Conformational changes of baseplate linked to tail contraction of *S. aureus* phage phi812

Short title: 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 mechanisms 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-positive *Staphylococcus aureus*, is formed of a core, wedge modules, and baseplate arms carrying receptor-binding proteins type 1 and 2 and tripod complexes. The baseplate of phi812 with extended tail has threefold symmetry but after tail contraction it transforms to sixfold symmetry. In this process, the receptor-binding proteins re-orient to bind to the host cell. The conformational changes to the tripod complexes trigger the release of the central spike and weld proteins from the baseplate, which exposes peptidoglycan-degrading domains of the hub proteins. 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 triggers the contraction of the tail sheath. The tail sheath of phi812, shortens from 200 to 96 nm, which pushes the tail tube 10-30 nm into the cytoplasm, depending on the local thickness of the *S. aureus* cell wall. Our results demonstrate how tail and baseplate enable phi812 genome delivery into Gram-positive *S. aureus*.

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[3–9]. Current knowledge regarding the mechanisms that are employed to breach the cell wall of Gram-positive bacteria is limited.

Phages of the genus *Kayvirus* of the family *Herelleviridae* exhibit high lytic activity against Gram-positive bacterium *Staphylococcus aureus*[10]. This human pathogen can cause skin and epithelial infections, as well as bacteremia that can develop into life-threatening sepsis[11,12]. *S. aureus* is frequently resistant to antibiotics, which has renewed interest in phage therapy[13–15]. The best studied *Kayvirus*, phage K, shares 98.5% of the genome identity with a polyvalent phage phi812. Host range mutants of phage phi812 were shown to infect more than 90 % of *Staphylococcus aureus* isolates[10,16]. Major components of the *S. aureus* cell wall are peptidoglycan and teichoic acids [17,18]. Phage phi812 and related *Kayvirus* phages recognize the backbone and side-chain moieties of the wall teichoic acid[10,19–22].

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 of the *Herelleviridae* family, which infects Gram-positive *Listeria* and is morphologically similar 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.

The naming convention for baseplate components was established in the pioneering studies of phage T4. The core of the baseplate is formed by central hub[3,25], which serves as a centerpiece for the baseplate assembly and a platform for the tail tube polymerization. Six wedge complexes encircle the core and provide foundation for tail sheath assembly[3,25]. Baseplate core proteins of contractile-tailed phages exhibit a high degree of structural conservation[3,25,26]. In addition, the structures of the baseplate core proteins are shared by CISs[27–33]. The outer baseplate of contractile-tailed phages can be either simple, containing one or two tail fibers to bind to host cellular receptors[4–9], or more intricate, such as that of T4[3].

Here, we used a combination of cryogenic electron microscopy (cryo-EM), X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR), and AlphaFold2 predictions to build the molecular structure of the phi812 tail and baseplate in extended and contracted conformations. 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.

Results and Discussion

The structure of baseplate of phi812 with extended tail

The structure of the phi812 baseplate with imposed sixfold symmetry was previously determined to a resolution of 13 Å using cryo-EM[23]. Here we show that the phi812 baseplate, similar to that of A511[24], is only approximately sixfold symmetric. Relaxation 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-C, SFig. 1, 2A and 3, Table S1). The baseplate of phi812 is lentil-shaped with a diameter of 655 Å and height of 310 Å, and its constituent proteins can be assigned to a core, six wedge modules, and six arms decorated with three types of putative receptor-binding proteins (Fig. 1C, D, SMovies 1 and 2).

Each of the baseplate arms is formed by the arm scaffold protein (gp117) reinforced by the arm segment protein (gp118) (Fig. 1C). The baseplate arm provides attachment sites for proteins with putative receptor binding functions: 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-C, 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-C). 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, C). Trimers of RBP1 and RBP2 are attached to the end of the baseplate arm (Fig. 1A-C). The RBP1 reaches from one baseplate arm to the neighboring one, where it interacts with RBP2 (Fig. 1B). The RBP2 points towards the tail sheath (Fig. 1B, C).

The threefold symmetry of the phi812 baseplate is most apparent in the distinct conformations of the alternating baseplate arms and the positions of the putative receptor-binding proteins attached to them (Fig. 1A-C, SMovie 2). Alternating arrangement of baseplate arms is caused by different angles at which the arms point away from the tail axis (Fig. 1C). The upper baseplate arms radiate from the tail axis at an angle of 72°, while the lower baseplate arms radiate at an angle of 62°. The differences in the baseplate arm angles bring the six inner tripods to form a shape of an equilateral triangle when viewed along the tail axis towards the phage head (Fig. 1A). The center of mass of the inner tripods attached to the lower baseplate arms is positioned 141 Å from the tail axis, whereas of those attached to the upper baseplate arms is 87 Å from the tail axis. The distinct angles of baseplate arms also influence the interactions of tripod proteins with the tail sheath. The outer tripod attached to the upper baseplate arm interacts with the first tail sheath disc (SFig. 4B, D). In contrast, the outer tripod bound to the lower baseplate arm does not interact with the tail sheath (SFig. 4A, C).

Each baseplate arm protrudes radially from a wedge module. A localized reconstruction of the wedge modules and baseplate core reached a resolution of 4.4 Å (Fig. 2, SFig. 1, 2C, 3, 5, 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. 2, SFig. 6). Two neighboring wedge modules that connect to upper and lower baseplate arms assume distinct conformations (SFig. 7A-C).

Six wedge modules encircle the baseplate core composed from tail tube initiator, tail sheath initiator, hub, assembly, central spike, and weld proteins (Fig. 2, SFig 6). 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), bound to the three

118 copies of weld protein (gp111), plugs the baseplate channel at the level of the hub proteins (Fig. 2A,
119 B). The hexamer of tube initiator proteins is surrounded by six copies of each of the sheath initiator
120 (gp115) and assembly proteins (gp114), which are placed above each other (Fig. 2A, SFig 6).
121 It is likely that the overall threefold symmetry of the phi812 baseplate is determined during
122 assembly by asymmetric addition of baseplate proteins to the trimer of hub proteins. The structural
123 differences propagate into the baseplate periphery, resulting in the alternating angles of the
124 baseplate arms relative to the tail axis (Fig. 1B, SMovie 2). The same mechanism is probably
125 employed to establish the threefold symmetry of the A511 baseplate[24]. In contrast, baseplate core
126 proteins of recently characterized phages were shown to be arranged with quasi-sixfold symmetry[3–9].
127
128
129 **Central spike protein blocks the baseplate channel and protrudes from the baseplate**

130 The trimer of central spike proteins of phi812 blocks the baseplate channel and protrudes more than
131 30 nm from the baseplate (Fig. 3A, B, SMovie 2). The central spike protein can be divided into
132 oligosaccharide-binding (OB), β-prism, coiled-coil, knob, and petal domains (SFig. 8A, Table S2). The
133 OB and β-prism domains have conserved structures shared by central spike proteins of other phages
134 and CISs[4,6–9,28–32]. The two domains bind to the inside of the channel formed by the hub proteins (Fig.
135 2A, B, SFig. 6). The coiled-coil, knob, and petal domains protrude from the phi812 baseplate and are
136 not resolved in the cryo-EM reconstruction of the baseplate, due to the coiled-coil domain's
137 flexibility (Fig. 3A). Structural prediction indicates that the coiled-coil domains from the three central
138 spike proteins form a 22-nm-long fiber (Fig. 2B). The structures of the knob and petal domains were
139 determined by cryo-EM and single-particle reconstruction of a recombinantly produced central spike
140 protein (Fig. 2B, SFig. 1, 2J, 3, Table S1). The knob domain has a tumor necrosis factor-like fold[34],
141 similar to that of the head domains of receptor-binding proteins from lactococcal phages Tuc2009
142 and 1358[35–37] and a lectin from *Burkholderia cenocepacia*[38] which indicates that the knob domain
143 may bind to saccharides from the bacterial cell wall. The petal domain of the central spike protein
144 has an α/β-barrel fold structurally similar to phosphodiesterases, such as the GlpQ protein from *B.*
145 *subtilis* that cleaves teichoic acid during phosphate starvation[39,40]. Additionally, the petal domain
146 includes two conserved histidines commonly found in active sites of phosphodiesterases, suggesting
147 the domain's putative cleavage activity (Fig. 3C).
148 Phages with contractile tails and CISs attacking Gram-negative bacteria pierce the outer
149 bacterial membranes using puncturing tips formed of proteins with β-prism or β-helix domains[41–43].
150 In contrast, to infect Gram-positive *S. aureus*, phage phi812 has to penetrate a 50-70-nm-thick layer
151 of teichoic acids and peptidoglycan. Therefore, we propose that the knob and petal domains of the
152 central spike of phi812 bind to and degrade the host teichoic acids to expose the peptidoglycan layer
153 for degradation by other proteins of the baseplate.
154
155 **Hub proteins facilitate penetration of the tail tube through the cell wall and membrane**

156 The three hub proteins form a ring, extending the tail channel into the phi812 baseplate (Fig. 2A, B).
157 Phi812 hub protein can be divided into channel-forming domains I to IV and clamp and cleaver
158 domains (SFig. 8B, Table S2). The channel-forming domains I and III have beta-barrel folds and are
159 positioned next to each other so that the head-proximal interface formed by the trimer of hub

proteins has pseudo-sixfold symmetry for binding the hexamer of tube initiator proteins (SFig. 6). In contrast, the channel-forming domains II and IV of the hub protein, positioned below domains I and III respectively, are structurally distinct and form interfaces for binding the trimer of central spike proteins to the inside of the hub channel (SFig. 6). The channel-forming domains I-IV are structurally similar to those of the gp27 of phage T4, for which the domain naming convention was established[42]. Homologs of the channel-forming domains can also be identified among phages with long non-contractile tails (Tal proteins)[26,44] and CISs (Vgrg-like proteins)[25,45,46] (Table S2).

The cleaver domain of the hub protein, the structure of which was determined using NMR (Fig. 3D, SFig. 8B, Table S3), has the characteristic fold of the cysteine- and histidine-dependent amido-hydrolase/peptidase (CHAP)[47], and contains the conserved catalytic triad Cys699-His759-Glu777[48] (Fig. 3D). Accordingly, zymogram analysis demonstrated that a recombinant cleaver domain degrades *S. aureus* cell walls (hence the name cleaver) (SFig. 9), which agrees with previous functional studies of hub protein homologs from the related phage K[49,50]. The clamp domain of the hub protein is formed by ten short α-helices that connect the central spike protein to weld proteins (Fig. 2B, C, and 3E, SFig. 6). The clamp domain is structurally similar to pore-forming saposin-like proteins, which employ positively charged residues to interact with the negatively charged surfaces of lipid membranes (Table S2)[51,52]. Since the clamp domain of the phi812 hub protein contains a long positively charged patch of residues, as the saposin-like proteins do (Fig. 3E), we speculate that it may bind the *S. aureus* membrane to facilitate its penetration by a tail tube. In addition to gp27 of T4 featuring a lysozyme domain, more hub proteins equipped with enzymatic domains have been reported recently[8,33]. Therefore, we propose that the hub proteins of phi812 and related phages enable genome delivery by degrading peptidoglycan to facilitate penetration of the tail tube through the cell wall.

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

Three weld proteins bind to the central spike, hub, and three inner tripod proteins, connecting them together (Fig. 2A-C, SFig. 6 and 10). Six assembly and six tail sheath initiator proteins surround the hub and tail tube initiator proteins to complete the structure of the baseplate core (Fig. 2A, B, E, F, SFig. 6). The tube initiator and sheath initiator proteins are organized with sixfold symmetry, whereas the central spike, weld, hub, and assembly proteins are organized with threefold symmetry that determines the overall organization of the baseplate when phi812 tail is in extended conformation.

The weld protein of phi812 can be divided into the N-terminal lid and C-terminal clip domains, which are connected by a 14-residue-long linker (SFig. 8D). The lid domain interacts with an inner tripod and serves as a lid covering the peptidase site of the cleaver domain of the hub protein (SFig. 10). Therefore, releasing the weld protein from the baseplate upon host cell binding exposes the active site of the cleaver domain of hub protein to hydrolyze cell wall peptidoglycan. The clip domain of the weld protein, located closer to the tail axis, binds the weld protein to the central spike protein (Fig. 2C). The clip domain has the CHAP protein fold similar to that of the cleaver domain of the hub protein (Fig. 3F). However, the clip domain lacks the conserved catalytic residues Cys-His-Glu required for the peptidoglycan cleavage[48] (Fig. 3F). Thus, the clip domain is likely enzymatically inactive, but may have retained peptidoglycan-binding ability, similar to the RipD protein from *M. tuberculosis*[53]. There is no homolog of the weld protein in the baseplate of the related phage A511[24,54]. The weld protein may be unnecessary for A511 because of the differences in the cell wall composition between *Staphylococcus* and *Listeria*[17].

The assembly protein of phi812 is formed by a compact bundle of nine α-helices. The assembly protein interacts with the wedge protein and C-terminal handle of the tail tube initiator protein (Fig. 2E, SFig. 6). There are proteins with a compact globular structure and in the equivalent position in the baseplates of pyocin R2 and phages Pam3, E217, A-1 and Milano (Table S2), indicating the structural role of these proteins in baseplate cores[4,6,8,9,28].

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. 2A). 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. 6, 8C). 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[2,26].

The tail sheath initiator protein of phi812 is composed of the LysM and handshake domains (SFig. 8E). 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. 11C). Therefore, the interactions of the tail sheath and tail sheath initiator proteins can initiate formation of the phi812 tail sheath. 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)[25,45,55].

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[3,56]. 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. 2A, SFig. 6, SMovie 1). Wedge protein can be divided into N-terminal α-helices, wing, trifurcation, and dimerization domains (SFig. 8F, G), 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. 4A, B). Individual parts of the phi812 wedge module are structurally conserved and can be found in the baseplate of phage T4[3]. The N-terminal helices and trifurcation domains mediate contacts between the proteins forming the wedge module (Fig. 2F, SFig. 7A-C). The wing domain of wedge protein A binds to the hub and central spike proteins (Fig. 2D). In contrast, the wing domain of wedge protein B binds to the tail sheath initiator protein (Fig. 2F). The six wedge modules form an iris around the baseplate core (SFig. 7A), as described in phage T4[3]. Dimerization domains of wedge proteins mediate interactions between neighboring wedge modules (SFig. 7A). Even though the neighboring wedge modules differ in their conformations due to the overall threefold symmetry of the baseplate, the contacts between the dimerization domains are similar (SFig. 7A-C, F).

Baseplate arms provide attachment sites for putative receptor-binding proteins

Baseplate arms of phi812 are formed by linker AB, bulge, linker BC, elbow, linker Z, and connector domains of the arm scaffold protein that are reinforced by dimers of the arm segment proteins A, B, and C (Fig. 4A-D). We employed X-ray crystallography to determine the structure of the arm segment

protein (Fig. 4C, Table S4), and AlphaFold2[57] 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. 4D, SFig. 12).

Each arm segment protein is composed of an N-terminal α-helix followed by a three-layered β-sandwich (Fig. 4C). The overall structure of arm segment protein is similar to those of gp8 from phage T4 and gp105 from phage A511[24,58]. When fitted into the cryo-EM map, the C-terminal β-strands of neighboring subunits of arm segments A and B were separated by approximately 10.2 Å. Structural prediction positions β-strand formed by residues 197-207 of linkerAB of the arm scaffold protein in the gap between arm segment proteins (Fig. 4E, SFig. 12B-E). Therefore, in accordance with previous hypothesis[24], the linker AB of the arm scaffold protein likely fills the gap between the arm segments A and B, 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. 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. 4A, B, D). Similarly, the bulge domain of the arm scaffold protein contains 40-residue-long docking loop 2, which probably enables attachment of the outer tripod complex (Fig 4A, B, D). 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 probably located at the distal end of the baseplate arm, providing attachment sites for RBP1 and RBP2, respectively (Fig. 4A, B, D).

Baseplate of phi812 contains twelve tripod protein complexes

The most prominent components of the phi812 baseplate are the twelve trimers of tripod proteins (Fig. 1A-C). We predicted structures of the tripod proteins in phi812 with extended tail using AlphaFold2[57] because the ~10Å resolution of the reconstruction did not allow building of their structure (Fig. 5A-C, SFig. 13). The tripod protein can be divided into the spine, flap, fiber, base, fin, leg, anchor, and hook domains (Fig. 5C). The trimer of tripod proteins resembles a three-sided pyramid (Fig. 5B). The spine domains are positioned along the axis of the pyramid. The flap and fiber domains form the edge, the base domain forms the base, and the fin, leg, and anchor domains are located at the faces of the imaginary pyramid (Fig. 5A, C). The flap domain contains an insertion loop, three of which play a role in the attachment of the tripod complex to the arm scaffold protein (see below) (Fig. 5C). 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. 14, Table S2).

Receptor-binding proteins

The baseplate of phi812 contains six trimers of each of RBP1 and RBP2 with putative receptor-binding activities (Fig. 1A-C). As the cryo-EM reconstruction of the pre-contraction baseplate did not resolve RBP1 and RBP2 to sufficient resolutions to enable structure determination, we employed AlphaFold2[57] prediction to obtain the structures (Fig. 5D-G). The predicted structure of the 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. 5E). The tower domains are structurally similar to each other and contain characteristic six-stranded β-sheets. The predicted structure of the RBP1 trimer could be directly fitted into the cryo-EM reconstruction of the baseplate (Fig. 5D). The antenna domain likely mediates the binding of RBP1 to the end of the baseplate arm and the tip domain of RBP1 points towards the neighboring baseplate arm (Fig. 1A, B and 5A). RBP2 has the same domain composition as RBP1 with the exception of the wing domain, which is missing (Fig. 5G). In the baseplate, the antenna domain of RBP2 binds near arm segment C and the tip domain is in contact with a wedge module (Fig. 1C and 5A).

Re-arrangement of RBPs and tripod complexes

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[1,3,23]. 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[2,3]. Baseplate of phi812 particles with contracted tail was reconstructed to an overall resolution of 4.4 Å (SFig. 1, 2D, 3, Table S1). The post-contraction baseplate has the shape of a six-pointed star with a diameter of 713 Å and a thickness of 344 Å (Fig. 6B, C). The conformational change transforms the baseplate to sixfold symmetry, with all six baseplate arms protruding at 83° relative to the tail axis (Fig. 6C, SMovie 3). In the post-contraction baseplate, the upper and lower baseplate arms are shifted 11 and 15 Å away from the tail axis, respectively, relative to their positions in the baseplate of the extended tail (Fig. 1C, 6C). Comparing the structures of the baseplates of phi812 particles with extended and contracted tails enabled us to describe conformational changes controlling the baseplate rearrangement and tail sheath contraction. Reorganization of phi812 baseplate is probably initiated by the interaction of the RBP1 with teichoic acid as its putative receptor. 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). In the post-contraction baseplate, RBP2 trimers are oriented parallel with the tail axis and their tip domains face away from the phage head (Fig. 6A, C, SMovie 3). Movement of RBP1 and RBP2 provides space for the outer tripod complexes to rotate 170° to point their receptor-binding domains away from the phage head (Fig. 6A-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 6A-C).

The interaction between the inner tripod complex and the arm scaffold protein is well resolved in the focused reconstruction of a contracted baseplate with a resolution reaching 3.1 Å (SFig. 1, 2F, 3, Table S1). 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.

329 15A-C). Since the bulge domain of the arm scaffold protein contains a docking loop 2 with
330 hydrophobic residues in equivalent positions to those of the docking loop 1 of the trifurcation
331 domain (SFig. 15D), the inner and outer tripod complexes are likely attached to the arm scaffold
332 protein using similar interactions. The docking loop 2 of the bulge domain, which binds the outer
333 tripod complex, is twelve residues longer than the docking loop 1 of the trifurcation domain (SFig.
334 15D). The extra residues of the docking loop 2 probably enable the ~170° rotation of the outer tripod
335 complexes during the baseplate transformation.

336 The tripod complexes not only rotate, but also undergo complex conformational changes
337 (Fig. 6A, D, E, SMovie 4): (i) The second and third spine helices move towards the first spine helix by
338 10 and 16 Å, respectively. (ii) The fiber domain moves 13 Å towards the flap domain and rotates 20°
339 counterclockwise around the tripod symmetry axis when looking from the bottom of the tripod. (iii)
340 The base domain shifts 24 Å along the third spine helix while rotating 174° around an axis
341 perpendicular to the tripod symmetry axis. (iv) The fin domain moves 56 Å to protrude laterally from
342 the tripod. (v) The three leg domains detach from the flap domains and move 88 Å to interact with
343 each other under the base domains. (vi) The anchor and hook domains were not resolved in the
344 cryo-EM reconstruction of the baseplate from a phage with contracted tail (Fig. 6D, E); however,
345 phage particles attached to the cell wall contain density at the bottom of the tripod complexes,
346 suggesting that the anchor and hook domains stretch along the tripod axis toward the cell wall (SFig.
347 16). Similar density under the tripod complexes was observed in micrographs of contracted phage
348 A511 attached to *Listeria* cells, and in the baseplate reconstructions of the spontaneously contracted
349 phi812[23] and A511[24]. Because of their positioning after the tripod conformational change and their
350 sequence similarity to peptidoglycan-binding proteins of LysM family (SFig. 14), we speculate that the
351 hook domains of phi812 and related phages could bind the host cell wall.

