Function of bacteriophage G7C esterase tailspike in host cell adsorption

Nikolai S. Prokhorov[1,7], Cristian Riccio[2,5], Evelina L. Zdorovenko[3], Mikhail M. Shneider[2,4], Christopher Browning[2,6], Yuriy A. Knirel[3], Petr G. Leiman[2,7]*, Andrey V. Letarov[1,8]*

[1] Winogradsky Institute of Microbiology, Research Center of Biotechnology, Russian Academy of Sciences, 7b2 pr. 60-letiya Oktyabrya, Moscow 117312, Russia.

[2] École Polytechnique Fédérale de Lausanne (EPFL), BSP-415, 1015 Lausanne, Switzerland.

[3] Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 47 Leninsky pr., Moscow 119991, Russia.

[4] Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Laboratory of Molecular Bioengineering, 16/10 Miklukho-Maklaya St., 117997 Moscow, Russia.

[5] Present address: University of Cambridge, Queens' College, Silver Street, Cambridge, CB3 9ET, UK.

[6] Present address: Vertex Pharmaceuticals (Europe) Ltd, 86–88 Jubilee Avenue, Milton Park, Abingdon, Oxfordshire OX14 4RW, UK.

[7] Present address: University of Texas Medical Branch, Department of Biochemistry and Molecular Biology, 301 University Blvd, Galveston, TX 77555-0647, USA.

[8] Faculty of Biology, Lomonosov Moscow State University, 1-12 Leninskie Gory, Moscow 119991, Russia

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/mmi.13710

* Corresponding authors: letarov@gmail.com, pgleiman@utmb.edu

Summary

Bacteriophages recognize and bind to their hosts with the help of receptor-binding proteins (RBPs) that emanate from the phage particle in the form of fibers or tailspikes. RBPs show a great variability in their shapes, sizes, and location on the particle. Some RBPs are known to depolymerize surface polysaccharides of the host while others show no enzymatic activity. Here we report that both RBPs of podovirus G7C – tailspikes gp63.1 and gp66 – are essential for infection of its natural host bacterium *E. coli* 4s that populates the equine intestinal tract. We characterize the structure and function of gp63.1 and show that unlike any previously described RPB, gp63.1 deacetylates surface polysaccharides of *E. coli* 4s leaving the backbone of the polysaccharide intact. We demonstrate that gp63.1 and gp66 form a stable complex, in which the N-terminal part of gp66 serves as an attachment site for gp63.1 and anchors the gp63.1-gp66 complex to the G7C tail. The esterase domain of gp63.1 as well as domains mediating the gp63.1-gp66 interaction is widespread among all three families of tailed bacteriophages.

Keywords

bacteriophage / podovirus / phage-host interaction / tailspike / esterase

Introduction

Adsorption of a tailed phage to its host begins with a reversible event of cell surface recognition, which is carried out by specialized appendages, then proceeds to irreversible binding, and culminates in DNA delivery into the host cytoplasm (Casjens & Molineux, 2012, Leiman & Shneider, 2012, Davidson *et al.*, 2012). Despite the fact that podoviruses (phages with short non-contractile tails) are considered to be simpler than other tailed phages, the mechanisms of adsorption are poorly understood even in long-studied models such as Salmonella phage P22, coliphages T3, T7, and N4. Receptor binding proteins (RBPs) of some of these viruses are long fibers devoid of any enzymatic activity e.g. T3 and T7. In many other cases (e.g. P22, K1F, SP6) they appear as small spikes in electron microscopy photographs (and hence are called tailspikes). Many tailspikes contain predicted enzymatic domains and an activity has been determined for a number of them. Glycosyl hydrolase and lyase tailspikes depolymerize cell surface polysaccharides and this activity is essential for phage infection (Casjens & Molineux, 2012, Pires *et al.*, 2016). Some phages were shown to deacetylate O-antigen chains although this activity has never been assigned to any protein experimentally (Taylor, 1965, Kwiatkowski, 1969, Kwiatkowski *et al.*, 1975, Iwashita & Kanegasaki, 1976, Beilharz *et al.*, 1978, Pickard *et al.*, 2010). Furthermore, the role of deacetylation has so far remained unclear.

Regardless of the diversity of RBP-cell surface interactions, host cell recognition must be coordinated with irreversible binding of the phage to the host cell surface that ultimately leads to phage DNA release. Furthermore, as the tail is conserved to a significantly higher degree than the RPBs, the interaction of RBPs with host cell receptors, independently of its nature, must result in similar conformational changes in the tail and portal vertex structure. At present, little is known about how these events are linked in any

podovirus.

We decided to address these questions by studying the interaction of bacteriophages related to coliphage N4 with their hosts. Such viruses are widely distributed in many natural habitats (Chan *et al.*, 2014, Cai *et al.*, 2015, Zhan *et al.*, 2015) and constitute a substantial fraction of the human intestinal virome (Waller *et al.*, 2014). Two tail proteins of N4 are thought to play a role in host recognition and adsorption – gp65 and gp66. Gp66 emanates from the head-proximal part of the tail and forms an umbrella-like structure (Choi *et al.*, 2008). The function of gp66 in adsorption has never been characterized however clear relatedness of its a.a. sequence to other phage RBPs indicates that gp66 may be involved into host recognition. Gp65 was reported to form so called non-contractile sheath of the N4 tail (although the current body of data suggests that the sheath is likely formed by downward-pointing fibers). This protein was shown to be essential for recognition of the bacterial NfrA protein located in the outer membrane (Kiino & Rothman-Denes, 1989, McPartland & Rothman-Denes, 2009).

A number of N4-like phages and their hosts have been isolated from complex multispecies environments recently, including phage G7C and its O22-like *E. coli* strain 4s host (Golomidova *et al.*, 2007, Isaeva *et al.*, 2010, Knirel *et al.*, 2015). Bioinformatic analysis shows that all known virion proteins of G7C and N4 are similar except for those responsible for host cell recognition and attachment. The N4 locus of genes 64, 65 and 66 is replaced with 63.1 and 66 in G7C (Kulikov *et al.*, 2012). Forty N-terminal residues of G7C can be aligned to N4 gp66 showing 50% a.a. Identity. This is comparable to the level of conservation of other virion components suggesting a possible interaction of this domain with the tail. The rest of phage G7C gp66 protein sequence is unrelated to the N4 counterpart. Gp63.1 shows no similarity to any N4 protein. Gp63.1 contains a predicted

SGNH lipase or esterase domain (residues 269-492), whereas gp66 contains a pectate lyase-like beta-helix (residues 629-923) that is often found in phage RBPs (Leiman *et al.*, 2007, Pires *et al.*, 2016) (Fig. S1).

Here, we report that proteins gp63.1 and gp66 of phage G7C are responsible for host cell recognition and attachment. We determined the structure of gp63.1 and showed that this protein deacetylates the O-antigen of *E. coli* 4s lipopolysaccharide (LPS). In contrast to any previously characterized viral host recognition protein, gp63.1 introduces a small modification to the O-antigen by removing one O-acetyl group per six sugar residues of each repeating unit while leaving the rest of the O-antigen structure fully intact. We demonstrate that gp63.1 is attached to the phage tail *via* gp66, which also participates in host cell binding, and that the conserved gp63.1-gp66-like module is found in a large number of phages of different complexity and infecting different hosts. Finally, our work gives insights into the structure and function of a host cell attachment apparatus of a phage that populates the equine intestinal tract, a system where the concentrations and diversity of both viruses and cells and the complexity of their mutual interactions are extremely high.

