

Structure of the T4 baseplate and its function in triggering sheath contraction

Nicholas M. I. Taylor¹, Nikolai S. Prokhorov^{1,2}, Ricardo C. Guerrero-Ferreira¹, Mikhail M. Shneider^{1,3}, Christopher Browning^{1,†}, Kenneth N. Goldie⁴, Henning Stahlberg⁴ & Petr G. Leiman¹

Several systems, including contractile tail bacteriophages, the type VI secretion system and R-type pyocins, use a multiprotein tubular apparatus to attach to and penetrate host cell membranes. This macromolecular machine resembles a stretched, coiled spring (or sheath) wound around a rigid tube with a spike-shaped protein at its tip. A baseplate structure, which is arguably the most complex part of this assembly, relays the contraction signal to the sheath. Here we present the atomic structure of the approximately 6-megadalton bacteriophage T4 baseplate in its pre- and post-host attachment states and explain the events that lead to sheath contraction in atomic detail. We establish the identity and function of a minimal set of components that is conserved in all contractile injection systems and show that the triggering mechanism is universally conserved.

Bacteriophages (viruses of bacteria) use a specialized organelle called a tail to deliver their genetic material and proteins across the cell envelope during infection^{1–4}. In phages with the most complex contractile tails, attachment to the host cell is accompanied by a substantial transformation of the tail structure: the external tail sheath contracts and drives a spike-tipped^{5,6}, rigid tube through the host cell membrane. Other macromolecular complexes, such as the type VI secretion system (T6SS)^{7–11}, metamorphosis-associated contractile (MAC) arrays¹², R-type pyocins^{13,14}, *Serratia* antifeeding prophage¹⁵, *Photorhabdus* virulence cassette¹⁶ and rhabidosomes^{1,17}, use a similar contractile sheath–rigid tube mechanism to breach the bacterial or eukaryotic cell envelope. However, the most complex part of these ‘contractile injection systems’¹⁸—the baseplate, which is responsible for coordinating host recognition or other environmental signals with sheath contraction—remains poorly characterized, and even its exact protein composition remains unknown for otherwise well studied systems such as bacteriophage T4 (refs 1, 8).

The T4 baseplate is currently thought to contain at least 15 different proteins with copy numbers ranging from 1 to 18 (ref. 19). Assembly of the T4 baseplate involves two large independent intermediates: a hub and a wedge^{20–22}. Several phage and possibly host cell chaperones mediate the joining of six wedges to the hub, which is then followed by the attachment of receptor-binding fibres to this structure. The nascent baseplate initiates tube assembly and subsequent polymerization of the sheath in the extended high-energy state^{23,24}.

The remarkable structure and transformation of the T4 tail and other contractile injection systems have received considerable attention^{2–5,10,12,15,25–32}. However, many questions remain unanswered, such as how the baseplates are organized at the molecular level, how they circularize, how the tail fibre network communicates with the central part of the baseplate, and how the signal of host cell surface binding leads to the conformational change that releases the tube and induces sheath contraction. To answer these questions, we used cryo-electron microscopy (cryo-EM) to determine the structure of the T4 baseplate–tail tube complex at a resolution of 3.8 Å for the universally conserved central part and tail tube and 4.1 Å overall, and the structure of the

post-attachment baseplate at a resolution of 6.8 Å. We have built and refined the atomic model of the baseplate in its pre- and post-host attachment conformations, allowing us to visualize the transformation of this 440,000-atom (not counting hydrogens) structure. We have analysed the signal transduction network of the baseplate by mutagenesis and examined the baseplate assembly pathway in the T6SS.

Overall structure of the T4 baseplate

We produced the pre-host attachment (also known as ‘hexagonal’³³) baseplate–tail tube complex *in vivo* by stopping phage assembly at the stage of sheath polymerization²⁶ (Extended Data Fig. 1). The complex switched to a post-attachment (also known as ‘star-shaped’³³) conformation upon interaction with the cell membrane and sometimes spontaneously; at this point it lost the tube and central spike complex (Extended Data Fig. 1).

The large size of the T4 baseplate (the diameter is 490 Å, about twice that of the ribosome) made particle alignment very accurate but contributed to a substantial difference in resolution at the centre and edges of the reconstructed volume (Extended Data Fig. 2). We used local mask-based particle alignment^{34,35} to improve the electron density around the periphery of the baseplate (Extended Data Fig. 2). Furthermore, because this region is made of mostly trimeric proteins, we were able to improve the maps by additional local three-fold averaging.

The final atomic model is around 96% complete and contains 56,082 amino acid residues (9,886 unique) (Supplementary Tables 1 and 2). These amino acids belong to 145 polypeptide chains of 15 different proteins (gene products, gps) that comprise the baseplate (gp5, gp5.4, gp6, gp7, gp8, gp9, gp10, gp11, gp12, gp25, gp27 and gp53) and the proximal region of the tail tube (gp19, gp48 and gp54, although gp48 and gp54 could also be considered to be baseplate components¹⁹; Fig. 1). The tape measure protein gp29, which is located inside the tube and is thought to span its entire length¹⁹, is not resolved in our cryo-EM map. Gp5, gp5.4 and gp27 form the central spike complex that creates an extension of the tube. The rest of the baseplate can be divided into inner (gp6, gp25, gp53 and part of gp7), intermediate (gp8 and the

¹École Polytechnique Fédérale de Lausanne (EPFL), BSP-415, 1015 Lausanne, Switzerland. ²Winogradsky Institute of Microbiology, Research Center of Biotechnology of the Russian Academy of Sciences, pr. 60-letiya Oktyabrya, 7 build. 2, 117312, Moscow, Russia. ³Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Laboratory of Molecular Bioengineering, 16/10 Miklukho-Maklaya St., 117997 Moscow, Russia. ⁴Center for Cellular Imaging and NanoAnalytics (C-CINA), Biozentrum, University of Basel, Mattenstrasse 26, 4058 Basel, Switzerland.
†Present address: Vertex Pharmaceuticals (Europe) Ltd, 86–88 Jubilee Avenue, Milton Park, Abingdon, Oxfordshire OX14 4RW, UK.

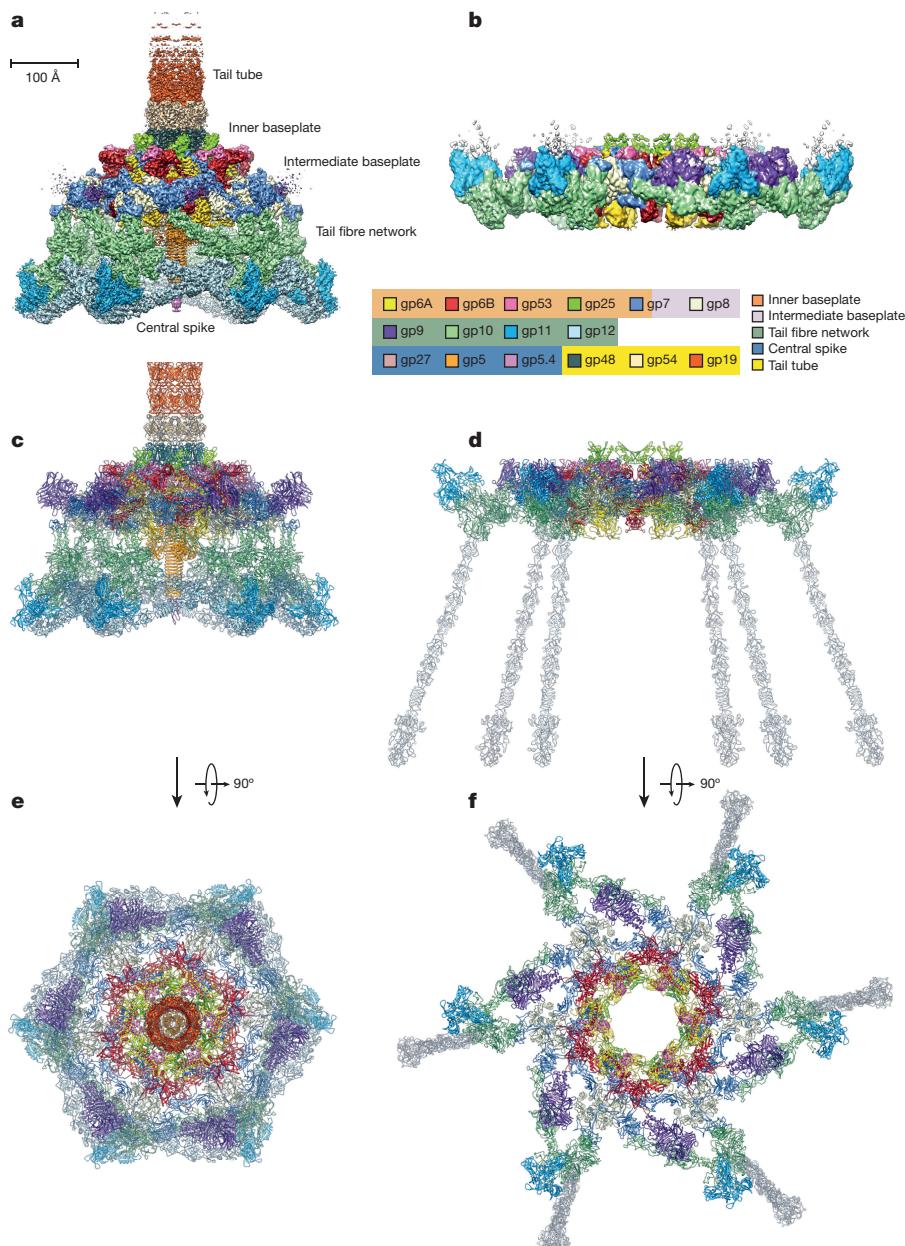


Figure 1 | Maps and atomic models of T4 baseplate in pre- and post-attachment states. **a, b**, Cryo-EM reconstructions of the pre- and post-attachment T4 baseplate, respectively. **c–f**, Atomic models of the two states

rest of gp7), and peripheral (gp9, gp10, gp11 and gp12) parts with the peripheral part constituting the tail fibre network (Fig. 1).

The inner baseplate

Twelve copies of gp6 form a continuous ring that makes up most of the inner baseplate (Extended Data Fig. 3). The asymmetric unit of the ring contains two gp6 molecules that are chemically identical but have markedly different conformations. One molecule (gp6A) lies closer to the six-fold axis whereas the other (gp6B) is further away (Extended Data Fig. 3a, b). Each gp6 molecule in the ring interacts with its two neighbours, forming an N-terminal dimer with one and a C-terminal dimer with the other (Extended Data Fig. 3c, d). An earlier study proposed a similar organization of the gp6 ring³⁶, and the crystal structure reported in that study matches the gp6 C-terminal dimer in the complete baseplate (Extended Data Fig. 3d). However, the interaction of the gp6 N-terminal regions is more complex than presumed earlier.

with component proteins shown as ribbon diagrams. In **d** and **f**, the STFs (gp12 trimers) are displayed semi-transparently to indicate that they are not present in the refined model of the post-attachment baseplate.

The N-terminal half of gp6 is part of a (gp6)₂-gp7 heterotrimeric module that consists of an α -helical ‘core bundle’ and a flat ‘trifurcation unit’ (Fig. 2). Gp25 and gp53 connect the core bundle to the central hub and the tube (Figs 2a, 3a), whereas the trifurcation unit allows the three polypeptide chains comprising it to reach different destinations (Fig. 2d). Two gp6 molecules extend in opposite directions to form C-terminal dimers that connect the trifurcation units into an iris-like structure (Fig. 2d). Gp7 follows a path towards the periphery of the baseplate, where it interacts with the tail fibre network proteins gp9 and gp10. The (gp6)₂-gp7 module plays a critical role in the assembly and function of the baseplate because it mediates both the circularization of the baseplate and the communication of the tail fibre network with the baseplate hub.

The (gp6)₂-gp7 heterotrimeric module contains large insertions that do not distort the gp6-gp7 iris (Fig. 2d). A wing-shaped domain (residues 101–315) is inserted between the core bundle and the trifurcation unit in gp6 (Fig. 2b). The wing domain of gp6A tightly cradles the hub,

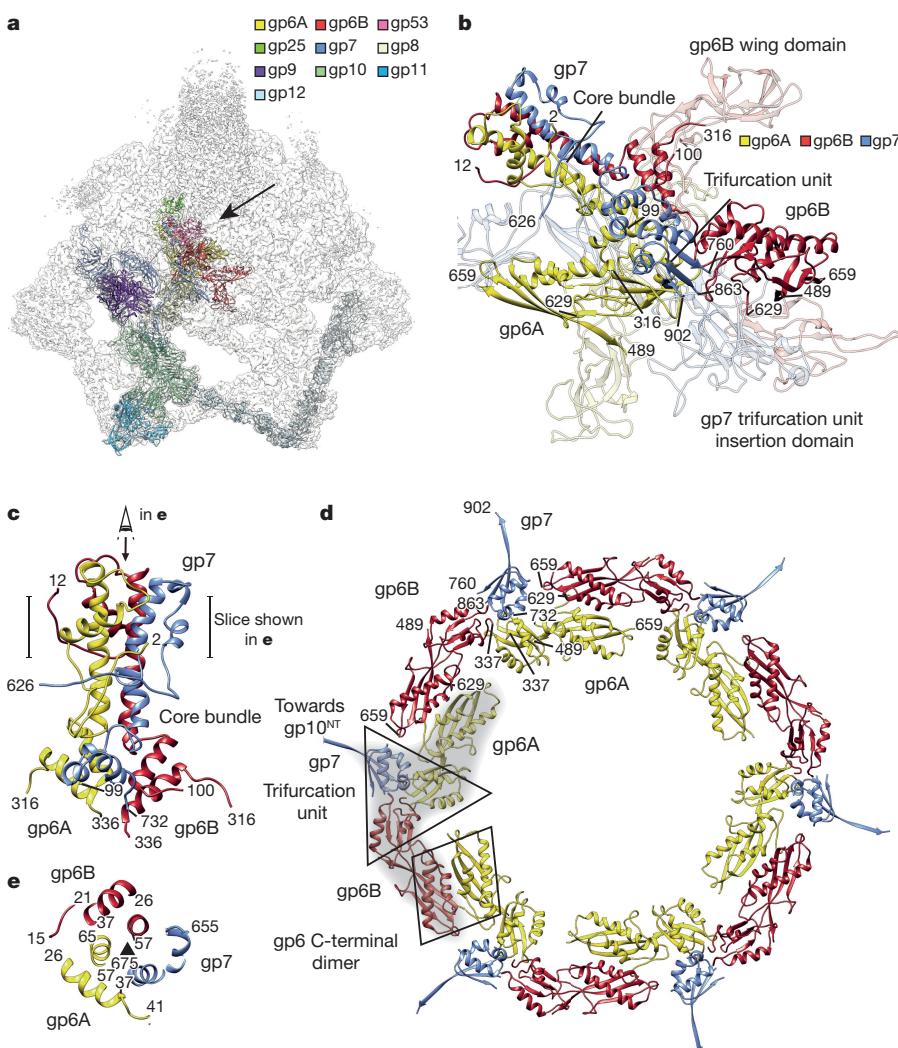


Figure 2 | The $(\text{gp6})_2$ -gp7 heterotrimeric unit. **a**, Composition of the largest assembly intermediate of the baseplate (the wedge) and its location in the cryo-EM map. Gp12 and gp9 are not part of the wedge and are shown as visual clues to help with interpretation. The arrow points to the $(\text{gp6})_2$ -gp7 heterotrimeric unit. **b**, The structure of the $(\text{gp6})_2$ -gp7 heterotrimer.