352 The tripod proteins' conformational change is required to form tripod-tripod interactions in
353 phages with contracted tails. In particular, the rotation and translation of the base domain are
354 essential to expose an interface to which a fiber domain of the neighboring tripod complex binds
355 through an area of 1,900 $Å^2$ (SFig. 17). Each inner tripod complex connects to two inner tripod from
356 neighboring baseplate arms and an outer tripod complex attached to the same baseplate arm.
357 Additionally, one fiber domain of each outer tripod complex interacts with tower domains 1 and 2 of
358 RBP2 (Fig. 6A-C). The contacts mediated by the fiber domains probably stabilize the baseplate
359 structure in phages with contracted tails.

360 Overall, the recognition and adhesion of phi812 to the host cell requires re-orientation of the
361 RBP1, RBP2, and conformational changes of tripod complexes, affecting the positions of the six arms
362 of the post-contraction baseplate. Less extensive conformational changes were described for the
363 receptor-binding proteins of other phages[4,63,64].

364

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

366 The structure of wedge modules and baseplate core is affected by the re-arrangements of
367 receptor-binding proteins and baseplate arms, which in turn triggers the contraction of the tail
368 sheath. The polypeptide chains of arm scaffold proteins link baseplate arms of phi812 to the wedge
369 modules (Fig. 4A, B). Therefore, the changes in the position and orientation of baseplate arms cause
370 the iris-like structure formed by the wedge modules to expand in diameter from 104 to 122 Å (SFig.
371 7A, D). The expansion of the iris is enabled by the movement of the dimerization and trifurcation
372 domains relative to each other, which brings the six wedge modules to an identical conformation

(SFig. 7D, E, SMovie 5). In contrast, the interactions between the dimerization domains of wedge proteins remain the same as in the baseplate of the extended tail (SFig. 7F). A similar expansion was described for gp6 and gp7 during the transition of the T4 baseplate from its extended to contracted state[3]. The expansion of the diameter 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. 11A, 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 baseplate of the extended tail (SFig. 11G, H). The handshake domain of the tail sheath initiator protein inclines 17° (SFig. 11B, E) and, together with the wedge iris expansion, triggers the contraction of the tail sheath, as discussed below.

The tail of phi812 in extended conformation can bend

Contractile tails of phages differ in their length and mechanical properties. A flexible contractile tail has been observed only in the *Herelleviridae* phages phi812[23] and A511[24] with ~200-nm-long tails that can bend up to 90°, and in phage Milano[9] (genus *Schmittlotzvirus*, unclassified family) with ~125-nm-long tail. On the other hand, the ~100-nm-long tails of *Straboviridae* phages, represented by T4, are rigid[65]. The flexibility of the longer tails was speculated to be an adaptation to withstand mechanical forces without breaking[66]. To explain the ability of the phi812 tail to bend without breaking, we determined the structure of the extended phi812 tail to a resolution of 4.2 Å (Fig. 6F, SFig. 1, 2G and 3, 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. 6F).

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. 18A). Whereas the outer surface of the phi812 tail tube features roughly equal distribution of positively and negatively charged residues, the inner surface of the tail tube is negatively charged, probably to facilitate the passage of the phage DNA during ejection (SFig. 18B, C). The ability of phage Milano to bend was attributed to the missing loop in the β-barrel domain of tail tube protein[9]. The phage Milano tail tube straightening, after contraction, was speculated to be caused by the negative electrostatic potential of its outer surface[9]. However, since the tail tube of phi812 contains the loop between the β-strands 6 and 7 (SFig. 18D) and its outer surface charges are balanced (SFig. 18B), its bending and straightening are likely enabled by different mechanisms.

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. 19A). In an extended tail, the domains I and II of phi812 tail sheath proteins interact with tail tube proteins and mediate substantial contacts between tail sheath proteins (SFig. 19B, C). Domains III and IV are positioned successively next to domain II at the tail surface (SFig. 19B). 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. 19C). 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[3–9,28–32,67,68]. 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. 19A). The domain III contacts domain IV of the tail sheath protein from the disc located closer to the baseplate (SFig. 19D). Additionally, domain IV interacts with domain II of the neighboring tail sheath protein from the same disc (SFig. 19D). 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 (Fig. 19E).

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 extended 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[4,8,23,24,28]. The contracted tail sheath of phi812 reconstructed to a resolution of 3.2 Å broadens 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. 6G, SFig. 1, 2H and 3, 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 extended conformation (Fig. 6H) while maintaining the inter-subunit handshake interactions among domains I from successive discs (SFig. 19F, G). During contraction, the individual domains of the tail sheath protein of phi812, unlike those of T4 and some CISs[32,69,71], move relative to each other, which enables the formation of the more extensive contacts between the subunits (SFig. 19H). The buried surface area of intersubunit contacts per tail sheath monomer increases from ~5,500 $Å^2$ in the extended tail sheath to ~9,500 $Å^2$ in the contracted tail sheath. We speculate that the enlarged contacts stabilize the straight structure of the contracted tail sheath, which is less flexible than the extended one.

It has been shown for phages T4 and A511 and R-type diffocin that the contraction of the tail sheath is gradual and occurs as a wave along the tail axis[24,33,70,72]. 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. 16). Furthermore, the cryo-EM reconstructions of baseplate-proximal parts of the tail sheath in the extended and contracted state provide structures of discrete intermediates of contraction (SFig. 1, 2B, E, and 3, SMovie 5). The first three baseplate-proximal tail sheath discs of the extended phi812 tail sheath differ in conformation from those in the central part of the extended tail sheath (SFig. 20A, B). The deviation occurs gradually, with the first tail sheath disc being the most affected (SFig. 20B). Similar deviations of baseplate-proximal tail sheath proteins were observed in the R2 pyocin and T6SS[28,73]. 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 tail sheath 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. 20C, D). Similar to the extended tail sheath, the first contracted sheath disc is the most affected and the sixth one the least affected (SFig. 20D). 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. 20D). The absence of a twist difference may be due to the more extensive contacts between consecutive tail sheath discs than in the extended 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.

461

462 **Changes in the baseplate structure required for cell wall penetration**

463 Phages with contractile tails infecting Gram-positive bacteria employ enzymes that degrade teichoic
464 acid and peptidoglycan to facilitate tail tube penetration through the cell wall[18,75]. Phage phi812 may
465 use the petal domain of the central spike protein to cleave teichoic acid (Fig. 3B, C); however, the
466 central spike protein needs to be released from the tail channel to enable the genome ejection. We
467 speculate that the conformational change to inner tripods, which disrupts their interaction with weld
468 proteins, triggers the release of the weld and central spike proteins from the baseplate. Dissociation
469 of the weld and central spike proteins is required to enable the cleaver domains of the hub protein to
470 interact with and degrade the cell wall peptidoglycan (Fig. 4D, E, SFig. 10). This indicates that the
471 weld and central spike proteins of phi812 are released before the cell wall penetration by the tail
472 tube. In contrast, the central spike proteins of phages T4 and E217, which infect Gram-negative
473 bacteria, are released from baseplates after puncturing the host outer membrane[4,76,77].
474 Relative rotation of consecutive tail sheath discs in extended phi812 tail is 21.2°, whereas
475 that of the contracted tail sheath is 30.6°. Since the tail contains 51 tail sheath discs, as the phi812
476 tail sheath contracts, the tip of the tail tube is pushed through the teichoic acid and peptidoglycan
477 layers while rotating about its axis by 475° (see Methods). The rotation of the hub proteins at the tip
478 of the tail tube can enhance their peptidoglycan degradation activity by bringing them into contact
479 with more substrate. Once the peptidoglycan layer is breached, the clamp domains of the hub
480 proteins can interact with the *S. aureus* cytoplasmic membrane. It was shown that after the
481 penetration of the outer membrane and peptidoglycan layer, gp27 from T4 (a homolog of the hub
482 protein from phi812) remains associated with the tip of the tail tube and probably interacts with
483 ejected tape measure protein to form a pore through the inner membrane of *E. coli*[78]. However, the
484 tail tube of phi812 penetrates 10-30 nm into the *S. aureus* cytoplasm, depending on the thickness of
485 the cell wall, which varies between 50-70 nm[23]. Therefore, unlike in T4, the hub proteins of phi812
486 are probably not required to form a pore in the *S. aureus* membrane, and their role in genome
487 delivery may be limited to facilitation of the penetration of the cell wall and cytoplasmic membrane
488 by the tail tube tip. Whether the hub and other baseplate core proteins remain associated with the
489 tip of the tail tube after the membrane penetration remains unknown.
490

491 **Conclusions**

492 In this study, we employed a combination of cryo-EM, X-ray crystallography, NMR, and structure
493 prediction methods to elucidate the intricate mechanisms employed by phage phi812 to penetrate
494 the *S. aureus* cell wall in preparation for transferring its genetic material into the host cytoplasm.
495 While most of the previously structurally characterized phages with contractile tails target
496 Gram-negative bacteria, our focus has been on phi812 attacking *S. aureus*, a representative of
497 Gram-positive bacteria, which are comparatively less explored. We show that the baseplate of the
498 phage phi812 comprises a core, wedge modules, and baseplate arms that carry RBP1, RBP2, and
499 tripod complexes (Fig. 1A-C, SMovies 1, 2). Upon binding to the *S. aureus* cell wall, the baseplate
500 symmetry transitions from threefold to sixfold, which is enabled by a series of conformational
501 changes that propagate from the baseplate periphery to the core (Fig. 6A-C, SMovies 3, 4). Based on
502 our results, we propose a model of phi812 cell adhesion and penetration (Fig. 7). During the

attachment, the receptor-binding proteins reorient and tripod complexes undergo conformational changes to enable host cell surface binding (Fig. 7A, B). Subsequently, the central spike proteins degrade cell wall teichoic acids (Fig. 7A, B). 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 peptidoglycan (Fig. 7C, D) and facilitate penetration of the tail tube through the cell membrane (Fig. 7D, E). Changes in the baseplate arms' positions relayed through wedge modules to the hexamer of tail sheath initiator proteins cause an expansion of its diameter, triggering a tail sheath contraction (Fig. 7C-E). The baseplate-proximal tail sheath discs of the extended phage particle have an altered structure, which may lower the energy barrier required to initiate the tail sheath contraction. The tail sheath shortens from 200 to 96 nm upon contraction[23], pushing the tail tube's tip 10-30 nm into the cytoplasm, depending on the local thickness of the *S. aureus* cell wall. The molecular characterization of the phi812 baseplate structure and the mechanism of cell wall binding open up the possibility of a rational design of phi812 to target specific strains of *S. aureus*.

Figures

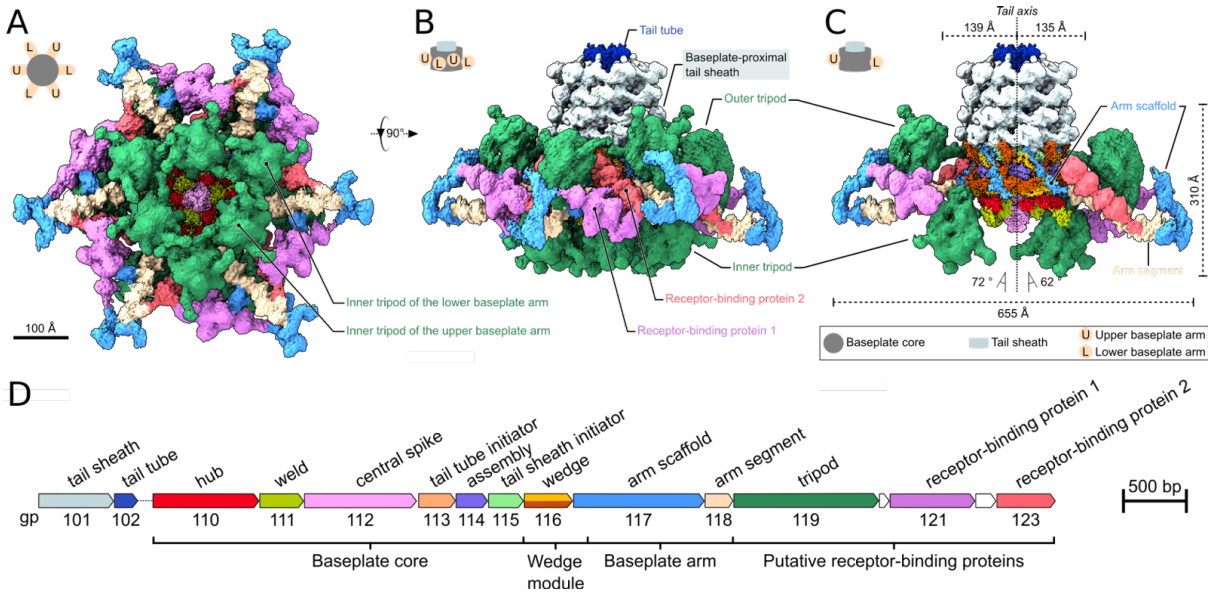


Fig. 1. Structure of baseplate of phi812 with extended tail. (A-C) The surface representation of the cryo-EM map of the phi812 baseplate is viewed along (A) and perpendicular to (B, C) the tail axis. Protein components of the baseplate are colored according to the scheme depicted in panel D. Selected baseplate components are indicated. Pictograms on the upper left indicate upper and lower baseplate arms that are showed and the legend in showed at the bottom of the panel C. **(C)** Side view of phi812 baseplate after segmentation, with two opposing baseplate arm showed to display the wedge modules and baseplate core. The height and width of the baseplate, tail axis, angles of baseplate arms relative to the tail axis, and distances of baseplate arms from the tail axis are indicated. **(D)** Scheme of part of the phi812 genome encoding tail and baseplate genes.

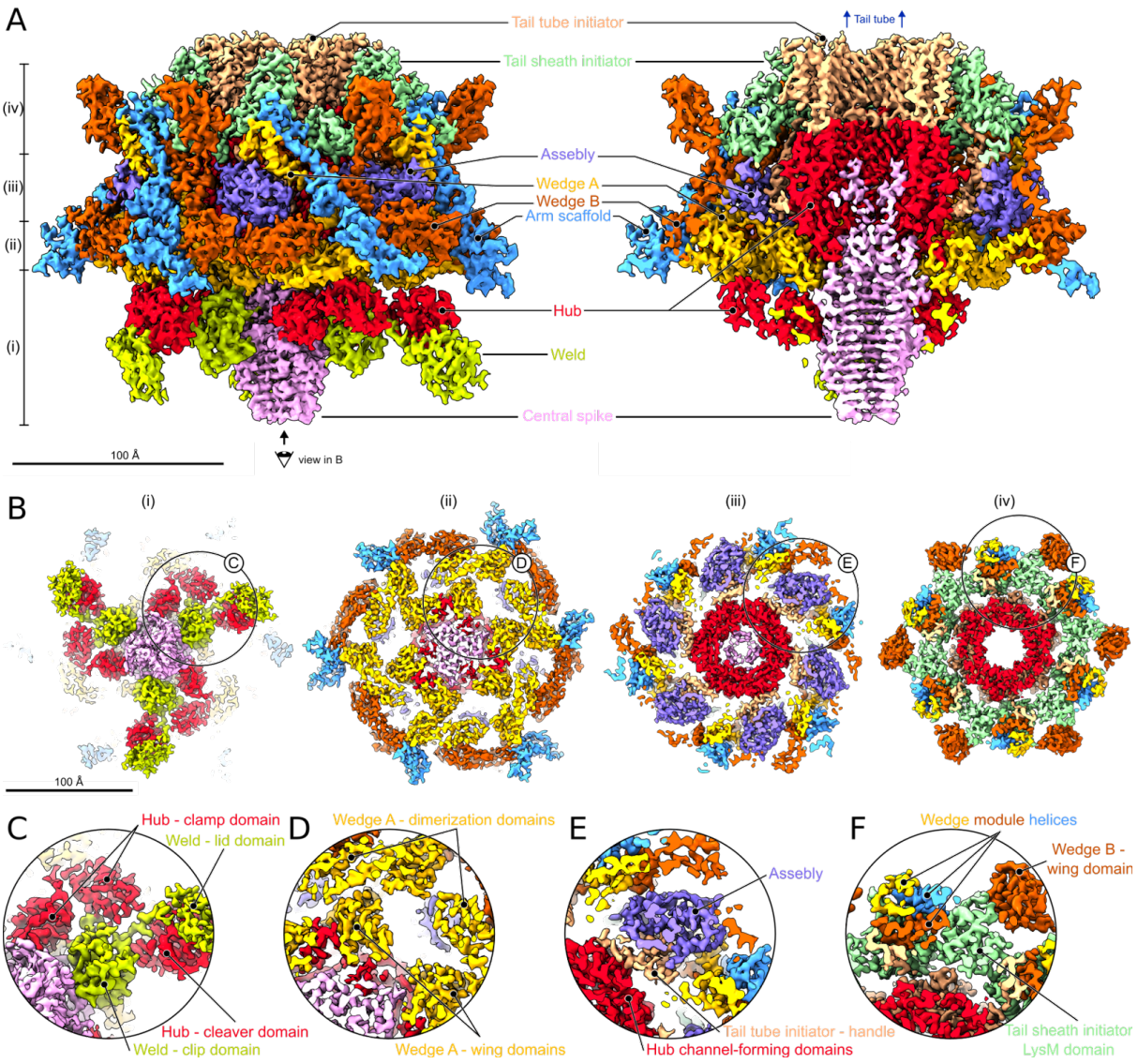


Fig. 2. Structure of wedge modules and baseplate core of phi812 with extended tail. (A-F) The surface representation of the composite cryo-EM map showing a complex of wedge module and baseplate core proteins of phi812 baseplate colored according to the Fig. 1D. Selected proteins and protein domains are indicated. The position of the tail tube is indicated. **(A)** Side views of the protein complex showed as whole (left) and clipped (right). The positions of sections shown in panel (B) are indicated on the left. Names of protein components are indicated. **(B)** Sections through the complex of wedge module and baseplate core proteins viewed from the bottom of the baseplate. Close views of selected region displayed in panels C-F are indicated. **(C-F)** Close views of selected regions from sections in the panel B. Names of proteins and protein domains are indicated.

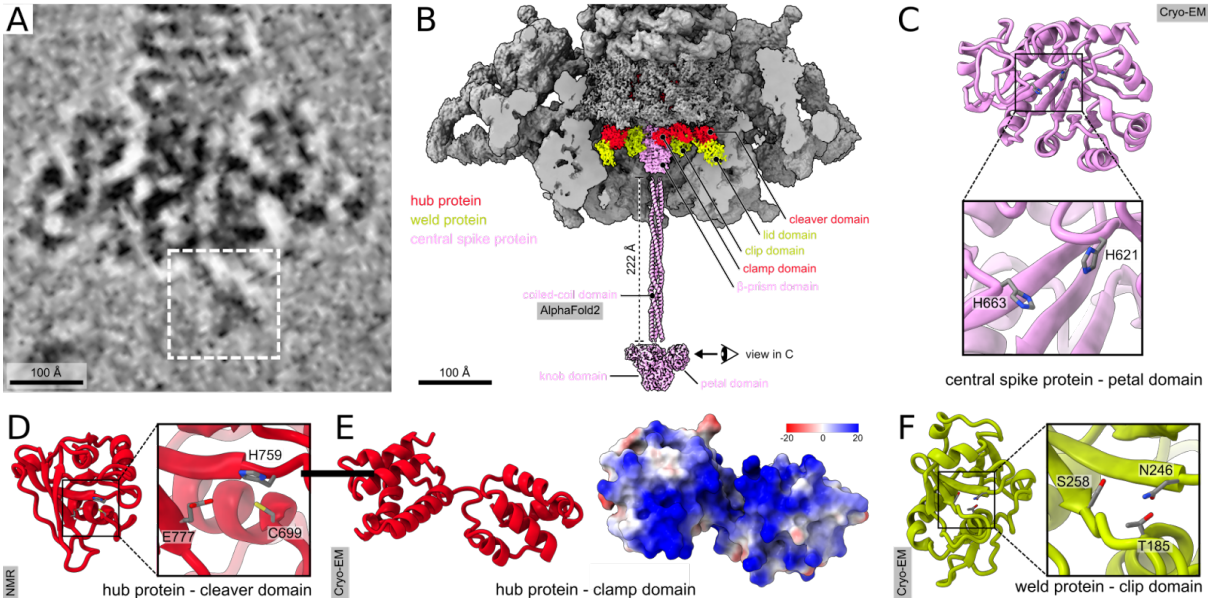


Fig. 3. Central spike, hub, and weld proteins. (A) Cryo-electron micrograph of the phi812 baseplate. The white dashed rectangle presumably contains part of the coiled-coil domain and knob and petal domains of the trimer of central spike proteins. **(B)** Surface representations of the cryo-EM reconstruction of the baseplate of phi812 with extended tail (top) and the cryo-EM reconstruction of the knob and petal domains of central spike protein trimer. Map of the baseplate is colored in gray except for central spike, hub, and weld proteins colored according to the legend on the left. The front part of the baseplate has been removed. Structure of the coiled-coil domain of central spike protein trimer obtained from AlphaFold2 prediction showed in cartoon indicates the length of central spike protein, which corresponds to the observation in panel A. Names of proteins and protein domains are indicated. **(C,D)** Cartoon representations of the petal domain of the central spike protein (C) and the cleaver domain of the hub protein (D). Side chains of residues forming putative hydrolytic sites are shown as sticks. Insets show details of the putative hydrolytic sites. **(E)** Cartoon (left) and surface (right) representation of clamp domain of the hub protein. Molecular surface is colored according to Coulombic electrostatic potential in the range from -20 to 20 kcal/(mol·*e*) at 298 K. **(F)** Cartoon representation of the clip domain of the weld protein. Side chains of residues forming a putative receptor-binding site shown as sticks. In panels C-F, showed structures were obtained either by cryo-EM or NMR, as indicated.

