Structure and Function of Bacteriophages

Marta Sanz-Gaitero, Mateo Seoane-Blanco, and Mark J. van Raaij

Contents

Abstract

Bacteriophages, or phages, are viruses with an exquisitely evolved structure to accomplish their goals. These goals are recognizing a suitable host bacterium, profiting from the host metabolism, and producing multiple progeny phages that are stable enough to survive until they find a new host bacterium to infect. Their

M. Sanz-Gaitero · M. Seoane-Blanco · M. J. van Raaij (✉)
Department of Macromolecular Structure, Centro Nacional de Biotecnologia (CNB-CSIC), Madrid, Spain
e-mail: Martasanz34@gmail.com; mseoane@cnb.csic.es; mjvanraaij@cnb.csic.es

genomes consist of single-stranded RNA, double-stranded RNA, single-stranded DNA, or double-stranded DNA, depending on phage type. They store their genome in highly symmetric protein capsids to protect it from degradation. Often these capsids are icosahedral, but helical and other shapes are also used. *Tectiviridae* and *Corticoviridae* have an internal lipid membrane, while *Cystoviridae* sport an outer membrane layer. Phages with tails, belonging to the *Caudovirales* order, are the most commonly encountered bacteriophages and have icosahedral or prolate capsids. In addition to the capsid, phages need a host cell recognition apparatus. The small icosahedral *Leviviridae* have a single minor capsid protein for this purpose. More complex phages dedicate multiple proteins to host cell recognition, and examples of this are the helical *Inoviridae* and the icosahedral *Tectiviridae*, *Corticoviridae*, and *Cystoviridae*. The *Caudovirales* have highly efficient tail protein complexes for DNA transfer. These tails are flexible (*Siphoviridae*), extensible (*Podoviridae*), or contractile (*Myoviridae*). Apart from elements designed for genome protection, host recognition, and genome transfer, more complicated phage particles may contain proteins for environmental sensing, binding to suitable matrices where host bacteria are likely to be encountered, and other functions.

Introduction

The replication cycle of a bacteriophage consists of several steps: (i) recognition of a suitable bacterium to infect, (ii) the transfer of the genomic material into the host, (iii) subversion of the host metabolic machinery to produce a multitude of new phage particles, (iv) escape from the confines of the host cell by lysis or secretion, and (v) the wait for an encounter with a new, suitable, host bacterium (see chapter ▶ "Phage Infection and Lysis"). To perform these tasks, bacteriophages have evolved metastable particles. They need to survive between infections, protecting their genomic material inside a sturdy protein capsid. At the same time, these capsids contain exposed or exposable sites for the recognition of new hosts. When this recognition happens, the capsids are poised for important conformational changes necessary to transfer their genomic material into the bacterium.

Due to their small size, from a few tens of nm to a few hundred nm in their largest dimension, phages have relatively small genomes which cannot code for a lot of proteins (see chapters ▶ "Genetics and Genomics of Bacteriophages" and ▶ "Bacteriophage Discovery and Genomics"). Therefore, phage capsids are generally symmetrical, with many copies of the same protein. Icosahedral shapes are the most common, but helical phages also exist. Some phages decorate their capsids with proteins that are thought to bind molecules present in matrices where host bacteria are likely to be found, for instance, mucins for binding in animal lungs or intestine. Phages may also have structures designed for environmental sensing; displaying certain proteins only when encountering a suitable host under the right conditions where successful infection is likely.

Another constraint on phage structure is the cell wall of the bacterial host, including glycans, outer membrane proteins, flagella, and other appendages. The phage must be adapted to recognize one or more of these molecules efficiently and conclusively. On the one hand, phages must avoid transferring their genomic material into a host cell that is unsuitable for replication, and on the other, for maximum evolutionary success, they should not miss any suitable hosts. The phage must also carry a mechanism to help its genetic material traverse the cell wall and membranes. Once the genome has entered the bacterium, the phage can take over host cell metabolism. Most phages only transfer a few proteins into the cell with their nucleic acid, relying on the host cell for production of the rest of their gene products.

In this chapter, we discuss the virion structures of bacteriophages of the *Leviviridae*, *Microviridae*, *Inoviridae*, *Cystoviridae*, *Tectiviridae*, *Corticoviridae*, *Siphoviridae*, *Podoviridae*, and *Myoviridae* families (Fig. 1). We will also relate these structures to function. A glossary of terms relevant to phage structure and function can be found in Table 1.

Overview of Phage Families

The *Leviviridae* (Fig. 1a) are a family of small icosahedral bacteriophages with a monopartite, single-stranded, linear, plus-strand RNA genome that serves as a messenger RNA, encoding only four proteins: the coat protein, the replicase, the maturation protein, and the lysis protein. There are two genera: *Levivirus* (containing the species MS2 and BZ13) and *Allolevivirus* (containing the species Qβ and F1).

The *Microviridae* (Fig. 1b) are a family of small icosahedral bacteriophages with a single-stranded, circular DNA genome. Virions contain the plus-strand of the DNA, which means the minus-strand has to be generated intracellularly to be used as a template for the generation of messenger RNAs by transcription. The *Microvirinae* are the most representative genus of the *Microviridae* and include the well-studied ϕX174 but also phages G4 and α3. Other genera are the less well-studied *Gokushovirinae*, *Pichovirinae*, *Aravirinae*, and *Stokavirinae*.

The *Inoviridae* (Fig. 1c) are a family of rod-shaped filamentous viruses with a circular, single-stranded DNA molecule (plus-strand). They have a helical structure and do not lyse the bacterium after infection but use the host cell to continually produce progeny virions by extrusion through the host membrane. Accepted genera are *Inovirus* (including the well-known cloning vehicle and phage display tool M13), *Fibrovirus*, *Habenivirus*, *Lineavirus*, *Plectrovirus*, *Saetivirus*, and *Vespertiliovirus*.

The *Cystoviridae* (Fig. 1d) are the only known family of dsRNA viruses that infect bacteria. They have a nucleocapsid containing three double-stranded RNA segments: a small, a medium, and a large RNA. The nucleocapsid is covered by a lipid membrane layer. They are only known to infect *Pseudomonas* bacteria, and only a few species are known, belonging to a single genus: *Cystovirus*.
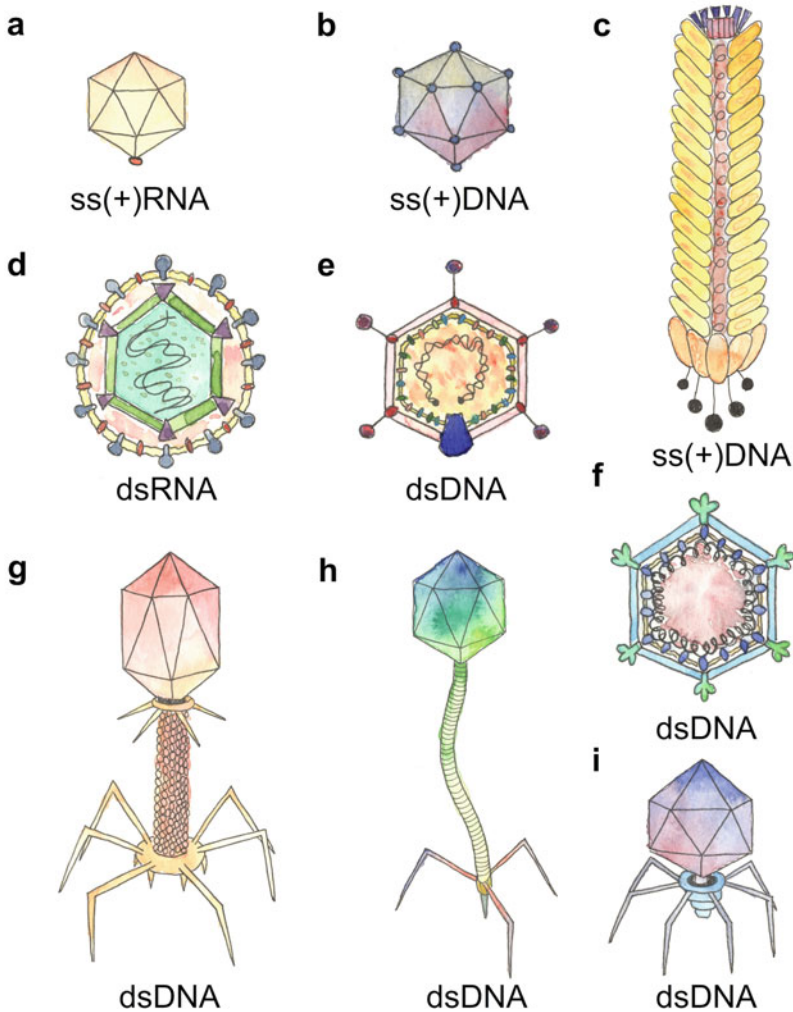
Fig. 1 Schematic drawing of the levivirus MS2 (**a**), the microvirus φX174 (**b**), the inovirus M13 (**c**), the cystovirus φ6 (**d**), the tectivirus PRD1 (**e**), the corticovirus PM2 (**f**), the myovirus T4 (**g**), the siphovirus T5 (**h**), and the podovirus T7 (**i**). Phages are not drawn to scale, and the inovirus in panel C should be longer, with a longer genome (red) and many more yellow major coat protein subunits. The type of genome of each phage is indicated under their drawings

The *Tectiviridae* (Fig. 1e) are a family of double-stranded DNA phages that infect Gram-negative bacteria. These phages have an icosahedral capsid (structurally related to adenovirus capsids) that is decorated with spikes at the fivefold vertices. They do not have tails, but the capsid encloses an internal host-derived membrane. Upon infection, this membrane, together with associated proteins, is extruded and functions as an extensible appendage for DNA transfer. The *Corticoviridae* (Fig. 1f) are also icosahedral viruses with a circular double-stranded DNA genome.

Table 1 Glossary of terms relevant to phage structure and function

Term	Definition
α-Helix	Secondary structure element of a protein chain consisting of a right-handed helix, in which the backbone nitrogen donates a hydrogen bond to the carbonyl oxygen of the amino acid four residues found before it
Asymmetric unit	Minimal unit of a symmetric structure that, when repeated in space according to the different symmetry elements, can generate the entire structure
Avidity	Increased binding strength due to multiple copies of the same protein or protein domain binding to multiple copies of a ligand attached to each other or to a surface (like a bacterial membrane)
β-Hairpin	Minimal β-sheet made up of two antiparallel β-strands connected to each other by a loop region
β-Sheet (also β-pleated sheet)	Flat or curved part of a protein structure made up of multiple β-strands interacting with each other via backbone hydrogen bonds; β-sheets can be parallel, with the strands running in the same direction, or antiparallel, with the strands running in alternate directions
β-Strand	Secondary structure element of a protein chain consisting of an extended region of the protein chain, interacting with other β-strands via backbone hydrogen bonds to form a β-sheet
Baseplate	Structure found at the end of a phage tail that orchestrates the interaction with host bacteria and the start of DNA transfer into the host
Capsid	Protein shell that contains the genetic material of the phage
Capsid expansion	The process of a phage capsid enlarging as it is filled by DNA and internal scaffolding proteins are removed. It is accompanied by a thinning of the capsid wall (see also capsid maturation)
Capsid maturation	The broader process of a phage capsid changing conformation after assembly and into an infectious phage. It entails capsid expansion, but also processing such as proteolytic processing of capsid proteins, when applicable (see also capsid expansion)
Cementing protein	Proteins that bind to the exterior part of the capsid at regular locations, usually between coat protein subunits, to stabilize the capsid wall (also called staple protein)
Chaperone protein	Protein that helps one or more other proteins fold and assemble into the correct arrangement but does not form part of the final structure
Chronic infection	Phage infection that does not lead to host cell lysis, but transforms the host into a phage factory, constantly extruding newly formed phage particles. Plaques can still form due to drastically reduced rates of bacterial cell division compared to uninfected cells
Coat protein	Protein making up the capsid of a phage. Multiple copies of the same coat protein subunits make up a symmetrical icosahedral, prolate, or helical capsid (see also major capsid protein)
Connector protein	Protein that connects the head to the tail of tailed bacteriophages. Twelve circularly arranged copies of the connector protein form the special or unique vertex in the head and function as the gateway for DNA entry and release (also called portal protein)
Contractile tail	Double-layered tail of the *Myoviridae*, of which the outer sheath layer can contract to drive the inner tail tube into the host cell wall
Decoration protein	Protein bound to the outside of the phage capsid or tail without an apparent function for the structural stability of the phage. They likely

(continued)

Table 1 (continued)

Term	Definition
	have a function in host binding or in helping the phage retain position in favorable environments such as an animal lung or gut
DNA transfer	Movement of DNA out of the phage and into the host bacterium during infection
DNA translocation	Assisted incorporation of newly synthesized phage DNA into the preformed capsid by packaging proteins
Domain swapping	Instance where one or more secondary structure elements (often a β-strand) insert themselves into the fold of a neighboring protein
Enveloped virus	Virus with a membrane covering its protein capsid. Additional viral proteins are inserted into the membrane
Extensile tail	Tail formed during the infection process by proteins that come from inside the capsid. *Podoviridae* and *Microviridae* use this mechanism
Fiber diffraction	Technique where aligned biomacromolecules are irradiated and the diffraction patterns recorded to make inferences about the underlying structure of sample
Filamentous phage	Phage that has a long, thin, and to some extent flexible shape, i.e., those belonging to the *Inoviridae* family. An interesting consequence of this shape is that the length of the virus varies with genome size
Fusogenic protein or peptide	Protein or peptide that mediates fusion between viral and host cell membranes
Head-full packaging	Mechanism where DNA is translocated into the phage capsid until a certain pressure is reached, i.e., the capsid (head) is full. It means that somewhat more than a complete genome is packaged and that different phage particles have different genome, terminally redundant, ends (in an alternative mechanism, packaging starts at a specific genome site, the packaging signal, and ends when a termination site is encountered)
HK97 fold	Basic fold of the major capsid proteins of all members of the *Caudovirales* order, suggesting that all these viruses, and including the *Herpesviridae* family, have a common ancestor
Icosahedron	Geometric figure with a high internal symmetry, consisting of 12 vertices, 20 triangular faces, and 30 ridges. Each vertex contains a fivefold symmetry axis, each face a threefold symmetry axis, and each ridge a twofold symmetry axis
Immunoglobulin domain	Protein domain with a fold similar to that encountered in antibodies and other immune system proteins. The fold consists of two β-sheets forming a sandwich with a specific topology (ABED-CFG)
Jelly roll	Fold encountered in the capsid protein of many viruses, including phages belonging to the *Microviridae*, *Tectiviridae*, and *Corticoviridae* families. The fold consists of two β-sheets forming a sandwich with a specific topology (BIDG-CHEF)
Lipopolysaccharide	Outer membrane component of Gram-negative bacteria that is also called endotoxin. They consist of a lipid A membrane anchor, an oligosaccharide inner and outer core, and an O-antigen
Major capsid protein	Viral capsid protein with the highest copy number, i.e., the protein that forms the capsid faces for icosahedral viruses (see also coat protein)

(continued)

Table 1 (continued)

Term	Definition
Metastable particles	Metastable particles are stable phage particles which break apart (usually in a specific way) in response to a certain trigger, such as receptor binding or pH
O-antigen	The O-antigen is the part of the bacterial lipopolysaccharide that is exposed to the environment. It is made up of repeating oligosaccharide units and is specific to the bacterial species. Phage receptor-binding proteins often specifically recognize the oligosaccharide units
Packaging signal	Site on the viral genome that serves to initiate the assembly of the capsid around it or as a recognition site for proteins to start translocating the genome into a preformed phage capsid
Peptidoglycan	Peptidoglycan, or murein, forms a meshed layer outside the bacterial plasma membrane. It is composed of glycan polymers cross-linked by peptide bridges. Bacterial species have different peptidoglycan structures
Pilus	A pilus (plural pili) is a thin appendage of bacteria used in DNA transfer between them (conjugation). Phages can use them as specific attachment sites and use the fact that they are regularly retracted to get close and enter the bacterial cell wall
Plasmid	Circular double-stranded DNA structure present in many bacteria that is small compared to the bacterial genome and that replicates independently. Plasmids often encode genes that are useful to bacteria only in certain circumstances
Portal protein	See connector protein
Procapsid	Immature capsid of a bacteriophage, generally not yet containing genomic material. Procapsids may still contain scaffolding proteins
Prolate capsid	Phage capsid with an elongated shape. The caps on the ends are icosahedral, but extra rings of protein subunits are inserted between them, yielding the elongated shape
Receptor-binding protein	Phages use receptor-binding proteins to specifically recognize elements on the surface of suitable host bacteria. These proteins are usually trimeric, i.e., made up of three identical protein chains
Scaffolding protein	Protein structures that serve to assemble coat protein units into a complete capsid. Phages may have internal and/or external scaffolding proteins
Special vertex	Vertex of icosahedral and prolate capsids where the portal protein is located; the other 11 vertices contain pentameric capsid proteins with the HK97 fold. This vertex initiates capsid assembly, is where DNA translocation into the capsid takes place and where DNA transfer into the host originates (also called unique vertex)
Spike protein	Proteins forming spikes on the viral surface. These proteins are generally homo-trimeric and involved in receptor binding and/or membrane fusion
Staple protein	Proteins that bind to the exterior part of the capsid at regular locations, locking together coat protein subunits, to stabilize the capsid wall (see also cementing protein)
Stem-loop	RNA secondary structure in which two nearby, short complementary regions pair up to form a small double helix (the stem), connected by a

(continued)

Table 1 (continued)

Term	Definition
	loop, formed by the nucleotides between the two complementary sequences
Structural biology	Science that studies the ultrastructure of biomacromolecules and their assemblies and complexes with ligands, for example, phages, phage proteins, and their complexes with host cell components
Tail fiber	Fibrous structures attached to the bottom of a phage tail that are responsible for host cell binding. Many phages may contain two kinds of tail fibers, long tail fibers for primary, reversible, host cell interaction and short tail fibers for secondary, irreversible binding
Tail sheath	The contractile outer layer of the double-layered tail of phages from the *Myoviridae* family
Tail tube	The inner tube of the double-layered tail of phages from the *Myoviridae* family. Phages from the *Siphoviridae* family only have a tail tube, without the contractile sheath
Tailspike	Stubby extensions to the tails of many members of the *Caudovirales* order that bind host receptor molecules (often oligosaccharides) and in many cases hydrolyze them to allow the phage tail access to the host cell membrane
Tape measure (protein)	Phages of the *Siphoviridae* and *Myoviridae* families have copies (presumably six) of a long α-helical tape measure protein spanning the length of their tail tubes. They appear to regulate tail length. PRD1, a phage of the *Tectiviridae* family, has tape measure proteins (P30) spanning the inside of its icosahedral capsid. P30 is proposed to regulate capsid size
Teichoic acid	Anionic glycopolymer on the outside of most Gram-positive bacteria. Gram-positive bacteria lack an outer membrane. Instead, their peptidoglycan layer is modified with teichoic acid chains (wall teichoic acid). Lipoteichoic acid, anchored in the membrane, is usually also present. Teichoic acids are copolymers of glycerol or ribitol phosphate and carbohydrates
Terminase proteins	Terminase proteins bind to the special vertex of a phage head for DNA packaging. They couple ATP hydrolysis to DNA translocation into the preformed capsid
Transmembrane domain	Part of a protein, usually an α-helix, that traverses the lipid bilayer
Triangulation number	Number that describes the arrangement of protein subunits into a symmetric icosahedral or prolate virus capsid
Uncoating	Removal of protein and/or membrane that covers the genetic material of a virus upon entering a host cell
Unique vertex	See special vertex

The *Caudovirales* order contains three different families of tailed bacteriophages, all with an icosahedral or prolate capsid along with a linear double-stranded DNA genome. The *Myoviridae* (Fig. 1g) have a long contractile tail; the *Siphoviridae* (Fig. 1h) a long, flexible tail; and the *Podoviridae* (Fig. 1i) a short tail. The outer tail sheath of a myovirus contracts upon infection to drive the inner tail tube through the

bacterial outer membrane and peptidoglycan layer and delivers the phage DNA directly into the cytoplasm. In the case of a podovirus, core proteins located inside the capsid pass through the short tail to form a protective tube traversing the periplasmic space, again allowing safe passage of the phage DNA. Siphoviral DNA transfer has not been studied that well.

The *Plasmaviridae* are a family of membrane-enveloped viruses that infect bacteria without a cell wall. Only one species, L2, is known. It has a genome consisting of a 12 kb molecule of circular, supercoiled double-stranded DNA. It is pleomorphic, i.e., of variable shape, but no detailed studies about the structure have been reported, and they will not be further discussed here.

Because many bacteriophages are relatively easy to produce, are innocuous to the experimenter, and give valuable general information about virus biology, they have been studied extensively by structural biology (see Box 1). In this chapter, we explain what is known about how the phage particles are assembled, what the final infectious particles look like in atomic detail, and how they are designed to efficiently infect the next available host bacterium.

Box 1: Structural Biology

Structural biology is dedicated to the structural determination of biological molecules and complexes. The main techniques employed are electron microscopy, X-ray crystallography, and NMR spectroscopy, which can lead to high-resolution atomic models. Small-angle X-ray scattering (SAXS) and atomic force microscopy (AFM) should also be mentioned. SAXS is used to obtain envelopes in solution (Korasick and Tanner 2018), and AFM allows the study and manipulation of individual molecules (Moreno-Madrid et al. 2017) but is limited to low resolution.

Negative staining electron microscopy is a relatively easy method to obtain two-dimensional images of a phage at resolutions of 5–10 nm, often allowing a rapid and reliable classification of the phage family (Ackermann and Prangishvili 2012). Obtaining many two-dimensional negatively stained electron microscopy images allows three-dimensional reconstruction, although distortions introduced by the stain limit the attainable resolution to 1–2 nm. Stainless cryo-electron microscopy avoids these distortions, and recent developments in detector technology and image processing have allowed a resolution revolution (Kühlbrandt 2014; Henderson 2015). Three-dimensional reconstruction of electron density maps of biological macromolecules, including bacteriophages, to resolutions where reliable atomic models can be built (0.2 to 0.5 nm), is now possible (see chapter ▶ "Detection of Bacteriophages: Electron Microscopy and Visualization").

X-ray crystallography is the oldest and still most-used technique to obtain atomic models of biological macromolecules. Well-diffracting crystals need to be obtained, which is not always straightforward or possible. Furthermore, the "phase problem" needs to be solved for accurate calculation of electron density

(continued)

Box 1: (continued)

maps (Llamas-Saiz and van Raaij 2013). Electron density maps are calculated by summing three-dimensional wave functions. These wave functions have a direction, an amplitude, and a phase, all necessary to sum them correctly. While the direction and the amplitude can be inferred from the diffraction pattern, the phase cannot and has to be estimated using special techniques involving related protein structures and/or heavy atom derivatives (Taylor 2010). Nevertheless, if these conditions are met, it leads to high-quality atomic models, hence its popularity. Several small bacteriophages have been crystallized, for example, the levivirus MS2 (Golmohammadi et al. 1993), allowing structure solution by X-ray crystallography (at 0.3 nm resolution or better). More often, however, structural proteins of bacteriophages have been crystallized outside the context of an intact phage particle. Usually, this means that higher resolution can be attained (0.1–0.2 nm). The resulting structures can be fitted into lower-resolution electron microscopy maps, allowing for atomic representation of entire phages or phage organelles like tails or baseplates (Taylor et al. 2016). X-ray fiber diffraction is useful for determining helical parameters and structural modeling (Marvin 2017).

NMR spectroscopy (Higman 2013) is a very powerful technique for solving detailed structures and especially the dynamics of monomeric proteins up to around 30 kD in size, but more difficult for the generally larger and multimeric structural proteins of bacteriophages. Solid-state NMR spectroscopy has been used to solve the structure of magnetically aligned inovirus particles (Thiriot et al. 2004).

The *Leviviridae* Family of Single-Stranded RNA Phages

The *Leviviridae* are spherical viruses with a genome consisting of a 4 kb linear single strand of RNA. The first complete genome sequenced was that of the levivirus MS2 (Fiers et al. 1976). Leviviruses gain entrance to the host cell cytoplasm via attachment to surface pili structures (Bollback and Huelsenbeck 2001). Their capsids have an icosahedral structure with a triangulation number $T = 3$ (for an explanation of triangulation numbers, see Box 2). *Leviviridae* genomes only encode four proteins: a maturation protein or minor capsid protein that is involved in host cell recognition, a coat protein that is the major capsid protein, an RNA-dependent RNA replicase subunit, and a lysis protein (in case of *Levivirus*) or a read-through protein in case of *Allolevivirus* (Bernhardt et al. 2001). The *Allolevivirus* read-through protein is the result of occasional read-through of the capsid protein stop codon. Each *Allolevivirus* is estimated to contain three to ten copies of the read-through protein (replacing the normal, shorter, coat protein). The reading frame for the lysis protein of leviviruses overlaps the coat protein and replicase genes in a different reading frame. In the case of *Allolevivirus*, the maturation protein is also responsible for the lysis function.

Assembly

The capsids of *Leviviridae* co-assemble with their genomic RNA molecules (Stockley et al. 2016). Each RNA molecule contains multiple packaging sites and specific sequences that recruit coat proteins. Folding of the RNA molecule into its secondary and tertiary structure then helps to orient coat protein molecules relative to each other and form the icosahedral capsid. Contacts between coat protein molecules probably also play an important role, but assembly is initiated by the genomic RNA molecule (Fig. 2). An RNA sequence at the start of the replicase subunit gene has a high affinity for the coat protein and initiates assembly (Rumnieks and Tars 2014). This interaction is phage-specific (i.e., the coat protein of a phage will only interact with the correct sequence in its own genome), so that only the right genome is coated with its protein. Additional coat protein dimers interact with A-X-X-A sequences in the loops of stem-loop structures of the RNA (A stands for adenosine, X for any nucleoside). This process is driven by the specific three-dimensional structure of the RNA molecule. Stem-loops just upstream and downstream of the putative initiator stem-loop are near each other in the structure, with most of the other well-ordered stem-loops also at the same side of the virus (Dai et al. 2017).

Leviviridae do not encode scaffolding proteins. Instead, the folded genomic RNA molecule functions as an internal scaffold, around which the capsid is assembled. In the final structure, the RNA molecule shows some flexibility, while the capsid is well-ordered. This suggests that the RNA molecule orients the coat protein dimers in roughly the right positions, avoiding off-pathway octahedral assemblies (Plevka et al. 2008). Protein-protein contacts between coat protein dimers then take care of the precise relative orientations, leading to a well-ordered icosahedral coat protein structure.
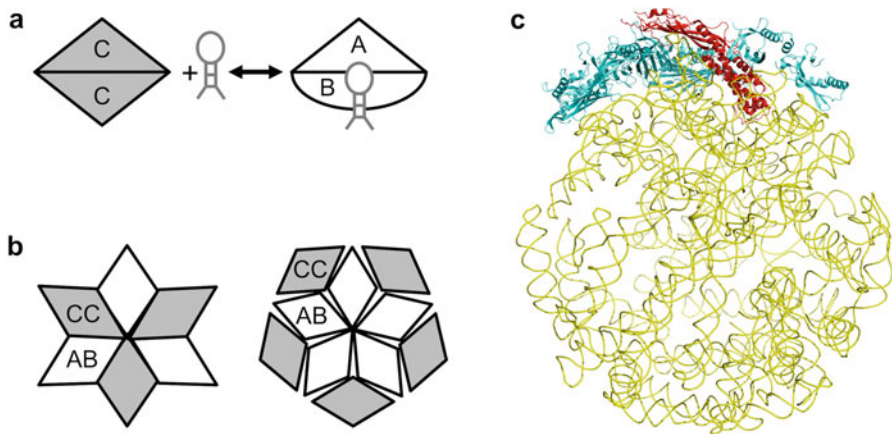

Fig. 2 Assembly of the levivirus MS2. (**a**) Conversion of the symmetric CC coat protein dimer to an asymmetric AB dimer through interaction with an RNA hairpin loop. (**b**) Threefold and fivefold symmetric capsid formation intermediates formed by combinations of asymmetric AB dimers and symmetric CC dimers interacting with genomic RNA (not shown). These intermediates can further assemble and grow into complete $T = 3$ capsids. (**c**) Three-dimensional model of the viral RNA (yellow). Part of the protein capsid is shown in cyan; the maturation protein is shown in red, from a model in the supplementary information of Dai et al. (2017)

Structure

The *Leviviridae* capsid contains 178 copies of the coat protein plus a single copy of the maturation protein (Gorzelnik et al. 2016). Each coat protein adopts one of three conformations (A, B, or C) (Golmohammadi et al. 1993) and organizes into AB and CC dimers in the capsid (Fig. 3). The α-helical part of the maturation protein takes up the position of one of the CC dimers of the coat protein in the otherwise icosahedrally symmetric $T = 3$ capsid. The β-structured part of the maturation protein sticks out into solution and is likely the domain that recognizes the host receptor. The insertion of the maturation protein in the capsid opens a gap right next to it; this probably helps the RNA to leave the capsid and enter the host bacterium (Gorzelnik et al. 2016).

