Structure of the receptor-binding carboxy-terminal domain of bacteriophage T7 tail fibers

Carmela Garcia-Doval and Mark J. van Raaij[1]

Department of Molecular Structure, Centro Nacional de Biotecnologia, Consejo Superior de Investigaciones Cientificas, E-28049 Madrid, Spain

Edited by F. William Studier, Brookhaven National Laboratory, Upton, NY, and approved April 30, 2012 (received for review November 30, 2011)

The six bacteriophage T7 tail fibers, homo-trimers of gene product 17, are thought to be responsible for the first specific, albeit reversible, attachment to *Escherichia coli* lipopolysaccharide. The protein trimer forms kinked fibers comprised of an amino-terminal tail-attachment domain, a slender shaft, and a carboxyl-terminal domain composed of several nodules. Previously, we expressed, purified, and crystallized a carboxyl-terminal fragment comprising residues 371–553. Here, we report the structure of this protein trimer, solved using anomalous diffraction and refined at 2 Å resolution. Amino acids 371–447 form a tapered pyramid with a triangular cross-section composed of interlocked β-sheets from each of the three chains. The triangular pyramid domain has three α-helices at its narrow end, which are connected to a carboxyl-terminal three-blade β-propeller tip domain by flexible loops. The monomers of this tip domain each contain an eight-stranded β-sandwich. The exact topology of the β-sandwich fold is novel, but similar to that of knob domains of other viral fibers and the phage Sf6 needle. Several host-range change mutants have been mapped to loops located on the top of this tip domain, suggesting that this surface of the tip domain interacts with receptors on the cell surface.

bacterial viruses | caudovirales | crystallography | infection | Podoviridae

Bacteriophages (bacterial viruses or phages) are important biological model systems and, because of the high specificity for their host bacteria, have found application in phage typing, food security, and phage therapy (1). *Escherichia coli* phage T7 is a member of the *Podoviridae* family of the *Caudovirales* (tailed phages) order (2). T7 is composed of an icosahedral capsid with a 20-nm short tail at one of the vertices (3, 4). The capsid is formed by the shell protein gene product (gp) 10 and encloses a DNA of 40 kb. A cylindrical structure composed of gp14, gp15, and gp16 is present inside the capsid (5), attached to the special vertex formed by the connector, a circular dodecamer of gp8 (6). Gp11 and gp12 form the tail; gp13, gp6.7, and gp7.3 have also been shown to be part of the virion and to be necessary for infection, although their location has not been established (7, 8). Although extensive electron microscopy studies have been performed on phage T7 (3–6, 9), crystallographic studies have so far been limited to its nonstructural proteins.

The main portion of the tail is composed of gp12, a large protein of which six copies are present (10); the small gp11 protein is also located in the tail (5). Attached to the tail are six fibers, each containing three copies of the gp17 protein. T7 tail fibers are elongated homo-trimers, which are responsible for initial, reversible, host cell recognition. A second, irreversible, decision-making interaction with the bacterial membrane is presumably mediated by one or more of the tail-tube proteins. DNA transfer into the host is then mediated by an extension formed by gp14–16 (7, 8, 11). Previously, we have reported the production of well-diffracting crystals of the phage-distal carboxyl-terminal domain of gp17 containing a 45-residue purification tag (12). Here we present its structure, solved using four-wavelength anomalous diffraction analysis of a mercury derivative and refined against data collected from crystals of two different forms.

Results

Structure Solution. Crystals of gp17(371–553) belonging to two different space groups were obtained, $P2_12_12_1$ and $C222_1$, and a multiwavelength anomalous diffraction dataset was collected on a crystal of form $P2_12_12_1$ derivatized with methylmercury chloride (12). The derivative was not isomorphous to the native crystal and the native dataset was not used in phase determination. Nine heavy atom sites were identified, six of which are located near the two cysteines present in each of the three chains (Cys408 and Cys499). Two sites are near Asp442 (of chains A and B) and one near His433 of chain A. After phasing and solvent flattening at 2.7 Å resolution, a readily interpretable map was obtained in which 527 residues could be automatically traced. This model was used as input in a molecular replacement with the 1.9 Å resolution dataset obtained of the $P2_12_12_1$ crystal form. Automatic tracing at this resolution produced a model containing 528 residues, which was completed and corrected manually and to which solvent atoms were added. An intermediate protein model was used to solve the $C222_1$ crystal form structure at 2.0 Å resolution, which was also completed. Both structures were refined to R-factors of 15% and free R-factors of 20% and have few residues in unlikely regions of the Ramachandran plot (Table S1). One amino acid, Gly522, is in a very uncommon conformation in all three chains of both structures (all have the backbone torsion angles ϕ around 165° and ψ around −90°); however, it has convincing electron density in each case and is thus likely forced into this conformation by the rest of the structure.

Overview of the Structure. Electron micrographs revealed gp17 to be an extended protein, with a proximal rod about 16-nm long and 2 nm in diameter, a sharp kink, and a distal rod about 15-nm long and with a diameter that varies between 3 and 5 nm (Fig. 1*A*) (3). The distal rod can be divided into four "nodules" of unequal size, which were estimated to contain residues 268–365, 366–432, 433–456, and 466–553, respectively. The crystal structure of the gp17(371–553) fragment corresponds to the most distal three of these nodules (Fig. 1). The structure can be divided into two parts: a globular "tip" domain (residues Ala465-Glu553), corresponding to the fourth and last nodule, and a tapered interlocked mainly β-structured pyramid domain (residues Gly371 to Trp454) which, from the electron microscopy analysis, was interpreted as forming the second and third nodules. The pyramid domain contains three short α-helices (one from each monomer) at its thinner end.