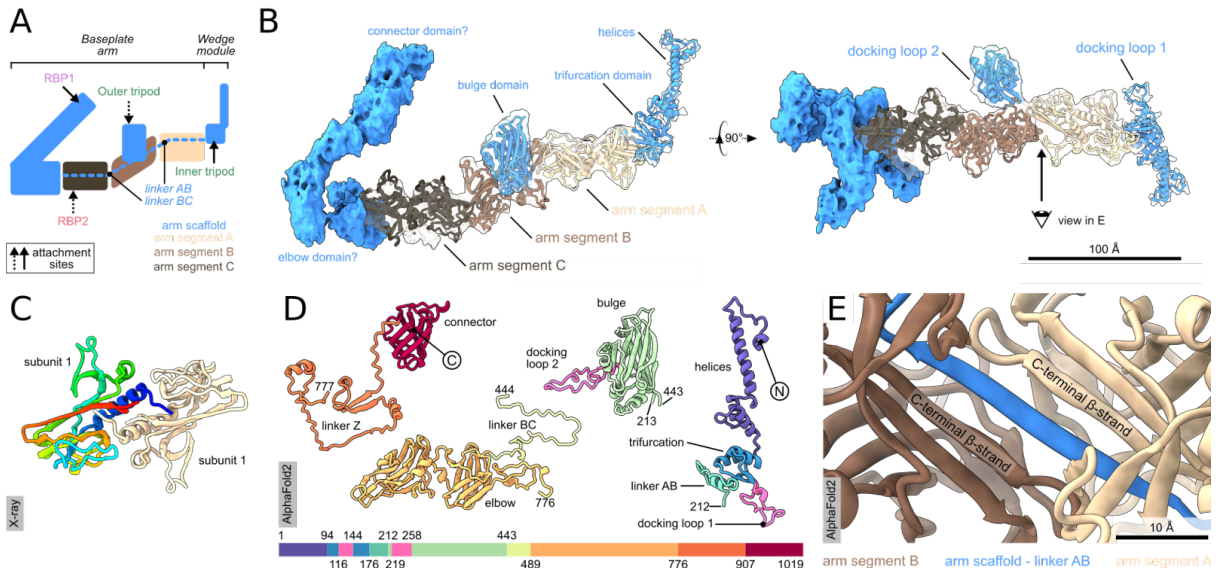


Fig. 4. Structure of baseplate arm phi812 with the extended tail. (A) Schematic representation of the baseplate arm and part of the wedge module. Attachment sites of tripods, RBP1 and RBP2 are indicated. **(B)** Surface representation of the cryo-EM map corresponding to the scheme at the bottom of the panel A. Distal part of the baseplate arm, probably corresponding to elbow and connector domains, is shown as blue opaque surface. Parts of the baseplate arm with fitted structures are shown as white semi-transparent surface. Arm segment proteins and N-terminal helices, trifurcation domain, bulge domain and docking loops of the arm scaffold protein are shown in cartoon representation. Approximate view in the panel E is indicated on the right. **(C)** Cartoon representation of the arm segment protein dimer obtained by X-ray crystallography, with one of the subunits colored in rainbow. **(D)** Cartoon representation of the arm scaffold protein structure predicted by AlphaFold2 colored according to domains. The structure was divided into several parts for better visualization. Domain names and chain termini are indicated. Linear diagram (bottom) representing arm scaffold protein domains is colored accordingly. **(E)** Interaction of C-terminal β-strands of two arm segment subunits with linker AB of the arm scaffold protein obtained by AlphaFold2 prediction and shown in cartoon.

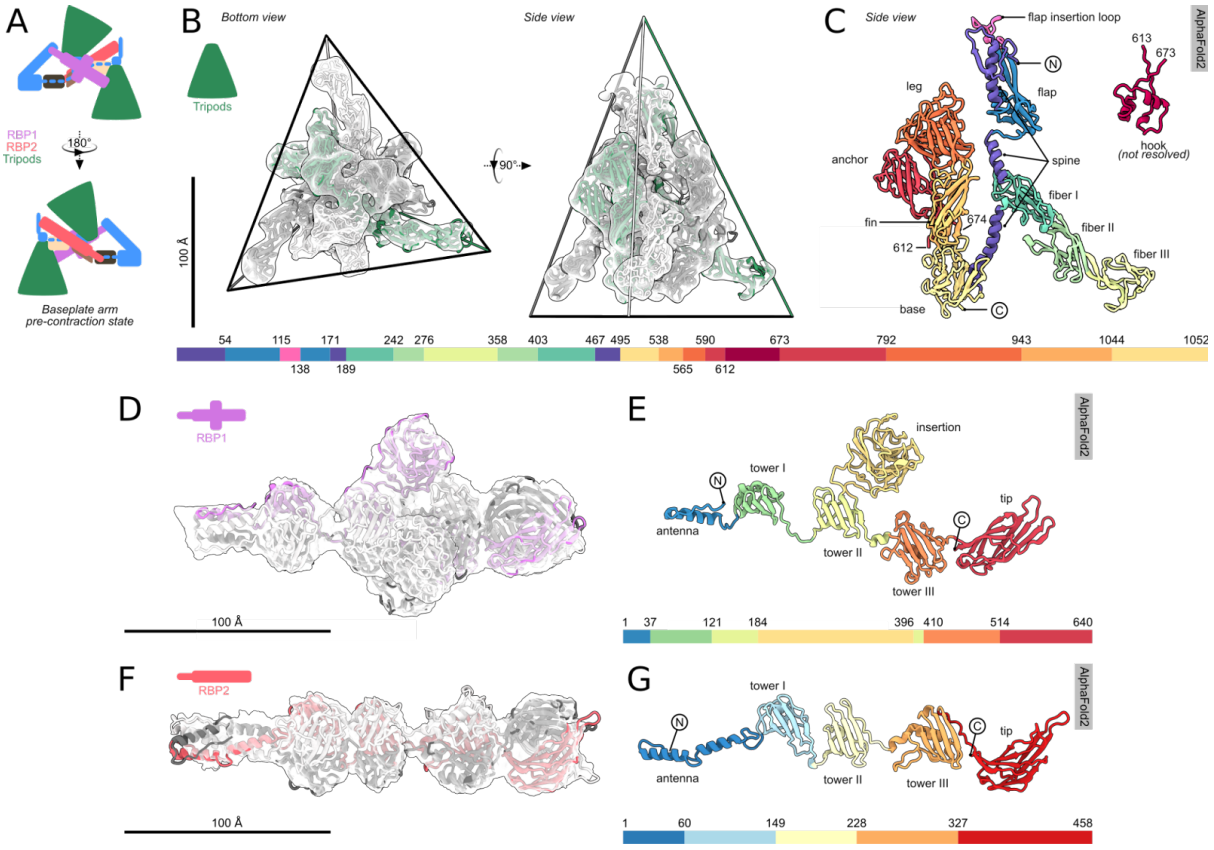


Fig. 5. Tripod and receptor-binding proteins. (A) Schematic representation of the baseplate arm and part of the wedge module of phi812 with the extended tail. **(B)** Cryo-EM map corresponding to the inner tripod complex segmented out from the reconstruction of the whole baseplate showed as white semi-transparent surface with the structure of the tripod protein trimer obtained by AlphaFold2 prediction showed in cartoon representation and colored by subunits. Only the domains of tripod protein for which a density was resolved are shown. Triangular cage highlights the shape of the tripod protein complex **(C)** Structure of the tripod protein subunit from panel B obtained by AlphaFold2 prediction is showed in cartoon representation colored by domains. Domain names and chain termini are indicated. Linear diagram the the bottom of panels B and C represents tripod protein domains and is colored accordingly. **(D)** Cryo-EM map corresponding to the RBP1 segmented out from the reconstruction of the whole baseplate showed as white semi-transparent surface with the structure of the RBP1 trimer obtained by AlphaFold2 prediction showed in cartoon representation and colored by subunits. **(E)** Structure of the RBP1 subunit obtained by AlphaFold2 prediction is showed in cartoon representation colored by domains. **(F)** Cryo-EM map corresponding to the RBP2 segmented out from the reconstruction of the whole baseplate showed as white semi-transparent surface with the structure of the RBP2 trimer obtained by AlphaFold2 prediction showed in cartoon representation and colored by subunits. **(G)** Structure of the RBP2 subunit obtained by AlphaFold2 prediction is showed in cartoon representation colored by domains. In panels E and G, the domain names and chain termini are indicated. Linear diagrams at the bottom of panels E and G represent domains of corresponding proteins and colored accordingly.

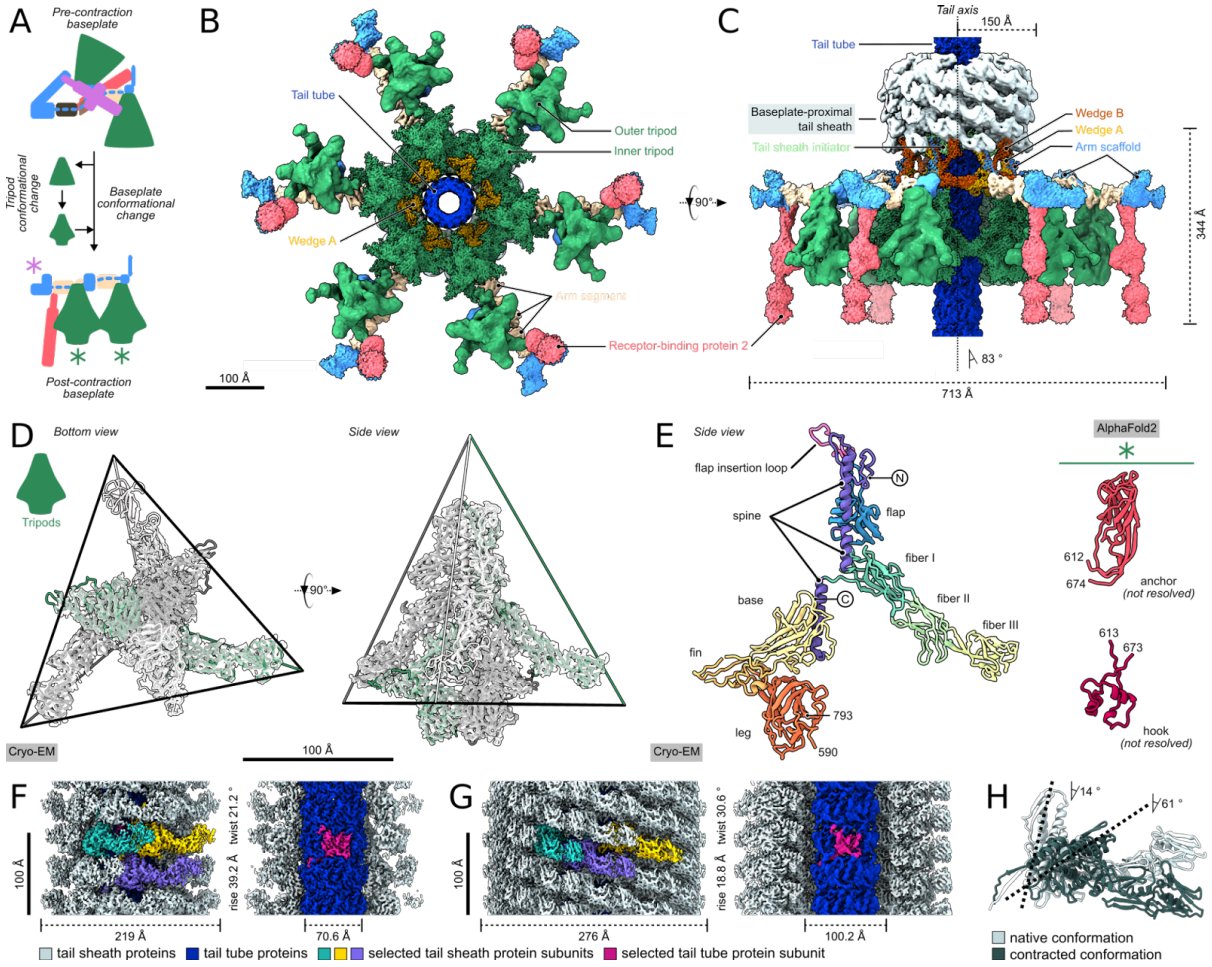


Fig. 6. Conformation of baseplate, tripods and tail sheath of phi812 after the tail sheath contraction. (A) Schematic representation of conformational change to the baseplate arm and tripod complexes of phi812. Pink asterisk indicates the putative attachment site of RBP1, which was not resolved in the reconstruction. Green asterisks indicate putative position of anchor and hook domains, which were not resolved in the reconstruction. **(B, C)** The surface representation of the cryo-EM map of the phi812 with contracted tail viewed along (B) and perpendicular to (C) the tail axis. Protein components are colored according to the Fig. 1D. Selected baseplate components are indicated. In panel C, the height and width of the baseplate, tail axis, angle of baseplate arms relative to the tail axis, and distance of baseplate arms from the tail axis are indicated. **(D)** Cartoon representation of the tripod protein trimer colored by subunits in the cryo-EM map corresponding to the inner tripod complex segmented out from the reconstruction of the baseplate of phi812 with contracted tail showed as white semi-transparent surface. Only tripod protein domains for which density was resolved are shown. Triangular cage highlights the shape of the tripod protein complex. **(E)** Structure of the tripod protein subunit showed in cartoon representation colored by domains. Tripod protein domains on the left were resolved in the cryo-EM reconstruction showed in panel D. The anchor and hook domains of the tripod protein were obtained from AlphaFold2 prediction and their putative position is indicated in the panel A by green asterisks. Domain names and chain termini are indicated. **(F, G)** Surface representation of cryo-EM map of the extended tail (F) and composite cryo-EM map of the contracted tail (G) 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

619 colored according to the legend below the panel. **(H)** Comparison of the structures of tail sheath
620 proteins in extended and contracted conformations. The structures are shown in cartoon
621 representation, and their domains II were superimposed. Angles of the long axes of domain I relative
622 to the tail axis are indicated for each conformation.

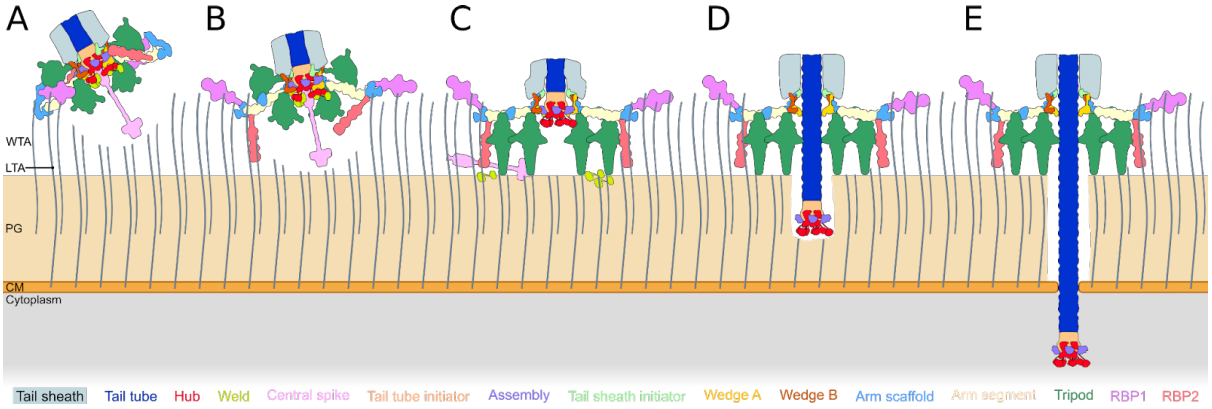
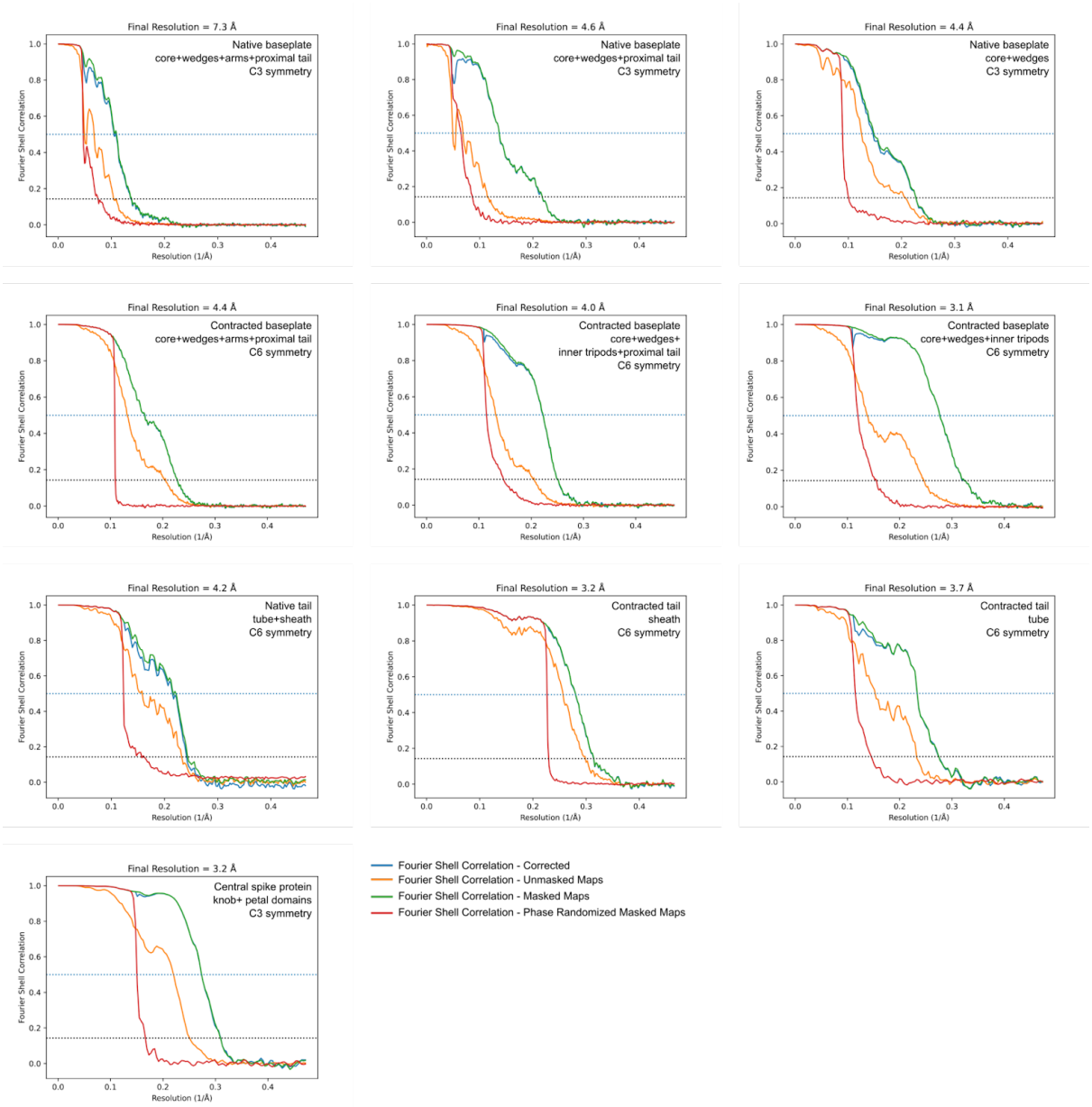
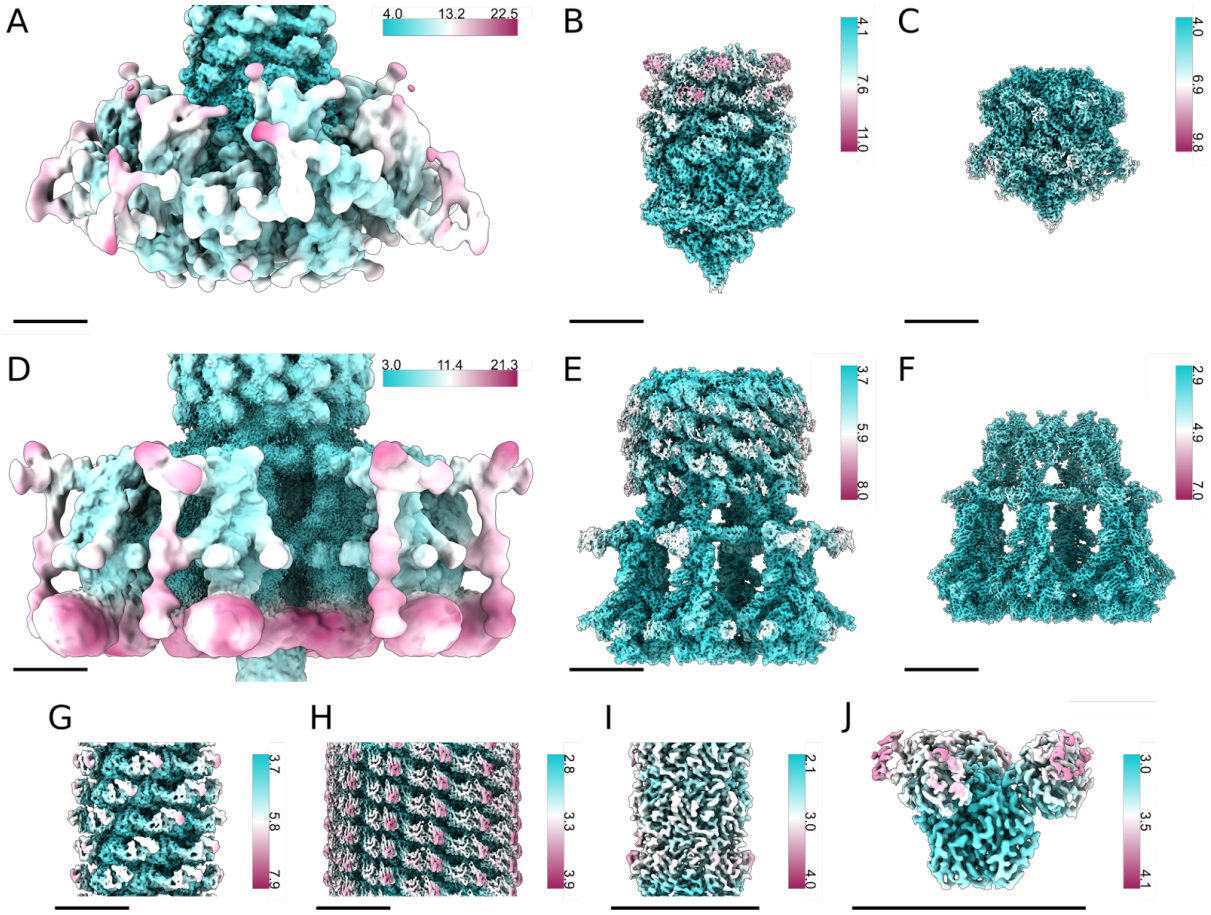