The central part of the MS2 coat protein (Ni et al. 1995) is a five-stranded antiparallel β-sheet that faces the interior of the virus (Fig. 3a). The amino-terminal region forms a β-hairpin that covers part of the β-sheet on the outside, while the carboxy-terminal end is helical and makes extensive interactions with the neighboring monomer, covering the part of its β-sheet not covered by its own β-hairpin. The
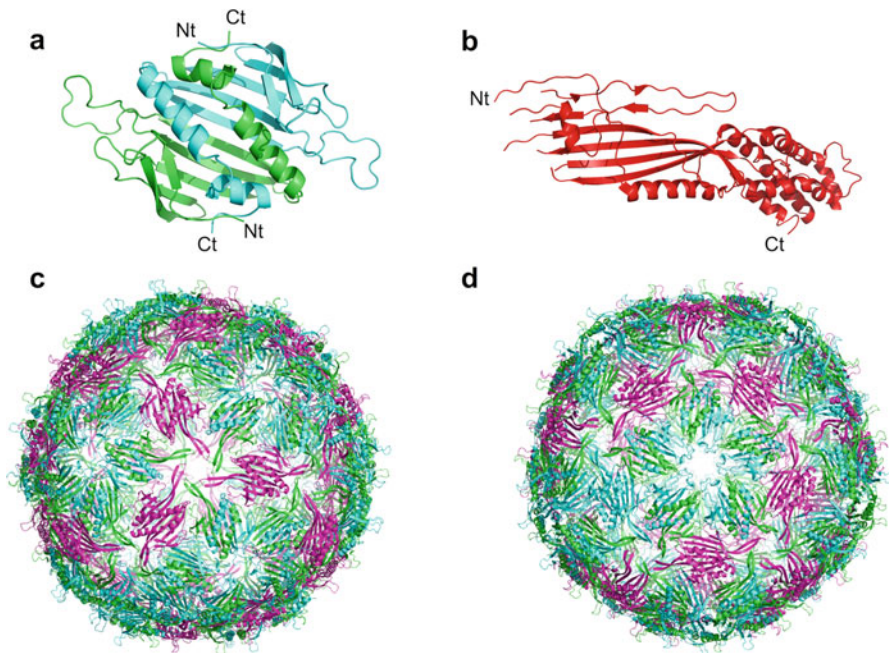

Fig. 3 Structure of the levivirus MS2. (**a**) Structure of the coat protein dimer viewed from the outside of the capsid (PDB entry 1MSC). Protein chains are colored green and cyan. Amino-termini (Nt) and carboxy-termini (Ct) are indicated. (**b**) Structure of the maturation protein (PDB entry 5TC1), viewed from the outside of the capsid (red). (**c**) Structure of the MS2 virus viewed down the fivefold symmetry axis (PDB entry 1MS2). The capsid has triangulation number $T = 3$. (**d**) Structure of the MS2 virus viewed down the threefold symmetry axis. AB trimers are in green/cyan, CC trimers in magenta. PDB is the Protein Data Bank (https://www.ebi.ac.uk/pdbe/)

resulting structure is a stable, interlocked dimer. Dimerization is also favored by the β-sheets of the two monomers aligning in such a way that a shared antiparallel, ten-stranded β-sheet results. The Qβ phage coat protein (*Allolevirus* genus) has an identical fold, although the sequence identity is only 25% (Golmohammadi et al. 1996). The coat proteins are assembled so that five AB dimers surround the fivefold symmetry axes (Fig. 3c). Three AB dimers and three CC dimers alternate around the threefold axes, making them pseudo-sixfold. Finally, the twofold symmetry axes coincide with the twofolds of each CC dimer.

The maturation protein consists of an α-helical domain and a β-sheet domain (Fig. 3b). The α-helical domain contains a bundle of six α-helices. The β-sheet domain has six antiparallel strands with an α-helix on the side and a helix-loop-helix motif covering the sheet on the outside of the virus (Dai et al. 2017). The α-helical domain is inserted into the protein capsid, while the β-sheet domain is exposed, presumably ready to interact with the bacterial pilus.

The single-stranded genomic RNA in the virus has a complicated secondary and tertiary structure (Fig. 2c). There are many short-range interactions, leading to most of the RNA molecule being double-helical. Long-range tertiary interactions (base pairing or kissing loops) are also present and important for the RNA molecule adopting its final shape. In total, most of the bases are involved in interactions with other nucleotides. Exceptions are the loop regions on the outside, which contact the coat protein capsid instead. The spherical shape of the genomic RNA is probably partially imposed by the RNA structure itself and partially by the coat protein.

The maturation protein interacts with an RNA stem-loop region at the 3′ end of the genome. It has been proposed that infection takes place when the pilus retracts, with a virus particle bound to it. The virus, being too big to enter the host, gets stuck on the surface, while the maturation protein is pulled inside, together with the bound genomic RNA (Dai et al. 2017).

Box 2: Symmetry in Viruses

Triangulation numbers of icosahedral capsids. Icosahedra can be thought to be made up of pentagonal tiles at each of the 12 vertices and a variable number of hexagonal tiles on the faces. A $T = 1$ capsid has no hexagonal tiles, while a virus with a high triangulation number has many hexagonal tiles. To determine the triangulation number, T, one starts at a fivefold symmetry axis (i.e., in the middle of a pentagonal tile) and jumps via the shortest route to the next pentagonal tile, via the hexagonal tiles (if they are present). Straight jumps count for **h** and jumps to the side for **k**. For a $T = 1$ capsid, only one jump is necessary, so $\mathbf{h} = 1$ and $\mathbf{k} = 0$. For a $T = 13laevo$ capsid, three straight jumps and one jump to the left are needed, so $\mathbf{h} = 3$ and $\mathbf{k} = 1$ (the mirror image of a $T = 13laevo$ capsid would be a $T = 13dextro$ capsid). The triangulation number is given as $T = \mathbf{h}^2 + \mathbf{h} \cdot \mathbf{k} + \mathbf{k}^2$ (Fig. 4a). The figure illustrates some examples. Further reading can be found in (Caspar and Klug 1962) and (Johnson and Speir 1997).
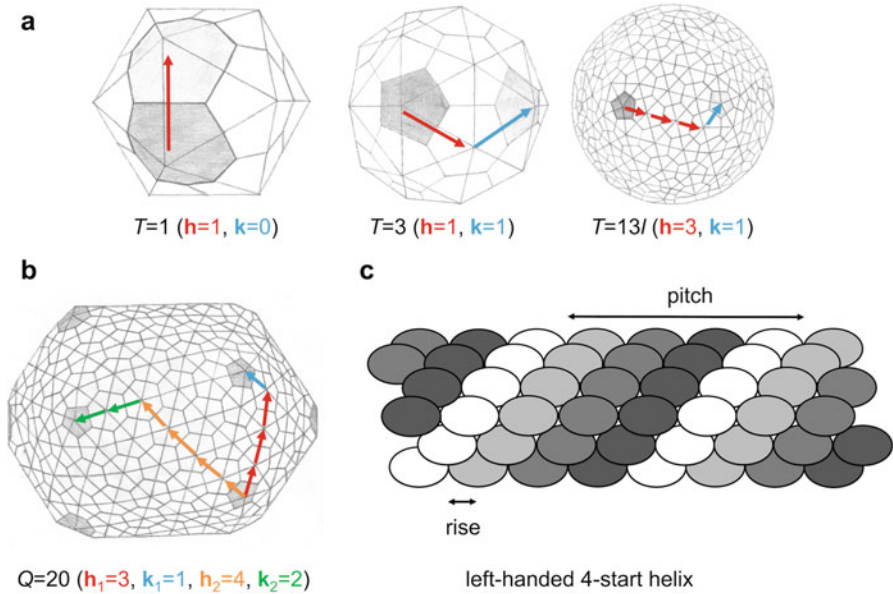
Fig. 4 Symmetry in viruses. (**a**) Triangulation numbers of icosahedral capsids. $T = 13\,l$ stands for $T = 13laevo$. (**b**) Triangulation numbers of prolate capsids. (**c**) Parameters of helical structures

Box 2: (continued)

Triangulation numbers of prolate capsids. Although the caps of prolate (elongated) heads can be described with triangulation numbers as above, the midsection is elongated and made up of nonsymmetric triangles. Calculating the triangulation number Q of the facets with uneven sides now needs h_1, h_2, k_1, and k_2, with $Q = h_1 h_2 + h_1 k_2 + k_1 k_2$. In the case of bacteriophage T4, $h_1 = 3$, $h_2 = 1$, $k_1 = 4$, and $k_2 = 2$ (Fig. 4b). In this case, for the caps, the triangulation number T equals 13. For the side facets, the triangulation number Q equals 20. For a more thorough explanation, see Prasad and Schmid (2012).

Parameters of helical structures. Helices can be right-handed (clockwise screw moving away from the observer looking along the helix) or left-handed (anticlockwise screw moving away from the observer looking along the helix). The helical pitch P is defined as the distance that the helix travels in one complete turn (Fig. 4c). A helix can consist of more than one intertwined helix of subunits, i.e., have more than one start. For instance, a dsDNA helix can be considered a two-start helix. The twist or tilt angle can be defined as the number of degrees the helix turns between subunits.

The *Microviridae* Family of Single-Stranded DNA Phages

The *Microviridae* are a family of small, icosahedral, bacteriophages with a plus-stranded, circular DNA genome (Doore and Fane 2016). They initiate infection via attachment to host cell lipopolysaccharide molecules (Inagaki et al. 2003). The genome is around 5 kb and codes for 11 gene products: A (DNA replication protein), A∗, B (internal scaffolding protein), C (DNA replication protein), D (external scaffolding protein), E (lysis protein), F (coat protein), G (spike protein), H (DNA pilot protein), J (DNA-binding protein), and K. Genes A∗, E, and K are nonessential when the virus is grown in the lab. The A∗ protein results from internal translation in gene A, in the same frame as the parent protein. Microvirus capsids have a triangulation number $T = 1$.

Assembly

The assembly of microviruses can be divided into early and late stages (Doore and Fane 2016). Early stage assembly is mediated by the internal scaffolding protein B. Five copies of the small α-helical B protein bind with their carboxy-terminus to the underside of a capsid protein F pentamer, recruiting one copy of the H protein per pentamer and inducing a conformational change (Fig. 5). The resulting $F_5B_5H_1$ particle then recruits a spike protein G pentamer, which binds on top of the F pentamer, forming the $G_5F_5B_5H_1$ particle (Fig. 5).

Late stage assembly is mediated by the external scaffolding protein D; 240 copies of the α-helical D protein organize 12 $G_5F_5B_5H_1$ particles into procapsids. Four copies of the D protein are arranged in two distinct asymmetric dimers (D1D2 and D3D4), each of which contacts a coat protein F molecule. The D1D2 and D3D4 dimers contact each other in the complex, making a tetramer. Each structural conformer of the D tetramer establishes different interactions with the coat protein beneath it and the D proteins next to it (Prevelige and Fane 2012). The lattice formed by the 240 copies of the D protein holds the procapsid together, allowing divisions between the coat protein pentamers at the threefold symmetry axes, forming 3 nm
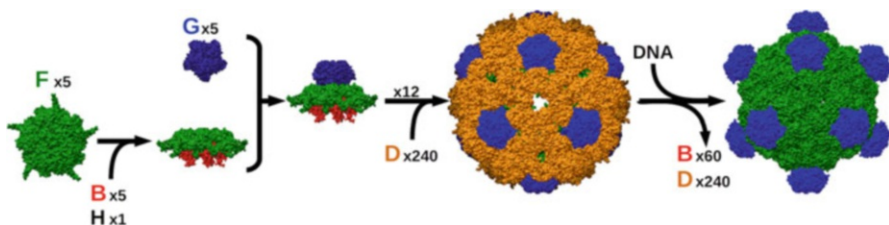

Fig. 5 Assembly of the microvirus ϕX174. Pentamers of the F coat protein form spontaneously and bind five copies of the internal scaffolding protein B on the inside and a pentamer of the spike protein G on the outside. A single copy of the pilot protein H also binds to the inside. The external scaffolding protein D helps to organize the complexes into icosahedral procapsids, with pores large enough to allow DNA entry (PDB entry 1CD3). After DNA entry, the scaffolding proteins leave and allow the collapse of the capsid into a stable virus particle without any visible pores (PDB entry 1RB8)

pores (Fig. 5). These pores are necessary for DNA entry and for the exit of the internal scaffolding protein B.

The single-stranded DNA is synthesized and transferred into the procapsids by the DNA packaging complex, which binds the procapsid at a twofold symmetry axis. This complex includes the bacterial host protein Rep and the viral proteins A and C. Along with the genome, 60 molecules of the small DNA-binding protein J enter the capsid and displace the internal scaffolding protein B (Bernal et al. 2003). Protein B displacement triggers its auto-proteolytic activity, targeting three Arg-Phe motifs at the carboxy-terminal end (positions 77, 93, and 109 in the case of ϕX174), which enables its escape through the 3 nm pores. The release of the internal and external scaffolding proteins facilitates the collapse of the 12 $G_5F_5B_5H_1$ tiles around the genome, resulting in mature virions without any major gaps (Fig. 5). *Microvirinae* such as ϕX174 are characterized by utilizing a dual scaffolding protein system, although other subfamilies of the *Microviridae* only have the internal scaffolding protein (Doore and Fane 2016).

Structure

Microvirus virions consist of a $T = 1$ icosahedral capsid of 25 nm in diameter. The capsid is formed by 12 pentamers of coat protein F. Each fivefold symmetry axis is decorated with a pentamer of the spike protein G, which acts as receptor-binding protein. Both protein F and protein G share an eight-stranded antiparallel β-barrel core structure with BIDG-CHEF topology (Fig. 6). This β-barrel jelly roll fold is common to many small plant and animal viruses, forming the Picorna-like virus structural lineage (Abrescia et al. 2012). The F protein has two extensive insertion loops absent from protein G (Doore and Fane 2016), extending the shape of the protein into a triangle suited to cover the icosahedral viral surface and covering the top of the β-barrel with several α-helices. The β-barrels of protein G associate into a pentamer, with the BIDG-sheets interacting with each other and the CHEF-sheets on the outside. Long AB- and EF-loops decorate the CHEF-sheet surface. The structural similarity between the F and G proteins suggests they may have originated by a gene duplication event.

Inside the capsid, the microvirus genome is associated with 60 copies of the positively charged DNA-binding protein J, which also attaches to the capsid protein F (Bernal et al. 2003). The virion also contains between 10 and 12 copies of the DNA pilot protein H. Protein H helps the DNA to get into the host cell, forming a decameric α-helical coiled-coil barrel (Fig. 6). This tube spans the periplasmic space of the host bacterium, enabling the translocation of the DNA into the cytoplasm (Sun et al. 2014). In the free virion, protein H has been suggested to be present as a monomer near the fivefold symmetry axis. It is not known if the coiled-coil barrel is pre-formed in the viral capsid just before DNA ejection or if it only forms once the H protein molecules traverse the outer membrane.

Apart from the *Microvirinae*, a member of the *Gokushovirinae* (SpV4, infecting *Spiroplasma* bacteria) is the only other microvirus that has been structurally characterized (Chipman et al. 1998). *Gokushovirinae* lack the spike protein. Instead, their F coat proteins have a sequence insertion, with the insertions of three neighboring
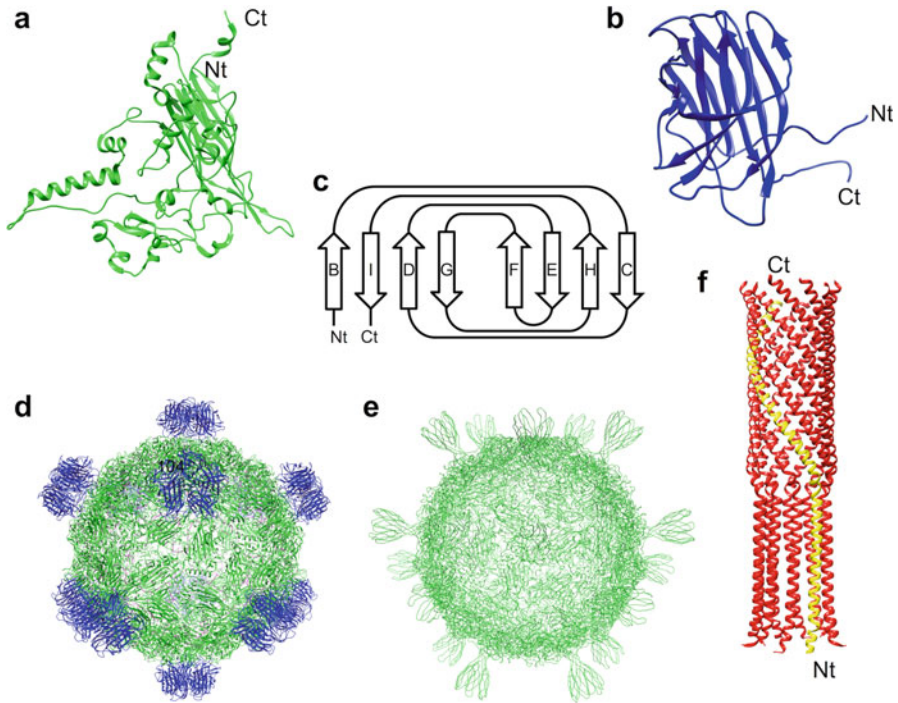
Fig. 6 Structure of microviruses. (**a**) Structure of the F coat protein, a jelly roll β-barrel with insertion leading to a triangle-shaped protein. Amino-termini (Nt) and carboxy-termini (Ct) are indicated. (**b**) Structure of the G spike protein, which also consists of a jelly roll, but with less decoration. (**c**) Topology diagram of the jelly roll β-barrel fold. (**d**) Overall structure of φX174 (PDB entry 1RB8). The coat protein F is shown in green, the spike protein G in blue, and partial structures of the internal H protein are in magenta. (**e**) Model of the *Spiroplasma* virus SpV4, a member of the *Gokushovirinae*, in the same orientation as in panel D (PDB entry 1KVP). Note the absence of the G spike protein at the fivefold symmetry axes; instead, insertions in the F sequence lead to symmetric protrusions at the threefold symmetry axes. (**f**) Membrane tube formed by the pilot protein H to allow DNA transfer into the bacterial cytoplasm (PDB entry 4JPP). In the crystal, the protein forms homodecameric tubes as shown; the exact stoichiometry in vivo is unknown. One of the monomers is shown in yellow, the other nine in red

coat protein molecules leading to protrusions at the threefold axes (Fig. 6). These trimeric protrusions presumably play the same receptor-binding role that the spike protein G does in the *Microvirinae* (Doore and Fane 2016).

Box 3: A Short Note About Phage Protein Nomenclature
Phage protein nomenclature, and virus protein nomenclature in general, can be bewildering. It changes between phage families, different phages in the same family, and even between research groups working on the same phage. Phage

(continued)

The *Inoviridae* Family of Filamentous, Single-Stranded DNA Phages

The *Inoviridae* are a family of long, thin, filamentous bacteriophages containing a
positive sense, single-stranded, circular DNA genome (Fig. 7). They reproduce
without killing their host, instead causing a chronic infection (Rakonjac et al.
2017). There are seven known genera, of which *Inovirus* is the best known.
Inoviruses are long flexible filamentous viruses, measuring about 7 nm in diameter
and 1 µm in length, with different types infecting Gram-negative and Gram-positive
bacteria. The type species is Enterobacteria phage M13, but f1 and fd are also well-
known. Their genomes are between 4 and 9 kb in length.

The *Escherichia coli* Ff phages (for F-pilus-specific filamentous phage) f1, fd,
and M13 are genetically more than 98% identical. They have been useful for
molecular biology and biotechnology applications, due to the ease with which
their genome can be modified. Unlike for most other viruses, capsid size is not
strictly limiting, because larger DNA molecules can generate longer viruses. Up to
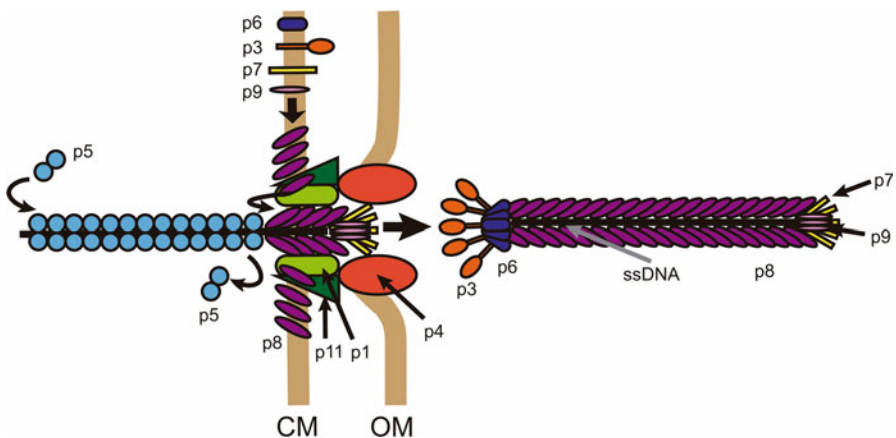12 kb of foreign DNA can be inserted in the viral genome (Marvin 1998). They also

Fig. 7 Overall structure and assembly of the *Inoviridae*. An inovirus part-way through assembly
(left) and a mature inovirus are depicted (right), separated by a horizontal arrow. The circular single-
stranded DNA is drawn as a black bar. Most of the phage-encoded proteins, the structural proteins
p3, p6, p7, p8, and p9, and the nonstructural proteins p1, p4, p5, and p11 implicated in the process
are indicated

provide a convenient way to generate specific single-stranded DNA molecules, which used to be essential for efficient DNA sequencing (see chapter ▶ "Bacteriophage Use in Molecular Biology and Biotechnology"). They have also been used as vectors for gene transfer and vehicles for phage display. In phage display, peptides are presented on the virion surface fused to the viral coat proteins, allowing their interaction with other molecules and the selection of strong binders from libraries containing many variants (see chapter ▶ "Bacteriophages in Nanotechnology: History and Future"). Their non-lytic infection mechanism allows the long-term maintenance of bacterial clones producing mutant viruses. Below, we will discuss the assembly and the structure of the *E. coli* Ff phages as an example for all members of the *Inoviridae* family.

Assembly

Filamentous phages recognize the host pilus via the receptor-binding protein p3. The p3 protein has three domains: N1, N2, and C1. The carboxy-terminal domain C1 is integrated into the virion. The second amino-terminal domain, N2, is the one responsible for interaction with the pilus (Lubkowski et al. 1999). After retraction of the pilus, the first amino-terminal domain N1 of p3 interacts with the TolA protein. The Tol complex then mediates close contact between the bacterial outer and inner membranes, allowing DNA transfer from the phage directly into the cytoplasm (Karlsson et al. 2003). The single-stranded DNA acts as a template for negative strand synthesis. This is a process independent of phage proteins, initiated by the host RNA polymerase. The RNA polymerase synthesizes a primer that is used by the host DNA polymerase III to synthesize the negative strand, obtaining a circular double-stranded DNA, which replicates by a rolling circle mechanism (Higashitani et al. 1997). The dsDNA form of the genome is known as the replicative form, while the single-stranded plus-chain is known as the infective form. To convert the replicative form to the infective form, the p2 protein is necessary (Rakonjac et al. 2011).

As soon as the ss(+)DNA chain is synthesized, it is covered with dimers of p5, which collapse the circular single strand into a flexible rod of about 8 nm in diameter, preventing the synthesis of the complementary strand (Marvin 1998). The finalized p5-DNA complex forms a left-handed helix with the DNA wrapped inside the protein. In the p5-DNA complex, the only exposed zone of the genome is the packaging signal, which is a hairpin loop. To initiate assembly, the packaging signal interacts with the assembly machinery in a sequence-specific way. The structure of p5 has been determined (Fig. 8a) and revealed a β-structured dimer with a positively charged side and a negatively charged side (Su et al. 1997), but exactly how the p5 dimers associate with the single-stranded DNA is not known yet.

Assembly takes place in the cytoplasmic membrane, and nascent virions are secreted from the cell as they assemble (Fig. 7). The eight phage-encoded proteins involved in assembly, including three proteins not present in the virion (p1, p4 and p11), but also the five proteins forming the virion coat (p3, p6, p7, p8, and p9), all have a transmembrane domain and are inserted in the cytoplasmic membrane before phage assembly (Rakonjac et al. 2017). Phage assembly occurs at distinct membrane
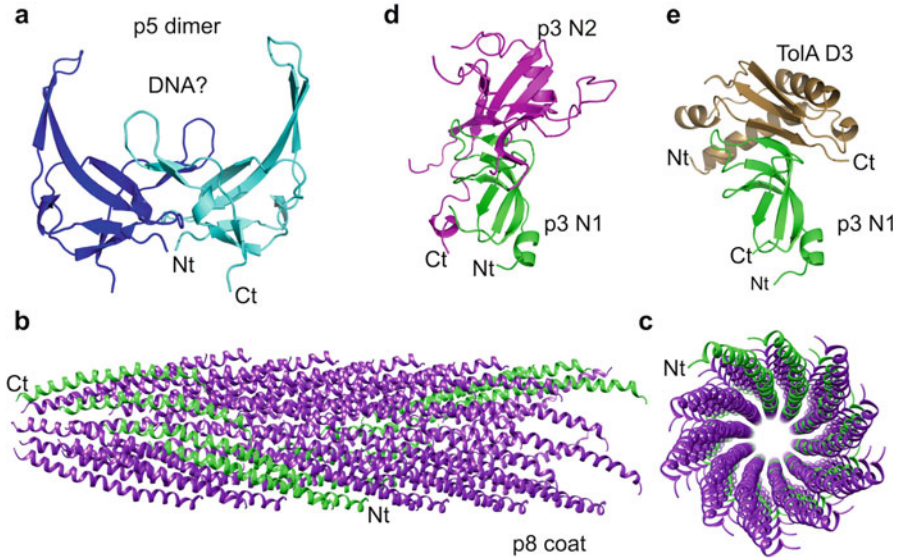
Fig. 8 Structure of inovirus proteins. (**a**) Structure of the p5 dimer (PDB entry 1GVP). Monomers are colored light and dark blue. The putative site of DNA binding is indicated. Amino-termini (Nt) and carboxy-termini (Ct) are indicated. Model of part of the helical inovirus Pf1 capsid formed by the major capsid protein p8 (PDB entry 2C0W) seen from the side (**b**) and from the end (**c**). While most copies of the p8 protein are shown in purple, a helical turn of five p8 molecules is shown in green. The orientation in panel B is the same as in Fig. 6, i.e., the p3 receptor-binding protein and the p6 protein would be on the left and the p7 and p9 proteins on the right. In panel C, p7 and p9 would be in the front and p6 and p3 in the back. (**d**) Structure of the p3 N1 and N2 domains (PDB entry 2G3P). The p3 N1 domain is shown in green, the N2 domain in magenta. (**e**) Structure of the p3 N1 domain bound to the carboxy-terminal domain of the inovirus co-receptor TolA (PDB entry 1TOL), shown in brown

assembly sites, at regions where the cytoplasmic and outer membranes are in close contact. The packaging signal located in the hairpin loop of the ssDNA-p5 complex is recognized by p7, p9, and p1, initiating the assembly. The assembly site is a transmembrane complex formed by the phage-encoded membrane proteins p1, p4, and p11 (Fig. 7). Proteins p1 and p11 are embedded in the cytoplasmic membrane, forming a multimeric complex composed of five or six copies each. Their carboxy-terminal domains protrude into the periplasm and contact the outer membrane protein p4. The protein p4 integrates into the outer membrane, forming a barrel-shaped homo-multimer composed of 12 to 14 subunits with a central cavity measuring 8 nm in diameter, enough to allow the assembled phage to pass through. The cytoplasmic amino-terminal domain of p1 contains a DNA-binding motif necessary for phage assembly. Phage assembly requires ATP hydrolysis (Feng et al. 1999). The protein p1 may also be necessary for the formation of the adhesion zones between the inner and outer membranes where assembly takes place (Russel et al. 1997). The host thioredoxin is also required for correct phage assembly and acts as a DNA-handling protein, not as a redox enzyme (Marvin et al. 2014).

The small proteins p7 and p9 cap the leading, blunt end of the secreted phage. The ssDNA starts to traverse the membrane through the assembly site, causing p5 to dissociate and to be replaced by major coat protein p8. When the DNA is completely coated with p8, the minor coat proteins p6 and p3 are added, resulting in the release of the assembled phage (Marvin et al. 2014). If one of the p3, p6, p7, or p9 proteins is absent, the phage continues to elongate and stays bound to the membrane (Rakonjac et al. 2011). The p3 and p6 proteins cap the trailing, sharp end of the virion. The protein p6 is small, but p3 is a large multi-domain protein that functions as the receptor-binding protein, entering first during infection and leaving last during assembly.

Structure

Inovirus virions are around 7 nm in diameter. The exact length is determined by the size of the genome (normally 5–8 kb of single-stranded DNA) and is in the order of 1 μm. The single-stranded circular DNA molecule is protected by a long cylindrical protein coat made of thousands of copies of the major coat protein p8, a small protein of only around 50 amino acids (Wang et al. 2006). This protein forms a tube around the DNA in an overlapping helical array, with the amino-terminal end of p8 located at the outside of the coat and its positively charged carboxy-terminal end at the inside. These positively charged residues interact with the DNA. The organization of the DNA is unknown; conflicting models for it have been presented (Marvin et al. 2014). What is known is that the DNA is circular, so it is flattened, and one strand goes upward in the phage and the other down. Except for five disordered negatively charged surface exposed amino-terminal residues, each p8 subunit forms a single, continuous α-helix. The central domain of p8 is hydrophobic, allowing the protein to interact with its neighboring subunits (Rakonjac et al. 2011).

