Crystal Structure of pb9, the Distal Tail Protein of Bacteriophage T5: a Conserved Structural Motif among All Siphophages

Ali Flayhan, Frédéric M. D. Vellieux, Rudi Lurz, Olivier Maury, Carlos Contreras-Martel, Eric Girard, Pascale Boulanger, Cécile Breyton

The tail of Caudovirales bacteriophages serves as an adsorption device, a host cell wall-perforating machine, and a genome delivery pathway. In Siphoviridae, the assembly of the long and flexible tail is a highly cooperative and regulated process that is initiated from the proteins forming the distal tail tip complex. In Gram-positive-bacterium-infecting siphophages, the distal tail (Dit) protein has been structurally characterized and is proposed to represent a baseplate hub docking structure. It is organized as a hexameric ring that connects the tail tube and the adsorption device. In this study, we report the characterization of pb9, a tail tip protein of Escherichia coli bacteriophage T5. By immunolocalization, we show that pb9 is located in the upper part of the cone of the T5 tail tip, at the end of the tail tube. The crystal structure of pb9 reveals a two-domain protein. Domain A exhibits remarkable structural similarity with the N-terminal domain of known Dit proteins, while domain B adopts an oligosaccharide/oligonucleotide-binding fold (OB-fold) that is not shared by these proteins. We thus propose that pb9 is the Dit protein of T5, making it the first Dit protein described for a Gram-negative-bacterium-infecting siphophage. Multiple sequence alignments suggest that pb9 is a paradigm for a large family of Dit proteins of siphophages infecting mostly Gram-negative hosts. The modular structure of the Dit protein maintains the basic building block that would be conserved among all siphophages, combining it with a more divergent domain that might serve specific host adhesion properties.

Received 30 August 2013 Accepted 14 October 2013 Published ahead of print 23 October 2013
Address correspondence to Cécile Breyton, Cecile.Breyton@ibs.fr.
* Present address: Ali Flayhan, Institute of Biology and Chemistry of Proteins, Lyon, France.
Copyright © 2014, American Society for Microbiology. All Rights Reserved. doi:10.1128/JVI.02135-13
We localized pb9 in the tail tip at the junction between the tail tube and the conical structure of the host adsorption device of T5. pb9 is composed of two domains, one of which shows structural similarity with the hexamerization domain of Dit tail proteins of phages p2, TP901-1, and SPP1. However, its second domain appears more divergent. Based on these data, we conclude that pb9 is the Dit protein of T5, and we thus propose that the Dit basic building block is a conserved structural motif among all siphophages infecting both Gram-negative and Gram-positive bacteria, which can be combined with a more divergent domain that serves specific adhesion and/or hub properties.

MATERIALS AND METHODS

Cloning, overexpression, and purification. The DNA sequence (GenBank accession number AAU02274.1) coding for the tail protein pb9 was cloned in the pLIM14 (His6-Nter fusion) or pLIM13 (His6-Cter fusion) vector (Noirclerc-Savoye et al., submitted for publication) (14). A tobacco etch virus protease cleavage site was inserted between the His6-Cter fusion and pb9. Positive plasmids were transformed into the chemically competent Escherichia coli BL21 (DE3) expression strain. Transformed cells were cultured for 72 h at 28°C in an autoinduction medium, supplemented with 50 μg/ml kanamycin. Cells were harvested and stored at −80°C. The frozen pellet was resuspended in 30 ml lysis buffer (50 mM Tris, pH 8.0, 150 mM NaCl, and 2 mM MgSO4), supplemented with 50 μl DNase (3 U/μl) and a cocktail of protease inhibitors (EDTA-Free; Roche). Cells were then broken with a microfluidizer at 14,000 lb/in2 and centrifuged for 20 min at 55,000 rpm in a 70 Ti rotor, at 4°C. A final concentration of 250 mM NaCl was added to the supernatant before loading it onto a nickel affinity column (HiTrap chelating; 5 ml; GE Healthcare), equilibrated with 15 ml of equilibration buffer (20 mM Tris, pH 8.0, and 150 mM NaCl). The protein was eluted with a 0.5 M imidazole gradient. pb9-containing fractions were pooled and loaded on an anion exchange column (HiTrap Q; 5 ml; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimer and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl. pb9 dimers and monomers were separated by size exclusion chromatography (SD200 10/300 GL column; GE Healthcare) equilibrated with 20 mM Tris, pH 8.0 and 250 mM NaCl.
tions were used for crystallization. Rabbit immunization against pb9-Nter was carried out according to standard protocols. The antiserum was depleted from E. coli antibodies by incubation with an E. coli cell lysate, and IgGs were purified by affinity chromatography using a HITrap protein A column as recommended by the supplier (GE Healthcare).