Author contributions: C.G.-D. and M.J.v.R. designed research; C.G.-D. and M.J.v.R. performed research; C.G.-D. contributed new reagents/analytic tools; C.G.-D. and M.J.v.R. analyzed data; and C.G.-D. and M.J.v.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Database deposition: Coordinates and structure factors for the $P2_12_12_1$ and $C222_1$ crystal forms have been submitted to the Protein Data Bank, www.pdb.org (PDB ID codes 4A0T and 4A0U, respectively).

[1]To whom correspondence should be addressed. E-mail: mjvanraaij@cnb.csic.es.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1119719109/-/DCSupplemental.

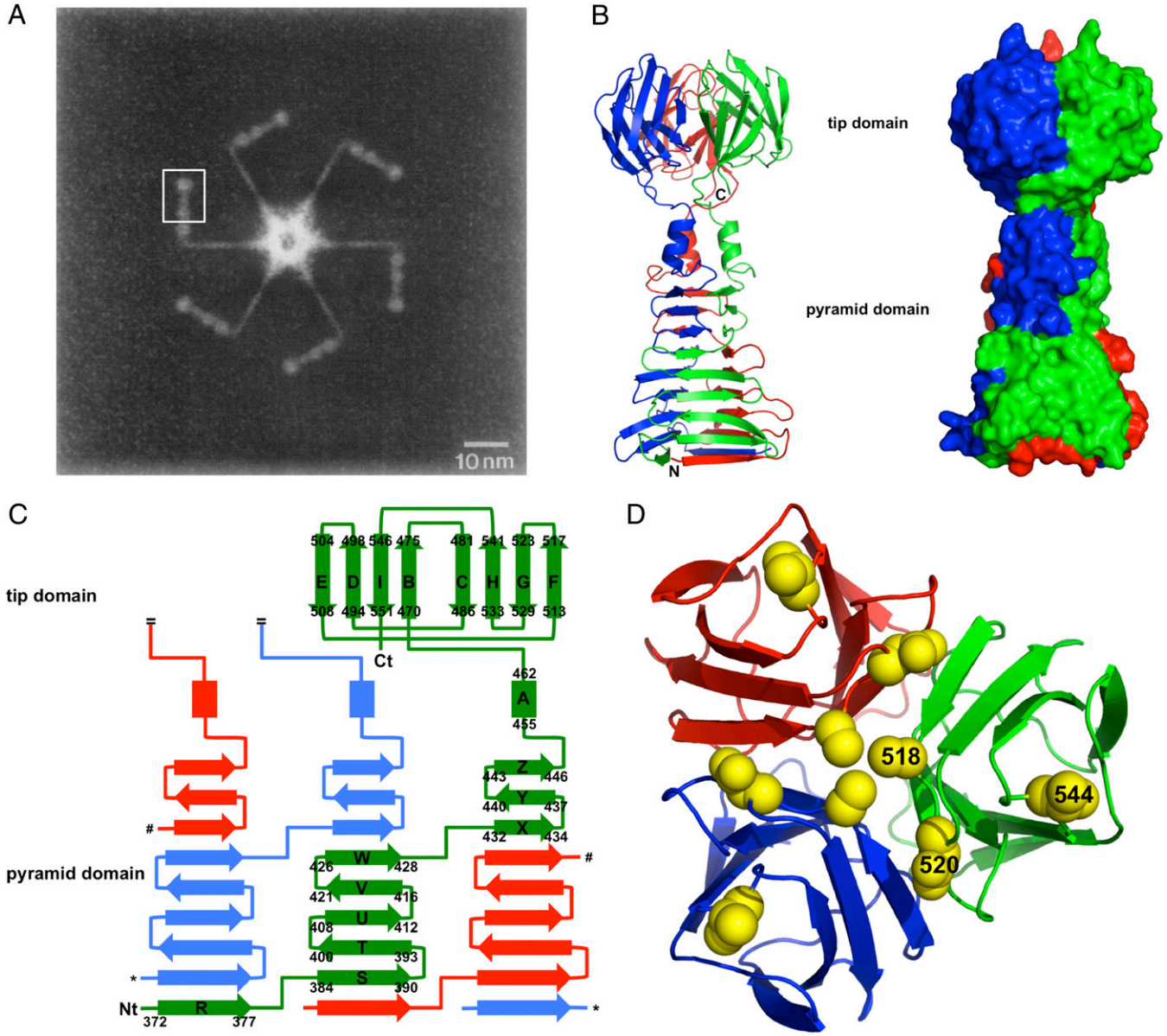

Fig. 1. Crystal structure of bacteriophage T7 gp17(371–553). (*A*) Composite negatively stained electron microscopy image of the tail-complex of bacteriophage T7. The region of one of the six symmetry-equivalent positions where our structure fits into this complex is boxed (reprinted from ref. 3, with permission from Elsevier, http://www.sciencedirect.com/science/journal/00222836). (*B*) Ribbon diagram (*Left*) and space-filled representation (*Right*) of gp17(371–553). The three monomers are colored red, green, and blue. The termini of the red monomer are indicated. (*C*) Topology diagram. β-Strands are shown as arrows, α-helices as rectangles. Secondary structure elements are labeled and their start and end residues are indicated for the green monomer, as are the termini. Only one of the three tip domain monomers is shown. Asterisks and hash signs indicate connections in the other monomers that were removed for clarity. (*D*) Top view of the tip domain with the side chains of residues Ala518, Asp520, and Val544 shown in space-filled representation (labeled in the green monomer only).

