Tail Structure and Dynamics

Shweta Bhatt, University of Copenhagen, Copenhagen, Denmark
Petr G Leiman, The University of Texas Medical Branch, Galveston, TX, United States
Nicholas MI Taylor, University of Copenhagen, Copenhagen, Denmark

Nomenclature

ATP Adenosine triphosphate
DNA Deoxyribonucleic acid
dsDNA Double-stranded DNA
gp Gene product (protein expressed by a particular gene)
ssDNA Single-stranded DNA
TMP Tape measure protein

Glossary

Caudovirales Order of dsDNA bacteriophages with an external tail structure. Includes the *Podoviridae*, *Siphoviridae* and *Myoviridae*.
Contractile injection system Protein complex related to (and including) the contractile tail apparatus of *Myoviridae*. Also includes the type VI secretion system, R-type pyocins, the antifeeding prophage/*Photorhabdus* virulence cassette and other tailocins.
Myoviridae Bacteriophages with a long, contractile tail.
N-fold symmetric assembly A protein complex comprised of N parts in which these parts are spatially related by a rotational axis of symmetry of order N is called an N-fold symmetric complex (e.g., a complex composed of three polypeptide chains that are related by a threefold is a threefold symmetric complex).
Podoviridae Bacteriophages with a short, non-contractile tail.
Siphoviridae Bacteriophages with a long, non-contractile tail.

Introduction

To replicate, bacterial viruses or (bacterio)phages have to infect their microbial hosts. Unlike eukaryotic viruses, which are usually taken up through endocytosis or membrane fusion, bacteriophages are required to translocate their genome and certain proteins across the host cell envelope. To address this challenge, an overwhelming number of phages use a structure known as a tail. The tail creates a conduit between the phage capsid and host cytoplasm and allows the phage particle to remain attached to the cell surface. Such phages are known as the tailed bacteriophages or *Caudovirales* (Fig. 1). The genome of all *Caudovirales* is a double-stranded DNA molecule, which is packaged into an icosahedral capsid. The proteinaceous tail structure is attached to a special vertex of the capsid through which the DNA is packaged during capsid assembly.

The tail is crucial in the infection process of *Caudovirales*: it is responsible for the initial host recognition and attachment that may include digestion or modification of the cell surface polysaccharides, for the irreversible attachment to the cell, for the degradation of the peptidoglycan, for sensing the "injection signal", and for the translocation of DNA and proteins into the host. Because of the multitude of challenges that need to be overcome, it is not surprising that these tails are large, multiprotein assemblies of extreme complexity. For example, the tail of bacteriophage T4, one of best studied and more complicated phages, consists of a total of at least twenty different gene products present in varying stoichiometries.

In gram-positive bacteria and archaea, the genome needs to be injected across the only membrane of the cell. In gram-negative bacteria however, both the inner and outer membranes (as well as the peptidoglycan layer) need to be crossed, generating some additional requirements for the tail structures of phages infecting these organisms.

It should be noted that other viruses outside the order of the *Caudovirales*, have been found to have structures that are reminiscent of tails, and are used to inject the genome: e.g., the ssRNA virus MS2, the ssDNA virus φX174, the dsDNA phage PRD1, the algal virus PBCV-1, certain archaeal viruses such as *Acidianus* two-tailed virus (ATV) and some eukaryotic viruses such as Herpes simplex. These viruses are discussed in other chapters of this collection.

Here, we will discuss the function and dynamics of the tail of the *Caudovirales*. We will examine the similarities and differences of all three families belonging to this order and point out specific differences between tails of bacteriophages targeting gram-positive and gram-negative bacteria.

Caudovirales

The order of the *Caudovirales* represents the largest group of bacteriophages, and therefore, of viruses. *Caudovirales* can infect either bacterial or archaeal hosts, suggesting that they share a common ancestor that is as old as the emergence of the bacteria circa

