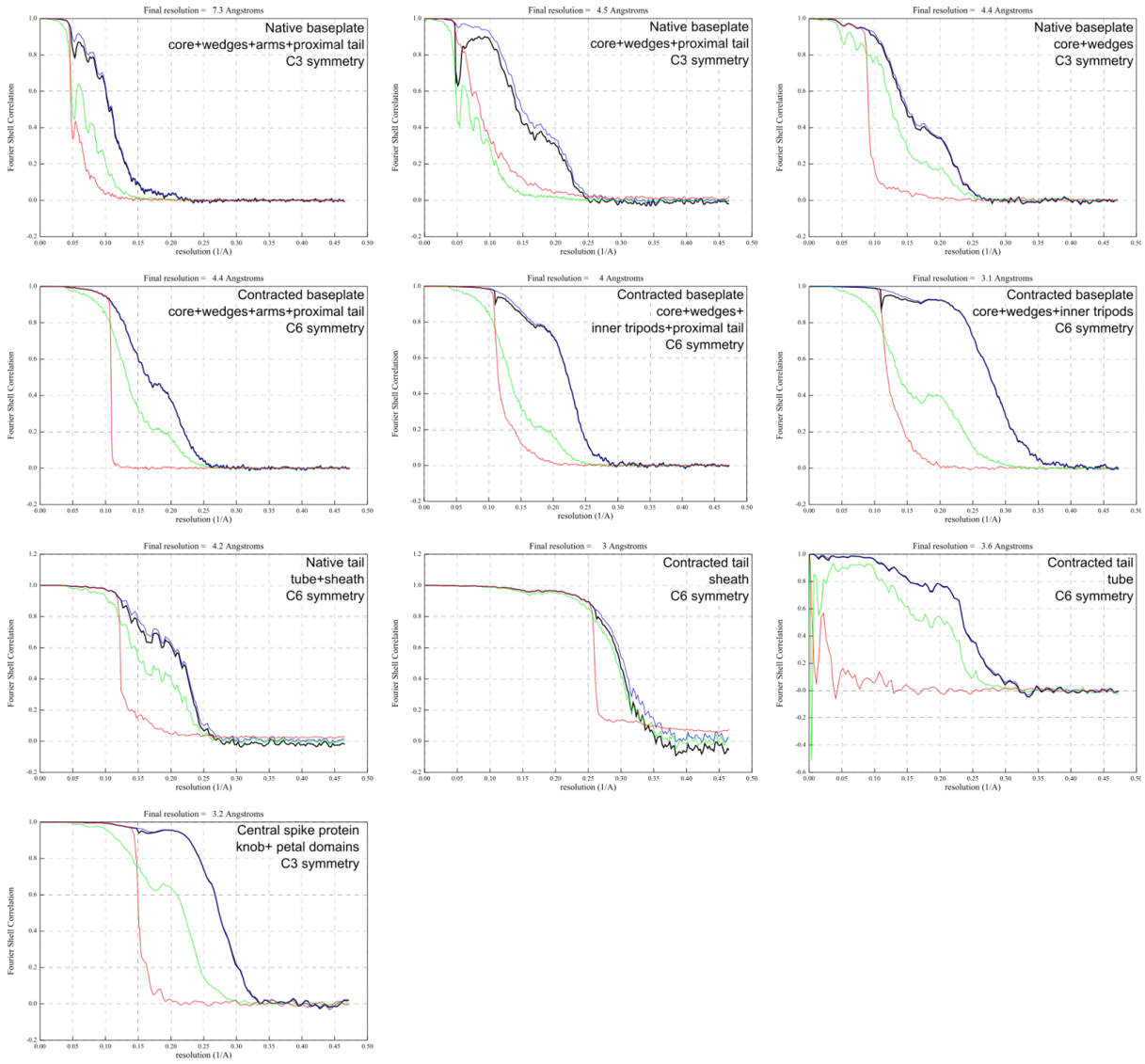
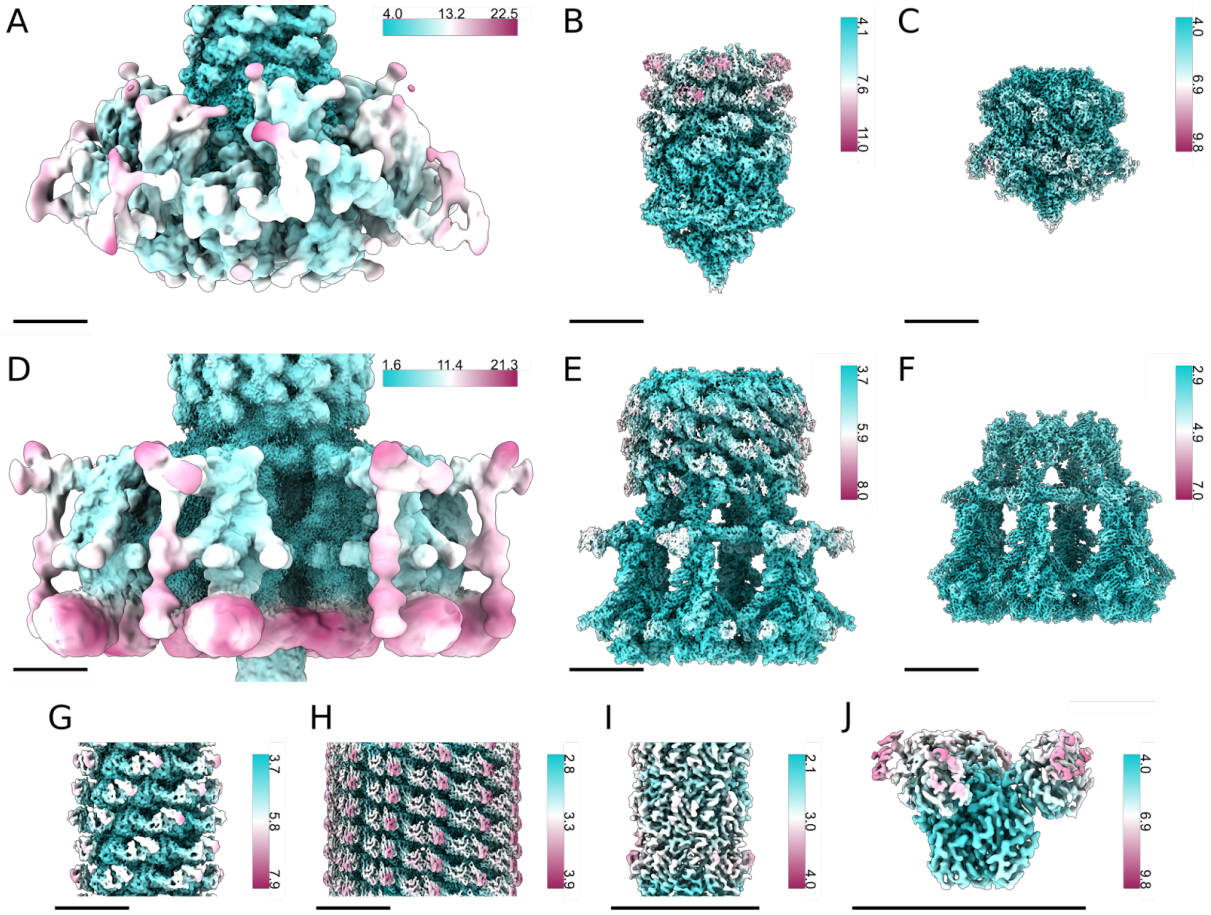



**Fig. 7. Scheme of phi812 cell adhesion and penetration. (A)** Upon contact of the phage phi812 baseplate with the cell wall of *S. aureus*, trimers of RBP1 bind to (lipo)teichoic acid. The petal domains of the central spike proteins cleave teichoic acid below the baseplate. **(B)** Trimers of RBP2 rotate and bind to teichoic acid. **(C)** Tripod complexes undergo a conformational change and anchor the baseplate to the cell wall. Central spike and weld proteins detach from the baseplate. The diameter of the iris formed by wedge modules increases, leading to a conformational change to tail sheath initiator proteins and baseplate-proximal sheath discs. The remaining baseplate core proteins detach from the sheath initiator and wedge module proteins. **(D)** As the contraction of the sheath discs propagates towards the phage head, the tail tube with tail tube initiator, hub, and assembly proteins at its tip is driven through the cell wall. The cleaver domain of the hub protein degrades the peptidoglycan. **(E)** The tail tube penetrates the cytoplasmic membrane, and its tip reaches the cytoplasm. However, whether the baseplate core proteins remain associated with the tip of the tail tube during the membrane penetration remains unknown.

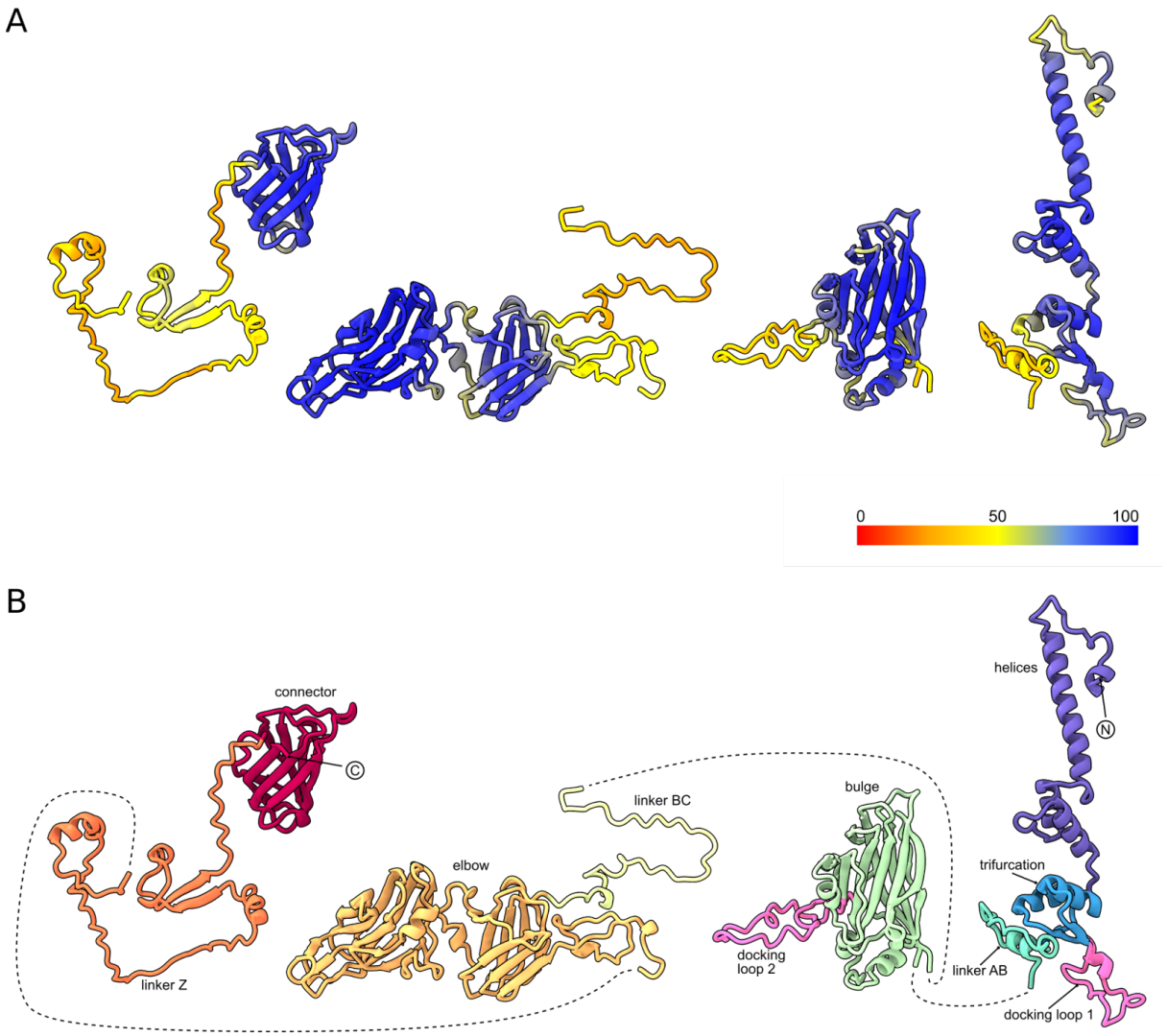
# Supplementary figures



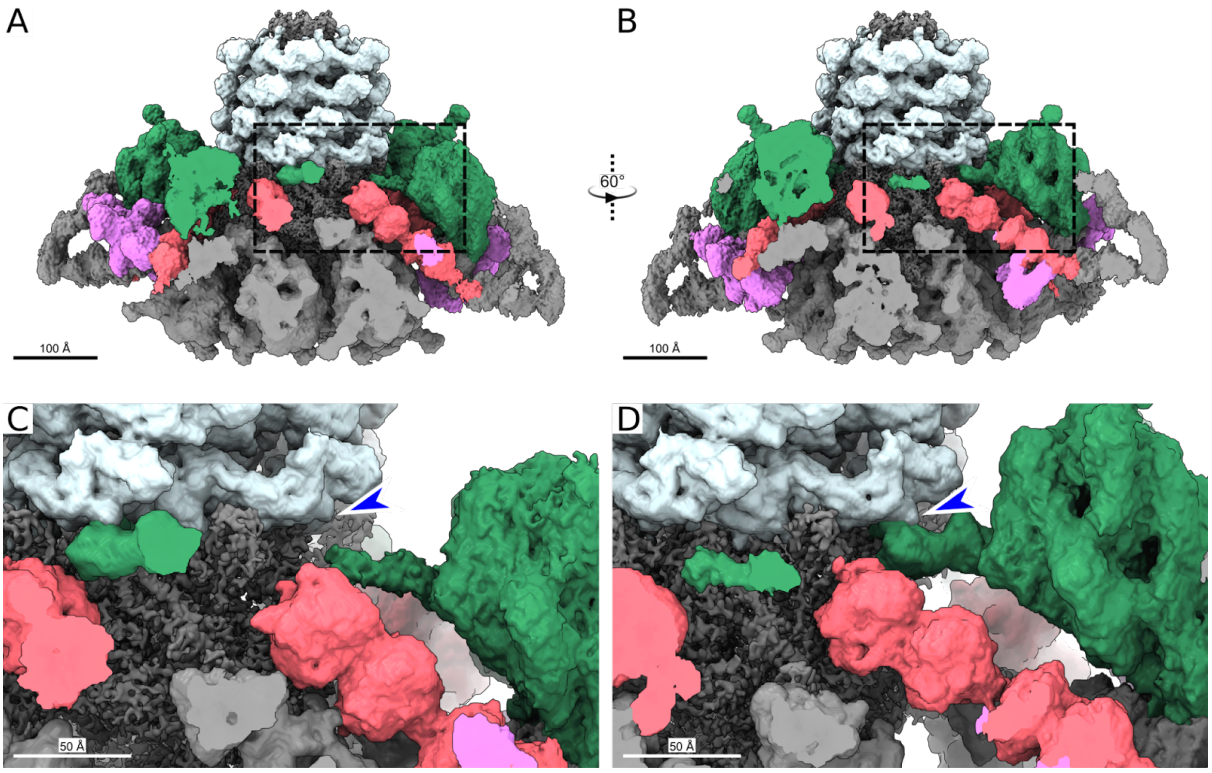

**SFig. 1. FSC curves of cryo-EM reconstructions.** Each panel shows a comparison of Fourier shell correlation curves of CTF-corrected half-maps (black), unmasked half-maps (green), masked half-maps (blue), and phase-randomized and masked half-maps (red).



