Molecular assembly and structure of the bacteriophage T4 tail

Fumio Arisaka¹ · Moh Lan Yap² · Shuji Kanamaru³ · Michael G. Rossmann²

Received: 27 July 2016 / Accepted: 3 October 2016 / Published online: 5 November 2016
© International Union for Pure and Applied Biophysics (IUPAB) and Springer-Verlag Berlin Heidelberg 2016

Abstract The tail of bacteriophage T4 undergoes large structural changes upon infection while delivering the phage genome into the host cell. The baseplate is located at the distal end of the contractile tail and plays a central role in transmitting the signal to the tail sheath that the tailfibers have been adsorbed by a host bacterium. This then triggers the sheath contraction. In order to understand the mechanism of assembly and conformational changes of the baseplate upon infection, we have determined the structure of an in vitro assembled baseplate through the three-dimensional reconstruction of cryo-electron microscopy images to a resolution of 3.8 Å from electron micrographs. The atomic structure was fitted to the baseplate structure before and after sheath contraction in order to elucidate the conformational changes that occur after bacteriophage T4 has attached itself to a cell surface. The structure was also used to investigate the protease digestion of the assembly intermediates and the mutation sites of the tail genes, resulting in a number of phenotypes.

Keywords Bacteriophage · Contractile tail · Tail baseplate · Assembly · Infection · Molecular recognition

This article is part of a Special Issue on 'Analytical Quantitative Relations in Biochemistry' edited by Damien Hall and Stephen Harding.

✉ Fumio Arisaka
arisaka.fumio@nihon-u.ac.jp

¹ Life Science Research Center, School of Bioresource Science, Nihon University, 1866 Kameino, Fujisawa 252-0880, Japan

² Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA

³ Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259, Nagatsuta, Midori-ku, Yokohama 226-8501, Japan

Introduction

The efficiency of infection of bacteriophage T4 (further referred to as T4) is almost 1.0, which means that almost every phage particle adsorbed by a bacterium will successfully infect and replicate in the cell. It is the contractile tail of bacteriophage which gives rise to the extremely high efficiency of infection. The T4 virion is very stable and can be stored in a refrigerator for many years. When exposed to a host cell, it will immediately proceed with the structural changes of the baseplate upon adsorption, followed by contraction of the sheath for infection. The baseplate consists of a hub and six surrounding wedges. In a recent study, the wedge, which is known to assemble in a strictly sequential manner, was formed by mixing *Escherichia coli* cells in which seven gene products, namely, gp10, gp7, gp8, gp6, gp53, gp25 and gp11, were separately expressed and then lysed (Yap et al. 2010). All of the assembly intermediates of the wedge were formed and isolated. However, when gp53, which is the penultimate gene product to be associated, was added to the wedge intermediate, six wedge intermediates spontaneously associated to form a baseplate-like structure, corresponding to the post-infection form of the baseplate after the sheath has contracted. The structure of the in vitro-assembled baseplate, which has a molecular weight of 3.3 MDa was elucidated by both X-ray crystallography and cryo-electron microscopy (cryoEM) three-dimensional (3D) reconstruction. The structure obtained by cryoEM achieved a resolution of 3.8 Å (Yap et al. 2016), which was better than the 4.2 Å resolution of the crystallographic data (Yap et al. 2014). The structure of the baseplate not only showed the hitherto unknown structures of some gene products, including gp7, gp53 and part of gp6 and gp10, but also showed why the association of gp53 spontaneously forms the hexagonal baseplate (Tables 1, 2).

Infection of phage T4 proceeds in two steps. The first step is host recognition by the tip of the long tailfibers. The receptor is a glucose residue of the core saccharides of the lipopoly-saccharide. The binding of the virus to the receptor is reversible. When the signal of the receptor binding, gene product 9 (gp9), is transmitted to the baseplate via the long tailfibers, the baseplate changes its conformation from “dome-shaped” to “star-shaped”, and the short tailfibers, which had been

accommodated under the baseplate, spring out and bind to the heptose residue of the core saccharide. This conformational change of the baseplate triggers the contraction of the tail sheath.

Contraction of the tail sheath was studied by Moody who first use formaldehyde to partially fix the tail sheaths and then induced sheath contraction by urea. He showed that sheath contraction starts at the end closest to the baseplate and

Table 1 Gene products involved in the assembly of the tail of phage T4

Gene product (gp)	Amino acid residue number	Stoichiometry in the tail	Localization or function	Protein Data Base	Reference
gp10	602	18 (3)	Wedge (short tailfiber attachment)	5HX2, 5IV5 ^a , 5IV7 ^b	Yap et al. 2016, Taylor et al., 2016
gp7	1032	6 (1)	Wedge	5HX2, 5IV5, 5IV7	Yap et al. 2016, Taylor et al., 2016
gp8	334	12 (2)	Wedge	1N80, 5HX2, 5IV5, 5IV7	Leiman et al. 2003; Yap et al. 2016, Taylor et al., 2016
gp6	660	12 (2)	Wedge	5HX2, 5IV5, 5IV7	Yap et al. 2016, Taylor et al., 2016
gp5	575	3 (3)	Hub (tail lysozyme)	1 K28	Kanamaru et al. 2002
gp5.4	97	1 (1)	Hub	4KU0	
gp27	391	3 (3)	Hub	1 K28	Kanamaru et al. 2002
gp29	590	ND	Hub (tape measure)		
gp26	208	ND	Hub (?)		
gp28	177	ND	Hub (?)		
gp9	288	18 (3)	Socket for long tailfibers	1S2E, 5IV5, 5IV7	Kostyuchenko et al. 1999, Taylor et al., 2016
gp11	219	18 (3)	Baseplate	1EL6, 5IV5, 5IV7	Leiman et al. 2000, Taylor et al., 2016
gp12	527	18 (3)	Short tailfiber	1OCY, 5IV5	Thomassen et al. 2003, Taylor et al., 2016
gp25	132	6 (1)	Baseplate (sheath initiator)	5IW9	Taylor et al., 2016
gp48	364	6 (ND)	Baseplate	5IV5, 5IV7	Taylor et al. 2016
gp53	196	6 (1)	Baseplate	5HX2, 5IV5, 5IV7	Yap et al. 2016, Taylor et al., 2016
gp54	320	6 (ND)	Baseplate (tube initiator)	5IV5, 5IV7	Taylor et al. 2016
gp18	659	138 (1*)	Sheath protein	3FOA	Aksyuk et al. 2009
gp19	163	138 (1)	Tube protein	5IV5	Taylor et al., 2016
gp3	176	6 (6)	Tube terminator		
gp15	272	6 (6)	Sheath terminator	4HUD	Fokine et al. 2013
gp57A	80	–	Chaperone		
gp51	249	ND	Catalytic		