Fig. 7. Schematic model of phi812 cell adhesion and penetration mechanism. (A) Baseplate of phage phi812 comes in contact with the cell wall of *S. aureus*, trimers of RBP1 bind to teichoic acid and the petal domains of the central spike proteins cleave teichoic acid below the baseplate. **(B)** Trimers 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. WTA - wall teichoic acid; LTA - lipoteichoic acid; PG - peptidoglycan; CM - cytoplasmic membrane. The scheme is colored according to the list of proteins at the bottom.

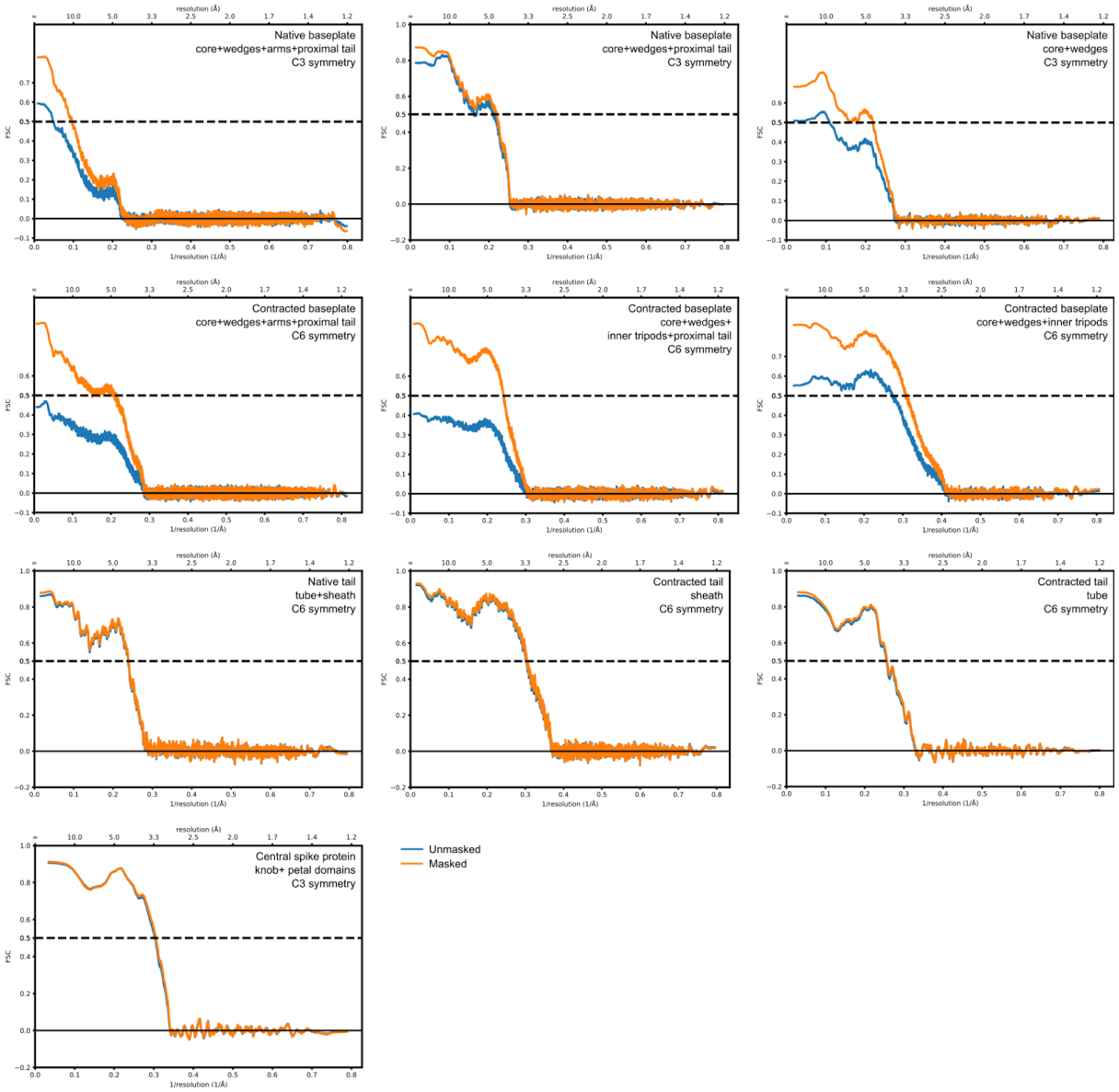
Supplementary figures



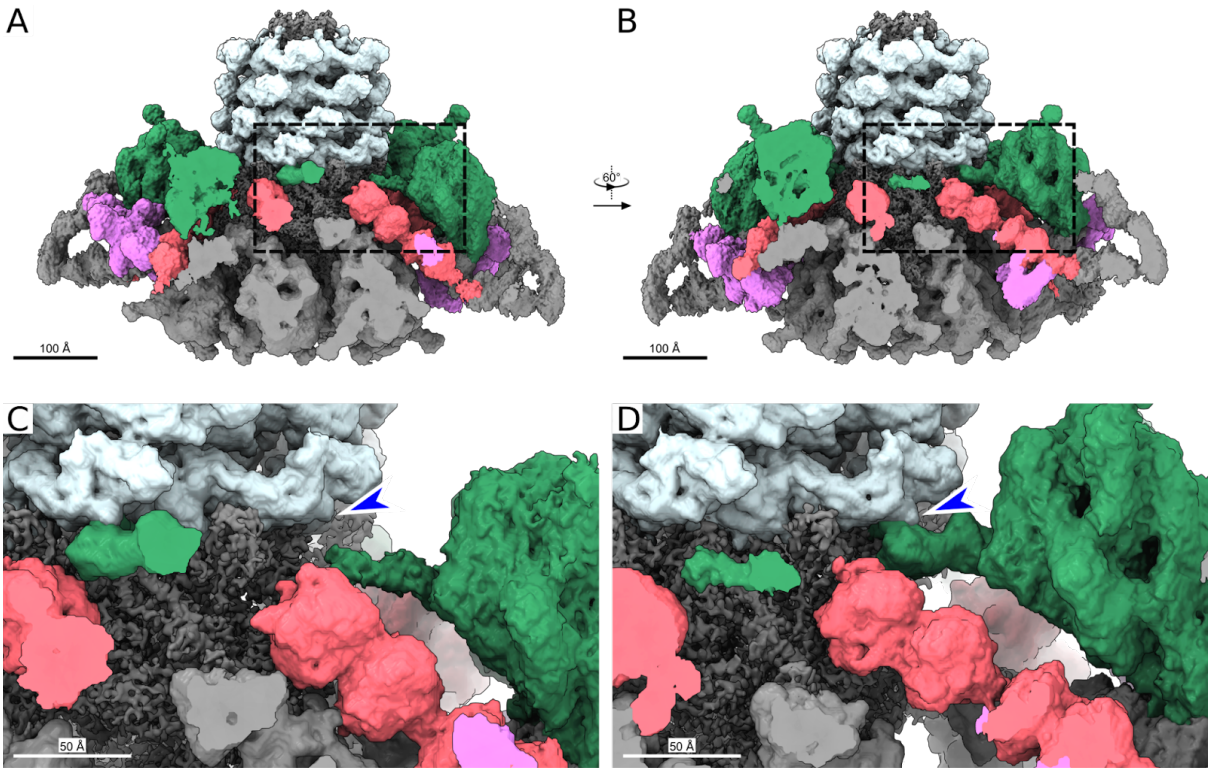
SFig. 1. FSC curves of cryo-EM reconstructions. Each panel shows a comparison of Fourier shell correlation curves color-coded according to the legend. The FSC threshold levels are indicated by a blue dotted line (threshold 0.5) and a black dotted line (threshold 0.143).



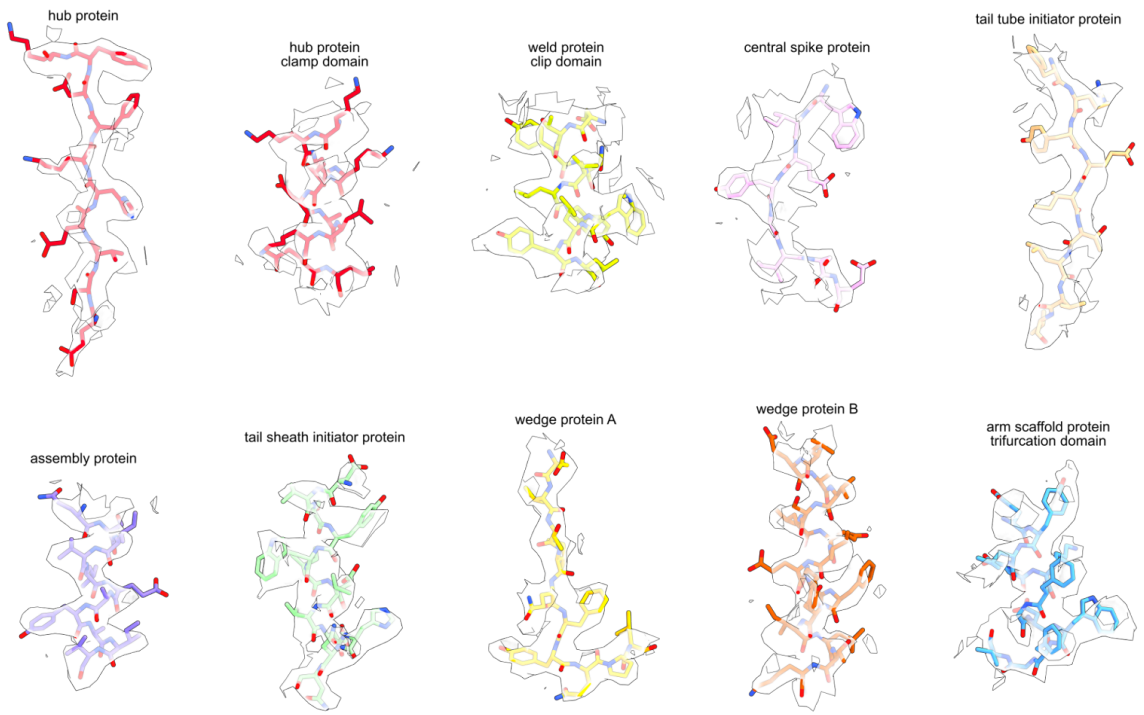
SFig. 2. Local resolution maps of cryo-EM reconstructions. (A-J) Surface representations of cryo-EM reconstructions colored according to the local resolution in Å. Scalebars = 100 Å. **(A)** Baseplate of phi812 with extended tail. **(B)** Core, wedge modules, and three tail sheath discs of the baseplate from phi812 with extended tail. **(C)** Core and wedge modules of the baseplate from phi812 with extended tail. **(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)** Tail sheath and tail tube from phi812 with extended tail. **(H)** Contracted tail sheath. **(I)** Tail tube from the contracted tail. **(J)** Knob and petal domains of the central spike protein.



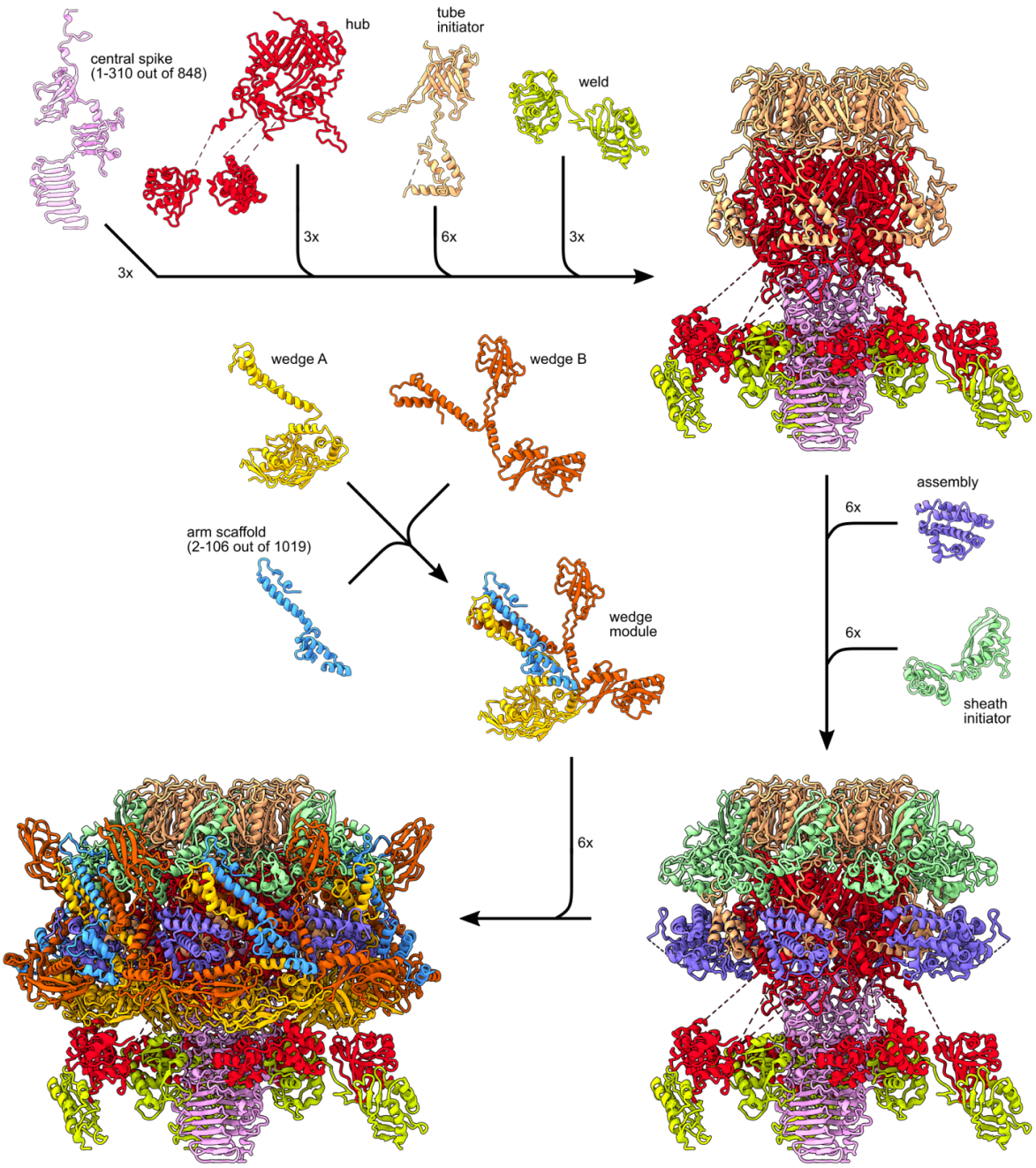
SFig. 3. Map to model FSC curves of cryo-EM reconstructions. Each panel shows a comparison of Fourier shell correlation curves color-coded according to the legend. The differences between the masked and unmasked FSC curves in some of the structures are due to the presence of unmodelled densities. The FSC threshold levels are indicated by a black dotted line (threshold 0.5) and a black line (threshold 0).



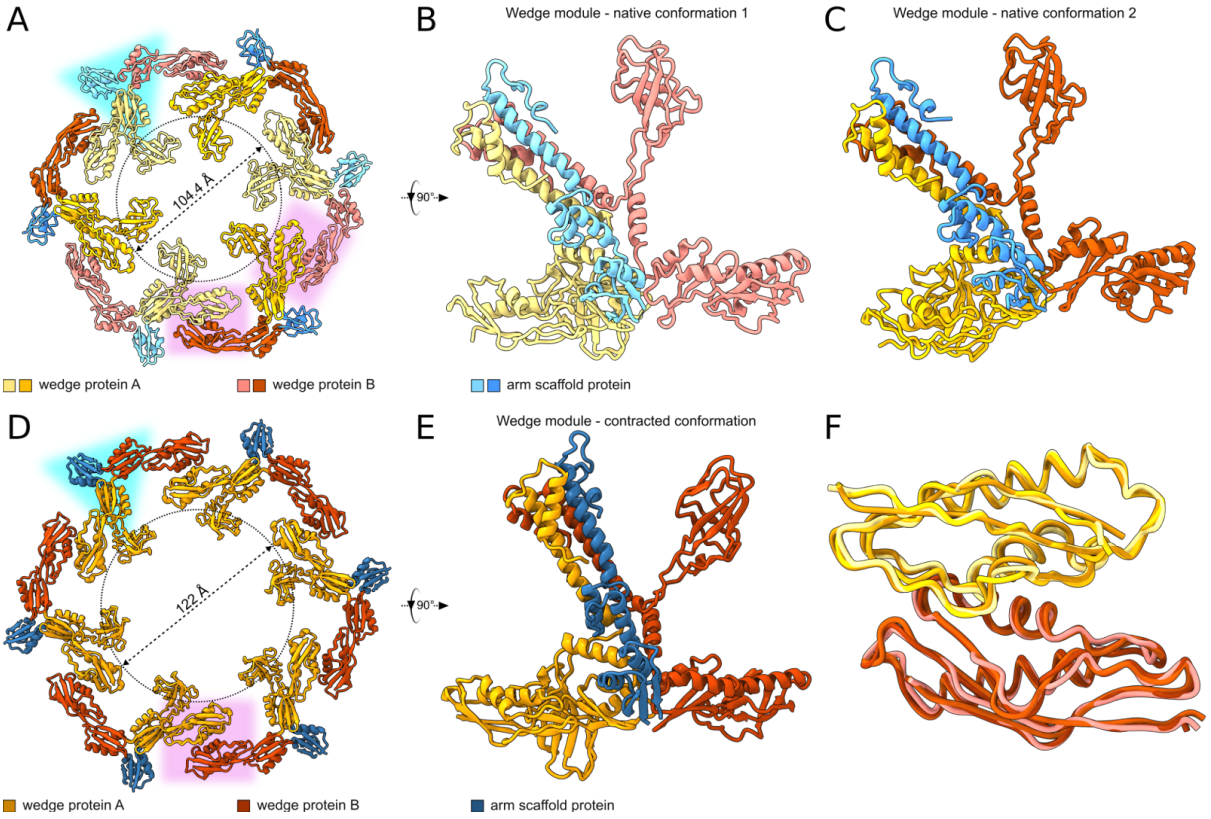
SFig. 4. Comparison of interactions of tripod proteins, RBP1, and RBP2 attached to upper and lower baseplate arms with baseplate core and tail sheath. Surface representation of a composite cryo-EM map of the baseplate of phi812 with extended tail. All proteins are colored gray except the tail sheath 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.