When the trimeric structures from the two crystal forms are superimposed, a rms difference of 1.2 Å between their C-α positions is observed (residues 373–553 of all chains were included in the calculation). Amino acids N-terminal to residue 373 show different conformations, probably as a result of the different crystal packing in the two structures. When individual chains are compared, another local difference becomes apparent in the region around β-turn Asp390-Arg393 in one of the chains, which forms a crystal contact in the $C222_1$ crystal form. The tip domains alone (amino acids 465–553) superpose with an rmsd of only 0.2 Å and the pyramid domain (amino acids 373–464) superposes with an rmsd of only 0.5 Å, suggesting the short loop (Val464-Lys466) between the tip and pyramid domains is somewhat flexible. This flexibility may be of importance in the infection process. Adenovirus fiber contains a hinge region between the shaft and head domains (13); mammalian reovirus fiber (σ1 protein) has a hinge region between the last and second-to-last triple β-spiral repeats of the shaft (14). In the case of adenovirus fiber, flexibility has been shown to be important for the infection process, presumably to allow secondary receptor interaction (15). Flexibility of the T7 tail fiber may also be necessary to allow for tail conformational changes during the infection process.

Pyramid Domain. The pyramid domain is composed of three concave nine-stranded mixed β-sheets stacked against each other, forming a tapered pyramid with a triangular cross-section. The

taper is the result of the β-strands at the bottom of the pyramid being longer (up to 6 amino acids) than those near the top (down to three residues). Each β-sheet involves strands contributed by all three chains; interactions between β-strands are antiparallel between strands from the same monomer, but interaction between β-strands from neighboring monomers are parallel. Each β-sheet consists of one strand, R, from the first monomer, followed by five strands, STUVW, from the next monomer, and capped by three strands, XYZ, from the third. The loops connecting the β-strands vary from short β-turns to longer loops; all loops are well ordered in both crystal forms and interact extensively with other loops from the same or from neighboring monomers. The center of the domain contains exclusively hydrophobic and aromatic side-chains; each β-strand contributes one or two central side-chains to this core. In the loops and at the beginning and end of the β-strands, interactions are mixed and, apart from hydrophobic interactions, many polar interactions are formed. At the top of the pyramid domain three α-helices are located, one from each protein chain. Leu455 and Leu459 residues from the α-helices project into the center, forming a small hydrophobic core, capped at the top by Phe463.

Carboxyl-Terminal Globular Tip Domain. The α-helices are connected to the tip domain by a loop (Val464-Lys466) containing both polar and apolar residues. This loop surrounds a solvent-filled central cavity between the tip and pyramid domains. The presence of this solvent-filled cavity is further indication of possible flexibility between the tip and pyramid domains.

The tip domain of each monomer (amino acids 464–553) forms a β-sandwich with the topology shown in Fig. 1*C*. The β-sandwich contains two sheets of four β-strands each, BIDE and CHGF. β-Strands B and C are on the outside of the trimer, E and F on the inside. A search using the program DALI (16) did not turn up β-sandwich domains with the same topology (see below). As in the β-structured pyramid domain, all loops between β-strands are well ordered in both crystal forms and interact extensively with neighboring loops. Interactions between the tip monomers are of mixed nature, involving salt-bridges, hydrogen bonds, solvent molecules, and van der Waals interactions. The end of the protein chain is locked firmly into place by two salt-bridges between Glu551 and Arg508 of a neighboring chain and between Glu553 (the very C-terminal residue) and Lys468 of the other neighboring chain. This position likely prevents attack of the C terminus by proteases and contributes to trimer stability.

Discussion

Phages belonging to the *Caudovirales* order attach to host bacteria with the end of their tails. Primary, generally reversible, recognition is via tail-spike proteins or tail fibers on the side of the tail; at this point the phage is not yet committed to DNA injection. Positive recognition leads to central tail proteins productively attaching to the host membrane and DNA injection. The tail-spike of the podovirus P22 has been studied extensively in terms of carbohydrate binding, hydrolysis, folding, and assembly (17, 18). Much is also known about assembly and function of the complex fibers of the myovirus T4 (19–21). Siphoviruses like T5 and λ contain less-studied side tail fibers (22, 23) which, like the T4 fibers, do not exhibit receptor-hydrolysis activity. Here we have presented unique high-resolution structural information on a podoviral tail fiber that also does not hydrolyze its receptor.

Stability and Folding of gp17(371–553). The surface area of a monomer is around 12×10^3 Å^2, of which more than 40% is buried upon trimer formation (5.2×10^3 Å^2). The calculated dissociation energy of the trimer is 125 kcal/mol (24) and contains more than 120 intermonomer hydrogen bonds and nearly 40 potential intermonomer salt-bridges. When the pyramid domain trimer (Gly371-Phe463) is considered, the surface area of a monomer is 7.4×10^3 Å^2, of which nearly half (3.6×10^3 Å^2) is buried. The calculated dissociation energy of the pyramid domain trimer is 90 kcal/mol and contains more than 90 intermonomer hydrogen bonds and 8 potential intermonomer salt-bridges. The surface area of a tip domain monomer (Val464-Glu553) is 4.7×10^3 Å^2, of which 1.4×10^3 Å^2 is buried (around 30%). The calculated dissociation energy of the tip domain trimer is 17 kcal/mol and contains 27 intermonomer hydrogen bonds and 30 potential intermonomer salt-bridges. The large buried surface area explains the stability of trimeric gp17 (371–553), as revealed by its high melting temperature (around 74 °C) (Fig. 2) and the fact that it does not dissociate into monomers in SDS/PAGE, unless previously boiled in SDS-containing buffer (12). Our structure shows there are no disulphide bonds in the C-terminal construct, and the crystallized fragment contains the only two cysteines of the gp17 sequence.