ND, Not determined; (), in solution; (*), in low ionic strength buffer, a: pre-attachment structure, b: post-attachment structure

Table 2 Gene products other than tail proteins discussed in this review

Gene product (gp)	Amino acid residue number	Localization or function
gp34	1289	Proximal long tailfiber protein
gp37	1026	Distal long tailfiber protein
gp13	309	Neck protein
gp14	256	Neck protein
gp20	512	Portal vertex
gp63	374	Facilitates tailfiber attachment, RNA ligase activity
gp38	183	Molecular chaperone for long tailfiber folding

propagates like a wave toward the end close to the head (Moody 1973; Caspar 1980). Micro-calorimetry has been used to show that the contraction is exothermic (Arisaka et al. 1981). Crowther et al. (1977) localized a number of gene products in the baseplate, including gp9, gp12 and gp11, by using mutant phages which lacked specific gene products. The results were later confirmed and expanded by cryoEM 3D reconstruction of the tail before and after sheath contraction (Kostyuchenko et al. 2003, 2005; Leiman et al. 2004).

Here we review assembly and conformational changes in the baseplate based on newly determined cryoEM reconstructions at 3.8 Å resolution (Yap et al. 2016) and protease digestion of the wedge intermediates (unpublished results).

The sequential assembly pathway of the tail

The baseplate consists of a hub and six surrounding wedges. The tail assembly pathway was established based on the model formulated by Kikuchi and King (1975) and was modified by Leiman et al. (2010) and Yap et al. (2016) with additional information on the in vitro-assembled complex determined by cryoEM 3D reconstruction to a resolution of 3.8 Å. Gp7, which binds to gp10 in the first step of wedge assembly, plays the central role in connecting baseplate proteins. Yap et al. (2010) also observed that as soon as gp53 was added to a wedge intermediate, consisting of gp6, gp7, gp8 and gp10, six of these complexes associated to form a baseplate-like structure, which corresponded to the contracted, or low energy, conformation of the tail baseplate. These authors showed that gp53 binds to the interface between two wedges to form the hexagonal baseplate. However, in the presence of the central hub, (gp27)₃(gp5)₃(gp5.4), the baseplate assembles into the high-energy, dome-shaped structure by bonding to gp27 and gp6 (Kanamaru et al. 2002), followed by the binding of (gp9)₃ and (gp11)₃(gp12)₃ to stabilize the high-energy, dome-shaped baseplate. Binding of gp48 and gp54 completes the baseplate assembly. Then the baseplate is ready to initiate the tail assembly (Fig. 1). A total of 138 copies of gp19 polymerize into 23 hexameric rings to form the tail tube. The tail tube

polymerization stops when gp29, the tape measure protein, is fully extended. Binding of (gp3)₆ prevents further polymerization. Polymerization of the sheath protein, gp18, around the tail tube starts as soon as the polymerization of the tail tube has been initiated.

Very recently (after submission of the present manuscript), Taylor et al. (2016) reported the 3D reconstruction of the T4 baseplate in pre- and post-host attachment conformation. This reconstruction not only elucidated the structure of a number of previously unsolved proteins, including gp48 and gp54, but also made it possible to discuss the change in the baseplate structure upon infection more precisely than previously possible (See Table 1). In their article, Taylor et al. (2016) proposed an alternative model for the circularization of six wedges in which gp6 C-terminal domain dimerization circularizes the wedge into the baseplate.

The in vivo assembly pathway of the tail has been studied by Ferguson and Coombs (2000). These authors confirmed the results of the in vitro studies, but also introduced the concept of “phage equivalent” or PE which reflects the number of subunits of each product per phage particle. They reported that the appearance of the gene product in their pulse-chase experiments reflected not only the assembly pathway, but also the PE and the pool size or concentration of the soluble gene product in the infected cell. For example, gp18 has a 23-fold larger PE than gp15, but has a much smaller pool size, indicating that the assembly pathway, where gp18 precedes that of gp15, is in agreement.

