Crystal structure of pb9, the distal tail protein of bacteriophage T5: A conserved structural motif among all siphophages

Ali Flayhan,a,b,c,1 Frédéric M. D. Vellieux,a,b,c Rudi Lurz,d Olivier Maury,e Carlos Contreras-Martel,a,b,c Eric Girard,a,b,c Pascale Boulanger,f and Cécile Breytona,b,c#

aUniv. Grenoble Alpes, Institut de Biologie Structurale (IBS), Grenoble, France.
bCNRS, UMR5075, IBS, F-38027 Grenoble, France.
cCEA, DSV, IBS, F-38027 Grenoble, France.
dMax Planck Institute for Molecular Genetics, Berlin, Germany
eUniversité de Lyon, Ecole Normale Supérieure de Lyon, Laboratoire de Chimie, UMR CNRS 5182, Lyon, France.
fInstitut de Biochimie et de Biophysique Moléculaire et Cellulaire, Université Paris-Sud, UMR CNRS 8619, Orsay, France.

RUNNING TITLE: Structure of pb9, the Distal tail protein of phage T5

KEYWORDS: Bacteriophage, crystal structure, Siphoviridae, Gram-negative bacteria, Dit protein

Abstract

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 infecting siphophages, the Distal tail (Dit) protein has been structurally characterised, and is proposed to represent a baseplate hub docking structure. It is organised as a hexameric ring that connects the tail tube and the adsorption device. In this study we report the characterisation of pb9, a tail tip protein of Escherichia coli bacteriophage T5. By immunolocalisation we show that pb9 is located in the upper cone of 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 infecting siphophages. 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.

**INTRODUCTION**

The order *Caudovirales*, tailed bacteriophages, comprises the vast majority (>95%) of bacteriophages. They all have in common a proteinaceous capsid enclosing the genome consisting of double-stranded DNA and a tail (1). Three families are distinguished by the morphology of their tail: *Myoviridae* (long contractile tail), *Podoviridae* (short non-contractile tail) and *Siphoviridae* (long flexible non-contractile tail). Tails of bacteriophages are complex supramolecular assemblies that specifically recognise the target bacteria, via the host absorption device located at the distal end of the tail tube, and efficiently deliver the genome into the cytoplasm of the cell. The infection process is initiated by the interaction between the Receptor Binding Proteins (RBPs) and their receptors at the host cell surface, leading ultimately to the injection of the phage DNA into the cytoplasm of the bacterium (reviewed in 2–4).

In siphophages, high-resolution structures of the host adsorption device of the *Lactococcus* phages p2 (5) and TP901.1 (6), and part of it for the *Bacillus subtilis* phage SPP1 (7, 8) are available: they are made up of a complex baseplate containing multiple copies of the saccharide-binding RBP (18 and 54 for phages p2 and TP901-1, respectively), or a tail spike containing three copies of the protein-binding RBP (SPP1) (9). In these bacteriophages, which infect Gram-positive bacteria, a common docking hub between the tail tube and the tail adsorption device is the Dit-Tal complex (8). The Dit protein (Distal tail protein) is composed of two domains, one of which forms an open hexameric ring at the extremity of the tail tube. The second, galectin-like domain was proposed to bear saccharide-binding properties in SPP1 (8), and serves as a platform for the attachment of the RBPs (5, 6). The trimeric Tal protein acts as a closing plug. In SPP1, binding of the RBPs to its receptor triggers a cascade of conformational changes that are transmitted along the tail to the capsid, allowing its opening (10) as well as tail-tip reorganisation and opening of the Tal trimer (5, 7). Perforation of the host cell envelope and the transfer of the genome into the host cytoplasm are mechanisms that remain poorly understood.
No structural information is as yet available for the adsorption device of siphophages infecting Gram-negative bacteria. In this context, the Siphoviridae coliphage T5 is as a very suitable model: its tail tip is composed of a limited number of proteins (11)(Fig. 1A,B), and its protein receptor has been identified as FhuA, the outer-membrane iron-ferrichrome transporter (12). Phage T5 adsorption device contains three L-shaped fibres attached to a conical structure that is extended by a straight fibre, at the tip of which is located only one copy of the RBP (Fig. 1A)(11). The overall structure of T5, as determined by electron cryo-microscopy at resolutions of 20Å for the capsid and 30Å for the tail tube, is available (13), and the analysis of the tail structural genes allowed the identification of all tail proteins (11)(Fig. 1B). In this study, we report the crystal structure of pb9, a tail protein encoded by a gene whose position within the tail morphogenesis gene cluster is the landmark of the Dit protein gene (11). We localised 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 hexamerisation 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, that can be combined with a more divergent domain that serve specific adhesion and/or hub properties.