**Immunoelectron microscopy.** One microliter of phage T5st0 or hd1 (10^13 PFU/ml) was mixed with 1 μl of purified IgG and complemented to 20 μl with T5 buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM MgCl₂, and 1 mM CaCl₂). The mixture was incubated for 1 h to over-night at 4°C or room temperature and then diluted twice with T5 buffer. Aggregates were discarded after centrifugation (2 min at 18,600 × g), and free IgGs were separated from the cross-linked phages by chromatography on a Sephacryl 500 MicroSpin column (400 μl, 75% slurry; spun for 5 min at 700 × g) (15) equilibrated with T5 buffer. The IgG-phage complexes were imaged after negative staining with 2% uranyl acetate or additionally labeled with anti-rabbit goat IgG–5-nm gold complexes (British Biocell). Free goat IgGs and unbound 5-nm gold were separated from phages by spin chromatography as described above before negative staining with 2% uranyl acetate. Electron microscopy was performed using a Tecnai G2 Spirit equipped with an Eagle charge-coupled device (CCD) camera (FEI).

**Crystallization, data collection, and processing.** Recombinant pb9 protein was concentrated to 10 mg/ml using an Amicon Ultra 10-kDa concentrator. The final concentration was determined by UV spectroscopy with E₅₄₀ = 1.181 and 1.154 (mg/ml)^⁻¹ cm⁻¹ for pb9-Cter and pb9-Nter, respectively. The first crystallization screening for the two constructs was carried out using commercial screens (Noirclerc-Savoye et al., submitted). The sitting drops, consisting of 100 nl protein and 100 nl crystallization buffer, were dispensed in 96-well plates (Greiner Crystal Quick plates) using a Cartesian PIXSYS 4200 robot (Genomic Solutions) and equilibrated at 20°C against 100 μl of crystallization buffer. Hits were then manually reproduced and improved using the vapor diffusion hanging drop technique. The drops, consisting of 0.8 μl protein and 0.8 μl crystallization buffer, were equilibrated against 250 μl of crystallization buffer at 20°C in 48-well plates (Hampton Research). Crystals were trans- ferred to the crystallization buffer supplemented with 20% (vol/vol) glycerol for 30 s, flash-cooled, and stored in liquid nitrogen. A lanthanide derivative was obtained by soaking a pb9-Nter pentagonal crystal for 5 min in a solution containing the crystallization buffer supplemented with 100 mM [Na₃][Eu(DPA)] (16). This crystal was back soaked for 30 s in 20% (vol/vol) glycerol-containing crystallization buffer and flash-cooled in liquid nitrogen. Diffraction data were collected at 100 K. A wavelength of 1.033 Å was used for collecting native data on the pb9-Cter crystals, and 1.776 Å, i.e., the L₃₄,₅ absorption edge of Eu as determined from an X-ray fluorescence scan, was used for collecting data on the Eu-soaked crystals. The total rotation angle was 360° for pb9-Cter crystals and 180° for the pb9-Nter derivative crystals, with 1° data frames. Reflections were in- dexed, integrated, and scaled with the XDS program suite (17). The high-resolution cutoffs were estimated according to ce1/2 (18). Data statistics and parameters are summarized in Table 1.