 doi:10.1016/B978-0-12-809633-8.20965-5

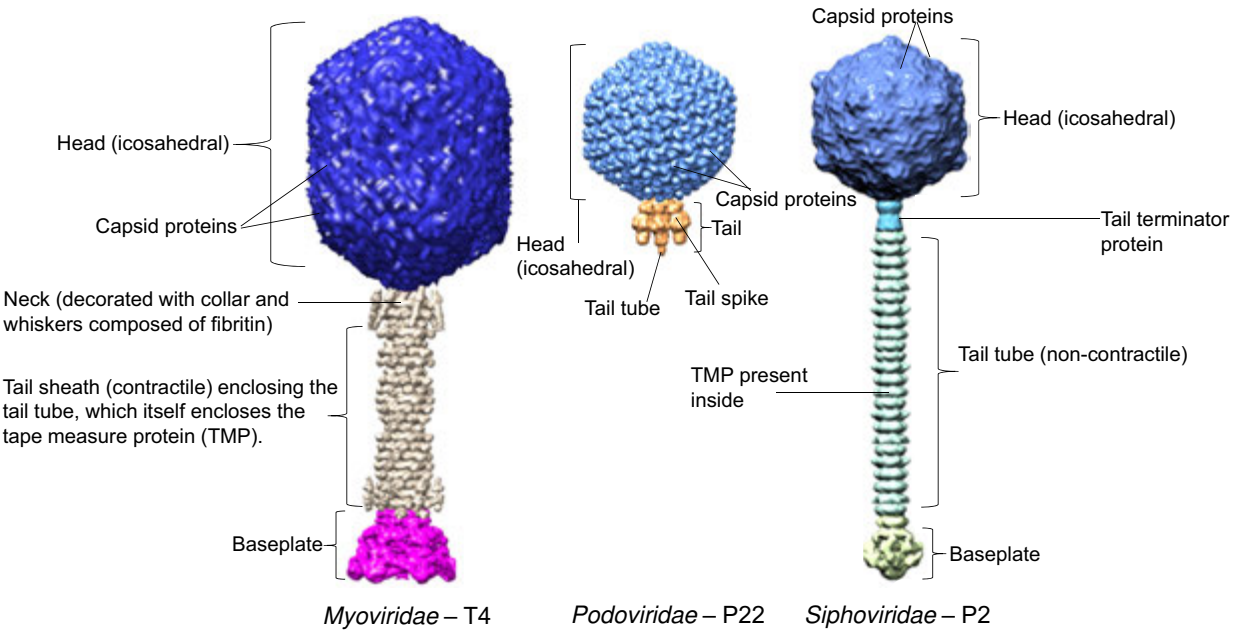

Fig. 1 Cryo-EM reconstructions of archetypal phages representing the three families in the order *Caudovirales*. From left to right: T4, which is characterized by the presence of a long tail surrounded by a contractile sheath and a terminal baseplate. P22, with a short tail and appendages. The lactococcal siphophage P2 with a long, flexible tail encompassed by a non-contractile sheath and a distal baseplate. Figure adapted from Veesler, D., Cambillau, C., 2011. A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. Microbiology and Molecular Biology Reviews 75, 423–433.

3.5 billion years ago. The common architecture and sheer complexity of these viruses makes it near impossible to have arisen from convergent evolution.

The tail is attached to a special vertex of the capsid that is occupied not by a capsid protein but by a 12-fold symmetric mushroom- or cone-shaped protein called the head-to-tail connector or portal (**Fig. 1**). Most tails have (mostly) sixfold symmetry, although pseudo-sixfold/true threefold symmetric tails exist. The portal protein is responsible for symmetry adjustment between the tail and the five-fold symmetric vertex of the capsid which houses the portal. During phage particle morphogenesis and prior to tail attachment, the portal protein is part of the phage-encoded machinery that uses ATP to package the DNA genome into the preformed capsid.

Tails can be divided into three different types, and tail morphology has been one of the obvious and widely used characteristics to classify phages. There are bacteriophages with a short, non-contractile tail (*Podoviridae*); with a long, non-contractile tail (*Siphoviridae*) and with a long, contractile tail (*Myoviridae*) (**Fig. 1**). This classification emphasizes host recognition, binding, attachment and genome delivery mechanisms – all the processes that precede host takeover by the phage and production of new phage particles. In that aspect, post-DNA delivery behavior of phages that have different tail morphologies, e.g., P22 and lambda (*Podoviridae* and *Siphoviridae*, respectively), could be more similar to each other than those of the same type, e.g., P2 and T4 (both *Myoviridae*).

Siphoviridae are the most prevalent family of tailed phages (**Fig. 2**). They are characterized by their long tail which is tube-like in appearance. The length of the tail is regulated by a tape measure protein (TMP). *Myoviridae* are also a very common family (**Fig. 3**). They contain a tube that is structurally related to the tube of the *Siphoviridae* and which is also of defined length, controlled by a TMP. However, *Myoviridae* tails are more complex as they also contain a sheath structure wrapped around the tail tube. The high complexity of the sheath, the conservation of the tail sheath protein structure and the fact that *Myoviridae* can be found both in bacteria as well as archaea strongly suggests that *Myoviridae* only evolved once. The *Podoviridae* on the other hand are less prevalent (**Fig. 4**). They only target certain types of bacteria, which suggests that they might have evolved multiple times from the *Siphoviridae* and/or *Myoviridae*.

Organization and Assembly of Long Tails

Certain features of *Siphoviridae* and *Myoviridae* are similar, which is why we discuss them here together. In these phages, the capsid and the tail assemble independently from each other.

Upon completion of DNA packaging, one or two head completion proteins bind to the portal vertex of the capsid making the latter competent for tail attachment. In the absence of tail completion proteins, the DNA can leak from the capsid.

Tail assembly starts from the assembly of its capsid-distant part, which is called the tail tip complex in *Siphoviridae* or the central hub-spike complex in *Myoviridae*. In *Myoviridae*, the central hub-spike complex forms the centerpiece of a roughly planar structure called the baseplate. In both systems, this complex interacts with a "ruler" protein called the TMP, its chaperone, and other