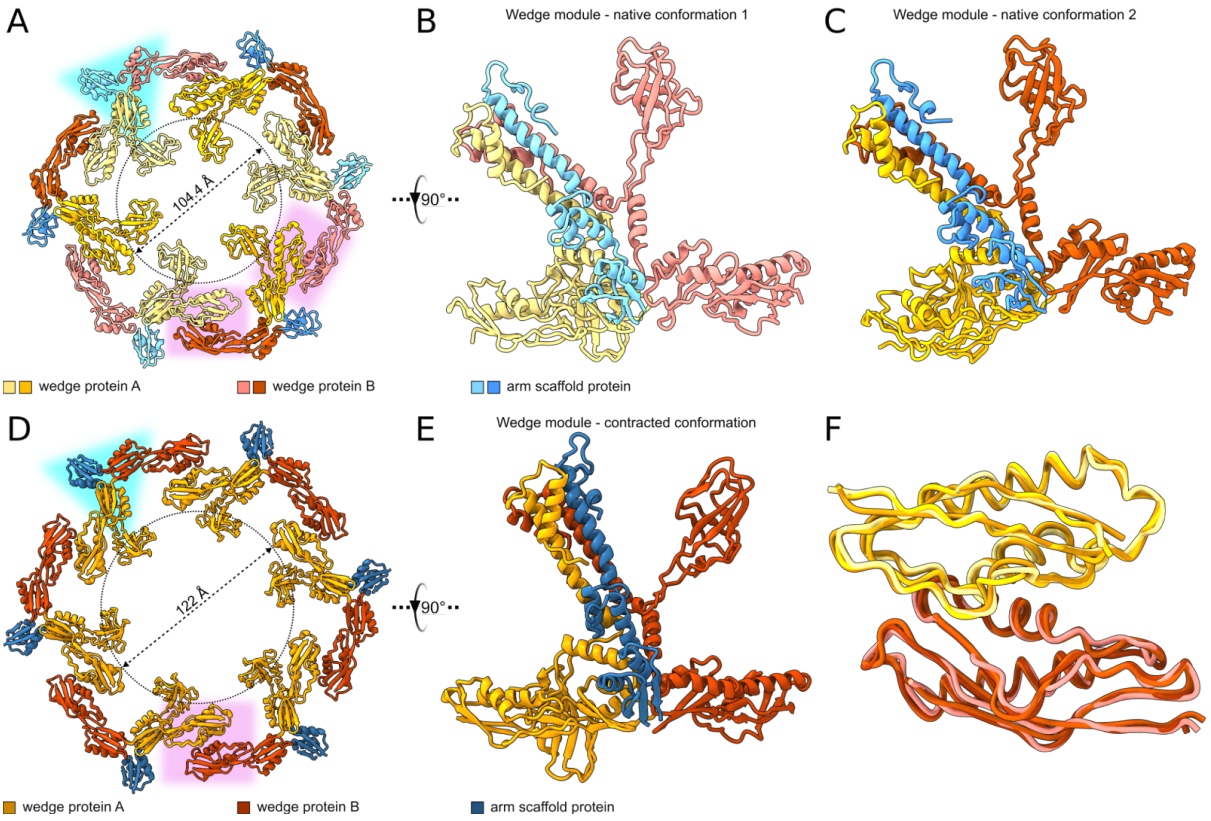

**SFig. 2. Local resolution maps of cryo-EM reconstructions. (A-J)** Surface representations of cryo-EM reconstructions colored according to the local resolution in Å. Scalebars = 100 Å. **(A)** Native baseplate. **(B)** Core, wedge modules, and three tail sheath discs of the native baseplate. **(C)** Core and wedge modules of the native baseplate. **(D)** Contracted baseplate. **(E)** Tail sheath initiator proteins, wedge modules, inner tripod proteins, arm segments A, and six tail sheath discs of the contracted baseplate. **(F)** Sheath initiator proteins, wedge modules, and inner tripod proteins of the contracted baseplate. **(G)** Native tail sheath and tail tube. **(H)** Contracted tail sheath. **(I)** Tail tube from the contracted tail. **(J)** Knob and petal domains of the central spike protein.



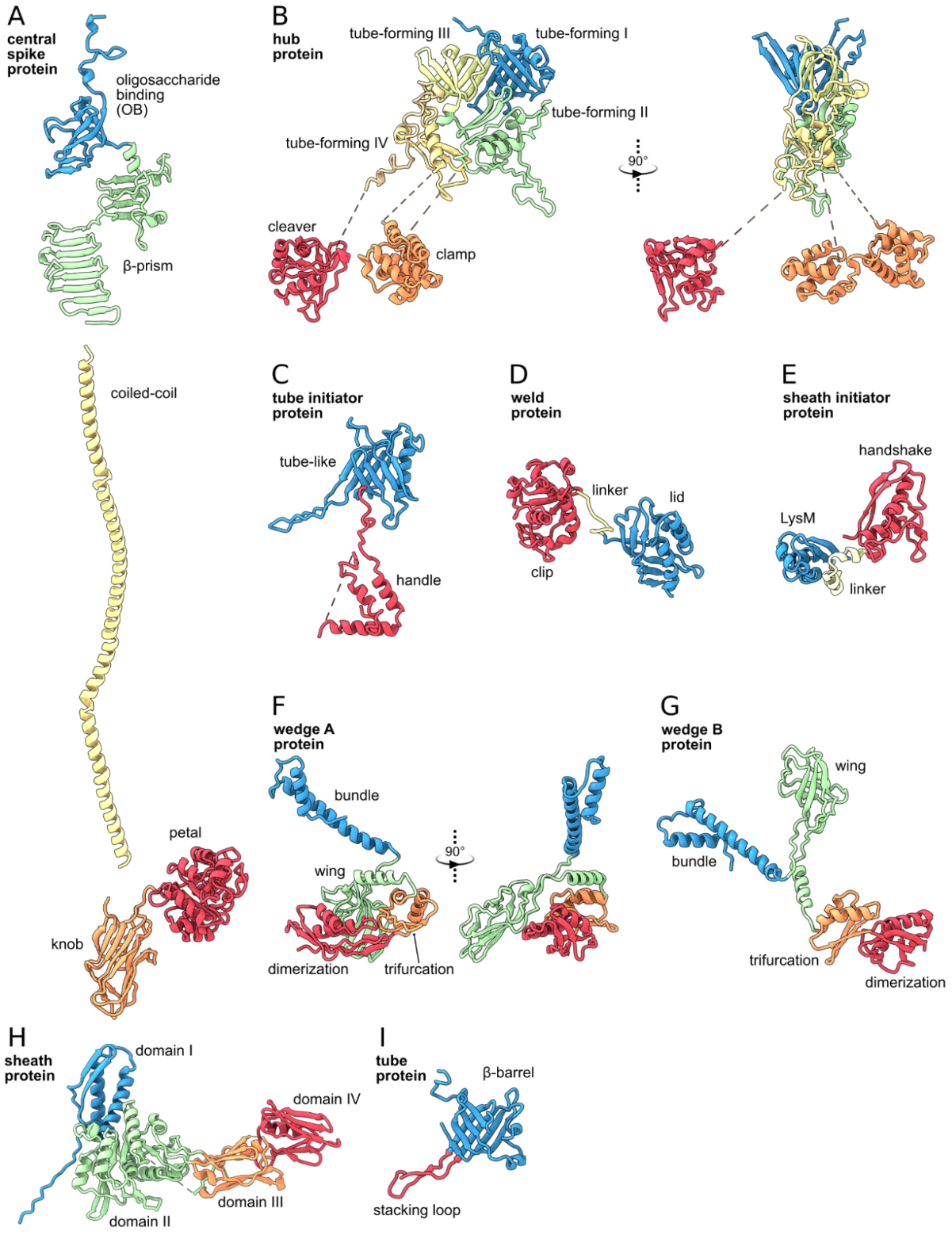

**SFig. 3. Arm scaffold protein. (A,B)** AlphaFold2-predicted structure of the arm scaffold protein shown in cartoon representation and colored according to pLDDT score[52] (A) and domain composition (B). The structure was divided into several parts for better visualization. In (B), domain names and chain termini are indicated. Dashed lines indicate connections among domains.



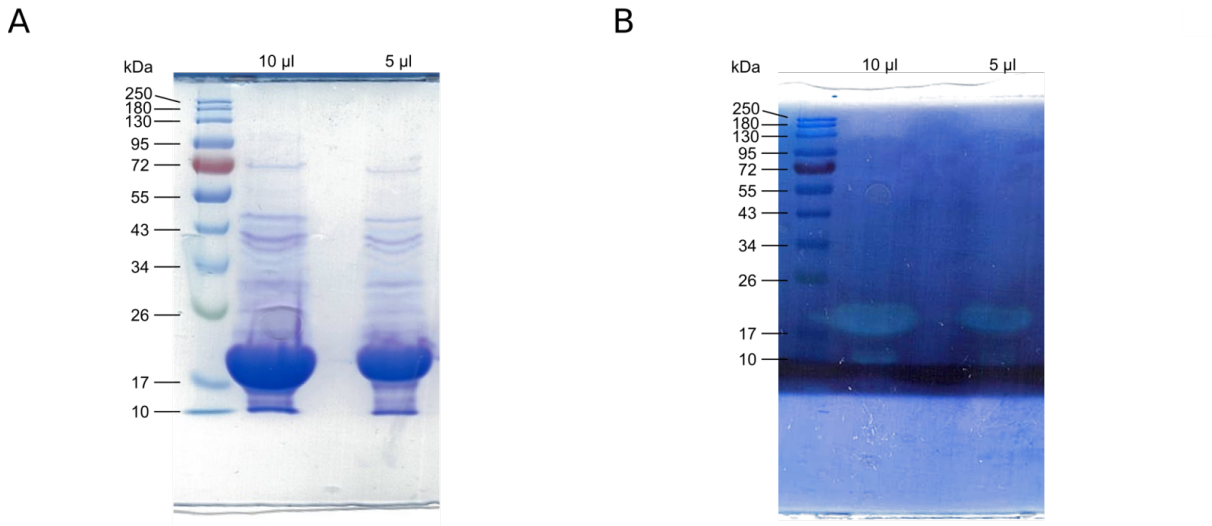

**SFig. 4. Comparison of interactions of tripod proteins, RBP1, and RBP2 attached to upper and lower baseplate arms with baseplate core and tail sheath.** Surface representation of a composite cryo-EM map of the native phi812 baseplate. All proteins are colored gray except the tail sheath proteins in white, tripod proteins in green, RBP1 in magenta, and RBP2 in firebrick red. The front part of the baseplate density has been removed to display the interactions of the receptor binding proteins from the lower **(A, C)** and upper **(B, D)** baseplate arm with the baseplate core and tail sheath. Rectangles in A and B indicate close views, as shown in C and D. Blue arrowheads point to the sites where the interactions of the lower (C) and upper (D) baseplate arms with the tail sheath differ.



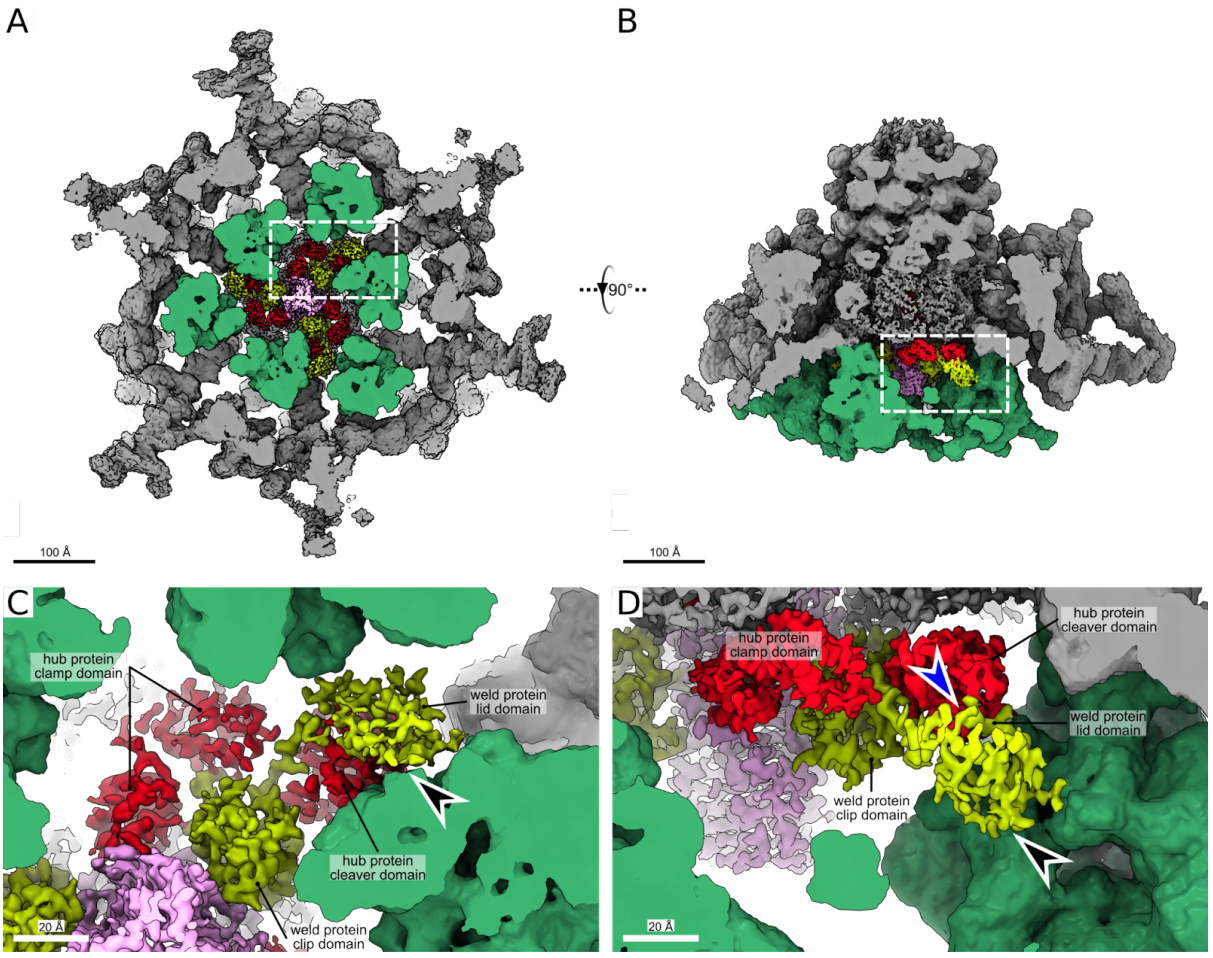

**SFig. 5. Expansion of iris formed by wedge modules upon phi812 tail contraction. (A-C)** Cartoon representation of wedge modules from the native phi812 baseplate is colored according to the legend at the bottom. (A) Six wedge modules forming the iris are viewed along the tail axis towards the phage head. Three wedge modules that enable the attachment of lower arms are shown in lighter colors, and the three wedge modules that enable the attachment of upper arms are shown in darker colors. The hetero-trimer interface formed by trifurcation domains is highlighted in cyan. Two conformationally different dimer interactions of dimerization domains are highlighted with magenta backgrounds. (B, C) A single wedge module in conformation 1 (B) and 2 (C). **(D-F)** Wedge modules from the contracted baseplate of phi812 are shown in cartoon representation and colored according to the legend at the bottom. (D) Six wedge modules forming the iris structure viewed along the tail axis towards the phage head. The hetero-trimer interface formed by trifurcation domains is highlighted in cyan. The interfaces of dimerization domains are highlighted in magenta. A dotted circle indicates the inner diameter of the wedge iris. (E) A single wedge module from the contracted baseplate. (F) Superimposition of three pairs of dimerization domains, which are highlighted in magenta in panels A and D.