**Structure solution and refinement.** Due to crystal sensitivity to radiation damage, a single-wavelength anomalous diffraction data set was estimated with Phenix xtriage. The data set was detwinned using the CCP4 program Detwin (20), and the phasing procedure (including solvent flattening) was repeated, leading to an average FOM of 0.252. In the resulting electron density map, density for two helices appeared in which a poly(Ala) partial model was built using Coot (21). Assuming from the secondary structure prediction that there should be a single long α-helix per monomer, the two helices were used to determine an initial noncrystallographic symmetry (NCS) operator using Find-NCS from Phenix. The solvent-flattened map was subjected to iterative 2-fold NCS averaging using the CCP4 program DM, providing an average FOM of 0.321 with density corresponding to two β-strands appearing in the averaged map. An iterative "bootstrapping" procedure was then used (22): refinement of the NCS operator, 2-fold NCS averaging, partial model rebuilding, and phase combination using partial model phases and heavy-atom phases. This led to an average FOM of 0.659. The resulting poly(Ala) model was used for molecular replacement calculations using the triclinic data set at 1.89 Å resolution with Phaser (23), in which four monomers were positioned with a translation function Z-score of 11.6. Iterative 4-fold NCS averaging was used to improve the electron density, which was then subjected to Phenix’s AutoBuild. An initial model consisting of 684 residues in four chains and 621 water molecules was obtained (Rwork = 0.25 and Rfree = 0.29). Model completion was done with sessions of model rebuild- ing using Coot interspersed with model refinement with Phenix, using the TLSMD web server for the generation of multigroup TLS models (24).

Hexamer pb9 was modeled by structurally aligning six pb9 monomers (domain A only) onto the six ring-forming molecules of p2 Dint, using the Dalilite server (25). Electrostatic surface potential calculations were performed using APBS (26) with the AMBER force field.

**Protein structure accession number.** Coordinates and structure fac- tors have been deposited with the Protein Data Bank as entry 4JMQ.

<table>
<thead>
<tr>
<th>TABLE 1 Crystallographic data and phasing and refinement statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection: pb9-Nter Eu derivative</td>
</tr>
<tr>
<td>Space group</td>
</tr>
<tr>
<td>Cell parameters (Å, °)</td>
</tr>
<tr>
<td>Beam line</td>
</tr>
<tr>
<td>Detector</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
</tr>
<tr>
<td>No. of observed (unique) reflections</td>
</tr>
<tr>
<td>Multiplicity</td>
</tr>
<tr>
<td>Completeness (%)</td>
</tr>
<tr>
<td>Rmerge</td>
</tr>
<tr>
<td>Polya(A) partial model</td>
</tr>
<tr>
<td>Molecular replacement phasing</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Refinement statistics</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
RESULTS

**pb9 immunolocalization.** Antibodies raised against pb9 were used to immunolocalize the protein within the phage structure. pb9 is located in the upper part of the cone, right under the collar onto which are grafted the L-fibers, as attested by the cross-linking of T5 bacteriophages when incubated with anti-pb9 IgG (Fig. 1C). Labeling specificity was confirmed by goat anti-rabbit IgG–gold conjugate (Fig. 1D). Immunolocalization was also performed on T5hd1, a T5 mutant lacking the L-shaped fibers and the associated collar, which allowed a better sighting of the cross-linking (Fig. 1C and D).

**pb9 characterization and crystallization.** Overproduction experiments yielded 150 and 180 mg per liter of culture of purified pb9-Nter and pb9-Cter, respectively. Purified proteins exhibited molecular masses of 23,728 Da for pb9-Nter and 24,454 Da for pb9-Cter, as determined by mass spectrometry, in complete agreement with the theoretical masses of 23,732 and 24,452 Da, respectively. Both proteins were >99% pure and mainly monomeric (ca. ~95%) in solution, as determined by size exclusion chromatography coupled to multilangle light scattering (Noirclerc-Savoye et al., submitted). However, regardless of the concentration, a small and constant part of dimer was always present. Plate-shaped crystals (400 by 400 by 30 μm2) were obtained for pb9-Cter with 10 to 14% (wt/vol) polyethylene glycol 3350 (PEG 3350), 0.05 M morpholineethanesulfonic acid (MES) (pH 6.0), and 0.05 to 0.15 M NaCl. PEG 5000 mono methyl ether, 0.05 MES (pH 6.0), and 0.05 to 0.15 M NaCl.