Regions with lower conservation are shown semi-transparently. **c**, A close-up view of the $(\text{gp6})_2$ -gp7 core bundle.

d, The iris created by the $(\text{gp6})_2$ -gp7 trifurcation domains and gp6 C-terminal dimers. **e**, A slice of the structure shown in **c** demonstrates three-fold symmetry.

whereas that of gp6B forms a segment of the platform for assembly of the first layer of the sheath (Extended Data Fig. 3c). In gp7, an insertion domain (residues 761–862) is part of the trifurcation unit. It is positioned close to the three-fold axis of the unit on the side opposite to the core bundle (Fig. 2b).

Gp25 and gp53 form extensive interactions with the core bundle that stabilize the entire assembly. Gp25 is firmly attached to the tip of the bundle (the buried surface area is $1,525 \text{ \AA}^2$) with its residues 1–28 capping the bundle (Fig. 3b). Gp53 consists of two domains, one of which (residues 1–103 and 155–193) clamps the core bundle and encircles it for about three-quarters of a turn (the buried surface area is $3,179 \text{ \AA}^2$) while the other extends the sheath assembly platform formed by the gp6B wing domain.

Both gp25 and gp53 contain conserved motifs that interact with each other and with other components of the core bundle. The most conserved region of gp53 and its orthologues is a LysM motif³⁷ (residues 46–82; Fig. 3b). The tip of a structurally conserved loop of gp25 (residues 47–49) mediates the interaction between LysM and the core bundle. Orthologues of gp25 across many contractile ejection systems contain an EPR motif (Glu-Pro-Arg, residues 85–87 of gp25)³⁸. This motif interacts with the core bundle (Fig. 3b) and points towards the region of the gp27–gp48 interface, which is not resolved in our cryo-EM map. The remarkable degree of conservation of this motif and

its location at the baseplate–tail tube interface suggest that it might perform an additional, possibly enzymatic, function, such as maturational proteolysis, which is widespread in virus assembly³⁹.

The central spike and the tube

The density of the trimeric part of the $(\text{gp5.4})_1$ – $(\text{gp5})_3$ – $(\text{gp27})_3$ central spike complex is a superposition of its two non-equivalent positions related by a 60° rotation around the six-fold axis of the baseplate. Nevertheless, the superimposed β -strands of the tandem tube domains of gp27 are perfectly resolved, confirming its three-to-six-fold symmetry adaptor function, as proposed in an earlier study⁵. Thanks to the prominent lysozyme domain of gp5, placement of the trimeric part of the spike in its two possible positions is unambiguous.

The tail tube is a smooth continuation of the spike complex, and gp48 and gp54 form its first two annuli (Fig. 3a). The tube body, as expected²³, is made of stacked gp19 hexamers. All the three proteins have a fold common to tube proteins from other contractile injection systems (for example, T6SS Hcp (ref. 11) or PA0623 of R-type pyocin¹⁴). The electron density is of sufficient quality to build the atomic model of the first two gp19 rings, allowing us to define the symmetry of the tube. Its helical rise and twist are 40.2 \AA and 17.9° , respectively. These parameters are very close to those of the extended sheath determined in previous, lower-resolution studies³ and probably represent the true

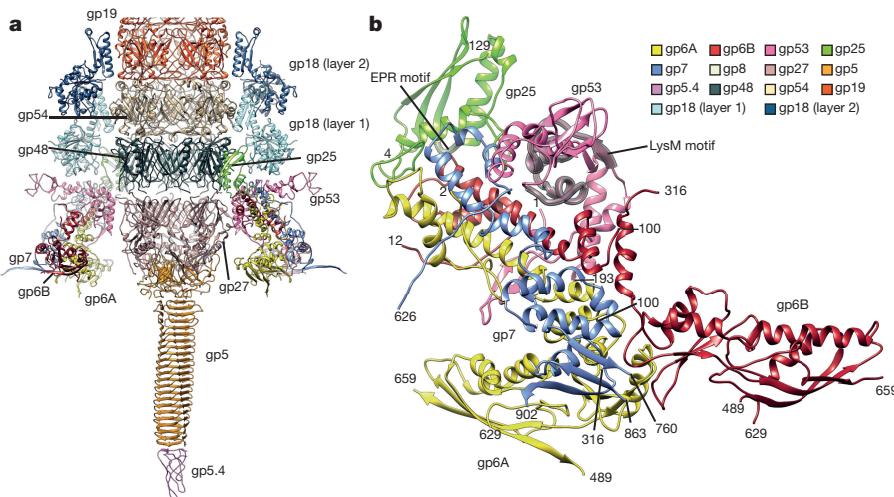


Figure 3 | Structure of the conserved inner baseplate. **a**, The minimal composition of a contractile injection system is derived from the T4 tail structure: the central spike complex (gp5, gp5.4, gp27), the conserved part of the wedge (gp6, gp7, gp25, gp53), the tail tube (gp19, gp48, gp54) and the conserved part of the sheath (gp18), some of which is modelled using

symmetry of the extended T4 tail. Gp54 and gp48 obey the helical symmetry of the tube. The rise and twist of superposition of the gp54 ring onto the gp19 and gp48 hexamers are 39.9 Å and 17.2°, and –38.8 Å and –22.1°, respectively.

In the complete tail, gp54 interacts with the first layer of sheath subunits (Fig. 3a), suggesting that gp54 is involved in the initiation of sheath assembly. However, most contractile injection systems lack gp54 orthologues and their first sheath layer must therefore bind to the main tube protein. Thus, the baseplate and the universally conserved gp25-like protein¹ in particular play a more important role in the initiation of sheath assembly in the extended state, as discussed further below.

the pycocin sheath¹⁴. **b**, A close-up of the structure of the conserved region of the wedge consisting of the (gp6)₂-gp7 heterotrimer, gp25 and gp53. The EPR motif of gp25 and the LysM motif of gp53 are highlighted with semi-transparent grey.

Notably, T6SS gene clusters do not appear to contain orthologues of either gp48 or gp54 (Supplementary Table 3), presumably leaving both functions to Hcp, the main tube protein of T6SS. Hcp could acquire new functions by binding an adaptor or effector protein to its ‘doughnut-hole’ surface⁴⁰. On the other hand, bacteria often carry several copies of *hcp* genes that can encode products with different functions.

The tail fibre network

The T4 baseplate distinguishes itself from its orthologues by being equipped with a very complex tail fibre network (Fig. 4). It consists of a gp9 trimer ((gp9)₃) to which a long tail fibre (LTF) comprising (gp34)₃,

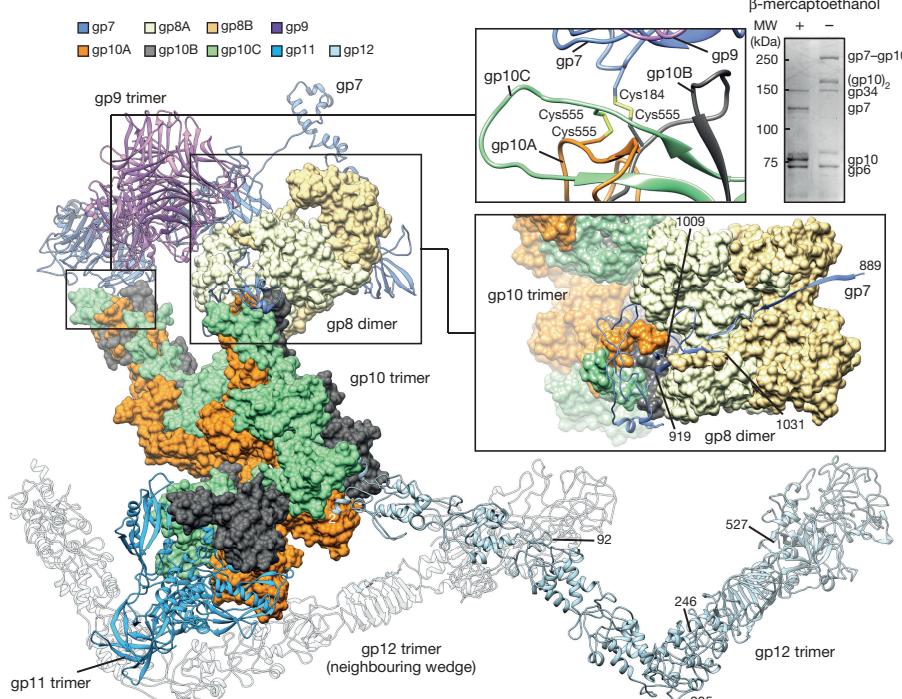


Figure 4 | The intermediate baseplate and the tail fibre network. Shown are gp7, gp8 (surface representation), gp9, gp10 (surface representation), gp11 and gp12. A symmetry-related gp12 trimer is displayed semi-transparently. Insets: 1) gp10–gp10 and gp10–gp7 disulfide bridges; 2)

SDS-PAGE of T4 baseplates in reducing and non-reducing conditions (the full gel is shown in Supplementary Fig. 1); 3) close-up of the interaction of gp7 with gp8. Residues 890–918 of gp7 form a long β-strand that runs across the surface of the gp8 dimer end-to-end.

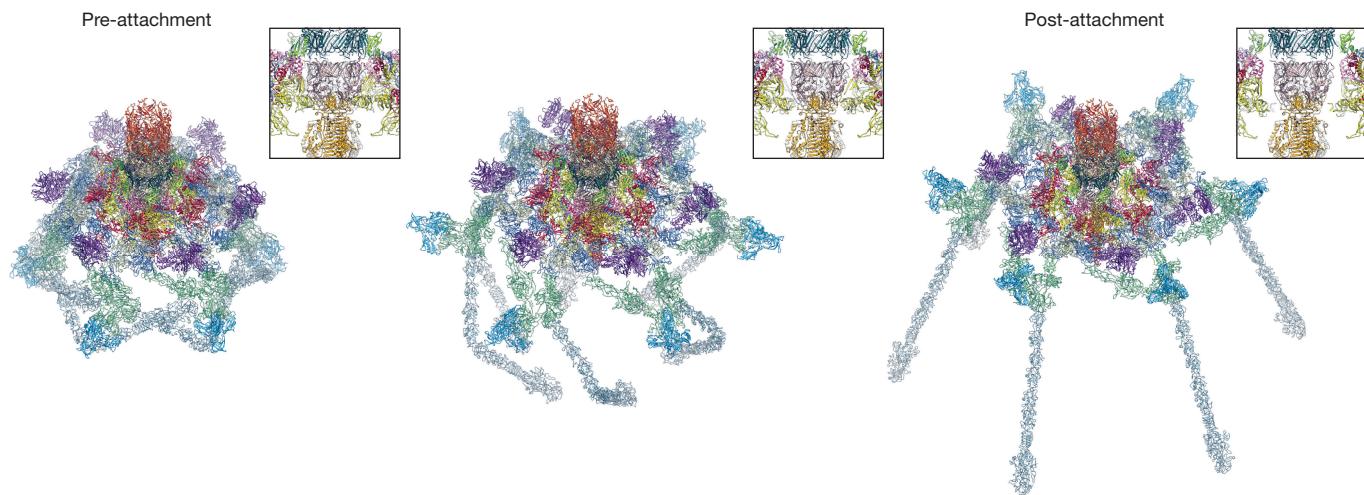


Figure 5 | Conformational change of the T4 baseplate upon host cell attachment. The left and right panels show the pre- and post-attachment structures that are derived from cryo-EM data. The middle image is a model of an intermediate with assumptions described in the

(gp35)₁, (gp36)₃ and (gp37)₃ is attached, and a (gp10)₃–(gp11)₃–(gp12)₃ module, in which (gp12)₃ is the short tail fibre (STF). The (gp10)₃–(gp11)₃–(gp12)₃ network forms a spring-like mechanism that extends upon interaction of the phage with a host cell. Our baseplate–tail tube complexes represent an intermediate of phage particle assembly and thus lack LTFs (except for a few aberrantly bound ones) because LTFs attach to otherwise complete phage particles at the very end of the particle assembly pathway²³. The cryo-EM density of the LTF-attachment protein gp9 is weak but could nevertheless be easily interpreted with the help of its crystal structure⁴¹ (Extended Data Fig. 2i).

Gp10 is key to understanding the function and evolution of the tail fibre network (Extended Data Fig. 4). This trimeric molecule has a distorted letter X shape, but each of its four domains (D1–D4) displays clear three-fold symmetry. A chain order swapping event takes place between D2 and D3 (Extended Data Fig. 4c). D1 is homologous to the N-terminal part of gp9 and this relationship could be detected at the level of the amino acid sequence. Unexpectedly, however, gp10 D2 and D3 are structurally similar to each other. They interact with the N-terminal motifs of gp12 and gp11, respectively, which themselves are also homologous (Extended Data Fig. 5a). Moreover, each of these units has a common three-fold axis (for example, gp10 D2–gp12) that orients the axes of the gp11 and gp12 trimers to be roughly perpendicular to each other. The X-like shape of gp10 allows each gp10–gp11 complex to interact with two STFs in the baseplate. One of the STFs extends D2 of gp10, and the other wraps around gp11 as described below.

Gp10 domains D1 and D4 attach the (gp10)₃–(gp11)₃–(gp12)₃ module to gp7 in the intermediate baseplate. Our structure shows, and mass spectrometry confirms, that apart from non-covalent interactions, the gp10 D4–gp7 interface involves a disulfide bridge between Cys555 of one gp10 chain and Cys184 of gp7 (Fig. 4). The Cys555 residues of the two remaining gp10 molecules form a disulfide bond (Extended Data Fig. 5b). This finding explains the gp10–gp10 inter-chain disulfide bridge found in the recombinant (gp10)₃–(gp11)₃ complex⁴². The gp7–gp10 and gp10–gp10 crosslinks are clearly seen in purified baseplates (Fig. 4). Cys555 of Gp10 and Cys184 of gp7 are conserved in all T4-like phages (Extended Data Fig. 5c, d), making the covalent attachment of the tail fibre network to the intermediate baseplate a hallmark of these viruses.

Using a multistep modelling procedure (see Methods), we were able to build the complete structure of the STF (Fig. 4 and Extended Data Fig. 6). Its previously unknown part (residues 2–245)^{43,44} contains six repeats (Extended Data Fig. 6a–c) that follow the N-terminal gp11-like motif (residues 2–47; Extended Data Fig. 5a). Two kinks allow the fibre to fold neatly around the periphery of the baseplate. The first bend

text. The insets show close-up views of the central part of the baseplate demonstrating a release of the tail tube–central spike complex, whose position is kept unchanged throughout the transformation.

(approximately 30°) is between repeats 1 and 2 where the fibre contacts the C-terminal domain of a symmetry-related mate. The second, much sharper bend (approximately 90°) is between repeats 3 and 4 where the fibre wraps around the gp11 trimer. In this region, the gp11-proximal chain of gp12 folds into an α -helix whereas the other two chains adopt an extended conformation.

The structural data presented above show that all trimeric proteins comprising the baseplate periphery and fibres contain structurally related domains (Extended Data Fig. 6d). Thus, despite their different functions in the T4 particle, it is possible that all these proteins evolved from a common ancestor by multiple gene duplication events. Notably, simpler baseplates contain only one trimeric protein (the host receptor binding fibre) that is attached to a (gp6)₂–gp7-like unit¹.