The C-terminal tip domain is the only domain of the current structures where the monomer has a potentially independent fold and where the three chains do not intertwine. The surface buried between monomers and calculated dissociation energy is also smaller. This finding suggests gp17 folding may begin with the spontaneous formation of a monomeric carboxyl-terminal β-barrel. Interaction of three β-barrels could then lead to trimer formation. The remaining region of the fiber would then "zip up" to form the intact trimer. This folding pathway would be the same as that proposed for adenovirus fiber (25) and consistent with the observation that amber mutants in the C-terminal domain do not incorporate truncated gp17 in the virions (10). Gp17 does not appear to require specific chaperones for folding like phage T4 fibers do (26).

Comparison with Other Proteins. When the sequence of phage T7 gp17 is compared with protein sequences in databases, a strong similarity (> 85% sequence identity) is observed with gp17 of *E. coli* phage T3, *Yersinia pestis* phage PhiA1122, and enterobacteria phage 13a. Sequence similarity is also observed with fibers of phages from other bacterial species (*Kluyvera*, *Salmonella*, *Pseudomonas*, *Vibrio*, *Klebsiella*, and others). The phage attachment domain (amino acids 1–150) belongs to domain family PHA00430 and, of the gp17 domains, has the most sequence homologs in the database, consistent with the fact that many T7-like phages have the same mechanism of attaching the fiber to the phage. The tip domain has fewer sequence homologs, only T3, PhiA1122, enterobacteria phages 13a, BA14, 285P, and *Yersinia* phages Yep-Phi, Berlin, and Yepe, consistent with the well-known fact that many phages have evolved their C-terminal receptor attachment domain to adapt to different hosts by exchange with other phages (27). An

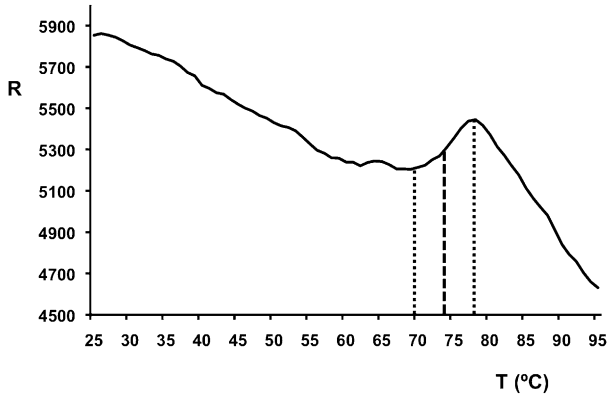

Fig. 2. Thermal stability assay. The relative fluorescence emission intensity (R) is plotted as a function of the temperature (*x* axis, °C). A melting temperature of around 74 °C was estimated (midpoint between the baseline and the point with maximum fluorescence intensity).

illustrative example is *E. coli* phage K1F, which has a homologous phage-attachment domain but a different C-terminal domain containing endo-*N*-acetylneuramidase activity to digest the *E. coli* K1 capsular polysaccharide (28, 29). Only a few phages appear to have evolved by mutation maintaining the same structural framework. Residues 359–524 of T7 gp17 (encompassing all but the final 29 residues of our structure) have been classified in Pfam domain gp37_C, which includes a subset of the highly diverged C-terminal regions of the gp37-type tail-fiber proteins of T4-like phages, the subset found in phages AR1 and Ac3 (30). Thus, amino acid sequence comparisons suggest that the tapered triangular pyramid and part of the tip structure of T7 tail fibers may be structural building blocks used in fibers of phages of the *Myoviridae* as well as the *Podoviridae* family.

When searches are conducted either with the whole structure presented here or with separate domains, no structural homologs with the same topology are found in the PDB database. Adenovirus fiber (13), reovirus fibers (14, 31), bacteriophage PRD1 P5 (32), and phage Sf6 cell-penetrating needle (33) all have a β-structured C-terminal domain of similar size and organization as gp17 has, but with a different topology. The Sf6 cell-penetrating needle knob domain (Fig. 3*A*) has the most similar topology, with two small extra β-strands (C1 and C2) and a straight swap of the E and G strands compared with our structure. However, their stalk domains contain only triple β-spiral folds or a triple coiled coil (34, 35), without an intervening β-helical domain, such as the pyramid domain in the present structure. The receptor-binding proteins of lactococcus phages p2 and TP901-1 also have a similar C-terminal domain attached to a short β-helical stalk (36, 37). The bacteriophage T4 short tail-fiber (gp12) and long tail-fiber (gp37) C-terminal domains have different folds, consisting of three intertwined monomers rather than composed of individually folded monomeric domains (21, 38), and T4 fibritin (gpwac) and the phage P22 cell-penetrating needle (gp26) have a much smaller trimerisation domain (39, 40).