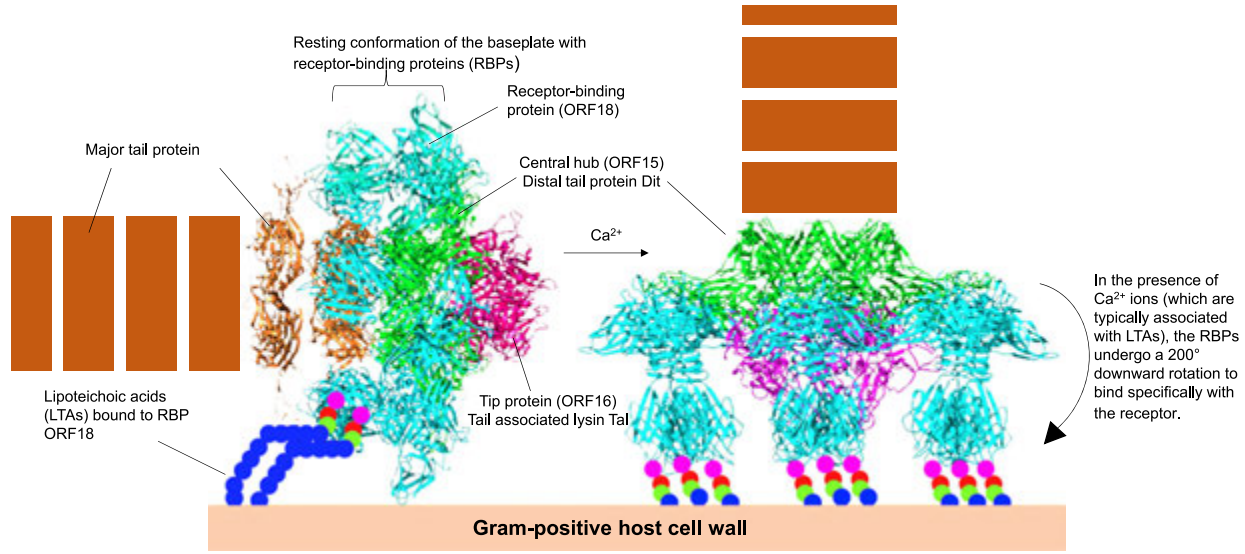

Fig. 2 Schematic representation of the conformational changes in the lactococcal phage P2 baseplate upon adsorption to the host cell wall components. The RBPs interact with the lipoteichoic acids through the host recognition domains and six RBP trimers undergo a 200° downward rotation, enabling them to interact with the host-specific "pellicle" phosphopolysaccharides. Adapted from Sciara, G., Bebeacua, C., Bron, P., *et al.*, 2010. Structure of lactococcal phage p2 baseplate and its mechanism of activation. Proceedings of the National Academy of Sciences of the United States of America 107 (15), 6852–6857.

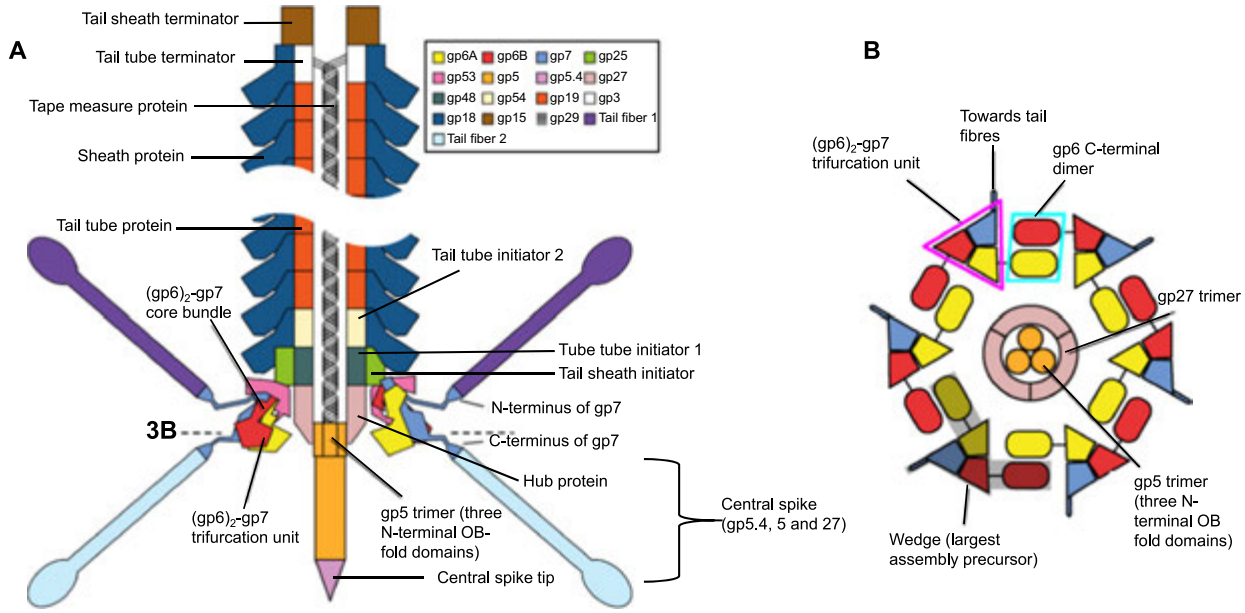

Fig. 3 (A) Organization and architecture of the *Myoviridae* tail. Schematic showing the components of the contractile tail, which have been labeled according to the gene products in bacteriophage T4. (B) Organization of the contractile tail baseplate. Figures have been adapted from Taylor, N.M.I., van Raaij, M.J., Leiman, P.G., 2018. Contractile injection systems of bacteriophages and related systems. Molecular Microbiology 108, 6–15.

chaperones whose conservation and functions are unclear. The TMP determines the eventual length of the tail. The tube is then assembled onto the tail tip/central hub-spike complex, first by binding one or two special tube initiator proteins and then by subsequent binding of multiple copies of the tube protein around the TMP (**Figs. 2** and **3**). The tube initiator proteins and the tube proteins are hexamers, so the trimeric tail tip/central hub-spike complex contains three-to-six-fold adapter domains. Binding of the tail tube terminator protein to the end of the growing tube completes tube assembly. The central hub protein, tube initiator proteins, tube protein and terminator protein are all structurally and evolutionary related.