Results

1. Interaction of gp63.1 and gp66 with host cells

To explore the role of gp63.1 and gp66 in G7C infection process and further characterize them both functionally and structurally, we produced them in a recombinant form suitable for affinity purification. Gp63.1 was expressed and purified with a C-terminal hexahistidine tag (the resulting protein will be referred to as gp63.1ht). Because hexahistidine tags on either end of gp66 did not work in affinity purification, we expressed gp66 with an N-terminal expression- and folding-enhancer SlyD chaperone domain that carried an N-terminal hexahistidine tag and can be cleaved off by the tobacco etch virus (TEV) protease (Taylor *et al.*, 2016). The resulting protein will be referred to as gpS66 further in the text.

We showed previously that untagged and not purified from the cell lysate gp63.1 could bind to *E. coli* 4s cells transiently and hypothesized that this property could be related to its function *in vivo* (Kulikov *et al.*, 2012). Crystallization grade purity gp63.1ht exhibited a very similar transient cell-binding pattern – an instant binding followed by a gradual release (Fig. 1A), which suggests that gp63.1ht retained its putative function. Notably, gp63.1ht did not bind any other *E. coli* strain tested including 4sR, a spontaneous mutant of 4s strain lacking O-antigen biosynthesis that is completely resistant to G7C infection (Fig. 1A) (Knirel *et al.*, 2015). Neither gpS66 nor unmodified gp66 showed any noticeable binding to *E. coli* 4s in this experimental setup despite our numerous attempts at modifying the experimental conditions (data not shown).

The presence of purified RBPs in a phage-cell sample inhibits phage attachment because the phage and RBPs compete for the same set of receptors on the host cell surface (Parent *et al.*, 2014, Lindberg, 1973). We used a similar competition assay to

assess the role of gp63.1 and gp66 in the G7C infection process (BSA served as a control of unspecific interactions). The phage was added to the cells, and after 8 minutes of incubation at 37°C we measured the titer of the phage that remained free in solution. Both gp63.1ht and gpS66 inhibited the binding indicating that both proteins participate in phage attachment to the host cell (Fig. 1B).

2. Gp63.1 interacts with the O-antigen of *E. coli* 4s

The transient nature of gp63.1ht binding to phage-sensitive cells suggested that gp63.1 has an enzymatic activity with which it modifies a cell surface component causing the protein to eventually dissociate. We therefore analyzed the influence of LPS, the most abundant cell surface entity, on G7C adsorption and viability. Highly purified *E. coli* 4s LPS not only inhibited phage adsorption (data not shown) but also caused irreversible virus inactivation (Fig. 2). This effect was not observed with LPS of G7C resistant strains 4sR and 4sI, isogenic to *E. coli* 4s. The 4sR lacks the O-antigen entirely, whereas the 4sI O-antigen unit is not O-acetylated (Knirel *et al.*, 2015). At the same time, preparations of the *E. coli* 4s outer membrane produced an even stronger inactivation effect suggesting that additional receptors could be present in the outer membrane (e.g. outer membrane proteins). Taking these findings together and considering that gp63.1 does not bind to 4sR cells (Fig. 1A) we concluded that the substrate of gp63.1 is the *E. coli* 4s O-antigen.

3. Gp63.1 is an esterase

Several phage tailspikes were shown to digest cell surface polysaccharides (Pires *et al.*, 2016) and we therefore examined whether gp63.1ht digests host LPS. Interestingly, the strain 4s LPS sample treated with gp63.1ht showed only a slight retardation of LPS bands

in SDS PAGE instead of a typical, nearly or fully complete degradation pattern (Fig. 3A). GpS66 did not modify LPS in a manner detectable on SDS PAGE (Fig. 3A).

Comparison of the NMR ^{1}H spectrum of the *E. coli* 4s O-antigen (O-polysaccharide, OPS) treated with gp63.1ht with that of the untreated O-antigen (Knirel *et al.*, 2015) shows that the former is missing a signal for the O-acetyl group at δ_H 2.13 (Fig. 3B). Thus, gp63.1 removes the O-acetyl group from *E. coli* 4s OPS and is therefore an O-antigen esterase (Fig. 3C). The spectra also show small differences in the signals of several sugar and N-acetyl group protons, which are most likely due to O-deacetylation. Notably, the spectrum of the *E. coli* 4s OPS treated with gp63.1 is essentially identical to that of the *E. coli* 4sI OPS that lacks the O-acetyl group at the third sugar residue of the repetitive O-unit (Fig. 3B,C), which explains the resistance of *E. coli* 4sI to G7C (Knirel *et al.*, 2015).

4. Gp63.1 crystal structure

To further characterize this novel enzyme, we crystalized gp63.1ht and determined its structure by X-ray crystallography to a resolution of 2.4 Å. Gp63.1ht consists of six domains (Fig. 4).

The first two N-terminal domains of gp63.1 (residues 1-88 and 89-160, respectively) have similar, but not identical folds (Fig. S2A). Depending on the superposition algorithm and the number of atoms participating in alignment, the root mean square deviation (RMSD) of the superposition spans the range of 2.3 Å to 3.6 Å, and the sequence identity is between 10% and 12%. As the two domains are close in space and follow each other in the amino acid sequence, it is very likely that they originate from a common ancestor as a result of an ancient gene (or domain) duplication event.

The two N-terminal domains show 70% sequence identity to the N-terminal region of

the CBA120 phage tailspike 1 (orf210, PDB code 4OJ5) (Chen *et al.*, 2014). The structures of the two proteins are very similar and can be superimposed (as rigid two-domain units) on each other with a RMSD of 0.5 Å between the 142 aligned C-alpha atoms. In both proteins, three symmetry-related His25 residues of the first N-terminal domain coordinate a buried Zn atom located on the protein's threefold axis. G7C gp63.1 displays a very prominent and large patch of positive charge at its N-terminus (Fig. S2B). This feature could be important for binding to gp66 (see below).

Domain 3 of gp63.1 (residues 170-250) is a good structural match (PDBeFold Z-score of 5.9 and RMSD of 2.2 Å between 65 aligned C-alpha atoms) to a small domain that connects the sialidase domain with the N-terminal particle-binding domain in phage tailspikes with sialidase activity (e.g. residues 245-312 of endoNF1, PDB code 3GVK, Fig. S3A). At the same time, the structure of gp63.1 domain 3 resembles that of the N-terminal part of a pectate lyase-like beta-helix found in numerous phage tailspikes. It consists of two beta-helical rungs capped by an N-terminal alpha-helix oriented perpendicular to the rungs' axis (Fig. S3B). This suggests that the junction between this domain and the following enzymatic domain in phage sialidases and in G7C gp63.1 is a recombination hotspot. It is possible that phage endosialidases and G7C gp63.1 and many other phage tailspikes with similar organization have evolved from an all-beta-helix tailspikes in which the catalytic part of the beta-helical domain was replaced with another domain (e.g. sialidase or esterase) by a horizontal gene transfer.

Domain 4 of gp63.1 (residues 250-495) has a fold of an SGNH-esterase (Fig. S4A). Despite displaying low sequence identity (11 – 23%), gp63.1 is a good structural match to proteins from this family (DALI Z-score of up to 19.8, RMSD of 2.1-3.5 Å between 170-190 aligned C-alpha atoms). The canonical serine-histidine-aspartic acid catalytic triad and the

conserved glycine and asparagine of SGNH-like lipases are formed by residues S276-H473-D470, G345 and N382, respectively (Fig. S4B) suggesting that the enzymatic mechanism gp63.1 is identical to that of other SGNH-like esterases (Lo *et al.*, 2005). Unlike its structurally orthologous SGNH-like lipases, which hide a large part of their substrates into a deep surface pocket, the hydrophilic substrate of gp63.1 binds to a shallow surface cavity (Fig. S4C).