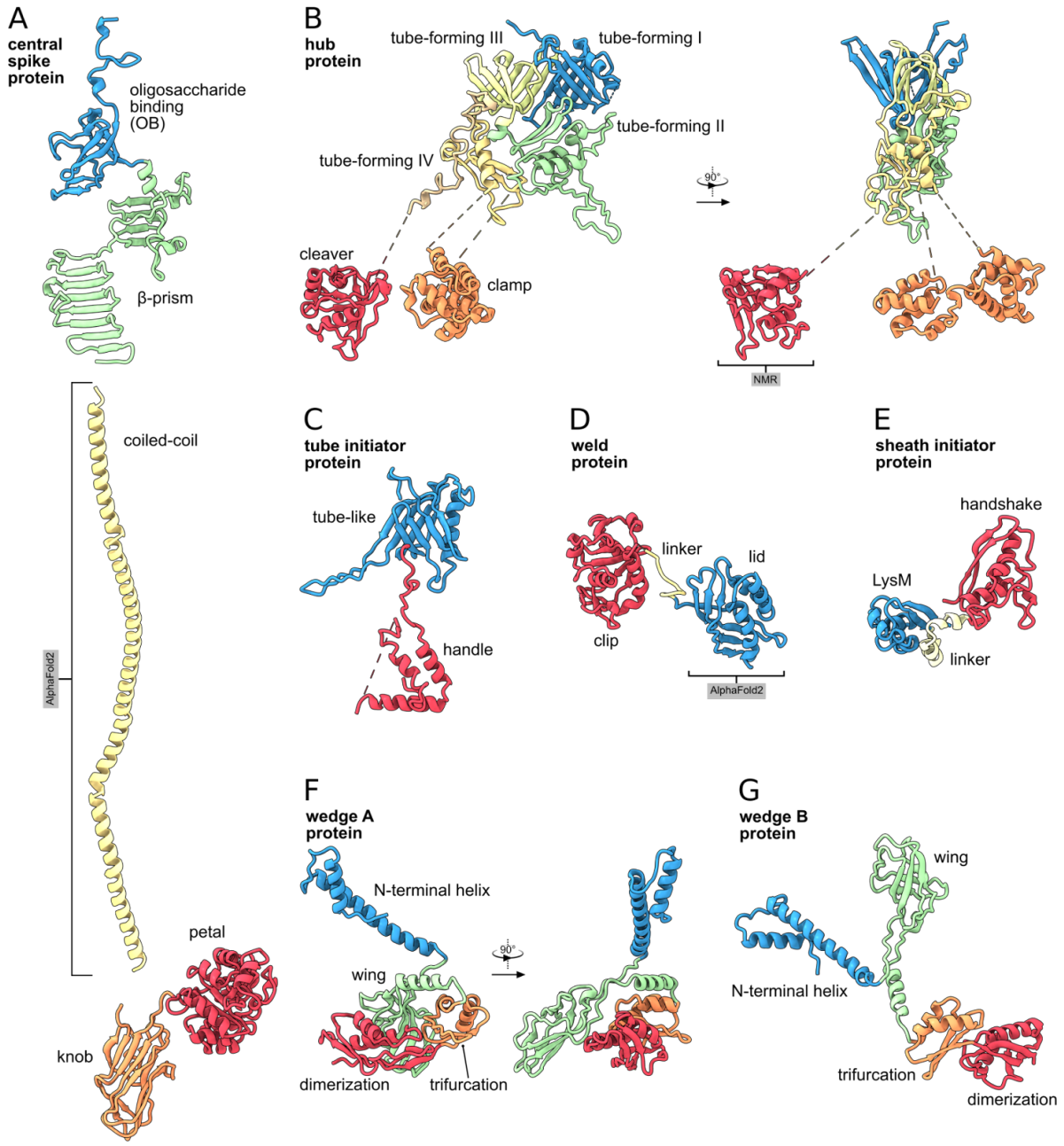
SFig. 5. Structures of the baseplate core and wedge module proteins. Representative segments of protein structures showed in stick representation and the corresponding segment of cryo-EM map from focused reconstruction of the baseplate core and wedge modules showed as white semi-transparent surface.



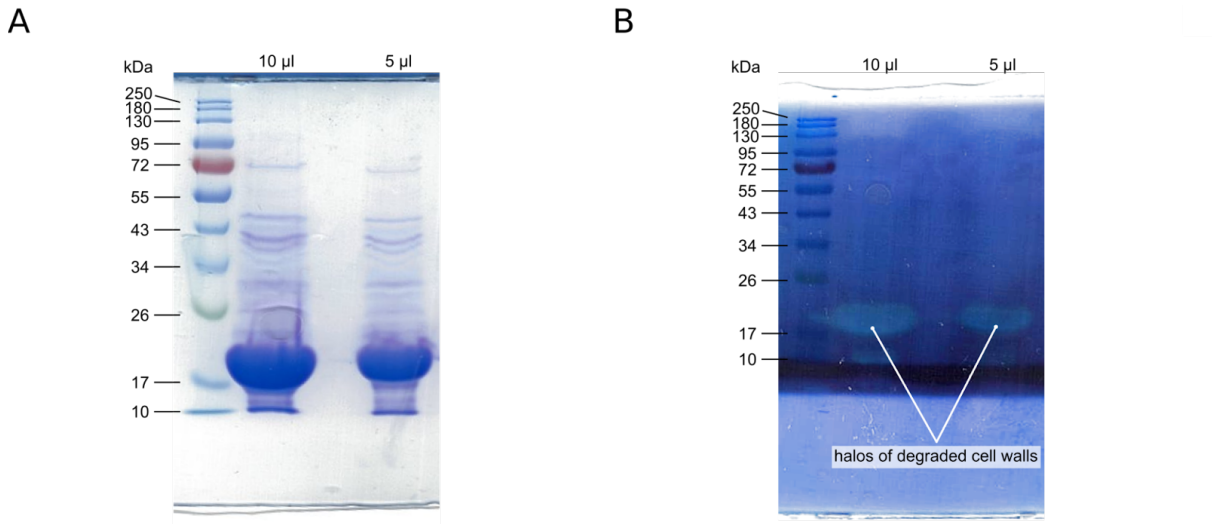
SFig. 6. Components of phi812 baseplate core and wedge modules. The baseplate core and wedge module proteins are shown in cartoon representation. Protein names and the numbers of subunits of each protein forming the baseplate core and wedge modules are indicated.



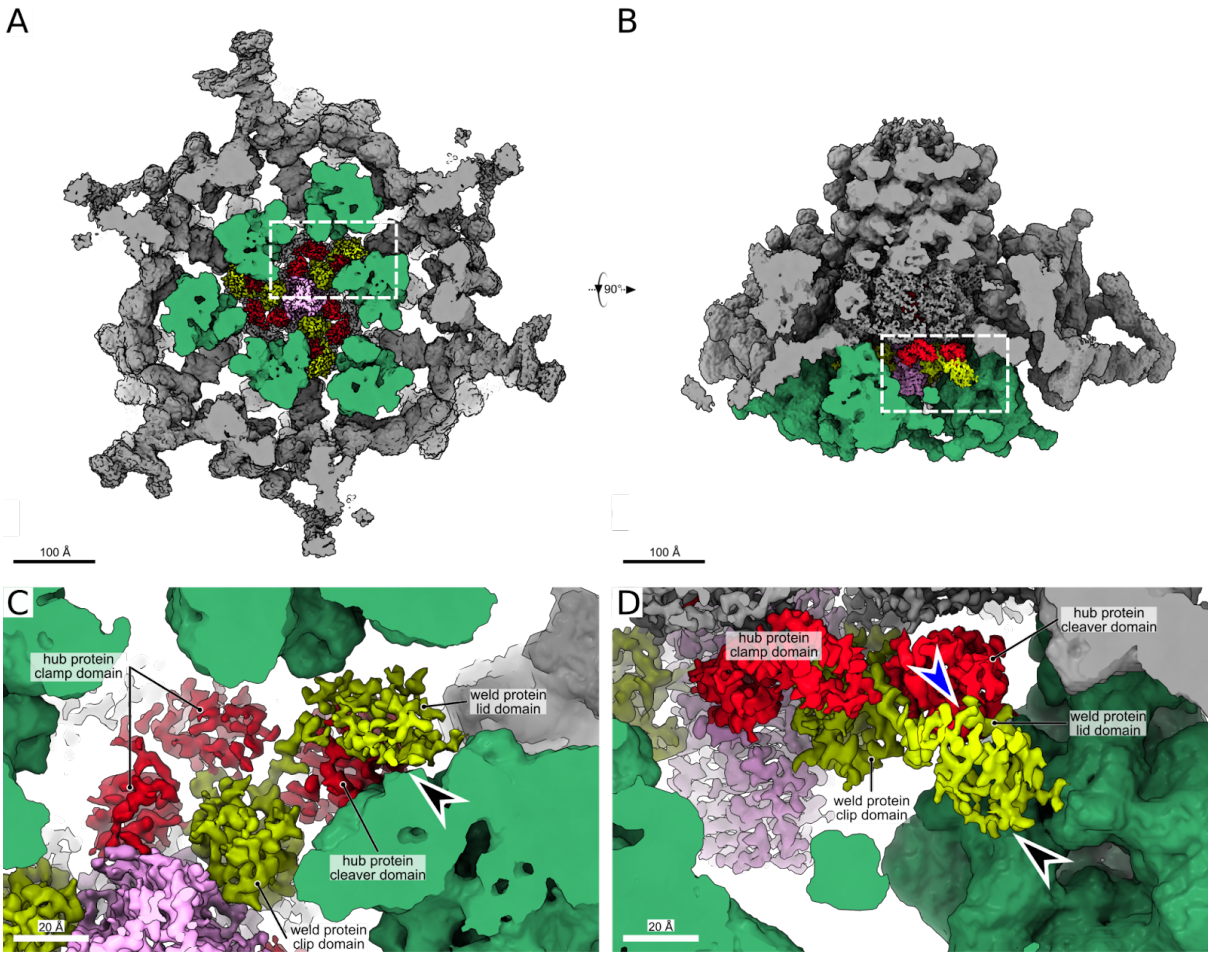
SFig. 7. Expansion of iris formed by wedge modules upon phi812 tail contraction. (A-C) Cartoon representation of wedge modules from the baseplate of phi812 with extended tail is colored according to the legend at the bottom. (A) Six wedge modules forming the iris are viewed along the tail axis towards 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.



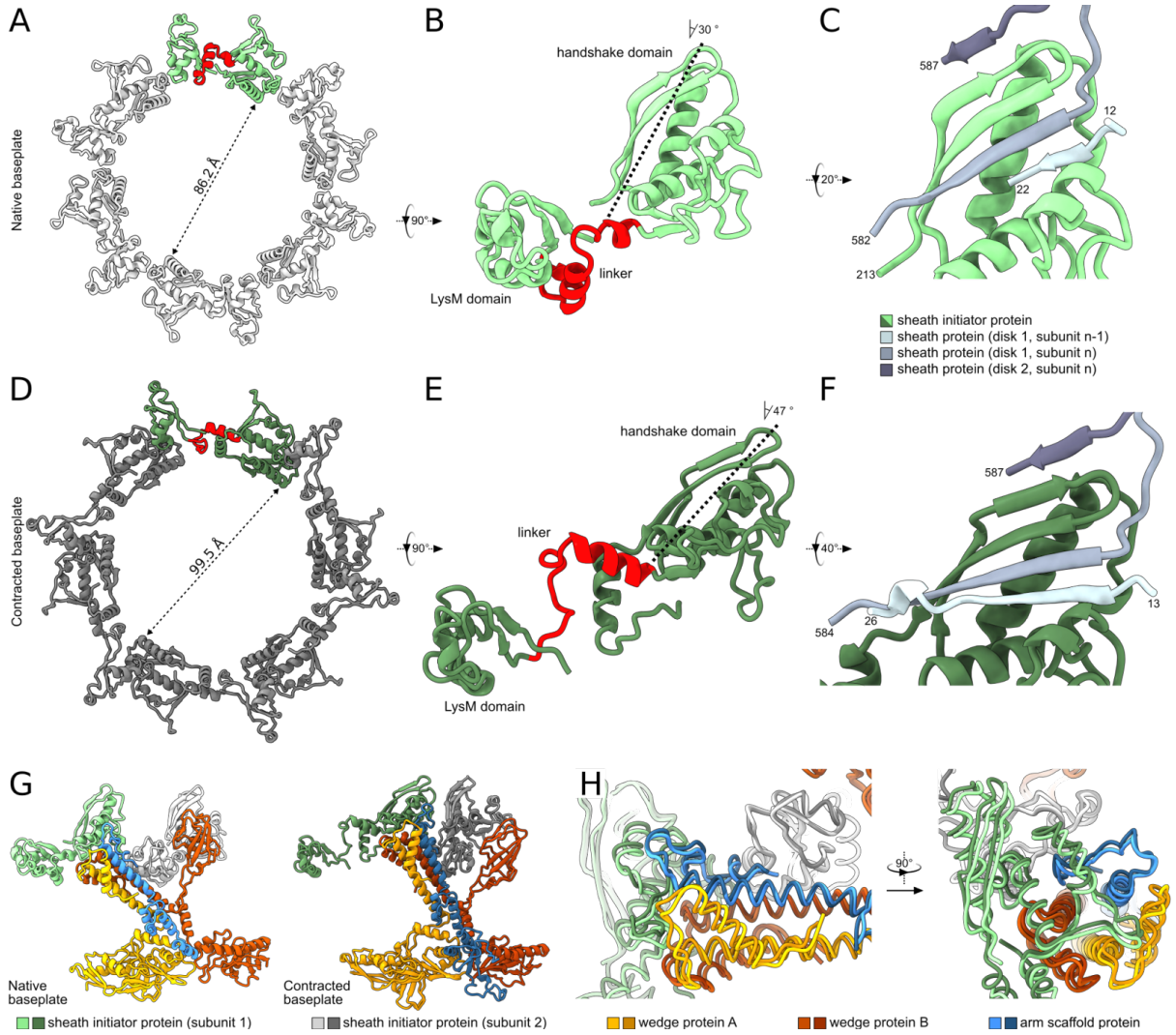
SFig. 8. Domain organization of proteins forming baseplate core, wedge modules, and tail of phi812. (A-G) The proteins are shown in cartoon representation and colored according to domains and functional elements. The names of proteins are in bold. Structures were determined by cryo-EM except the coiled-coil domain of the central spike protein (AlphaFold2), cleaver domain of the hub protein (NMR), and lid domain of the weld protein (AlphaFold2), as indicated.



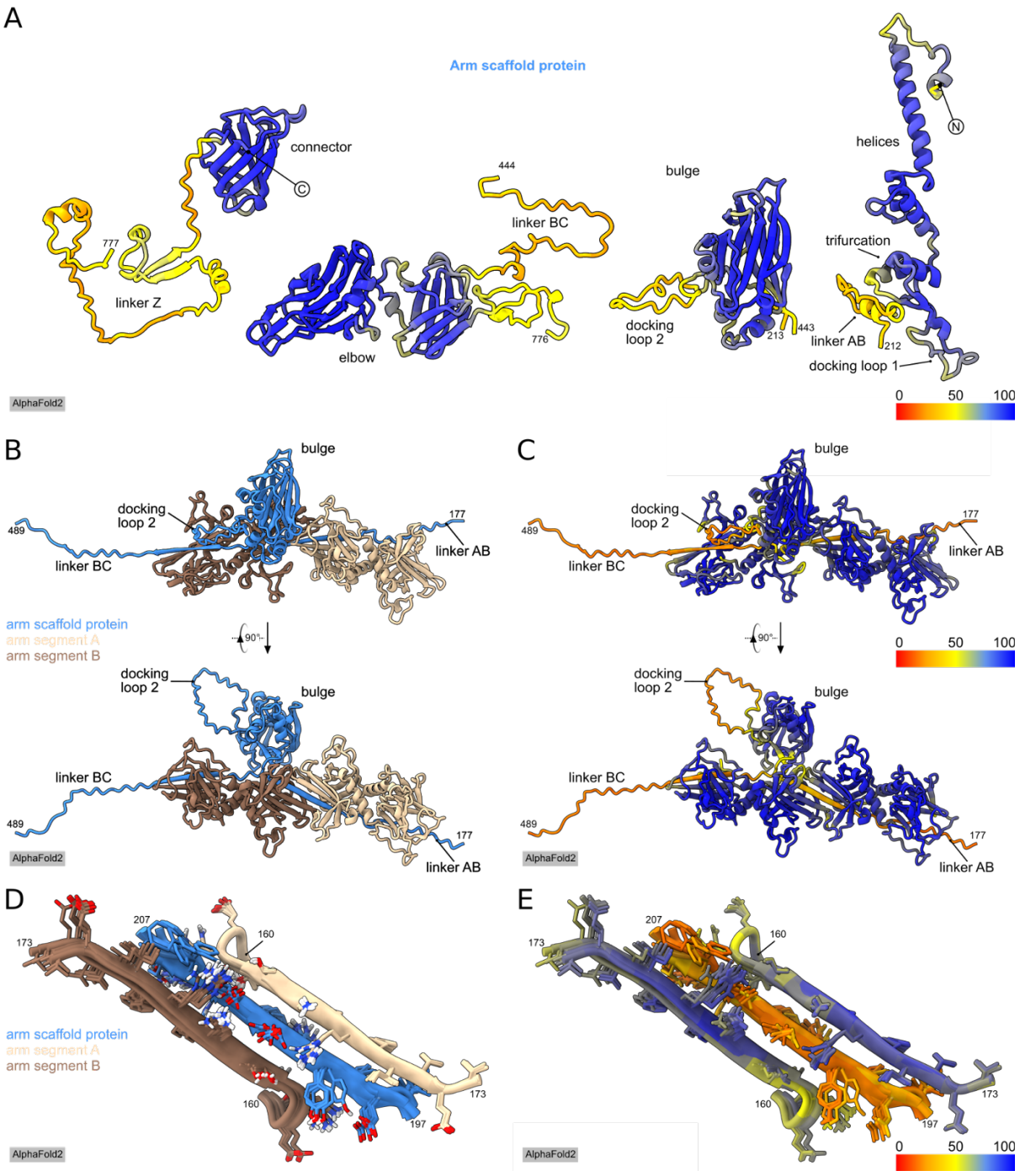
SFig. 9. SDS-PAGE and zymogram analysis of cleaver domain of hub protein. Scan of an SDS-PAGE gel **(A)** and zymogram **(B)** of recombinantly expressed cleaver domain of the hub protein after Histrap purification. As indicated, ten and five µl of the purified protein sample concentrated to 10 mg/µl were loaded. Halos on the zymogram resulting from cell wall degradation activity of cleaver domain of the hub protein are indicated.



SFig. 10. 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.

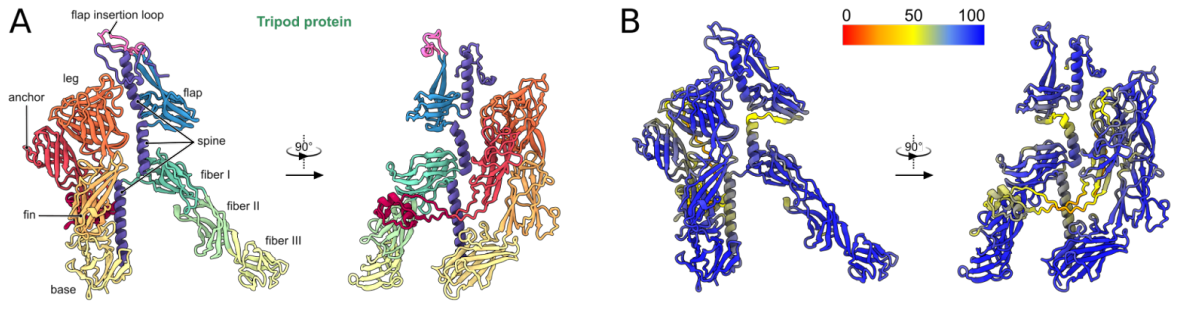


SFig. 11. Conformational changes and β-sheet augmentation of handshake domain of tail sheath initiator protein. (A-C) Cartoon representation of the tail sheath initiator protein from the baseplate of phi812 with extended tail. (A) Five subunits from the hexamer of the tail sheath initiator proteins are 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 baseplate of phi812 with extended tail 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 baseplates from phage particles with extended and contracted tails. **(H)** Differences between the interactions of tail sheath initiator proteins and wedge modules in baseplates from phage particles with extended and contracted tails.