**MATERIALS AND METHODS**

*Cloning, overexpression and purification*–The DNA sequence (GenBank accession AAU05274.1) coding for the tail protein pb9 was cloned in the pLIM14 (His\textsubscript{6}-Nter fusion) or pLIM13 (His\textsubscript{6}-Cter fusion) vectors (14). A tobacco etch virus protease cleavage site was inserted between the His\textsubscript{6}-Cter and pb9. Positive plasmids were transformed into chemically competent E. coli BL21(DE3) expression strain. Transformed cells were cultured for 72 hours at 28°C in an auto-induction medium, supplemented with 50 µg/ml kanamycin. Cells were harvested and stored at -80°C. The frozen pellet (~11 g) was resuspended in 30 ml lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl and 2 mM MgSO\textsubscript{4}), supplemented with 50 µl DNAse (3 U/µL) and a cocktail of protease inhibitors (EDTA-Free Roche). The cells were then broken with a microfluidizer at 14000 PSI and centrifuged for 20 min at 55,000 rpm in a 70Ti 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 to 0.5 M Imidazole gradient. Pb9 containing fractions were pooled and loaded on an anion exchange column (HiTrap Q, 5 mL, GE Healthcare) previously equilibrated with 20 mM Tris pH 8.0. Pb9 was eluted by a 0 to 0.2 M NaCl gradient. 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. Monomer-containing fractions were pooled and desalted (HiTrap desalting column 5ml, GE Healthcare). Both pb9 constructs, with a hexa-histidine tag at either the N-terminus (pb9-Nter) or at the C-terminus (pb9-Cter), were purified in the same manner and exhibited the same behaviour during purification. Monomer containing fractions were used for crystallisation. Rabbit immunisation against pb9-Nter was carried out according to standard protocols. The antiserum was depleted from \textit{E. coli} antibodies by incubation with an \textit{E. coli} cell lysate, and IgGs were purified by affinity chromatography using a HiTrap Protein A column as recommended by the supplier (GE Healthcare).

\textbf{Immuno-electron microscopy}—1 µL of phages 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 MgCl\textsubscript{2} and 1 mM CaCl\textsubscript{2}). The mixture was incubated from 1h to overnight 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 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 labelled 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 stain with 2% uranyl acetate. Electron microscopy was performed using a Tecnai G2 Spirit equipped with an Eagle CCD camera (FEI).

\textbf{Crystallisation, 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 $\varepsilon_{280nm} = 1.181$ and 1.154 (mg/mL)$^{-1}$cm$^{-1}$ for pb9-Cter and pb9-Nter, respectively. The first crystallisation screening for the two constructs was carried out using commercial screens. The sitting drops, consisting of 100 nL protein and 100 nL crystallisation buffer, were dispensed in 96 wells plates (Greiner Crystal Quick plates) using a Cartesian PIXSYS 4200 robot (Genomic Solutions) and equilibrated at 20°C against 100 µL of crystallisation buffer. Hits were then manually reproduced and improved using the vapour diffusion hanging drop technique. The drops, consisting of 0.8 µL protein and 0.8 µL crystallisation buffer, were equilibrated against 250 µL of crystallisation buffer at 20°C in 48 well plates (Hampton Research). Crystals were transferred to the crystallisation buffer supplemented with 20% (v/v) glycerol for 30 seconds, flash-cooled and stored in liquid nitrogen. A lanthanide derivative was obtained by soaking a pb9-Nter pentagonal crystal for 5 minutes in a solution containing the crystallisation buffer supplemented with 100 mM $[\text{Na}_3][\text{Eu(DPA)_3}]$ (16). This crystal was back soaked for 30 seconds in 20% (v/v) glycerol-containing crystallisation 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 L$_{III}$ 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 indexed, integrated and scaled with the XDS program suite (17). The high-resolution cut-offs were estimated according to cc1/2 (18). Data statistics and parameters are summarised in table 1.