Domain 5 of gp63.1 (residues 496-667) is a good structural match – DALI Z-score of 9.2-9.6 and RMSD of 3.0-3.4 Å between 130 aligned C-alpha atoms – to several families of carbohydrate-binding modules (CBM22, CBM35, CBM61, and perhaps to others; Fig. S5). These non-catalytic CBM domains are found in many multidomain polysaccharide depolymerases where they perform substrate-binding function and either enhance the activity of the catalytic domain through the avidity mechanism or determine the substrate specificity of the entire enzyme (Cuskin *et al.*, 2012). Notably, almost all DALI hits, including the most similar ones, show sequence identity to gp63.1 below 10% with only a few at 12-15%. Nevertheless, the substrate-binding site in several best structural matches (PDB code 2XON (Cid *et al.*, 2010) and 2W87 (Montanier *et al.*, 2009)) roughly overlaps with a shallow cavity of gp63.1 that binds a di-ethylene glycol molecule in an extended conformation (this compound is likely to be a result of decomposition of polyethylene glycol used in crystallization). The di-ethylene glycol is likely to mimic an extended polysaccharide chain.

Domain 6 (residues 668-851) is also a CBM domain (CBM 66). Two of its best structural matches identified by DALI are the C-terminal domain (residues 516-667) of the *Bacillus subtilis* exo-acting β-fructosidase SacC (Cuskin *et al.*, 2012) and the C-terminal domain (residues 373-520) of levan fructotransferase from *Arthrobacter*

ureafaciens (PDB codes 4B1M and 4FFI, respectively) (Fig. S6) (Park *et al.*, 2012). DALI gives Z-scores of 12.1-12.3, RMSD of 2.5 Å between 130-135 aligned C-alpha atoms with 10% sequence identity in these superpositions. The concave surface of this beta-sandwich domain presents ligand-binding pockets in SecC and levan fructotransferase. The surface of gp63.1 also has a depression that can potentially bind a sugar substrate (Fig. S6), although its surface charge distribution is different from that of the two other proteins. Similar to domain 5, domain 6 of gp63.1 also picks up a di-ethylene glycol molecule from the crystallization solution, but it does not overlap with the substrate binding sites of its two closest structural homologs. Instead, it binds to a large extra loop that is not found in the other two proteins.

Three domains of gp63.1 that comprise the host cell surface proximal part of the protein – the SGNH-esterase domain and domains 5 and 6 – contain large insertions (residues 285-311, 591-613 and 692-711, respectively) that are not found in their closest structural homologs (Fig. 4). These insertions form a continuous ridge on the surface of gp63.1. The putative saccharide binding sites of domains 5 and 6 are thus situated in a valley that starts at the tip of the protein and ends in the SGNH-esterase active site. This valley and possibly the ridge itself (the loop of domain 6, for example) are likely to be important for binding of an extended LPS substrate. Interestingly, the valley is created by domains 5 and 6 of one polypeptide chain and the SGNH-esterase domain of another chain in the trimer. The involvement of two chains is unclear, but it might be important for cooperative substrate binding or possible processivity.

5. The gp63.1 esterase activity is necessary for host binding

To ascertain our hypothesis that enzymatic activity of the esterase domain of gp63.1

plays a crucial role in cell surface recognition and binding, we carried out site-directed mutagenesis of the active site of the enzyme. Structure and sequence comparison of gp63.1ht with those of other SGNH-enzymes shows that residues S276, D470, and H473 comprise an esterase catalytic triad (Fig. S7) (Marchler-Bauer *et al.*, 2011). The histidine is absolutely essential for the activity of SGNH-esterases (Brumlik & Buckley, 1996, Fenster *et al.*, 2000), whereas the aspartate is the least conserved and its role in the catalysis is somewhat disputable (Wei *et al.*, 1995, Brumlik & Buckley, 1996). We, therefore, produced two alanine mutants – H473A and D470A – with a fully or partially impaired catalytic ability.

Both mutants were soluble and we purified them in a procedure developed for the wild type protein. The cell-binding test was performed as it was described above. The H473A mutant did not bind to *E. coli* cells, which strongly supports the hypothesis that a series of consecutive deacetylation events keeps the protein bound to the cell surface. On the other hand, the behavior of the D470A was identical to the wild type protein (data not shown), which suggests that this residue is not essential for the deacetylation in gp63.1. The dispensability of D470 residue for the function of the esterase domain is consistent with a mechanism of catalysis suggested for the homologous SGNH-esterase estA of *Streptomyces scabiei* (Wei *et al.*, 1995). In this protein, the catalytic histidine forms a hydrogen bond with a carbonyl oxygen of the backbone chain instead of the carboxylic group of an aspartate. It is possible that the side chain of the T472 residue in the D470A gp63.1ht mutant plays a similar role. In the gp63.1ht crystal structure, its side chain configuration is close to the second most common rotamer. Adjusting it to the most common rotamer, which would now occupy the space of the carboxyl group of D470, brings the T472 $O^{\gamma 1}$ atom to within hydrogen bond distance of the $N^{\delta 1}$ of H473 (Fig. S7A).

To confirm that the altered cell-binding properties of the mutants were not due to

severe folding defects, we examined their thermal stability and the amount of exposed hydrophobics with the help of the SYPRO orange dye that fluoresces when it binds to hydrophobic substrates. The melting curve of the WT gp63.1ht had a mid-transition denaturation temperature of ~67°C (Fig. S7B). Both mutants showed low fluorescence at room temperature and similar heights of the transition peak suggesting that the proteins were folded correctly. However, the mutations affected the overall stability as the mutants' denaturation temperatures were lower by ~20°C (Fig. S7C).

6. Interaction of gp63.1 and gp66 proteins

HHpred analysis (Soding, 2005) shows residues 295-467 of gp66 have a structure that is similar to that of domains 1 and 2 of gp63.1 (Fig. S1, Fig. 4). Orthologous domains are also found in a variety of unrelated viruses including putative tailspikes of Vil-like phages (Fig. S1B) that have a branched architecture (Adriaenssens *et al.*, 2012). We surmised that in all these proteins, the conserved domains are responsible for forming the branched structure. As the very N-terminal regions of G7C and N4 gp66 are very similar, this domain could attach the gp63.1-gp66 complex to the tail (Kulikov *et al.*, 2012).

To test these hypotheses we introduced a stop codon in the 12th and 481st amino acid position of genes 63.1 and 66 of the G7C genome and grew both mutants (63.1S12am and 66S481am, respectively) on a non-permissive host. Their virion protein composition and that of the WT phage was analyzed by SDS-PAGE (Fig. 5A). The 66S481am mutant did not contain either of the tailspikes, whereas the 63.1S12am was missing only gp63.1. Thus, gp63.1 is attached to gp66 and the latter binds to the particle directly.

To confirm that gp63.1 incorporation depends on gp66 because of the direct

interaction between the two proteins, we performed pull-down experiments with the help of an IMAC resin and a combination of tagged and untagged proteins. We found that unmodified gp66 binds to IMAC-immobilized His-tagged gp63.1ht (Fig. 5B) and vice versa – His-tagged gpS66 captures tagless gp63.1 (data not shown).

To identify the region of gp66 responsible for gp63.1 binding, we produced a series of His-tagged N- and C-terminally truncated mutants of gp66 lacking one or more of its predicted domains (Fig. 6A) and analyzed their ability to bind WT gp63.1 by the IMAC-based pull-down assay described above. To increase the yield and solubility, all gp66 fragments carried a SlyD folding enhancer domain at their N-terminus. All mutants except 66C653 (residues 653-1063) formed soluble SDS-resistant trimers suitable for testing. This analysis showed that residues 138-375 of gp66 are required for *in vitro* complex formation (Fig. 6B).