Protease digestion of the wedge Assembly intermediates can be used as a probe for following the tail assembly

Protease digestion has been used to probe the interaction of the subunits that form the tail assembly. Initially, gp10 and gp7 form the (gp10)₃(gp7) hetero-tetramer. Gp10 itself is readily digested by lysylendopeptidase. However, in the presence of gp7, gp10 becomes protease resistant whereas gp7 is digested into small fragments. Gp10 has been found to bind to the 14-kDa C-terminal fragment

Fig. 1 Assembly of a baseplate based on present and earlier results. **a** Wedge assembly. Gp10, gp8 and gp6 bind sequentially to the gp7 backbone protein. The central hub of the baseplate is assembled independently. **b** Baseplate and tail assembly. Six wedges assemble around the central hub to form a baseplate. Gp53 binds adjacent wedges together, and gp9 and the gp11–gp12 complex subsequently

bind to the baseplate, further stabilizing the dome-shaped configuration. Then, gp48 and gp54 bind to the top of the central hub and initiate polymerization of the tail tube. Gp25 attaches to the gp48–gp54 complex, initiating polymerization of the tail sheath. For clarity, only three rings of the tail sheath are shown (adapted from Fig. 4 of Yap et al. 2016). *gp* Gene product

(residues 908–1032) of gp7, thus protecting it from protease digestion (T. Nakao, unpublished results). When gp8

was added to $(gp10)_3(gp7)_1$ and subjected to protease digestion, a larger 18-kDa fragment (residues 877–1032) of

Fig. 2 Lysylendopeptidase digestion of the wedge intermediates. **a** $(gp10)_3$ before (lane 1) and after (lane 2) digestion. **b** $(gp7)(gp10)_3$ before (lane 1) and after (lane 2) digestion. In the presence of gp7, gp10 becomes protease resistant and binds the 14-kDa C-terminal fragment of gp7 (gp7C residues 908–1032). **c** $(gp7)(gp10)_3(gp8)_2$ before (lane 1) and after (lane 2) digestion; lane 3 Size Exclusion Chromatography purified digested product. In the presence of gp8, the associated C-terminal fragment of gp7 to gp10 becomes larger to be 18 kDa (gp7C residues 877–1032). **d** $(gp10)_3(gp11)_3$ before (lane 1)

and after (lane 2) digestion. In the presence of gp11, gp10 becomes partially protease resistant and gives larger C-terminal fragment (gp10C residues 196–602). The rationale of this difference is now explained based on the quaternary structure of the complex and discussed in the text. Lane M Molecular weight marker. In **a** and **d**, the molecular weights are indicated in the figure; in **b**, the molecular weights are the same as in **a**; in **c**, the molecular weights are 200, 116, 97, 66, 45, 31, 21, 14 and 6 kDa, respectively, beginning from the top of the gel

Fig. 3 **a** The (gp10)₃(gp7)(gp8)₂ complex in the in vitro-assembled baseplate complex. **b** The (gp10)₃(gp7) complex after lysylendopeptidase cleavage. **c** The (gp10)₃(gp7)(gp8)₂ complex after digestion by lysylendopeptidase. Colors: yellow:gp10, red:gp7, blue:gp8.

gp7 remained attached to the gp10gp8 complex (Yap et al. 2010) (Fig. 2).

The 14-kDa fragment, residues 908–1032, corresponds to domain VI of gp7 (Yap et al. 2016). The larger fragment, residue 877–1032, obtained in the presence of gp8, is domain VI plus a long β -strand, which is connected to domain IV of gp7. These results together with the structure of the wedge suggest that gp8 binds to the long gp7 β -strand, thus protecting a protease cutting site on this strand (Fig. 3).

Gp10 alone can be cleaved by lysylendopeptidase at K289, giving rise to the S290–A602 fragment, whereas digestion in the presence of gp11 gives a larger C-terminal fragment, H195–A602, with the cleavage site between K194 and H195. K289 belongs to domain III, where gp11 binds so that only K194 is cleaved. The details of this interaction cannot yet be observed as the atomic structure of domain III has not yet been determined.

Formation of the contractile tail sheath

Both the tail tube and sheath subunits start to polymerize from the baseplate. The head capsid is eventually added to the far end of the assembled tube and sheath. The final tail tube assembly consists of 138 gp19 tail tube subunits and 138 gp18 tail sheath subunits. Gp25 initiates sheath assembly and is likely to bind to the tube–baseplate assembly prior to gp18 polymerization. The structure of gp25 (Protein Data Base ID 5IW9) is similar to the C-terminal domain of gp18 (Fig. 4) and has been proposed to initiate gp18 polymerization (King 1971).

When the sheath assembly is completed, a gp15 hexamer binds at the end distal from the baseplate and stabilizes the sheathed tail. The tail is now ready for head attachment. The resultant sheath, however, is rather

Fig. 4 Baseplate completion and sheath formation. **a** Interaction between gp6 and gp27, a hub protein. **b** Formation of the first and second annuli of gp18. The C-terminal structure of gp18 which resembles gp25 is indicated by red. **c** Interactions between gp18 and the baseplate proteins gp6, gp53

and gp25. Tube polymerization starts with (gp48)₆(gp54)₆ binding. Then after gp25 binds, tail sheath association or polymerization will start. C-terminal structure of gp18 which resembles gp25 is shown in red

Fig. 5 Tails (a), tails isolated from 23 infections (b) and “necked tail” (c) Images of contracted “necked tail” (b, c) from Coombs and Arisaka 1994)

unstable as demonstrated by the sheath protein depolymerizing when the assembly is dialyzed against low ionic strength buffer or water (Tschopp et al. 1977). As the sheath does not dissociate from the tail tube after the head has been attached, the stabilization or irreversible binding of the sheath proteins could be due to the binding of the neck protein(s), gp13 and/or gp14, to the tail or some other conformational change upon head attachment. The polymerization of the sheath protein on the tube–baseplate is very cooperative in that there is a nucleation difficulty. However, once the polymerization starts, it proceeds to the end. Sheaths with intermediate lengths are rarely observed (Arisaka et al. 1979).