Different models have been proposed for the helical array of p8 subunit, based on fiber diffraction and solid-state NMR spectroscopy (Marvin et al. 2014). Cryo-electron microscopy has also been performed (Wang et al. 2006), but did not lead to an exact model. This is likely due to the difficulty in correctly averaging flexible particles and the presence of structural transitions in the particle. It has been shown that temperature affects the fiber diffraction spectra and thus the structure of the virus. In Fig. 8, a model for the fd phage based on X-ray fiber diffraction and solid-state NMR spectroscopy is shown (Marvin et al. 2006). This model is a right-handed five-start helix with a rise of about 1.6 nm and a pitch of 16 nm; there are ten subunits in a complete turn.

The two ends of the virion are capped by three to five copies of each of the four minor capsid proteins, specific for infection and virus assembly (Fig. 7). The small p7 and p9 proteins are located at the end that is extruded first from the bacterial cell. The proteins p3 and p6 are located at the end that is extruded last (Marvin 1998). Once incorporated into the virion, p3 is required for the stability of the proximal end of the p8 array, and p6 is necessary for the incorporation of p3 into the virion. These minor proteins have apolar domains in their primary structure similar in length to the hydrophobic domain of p8, allowing for association with each other and with p8.

Crystal structures of the host cell interaction domains N1 and N2 of p3 are known (Holliger et al. 1999; Fig. 8). It is also known how the N1 interacts with the carboxy-terminal D3 domain of the inovirus co-receptor TolA (Fig. 8), but not exactly how the N2 domain binds to the pilus. In fact, different filamentous phages may bind to different pili, which are correlated to different N2 structures. In phage fd, it appears that the N2 domain shields the N1 domain from interacting with TolA before N2 binds to the F-pilus, while in phage IF1, binding of N1 to TolA is not shielded before N2 binds to the the I-pilus.

The *Cystoviridae* Family of Double-Stranded RNA Phages

The *Cystoviridae* are a family of enveloped viruses with a diameter of about 85 nm. Virions have a double capsid structure with an external membrane. Embedded in the membrane are trimeric spikes that decorate the particle. The genome consists of three double-stranded RNA segments, ranging from over 6 to just under 3 kb. Each genome segment encodes several proteins. So far, only one genus, *Cystovirus*, has been identified. The type species is φ6, but species φ7 through to φ13, φ2954, φNN, and φYY have also been identified. All known cystoviruses infect *Pseudomonas* species, although φ8 also infects *E. coli* and other hosts. In most cases, *Pseudomonas* bacteria pathogenic to plants are the host, although φNN has been isolated from lake water and φYY from hospital sewage (Mäntynen et al. 2018).

Cystoviruses consist of a double-shelled capsid and an external membrane envelope. The largest RNA segment encodes the nonstructural protein P14 and the RNA-dependent RNA polymerase complex proteins P1, P2, P4, and P7; the medium-sized RNA segment codes for the membrane proteins P3, P6, P10, and P13; and the smallest RNA segment codes for the lytic protein P5 and the nucleocapsid protein P8, the nonstructural protein P12, and the membrane protein P9.

Assembly

Cystoviruses commence infection by the trimeric P3 protein attaching to host cell pili, for example, φ6 (Bamford et al. 1976), or rough lipopolysaccharide, for example, φ8, φ12, and φ13 (Hu et al. 2008). The fusogenic P6 protein mediates fusion of viral membrane with the bacterial outer membrane, releasing the capsid into the periplasmic space. The P5 capsid protein then digests the peptidoglycan layer of the host, leading to endocytosis of the capsid into the cytoplasm, now covered by the host inner membrane. This membrane and the P8 protein shell are subsequently lost, but the rest of the capsid remains intact, hiding the viral genome from antiviral host factors. Polycistronic mRNAs are transcribed by the viral RNA-dependent RNA polymerase, P2, found inside the core particle and released into the cell cytoplasm. Newly synthesized proteins assemble into capsids. The P4 packaging protein translocates three plus-stranded RNA segments into the capsids, where they are converted into dsRNA by the P2 protein. Capsid maturation involves enveloping by the viral membrane (including its associated proteins), after which the host cell is lysed and mature virions are released. Presumably, free

P5 protein serves as the phage endolysin necessary for host lysis (Caldentey and Bamford 1992).

The best-studied cystovirus in terms of assembly is φ6 (Poranen and Tuma 2004). The assembly of phage φ6 consists of the generation of a procapsid, the packaging and replication of its RNA genome, the joining of the outer shell of the nucleocapsid, and the formation of its lipid envelope. Its major capsid protein, P1, when expressed in vitro, forms spherical particles and even dodecahedral cages, although they are unstable. In the presence of P4 (the single-stranded RNA packaging protein), P1 forms more stable and regularly ordered dodecahedral particles, containing 120 copies, and with P4 hexamers on the outside of the fivefold symmetry axes (Fig. 9). P4 also allows assembly at much lower concentrations of P1, suggesting it probably is important for nucleation of the assembly. Five P1 dimers assemble around the P4 hexamer. P1 dimers interact with each other to form a collapsed dodecahedral procapsid. This is facilitated by the P7 assembly cofactor which is incorporated, together with P2, at the threefold symmetry axes of the procapsid (Sun et al. 2017). A monomer of the RNA-dependent RNA polymerase, P2, associates with the assembling shell and ends up on the inside, near a threefold symmetry axis. An estimated 60 copies of P7 associate with the $T = 1$ P1 shell. P7 has a regulatory function in assembly and RNA packaging.

The assembled procapsid has a deflated shape with deeply recessed vertices, giving the appearance of a dodecahedron (Nemecek et al. 2013). The hexameric P4 NTPases package a small (3 kb), a medium (4 kb), and a large (6 kb) positive-strand RNA molecule into the capsid sequentially (Frilander and Bamford 1995). The RNA packaging reaction is dependent on magnesium ions and requires a nucleoside triphosphate as an energy source. The packaging of the small and medium segments is efficient when they are packaged on their own, but the packaging of the large segment is very inefficient alone and appears to be dependent on the medium segment being packaged first. P2, the RNA-dependent RNA polymerase, synthesizes the complementary strand, generating the three double-stranded genome segments, some of which migrate from the threefold to the fivefold symmetry axes (Oliveira et al. 2018). During RNA incorporation and replication, the procapsid expands to reach its final icosahedral shape, increasing its volume by about 250%. The expansion of the procapsid is principally due to changes in interdimer interactions, especially at the P1B subunit.
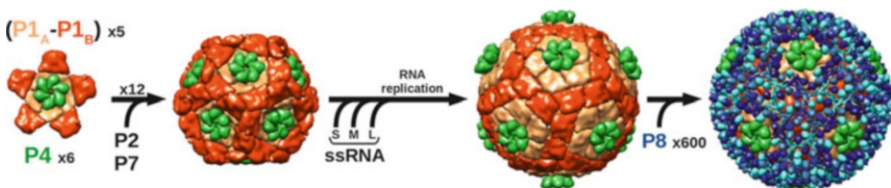

Fig. 9 Assembly of the nucleocapsid of cystovirus φ6. Up to five P1 dimers (light and dark orange) associate with a P4 hexamer (green). The P1 dimers interact to form a deflated dodecahedral procapsid (PDB entry 4BTG). After RNA translocation into the capsid and synthesis of the complementary strand, the procapsid expands. Then 200 trimers of P8 assemble (light and dark blue) onto the P1 shell to complete the nucleocapsid (PDB entry 5MUU)

The nucleocapsid surface protein, P8, can assemble as open, irregular, shell-like structures on its own. However, onto the inflated core, P8 assembles as a regular $T = 13$ shell (Sun et al. 2017). The assembly is promoted by calcium ions. No phage or host cell assembly factors appear to be necessary to yield an infectious nucleocapsid particle. Around the complete nucleocapsid particle, a lipid envelope assembles afterward. The phage membrane protein P9 and the nonstructural phage protein P12 are necessary for this. The lipids of the membrane are derived from the host cytoplasmic membrane, although the assembly of the envelope occurs in the cytoplasm and does not involve budding (Poranen and Tuma 2004). The host recognition and attachment protein, P3, which is anchored to the envelope through protein P6, is expressed as a soluble protein and is the last component to be assembled onto the virions.

Structure

The cystovirus internal capsid is made up of 60 dimers of the P1 protein (Sun et al. 2017), in a $T = 1$ geometry. These dimers are asymmetrical; they are formed by two subunits of the P1 protein (P1A and P1B) with the same α-helical fold but with small differences in tertiary structure (Oliveira et al. 2018). The P1 protein is shaped like a trapezoid tile (Fig. 10b). Most residues are in an α-helical conformation, although small β-strands are also present. Differences between the $P1_A$ and $P1_B$ subunits and between the conformations of $P1_A$ and $P1_B$ before and after capsid expansion are located mainly in a hinge region in the carboxy-terminal domain, although these differences are small compared to the differences in relative orientations of the protein subunits upon capsid expansion (Nemecek et al. 2013). The subunits of the P1 dimer are stabilized by the C-terminal tail of P4, which explains the relative rigidity of the dimer during expansion of the procapsid (Sun et al. 2017).

The fivefold vertices of the P1 icosahedral inner shell are covered by hexamers of the P4 protein. The P4 hexamer is attached to the P1 dimers by up to five of the six P4 carboxy-terminal tails, circumventing the symmetry mismatch of a hexamer binding on a fivefold symmetry axis (Sun et al. 2017). The P4 hexamers are responsible for RNA packaging into the capsid. P4 is an NTPase with a nucleotide-binding Rossmann fold (Fig. 10c). The amino- and carboxy-terminal regions of the P4 protein vary between different cystoviruses, but the central enzymatic domain is conserved (El Omari et al. 2013a).

The outer shell of φ6 is formed by 200 trimers of the P8 protein arranged as a $T = 13 \, l$ shell. The P8 protein consists of two α-helical domains (an amino-terminal peripheral domain and a carboxy-terminal core domain) bound by a linker (Fig. 10d). The amino-terminal peripheral domain interacts either with the core domain of its own trimer (closed conformation) or with the core domain of an adjacent trimer (open conformation). The open conformation is the most common (540 out of the 600 P8 subunits of the fully formed P8 shell), while the closed conformation is restricted to one of the monomers of the peri-pentonal P8 trimers to avoid clashes with the P4 hexamer. The open and closed conformations differ in the linker bending, although the interactions between the peripheral and core domains are analogous (Sun et al. 2017). This feature is known as domain swapping and has been described as a mechanism for protein oligomerization (Liu and Eisenberg
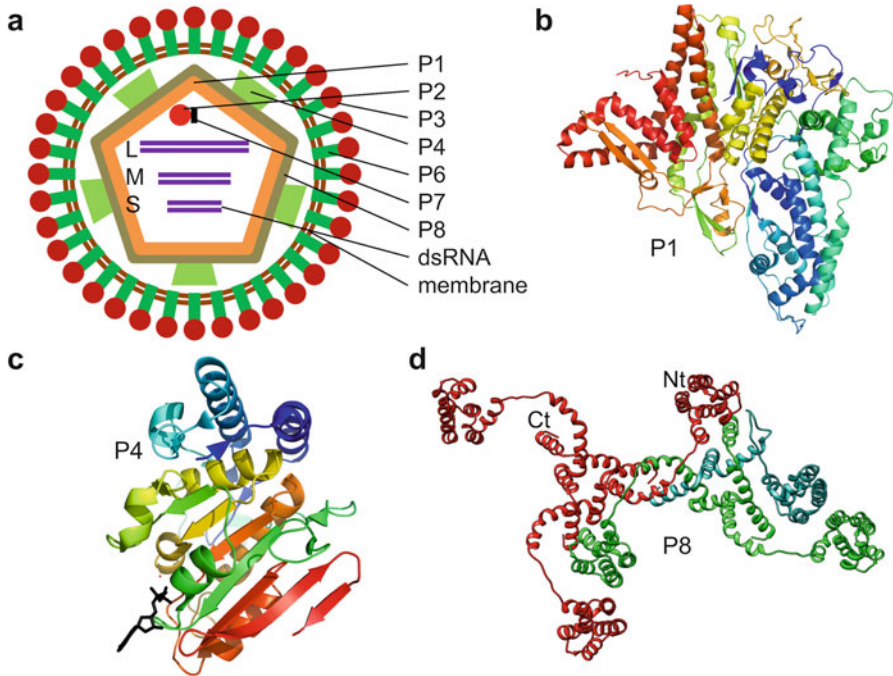
Fig. 10 (**a**) Schematic structure of the cystovirus φ6. The locations of the structural proteins P1, P2, P3, P4, P6, P7, and P8, the viral membrane, and the large (L), medium (M), and small (S) dsRNA genome segments are indicated. (**b**) Crystal structure of the P1 capsid protein seen from the outside of the capsid (PDB entry 4K7H). The chain is colored as a rainbow from dark blue (amino-terminal) to red (carboxy-terminal, Ct). (**c**) Structure of a P4 monomer (PDB entry 4BLO). The chain is colored as a rainbow from dark blue (amino-terminal) to red (carboxy-terminal). The nucleotide (ADP) is shown in black. (**d**) Arrangement of the P8 trimers in the nucleocapsid. A trimer with all its monomers in open conformation (red) and a trimer with two monomers in open conformation (green) and one in closed conformation (cyan) are shown (PDB entry 5MUU). The amino- and carboxy-terminal ends of one of the red monomers are indicated

2002). In phage φ8, the P8 shell appears to be missing (El Omari et al. 2013b). The nucleocapsid of φ6 is enclosed by a lipid bilayer, which interacts with P4 (and P8 when present). The membrane contains four viral integral membrane proteins: P6, P9, P10, and P13. In addition, the spike protein P3 is attached to P6 and protrudes 2 nm from the membrane surface (Jäälinoja et al. 2007; Fig. 10a).

Overview of Bacteriophages Containing Double-Stranded DNA Genomes

Bacteriophages with double-stranded DNA genomes comprise more than 95% of all currently identified phages. Families for which the assembly and structure have been studied are the *Tectiviridae*, the *Corticoviridae*, and those of the *Caudovirales* order.

The *Tectiviridae* and *Corticoviridae* are icosahedral phages with an internal membrane. Phages of the *Caudovirales* order have icosahedral or prolate capsids and a tail. The *Caudovirales* are divided into three families with different tail morphologies. The *Podoviridae* family consists of phages with a short noncontractile tail, and the *Siphoviridae* family consists of phages with a long flexible tail. The last family, the *Myoviridae* family, consists of phages with a long, contractile tail. All these bacteriophages appear to have one vertex that is different from the other 11 (although this hasn't been studied yet for the *Corticoviridae*). This specialized or unique vertex is where the DNA translocation apparatus attaches temporarily to transfer the phage DNA into the capsid. In the tailed phages, it is also where the tail is attached and where the DNA leaves the capsid to enter the host bacterium.

The *Tectiviridae* Family

The *Tectiviridae* are icosahedral viruses. There are three known genera; *Alphatectivirus*, infecting Gram-negative bacteria; *Betatectivirus*, infecting Gram-positive bacteria; and the newly discovered *Gammatectivirus* GC1 (Philippe et al. 2018). The most well-known tectivirus is PRD1, which has a pseudo $T = 25$ capsid of around 65 nm in diameter. PRD1 infects several Gram-negative bacterial species containing Inc-type conjugative plasmids, including *Salmonella* and *E. coli*. These Inc plasmids encode the phage receptor and DNA translocation machinery. PRD1 has a linear 15 kb double-stranded DNA genome. Upon infection, the inner membrane forms a tube, allowing for DNA transfer directly into the host cell cytoplasm.

Assembly

The early operons of the viral DNA newly transferred into the host bacterium direct the production of the terminal protein P8, the viral DNA polymerase P1, and the two single-stranded DNA-binding proteins (P12 and P19) for efficient DNA replication (Butcher et al. 2012). The terminal protein P8 functions as a primer for complementary strand DNA synthesis. It also localizes the DNA replication complex to the bacterial nucleoid.

The late operons include genes for the structural proteins and chaperone proteins important for their assembly (Butcher et al. 2012). PRD1 assembly begins with the expression of soluble capsid proteins such as trimers of P3, trimers of P5, and pentamers of P31. At the same time, phage membrane proteins (P7, P11, P14, P16, and P18) are produced and inserted into the cytoplasmic membrane of the host. The proper folding of some viral proteins, such as the major capsid protein P3, requires the action of either the host GroEL-GroES chaperone complex or a complex of GroEL and the viral chaperone P33. Virion assembly is initiated by pinching off a membrane patch containing phage proteins from the host membrane (Fig. 11). This budding-off from the membrane involves both nonstructural phage proteins (membrane bound P10 and soluble P17) and structural phage proteins (P3, P30, P31, and P6). Curvature is introduced into the capsid by the P31 pentamers and the membrane protein P16. Extended dimers of the protein P30, on the inside of the P3/P31 capsid,
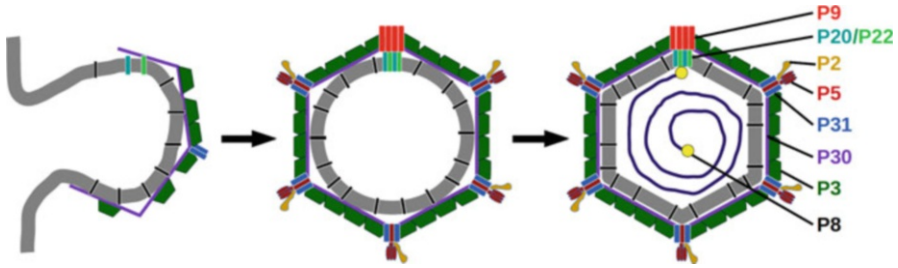
Fig. 11 Assembly of the tectivirus PRD1. In the left panel, a segment of the host membrane (gray) containing phage proteins is shown, already partially covered with capsid proteins. After completion of the capsid, the membrane segment is pinched off and the unique vertex closed (middle panel, unique vertex at the top). The preformed capsid is now ready for translocating the phage genome into it. After DNA translocation, the lagging copy of the covalently DNA-attached protein P8 plugs the hole in the unique vertex, and the internal pressure pushes the membrane against the inside of the capsid (right panel). Phage proteins are color-coded

are thought to be important for regulating the length of the icosahedral edges (Abrescia et al. 2004). The receptor-binding fivefold vertices are completed with the joining of the spike proteins P5 and P2, whose incorporation is P31-dependent.

At the unique vertex, a complex of proteins P20 and P22 is the nucleating site for the assembly of the packaging efficiency factor P6 and packaging ATPase P9 (Hong et al. 2014). The genome is probably recruited to the unique vertex by protein P8, which is linked to the 5′ ends of the DNA. ATP hydrolysis by P9 drives the translocation of the genome and the leading copy of P8 through the P20/P22 channel and into the procapsid. After packaging, the pore in the vertex is sealed by the lagging copy of P8, and the increased internal pressure expands the membrane to reach its final shape (Fig. 11).

Structure

The mature virion of PRD1 has a mass of 66 MDa and contains about 18 different proteins (Abrescia et al. 2004; Fig. 12a). The external pseudo $T = 25$ icosahedral capsid shell is formed by 235 trimers of the major capsid protein P3. The P3 monomer contains two jelly roll domains, endowing the trimers with an almost hexagonal shape (Fig. 12b). Four P3 trimers constitute the icosahedral asymmetric unit, except for the five asymmetric units surrounding the unique vertex which consist of three P3 trimers. The amino-terminal and carboxy-terminal extensions of the 12 P3 copies in the regular asymmetric unit adopt different conformations of the amino-terminal and carboxy-terminal extensions, depending on the location within the asymmetric unit. These conformations allow differential interaction with the membrane, other subunits of the trimer, other trimers, the vertex proteins, and the protein P30. P3 trimers are arranged on a framework of 60 copies of the tape measure protein P30. P30 forms extended dimers locked together by amino-terminal hooks. These dimers cement P3 trimers along the icosahedral facet edges and interact with P16 at the vertices (Butcher et al. 2012).
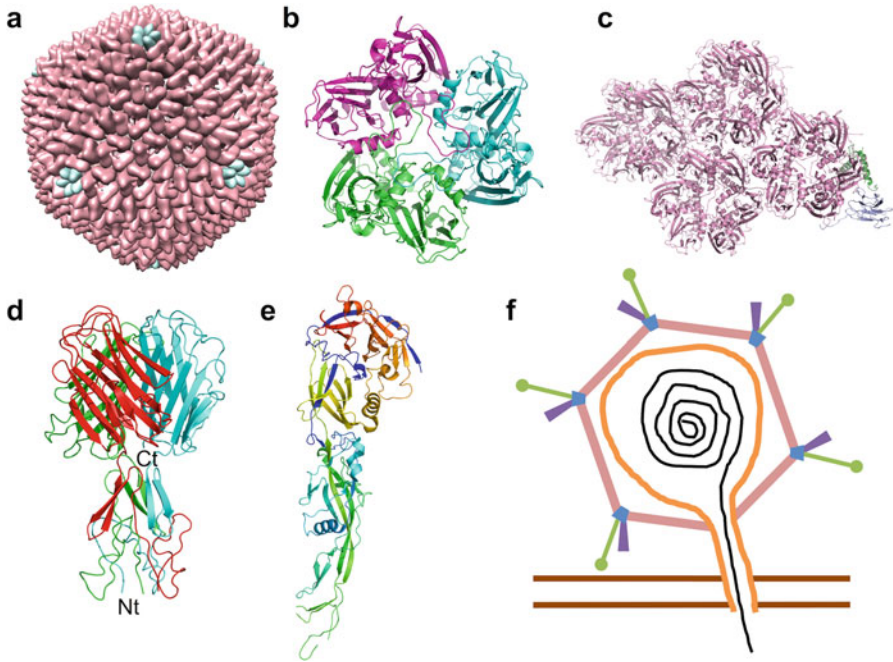
Fig. 12 Structure of the tectivirus PRD1. (**a**) Overall structure of the icosahedral protein capsid viewed down a threefold axis (PDB entry 1W8X). The double-barrel structure of the P3 monomers is shown in pink and the single-barrel structure of P31 in light blue. P3 forms trimers (pseudo-hexagons) and P31 forms pentamers. (**b**) Structure of the major capsid protein P3 trimer (PDB entry 1CJD) as seen from the outside of the capsid. The three protein chains are colored green, cyan, and magenta. The amino- and carboxy-termini are at the back. (**c**) Structure of the asymmetric unit of the icosahedron seen from the outside with four trimers of P3 in pink, a monomer of P31 in light blue, and part of a monomer of P16 in green. (**d**) Structure of the carboxy-terminal end of the trimeric central spike P5 (PDB entry 1YQ8). The three monomers are colored green, red, and light blue. The amino-terminal and carboxy-terminal ends of the red monomer are indicated. (**e**) Structure of the lateral receptor-binding spike protein P2 (PDB entry 1N7V). The chain is colored as a rainbow from dark blue (amino-terminal) to red (carboxy-terminal). (**f**) Schematic representation of the viral membrane (orange) tube "injecting" the DNA (black) into the host cell. The viral capsid is shown with P3 in pink, P31 in light blue, P5 in green, and P2 in purple. The host cell inner and outer membranes are in brown

The regular fivefold vertex consists of the membrane anchor protein P16, the penton base protein P31, the receptor recognition protein P2, and the spike protein P5 (Hong et al. 2014; Fig. 12c). The penton base protein P31 has a single jelly roll domain that packs with other subunits to form the base of the vertex complex. P31 is linked to P3, the carboxy-terminal end of P30, and to P16. P2 and P5 constitute two separate spikes. The spike protein P5 is a trimer with a carboxy-terminal TNF-like domain and a stalk with, in part, a triple β-spiral fold (Merckel et al. 2005). The amino-terminal domain of the homotrimeric P5 protein is embedded in the

pentameric penton base of the vertex, P31. P2 is an elongated monomer attached to the vertex complex at an angle. The P2 protein has a seahorse shape with an extended β-sheet tail and a β-propeller head domain (Xu et al. 2003). The head domain binds to the host receptor complex (Huiskonen et al. 2007). P3, P31, and P5 are structurally related to the adenovirus hexon, penton, and fiber proteins, respectively, but P2 is unique to the *Tectiviridae* (Merckel et al. 2005).

The unique vertex contains the transmembrane proteins P20 and P22, the ATPase P9, and the packaging efficiency factor P6. A hexamer of the transmembrane heterodimer P20/P22 forms the central genome delivery channel (Hong et al. 2014). Part of the packaging efficiency factor P6 is anchored to the center of the transmembrane channel. The remaining region of P6 remains exterior to the membrane, associating with P9 and forming a twelvefold symmetry portal complex surrounded by ten P3 trimers.

Beneath the capsid lies a membrane which, in the mature virion, follows the shape of the capsid due to the pressure of the packaged genome, the presence of viral transmembrane proteins, and the interactions with P3 (Hong et al. 2014). Half of the mass of the internal membrane has been attributed to membrane-associated proteins, either integral membrane proteins (P7, P11, P14, P16, P18, P20, P22, P32, and P34) or peripheral membrane proteins (P15). The lipid composition of the viral membrane includes predominantly phosphatidylglycerol (43%), phosphatidylethanolamine (53%), and cardiolipin (4%). The distribution of these lipids within the membrane is asymmetric. The inner leaflet contains more phosphatidylethanolamine, whose zwitterionic nature might stabilize the negative charge of the genome. The outer leaflet is enriched in phosphatidylglycerol and cardiolipin, whose negative charge might interact with the positively charged base of P3 (Cockburn et al. 2004). Inside the PRD1 virion, the double-stranded DNA is covalently bound to the terminal protein P8 at both ends through a 5′-linkage. As mentioned before, P8 acts as a primer for replication, but it is also a recognition signal for packaging. The PRD1 genome is presumably tightly wound, which results in a highly pressurized capsid interior. The resulting pressure likely facilitates the formation of the membrane tube responsible for genome transfer (Santos-Pérez et al. 2017; Fig. 12f).

The *Corticoviridae* Family

The *Corticoviridae* are icosahedral viruses containing an internal lipid membrane like the *Tectiviridae* (Fig. 13). Only one species has so far been included in this family: *Pseudoalteromonas* phage PM2. PM2 particles are formed by an icosahedral capsid with an approximate diameter of 60 nm and a triangulation number $T = 21d$, containing spikes at the vertices. This capsid surrounds an inner lipid bilayer (the lipid core), in which eight different proteins are embedded (Kivelä et al. 2008). The lipid-protein complex is known as the lipid core. In turn, the lipid core encloses a 10 kb, circular, supercoiled molecule of double-stranded DNA (Espejo et al. 1969). The DNA encodes 21 potential open reading frames (Männistö et al. 1999).

Fig. 13 Simplified view of a corticovirus. The major capsid proteins P1 and P2, as well as the lipid core, are indicated. The membrane proteins P3 to P10 are shown as colored shapes, but are not distinguished individually
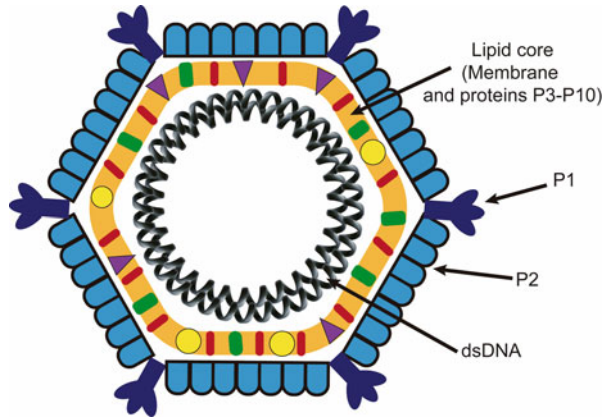
Assembly

The corticovirus phage PM2 recognizes its Gram-negative *Pseudoalteromonas* hosts using its P1 spike protein (Kivelä et al. 2002). Binding of P1 to the host triggers the uncoating of the virion, allowing the lipid core to interact and fuse with the host outer membrane (Huiskonen et al. 2004). Then, protein P7 degrades the periplasmic peptidoglycan layer (Kivelä et al. 2004), allowing the highly supercoiled DNA to reach, and then pass through, the cytoplasmic membrane. It is thought that DNA replication and assembly of the virus occur near the point of infection, at a site anchored to the inside of the host cytoplasmic membrane (Brewer 1978). The DNA replicates using a rolling circle mechanism (Canelo et al. 1985). Transcription of PM2 genes is carried out by the host RNA polymerase, using the highly supercoiled PM2 DNA as a template (Zimmer and Millette 1975).

One model for assembly proposes that two dimers of the transmembrane protein P3 interact with a monomer of the P6 protein, forming the protein scaffold building block (Fig. 14). Triggered by interaction with supercoiled DNA, and probably involving the P4 protein, three building blocks associate to form a subassembly corresponding to an icosahedral facet. The P6 protein determines the angle between adjacent facets and probably drives the membrane curvature. Recruitment of further protein subassemblies leads to the formation of DNA-containing vesicles covered in P3 and P6 proteins, to which the major capsid protein P2 would bind to form the virion. Specifically, in every virus facet, P3 dimers form a planar triangle of helices at the icosahedral threefold axis, to which trimers of the P2 proteins are attached. For this interaction, calcium ions are required. Finally, the pentameric P1 protein is incorporated (Abrescia et al. 2008). P3 and P6 are maintained in the mature particles (Kivelä et al. 1999), so it is thought that they are not only used for phage assembly but also to stabilize the mature virion.