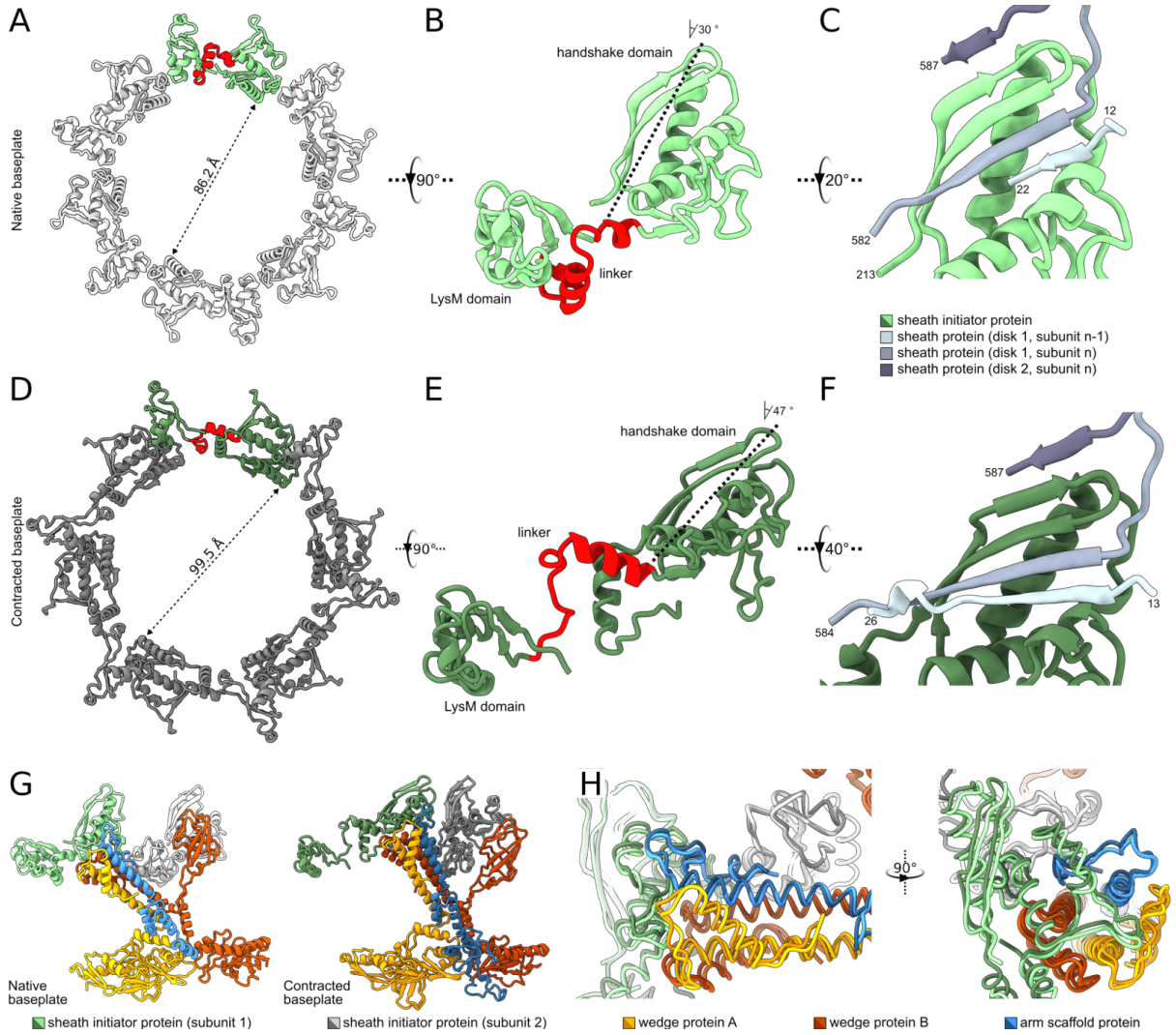

**SFig. 6. Domain organization of proteins forming baseplate core, wedge modules, and tail of phi812. (A-I)** The proteins are shown in cartoon representation and colored according to domains and functional elements. The names of proteins are in bold.



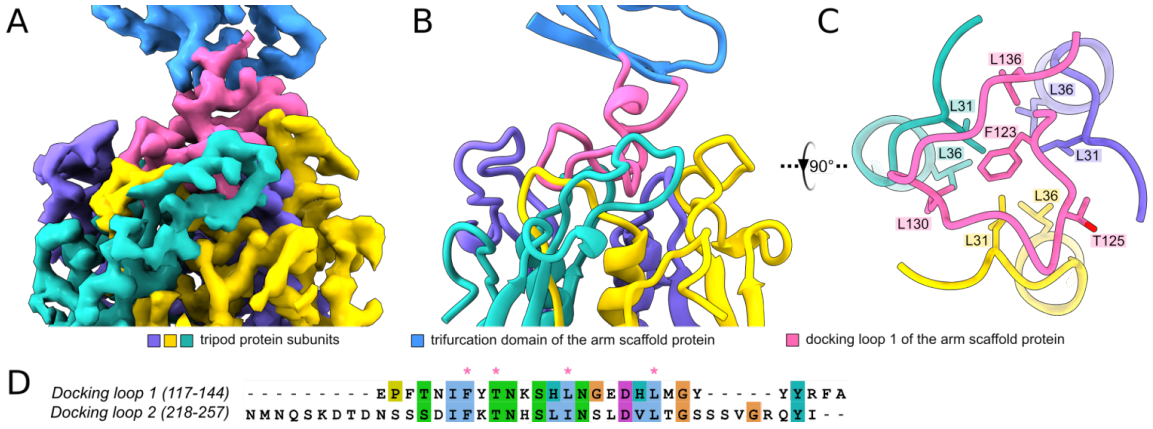

**SFig. 7. SDS-PAGE and zymogram analysis of cleaver domain of hub protein.** Scan of an SDS-PAGE gel **(A)** and zymogram **(B)** of recombinantly expressed cleaver domain of the hub protein after Histrap purification. As indicated, ten and five µl of the purified protein sample concentrated to 10 mg/µl were loaded.




**SFig. 8. Interactions of weld proteins with baseplate.** The surface representation of the composite cryo-EM map of the phi812 baseplate is viewed along the tail axis towards the phage head **(A, C)**, and perpendicular to the tail axis **(B, D)**. Baseplate proteins are shown in gray except for the hub protein in red, the weld protein in light green, the central spike protein in pink, and the inner tripod proteins in dark green. The front density in panels A and B has been clipped away to show the interactions of the weld protein with the rest of the baseplate. Rectangles in A and B indicate close views, as shown in (C) and (D). Black arrowheads in C and D point to the interaction between the lid domain of the weld protein and the inner tripod protein. The blue arrowhead in D points to the site where the lid domain of the weld protein covers the active site of the cleaver domain of the hub protein. The names of domains of hub and weld proteins are indicated.



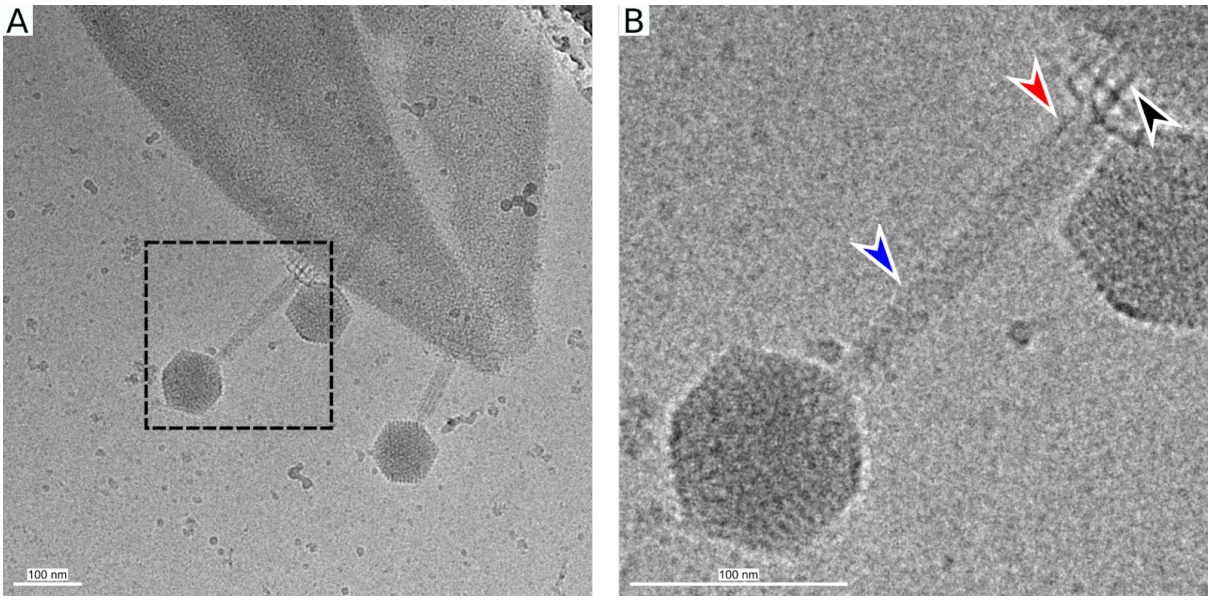

**SFig. 9. Conformational changes and β-sheet augmentation of handshake domain of tail sheath initiator protein. (A-C)** Cartoon representation of the tail sheath initiator protein from the native phi812 baseplate. (A) Five subunits from the hexamer of the tail sheath initiator proteins are shown in light gray, and one subunit in light green, with the linker between the LysM and handshake domains in red. (B) The structure of the tail sheath initiator protein with the linker between the LysM and handshake domains is shown in red. The angle between the tail axis and the axis of the handshake domain is indicated. (C) Augmentation of the β-sheet of handshake domain of the tail sheath initiator protein in the native phi812 baseplate with β-strands formed by an N-terminus and C-termini from three tail sheath proteins colored in three shades of gray. **(D-F)** Structure of tail sheath initiator proteins from the baseplate of phi812 with a contracted tail. (D) The hexamer of the tail sheath initiator proteins with five subunits is shown in dark gray, and one subunit in dark green with the linker between the LysM and handshake domains in red. (E) Structure of the tail sheath initiator protein. (F) Augmentation of the β-sheet of handshake domain of tail sheath initiator protein in phi812 with a contracted tail with β-strands formed by an N-terminus and C-termini from three tail sheath proteins colored in three shades of gray. **(G)** Interactions of tail sheath initiator proteins and a wedge module in native and contracted baseplates. **(H)** Differences between the interactions of tail sheath initiator proteins and wedge modules in native and contracted baseplates.



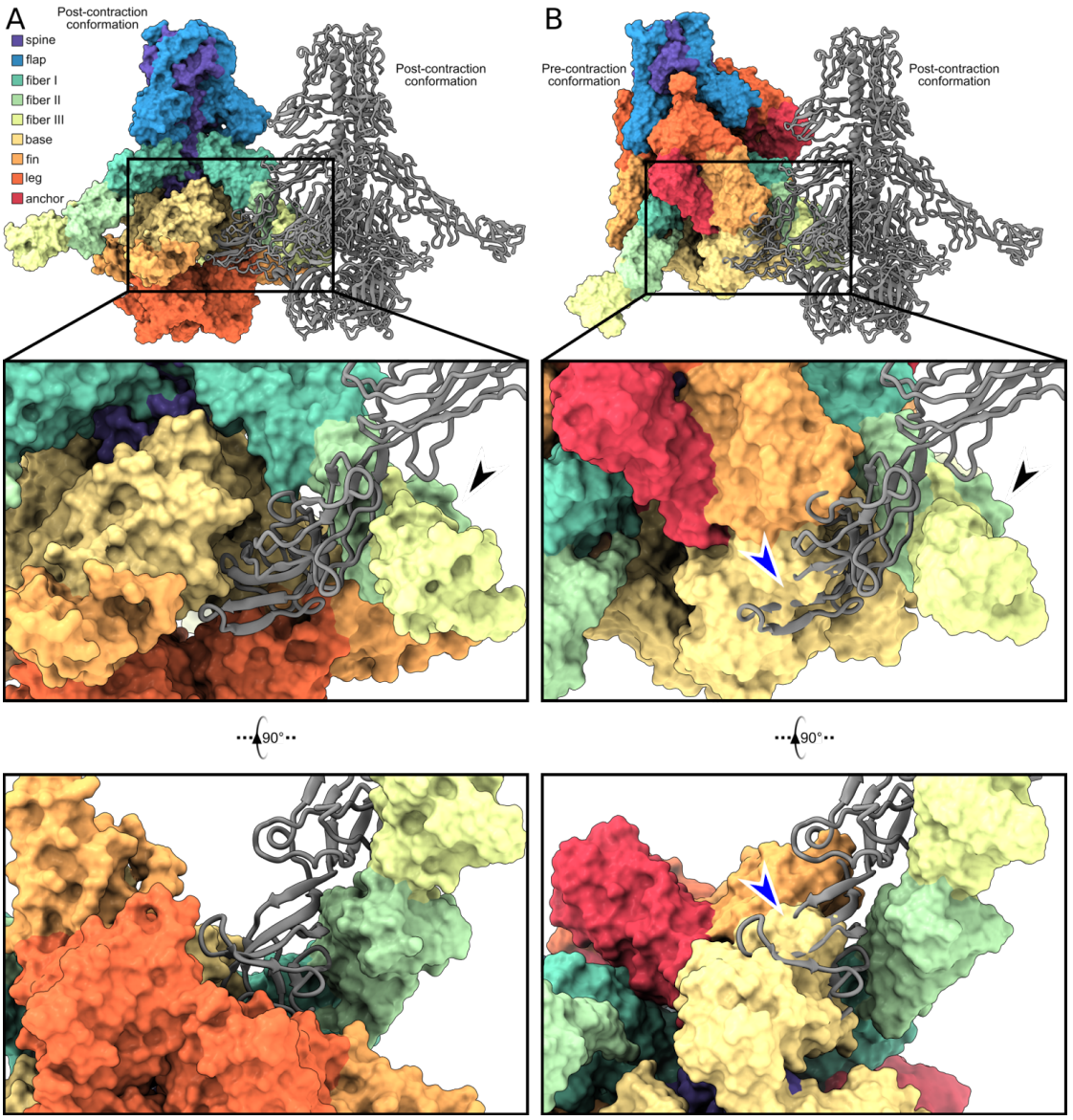

**SFig. 10. Mechanism of binding of inner tripod trimer to arm scaffold protein. (A)** Surface representation of cryo-EM map of the interface between the inner tripod complex and the arm scaffold protein from phi812 with a contracted tail. Tripod proteins and parts of the arm scaffold protein are colored according to the legend at the bottom. **(B)** The same as A, with the interaction partners shown in cartoon representation. **(C)** Top view of the interaction between the three tripod proteins and docking loop 1 of the arm scaffold protein. Side chains of the residues contributing to the hydrophobic interaction are shown as sticks. **(D)** Sequence alignment of docking loops 1 and 2 of the arm scaffold protein. Pink asterisks indicate residues of docking loop 1 participating in the interaction with the tripod proteins, as shown in panel C.

```
Tripod hook domain (613-670) ----TEQKTKKVTMKRFWRHNESYYGETEHYLYSNLQLDAG---QKWKGETYAYDR---EHGRTGTLLESG---------------
2mpw LTHNSYQNRLFYTLKT---------GETVADLSKSQDINL--------STIWSLNKHLYSSESEMMKAAPGQQIILPLKKL-----
2mtz -----GRKLLTYQVKQ---------GDTLNSIAADFRIST--------AALLQAN----PSLQAG--LTAGQSIVIPGLPD-PYTI
3zqd -----GRKLLTYQVKQ---------GDTLNSIAADFRIST--------AALLQAN----PSLQAG--LTAGQSIVIPGLPD-PYT-
4a1i -------GMLTYQVKQ---------GDTLNSIAADFRIST--------AALLQAN----PSLQAG--LTAGQSIVIPGLPD-PYTI
4b8v C---GSTSNIKYTVVK---------GDTLTSIAKKFKSGI--------CNIVSVN-----KLANPNLIELGATLIIPENCSNPDN-
7pvc -------ASQFYTVKS---------GDTLSAISKQVYGNAN-----LYNKIFEANK---PMLKSPDKIYPGQVLRIPEEL------
7vcm ------TASQFYTVKS---------GDTLSAISKQVYGNAN-----LYNKIFEANK---PMLKSPDKIYPGQVLRIP---------
7r1l ----------MYTVKP---------GDTMWKIAVKYQIGI--------SEIIAAN----PQIKNPNLIYPGQKINIP---------
4s3k -----PINGTYYTVKA---------GDTIWKVGRKLGVSY--------QAIANAN-----NVSVTAPLTPGRRILIPPSPN-----
4pxv -------CT-TYTIKS---------GDTCYAISQARGISL--------SDFESWN-----AGIDCNNLQIGQVVCV----------
5ylg -------CT-TYTIKS---------GDTCYAISQARGISL--------SDFESWN-----AGIDCNNLQIGQVVCVSK--------
5bum -------CTSYYTVKS---------GDICYNIAQTYGIDV--------ATLQSYN-----PGLQCDNLQIGQQLCVAD--------
4uz3 --------QATYTVAP---------GDTLYSIARRYGTTV--------EELMRLN------GLESFLLQPGQVLKLPS--------
5k2l -------MGCTYTIQP---------GDTFWAIAQRRGTTV--------DVIQSLN-----PGVNPARLQVGQVINVPC--------
5c8p -------STATVTVQQ---------GDTLRDIGRRFDCDF--------HEIARRN-----NIQNEDLIYPGQVLQVGG--------
5jcd ----QSDRLPIYVVQP---------QDGLDAIARNVFNAFV-----TYQEIAAAN-----NIPDPNKINVSQTLWIPL----P---
4eby ----CPLALASYYLEN---------GTTLSVINQNLNSSIAPYDQINFDPILRYN----SNIKDKDRIQMGSRVLVPF----PC--
7bax ----NSFSNISHMIKE---------GESYYYLSTTSYENLT-----NWETVQDSN-----PNYNPYLLPVGIKVVIPLFCK-----
```

**SFig. 11. Hook domain of tripod protein shares conserved residues with proteins from LysM family.** Primary sequence alignment of residues 613-670 forming the hook domain, with LysM protein family members (PDB codes are listed).