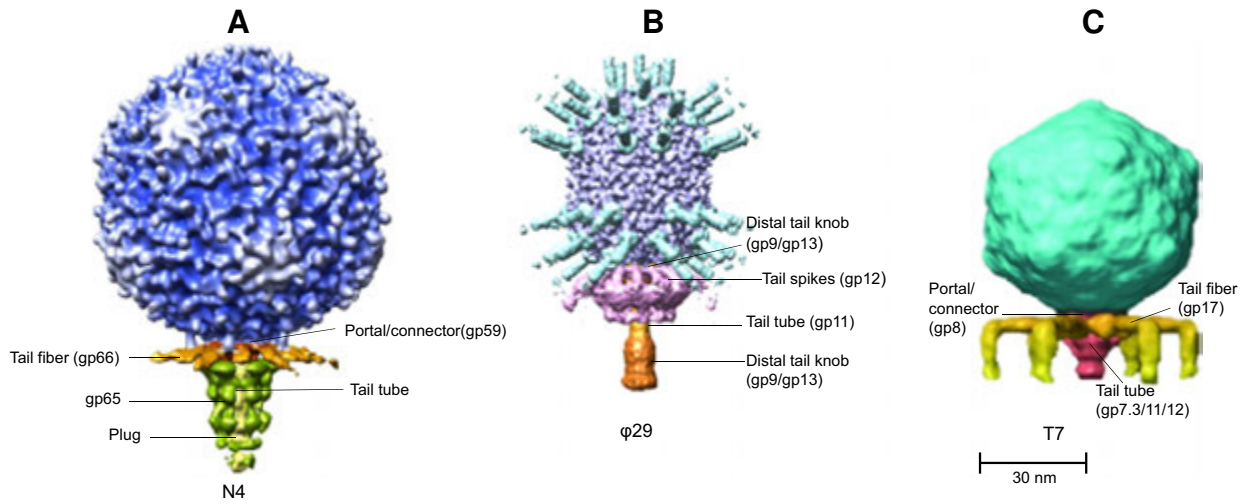

Fig. 4 Cryo-EM reconstructions of different phages of the *Podoviridae* family showcasing the diversity in size, tail structure and components. All the phages share a common structural feature in the form of fibers/appendages attached generally to the proximal part of the tail. Fig. A is adapted from Choi, K.H., McPartland, J., Kaganman, I., *et al*., 2008. Insight into DNA and protein transport in double-stranded DNA viruses: The structure of bacteriophage N4. Journal of Molecular Biology 378, 726–736. Fig. B is adapted from Xiang, Y., Morais, M.C., Cohen, D.N., Bowman, V.D., Anderson, D.L., Rossmann, M.G., 2008. Crystal and Cryo-EM structural studies of a cell wall degrading enzyme in the bacteriophage φ29 tail. Proceedings of the National Academy of Sciences of the United States of America 105 (28), 9552–9557. Fig. C is adapted from Hu, B., Margolin, W., Molineux, I.J., Liu, J., 2013. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. Science 339 (6119), 576–579

In the *Siphoviridae*, binding of the tail tube terminator marks the end of tail synthesis. However, in the *Myoviridae*, tube completion is followed by the attachment of the tail sheath. One of the baseplate proteins (gp25 in bacteriophage T4) is homologous to one of the domains of the sheath protein (**Fig. 3**). It forms an integral part of the baseplate-proximal layer of the sheath and can therefore be thought of as the tail sheath initiator protein. The sheath itself is assembled as a six-start helix with a symmetry matching that of the tube. In the very last step of contractile tail assembly, the tail sheath completion protein binds to the tube terminator and to the last disk of the sheath.

In *Siphoviridae*, the head–tail interface is formed by the head completion proteins and the tail tube terminator. In *Myoviridae*, the head completion proteins instead bind to the tail sheath terminator. In bacteriophage T4, the final step in the assembly is the attachment of the long tail fibers to the tail. However, in some other systems, such as e.g., the R-type pyocin, the tail fiber is required for assembly of the baseplate showing that it acts early in the assembly process.

The Tube

The tube is organized around the TMP. The TMP is mainly alpha-helical in structure and assembles as a multimer of unknown stoichiometry. Given the long size of the tail, the TMP is likely crucial to transmit the "injection signal" from the tail tip complex to the capsid. It has been shown that the length (in number of amino acids) of the TMP correlates with the length of the tail across different phages. Furthermore, deletion of fragments of the tape-measure protein results in a concomitant shortening of the tail. The TMP likely interacts with both the tail tip complex and the tube terminator protein (**Fig. 3**). As mentioned previously, structural homology of these proteins positioned at both ends of the tube, as well as for the major tube protein, has been established for both *Myoviridae* and *Siphoviridae*. For bacteriophage T4, a myovirus, the structure of the pseudo-hexameric trimeric hub protein gp27, the two tube initiator proteins gp48 and gp54 (each forming a hexamer), and the tube protein gp19 (forming a tube of hexameric rings) have been determined experimentally. Even though the experimental structure of the tail tube terminator gp3 is unknown, its sequence is similar to that of the tube protein. For the *Siphoviridae* there are also several structures of the tube protein, exemplified by p2 ORF18 (pseudohexameric trimeric hub), p2 ORF15 (hexameric tube initiator protein), the N-terminal domain of bacteriophage λ gpV (major tail protein) and bacteriophage λ gpU (tail terminator protein).

The tube of most *Siphoviridae* phages is rather flexible. In contrast, the tube of bacteriophage T4 (and probably most myoviruses) is rigid, even in the absence of the tail sheath. One exception that has been described however is bacteriophage A511 which has a flexible tail although it is a myovirus.

The Sheath

The tail sheath of *Myoviridae* is a complex and intricate molecular machine. Its initial extended state is a high energy conformation whereas the contracted state found in the particle in the post-injection state is a low energy conformation. The contraction powers a drill-like

motion of the tube that results in breaching the integrity of the cell envelope of the host. The contraction is an irreversible process: this makes it especially important that it is only initiated at the right time, when the bacteriophage is correctly positioned on the cell.

The sheath is assembled starting from the baseplate, and the tube–baseplate complex is a necessary primer for sheath assembly. In T4, the sheath consists of 23 hexameric layers of sheath protein gp18 (six per layer). The extended sheath is a six-start helix, with a rise of 40.6 Å and a rotation of 17.2°, is 240 Å wide and 925 Å long. The symmetry of the sheath repeats the symmetry of the tube. This is an important fact, as it suggests the mechanistic basis for sheath assembly. The tube acts as a template for the assembly of the sheath in the extended, high-energy state: indeed, when gp18 is overexpressed in the absence of the baseplate–tail tube complex, the sheath is assembled into a low-energy state similar to the contracted tail sheath, which is only 420 Å long and 330 Å wide. The rise of the contracted sheath is 16.4 Å, and the rotation is 32.9°.