The inner baseplate is universally conserved

The structure of the T4 baseplate allows us to define the composition of baseplates in other contractile injection systems. Using HMM profiles and structures of T4 baseplate proteins in HHsuite^{45,46}, we found orthologues of all inner baseplate components in all systems except for a gp53-like component in T6SS (Supplementary Table 3). This component could nevertheless constitute a domain in one of the T6SS core bundle proteins (TssF or TssG) as both are larger than some of their phage orthologues. Such a rearrangement of functional domains is not uncommon and, for example, gp102 of bacteriophage A511 combines the functions of gp25 and gp53 into a single protein (Supplementary Table 3).

To test the validity of our bioinformatics-based assignment of functions, we examined the assembly of the T6SS baseplate, as it appears to be more distantly related to T4 than other systems. We found that, similar to T4, the assembly pathway of the T6SS baseplate involves a wedge-like intermediate that contains a (gp6)₂–gp7-like heterotrimer (Extended Data Fig. 7a–d; for additional details regarding T4 baseplate assembly, see Supplementary Information). The T6SS orthologues of T4 gp25, gp6 and gp7 (TssE, TssF and TssG, respectively) form a complex with TssK that has a stoichiometry of (TssE)₁–(TssF)₂–(TssG)₁–(TssK)₃ (Extended Data Fig. 7e, f). No complex was detected in the absence of TssK.

Considering the strength of bioinformatic predictions and the biochemical data, we conclude that the TssE–(TssF)₂–TssG unit forms the core bundle, the trifurcation unit, and the sheath polymerization platform in the T6SS baseplate. Furthermore, similar to the function of the gp10 trimer in T4 (ref. 23), trimeric TssK (ref. 47) appears to initiate the assembly of the T6SS wedge.

Conformational switch of the baseplate

Although the structure of a single state of a macromolecular machine as complex as the T4 baseplate can provide a plethora of information, to understand fully how it works one needs to describe the machine in its multiple states. Fortunately, a small fraction of the baseplates in our study switched to the post-host attachment state (Extended Data Fig. 1). We selected all such particles and calculated a reconstruction to 6.8 Å resolution. Virtually all domains from the pre-attachment hexagonal baseplate fit into this map as rigid bodies (see Methods). One notable exception is a loop formed by residues 841–862 of gp7. Cryo-EM density convincingly places this loop on two opposite sides of the gp7 trifurcation unit insertion domain in the two conformations (Extended Data Fig. 8a, b), indicating that it performs a ‘jump rope’ type of motion around this domain during the conformational switch. The location of this loop in our post-attachment structure is similar to that found in the 3.8 Å resolution reconstruction of the recombinantly assembled star-shaped baseplate⁴⁸.

To better understand and visualize structural changes that relate the two conformations, we aligned the two structures using gp53 molecules as a reference (see Methods) and performed interpolation with the help of UCSF Chimera⁴⁹ (Fig. 5 and Supplementary Video 1). This procedure resulted in intermediates with no grave clashes or chain breaks although it did not work for the STFs or the jump rope loop of gp7. Several intermediates had to be created as described in Methods.

All baseplate proteins participate in the conformational switch (Fig. 5 and Supplementary Video 1). The tail fibre network displays the largest amount of motion. The STFs unravel and straighten, as can be seen in raw images of baseplates attached to membranes (Extended Data Fig. 2j). The multivalent structure of the gp10 trimer plays a critical role in this transformation. Three-fold symmetrical gp10–gp11 and gp10–gp12 interfaces are preserved in the two conformations, suggesting that the gp10–gp11–gp12 complex rotates as a unit during the transformation. At the end of this rotation, the N-terminal domain of gp12 points straight towards the host cell surface. On the other hand, many interactions that lack three-fold symmetry in the pre-attachment state are broken: gp11 dissociates from the gp12 fibre that was originally wrapped around it (gp11 then binds the LTF protein gp34 in the post-attachment baseplate⁴); and the C-terminal domain gp12 dissociates from a symmetry-related gp12 fibre and binds to a receptor on the host cell surface.

The reorganization of the inner baseplate can be best understood by examining the motion of (gp6)₂–gp7 heterotrimers (Extended Data Fig. 9 and Supplementary Videos 2, 3). Gp7 transfers the rotation of the gp10–gp11–gp12 module to the (gp6)₂–gp7 heterotrimer, causing the latter also to rotate as a unit and tilt its three-fold axis by about 30°. As a result, the diameter of the (gp6)₂–gp7 iris increases and the proteins bound to the core bundle (gp25 and gp53) dissociate from the tube and the hub (Fig. 5), preparing the latter for subsequent sheath-driven translocation.

To analyse how the relative positions of different baseplate domains influence the baseplate conformational change and phage infectivity, we subjected gp7 to site-directed mutagenesis by extending or shortening some of its domain-linking regions by one amino acid (Extended Data Fig. 8 and Supplementary Information). We also attempted to remove the gp7–gp10 disulfide bridge. The region preceding the insertion domain of the trifurcation unit did not tolerate the two desired mutations, whereas small changes in the length of the linker upstream of the trifurcation unit and the jump rope loop did not affect phage infectivity (Extended Data Fig. 8). On the other hand, removal of the gp7–gp10 covalent linkage had a profound effect on infectivity, confirming that this bond is important for maintaining the structure of the baseplate during the conformational switch.

To extrapolate our results to an even larger framework, we used all available structural and functional data^{19,50} to construct an atomic model of the T4 tail in the extended and contracted states (see Methods, Extended Data Fig. 10). Gp25, gp54 and the gp6B–gp53 platform create

a complex interface that promotes the assembly of the first layer of the sheath in the extended conformation (Extended Data Fig. 10a). Gp25 serves as a nucleus for sheath polymerization because the structure and symmetry of the gp25–gp48 complex are repeated in every sheath–tube subunit interaction. In the assembled structure, gp25 is a seamless continuation of the lattice formed by gp25-like tube-interacting domains of the sheath. To achieve this, gp25 must accept the long N- and C-terminal arms of the first sheath subunit^{14,31,32}, thus firmly attaching the sheath to the baseplate (Extended Data Fig. 10; for additional details see Supplementary Information). In the contracted tail and the post-attachment baseplate, an outward motion and rotation of the core bundle with the associated proteins (gp25 and gp53) causes an even larger outward motion and around 45° rotation of sheath subunits²⁷ (Extended Data Fig. 10 and Supplementary Video 2). The sheath appears to maintain its original interaction with gp25, but now finds a different, presumably more energetically favourable, interface on the gp6B–gp53 platform (Extended Data Fig. 10b). Thus, in agreement with its high conservation in contractile injection systems¹, gp25 plays a critical role in sheath assembly and contraction.

The structural transformation of the baseplate described here involves proteins that are universally conserved in all contractile injection systems (except for the unknown T6SS gp53 orthologue). A similar sequence of events is likely to occur in any such system, regardless of the complexity of its peripheral baseplate or tail fibre network. Thus, the conformational switch consists of the following steps: an external factor triggers reorganization of the peripheral baseplate; the gp7 orthologue transfers this signal to a gp25–(gp6)₂–gp7-like module; and this module reorients, releasing the hub complex while simultaneously initiating sheath contraction.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 February; accepted 5 April 2016.

- Leiman, P. G. & Shneider, M. M. Contractile tail machines of bacteriophages. *Adv. Exp. Med. Biol.* **726**, 93–114 (2012).
- Hu, B., Margolin, W., Molineux, I. J. & Liu, J. Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proc. Natl Acad. Sci. USA* **112**, E4919–E4928 (2015).
- Kostyuchenko, V. A. et al. The tail structure of bacteriophage T4 and its mechanism of contraction. *Nature Struct. Mol. Biol.* **12**, 810–813 (2005).
- Leiman, P. G., Chipman, P. R., Kostyuchenko, V. A., Mesyanzhinov, V. V. & Rossmann, M. G. Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell* **118**, 419–429 (2004).
- Kanamaru, S. et al. Structure of the cell-puncturing device of bacteriophage T4. *Nature* **415**, 553–557 (2002).
- Browning, C., Shneider, M. M., Bowman, V. D., Schwarzer, D. & Leiman, P. G. Phage pierces the host cell membrane with the iron-loaded spike. *Structure* **20**, 326–339 (2012).
- Leiman, P. G. et al. Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc. Natl Acad. Sci. USA* **106**, 4154–4159 (2009).
- Basler, M. Type VI secretion system: secretion by a contractile nanomachine. *Phil. Trans. R. Soc. Lond. B* **370**, 20150021 (2015).
- Shneider, M. M. et al. PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. *Nature* **500**, 350–353 (2013).
- Basler, M., Pilhofer, M., Henderson, G. P., Jensen, G. J. & Mekalanos, J. J. Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* **483**, 182–186 (2012).
- Mougous, J. D. et al. A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science* **312**, 1526–1530 (2006).
- Shikuma, N. J. et al. Marine tubeworm metamorphosis induced by arrays of bacterial phage tail-like structures. *Science* **343**, 529–533 (2014).
- Nakayama, K. et al. The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. *Mol. Microbiol.* **38**, 213–231 (2000).
- Ge, P. et al. Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states. *Nature Struct. Mol. Biol.* **22**, 377–382 (2015).
- Heymann, J. B. et al. Three-dimensional structure of the toxin-delivery particle antifeeding prophage of *Serratia entomophila*. *J. Biol. Chem.* **288**, 25276–25284 (2013).
- Yang, G., Dowling, A. J. & Gerike, U. fffrench-Constant, R. H. & Waterfield, N. R. *Photobacterium* virulence cassettes confer injectable insecticidal activity against the wax moth. *J. Bacteriol.* **188**, 2254–2261 (2006).

17. Yamamoto, T. Presence of rhabdosomes in various species of bacteria and their morphological characteristics. *J. Bacteriol.* **94**, 1746–1756 (1967).
18. Bönenmann, G., Pietrosiuk, A. & Mogk, A. Tubules and donuts: a type VI secretion story. *Mol. Microbiol.* **76**, 815–821 (2010).
19. Leiman, P. G. et al. Morphogenesis of the T4 tail and tail fibers. *Virol. J.* **7**, 355 (2010).
20. Kikuchi, Y. & King, J. Genetic control of bacteriophage T4 baseplate morphogenesis. I. Sequential assembly of the major precursor, *in vivo* and *in vitro*. *J. Mol. Biol.* **99**, 645–672 (1975).
21. Kikuchi, Y. & King, J. Genetic control of bacteriophage T4 baseplate morphogenesis. II. Mutants unable to form the central part of the baseplate. *J. Mol. Biol.* **99**, 673–694 (1975).
22. Kikuchi, Y. & King, J. Genetic control of bacteriophage T4 baseplate morphogenesis. III. Formation of the central plug and overall assembly pathway. *J. Mol. Biol.* **99**, 695–716 (1975).
23. Coombs, D. H. & Arisaka, F. In *Molecular Biology of Bacteriophage T4* (ed. Karam, J. D.) 259–281 (American Society for Microbiology, 1994).
24. Kikuchi, Y. & King, J. Assembly of the tail of bacteriophage T4. *J. Supramol. Struct.* **3**, 24–38 (1975).
25. De Rosier, D. J. & Klug, A. Reconstruction of three dimensional structures from electron micrographs. *Nature* **217**, 130–134 (1968).
26. Kostyuchenko, V. A. et al. Three-dimensional structure of bacteriophage T4 baseplate. *Nature Struct. Biol.* **10**, 688–693 (2003).
27. Aksyuk, A. A. et al. The tail sheath structure of bacteriophage T4: a molecular machine for infecting bacteria. *EMBO J.* **28**, 821–829 (2009).
28. Amos, L. A. & Klug, A. Three-dimensional image reconstructions of the contractile tail of T4 bacteriophage. *J. Mol. Biol.* **99**, 51–64 (1975).
29. Effantin, G. et al. Cryo-electron microscopy three-dimensional structure of the jumbo phage PhiRS1 infecting the phytopathogen *Ralstonia solanacearum*. *Structure* **21**, 298–305 (2013).
30. Liu, J., Chen, C. Y., Shiomii, D., Niki, H. & Margolin, W. Visualization of bacteriophage P1 infection by cryo-electron tomography of tiny *Escherichia coli*. *Virology* **417**, 304–311 (2011).
31. Kudryashev, M. et al. Structure of the type VI secretion system contractile sheath. *Cell* **160**, 952–962 (2015).
32. Clemens, D. L., Ge, P., Lee, B. Y., Horwitz, M. A. & Zhou, Z. H. Atomic structure of T6SS reveals interlaced array essential to function. *Cell* **160**, 940–951 (2015).
33. Crowther, R. A., Lenk, E. V., Kikuchi, Y. & King, J. Molecular reorganization in the hexagon to star transition of the baseplate of bacteriophage T4. *J. Mol. Biol.* **116**, 489–523 (1977).
34. Amunts, A., Brown, A., Toots, J., Scheres, S. H. & Ramakrishnan, V. Ribosome. The structure of the human mitochondrial ribosome. *Science* **348**, 95–98 (2015).
35. Yan, C. et al. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **349**, 1182–1191 (2015).
36. Akxyuk, A. A., Leiman, P. G., Schneider, M. M., Mesyanzhinov, V. V. & Rossmann, M. G. The structure of gene product 6 of bacteriophage T4, the hinge-pin of the baseplate. *Structure* **17**, 800–808 (2009).
37. 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); correction **196**, 2122 (2014).
38. Lossi, N. S., Dajani, R., Freemont, P. & Filloux, A. Structure-function analysis of Hsif, a gp25-like component of the type VI secretion system, in *Pseudomonas aeruginosa*. *Microbiology* **157**, 3292–3305 (2011).
39. Veesler, D. & Johnson, J. E. Virus maturation. *Annu. Rev. Biophys.* **41**, 473–496 (2012).
40. Silverman, J. M. et al. Haemolysin coregulated protein is an exported receptor and chaperone of type VI secretion substrates. *Mol. Cell* **51**, 584–593 (2013).
41. Kostyuchenko, V. A. et al. The structure of bacteriophage T4 gene product 9: the trigger for tail contraction. *Structure* **7**, 1213–1222 (1999).
42. Zhao, L., Takeda, S., Leiman, P. G. & Arisaka, F. Stoichiometry and inter-subunit interaction of the wedge initiation complex, gp10-gp11, of bacteriophage T4. *Biochim. Biophys. Acta* **1479**, 286–292 (2000).
43. Thomassen, E. et al. The structure of the receptor-binding domain of the bacteriophage T4 short tail fibre reveals a knitted trimeric metal-binding fold. *J. Mol. Biol.* **331**, 361–373 (2003).
44. van Raaij, M. J., Schoehn, G., Burda, M. R. & Miller, S. Crystal structure of a heat and protease-stable part of the bacteriophage T4 short tail fibre. *J. Mol. Biol.* **314**, 1137–1146 (2001).
45. Söding, J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951–960 (2005).
46. Remmert, M., Biegert, A., Hauser, A. & Söding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods* **9**, 173–175 (2012).
47. Zoued, A. et al. TssK is a trimeric cytoplasmic protein interacting with components of both phage-like and membrane anchoring complexes of the type VI secretion system. *J. Biol. Chem.* **288**, 27031–27041 (2013).
48. Yap, M. L. et al. Role of bacteriophage T4 baseplate in regulating assembly and infection. *Proc. Natl. Acad. Sci. USA* **113**, 2654–2659 (2016).
49. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
50. Fokine, A. et al. The molecular architecture of the bacteriophage T4 neck. *J. Mol. Biol.* **425**, 1731–1744 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank V. Mesyanzhinov for developing the initial baseplate–tail tube complex purification protocol; C. Maillard for technical support; A. Brown for advice on structure refinement involving cryo-EM data; D. Demurtas for sample screening by negative stain EM; R. McLeod for help with data transfer; S. Nazarov, M. Plattner, and V. Kostyuchenko for discussions; and M. Basler for reading the manuscript. We acknowledge support from the EPFL SCITAS (high performance computing), the EPFL Centre for Interdisciplinary Electron Microscopy and the EPFL Proteomics Core Facility. The work was supported by the Swiss National Science Foundation grant 310000_144243.