Structure solution and refinement—Due to crystal sensitivity to radiation damage, a single wavelength anomalous diffraction data set was recorded for the Eu-derivative at the L$_{III}$ absorption edge wavelength of 1.776 Å. The initial heavy-atom sites were located using AutoSol from the PHENIX program suite (19). The initial phases after solvent flattening had a low average figure of merit (FOM) of 0.151. A re-examination of the data indicated that the crystal was twinned. A twinning fraction of 0.3 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 Non-Crystallographic 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; 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 (R_{work} = 0.25 and R_{free} = 0.29). Model completion was done with sessions of model rebuilding using COOT interspersed with model refinement with PHENIX, using the TLSMD web server for the generation of multi-group TLS models (24).

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

RESULTS

Pb9 immunolocalisation–Antibodies raised against pb9 were used to immunolocalise 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-fibres, as attested by the cross-linking of T5 bacteriophages when incubated with anti-pb9 IgG (Fig. 1C). Labelling specificity was confirmed by goat anti-rabbit IgG gold conjugate (Fig. 1D). Immunolocalisation was also performed on T5hd1, a T5 mutant lacking the L-shaped fibres and the associated collar, which allowed a better sighting of the cross-linking (Fig. 1C,D).

Pb9 characterisation and crystallisation–Over-production experiments yielded 150 and 180 mg per L of culture of purified pb9-Nter and -Cter, respectively. Purified
proteins exhibited molecular masses of 23728 Da for pb9-Nter and 24454 Da for pb9-Cter, as determined by mass spectrometry, in complete agreement with the theoretical masses of 23732 and 24452 Da, respectively. Both proteins were > 99% pure and mainly monomeric (ca. ~95%) in solution, as determined by size exclusion chromatography coupled to multi-angle light scattering. However, regardless of the concentration, a small and constant proportion (ca. ~5%) of dimer was always present. Plate shaped crystals (400 x 400 x 30 µm³) were obtained for pb9-Cter with 10 to 14% (w/v) PEG3350, 0.05 M MES pH 6.0 and 0.1 to 0.2 M MgCl₂ and pentagonal crystals (200 x 50 x 20 µm³) for pb9-Nter with 5 to 8 % (w/v) PEG5000 MME, 0.05 MES pH 6.0 and 0.05 to 0.15 M NaCl.

Structure of pb9, a two-domain protein–Native and derivative pb9-Nter crystals belong to space group P3₂₁2₁ with two molecules in the asymmetric unit. Pb9-Cter crystals belong to space group P1 and the asymmetric unit contains four monomers (1 to 4) with an overall root mean square deviation (r.m.s.d.) between different monomers ranging from 0.185 to 0.405 Å. The pb9-Cter model was refined at 1.89 Å resolution, with R_work and R_free of 0.205 and 0.254, respectively. The C-terminal His-tag was seen in the electron density of monomer 3, where it is located in a crystal contact region. Residues 28-42 of monomer 1, 27-42 of monomer 2 and 29-44 of monomer 3, could not be seen in the electron density, and most likely form a flexible and unstructured loop within the crystal.

Pb9 is composed of two domains, named A and B. Domain A (residues 1-82/172-205) adopts a split barrel-like fold (SCOP:50475) and is formed of one α-helix, two helical turns and a five-stranded anti-parallel twisted β-sheet (β₁.A to β₅.A; Fig. 2). Domain B is a small five-stranded open β-barrel (residues 90-169, β₁.B to β₅.B, Fig. 2). Two small anti-parallel β strands, β₁ and β₂, and one helical turn are located on one side of the barrel without obstructing it (Fig. 2). Domain B belongs to the Reductase/isomerase/elongation factor common domain (R/I/EFCD) fold (SCOP:40512). Domain B is inserted in a loop of domain A, connecting β₃.A to β₁ and β₅B to β₄A.

