High-resolution structure of podovirus tail adaptor suggests repositioning of an octad motif that mediates the sequential tail assembly

Lingfei Liang, Haiyan Zhao, Bowen An, and Liang Tang

*Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66045

Edited by Michael G. Rossmann, Purdue University, West Lafayette, IN, and approved November 29, 2017 (received for review April 25, 2017)

The sophisticated tail structures of DNA bacteriophages play essential roles in life cycles. Podoviruses P22 and Sf6 have short tails consisting of multiple proteins, among which is a tail adaptor protein that connects the portal protein to the other tail proteins. Assembly of the tail has been shown to occur in a sequential manner to ensure proper molecular interactions, but the underlying mechanism remains to be understood. Here, we report the high-resolution structure of the tail adaptor protein gp7 from phage Sf6. The structure exhibit distinct distribution of opposite charges on two sides of the molecule. A gp7 dodecameric ring model shows an entirely negatively charged surface, suggesting that the assembly of the dodecamer occurs through head-to-tail interactions of the bipolar monomers. The N-terminal helix-loop structure undergoes rearrangement compared with that of the P22 homolog complexed with the portal, which is achieved by repositioning of two consecutive repeats of a conserved octad sequence motif. We propose that the conformation of the N-terminal helix-loop observed in the Sf6-gp7 and P22 portal-gp4 complex represents the pre- and post-assembly state, respectively. Such motif repositioning may serve as a conformational switch that creates docking sites for the tail nozzle only after the assembly of adaptor protein to the portal. In addition, the C-terminal portion of gp7 shows conformational flexibility, indicating an induced fit on binding to the portal. These results provide insight into the mechanistic role of the adaptor protein in mediating the sequential assembly of the phage tail.

Significance

Many dsDNA bacteriophages possess a tail that encodes functions essential for viral life cycles. Assembly of such a molecular machine has been shown to occur in a sequential manner. Here, we report the high-resolution crystal structure of the tail adaptor protein gp7 from Shigella phage Sf6. Comparative structural studies reveal that the N-terminal portion undergoes structural rearrangement by repositioning two consecutive repeats of a conserved octad sequence motif, turning the molecule from the preassembly state to the postassembly state, which creates the binding site for the next tail component to attach to. These results provide a structural basis for a mechanism of repositioning of sequence motifs by which the adaptor protein mediates the sequential assembly of the phage tail.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

Data deposition: Atomic coordinates have been deposited with the RCSB PDB with the entry code 5VGT.

1To whom correspondence should be addressed. Email: liang.w.tang@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1706846115/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1706846115

PNAS | January 9, 2018 | vol. 115 | no. 2 | 313–318
understood. The X-ray structure of phage P22 portal:adaptor complex showed the dodecameric ring-like architecture of the tail adaptor (P22-gp4) that binds to the portal dodecamer with elongated C-terminal portions (21). This and the subnanometer-resolution cryo-EM structures of the isolated P22 tail (22) and the P22 virion (23, 24) shed light on the assembly of the multiple components in the context of the tail and the virion. Here we report the high-resolution X-ray structure of phage Sf6 tail adaptor gp7 in its preassembly state. Comparison with the P22 tail adaptor structure reveals a conformational switch at the N-terminal segment upon tail assembly that ensures the subsequent attachment of the tail nozzle. Such conformational switch is enabled by repositioning of two consecutive repeats of a conserved sequence motif that are capable of binding to the same protein surface.