Coombs and Eiserling (1977) reported the “necked tail” which could be isolated by osmotic shock. The sheath of the tails isolated from 23[−] infections do not contract, but the sheath is dissociated upon heat treatment or in 2 M urea. In contrast, the sheath of the “necked tails” can contract under the same conditions. Coombs and Eiserling (1977) isolated the “necked tails” using the following procedure. T4 phage particles were isolated and concentrated by a cesium chloride (CsCl) step gradient; the phage in equilibrium with CsCl was then subjected to osmotic shock by injection into a buffer without CsCl. As a result, a small percentage of the viruses, their tails are detached from the head with their portal protein, gp20, kept attached (Fig. 5b). The sheath of the isolated necked tails contracts upon heating or adding high concentrations of urea. The heat treatment led to tail sheath contraction with no conformational change of the baseplate, whereas a high concentration of urea caused normal sheath contraction accompanied by the conformational change of the baseplate. Caspar (1980) pointed out based on an entity model that the wave-like transmission of the contraction can be explained by the sequential detachment of the originally interacting gp18 and gp19 from the baseplate side toward the head.

Thermodynamics of the tail sheath contraction was studied by micro-calorimetry (Arisaka et al. 1981). By using differential scanning micro-calorimetry for ghosts, these authors observed an endothermic peak at 68 °C followed by a negative peak at 72 °C with the exothermic heat of −44 kcal/mole of gp18, a sheath subunit. The endothermic peak corresponds to the denaturation of some proteins which triggers the sheath contraction. In this case, the baseplate did not change its conformation and stayed at its original position of the distal end from the head. The measurement of the heat of urea-induced sheath contraction which resembles the sheath contraction upon phage infection, where the baseplate changes its conformation from “dome-shaped” to “star-shaped”, was more difficult due to the large endothermic heat from the mixing with urea, but the minimum value of −25 kcal/mole was obtained by using a calorimeter with a mixing chamber. In any event, the contraction of the sheath is exothermic.

DNA ejection is not coupled to tail sheath contraction. In fact, treatment of the virion with urea causes sheath contraction, but the phage retains the entire DNA. It has been shown that the phage with a contracted sheath due to urea treatment can infect host spheroplasts (Furukawa and Mizushima 1982; Furukawa et al. 1979, 1983).

Injection of DNA into the cytoplasm requires electrochemical potential of the inner membrane (Labedan et al. 1980). The electrochemical potential consists of two terms; one is the asymmetric charge distribution on two sides of the membrane, and the other is the concentration difference of the proton.

Tail sheath protein gp18

Tail sheath protein gp18 has been difficult to crystallize due to its tendency to form a polysheath. Vadim Mesyanzhinov’s group designed and produced a series of gp18 mutant proteins in which parts of the sequence were deleted. A number of the mutant proteins were crystallized and the structures determined. The structure of the largest gp18 fragment (residues 1–510) was determined by Aksyuk et al. (2009; Fig. 4). In this crystal structure the 20 N-terminal amino acid residues as well as residues 484–496 were not ordered. The structure comprises about 74 % of the full-length wild-type gp18 and consists of three domains. Domains I and II form the protease-resistant fragment. Domain I is an insert into domain II and Domain I and II are an insert into domain III like a Russian nesting doll (matryoshka).

We previously mapped various mutation sites in the T4 genome (Figs. 6, 7) and characterized their phenotypes (Takeda et al. 2004). There are five phenotypes, a nonsense mutation, amber (am), and four missense mutations, namely, temperature-sensitive (ts), cold-sensitive (cs), heat-sensitive (hs) and carbowax mutants (CBW). Regarding the amber

Fig. 6 Crystal structure of gp18 (residues 20–510) (Aksyuk et al. 2009) in combination with the predicted C-terminal domain (Fokine et al. 2013). Four domains of gp18 are shown in blue (Domain I), olive green (Domain II), orange-red (Domain III) and green (C-terminal domain). The N-terminal 20 residues and residues 484–496 are not ordered in the crystal structure and thus not observed. Residue 510 is the last residue in the crystal structure. Mutation sites with a number of phenotypes were mapped by DNA sequencing of the mutants (Takeda et al. 2004). The temperature-sensitive mutants (*ts*; grow at 30 °C, but not at 42 °C) were

mapped as W34R (or P), Q39P, P219L, G257D, Q277Y (or C, R, F), W487R, G587N and Q589Y. The heat-sensitive mutants (*hs*; lose viability upon incubation for 30 min at 55 °C) were mapped as P5S, R565C and G598S. The cold-sensitive mutants (*cs*; grow at 37 °C, but not at 25 °C) were mapped as R584H, Q589E and Q592E (or R); they are in the C-terminal domain circled). The carbowax mutants (CBW; can infect in the presence of a high concentration of polyethylene glycol) were mapped at G106S, S175F and A178V, in Domain I, and the mutation sites are circled

nonsense mutation, a number of amino acids could be inserted at the amber site by using amber suppressors. *ts* mutants can grow at 30 °C, but not at 42 °C; *cs* mutants can grow at 37 °C, but not at 25 °C (Scotti 1968); *hs* mutants can be inactivated at 55 °C for 30 min (Yamamoto and Uchida 1975); CBW mutants can adsorb to the host bacterium at a high concentration

of polyethylene glycol, where the wild-type phage cannot (Follansbee et al. 1974).