The release of progeny is mediated by two proteins, P17 and P18. They are synthesized at about half an hour post-infection. P17 is a holin that gets inserted in the cell cytoplasmic membrane, forming pores through which the cellular lytic factor can reach the periplasmic space and digest the peptidoglycan layer. Through these
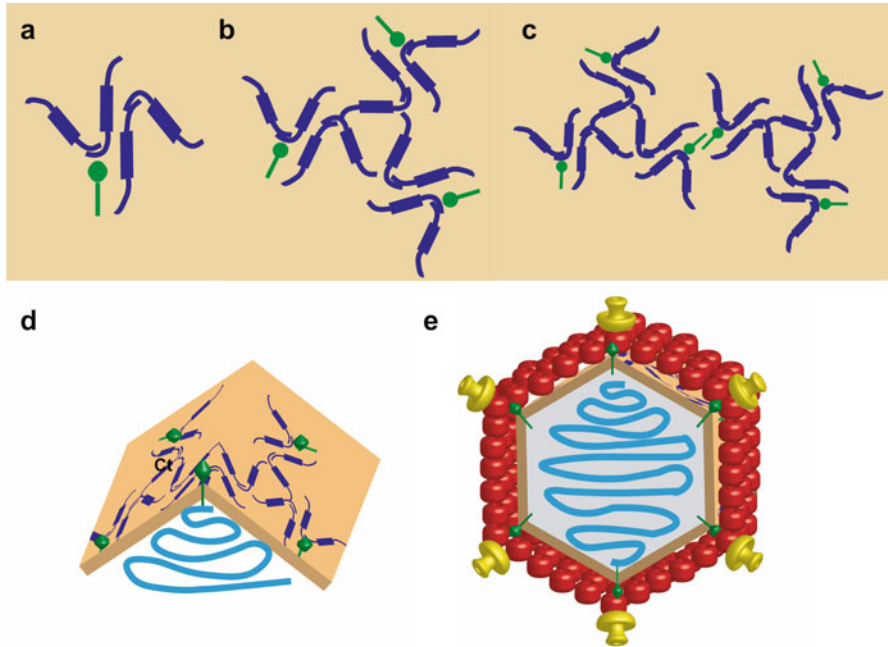
Fig. 14 Corticovirus assembly. (**a**) The protein scaffold building block is formed by two dimers of the P3 protein (dark blue) interacting with a monomer of the P6 protein (green), in the membrane. (**b**) Three building blocks associate to form a subassembly corresponding to an icosahedral facet. P6 proteins from two different subassemblies bind to each other, probably influenced by its interaction with supercoiled DNA. (**d**) The P6 protein interaction drives the membrane curvature and determines the angle between adjacent facets. DNA is shown in light blue. (**e**) Further protein subassemblies associate to form DNA-filled vesicles covered in P3 and P6 proteins, to which the P2 major capsid protein (red) and the P1 vertex protein (yellow) bind to form the virion

pores, the P18 protein also penetrates, reaching the outer membrane and disrupting it (Krupovic et al. 2007).

Structure

The mature PM2 virion is formed by an icosahedral capsid. The facet-to-facet and vertex-to-vertex capsid dimensions are 57 and 64 nm, respectively. The surface of the capsid is formed by 600 copies of the P2 protein organized in crown-shaped trimers (Fig. 15). Each P2 monomer is composed of two jelly roll domains, so each trimer occupies the quasi-sixfold position of a pseudo $T = 21$ icosahedral lattice. Each vertex of the icosahedral capsid is occupied by protruding spikes formed by a pentamer of protein P1. Every P1 monomer has three domains, a protruding globular distal domain through which the phage interacts with its host, a central domain, and a proximal jelly roll domain which interdigitates with the surrounding P2 trimers to form the base of the vertex. One icosahedral asymmetric unit contains one P1 monomer and ten P2 monomers (Huiskonen et al. 2004). Underneath the protein
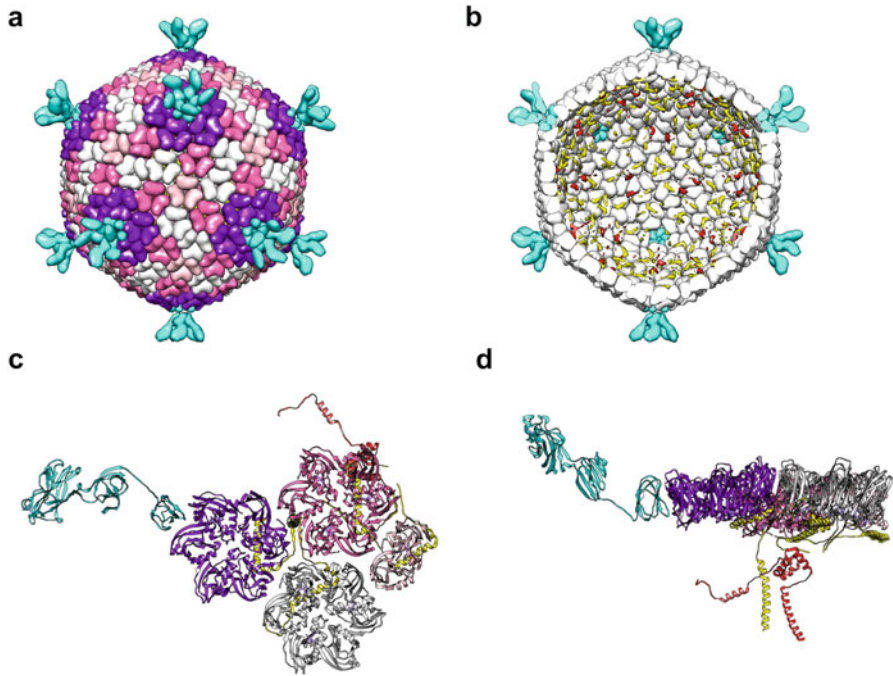
Fig. 15 Corticovirus structure. (**a**) Overall structure of the virion seen from the outside. The P1 protein pentamers are shown in light blue and the P2 trimers in white, light pink, pink, and purple. (**b**) Structure as seen from the inside of the capsid. Here, all the P2 trimers are shown in white, while the ordered parts of P3 are shown in yellow and the ordered parts of P6 in red. (**c**) Structure of the icosahedral asymmetric unit (PDB entry 2W0C) seen from the outside of the capsid, with the same coloring as in the previous panels. (**d**) Structure of the icosahedral asymmetric unit seen from the side

capsid, there is a lipid bilayer which is associated with eight different proteins (P3–P10). The lipids and the proteins form a particle filled with the dsDNA genome, called the lipid core (Kivelä et al. 2002).

There are 240 copies of P3 and 60 copies of P6 in the lipid core (Abrescia et al. 2008). The P4, P5, P7, P8, and P10 proteins could not be localized in the icosahedrally averaged electron density map, so their stoichiometry has not been determined (Kivelä et al. 2008). There are four copies of P3 and one of P6 per icosahedral asymmetric unit. Both P3 and P6 comprise an ectodomain and a transmembrane helix. Protein P3 is arranged in 120 asymmetric dimers. Groups of P3 subunits are linked due to the interaction of the amino-terminal end of a P3 dimer with the protein P6, which inserts into the lipid layer along the edges of the virus facet. In each icosahedral facet, P3 dimers form planar triangles of helices at threefold axes to which P2 trimers attach, connecting the lipid core to the outer capsid (Abrescia et al. 2008). Inside the lipid core, it seems that the strongly supercoiled DNA is organized by interactions with the membrane and membrane

proteins. The P6 transmembrane domain together with other components such as the P4 protein could mediate this interaction.

Overview of Order *Caudovirales*

Of all identified viruses, the *Caudovirales* are the most numerous. They consist of an icosahedral or prolate head with a tail connected to one of the 12 vertices (Fig. 16). The *Caudovirales* contain bacteriophages of different sizes, both in regard of their physical size and genome length. Correspondingly, the triangulation numbers of their head domains vary. The double-stranded DNA in the capsid is densely packed into a condensate at around 0.5 g/ml (Black and Thomas 2012). The DNA does not follow the icosahedral symmetry of the outside capsid and is organized in shells with

Fig. 16 Overall structure and general infection mechanism of the three kinds of *Caudovirales* members. Schematic structures are shown of a podovirus before (**a**) and after (**b**) genome transfer, a siphovirus before (**c**) and after (**d**) genome transfer, and a myovirus before (**e**) and after (**f**) genome transfer. Capsid proteins are in black, phage genomes in blue, and phage or host proteins that allow genome transfer are in red. The bacterial membrane is shown in brown
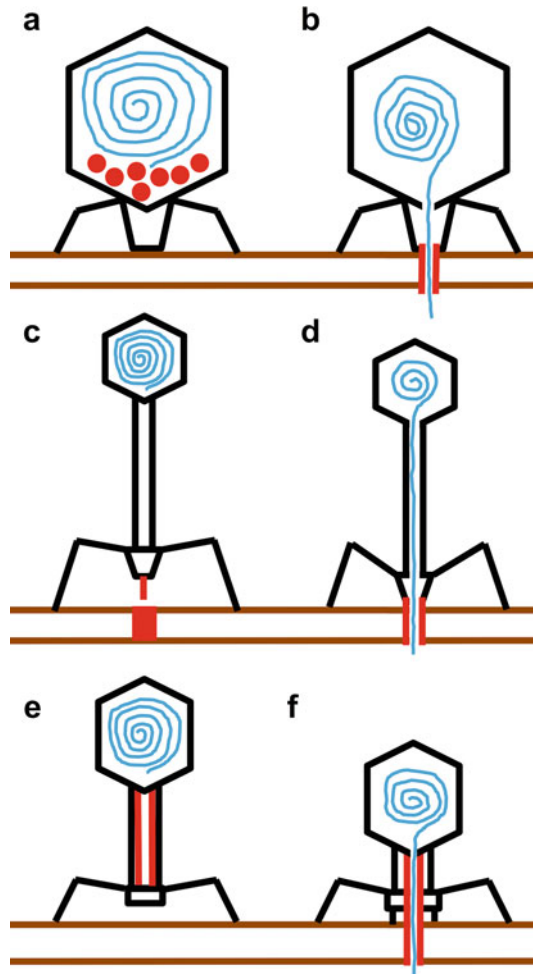
an approximate 2.5 nm spacing. The DNA may be naked, have dispersed unstructured proteins embedded within the DNA, have a small number of localized proteins, or have a significant protein core that functions as a DNA translocation device or is itself translocated into the bacterium upon infection.

The tail functions as a device to recognize a suitable host cell and a conduit for efficient transfer of their double-stranded genome into it. In many cases, fibers or spikes are part of the tail complex and function to recognize receptors on the host cell (Garcia-Doval and van Raaij 2013). This receptor binding is generally reversible and is followed by an irreversible interaction with a secondary receptor. After receptor binding, the phage DNA is ejected into the cell cytoplasm and starts directing the generation of progeny phage particles. Members of the *Podoviridae* have an extensile tail: proteins from inside the capsid are extruded and form a tube connecting to the cytoplasm of the bacterium (Fig. 16). The *Podoviridae* account for about 15% of all identified *Caudovirales* (Ackermann and Prangishvili 2012). Members of the *Siphoviridae* have a noncontractile tail; proteins at the end of this tail probably contact specific complexes on the bacterial cell wall where it is close to the cytoplasm, so they can eject their DNA directly into it. The *Siphoviridae* family comprises more than half of the viruses in the *Caudovirales* order. Finally, members of the *Myoviridae* have a contractile outer tail sheath, driving the inner tail tube through the bacterial cell wall and making a direct connection with the cytoplasm. The *Myoviridae* account for about a quarter of the known members of the order *Caudovirales*. Hereafter, the assembly processes and structures will be described in more detail and illustrated for some well-known bacteriophages of the *Caudovirales* order.

Caudovirales Head Assembly

Capsid formation in the *Caudovirales* starts with the portal protein, of which 12 copies form a dodecameric ring. Scaffold proteins assemble onto the portal protein (also called connector protein), and the capsid proteins assemble around this scaffold (Casjens and King 1975) to form the immature phage head (Fig. 17). When the head is complete, terminase subunits bind to the portal ring, forming a DNA packaging motor. This motor translocates the genome into the capsid. During this packaging, the capsid expands, thinning out its wall, and allowing the entry of more double-stranded DNA. During this maturation, scaffolding proteins are digested and leave the capsid in a process associated to the DNA packaging (Suhanovsky and Teschke 2015). When the head is full or a terminase signal is encountered (depending on the phage species), the DNA is cleaved, the portal shuts, the terminase subunits dissociate, and a connector complex binds to the portal ring.

For the *E. coli* siphovirus HK97, atomic models of maturation intermediates and of the mature empty capsid have been determined (Fig. 18) (Veesler et al. 2012a). Here, the capsid proteins covalently cross-link to each other during the final expansion step, providing extra stability (Ross et al. 2005). In *Salmonella* phage ε15 and coliphage K1-5, there are minor capsid proteins, necessary to keep the capsid stable.
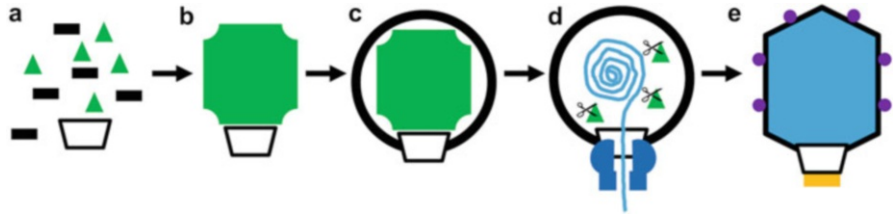
Fig. 17 General mechanism of *Caudovirales* head assembly. The portal protein ring (also called connector protein, in white), scaffolding proteins (green), and capsid proteins (black) are shown (**a**). The portal protein ring serves as the base, onto which the scaffolding proteins (green, **b**) and then the capsid proteins (black, **c**) assemble. (**d**) Proteases (shown as scissors) cleave the scaffolding proteins, and the terminase complex (dark blue) translocates the phage genomic double-stranded DNA (light blue) into the expanded capsid (**e**). When the capsid is full, head completion proteins (orange) take the place of the terminase complex, and decoration proteins (purple) may bind to the capsid
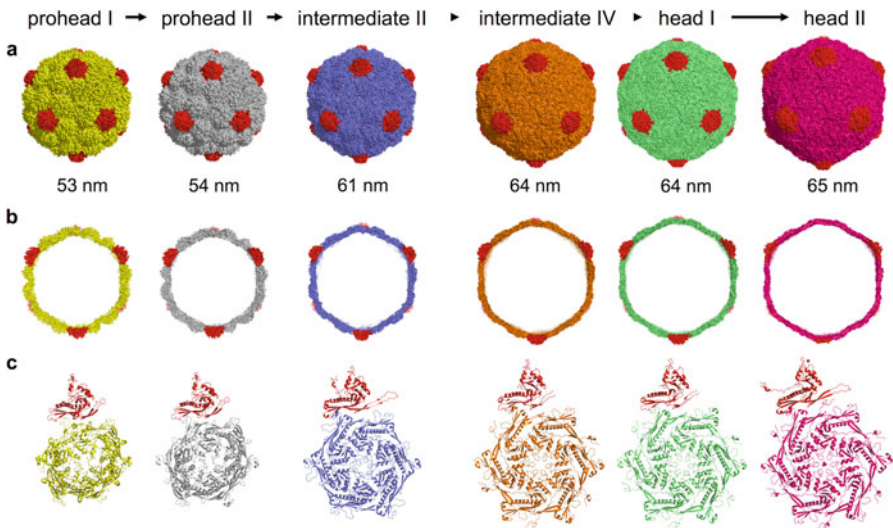

Fig. 18 Assembly intermediates of the bacteriophage HK97 head. (**a**) View along the icosahedral threefold axes of the different intermediates. The hexamers of the capsid protein are colored differently for each intermediate, while the pentamers are colored red. (**b**) Cross section of the intermediates shown above. Note the increase in internal volume and the thinning of the capsid wall during maturation. (**c**) Ribbon diagrams of the icosahedral asymmetric unit, consisting in each case of one hexamer (colored differently for each intermediate) and one subunit of the pentamer (red). Prohead I, prohead II, intermediate II, intermediate IV, head I, and head II are PDB entries 3QPR, 3E8K, 3DDX, 2FRP, 2FS3, and 2FT1, respectively

They are also called staple proteins (i.e., as metaphorically equivalent to the *stapling* of sheets of paper together) or instead cementing proteins and play a key role in the stability of the phage capsid against extreme pH or temperature. In other *Caudovirales* capsids, there are no cementing proteins or cross-linking, but stable

non-covalent interlocking interactions between the capsid proteins lead to a similar result (Morais et al. 2005).

To describe head assembly in more detail, we will discuss the example of bacteriophage T4 and note some interesting differences with other phages. To start assembly of the T4 capsid, a circular dodecamer of portal proteins assembles on the inner side of the host cytoplasmic membrane with the help of gp40. Then, prohead core proteins, gp21, gp22, gp67, and gp68; the initiation proteins, IPI, IPII, and IPIII; and gpalt form what is to be an internal scaffold, on the portal complex. The capsid proteins gp23 and gp24 assemble around this core to establish the phage prohead. Gp23 forms the hexameric tiles of the facets, while gp24 forms the pentameric vertices (in many other phages, the same protein is used to assemble the hexameric and the pentameric tiles). The prohead protease gp21 cleaves amino-terminal residues from gp23, gp24, gp67, and gp68 and digests gp21, gp22, the internal proteins, and gpalt into small fragments. These small fragments are then totally digested and leave the capsid (Fokine and Rossmann 2014).

Cleavage of the N-terminus of the capsid proteins triggers a rearrangement, flattening the capsid walls. This increases the phage head volume by about 50%. The immature capsid leaves the host membrane, and a complex of terminase proteins and DNA can now bind to the portal protein and form the genome-packaging machine. The small terminase subunit triggers the activity of large terminase protein, which is the ATP-consuming motor that pushes the DNA into the capsid (Yap and Rossmann 2014). The *Bacillus subtilis* phage ϕ29 lacks the small terminase subunit and instead has a 174-nucleotide pRNA that helps TerL to package the genome (Rao and Feiss 2015). Packaging stops when the portal complex detects a certain pressure. This is the head-full packaging mechanism that phages like T4 or Mu use. For other phages, such as P2, the terminase complex detects a conserved sequence at the beginning of a full genome copy. Packaging is judged complete when the second conserved sequence is detected at the end of the genome. Whatever the mechanism, the large terminase stops it by cutting the DNA (Nemecek et al. 2007). Neck proteins may now bind to the portal vertex. In the case of bacteriophage T4, the neck proteins gp13 and gp14 bind to the portal complex after DNA packaging. They serve as adaptors to bind the portal dodecamer to the gp15 hexamer on the top of the tail (see below). Fibritin proteins (gpwac) are placed around the neck. They make up both the collar and the whiskers. Now, the full head can bind to the pre-assembled tail (Yap and Rossmann 2014).

Caudovirales Head Structure

All members of the *Caudovirales* order have a capsid, or head, which is full of double-stranded DNA. These heads are made up of the same basic building blocks, although some have additional stabilization or decoration proteins (Fig. 19). They vary in size, and the size is correlated with the genome length (Lavigne et al. 2009). A very large known member is bacteriophage G of *Bacillus megaterium*, with a genome size of nearly 500 kb, several hundred genes, a head domain of 160 nm in
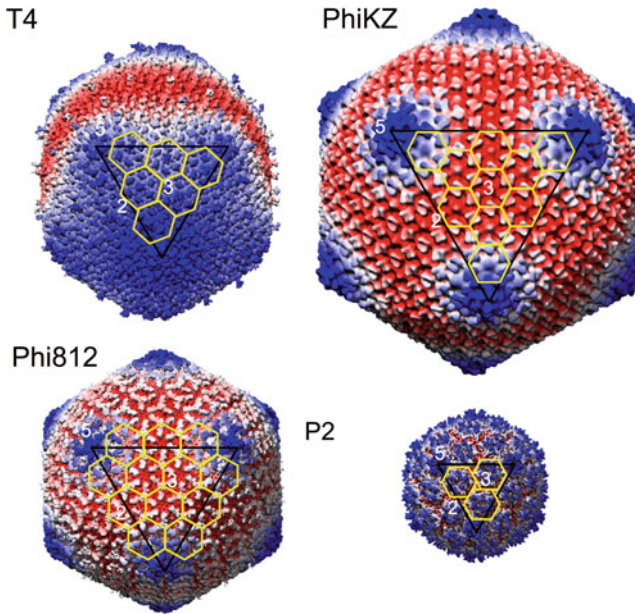
Fig. 19 Representation of several different *Caudovirales* capsids. The name of the virus is shown above each capsid. For coliphage T4 (top view; EMDB entry EMD-6323), *Pseudomonas* phage φKZ (EMDB entry EMD-1392), *Staphylococcus* phage φ812 (EMDB entry EMD-8304), and coliphage P2 (EMDB entry EMD-5406), a surface-rendered capsid is shown, where the distance from the icosahedral center is color-coded, red for closer to the center and blue for more distal, with white for in-between distances. The icosahedral facet nearest to the reader is indicated with a black triangle and some of the capsid protein hexamers with yellow hexagons. The approximate locations of the fivefold, threefold, and twofold symmetry axes are indicated with white numbers. EMBD is the Electron Microscopy Data Bank (https://www.ebi.ac.uk/pdbe/emdb/)

diameter, and a tail over 450 nm long (Ageno et al. 1973). On the other hand, at just over 14 kb and with a capsid diameter of 43 nm, the *Rhodococcus* phage RRH1 is an exceptionally small phage, with only 20 genes (Petrovski et al. 2012). Many phage heads are icosahedral, but other phage capsids are prolate, i.e., elongated, like those of *Bacillus* phage φ29 and coliphage T4. The φ29 head is 54 nm long by 45 nm wide, while the head of T4 has a length of 115 nm and a width of 85 nm (Fokine et al. 2004).

All the *Caudovirales* main capsid proteins have the HK97 fold (Fig. 20). The HK97 fold allows capsid proteins to be flexible, so the capsid can change conformation during maturation. However, it is also sufficiently strong to keep a large amount of DNA in the capsid at a high internal pressure (Suhanovsky and Teschke 2015). The main capsid proteins make up the hexamers on the facets or the pentamers on the vertices, either with the same protein making up both the hexamers and pentamers (e.g., gpN in phage P2) or different but structurally related proteins making up the hexamers and pentamers (e.g., gp23 and gp24, respectively, in phage T4). Sometimes they have added domains and loops or a partially altered topology.
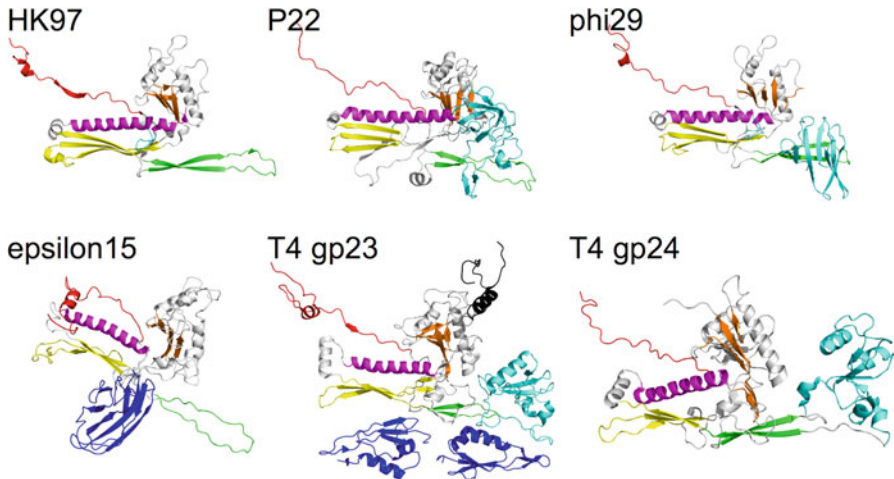
Fig. 20 *Caudovirales* major capsid proteins: the HK97 fold. Capsid proteins of phages HK97 (PDB entry 2FT1), P22 (PDB entry 5UU5), φ29 (PDB entry 1YXN), ε15 (PDB entry 3 J40), and T4 major capsid proteins gp23 and gp24 (PDB entry 5VF3). The N-arm is shown in red, the E-loop in green, the P-loop in yellow, the spine helix in magenta, and the five-stranded β-sheet in orange. Phage ε15 and T4 capsids have cementing proteins, which are shown in dark blue. The carboxy-terminal anchor of the T4 Hoc decoration protein is shown in black. In phages P22, φ29, and T4, there is an extra domain (cyan) in the capsid protein that may have the same function as the cementing proteins have in other phages

The size and shape (icosahedral or prolate) of the capsid are probably largely determined by the scaffolding structure around which it assembles. In larger capsids, more hexameric tiles of the main capsid protein are incorporated, leading to higher triangulation numbers, up to $T = 52$ for very large bacteriophages (Hua et al. 2017).

Proteins with the HK97 fold have several common features (Fig. 20). A five-stranded β-sheet and two helices form the A-domain, which is at the center of hexamers and pentamers. The E-loop, the P-loop, and the spine helix form the P-domain, which makes up the outer part of hexamers and pentamers. An amino-terminal arm (N-arm) with α-helical and β-sheet content connects with a neighboring capsid protein hexamer or pentamer. In addition, some *Caudovirales* have cementing proteins (Fig. 20; e.g., Soc in phage T4). In addition to the cementing proteins, or instead of them, the major capsid proteins may have an extra domain in the capsid protein that may also have a stabilization function.

All *Caudovirales* capsids have one special vertex. At one of the 12 vertices, a dodecameric head-tail connector or portal protein is present instead of a pentamer of capsid proteins (Prevelige and Cortines 2018). The dodecameric portal initiates capsid assembly. The genomic DNA is packaged into the assembled capsid through this portal. The portal is essential for tail assembly, and DNA again passes through it upon adsorption. The basic structure of all portal proteins is the same, despite of, in many cases, little sequence homology (Fig. 21; Parent et al. 2018).
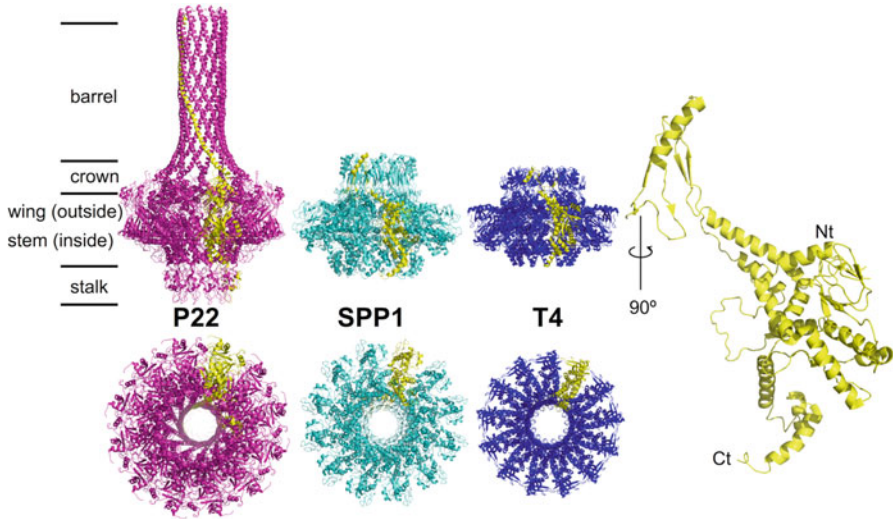
Fig. 21 Structure of *Caudovirales* portal proteins. Side (top) and bottom views (bottom, viewed from outside the phage) of portal complexes of phage P22 (PDB entry 3LJ5), SPP1 (PDB entry 2JES), and T4 (PDB entry 3JA7). One monomer is highlighted in yellow. The rough locations of the barrel, crown, wing, stem, and stalk domains are indicated on the left. Right: Single monomer of the T4 portal complex with the amino- and carboxy-termini indicated. It is rotated 90° with respect to the yellow monomer in the side view of the dodecamer next to it

Portal proteins contain mainly α-helices and coils. They have up to five domains: the barrel, the crown, the wing, the stem, and the stalk (Fig. 21; Prevelige and Cortines 2018). Some but not all portals have an α-helical barrel on the inside of the virus (Fig. 21; Olia et al. 2011). The barrel changes conformation upon genome packaging, going from unstructured to helical, acting as a pressure sensor. The portals of phages ϕ29 and T7 do not have barrels. These phages do not perform head-full assembly and so presumably do not need these barrels. Facing the inside of the viral capsid is the crown. The crown is flexibly linked to the wing. The crown and wing display the greatest variability between portals and allow conformation changes during packaging, head-full sensing, and opening for DNA exit. The wing contacts nearby capsid proteins directly and is important for transmitting conformational changes between DNA translocation and capsid structure. The wing is connected to the stem region by a flexible loop. The loops of the 12 portal monomers extend into the central portal channel and prevent leakage of the packaging DNA. The stem is formed by 12 helices, tilted by 30 to 50° with respect to the direction of the channel. The stalk forms the initial channel for DNA packaging. The central channel is about 3 nm in diameter, just enough to allow passage of double-stranded DNA. Upon maturation, conformational changes widen the channel to about 4 nm, presumably to allow smooth DNA delivery. The stalk is on the outer surface of the capsid and interacts with the terminase proteins during DNA translocation and with adaptor proteins to bind the tail in the mature virion.
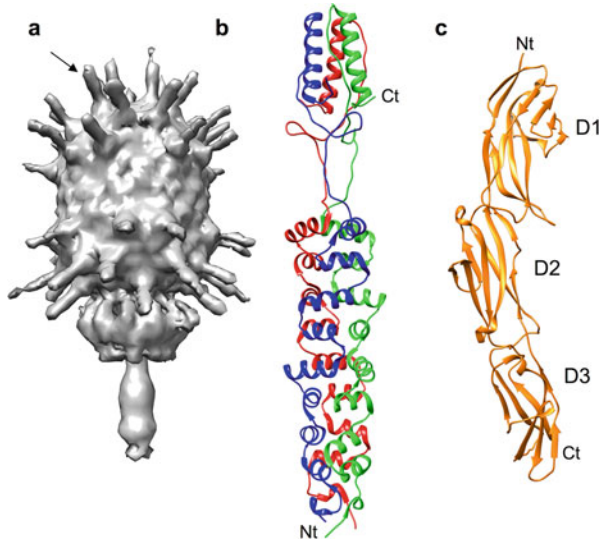
Fig. 22 Decoration proteins of *Caudovirales*. (**a**) Low-resolution cryo-electron microscopy structure of *Bacillus* phage ϕ29 (gray; EMDB entry EMD-1506). One of the head fibers is indicated with an arrow. (**b**) Fibrous part of the ϕ29 head fiber trimer colored in red, green, and blue (PDB entry 3QC7). The amino-terminal part forms a trimeric super helix-turn-helix-coiled coil, while the carboxy-terminal part forms a small tip domain to which each monomer contributes an α-helix. The amino- and carboxy-termini of the green chain are indicated. (**c**) Hoc from coliphage RB49, a close relative to phage T4 (orange; PDB entry 3SHS). The three amino-terminal immunoglobulin-like domains are indicated (D1-3); the fourth domain, which in the intact phage would bind to the capsid, is not resolved. The amino- and carboxy-termini are indicated

Apart from capsid proteins and cement proteins, *Caudovirales* capsids sometimes have what are called decoration proteins (Fig. 22). Decoration proteins are not necessary for capsid assembly or phage infectivity, but may help increasing phage binding to their bacterial host. They may also help to attach to surfaces where host bacteria are likely to pass, such as the lung or gut epithelia. Head decoration proteins can be used in phage display applications, for example, for displaying antigens when using phages as vaccination vehicles (Tao et al. 2018).