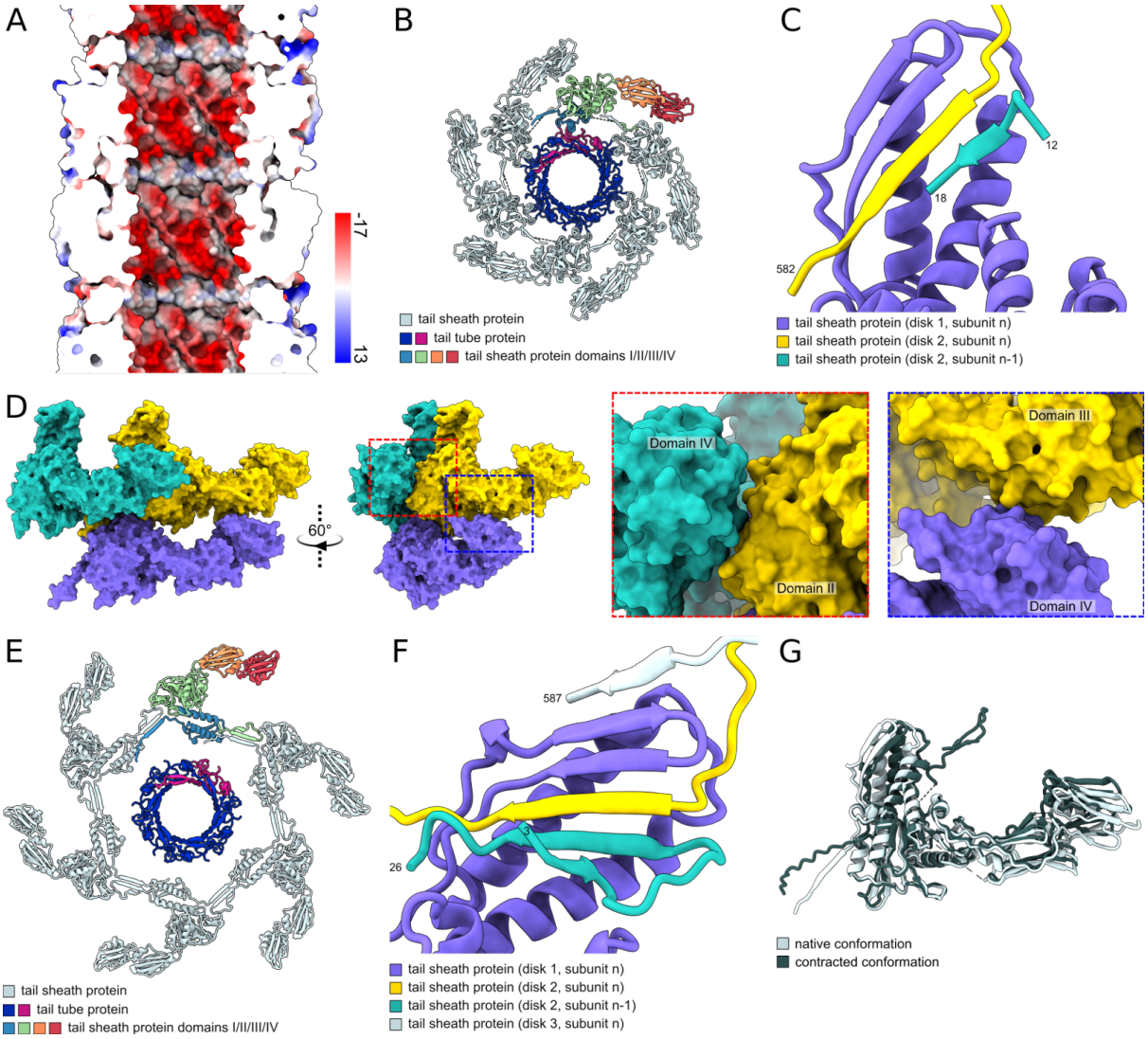



**SFig. 12. Particle of phage phi812 captured in intermediate state of tail contraction. (A)** Cryo-electron micrograph of phi812 attached to a cell wall of *S. aureus*. The black rectangle indicates the detail in B. **(B)** phi812 particle with a semi-contracted tail sheath. The blue arrowhead points to the sheath in the native conformation, whereas the red arrowhead points to the sheath in the contracted conformation. The black arrowhead points to the densities at the bottom of the contracted baseplate, presumably belonging to the anchor and hook domains of tripod proteins.

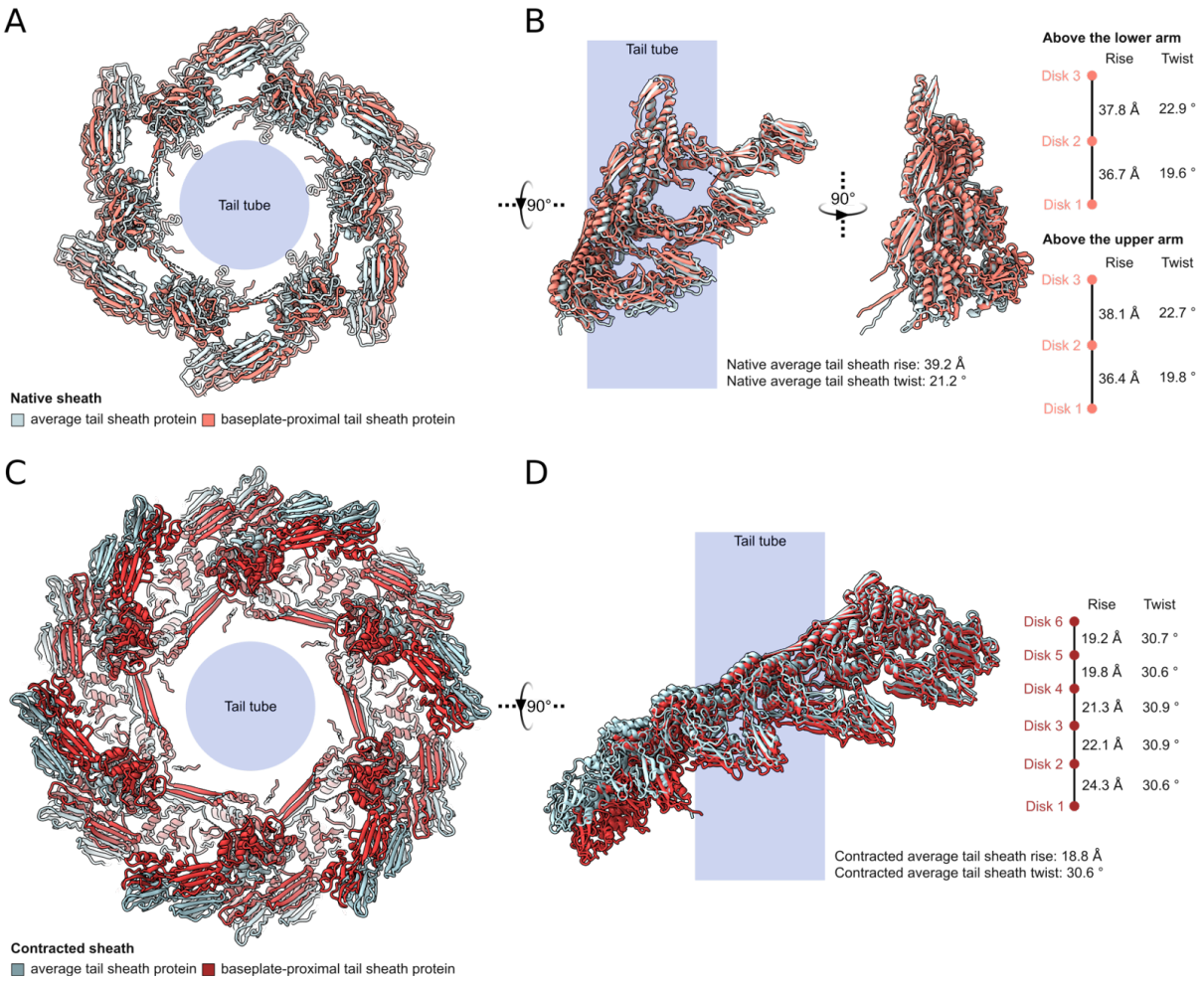



**SFig. 13. Interactions between tripod complexes in baseplate of phi812 particle with contracted tail. (A)** Interactions between two neighboring inner tripod complexes. One tripod complex is shown as a molecular surface colored according to domain, and the other is shown in cartoon representation in gray. The black rectangle indicates the magnified view shown in the insets below. The upper inset shows the interactions between fiber domains viewed perpendicular to the tail axis, and the lower inset shows the view along the tail axis toward the phage head. The black arrowhead points to the fiber domains of a tripod protein subunit used to superimpose the tripod complex in the pre-contraction state, as shown in B. **(B)** The inner tripod complex in the pre-contraction conformation is shown as a molecular surface colored according to domain, and the tripod complex in the post-contraction state is shown in gray cartoon representation. The pre-contraction tripod complex was superimposed onto the post-contraction tripod complex using the fiber domains indicated by the black arrowhead. The black rectangle indicates the close view shown in the insets below. The upper inset shows the view perpendicular to the tail axis, and the lower inset shows the view along the tail axis and towards the phage head. Blue arrowheads point to the site where the

fiber III domains of the post-contraction tripod complex clash with the base domain of the pre-contraction tripod complex.




**SFig. 14. Tail tube and tail sheath proteins. (A)** Molecular surface representation of the phi812 tail tube with the front half removed and colored according to Coulombic electrostatic potential in the range from -17 to 13 kcal/(mol·*e*-) at 298 K. **(B)** Cartoon representation of a disc of tail tube and tail sheath proteins from the native phi812 tail viewed along the tail axis towards the phage head. Protein subunits are colored according to the legend. **(C)** β-sheet augmentation of the handshake domain of the tail sheath protein in native conformation with the N-terminus and two C-termini from other tail sheath proteins. The number of terminal residues of the displayed structures are indicated. **(D)** Molecular surface representations of three tail sheath proteins in their native conformation. Red and blue dashed rectangles indicate views of interactions among the subunits shown on the right at higher magnification. **(E)** Cartoon representation of a disc of the tail tube and tail sheath proteins from the contracted tail viewed along the tail axis towards the phage head. The relative rotation of the tail tube disc against the sheath disc is arbitrary. The proteins are colored according to the legend. **(F)** β-sheet augmentation of the handshake domain of the tail sheath protein from the contracted tail. Terminal residues are indicated. **(G)** Comparison of the tail sheath proteins from native (light gray) and contracted (black) tails. The subunits were superimposed by aligning their domains II.




**SFig. 15. Baseplate-proximal discs of tail sheath proteins of phi812 in native and contracted states deviate from respective tail sheath structures at the center of the tail. (A)** Comparison of the cartoon representations of three baseplate-proximal tail sheath discs to three tail sheath discs from the central part of the tail. The structures are viewed along the tail axis towards the phage head. The position of the tail tube is indicated. Rise and twist differences were determined by superimposing domains I of tail sheath proteins from the uppermost discs (see Materials and Methods section for details). **(B)** Comparison of distances between tail sheath proteins from three baseplate-proximal tail sheath discs to three tail sheath discs from the central part of the tail. The twist and pitch parameters are indicated. **(C)** Comparison of the cartoon representations of six baseplate-proximal contracted tail sheath discs to six tail sheath discs from the central part of the tail. The structures are viewed along the tail axis towards the phage head. The position of the tail tube is indicated. **(D)** The scene from panel C viewed perpendicularly to the tail axis. Only six tail sheath proteins from each average and baseplate-proximal tail sheath protein discs are shown. The rise and twist difference between tail sheath subunits from successive discs is indicated on the right.

# Supplementary movies

**SMovie 1. Core proteins and wedge modules of the baseplate of phi812 in native conformation.** The baseplate core and wedge module proteins are shown in cartoon representation and colored according to Fig. 1E. The wedge module in the native conformation 1 is shown in lighter colors to emphasize two distinct conformations of the wedge modules in the native phi812 baseplate. The superposition of two neighboring wedge modules in native conformation 1 and native conformation 2 is shown. Protein names are indicated.

**SMovie 2. Baseplate and baseplate-proximal tail of phi812 in native conformation.** The baseplate and baseplate-proximal part of the tail of phi812 in native conformation are shown in cartoon representation and colored according to Fig. 1E.

**SMovie 3. Conformational changes of the baseplate arm of phi812.** The baseplate and baseplate-proximal part of the tail of phi812 in native and contracted conformations are shown in cartoon representation. Conformational changes to a single baseplate arm are shown.

**SMovie 4. Conformational changes of the tripod complexes.** The tripod protein complexes of phi812 in native and contracted conformations are shown in cartoon representations colored by subunits and by domain.

**SMovie 5. Conformational changes of the wedge modules, tail sheath initiator, and baseplate-proximal tail sheath proteins of phi812.** The baseplate and baseplate-proximal tail of phi812 in native and contracted conformations are shown in cartoon representation. Conformational changes of the wedge module, tail sheath initiator, and baseplate-proximal tail sheath proteins are shown.

# Supplementary Tables

**Table S1. Cryo-EM data and structure quality indicators**.

| | **Native tail** | **Native baseplate (core, wedge modules, arms, proximal tail)** | **Native baseplate (core, wedge modules, proximal tail)** | **Native baseplate (core, wedge modules)** | **Contracted baseplate (core, wedge modules, arms, proximal tail)** | **Contracted baseplate (core, wedge modules, arm segment, inner tripods, proximal tail)** | **Contracted baseplate (core, wedge modules, inner tripods)** | **Contracted tail tube** | **Contracted tail sheath** | **Central spike protein (knob and petal domains)** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Sample** | Native particles | Native particles | Native particles | Native particles | Contracted particles | Contracted particles | Contracted particles | Contracted particles | Contracted particles | Recombinant central spike protein |
| **Data collection and processing** | | | | | | | | | | |
| Detector | Falcon 2 | Falcon 2 | Falcon 2 | Falcon 2 | K2 Summit | K2 Summit | K2 Summit | K2 Summit | Falcon 2 | K2 Summit |
| Magnification | 75,000 | 75,000 | 75,000 | 75,000 | 130,000 | 130,000 | 130,000 | 130,000 | 75,000 | 130,000 |
| Voltage [kV] | 300 | 300 | 300 | 300 | 300 | 300 | 300 | 300 | 300 | 300 |
| Exposure [$e^-/Å^2$] | 48 | 48 | 48 | 48 | 42 | 42 | 42 | 42 | 48 | 42 |
| Pixel size [Å] | 1.061 | 1.061 | 1.061 | 1.061 | 1.057 | 1.057 | 1.057 | 1.057 | 1.061 | 1.04 |
| Symmetry | C6 | C3 | C3 | C3 | C6 | C6 | C6 | C6 | C6 | C3 |
| Initial number of particles | 43,718 | 23,421 | 23,421 | 23,421 | 54,841 | 54,841 | 54,841 | 25,203 | 106,091 | 226,791 |
| Final number | 18,188 | 9,709 | 9,709 | 9,709 | 25,203 | 25,203 | 25,203 | 12,906 | 69,969 | 135,736 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| of particles | | | | | | | | | | |
| Initial model | Previous reconstruction* | Previous reconstruction* | Previous reconstruction* | Reconstructed particles with subtracted signal | Previous reconstruction* | Previous reconstruction* | Previous reconstruction* | Reconstructed particles with subtracted signal | Previous reconstruction* | De novo |
| Map resolution [Å] | 4.2 | 7.3 | 4.5 | 4.4 | 4.4 | 4.0 | 3.1 | 3.6 | 3.0 | 3.2 |
| FSC threshold | 0.143 | 0.143 | 0.143 | 0.143 | 0.143 | 0.143 | 0.143 | 0.143 | 0.143 | 0.143 |
| **Accession codes** | | | | | | | | | | |
| EMDB | EMD-50093 | EMD-19968 | EMD-19969 | EMD-19970 | EMD-19971 | EMD-19972 | EMD-19973 | EMD-50094 | EMD-50095 | EMD-19974 |
| PDB | 9F04 | 9EUF | 9EUG | 9EUH | 9EUI | 9EUJ | 9EUK | 9F05 | 9F06 | 9EUL |
| **Model composition** | | | | | | | | | | |
| Atoms (except hydrogens) | 20,968 | 160,501 | 62,532 | 30,906 | 92,222 | 56,898 | 29,740 | 3,330 | 25,926 | 9,132 |
| Residues | 2,680 | 30,231 | 7,900 | 3,856 | 12,485 | 7,232 | 3,756 | 423 | 4,424 | 1,146 |
| B-factors | 89.62 | 48.40 | 69.46 | 49.03 | 64.71 | 71.54 | 49.55 | 54.95 | 98.49 | 69.58 |
| FSC model 0.5 (masked) | 4.19 | 10.38 | 4.61 | 6.21 | 6.65 | 4.14 | 3.25 | 3.86 | 3.05 | 3.27 |
| **RMSD** | | | | | | | | | | |
| Bond length [Å] | 0.005 | 0.012 | 0.007 | 0.003 | 0.008 | 0.006 | 0.005 | 0.001 | 0.002 | 0.002 |
| Bond angles [°] | 0.726 | 1.998 | 1.046 | 0.471 | 1.400 | 0.881 | 0.781 | 0.413 | 0.494 | 0.438 |

| **Validation** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| MolProbity score [percentile] | 0.94 | 1.51 | 1.35 | 1.18 | 1.04 | 1.16 | 0.94 | 1.08 | 1.03 | 1.08 |
| ClashScore [percentile] | 0.96 | 0.99 | 1.55 | 2.67 | 1.05 | 1.23 | 0.90 | 2.86 | 2.49 | 2.71 |
| Poor rotamers [%] | 0.18 | 1.45 | 1.20 | 0.32 | 0.44 | 0.43 | 0.12 | 0 | 0 | 0 |
| C-β outliers [%] | 0 | 0.15 | na | na | 0.04 | na | na | na | na | na |
| **Ramachandran plot** | | | | | | | | | | |
| Outliers [%] | 0 | 1.98 | 0.13 | 0.03 | 0.05 | 0.07 | 0 | 0 | 0 | 0 |
| Favored [%] | 97.13 | 88.30 | 94.27 | 97.35 | 96.42 | 95.40 | 97.07 | 98.80 | 98.17 | 97.89 |

na = not applicable

* = [23]