Author Contributions N.M.I.T. purified the baseplates, calculated cryo-EM reconstructions, built and refined all atomic models, performed bioinformatic analyses and wrote the first draft of the paper. R.C.G.-F., N.M.I.T., K.N.G. and H.S. collected cryo-EM data. N.M.I.T. and N.S.P. designed, and N.S.P. performed, site-directed mutagenesis of T4 phage. M.M.S. designed, produced and analysed T6SS samples and T4 gp25. C.B. crystallized T4 gp25. P.G.L. solved the T4 gp25 structure, analysed the data from all sources, and integrated all the information into a single manuscript.

Author Information Cryo-EM maps have been deposited in the Electron Microscopy Data Bank under the following accession numbers: EMD-3374 and EMD-3396 for the pre-attachment and post-attachment baseplate, respectively, and EMD-3397, EMD-3392, EMD-3393, EMD-3394 and EMD-3395 for the locally masked reconstructions of the pre-attachment baseplate for the tail tube, inner, intermediate, upper and lower peripheral baseplate regions, respectively. Atomic coordinates have been deposited in the Protein Data Bank under the following accession numbers: 5IV5 and 5IV7 for the pre-attachment and post-attachment baseplate, respectively, and 4HRZ for the crystal structure of T4 gp25. The vector with the TEV-cleavable His-SlyD expression tag derived from pET-23d(+) (Novagen) has been deposited to the NCBI database under the accession number KU314761. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.G.L. (petr.leiman@epfl.ch).

METHODS

Purification of the T4 baseplate-tail tube complex. We purified T4 baseplates by improving a protocol established in a previous study²⁶ by implementing new tools, resins and media (Extended Data Fig. 1a). This protocol involves a phage mutant that carries the amber stop codon in the 5' region of the sheath and major capsid protein genes (T4-18am-23am). T4 particle assembly is ordered and infection of a non-amber-suppressing (Sup^0) cell strain by such a mutant results in accumulation of assembly intermediates up to the point of incorporation of the defective gene product (that is, baseplate-tail tube complexes and fibres in the case of T4-18am-23am; Extended Data Fig. 1b, c).

Six litres of Sup^0 *Escherichia coli* B^E cells were grown at 37 °C to an attenueance of 0.6 at 600 nm ($A_{600} = 0.6$) and infected by T4-18am-23am with a multiplicity of infection (M.O.I. or phage-to-cell ratio) of 5. Eight minutes later, the same amount of phage particles were added, resulting in lysis delay⁵¹ or 'superinfection'. The infected cells were incubated at 30 °C for an additional 150 min, collected by centrifugation and flash frozen in liquid nitrogen. The cells were then thawed and resuspended in 120 ml of lysis buffer (50 mM Tris-HCl pH 8.0, 1 mM MgCl₂, 30 µg ml⁻¹ DNase I, 100 µg ml⁻¹ lysozyme), and incubated at 30 °C for 30 min. Then, an equal volume of RNase buffer (50 mM Tris-HCl pH 8.0, 40 mM EDTA) containing RNase at 50 µg ml⁻¹ was added, and the cells were incubated at 30 °C for an additional 30 min.

This sample was concentrated to ~20 ml with the help of Centricon Plus-70 (30-kDa cut-off) filtration devices (Merck Millipore) and centrifuged at 50,000g overnight. The pellet was resuspended in purification buffer (50 mM Tris-HCl pH 8.0, 1 mM MgCl₂), loaded on a sucrose gradient (5–30% (w/v) with a 50% cushion in the same buffer), and centrifuged at 35,000g for 4 h. Fractions containing baseplates were pooled, diluted with the same buffer and centrifuged at 50,000g overnight. The pellet was resuspended in SM8 buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 8 mM MgSO₄), loaded on an OptiPrep (PROGEN Biotechnik) gradient (10–40% (w/v) with a 50% cushion) buffered by SM8 buffer and centrifuged at 35,000g for 24 h. Fractions containing baseplates were pooled and subjected to size-exclusion chromatography to remove the OptiPrep compound. A Superose 6 Increase 10/300 GL column (GE Healthcare Life Sciences) pre-equilibrated with SM8 buffer was used. Peak fractions were pooled and centrifuged at 50,000g for 4 h. The pellet was resuspended in about 200 µl of SM8 buffer, to a final concentration of around 1 mg ml⁻¹.

Cryo-EM sample preparation and data acquisition. T4 baseplates were applied to glow-discharged, 300-mesh, copper Quantifoil grids (Quantifoil). The grids were blotted for 3 s and vitrified by plunging into liquid ethane cooled down to liquid nitrogen temperature using a Vitrobot Mark IV (FEI). Cryo-EM samples were imaged using a K2 summit direct electron detector (Gatan) on an FEI Titan Krios at an acceleration voltage of 300 kV and equipped with a Gatan image filter (slit width 20 eV) (Supplementary Table 2). Images were collected at a defocus of around 1.5 µm and recorded in dose fractionation mode (40 frames) at a dose rate between 6 and 8 e⁻ per pixel per s, dose per image sub-frame between 1 and 1.5 e⁻ Å⁻² and a cumulative dose between 40 and 60 e⁻ Å⁻². The effective pixel size of the images was 1.32 Å per pixel at a nominal magnification of 105,000×. Frames 3–40 were aligned by motion correction⁵² as implemented in 2dx_automator⁵³.

Cryo-EM data processing. For all 1,699 micrographs, we calculated the defocus parameters using CTFFind4 (ref. 54) and discarded 78 micrographs that fell outside a 0.5–4-µM defocus window, had a cross-correlation of lower than 0.006 or did not have Thon rings extending to at least 8 Å. 47,516 pre-attachment and 6,900 post-attachment particles were picked using e2boxer.py in EMAN2 (ref. 55).

For the pre-attachment baseplate, particles were classified into 250 classes using reference-free 2D classification with RELION 1.4 (ref. 56). The 47,362 particles belonging to the best 41 classes were regrouped into 421 groups, and used to calculate the initial reconstruction using RELION 3D auto-refine. The previously published 12 Å resolution baseplate-tail tube complex (EMDB code 1048)²⁶ was filtered to 60 Å and used as an initial reference. This procedure gave a reconstruction with a resolution of 5.35 Å before and 4.36 Å after postprocessing employing automatic masking, with an empirically optimized binarization threshold of 0.11, automated B-factor sharpening and correction of the modulation transfer function (MTF)⁵⁷. To further improve the resolution, we performed alignment of the individual particle movie frames ('particle polishing'), using 47,297 particles after discarding those that had missing frames. This was followed by a new refinement using these polished ('shiny') particles. We regrouped the particles into 19 groups and repeated the 3D refinement. Then, particles with a maximum probability distribution value lower than 0.09 were discarded and the final round of refinement was run using 37,913 particles. This resulted in a reconstruction with a resolution of 4.65 Å. This map was postprocessed using automatic masking, an empirically optimized binarization threshold of 0.11, automated B-factor sharpening and

correction of the modulation transfer function (MTF). The postprocessed map had an overall resolution of 4.11 Å (ref. 57).

To perform localized reconstructions, we low-pass filtered the map from the final round of refinement to 8 Å, generated local masks by segmentation with Segger⁵⁸ in UCSF Chimera⁴⁹, and processed them with relion_mask_create. The subsequent refinement was performed by applying these local masks⁵⁴, similar to Yan *et al.*³⁵. This procedure resulted in maps with generally improved interpretability for most regions of the baseplate, with the exception of the central non-six-fold-symmetric central spike and the gp9 protein, which displays a very high conformational heterogeneity. These maps after postprocessing, and their FSC curves, are shown in Extended Data Fig. 2g, h.

Reconstruction of the post-attachment baseplate was initiated by grouping all images into 70 classes with RELION while limiting the high-resolution terms to 10 Å and performing phase flipping instead of the full CTF correction. The 5,188 particles belonging to the best 25 classes were selected and regrouped into 25 groups. These particles were then used for RELION 3D refinement. We created the starting model by extracting the baseplate from the previously published 17 Å reconstruction of the contracted T4 phage tail (EMDB code 1089)⁴ with the help of Segger⁵⁸ in UCSF Chimera⁴⁹. This map was filtered to 60 Å and subjected to the RELION 3D auto-refine procedure, resulting in a 9.50 Å resolution reconstruction. Postprocessing with an optimized binarization threshold of 0.15 improved the resolution to 7.49 Å. A total of 5,176 particles had complete stacks of movie frames, which allowed us to 'polish' them by aligning individual movie frames. The resulting map had a resolution of 6.77 Å. We further improved the interpretability of this map by postprocessing using the mask from the polishing procedure, MTF correction and automatic B-factor correction.

Model building: pre-attachment baseplate. All available crystal structures of baseplate proteins and their domains^{5,36,41,43,44,59–61} were fitted into the best post-processed map using Coot⁶² (Supplementary Table 1). Reliable homology models were generated with the help of HHpred⁶³ and Phyre2⁶⁴, fitted into the map with Coot and rebuilt. The rest of the models were built *de novo* also in Coot. Ramachandran restraints were used for local real space refinement, which was performed against the postprocessed overall map or the postprocessed localized refinement maps (which were aligned to the postprocessed overall map, and resampled using Chimera). For the tail tube protein gp19, we generated a model using MODELLER⁶⁵ based on HHalign⁴⁵ alignment with gp54. Finally, taking advantage of the technological development of the last 15 years⁶⁶, we re-refined the published structure of the gp12 fragment⁴⁴. The improved electron density allowed us to extend the existing model and complete the structure of the gp12 repeat unit (residues 220–253; Extended Data Fig. 6b, c). This structure was then used to model the other repeats with MODELLER, imposing three-fold symmetry restraints. The N-terminal residues of gp12 were modelled on the basis of alignment with the gp11 N-terminal residues found by HHpred, using MODELLER and three-fold symmetry restraints.

The cryo-EM map of gp10 domains 2 and 3 was one of the most challenging regions for model building. Therefore, we used mass spectroscopy to verify the *de novo* built model. A disulfide bond between cysteines 331 and 342 was first identified in our model and then confirmed by mass spectrometry, indicating the correctness of the atomic model even in difficult parts of electron density.

Model building: post-attachment baseplate. The model of the post-attachment baseplate was constructed with the help of the refined coordinates of the pre-attachment baseplate in the following way. Initially, the inner baseplate (gp6 (residues 2–104, 170–278 and 317–411), gp7 (residues 638–753), gp25 and gp53) was fitted as a rigid body. Then, gp8 and gp9 were located and the remaining domains of gp6 and gp7 were fitted with the help of the secondary structure elements visible in the map. Residues 841–862 of gp7 (the jump rope loop) had to be completely rebuilt.

In the peripheral baseplate, the following modules were fitted as rigid units: gp10 domain 1, (gp10 domain 2)–(gp12 N-terminal domain), (gp10 domain 3)–gp11, and gp10 domain 4. Subsequently, linker regions and a few loops were remodelled by rigid body fitting, regularization and real space refinement.

The final refined model does not contain gp12. The structure of the extended gp12 STF is obtained by assuming that the interaction between gp10 domain 2 and the N-terminal domain of gp12 is preserved and that gp12 simply straightens upon interaction with the host cell membrane (Extended Data Fig. 2j).

Refinement. Both structures (pre- and post-attachment baseplate) were refined with phenix.real_space_refine⁶⁷ using global real space refinement and reciprocal space B-factor refinement (two B-factors per residue; Supplementary Table 2). The following restraints were applied in the refinement: standard C_β, secondary structure, Ramachandran, global six-fold symmetry and local two- and three-fold symmetry for the following proteins: gp9, gp11, and gp19. In the pre-attachment structure, the central spike (gp5.4)₁–(gp5)₃–(gp27)₃ was modelled as alternate conformers in its six different symmetry-related positions. The pre-attachment

structure served as a reference model in the refinement of the post-attachment structure providing additional restraints.

To detect overfitting, we performed a limited random displacement of atoms in the final model with a mean of 0.5 Å and refined it against one of the two maps from the final round of cryo-EM image processing, each calculated with half of the data. Both models showed similar agreement with both maps confirming that no overfitting took place⁶⁸ (Extended Data Fig. 2c, f).

Baseplate conformational change modelling. The post-attachment baseplate was aligned to the pre-attachment baseplate by least-squares superposition of the centres of mass of six gp53 molecules in the two conformations using LSQKAB⁶⁹ from the CCP4 suite of programs⁷⁰. Then, we calculated a morph between the pre- and post-attachment baseplate by means of the corkscrew interpolation algorithm as implemented in UCSF Chimera⁴⁹. Only two regions of the entire baseplate structure had physically impossible intermediates: the three chains comprising the STF had independent trajectories resulting in their spatial separation, and the jump rope loop of gp7 (residues 841–862) overlapped with other atoms of the gp7 trifurcation unit insertion domain (residues 761–862). Splitting the STF into two separate trimeric parts comprising residues 2–98 and 99–527 resulted in a trajectory in which all three chains maintained their native association, but the fibre was now separated at its sharp kink region (residues 180–185). To overcome this problem, we created intermediates in which the kink region was manually straightened and regularized. Finally, the two ends of the jump rope loop of gp7 were rebuilt to allow its large motion around the trifurcation unit insertion domain, and the insertion domain had to be slightly displaced as a rigid body to accommodate this movement.

Pseudoatomic tail models: extended tail. The pseudo-atomic model of the T4 tail presented in Extended Data Fig. 10 was produced by fitting the structures—as rigid bodies—of tail proteins obtained in this and other studies^{5,19,27,36,41,43,44,50,59–61,71} into the cryo-EM map of the extended T4 tail (EMDB code 1126)³ with the help of UCSF Chimera⁴⁹. A crystal structure of the sheath protein gp18 lacking gp25-like Domain I¹ and a part of Domain II¹ is available (PDB code 3FOA)²⁷. We used HHpred⁶³, MODELLER⁶⁵, and the structure of the R-type pyocin sheath (PDB code 3J9Q)¹⁴ to model the missing fragments giving rise to a nearly complete atomic model of the sheath protein subunit. This structure was then fitted into the tail cryo-EM map. The tail tube was constructed by repetitively applying the symmetry transformation relating the first two gp19 hexamers in the atomic structure described in this work. The atomic model of the tail tube terminator protein gp3 (ref. 72) was derived from the structure of gp19 using the same tools^{63,65} and was assumed to obey the overall symmetry of the tube.