Mutation sites with the same phenotype tend to be mapped close to each other in the 3D structure. For example, mutation sites of CBW mutants are close together in Domain I, and *hs* and *cs* mutation sites are in Domain

Fig. 7 Schematic map of amino acid replacements and phenotypes on the primary structure of gp18, the tail sheath protein. Black and white lines correspond to protease-resistant domain and PS17 phage sheath-homologous regions, respectively. (Takeda et al. 2004). See section Tail sheath protein gp18 for definition of mutations

Fig. 8 **a.** Structure of $(gp27)_3(gp5)_3(gp5.4)_1$ to show the location of the pseudo-sixfold symmetry of gp27 and pseudo threefold symmetry of gp5.4. **b.** Four domains of gp27, where Domains I and III constitute part of the pseudo sixfold symmetry as shown in **c.** **c.** Overview of gp27 and gp5.4 to show the pseudo sixfold symmetry of $(gp27)_3$ and pseudo threefold symmetry of gp5.4. **d.** Parallel β -sheets formed between gp5.4 and gp5C. #1, #2, #3 Three identical subunits (numbers are arbitrary)

III of gp18. The sheath of hs mutants in gene 18 contract upon heat treatment, but upon contraction, the baseplate retains their conformation as dome-shape and stays at the original position of the tail tube, indicating that these mutation sites reside in the interaction sites with the baseplate. It should be noted that hsP5S mutation is far from the other hs sites, but actually close in 3D structure. The cs mutation is very likely related to the hydrophobic inter-subunit interaction site, as expected from the temperature

dependence of the hydrophobic nature. Fokine et al. (2013) reported the structure of the gp18 C-terminal domain (Fokine et al. 2013). The previously reported structure of Domains I through III together with the C-terminal structure is shown in Fig. 6.

The structure of subunit gp18 does not change much upon contraction or structure transformation of the baseplate from dome-shaped to star-shaped. Upon contraction, a hexamer ring of gp18 slides into the next hexamer ring below. The

Fig. 9 Crystal structure of gp5 structure (residues 1–575) (Kanamaru et al. 2002). Only one polypeptide chain of the trimer is shown for clarity. ts and hs mutation sites were mapped by DNA sequencing of the mutants, with ts mutants mapped as A65T, T80A, A276T, G322D(5ts1) and E337K, and hs mutants mapped as P122S and A352T (Takeda et al. 1998). See section Tail sheath protein gp18 for definition of mutations

Fig. 10 Schematic map of amino acid replacements and phenotypes on the primary structure of gp5, the tail lysozyme protein. *Black region of bar* corresponds to T4-lysozyme homologous region. Five ts mutation sites and two hs sites as well as three amber sites are listed (Takeda et al. 1998). See section [Tail sheath protein gp18](#) for definition of mutations

resultant contracted sheath is stable in 4 M Gdn-HCl (Brenner and Horne 1959).

Assembly and structure of the hub

Six gene products, namely, gp5, gp27, gp26, gp28, gp51 and gp29, are required for the formation of the baseplate hub. Among these, gp51 has been reported to have catalytic activity as opposed to stoichiometric (Snustad 1968). Gp26 and gp28 appear to be components of the baseplate, but their location and specific functions are unknown (Nivinskas et al. 1990, 1992; Nieradko and Koszałka 1999; Nieradko et al. 1998). Gp29 has been identified as a tape measure protein (Abuladze et al. 1994), but it has been difficult to isolate this protein.

The main structural component of the hub, (gp27)₃(gp5)₃, was determined by Kanamaru et al. (2002). Kanamaru et al. had found that gp5 undergoes maturational processing in which the peptide bond between Ser351 and Ala352 is cleaved to generate gp5* and (gp5C)₃ (Kanamaru et al. 1999).

Three C-terminal domains of gp5, (gp5C)₃, assume a triple β-helix. This is the needle which punctures the host cell upon infection. Gp5*, which enters the periplasm together with the needle, has a lysozyme domain that hydrolyzes the peptidoglycan or cell wall to dig a hole. The tail tube then penetrates to the inner membrane through the hole, and the phage DNA in the head is injected into the cytoplasm through the tube. The gp27 trimer probably stays at the tip of the tail tube and functions as a channel for DNA entry (Hu et al. 2015). It has been reported that DNA injection requires the electrochemical potential of the inner membrane (Labedan et al. 1980).

The electron density at the tip of (gp5C)₃ (Kostychenko et al. 2003), which was at the time of the study unidentified, was recently identified as gp5.4 and the structure has been resolved by Leiman and his group (Shneider et al. 2013; Fig. 4a). The stoichiometry of gp5.4 against gp27 and gp5 in the complex (gp27)₃(gp5)₃(gp5.4) is 3:1. The complex is

Fig. 11 Structure of gp15-18 before (a, c) and after (b, d) after sheath contraction) and schematic longitudinal cross-section of the neck area (adapted from Figs. 1 c, d and 6 c–f by Fokine et al. 2013)

associated with the tail tube (gp19) via the (gp48)₆(gp54)₆ complex of which the structure has not been resolved. In the dome-shaped baseplate, the hub protein gp27 and the wedge protein gp6 are associated and stabilize the structure before contraction (Fig. 4a). At the interface between gp5C and gp5.4, parallel β -sheets are formed. As shown in Fig. 8d, residues 28–32, 66–70 and 93–97 of gp5.4 each associates with the residues 571–575 from the last triple β -strand of gp5C. On the other hand, the pseudo-hexameric symmetry of gp27 (Fig. 8c) has a connection to gp48. The structure of gp48 and gp54 has been recently reported (Taylor et al. 2016).

Tail lysozyme gp5

Gp5 is a structural protein, but it has lytic activity which is ascribed to the lysozyme domain. The lysozyme domain has 43 % sequence identity with the T4 lysozyme (gpe). The crystal structure of gp5 has been determined as a complex, (gp27)₃(gp5)₃ (Kanamaru et al. 2002). As shown in Fig. 9, this structure consists of three domains, namely, the N-terminal domain connected to the gp27 trimer, the lysozyme domain and the C-terminal β -helix (“needle”) domain. There are a number of ts and cs mutations.