No atomic structure of an assembled bacteriophage sheath is available, but the structures of the sheath in the extended and contracted state for the related R-type pyocin and the type VI secretion system are known. Sequence similarity shows that the structure of the sheath subunit is conserved in all contractile injection systems, although the size of the subunit varies thanks to additional surface exposed domains in some systems. Also of note is that the sheath subunit of the type VI secretion system is encoded by two sequential genes. Each sheath subunit donates and accepts two long linker arms that interconnect all the subunits into a cylindrical mesh resembling a protective bottle sleeve. The connectivity of this mesh is the same in the extended and contracted state. In the contracted state, the entire mesh is simply widened and twisted. The linker arms of the first layer of the sheath subunits extend into the baseplate (where they interact with a T4 gp25-like protein) whereas the tail sheath terminator protein donates its linker arms to the last layer of the sheath subunits.

Baseplates of Long Contractile Tails

The baseplate is by far the most complex part of the contractile bacteriophage (**Fig. 3**). The best-characterized baseplate is the one from bacteriophage T4. It is believed that all contractile bacteriophages share a common architecture similar to the one of T4. However, each myophage will have very specific adaptations which are important to infect its host. These especially include enzymes to degrade cell walls and tail fibers to specifically bind cells of choice.

All known contractile tail baseplates are hexameric. They are built up of six "wedges". In T4, the wedge consists of seven gene products, in different stoichiometries, which assemble in a specific order: gp11, gp10, gp7, gp8, gp6, gp53, and gp25. Once wedge synthesis is complete, the wedges can assemble around the central hub complex, consisting of gp27 (the central hub protein), gp29 (the TMP), gp5 and gp5.4. Gp26 and gp28 likely perform a chaperone function as they are not found in the fully assembled baseplate-tube complex. The formed, high energy baseplate has a "dome" shape. In vitro, in the absence of the central hub complex, the baseplates will assemble in a lower energy "star-shaped" form (see next section). After assembly of the dome-shape baseplate, the tube initiators (gp48 and gp54) can bind, followed by assembly of the tail tube and incorporation of the tail tube terminator as described in the previous section.

A minimal baseplate wedge has been proposed to consist of homologs of the following T4 proteins: gp6, gp7, gp25, gp53, and at least one type of tail fiber protein (usually trimeric). Two chemically identical copies of gp6 interact with gp7 to form a heterotrimer that comprises the main part of the wedge. Starting from the center and moving towards the periphery, the gp6–gp7 heterotrimer consists of a part where the three chains interdigitate or interact extensively (the core bundle and the trifurcation unit) and the part where they go their separate ways (two opposing dimerization domains of gp6 and fiber attachment domain of gp7). Gp25 is positioned at the tip of the core bundle and gp53 is wrapped around the bundle. The trifurcation unit sends the three chains comprising it in three different directions. The dimerization domains of two neighboring wedges interact to link the baseplate into a ring. The fiber attachment domain of gp7 forms a radial protrusion on this ring. The N-terminal part of T4 gp7 (which is upstream of the core bundle domain) also interacts with the tail fibers. However, in simpler baseplates (e.g., in phage P2, phage Mu, the R-type pyocin) but even in some complex baseplates (e.g., of SPO1-like phages), orthologs of gp7 do not contain equivalent N-terminal extensions. To summarize, gp6 has an important role in the circularization of the baseplate and gp7 is crucial for the interaction with the tail fiber network (**Fig. 3(B)**).

Apart from these conserved baseplate proteins, each bacteriophage has a specific set of tail fibers and other structural proteins and/or adaptations. In the case of T4, gp7 is an extended protein, and together with gp8 and gp9 (a dimer and a trimer, respectively), it makes up the intermediate or less conserved part of the baseplate, which is contracting the tail fiber network. In T4, the tail fiber network consists of two different types of tail fibers: the short and the long tail fibers. The short tail fibers (trimers of gp12) are "curled up" around the periphery of the baseplate and form part of the short tail fiber network, which also consists of two other trimeric proteins (gp10 and gp11). The long tail fibers (which have a stoichiometry $(gp34)_3(gp35)_1(gp36)_3(gp37)_3$) bind to the gp9 trimer. These viral adhesins have a tremendous length ($\sim$1600 Å) and recognize both LPS as well as the outer membrane protein (OmpC). In the assembled T4 particle, the long tail fibers can interact with the neck protein fibritin (gp *wac*) when the former are in the retracted state. Upon infection, the long tail fibers need to become extended. This regulation might prevent the inadvertent binding of the tail fibers to the host in conditions that are unfavorable. In vitro, retraction and extension can be controlled by pH, ionic strength, temperature and polyethylene glycol concentration.

During the infection process, it is thought that the binding of the long tail fibers to the cell surface induces a conformational change in the baseplate (going from the high energy, dome-shaped state to the low energy, star-shaped state) which also allows binding of the short tail fibers. Although the precise sequence of events is not clear, the "end state" of this process is structurally well understood from the