Finally, we examined cell-binding properties of the gp63.1-gp66 complex (Fig. S8A). We found that the complex behaves identically to gp63.1ht.

Discussion

1. Structure of gp63.1-gp66 complex and its role in infection

Residues 138-294 of gp66 display a weak but not insignificant similarity (HHpred probability of 38% and p-value of 2.3×10^{-4}) to domains 2 and 3 of phage T4 gp10 baseplate protein (Taylor *et al.*, 2016). These domains form an on- and off-axis attachment site for the T4 gp12 and gp11 proteins, respectively, a *bona fide* branched structure. The gp63.1-gp66 complex is likely to have a similar organization in which the C-terminal part of gp66 forms an on-axis extension of its N-terminal domain, and gp63.1 binds to its off-axis gp10-like domain (Fig. 7). The strong positive charge of the gp63.1 N-terminus (Fig. S2)

likely mediates this interaction.

Incubation of G7C with a highly purified *E. coli* 4s LPS sample leads to virus inactivation. At the same time, LPS that lacks either the complete O-antigen (*E. coli* 4sR) or only its O-acetyl group (*E. coli* 4sI) have no effect on G7C viability (Fig. 2). A similar effect – LPS-induced inactivation – has been reported for Salmonella phage P22 (Andres *et al.*, 2010). Interestingly, inactivation of Sf6, a close relative of P22, takes place only in the presence of both LPS and OmpA (Parent *et al.*, 2014). We can speculate that the event of LPS or LPS+OmpA binding causes the particle to adopt a conformation associated with irreversible attachment to the cell surface. These structural changes make the phage particle non-infectious.

In bacteriophage T4 interaction of the RBPs with the host cell surface involves a rotation of gp10, and this structural transformation commits the phage particle to irreversible host cell binding (i.e. a conformational switch of the baseplate that is followed by sheath contraction) (Taylor *et al.*, 2016). A similar architecture of the gp63.1-gp66 complex can also be important for transmitting the signal of irreversible host cell attachment to the rest of G7C tail by means of a global conformational change of the complex (i.e. rotation or reorientation). This will be similar to rotation of the fibers in phage T7 upon binding to the host cell surface (Hu *et al.*, 2013).

A number of G7C-like viruses carry a truncated version of gp66 (Fig. S8B). In these phages, the putative beta-helical and C-terminal domains are absent, but the part responsible for gp63.1 binding (the gp10-like domain) is always conserved (Fig. S8B). This analysis supports several assertions stated above: 1) the N-terminal part of gp66 binds the complex to the phage particle; 2) the C-terminal domain of gp66 may play an auxiliary or redundant role in infection of *E. coli* 4s; and 3) the architecture of the gp63.1-gp66 complex

with its putative off-axis domain is likely to be conserved because it is important for activating the rest of the tail upon host cell binding.

2. G7C alters non-specific phage defense of the host

It has been recently shown that the presence of acetyl groups on the O-antigen protects *E. coli* 4s from infection of certain T5-like bacteriophages. These viruses (e.g. DT571/2 and fiberless mutants of phage DT57C) possess long non-contractile tails and recognize membrane proteins, but not LPS, on the cell surface (Golomidova *et al.*, 2016). They cannot infect WT 4s but grow well on *E. coli* 4sI or 4sR that carry a non-acetylated O-antigen or no O-antigen at all (Golomidova *et al.*, 2016). The mechanism of this effect is unclear, but one can speculate that the O-acetyl groups enhance the interaction between the neighboring O-antigen chains thus creating a less penetrable surface layer. Furthermore, phages with esterase tailspikes are likely to play an important role in maintaining the diversity in dense bacterial populations where their selective pressure leads to a dynamic coexistence of isogenic strains with different O-antigen acetylation variants. The importance of O-acetyl modifications of the O-antigen for host recognition and their impact on bacteriophage lifestyle were shown in experiments on serotype converting temperate phages infecting *Shigella flexnery* (Lindberg *et al.*, 1978, Allison & Verma, 2000, Sun *et al.*, 2014).

External polysaccharides are thought to protect the host cell from infection by bacteriophages that cannot degrade the polysaccharide substrate (Scholl *et al.*, 2005). Our findings question this paradigm. During infection, G7C removes only one acetyl group per six sugar residues (~4% of material) and leaves the rest of the O-antigen structure, and most of the protection effect that comes with it, intact (Fig. 3C).

3. G7C deacetylates the O-antigen in a processive manner

We showed that G7C could not attach to the *E. coli* 4sI mutant strain that carries non-acetylated O-antigen or to its original host strain 4s pre-treated with gp63.1 (Fig. 1A and 2). This shows that the deacetylation of O-antigen has to take place in the course of infection but not before it, and that the process of deacetylation itself is required for irreversible attachment. The dependence of the current step of a chemical reaction on the outcome of the previous, chemically identical, step is a necessary condition for processivity. We, therefore, propose here that gp63.1 deacetylates its substrate in a processive manner. The prolonged and cyclic binding of gp63.1 to *E. coli* 4s (Fig. 1A, Fig. S8A and (Kulikov *et al.*, 2012)) puts forward a model in which the protein binds to one or more O-antigen chains and is not released until it fully deacetylates its substrate. The structure of gp63.1 also contains features supporting processivity. The trimeric gp63.1 carries an O-antigen-binding site in each of its three esterase domains and six putative saccharide-binding sites in the three copies of CBM-like domains 5 and 6, giving a total of 9 putative binding sites per trimer (Fig. 4B). These components are likely to have evolved to function together in the context of the LPS layer where they drive the enzyme and the phage particle towards the host cell surface.

3. Concluding remarks

In this work, we characterized the mechanism of phage G7C recognition and attachment to its *E. coli* 4s host cell. We showed that G7C carries two tailspike proteins gp63.1 and gp66. Gp63.1 is an esterase that deacetylates the O-antigen of *E. coli* 4s while leaving the rest of its structure unchanged. Not only the end result but also the process of

deacetylation of the O-antigen by gp63.1 is essential for G7C infection. Gp63.1 is attached to gp66, which anchors the gp63.1-gp66 complex to the phage particle by means of its 130 residue-long N-terminal domain. Residues 138-294 of gp66 are likely to form an off-axis T4 gp10-like domain that binds gp63.1. This module is widely distributed in host cell-binding proteins of unrelated bacteriophages belonging to at least two families – Podoviridae and Myoviridae that may reflect the ancient evolutionary origin of the signal transduction mechanism in bacteriophage adsorption.

Experimental Procedures

Phages, bacterial strains and growth conditions

Bacteriophage G7C (NCBI database accession number NC_015933), host strain *E. coli* 4s and its G7C-resistant derivatives 4sI and 4sR were available from earlier studies (Table S1) (Kulikov *et al.*, 2012, Knirel *et al.*, 2015). Bacteria were cultivated in liquid Lennox LB medium or on Lennox LB agar plates (Invitrogen, Carlsbad, California). For G7C lysate preparation, strain 4s was grown at 37°C from a night culture diluted 1:100 in fresh LB to an OD_{600} = 0.2 and infected with a multiplicity of infection (MOI) of 0.1. After the complete lysis of the culture, a few drops of chloroform were added, the lysate was treated with 1 μg/ml DNAse I for 30 minutes at 37°C (Sigma-Aldrich, St. Louis, Missouri), cleared by low speed centrifugation and stored at 4°C. Bacteriophages were counted by conventional double-layer plating (Sambrook & Russell, 2001, Clokie & Kropinski, 2009).