**Structure of pb9, the Distal Tail Protein of Phage T5.**

(A) Ribbon representation of the pb9 monomer. Domains A and B are colored in red and blue, respectively. The two linkers are colored in green. The N and C termini are labeled. (B) Topological diagram of pb9, using the same color code as in panel A. The missing unstructured loop is represented as a dashed line.

*Fig. 2* (A) Ribbon representation of the pb9 monomer. Domains A and B are colored in red and blue, respectively. The two linkers are colored in green. The N and C termini are labeled. (B) Topological diagram of pb9, using the same color code as in panel A. The missing unstructured loop is represented as a dashed line.
therefore, that the “OB-fold” in fact consists of several architectural classes, all of which are based on \( /H9252 \)-barrels. This fold is known to bind oligonucleotides or oligosaccharides (27). No interaction of purified pb9 with the DNA of T5 could be detected by electrophoretic mobility shift assay (data not shown).

Sequence homologues of pb9. We have shown that pb9 shares the same fold as the Dit proteins of siphophages infecting Gram-positive bacteria. Is this feature extendable to all siphophages infecting Gram-negative bacteria? A PSI-BLAST search with 4 iterations links pb9 to phage proteins of T5-related phages H8, EPS7, and SCP35; \textit{Vibrio} phages pVp-1, SSP002, My1, and AKFV33; and numerous siphophages, including \textit{Yersinia} phages Phi201 and PY54, EBPR siphovirus1, and \textit{Rhizobium} phage 16-3, as well as \textit{Salmonella}, \textit{Citrobacter}, and \textit{Shigella} phages, and the myophages EcoM-FV3 and EcoM-VR5 (Fig. 4), and to many “hypothetical phage tail proteins” identified in the genomes of Gram-negative bacteria. Searches through HHpred also links residues 35 to 85 of pb9 to a family of “phage minor tail proteins,” represented by the gpM protein of phage \( /H9252 \) (Fig. 4; probability 70.4%). Sequence similarity between gpM and pb9 in its N terminus would suggest that the two proteins share a similar fold. However, gpM is only 109 residues long, whereas pb9 is composed of 204 residues. Sequence alignment based on secondary structure prediction shows that the gpM C terminus aligns well with the C termini of pb9 and other Dit proteins (Fig. 4). Thus, gpM would be composed of a unique domain that would share the domain A fold, and domain B would be absent in gpM-like proteins.

DISCUSSION

pb9, the Dit protein of bacteriophage T5. Topological and structural evidence indicates that pb9 is the phage T5 Dit protein. Dit proteins provide a hub for assembly of the adsorption device of long-phage tails forming an open channel located between the tail tube and the host adsorption device. pb9 was immunolocalized in the upper part of the tail tip conical structure, just below the at-