**Table S2. List of structural proteins from phi812 tail and baseplate.**

| ORF | protein | residues | domains (residue range) | representative homologs |
|---|---|---|---|---|
| 101 | Tail sheath | 587 | I (491-587) | Phage T4, gp18[4]; phage E217, gp31[10]; *V. cholerae* Type VI Secretion System, VipA/B[67] |
| | | | II (34-96,313-490) | |
| | | | III (97-151,251-312) | Prophage *L. innocua*, Lin1278, PDB: 3LML |
| | | | IV (152-250) | |
| 102 | Tail tube | 142 | β-barrel (1-34,59-142) | Phage T4, gp19[4]; *P. aeruginosa* R2 pyocin, PA0626[7], *C. difficile* diffocine, CD1364[46] |
| | | | stacking loop (35-58) | |
| 110 | Hub | 810 | tube-forming I (1-125) | Phage T4, gp27[35]; *E. coli* Type VI Secretion System, VgrG[39]; *P. aeruginosa* R2 pyocin, PA0628[7]; phage 80alpha, ORF59[37] |
| | | | tube-forming II (126-266) | |
| | | | tube-forming III (267-285,546-641) | |
| | | | tube-forming IV (286-334,527-545) | |
| | | | clamp (358-511) | *V. vulnificus*, PAS factor[45] |
| | | | cleaver (642-808) | Phage K, ORF56 (tail-associated lysin)[43]; phage K, LysK (endolysin)[42] |
| 111 | Weld | 295 | clip (1-128) | *R. typhi* Type IV Secretion System, VirB8[79]; *H. pylori* Type IV Secretion System, CagV[80] |
| | | | linker (129-142) | |
| | | | lid (143-295) | *B. cereus*, YkfC[81]; *M. tuberculosis*, RipD[47] |
| 112 | Central spike | 848 | oligosaccharide-binding (OB) (1-110) | Phage T4, gp5[35]; phage phi92, gp138 (spike)[34] |
| | | | β-prism (111-310) | |
| | | | coiled-coil (311-466) | Phage P22, gp26 (tail needle)[82] |
| | | | knob (467-610) | Phage Sf6, tail needle[83]; phage 1358, receptor-binding protein[29]; *B. cenocepacia,* BC2L-C (lectin)[31] |
| | | | petal (611-848) | *B. subtilis*, GlpQ (phosphodiesterase)[33] |
| 113 | Tail tube initiator | 263 | tube-like (1-156) | Phage T4, gp54/gp48[4]; *P. aeruginosa* R2 pyocin, PA0626 (ripcord)[7]; *C. difficile* diffocine, CD1367[46] |

| | | | | |
|---|---|---|---|---|
| | | | handle (157-263) | *P. aeruginosa* R2 pyocin, PA0626 (ripcord)[7]; phage E217, gp37/gp38[10]; phage Pam3, gp17[12] |
| 114 | Assembly | 173 | - | *P. aeruginosa* R2 pyocin, PA0626 (ripcord)[7]; phage Milano, BCP (gp25)[15] |
| 115 | Tail sheath initiator | 234 | LysM (1-66) | Phage T4, gp53[4]; *P. aeruginosa* R2 pyocin, PA0627 (glue)[7]; phage Milano, BCP (gp25)[15] |
| | | | linker (67-92) | - |
| | | | handshake (93-234) | Phage T4, gp25[4]; *P. aeruginosa* R2 pyocin, PA0617 (sheath initiator)[7]; phage E217, gp34[10]; phage Pam3, gp17[12] |
| 116 | Wedge A, B | 348 | helices (1-67) | Phage T4, gp6[4]; *E. coli* Type VI Secretion System, TssFa/b[6]; *C. difficile* diffocine, CD1364[46] |
| | | | wing (68-192) | *C. difficile* diffocine, CD1364[46]; phage E217, gp44[10]; phage Milano, BCP (gp25)[15] ; phage A1(L), gp31[14] |
| | | | trifurcation unit (193-256) | Phage T4, gp6[4]; *E. coli* Type VI Secretion System, TssFa/b[6]; *C. difficile* diffocine, CD1364[46] |
| | | | dimerization (257-348) | |
| 117 | Arm scaffold | 1019 | helices (1-94) | Phage T4, gp7[4]; *E. coli* Type VI Secretion System, TssG[6]; *P. aeruginosa* R2 pyocin, PA0619 (Tri2)[7]; phage E217, gp45[10]; phage Pam3, gp23[12] |
| | | | trifurcation (95-116,145-176) | |
| | | | docking loop 1 (117-144) | |
| | | | linker AB (177-212) | - |
| | | | bulge (213-219,259-443) | *A. thermocellus*, Xyn10B (carbohydrate-binding module 22)[84]; *P. barcinonensis*, Xyn10C (carbohydrate-binding module 22)[85] |
| | | | docking loop 2 (220-258) | |
| | | | linker BC (444-489) | - |
| | | | elbow (490-776) | Phage Tuc2009, BppA (putative receptor-binding protein)[28]; *P. barcinonensis*, Xyn10C (family 22 carbohydrate-binding module)[85] |
| | | | linker Z (777-907) | - |
| | | | connector (908-1019) | *B. ovatus*, putative alpha-L-fucosidase, (4ZRX); *C. polysaccharolyticus*, family 16 carbohydrate-binding module[60] |
| 118 | Arm segment | 173 | - | Phage T4, gp8[4]; A511, gp105[24] |
| 119 | Tripod | 1152 | spine helices (1-54,172-189,468-495) | - |
| | | | flap (55-115,139-171) | Phage 1358, receptor-binding protein (shoulder)[29]; *E. coli* Type VI Secretion System, TssK[6] |

| | | | | |
|---|---|---|---|---|
| | | | flap insertion loop (116-138) | |
| | | | fiber I (190-242,404-467) | |
| | | | fiber II (243-276,359-403) | *R. capsulatus* Gene Transfer Agent, adaptor[86]; phage YSD1, major tail protein (domain 2)[87]; phage CBA120, central spike (domain XD2/XD3)[88] |
| | | | fiber III (277-358) | |
| | | | base (496-538,1045-1152) | *C. polysaccharolyticus*, family 16 carbohydrate-binding module[60]; *T. maritima*, family 61 carbohydrate-binding module[59] |
| | | | fin (539-565,944-1044) | |
| | | | leg (566-590,793-943) | *C. difficile*, Cwp84 (lectin-like)[61] |
| | | | anchor (591-612,674-792) | SSV19, B120 (adaptor)[89]; *B. ovatus*, $BoSGBP_{MLG}$-B (Surface glycan-binding protein)[90] |
| | | | hook (613-673) | - |
| 121 | Receptor-binding protein 1 | 640 | antenna (1-37) | - |
| | | | tower I (38-121) | Phage T4, gp34 (long tail fiber)[91]; *P. aeruginosa* R1 pyocin, PA0620 (fiber)[92] |
| | | | tower II (122-184,397-410) | |
| | | | wing (185-396) | Phage CBA120, tail spike protein 4[93]; *T. polysaccharolyticum*, Family 16 Carbohydrate-binding Module[60] |
| | | | tower III (411-514) | Phage T4, gp34 (long tail fiber)[91]; *P. aeruginosa* R1 pyocin, PA0620 (fiber)[92] |
| | | | tip (515-640) | Phage T7, gp17 (fiber)[94] |
| 123 | Receptor-binding protein 2 | 458 | antenna (1-60) | - |
| | | | tower I (61-149) | Phage T4, gp34 (long tail fiber)[91]; *P. aeruginosa* R1 pyocin, PA0620 (fiber)[92] |
| | | | tower II (150-228) | |
| | | | tower III (229-327) | |
| | | | tip (328-458) | T7, gp17 (fiber)[94] |

801

802 **Table S3. NMR data and structure quality indicators for cleaver domain of hub protein.**

| **Table** Structural statistics of 20 models | **CHAP** |
|---|---|
| **NMR distance and dihedral restraints** | |
| Distance restraints | |
| Total NOE | 1632 |
| Intra-residue (*i*=*j*) | 460 |
| Sequential (\|*i* – *j*\| = 1) | 439 |
| Medium-range (1 < \|*i* – *j*\| < 5) | 223 |
| Long-range (\|*i* – *j*\| > = 5) | 510 |
| Total dihedral angle restraints | 195 |
| $\varphi$ | 95 |
| $\psi$ | 100 |
| Violations (mean ± s.d.) | |
| Distance restraints (Å) | 0.020 ± 0.002 |
| Dihedral angle restraints (º) | 0.808 ± 0.085 |
| Max. dihedral angle violation (º) | 5.72 |
| Max. distance constraint violation (Å) | 0.946 |
| Deviations from idealized geometry | |
| Bond lengths (Å) | 0.011± 0.000 |
| Bond angles (º) | 1.12± 0.032 |
| Impropers (º) | 1.88 ± 0.071 |
| Average pairwise r.m.s. deviation* (Å) | |
| Backbone | 0.99 ± 0.26 |
| Heavy | 1.36 ± 0.23 |
| Ramachandran plot statistics* (%) | |
| Residues in most favoured regions | 96.9 |
| Residues in additionally allowed regions | 2.7 |
| Residues in generously allowed regions | 0.1 |
| Residues in disallowed regions | 0.3 |

803 * Ordered residues: 25-42, 51-100, 103-129, 135-159

**Table S4. X-ray data and structure quality indicators.**

| **Data collection** | Arm segment protein |
|---|---|
| Space group | P2(1) |
| Cell axes a, b, c [Å] | 92.93, 75.14, 94.45 |
| Cell angles α, β, γ [°] | 90.00, 109.87, 90.00 |
| Beamline | Soleil, Proxima 2A |
| Resolution range [Å] | 46.49-1.80 |
| Reflections | 109 411 |
| Completeness [%] | 96.07 |
| Mean <I/s(I)> | 1.13 |
| $R_{merge}$ [%] | 0.146 |
| Wilson B [$Å^2$] | 38.12 |
| **Refinement** | |
| Reflections $R_{free}$ | 5471 |
| *R*-factor | 0.245 |
| *R*-free | 0.275 |
| Overall B-factor [$Å^2$] | 43.51 |
| Ramachandran stats [%] | 98.16 |
| rmsd bonds (Å)/angles [°] | 0.011; 1.146 |
| PDB code | 9FKO |

# Materials and Methods

**Bacteriophage propagation and purification**

Phage phi812, strain K1/420 (GenBank Accession Number KJ206563.2), was propagated in *S. aureus* SA 812 planktonic culture and purified on a CsCl gradient as described before[23], then dialyzed against phage buffer (50 mM Tris pH 8, 10 mM NaCl, 10 mM $CaCl_2$).

**Induction of phi812 tail contraction and genome release *in vitro***

To induce contraction of the phi812 tail, a previously reported protocol[95]o was modified. The purified phage sample ($10^{12}$ PFU/ml) in phage buffer was supplemented with lipoteichoic acid from *Staphylococcus aureus* (SIGMA) at a concentration of 100 µg/ml, then diluted 10x using a solution of 3M urea in the phage buffer and incubated at 42°C for 45 min. After incubation, the sample was diluted 3x in the phage buffer and treated with Turbonuclease (Abnova) at a final concentration of 3.5 U/ml for 15 min at room temperature. Phage particles were pelleted by centrifugation at 75,000 x *g* at 4°C for 1 h using a 50.2 Ti rotor (Beckman Coulter) and resuspended in the phage buffer.

**Identification of structural proteins of phage baseplate**

The purified phi812 phage was resuspended in Laemmli buffer and boiled for 3 min. Subsequently, the sample was treated with Turbonuclease (Abnova) at a final concentration of 2 U/µl for 15 min at room temperature. Proteins from the phage particles were separated using tricine gradient gel electrophoresis. Major protein bands were cut from the gel and used for mass spectrometry analysis. Mass spectrometry data processing and analyses were performed using FlexAnalysis 3.4 and MS BioTools (Bruker Daltonics). Mascot software (Matrix Science, London, UK) was used for sequence searches in exported MS/MS spectra against the National Center for Biotechnology Information database and a local database supplied with the expected protein sequences. The mass tolerance of peptides and MS/MS fragments for MS/MS ion searches were 50 parts per million and 0.5 Da, respectively. The oxidation of methionine and propionyl-amidation of cysteine as optional modifications and one enzyme mis-cleavage were set for all searches. Peptides with a statistically significant peptide score ($P < 0.05$) were considered.

**Cryo-electron microscopy sample preparation, data collection, motion correction, and CTF estimation**

Purified phi812 ($10^{11}$ PFU/ml) in phage buffer (4 µl), either in the native or contracted state, was pipetted onto a holey carbon-coated copper grid (R2/1, mesh 200; Quantifoil), blotted and vitrified by plunging into liquid ethane using a Vitrobot Mark IV (FEI). The vitrified sample was transferred to a Titan Krios electron microscope (Thermo Fisher) operated under cryogenic conditions and at an acceleration voltage of 300 kV. The beam was aligned for parallel illumination in NanoProbe mode, and coma-free alignments were performed to remove the residual beam tilt.

Micrographs of phi812 virions were collected at a magnification of 75,000 x using a Falcon 2 direct electron detector (FEI) operating in linear mode, resulting in a calibrated pixel size of 1.061 Å/pix. Imaging was done under low-dose conditions (total dose 48 e-/Å2) and with defocus values ranging from −0.8 to −3.0 µm. A one-second exposure was fractionated into 18 frames and saved as a movie.

Micrographs of phi812 in a contracted state were collected at a magnification of 130,000 x using a K2 Summit direct electron detector (Gatan) operating in counting mode, resulting in a calibrated pixel size of 1.057 Å/pix. Imaging was done under low-dose conditions (total dose 42 e-/Å2) and with defocus values ranging from −1.0 to −2.4 µm. A seven-second exposure was fractionated into 40 frames and saved as a movie.

Automated data acquisitions were performed using the software EPU (Falcon 2) or Serial EM (K2 Summit). The movies were corrected for global and local motion during acquisition using MotionCor2[96], and saved as dose-weighted micrographs. Defocus values were estimated from aligned non-dose-weighted micrographs using the program Gctf[97].