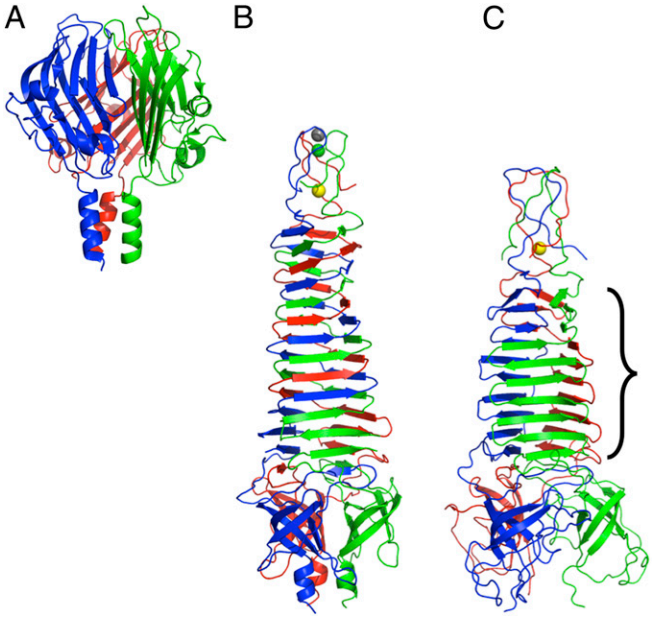

Fig. 3. Structures of the cell-penetrating tail needle knob domain of phage Sf6, PDB-code 3RWN (*A*), gpV of phage P2, PDB entries 3QR7 and 3QR8 (*B*), and gp138 of phage phi92, PDB codes 3PQH and 3PQI (*C*). Iron, calcium, and chloride ions are shown as yellow, gray, and green balls, respectively. The region of gp138 that is topologically the same as amino acids 384–446 of phage T7 gp17 is indicated.

The pyramid domain, with its interlocked trimeric β-helix, also does not have any exact topological equivalent in the structure database. Comparable structures include the triple β-helix regions of phage T4 gp12 (41) and gp5 (42), the K1F endosialidase (43) and endo-*N*-acetylneuramidase (29), streptococcal (pro)phage HylP1 and Hylp2 (44, 45), the phage P22 cell penetrating needle gp26 (40), and phage Phi29 tail-spike (46). However, the most similar structures are those of phage P2 gpV (Fig. 3*B*) and phage Phi92 gp138 (Fig. 3*C*) (47, 48), which form the tip of inner tail tube and are the first proteins to pierce the membrane in these myoviruses. Both proteins contain a tapered β-helix, strongly intertwined in the case of gpV and interlocked like gp17 in the case of gp138. Eight β-strands (from sheets STUVW and XYZ) have structural equivalents in gp138. However, both gpV and gp138 contain a small apex domain with a central iron ion instead of the α-helical region, and lack a globular tip domain.

Receptor Binding. Initial, reversible, binding of phage T7 to bacteria is mediated by the interaction of its six gp17 tail fibers with LPS (7). This interaction is presumably followed by a secondary, irreversible attachment of the tail to an unknown receptor. The relative importance of the two interactions in host-range determination is not known; gp17 may just keep the phage near the bacterial surface to make productive tail interactions with its receptor more likely (i.e., 2D diffusion vs. 3D diffusion). However, Heineman et al. (49) identified two mutants of bacteriophage T7 gp17 involved in host avoidance: a Asp520 to Glu change adapted the phage to avoid *E. coli* B and a Val544 to Ala change adapted the phage to avoid *E. coli* K12. Furthermore, Garcia et al. (50) identified a spontaneous host-range change mutant in the highly homologous phage PhiA1122 fiber in this region (Leu523 to Ser, which aligns to Ala518 of bacteriophage T7 gp17). Ala518 and Asp520 are located in the EF-loop and Val544 in the GH-loop of the gp17 tip domain; both are located on the top of the tip trimer in our structure (Fig. 1*D*). When the sequence of bacteriophage T7 gp17 is aligned with that of tail fibers of *E. coli* phage T3, *Salmonella enteridis* phage 13a and *Y. pestis* phage PhiA1122 (Fig. 4), the main differences are also observed in residues that are located in the four loops at the top of the tip domain (BC-, DE-, FG-, and HI-loops). Taken together, this information strongly suggests that the tip domain of gp17 has an important role for host-range determination, presumably by binding to a specific LPS region that differs between bacterial strains or that may be occluded in some strains and available for interactions in others.

Another possible site for receptor interaction is the concave eight-stranded β-sheet of the pyramid domain. Although there is no biochemical or mutation evidence, the concave shape and the presence of several aromatic residues is suggestive. In this region, in the $P2_12_12_1$ crystal form, Tyr385, Tyr397, and Tyr413 interact with a trimethylamine-*N*-oxide molecule from the crystallization solution in each of the three monomers, but in the $C222_1$ crystal form Tyr425 interacts with a carbonate ion in two of three monomers. Site-directed mutagenesis experiments combined with competition experiments, identification of the LPS region to which T7 binds, and cocrystallization studies with the relevant LPS fragments will be needed to identify the specific gp17-residues from the tip or pyramid domain responsible for receptor binding. If and how this initial interaction is communicated to the phage to trigger irreversible binding and DNA injection is also an interesting future subject of study.