---

**FIG 3** (A) Ribbon tracing of the superimposed domain A of pb9 (yellow) and the Dit N-terminal domains of bacteriophages p2 (green), TP901-1 (pink), and SPP1 (cyan). Overall C\( /H9252 \) RMSD is between 3.0 and 3.3 Å. (B) C\( /H9252 \) tracing of the superimposed domain B of pb9 (cyan) and domain II of SelB (green). RMSD between equivalent C\( /H9252 \) positions is 2.3 Å for 69 residues. (C) Ribbon tracing of the model of the hexamer of pb9 domain A (green), superimposed with a pb9 monomer, including domain B (blue). (D) Electrostatic potential at the surface of the homohexamer model of domain A of pb9. Domain B was removed from the set of coordinates to allow modeling of the ring. Red and blue colors correspond to the negative and positive potentials, respectively. (Left) Putative tail tube-facing negatively charged surface. (Middle) Putative straight fiber-facing positively charged surface (rotation of 180° relative to the left panel). (Right) Clipped view after a 90° rotation relative to the middle panel. The clipping allows viewing of the central channel and its overall negatively charged surface. (E) Ribbon representation of the model of domain A homohexamer (left) and pb9 monomer (right), colored according to the temperature factor of C\( /H9252 \) atoms, PyMol scale. The N and C termini are indicated, the asterisks indicate the last ordered residues from the disordered loop, and the black and white diamonds indicate the boundary residues between domain A and domain B in the left panel. Figures were generated with PyMol.
FIG 4 Sequence alignment of pb9 with proteins of the Myoviridae coliphage EcoM-VR5, the Siphoviridae Yersinia phages PhiR201 and PY54, Salmonella phage FSL SP-016, Rhizobium phage 16-3, EBPR siphovirus 1, and coliphage 1/H9261 and with phage proteins of the DUF2460 family identified in the genome of Gram-negative bacteria (YP_002518238, Caulobacter crescentus; YP_207656.1, Neisseria meningitidis). Domain B of pb9 is from residue 83 to residue 171. Sequence alignment of pb9 with proteins of the Myoviridae coliphage EcoM-VR5, the Siphoviridae Yersinia phages PhiR201 and PY54, Salmonella phage FSL SP-016, Rhizobium phage 16-3, EBPR siphovirus 1, and coliphage 1/H9261 and with phage proteins of the DUF2460 family identified in the genome of Gram-negative bacteria (YP_002518238, Caulobacter crescentus; YP_207656.1, Neisseria meningitidis), presented using ESPript (42). Secondary structures of pb9 are indicated (4JMQ). Domain B of pb9 is from residue 83 to residue 171. Sequence alignment of pb9 with proteins of the Myoviridae coliphage EcoM-VR5, the Siphoviridae Yersinia phages PhiR201 and PY54, Salmonella phage FSL SP-016, Rhizobium phage 16-3, EBPR siphovirus 1, and coliphage 1/H9261 and with phage proteins of the DUF2460 family identified in the genome of Gram-negative bacteria (YP_002518238, Caulobacter crescentus; YP_207656.1, Neisseria meningitidis), presented using ESPript (42). Secondary structures of pb9 are indicated (4JMQ). Domain B of pb9 is from residue 83 to residue 171. Sequence alignment of pb9 with proteins of the Myoviridae coliphage EcoM-VR5, the Siphoviridae Yersinia phages PhiR201 and PY54, Salmonella phage FSL SP-016, Rhizobium phage 16-3, EBPR siphovirus 1, and coliphage 1/H9261 and with phage proteins of the DUF2460 family identified in the genome of Gram-negative bacteria (YP_002518238, Caulobacter crescentus; YP_207656.1, Neisseria meningitidis), presented using ESPript (42). Secondary structures of pb9 are indicated (4JMQ). Domain B of pb9 is from residue 83 to residue 171. Sequence alignment of pb9 with proteins of the Myoviridae coliphage EcoM-VR5, the Siphoviridae Yersinia phages PhiR201 and PY54, Salmonella phage FSL SP-016, Rhizobium phage 16-3, EBPR siphovirus 1, and coliphage 1/H9261 and with phage proteins of the DUF2460 family identified in the genome of Gram-negative bacteria (YP_002518238, Caulobacter crescentus; YP_207656.1, Neisseria meningitidis), presented using ESPript (42). Secondary structures of pb9 are indicated (4JMQ). Domain B of pb9 is from residue 83 to residue 171. Sequence alignment of pb9 with proteins of the Myoviridae coliphage EcoM-VR5, the Siphoviridae Yersinia phages PhiR201 and PY54, Salmonella phage FSL SP-016, Rhizobium phage 16-3, EBPR siphovirus 1, and coliphage 1/H9261 and with phage proteins of the DUF2460 family identified in the genome of Gram-negative bacteria (YP_002518238, Caulobacter crescentus; YP_207656.1, Neisseria meningitidis), presented using ESPript (42). Secondary structures of pb9 are indicated (4JMQ). Domain B of pb9 is from residue 83 to residue 171. Sequence alignment of pb9 with proteins of the Myoviridae coliphage EcoM-VR5, the Siphoviridae Yersinia phages PhiR201 and PY54, Salmonella phage FSL SP-016, Rhizobium phage 16-3, EBPR siphovirus 1, and coliphage 1/H9261 and with phage proteins