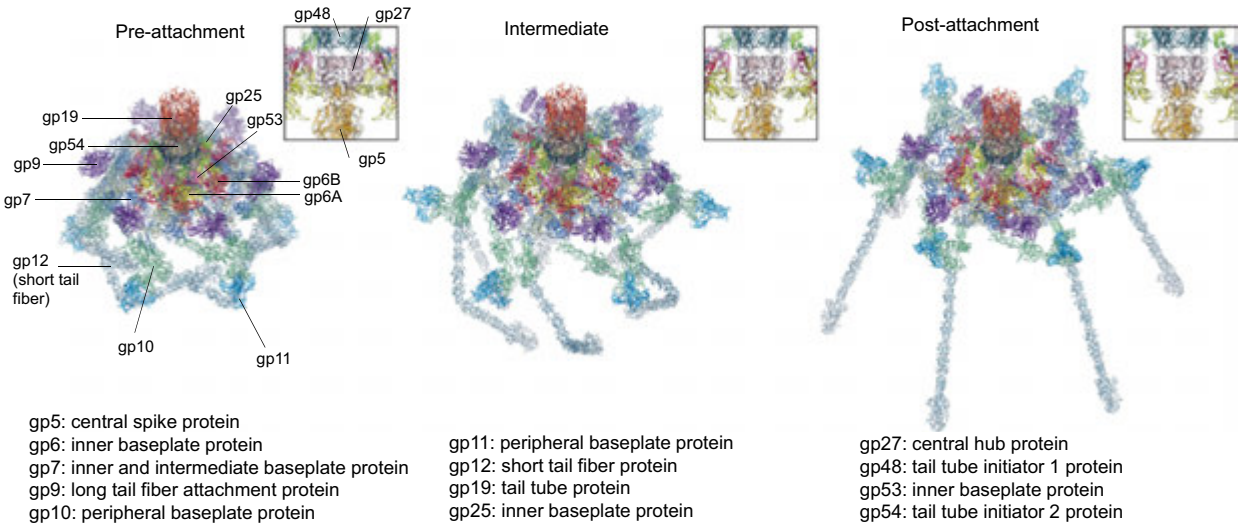

Fig. 5 Structure of the complex baseplate of phage T4 in a pre-attachment (from cryo-EM), intermediate (modeled) and a post-host-attachment state (based on cryo-EM). The insets are focused on the central part of the baseplate and demonstrate a release of the of the tail tube–central spike complex, which remains in a fixed position throughout the transformation. Figure reproduced from Taylor, N.M.I., Prokhorov, N.S., Guerrero-Ferreira, R.C., *et al.*, 2016. Structure of the T4 baseplate and its function in triggering sheath contraction. Nature 533, 346–352.

structure of baseplates in the "star-shaped" state (Fig. 5). In this post-attachment-mimicking state, the long tail fibers make extended interactions with the baseplate, especially gp11. The short tail fibers are then extended and can interact with the LPS. This is accompanied by a slight opening of the iris formed by gp6 and gp7 and results in an outward movement of the gp6-gp7 core bundles. As gp25 is attached to the tip of the bundles, it also moves outward, pulling on the sheath, initiating its contraction.

As already alluded to, even though the architecture of the baseplate is universally conserved, adaptation to different hosts results in tail fibers and receptor binding proteins (RBPs) of extreme complexity. In fact, the size and complexity of RBPs dwarf the actual baseplate in many large bacteriophages, e.g., in Twort-like phages (e.g., *Listeria monocytogenes* phage A511, *Staphylococcus aureus* phage φ812) or Vil-like phages (e.g., *Escherichia coli* phage CBA120).

Tail Tip Complex of Long Non-Contractile Tails

The tail tip complex and the adjacent part of the tail in siphophages is similar to the basal part of the tail tube and the baseplate hub of myophages. The tip protein, called Tal (tail-associated lysin) in siphophages, is homologous to T4 gp27 (Fig. 2). The protein that forms the interface between Tal and the rest of the tail is called Dit (distal tail protein). Dit is homologous to the T4 tube initiator protein gp48. The tail fibers are attached directly to this complex. In certain phages infecting *Lactococcus lactis* the RBP network is so extensive that it forms a large planar structure that by analogy with myophages is called a 'baseplate'. However, this 'baseplate' appears to be structurally and evolutionary unrelated to myophage baseplate. In some of these phages (e.g., lactophage p2) the RBPs perform an acrobatic move during attachment to the host cell surface – they rotate by 200° upon binding to the cell surface polysaccharides (Fig. 2). This conformational change is accompanied by opening of the Tal trimer, which allows the TMP to exit the tail. The TMP forms a channel across the cell envelope through which the DNA is translocated into the host cytoplasm. Of note, whether the TMP participates in forming a membrane channel in myophages is unclear.

In the case of the coliphage T5, three L-shaped fibers (pb1) are attached to sides of the baseplate-hub protein (BHP) pb3 and one straight fiber containing a single RBP is attached to its center. Pb3 is the Tal protein of T5, as it adopts the same protein fold as that seen with the BHP (gp27) of the T4 myophage. At the end of the tail tube, a two-domain tail tip protein (TTP) pb9 is located in the upper region of the tail tip cone. X-ray crystallography studies have revealed structural similarities between domain A of pb9 and the N-terminal domains of known Dit proteins, such as those present in the lactophage p2. It has hence been proposed that Dit protein modules are likely to be a conserved structural motif amongst both Gram positive and gram-negative host siphophages. It has also been proposed that pb2 forms the TMP and possesses fusogenic and muralytic activity.

In T5, host recognition occurs via the reversible binding of three L-shaped fibers (pb1) to the O-antigen of the lipopolysaccharide. The irreversible interaction of the RBP (pb5) that forms the tip of the central tail fiber with an outer membrane iron-siderophore receptor FhuA initiates a cascade of conformational changes, leading to the expulsion of TMP which digests the peptidoglycan and perforates the cell membrane. TMP plays a role in signal transduction by initiating the opening of the head-tail connector, which leads to DNA release from the capsid. The TMP likely forms a channel that extends the tail through the periplasm and allows the DNA to reach the host cytoplasm.

Short Non-Contractile Tails: Tube-Like Tail Structure

Bacteriophages belonging to the family *Podoviridae* are characterized by the presence of short non-contractile tails varying in lengths in the range of 10–30 nm (Fig. 4). They are the smallest family in the *Caudovirales*, considering the number of different phages discovered. In podophages, the tail is assembled directly onto the portal protein of the capsid after the DNA packaging is complete. The *Podoviridae* tail is generally comprised of two different proteins – a dodecameric protein that interacts with the portal protein (also called, confusingly, the head-to-tail connector in some phages), and a second hexameric protein which forms the rest of the tail tube.