In phage T4 and T4-like phages, the Hoc protein (highly immunogenic outer capsid protein) protrudes from the center of the gp23 hexamers (Fig. 19; Fokine et al. 2011). Hoc is anchored to the capsid with its carboxy-terminal end and consists of four consecutive immunoglobulin-like domains (Fig. 22), exposing the amino-terminal domains to the medium. *Bacillus* phage ϕ29 has 55 head fibers bound to the capsid (Xiang and Rossmann 2011). The ϕ29 prolate capsid ($T = 3/Q = 5$) consists of the major capsid protein gp8 and the head fiber, which is a trimer of gp8.5. Gp8 has an additional domain (Fig. 20), which provides attachment sites for

the head fibers at quasi-threefold symmetry positions. The head fibers have two domains: a base that attaches to the capsid and a protruding fibrous domain. The fibrous domain has a unique helix-turn-helix supercoil fold capped with a small head domain that contains a short triple coiled coil (Fig. 22). In this case, the amino-terminal end of the protein binds to the phage capsid, and the carboxy-termini are exposed to the medium, like for bacteriophage fibers and tailspikes (see below).

Podoviridae Tail Assembly and Structure

Bacteriophages of the *Podoviridae* family have a relatively short, noncontractile tail which is not flexible (Casjens and Molineux 2012). The exact length and shape of these tails vary, presumably in relation to the phage host. A dodecameric adaptor protein binds to the portal, and one or more tail proteins bind to this adaptor protein. These tail proteins are often hexamers. Many podoviruses have six trimeric receptor-binding tailspikes or tail fibers that attach to the tail. When they bind to the phage receptor, this interaction leads to a conformational change in the tail to initiate DNA transfer into the host (Garcia-Doval and van Raaij 2013).

In the *Podoviridae*, after the head is assembled and filled with DNA, tail proteins are added to the portal sequentially, until the tail, and thus the phage, is complete (Fig. 23). In the case of coliphage T7, a dodecamer of gp11 binds first and functions as a gatekeeper complex, to retain the DNA in the capsid (Cuervo et al. 2013). Six copies of gp12 bind to gp11, forming a nozzle. Finally, six trimeric tail fibers (made up of the gp17 protein) are bound to the gatekeeper and retracted toward the capsid (Hu et al. 2013). During infection, the C-termini of the fibers dislodge and bind to the host lipopolysaccharide (Fig. 16; González-García et al. 2015).

The T7 tail is about 30 nm long and 17 nm wide, including the portal protein. An isolated tail-portal complex has been studied (Cuervo et al. 2013). It consists of a dodecamer of the portal protein gp8, a dodecamer of the gatekeeper protein gp11, a hexamer of the nozzle protein gp12, and six trimers of the fiber protein gp17. The carboxy-terminal half of the T7 tail fiber forms a threefold symmetric pyramid
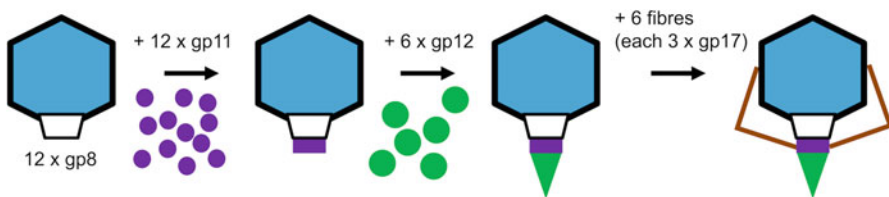

Fig. 23 Schematic overview of the construction of the bacteriophage T7 tail as an example of podovirus tail assembly. The DNA-filled head, previously assembled around the dodecameric portal complex (gp8, in white), forms the starting point. Twelve copies of gp11 (purple) join first, forming a circular gatekeeper complex. Subsequently, six gp12 proteins (green) bind and make up the nozzle. Finally, six pre-assembled trimers of gp17 fibers (brown) are incorporated

domain and a globular tip domain (Fig. 24c; Garcia-Doval and van Raaij 2012). Before infection, phage T7 particles have the fiber pointing upward, interacting with the icosahedral capsid (Hu et al. 2013). In the isolated tail, the amino-terminal part of the tail fiber can be clearly seen pointing upward (Fig. 24a; González-García et al. 2015). In this conformation, the nozzle is closed at the tip of the tail, with the six copies of the gp12 protein pointing inward. When the carboxy-terminal distal parts of the fiber contact the host membrane, the proximal amino-terminal ends of the fiber initiate a conformational change, leading to a straightening of the gp12 monomer and an opening of the nozzle (Fig. 24b). During infection, the core complex, consisting of multiple copies of the gp14, gp15, and gp16 proteins, moves from inside the capsid, where they sit just above the portal ring, to form a tube spanning the periplasmic space of the bacterial host (Hu et al. 2013). Presumably, the core complex proteins unfold to a large extent, pass through the gp8-gp11-gp12 complex, and refold in the periplasmic space. The core protein gp16 has been shown to have peptidoglycan digestion activity (Moak and Molineux 2004), so this probably helps the process. The phage genomic DNA then passes through the tail, through the periplasmic tube, and directly into the host cytoplasm. Structural studies on the *Prochlorococcus* phage P-SPP7 (Liu et al. 2010) suggest it has the same mechanism of infection as coliphage T7.

The tail of the *Salmonella* phage P22 (Fig. 24d) has an organization similar to the tail of coliphage T7 (Bhardwaj et al. 2014). Here, the dodecameric portal protein (gp1) binds to the dodecameric α-helical adaptor protein gp4 and the hexameric nozzle protein gp10. However, the nozzle is not closed, but its channel is occupied by amino-terminal part of the trimeric needle protein gp26. The rest of gp26 protrudes by about 14 nm (Tang et al. 2011). Gp26 is composed of a long α-helical coiled coil, interspersed with β-structure (Fig. 24f; Olia et al. 2007). In phage SF6, gp26 is capped with a globular tip domain (Bhardwaj et al. 2011), but in P22, this tip is absent.

As primary receptor-binding proteins and instead of thin, L-shaped, fibers, phage P22 has six stubby trimeric tailspikes, each made up of trimers of gp9 (Fig. 24f). Each tailspike has a small trimeric amino-terminal β-structured domain with which they attach to the phage neck (Seul et al. 2014). The distal carboxy-terminal domain is intertwined and has been shown to be important for the correct trimeric assembly and folding of the tailspikes (Takata et al. 2012). In the central part of the tailspikes, each protomer contains parallel β-helix domains. The β-helix domains function in adhesion to the O-antigen repeating units of the host lipopolysaccharide and cleavage of the O-antigen repeating units. Multiple rounds of cleavage and adhesion allow the phage to approach the host membrane. Presumably, the tip of gp26 then gets pushed against the membrane and leads to a conformational change in the tail, setting off the events necessary for DNA-transfer into the host. Tailspikes with a β-helical domain that cleaves the host O-antigen are a common occurrence in the adsorption devices of *Caudovirales*, not only in the *Podoviridae* but also in *Myoviridae* (Walter et al. 2008), so this method of approaching the membrane is apparently successful and necessary for infecting certain host bacteria. The exact
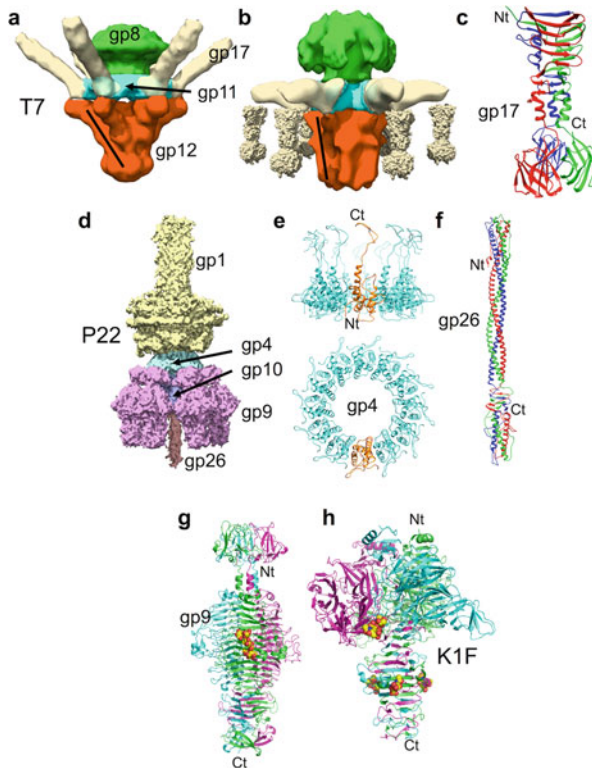
Fig. 24 Structure of tails of podoviruses that infect Gram-negative bacteria. (**a**) Structure of the coliphage T7 tail before DNA ejection (EMDB entry EMD-1163). Approximate volumes of the portal protein (gp8; green), adaptor protein (gp11; blue), nozzle protein (gp12; orange), and fiber (gp17; beige) are colored. (**b**) Structure of the coliphage T7 tail after DNA ejection (EMDB entry EMD-2717). The approximate positions of the carboxy-terminal domains of the fiber are indicated. Black lines indicate the positions of an individual gp12 molecule in each of the two structures, which straightens out to allow DNA ejection. (**c**) Structure of the carboxy-terminal domain of the phage T7 fiber (gp17; PDB entry 4A0U). The three chains of the protein trimer are colored differently, and the amino- and carboxy-termini of the green chain are indicated. (**d**) Structure of the *Salmonella* phage P22 tail (EMDB entry EMD-5348). The portal protein (gp1) is shown in beige, the adaptor protein (gp4) in light blue, the nozzle protein (gp10) in purple, the tailspikes (gp9) in magenta, and the tail needle (gp26) in brown. (**e**) Structure of the dodecameric P22 gp4 adaptor ring (PDB entry 4V4K) in side view (top) and bottom view (bottom). One of the monomers is highlighted in orange. The carboxy-terminal end interacts with the bottom of the dodecameric portal (gp1). (**f**) Structure of the phage P22 needle (gp26; PDB entry 2POH). The three chains of the protein trimer are colored differently, and the amino- and carboxy-termini of the red chain are indicated. (**g**) Phage P22 tailspike structure (gp9; PDB entry 2XC1) with the three chains of the protein trimer colored differently. A fragment of O-antigen receptor (PDB entry 1TYX) is super-posed in the binding site facing the reader. (**h**) Phage K1F endosialidase tailspike structure (PDB entry 1V0F) with the three chains of the protein trimer colored differently. Fragments of oligomeric α-2,8-sialic acid are shown

receptor-binding sites and enzymatic mechanisms of different tailspikes vary, but their structural framework is very similar.

Some podoviruses have a more complex tail with multiple tailspikes or fibers. An example is coliphage K1-5 (Leiman et al. 2007), in which the single receptor-binding protein is replaced with a protein that binds to two different tailspikes. One of these tailspikes is an endosialidase (Fig. 24h; Schulz et al. 2010b), which trimerises using an intramolecular chaperone (Schulz et al. 2010a), just like the T5 side tail fibers described in the next section. The endosialidase, which has multiple binding sites for sialic acid moieties (Fig. 24h) and an active site in each monomer, allows the phage to tunnel its way through the poly-sialic acid capsule of host bacteria. The other tailspike is more like the P22 tailspike.

Phage φ29 infects the Gram-positive bacterium *B. subtilis*, so its DNA does not need to traverse an outer membrane. Phage φ29 has a longer tail than many other podoviruses (about 50 nm). Here, the portal is a dodecamer of gp10, occupying the special vertex and forming the upper part of the collar. A tubular protein, dodecamer of gp11, binds to the portal forming the lower collar and the tail tube. A hexamer of gp9 and two copies of gp13 bind to the tail tube, forming the distal knob (Tang et al. 2008). Twelve appendages are attached to the phage collar (Fig. 25). Each appendage folds using an intramolecular chaperone (Xiang et al. 2009), like the endosialidase of coliphage K1-F and the fibers of the siphovirus T5 (see below). The appendages are homo-trimers of gp12 and can be in the up or down position (Farley et al. 2017). They are responsible for the digestion of the teichoic acid layer of the bacterium.

At the end of the tail, a knob complex is located, consisting of a hexamer of gp9 (Xu et al. 2016) and probably two molecules of gp13 (Xiang et al. 2008). Gp13 is a peptidoglycan-degrading enzyme that helps the tail knob to reach the bacterial cytoplasmic membrane. Gp13 has an amino-terminal domain with a lysozyme fold and a carboxy-terminal domain with an endopeptidase fold (Fig. 25e). The domains are linked by an oligo-glycine linker which allows flexibility and was not resolved in the crystal structure. Once the membrane has been reached, gp9 is able to form a pore through it to allow DNA transfer directly into the host cytoplasm (Xu et al. 2016). The arrangement of a longer tail and 12 appendages is not limited to podoviruses infecting Gram-positive bacteria. Coliphage N4 also has 12 appendages and an overall similar structure to phage φ29 (Choi et al. 2008).

Siphoviridae Tail Assembly

In the *Siphoviridae* (like *Myoviridae*), tail assembly takes place in parallel to head assembly, starting with the baseplate, i.e., the distal end of the tail (Davidson et al. 2012). Coliphage λ has been the model for siphovirus assembly (Katsura 1990). To start assembly, a trimer of gpJ makes up the distal tip structure of the tail (Xu et al.
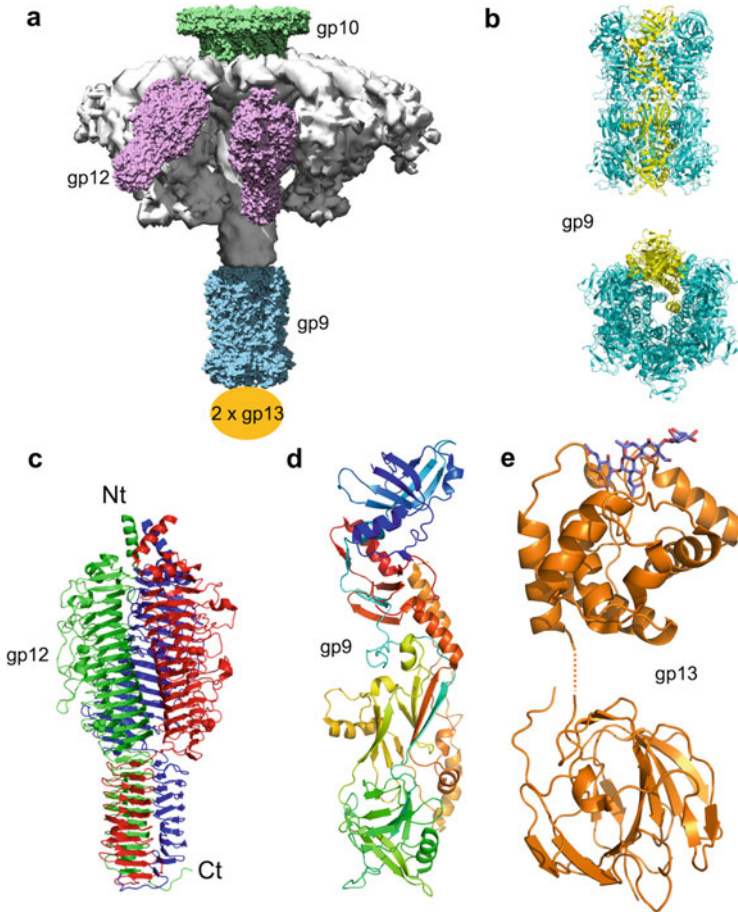
Fig. 25 Structure of bacteriophage φ29 tail. (**a**) Cryo-electron microscopy map of the φ29 bacteriophage (gray; EMDB entry EMD-1420) with the atomic structures of the connector protein gp10 (green), two appendages, each consisting of a trimer of gp12 (violet), and the tail knob, a hexamer of gp9 (cyan). The atomic structure of the appendages only contains the C-terminal region of the trimer. Hence, the N-terminal region is still in gray, as part of the cryo-electron microscopy map. Missing from the tip of the tail is gp13, an enzyme that degrades the peptidoglycan layer and facilitates access of the tail tip to the cytoplasmic membrane. (**b**) Side (top) and bottom (bottom) views of the knob complex. One monomer is highlighted in yellow. (**c**) Structure of the carboxy-terminal domain of the φ29 appendage (PDB entry 3GQ7). The three monomers are colored in green, blue, and red. The amino- and carboxy-termini of the green chain are indicated. (**d**) Structure of the gp9 knob monomer (PDB entry 5FB5) in the same position as the yellow monomers in B. The protein is colored in a rainbow representation, amino-terminus to carboxy-terminus in blue to red. (**e**) Structures of the amino-terminal lysozyme domain (PDB entry 3CT0) and the carboxy-terminal endopeptidase domain (PDB entry 3CSQ) of gp13. An N-acetyl glucosamine oligomer bound to the amino-terminal domain is shown in stick representation
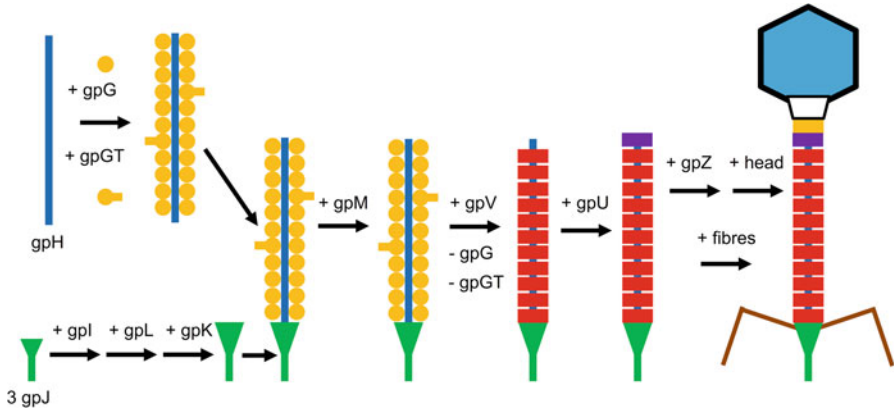
Fig. 26 Schematic overview of the assembly of bacteriophage λ as an example of siphovirus assembly. The assembly of the tail is shown in greater detail. The tail tip complex (shown in green, at the bottom of the figure), consisting of a trimer of gpJ and the gpI, gpL, and gpK proteins, joins up with a pre-assembled complex of the tape measure protein gpH (in blue, copy number unknown), covered with the tail assembly chaperone proteins gpG and gpGT (both shown in orange). GpM joins at an unspecified position. Incorporation of the tail tube protein gpV (in red) displaces gpG and gpGT. When the tail is complete, it is capped by gpU (purple) and gpZ also incorporates. To complete phage assembly, the DNA-filled head (colored as in Fig. 17) and the fibers (brown) join

2014). Subsequently, copies of gpI, gpL, and gpK are added (Fig. 26). At the same time, a sixfold helical coiled coil of the tape measure protein gpH gets covered with the tail chaperone proteins gpG and gpGT (gpGT is a read-through product of gpG gene due to a frameshift regulated to give a gpG-to-gpGT ratio of 30:1). GpG is thought to bind to gpH, increasing gpH's solubility, while gpGT may guide the incorporation of gpV building blocks. The C-terminal end of gpH binds to the top of the tail tip complex, and the tail tube initiator protein gpM binds at an unspecified position. Multiple copies of the tail tube protein gpV assemble onto the tip and around gpH, displacing gpG and gpGT. The tail termination protein gpU binds to the top of the tail, while gpZ also joins at an unknown location. The complete tail can now bind to the assembled phage head.

Not all siphoviruses have side tail fibers, but in the case of phage λ, side tail fibers – each consisting of trimers of the side tail fiber protein capped by trimers of the tail fiber assembly protein – also bind to the tail tip. The tail fiber assembly protein is expressed in large amounts and probably also functions as a chaperone for correct assembly of the side tail fibers. The exact number of side tail fibers that bind to each λ virion is uncertain. In the case of coliphage T5, another siphovirus, the three side tail fibers, which are each trimers of the pb1 protein, also assemble with the help of a chaperone. However, this chaperone is intramolecular and a carboxy-terminal extension of the pb1 protein. The intramolecular chaperone helps the fibers to fold, and once folded, a proteolytic site forms and the chaperone cleaves itself off (Garcia-Doval et al. 2015).

Siphoviruses infecting Gram-positive bacteria probably share many of the assembly features described above. This is confirmed by a model for the assembly of the lactococcal phage TP901-1 (Mahony et al. 2016). Here, the Dit protein is the hub to which the C-terminus of the tape measure protein and the N-terminus of Tal bind. On the top of this platform, tail construction takes place, while on the lower side, baseplate proteins involved in host recognition bind. Like in phage λ, gpG and gpGT chaperone the incorporation of the multiple tail tube protein subunits. The top of the tail is capped by Ttp. Tap, HTC1, and HTC2 are probably involved in head-to-tail assembly.

Siphoviridae Tail Structure

Phages of the *Siphoviridae* family have tails from just under 100 nm up to more than 400 nm long, depending on the species. These tails may be flexible and are not contractile. The tails are connected to the head portal via small adaptor proteins (Tavares et al. 2012), which generally form dodecameric rings. These adaptor proteins are of different kinds. Some are α-helical and have the same fold as the P22 gp4 protein (Fig. 24e). Examples are gp6 of the coliphage HK97 and gp15 of *Bacillus* phage SPP1. A second kind, including gpW of phage λ, has a different fold (Fig. 27b), consisting of two α-helices and a β-hairpin (Maxwell et al. 2001). GpW acts as a stopper that prevents exit of the genome (Perucchetti et al. 1988). Phages λ and SPP1 (Fig. 27a) have a second ring of proteins below the adaptor proteins, made up of dodecamers of gpFII and gp16, respectively. These are small β-structured proteins (Fig. 27c; Maxwell et al. 2002). In phage SPP1, this second ring of gp16 dodecamers acts as the stopper (Lhuillier et al. 2009). A loop from each of the 12 gp16 subunits extends into the tail channel, blocking DNA exit (Chaban et al. 2015).

The lower ring of connector proteins attaches to a hexameric ring of the tail terminator protein. The structure of the hexameric ring of the tail terminator protein of phage λ, gpU, has been determined (Fig. 27d; Pell et al. 2009a). Each monomer of the tail terminator protein is made up of a five-stranded β-sheet covered with α-helices on the outside. The main part of the tail is made up of a singular tubular structure, the tail tube, which generally has sixfold symmetry (Davidson et al. 2012). The tube is made up of stacked hexameric rings of the tail tube protein (shown in Fig. 27g for the myovirus T4, which has a homologous structure). The tail can be lengthened or shortened by incorporating more or less hexameric rings; the phage λ tail tube contains 32 stacked rings. The tail tube protein fold can be described as a β-sandwich or a folded-over β-sheet (Fig. 27e; Pell et al. 2009b). Like for the tail termination protein, the part facing inside is a β-sheet, and the inner diameter is around 4 nm. The outer diameter of the tail tube is about 9 nm. This diameter may be increased in siphoviruses if there are decoration domains present. For example, each copy of the tail tube protein of phage λ, gpV, has a carboxy-terminal immunoglobulin domain decorating the outside of the tail tube (Fig. 27e; Pell et al. 2010). The tape measure protein, which is located inside the tail tube, probably forms a
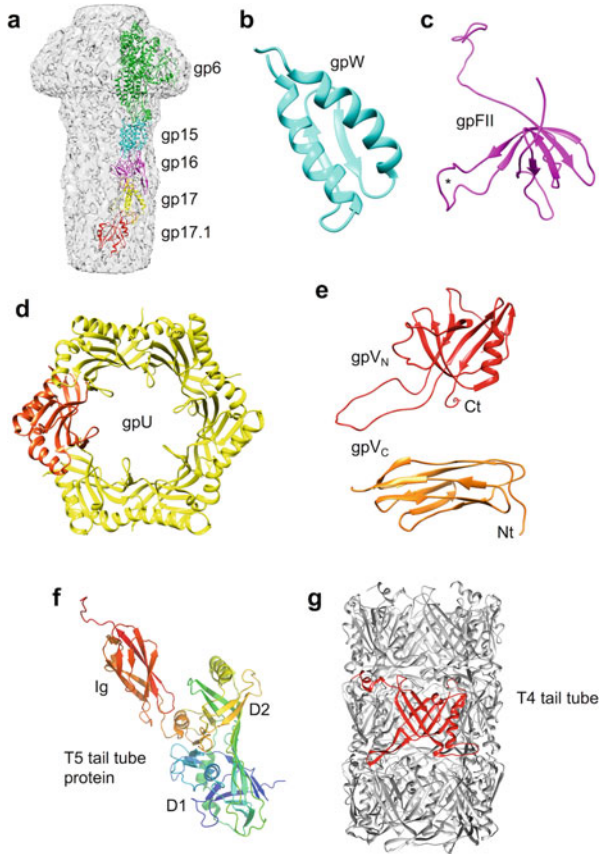
Fig. 27 Structural details of siphovirus tail neck and tube. (**a**) Cryo-electron microscopy map of the *Bacillus* phage SPP1 neck region (EMDB entry EMD-2993) with models of the portal protein gp6 (green, 2 of 12 total copies), the adaptor protein gp15 (cyan, 2 of 12 total copies), the stopper protein gp16 (magenta, 2 of 12 total copies), the tail terminator protein gp17 (yellow, 1 of 6 total copies), and the tail tube protein gp17.1 (red, 1 of 6 total copies) fitted (PDB entry 5A20). (**b**) Solution structure of the adaptor protein GpW of coliphage λ (cyan, PDB entry 1HYW). (**c**) Solution structure of the adaptor protein GpFII of coliphage λ (magenta, PDB entry 1K0H). An asterisk indicates the loop that in the homologous protein gp16 of phage SPP1 points into the inner channel. (**d**) Phage λ tail terminator protein gpU hexamer (PDB entry 3FZ2) seen from the top (head-binding) side. One of the six monomers is colored orange. (**e**) Phage λ gpV$_N$ tail tube protein domain (PDB entry 2K4Q; top; red) and the immunoglobulin decoration domain gpV$_C$ (PDB entry 2 L04; bottom; orange). Amino- and carboxy-termini are indicated. (**f**) Coliphage T5 tail tube protein gp6, containing a duplicated tail tube domain (PDB entry 5NGJ). The protein is colored in a rainbow spectrum: blue to red from amino- to carboxy-terminus. Tail tube domain 1 (D1) is shown in blue-green, tail tube domain 2 (D2) in green-yellow, and the immunoglobulin-like decoration domain in orange-red. (**g**) Bacteriophage T4 inner tail tube (PDB entry 5V5F). Three stacked hexamers are shown in gray; one monomer is shown in red

hexameric α-helical coiled coil along the length of the tail tube. Bacteriophage T5 has a threefold, rather than sixfold, symmetric tail tube (Arnaud et al. 2017). Its tail tube protein, pb6, is larger than the tail tube proteins of other siphoviruses and contains two β-barrel tail tube domains instead of one, leading to a pseudo-sixfold symmetry (Fig. 27f).