Conclusion

We have solved the structure of the C-terminal domain of the phage T7 tail fiber, gp17(371–553). The structure contains two domains with hitherto unseen topologies, provides insight into the reason for the stability of the protein, and suggests regions that may be involved in receptor binding. The structural data may also contribute to applications of bacteriophages in, for example,

BIOCHEMISTRY

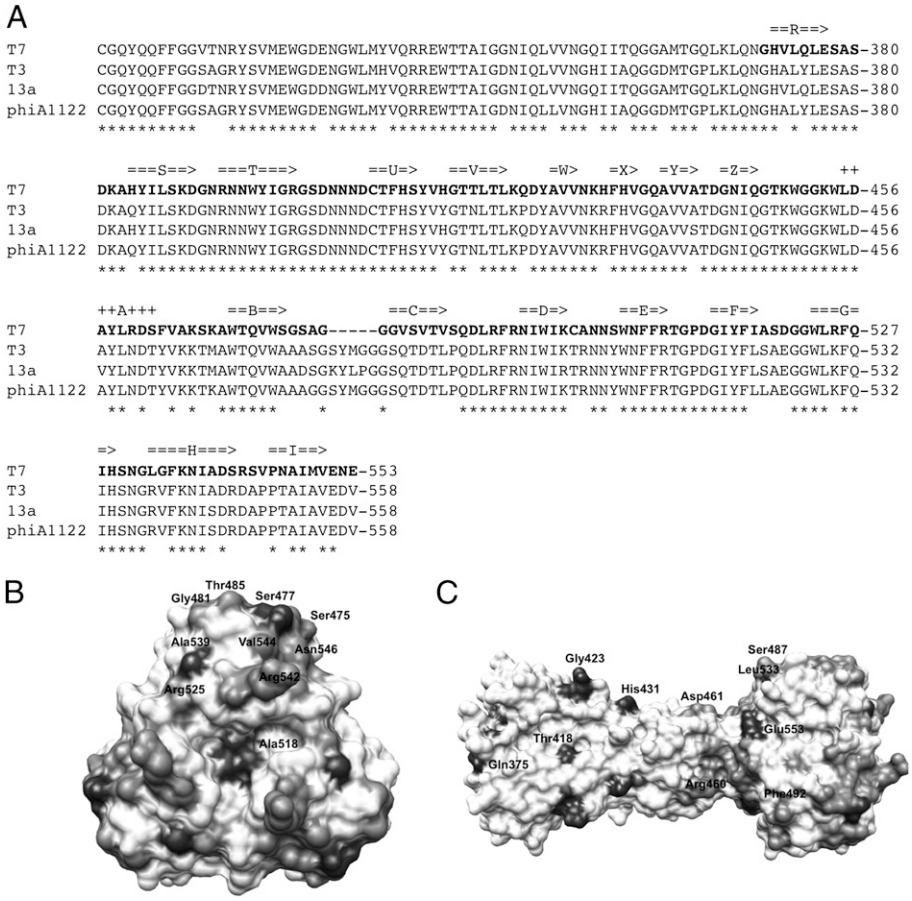

Fig. 4. Sequence conservation of gp17. (*A*) Alignment of the sequence of bacteriophage T7 gp17 with its homologs from *E. coli* phage T3, *S. enteridis* phage 13a, and *Y. pestis* phage PhiA1122. Amino acids present in our structures are shown in bold. Residues that are identical in all four proteins are marked with asterisks. Secondary structure elements identified in our structure are also indicated, beta-strands with arrows and α-helices with plus-signs. (*B* and *C*) Sequence conservation mapped on the structure. The color scale is from white (absolutely conserved) to black (no conservation). A top view (*B*) and a side view (*C*) are shown. Residues that are not conserved, and thus may be important for host range discrimination, are indicated.

bacteria typing or phage therapy, through the rational design of mutants binding different receptors or the engineering of artificial, chimeric phage fibers.

Materials and Methods

Thermal Stability Assay. Thermal stability of the protein was measured by following the fluorescence of the dye Sypro orange in the presence of 0.01 mM gp17(371–553) as a function of temperature as described in Dupeux et al. (51). The melting temperature was estimated as the temperature corresponding to the midpoint between the baseline and the point with maximum fluorescence intensity (Fig. 2).

Crystallographic Structure Solution and Refinement. Crystallogenesis and crystallographic data collection have been described previously (12). Datasets of the $P2_12_12_1$ mercury derivative collected at four different wavelengths (peak, inflection point, high energy remove, low energy remote) were input into AUTOSHARP (52). Three mercury sites were identified by the SHELXC/D programs (53) and phases were refined using data between 20.0 and 2.7 Å with AUTOSHARP, which found additional sites and rejected sites, settling on a final list of nine heavy atom sites. Further solvent flattening and histogram matching was done with SOLOMON (54) and automated building proceeded using BUCCANEER (55) at 2.7 Å resolution. This model was used for molecular replacement into the higher resolution, native data using the program MOLREP (56). The model was then input into the ARP-WARP (57) auto-trace mode using the data to 1.9 Å resolution for the $P2_12_12_1$ crystal form. Adjustment of the model, addition of extra amino acids and solvent molecules was done with COOT (58). Refinement was done using the REFMAC5 program (59). The final model contains 548 amino acids, 1 poly-ethyleneglycol fragment, 1 Tris molecule, 3 trimethylamine-*N*-oxide molecules, and 754 water molecules. A partially complete model refined against the high-resolution $P2_12_12_1$ data were used to solve the $C222_1$ structure by molecular replacement, which was extended and refined as above. This final model contains 552 amino acids (including one amino acid from the purification tag; the rest of the purification tag is apparently disordered), 2 CO_3 molecules, and 910 water molecules. Data statistics were published previously (12); phasing, refinement, and model statistics can be found in Table S1. Validation was performed with MOLPROBITY (60) and protein structure figures were prepared using PYMOL (Schroedinger) and University of California at San Francisco CHIMERA (61).