Structural homologues of the two pb9 domains–A DALI search revealed that despite a low sequence identity (< 12 %), domain A of pb9 exhibits remarkable structural
similarity with the N-terminal domain (N-domain) of Dit proteins ORF15 of p2 (PDB 2WZP, Z-score = 6.1), ORF46 of TP901.1 (PDB 4DIV, Z-score = 5.6) and gp19.1 of SPP1 (PDB 2X8K, Z-score = 5.0) (Fig. 3A). These proteins have been shown to form a hexameric ring that occupies the central core of the baseplate (5–8). The missing unstructured loop (residues 27-44) from the structure of pb9 corresponds, in the homologous structures, to a β-hairpin that ensures the connection between adjacent monomers within the hexameric ring. A model of the ring formed by domain A of pb9 could be obtained by structural superposition using the Dit ring of phage p2 as a template (Fig. 3C,D). Domain B however had to be removed, as steric hindrance between domain B and domain A of the neighbouring monomer occurred upon building of the hexamer with full-length pb9.

Domain B of pb9 does not share any structural homology with the galectin-like domain of other Dit proteins. Furthermore, unlike what is observed in previously determined Dit protein structures, where the galectin-like domain folds in C-terminus of the N-domain, domain B of pb9 is inserted in a loop of domain A. A DALI search points to a structural relationship between domain B and domain II of SelB (PDB 2WZP, Z-score = 6.4), a specialised translation elongation factor responsible for the co-translational incorporation of selenocysteine into proteins (Fig. 3B), and domain II of the LepA protein (PDB 3CB4, residues 189-281, Z-score 5.1). This latter domain adopts an oligonucleotide/oligosaccharide-binding fold (OB-fold). A search of the Protein Data Bank using the coordinates of the OB-fold domain of LepA indicates that it belongs to the R/I/EFCD family, i.e. not classified as an OB-fold by SCOP. In the SCOP database, OB-fold proteins are classified as belonging to several families. Our current view is therefore that the “OB-fold” in fact consists of several architectural classes, all of which based on β-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-relatives phages, H8, EPS7, SCP35, Vibrio phages pVp-1 and SSP002, My1 and AKFV33, and to numerous Siphophages including Yersinia phages Phi201 and PY54, EBPR siphovirus1,
Structure of pb9, the Distal tail protein of phage T5

279  Rhizobium phage 16-3, as well as Salmonella, Citrobacter and Shigella phages, and
280  the Myophages EcoM-FV3 and EcoM-VR5 (Fig. 4), and to many "hypothetical phage
281  tail proteins" identified in the genome of Gram-negative bacteria. Searches through
282  HHpred link pb9 (residues 24 to 203, i.e. on nearly all its length) to the family
283  DUF2460 of conserved hypothetical proteins found in phage-derived regions of
284  Gram-negative bacterial chromosomes (probability 96.4%), and including 4 tail
285  proteins of recognised prophages (Fig. 4). Thus, pb9 would be the representative of
286  a large family of Dit proteins found in numerous siphophages, but also in myophages,
287  infecting Gram-negative bacteria. Most interestingly, HHpred also links residues 35 to
288  85 of pb9 to a family of "phage minor tail protein", represented by the gpM protein of
289  phage λ (Fig. 4, probability 70.4%). Sequence similarity between gpM and pb9 in its
290  N-terminus would suggest that both proteins share a similar fold. However, gpM is
291  only 109 residues long, whereas pb9 is composed of 204 residues. Sequence
292  alignment based on secondary structure prediction shows that gpM C-terminus aligns
293  well with the C-termini of pb9 and other Dit proteins (Fig. 4). Thus, gpM would be
294  composed of a unique domain that would share Domain A fold, and domain B would
295  be absent in gpM-like proteins.

296  DISCUSSION

297  Pb9, the Dit of bacteriophage T5—Topological and structural evidences indicate
298  that pb9 is phage T5 Dit protein. Dit proteins provide a hub for assembly of the
299  adsorption device of long-phage tails forming an open channel located between the
tail tube and the host adsorption device. Pb9 was immuno-localised in the upper part
of the tail tip conical structure, just below the attachment point of the L-shaped fibres
(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 pb9 gene is located
downstream of pb2 gene encoding the tape measure protein (TMP) of T5, and
upstream of pb3 gene encoding a large protein that forms the bottom of the cone,
connecting with the straight fibre (11). We proposed pb3 to be the Tal/BHP protein of
T5 (11), as this protein is predicted to adopt the same fold as gp27 protein of phage
T4, the baseplate hub (BHP) protein that connects the tail tube and the central cell-
puncturing device in Myophages. A similar gene organisation 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,D). The modelled homo-hexameric
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 modelled 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 fibre-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
dodecamer that form the SPP1 head-to-tail connection (30), and for the binding of
TP901-1 RBP to the Upper baseplate protein BppU, where a negatively charged loop
of BppU is inserted in a positively charged crevice in the interacting surface of the
RBP (6).