Results and Discussion

Overall Structure. The Sf6-gp7 has 160 amino acid residues. We designed a truncated gp7 construct lacking 36 C-terminal residues which were predicted as disordered. A C-terminal His-tag was added to the construct to facilitate the protein purification. The structure was determined at 1.78 Å resolution (Table 1), with two molecules in the asymmetric unit related by a non-crystallographic twofold symmetry (Fig. 1). One magnesium ion and four calcium ions were found in the structure. The structure is composed of four α-helices, namely α1–α4, connected by three loops. Despite low sequence homology, this protein fold is common in tail adaptor proteins of podoviruses, siphoviruses, and probably myoviruses as well. Structural superimposition shows a

| Table 1. X-ray data collection and structure refinement statistics |
|-------------------------|-------------------------|
| Data collection         | Native                  |
| Beamsline               | APS 23ID-D              |
| Wavelength, Å           | 1.03324                 |
| Resolution, Å           | 29.6–1.78(1.81–1.78)*   |
| No. of measurements     | 235,103                 |
| Completeness, %         | 98.9 (89.2)*            |
| I/σ, %                  | 17.5 (1.7)*             |
| Rmerge, %               | 5.7 (67.9)*             |
| Space group             | P422                    |
| Unit cell, Å            | a = 101.30, c = 76.46   |

Structure refinement

Resolution, Å     29.55–1.78
Rwork/Rfree     0.18/0.20
No. of atoms     Protein/water 1,838/230
B-factors        Protein/water 40.94/48.23
rmsd              Bond lengths, Å 0.007
                   Bond angles, ° 0.938
Ramachandran plot, %
Most favored     98.28
Allowed          1.72
Disallowed       0.00

APS, Advanced Photon Source.

*A values in parentheses are for the outermost resolution shells.

Rmerge = Σ(|I(hkl)|−<|I(hkl)|>/Σ(|I(hkl)|), where |I(hkl)| is the observed intensity of reflection hkl and <|I(hkl)|> is the averaged intensity of symmetry-equivalent measurements.

Rwork = Σ(|Iobs|−|Icalc|)/Σ(|Iobs|), where Fobs and Fcalc are structure factors of the observed reflections and those calculated from the refined model, respectively. Rwork has the same formula as Rwork except that it was calculated against a test set of the data that were not included in the refinement.

A Distinct Conformation of the N-Terminal Portion. Sf6-gp7 and P22-gp4 share 37% identity and 56% similarity in amino acid sequences (Fig. 2). We extracted a gp4 monomer from the P22 portal:gp4 complex (21) and compared its structure with Sf6-gp7. The protein folds were highly conserved, while the N-terminal and C-terminal portions showed significant differences (Fig. 2B). In the N-terminal segment, Sf6-gp7 has a long α1-helix and a nearly vertically oriented α1α2 loop, while P22-gp4 has the α1-helix half in length than that of Sf6-gp7 and an essentially horizontal α1α2 loop (Fig. 2B). The X-ray structure of Sf6-gp7 fits well into the isolated cryo-EM map of the P22 tail (Fig. S3). However, the α1-helix and the α1α2 loop of Sf6-gp7 penetrate into the portion of the cryo-EM map corresponding to the tail nozzle (P22-gp10 or Sf6-gp8), which locates right underneath the adaptor in the tail complex, creating a steric clash (Fig. 2C). This indicates that the Sf6-gp7 α1-helix and the α1α2 loop must undergo conformational changes upon assembly from the monomers free in solution into the tail.

Interestingly, structural superposition shows that the first half (residue nos. 6–18) of the Sf6-gp7 α1-helix overlays the P22-gp4 α1-helix (residue nos. 12–23) structurally (Fig. 2B), while sequence alignment shows that the second half (residue nos. 14–24) of the
Sf6-gp7 α1-helix aligns well with the P22-gp4 α1-helix (residue nos. 12–23) at the amino acid sequence level (Fig. 2A), indicating a correlation between the two structures. Notably, the amino acid sequences of tail adaptor loop α1α2 are highly conserved between phage Sf6 and P22. Moreover, the tail nozzle proteins in the two phages share as high as 93% sequence identity (3). Thus, it is reasonable to infer that the tail adaptor loop α1α2 may physically interact with the essentially identical tail nozzle in phages Sf6 and P22, and the modes of interactions between the tail adaptor loop α1α2 and the tail nozzle are largely identical, which indicates consistent conformations for the loop α1α2 of Sf6-gp7 and P22-gp4 in the context of the tail. Given that Sf6-gp7 is crystallized in solution while P22-gp4 is in complex with the portal protein, these results together suggest that the distinct conformation of the N-terminal portion of Sf6-gp7 represents the preassembly state, whereas that observed in the P22 portal-gp4 complex structure represents the postassembly state. While free in the infected cell cytoplasm after protein synthesis, the tail adaptor protein may adopt the conformation as observed in the present Sf6-gp7 X-ray structure—that is, its loop α1α2 and α1-helix adopt the vertical orientation. Upon assembly onto the portal ring, the tail adaptor undergoes conformational changes in its loop α1α2 and α1-helix, which move up so that the second half of the α1-helix takes the place of the first half and the loop α1α2 adopts a horizontal orientation. Such conformational changes would eliminate the steric clash with the tail nozzle and create a specific binding site for the tail nozzle.