of the DUF2460 family identified in the genome of Gram-negative bacteria (YP_002518238, Caulobacter crescentus; YP_207656.1, Neisseria meningitidis), presented using ESPript (42). Secondary structures of pb9 are indicated (4JMQ). Domain B of pb9 is from residue 83 to residue 171. Sequence alignment of pb9 with proteins of the Myoviridae coliphage EcoM-VR5, the Siphoviridae Yersinia phages PhiR201 and PY54, Salmonella phage FSL SP-016, Rhizobium phage 16-3, EBPR siphovirus 1, and coliphage 1/H9261 and with phage proteins of the DUF2460 family identified in the genome of Gram-negative bacteria (YP_002518238, Caulobacter crescentus; YP_207656.1, Neisseria meningitidis), presented using ESPript (42). Secondary structures of pb9 are indicated (4JMQ). Domain B of pb9 is from residue 83 to residue 171. Sequence alignment of pb9 with proteins of the Myoviridae coliphage EcoM-VR5, the Siphoviridae Yersinia phages PhiR201 and PY54, Salmonella phage FSL SP-016, Rhizobium phage 16-3, EBPR siphovirus 1, and coliphage 1/H9261 and with phage proteins of the DUF2460 family identified in the genome of Gram-negative bacteria (YP_002518238, Caulobacter crescentus; YP_207656.1, Neisseria meningitidis), presented using ESPript (42). Secondary structures of pb9 are indicated (4JMQ). Domain B of pb9 is from residue 83 to residue 171. Sequence alignment of pb9 with proteins of the Myoviridae coliphage EcoM-VR5, the Siphoviridae Yersinia phages PhiR201 and PY54, Salmonella phage FSL SP-016, Rhizobium phage 16-3, EBPR siphovirus 1, and coliphage 1/H9261 and with phage proteins of the DUF2460 family identified in the genome of Gram-negative bacteria (YP_002518238, Caulobacter crescentus; YP_207656.1, Neisseria meningitidis), presented using ESPript (42). Secondary structures of pb9 are indicated (4JMQ). Domain B of pb9 is from residue 83 to residue 171. Sequence alignment of pb9 with proteins of the Myoviridae coliphage EcoM-VR5, the Siphoviridae Yersinia phages PhiR201 and PY54, Salmonella phage FSL SP-016, Rhizobium phage 16-3, EBPR siphovirus 1, and coliphage 1/H9261 and with phage proteins of the DUF2460 family identified in the genome of Gram-negative bacteria (YP_002518238, Caulobacter crescentus; YP_207656.1, Neisseria meningitidis), presented using ESPript (42). Secondary structures of pb9 are indicated (4JMQ). Domain B of pb9 is from residue 83 to residue 171. Sequence alignment of pb9 with proteins of the Myoviridae coliphage EcoM-VR5, the Siphoviridae Yersinia phages PhiR201 and PY54, Salmonella phage FSL SP-016, Rhizobium phage 16-3, EBPR siphovirus 1, and coliphage 1/H9261 and with phage proteins of the DUF2460 family identified in the genome of Gram-negative bacteria (YP_002518238, Caulobacter crescentus; YP_207656.1, Neisseria meningitidis), presented using ESPript (42). Secondary structures of pb9 are indicated (4JMQ). Domain B of pb9 is from residue 83 to residue 171. Sequence alignment of pb9 with proteins of the Myoviridae coliphage EcoM-VR5, the Siphoviridae Yersinia phages PhiR201 and PY54, Salmonella phage FSL SP-016, Rhizobium phage 16-3, EBPR siphovirus 1, and coliphage 1/H9261 and with phage proteins of the DUF2460 family identified in the genome of Gram-negative bacteria (YP_002518238, Caulobacter crescentus; YP_207656.1, Neisseria meningitidis), presented using ESPript (42). Secondary structures of pb9 are indicated (4JMQ). Domain B of pb9 is from residue 83 to residue 171. Sequence alignment of pb9 with proteins of the Myoviridae coliphage EcoM-VR5, the Siphoviridae Yersinia phages PhiR201 and PY54, Salmonella phage FSL SP-016, Rhizobium phage 16-3, EBPR siphovirus 1, and coliphage 1/H9261 and with phage proteins of the DUF2460 family identified in the genome of Gram-negative bacteria (YP_002518238, Caulobacter crescentus; YP_207656.1, Neisseria meningitidis), presented using ESPript (42). Secondary structures of pb9 are indicated (4JMQ). Domain B of pb9 is from residue 83 to residue 171. Sequence alignment of pb9 with proteins of the Myoviridae coliphage EcoM-VR5, the Siphoviridae Yersinia phages PhiR201 and PY54, Salmonella phage FSL SP-016, Rhizobium phage 16-3, EBPR siphovirus 1, and coliphage 1/H9261 and with phage proteins of the DUF2460 family identified in the genome of Gram-negative bacteria (YP_002518238, Caulobacter crescentus; YP_207656.1, Neisseria meningitidis), presented using ESPript (42). Secondary structures of pb9 are indicated (4JMQ). Domain B of pb9 is from residue 83 to residue 171. Sequence alignment of pb9 with proteins of the Myoviridae coliphage EcoM-VR5, the Siphoviridae Yersinia phages PhiR201 and PY54, Salmonella phage FSL SP-016, Rhizobium phage 16-3, EBPR siphoviru
tachment point of the L-shaped fibers (Fig. 1). Whereas sequence similarity is poor within phage proteins, the arrangement of structural genes within the genomes is remarkably conserved (11, 28).