Plasmids and recombinant proteins

All plasmids used in this study are listed in Table S1, all primers – in Table S2.

Plasmids p2363.1 and p1966 for expression of untagged gp63.1 and gp66 are described elsewhere (Kulikov *et al.*, 2012).

Details of cloning procedures, production, and purification of recombinant proteins described in Supplementary Information.

Purification of lipopolysaccharides, O-polysaccharides and outer membranes

Lipopolysaccharide (LPS) samples were isolated from overnight cultures of *E. coli* 4s, 4sI, 4sR by the phenol-water extraction method with consecutive pretreatment with RNAse A, DNAse I and proteinase K (Sigma-Aldrich) as described elsewhere (Knirel *et al.*, 2015, Westphal & Jann, 1965). The purity of LPS samples was checked by SDS-PAGE separation (12% PAAG) and staining with Pierce Silver Stain Kit (Thermo Fisher Scientific). For nuclear magnetic resonance (NMR) spectroscopy, the LPS samples were degraded with 2% acetic acid for 3 hours at 100°C. The released O-polysaccharides (OPS) were isolated by size-exclusion chromatography on Sephadex G-50 Superfine (GE Healthcare, Little Chalfont, United Kingdom) with 50 mM pyridinium acetate buffer pH 4.5 as eluent monitoring elution with a differential refractometer (Knauer, Berlin, Germany).

The outer membrane (OM) of *E. coli* 4s was purified as described by Hobb *et al*. by differential solubilization with N-lauroylsarcosine sodium salt (Sigma-Aldrich) with some modifications (Hobb *et al.*, 2009). Cells were disrupted by sonication, and instead of HEPES buffer, 10 mM Tris-HCl pH 7.5 was used in all the subsequent procedures. The purified OM sample was lyophilized and weighted, taking into the account the presence of the Tris buffer in the sample.

Phage adsorption inhibition assay

The effect of gp63.1ht and gpS66 on the adsorption of G7C to *E. coli* 4s cells was assessed as follows. Aliquots of cells at $3.5x10^8$ cfu/ml containing gp63.1ht, gpS66 or bovine serum albumin (BSA, for negative control) at 100 μg/ml and a phage sample at $5.0x10^4$ pfu/ml were first brought to 37°C. Then equal volumes of the phage and cells samples were mixed, incubated at 37°C for 8 minutes, and centrifuged at 14000 g for 1 min. The resulting supernatants containing unbound phage were then titrated on *E. coli* 4s lawns. All the adsorption experiments were performed in SM buffer – 8 mM $MgCl_2$, 100 mM NaCl, 50 mM Tris-HCl pH 7.5 (Sambrook & Russell, 2001).

Phage inactivation assay

Inactivation of G7C by purified LPS and OM fraction was evaluated similar to the phage adsorption inhibition assay described above with a notable exception – samples were not subjected to centrifugation prior to plating. In brief, equal volumes of pre-warmed phage samples and LPS or OM suspensions at 0.2 mg/ml were mixed, aliquots were taken after 8 minutes of incubation and titrated (Clokie & Kropinski, 2009).

Whole cell binding assay

Binding of recombinant gp63.1ht and gpS66 to intact host cells was measured according to the following procedure. *E. coli* 4s cells in the exponential growth phase were pelleted by centrifugation in microtubes (14000 g for 1 min), washed with SM buffer once and resuspended in SM buffer containing either gp63.1ht or gpS66 at 25 μg/ml and BSA (as a control for subsequent SDS-PAGE). The final concentration of cells was $5.8x10^9$ cfu/ml. Aliquots were then taken at various times (ranging from 0.1 to 32 minutes at 37°C). Cells with bound proteins were briefly spined down at 15000 g and the supernatants were

then visualized on 10% SDS-PAGE in denaturing conditions according to a standard procedure (Sambrook & Russell, 2001). Proteins were stained with Coomassie Blue.

Nuclear magnetic resonance spectroscopy of gp63.1-treated O-polysaccharide

NMR analysis was performed as described elsewhere (Knirel *et al.*, 2015). Prior to O-antigen isolation and NMR examination, the LPS was treated with purified gp63.1ht in mass ratio 500 to 1 (LPS to protein) in 150 mM NaCl, 20 mM Tris-HCl pH 7.5 for 2 hours at 37°C.

Crystallization and structure determination of gp63.1

Gp63.1ht was crystallized at a concentration of 17.7 mg/ml in 150 mM ammonium sulfate, 34% PEE 797, 100 mM Bis-Tris pH 5.5. The crystallization conditions were found with the help of Jena Bioscience (Jena, Germany) protein crystallization kits. The selenomethionine (SeMet) derivative protein was produced in SelenoMet Medium (Molecular Dimensions, Suffolk, UK), and purified similarly to the wild type protein. Crystallization conditions of the SeMet derivative were similar to those of the wild type, except for a slightly lower protein concentration (14.3 mg/ml).

Crystallographic data were collected at the Swiss Light Source PX-I (X06SA) beamline using a PILATUS-6M (DECTRIS Ltd., Baden-Daettwil, Switzerland) pixel detector. For data collection, a crystal was picked up from the mother liquor solution and flash frozen in vaporized liquid nitrogen stream at 100 K. To determine the wavelength corresponding to the maximum of the *f″* scattering factor (i.e. the peak, Table S3), an X-ray fluorescence spectrum of a gp63.1ht SeMet crystal was first recorded around the Kα absorption edge of a free Se atom. A continuous rotation data collection protocol

comprising a complete 360° rotation with a frame width of 0.25° was then performed at this peak *f''* wavelength. The diffraction data were integrated using the XDS program (Kabsch, 2010).

The positions of the Se atoms were found with the help of the SHELXD program (Sheldrick, 2008) and the initial phases were generated by SHELXL (Sheldrick, 2010). The model was built with the help of BUCCANEER (Cowtan, 2006) and ARP/wARP (Langer *et al.*, 2008). The structure was refined with COOT (Emsley *et al.*, 2010), Refmac5 (Murshudov *et al.*, 2011) and PHENIX (Adams *et al.*, 2010). The structure was deposited to the Protein Data Bank under the accession number 4QNL.

Analysis of gp63.1 thermal denaturation

A SYPRO Orange-based thermal shift assay (Lo *et al.*, 2004) was performed in a 384-well plate with the help of a 7900HT qPCR instrument (Applied Biosystems, Foster City, California) (Lavinder *et al.*, 2009), on the following gp63.1ht variants: wild-type, D470A, H473A. All proteins had a concentration of 0.5 mg/ml. The stock solution of the SYPRO Orange dye (Thermo Fisher Scientific) was diluted 5,000 times to yield a nominal concentration (1 X dilution). Two negative controls that either did not contain the protein or both the protein and the dye were tested. Both controls showed no change in fluorescent signal associated with protein unfolding. All statistical analyses were conducted in R 3.0.2. For the purpose of visualization, fluorescence intensity was normalized within each well to values between 0 and 1 using the formula: $F_{norm} = (F_{raw} - F_{min})/(F_{max} - F_{min})$. The normalized fluorescences were then averaged, smoothed and plotted against the temperature. The melting temperatures were defined as the inflexion points of a Boltzmann sigmoidal curve fitted using a non-linear least-square regression on the raw fluorescence

data.

In vitro protein interaction studies

All *in vitro* protein interactions were derived from co-purification experiments involving metal-affinity tags and an IMAC resin and buffers identical to the ones described earlier for gp63.1ht purification. All the proteins as well as empty vectors (for negative control) were expressed in BL21(DE3) *E. coli*, the cells were lysed by sonication, and the soluble fraction was retained for subsequent experiments.