Hexameric pb9 was neither evidenced in solution nor in crystal structures of
different constructions (His-tag on the C- or N-terminus or cleaved). As for pb9,
expression of isolated Dit genes TP901-1 orf46, p2 orf15 and Tuc2009 orf49 yielded
monomeric proteins in solution (31, 32), while only SPP1 gp19.1 was crystallised as a
dodecamer (two head-to-tail hexameric rings)(8). For the former phages,
hexamerisation of the Dit protein is induced by its interaction with the Tal trimer, as
evidenced by mass spectrometry (33) and crystallography (5, 31, 33). This is a
common phenomenon among phage proteins, where the oligomerisation of a protein
is regulated by the interaction with its partners (see e.g. 28). This pb9 structure is the
first one available for a monomeric Dit protein. As its domain A superimposes well with
the core of the N-domain of Dit hexamers of known 3D structure, it is very likely that
the core of the pb9 hexamerisation domain remains largely unchanged upon hexamer formation. The pb9 monomer contains a disordered loop (residues 28-43). The homologous loop of Dit proteins of known 3D structure is ordered and connects a neighbouring monomer within the hexameric Dit rings. Another common feature of phage structural proteins is the presence of flexible loops that probably prevent aberrant oligomerisation of individual proteins and promote the concerted assembly of the phage particle upon encountering suitable partners (see e.g. 29, 33). In this context, figure 3E shows the mean temperature factors of the Cα of pb9 domain A hexamer (left) and of the pb9 monomer (right). We note higher average temperature factors for the C-terminal end and loops that form the putative interacting surfaces with the tail tube (residues 203-205 and 69-78) and with the pb3 protein (residues 186-192, 55-60): these could become more ordered upon interaction with their respective partners. Another conformational change induced by the association with partners would be the displacement of domain B, which, in the pb9 monomer, would prevent spontaneous hexamer formation. Interaction with partners would displace domain B and enable oligomerisation. Interestingly, the linker between the two domains (residues 85-97) also has a higher than average temperature factor, suggesting that it is flexible and can be subjected to conformational changes (Fig. 3E, right).

Based on the extensive relatedness to Dit proteins, we propose that pb9 also adopts a hexameric quaternary fold (35). Dit hexamers interact directly with Tal proteins in the siphophage tail adsorption apparatus (5–7). The Tals of Gram-positive siphophages (5–7) and their homologue BHP gp27 of myophage T4 (36) assemble as trimers. It is thus reasonable to assume that phage T5 pb3 adopts the same fold and symmetry (11, 35). There would thus be a break in symmetry at the pb9-pb3 level, as in other siphophages. The T5 tail tube clearly displays 3-fold symmetry (13), unlike the tail tube of most other sipho- and myophages. There would thus be an additional symmetry break at this position of the tail of phage T5, from three monomers (Tail Tube Protein) to six (Dit)(Fig. 1A). We can rule out the possibility of pb9 forming a trimer, as the channel it would delineate would be too small to allow DNA passage.

The DALI search 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 24KQ, Z-score = 4.3) and of the tail terminator protein gpU (PDB 3FZ2, Z-score = 3.9) of phage λ, but
Structure of pb9, the Distal tail protein of phage T5

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 that 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 a common ancestor, and that structural tail proteins evolved from a unique ancestral protein module (8, 35). 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 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, were suggested to be the nucleating complex of the phage tail assembly (37). The major difference between the Dit proteins of Gram-negative infecting siphophages and those of Gram-positive 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 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 infecting siphophages, Dit proteins bare 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 (38, 39). It is interesting to note that an OB-fold domain is also observed and is a conserved feature of the central spikes of myophages and type VI secretion system (36, 40). 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 location of the protein 3D structure would argue in favour 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 SPP1 (8). These additional functions could
provide adaptation to specific surface sugars of different host cells. Saccharide-
binding properties have also shown to be important in the symbiosis that relates
phages to metazoan hosts mucus (41). In the case of Gram-positive 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). In lambdoid
phages, an oligosaccharide-binding domain may also be present as an independent
protein: indeed phage λ encodes three small proteins, gpL (42), gpK and gpl, which
genes are located between the putative Dit protein gpM and the central fibre/host
recognition protein gpJ, which we propose to be the Tal/BHP protein of λ (11) (Fig
1B). Both structure and function of the three former proteins are unknown, however,
they may posses an oligosaccharide-binding fold.