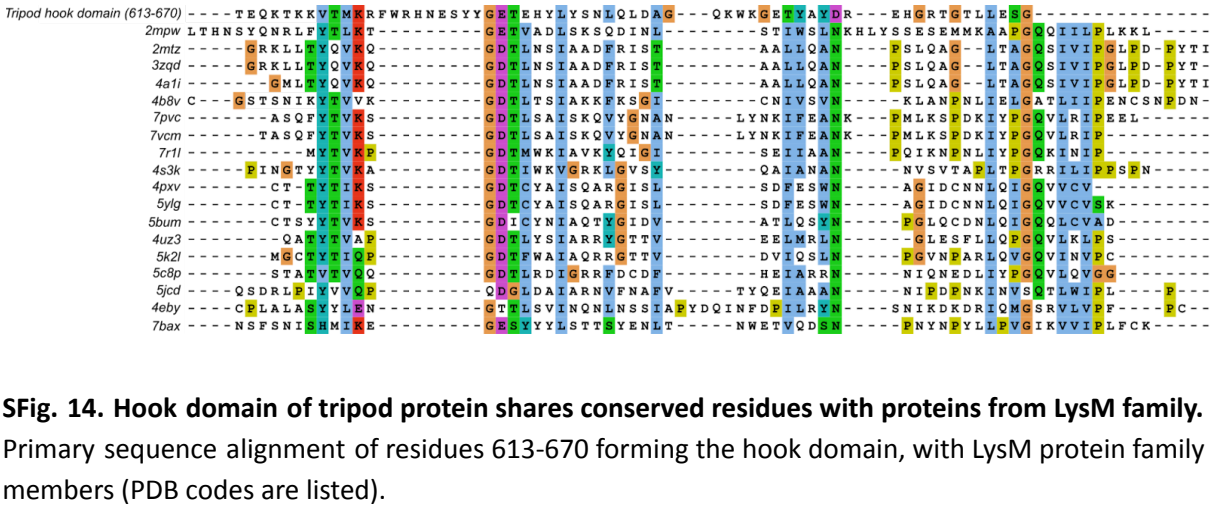


SFig. 12. Arm scaffold and arm segment proteins. (A) AlphaFold2-predicted structure of the arm scaffold protein shown in cartoon representation and colored according to pLDDT score[57]. The structure was divided into several parts for better visualization. Domain names and chain termini are indicated. **(B, C)** AlphaFold2-predicted complex of arm scaffold protein residues 177-489 (linker AB, bulge domain, docking loop 2, linker BC) and four subunits of arm segment protein. Structures are showed in cartoon representation colored according to the legend on the left (B) or according to pLDDT score[57] (C). Domain names and chain termini of arm scaffold protein are indicated. One model out of twenty is showed. **(D, E)** Close view on selected parts of the AlphaFold2-predicted complex described above. Main chain is showed in cartoon and side chains are showed in sticks representations. Structures in D and E are colored as in B and C, respectively. Numbers of first and

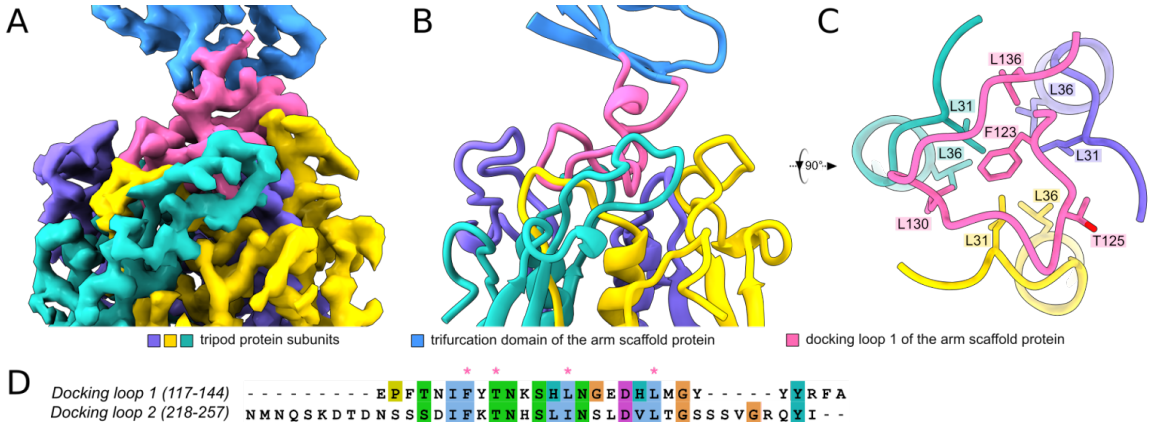
761 last residues of showed chains are indicated. Seven predicted models out of twenty are showed and
762 superimposed by the residues 197-207 of the arm scaffold protein. For more information about the
763 predictions, see Methods.



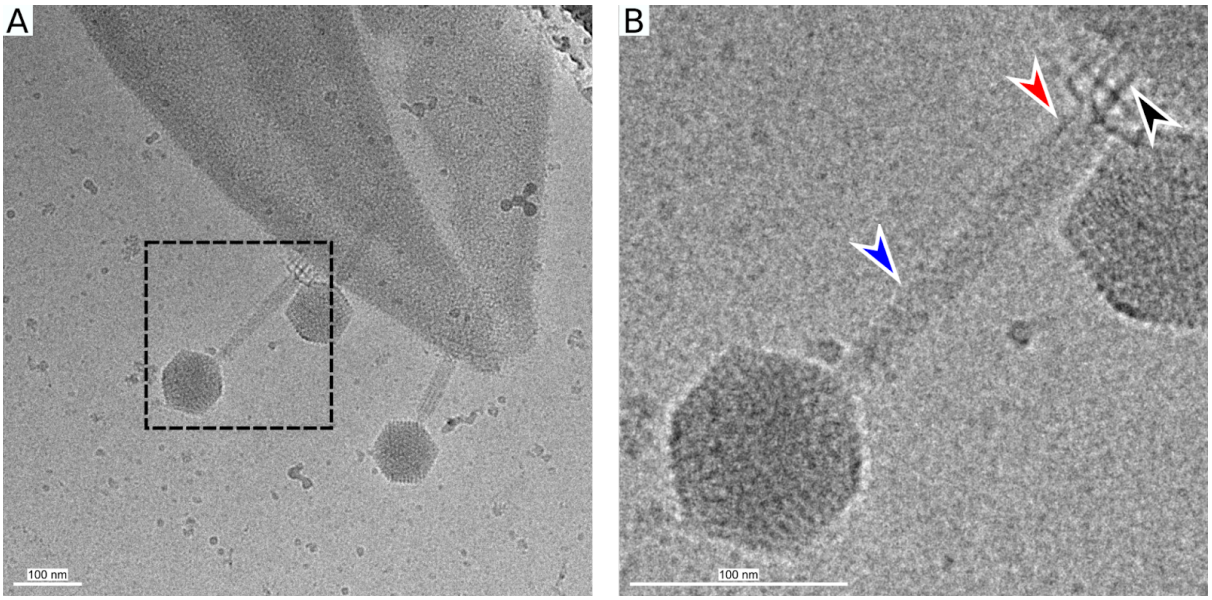
SFig. 13. Prediction of the tripod protein structure. (A, B) AlphaFold2-predicted structure of the tripod protein shown in cartoon representation and colored according to domains (A) and according to the pLDDT score[57] (B). In panel A, domain names are indicated.



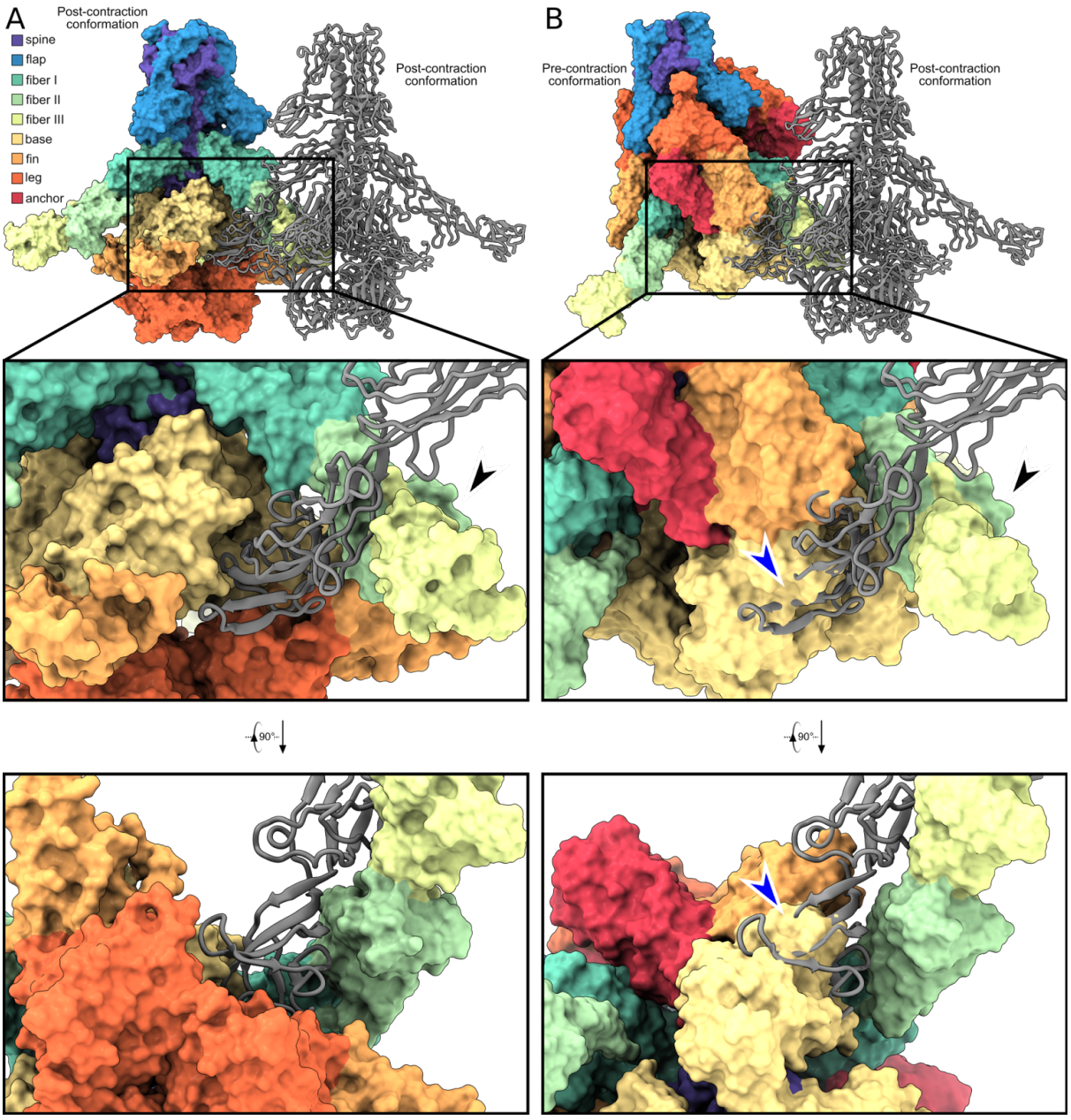
SFig. 14. Hook domain of tripod protein shares conserved residues with proteins from LysM family. Primary sequence alignment of residues 613-670 forming the hook domain, with LysM protein family members (PDB codes are listed).



SFig. 15. 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.

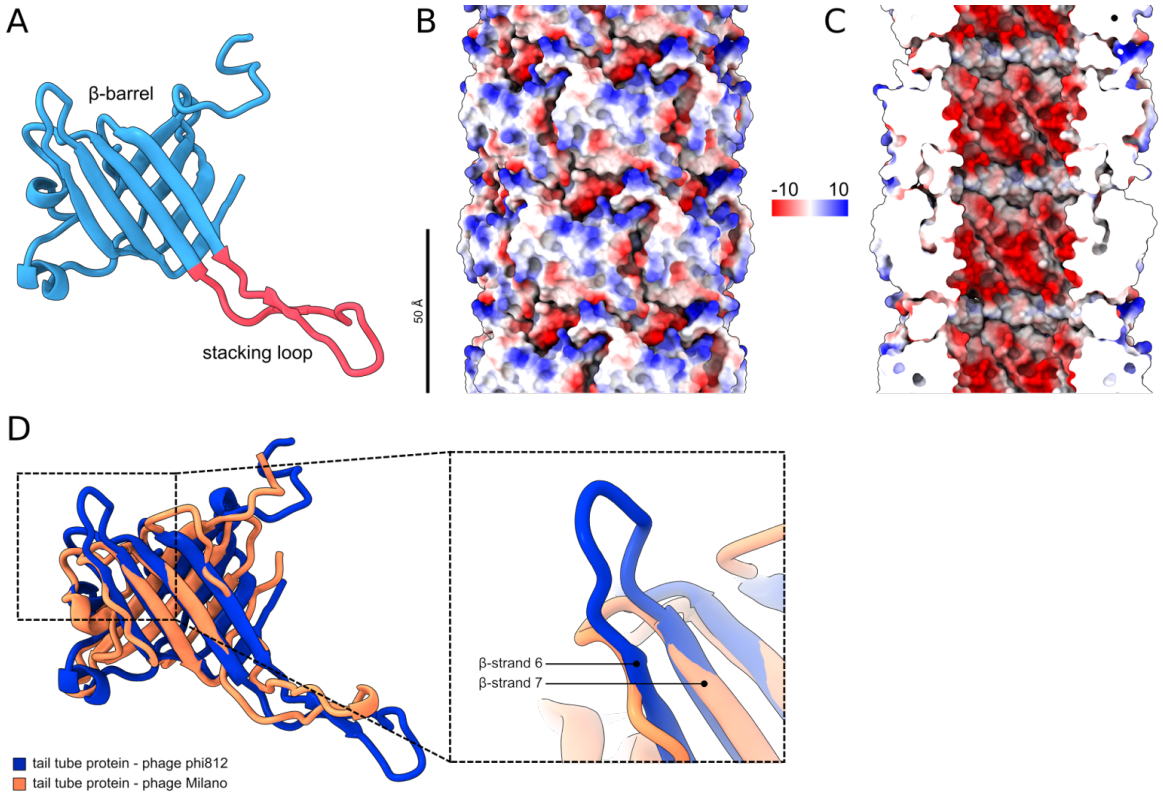


SFig. 16. Particle of phage phi812 captured in intermediate state of tail contraction. (A) Cryo-electron micrograph of phi812 attached to a cell wall of *S. aureus*. The black rectangle indicates the detail in B. **(B)** phi812 particle with a semi-contracted tail sheath. The blue arrowhead points to the sheath in the extended 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.

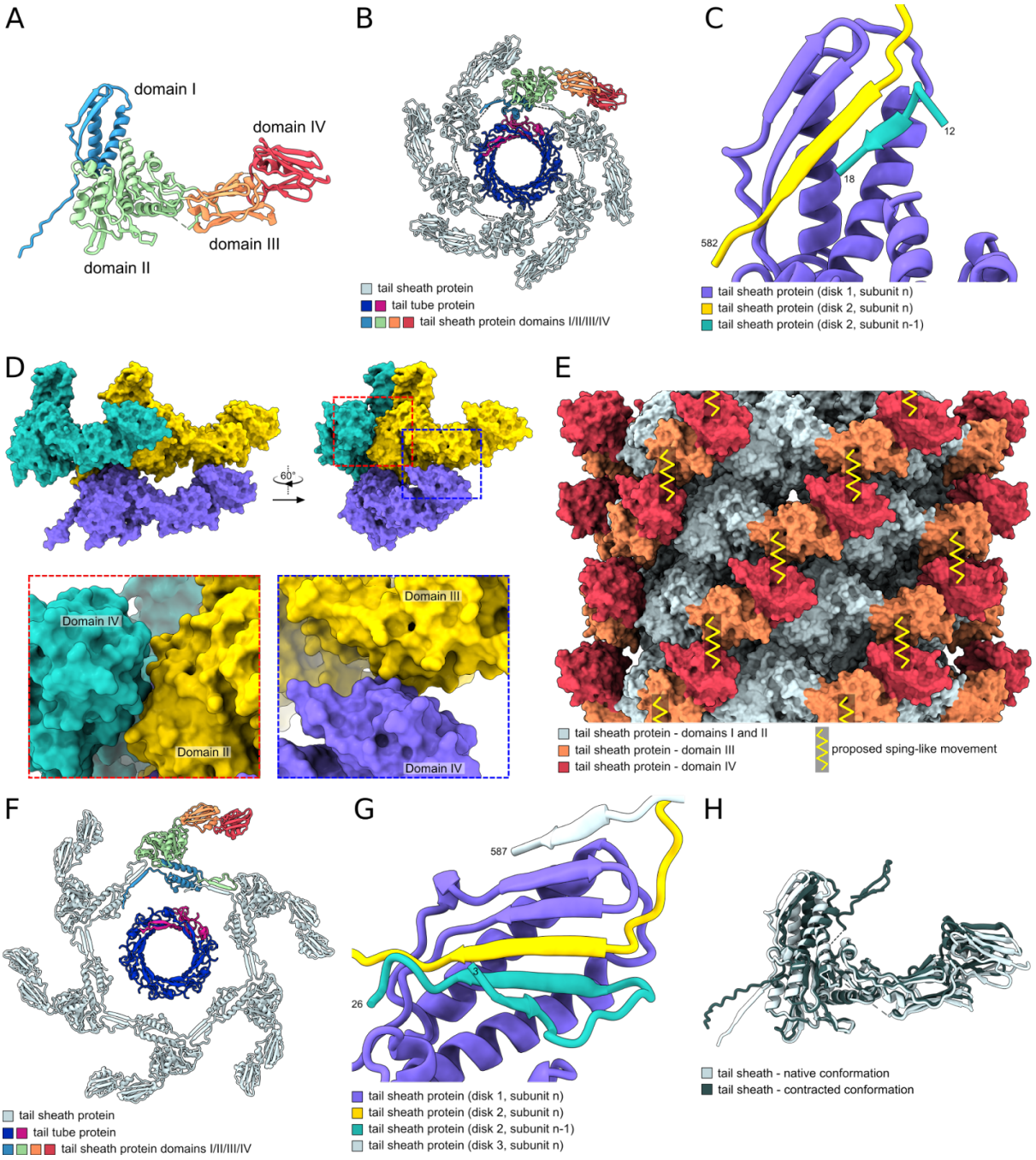


SFig. 17. Interactions between tripod complexes in baseplate of phi812 particle with contracted tail. (A) Interactions between two neighboring inner tripod complexes. One tripod complex is shown as a molecular surface colored according to domain, and the other is shown in cartoon representation in gray. 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

808 view along the tail axis and towards the phage head. Blue arrowheads point to the site where the
809 fiber III domains of the post-contraction tripod complex clash with the base domain of the
810 pre-contraction tripod complex.