Pseudoatomic tail models: contracted tail. Several cryo-EM reconstructions of the T4 tail in the contracted state are available (EMDB codes 1086, 1089, 5528)^{4,50}. Of these, the highest resolution (EMDB code 1086) is at 17 Å resolution⁴, but it lacks the fibritin whisker proteins, which are present in a more recent study that has a resolution of 25 Å (EMDB code 5528)⁵⁰. To facilitate the interpretation of these complex maps, they were brought to a common origin with the structure of the post-attachment baseplate (which was itself aligned to the pre-attachment baseplate, as described above). We then fitted the atomic structure of the post-attachment baseplate, all the sheath subunits, and the tube from the extended tail into the 17 Å resolution map. The tail- and tube-terminating proteins (gp15 and gp3, respectively), were assumed to preserve their relative orientation in the contracted tail, thus defining the position of the tube. The fibritin model was then taken from the 25 Å resolution study.

Purification, crystallization and structure determination of T4 gp25. Our original attempt to express gp25 resulted in an insoluble protein (Extended Data Fig. 7g, lanes 2 and 3). To produce gp25 in soluble form, we expressed it as a downstream fusion to SlyD, a known metal-affinity column contaminant, peptidyl-prolyl isomerase and expression enhancer⁷³ (Extended Data Fig. 7g, lanes 4–8). The SlyD tag carried a polyhistidine sequence at its N terminus to increase its affinity to IMAC resin. The linker between SlyD and gp25 contained a cleavage site for the tobacco etch virus (TEV) protease. We used the TEV protease with a polyhistidine tag at its C terminus that we produce recombinantly in-house (Extended Data Fig. 7g, lane 9). The vector with the TEV-cleavable His-SlyD expression tag was derived from pET-23d(+) (Novagen). We named it pTSL and deposited it to the NCBI database under accession number KU314761.

For cloning into pTSL, gene 25 was PCR amplified using the primers given in Supplementary Table 4 and T4 DNA as a template. The expression vector was transformed into methionine-auxotrophic B834/DE3 cells with ampicillin as a selection marker. Protein expression was performed in 2 × TY medium supplemented with ampicillin at 200 mg ml⁻¹. The plasmid-carrying cells were grown to A_{600} of 0.6 and induced by addition of IPTG to a final concentration of 0.5 mM (Extended Data Fig. 7g, lanes 4, 5). The expression was performed at 18 °C with rigorous shaking overnight. The cells were pelleted at 8,000g, resuspended in lysis buffer

(50 mM Tris-HCl pH 8.0, 300 mM NaCl), and disrupted by sonication at 4–16 °C. The insoluble fraction was removed by centrifugation at 27,000g for 20 min. The clarified cell lysate was loaded onto a 5-ml GE HisTrap FF Ni-charged column, and the protein was eluted with imidazole-containing buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 200 mM imidazole) using a step gradient (Extended Data Fig. 7g, lane 6). Fractions containing the protein were pooled and the TEV protease was added to give a protease/protein ratio of 1/100 (w/w). This mixture was dialysed into 10 mM Tris-HCl pH 8.0 by overnight incubation at 20 °C resulting in cleavage of the His-SlyD expression tag (Extended Data Fig. 7g, lane 7). The digested protein was filtered and purified by ion-exchange chromatography using a MonoQ 10/100 GL column and 0 to 1 M NaCl gradient in 20 mM Tris-HCl pH 8.0 buffer. Fractions containing gp25 were pooled and loaded onto a Superdex 75 HiLoad 16/60 size-exclusion column equilibrated with 10 mM Tris-HCl pH 8.0 and 150 mM NaCl.

For crystallization, the protein was dialysed into crystallization buffer (20 mM Tris-HCl at pH 8.0, 50 mM NaCl), concentrated to 20 mg ml⁻¹, and crystallized in 0.9–1.0 M MgSO₄, 100 mM MES pH 6.2. The structure was solved by the single-wavelength anomalous dispersion method with the help of a SelenoMethionine (SeMet) derivative. The latter was produced using the SelenoMethionine Medium Complete kit (Molecular Dimensions). The SeMet derivative was purified and crystallized in a similar way to the native protein. Crystallographic data were collected at the Swiss Light Source PX-I (X06SA) beamline using a PILATUS-6 M (DECTRIS Ltd.) pixel detector. For data collection, a crystal was briefly washed in cryoprotectant solution that contained all the components of the well solution at the same concentrations and 25% ethylene glycol and flash frozen in a vaporized liquid N₂ stream at 100 K. The X-ray data were processed with MOSFLM⁷⁴, SCALA⁷⁵ and POINTLESS⁷⁶ (Supplementary Table 5). The selenium atoms were located with the help of SHELXD⁷⁷. The atomic model was built and refined with COOT⁶² and PHENIX⁷⁸.

Purification of TssE-TssF-TssG-TssK complex. Recombinant expression of *E. coli* CFT073 gene c3400, which was proposed to encode a TssF-TssG fusion⁷⁹, results in two proteins with molecular masses of ~65 kDa and ~40 kDa. The sequence of CFT073 'gene' c3400 is identical to that of genes CP009072 of *E. coli* ATCC 25922 and CP007799 of *E. coli* Nissle 1917, except for two missing G nucleotides after positions 3245777 and 3245794. We resequenced this fragment and confirmed that the missing nucleotides represent sequencing or genome assembly errors. We named the two correct genes c3400.1 and c3400.2. These genes encode proteins that show 100% sequence identity to *E. coli* TssF and TssG with NCBI references WP_000342476.1 and WP_000373314.1, respectively.

Two PCR product-containing *E. coli* CFT073 genes, c3402 (TssE) and c3400.1 (TssF), c3400.2 (TssG) were fused by overlapping PCR using the primers given in Supplementary Table 4 to create a single construct with c3402 upstream of c3400.1 and c3400.2. The resulting PCR product was cloned into a modified pBAD33 vector that resembled the pTSL vector described above. TssE carried a TEV protease-cleavable His-SlyD tag at its N terminus. Gene c3387 (TssK) was cloned into the vector pBAD24 without any tags. Both plasmids were co-transformed into *E. coli* TOP10 cells using appropriate antibiotic selection. These cells were then used for protein expression.

Protocols used for expression and purification of the TssE-TssF-TssG-TssK complex were similar to those used for gp25 with small modifications. The 2 × TY medium was supplemented with ampicillin at 200 mg ml⁻¹ and kanamycin at 50 mg ml⁻¹. The plasmid-carrying cells were grown to an A_{600} of 0.8 and induced with arabinose to a final concentration of 0.1% (w/v). The expression was performed at 18 °C with rigorous shaking overnight. The cells were pelleted at 8000g, resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl), and disrupted by sonication at 4–16 °C. The insoluble fraction was removed by centrifugation at 27,000g for 20 min. The clarified cell lysate was loaded onto a 5-ml GE HisTrap FF Ni-charged column. The protein was eluted with imidazole-containing buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 200 mM imidazole) using a step gradient. Fractions containing the protein were pooled and the TEV protease was added to give a protease/protein ratio of 1/100 (w/w). This mixture was dialysed into 10 mM Tris-HCl pH 8.0 by overnight incubation at 20 °C resulting in cleavage of the His-SlyD expression tag. The digested protein was filtered and purified by ion-exchange chromatography using a MonoQ 10/100 GL column and 0 to 1 M NaCl gradient in 20 mM Tris-HCl pH 8.0 buffer. Fractions containing the TssE-TssF-TssG-TssK complex were pooled and loaded onto a Superdex 200 HiLoad 16/60 size-exclusion column equilibrated with 10 mM Tris-HCl pH 8.0 and 150 mM NaCl.

The TssE-TssF-TssG-TssK complex had an apparent molecular mass of 260 kDa (90% confidence interval is 203–331 kDa) in size-exclusion chromatography on a Superdex 200 10/30 GL column in 10 mM Tris-HCl, pH 8.0, and 150 mM NaCl (Extended Data Fig. 7e, f).

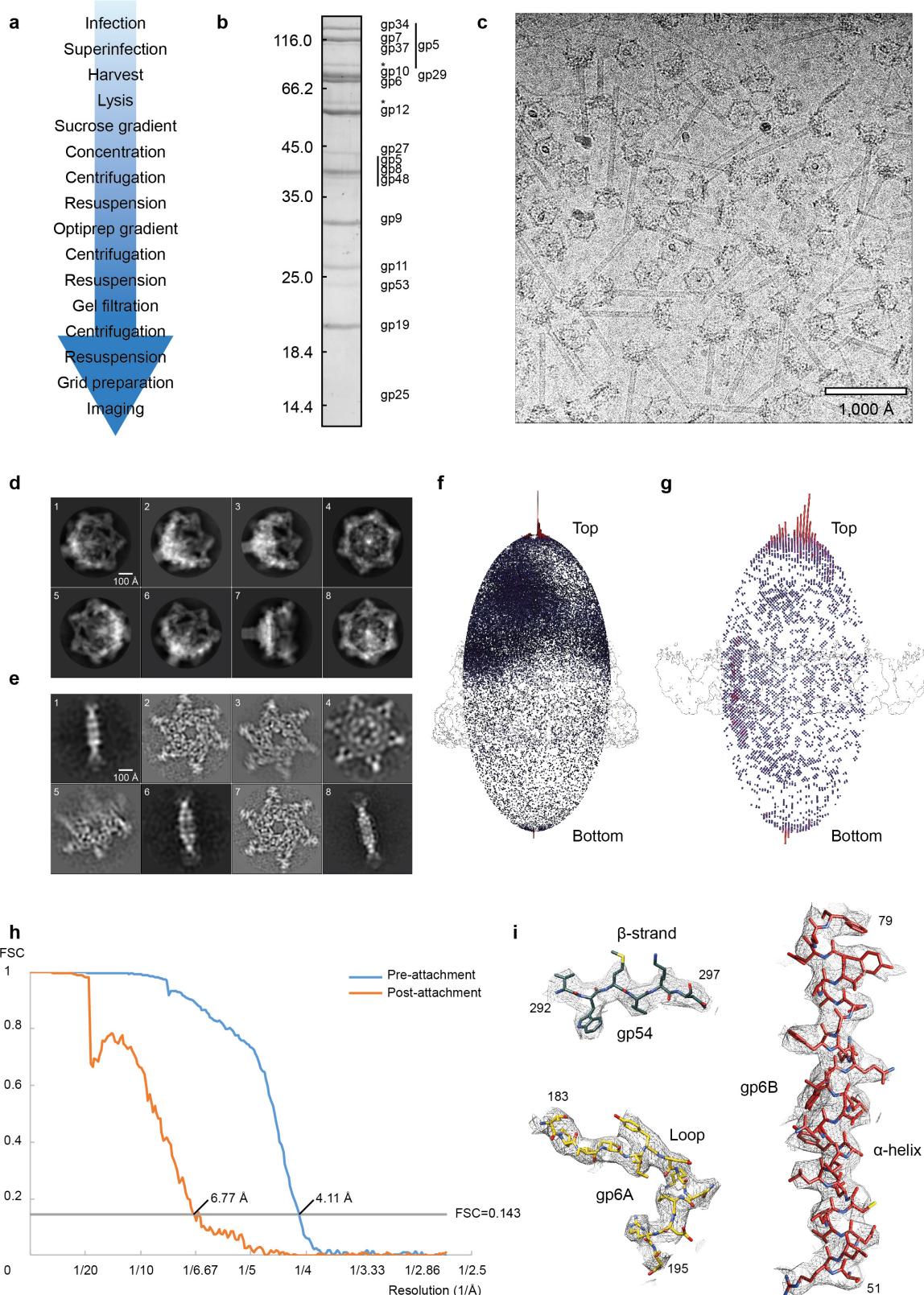
Construction of T4-7am mutant phage. T4-7am mutant phage was constructed by replacing Ser45 of gene 7 with the amber stop codon by means of homologous recombination with a plasmid carrying this mutation. First, an overlap PCR with primers 7S45AmF, 7S45AmRM, 7S45AmFM and 7S45AmR (Supplementary Table 4) and T4 genomic DNA as a template was performed as described elsewhere⁸⁰. The resulting PCR product was then ligated into the pGem-T vector (Promega) following the manufacturer's recommendations, giving rise to a pG7-45Am plasmid that carried the desired mutation.

For homologous recombination, pG7-45Am was electroporated into the amber suppressor *supD* *E. coli* strain CR63 (ref. 81) (serU60(AS)), and the cells were infected with phage T4 with a M.O.I. of 5. After the completion of lysis, the lysate was diluted to obtain an appropriate plaque count when plated onto a CR63 lawn. Several hundred plaques were tested for their ability to grow on CR63 and Sup⁰ BE lawns and a T4-7am mutant phage was found. The presence of the mutation was confirmed by Sanger sequencing.

Complementation assay. The complementation was performed by infecting BE^E cells, which expressed the wild-type gp7 protein and its mutants, with the T4-7am mutant phage. First, gene 7 was PCR amplified using T4 genomic DNA as a template and primers T4-7F1 and T4-7R1 (Supplementary Table 4). The PCR fragment was then cloned into the pBAD24 vector to give rise to a plasmid in which gene 7 was under the control of an arabinose promoter. The plasmids for expression of gp7 mutants were constructed by overlap extension with pBAD forward and reverse primers (Invitrogen) as flanking oligonucleotides and the primers given in Supplementary Table 4 as mutagenic oligonucleotides⁸⁰. The wild-type gene 7 plasmid served as a template for all these PCR reactions. The mutations were confirmed by partial sequencing, and the expression levels were analysed by SDS-PAGE⁸⁰.