Three notable mutations are worthy of mention: A65T (5tsDH6318), G322D (5ts1) and A352T (5hsY204). A65T (5tsDH6318) has been isolated as a pseudo-revertant of the gene 63[−] mutant. Gp63 has a function to facilitate tailfiber attachment, and A65T is a bypass mutant of gene 63 deletion (Hall et al. 1980). The mechanism of the bypass function is not known. G322D (5ts1) is a bypass mutant of gene e, namely T4 lysozyme gene. This mutant was used to identify the gene encoding the phage-associated lysozyme activity for the first time (Kao and McClain 1980). Gpsp (61.3) does not bind to gpe, but does bind to gp5, and the structure of gpsp/gp5Lys has been determined (Kanamaru et al., in preparation). Gp sp has been assumed to be responsible for preventing superinfection. It is surmised that gp sp does not bind the lysozyme domain of the gp5ts1(G322D) mutant because of the amino acid substitution from Gly to bulky Asp residue at residue 322, which. Asp is the corresponding amino acid residue in wild-type gpe (Fig. 10). The mutation site of 5hsY203(A352T) is in the linker between the lysozyme domain and the C-terminal domain. After assembly of the tail, the linker is cleaved between Ser351 and Ala352, but the C-terminal domain stays in the tail until infection.

No mutants have been mapped in the β -helix needle domain of gp5C. It is possible that gp5C is essential for infection and that any mutation in that domain might not be viable.

Gp9 (tailfiber socket) and gp12 (short tailfiber) have a stabilizing effect on the baseplate. Crowther (1980) utilized 9[−]/34[−] or 9[−]/X4E multiple mutants which possess neither gp9 nor

tailfibers to isolate temperature-sensitive pseudo-revertants. These mutants can infect without the presence of tailfibers and have been named pfp (permit fiberless plating). The mutation sites were identified as being in gene 27, gene 6 or gene 7. All three genes code for a baseplate protein, and their mis-sense mutations destabilize the baseplate.

Head joining and tailfiber attachment

After tail assembly has been completed with the attachment of the connector (Fig. 11), gp15, the tails are ready for being joined to the head. Tails are made in considerable excess over the number of available heads, probably because heads cost much more energy to assemble. The gp13–gp14 complex attaches to the portal end of the head after DNA packaging is completed and forms a part of the neck, and attaches directly to gp15 in the tail. The structure of the gp13–gp14 complex has not yet been determined, but it presumably has a sixfold symmetry and binds to the 12-fold symmetric portal complex, (gp20)₁₂ (Driedonks and Caldentey 1983; Sun et al. 2015). It has been reported that the molar ratio of gp13 and gp14 is 2:1 (Akhter et al. 2007), which may imply that the gp13–gp14 complex may connect the hexameric gp15 and dodecameric gp20. Gpwac (whisker or fibrin) forms the whiskers. It is intriguing that the two hexameric rings of identical subunits, gpwac, assume a somewhat different conformation (Kostychenko 2005).

Finally, six long tailfibers attach to the baseplate through gp9 to complete the infectious phage assembly. There are a couple of proteins that facilitate the tailfiber attachment. One of these is the whisker. The hinge region of the tailfiber first binds to the tip or the C-terminal end of the whisker, which orients the tailfibers so that the N-terminal end of gp34 binds to gp9 in the upper edge of the dome-shaped baseplate. The other gene product which facilitates the tail attachment is gp63. Gp63 has a dual function; one is RNA ligase activity and the other is to facilitate the tailfiber attachment (TFA). It has been shown that the RNA ligase has no relation to facilitating attachment by deleting the ligase activity of gp63 yet retaining the facilitating function (Snopek et al. 1977). The crystal structure of gp63 has been determined (El Omari et al. 2006), but the mechanism of the TFA activity is still unknown. The sheath and whiskers play important roles in regulating the tailfiber positioning between upregulation (retracted; non-infectious) and downregulation (infectious) for infection.

Conclusions

1. Gp7 plays the central role in connecting baseplate proteins in the wedge. The sequential wedge assembly is due to the

increasing stability and binding affinity of gp7 upon attachment of the wedge proteins in a specific order. (Section [Formation of the contractile tail sheath](#))

2. Gp53 plays an essential role in association of six wedges to form the hexagonal baseplat. (Section [The sequential assembly pathway of the tail](#))
3. Gp25 resembles the C-terminal part of gp18 so that it functions as part of the initiator of gp18 polymerization on the tube baseplate. (Section [Formation of the contractile tail sheath](#))
4. In order to form a complicated and intricate structure of the tail, domains or motifs are utilized to form a pseudo-symmetrical structure observed in the monomer–trimer interaction (gp5.4–gp5C) or trimer–hexamer interactions (gp27–gp6). In the retracted form of the tailfibers, six of the distal half fibers, gp37, make contact with the bottom part of the head which has a fivefold (head) symmetry, thus causing a symmetry mismatch. (Sections [Tail sheath protein gp18](#), [Assembly and structure of the hub](#), [Tail lysozyme gp5](#))
5. Together with the hub component, gp27, gp6 plays the key role in forming and maintaining the dome-shape conformation of the baseplate. (Section [Assembly and structure of the hub](#))
6. The C-terminal domain of gp18 interacts with gp19 in polymerization of the sheath surrounding the tail tube. Losing contacts of this domain to gp19 leads to sheath contraction, which can be triggered by conformational changes of the baseplate. (Section [Formation of the contractile tail sheath](#))
7. The cell-puncturing device is formed by gp27–gp5–gp5.4. Gp5.4 has a sharp tip end to punch a hole on the cell membrane, whereas gp5 has a lysozyme domain to degrade the cell membrane for inserting the tail tube to deliver the phage genome into the cell cytoplasm. (Sections [Assembly and structure of the hub](#), [Tail lysozyme gp5](#))
8. Gp15 is the sheath terminator to stop and stabilize the sheathed tail. It also interacts with the neck protein, gp14, when the tail is joined to the fully packaged head. (Section [Head joining and tailfiber attachment](#))