When transitioning from the preassembly state to the postassembly state, the α1-helix repositions so that its second half takes the place of the first half in the preassembly state. Analysis of amino acid sequences reveals a sequence motif that supports such replacement. In Sf6-gp7, the amino acid region 7–24 (α1-helix) contains two repeats of an octad motif KXXΦΦXXA (Fig. 2A), where K (position 1) is lysine, X is any residue, Φ (positions 4 and 5) represents a hydrophobic residue, and A (position 8) is alanine. This motif is conserved in P22-gp4 (Fig. 2A) and many P22-like phages as well (Fig. S1A). In this motif, Lys, the two hydrophobic residues, and Ala are located on the same side of the α1-helix and thus define a common surface (Fig. 2A). In the X-ray structures of Sf6-gp7 and the P22 portal-gp4 complex, this surface is the one that forms the interface between α1-helix with the rest of the molecule (mainly α2- and α3-helices) (Fig. 2A). Thus, repositioning of the α1-helix during transition from the preassembly state to the postassembly state is enabled by swapping of the two repeats of such octad sequence motif, which utilizes the essentially identical interface and would maintain essentially the same interactions between this helix to the binding interface on the molecule.

The C-Terminal Segment Displays Conformational Flexibility. Superposition of molecules A and B gives an rmsd of 0.468 Å. The significant structural difference is at the C-terminal segment. The eight C-terminal residues in molecule B (residues T117–R124) show no electron density in the X-ray map and thus are disordered, while this region in molecule A forms a well-defined extended loop that is immobilized by interactions with adjacent molecules in the crystal (Fig. 1B). These data indicate conformational flexibility of the Sf6-gp7 C-terminal segment. In addition, the conformation of the Sf6-gp7 C-terminal portion differs from that of P22-gp4, as seen in the P22 portal-gp4 complex structure. A part of this portion in
The Sf6-gp7 monomer shows bipolar distribution of electric charges.

Bipolar Distribution of Electric Charges on the Surface of the Sf6-gp7 Monomer. Sf6-gp7 contains 19 negatively and 10 positively charged residues. While negatively charged residues are spread over the molecule essentially uniformly, 7 out of 10 positively charged residues are concentrated on one face of the molecule. Calculation of the electrostatic potential using USCF Chimera (30) reveals a distinct bipolar feature—that is, a convex, positively charged face on one side and a negatively charged face covering the rest of the molecule (Fig. 3A). During assembly of the ring-like gp7 dodecamer (see below) in the tail, the positively charged ridge of each monomer packs into the negatively charged groove of the adjacent subunit (Fig. 3C). This indicates that assembly of gp7 monomers into the ring-like dodecamer in the tail is mediated by electrostatic interactions. Each monomer mimics a molecular “magnet” that joins the next one in a head-to-tail manner via interactions between its positively charged face and the negatively charged face of the next monomer. The similar bipolar feature is observed on gp7 homologs P22-gp4 and HK97-gp6 (Figs. S4 and S5) (26, 31), suggesting that such charge/charge interactions between bipolar molecules may be a common mechanism for assembly of the adaptor protein monomers into ring-like oligomers in those phage tails.