We conclude that all siphophages posses a homologous Dit building block
domain, which forms a hexameric ring connecting the tail tube to the adsorption
device including the straight fibre/spike and/or multiple RBPs. This building block
domain bears a flexible loop in the monomer, which becomes ordered upon formation
of the hexamer. Hexamerisation of the protein does not occur spontaneously and
would be induced by interaction with tail tip partners. To 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).

Protein structure accession number. Coordinates and structure factors have been
deposited with the Protein Data Bank as entry 4JMQ.

ACKNOWLEDGEMENTS:

This work used the RoBioMol, Mass spectrometry, Protein Analysis On Line
and High throughput crystallisation platforms of the Grenoble Instruct centre (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
agreements n° 211800 (SBMP) and from FINOVI.
We thank Christine Ebel for constant support, Monika Spano and Jacques-Philippe Colletier for very useful crystallisation 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 beamtime on ID23-1, ID23-2, ID14-4, BM30 and PROXIMA1.

REFERENCES:


Structure of pb9, the Distal tail protein of phage T5


Structure of pb9, the Distal tail protein of phage T5


Structure of pb9, the Distal tail protein of phage T5


LEGEND TO FIGURES

Figure 1: A. Schematic representation of the tail tip of phage T5 (see also 11). B. Arrangement of the tail tip genes in the Siphophages T5, TP901-1, p2, SPP1 and λ. Genes or part of them predicted to encode the same functions are depicted in the same gray tone (see also 11). C. and D. Localisation of pb9 in the upper part of the cone of the tail adsorption device in phages T5st0, a heat stable mutant, or T5hd1, devoid of L-shaped fibres (hd1). Phages were incubated with purified IgG raised against pb9 and observed by negative stain EM. The position of the protein was identified by IgG cross-linking (C) or localisation of the IgG molecules associated to goat anti-rabbit IgG gold conjugate (D). Fields of phages are shown (C), together with a gallery of blow-ups of the tail tips (C and D) to highlight the cross-linking, which are indicated with an arrow, and immuno-labelling. Isolated IgG molecules, giving the scale of the cross-linking distance, are circled. Diameter of the tail tube: 12 nm.

Figure 2: A. Ribbon representation of the pb9 monomer. Domain A and B are coloured in red and blue, respectively. The two linkers are coloured in green. The N and C termini are labelled. B. Topological diagram of pb9, using the same colour code as in A. The missing unstructured loop is represented as a dashed line.

Figure 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 Cα r.m.s.d. between 3.0 and 3.3 Å. B. Ca-tracing of the superimposed domain B of pb9 (cyan) and domain II of SelB (green). r.m.s.d. between equivalent Ca positions is 2.3 Å for 69 residues. C. Ribbon tracing of the model of the hexamer of pb9 domainA (green), superimposed with a pb9 monomer, including domain B (blue). D. Electrostatic potential at the surface of the homo-hexamer model of the domain A of pb9. Domain B was removed from the set of coordinates to allow modelling of the ring. Red and blue colours correspond to the negative and positive potentials, respectively. Left panel: the putative tail tube-facing negatively charged surface. Middle panel: the putative straight fibre-facing positively charged surface (rotation of 180° relative to the left panel). Right panel: clipped view after a 90° rotation relative to the middle panel. The clipping allows to view the central channel and its overall negatively charged surface. E. Ribbon representation of the model of the domain A homo-hexamer (left) and pb9 monomer (right), coloured according to the temperature factor of Ca atoms, PyMol scale. The N and C-termini are indicated, the stars indicate last ordered residues from the disordered loop, and the black and white diamond the boundary residues between domain A and domain B in the left pannel. Figures were generated with PyMol.