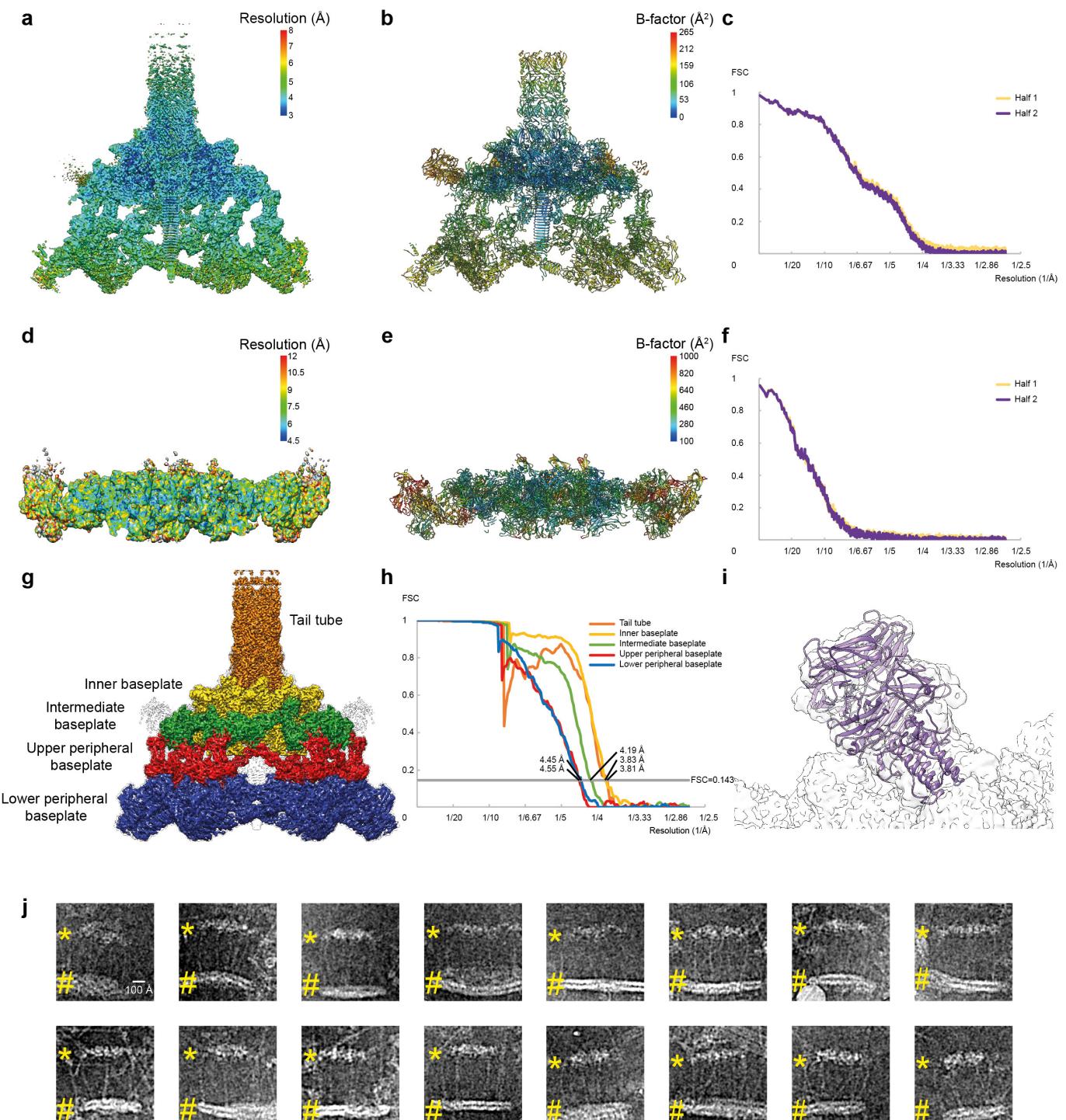
The plasmids expressing all variants of gp7 and an empty pBAD24 vector (for negative control) were transformed into the BE^E strain by electroporation. Night cultures were diluted 100 times into an LB medium supplemented with ampicillin (Sigma) at 100 mg ml⁻¹ and then incubated at 37 °C with shaking at 220 r.p.m. When the cultures reached an *A*₆₀₀ of 0.150 they were induced by the addition of arabinose (Sigma) to a concentration of 0.02% (w/v). When the cells reached an *A*₆₀₀ of 0.600, T4-7am was added with a M.O.I. of 5. Superinfection was performed 10 min later at the same M.O.I. After 4 h, a few drops of chloroform were added to complete cell lysis. The phages were harvested and purified by centrifugation in a stepped gradient of caesium chloride⁸² and then dialysed into SM buffer (50 mM TrisCl pH 7.5, 100 mM NaCl, 8 mM MgSO₄). The concentration of the samples was normalized according to their absorbance at 260 nm and then examined using SDS-PAGE (Extended Data Fig. 9b). The infectivity was tested by plaque assay on *E. coli* CR63 (Extended Data Fig. 9c) and *E. coli* BE^E (for negative control).

51. Doermann, A. H. Lysis and lysis inhibition with *Escherichia coli* bacteriophage. *J. Bacteriol.* **55**, 257–276 (1948).
52. Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**, 584–590 (2013).
53. Scherer, S. et al. 2dx_automator: implementation of a semiautomatic high-throughput high-resolution cryo-electron crystallography pipeline. *J. Struct. Biol.* **186**, 302–307 (2014).
54. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
55. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
56. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
57. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
58. Pintilie, G. D., Zhang, J., Goddard, T. D., Chiu, W. & Gossard, D. C. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J. Struct. Biol.* **170**, 427–438 (2010).
59. Leiman, P. G. et al. Structure of bacteriophage T4 gene product 11, the interface between the baseplate and short tail fibers. *J. Mol. Biol.* **301**, 975–985 (2000).
60. Leiman, P. G., Shneider, M. M., Mesyanzhinov, V. V. & Rossmann, M. G. Evolution of bacteriophage tails: Structure of T4 gene product 10. *J. Mol. Biol.* **358**, 912–921 (2006).
61. Leiman, P. G. et al. Structure and location of gene product 8 in the bacteriophage T4 baseplate. *J. Mol. Biol.* **328**, 821–833 (2003).
62. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
63. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).
64. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845–858 (2015).
65. Eswar, N. et al. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* **50**, 2.9:2.9.1–2.9.31 (2007).
66. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463 (1999).
67. Afonine, P. V., Headd, J. J., Terwilliger, T. C. & Adams, P. D. New tool: phenix_real_space_refine. *Comput. Crystallogr. Newslett.* **4**, 43–44 (2013).
68. Brown, A. et al. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D* **71**, 136–153 (2015).
69. Kabsch, W. Solution for best rotation to relate 2 sets of vectors. *Acta Crystallogr. A* **32**, 922–923 (1976).
70. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
71. Tao, Y., Strelkov, S. V., Mesyanzhinov, V. V. & Rossmann, M. G. Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. *Structure* **5**, 789–798 (1997).
72. Abuladze, N. K., Gingery, M., Tsai, J. & Eiserling, F. A. Tail length determination in bacteriophage T4. *Virology* **199**, 301–310 (1994).
73. Han, K. Y. et al. Solubilization of aggregation-prone heterologous proteins by covalent fusion of stress-responsive *Escherichia coli* protein, SlyD. *Protein Eng. Des. Sel.* **20**, 543–549 (2007).
74. Leslie, A. G. W. & Powell, H. R. In *Evolving Methods for Macromolecular Crystallography* (eds Read, R. J. & Sussman, J. L.) vol. 245, 41–51 (Springer, 2007).
75. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
76. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D* **67**, 282–292 (2011).
77. Sheldrick, G. M. A short history of SHELDXL. *Acta Crystallogr. A* **64**, 112–122 (2008).
78. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
79. Brunet, Y. R., Zoued, A., Boyer, F., Douzi, B. & Cascales, E. The type VI secretion TssEFGK-VgrG phage-like baseplate is recruited to the TssJLM membrane complex via multiple contacts and serves as assembly platform for tail tube/sheath polymerization. *PLoS Genet.* **11**, e1005545 (2015).
80. Sambrook, J. & Russell, D. W. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory Press, 2001).
81. Appleyard, R. K., McGregor, J. F. & Baird, K. M. Mutation to extended host range and the occurrence of phenotypic mixing in the temperate coliphage lambda. *Virology* **2**, 565–574 (1956).
82. Clokie, M. R. J. & Kropinski, A. M. *Bacteriophages: Methods and Protocols*, (Humana Press, 2009).
83. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature Methods* **11**, 63–65 (2014).
84. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
85. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).



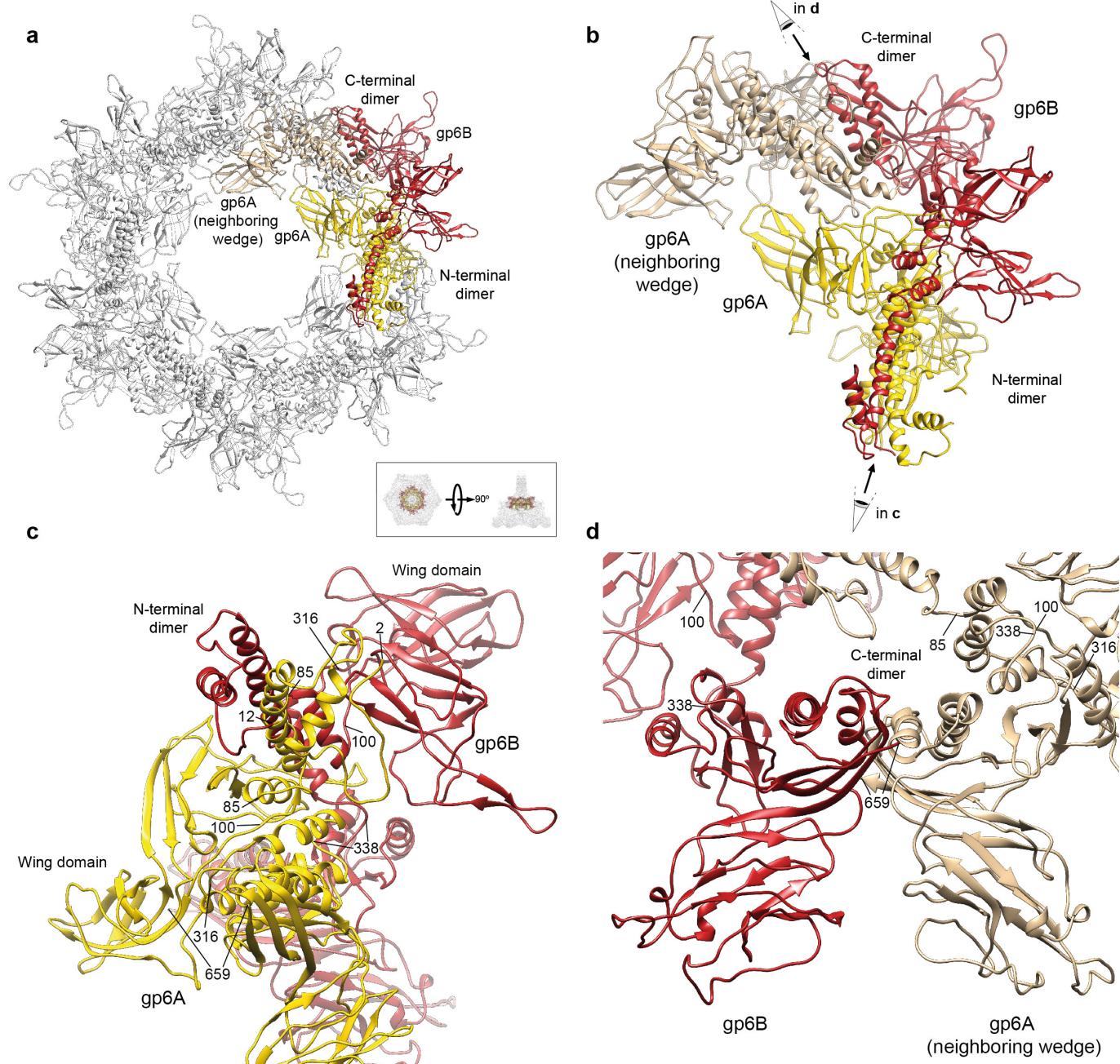
Extended Data Figure 1 | Sample purification, cryo-EM imaging and reconstruction. **a**, Purification scheme of the T4 baseplate–tail tube complexes. **b**, SDS–PAGE of the sample used in cryo-EM imaging (the full gel is shown in Supplementary Fig. 1). **c**, Raw cryo-EM image of T4 baseplates. **d**, **e**, Representative reference-free 2D class averages of T4 baseplates in pre- and post-attachment conformations, respectively. The number of particles in each class in **d** is as follows: 1: 534; 2: 813; 3: 689;

4: 655; 5: 292; 6: 858; 7: 2,223; 8: 977. The number of particles in each class in **e** is as follows: 1: 315; 2: 68; 3: 494; 4: 60; 5: 263; 6: 387; 7: 62; 8: 566. **f**, **g**, Distribution of refined angles of the baseplate in both conformations. **h**, Fourier shell correlation (FSC) between independently refined maps calculated using half of the data (gold-standard refinement) after post-processing for both conformations. **i**, Fragments of the pre-attachment baseplate cryo-EM reconstruction map with fitted atomic model.



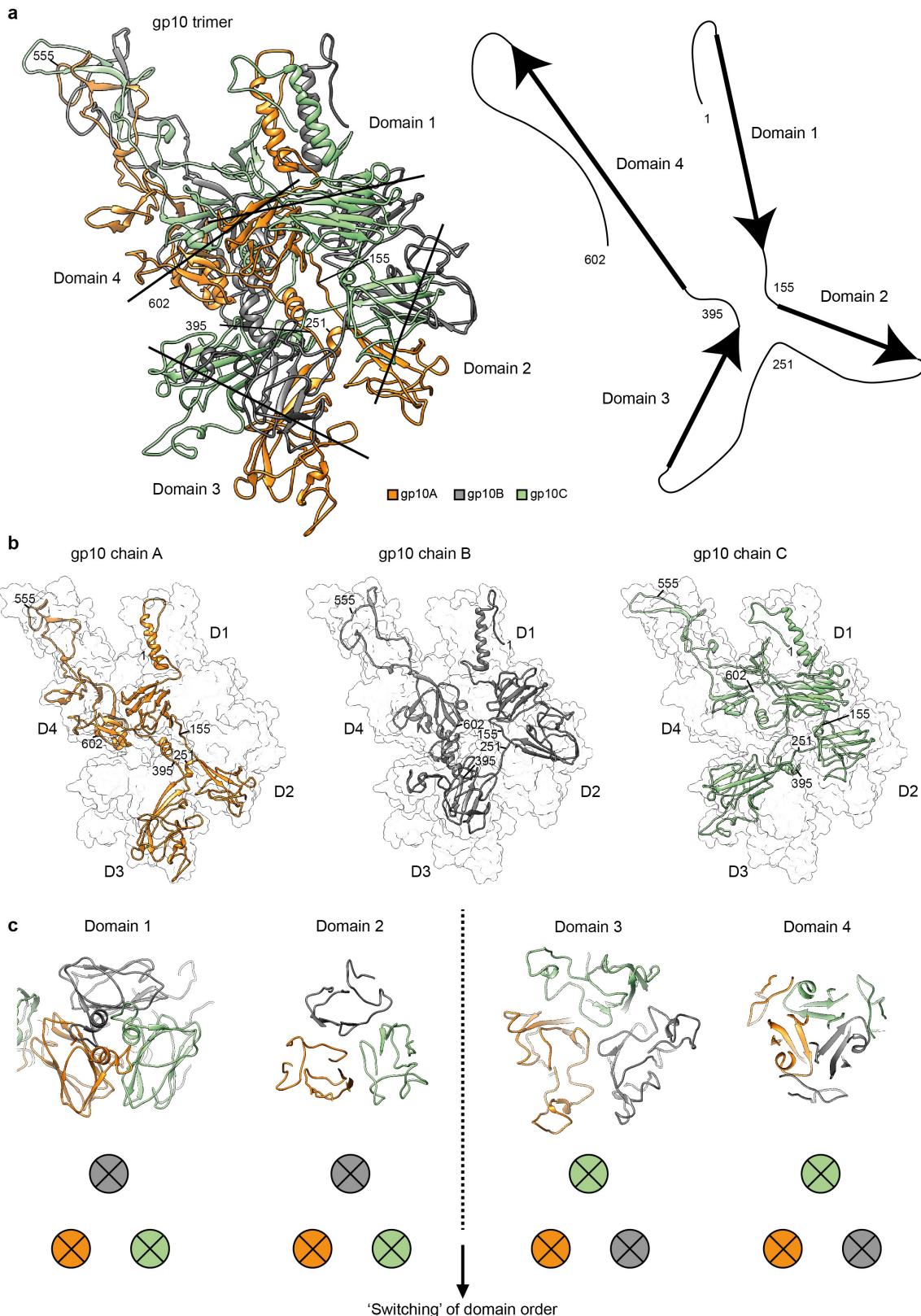
Extended Data Figure 2 | Details of local resolution estimation and baseplates attached to membranes. **a, d**, Resolution of pre- and post-attachment reconstructions analysed with ResMap⁸³. **b, e**, Atomic models of pre- and post-attachment baseplates coloured by B-factors. **c**, Plots of the model-map FSC of the scrambled pre-attachment baseplate structure refined against half data map 1 from the gold-standard refinement versus half data map 1 (the map it was refined against) and half data map 2 (against which it was not refined). The absence of a large gap between both

curves indicates that no excessive overfitting took place. **f**, As in **c**, but for the post-attachment baseplate. **g**, Localized cryo-EM refinement maps of the pre-attachment baseplate. **h**, FSC graphs of the localized refinements in **g**. **i**, Fit of gp9 into the pre-attachment baseplate map (the contour level is lower than that in Fig. 1a). **j**, Gaussian low-pass filtered ($1/25 \text{\AA}$) raw images of individual baseplates attached to cell membranes selected from $n = 243$ similar images. Asterisks, baseplates; hash symbols, membranes. The extended STFs connecting the two can be seen.