Acknowledgment The authors would like to acknowledge Professor Donald J. Winzor for his long-standing contribution to analytical ultracentrifugation on the occasion of his 80th birthday. This work was supported by JSPS KAKENHI Grant JP16087204 to FA and NIH grant AI081726 to MGR.

Compliance with ethical standards

Conflict of interest Fumio Arisaka declares that none of the authors have any conflicts of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by the author.

References

- Abuladze NK, Gingery M, Tsai I, Eiserling FA (1994) Tail length determination in bacteriophage T4. *Virology* 199:301–310
- Akhter T, Zhao L, Kohda A, Mio K, Kanamaru S, Arisaka F (2007) The neck of bacteriophage T4 is a ring-like structure formed by a hetero-oligomer of gp13 and gp14. *Biochim Biophys Acta* 1774(8):1036–1043
- Aksyuk AA, Leiman PG, Kurochkina LP, Schneider MM, Kostyuchenko VA, Mesyanzhinov VV, Rossmann MG (2009) The tail sheath structure of bacteriophage T4: a molecular machine for infecting bacteria. *EMBO J* 28:821–829. doi:10.1038/emboj.2009.3
- Arisaka F, Tschopp J, van Driel R, Engel J (1979) Reassembly of the bacteriophage T4 tail from the corebaseplate and the monomeric sheath protein P18: a cooperative association process. *J Mol Biol* 132:369–386
- Arisaka F, Engel J, Klump J (1981) Contraction and dissociation of the bacteriophage T4 tail sheath induced by heat and urea. In: DuBow M (ed) *Bacteriophage assembly*. Alan R. Liss, New York, pp 365–379
- Brenner S, Horne RW (1959) A negative staining method for high resolution electron microscopy of viruses. *Biochim Biophys Acta* 34:103–110
- Caspar DLD (1980) Movement and self-control in protein assemblies. Quasi-equivalence revisited. *Biophys J* 32:103135
- Coombs DH, Arisaka F (1994) T4 tail structure and function. In: Karam JD (ed) *Molecular biology of bacteriophage T4*. American Society for Microbiology, Washington DC, pp 259–281
- Coombs DH, Eiserling FA (1977) Studies on the structure, protein composition and assembly of the neck of bacteriophage T4. *J Mol Biol* 116(3):375–405.
- Crowther RA, Lenk EV, Kikuchi Y, King J (1977) Molecular reorganization in the hexagon to star transition of the baseplate of bacteriophage T4. *J Mol Biol* 116:489–523
- Crowther RA (1980) Mutants of bacteriophage T4 that produce infective fibreless particles. *J Mol Biol* 137(2):159–74
- Driedonks RA, Caldentey J (1983) Gene 20 product of bacteriophage T4. II. Its structural organization in prehead and bacteriophage. *J Mol Biol* 166:341–360
- El Omari K, Ren J, Bird LE, Bona MK, Klarmann G, LeGrice SF, Stammers DK (2006) Molecular architecture and ligand recognition determinants for T4 RNA ligase. *J Biol Chem* 281(3):1573–1579
- Ferguson PL, Coombs DH (2000) Pulse-chase analysis of the in vivo assembly of the Bacteriophage T4 tail. *J Mol Biol* 297:99–117
- 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. doi:10.1016/j.jmb.2013.02.012
- Follansbee SE, Vanderslice RW, Chavez LG, Yegian CD (1974) A new set of adsorption mutants of bacteriophage T4D: identification of a new gene. *Virology* 58:180–199
- Furukawa H, Mizushima S (1982) Roles of cell surface components of *Escherichia coli* K-12 in bacteriophage T4 infection: interaction of tail core with phospholipids. *J Bacteriol* 150:916–924
- Furukawa H, Yamada H, Mizushima S (1979) Interaction of bacteriophage T4 with reconstituted cell envelopes of *Escherichia coli* K-12. *J Bacteriol* 140:1071–1080
- Furukawa H, Kuroiwa H, Mizushima S (1983) DNA injection during bacteriophage T4 infection of *Escherichia coli*. *J Bacteriol* 154:938–945
- Hall DH, Sargent RG, Trofatter KF, Russell DL (1980) Suppressors of mutations in the bacteriophage T4 gene coding for both RNA ligase and tail fiber attachment activities. *J Virol* 36:103–108
- Hu B, Margolin W, Molineux IJ, Liu J (2015) Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proc Natl Acad Sci USA* 112:E4919–E4928. doi:10.1073/pnas.1501064112