To obtain the gp63.1-gp66 complex, an IMAC column was first loaded with a lysate of gp63.1ht-expressing cells (C-terminally His-tagged gp63.1). All non-specifically bound proteins were washed with the wash buffer. A lysate containing untagged gp66 was applied onto the column, and it was then washed again with the wash buffer. All specifically bound proteins were eluted with the elution buffer and analyzed with SDS-PAGE and Coomassie staining. As a negative control, the procedure was repeated with cells carrying an empty pET23a vector instead of p2363.1HT.

To determine which part of gp66 is responsible for interaction with gp63.1, a series of gp66 deletion mutants (66C137, 66C294, 66C375, 66C460, 66N293, 66N375, 66N460, 66N537, 66N650) were produced as described above. All these mutants carried an N-terminal SlyD expression/folding enhancer tag, which was further enhanced with an N-terminal His-tag. Gp63.1 was expressed without any tags. The complex formation procedure was identical to the one described above, but in this case, the mutants of gp66 were loaded onto the IMAC column first and gp66 was loaded after. A lysate of an empty vector that expressed only His-tagged SlyD was used as a negative control (Fig. 6).

G7C 63.1 and 66 amber mutants construction

The plasmid pAcBSD that expresses arabinose-inducible amber-suppressing seryl tRNA was transformed into *E. coli* 4s. A recombineering procedure was used to create amber mutations in the genome of phage G7C. Two multicopy donor plasmids – pG63.1S12Am and pG66S481Am – were constructed as described above. Each plasmid was electroporated into the strain 4s alongside with plasmid pACBSR encoding the λ Red recombination system under the control of arabinose promoter (Herring *et al.*, 2003). Transformants that acquired resistance to ampicillin (50 μg/ml) and chloramphenicol (34 μg/ml) were selected. An overnight culture of cells grown from a single colony in LB medium supplemented with ampicillin, chloramphenicol and 0.2% arabinose was diluted 100 times in fresh medium and incubated for 1 hour at 37°C with shaking. The cells were then infected with phage G7C and the lysates were prepared as described for WT G7C above. The lysates were diluted to obtain an appropriate plaque count when plated on induced 4s:pAcBSD lawn. One thousand plaques from each sample were tested for their ability to grow on induced 4s:pAcBSD and WT 4s, and mutant phages 63.1S12Am and 66S481Am were found. The presence of the mutations was confirmed by Sanger sequencing. To analyze the protein composition of mutants' virions, the mutants were grown on the non-permissive 4s strain, which was infected with MOI of 5. Upon completion of cell lysis, mutant phages were purified by PEG precipitation and centrifugation in a stepped gradient of cesium chloride, dialyzed into SM buffer and analyzed with SDS-PAGE and subsequent Coomassie staining (Clokie & Kropinski, 2009).

Acknowledgements

We thank Prof. Jamila Horabin from the Florida State University for providing

plasmid pDS3, Hannes Richter from the Genomics Technologies Facility at the Center for Integrative Genomics of the University of Lausanne for letting us use the Applied Biosystems 7900HT qPCR machine for the thermal stability experiment, Alexander S. Shashkov from the Zelinsky Institute for help with NMR spectroscopy. The work of the laboratory in the Winogradsky Institute was partially supported by Russian Science Foundation (RSF) grant #15-15-00134 (generation of the active site mutants, mapping of the gp63.1-gp66 interaction site, phage inactivation and adsorption inhibition assays, the bioinformatic work, the isolation and structure elucidation of the O-polysaccharide). The crystallographic experiments were performed on the X06SA beamline at the Swiss Light Source, Paul Scherrer Institut, Villigen, Switzerland. We thank Dr. Meitian Wang, Dr. Takashi Tomizaki and Dr. Vincent Olieric for their continuous support of the protein crystallography beamlines.

Author contributions

N.S.P., C.R., Y.A.K., P.G.L and A.V.L. designed experiments. N.S.P. did cloning and protein expression for functional tests, purified LPS and outer membrane samples, designed and performed site-directed mutagenesis of gp63.1, gp66 and phage G7C, designed and performed functional studies. C.R., M.M.S. and C.B. purified gp63.1 samples for structural analysis. C.R. and C.B. crystallized gp63.1. C.R., C.B. and P.G.L. solved the structure. C.R. performed calorimetric analysis of gp63.1. E.L.Z. and Y.A.K. prepared OPS samples and performed structural analysis. N.S.P., P.G.L. and C.R. prepared the figures. A.V.L., P.G.L., N.S.P. and Y.A.K. analyzed the data from all sources and integrated all information into a single manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

Adams, P.D., P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger & P.H. Zwart, (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**: 213-221.

Adriaenssens, E.M., H.W. Ackermann, H. Anany, B. Blasdel, I.F. Connerton, D. Goulding, M.W. Griffiths, S.P. Hooton, E.M. Kutter, A.M. Kropinski, J.H. Lee, M. Maes, D. Pickard, S. Ryu, Z. Sepehrizadeh, S.S. Shahrbabak, A.L. Toribio & R. Lavigne, (2012) A suggested new bacteriophage genus: "Viunalikevirus". *Archives of virology* **157**: 2035-2046.

Allison, G.E. & N.K. Verma, (2000) Serotype-converting bacteriophages and O-antigen modification in Shigella flexneri. *Trends in microbiology* **8**: 17-23.

Andres, D., C. Hanke, U. Baxa, A. Seul, S. Barbirz & R. Seckler, (2010) Tailspike interactions with lipopolysaccharide effect DNA ejection from phage P22 particles in vitro. *J Biol Chem* **285**: 36768-36775.

Beilharz, H., B. Kwiatkowski & S. Stirm, (1978) On the deacetylase activity of Vi bacteriophage III particles. *Acta Biochim Pol* **25**: 207-219.

Brumlik, M.J. & J.T. Buckley, (1996) Identification of the catalytic triad of the lipase/acyltransferase from Aeromonas hydrophila. *Journal of bacteriology* **178**: 2060-2064.

Cai, L., Y. Yang, N. Jiao & R. Zhang, (2015) Complete genome sequence of vB_DshP-R2C, a N4-like lytic roseophage. *Mar Genomics* **22**: 15-17.

Casjens, S.R. & I.J. Molineux, (2012) Short noncontractile tail machines: adsorption and

DNA delivery by podoviruses. *Advances in experimental medicine and biology* **726**: 143-179.

Chan, J.Z., A.D. Millard, N.H. Mann & H. Schafer, (2014) Comparative genomics defines the core genome of the growing N4-like phage genus and identifies N4-like Roseophage specific genes. *Frontiers in microbiology* **5**: 506.

Chen, C., P. Bales, J. Greenfield, R.D. Heselpoth, D.C. Nelson & O. Herzberg, (2014) Crystal structure of ORF210 from E. coli O157:H1 phage CBA120 (TSP1), a putative tailspike protein. *PloS one* **9**: e93156.

Choi, K.H., J. McPartland, I. Kaganman, V.D. Bowman, L.B. Rothman-Denes & M.G. Rossmann, (2008) Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. *Journal of molecular biology* **378**: 726-736.

Cid, M., H.L. Pedersen, S. Kaneko, P.M. Coutinho, B. Henrissat, W.G. Willats & A.B. Boraston, (2010) Recognition of the helical structure of beta-1,4-galactan by a new family of carbohydrate-binding modules. *J Biol Chem* **285**: 35999-36009.