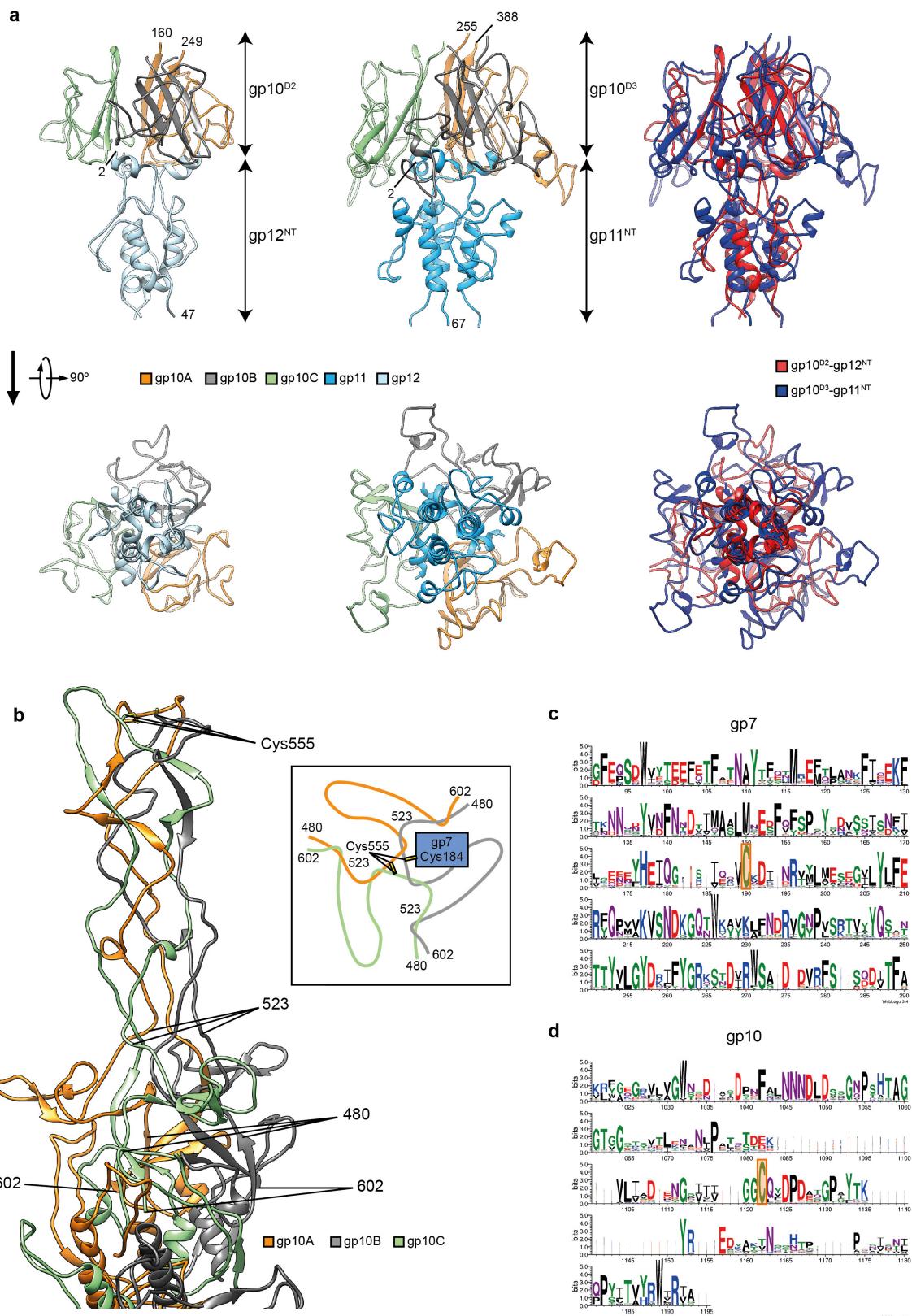
Extended Data Figure 3 | Structure of the gp6 ring. **a**, Two structurally non-identical copies comprising the asymmetric unit of the ring are coloured yellow (gp6A) and red (gp6B). A neighbouring gp6A copy is coloured light brown. The rest of the ring is shown in light grey. The

inset shows the position of the ring in the baseplate map. **b**, A close-up view of the gp6 subunits coloured in **a**. **c**, **d**, Close-up views of the N- and C-terminal gp6 dimers, respectively. In both views, the six-fold axis of the baseplate is roughly vertical.



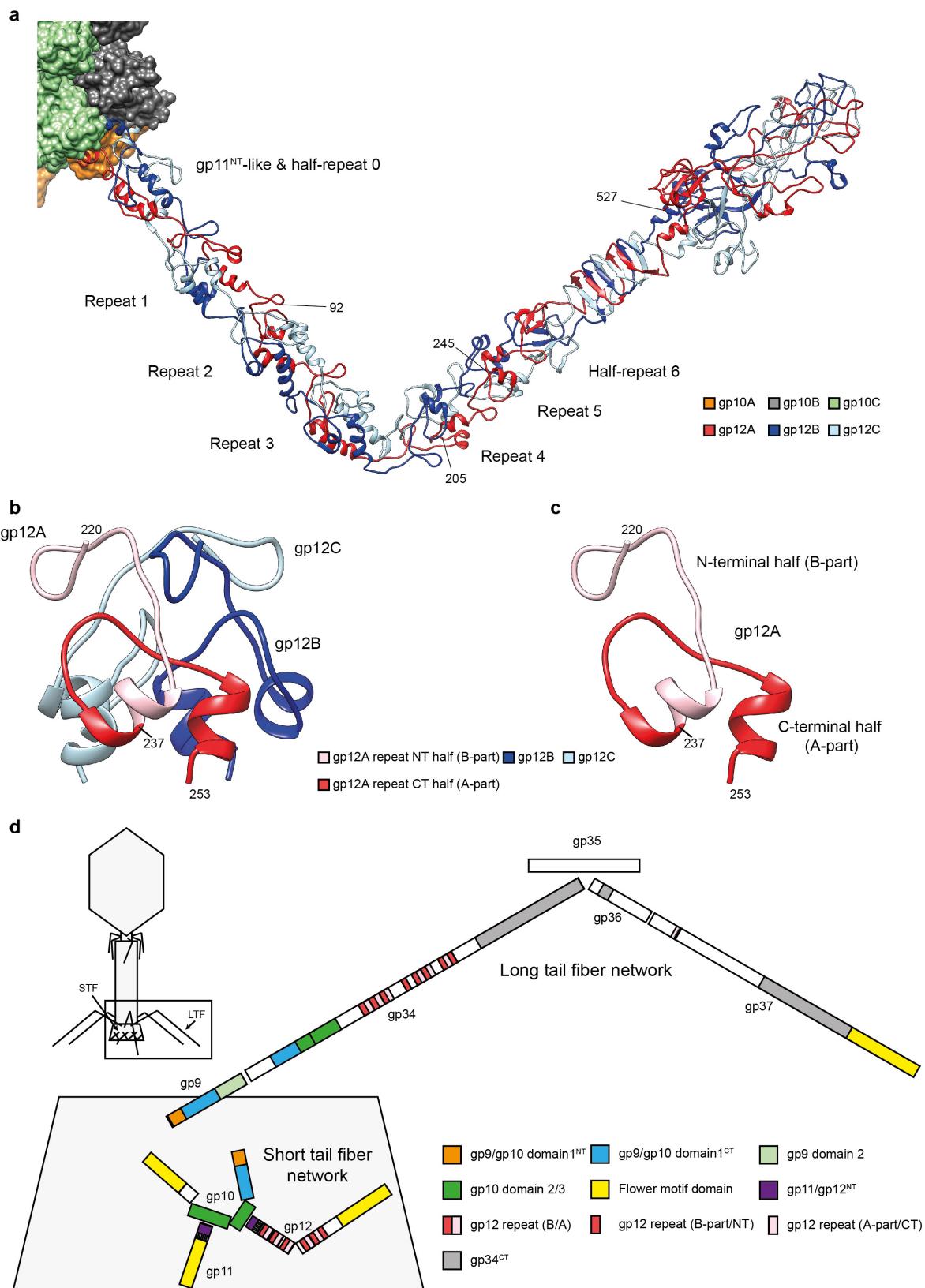
Extended Data Figure 4 | Structure of gp10. **a**, Ribbon diagram of gp10 trimer and a schematic showing the general direction of the polypeptide chain in its four domains. The arrows also indicate the viewing direction in **c**. **b**, Traces of the three chemically identical polypeptide chains making

up the complete gp10 molecule, which is shown as a semitransparent surface. **c**, Cross-sections through the four domains of gp10 in positions indicated with black lines in **a**. Note the switch of chain order around the three-fold axis between domains 2 and 3.



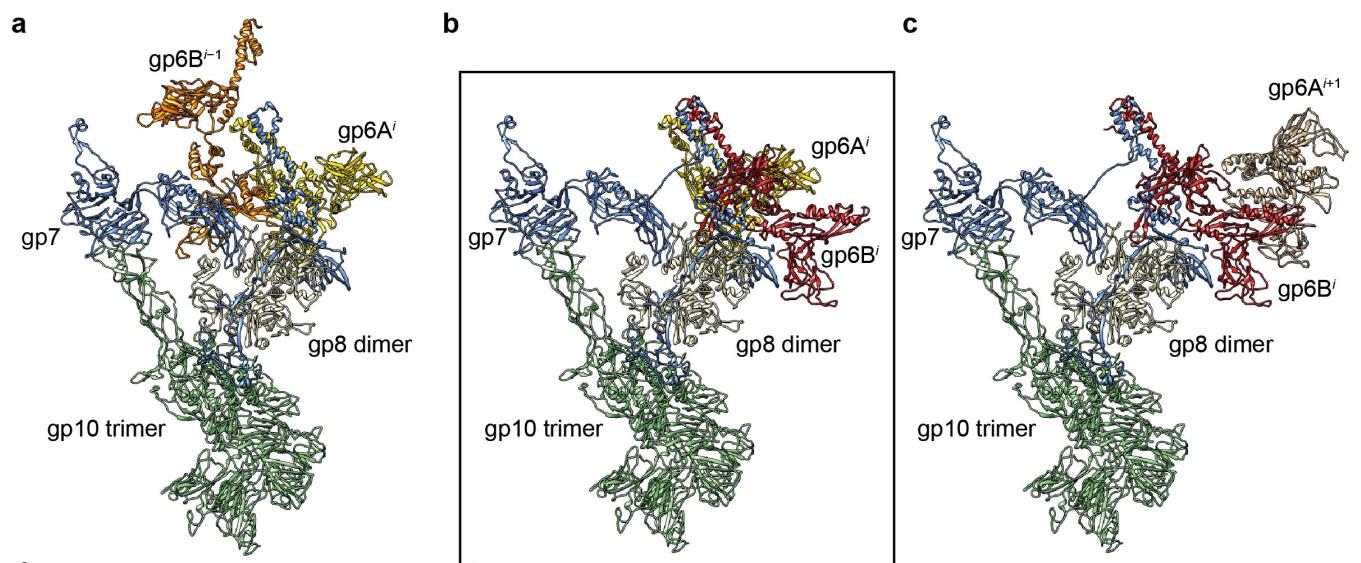
Extended Data Figure 5 | Interaction of gp10 with gp11 and gp12.
a, Structure and binding of gp10 domains 2 and 3 to the N-terminal domains of gp12 and gp11, respectively. The rightmost panel shows a superposition of the two complexes. **b**, The structure of gp10 domain 4. Cys555 and several other residues at strategic locations are labelled. Inset: a diagram explaining the complex knotted topology of gp10 domain 4.

The gp10-gp10 and gp10-gp7 inter-chain disulfide bridges are indicated. **c, d**, Conservation of gp7 and gp10 amino acid sequences, respectively, is represented in the WebLogo format with letter heights proportional to the degree of conservation⁸⁴. The conserved cysteines are highlighted with orange boxes. The numbers below the letters are positions in a multiple sequence alignment.

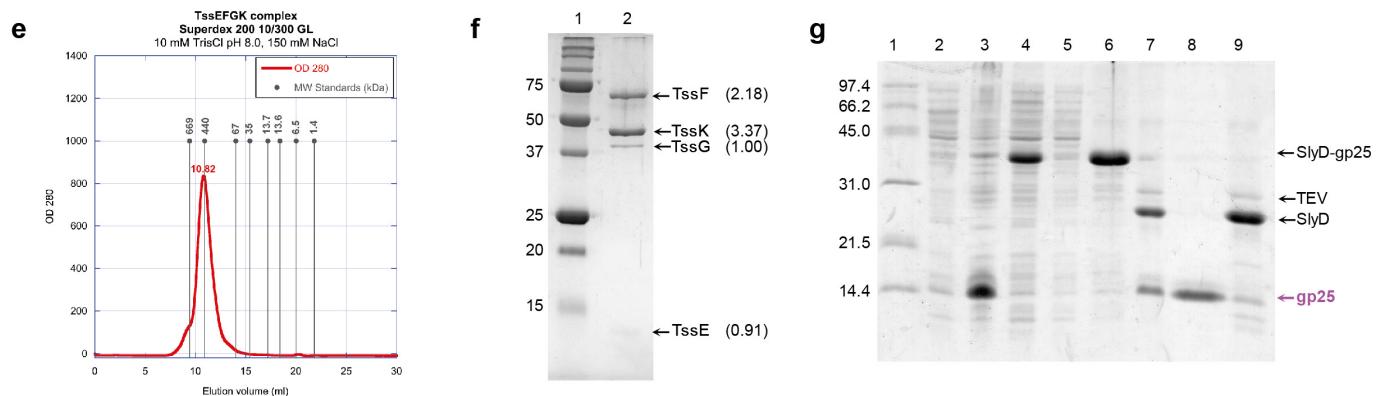


Extended Data Figure 6 | Structure of gp12 and domain organization of T4 fibres. **a**, Ribbon diagram of the gp12 trimer, anchored to gp10 domain 2 (shown in surface representation). The N-terminal part of the fibre (residues 2–245) was built *de novo*. **b**, Structure of the gp12 repeat. **c**, Fold of the polypeptide chain making up the repeat. **d**, Evolutionary

relationships between different T4 proteins comprising the baseplate's periphery and fibres. The size of each bar is proportional to the amino acid sequence length. The gp12 repeat shown in **b** and **c** constitutes a major part of the proximal LTF protein gp34.