The comprehensive analysis of T5 structural genes shows that the pb9 gene is located downstream of the pb2 gene encoding the tape measure protein (TMP) of T5, and upstream of the pb3 gene encoding a large protein that forms the bottom of the cone, connecting with the straight fiber (Fig. 1B) (11). We proposed pb3 to be the Tal/baseplate hub protein (BHP) of T5 (11), as this protein is predicted to adopt the same fold as gp27 protein of phage T4, the BHP that connects the tail tube and the central cell-puncturing device in myophages. A similar gene organization has been observed in siphophages infecting Gram-positive bacteria (Fig. 1B), where the Dit coding gene is located between the TMP and the Tal/BHP/gp27-like coding gene. Finally, domain A of pb9 shows striking structural similarity with the N-domain of Dit proteins ORF15, ORF46, and gp19.1, of p2, TP901-1, and SPP1, respectively.

A model of the hexamer of domain A of pb9 could be built, by homology with that of the hexamer of the N-domain of p2 (Fig. 3C and D). The modeled homohexameric ring of pb9 domain A delineates a wide central channel of ~45 Å in diameter. This is consistent with the diameter of the internal channel of the tail tube of phage T5, estimated to be ca. 50 Å (13), and would allow the passage of the DNA. The surface of the internal channel of the modeled pb9 domain A ring displays a strong negative electrostatic potential, due to the abundance of acidic residues (Fig. 3D, right panel). This is also observed in the other Dit rings and would ease DNA transfer through the tail (8). This characteristic is often observed in phage proteins that channel DNA during infection (e.g., gp6 and gp16 of the SPP1 head-to-tail connector [29] and the tail terminator of phage λ [30]). The pb9 domain A ring has two oppositely charged surfaces: its putative tail tube-facing surface displays a completely negative electrostatic surface (Fig. 3D, left panel), whereas the putative straight fiber-facing surface is mainly positively charged (Fig. 3D, middle panel). This suggests that within the T5 tail, pb9 interacts with its partners via strong electrostatic potential complementarities. Such a situation has been described for the gp15 and gp16 dodecamers that form the SPP1 head-to-tail connector (29) and the tail terminator of phage λ (30).