A Model for the Sf6-gp7 Dodecameric Ring. The P22 portal:gp4 complex X-ray structure (21) and the subnanometer-resolution cryo-EM map of P22 isolated tail (22) reveal how the P22-gp4 assemblies with other components of the tail. Given the apparent sequence identity and common folds between tail component proteins, the overall assembly of the tail may be conserved between the two phages, which is also supported by the similarities of the tail structures as shown in the cryo-EM asymmetric reconstructions of P22 (23) and Sf6 (3). The X-ray structure of the Sf6-gp7 fits well into the 9.4 Å-resolution cryo-EM map of the P22 tail, except for the N-terminal and C-terminal portions where conformational changes occur as discussed above (Fig. S24). The biological assembly of P22-gp4 is a ring-like dodecamer. To model the Sf6-gp7 dodecameric ring, we first docked one Sf6-gp7 monomer to the portal, the C-terminal portions of Sf6-gp7 and P22-gp4 both contain several glycine and proline residues that are conserved between the two phages and are well aligned (Fig. 2A), which may facilitate such conformational changes upon postassembly with the portal. In fact, folding of unstructured segments upon binding to viral partners appears as a general strategy for the control of sequential tail assembly (26, 28, 29).

The peripheral surface of the Sf6-gp7 dodecamer must provide binding sites for the tailspike and the tail nozzle. Thus, electrostatic interactions may play a major role in mediating assembly of the tail adaptor with the tailspike and tail nozzle. This may, at least partially, contribute to the ultrastability of the tail, which is able to withstand heating to 60 °C in 2 M urea and in the presence of detergent in case of P22 (17). Additionally, the highly negatively charged surface of the tail adaptor ring assembly may be suitable for the need of switching from the terminase-bound state of the capsid at the end of DNA-packaging process to the state for terminase dissociation and tail assembly. Interestingly, the highly negatively charged surface of the tail adaptor ring is also observed in P22-gp4, HK97-gp6, and Spp1-gp15 (Figs. S4–S6). SPP1-gp15 shows some variance in that it contains a positively charged proteins as a common characteristic and was thought to help to avoid nonspecific DNA binding during DNA passage through these molecular channels (21, 25, 32). Hence, such a feature in Sf6-gp7 may facilitate translocation of phage dsDNA from inside of the phage into host cell cytoplasm.

Fig. 3. The Sf6-gp7 monomer shows bipolar distribution of electric charges on its surface, whereas the dodecameric gp7 ring has the remarkably negative-charged surface. (A) Three neighboring gp7 monomers from the pseudo-atomic dodecamer model (shown in B) are shown as molecular surface colored according to electrostatic potential (middle) or ribbon (sides). The molecule is positively charged (blue) on one side, and negatively charged (red) on the other side. The positively charged ridge packs into the adjacent, negatively charged groove during assembly. (B) Ribbon diagram of the pseudo-atomic model of gp7 dodecameric ring with chains in different colors. (C) Surface representation of the gp7 dodecameric ring colored according to the electrostatic potential calculated with UCSF Chimera. (D) A cut-away view shows the negatively charged inner surface of gp7 dodecameric ring. For the color scheme of molecular surfaces, the blue area corresponds to an electrostatic potential of +10 kcal/(mol*e), and the red area corresponds to an electrostatic potential of −10 kcal/(mol*e).
rim at the bottom (Fig. S6C). Thus, such a feature may be a common characteristic among those phages.

Implications for the gp7-Mediated Sequential Assembly of the Tail.