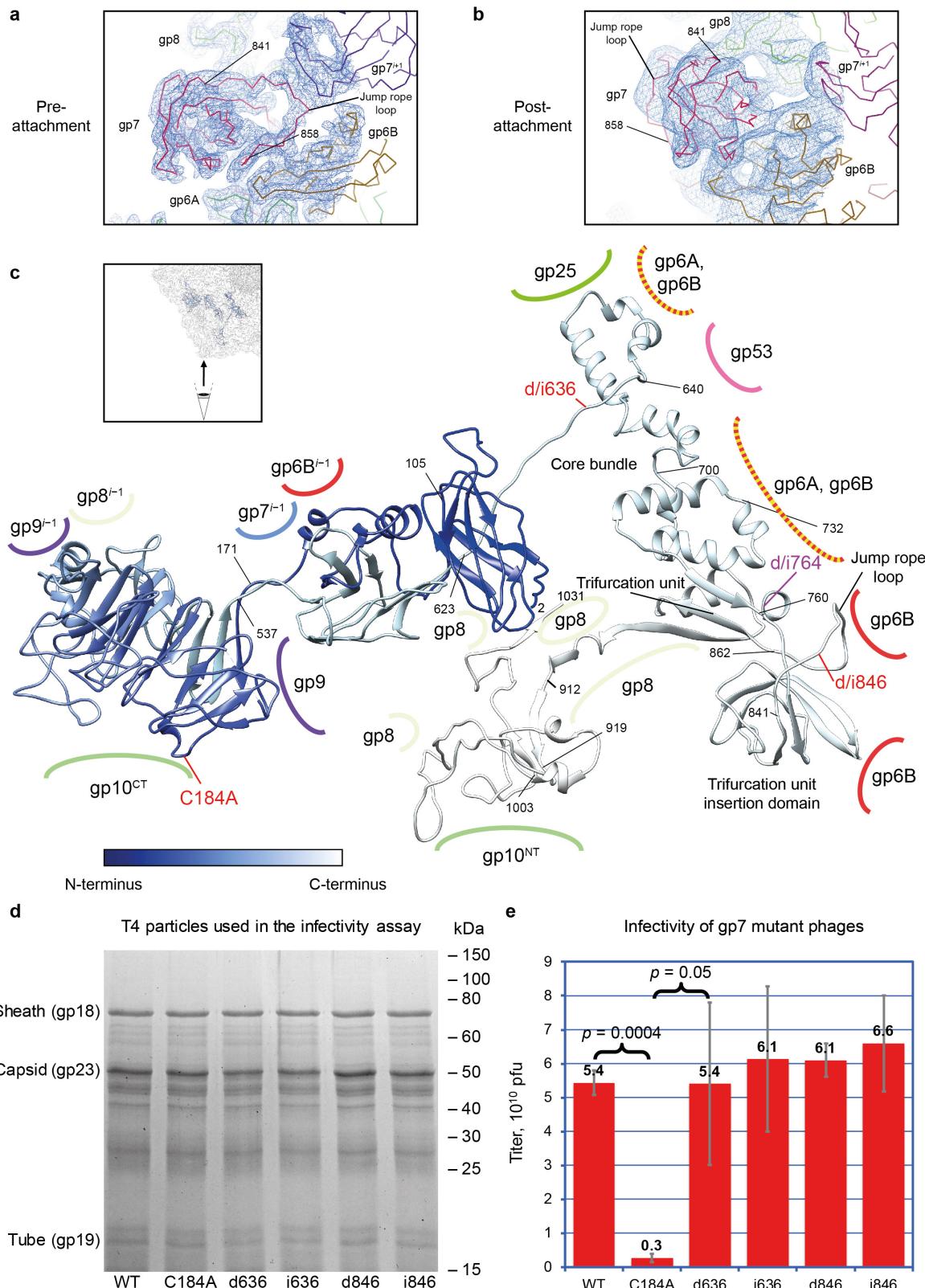


	Potential wedge intermediate A	Potential wedge intermediate B	Potential wedge intermediate C
BSA (\AA^2)	73,018.2	81,929.6	72,413.9
ΔG_{diss}^0 (kcal/mol)	-2.3	36.0	-6.3



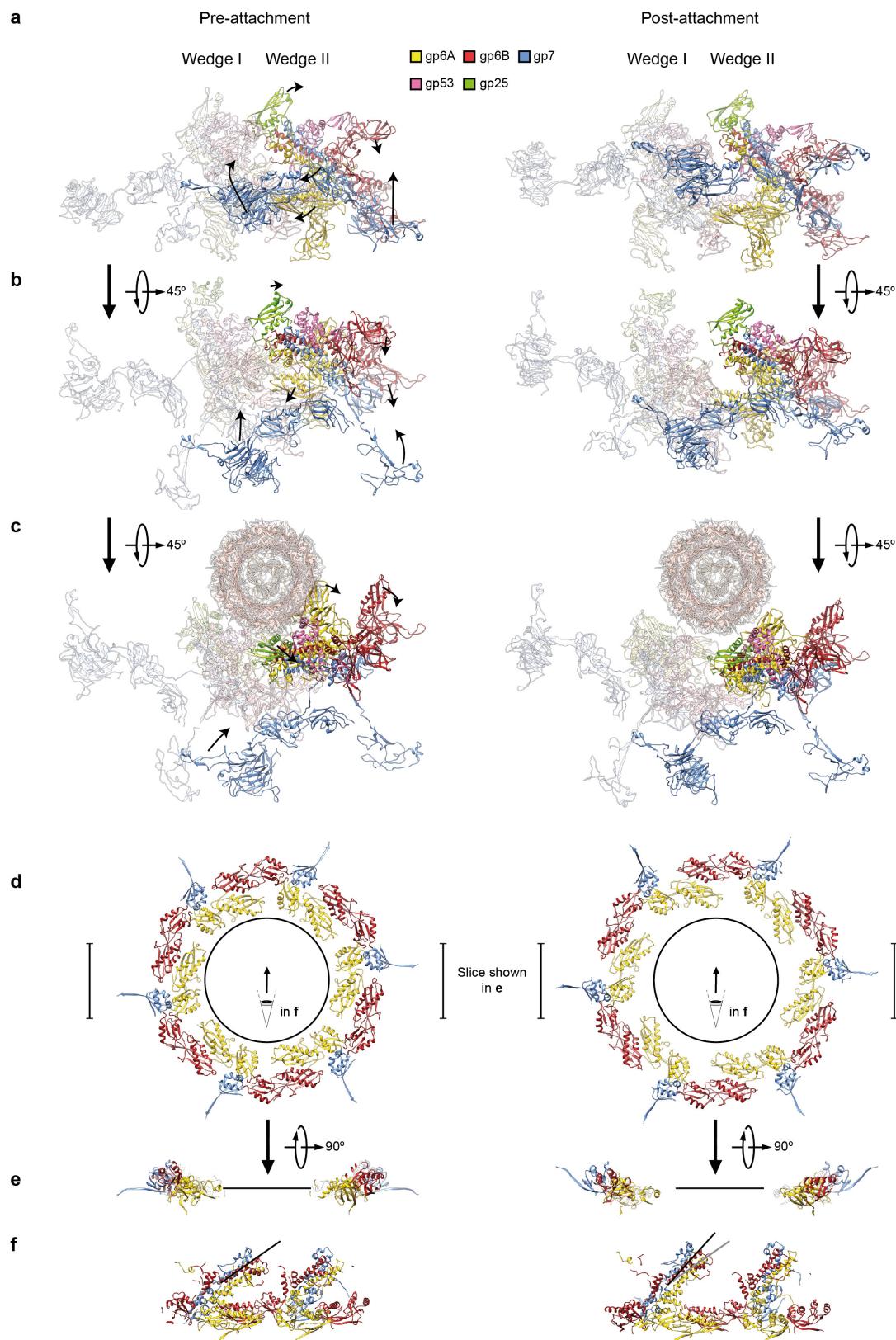
Extended Data Figure 7 | T4 and T6SS baseplate assembly.
Determination of the structure of the wedge precursor on biophysical grounds (see Supplementary Information). **a–c**, Three possible precursors with a composition of $(\text{gp6})_2-\text{gp7}-(\text{gp8})_2-(\text{gp10})_3$ and association as found in the pre-attachment baseplate. **d**, Calculations with PISA⁸⁵ favour the assembly shown in **b**. **e**, Size-exclusion chromatography of $(\text{TssE})_1-(\text{TssF})_2-(\text{TssG})_1-(\text{TssK})_3$ complex. **f**, 12% SDS-PAGE of

$(\text{TssE})_1-(\text{TssF})_2-(\text{TssG})_1-(\text{TssK})_3$ complex. Numbers on the left are MW standards in kDa. Numbers in parentheses to the right of protein names are relative band intensities as quantified by Image Studio Lite (LI-COR) (the full gel is shown in Supplementary Fig. 1). **g**, Expression and purification of T4 gp25 in soluble form. Fraction 8 contains pure gp25 that was used for crystallization and structure determination. See Methods.



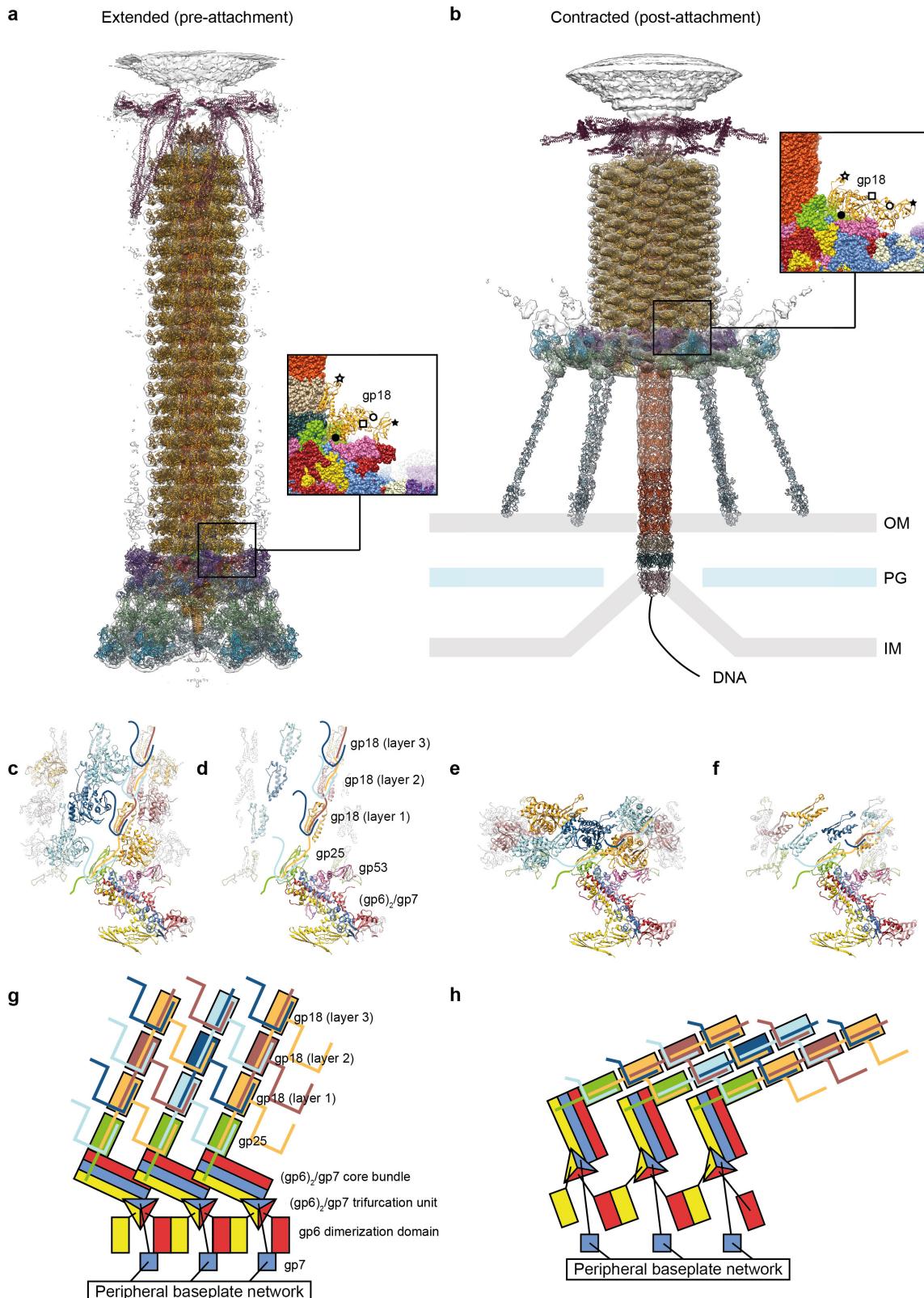
Extended Data Figure 8 | Contacts of gp7 with other baseplate proteins and gp7 mutagenesis. **a, b**, Location of the gp7 jump rope loop in the two conformations of the baseplate. **c**, The structure of gp7 is coloured by amino acid number from blue N terminus to white C terminus. The inset shows the position of gp7 (coloured blue) within the baseplate map. Interactions of gp7 with other baseplate proteins are shown schematically using the colour code of Fig. 1. The superscript $i - 1$ denotes interactions with a symmetry-related copy of a given protein. Red letters indicate sites of mutagenesis (for example, d/i636: deletion/insertion at position 636) that resulted in viable phage particles (see Supplementary Information).

The purple label shows the site that tolerated neither residue deletion nor insertion. **d**, SDS-PAGE of purified T4-7am particles carrying the wild-type (WT) and five mutant gp7 proteins after their concentrations were brought to a common scale according to their absorbance at 260 nm (the full gel is shown in Supplementary Fig. 1). **e**, Infectivity of the mutant particles shown in **d**. The error bar indicates the 95% confidence interval obtained in three independent experiments ($n = 3$). Only one mutation was statistically significantly different from the rest and P -values (two-tailed Student's t-test) comparing it to the wild-type phage and to the mutant with the largest experimental error are given.



Extended Data Figure 9 | Transformation of the conserved part of the baseplate. **a–c**, Three different views of the inner and intermediate parts of two adjacent wedges (gp25, gp53, gp6 and gp7). Complete proteins are shown, including weakly conserved domains. The left and right columns represent the pre- and post-attachment baseplates, respectively. One of the two wedges (wedge I) is semi-transparent. In **c**, the central spike-tail tube

complex is also displayed (semi-transparent). **d, e**, Transformation of the ring of the $(\text{gp6})_2$ -gp7 trifurcation and gp6 dimerization domains from the pre- to the post-attachment state (as in Fig. 2d). The bars in **d** indicate the size of the fragment shown in **e** after rotation. **f**, Reorientation of the $(\text{gp6})_2$ -gp7 core bundles.



Extended Data Figure 10 | Model for baseplate-induced sheath contraction. **a, b**, Pseudoatomic model of the complete T4 tail in the extended (pre-attachment) and contracted (post-attachment) conformations. The insets show a close-up view (labelled with a black box) of the position of the gp18 subunit on the baseplate in the extended and contracted conformations of the sheath. The white geometrical shapes label the same regions of the sheath subunit in both conformations. **c**, Interaction of the two conserved domains of the gp18 sheath protein

with the conserved components of the T4 baseplate wedge (as in Fig. 3). Coloured lines indicate the putative topology of the N- and C-terminal gp18 extensions, as well as the gp25 C-terminal strand. **d**, The same view as in **c**, but the external domain is now not shown for clarity to demonstrate the interaction of gp25-like sheath domains with each other and with gp18. **e, f**, The same as **c** and **d** but in the contracted state. **g, h**, Two diagrams demonstrating the motion of baseplate components that results in sheath contraction.