Clokie, M.R.J. & A.M. Kropinski, (2009) *Bacteriophages : methods and protocols*. Humana Press, New York.

Cowtan, K., (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**: 1002-1011.

Cuskin, F., J.E. Flint, T.M. Gloster, C. Morland, A. Basle, B. Henrissat, P.M. Coutinho, A. Strazzulli, A.S. Solovyova, G.J. Davies & H.J. Gilbert, (2012) How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 20889-20894.

Davidson, A.R., L. Cardarelli, L.G. Pell, D.R. Radford & K.L. Maxwell, (2012) Long noncontractile tail machines of bacteriophages. *Advances in experimental medicine and biology* **726**: 115-142.

Emsley, P., B. Lohkamp, W.G. Scott & K. Cowtan, (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**: 486-501.

Fenster, K.M., K.L. Parkin & J.L. Steele, (2000) Characterization of an arylesterase from Lactobacillus helveticus CNRZ32. *Journal of applied microbiology* **88**: 572-583.

Golomidova, A., E. Kulikov, A. Isaeva, A. Manykin & A. Letarov, (2007) The diversity of coliphages and coliforms in horse feces reveals a complex pattern of ecological interactions. *Applied and environmental microbiology* **73**: 5975-5981.

Golomidova, A.K., E.E. Kulikov, N.S. Prokhorov, R.C. Guerrero-Ferreira, Y.A. Knirel, E.S. Kostryukova, K.K. Tarasyan & A.V. Letarov, (2016) Branched Lateral Tail Fiber Organization in T5-Like Bacteriophages DT57C and DT571/2 is Revealed by Genetic and Functional Analysis. *Viruses* **8**.

Herring, C.D., J.D. Glasner & F.R. Blattner, (2003) Gene replacement without selection: regulated suppression of amber mutations in Escherichia coli. *Gene* **311**: 153-163.

Hobb, R.I., J.A. Fields, C.M. Burns & S.A. Thompson, (2009) Evaluation of procedures for outer membrane isolation from Campylobacter jejuni. *Microbiology* **155**: 979-988.

Hu, B., W. Margolin, I.J. Molineux & J. Liu, (2013) The bacteriophage t7 virion undergoes extensive structural remodeling during infection. *Science* **339**: 576-579.

Isaeva, A.S., E.E. Kulikov, K.K. Tarasyan & A.V. Letarov, (2010) A Novel High-Resolving Method for Genomic PCR-Fingerprinting of Enterobacteria. *Acta naturae* **2**: 82-88.

Iwashita, S. & S. Kanegasaki, (1976) Deacetylation reaction catalyzed by Salmonella phage c341 and its baseplate parts. *J Biol Chem* **251**: 5361-5365.

Kabsch, W., (2010) Xds. *Acta Crystallogr D Biol Crystallogr* **66**: 125-132.

Kiino, D.R. & L.B. Rothman-Denes, (1989) Genetic analysis of bacteriophage N4 adsorption. *Journal of bacteriology* **171**: 4595-4602.

Knirel, Y.A., N.S. Prokhorov, A.S. Shashkov, O.G. Ovchinnikova, E.L. Zdorovenko, B. Liu, E.S. Kostryukova, A.K. Larin, A.K. Golomidova & A.V. Letarov, (2015) Variations in O-antigen biosynthesis and O-acetylation associated with altered phage sensitivity in Escherichia coli 4s. *Journal of bacteriology* **197**: 905-912.

Kulikov, E.E., A.M. Kropinski, A. Golomidova, E. Lingohr, V. Govorun, M. Serebryakova, N. Prokhorov, M. Letarova, A. Manykin, A. Strotskaya & A. Letarov, (2012) Isolation and characterization of a novel indigenous intestinal N4-related coliphage vB_EcoP_G7C. *Virology* **426**: 93-99.

Kwiatkowski, B., (1969) Location of the VI-phage II enzyme causing deacetylation of VI-polysaccharide. *Biul Inst Med Morsk Gdansk* **20**: 235-242.

Kwiatkowski, B., H. Beilharz & S. Stirm, (1975) Disruption of Vi bacteriophage III and localization of its deacetylase activity. *The Journal of general virology* **29**: 267-280.

Langer, G., S.X. Cohen, V.S. Lamzin & A. Perrakis, (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **3**: 1171-1179.

Lavinder, J.J., S.B. Hari, B.J. Sullivan & T.J. Magliery, (2009) High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. *J Am Chem Soc* **131**: 3794-3795.

Leiman, P.G., A.J. Battisti, V.D. Bowman, K. Stummeyer, M. Muhlenhoff, R. Gerardy-Schahn, D. Scholl & I.J. Molineux, (2007) The structures of bacteriophages K1E and K1-5 explain processive degradation of polysaccharide capsules and evolution of

new host specificities. *Journal of molecular biology* **371**: 836-849.

Leiman, P.G. & M.M. Shneider, (2012) Contractile tail machines of bacteriophages. *Advances in experimental medicine and biology* **726**: 93-114.

Lindberg, A.A., (1973) Bacteriophage receptors. *Annual review of microbiology* **27**: 205-241.

Lindberg, A.A., R. Wollin, P. Gemski & J.A. Wohlhieter, (1978) Interaction between bacteriophage Sf6 and Shigella flexner. *J Virol* **27**: 38-44.

Lo, M.C., A. Aulabaugh, G. Jin, R. Cowling, J. Bard, M. Malamas & G. Ellestad, (2004) Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Analytical biochemistry* **332**: 153-159.

Lo, Y.C., S.C. Lin, J.F. Shaw & Y.C. Liaw, (2005) Substrate specificities of Escherichia coli thioesterase I/protease I/lysophospholipase L1 are governed by its switch loop movement. *Biochemistry* **44**: 1971-1979.

Marchler-Bauer, A., S. Lu, J.B. Anderson, F. Chitsaz, M.K. Derbyshire, C. DeWeese-Scott, J.H. Fong, L.Y. Geer, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, J.D. Jackson, Z. Ke, C.J. Lanczycki, F. Lu, G.H. Marchler, M. Mullokandov, M.V. Omelchenko, C.L. Robertson, J.S. Song, N. Thanki, R.A. Yamashita, D. Zhang, N. Zhang, C. Zheng & S.H. Bryant, (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic acids research* **39**: D225-229.

McPartland, J. & L.B. Rothman-Denes, (2009) The tail sheath of bacteriophage N4 interacts with the Escherichia coli receptor. *Journal of bacteriology* **191**: 525-532.

Montanier, C., A.L. van Bueren, C. Dumon, J.E. Flint, M.A. Correia, J.A. Prates, S.J. Firbank, R.J. Lewis, G.G. Grondin, M.G. Ghinet, T.M. Gloster, C. Herve, J.P. Knox, B.G. Talbot, J.P. Turkenburg, J. Kerovuo, R. Brzezinski, C.M. Fontes, G.J. Davies,

A.B. Boraston & H.J. Gilbert, (2009) Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 3065-3070.

Murshudov, G.N., P. Skubak, A.A. Lebedev, N.S. Pannu, R.A. Steiner, R.A. Nicholls, M.D. Winn, F. Long & A.A. Vagin, (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**: 355-367.

Parent, K.N., M.L. Erb, G. Cardone, K. Nguyen, E.B. Gilcrease, N.B. Porcek, J. Pogliano, T.S. Baker & S.R. Casjens, (2014) OmpA and OmpC are critical host factors for bacteriophage Sf6 entry in Shigella. *Mol Microbiol* **92**: 47-60.