The tail tube structure does not change upon DNA ejection (Arnaud et al. 2017), so it is likely that the signal that a suitable host is encountered is transmitted through the tape measure protein. When the tail tip contacts its receptor on the bacterial membrane, it opens and the tape measure protein leaves first. Its release must then lead to a conformational change in the neck and connector region of the tail (Tavares et al. 2012). Some parts of the tape measure protein have homology with peptidase- and peptidoglycan-degrading domains, suggesting that, once ejected, the tape measure protein may refold to facilitate passage of the DNA through the periplasm (Davidson et al. 2012). The tape measure protein may even form a channel across the periplasmic space, like the podovirus core proteins.

The end of the siphovirus tail is the part of the structure where most structural variation between phages is observed. Here, a tail tip complex or a baseplate is located (Fig. 28a). These complexes recognize the host bacterium and initiate conformational changes that allow successful infection (Davidson et al. 2012). The tail tip complex is a narrow conical tip, while other phages have a more platelike assembly, i.e., a baseplate. Cell attachment is by tail fibers or spike-shaped receptor-binding proteins, which project from the tail tip or baseplate.

The central element of the tail tip is a trimeric hub protein. Each monomer of the trimeric hub protein contains two domains with the same fold as the tail tube protein, a folded-over β-sheet (Fig. 28b), making a pseudo-hexameric ring. The distal tail protein is a hexamer and adapts the hub to the tail. It also contains the folded-over β-sheet and so also leads to a hexameric ring of the same structure (Fig. 28c; Veesler et al. 2010). Coliphages λ and T5 have several trimeric side tail fibers and a central tail fiber (Flayhan et al. 2014). The correct trimerization and folding of the side tail fibers are mediated by chaperone proteins. These chaperone proteins may be intra-molecular, like for T5 (Fig. 28e; Garcia-Doval et al. 2015), or they may be separate proteins, like for phage λ. Interestingly, for phage T5, the intramolecular chaperone has also been shown to shield the receptor-binding site. The side tail fibers are responsible for reversible attachment to a common component of the bacterial cell wall, such as the lipopolysaccharide in case of Gram-negative bacteria. This allows for lateral or two-dimensional diffusion until the central tail fiber encounters its receptor, which is usually a protein to which it binds irreversibly (Garcia-Doval and van Raaij 2013). Bacteriophage SPP1, which infects the Gram-positive *B. subtilis*, lacks side tail fibers, but also binds to a major cell wall component reversibly, in this case to teichoic acids (Baptista et al. 2008). It is not known which protein is responsible for binding to teichoic acids, but its central tail fiber then binds irrevers-ibly to the protein YueB (Vinga et al. 2012).
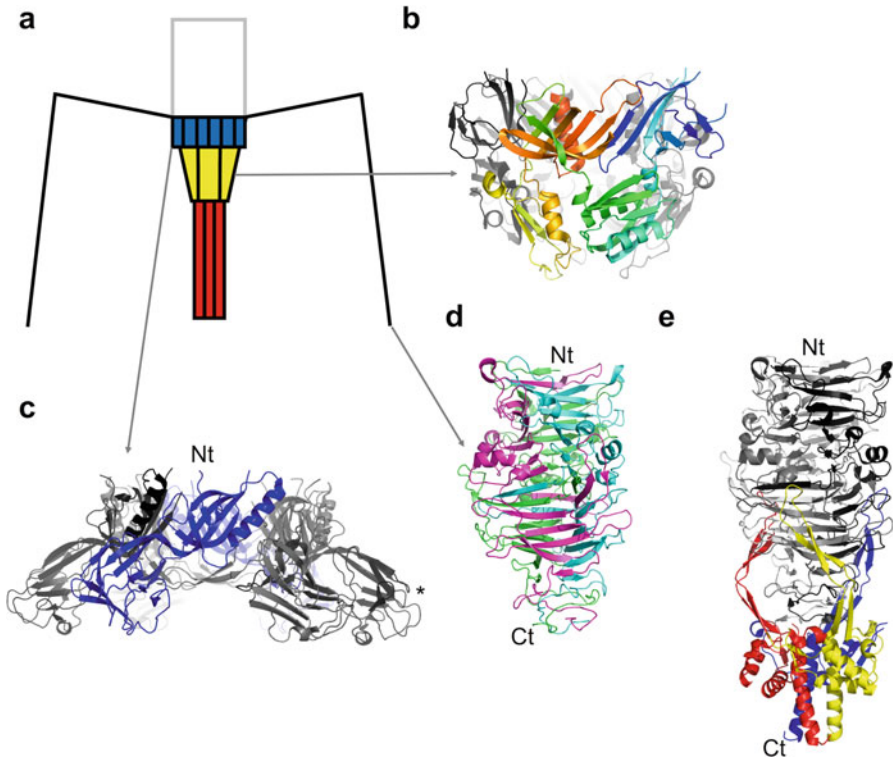
Fig. 28 Structures of siphovirus tail tips and receptor-binding proteins. (**a**) Schematic structure of a siphovirus tail tip with the distal tail protein in blue, the trimeric hub protein in yellow, and the trimeric straight tail fiber in red. The side tail fibers are also shown, and the tail tube outline is shown in gray. (**b**) Trimeric baseplate hub protein of a *Listeria* prophage (PDB entry 3GS9). One monomer is shown in rainbow color, from blue (amino-terminus) to red (carboxy-terminus); the other two monomers are shown in gray and black. Note the trimeric bottom part and pseudo-hexameric top part. (**c**) Distal tail protein hexamer of phage SPP1 (PDB entry 2X8K). One of the monomers in the front is shown in blue; the others are gray. The amino-terminus of the protein is indicated. One of the carboxy-terminal galectin domains is indicated with an asterisk. (**d**) Structure of the distal, carboxy-terminal end of the phage T5 side tail fiber (PDB entry 4UW7). The protein chains are colored green, magenta, and cyan. The amino- and carboxy-termini are indicated. (**e**) Structure of the distal, carboxy-terminal end of the phage T5 side tail fiber with the intramolecular chaperone is still attached (PDB entry 4UW8). Here, the part corresponding to the mature protein is colored black, gray, and white and the intramolecular chaperone red, blue, and yellow. The amino- and carboxy-termini are indicated

It has been proposed that siphoviruses with a narrow tail tip bind a protein receptor with very high affinity, while siphoviruses with a more elaborate baseplate bind saccharidic receptors (Veesler et al. 2010). Examples of siphophages with a structurally studied baseplate are the *Lactococcus* phages TP901-1 and p2 and the *Staphylococcus* phage φ11. The baseplate of phage TP901-1 contains four different proteins (Fig. 29a,b; Veesler et al. 2012b). The center is a hexameric ring of the distal
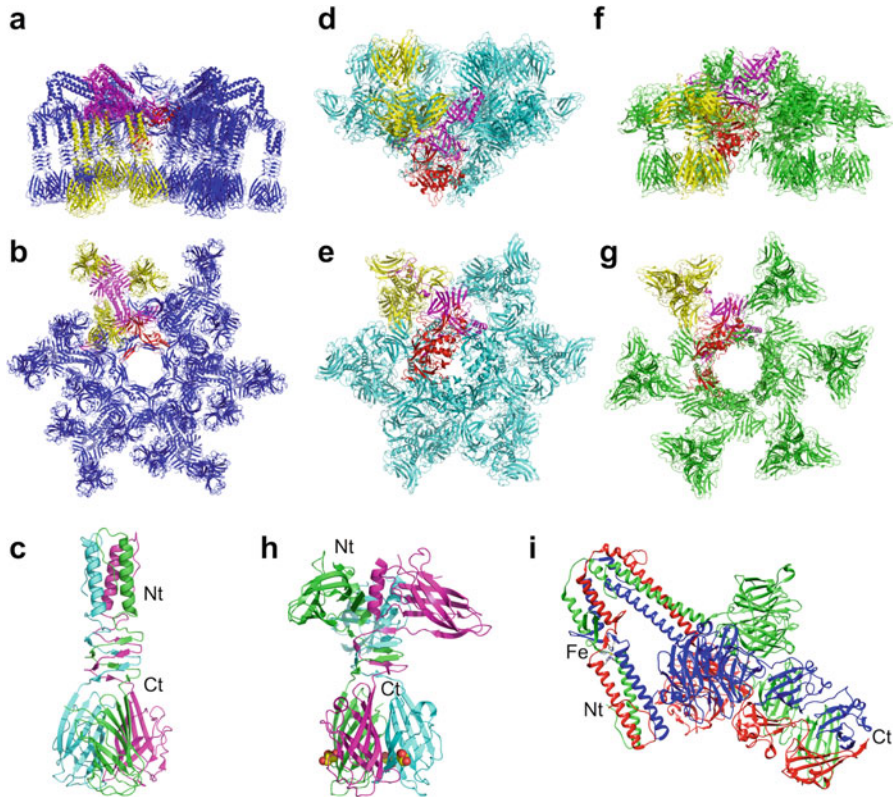
Fig. 29 Structures of siphovirus baseplates and receptor-binding proteins. (**a**) *Lactococcus* phage TP901-1 baseplate seen from the side, colored in blue (PDB entry 4 V96). One of the distal tail proteins of the central ring is colored red, the trimeric arm bound to it is colored magenta, and the three receptor-binding protein trimers bound to the arm are in yellow. (**b**) *Lactococcus* phage TP901-1 baseplate seen from the bottom colored as in part a. (**c**) Phage TP901-1 receptor-binding protein trimer (PDB entry 3EJC). The amino- and carboxy-termini of the magenta chain are indicated. (**d**) Unactivated *Lactococcus* phage p2 baseplate seen from the side, colored in cyan (PDB entry 2WZP). One of the distal tail proteins of the central ring is colored magenta, one of the central hub proteins is colored red, and one of the receptor-binding protein trimers bound is in yellow. (**e**) Unactivated *Lactococcus* phage p2 baseplate seen from the bottom colored as in part d. (**f**) Calcium-activated *Lactococcus* phage p2 baseplate seen from the side, colored in green (PDB entry 2X53). One of the distal tail proteins of the central ring is colored magenta, one of the central hub proteins is colored red, and one of the receptor-binding protein trimers bound is in yellow. (**g**) Calcium-activated *Lactococcus* phage p2 baseplate seen from the bottom colored as in part f. (**h**) *Lactococcus* phage p2 receptor-binding protein trimer with each monomer colored differently (PDB entry 1ZRU). The amino- and carboxy-termini of the green chain are indicated. (**i**) Receptor-binding protein of *Staphylococcus* phage φ11 (PDB entry 5EFV). The amino- and carboxy-termini of the red chain are indicated, as is the location of the iron ion (Fe)

tail protein, which surrounds a central spike (not shown in the figure). The distal tail protein has the same folded-over β-sheet fold as the tail tube, while the structure of the central spike has not been determined. However, it may well be the equivalent of

the trimeric tail tip hub. From the distal tail protein, six trimeric α-helical arms project sideways. The rest of the arm points downward, and each monomer forms an adaptor domain. To each adaptor domain, three trimeric receptor-binding proteins are attached. The amino-terminal end of the protein forms a short α-helical coiled coil, followed by a short β-helical domain and a carboxy-terminal head domain oriented toward the bacterial host (Fig. 29c; Bebeacua et al. 2010), leading to a total of 54 receptor-binding sites. The receptor-binding proteins point downward toward the host bacterium, ready for adhesion. How receptor binding is related to DNA transfer is less clear, perhaps the strong binding with up to 54 receptor molecules pushes the spike against the cell wall, and this force is sensed by the baseplate, which then opens to let the tape measure protein leave, followed by the viral DNA.

The baseplate of phage p2 is composed of three different proteins (Fig. 29d,e; Sciara et al. 2010). The central part of the baseplate is formed by the distal tail protein, which forms a hexameric ring with a central hole. Each monomer in this ring contains the tail tube fold in its N-terminal domain and also has a carboxy-terminal galectin domain. Below this ring is the trimeric hub, with each monomer containing two tail tube domains to adapt to the hexameric ring above. The hub forms a closed dome, blocking passage of the tape measure protein and the phage DNA. From the galectin domain of the distal tail protein, an adapter arm protrudes which interacts with a trimer of the receptor-binding protein. In the unactivated baseplate, the six trimeric receptor-binding proteins point upward, away from the host. Calcium causes a large conformational change in the baseplate, activating it and leading to rotation of receptor-binding domains to point downward (Fig. 29f,g; Sciara et al. 2010). The receptor-binding protein has an amino-terminal β-sandwich shoulder domain, which binds to the adaptor protein. It is followed by a short triple β-helical neck and a carboxy-terminal head domain (Fig. 29h; Spinelli et al. 2006; Tremblay et al. 2006). The structure of the receptor-binding protein was determined bound to glycerol (Fig. 29h), which may be mimicking part of the teichoic acid of the bacterial cell wall. During the conformation change, the monomers of the dome protein separate, allowing passage of the tape measure protein and DNA into the host.

The structure of the receptor-binding protein of the *Staphylococcus* phage φ11 revealed a trimer that can be divided, from the amino- to carboxy-terminus, into stem, platform, and tower domains (Fig. 29i; Li et al. 2016; Koç et al. 2016). The stem is formed by several triple α-helical coiled coils. The first two coiled coils are colinear and interrupted by a region where the three protein chains intertwine, and each chain forms a small β-hairpin. In the center of the intertwined region, an iron ion is located, coordinated by six histidine residues, two from each protein chain. After the second coiled coil, a sharp bend leads into the third coiled coil. The stem is followed by a platform of three β-propellers, and the carboxy-terminal part is a tower formed by two β-prism domains. The β-propeller platform and tower are interconnected by another short triple α-helical coiled coil inside the protein. The β-propeller domain is structurally related to carbohydrate degradation proteins and may thus be involved in receptor binding.

Myoviridae Tail Assembly

As for the *Siphoviridae*, in the *Myoviridae*, tail assembly takes place in parallel to head assembly, starting with the baseplate, i.e., the distal end of the tail. Onto the baseplate, the helical inner tail tube and the outer tail sheath are assembled. The length of the tail is controlled by the tape measure protein. When tail assembly is complete, the tail is joined to the full phage head by the connector complex. Long tail fibers, if present, form separately and are joined to the head-tail assembly afterward. Here, we illustrate the assembly of myovirus tails using as examples the relatively simple coliphage Mu and the more complicated phage T4. The general assembly mechanism of these phages is likely to be extensible to myoviruses infecting Gram-positive bacteria also.

Assembly of the myovirus T4 has been well-studied, and the tail is no exception (Kikuchi and King 1975; Leiman et al. 2010; Arisaka et al. 2016). At the same time, bacteriophage Mu is a relatively simple myovirus and a good general model for
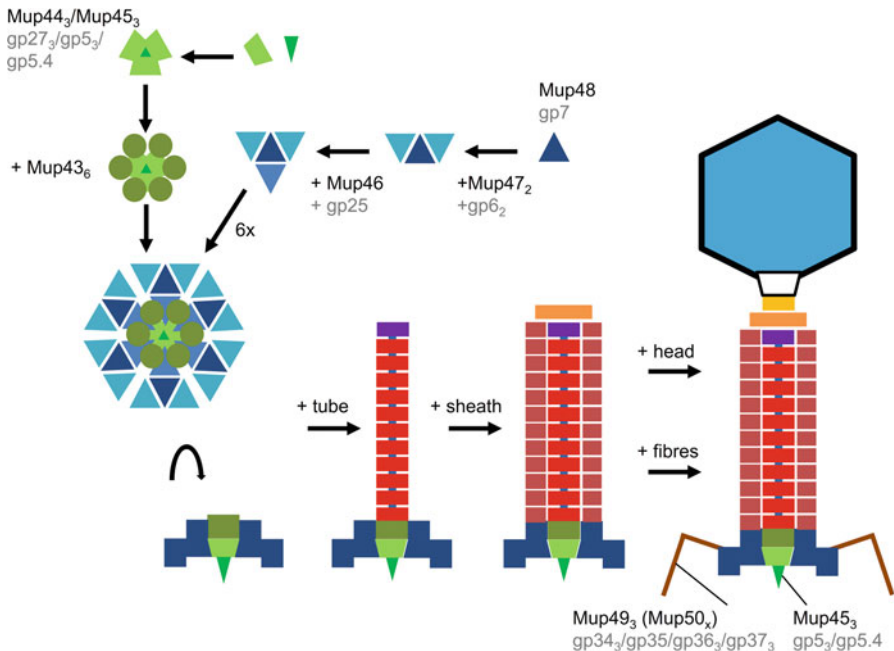

Fig. 30 Myovirus Mu tail assembly. Six wedges, shown in different shades of blue, assemble around a central hub, shown in shades of green, to form a dome-shaped baseplate. Onto the baseplate, first the tail tube assembles in the same way as for the *Siphoviridae*, followed by the contractile tail sheath. Tail completion proteins then bind and the tail joins to the head. Fibers also incorporate. The names of some of the implicated Mu proteins are shown in black lettering, and the corresponding T4 proteins in gray

contractile tail assembly (Büttner et al. 2016). To form the baseplate, wedge proteins are assembled into a wedge complex. The baseplate is then constructed through assembly of wedges around a central hub complex (Fig. 30). A dimer of Mu protein 47 (Mup47) initiates the formation of the wedge by binding to Mup48 (gp7) and subsequently to Mup46. Six wedge complexes assemble around the trimeric central hub complex made up out of Mup44, Mup45, and Mup43, forming the baseplate.

In phage T4, a dimer of gp6 binds to a complex made up of a single copy of gp7, a trimer of gp10, and a dimer of gp8. Six of these wedge complexes come together to form the dome-shaped baseplate, around a hub complex consisting of a trimer of gp27, a trimer of gp5, and a monomer of gp5.4, plus some additional proteins (gp26, gp28, gp51, and gp29). The $(gp5)_3(gp5.4)$ complex is the spike. The six gp6 dimers form a tight ring around the trimer of gp27. The baseplate is completed by the addition of six copies of gp53, six trimers of gp9, and six $(gp11)_3(gp12)_3$ complexes (trimers of gp12 form the short tail fibers; Leiman et al. 2010; Arisaka et al. 2016).

Once the T4 baseplate is complete, a complex consisting of a hexamer of gp48 and a hexamer of gp54 binds to the top of the central hub (Leiman et al. 2010; Arisaka et al. 2016). This complex is the equivalent of Mup43. The $(gp48)_6(gp54)_6$ complex primes the assembly of the inner tail tube, which is assembled in the same way as for the *Siphoviridae*, first assembling a tube consisting of chaperone proteins, which are then replaced by tail tube proteins. In phage T4, the tail tube consists of 138 copies of gp19, forming a six-start helix. The tape measure protein gp29 controls the length of the tail tube, and a hexameric ring of the tail tube terminator protein gp3 stabilizes the completed structure (purple in Fig. 30). Six copies of gp25 then bind to gp53, gp6, gp48, and gp54, to prime the assembly of the tail sheath, which is a six-start helix of gp18. The assembled tail sheath is in a high-energy conformation. A hexamer of gp15 binds to the top of the assembled tail and stabilizes it (orange in Fig. 30). Gp15 then binds to the neck proteins gp13 and gp14 of the capsid, joining the head to the tail.

After the head and tail have joined together, in phage T4, the pre-assembled long tail fibers bind to the phage particle. A homo-trimer of gp34 makes up the proximal half fiber (nearest to the baseplate). The distal half fiber is composed of a trimer of gp36 and a trimer of gp37. For the correct folding of gp34 and gp37, but also for the short tail fiber gp12, the chaperone protein gp57 is necessary. In addition, in T4, gp37 also needs gp38 for correct folding (Bartual et al. 2010a). In other phages, like the *Salmonella* phage S16, gp38 stays bound to gp37 and is the de facto receptor-binding protein. The long tail fiber is completed when the amino-terminal end of the distal half-fiber and carboxy-terminal end of the proximal half-fiber complex are attached through a monomer of gp35 (Hyman and van Raaij 2018). In the case of phage T4, the completed long tail fiber is joined to the tail with the aid of the assembly protein gp63 and the neck fibers (fibritin). The amino-terminal domain of the gp34 homo-trimer binds to the gp9 trimer on the outer ring of the baseplate (Taylor et al. 2016). The long tail fibers are folded up along the tail and capsid, presumably to avoid strong interactions with host membrane fragments and to allow faster diffusion. The neck fibers hold the long tail fibers "up" most of the time, by binding the gp35 knee to the neck (Arisaka et al. 2016). In other myoviruses, like the coliphages Mu and P2, the long tail fibers are simpler and composed of a homo-

trimeric protein, which needs a chaperone protein for correct folding (Haggård-Ljungquist et al. 1992). This chaperone protein may dissociate after assembly, like in T4, or may stay bound to the distal end of the fiber, like in phage λ or *Salmonella* phage S16. Many myoviruses, especially those infecting Gram-positive bacteria, do not have tail fibers, but shorter trimeric receptor-binding proteins instead, as will be seen in the structure section below.

Myoviridae Tail Structure

Myovirus tails show very variable lengths (between about 100 and 4500 nm), but their width is more conserved (between 18 and 24 nm) (Leiman and Shneider 2012). Sometimes, the phage tail is shorter than its head, like in T4 (Yap and Rossmann 2014); for other phages, it is considerably longer. The length of the tail is probably adapted to the host and may be related to the thickness and/or the toughness of the bacterial cell wall. Like for siphovirus tails, the length of the tail is regulated by the tape measure protein. The width of the tails is more conserved, because the folds of the tail tube and tail sheath proteins are conserved. In myoviruses, potential decoration domains are on the outside of the sheath. Here, we will discuss as an example the tail structure of coliphage T4, which is the most studied myovirus tail, and compare it with the *Staphylococcus* phage ϕ812 and the *Listeria* phage A551, as two examples of phages infecting Gram-positive bacteria. The receptor-binding fibers are discussed in a bit more detail.

The tail tube structure of T4 and other myoviruses is basically the same as that of a siphovirus tail and consists of stacked hexameric rings of the tail tube protein gp19 (Zheng et al. 2017; Fig. 28g). In the case of the myoviruses, the tail tube protein is between 15 and 19 kDa, and, unlike for many siphoviruses, it is not decorated, due to the presence of the tail sheath around it. The phage T4 tail tube has 23 hexamers of gp19 and is capped on the top by a hexamer of gp3, which probably has the same fold and structural organization as the phage λ tail tube capping protein gpU (Fig. 27d).

The phage T4 tail sheath is 24 nm wide and 93 nm long. It has 138 copies of gp18, organized in a six-start helix with a pitch of 4.1 nm and a twist of 17° (Fig. 31a, b; Leiman et al. 2004; Kostyuchenko et al. 2005). The contracted T4 sheath is 9 nm wider and 51 nm shorter (Fig. 31c; Arisaka et al. 2016). It is still a six-start helix, but the pitch has decreased to 1.6 nm and the twist has increased to 33°. During contraction, the gp18 subunits move, as rigid bodies, about 5 nm away from the tail center and tilt about 45° (Aksyuk et al. 2009a). Contraction of the sheath assembly is triggered by the baseplate, starts there, and propagates through the sheath in a wavelike motion (Guerrero-Ferreira et al. 2019).

The phage T4 tail sheath is made up of the tail sheath protein gp18. Tail sheath proteins are between 40 and 80 kDa in size, with most around 45 kDa, and have a conserved fold (Leiman and Shneider 2012). The larger tail sheath proteins usually have a decoration domain that is displayed on the outside of the sheath. The T4 gp18 protein has four domains which are inserted into each other like Russian dolls
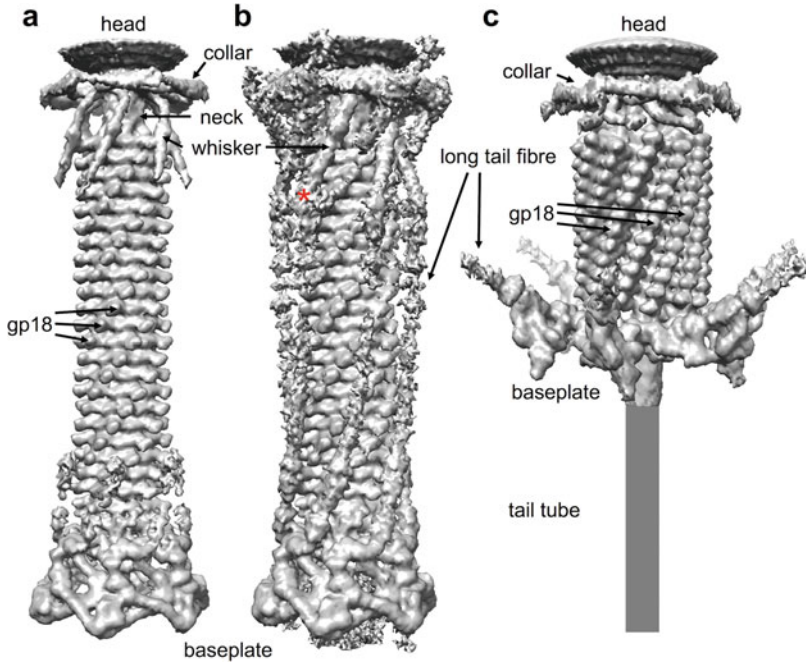
Fig. 31 Overall structure of the myovirus T4 extended and contracted tail. (**a**) Extended tail (EMDB entry EMD-1126). Three gp18 domain IV knobs belonging to the same helical strand are highlighted. (**b**) Extended tail as in panel a but with a reduced contour level to visualize the long tail fibers folded back against the tail. A red asterisk indicates where the carboxy-terminal domain of the fibritin contacts the knee of the long tail fiber. (**c**) Contracted tail (EMDB entry EMD-5528). Three gp18 domain IV knobs belonging to the same helical strand are highlighted, and the protruding part of the tail tube is shown as a gray rectangle

(Fig. 32a; Aksyuk et al. 2009a; Leiman and Shneider 2012): domain II is inserted into a loop of domain I, domain III is inserted into a loop of domain II, and domain IV, which is not present in tail sheath proteins of many other phages, is inserted into a loop of domain III. Although domain I was absent from the gp18 crystal structure, the structure can be inferred from a structurally homologous protein (PDB entry 3HXL). In the structure of domain I, one β-strand is donated by the amino-terminus of a molecule in the next row. Topological considerations suggest that this might also occur in the intact sheath. Furthermore, in the contractile tail sheath of the type VI secretion system, the carboxy-terminal end of the tail sheath homologue inserts into a molecule of the same row (Kudryashev et al. 2015), and this arrangement is likely conserved universally. This fishnet-like organization might be necessary to keep the sheath stable and help it to not fall apart during contraction (Leiman 2018). The conformation of an individual tail sheath protein does not change upon contraction. Domain I maintains the same interactions with neighboring subunits, but domains II and III change partners and actually increase their interaction surface by about four times (Aksyuk et al. 2009a), explaining why the contracted state is more stable.
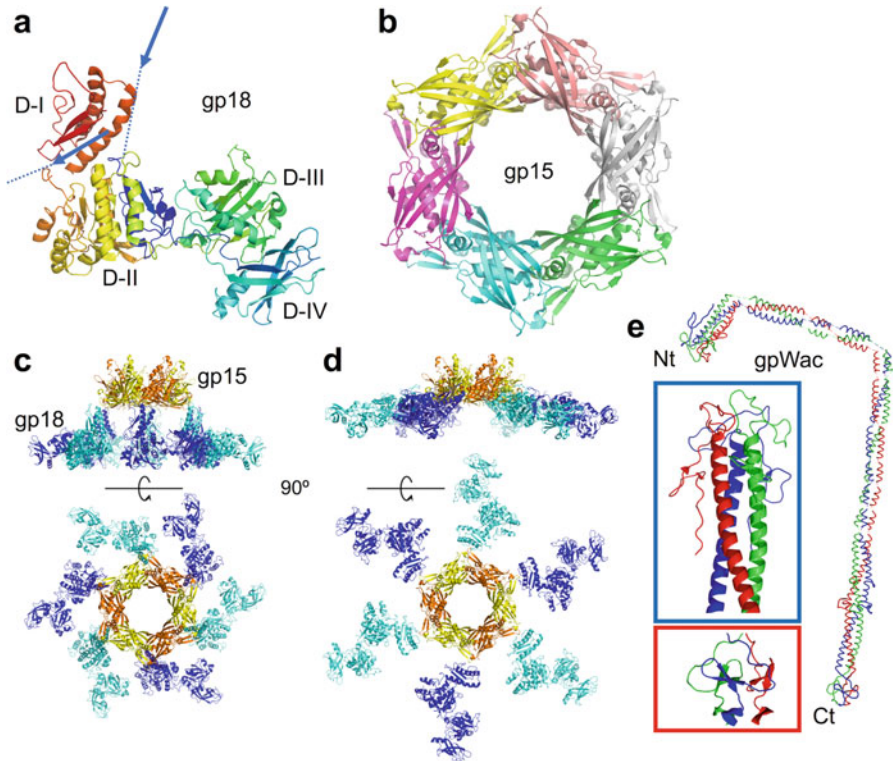
Fig. 32 Structure of the myovirus T4 tail proteins. (**a**) Structure of the tail sheath protein gp18 (PDB entry 3J2O; domains II–IV are from PDB entry 3FOA a domain I is modeled from PDB entry 3HXL). The strand exchange is illustrated: domain I contains a strand (blue arrow) from a neighboring gp18 molecule, while another strand that contributes to domain I of a neighbor on the other side is shown as a blue arrow connected with a dotted line. (**b**) Structure of the gp15 cap seen from the bottom, i.e., the side of the tail tube (PDB entry 4HUD). The six monomers of the hexameric ring are colored differently. (**c, d**) Relative positions of the tail sheath protein gp18 (blue and cyan) and the tail sheath capping protein gp15 (yellow and orange) in the extended (**c**; PDB entry 3J2M) tail and the contracted (**d**; PDB entry 3J2N) tail. (**e**) Model of the fibritin (gpWac) trimer (PDB entry 3J2O) with high-resolution structure of the amino-terminal and carboxy-terminal domains highlighted in blue and red boxes, respectively (PDB entry 1OX3). The three chains of the trimer are colored differently. The amino- and carboxy-termini for the full-length protein model are indicated

Domain I interacts with the tail tube, while domains II and III are partially exposed to solution, and domain IV is on the outside surface of the sheath.