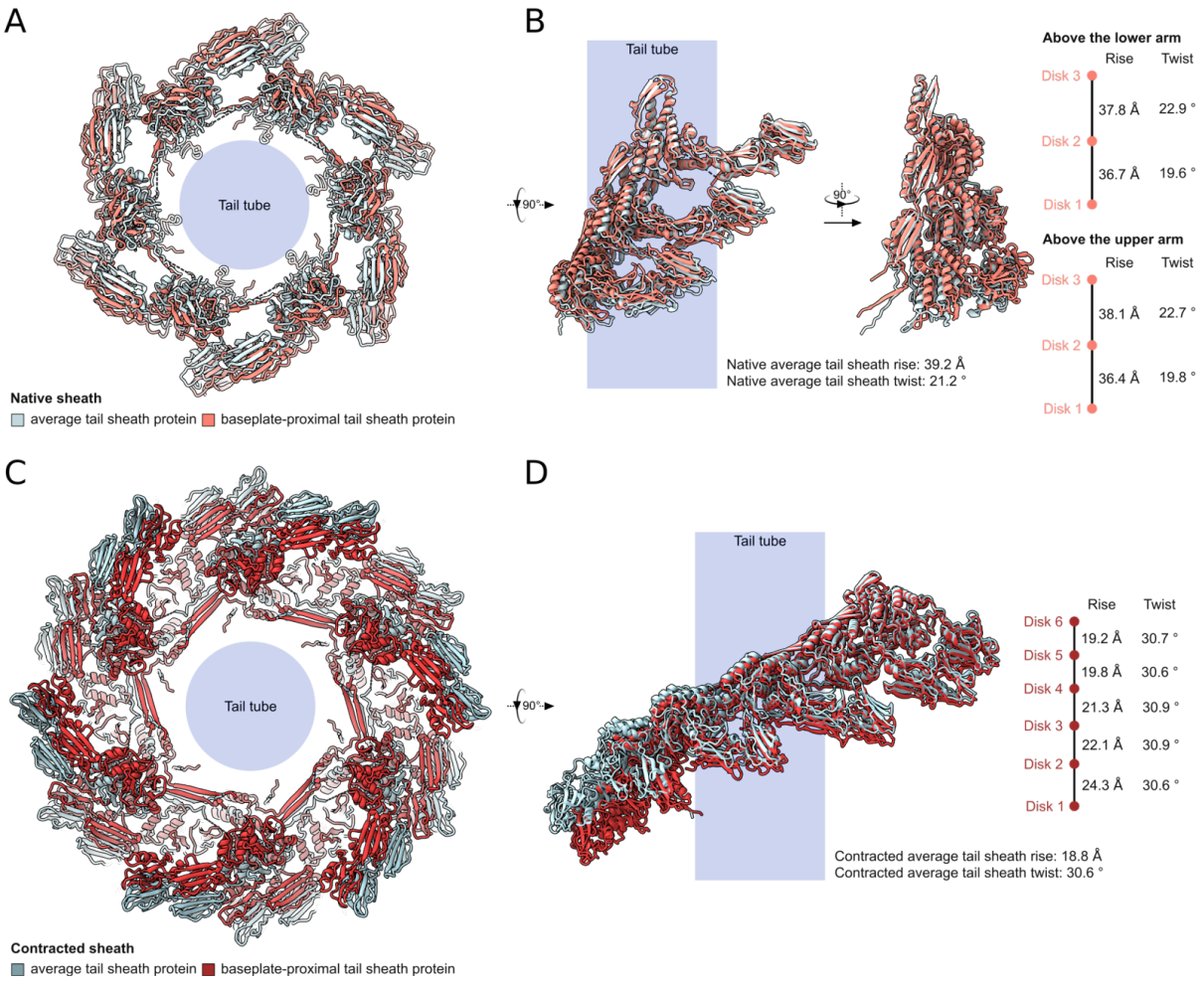


SFig. 18. Tail tube proteins. (A) Cartoon representation of the phi812 tail tube protein colored according to domains. **(B,C)** Molecular surface representation of the phi812 tail tube viewed as a whole (B) and clipped (C), colored according to Coulombic electrostatic potential in the range from -10 to 10 kcal/(mol·*e*-) at 298 K. **(D)** Cartoon representation of the tail tube proteins from phage phi812 and phage Milano superposed by the β-strands 6 and 7. The inset shows a close view on the loops between β-strands 6 and 7.



SFig. 19. Tail sheath proteins. (A) Cartoon representation of the phi812 tail sheath protein colored according to domains. **(B)** Cartoon representation of a disc of tail tube and tail sheath proteins from the extended 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 extended 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 extended conformation. Red and blue dashed rectangles indicate views of interactions among the subunits shown as insets. **(E)** Molecular surface representation of four tail sheath disks colored according to the legend. The proposed vertical spring-like movement of domains III and IV of the tail sheath protein is indicated. **(F)** 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. **(G)** β-sheet augmentation of the handshake domain of the tail sheath protein from the contracted tail. Terminal residues are indicated. **(H)** Comparison of the tail sheath proteins from extended (light gray) and contracted (black) tails. The subunits were superimposed by aligning their domains II.



SFig. 20. Baseplate-proximal discs of tail sheath proteins of phi812 in extended 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 from phi812 with extended tail. The baseplate core and wedge module proteins are shown in cartoon representation and colored according to Fig. 1E. The wedge module in the conformation 1 is shown in lighter colors to emphasize two distinct conformations of the wedge modules in the baseplate of phi812 with extended tail. The superposition of two neighboring wedge modules in conformation 1 and conformation 2 is shown. Protein names are indicated.

SMovie 2. Baseplate and baseplate-proximal part of extended phi812 tail. The baseplate and baseplate-proximal part of the tail of phi812 in extended 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 extended 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 with extended and contracted tails 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 extended 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.

	Extended tail	**Baseplate of extended tail (core, wedge modules, arms, proximal tail)**	**Baseplate of extended tail (core, wedge modules, proximal tail)**	**Baseplate of extended tail (core, wedge modules)**	**Baseplate of contracted tail (core, wedge modules, arms, proximal tail)**	**Baseplate of contracted tail (core, wedge modules, arm segment, inner tripods, proximal tail)**	**Baseplate of contracted tail (core, wedge modules, inner tripods)**	**Contracted tail tube**	**Contracted tail sheath**	**Central spike protein (knob and petal domains)**
Sample	Particles with extended tails	Particles with extended tails	Particles with extended tails	Particles with extended tails	Particles with contracted tails	Particles with contracted tails	Particles with contracted tails	Particles with contracted tails	Particles with contracted tails	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 of particles	18,188	9,709	9,709	9,709	25,203	25,203	25,203	12,906	69,969	135,736
Initial model	EMD-4051	EMD-8210	EMD-8210	EMD-8210	EMD-8212	EMD-8212	EMD-8212	Central part of the contracted baseplate recon.	EMD-4052	De novo
Map resolution [Å]	4.2	7.3	4.6	4.4	4.4	4.0	3.1	3.7	3.2	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	34,568	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) at 0.5 (masked)	4.19	10.38	4.57	6.21	6.65	4.14	3.25	3.86	3.30	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	1.027	0.471	1.400	0.881	0.781	0.404	0.437	0.438
Validation										
MolProbity score [percentile]	0.94	1.51	1.44	1.18	1.04	1.16	0.94	0.96	0.99	1.08
ClashScore [percentile]	0.96	0.99	1.72	2.67	1.05	1.23	0.90	1.96	2.19	2.71
Poor rotamers [%]	0.18	1.45	1.41	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	0.15	0.03	0.05	0.07	0	0	0	0
Favored [%]	97.13	88.30	94.06	97.35	96.42	95.40	97.07	99.04	98.35	97.89

na = not applicable

881 **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[3]; phage E217, gp31[4]; *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[3]; *P. aeruginosa* R2 pyocin, PA0626[28], *C. difficile* diffocine, CD1364[33]
			stacking loop (35-58)	
110	Hub	810	tube-forming I (1-125)	Phage T4, gp27[42]; *E. coli* Type VI Secretion System, VgrG[46]; *P. aeruginosa* R2 pyocin, PA0628[28]; phage 80alpha, ORF59[44]
			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[52]
			cleaver (642-808)	Phage K, ORF56 (tail-associated lysin)[50]; phage K, LysK (endolysin)[49]
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[53]
112	Central spike	848	oligosaccharide-binding (OB) (1-110)	Phage T4, gp5[42]; phage phi92, gp138 (spike)[41]
			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[36]; *B. cenocepacia,* BC2L-C (lectin)[38]
			petal (611-848)	*B. subtilis*, GlpQ (phosphodiesterase)[40]
113	Tail tube initiator	263	tube-like (1-156)	Phage T4, gp54/gp48[3]; *P. aeruginosa* R2 pyocin, PA0626 (ripcord)[28]; *C. difficile* diffocine, CD1367[33]

			handle (157-263)	*P. aeruginosa* R2 pyocin, PA0626 (ripcord)[28]; phage E217, gp37/gp38[4]; phage Pam3, gp17[6]
114	Assembly	173	-	*P. aeruginosa* R2 pyocin, PA0626 (ripcord)[28]; phage Milano, BCP (gp25)[9]
115	Tail sheath initiator	234	LysM (1-66)	Phage T4, gp53[3]; *P. aeruginosa* R2 pyocin, PA0627 (glue)[28]; phage Milano, BCP (gp25)[9]
			linker (67-92)	-
			handshake (93-234)	Phage T4, gp25[3]; *P. aeruginosa* R2 pyocin, PA0617 (sheath initiator)[28]; phage E217, gp34[4]; phage Pam3, gp17[6]
116	Wedge A, B	348	helices (1-67)	Phage T4, gp6[3]; *E. coli* Type VI Secretion System, TssFa/b[27]; *C. difficile* diffocine, CD1364[33]
			wing (68-192)	*C. difficile* diffocine, CD1364[33]; phage E217, gp44[4]; phage Milano, BCP (gp25)[9] ; phage A1(L), gp31[8]
			trifurcation unit (193-256)	Phage T4, gp6[3]; *E. coli* Type VI Secretion System, TssFa/b[27]; *C. difficile* diffocine, CD1364[33]
			dimerization (257-348)	
117	Arm scaffold	1019	helices (1-94)	Phage T4, gp7[3]; *E. coli* Type VI Secretion System, TssG[27]; *P. aeruginosa* R2 pyocin, PA0619 (Tri2)[28]; phage E217, gp45[4]; phage Pam3, gp23[6]
			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)[35]; *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[3]; 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)[36]; *E. coli* Type VI Secretion System, TssK[27]

			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]

882

883 **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
φ	95
ψ	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

884 * 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 teichoic 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-EM sample preparation, data collection, motion correction, and CTF estimation

Purified phi812 (10^{11} PFU/ml) in phage buffer (4 µl), either in the extended 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 baseplate of bacteriophage phi812 with extended tail

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 extended 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 extended 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.

Calculation of the tail rotation during the tail sheath contraction

Electron micrographs of the phages with extended tails were used to determine that the tail contains 51 tail sheath disks, The rotation between consecutive disks of tail sheath proteins is 21.1° in extended tail sheath and 30.6° in contracted tail. The rotation of the tail tip upon tail contraction was calculated as follows. Taking the first tail sheath disk as a reference point, the rotation of 50 tail sheath disks with the 21.1° twist equals to overall rotation of 1055°. For the contracted tail sheath, the rotation of 50 tail sheath disks with the 30.6° twist equals to overall rotation of 1530°. The overall rotation of contracted tail sheath was subtracted by the overall rotation of the extended tail sheath, resulting in the difference of 475°.

Cryo-EM of phi812 attached to *S. aureus* cell wall fragments

S. aureus SA 812 was cultured in 100 ml of meat peptone broth (1 g of Lab-Lemco (Difco), 5 g of yeast extract powder L21 (Oxoid), 10 g of peptone L37 (Oxoid), and 5 g of NaCl were dissolved in distilled water to a final volume of 1000 ml, and the pH was adjusted to 7.4 using 10 M NaOH) at 37°C until OD_{600}=0.3. Cells were harvested by centrifugation at 3,000 x *g* for 15 min. The resulting pellet was resuspended in 80% isopropanol to a final volume of 10 ml, and incubated for 1 h at room temperature. The suspension was centrifuged at 1,000 x *g* at room temperature for 15 min. The resulting supernatant was transferred to another tube, and cell wall fragments were harvested by centrifugation at 20,000 x *g* for 30 min. The resulting pellet was resuspended in 100 µl of the phage

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

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

1090 onto SDS-PAGE gels and zymograms. The SDS-gel and zymogram were separated using gel
1091 electrophoresis. After the gel electrophoresis, the zymogram was washed for 30 min with distilled
1092 water to remove SDS. To renature the cleaver domain, the zymogram was soaked at room
1093 temperature in distilled water for 2-4 h, and then for 1 h in the buffer containing 150 mM
1094 Na-phosphate, pH 7.0, 0.1% Triton X-100 and 10 mM $MgCl_2$. The zymogram was stained for 3
1095 h with 0.1% methylene blue in 0.001% KOH and washed with water. The peptidoglycan-hydrolytic
1096 activity of the cleaver domain of the hub protein was detected as a clear zone on a dark blue
1097 background.
1098

1099 **NMR structure determination of cleaver domain of hub protein**

1100 NMR spectra were recorded at 298 K in an 850 MHz Bruker Avance III spectrometer equipped with a
1101 $^{1}H/^{13}C/^{15}N$ TCI cryogenic probe head with z-axis gradients. The composition of the NMR sample was
1102 1.8 mM uniformly ^{13}C-, ^{15}N-labeled cleaver domain in a buffer containing 25 mM sodium phosphate,
1103 pH 6.6, 100 mM NaCl, 1.5 mM DTT, and 10% D_2O for the lock. Three sparsely sampled 4D NMR
1104 experiments were acquired: 4D HC(CC-TOCSY(CO))NH, 4D $^{13}C,^{15}N$ edited HMQC-NOESY-HSQC (HCNH),
1105 and 4D $^{13}C,^{13}C$ edited HMQC-NOESY-HSQC (HCCH). Sequential and aliphatic side chain assignments
1106 were obtained automatically using the 4D-CHAINS algorithm[105] and checked manually. Backbone
1107 dihedral angle restraints were derived from TALOS[106] using a combination of five kinds (H^N, H^α, C^α, C^β,
1108 N) of chemical shift assignments for each residue in the sequence. NOE cross-peaks from the two 4D
1109 NOESY spectra (HCNH and HCCH) were assigned automatically by CYANA 3.0[107] in structure
1110 calculations with torsion angle dynamics. Unambiguous distance and torsion angle restraints (Table
1111 S3) were used in a water refinement calculation[108] applying the RECOORD protocol[109]. The quality of
1112 the NMR-derived structure ensembles was validated using PSVS[110].
1113

1114 **Production of recombinant central spike protein**

1115 The gene for central spike protein was amplified from the genomic DNA of phage phi812 using
1116 forward and reverse primers (5'AGATTGGTGGCTCGGATGATTTAAATGTAAAAGGATTAGTTTTA3' and
1117 5'GAGGAGAGTTAAGACTTATTAAATCTTTTTATATTCATCTAGGTAGTTTGTAAA3'). Compared to the
1118 full-length gene, the first 17 nucleotide triplets from the 5' end of the amplified gene were omitted,
1119 as they code for the first 17 residues of the central spike protein, which were predicted to be
1120 disordered by the secondary structure prediction using Phyre[111]. The truncated gene was cloned into
1121 a vector derived from pET22b, which was modified to provide a His6-SUMO-tag fusion followed by
1122 the Ulp1 cleavage site at the N-terminus of the protein. The *E. coli* expression strain BL21(DE3) was
1123 transformed with the construct using a standard heat shock procedure (45 s at 42°C). An overnight
1124 culture was used to inoculate 2% Luria-Bertani broth (Sigma) supplemented with ampicillin (50
1125 µg/ml). Cells were cultivated at 37°C and 250 rpm until OD_{600}=0.5, followed by cultivation at 16°C and
1126 250 rpm for 24 h without the induction of expression, as soluble central spike protein was only
1127 produced after non-induced leaky expression. Cells were harvested in a typical yield of 8 g per liter of
1128 cell culture, resuspended in 2.5 ml of the loading buffer (50 mM Na/K phosphate, pH 7, 200 mM
1129 NaCl) per g of cell pellet, and subjected to chemical disintegration (0.5 % octylthioglucoside, 1 mg/ml
1130 lysozyme, 2 U/ml Turbonuclease, 1 mM PMSF, 5 mM TCEP) at 6°C overnight. The lysate was
1131 centrifuged at 20,000 x *g* and 8°C for 30 min. The resulting supernatant was filtered (0.45 µm cut-off,
1132 Corning Inc.) and loaded into a 5 ml Histrap column (GE Healthcare) pre-equilibrated with the loading

1133 buffer and connected to an AKTA FPLC system (GE Healthcare). After the column was washed with 15
1134 column volumes of the loading buffer, gradient elution was performed using the elution buffer (50
1135 mM Na/K phosphate, pH 7, 200 mM NaCl, 500 mM imidazole). Eluted fractions containing the central
1136 spike protein, as determined by SDS-PAGE analysis using 12.5 % polyacrylamide gels, were pooled
1137 and supplemented with 5 mM TCEP before the cleavage of the purification tag overnight at 6°C using
1138 the in-house prepared Ulp1 protease. The cleavage solution was desalted using a HiPrep™ 26/10
1139 Desalting column (Cytiva) before reverse affinity chromatography using the 5 ml Histrap column (GE
1140 Healthcare). Cleaved protein collected in the flow-through fraction was concentrated (30 kDa cut-off;
1141 Amicon® Ultra; Merc Millipore) and loaded into a Superdex 16/600 200 pg column (GE Healthcare)
1142 pre-equilibrated with the loading buffer (50 mM Na/K phosphate, pH 7, 200 mM NaCl), followed by
1143 an isocratic elution. Monodisperse fractions were analyzed by SDS-PAGE and concentrated to a final
1144 concentration of 5 mg/ml.
1145
1146 **Cryo-EM reconstruction of recombinant central spike protein**

1147 The purified central spike protein sample was diluted in 50 mM Na/K phosphate, pH 7, 200 mM NaCl
1148 buffer to a concentration of 0.8 mg/ml, and then diluted 10x in water. Four µl of the central spike
1149 protein at a concentration of 0.08 mg/ml were applied onto a Quantifoil R1.2/1.3 200-mesh Cu grid
1150 glow-discharged for 60 s from both sides. The excess of the sample was blotted for 1 s with filter
1151 paper (Whatman No. 1) using blot force -2, and immediately plunge-frozen in liquid ethane using a
1152 Vitrobot Mark IV (FEI). The instrument chamber was kept equilibrated to 100 % humidity and at 4°C.
1153 Cryo-EM data were collected using a Titan Krios TEM (Thermo Fisher Scientific) operated at 300 ke^-V
1154 and equipped with a K2 Summit direct electron detector (Gatan). The beam was aligned for parallel
1155 illumination in NanoProbe mode, and the residual beam tilt was removed by coma-free alignments.
1156 The software SerialEM was used to acquire the data in the electron counting mode at 130,000 x
1157 magnification, corresponding to a calibrated pixel size of 1.04 Å/px over a defocus range of -0.3 to
1158 -3.0 µm. The exposure of 8 s with a dose rate of 6.6 e^-/px/s was fractionated into 40 frames (total
1159 dose 57 $e^-/Å^2$, 1.425 $e^-/Å^2$ per frame). A total of 3120 micrographs were collected. For each movie,
1160 the drift and gain correction was performed with MotionCor2[96], while the contrast-transfer function
1161 was estimated using the program Gctf[97]. crYOLO[102] was used to pick particles from micrographs
1162 denoised using the program janni. Particles picked manually from 10 micrographs were used to train
1163 the neural network. The resulting 487,160 auto-picked particles were extracted and binned to a 128
1164 px box size using Relion[100]. Several rounds of 2D classification sorted out classes of the central spike
1165 protein's C-terminal domains containing 226,791 particles. These particles were re-extracted and
1166 binned to a 128 px box size (2.08 Å/px) using Relion 5.0 beta (github.com/3dem/relion/tree/ver5.0).
1167 Three more rounds of 2D classification resulted in 135,736 particles, which were used for 3D
1168 reconstruction using the initial model generated by the stochastic gradient descent algorithm within
1169 Relion. Subsequent 3D classifications did not improve the quality or resolution of the reconstruction.
1170 Therefore, all 135,736 particles were re-extracted using a 256 px box size (1.04 Å/px) and assigned to
1171 their respective optics groups. Several rounds of CTF refinements, including anisotropic magnification
1172 estimation, beam tilt and trefoil aberrations estimation, and fitting of the defocus and astigmatism
1173 parameters per micrograph, improved the resolution of the post-processed map to 3.65 Å. Then, two
1174 rounds of Bayesian polishing and subsequent 3D auto-refinements and CTF refinements were
1175 performed. The final map reached a resolution of 3.23 Å. The resolution after post-processing was

1176 calculated according to the gold standard FSC = 0.143 criterion. Local resolution maps were
1177 computed using Relion[101].
1178

1179 **Production of recombinant Arm segment protein**

1180 The gene encoding the arm segment protein was cloned into the pET-28b(+) vector
1181 (Novagen/Merck), providing a His6 tag at the C-terminus of the protein. An *E. coli* expression strain
1182 BL21(DE3) was transformed with the construct using a standard heat shock procedure (45 s at 42°C).
1183 An overnight culture was used to inoculate 2% Luria-Bertani broth (Sigma) supplemented with 1 mM
1184 MgCl2, 1 % glucose, and ampicillin (50 µg/ml), followed by cultivation at 37°C and 250 rpm until
1185 OD_{600}=0.5. The expression was induced with 0.4 mM IPTG at 18°C and was allowed to progress for 17
1186 h. Cells were harvested and lysed using the lysis buffer (0.5 % octyl-thioglucoside, 1 mg/ml lysozyme,
1187 2 U/ml Turbonuclease, 1 mM PMSF, 5 mM TCEP) at room temperature for 30 minutes, and
1188 subsequently disintegrated with an Emulsiflex C3 (Avestin). After centrifugation at 20,000 x *g* and 8°C
1189 for 30 min, the supernatant was filtered (0.45 µm cut-off, Corning Inc.) and loaded into a 5 ml Histrap
1190 column (GE Healthcare) pre-equilibrated with the loading buffer (50 mM HEPES, pH 8, 200 mM NaCl,
1191 40 mM imidazole) connected to an AKTA Pure FPLC system (GE Healthcare). The protein was eluted
1192 in one step using the elution buffer (50 mM HEPES, pH 8, 200 mM NaCl, 500 mM imidazole). Eluted
1193 fractions containing the arm segment protein, detected by SDS-PAGE analysis on 12 %
1194 polyacrylamide gels, were loaded into a Superdex 16/600 75 pg column (GE Healthcare)
1195 pre-equilibrated with the loading buffer (50 mM HEPES, pH 8, 200 mM NaCl), followed by an isocratic
1196 elution. Monodisperse fractions were analyzed by SDS-PAGE and concentrated to a final
1197 concentration of 14 mg/ml.
1198

1199 **Crystallization and X-ray structure determination of recombinant arm segment protein**

1200 The arm segment protein was crystallized by vapor diffusion using the sitting-drop method at 21°C.
1201 The protein sample at a concentration of 14 mg/ml was used to prepare drops with protein to
1202 precipitant ratios of 200 nl : 100 nl, 100 nl : 100 nl, and 100 nl : 200 nl in MRC 3-drop 96-well plates
1203 using a Dragonfly robot (SPT Labtech) with a range of commercially available screens (Molecular
1204 Dimensions, Hampton Research, Qiagen-Nextal). Thick needle-like crystals were observed in the
1205 drops containing the solution of 1.6 M LiCl, 0.1 M Tris, pH 7.5, 25 % w/v PEG 4,000. The same
1206 crystallization solution was used to prepare hanging drops for vapor diffusion. Crystals were
1207 harvested from the hanging drops and flash-vitrified in liquid nitrogen.
1208 Diffraction data were recorded at 100 K on the Proxima-2A beamline at the Soleil synchrotron (Paris,
1209 France). Native crystals were soaked in I3C using the I3C phasing kit (Hampton Research), and the
1210 dataset was phased using the multiple wavelength anomalous dispersion method. The diffraction
1211 data from the crystals soaked with I3C were collected at 12.67 keV and 5.00 keV. The structure of
1212 the arm segment protein was determined using the program SHELXD[112]. The native diffraction data
1213 recorded from the crystals of arm segment protein were phased using the program Phaser[113], and the
1214 structure was determined using the anomalous data as a molecular replacement model. Refinement
1215 was performed manually using the program *Coot*[114] and automatically in *REFMAC5*[115]. The structures
1216 were validated using *MolProbity*[116].
1217

Structure prediction, determination and refinement using cryo-EM maps

The structures of phage proteins were predicted using the program Alphafold2[57] and fitted into their densities as rigid bodies or by individual domains using the programs Chimera[117] or ChimeraX[118], and then refined interactively using ISOLDE[119]. The structures were improved using the real-space refinement implemented in the *Phenix* package[120], and corrected manually using the programs *Coot*[114] and ISOLDE[119]. The geometry evaluation of models was performed using *MolProbity*[116]. Map to model FSC curves were calculated using Comprehensive Validation tool in *Phenix* package[120].