The X-ray structure of Sf6 tail adaptor gp7 and the pseudo-atomic model of its dodecameric assembly provide insight into the structural basis of the sequential assembly of the tails in P22-like phages. We propose that the Sf6-gp7 reported here and the P22-gp4 structure in the context of the portal-gp4 complex represent the pre- and postassembly states of the tail adaptor, respectively. Our proposal is supported by the presence of the repeated octad motif KXXΦXXA in the α1-helix of Sf6-gp7 and P22-gp4. Comparison of the two X-ray structures shows that the first motif of the Sf6-gp7 α1-helix is superimposed on the second motif of the P22-gp4 α1-helix (Fig. 2B). Due to the distributions of the four conserved residues in the motif, such a positional replacement between the first and second motifs in α1-helix still retains interactions of this α1-helix with the rest of the molecule. That is to say that upon assembly of the tail adaptor to the portal, the second motif replaces the first motif, generating a different conformation for the highly conserved loop α1α2. Such a conformational change in the loop α1α2 creates the binding site for the next tail component (i.e., the tail nozzle) and eliminates the steric clash between the tail adaptor and the tail nozzle as observed in fitting the Sf6-gp7 structure in the P22 tail cryo-EM map (Fig. 2C).

Based on these results, we propose a model for the sequential assembly of the phage tail in Sf6 (Fig. 4). In the infected cell cytoplasm, the newly synthesized tail adaptor protein molecules exist as monomers (33) that adopt the preassembly conformation as observed in the Sf6-gp7 X-ray structure, with the first octad motif occupying the interface with the protein core and a largely vertically oriented loop α1α2. The C-terminal portion may be flexible and unstructured. Upon completion of the DNA-packaging process, the terminase dissociates from the capsid and a tail adaptor subunit then binds to the portal dodecamer via its elongated C-terminal portion, which undergoes an induced fit from the flexible conformation to an extended and immobilized conformation that makes extensive contact with the portal. The binding probably triggers the upward repositioning of the second octad sequence motif in the α1-helix by ~10 Å, which takes the place of the first motif, and this likely favors the binding of the next tail adaptor subunit and is supported by that facts that P22-gp4 exists as monomer in solution (33) and binds to the portal ring in a highly cooperative manner (19). Accompanying the repositioning of the motif, the loop α1α2 moves away from protein core to a horizontal conformation, generating the binding site for the tail nozzle. This structural rearrangement allows assembly of Sf6-gp7 to the portal in the post-DNA-packaging capsid, followed by attachment of the tail nozzle. The tail adaptor protein free in solution, which is in the preassembly conformation, would not be able to bind to the tail nozzle until it binds to the portal and undergoes the structural rearrangement at the α1-helix and the loop α1α2. After that, tailspike and tail needle will bind sequentially.

Our structural analysis suggests that the sequential tail assembly of P22-like podovirus is enabled by a conserved conformational switch in the tail adaptor via replacement of two consecutive repeats of a sequence motif, which creates the binding site for the tail nozzle only when it is time. Additional studies, such as a mutagenesis approach, can further demonstrate the proposed mechanism, and a high-resolution structure of the entire phage or the isolated tail assembly of Sf6 would unveil the conformation of the postassembly state and ultimately confirm our proposed mechanism. Although Siphoviridae and Myoviridae tail and head assemblies occur independently, assembly of tails in those phages may also follow a sequential order such as in Podoviridae (34). For example, the baseplate wedge of the myovirus T4 is assembled sequentially, which is mediated by the induced conformational change caused by addition of new components (35). SPP1-gp15 has been shown to undergo conformational change during assembly to allow appropriate interaction between its α0-helix and loop α2α3 with the next tail protein gp16 (26, 31). It shares certain similarity with Sf6-gp7, in which the structural change allows the loop α1α2 to interact with gp8. The unique feature in Sf6-gp7 is that the conformational change is achieved by repositioning of the two octad motifs. Such localized structural change may be more efficient and genetically stable compared with the large-scale conformational change as in SPP1-gp15 because only several residues (the conserved residues of the octad motif) are essential. The tandem octad motif is highly conserved in P22-like phages, suggesting a conserved mechanism adopted by this group of phages. It will be interesting to see if the sequence motif-repositioning mechanism is a recent evolutionary result adopted only among P22-like phages or if it also exists in other Caudoviruses.