Figure 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 siphovirus1, and coliphage λ, and with phage proteins of the DUF2460 family identified in the genome of Gram negative bacteria (YP_002518238, Caulobacter crescentus; YP_207656.1, Neisseria gonorrhoeae; ZP_08868130, Azospirillum amazonense; ZP_07368836.1, Neisseria meningitides), presented using Espript (43). Secondary structures of pb9 are indicated (JMJ), and domain B of pb9 is underlined. Sequence alignment of pb9 with phage proteins (apart from λ) was performed by PSI-Blast.
and with phage members of the DUF2460 family and the N-terminus of λ-gpM, by HHPred (44). For λ-gpM, pairwise alignment with the domain A of pb9 was performed with PromalsS3D and the alignment of gpM C-terminus manually inserted in the HHpred alignment.
Table 1: Crystallographic data, phasing and refinement statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Pb9-Nter Eu-derivative</th>
<th>Pb9-Cter native</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>( P3_{2}1 )</td>
<td>( P1 )</td>
</tr>
<tr>
<td>Cell parameters (Å, °)</td>
<td>( a = b = 73.99 )</td>
<td>( a = 55.66, b = 70.03, c = 71.01 )</td>
</tr>
<tr>
<td>( c = 166.39 ) (Å)</td>
<td>( a = 91.05, \beta = 107.66, \gamma = 112.03 )</td>
<td></td>
</tr>
<tr>
<td>Beam line</td>
<td>ESRF ID23-1</td>
<td>ESRF ID23-1</td>
</tr>
<tr>
<td>Detector</td>
<td>ADSC Q315r CCD</td>
<td>ADSC Q315r CCD</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.776</td>
<td>1.033</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>63.80-3.00 (3.21-3.00)</td>
<td>66.9-1.89 (2.01-1.89)</td>
</tr>
<tr>
<td>Observed (unique) reflections</td>
<td>110205 (19783)</td>
<td>209072 (71120)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>5.57 (5.29)</td>
<td>2.94 (2.82)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.3 (96.0)</td>
<td>95.0 (83.3)</td>
</tr>
<tr>
<td>( R_{sym} )</td>
<td>0.134 (0.663)</td>
<td>0.159 (1.175)</td>
</tr>
<tr>
<td>( (I/\bar{I}) )</td>
<td>12.43 (3.46)</td>
<td>6.66 (1.09)</td>
</tr>
<tr>
<td>SAD phasing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range used</td>
<td>63.8-3.2</td>
<td></td>
</tr>
<tr>
<td>Eu sites</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Refined occupancies</td>
<td>0.82, 0.68</td>
<td></td>
</tr>
<tr>
<td>&lt;Figure of merit&gt;</td>
<td></td>
<td>0.252</td>
</tr>
<tr>
<td>(after detwinning)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poly-Ala partial model</td>
<td></td>
<td>0.71</td>
</tr>
<tr>
<td>Molecular replacement phasing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translation function Z-score</td>
<td>11.6</td>
<td></td>
</tr>
<tr>
<td>&lt;Figure of merit&gt;</td>
<td>0.255</td>
<td></td>
</tr>
<tr>
<td>Refinement statistics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range used (Å)</td>
<td>27.69-1.89 (1.96-1.89)</td>
<td></td>
</tr>
<tr>
<td>No. of reflections used</td>
<td>67405 (5652)</td>
<td></td>
</tr>
<tr>
<td>No. of reflections used for R-free</td>
<td>2826 (297)</td>
<td></td>
</tr>
<tr>
<td>( R )-factor</td>
<td>0.20 (0.33)</td>
<td></td>
</tr>
<tr>
<td>( R_{free} )</td>
<td>0.25 (0.39)</td>
<td></td>
</tr>
<tr>
<td>No. of protein/water atoms</td>
<td>6106/862</td>
<td></td>
</tr>
<tr>
<td>Wilson B-factor (Å²)</td>
<td>31.9</td>
<td></td>
</tr>
<tr>
<td>Average isotropic B-value</td>
<td>38.3/37.9/40.8</td>
<td></td>
</tr>
<tr>
<td>overall/protein/solvent (Å²)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ramachandran favoured/allowed (%)</td>
<td>98.0/1.9</td>
<td></td>
</tr>
<tr>
<td>RMS (bonds, Å; angles, °)</td>
<td>0.007/1.140</td>
<td></td>
</tr>
</tbody>
</table>