**Reconstruction of native baseplate of bacteriophage phi812**

A total of 7,187 micrographs of phi812 virions were used to pick 23,421 particles using the e2boxer program from the EMAN2 suite[98]. Particles were then extracted and binned 4x to a 210 px box size (4.244 Å/px) using Relion[99]. The particles were subjected to several rounds of 2D and 3D classification with three-fold symmetry imposed, using the previously reconstructed native baseplate (EMD-8210) as an initial model lowpass-filtered to 60 Å.

The resulting 9,705 particles were re-extracted using an 840px box (1.061 Å/px) and auto-refined with C3 symmetry imposed and masking using Relion[100]. Next, several rounds of CTF refinements were performed, including an anisotropic magnification estimation and fitting of the defocus and astigmatism parameters per particle and micrograph, respectively. The resulting map of the baseplate reached a resolution of 7.3 Å after the post-processing, with a calibrated pixel size of 1.075 Å/px applied. Then, alternating rounds of 3D auto-refinement and CTF refinement were performed using a mask including the first three tail sheath rings, core, and wedge proteins, yielding a map with a resolution of 4.5 Å and a calibrated pixel size of 1.075 Å/px.

To improve the reconstruction of the core and wedge proteins, particles were re-extracted into a 300 px box (1.061 Å/px), and the reconstructed map was segmented to retain only the density corresponding to the core and wedge proteins. The sub-particles were then re-extracted into the same box with the offsets centered to x, y, and z coordinates equal to 0,0,-37 to account for the shift of the calculated center of mass along the Z axis. Re-centered particles were reconstructed, and an inverted mask was used to subtract the signal around the reconstructed map. After a 3D auto-refinement, additional 3D classification or CTF refinement did not improve the resolution. The final map was post-processed with a calibrated pixel size of 1.075 Å/px applied, reaching a resolution of 4.4 Å. In all the above cases, the resolution after post-processing was calculated according to the gold standard FSC = 0.143 criterion. The local resolution for each map was computed using Relion[101].

**Reconstruction of phage phi812 with contracted tail**

Out of 15,371 micrographs of phages in the post-contraction state, the first 1,000 micrographs were used to manually pick 7,539 particles using crYOLO[102]. Particles were then extracted and binned 4x to a 208 px box size (4.228 Å/px) using Relion[100]. After a single round of 2D classification, all aligned particles were re-extracted unbinned to a 832 px box size (1.057 Å/px) with re-centering of the offsets to x,y,z coordinates equal to 0,0,0. Then, particles were imported into Topaz[103] with a downsampling factor of 8. After training the neural network, the positions of particles were predicted using the radius r = 36 and picking threshold t = 0. The inspection of one hundred randomly selected micrographs with the respective particle coordinates in the crYOLO program revealed a negligible number of false negatives and false positives. After the picking coordinates were corrected for the sampling factor, 54,841 particles were extracted using a 960px box size and binned 4x (4.228 Å/px) using Relion[101]. After several rounds of 2D classification, 31,308 particles were subject to 3D reconstruction and classification with imposed C6 symmetry, using the previously reconstructed post-contraction baseplate (EMD-8212) lowpass-filtered to 60 Å as an initial model. The resulting 25,487 particles were re-extracted and binned 2x to obtain a 480 px box size (2.114 Å/px). After a 3D classification, one of the three classes exhibited the wrong handedness. Particles from this class were corrected using the xflip_particles_star.py script (github.com/fuzikt/starpy). After another round of 3D classification, 25,203 particles were used for the reconstruction, yielding a final map that reached the Nyquist resolution. Next, particles were re-extracted using a 960px box size (1.057 Å/px) and subjected to 3D auto-refinement using a whole-map mask. The final map reached a resolution of 4.4 Å after the post-processing. Next, 3D auto-refinement was performed using a focused mask that included the core and wedge proteins and an inner ring of tripod proteins. The resulting map reached a resolution of 4.11 Å after the post-processing. The particles were assigned to their respective optics groups, and several rounds of CTF refinements (anisotropic magnification estimation, beam tilt, trefoil, and tetrafoil aberrations estimation, and fitting of the defocus and astigmatism parameters per particle) improved the resolution of the map to 3.6 Å. Then, two rounds of Bayesian polishing and subsequent 3D auto-refinements and CTF refinements were performed. In the final post-processing step, a negative B-factor of -25 was applied to produce the final map with an average resolution of 3.1 Å. The final post-processing step was also performed using a mask encompassing the first six tail sheath discs, core and wedge proteins, inner ring of tripod proteins, and a subunit of the arm segment protein, resulting in a map with a resolution of 4 Å. The resolution was calculated according to the gold standard FSC = 0.143 criterion in all cases. The local resolution for each map was computed using Relion[101].

**Reconstruction of native phi812 tail**

The phi812 tails were manually boxed using the program e2helixboxer.py from the EMAN2 suite[98], and the helical segments were extracted in a 360 px box using Relion[100]. Tail edges and bent and damaged particles were removed from the dataset by multiple rounds of reference-free 2D classification. A map of the phi812 tail in its native conformation (EMD-4051) low-pass filtered to 50 Å was used as an initial model for the 3D auto-refinement of 18,188 particles using a helical refinement protocol with C6 symmetry imposed. The helical refinement and the per-particle CTF refinement were iterated until convergence. The final map reached a resolution of 4.2 Å. The

resolution after post-processing was calculated according to the gold standard FSC = 0.143 criterion. Local resolution maps were computed using Relion[101].

**Reconstruction of contracted tail**

A total of 106,091 particles were manually picked from 4,833 micrographs using the program e2helixboxer.py from the EMAN2 suite[98], and the helical segments were extracted using a 400 px box size (1.061 Å/px) in Relion[100]. Tail edges and bent and damaged particles were removed from the dataset by multiple rounds of reference-free 2D classification. Then, 85,549 particles were subjected to several rounds of 3D reconstruction and classification with C6 symmetry imposed. A map of the contracted tail determined previously (EMD-4052) filtered to 50 Å was used as an initial model for the 3D auto-refinement of 69,969 particles using a helical refinement protocol with C6 symmetry imposed. The final map reached an average resolution of 3.0 Å.

**Localized reconstruction of tail tube from contracted baseplate**

Information about particle positions and orientations from the final reconstruction of the baseplate of phages with contracted tails was used to extract sub-particle images of the tail tube with a modified localized reconstruction protocol[104] (github.com/fuzikt/localrec)). The initial set of 25,203 sub-particles (box size 256 px, 1.057 Å/pixel) centered on the contact region between the tail tube and wedge A proteins was subjected to several rounds of 2D classifications in Relion[101]. After the initial 3D auto-refinement with imposed C6 symmetry, a single 3D classification was performed. The selected 12,906 particles belonging to the best class were further auto-refined with imposed C6 symmetry and maximum allowed deviations from previous orientations of 10°. The final map reached an average resolution of 3.6 Å.

**Production of recombinant cleaver domain of hub protein**

The DNA segment encoding the cleaver domain of hub protein (residues 650-808) was amplified from the genomic DNA of phage phi812 by PCR using forward and reverse primers (5'GTAGCCATGGAACAAAGTAGTGGA3' and 5'TAGGATCCTTTATTTCTTATCGTAAA3'). Using restriction cloning, the amplicon was cloned into the vector pET28a(+) (Novagen/Merck), providing a T7 tag and a His6 tag at the C-terminus of the protein. An *E. coli* expression strain BL21(DE3) was transformed with the construct using a standard heat shock procedure (45 s at 42 °C). An overnight culture was used to inoculate 1 l of 2% Luria-Bertani broth (Oxoid) supplemented with kanamycin (50 µg/mL) and cultivated at 37 °C until $OD_{600}$=0.5. The protein expression was induced with 0.4 mM IPTG, and was allowed to progress for 4-6 h at 30°C.
For nuclear magnetic resonance spectroscopy (NMR), the cleaver domain was produced in M9 minimal medium supplemented with traces of metals, vitamins, and 2 g/L [$^{13}$C]glucose or 0.5 g/L $^{15}NH_4Cl$ (Cambridge Isotope Laboratories). *E. coli* BL21 (DE3) was grown to $OD_{600}$ = 0.5, and the gene expression was induced with 0.4 mM IPTG and continued at 30 ºC for 12 h. The cultures were harvested, resuspended in the sonication buffer (100 mM Tris-Cl, pH 8, 500 mM NaCl, 4 mM DTT, 10 mM imidazole, 0.1 % TritonX-100), and subjected to sonication (Sartorius Labsonic® M), followed by centrifugation at 14,000 x *g* and 4 °C for 45 min. The supernatant was filtered (0.45 µm cut-off,

Corning Inc.) and loaded into a 5 ml Histrap column (GE Healthcare) pre-equilibrated with the loading buffer and connected to an AKTA FPLC system (GE Healthcare). After the column was washed with 15 column volumes of the loading buffer (50 mM Tris-Cl, 300 mM NaCl, 20 mM imidazole), gradient elution was performed using the elution buffer (50 mM Tris-Cl, 300 mM NaCl, 500 mM imidazole). Fractions with protein were pooled, concentrated, and loaded onto a HiLoad 16/600 Superdex 75 pg size-exclusion chromatography column (GE Healthcare Life Sciences), followed by an isocratic elution with 100 mM Na-phosphate, pH 6.8, 150 mM NaCl. Purified protein was pooled, and the buffer was exchanged by performing two washes with 20 mM Na-phosphate buffer, pH 6.8, using an Amicon Ultra 15 filter (10 kDa cut-off, Millipore). While exchanging the buffer, the sample was concentrated to 20 mg/ml. For the NMR studies, the protein was concentrated at 35 mg/ml while buffer-exchanged to 25 mM Na-phosphate, pH 6.6, and 100 mM NaCl.

**Zymogram analysis**

The cleaver domain of the hub protein after Histrap purification was separated on conventional 12.5% polyacrylamide gel using SDS-PAGE. Polyacrylamide gels (12,5 %) for zymograms contained 0.5 % bacterial cell walls, prepared from *S. aureus* CCM 885 cells, autoclaved at 121 ºC for 5 min. Samples of the cleaver domain were mixed with loading buffer (1% SDS, 6% sucrose, 100 mM dithiothreitol, 10 mM Tris, pH 6.8, 0.0625% bromophenol blue) and boiled for 5 min before loading onto SDS-PAGE gels and zymograms. The SDS-gel and zymogram were separated using gel electrophoresis. After the gel electrophoresis, the zymogram was washed for 30 min with distilled water to remove SDS. To renature the cleaver domain, the zymogram was soaked at room temperature in distilled water for 2-4 h, and then for 1 h in the buffer containing 150 mM Na-phosphate, pH 7.0, 0.1% Triton X-100 and 10 mM $MgCl_2$. The zymogram was stained for 3 h with 0.1% methylene blue in 0.001% KOH and washed with water. The peptidoglycan-hydrolytic activity of the cleaver domain of the hub protein was detected as a clear zone on a dark blue background.

**NMR structure determination of cleaver domain of hub protein**

NMR spectra were recorded at 298 K in an 850 MHz Bruker Avance III spectrometer equipped with a $^{1}H/^{13}C/^{15}N$ TCI cryogenic probe head with z-axis gradients. The composition of the NMR sample was 1.8 mM uniformly $^{13}C$-, $^{15}N$-labeled cleaver domain in a buffer containing 25 mM sodium phosphate, pH 6.6, 100 mM NaCl, 1.5 mM DTT, and 10% $D_2O$ for the lock. Three sparsely sampled 4D NMR experiments were acquired: 4D HC(CC-TOCSY(CO))NH, 4D $^{13}C,^{15}N$ edited HMQC-NOESY-HSQC (HCNH), and 4D $^{13}C,^{13}C$ edited HMQC-NOESY-HSQC (HCCH). Sequential and aliphatic side chain assignments were obtained automatically using the 4D-CHAINS algorithm[105] and checked manually. Backbone dihedral angle restraints were derived from TALOS[106] using a combination of five kinds ($H^N$, $H^\alpha$, $C^\alpha$, $C^\beta$, N) of chemical shift assignments for each residue in the sequence. NOE cross-peaks from the two 4D NOESY spectra (HCNH and HCCH) were assigned automatically by CYANA 3.0[107] in structure calculations with torsion angle dynamics. Unambiguous distance and torsion angle restraints (Table S3) were used in a water refinement calculation[108] applying the RECOORD protocol[109]. The quality of the NMR-derived structure ensembles was validated using PSVS[110].

1008 **Production of recombinant central spike protein**

1009 The gene for central spike protein was amplified from the genomic DNA of phage phi812 using
1010 forward and reverse primers (5'AGATTGGTGGCTCGGATGATTTAAATGTAAAAGGATTAGTTTTA3' and
1011 5'GAGGAGAGTTAAGACTTATTAAATCTTTTTATATTCATCTAGGTAGTTTGTAAA3'). Compared to the
1012 full-length gene, the first 17 nucleotide triplets from the 5' end of the amplified gene were omitted,
1013 as they code for the first 17 residues of the central spike protein, which were predicted to be
1014 disordered by the secondary structure prediction using Phyre[111]. The truncated gene was cloned into
1015 a vector derived from pET22b, which was modified to provide a His6-SUMO-tag fusion followed by
1016 the Ulp1 cleavage site at the N-terminus of the protein. The *E. coli* expression strain BL21(DE3) was
1017 transformed with the construct using a standard heat shock procedure (45 s at 42°C). An overnight
1018 culture was used to inoculate 2% Luria-Bertani broth (Sigma) supplemented with ampicillin (50
1019 µg/ml). Cells were cultivated at 37°C and 250 rpm until $OD_{600}$=0.5, followed by cultivation at 16 °C
1020 and 250 rpm for 24 h without the induction of expression, as soluble central spike protein was only
1021 produced after non-induced leaky expression. Cells were harvested in a typical yield of 8 g per liter of
1022 cell culture, resuspended in 2.5 ml of the loading buffer (50 mM Na/K phosphate, pH 7, 200 mM
1023 NaCl) per g of cell pellet, and subjected to chemical disintegration (0.5 % octylthioglucoside, 1 mg/ml
1024 lysozyme, 2 U/ml Turbonuclease, 1 mM PMSF, 5 mM TCEP) at 6 °C overnight. The lysate was
1025 centrifuged at 20,000 x *g* and 8°C for 30 min. The resulting supernatant was filtered (0.45 µm cut-off,
1026 Corning Inc.) and loaded into a 5 ml Histrap column (GE Healthcare) pre-equilibrated with the loading
1027 buffer and connected to an AKTA FPLC system (GE Healthcare). After the column was washed with 15
1028 column volumes of the loading buffer, gradient elution was performed using the elution buffer (50
1029 mM Na/K phosphate, pH 7, 200 mM NaCl, 500 mM imidazole). Eluted fractions containing the central
1030 spike protein, as determined by SDS-PAGE analysis using 12.5 % polyacrylamide gels, were pooled
1031 and supplemented with 5 mM TCEP before the cleavage of the purification tag overnight at 6°C using
1032 the in-house prepared Ulp1 protease. The cleavage solution was desalted using a HiPrep™ 26/10
1033 Desalting column (Cytiva) before reverse affinity chromatography using the 5 ml Histrap column (GE
1034 Healthcare). Cleaved protein collected in the flow-through fraction was concentrated (30 kDa cut-off;
1035 Amicon ® Ultra; Merc Millipore) and loaded into a Superdex 16/600 200 pg column (GE Healthcare)
1036 pre-equilibrated with the loading buffer (50 mM Na/K phosphate, pH 7, 200 mM NaCl), followed by
1037 an isocratic elution. Monodisperse fractions were analyzed by SDS-PAGE and concentrated to a final
1038 concentration of 5 mg/ml.
1039

1040 **Cryo-EM reconstruction of recombinant central spike protein**

1041 The purified central spike protein sample was diluted in 50 mM Na/K phosphate, pH 7, 200 mM NaCl
1042 buffer to a concentration of 0.8 mg/ml, and then diluted 10x in water. Four µl of the central spike
1043 protein at a concentration of 0.08 mg/ml were applied onto a Quantifoil R1.2/1.3 200-mesh Cu grid
1044 glow-discharged for 60 s from both sides. The excess of the sample was blotted for 1 s with filter
1045 paper (Whatman No. 1) using blot force -2, and immediately plunge-frozen in liquid ethane using a
1046 Vitrobot Mark IV (FEI). The instrument chamber was kept equilibrated to 100 % humidity and at 4°C.
1047 Cryo-EM data were collected using a Titan Krios TEM (Thermo Fisher Scientific) operated at 300 keV
1048 and equipped with a K2 Summit direct electron detector (Gatan). The beam was aligned for parallel
1049 illumination in NanoProbe mode, and the residual beam tilt was removed by coma-free alignments.
1050 The software SerialEM was used to acquire the data in the electron counting mode at 130,000 x

magnification, corresponding to a calibrated pixel size of 1.04 Å/px over a defocus range of -0.3 to -3.0 µm. The exposure of 8 s with a dose rate of 6.6 $e^-$/px/s was fractionated into 40 frames (total dose 57 $e^-/Å^2$, 1.425 $e^-/Å^2$ per frame). A total of 3120 micrographs were collected. For each movie, the drift and gain correction was performed with MotionCor2[96], while the contrast-transfer function was estimated using the program Gctf[97]. crYOLO[102] was used to pick particles from micrographs denoised using the program janni. Particles picked manually from 10 micrographs were used to train the neural network. The resulting 487,160 auto-picked particles were extracted and binned to a 128 px box size using Relion[100]. Several rounds of 2D classification sorted out classes of the central spike protein's C-terminal domains containing 226,791 particles. These particles were re-extracted and binned to a 128 px box size (2.08 Å/px) using Relion 5.0 beta (github.com/3dem/relion/tree/ver5.0). Three more rounds of 2D classification resulted in 135,736 particles, which were used for 3D reconstruction using the initial model generated by the stochastic gradient descent algorithm within Relion. Subsequent 3D classifications did not improve the quality or resolution of the reconstruction. Therefore, all 135,736 particles were re-extracted using a 256 px box size (1.04 Å/px) and assigned to their respective optics groups. Several rounds of CTF refinements, including anisotropic magnification estimation, beam tilt and trefoil aberrations estimation, and fitting of the defocus and astigmatism parameters per micrograph, improved the resolution of the post-processed map to 3.65 Å. Then, two rounds of Bayesian polishing and subsequent 3D auto-refinements and CTF refinements were performed. The final map reached a resolution of 3.23 Å. The resolution after post-processing was calculated according to the gold standard FSC = 0.143 criterion. Local resolution maps were computed using Relion[101].