A hexameric ring of the bacteriophage T4 tail terminator protein, gp15, attaches to the top of the phage tail, covering the tail tube capping hexamer gp3 and stabilizing the contractile sheath by interacting with six gp18 molecules and preventing further polymerization of the tail sheath. The hexameric gp15 ring forms the interface for binding the phage head. Each gp15 monomer contains a curled-up eight-stranded antiparallel β-sheet, the center of which faces the inside of the ring

(Fig. 32b). The outside of the ring and the top, where the head-binding site is, are covered with α-helices. From the head side, a dodecamer of the adaptor protein gp13 is attached to the portal and to that a hexamer of gp14. When the head joins the tail, gp14 and gp15 form stable interactions.

Once the phage head is attached to the tail, the neck of T4 virions is decorated by the collar and by whiskers. Twelve trimers of the fibritin protein (gpWac) bind to the neck; six of them are folded sideways around the neck and make up the collar, and another six point downward and form the whiskers (Fokine et al. 2013). Fibritin trimers forming the collar and the whiskers alternate. Each gpWac fiber consists of a long segmented, α-helical triple coiled coil with a small carboxy-terminal trimerization domain (Fig. 32e; Boudko et al. 2002; Boudko et al. 2004). The amino-terminal domain is bound to the phage neck, while the carboxy-terminal domain interacts with the long tail fiber knee (Fig. 31b). The whiskers act as chaperones, helping to attach the long tail fibers to the virus during the assembly
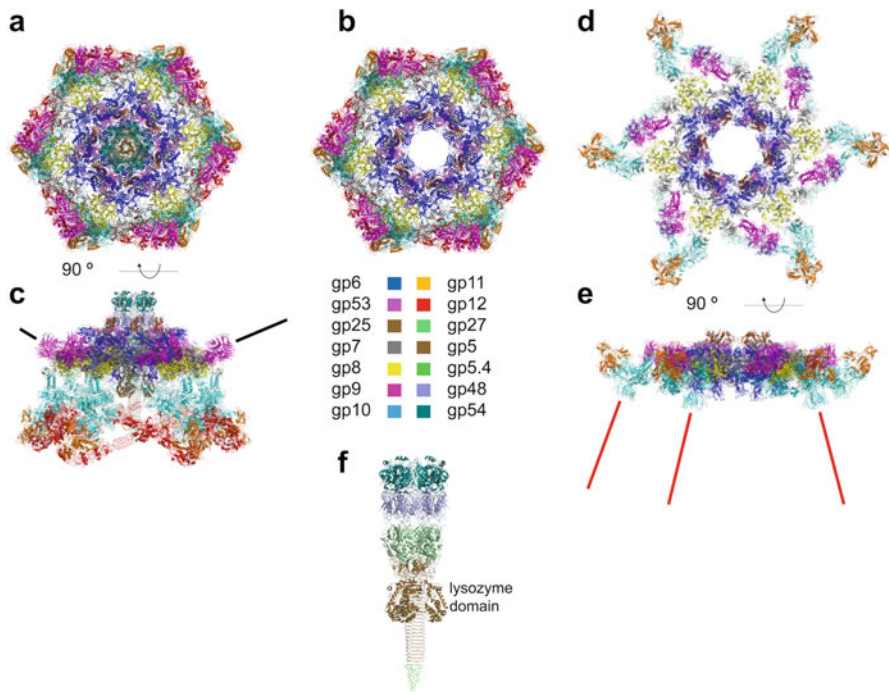

Fig. 33 The baseplate of bacteriophage T4. (**a**) Top view of the pre-attachment dome-shaped baseplate (PDB entry 5IV5). (**b**) Top view of the pre-attachment baseplate without the central tail tube and spike. (**c**) Side view of the pre-attachment dome-shaped baseplate. Black lines indicate the estimated location of the start of the long tail fibers. (**d**) Top view of the post-attachment T4 baseplate (PDB entry 5IV7). (**e**) Side view of the post-attachment T4 baseplate. The estimated locations of the three front-facing short tail fibers are indicated as red lines. (**f**) Side view of the T4 tail tube and its spike. A legend relating protein names to their colors is included

process. The collar and whiskers are also environment-sensing devices, regulating the retraction or deployment of the long tail fibers under unfavorable or favorable conditions, thus preventing or promoting infection, respectively (Arisaka et al. 2016).

The myovirus baseplate is responsible for coordinating correct host recognition with sheath contraction. In bacteriophage T4, the pre-attachment baseplate is dome-shaped (Fig. 33a, b, c) and in a metastable high-energy state. It has overall pseudo-sixfold symmetry with a central trimeric hub. The baseplate central part or hub is surrounded by six wedges, to each of which a receptor-binding long tail fiber is bound. Each wedge also contains a copy of the short tail fiber. At the bottom end of the polymeric gp19 tail tube, a hexameric ring of gp54 is located and below that another hexameric ring of gp48 (Fig. 33f). Both these proteins have the typical folded-over β-sheet tail tube motif. Attached to the gp48 ring is the trimeric gp27 hub. Each of the gp27 monomers has two folded-over β-sheet tail tube domains (Kanamaru et al. 2002), adapting perfectly to the gp48 ring. The central spike is composed of three copies of gp5 and one of gp5.4. Each gp5 monomer has a lysozyme domain on the side, which hydrolyzes the peptidoglycan layer during cell wall penetration. The tip of the tail tube is a triple β-helical tube of gp5, capped by the pointed gp5.4 monomer. This sharp point helps penetrate the membrane (Browning et al. 2012).

The baseplate can be divided into several parts (Taylor et al. 2016). The inner baseplate consists of a ring of 12 gp6 molecules (Aksyuk et al. 2009b). Six of these gp6 molecules are in one conformation and lie more on the inside, forming a constricted iris around the top of the hub and the bottom of the tube. Another six molecules are in a different conformation and alternate with the former, lying more toward the outside of the ring. The amino-terminal halves of the gp6 molecules are part of a $(gp6)_2gp7$ heterotrimeric module, together forming an α-helical core bundle. Gp25 and gp53 also connect the core bundle to the central hub. Gp7 extends outward to the peripheral baseplate and connects with gp9 and gp10. A homotrimer of gp9 is the base to which the long tail fibers are connected (Kostyuchenko et al. 1999).

Gp10 is a trimeric protein with a distorted X-shape. Each of its four domains (D1–D4) has threefold symmetry. D2 and D3 resemble each other. They interact with the amino-terminal regions of gp12 and gp11, respectively, orienting these proteins perpendicular to each other. The short tail fibers are extended trimers of the gp12 protein (van Raaij et al. 2001; Thomassen et al. 2003). Their amino-termini are bound to gp10 (Leiman et al. 2006). In the pre-attachment baseplate, the short tail fibers are bent and their knee region connects to a trimer of gp11, while their carboxy-terminal head domains interact with the amino-terminal part of a gp12 trimer and with gp10 from a neighboring wedge (Leiman et al. 2000). The D1 and D4 domains of gp10 interact with gp7 in the intermediate baseplate, including a disulfide bridge. There is also an intermolecular gp10-gp10 disulfide bridge. These covalent interactions provide extra stability to the baseplate (Taylor et al. 2016). Gp10, gp11, and gp12 share folds, so it is likely they resulted from each other by gene duplication events.

In T4, when at least three long tail fibers have bound a receptor molecule, a signal is transferred to the baseplate of the phage, which then changes conformation. The binding information transfer is likely related to the angle of attachment of the long tail fiber to the baseplate. In the free phage, this angle is variable and the fibers are flexible up to certain limits. When several fibers are attached to their receptor, external forces on the phage may force these angles to values outside this range, pushing proteins near the long fiber attachment site to a different conformation and triggering a sequential conformation change (Taylor et al. 2016). The baseplate flattens and acquires a six-pointed star shape (Fig. 33d, e). This conformational change is very extensive, involving changes of location and interaction partners for several baseplate proteins.

During the transformation, the gp10/gp11/gp12 complex rotates as a unit (Taylor et al. 2016). At the end of this rotation, gp10, and thus the short tail fibers, point straight toward the host cell surface. Gp11 rotates upward, releasing its grip on the knee of the short tail fibers. The interaction of the carboxy-terminal head domain of gp12 with the amino-terminal part of gp12 and with gp10 of the neighboring wedge is also broken. This allows the gp12 short tail fibers to extend fully and reach the cell surface. Only the proximal (amino-terminal) end of the gp12 trimer remains bound to the baseplate, and the distal carboxy-terminal part forms a tight, irreversible interaction with the core region of the host lipopolysaccharide. Gp7 transfers the rotation of the gp10/gp11/gp12 complex to the $(gp6)_2gp7$ heterotrimer. As a result, the diameter of the gp6 iris increases, and gp25 and gp53 dissociate from the tube and the hub. The movement of gp25 away from the tube also initiates sheath contraction. Gp18 subunits from the first ring, which are stacked on gp25, are pushed outward, and the whole sheath contracts as a wave, starting from the baseplate (Guerrero-Ferreira et al. 2019). The widening of the iris and the loosening of gp25 and gp53 allow the passage of a large part of the tail tube through the baseplate and through the bacterial cell wall, driven by sheath contraction. The tailspike, with its lysozyme domains having helped by locally degrading the peptidoglycan, probably falls off during this process and remains in the bacterial periplasm, leaving the end of the tail tube open, so the phage DNA can transfer directly into the host cytoplasm.

Myoviruses infecting Gram-positive hosts tend not to have long side tail fibers. The structure of the *Listeria* phage A511 baseplate is simpler that of phage T4 (Guerrero-Ferreira et al. 2019). It contains the conserved tube-baseplate core complex, consisting of the tube proteins (gp19 and gp54 in T4), the baseplate hub proteins (gp48 and gp27 in T4), the tailspike protein (gp5 in T4), and the four wedge proteins (gp6, gp7, gp25, and gp56 in T4). These proteins are likely to be also conserved in simpler myoviruses that infect Gram-negative bacteria, like the phage P2 and Mu (Guerrero-Ferreira et al. 2019). Simpler baseplates contain only one trimeric protein (the host receptor-binding fiber) that is attached to gp7 (Leiman and Shneider 2012). More complicated phages like T4 developed gp9 and the long tail fibers for reversible host selection on the one hand and irreversible binding via gp10, gp11, and gp12 on the other hand.

The low-resolution structures of the pre-attachment and post-attachment states of the *Staphylococcus* phage ϕ812 have also been determined (Nováček et al. 2016),

and this baseplate is likely to contain the same basic framework also. In both A511 and ϕ812 phages, in the pre-attachment state, the receptor-binding proteins are pointing upward, toward the head of the phage. After DNA ejection, the baseplates transform into a double-layered structure, and the receptor-binding proteins change to a downward orientation, i.e., toward the bacterium, in a similar process as observed for the *Lactococcus* siphovirus p2.

The most variable parts of myovirus baseplates are the receptor-binding proteins, presumably because they need to adapt to evolving host receptor molecules. They are also subject to domain exchange by horizontal gene transfer, which allows phages to change their host range (see chapter ▶ "Bacteriophage-Mediated Horizontal Gene Transfer: Transduction"). Receptor-binding proteins may be stubby proteins such as tailspikes, an example is the P22 gp9-like tailspike in the myovirus Det7 (Walter et al. 2008). As for the other *Caudovirales* members, enzymatic activities may be associated with these proteins, to allow hydrolysis of a bacterial capsule and access to the membrane. In other myoviruses, spindly fibers emanate from the baseplate. Some general features in receptor-binding proteins are conserved within their wide structural variation: they are usually trimeric, are anchored to the virus with their amino-terminal domains, and have carboxy-terminal receptor-binding domains. They are also largely composed of intertwined β-strands, which are likely important for their stability (Mitraki et al. 2006).

For myoviruses with fibers, these may consist of single trimeric proteins, like in phages P2 and Mu (Haggård-Ljungquist et al. 1992). These proteins have their specific chaperones, which are required for correct folding and may remain attached to the distal end of the fiber in mature virions. In coliphage T4 and the *Salmonella* phage S16, the fibers are a complex of four or five different proteins (Fig. 34). In the mature virion, the long tail fibers are retracted upward and anchored to the collar by the knee (Hu et al. 2015). In bacteriophage T4, the structures of the carboxy-terminal parts of gp34 (Granell et al. 2017) and gp37 (Bartual et al. 2010b) are known (Fig. 34). Gp34 contains repeats of a mixed α/β fibrous domain in the amino-terminal two thirds of the protein. The carboxy-terminal third is composed of a triple β-helix domain punctuated by three β-prism domains, the last of which is decorated with long β-hairpins that may be involved in binding gp35. The structure of the receptor-binding tip of gp37 contains an elongated six-stranded antiparallel β-strand needle domain containing seven iron ions coordinated by histidine residues. At the end of the tip, the three chains intertwine to form a small head domain, which contains the putative receptor interaction site. For *Salmonella* phage S16, the crystal structure of the adhesin gp38 attached to the trimeric β-helical tip of gp37 has been determined (Dunne et al. 2018). The monomeric gp38 contains a small α-helical adaptor domain, a β-barrel domain, and a PGII sandwich with three layers of four polyglycine type II helices. The $(gp37)_3gp38$ structure of phage S16 is conserved in other T4-like phages. The iron ion-containing needle motif can also be detected in gp37 of other T4-like phages, sometimes with eight or even nine putative iron ion sites. Other T4-like phages contain receptor-binding domains with yet to be discovered folds. Interestingly, the needle structure of the receptor-binding tip of the T4 gp37 is also conserved in the tip of the siphovirus λ side tail fibers.
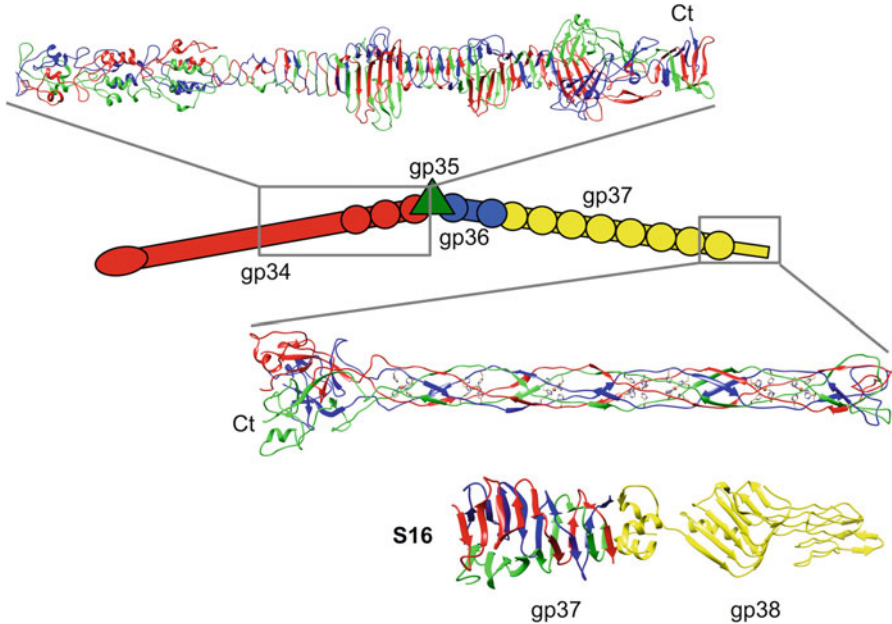
Fig. 34 Long tail fibers. A schematic overview of the long tail fiber of bacteriophage T4 is shown, with gp34 in red, gp35 in green, gp36 in blue, and gp37 in yellow. Gray boxes show parts for which a crystal structure has been determined (PDB entries 4NXH for gp34 and 2XGF for gp37). These structures are shown as ribbon diagrams with their carboxy-termini indicated. At the bottom, the tip of the long tail fibers of the *Salmonella* phage S16 is shown (PDB entry 6F45). This crystal structure contains the C-terminal end of the trimeric gp37 protein bound to a single copy of gp38

Conclusions and Perspectives

It has been proposed that viruses can be divided into a small number of structural lineages (Abrescia et al. 2012). All *Caudovirales*, plus the *Herpesviridae*, have the HK97 fold, which, given the abundance of tailed phages, may well be the most common protein fold on Earth. All *Caudovirales*, the *Podoviridae*, *Siphoviridae*, and *Myoviridae*, must be evolutionary related. They share capsids with the same organization and major capsid and portal protein folds. Adaptor proteins also have similar folds. The *Podoviridae* are structurally the most simple, so perhaps they existed first, and from them, the *Siphoviridae* evolved by acquiring the tail tube, followed by the *Myoviridae* by acquiring the contractile tail sheath. Interestingly, the contractile tail of the *Myoviridae* occurs in bacterial secretion systems and tailocins (Taylor et al. 2018), and it is possible that some bacteria have adopted a phage tail for their own purposes.

The *Tectiviridae* and *Corticoviridae*, but also eukaryotic viruses like the *Adenoviridae*, form the second lineage, characterized by the double jelly roll fold. In these viruses, the jelly rolls are perpendicular to the capsid surface. The *Microviridae*, as well as many eukaryotic RNA viruses, have a capsid protein with a single jelly roll, which lies parallel to the capsid surface. These viruses form the third lineage. A fourth lineage contains the *Cystoviridae* but also the eukaryotic *Reoviridae*. The *Leviviridae* are a distinct lineage, infecting bacteria only. Recent studies, especially using metagenomics, suggest that additional lineages may exist and that the relative abundance of the known lineages may be different than currently assumed. For example, double jelly roll phages may be much more abundant than assumed (Yutin et al. 2018).

Independent of how many different structural lineages and variants of them there are, there will be much more work for structural biologists to do in deciphering the structural diversity of phages, both in their overall structure and the detailed folds of their proteins. For many phage families, details about the assembly and structure are only known for a single or a few members. Future research will show whether the assembly mechanisms are general or whether interesting variations exist. The resolution revolution experienced in cryo-electron microscopy (see chapter ▶ "Detection of Bacteriophages: Electron Microscopy and Visualization") will allow many detailed phage structures to be determined from purified whole phage particles, obviating the need for expressing and purifying all the structural proteins separately. However, crystallography and NMR spectroscopy will remain important for determining detailed structures of flexible phage proteins and their protein-ligand complexes.

New and more precise data on the assembly and the structure of bacteriophages will have important implications for phage applications. Detailed knowledge of their assembly may allow for more efficient production of natural and synthetic phages (see chapter ▶ "Bacteriophage Manufacturing: From Early Twentieth-Century Processes to Current GMP"), as well as the design of phage variants as vaccination vehicles, and for drug delivery for gene therapy. Atomic models of phage structural proteins will allow targeted modification of these natural nanoparticles (see chapter ▶ "Bacteriophages in Nanotechnology: History and Future").

Structures of bacteriophage receptor-binding proteins bound to their receptor or a suitable analogue are relatively rare. This may be because the affinity of individual binding sites is low, and it is difficult to study these complexes in solution or in a crystal. However, bacteriophages, just like other viruses, recognize their host cells with multiple receptor-binding proteins that are often trimeric, and each can bind three receptor molecules simultaneously, leading to a strong avidity effect (Lortat-Jacob et al. 2001). For example, 54 potential receptor-binding sites exist in *Lactococcus* phage TP901-1 (Veesler et al. 2012b). Detailed structural knowledge of the proteins that phages use for receptor recognition and their complexes with receptor analogues may allow the generation of phage mutants with the desired altered host ranges.

Cross-References

References

Abrescia NG, Cockburn JJ, Grimes JM, Sutton GC, Diprose JM, Butcher SJ, Fuller SD, San Martín C, Burnett RM, Stuart DI, Bamford DH, Bamford JKH (2004) Insights into assembly from structural analysis of bacteriophage PRD1. Nature 432:68–74

Abrescia NG, Grimes JM, Kivelä HM, Assenberg R, Sutton GC, Butcher SJ, Bamford JK, Bamford DH, Stuart DI (2008) Insights into virus evolution and membrane biogenesis from the structure of the marine lipid-containing bacteriophage PM2. Mol Cell 31:749–761

Abrescia NG, Bamford DH, Grimes JM, Stuart DI (2012) Structure unifies the viral universe. Annu Rev Biochem 81:795–822

Ackermann HW, Prangishvili D (2012) Prokaryote viruses studied by electron microscopy. Arch Virol 157:1843–1849

Ageno M, Donelli G, Guglielmi F (1973) Structure and physico-chemical properties of bacteriophage G. II, the shape and symmetry of the capsid. Micron 4:376–403

Aksyuk AA, Leiman PG, Kurochkina LP, Shneider MM, Kostyuchenko VA, Mesyanzhinov VV, Rossmann MG (2009a) The tail sheath structure of bacteriophage T4: a molecular machine for infecting bacteria. EMBO J 28:821–829

Aksyuk AA, Leiman PG, Shneider MM, Mesyanzhinov VV, Rossmann MG (2009b) The structure of gene product 6 of bacteriophage T4, the hinge-pin of the baseplate. Structure 17:800–808

Arisaka F, Yap ML, Janamaru S, Rossmann MG (2016) Molecular assembly and structure of the bacteriophage T4 tail. Biophys Rev 8:385–396

Arnaud CA, Effantin G, Vivès C, Engilberge S, Bacia M, Boulanger P, Girard E, Schoehn G, Breyton C (2017) Bacteriophage T5 tail tube structure suggests a trigger mechanism for Siphoviridae DNA ejection. Nat Commun 8:1953

Bamford DH, Palva ET, Lounatmaa K (1976) Ultrastructure and life cycle of the lipid-containing bacteriophage ϕ6. J Gen Virol 32:249–259

Baptista C, Santos MA, São-José C (2008) Phage SPP1 reversible adsorption to Bacillus subtilis cell wall teichoic acids accelerates virus recognition of membrane receptor YueB. J Bacteriol 190:4989–4996

Bartual SG, Garcia-Doval C, Alonso J, Schoehn G, van Raaij MJ (2010a) Two-chaperone assisted soluble expression and purification of the bacteriophage T4 long tail fibre protein gp37. Protein Expr Purif 70:116–121

Bartual SG, Otero JM, Garcia-Doval C, Llamas-Saiz AL, Kahn R, Fox GC, van Raaij MJ (2010b) Structure of the bacteriophage T4 long tail fiber receptor-binding tip. Proc Natl Acad Sci U S A 107:20287–20292

Bebeacua C, Bron P, Lai L, Vegge CS, Brøndsted L, Spinelli S, Campanacci V, Veesler D, van Heel M, Cambillau C (2010) Structure and molecular assignment of lactococcal phage TP901-1 baseplate. J Biol Chem 285:39079–39086

Bernal RA, Hafenstein S, Olson NH, Bowman VD, Chipman PR, Baker TS, Fane BA, Rossmann MG (2003) Structural studies of bacteriophage α3 assembly. J Mol Biol 325:11–24

Bernhardt TG, Wang IN, Struck DK, Young R (2001) A protein antibiotic in the phage Qβ virion: diversity in lysis targets. Science 292:2326–2327

Bhardwaj A, Molineux IJ, Casjens SR, Cingolani G (2011) Atomic structure of bacteriophage Sf6 tail needle knob. J Biol Chem 286:30867–30877

Bhardwaj A, Olia AS, Cingolani G (2014) Architecture of viral genome-delivery molecular machines. Curr Opin Struct Biol 25:1–8

Black LW, Thomas JA (2012) Condensed genome structure. Adv Exp Med Biol 726:469–487

Bollback JP, Huelsenbeck JP (2001) Phylogeny, genome evolution, and host specificity of single-stranded RNA bacteriophage (family *Leviviridae*). J Mol Evol 52:117–128

Boudko SP, Londer YY, Letarov AV, Sernova NV, Engel J, Mesyanzhinov VV (2002) Domain organization, folding and stability of bacteriophage T4 fibritin, a segmented coiled-coil protein. Eur J Biochem 269:833–841

Boudko SP, Strelkov SV, Engel J, Stetefeld J (2004) Design and crystal structure of bacteriophage T4 mini-fibritin NCCF. J Mol Biol 339:927–935

Brewer GJ (1978) Membrane-localized replication of bacteriophage PM2. Virology 84:242–245

Browning C, Shneider MM, Bowman VD, Schwarzer D, Leiman PG (2012) Phage pierces the host cell membrane with the iron-loaded spike. Structure 20:326–339

Butcher SJ, Manole V, Karhu NJ (2012) Lipid-containing viruses: bacteriophage PRD1 assembly. Adv Exp Med Biol 726:365–377

Büttner CR, Wu Y, Maxwell KL, Davidson AR (2016) Baseplate assembly of phage mu: defining the conserved core components of contractile-tailed phages and related bacterial systems. Proc Natl Acad Sci U S A 113:10174–10179

Caldentey J, Bamford DH (1992) The lytic enzyme of the *Pseudomonas* phage ϕ6. Purification and biochemical characterization. Biochim Biophys Acta 1159:44–50

Canelo E, Phillips OM, del Roure RN (1985) Relating cistrons and functions in bacteriophage PM2. Virology 140:364–367

Casjens S, King J (1975) Virus assembly. Annu Rev Biochem 44:555–611

Casjens SR, Molineux IJ (2012) Short noncontractile tail machines: adsorption and DNA delivery by podoviruses. Adv Exp Med Biol 726:143–179

Caspar DL, Klug A (1962) Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol 27:1–24

Chaban Y, Lurz R, Brasiles S, Cornilleau C, Karreman M, Zinn-Justin S, Tavares P, Orlova EV (2015) Structural rearrangements in the phage head-to-tail interface during assembly and infection. Proc Natl Acad Sci U S A 112:7009–7014

Chipman PR, Agbandje-McKenna M, Renaudin J, Baker TS, McKenna R (1998) Structural analysis of the Spiroplasma virus, SpV4: implications for evolutionary variation to obtain host diversity among the *Microviridae*. Structure 6:135–145

Choi KH, McPartland J, Kaganman I, Bowman VD, Rothman-Denes LB, Rossmann MG (2008) Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. J Mol Biol 378:726–736

Cockburn JJ, Abrescia NG, Grimes JM, Sutton GC, Diprose JM, Benevides JM, Thomas GJ, Bamford DH, Bamford JK, Stuart DI (2004) Membrane structure and interactions with protein and DNA in bacteriophage PRD1. Nature 432:122–125

Cuervo A, Pulido-Cid M, Chagoyen M, Arranz R, González-García V, Garcia-Doval C, Castón JR, Valpuesta JM, van Raaij MJ, Martín-Benito K, Carrascosa JL (2013) Structural characterization of the bacteriophage T7 tail. J Biol Chem 288:26290–26299

Dai X, Li Z, Lai M, Shu S, Du Y, Zhou ZH, Sun R (2017) In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. Nature 541:112–116

Davidson AR, Cardarelli L, Pell LG, Radford DR, Maxwell KL (2012) Long noncontractile tail machines of bacteriophages. Adv Exp Med Biol 726:115–142

Doore SM, Fane BA (2016) The *Microviridae*: diversity, assembly, and experimental evolution. Virology 461:45–55

Dunne M, Denyes JM, Arndt H, Loessner MJ, Leiman PG, Klumpp J (2018) *Salmonella* phage S16 tail fiber adhesin features a rare polyglycine rich domain for host recognition. Structure 26:1573–1582

El Omari K, Meier C, Kainov D, Sutton G, Grimes JM, Poranen MM, Bamford DH, Tuma R, Stuart DI, Mancini EJ (2013a) Tracking in atomic detail the functional specializations in viral RecA helicases that occur during evolution. Nucleic Acids Res 41:9396–9410

El Omari K, Sutton G, Ravantti JJ, Zhang HW, Walter TS, Grimes JM, Bamford DH, Stuart DI, Mancini EJ (2013b) Plate tectonics of virus shell assembly. Structure 21:1384–1395

Espejo RT, Canelo ES, Sinsheimer RL (1969) DNA of bacteriophage PM2: a closed circular double-stranded molecule. Proc Natl Acad Sci U S A 63:1164–1168

Farley MM, Tu J, Kearns DB, Molineaux IJ, Liu J (2017) Ultrastructural analysis of bacteriophage ϕ29 during infection of Bacillus subtilis. J Struct Biol 197:163–171

Feng JN, Model P, Russel M (1999) A trans-envelope protein complex needed for filamentous phage assembly and export. Mol Microbiol 34:745–755

Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, Van den Berghe A, Volckaert G, Ysebaert M (1976) Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature 260:500–507