- Kanamaru S, Gassner NC, Ye N, Takeda S, Arisaka F (1999) C-terminal fragment of the precursor tail lysozyme of bacteriophage T4 stays as a structural component of the baseplate after cleavage. *J Bacteriol* 181:2739–2744
- Kanamaru S, Leiman PG, Kostychenko VA, Chipman PR, Mesyanzhinov VV, Arisaka F, Rossmann MG (2002) Structure of the cell-puncturing device of bacteriophage T4. *Nature* 415:553–557
- Kao S-H, McClain WH (1980) Roles of bacteriophage T4 gene 5 and gene 8 products in cell lysis. *J Virol* 34:104–107
- Kikuchi Y, King J (1975) Genetic control of bacteriophage T4 baseplate morphogenesis. I. Sequential assembly of the major precursor, in vivo and in vitro. *J Mol Biol* 99:645–672
- King J (1971) Bacteriophage T4 tail assembly: four steps in core formation. *J Mol Biol* 58(3):693–709
- Kostyuchenko VA, Navruzbekov GA, Kurochkina LP, Strelkov SV, Mesyanzhinov W., Rossmann MG (1999) The structure of bacteriophage T4 gene product 9: the trigger for tail contraction. *Structure* 7(10):1213–22
- Kostyuchenko VA, Leiman PG, Chipman PR, Kanamaru S, van Raaij M, Arisaka F, Mesyanzhinov VV, Rossmann MG (2003) Three-dimensional structure of bacteriophage T4 baseplate. *Nat Struct Biol* 10:688–693. doi:10.1038/nsb970
- 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
- Labadan B, Heller KB, Jasaitis AA, Wilson TH, Goldberg EB (1980) A membrane potential threshold for phage T4 DNA injection. *Biochem Biophys Res Commun* 93:625–630
- Leiman PG, Kostyuchenko VA, Shneider MM, Kurochkina LP, Mesyanzhinov W, Rossmann MG (2000) Structure of bacteriophage T4 gene product 11, the interrace between the baseplate and short tail fibers. *J Mol Biol* 301(4):975–85
- Leiman PG, Shneider MM, Kostyuchenko VA, Chipman PR, Mesyanzhinov W, Rossmann MG. (2003) *J Mol Biol* 328(4):821–33
- 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. doi:10.1016/j.cell.2004.07.022
- 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* 7:355–382
- Moody MF (1973) Sheath of bacteriophage T4. III. Contraction mechanism deduced from partially contracted sheaths. *J Mol Biol* 80:613–635
- Nieradko J, Koszałka P (1999) Evidence of interactions between Gp27 and Gp28 constituents of the central part of bacteriophage T4 baseplate. *Acta Microbiol Pol* 48:233–242
- Nieradko J, Koszałka P, Krzywicka A (1998) Characteristics of gene 28 product, the constituent of the central part of bacteriophage T4 baseplate. *Acta Microbiol Pol* 47:243–252
- Nivinskas R, Raudonikiene A, Vaishkunaite R (1990) Bacteriophage T4 gene 26. *Nucleic Acids Res* 18:1913
- Nivinskas R, Vaishkunaite R, Dagyte R, Raudonikiene A, Klaus V (1992) Cloning, sequence, and overexpression of bacteriophage T4 gene 51. *Virology* 188:887–889
- Scotti PD (1968) A new class of temperature conditional lethal mutants of bacteriophage T4D. *Mutant Res* 6:1–14
- Shneider MM, Buth SA, Ho BT, Basler M, Mekalanos JJ, Leiman PG (2013) PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. *Nature* 500:350–353
- Snopek TJ, Wood WB, Conley MP, Chen P, Cozzarelli NR (1977) Bacteriophage T4 RNA ligase is gene 63 product, the protein that promotes tail fiber attachment to the baseplate. *Proc Natl Acad Sci U S A* 74:3355–3359
- Snustad DP (1968) Dominance interactions in *Escherichia coli* cells mixedly infected with T4D wild-type and amber mutants and their possible implications as to type of gene-product function: catalytic vs. stoichiometric. *Virology* 35:550–563
- Sun L, Zhang X, Gao S, Rao PA, Padilla-Sanchez V, Chen Z, Sun S, Xiang Y, Subramaniam S, Rao VB, Rossmann MG (2015) Cryo-EM structure of the bacteriophage T4 portal protein assembly at near-atomic resolution. *Nat Commun* 6:7548. doi:10.1038/ncomms8548
- Takeda S, Hoshida K, Arisaka F (1998) Mapping of functional sites on the primary structure of the tail lysozyme of bacteriophage T4 by mutational analysis. *Biochim Biophys Acta* 1384(2):243–52
- Takeda Y, Suzuki M, Yamada T, Kageyama F, Arisaka F (2004) Mapping of functional sites on the primary structure of the contractile tail sheath protein of bacteriophage T4 by mutation analysis. *Biochim Biophys Acta* 1699:163–171
- 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(7603):346–352
- Tschopp J, Arisaka F, van Driel R, Engel J (1977) Purification, characterization and reassembly of the bacteriophage T4D tail sheath protein, P18. *J Mol Biol* 128:247–258
- Yamamoto Y, Uchida H (1975) Organization and function of the tail of bacteriophage T4. II. Structural control of the tail contraction. *J Mol Biol* 92:207–223
- Yap ML, Mio K, Leiman PG, Kanamaru S, Arisaka F (2010) The baseplate wedges of bacteriophage T4 spontaneously assemble into hubless baseplate-like structure in vitro. *J Mol Biol* 395:349–360
- Yap ML, Klose T, Plevka P, Aksyuk A, Zhang X, Arisaka F, Rossmann MG (2014) Structure of the 3.3MDa, in vitro assembled, hubless bacteriophage T4 baseplate. *J Struct Biol* 187:95–102
- Yap ML, Klose T, Arisaka F, Speirc JA, Veelslerc D, Fokine A, Rossmann MG (2016) The role of bacteriophage T4 baseplate in regulating assembly and infection. *Proc Natl Acad Sci USA* 113(10):2654–2659