Materials and Methods

Production of Sf6-gp7. The DNA fragment encoding Sf6-gp7 (residues 1–124) was cloned into pET28b (Novagen) between NcoI/XhoI. The region after residue 124 was predicted as disordered by Phyre2 therefore removed. An His-tag followed the protein at the C-terminus. Protein overexpression was achieved by growth of E. coli strain B834(DE3) at 37 °C in LB until OD600 nm = 0.6, followed by induction at 30 °C by adding isopropyl β-D-thiogalactopiranoside to a final concentration of 1 mM. Cells were harvested after 3 h of growth and lysed on a French press in the resuspension buffer.
buffer (20 mM Tris-HCl, pH 8.5; 500 mM NaCl; 10 mM J-mercaptoethanol). The protein was purified with a Ni-NTA column (Qiagen) followed by gel filtration chromatography on a HiLoad 16/60 Superdex 75 column (GE Healthcare) in the gel filtration buffer (20 mM Tris HCl, pH 8.5; 150 mM NaCl; 1 mM DTT; 1 mM EDTA), which showed an elution volume of 63.82 ml, corresponding to a dimer. The eluted fractions were collected and concentrated to 13 mg/ml using a Millipore centrificon (molecular weight cutoff 10 kDa) before crystallization.

Crystallization, X-Ray Data Collection, and Structure Determination. The purified Sf6-gp7 was crystallized with the hanging drop-vapor diffusion method by mixing 1 μL of the protein solution with 1 μL of the well solution containing 15% (v/v)PEG8000, 0.1 M Mes pH 6.0, and 0.28 M Ca(OAc)2. Crystals were dipped into well solution plus 15% and 30% ethylene glycol in succession before flash freezing in liquid nitrogen. The heavy atom derivativation was obtained by soaking the native crystals in well solution with 5 mM HgCl2 for 24 h before flash freezing.

Strups H, King J, and their coworkers on the heavy atom derivativation X-ray data were collected at the Advanced Photon Source. All data were indexed and integrated using XDS (36, 37), scaled using Aimless (38). The structure was determined by the single anomalous dispersion method using the mercury derivative data (Table 1). The experimental electron density map calculated with Autosol (39) in PHENIX (40) was of excellent quality, which allowed automated building of most of the model of using PHENIX. Small loops of regions of residues 26–29 in chain A and 24–29 and 96–101 in chain B were manually built using COOT (41). Structure refinement coupled with manual model building was performed with the programs PHENIX and COOT, respectively.

Generation of gp7 Dodecamer. Sf6 is a P22-like phage and therefore has conserved architectures with P22 (3). The 9.4 Å cryo-EM reconstruction of the P22 phage tail (EMD-5051) (22) and the 3.25 Å X-ray structure of P22-gp4 dodecamer bound to the portal core (PDB-3L74) (21) have been determined, which offers the possibility to establish a Sf6-gp7 dodecamer model by symmetry operations. Two approaches were employed to achieve the Sf6-gp7 model. First, the P22-gp4 dodecamer was separated from the portal core and fitted into the corresponding volume of P22 tail cryo-EM map. The first approach was to superimpose one Sf6-gp7 monomer (peptide chain A was chosen) onto the P22-gp4 dodecamer, giving an rmsd of 1.26 Å, while the second approach was to directly dock one Sf6-gp7 monomer onto the P22 tail map. This led to slight downward movement of Sf6-gp7 compared with the one superimposed on P22-gp4. For both approaches, the P22-gp4 dodecamer symmetry was applied to Sf6-gp7 by using the CCP4 (42) program lsqkab (43), giving gp4-based and map-based gp7 dodecamer models, respectively.

ACKNOWLEDGMENTS. We thank the General Medical Sciences and Cancer Institutes Structural Biology Facility at the Advanced Photon Source (GMICA @ APS) for beamline access for X-ray data collection. Molecular graphs and molecular modeling were performed with the UCSF Chimera package. This work was partially supported by Grant R01GM090010 and a subproject of the Kansas COBRE Grant P30 GM103326, both from the National Institute of General Medical Sciences of the National Institutes of Health (to L.T.).


Liang et al.