ACKNOWLEDGMENTS. We thank Ana Cuervo and José L. Carrascosa for bacteriophage T7 DNA and introducing us to T7 biology; Antonio L. Llamas-Saiz, José M. Otero, and José-Ignacio Baños-Sanz for diffraction testing of crystals; Silvia Russi and Hassan Belrhali of the European Molecular Biology Laboratory for help with data collection at BM14; Petr Leiman for careful reading of the manuscript; and Florine Dupeux, Martin Rower, Gael Seroul, Delphine Blot, and José A. Marquez for performing the thermofluor and high-throughput crystallization experiments (https://embl.fr/htxlab/). We acknowledge the use of the High-Throughput Crystallization Laboratory at the European Molecular Biology Laboratory Grenoble outstation (France), which receives funding from the European Community's Seventh Framework Program under Contract 227764. This research was funded by research Grants BFU2008-01588 and BFU2011-24843 from the Spanish Ministry of Science and Innovation; a grant from the Bill and Melinda Gates Foundation through the Grand Challenges Exploration initiative (to M.J.v.R.); and a Formación del Profesorado Universitario predoctoral contract of the Spanish Ministry of Education (to C.G.-D.).

1. Marks T, Sharp R (2000) Bacteriophages and biotechnology: A review. *J Chem Technol Biotechnol* 75:6–17.
2. Maniloff J, Ackermann HW (1998) Taxonomy of bacterial viruses: Establishment of tailed virus genera and the order Caudovirales. *Arch Virol* 143:2051–2063.
3. Steven AC, et al. (1988) Molecular substructure of a viral receptor-recognition protein. The gp17 tail-fiber of bacteriophage T7. *J Mol Biol* 200:351–365. http://www.sciencedirect.com/science/journal/00222836. Accessed May 14, 2012.
4. Cerritelli ME, Conway JF, Cheng N, Trus BL, Steven AC (2003) Molecular mechanisms in bacteriophage T7 procapsid assembly, maturation, and DNA containment. *Adv Protein Chem* 64:301–323.
5. Cerritelli ME, et al. (2003) A second symmetry mismatch at the portal vertex of bacteriophage T7: 8-Fold symmetry in the procapsid core. *J Mol Biol* 327:1–6.
6. Agirrezabala X, et al. (2005) Structure of the connector of bacteriophage T7 at 8A resolution: Structural homologies of a basic component of a DNA translocating machinery. *J Mol Biol* 347:895–902.
7. Molineux IJ (2001) No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. *Mol Microbiol* 40:1–8.
8. Serwer P, Wright ET, Hakala KW, Weintraub ST (2008) Evidence for bacteriophage T7 tail extension during DNA injection. *BMC Res Notes* 1:36.
9. Ionel A, et al. (2011) Molecular rearrangements involved in the capsid shell maturation of bacteriophage T7. *J Biol Chem* 286:234–242.
10. Kemp P, Garcia LR, Molineux IJ (2005) Changes in bacteriophage T7 virion structure at the initiation of infection. *Virology* 340:307–317.
11. Chang CY, Kemp P, Molineux IJ (2010) Gp15 and gp16 cooperate in translocating bacteriophage T7 DNA into the infected cell. *Virology* 398:176–186.
12. Garcia-Doval C, van Raaij MJ (2011) Crystallization of C-terminal domain of the bacteriophage T7 fibre gp17. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 68: 166–171.
13. van Raaij MJ, Mitraki A, Lavigne G, Cusack S (1999) A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* 401: 935–938.
14. Chappell JD, Prota AE, Dermody TS, Stehle T (2002) Crystal structure of reovirus attachment protein sigma1 reveals evolutionary relationship to adenovirus fiber. *EMBO J* 21:1–11.
15. Wu E, et al. (2003) Flexibility of the adenovirus fiber is required for efficient receptor interaction. *J Virol* 77:7225–7235.
16. Holm L, Kääriäinen S, Rosenström P, Schenkel A (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics* 24:2780–2781.
17. Andres D, Baxa U, Hanke C, Seckler R, Barbirz S (2010) Carbohydrate binding of *Salmonella* phage P22 tailspike protein and its role during host cell infection. *Biochem Soc Trans* 38:1386–1389.
18. Betts S, King J (1999) There's a right way and a wrong way: In vivo and in vitro folding, misfolding and subunit assembly of the P22 tailspike. *Structure* 7:R131–R139.
19. Kikuchi Y, King J (1975) Assembly of the tail of bacteriophage T4. *J Supramol Struct* 3: 24–38.
20. Leiman PG, et al. (2010) Morphogenesis of the T4 tail and tail fibers. *Virol J* 7:355.
21. Bartual SG, et al. (2010) Structure of the bacteriophage T4 long tail fiber receptor-binding tip. *Proc Natl Acad Sci USA* 107:20287–20292.
22. Heller K, Braun V (1982) Polymannose O-antigens of *Escherichia coli*, the binding sites for the reversible adsorption of bacteriophage T5+ via the L-shaped tail fibers. *J Virol* 41:222–227.
23. Hendrix RW, Duda RL (1992) Bacteriophage lambda PaPa: Not the mother of all lambda phages. *Science* 258:1145–1148.
24. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372:774–797.
25. Mitraki A, et al. (1999) Unfolding studies of human adenovirus type 2 fibre trimers. Evidence for a stable domain. *Eur J Biochem* 264:599–606.
26. Bartual SG, Garcia-Doval C, Alonso J, Schoehn G, van Raaij MJ (2010) Two-chaperone assisted soluble expression and purification of the bacteriophage T4 long tail fibre protein gp37. *Protein Expr Purif* 70:116–121.