One of the most extensively studied podophages is the gram-positive *Bacillus* phage φ29 (Fig. 4). In φ29, the tail is approximately 380 Å long. Its portal-proximal part consists of dodecameric gp11 and is called the lower collar. It is a funnel-like structure to which twelve receptor-binding tailspikes (called 'appendages' in the φ29 literature) are connected. In another podophage, P22, X-ray crystallography has revealed a dodecameric toroidal gp4 complex (called the 'head-tail connector' in the P22 literature) to which six P22 tailspike complexes are attached. Phage RBPs that carry enzymatic domains and thus are stockier in their appearance than fibers in electron microscopy images own their 'tailspike' name to the P22 tailspikes as they looked like small spikes attached to the P22 tail. A combination of the location, the stoichiometry and the predicted high α-helical content suggests that gp11 of φ29, gp4 of P22 and the head completion proteins of the long-tailed phages could be evolutionarily related.

The distal part of the φ29 tail tube is thicker than the portal-proximal part. It is called the tail "knob" and is formed by a hexamer of gp9. It plays a key role in the infection process because the distal end of the φ29 tube is blocked by a long loop of gp9 prior to DNA ejection. Interestingly, not only the structure of the gp12 tail knob of the *Streptococcus* podophage C1 is similar, but the tubular domain of φ29 gp9 and C1 gp12 is found in the tail tube, neck and baseplate proteins of myophages and siphophages – such as in the gp27 baseplate hub protein of T4 and in the λ tail tube protein gpV. This is an indication that the tails of all three families of phages (*Myoviridae*, *Siphoviridae* and *Podoviridae*) have evolved from a common ancestor.

The tail tip or distal part of the tail tube typically contains proteins that aid either in digesting the host cell wall, or in host recognition or both. φ29 degrades the *Bacillus* cell wall with the help of an enzyme located at the distal end of its tail knob (gp13), capable of cleaving the polysaccharide backbone and the peptide cross-links of the peptidoglycan layer. In phage P22, a long needle extending outwards from the center of the tail base is constructed by a single trimer of gp26. It has been observed that in addition to making the first contact with the host surface, the N-terminal end of the needle acts as a "plug" and prevents DNA leakage.

The Tailspikes, Tail Fibers and Phage-Host Interaction

As the host receptors continuously evolve, the genes present in the tail fibers, tailspikes or tail appendages are selectively pressured to adapt to the ever-changing target, which can range from peptide sequences to polysaccharide moieties. The thin, long tail fibers are rigid and are flexibly attached to the tail. They scout the cell surface for a specific binding target and initiate adsorption. Tailspike proteins on the other hand are shorter, stubbier and possess enzymatic activity (although the substrates are known only for a subset of tailspikes). Tailspikes digest or modify cell surface polysaccharides, thus irreversibly binding the phage particle to the bacterial surface and allowing the tail tube to reach the cell membrane. This event triggers conformational changes in the phage, leading to the formation of a channel or a pore through which genetic material and proteins located in the capsid can traverse across the bacterial cell wall. The tailspikes or fibers hence play a crucial role in the initial host recognition and subsequently in triggering the transfer of the genome.

There is considerable diversity in terms of length, shape and specific functionality not only in fibers and tailspikes of myo-, sipho-, and podophages, but even within each group. The majority of these spikes and fibers are homotrimeric – such as gp12 of φ29 and gp17 of T7. In phages such as φ29, the tailspike appendages recognize and cleave the glucosylated poly-glycerol phosphate teichoic acid by a mechanism similar to that seen in some ribonucleases. In phage C1, the appendages are more tail fiber-like and form a skirt around the tail. These appendages carry a flexible globular domain at their distal end.

The dynamics of phage-host interaction depend on the nature of the bacterial host. Gram-positive and gram-negative bacteria differ in cell wall thickness and components, which have led phages to the adoption of different infection strategies. Gram-positive bacteria lack an outer membrane and phages typically interact with cell wall-associated teichoic acids and other glycopolymers embedded in the thick peptidoglycan layer. They must overcome this physical barrier to reach the cell membrane and deliver their genome into the bacterial host cytoplasm. After the phage has reached the cell membrane, it is thought that there needs to be some protein–membrane fusion event. The archetypal phage φ29 relies on the formation of a cone-shaped structure constructed out of the long loop of gp9 that exits from the tail knob upon genome release. This structure functions as a membrane pore, similar to a hydrophobic fusion peptide, which is found in enveloped eukaryotic viruses and of a membrane-active peptide that is present in non-enveloped viruses. This shared mechanism employed by eukaryotic and prokaryotic viruses is likely a consequence of convergent evolution.

For phages of gram-negative hosts, the DNA needs to cross the peptidoglycan and the inner membrane layers. As discussed above, the tails of podo- and siphophages have to be extended by phage and, possibly, host proteins that connect the tail with the cytoplasm. In myophages, the tail tube crosses the outer membrane and the periplasmic space to contact the cytoplasmic membrane, which is invaginated near the tip of the tube. The involvement of other phage proteins (e.g., the TMP) or host proteins in creating the channel spanning the cytoplasmic membrane is unclear and the tip of the tube might be capable of crossing the membrane.