Prediction of arm segment protein and arm scaffold protein interaction

Structure predictions of the arm scaffold and arm segment protein complex were performed using AlphaFold2 Multimer with default parameters refxxx. For the arm scaffold protein, residues 177–489, encompassing the linkerAB, bulge domain, and linkerBC regions, were used as input. For the arm segment protein, four times full-length sequence was used as input to model four protein subunits.

Visualization and structure analysis

Chimera[117] and ChimeraX[118] were used to visualize maps and structures. PISA web server[121] was used to calculate interface areas. Sequences were analyzed using Clustal Omega[122], available via the EMBL-EBI server[123] and visualized with Jalview[124].

Construction of the composite maps

In Fig. 2, the composite map includes densities of the arm scaffold protein, weld protein, and cleaver domain of the hub protein, which were segmented out from the reconstruction of the whole baseplate of phi812 with the extended tail (SFig. 2A). The densities of rest of the proteins belong to a map from focused reconstruction of wedge modules and baseplate core proteins of phi812 with extended tail (SFig. 2C).
In Fig. 3B, the composite map includes reconstruction of the whole baseplate of phi812 with the extended tail (SFig. 2A) and reconstruction of the knob and petal domains of the central spike protein (SFig. J).
In Fig. 6B, C, the composite map includes densities from three separate reconstructions. First, the density of the tail tube belongs to the reconstruction of the tail tube of phi812 with contracted tail (SFig. 2I). Second, the densities of the tail sheath initiator protein, wedge module proteins and inner tripod proteins belong to the focused reconstruction of the baseplate of phi812 with contracted tail (SFig. 2F). Third, the densities of the tail sheath protein, bulge, elbow and connector domains of the arm segment protein, outer tripod protein, and receptor-binding protein 2 belong to the reconstruction of the whole baseplate of phi812 with contracted tail (SFig. 2D).
In Fig. 6G, the composite map includes density of the tail sheath from the reconstruction of the tail sheath of phi812 with contracted tail (SFig. 2H) and the density of the tail tube from the reconstruction of the tail tube of phi812 with contracted tail (SFig. 2I).

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 and the project New Technologies for Translational Research in Pharmaceutical Sciences /NETPHARM, project ID OP JAC CZ.02.01.01/00/22_008/0004607, which is co-funded by the European Union. 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 phage phi812 with extended 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 an extended 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 an extended conformation computed with imposed threefold symmetry: EMDB-19969 and PDB-9EUG. Structures of baseplate core and wedge modules of phi812 tail in an extended 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 an extended 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. Taylor, N. M. I. *et al.* Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* **533**, 346–352 (2016).
4. Li, F. *et al.* High-resolution cryo-EM structure of the Pseudomonas bacteriophage E217. *Nat. Commun.* **14**, 4052 (2023).
5. Wang, Z. *et al.* Structure of Vibrio Phage XM1, a Simple Contractile DNA Injection Machine. *Viruses* **15**, 1673 (2023).
6. Yang, F. *et al.* Fine structure and assembly pattern of a minimal myophage Pam3. *Proc. Natl. Acad. Sci.* **120**, e2213727120 (2023).
7. 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).
8. Yu, R.-C. *et al.* Structure of the intact tail machine of Anabaena myophage A-1(L). *Nat. Commun.* **15**, 2654 (2024).
9. Sonani, R. R. *et al.* An extensive disulfide bond network prevents tail contraction in Agrobacterium tumefaciens phage Milano. *Nat. Commun.* **15**, 756 (2024).
10. 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).
11. Kwiecinski, J. M. & Horswill, A. R. Staphylococcus aureus bloodstream infections: pathogenesis and regulatory mechanisms. *Curr. Opin. Microbiol.* **53**, 51–60 (2020).
12. Grunenwald, C. M., Bennett, M. R. & Skaar, E. P. Nonconventional Therapeutics against Staphylococcus aureus. *Microbiol. Spectr.* **6**, (2018).

1319 13. Venturini, C., Petrovic Fabijan, A., Fajardo Lubian, A., Barbirz, S. & Iredell, J. Biological
1320 foundations of successful bacteriophage therapy. *EMBO Mol. Med.* **14**, (2022).
1321 14. Petrovic Fabijan, A. *et al.* Safety of bacteriophage therapy in severe Staphylococcus aureus
1322 infection. *Nat. Microbiol.* **5**, 465–472 (2020).
1323 15. Uyttebroek, S. *et al.* Safety and efficacy of phage therapy in difficult-to-treat infections: a
1324 systematic review. *Lancet Infect. Dis.* **22**, e208–e220 (2022).
1325 16. Pantůček, R. *et al.* The Polyvalent Staphylococcal Phage φ812:Its Host-Range Mutants and
1326 Related Phages. *Virology* **246**, 241–252 (1998).
1327 17. Schaechter, M. *Encyclopedia of Microbiology*. (Elsevier/Academic Press, Amsterdam Boston,
1328 2009).
1329 18. van Dalen, R., Peschel, A. & van Sorge, N. M. Wall Teichoic Acid in Staphylococcus aureus Host
1330 Interaction. *Trends Microbiol.* **28**, 985–998 (2020).
1331 19. Xia, G. *et al.* Wall Teichoic Acid-Dependent Adsorption of Staphylococcal Siphovirus and
1332 Myovirus. *J. Bacteriol.* **193**, 4006–4009 (2011).
1333 20. Takeuchi, I. *et al.* The Presence of Two Receptor-Binding Proteins Contributes to the Wide Host
1334 Range of Staphylococcal Twort-Like Phages. *Appl. Environ. Microbiol.* **82**, 5763–5774 (2016).
1335 21. Azam, A. H., Hoshiga, F., Takeuchi, I., Miyanaga, K. & Tanji, Y. Analysis of phage resistance in
1336 Staphylococcus aureus SA003 reveals different binding mechanisms for the closely related
1337 Twort-like phages ϕSA012 and ϕSA039. *Appl. Microbiol. Biotechnol.* **102**, 8963–8977 (2018).
1338 22. Peng, C. *et al.* Silviavirus phage ϕMR003 displays a broad host range against methicillin-resistant
1339 Staphylococcus aureus of human origin. *Appl. Microbiol. Biotechnol.* **103**, 7751–7765 (2019).
1340 23. Nováček, J. *et al.* Structure and genome release of Twort-like Myoviridae phage with a
1341 double-layered baseplate. *Proc. Natl. Acad. Sci.* **113**, 9351–9356 (2016).
1342 24. Guerrero-Ferreira, R. C. *et al.* Structure and transformation of bacteriophage A511 baseplate and
1343 tail upon infection of *Listeria* cells. *EMBO J.* **38**, (2019).
1344 25. 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).

26. Veesler, D. & Cambillau, C. A Common Evolutionary Origin for Tailed-Bacteriophage Functional Modules and Bacterial Machineries. *Microbiol. Mol. Biol. Rev.* **75**, 423–433 (2011).
27. 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).
28. Ge, P. *et al.* Action of a minimal contractile bactericidal nanomachine. *Nature* **580**, 658–662 (2020).
29. Weiss, G. L. *et al.* Structure of a thylakoid-anchored contractile injection system in multicellular cyanobacteria. *Nat. Microbiol.* **7**, 386–396 (2022).
30. 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).
31. Desfosses, A. *et al.* Atomic structures of an entire contractile injection system in both the extended and contracted states. *Nat. Microbiol.* **4**, 1885–1894 (2019).
32. Jiang, F. *et al.* Cryo-EM Structure and Assembly of an Extracellular Contractile Injection System. *Cell* **177**, 370-383.e15 (2019).
33. Cai, X. *et al.* Atomic structures of a bacteriocin targeting Gram-positive bacteria. *Nat. Commun.* **15**, 7057 (2024).
34. Eck, M. J. & Sprang, S. R. The Structure of Tumor Necrosis Factor-α at 2.6 Å Resolution. *J. Biol. Chem.* **264**, 17595–17605 (1989).
35. Legrand, P. *et al.* The Atomic Structure of the Phage Tuc2009 Baseplate Tripod Suggests that Host Recognition Involves Two Different Carbohydrate Binding Modules. *mBio* **7**, (2016).
36. Farenc, C. *et al.* Molecular Insights on the Recognition of a Lactococcus lactis Cell Wall Pellicle by the Phage 1358 Receptor Binding Protein. *J. Virol.* **88**, 7005–7015 (2014).
37. McCabe, O. *et al.* The targeted recognition of *L actococcus lactis* phages to their polysaccharide receptors: Receptor binding to lactococcal phage 1358. *Mol. Microbiol.* **96**, 875–886 (2015).
38. Šulák, O. *et al.* A TNF-like Trimeric Lectin Domain from *Burkholderia cenocepacia* with Specificity

1371 for Fucosylated Human Histo-Blood Group Antigens. *Structure* **18**, 59–72 (2010).

1372 39. Myers, C. L. *et al.* Identification of Two Phosphate Starvation-induced Wall Teichoic Acid
1373 Hydrolases Provides First Insights into the Degradative Pathway of a Key Bacterial Cell Wall
1374 Component. *J. Biol. Chem.* **291**, 26066–26082 (2016).

1375 40. Shi, L., Liu, J.-F., An, X.-M. & Liang, D.-C. Crystal structure of glycerophosphodiester
1376 phosphodiesterase (GDPD) from *Thermoanaerobacter tengcongensis* , a metal ion-dependent
1377 enzyme: Insight into the catalytic mechanism: Crystal Structure of GDPD. *Proteins Struct. Funct.*
1378 *Bioinforma.* **72**, 280–288 (2008).

1379 41. Browning, C., Shneider, M. M., Bowman, V. D., Schwarzer, D. & Leiman, P. G. Phage Pierces the
1380 Host Cell Membrane with the Iron-Loaded Spike. *Structure* **20**, 326–339 (2012).

1381 42. Kanamaru, S. *et al.* Structure of the cell-puncturing device of bacteriophage T4. *Nature* **415**,
1382 553–557 (2002).

1383 43. Shneider, M. M. *et al.* PAAR-repeat proteins sharpen and diversify the type VI secretion system
1384 spike. *Nature* **500**, 350–353 (2013).

1385 44. Kizziah, J. L., Manning, K. A., Dearborn, A. D. & Dokland, T. Structure of the host cell recognition
1386 and penetration machinery of a Staphylococcus aureus bacteriophage. *PLOS Pathog.* **16**,
1387 e1008314 (2020).

1388 45. Büttner, C. R., Wu, Y., Maxwell, K. L. & Davidson, A. R. Baseplate assembly of phage Mu: Defining
1389 the conserved core components of contractile-tailed phages and related bacterial systems. *Proc.*
1390 *Natl. Acad. Sci.* **113**, 10174–10179 (2016).

1391 46. Leiman, P. G. *et al.* Type VI secretion apparatus and phage tail-associated protein complexes
1392 share a common evolutionary origin. *Proc. Natl. Acad. Sci.* **106**, 4154–4159 (2009).

1393 47. Bateman, A. & Rawlings, N. D. The CHAP domain: a large family of amidases including GSP
1394 amidase and peptidoglycan hydrolases. *Trends Biochem. Sci.* **28**, 234–237 (2003).

1395 48. Sanz-Gaitero, M., Keary, R., Garcia-Doval, C., Coffey, A. & van Raaij, M. J. Crystal structure of the
1396 lytic CHAPK domain of the endolysin LysK from Staphylococcus aureus bacteriophage K. *Virol. J.*

11, 133 (2014).

49. 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).

50. 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).

51. 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).

52. 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).

53. 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).

54. Habann, M. *et al.* Listeria phage A511, a model for the contractile tail machineries of SPO1-related bacteriophages. *Mol. Microbiol.* **92**, 84–99 (2014).

55. 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).

56. Watts, N. R. & Coombs, D. H. Structure of the bacteriophage T4 baseplate as determined by chemical cross-linking. *J. Virol.* **64**, 143–154 (1990).

57. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

58. Leiman, P. G. *et al.* Structure and Location of Gene Product 8 in the Bacteriophage T4 Baseplate. *J. Mol. Biol.* **328**, 821–833 (2003).

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).

1449 72. Moody, M. F. Sheath of bacteriophage T4. *J. Mol. Biol.* **80**, 613–635 (1973).

1450 73. Nazarov, S. *et al.* Cryo- EM reconstruction of Type VI secretion system baseplate and sheath distal
1451 end. *EMBO J.* **37**, e97103 (2018).

1452 74. Fraser, A. *et al.* Quantitative description of a contractile macromolecular machine. *Sci. Adv.* **7**,
1453 eabf9601 (2021).

1454 75. Fernandes, S. & São-José, C. Enzymes and Mechanisms Employed by Tailed Bacteriophages to
1455 Breach the Bacterial Cell Barriers. *Viruses* **10**, 396 (2018).

1456 76. Nishima, W., Kanamaru, S., Arisaka, F. & Kitao, A. Screw Motion Regulates Multiple Functions of
1457 T4 Phage Protein Gene Product 5 during Cell Puncturing. *J. Am. Chem. Soc.* **133**, 13571–13576
1458 (2011).

1459 77. Wenzel, S., Shneider, M. M., Leiman, P. G., Kuhn, A. & Kiefer, D. The Central Spike Complex of
1460 Bacteriophage T4 Contacts PpiD in the Periplasm of Escherichia coli. *Viruses* **12**, 1135 (2020).

1461 78. Hu, B., Margolin, W., Molineux, I. J. & Liu, J. Structural remodeling of bacteriophage T4 and host
1462 membranes during infection initiation. *Proc. Natl. Acad. Sci.* **112**, E4919–E4928 (2015).

1463 79. Gillespie, J. J. *et al.* Structural Insight into How Bacteria Prevent Interference between Multiple
1464 Divergent Type IV Secretion Systems. *mBio* **6**, e01867-15 (2015).

1465 80. Wu, X. *et al.* Crystal structure of CagV, the *Helicobacter pylori* homologue of the T4 SS protein
1466 VirB8. *FEBS J.* **286**, 4294–4309 (2019).

1467 81. Xu, Q. *et al.* Structure of the γ- D -glutamyl- L -diamino acid endopeptidase YkfC from *Bacillus*
1468 *cereus* in complex with L -Ala-γ- D -Glu: insights into substrate recognition by NlpC/P60 cysteine
1469 peptidases. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **66**, 1354–1364 (2010).

1470 82. Bhardwaj, A., Walker-Kopp, N., Casjens, S. R. & Cingolani, G. An Evolutionarily Conserved Family
1471 of Virion Tail Needles Related to Bacteriophage P22 gp26: Correlation between Structural
1472 Stability and Length of the α-Helical Trimeric Coiled Coil. *J. Mol. Biol.* **391**, 227–245 (2009).

1473 83. Bhardwaj, A., Molineux, I. J., Casjens, S. R. & Cingolani, G. Atomic Structure of Bacteriophage Sf6
1474 Tail Needle Knob. *J. Biol. Chem.* **286**, 30867–30877 (2011).

1475 84. Najmudin, S. *et al.* Putting an N-terminal end to the Clostridium thermocellum xylanase Xyn10B
1476 story: Crystal structure of the CBM22-1–GH10 modules complexed with xylohexaose. *J. Struct.*
1477 *Biol.* **172**, 353–362 (2010).
1478 85. Sainz-Polo, M. A., González, B., Menéndez, M., Pastor, F. I. J. & Sanz-Aparicio, J. Exploring
1479 Multimodularity in Plant Cell Wall Deconstruction. *J. Biol. Chem.* **290**, 17116–17130 (2015).
1480 86. Bárdy, P. *et al.* Structure and mechanism of DNA delivery of a gene transfer agent. *Nat. Commun.*
1481 **11**, 3034 (2020).
1482 87. Hardy, J. M. *et al.* The architecture and stabilisation of flagellotropic tailed bacteriophages. *Nat.*
1483 *Commun.* **11**, 3748 (2020).
1484 88. Chao, K. L. *et al.* Structure of Escherichia coli O157:H7 bacteriophage CBA120 tailspike protein 4
1485 baseplate anchor and tailspike assembly domains (TSP4-N). *Sci. Rep.* **12**, 2061 (2022).
1486 89. Han, Z. *et al.* Structural insights into a spindle-shaped archaeal virus with a sevenfold
1487 symmetrical tail. *Proc. Natl. Acad. Sci.* **119**, e2119439119 (2022).
1488 90. Tamura, K. *et al.* Surface glycan-binding proteins are essential for cereal beta-glucan utilization by
1489 the human gut symbiont Bacteroides ovatus. *Cell. Mol. Life Sci.* **76**, 4319–4340 (2019).
1490 91. Granell, M., Namura, M., Alvira, S., Kanamaru, S. & van Raaij, M. Crystal Structure of the
1491 Carboxy-Terminal Region of the Bacteriophage T4 Proximal Long Tail Fiber Protein Gp34. *Viruses*
1492 **9**, 168 (2017).
1493 92. Salazar, A. J., Sherekar, M., Tsai, J. & Sacchettini, J. C. R pyocin tail fiber structure reveals a
1494 receptor-binding domain with a lectin fold. *PLOS ONE* **14**, e0211432 (2019).
1495 93. Plattner, M. *et al.* Structure and Function of the Branched Receptor-Binding Complex of
1496 Bacteriophage CBA120. *J. Mol. Biol.* **431**, 3718–3739 (2019).
1497 94. Garcia-Doval, C. & van Raaij, M. J. Structure of the receptor-binding carboxy-terminal domain of
1498 bacteriophage T7 tail fibers. *Proc. Natl. Acad. Sci.* **109**, 9390–9395 (2012).
1499 95. Fokine, A. *et al.* Molecular architecture of the prolate head of bacteriophage T4. *Proc. Natl.*
1500 *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. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).

117. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

118. Meng, E. C. *et al.* UCSF ChimeraX : Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).

119. 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).

120. 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).

121. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).

122. Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein

1553 sequences. *Protein Sci.* **27**, 135–145 (2018).

1554 123.Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic*

1555 *Acids Res.* **50**, W276–W279 (2022).

1556 124.Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a

1557 multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191

1558 (2009).