The Dali search that we performed also revealed a noteworthy structural similarity between pb9 and tail proteins from other bacteriophages. The 3D structure of domain A is similar to those of the N-domain of the major tail protein gpV (PDB 2K4Q, Z-score = 4.3) and of the tail terminator protein gpU (PDB 3F2Z, Z-score = 3.9) of phage λ, but also to those of the HD1 domain of the BHP tail protein from *Shewanella oneidensis* MR-1 prophage MuSO2 (PDB 3CDD, Z-score = 4.9) and of type VI secretion system proteins EPVC of *Edwardsiella tarda* (PDB 3EAA, Z-score = 3.9) and Hcp3 (PDB 3EH1, Z-score = 3.9) of *Pseudomonas aeruginosa*. Except for BHP of MuSO2, which forms a trimer, these proteins form hexameric rings similar in fold and dimensions to those formed by Dit proteins of known 3D structure. These observations further support the widely accepted idea that long-tailed phages share an ancestor and that structural tail proteins evolved from a unique ancestral protein module (8, 34). It also adds one more brick to the growing wall of evidence showing that type VI secretion system and phage tails are evolutionarily connected.

**Gram-negative and Gram-positive-bacterium-infecting siphophages: separate but similar evolutionary pathways for Dit proteins.** Sequence alignments that relate pb9 to distant phages bring evidence that the Dit structural motif is conserved among all
bacteriophages belonging to the *Siphoviridae* family. The Dit protein, together with the Tal protein, was suggested to be the nucleating complex of the phage tail assembly (36). The major difference between the Dit proteins of Gram-negative-bacterium-infecting siphophages and those of Gram-positive-bacterium-infecting siphophages is the presence of a different additional domain in the two classes of proteins. The galectin-fold domain is present at the C terminus of the N-domain in the Dit protein of Gram-positive-bacterium-infecting siphophages, while domain B of pb9, inserted in a loop of domain A, adopts an OB-like fold. We also note that in some Gram-positive-bacterium-infecting siphophages, Dit proteins bear an additional large C-terminal extension of unknown structure and function (8). A common feature between the galectin- and the OB-like domains is their putative oligosaccharide-binding characteristic (37, 38). It is interesting that an OB-fold domain is also observed and is a conserved feature of the central spikes of myophages and type VI secretion systems (35, 39). Such an oligosaccharide-binding domain seems, however, to be absent in the gpM proteins of lambdoid phages. From an evolutionary point of view, the presence of one-domain Dit proteins and of Dit proteins bearing two different domains inserted at different locations of the protein 3D structure would argue in favor of the ancestor Dit protein being formed of the main domain building block, which can be elaborated with new domains, inserted at different positions in the protein, for additional functions. The acquisition of additional domains would result from horizontal transfer, presumably host specific, their saccharide-binding property enhancing cell adhesion, as was previously suggested for the galectin domain of the Dit protein SPPI (8). These additional functions could provide adaptation to specific surface sugars of different host cells. Saccharide-binding properties have also been shown to be important in the symbiosis that relates phages to metazoan host mucus (40). In the case of Gram-positive-bacterium-infecting phages, the galectin domain would have further evolved as a “hub” to accommodate a more complex baseplate and a higher number of saccharide-RBPs (5, 6). From an evolutionary point of view, the presence of multiple RBPs. This building block domain bears a spike and/or multiple RBPs. This building block domain can be added new domains for additional functions, which would provide adaptation to different host types (additional saccharide-binding domain and/or hub to a more complex baseplate structure).

**ACKNOWLEDGMENTS**

This work used the RoBioMol, Mass Spectrometry, Protein Analysis On Line, and High Throughput crystallization platforms of the Grenoble Instruct center (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). The research leading to these results has received funding from the European Community’s Seventh Framework Program (FP7/2007-2013) under grant agreement 211800 (SBMP) and from Finovi.

We thank Christine Ebel for constant support, Monika Spano and Jacques-Philippe Colletier for very useful crystallization and crystallography advice, Paulo Tavares for fruitful discussions and critical reading of the manuscript, and Knut Heller for providing the T5hd1 mutant. We also thank the ESRF and SOLEIL for BAG beam time on ID23-1, ID23-2, ID14-4, BM30, and Proxima1.

**REFERENCES**


Downloaded from http://j.asm.org on November 13, 2017 by guest