Flayhan A, Vellieux FM, Lurz R, Maury O, Contreras-Martel C, Girard E, Boulanger P, Breyton C (2014) Crystal structure of pb9, the distal tail protein of bacteriophage T5: a conserved structural motif among all siphophages. J Virol 88:820–828

Fokine A, Rossmann MG (2014) Molecular architecture of tailed double-stranded DNA phages. Bacteriophage 4:e28281

Fokine A, Chipman PR, Leiman PG, Mesyanzhinov VV, Rao VB, Rossmann MG (2004) Molecular architecture of the prolate head of bacteriophage T4. Proc Natl Acad Sci 101:6003–6008

Fokine A, Islam MZ, Zhang Z, Bowman VD, Rao VB, Rossmann MG (2011) Structure of the three N-terminal immunoglobulin domains of the highly immunogenic outer capsid protein from a T4-like bacteriophage. J Virol 85:8141–8148

Fokine A, Zhang Z, Kanamaru S, Bowman VD, Aksyuk AA, Arisaka F, Rao VB, Rossmann MG (2013) The molecular architecture of the bacteriophage T4 neck. J Mol Biol 425:1731–1744

Frilander M, Bamford DH (1995) In vitro packaging of the single-stranded RNA genomic precursors of the segmented double-stranded RNA bacteriophage ϕ6: the three segments modulate each other's packaging efficiency. J Mol Biol 246:418–428

Garcia-Doval C, van Raaij MJ (2012) Structure of the receptor-binding carboxy-terminal domain of bacteriophage T7 tail fibers. Proc Natl Acad Sci U S A 109:9390–9395

Garcia-Doval C, van Raaij MJ (2013) Bacteriophage receptor recognition and nucleic acid transfer. Subcell Biochem 68:489–518

Garcia-Doval C, Castón JR, Luque D, Granell M, Otero JM, Llamas-Saiz AL, Renouard M, Boulanger P, van Raaij MJ (2015) Structure of the receptor-binding carboxy-terminal domain of the bacteriophage T5 L-shaped tail fibre with and without its intra-molecular chaperone. Viruses 7:6424–6440

Golmohammadi R, Valegard K, Fridborg K, Liljas L (1993) The refined structure of bacteriophage MS2 at 2.8 Å resolution. J Mol Biol 234:620–639

Golmohammadi R, Fridborg K, Bundule M, Valegard K, Liljas L (1996) The crystal structure of bacteriophage Qβ at 3.5 Å resolution. Structure 4:543–554

González-García VA, Pulido-Cid M, Garcia-Doval C, Bocanegra R, van Raaij MJ, Martin-Benito J, Cuervo A, Carrascosa JL (2015) Conformational changes leading to T7 DNA delivery upon interaction with the bacterial receptor. J Biol Chem 290:10038–10044

Gorzelnik KV, Cui Z, Reed CA, Jakana J, Young R, Zhang J (2016) Asymmetric cryo-EM structure of the canonical *Allolevivirus* Qβ reveals a single maturation protein and the genomic ssRNA in situ. Proc Natl Acad Sci U S A 113:11519–11524

Granell M, Namura M, Alvira S, Kanamaru S, van Raaij MJ (2017) Crystal structure of the carboxy-terminal region of the bacteriophage T4 proximal long tail fiber protein gp34. Viruses 9:E168

Guerrero-Ferreira RC, Hupfeld M, Nazarov S, Taylor NM, Shneider MM, Obbineni JM, Loessner MJ, Ishikawa T, Klumpp J, Leiman PG (2019) Structure and transformation of bacteriophage A511 baseplate and tail upon infection of *Listeria* cells. EMBO J 38:e99455

Haggård-Ljungquist E, Halling C, Calendar R (1992) DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. J Bacteriol 174:1462–1477

Henderson R (2015) Overview and future of single particle electron cryomicroscopy. Arch Biochem Biophys 581:19–24

Higashitani A, Higashitani N, Horiuchi K (1997) Minus-strand origin of filamentous phage versus transcriptional promoters in recognition of RNA polymerase. Proc Natl Acad Sci U S A 94:2909–2914

Higman VA (2013) Nuclear magnetic resonance methods for studying soluble, fibrous, and membrane-embedded proteins. In: Rusa JM, Piñeiro A (eds) Proteins in solution and at interfaces: methods and applications in biotechnology and materials science. Wiley, New York, pp 23–48

Holliger P, Riechmann L, Williams RL (1999) Crystal structure of the two N-terminal domains of g3p from filamentous phage fd at 1.9 Å: evidence for conformational lability. J Mol Biol 288:649–657

Hong C, Oksanen HM, Liu X, Jakana J, Bamford DH, Chiu W (2014) A structural model of the genome packaging process in a membrane-containing double stranded DNA virus. PLoS Biol 12:e1002024

Hu GB, Wei H, Rice WJ, Stokes DL, Paul Gottlieb P (2008) Electron cryo-tomographic structure of cystovirus φ12. Virology 372:1–9

Hu B, Margolin W, Molineaux IJ, Liu J (2013) The bacteriophage T7 virion undergoes extensive structural remodeling during infection. Science 339:577–579

Hu B, Margolin W, Molineux IJ, Liu J (2015) Structural remodelling of bacteriophage T4 and host membranes during infection initiation. Proc Natl Acad Sci U S A 112:E4919–E4928

Hua J, Huet A, Lopez CA, Toropova K, Pope WH, Duda RL, Hendrix RW, Conway JF (2017) Capsids and genomes of jumbo-sized bacteriophages reveal the evolutionary reach of the HK97 fold. MBio 8:e01579-01517

Huiskonen JT, Kivelä HM, Bamford DH, Butcher SJ (2004) The PM2 virion has a novel organization with an internal membrane and pentameric receptor binding spikes. Nat Struct Mol Biol 11:850–856

Huiskonen JT, Manole V, Butcher SJ (2007) Tale of two spikes in bacteriophage PRD1. Proc Natl Acad Sci U S A 104:6666–6671

Hyman P, van Raaij MJ (2018) Bacteriophage T4 long tail fiber domains. Biophys Rev 10:463–471

Inagaki M, Kawaura T, Wakashima H, Kato M, Nishikawa S, Kashimura N (2003) Different contributions of the outer and inner R-core residues of lipopolysaccharide to the recognition by spike H and G proteins of bacteriophage ϕX174. FEMS Microbiol Lett 226:221–227

Jäälinoja HT, Huiskonen JT, Butcher SJ (2007) Electron cryomicroscopy comparison of the architectures of the enveloped bacteriophages ϕ6 and ϕ8. Structure 15:157–167

Johnson JE, Speir JA (1997) Quasi-equivalent viruses: a paradigm for protein assemblies. J Mol Biol 269:665–675

Kanamaru S, Leiman PG, Kostyuchenko VA, Chipman PR, Mesyanzhinov VV, Arisaka F, Rossmann MG (2002) Structure of the cell-puncturing device of bacteriophage T4. Nature 415:553–557

Karlsson F, Borrebaeck CA, Nilsson N, Malmborg-Hager AC (2003) The mechanism of bacterial infection by filamentous phages involves molecular interactions between TolA and phage protein 3 domains. J Bacteriol 185:2628–2634

Katsura I (1990) Mechanism of length determination in bacteriophage lambda tails. Adv Biophys 26:1–18

Kikuchi Y, King J (1975) Assembly of the tail of bacteriophage T4. J Supramol Struct 3:24–38

Kivelä HM, Männistö RH, Kalkkinen N, Bamford DH (1999) Purification and protein composition of PM2, the first lipid-containing bacterial virus to be isolated. Virology 262:364–374

Kivelä HM, Kalkkinen N, Bamford DH (2002) Bacteriophage PM2 has a protein capsid surrounding a spherical proteinaceous lipid core. J Virol 76:8169–8178

Kivelä HM, Daugelavičius R, Hankkio RH, Bamford JKH, Bamford DH (2004) Penetration of membrane-containing double-stranded-DNA bacteriophage PM2 into *Pseudoalteromonas* hosts. J Bacteriol 186:5342–5354

Kivelä HM, Abrescia NGA, Bamford JKH, Grimes JM, Stuart DI, Bamford DH (2008) Selenomethionine labeling of large biological macromolecular complexes: probing the structure of marine bacterial virus PM2. J Struct Biol 161:204–210

Koç C, Xia G, Kühner P, Spinelli S, Roussel A, Cambillau C, Stehle T (2016) Structure of the host-recognition device of *Staphylococcus* aureus phage ϕ11. Sci Rep 6:27581

Korasick DA, Tanner JJ (2018) Determination of protein oligomeric structure from small-angle X-ray scattering. Protein Sci 27:814–824

Kostyuchenko VA, Navruzbekov GA, Kurochkina LP, Strelkov SV, Mesyanzhinov VV, Rossmann MG (1999) The structure of bacteriophage T4 gene product 9: the trigger for tail contraction. Structure 7:1213–1222

Kostyuchenko VA, Chipman PR, Leiman PG, Arisaka F, Mesyanzhinov VV, Rossmann MG (2005) The tail structure of bacteriophage T4 and its mechanism of contraction. Nat Struct Mol Biol 12:810–813

Krupovic M, Daugelavicius R, Bamford DH (2007) A novel lysis system in PM2, a lipid-containing marine double-stranded DNA bacteriophage. Mol Microbiol 64:1635–1648

Kudryashev M, Wang RY, Brackmann M, Scherer S, Maier T, Baker D, DiMaio F, Stahlberg H, Egelman EH, Basler M (2015) Structure of the type VI secretion system contractile sheath. Cell 160:952–962

Kühlbrandt W (2014) Cryo-EM enters a new era. elife 3:e03678

Lavigne R, Darius P, Summer EJ, Seto D, Mahdevan P, Nilson AS, Ackermann KAM (2009) Classification of *Myoviridae* bacteriophages using protein sequence similarity. BMC Microbiol 9:224

Leiman PG (2018) Stretching the arms of the type VI secretion sheath protein. EMBO Rep 19:191–193

Leiman PG, Shneider MM (2012) Contractile tail machines of bacteriophages. Adv Exp Med Biol 726:93–114

Leiman PG, Kostyuchenko VA, Shneider MM, Kurochkina LP, Mesyanzhinov VV, Rossmann MG (2000) Structure of bacteriophage T4 gene product 11, the interface between the baseplate and short tail fibers. J Mol Biol 301:975–985

Leiman PG, Chipman PR, Kostyuchenko VA, Mesyanzhinov VV, Rossmann MG (2004) Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. Cell 118:419–429

Leiman PG, Shneider MM, Mesyanzhinov VV, Rossmann MG (2006) Evolution of bacteriophage tails: structure of T4 gene product 10. J Mol Biol 358:912–921

Leiman PG, Battisti AJ, Bowman VD, Stumeyer K, Mühlenhoff M, Gerardy-Schahn R, Scholl D, Molineaux IJ (2007) The structures of bacteriophages K1E and K1-5 explain processive degradation of polysaccharide capsules and evolution of new host specificities. J Mol Biol 371:836–849

Leiman PG, Arisaka F, van Raaij MJ, Kostyuchenko VA, Aksyuk AA, Kanamaru S, Rossmann MG (2010) Morphogenesis of the T4 tail and tail fibers. Virol J 7:355

Lhuillier S, Gallopin M, Gilquin B, Brasilès S, Lancelot N, Letellier G, Gilles M, Dethan G, Orlova EV, Couprie J, Tavares P, Zinn-Justin S (2009) Structure of bacteriophage SPP1 head-to-tail connection reveals mechanism for viral DNA gating. Proc Natl Acad Sci U S A 106:8507–8512

Li X, Koç C, Kühner P, Stierhof YD, Krismer B, Enright MC, Penadés JR, Wolz C, Stehle T, Cambillau C, Peschel A, Xia G (2016) An essential role for the baseplate protein gp45 in phage adsorption to *Staphylococcus* aureus. Sci Rep 6:26455

Liu Y, Eisenberg D (2002) 3D domain swapping: as domains continue to swap. Protein Sci 11:1285–1299

Liu X, Zhang Q, Murata K, Baker ML, Sullivan MB, Fu C, Dougherty MT, Schmid MF, Osburne MS, Chisholm SW, Chiu W (2010) Structural changes in a marine podovirus associated with release of its genome into *Prochlorococcus*. Nat Struct Mol Biol 17:830–836

Llamas-Saiz AL, van Raaij MJ (2013) X-ray crystallography of biological macromolecules: fundamentals and applications. In: Rusa JM, Piñeiro A (eds) Proteins in solution and at interfaces: methods and applications in biotechnology and materials science. Wiley, New York, pp 3–22

Lortat-Jacob H, Chouin E, Cusack S, van Raaij MJ (2001) Kinetic analysis of adenovirus fiber binding to its receptor reveals an avidity mechanism for trimeric receptor-ligand interactions. J Biol Chem 276:9009–9015

Lubkowski J, Hennecke F, Plückthun A, Wlodawer A (1999) Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. Structure 7:711–722

Mahony J, Stockdale SR, Collins B, Spinelli S, Douillard FP, Cambillau C, van Sinderen D (2016) Lactococcus lactis phage TP901-1 as a model for *Siphoviridae* virion assembly. Bacteriophage 6:e1123795

Männistö RH, Kivelä HM, Paulin L, Bamford DH, Bamford JK (1999) The complete genome sequence of PM2, the first lipid-containing bacterial virus to be isolated. Virology 262:355–336

Mäntynen S, Sundberg LR, Poranen MM (2018) Recognition of six additional cystoviruses: *Pseudomonas virus phi6* is no longer the sole species of the family *Cystoviridae*. Arch Virol 163:1117–1124

Marvin DA (1998) Filamentous phage structure, infection and assembly. Curr Opin Struct Biol 8:150–158

Marvin DA (2017) Fibre diffraction studies of biological macromolecules. Prog Biophys Mol Biol 127:43–87

Marvin DA, Welsh LC, Symmons MF, Scott WR, Straus SK (2006) Molecular structure of fd (f1, M13) filamentous bacteriophage refined with respect to X-ray fibre diffraction and solid-state NMR data supports specific models of phage assembly at the bacterial membrane. J Mol Biol 355:294–309

Marvin DA, Symmons MF, Straus SK (2014) Structure and assembly of filamentous bacteriophages. Prog Biophys Mol Biol 114:80–122

Maxwell KL, Yee AA, Booth V, Arrowsmith CH, Gold M, Davidson AR (2001) The solution structure of bacteriophage λ protein W, a small morphogenetic protein possessing a novel fold. J Mol Biol 308:9–14

Maxwell KL, Yee AA, Arrowsmith CH, Gold M, Davidson AR (2002) The solution structure of the bacteriophage λ head-tail joining protein, gpFII. J Mol Biol 318:1395–1404

Merckel MC, Huiskonen JT, Bamford DH, Goldman A, Tuma R (2005) The structure of the bacteriophage PRD1 spike sheds light on the evolution of viral capsid architecture. Mol Cell 18:161–170

Mitraki A, Papanikolopoulou K, van Raaij MJ (2006) Natural triple β-stranded fibrous folds. Adv Protein Chem 73:97–124

Moak M, Molineux IJ (2004) Peptidoglycan hydrolytic activities associated with bacteriophage virions. Mol Microbiol 51:1169–1183

Morais MC, Choi KH, Koti JS, Chipman PR, Anderson DL, Rossmann MG (2005) Conservation of the capsid structure in tailed dsDNA bacteriophages: the pseudoatomic structure of ϕ29. Mol Cell 18:149–159

Moreno-Madrid F, Martín-González N, Llauró A, Ortega-Esteban A, Hernando-Pérez M, Douglas T, Schaap IA, de Pablo PJ (2017) Atomic force microscopy of virus shells. Biochem Soc Trans 45:499–511

Nemecek D, Gilcrease EB, Kang S, Prevelige PE, Casjens S, Thomas GJ (2007) Subunit conformations and assembly states of a DNA-translocating motor: the terminase of bacteriophage P22. J Mol Biol 374:817–836

Nemecek D, Boura E, Wu W, Cheng N, Plevka P, Qiao J, Mindich L, Heymann JB, Hurley JH, Steven AC (2013) Subunit folds and maturation pathway of a dsRNA virus capsid. Structure 21:1374–1383

Ni CZ, Syed R, Kodandapani R, Wickersham J, Peabody DS, Ely KR (1995) Crystal structure of the MS2 coat protein dimer: implications for RNA binding and virus assembly. Structure 3:255–263

Nováček J, Šiborová S, Benešík M, Pantůček R, Doškař J, Plevka P (2016) Structure and genome release of Twort-like *Myoviridae* phage with a double-layered baseplate. Proc Natl Acad Sci U S A 113:9351–9356

Olia AS, Casjens S, Cingolani G (2007) Structure of phage P22 cell envelope–penetrating needle. Nat Struct Mol Biol 14:1221–1226

Olia AS, Prevelige PE, Johnson JE, Cingolani G (2011) Three-dimensional structure of a viral genome-delivery portal vertex. Nat Struct Mol Biol 18:597–603

Oliveira LM, Ye Z, Katz A, Alimova A, Wei H, Herman GT, Gottlieb P (2018) Component tree analysis of cystovirus ϕ6 nucleocapsid cryo-EM single particle reconstructions. PLoS One 13: e0188858

Parent KN, Schrad JR, Cingolani G (2018) Breaking symmetry in viral icosahedral capsids as seen through the lenses of X-ray crystallography and cryo-electron microscopy. Viruses 10:67

Pell LG, Liu A, Edmonds L, Donaldson LW, Howell PL, Davidson AR (2009a) The X-ray crystal structure of the phage λ tail terminator protein reveals the biologically relevant hexameric ring structure and demonstrates a conserved mechanism of tail termination among diverse long-tailed phages. J Mol Biol 389:938–951

Pell LG, Kanelis V, Donaldson LW, Howell PL, Davidson AR (2009b) The phage λ major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. Proc Natl Acad Sci U S A 106:4160–4165

Pell LG, Gasmi-Seabrook GM, Morais M, Neudecker P, Kanelis V, Bona D, Donaldson LW, Edwards AM, Howell PL, Davidson AR, Maxwell KL (2010 Oct 29) The solution structure of the C-terminal Ig-like domain of the bacteriophage λ tail tube protein. J Mol Biol 403 (3):468–479

Perucchetti R, Parris W, Becker A, Gold M (1988) Late stages in bacteriophage λ head morphogenesis: in vitro studies on the action of the bacteriophage λ D-gene and W-gene products. Virology 165:103–114

Petrovski S, Dyson ZA, Seviour RJ, Tillet D (2012) Small but sufficient: the *Rhodococcus* phage RRH1 has the smallest known *Siphoviridae* genome at 14.2 kilobases. J Virol 86:358–363

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera, a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612

Philippe C, Krupovic M, Jaomanjaka F, Claisse O, Petrel M, le Marrec M (2018) Bacteriophage GC1, a novel Tectivirus infecting *Gluconobacter cerinus*, an acetic acid bacterium associated with wine-making. Viruses 10:39

Plevka P, Tars K, Liljas L (2008) Crystal packing of a bacteriophage MS2 coat protein mutant corresponds to octahedral particles. Protein Sci 17:1731–1739

Poranen MM, Tuma R (2004) Self-assembly of double-stranded RNA bacteriophages. Virus Res 101:93–100

Prasad BV, Schmid MF (2012) Principles of virus structural organization. Adv Exp Med Biol 726:17–47

Prevelige PE, Cortines JR (2018) Phage assembly and the special role of the portal protein. Curr Opin Virol 31:66–73

Prevelige PE, Fane BA (2012) Building the machines: scaffolding proteins functions during bacteriophage morphogenesis. Adv Exp Med Biol 726:325–350

Rakonjac J, Bennett NJ, Spagnuolo J, Gagic D, Russel M (2011) Filamentous bacteriophage: biology, phage display and nanotechnology applications. Curr Issues Mol Biol 13:51–57

Rakonjac J, Russel M, Khanum S, Brooke SJ, Rajič M (2017) Filamentous phage: structure and biology. Adv Exp Med Biol 1053:1–20

Rao VB, Feiss M (2015) Mechanisms of DNA packaging by large double-stranded DNA viruses. Annu Rev Virol 9:351–378

Ross PD, Cheng N, Conway JF, Firek BA, Hendrix RW, Duda RL, Steven AC (2005) Crosslinking renders bacteriophage HK97 capsid maturation irreversible and effects an essential stabilization. EMBO J 24:1352–1363

Rumnieks J, Tars K (2014) Crystal structure of the bacteriophage Qβ coat protein in complex with the RNA operator of the replicase gene. J Mol Biol 426:1039–1049

Russel M, Linderoth NA, Sali A (1997) Filamentous phage assembly: variation on a protein export theme. Gene 192:23–32

Santos-Pérez I, Oksanen HM, Bamford DH, Goñi FM, Reguera D, Abrescia NGA (2017) Membrane-assisted viral DNA ejection. Biochim Biophys Acta 1861:664–672

Schulz EC, Dickmanns A, Urlaub H, Schmitt A, Mühlenhoff M, Stummeyer K, Schwarzer D, Gerardy-Schahn R, Ficner R (2010a) Crystal structure of an intramolecular chaperone mediating triple-β-helix folding. Nat Struct Mol Biol 17:210–215

Schulz EC, Schwarzer D, Frank M, Stummeyer K, Mühlenhoff M, Dickmanns A, Gerardy-Schahn-R, Ficner R (2010b) Structural basis for the recognition and cleavage of polysialic acid by the bacteriophage K1F tailspike protein EndoNF. J Mol Biol 397:341–351

Sciara G, Bebeacua C, Bron P, Tremblay D, Ortiz-Lombardia M, Lichière J, van Heel M, Campanacci V, Moineau S, Cambillau C (2010) Structure of lactococcal phage p2 baseplate and its mechanism of activation. Proc Natl Acad Sci U S A 107:6852–6857

Seul A, Müller JJ, Andres D, Stettner E, Heinemann U, Seckler R (2014) Bacteriophage P22 tailspike: structure of the complete protein and function of the interdomain linker. Acta Cryst D70:1336–1345

Spinelli S, Desmyter A, Verrips CT, de Haard HJ, Moineau S (2006) Cambillau C (2006) Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. Nat Struct Mol Biol 13:85–89

Stockley PG, White SJ, Dykeman E, Manfield I, Rolfsson O, Patel N, Bingham R, Barker A, Wroblewski E, Chandler-Bostock R, Weiss EU, Ranson NA, Tuma R, Twarock R (2016) Bacteriophage MS2 genomic RNA encodes an assembly instruction manual for its capsid. Bacteriophage 6:e1157666

Su S, Gao YG, Zhang H, Terwilliger TC, Wang AH (1997) Analyses of the stability and function of three surface mutants (R82C, K69H, and L32R) of the gene V protein from Ff phage by X-ray crystallography. Protein Sci 6:771–780

Suhanovsky MM, Teschke CM (2015) Nature's favorite building block: deciphering folding and capsid assembly of proteins with the HK97-fold. Virology 479-480:487–497

Sun L, Young LN, Zhang X, Boudko SP, Fokine A, Zbornik E, Roznowski AP, Molineux IJ, Rossmann MG, Bentley A, Fane BA (2014) Icosahedral bacteriophage φX174 forms a tail for DNA transport during infection. Nature 505:432–435

Sun Z, El Omari K, Sun X, Ilca SL, Kotecha A, Stuart DI, Poranen MM, Huiskonen JT (2017) Double-stranded RNA virus outer shell assembly by bona fide domain-swapping. Nat Commun 8:14814

Takata T, Haase-Pettingell C, King J (2012) The C-terminal cysteine annulus participates in auto-chaperone function for Salmonella phage P22 tailspike folding and assembly. Bacteriophage 2:36–49

Tang J, Olson N, Jardine PJ, Grimes S, Anderson DL, Baker TS (2008) DNA poised for release in bacteriophage φ29. Structure 16:935–943

Tang J, Lander GC, Olia A, Li R, Casjens S, Prevelige P, Cingolani G, Baker TS, Johnson JE (2011) Peering down the barrel of a bacteriophage portal: the genome packaging and release valve in P22. Structure 19:496–502

Tao P, Mahalingam M, Zhu J, Moayeri M, Sha J, Lawrence WS, Leppla SH, Chopra AK, Rao VB (2018) A bacteriophage T4 nanoparticle-based dual vaccine against anthrax and plague. MBio 9:e01926-18

Tavares P, Zinn-Justin S, Orlova EV (2012) Genome gating in tailed bacteriophage capsids. Adv Exp Med Biol 726:585–600

Taylor GL (2010) Introduction to phasing. Acta Cryst D66:325–338

Taylor NM, Prokhorov NS, Guerrero-Ferreira RC, Shneider MM, Browning C, Goldie KN, Stahlberg H, Leiman PG (2016) Structure of the T4 baseplate and its function in triggering sheath contraction. Nature 533:346–352

Taylor NMI, van Raaij MJ, Leiman PG (2018) Contractile injection systems of bacteriophages and related systems. Mol Microbiol 108:6–15

Thiriot DS, Nevzorov AA, Zagyanskiy L, Wu CH, Opella SJ (2004) Structure of the coat protein in Pf1 bacteriophage determined by solid-state NMR spectroscopy. J Mol Biol 341:869–879

Thomassen E, Gielen G, Schütz M, Schoehn G, Abrahams JP, Miller S, van Raaij MJ (2003) 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

Tremblay DM, Tegoni M, Spinelli S, Campanacci V, Blangy S, Huyghe C, Desmyter A, Labrie S, Moineau S, Cambillau C (2006) Receptor-binding protein of Lactococcus lactis phages: iden-tification and characterization of the saccharide receptor-binding site. J Bacteriol 188:2400–2410

van Raaij MJ, Schoehn G, Burda MR, Miller S (2001) Crystal structure of a heat and protease-stable part of the bacteriophage T4 short tail fibre. J Mol Biol 314:1137–1146

Veesler D, Robin G, Lichière J, Auzat I, Tavares P, Bron P, Campanacci V, Cambillau C (2010) Crystal structure of bacteriophage SPP1 distal tail protein (gp19.1): a baseplate hub paradigm in gram-positive infecting phages. J Biol Chem 285:36666–36673

Veesler D, Quispe J, Grigorieff N, Potter CS, Carragher B, Johnson JE (2012a) Maturation in action: CryoEM study of a viral capsid caught during expansion. Structure 20:1384–1390

Veesler D, Spinelli S, Mahony J, Lichiere J, Blangy S, Bricogne G, Legrand P, Ortiz- Lombardia M, Campanacci V, van Sinderen D, Cambillau C (2012b) Structure of the phage TP901-1 1.8 MDa baseplate suggests an alternative host adhesion mechanism. Proc Natl Acad Sci U S A 109:8954–8958

Vinga I, Baptista C, Auzat I, Petipas I, Lurz R, Tavares P, Santos MA, São-José C (2012) Role of bacteriophage SPP1 tail spike protein gp21 on host cell receptor binding and trigger of phage DNA ejection. Mol Microbiol 83:289–303

Walter M, Fiedler C, Grassl R, Biebl M, Rachel R, Hermo-Parrado XL, Llamas-Saiz AL, Seckler R, Miller S, van Raaij MJ (2008) Structure of the receptor-binding protein of bacteriophage Det7: a podoviral tail spike in a myovirus. J Virol 82:2265–2273

Wang YA, Yu X, Overman S, Tsuboi M, Thomas GJ, Egelman EH (2006) The structure of a filamentous bacteriophage. J Mol Biol 361:209–215

Xiang Y, Rossmann MG (2011) Structure of bacteriophage φ29 head fibers has a supercoiled triple repeating helix-turn-helix motif. Proc Natl Acad Sci U S A 108:4806–4810

Xiang Y, Morais MC, Cohen DN, Bowman VD, Anderson DL, Rossmann MG (2008) Crystal and
   cryoEM structural studies of a cell wall degrading enzyme in the bacteriophage φ29 tail. Proc
   Natl Acad Sci U S A 105:9552–9557
Xiang Y, Leiman PG, Li L, Grimes S, Anderson DL, Rossmann MG (2009) Crystallographic
   insights into the autocatalytic assembly mechanism of a bacteriophage tail spike. Mol Cell
   34:375–386
Xu L, Benson SD, Butcher SJ, Bamford DH, Burnett RM (2003) The receptor binding protein P2 of
   PRD1, a virus targeting antibiotic-resistant bacteria, has a novel fold suggesting multiple
   functions. Structure 11:309–322
Xu J, Hendrix RW, Duda RL (2014) Chaperone-protein interactions that mediate assembly of the
   bacteriophage λ tail to the correct length. J Mol Biol 426:1004–1018
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
Yap ML, Rossmann MG (2014) Structure and function of bacteriophage T4. Future Microbiol
   9:1319–1327
Yutin N, Backstrom D, Ettema TJG, Krupovic M, Koonin EV (2018) Vast diversity of prokaryotic
   virus genomes encoding double jelly-roll major capsid proteins uncovered by genomic and
   metagenomic sequence analysis. Virol J 15:67
Zheng W, Wang F, Taylor NMI, Guerrero-Ferreira RC, Leiman PG, Egelman EH (2017) Refined
   cryo-EM structure of the T4 tail tube: exploring the lowest dose limit. Structure 25:1436–1441
Zimmer SG, Millette RL (1975) DNA-dependent RNA polymerase from *Pseudomonas* BAL-31.
   II. Transcription of the allomorphic forms of bacteriophage PM2 DNA. Biochemistry
   14:300–307