27. Veesler D, Cambillau C (2011) A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. *Microbiol Mol Biol Rev* 75:423–433.
28. Petter JG, Vimr ER (1993) Complete nucleotide sequence of the bacteriophage K1F tail gene encoding endo-N-acylneuraminidase (endo-N) and comparison to an endo-N homolog in bacteriophage PK1E. *J Bacteriol* 175:4354–4363.
29. Schulz EC, et al. (2010) Structural basis for the recognition and cleavage of polysialic acid by the bacteriophage K1F tailspike protein EndoNF. *J Mol Biol* 397:341–351.
30. Yu SL, Ko KL, Chen CS, Chang YC, Syu WJ (2000) Characterization of the distal tail fiber locus and determination of the receptor for phage AR1, which specifically infects *Escherichia coli* O157:H7. *J Bacteriol* 182:5962–5968.
31. Guardado Calvo P, et al. (2005) Structure of the carboxy-terminal receptor-binding domain of avian reovirus fibre sigmaC. *J Mol Biol* 354:137–149.
32. Merckel MC, Huiskonen JT, Bamford DH, Goldman A, Tuma R (2005) The structure of the bacteriophage PRD1 spike sheds light on the evolution of viral capsid architecture. *Mol Cell* 18:161–170.
33. Bhardwaj A, Molineux IJ, Casjens SR, Cingolani G (2011) Atomic structure of bacteriophage Sf6 tail needle knob. *J Biol Chem* 286:30867–30877.
34. Guardado-Calvo P, Fox GC, Llamas-Saiz AL, van Raaij MJ (2009) Crystallographic structure of the alpha-helical triple coiled-coil domain of avian reovirus S1133 fibre. *J Gen Virol* 90:672–677.
35. Reiter DM, et al. (2011) Crystal structure of reovirus attachment protein σ1 in complex with sialylated oligosaccharides. *PLoS Pathog* 7:e1002166.
36. Spinelli S, et al. (2006) Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. *Nat Struct Mol Biol* 13:85–89.
37. Spinelli S, et al. (2006) Modular structure of the receptor binding proteins of *Lactococcus lactis* phages. The RBP structure of the temperate phage TP901-1. *J Biol Chem* 281:14256–14262.
38. Thomassen E, et al. (2003) The structure of the receptor-binding domain of the bacteriophage T4 short tail fibre reveals a knitted trimeric metal-binding fold. *J Mol Biol* 331:361–373.
39. Tao Y, Strelkov SV, Mesyanzhinov VV, Rossmann MG (1997) Structure of bacteriophage T4 fibritin: A segmented coiled coil and the role of the C-terminal domain. *Structure* 5:789–798.
40. Olia AS, Casjens S, Cingolani G (2007) Structure of phage P22 cell envelope–penetrating needle. *Nat Struct Mol Biol* 14:1221–1226.
41. van Raaij MJ, Schoehn G, Burda MR, Miller S (2001) Crystal structure of a heat and protease-stable part of the bacteriophage T4 short tail fibre. *J Mol Biol* 314: 1137–1146.
42. Kanamaru S, et al. (2002) Structure of the cell-puncturing device of bacteriophage T4. *Nature* 415:553–557.
43. Stummeyer K, Dickmanns A, Mühlenhoff M, Gerardy-Schahn R, Ficner R (2005) Crystal structure of the polysialic acid-degrading endosialidase of bacteriophage K1F. *Nat Struct Mol Biol* 12:90–96.
44. Smith NL, et al. (2005) Structure of a group A streptococcal phage-encoded virulence factor reveals a catalytically active triple-stranded beta-helix. *Proc Natl Acad Sci USA* 102:17652–17657.
45. Mishra P, et al. (2009) Polysaccharide binding sites in hyaluronate lyase—Crystal structures of native phage-encoded hyaluronate lyase and its complexes with ascorbic acid and lactose. *FEBS J* 276:3392–3402.
46. Xiang Y, et al. (2009) Crystallographic insights into the autocatalytic assembly mechanism of a bacteriophage tail spike. *Mol Cell* 34:375–386.
47. Yamashita E, et al. (2011) The host-binding domain of the P2 phage tail spike reveals a trimeric iron-binding structure. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 67: 837–841.
48. Browning C, Shneider MM, Bowman VD, Schwarzer D, Leiman PG (2012) Phage pierces the host cell membrane with the iron-loaded spike. *Structure* 20:326–339.
49. Heineman RH, Springman R, Bull JJ (2008) Optimal foraging by bacteriophages through host avoidance. *Am Nat* 171:E149–E157.
50. Garcia E, et al. (2003) The genome sequence of *Yersinia pestis* bacteriophage phiA1122 reveals an intimate history with the coliphage T3 and T7 genomes. *J Bacteriol* 185:5248–5262.
51. Dupeux F, Röwer M, Seroul G, Blot D, Márquez JA (2011) A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta Crystallogr D Biol Crystallogr* 67:915–919.
52. Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* 364:215–230.
53. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64:112–122.
54. Abrahams JP, Leslie AG (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr* 52:30–42.
55. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62:1002–1011.
56. Vagin AA, Isupov MN (2001) Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr D Biol Crystallogr* 57:1451–1456.
57. Langer G, Cohen SX, Lamzin VS, Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* 3:1171–1179.
58. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486–501.
59. Murshudov GN, et al. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67:355–367.
60. Chen VB, et al. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66:12–21.
61. Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE (2006) Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics* 7:339.