The outer membrane of gram-negative bacteria is dotted with highly conserved porin proteins that serve as reversible binding sites for phage tail fibers. These fibers are also known to interact with the inner (e.g., fibers of bacteriophage T4) or outer (e.g., fibers of bacteriophage φCTX from the *Myoviridae* family) core of LPS. Phage tail spikes on the other hand recognize and degrade the O-antigen (the polysaccharide part of LPS), which brings the particle closer to the bacterial cell surface, thus enabling the recognition of a second receptor in the outer-membrane, which paves the way for irreversible binding and triggers DNA release upon the proper orientation of the phage particle on the cell surface. A number of phages carry multiple sets of tailspike proteins on the particle which allow them to expand their host range from different bacterial strains to different bacterial species (e.g., phage K1–5 infects *E. coli* K1 and K5, phage SP6 infects different *Salmonella enterica* serogroups, and phage SFP10 infects *S. enterica* and *E. coli* O157:H7).

Phages from the T7 supergroup infect most *E. coli* strains (**Fig. 4**). The T7 particle contains six trimeric fibers of gp17 attached to the proximal end of the tail tube (to the dodecameric gp11 ring). Even though the other end of the fiber is free, and that part will eventually interact with a cell surface receptor, in the free state of the phage most of the fibers are bound to the capsid. Similar to the more complex T4 fibers and the side fibers of the siphophage T5, the T7 fibers are not straight but are bent roughly in the middle. In fact, the T7 fibers resembles the letter "L". The rough LPS forming the bacterial outer membrane is the main receptor for T7 fiber binding. This interaction is accomplished by rigid body rotation of the fibers and leads to ejection of the capsid core proteins. The latter form an extension to the tail and connect it to the cytoplasm. The extended tail then acts as the conduit for DNA transfer across the bacterial membranes while also protecting it from periplasmic nucleases.

N4-like bacteriophages usually encode at least two different tailspike/tail fiber proteins with the larger protein directly attached to the phage tail and the smaller protein bound to the longer one (**Fig. 4**). The shorter tailspike (gp63.1) of the N4-like phage G7C is a deacetylase, and the host recognition of G7C depends not only on the main chain structure of the O-antigen, but on the presence of the acetyl group at a certain position of that particular O-antigen. N4 was the first phage that was shown to contain a large virion-encapsidated RNA polymerase (vRNAP) inside its capsid. This enzyme exits the capsid through the tail during infection and participates in genome transfer. The accumulating body of data shows that many phages contain vRNAPs in their capsids that are delivered (together with the genome) into the host cell cytoplasm where they initiate RNA synthesis, although such phages are usually much larger than N4 (e.g., AR9 and PBS1, giant phages of *Bacillus subtilis*).

Acknowledgments

The Novo Nordisk Foundation Center for Protein Research is supported financially by the Novo Nordisk Foundation (NNF14CC0001). This work was also supported by an NNF Hallas-Møller Emerging Investigator grant (NNF17OC0031006) to NMIT. NMIT is a member of the Integrative Structural Biology Cluster (ISBUC) at the University of Copenhagen.

Further Reading

Ackermann, H.W., 1998. Tailed bacteriophages: The order caudovirales. Advances in Virus Research 51, 135–201.
Aksyuk, A.A., Bowman, V.D., Kaufmann, B., *et al.*, 2012. Structural investigations of a Podoviridae streptococcus phage C1, implications for the mechanism of viral entry. Proceedings of the National Academy of Sciences of the United States of America 109, 14001–14006.
Arnaud, C.-A., Effantin, G., Vivès, C., *et al.*, 2017. Bacteriophage T5 tail tube structure suggests a trigger mechanism for Siphoviridae DNA ejection. Nature Communications 8, 1953.
Bebeacua, C., Tremblay, D., Farenc, C., *et al.*, 2013. Structure, adsorption to host, and infection mechanism of virulent lactococcal phage p2. Journal of Virology 87, 12302–12312.
Choi, K.H., McPartland, J., Kaganman, I., *et al.*, 2008. Insight into DNA and protein transport in double-stranded DNA viruses: The structure of bacteriophage N4. Journal of Molecular Biology 378, 726–736.
Fokine, A., Rossmann, M.G., 2014. Molecular architecture of tailed double-stranded DNA phages. Bacteriophage 4, e28281.
Hu, B., Margolin, W., Molineux, I.J., Liu, J., 2013. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. Science 339, 576–579.
Lander, G.C., Khayat, R., Li, R., *et al.*, 2009. The P22 tail machine at subnanometer resolution reveals the architecture of an infection conduit. Structure 17, 789–799.
Leiman, P.G., Arisaka, F., van Raaij, M.J., *et al.*, 2010. Morphogenesis of the T4 tail and tail fibers. Virology Journal 7, 355.
Prokhorov, N.S., Riccio, C., Zdorovenko, E.L., *et al.*, 2017. Function of bacteriophage G7C esterase tailspike in host cell adsorption. Molecular Microbiology 105, 385–398.
Sokolova, M., Borukhov, S., Lavysh, D., *et al.*, 2017. A non-canonical multisubunit RNA polymerase encoded by the AR9 phage recognizes the template strand of its uracil-containing promoters. Nucleic Acids Research 45, 5958–5967.
Taylor, N.M.I., Prokhorov, N.S., Guerrero-Ferreira, R.C., *et al.*, 2016. Structure of the T4 baseplate and its function in triggering sheath contraction. Nature 533, 346–352.
Taylor, N.M.I., van Raaij, M.J., Leiman, P.G., 2018. Contractile injection systems of bacteriophages and related systems. Molecular Microbiology 108, 6–15.
Veesler, D., Cambillau, C., 2011. A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. Microbiology and Molecular Biology Reviews 75, 423–433.
Xu, J., Gui, M., Wang, D., Xiang, Y., 2016. The bacteriophage φ29 tail possesses a pore-forming loop for cell membrane penetration. Nature 534, 544–547.

Relevant Website

https://www.khanacademy.org/science/biology/biology-of-viruses/virus-biology/a/bacteriophages
Bacteriophages – Viruses – Khan Academy.