**Production of recombinant Arm segment protein**

The gene encoding the arm segment protein was cloned into the pET-28b(+) vector (Novagen/Merck), providing a His6 tag at the C-terminus of the protein. An *E. coli* expression strain BL21(DE3) was transformed with the construct using a standard heat shock procedure (45 s at 42°C). An overnight culture was used to inoculate 2% Luria-Bertani broth (Sigma) supplemented with 1 mM MgCl2, 1 % glucose, and ampicillin (50 µg/ml), followed by cultivation at 37 °C and 250 rpm until $OD_{600}$=0.5. The expression was induced with 0.4 mM IPTG at 18°C, and was allowed to progress for 17 h. Cells were harvested and lysed using the lysis buffer (0.5 % octyl-thioglucoside, 1 mg/ml lysozyme, 2 U/ml Turbonuclease, 1 mM PMSF, 5 mM TCEP) at room temperature for 30 minutes, and subsequently disintegrated with an Emulsiflex C3 (Avestin). After centrifugation at 20,000 x *g* and 8°C for 30 min, the supernatant was filtered (0.45 µm cut-off, Corning Inc.) and loaded into a 5 ml Histrap column (GE Healthcare) pre-equilibrated with the loading buffer (50 mM HEPES, pH 8, 200 mM NaCl, 40 mM imidazole) connected to an AKTA Pure FPLC system (GE Healthcare). The protein was eluted in one step using the elution buffer (50 mM HEPES, pH 8, 200 mM NaCl, 500 mM imidazole). Eluted fractions containing the arm segment protein, detected by SDS-PAGE analysis on 12 % polyacrylamide gels, were loaded into a Superdex 16/600 75 pg column (GE Healthcare) pre-equilibrated with the loading buffer (50 mM HEPES, pH 8, 200 mM NaCl), followed by an isocratic elution. Monodisperse fractions were analyzed by SDS-PAGE and concentrated to a final concentration of 14 mg/ml.

1093 **Crystallization and X-ray structure determination of recombinant arm segment protein**

1094 The arm segment protein was crystallized by vapor diffusion using the sitting-drop method at 21 °C.
1095 The protein sample at a concentration of 14 mg/ml was used to prepare drops with protein :
1096 precipitant ratios of 200 nl : 100 nl, 100 nl : 100 nl, and 100 nl : 200 nl in MRC 3-drop 96-well plates
1097 using a Dragonfly robot (SPT Labtech) with a range of commercially available screens (Molecular
1098 Dimensions, Hampton Research, Qiagen-Nextal). Thick needle-like crystals were observed in the
1099 drops containing the solution of 1.6 M LiCl, 0.1 M Tris, pH 7.5, 25 % w/v PEG 4,000. The same
1100 crystallization solution was used to prepare hanging drops for vapor diffusion. Crystals were
1101 harvested from the hanging drops and flash-vitrified in liquid nitrogen.
1102 Diffraction data were recorded at 100 K on the Proxima-2A beamline at the Soleil
1103 synchrotron (Paris, France). Native crystals were soaked in I3C using the I3C phasing kit (Hampton
1104 Research), and the dataset was phased using the multiple wavelength anomalous dispersion method.
1105 The diffraction data from the crystals soaked with I3C were collected at 12.67 ke-V and 5.00 ke-V. The
1106 structure of the arm segment protein was determined using the program SHELXD[112]. The native
1107 diffraction data recorded from the crystals of arm segment protein were phased using the program
1108 Phaser[113], and the structure was determined using the anomalous data as a molecular replacement
1109 model. Refinement was performed manually using the program *Coot*[114] and automatically in
1110 *REFMAC5*[115]. The structures were validated using MOLPROBITY.
1111
1112 **Structure determination and refinement using cryo-EM maps**

1113 The structures of phage proteins were predicted using the programs HHpred[116], RaptorX[117], or
1114 Alphafold2[52], and fitted into their densities as rigid bodies or by individual domains using the
1115 programs Chimera[118] or ChimeraX[119], and then refined interactively using ISOLDE[120]. The structures
1116 were improved using the real-space refinement implemented in the *Phenix* package[121], and corrected
1117 manually using the programs *Coot* and ISOLDE. The geometry evaluation of models was performed
1118 using *MolProbity*[122].
1119
1120 **Visualization and structure analysis**

1121 Chimera[118] and ChimeraX[119] were used to visualize maps and structures. PISA web server[123] was used
1122 to calculate interface areas. Sequences were analyzed using Clustal Omega[124], available via the
1123 EMBL-EBI server[125] and visualized with Jalview[126].
1124
1125 **Cryo-EM of phi812 attached to *S. aureus* cell wall fragments**

1126 *S. aureus* SA 812 was cultured in 100 ml of meat peptone broth (1 g of Lab-Lemco (Difco), 5 g of yeast
1127 extract powder L21 (Oxoid), 10 g of peptone L37 (Oxoid), and 5 g of NaCl were dissolved in distilled
1128 water to a final volume of 1000 ml, and the pH was adjusted to 7.4 using 10 M NaOH) at 37°C until
1129 $OD_{600}$=0.3. Cells were harvested by centrifugation at 3,000 x *g* for 15 min. The resulting pellet was
1130 resuspended in 80% isopropanol to a final volume of 10 ml, and incubated for 1 h at room
1131 temperature. The suspension was centrifuged at 1,000 x *g* at room temperature for 15 min. The
1132 resulting supernatant was transferred to another tube, and cell wall fragments were harvested by

1133 centrifugation at 20,000 x *g* for 30 min. The resulting pellet was resuspended in 100 µl of the phage
1134 buffer and mixed with phi812 at MOI 10, relative to the starting bacterial culture. After 5 min of
1135 incubation at room temperature, the sample was applied to a holey carbon-coated copper grid for
1136 electron microscopy (R2/1, mesh 200; Quantifoil), blotted and vitrified by plunging into liquid ethane
1137 using a Vitrobot Mark IV (FEI), blot force 2, blotting time 2 s. Grids were imaged at a magnification of
1138 17,000 x, using a Tecnai TF20 operated (FEI) under cryogenic conditions and at an acceleration
1139 voltage of 200 kV.

**Acknowledgments**

We gratefully acknowledge the core facilities Cryo-electron microscopy and tomography, Biomolecular Interactions and Crystallography, Proteomics, and Josef Dadok National NMR Centre of CIISB, Instruct-CZ Centre, supported by MEYS CR (LM2023042) and European Regional Development Fund-Project „UP CIISB“ (No. CZ.02.1.01/0.0/0.0/18_046/0015974). We gratefully acknowledge support from the National Institute of Virology and Bacteriology (Program EXCELES, Project ID No. LX22NPO5103) - Funded by the European Union - Next Generation EU. This work received funding from ERC Consolidator Grant No. 101043452 to P.P.

**Data availability**

Cryo-EM maps and structure coordinates were deposited into the Electron Microscopy Data Bank (EMDB) and Protein Data Bank (PDB), respectively, with the following accession numbers: Structure of native phage phi812 tail computed with imposed sixfold symmetry: EMDB-50093 and PDB-9F04. Structures of baseplate core, wedge modules, arms, and baseplate-proximal part of phi812 tail in a native conformation computed with imposed threefold symmetry: EMDB-19968 and PDB-9EUF. Structures of baseplate core, wedge modules, and baseplate-proximal part of phi812 tail in a native conformation computed with imposed threefold symmetry: EMDB-19969 and PDB-9EUG. Structures of baseplate core and wedge modules of phi812 tail in a native conformation computed with imposed threefold symmetry: EMDB-19970 and PDB-9EUH. Structures of baseplate core, wedge modules, arms, and baseplate-proximal part of phi812 tail in a contracted conformation computed with imposed sixfold symmetry: EMDB-19971 and PDB-9EUI. Structures of baseplate core, wedge modules, arm segment, inner tripod complexes, and baseplate-proximal part of phi812 tail in a contracted conformation computed with imposed sixfold symmetry: EMDB-19972 and PDB-9EUJ. Structures of baseplate core, wedge modules, and inner tripod complexes of phi812 tail in a contracted conformation computed with imposed sixfold symmetry: EMDB-19973 and PDB-9EUK. Structure of tail tube of phage phi812 with contracted tail computed with imposed sixfold symmetry: EMDB-50094 and PDB-9F05. Structure of tail sheath of phage phi812 with contracted tail computed with imposed sixfold symmetry: EMDB-50095 and PDB-9F06. Structure of petal and knob domains of the central spike protein of phage phi812 in a native conformation, computed with imposed threefold symmetry: EMDB-19974 and PDB-9EUL. The NMR structure of the cleaver domain of the hub protein was deposited to PDB under the accession number 9EUM. The X-ray structure of the arm segment protein was deposited to PDB under the accession number 9FKO.

**Code availability**

All custom scripts used during the single-particle data reconstruction are available at https://github.com/fuzikt/starpy.

# References


1. Nobrega, F. L. *et al.* Targeting mechanisms of tailed bacteriophages. *Nat. Rev. Microbiol.* **16**, 760–773 (2018).
2. Leiman, P. G. & Shneider, M. M. Contractile Tail Machines of Bacteriophages. in *Viral Molecular Machines* (eds. Rossmann, M. G. & Rao, V. B.) vol. 726 93–114 (Springer US, Boston, MA, 2012).
3. Veesler, D. & Cambillau, C. A Common Evolutionary Origin for Tailed-Bacteriophage Functional Modules and Bacterial Machineries. *Microbiol. Mol. Biol. Rev.* **75**, 423–433 (2011).
4. Taylor, N. M. I. *et al.* Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* **533**, 346–352 (2016).
5. Taylor, N. M. I., van Raaij, M. J. & Leiman, P. G. Contractile injection systems of bacteriophages and related systems. *Mol. Microbiol.* **108**, 6–15 (2018).
6. Park, Y.-J. *et al.* Structure of the type VI secretion system TssK–TssF–TssG baseplate subcomplex revealed by cryo-electron microscopy. *Nat. Commun.* **9**, 5385 (2018).
7. Ge, P. *et al.* Action of a minimal contractile bactericidal nanomachine. *Nature* **580**, 658–662 (2020).
8. Jiang, F. *et al.* Cryo-EM Structure and Assembly of an Extracellular Contractile Injection System. *Cell* **177**, 370-383.e15 (2019).
9. Xu, J. *et al.* Identification and structure of an extracellular contractile injection system from the marine bacterium Algoriphagus machipongonensis. *Nat. Microbiol.* **7**, 397–410 (2022).
10. Li, F. *et al.* High-resolution cryo-EM structure of the Pseudomonas bacteriophage E217. *Nat. Commun.* **14**, 4052 (2023).
11. Wang, Z. *et al.* Structure of Vibrio Phage XM1, a Simple Contractile DNA Injection Machine. *Viruses* **15**, 1673 (2023).
12. Yang, F. *et al.* Fine structure and assembly pattern of a minimal myophage Pam3. *Proc. Natl. Acad. Sci.* **120**, e2213727120 (2023).
13. Yang, F., Wang, L., Zhou, J., Xiao, H. & Liu, H. In Situ Structures of the Ultra-Long Extended and Contracted Tail of Myoviridae Phage P1. *Viruses* **15**, 1267 (2023).
14. Yu, R.-C. *et al.* Structure of the intact tail machine of Anabaena myophage A-1(L). *Nat. Commun.* **15**, 2654 (2024).
15. Sonani, R. R. *et al.* An extensive disulfide bond network prevents tail contraction in Agrobacterium tumefaciens phage Milano. *Nat. Commun.* **15**, 756 (2024).
16. Pantůček, R. *et al.* The Polyvalent Staphylococcal Phage φ812:Its Host-Range Mutants and Related Phages. *Virology* **246**, 241–252 (1998).
17. Botka, T. *et al.* Lytic and genomic properties of spontaneous host-range Kayvirus mutants prove their suitability for upgrading phage therapeutics against staphylococci. *Sci. Rep.* **9**, 5475 (2019).
18. Kwiecinski, J. M. & Horswill, A. R. Staphylococcus aureus bloodstream infections: pathogenesis and regulatory mechanisms. *Curr. Opin. Microbiol.* **53**, 51–60 (2020).
19. Grunenwald, C. M., Bennett, M. R. & Skaar, E. P. Nonconventional Therapeutics against Staphylococcus aureus. *Microbiol. Spectr.* **6**, (2018).
20. Venturini, C., Petrovic Fabijan, A., Fajardo Lubian, A., Barbirz, S. & Iredell, J. Biological foundations of successful bacteriophage therapy. *EMBO Mol. Med.* **14**, (2022).
21. Petrovic Fabijan, A. *et al.* Safety of bacteriophage therapy in severe Staphylococcus aureus

1218 infection. *Nat. Microbiol.* **5**, 465–472 (2020).
1219 22. Uyttebroek, S. *et al.* Safety and efficacy of phage therapy in difficult-to-treat infections: a
1220 systematic review. *Lancet Infect. Dis.* **22**, e208–e220 (2022).
1221 23. Nováček, J. *et al.* Structure and genome release of Twort-like Myoviridae phage with a
1222 double-layered baseplate. *Proc. Natl. Acad. Sci.* **113**, 9351–9356 (2016).
1223 24. Guerrero-Ferreira, R. C. *et al.* Structure and transformation of bacteriophage A511 baseplate and
1224 tail upon infection of *Listeria* cells. *EMBO J.* **38**, (2019).
1225 25. Desfosses, A. *et al.* Atomic structures of an entire contractile injection system in both the
1226 extended and contracted states. *Nat. Microbiol.* **4**, 1885–1894 (2019).
1227 26. Weiss, G. L. *et al.* Structure of a thylakoid-anchored contractile injection system in multicellular
1228 cyanobacteria. *Nat. Microbiol.* **7**, 386–396 (2022).
1229 27. Eck, M. J. & Sprang, S. R. The Structure of Tumor Necrosis Factor-α at 2.6 Å Resolution. *J. Biol.*
1230 *Chem.* **264**, 17595–17605 (1989).
1231 28. Legrand, P. *et al.* The Atomic Structure of the Phage Tuc2009 Baseplate Tripod Suggests that Host
1232 Recognition Involves Two Different Carbohydrate Binding Modules. *mBio* **7**, (2016).
1233 29. Farenc, C. *et al.* Molecular Insights on the Recognition of a Lactococcus lactis Cell Wall Pellicle by
1234 the Phage 1358 Receptor Binding Protein. *J. Virol.* **88**, 7005–7015 (2014).
1235 30. McCabe, O. *et al.* The targeted recognition of *L actococcus lactis* phages to their polysaccharide
1236 receptors: Receptor binding to lactococcal phage 1358. *Mol. Microbiol.* **96**, 875–886 (2015).
1237 31. Šulák, O. *et al.* A TNF-like Trimeric Lectin Domain from *Burkholderia cenocepacia* with Specificity
1238 for Fucosylated Human Histo-Blood Group Antigens. *Structure* **18**, 59–72 (2010).
1239 32. Myers, C. L. *et al.* Identification of Two Phosphate Starvation-induced Wall Teichoic Acid
1240 Hydrolases Provides First Insights into the Degradative Pathway of a Key Bacterial Cell Wall
1241 Component. *J. Biol. Chem.* **291**, 26066–26082 (2016).
1242 33. Shi, L., Liu, J.-F., An, X.-M. & Liang, D.-C. Crystal structure of glycerophosphodiester
1243 phosphodiesterase (GDPD) from *Thermoanaerobacter tengcongensis* , a metal ion-dependent
1244 enzyme: Insight into the catalytic mechanism: Crystal Structure of GDPD. *Proteins Struct. Funct.*
1245 *Bioinforma.* **72**, 280–288 (2008).
1246 34. Browning, C., Shneider, M. M., Bowman, V. D., Schwarzer, D. & Leiman, P. G. Phage Pierces the
1247 Host Cell Membrane with the Iron-Loaded Spike. *Structure* **20**, 326–339 (2012).
1248 35. Kanamaru, S. *et al.* Structure of the cell-puncturing device of bacteriophage T4. *Nature* **415**,
1249 553–557 (2002).
1250 36. Shneider, M. M. *et al.* PAAR-repeat proteins sharpen and diversify the type VI secretion system
1251 spike. *Nature* **500**, 350–353 (2013).
1252 37. Kizziah, J. L., Manning, K. A., Dearborn, A. D. & Dokland, T. *Structure of the Host Cell Recognition*
1253 *and Penetration Machinery of a* Staphylococcus Aureus *Bacteriophage*.
1254 http://biorxiv.org/lookup/doi/10.1101/721746 (2019) doi:10.1101/721746.
1255 38. Büttner, C. R., Wu, Y., Maxwell, K. L. & Davidson, A. R. Baseplate assembly of phage Mu: Defining
1256 the conserved core components of contractile-tailed phages and related bacterial systems. *Proc.*
1257 *Natl. Acad. Sci.* **113**, 10174–10179 (2016).
1258 39. Leiman, P. G. *et al.* Type VI secretion apparatus and phage tail-associated protein complexes
1259 share a common evolutionary origin. *Proc. Natl. Acad. Sci.* **106**, 4154–4159 (2009).
1260 40. Bateman, A. & Rawlings, N. D. The CHAP domain: a large family of amidases including GSP
1261 amidase and peptidoglycan hydrolases. *Trends Biochem. Sci.* **28**, 234–237 (2003).
1262 41. Sanz-Gaitero, M., Keary, R., Garcia-Doval, C., Coffey, A. & van Raaij, M. J. Crystal structure of the

lytic CHAPK domain of the endolysin LysK from Staphylococcus aureus bacteriophage K. *Virol. J.* **11**, 133 (2014).
42. Becker, S. C., Foster-Frey, J. & Donovan, D. M. The phage K lytic enzyme LysK and lysostaphin act synergistically to kill MRSA. *FEMS Microbiol. Lett.* **287**, 185–191 (2008).
43. Paul, V. D. *et al.* A novel bacteriophage Tail-Associated Muralytic Enzyme (TAME) from Phage K and its development into a potent antistaphylococcal protein. *BMC Microbiol.* **11**, 226 (2011).
44. Willis, C. *et al.* Insights into the Membrane Interactions of the Saposin-Like Proteins Na-SLP-1 and Ac-SLP-1 from Human and Dog Hookworm. *PLOS ONE* **6**, e25369 (2011).
45. Lee, J. H. *et al.* Crystal Structure and Functional Studies Reveal that PAS Factor from Vibrio vulnificus is a Novel Member of the Saposin-fold Family. *J. Mol. Biol.* **355**, 491–500 (2006).
46. Cai, X. *et al.* Atomic structures of a bacteriocin targeting Gram-positive bacteria. *Nat. Commun.* **15**, 7057 (2024).
47. Böth, D., Steiner, E. M., Izumi, A., Schneider, G. & Schnell, R. RipD (Rv1566c) from *Mycobacterium tuberculosis* : adaptation of an NlpC/p60 domain to a non-catalytic peptidoglycan-binding function. *Biochem. J.* **457**, 33–41 (2014).
48. Habann, M. *et al.* Listeria phage A511, a model for the contractile tail machineries of SPO1-related bacteriophages. *Mol. Microbiol.* **92**, 84–99 (2014).
49. Schaechter, M. *Encyclopedia of Microbiology*. (Elsevier/Academic Press, Amsterdam Boston, 2009).
50. Maxwell, K. L. *et al.* Structural and Functional Studies of gpX of Escherichia coli Phage P2 Reveal a Widespread Role for LysM Domains in the Baseplates of Contractile-Tailed Phages. *J. Bacteriol.* **195**, 5461–5468 (2013).
51. Watts, N. R. & Coombs, D. H. Structure of the bacteriophage T4 baseplate as determined by chemical cross-linking. *J. Virol.* **64**, 143–154 (1990).
52. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
53. Leiman, P. G. *et al.* Structure and Location of Gene Product 8 in the Bacteriophage T4 Baseplate. *J. Mol. Biol.* **328**, 821–833 (2003).
54. van Dalen, R., Peschel, A. & van Sorge, N. M. Wall Teichoic Acid in Staphylococcus aureus Host Interaction. *Trends Microbiol.* **28**, 985–998 (2020).
55. Xia, G. *et al.* Wall Teichoic Acid-Dependent Adsorption of Staphylococcal Siphovirus and Myovirus. *J. Bacteriol.* **193**, 4006–4009 (2011).
56. Takeuchi, I. *et al.* The Presence of Two Receptor-Binding Proteins Contributes to the Wide Host Range of Staphylococcal Twort-Like Phages. *Appl. Environ. Microbiol.* **82**, 5763–5774 (2016).
57. Azam, A. H., Hoshiga, F., Takeuchi, I., Miyanaga, K. & Tanji, Y. Analysis of phage resistance in Staphylococcus aureus SA003 reveals different binding mechanisms for the closely related Twort-like phages ϕSA012 and ϕSA039. *Appl. Microbiol. Biotechnol.* **102**, 8963–8977 (2018).
58. Peng, C. *et al.* Silviavirus phage ϕMR003 displays a broad host range against methicillin-resistant Staphylococcus aureus of human origin. *Appl. Microbiol. Biotechnol.* **103**, 7751–7765 (2019).
59. Cid, M. *et al.* Recognition of the Helical Structure of β-1,4-Galactan by a New Family of Carbohydrate-binding Modules. *J. Biol. Chem.* **285**, 35999–36009 (2010).
60. Bae, B. *et al.* Molecular Basis for the Selectivity and Specificity of Ligand Recognition by the Family 16 Carbohydrate-binding Modules from Thermoanaerobacterium polysaccharolyticum ManA. *J. Biol. Chem.* **283**, 12415–12425 (2008).
61. Bradshaw, W. J. *et al.* The structure of the cysteine protease and lectin-like domains of Cwp84, a

surface layer-associated protein from *Clostridium difficile*. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 1983–1993 (2014).
62. Buist, G., Steen, A., Kok, J. & Kuipers, O. P. LysM, a widely distributed protein motif for binding to (peptido)glycans. *Mol. Microbiol.* **68**, 838–847 (2008).
63. Šiborová, M. *et al.* Tail proteins of phage SU10 reorganize into the nozzle for genome delivery. *Nat. Commun.* **13**, 5622 (2022).
64. Sciara, G. *et al.* Structure of lactococcal phage p2 baseplate and its mechanism of activation. *Proc. Natl. Acad. Sci.* **107**, 6852–6857 (2010).
65. Kostyuchenko, V. A. *et al.* The tail structure of bacteriophage T4 and its mechanism of contraction. *Nat. Struct. Mol. Biol.* **12**, 810–813 (2005).
66. Zinke, M., Schröder, G. F. & Lange, A. Major tail proteins of bacteriophages of the order Caudovirales. *J. Biol. Chem.* **298**, 101472 (2022).
67. Kudryashev, M. *et al.* Structure of the Type VI Secretion System Contractile Sheath. *Cell* **160**, 952–962 (2015).
68. Ouyang, R. *et al.* High-resolution reconstruction of a Jumbo-bacteriophage infecting capsulated bacteria using hyperbranched tail fibers. *Nat. Commun.* **13**, 7241 (2022).
69. Aksyuk, A. A. *et al.* The tail sheath structure of bacteriophage T4: a molecular machine for infecting bacteria. *EMBO J.* **28**, 821–829 (2009).
70. Maghsoodi, A., Chatterjee, A., Andricioaei, I. & Perkins, N. C. How the phage T4 injection machinery works including energetics, forces, and dynamic pathway. *Proc. Natl. Acad. Sci.* **116**, 25097–25105 (2019).
71. Ge, P. *et al.* Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states. *Nat. Struct. Mol. Biol.* **22**, 377–382 (2015).
72. Moody, M. F. Sheath of bacteriophage T4. *J. Mol. Biol.* **80**, 613–635 (1973).
73. Nazarov, S. *et al.* Cryo- EM reconstruction of Type VI secretion system baseplate and sheath distal end. *EMBO J.* **37**, e97103 (2018).
74. Fraser, A. *et al.* Quantitative description of a contractile macromolecular machine. *Sci. Adv.* **7**, eabf9601 (2021).
75. Fernandes, S. & São-José, C. Enzymes and Mechanisms Employed by Tailed Bacteriophages to Breach the Bacterial Cell Barriers. *Viruses* **10**, 396 (2018).
76. Nishima, W., Kanamaru, S., Arisaka, F. & Kitao, A. Screw Motion Regulates Multiple Functions of T4 Phage Protein Gene Product 5 during Cell Puncturing. *J. Am. Chem. Soc.* **133**, 13571–13576 (2011).
77. Wenzel, S., Shneider, M. M., Leiman, P. G., Kuhn, A. & Kiefer, D. The Central Spike Complex of Bacteriophage T4 Contacts PpiD in the Periplasm of Escherichia coli. *Viruses* **12**, 1135 (2020).
78. Hu, B., Margolin, W., Molineux, I. J. & Liu, J. Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proc. Natl. Acad. Sci.* **112**, E4919–E4928 (2015).
79. Gillespie, J. J. *et al.* Structural Insight into How Bacteria Prevent Interference between Multiple Divergent Type IV Secretion Systems. *mBio* **6**, e01867-15 (2015).
80. Wu, X. *et al.* Crystal structure of CagV, the *Helicobacter pylori* homologue of the T4 SS protein VirB8. *FEBS J.* **286**, 4294–4309 (2019).
81. Xu, Q. *et al.* Structure of the γ- D -glutamyl- L -diamino acid endopeptidase YkfC from *Bacillus cereus* in complex with L -Ala-γ- D -Glu: insights into substrate recognition by NlpC/P60 cysteine peptidases. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **66**, 1354–1364 (2010).
82. Bhardwaj, A., Walker-Kopp, N., Casjens, S. R. & Cingolani, G. An Evolutionarily Conserved Family

of Virion Tail Needles Related to Bacteriophage P22 gp26: Correlation between Structural Stability and Length of the α-Helical Trimeric Coiled Coil. *J. Mol. Biol.* **391**, 227–245 (2009).

83. Bhardwaj, A., Molineux, I. J., Casjens, S. R. & Cingolani, G. Atomic Structure of Bacteriophage Sf6 Tail Needle Knob. *J. Biol. Chem.* **286**, 30867–30877 (2011).
84. Najmudin, S. *et al.* Putting an N-terminal end to the Clostridium thermocellum xylanase Xyn10B story: Crystal structure of the CBM22-1–GH10 modules complexed with xylohexaose. *J. Struct. Biol.* **172**, 353–362 (2010).
85. Sainz-Polo, M. A., González, B., Menéndez, M., Pastor, F. I. J. & Sanz-Aparicio, J. Exploring Multimodularity in Plant Cell Wall Deconstruction. *J. Biol. Chem.* **290**, 17116–17130 (2015).
86. Bárdy, P. *et al.* Structure and mechanism of DNA delivery of a gene transfer agent. *Nat. Commun.* **11**, 3034 (2020).
87. Hardy, J. M. *et al.* The architecture and stabilisation of flagellotropic tailed bacteriophages. *Nat. Commun.* **11**, 3748 (2020).
88. Chao, K. L. *et al.* Structure of Escherichia coli O157:H7 bacteriophage CBA120 tailspike protein 4 baseplate anchor and tailspike assembly domains (TSP4-N). *Sci. Rep.* **12**, 2061 (2022).
89. Han, Z. *et al.* Structural insights into a spindle-shaped archaeal virus with a sevenfold symmetrical tail. *Proc. Natl. Acad. Sci.* **119**, e2119439119 (2022).
90. Tamura, K. *et al.* Surface glycan-binding proteins are essential for cereal beta-glucan utilization by the human gut symbiont Bacteroides ovatus. *Cell. Mol. Life Sci.* **76**, 4319–4340 (2019).
91. Granell, M., Namura, M., Alvira, S., Kanamaru, S. & van Raaij, M. Crystal Structure of the Carboxy-Terminal Region of the Bacteriophage T4 Proximal Long Tail Fiber Protein Gp34. *Viruses* **9**, 168 (2017).
92. Salazar, A. J., Sherekar, M., Tsai, J. & Sacchettini, J. C. R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold. *PLOS ONE* **14**, e0211432 (2019).
93. Plattner, M. *et al.* Structure and Function of the Branched Receptor-Binding Complex of Bacteriophage CBA120. *J. Mol. Biol.* **431**, 3718–3739 (2019).
94. Garcia-Doval, C. & van Raaij, M. J. Structure of the receptor-binding carboxy-terminal domain of bacteriophage T7 tail fibers. *Proc. Natl. Acad. Sci.* **109**, 9390–9395 (2012).
95. Fokine, A. *et al.* Molecular architecture of the prolate head of bacteriophage T4. *Proc. Natl. Acad. Sci.* **101**, 6003–6008 (2004).
96. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
97. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
98. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
99. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
100. Zivanov, J., Nakane, T. & Scheres, S. H. W. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ* **7**, 253–267 (2020).
101. Kimanius, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S. H. W. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem. J.* **478**, 4169–4185 (2021).
102. Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 1–13 (2019).
103. Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).

104. Ilca, S. L. *et al.* Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nat. Commun.* **6**, 8843 (2015).
105. Evangelidis, T. *et al.* Automated NMR resonance assignments and structure determination using a minimal set of 4D spectra. *Nat. Commun.* **9**, 384 (2018).
106. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009).
107. Güntert, P. Automated structure determination from NMR spectra. *Eur. Biophys. J.* **38**, 129–143 (2009).
108. Linge, J. P., Williams, M. A., Spronk, C. A. E. M., Bonvin, A. M. J. J. & Nilges, M. Refinement of protein structures in explicit solvent. *Proteins Struct. Funct. Bioinforma.* **50**, 496–506 (2003).
109. Nederveen, A. J. *et al.* RECOORD: A recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins Struct. Funct. Bioinforma.* **59**, 662–672 (2005).
110. Bhattacharya, A., Tejero, R. & Montelione, G. T. Evaluating protein structures determined by structural genomics consortia. *Proteins Struct. Funct. Bioinforma.* **66**, 778–795 (2007).
111. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
112. Schneider, T. R. & Sheldrick, G. M. Substructure solution with *SHELXD*. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772–1779 (2002).
113. McCoy, A. J. *et al. Phaser* crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
114. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
115. Vagin, A. A. *et al. REFMAC*5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2184–2195 (2004).
116. Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
117. Källberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **7**, 1511–1522 (2012).
118. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
119. Meng, E. C. *et al.* UCSF ChimeraX : Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).
120. Croll, T. I. *ISOLDE* : a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. Sect. Struct. Biol.* **74**, 519–530 (2018).
121. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallogr. Sect. Struct. Biol.* **75**, 861–877 (2019).
122. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
123. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).
124. Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135–145 (2018).
125. Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276–W279 (2022).
126. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a

multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).