Park, J., M.I. Kim, Y.D. Park, I. Shin, J. Cha, C.H. Kim & S. Rhee, (2012) Structural and functional basis for substrate specificity and catalysis of levan fructotransferase. *J Biol Chem* **287**: 31233-31241.

Pickard, D., A.L. Toribio, N.K. Petty, A. van Tonder, L. Yu, D. Goulding, B. Barrell, R. Rance, D. Harris, M. Wetter, J. Wain, J. Choudhary, N. Thomson & G. Dougan, (2010) A conserved acetyl esterase domain targets diverse bacteriophages to the Vi capsular receptor of Salmonella enterica serovar Typhi. *Journal of bacteriology* **192**: 5746-5754.

Pires, D.P., H. Oliveira, L.D. Melo, S. Sillankorva & J. Azeredo, (2016) Bacteriophage-encoded depolymerases: their diversity and biotechnological applications. *Applied microbiology and biotechnology* **100**: 2141-2151.

Sambrook, J. & D.W. Russell, (2001) *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Scholl, D., S. Adhya & C. Merril, (2005) Escherichia coli K1's capsule is a barrier to bacteriophage T7. *Applied and environmental microbiology* **71**: 4872-4874.

Sheldrick, G.M., (2008) A short history of SHELX. *Acta Crystallogr A* **64**: 112-122.

Sheldrick, G.M., (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66**: 479-485.

Soding, J., (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**: 951-960.

Sun, Q., Y.A. Knirel, J. Wang, X. Luo, S.N. Senchenkova, R. Lan, A.S. Shashkov & J. Xu, (2014) Serotype-converting bacteriophage SfII encodes an acyltransferase protein that mediates 6-O-acetylation of GlcNAc in Shigella flexneri O-antigens, conferring on the host a novel O-antigen epitope. *Journal of bacteriology* **196**: 3656-3666.

Taylor, K., (1965) Enzymatic deacetylation of Vi-polysaccharide by Vi-phage. II. *Biochem Biophys Res Commun* **20**: 752-756.

Taylor, N.M., N.S. Prokhorov, R.C. Guerrero-Ferreira, M.M. Shneider, C. Browning, K.N. Goldie, H. Stahlberg & P.G. Leiman, (2016) Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* **533**: 346-352.

Waller, A.S., T. Yamada, D.M. Kristensen, J.R. Kultima, S. Sunagawa, E.V. Koonin & P. Bork, (2014) Classification and quantification of bacteriophage taxa in human gut metagenomes. *The ISME journal* **8**: 1391-1402.

Wei, Y., J.L. Schottel, U. Derewenda, L. Swenson, S. Patkar & Z.S. Derewenda, (1995) A novel variant of the catalytic triad in the Streptomyces scabies esterase. *Nature structural biology* **2**: 218-223.

Westphal, O. & K. Jann, (1965) Bacterial lipopolysaccharides. Extraction with phenol-water and further applications of the procedure. *Methods Carbohydr Chem* **5**: 83–91.

Zhan, Y., A. Buchan & F. Chen, (2015) Novel N4 Bacteriophages Prevail in the Cold Biosphere. *Applied and environmental microbiology* **81**: 5196-5202.

Figure legends

Fig. 1. Gp63.1 and gp66 participate in phage G7C host recognition.

A. Gp63.1ht cell binding assay. Purified gp63.1ht along with BSA (as a negative control) were added to concentrated cell suspensions, aliquots were taken at various times, the cells and cell-bound protein were pelleted down by centrifugation at 15000 x g for 30 s, and supernatants were analyzed by SDS-PAGE and Coomassie staining. Gp63.1ht binds transiently to G7C sensitive host 4s strain cells (top panel). Gp63.1ht does not bind to G7C resistant 4sR cells deficient in O-antigen production (bottom panel) and any other non-isogenic *E. coli* strain tested (data not shown). The images shown are representatives of three independent experiments (n = 3) with the same outcomes.

B. Phage G7C adsorption inhibition assay. A titer of the original phage sample (left column) and titers of the free phage particles after the incubation with G7C-sensitive *E. coli* 4s cells in the presence of BSA, gp63.1ht or gpS66 (columns 2 – 4). Average titers from three independent experiments are shown and error bars indicate the 0.95 confidence intervals for each mean.

Fig. 2. Phage G7C inactivation assay. Purified LPS from *E. coli* 4s and isogenic G7C resistant mutants 4sI and 4sR alongside with the same amount of purified outer membranes of WT 4s were added to identical phage samples and, after 8 minutes of incubation, aliquots were taken and titrated. Pure SM buffer was used as a negative control. LPS from 4s alone inactivated phages nearly as effectively as outer membranes presumably containing all host receptors. Average titers from three independent

experiments are shown and error bars indicate the 0.95 confidence intervals for each mean.

Fig. 3. Gp63.1 deacetylates the O-antigen of the G7C host *E. coli* 4s.

A. Treatment of 4s LPS with recombinant gp63.1ht, gpS66 and combination of both results in a subtle retardation of LPS bands in SDS-PAGE in samples incubated with gp63.1ht and both proteins compared to the original LPS sample and LPS treated with gpS66 alone. LPS were visualized by silver staining. The result was obtained in triplicate and a representative image is reported.

B. ^{1}H NMR spectra of *E. coli* strain 4s OPS (top), 4s OPS treated with gp63.1ht (middle), and mutant strain 4sI OPS devoid of O-acetylation of a repeating unit (bottom). The spectrum of 4s OPS treated with gp63.1ht lacks the characteristic peak of the O-acetyl group.

C. A scheme of the 4s LPS modification performed by gp63.1 tailspike protein. The single hexasaccharide repeating unit of the original and modified *E. coli* 4s O-antigen are presented. The units contain the following monomers: A – α-Glc*p*, B – β-Glc*p*A, C – 3-OAc-β-Gal*p*NAc, D – α-Gal*p*, E – β-Gal*p*NAc, F – α-Glc*p*, C' – β-Gal*p*NAc. Gp63.1 removes the only O-acetyl group presented in the O-antigen unit.

Fig. 4. Crystal structure of gp63.1. A ribbon diagram (A) and a molecular surface (B) are shown. Amino acid residues representing the inter-domain junctions are indicated in bold. In panel A, domains 1 and 2 of chain A are colored purple, domain 3 - dark gray, SGNH-esterase domain – yellow, two CBM-like domains 5 and 6 – green and light green, respectively. In panel B, each chain is depicted in different shades of gray, except for three

domains – SGNH-esterase of chain A and CBM-like domains 5 and 6 of chain C – representing the proposed O-polysaccharide-binding interface. The domains are colored according to the color scheme of panel A. Large insertions into the SGNH-esterase domain (residues 285-311), domain 5 (591-613) and domain 6 (692-711) that are not found in closest structural homologs are shown in different shades of red.

Fig. 5. Gp63.1 and gp66 interact *in vivo* and *in vitro*.

A. Protein composition of WT G7C and 63.1S12Am and 66S481Am mutants grown in non-permissive conditions. Purified phage samples were separated by SDS-PAGE and Coomassie stained. 63.1S12Am does not contain gp63.1 whereas 66S481Am is missing both proteins – gp63.1 and gp66. Molecular weight markers (kDa) are indicated on the right.

B. IMAC-based pull-down assay with His-tagged gp63.1ht and tagless gp66. An IMAC column was loaded with lysates of gp63.1ht- and gp66-expressing cells (gp63.1ht lys. and gp66 lys.) in succession and specifically bound proteins were eluted. Coomassie-stained SDS PAGE documents the different stages of the purification procedure. A lysate of cells with an empty pET23a vector (C. lys.) instead of the gp63.1 expressing plasmid was used in the same procedure as a negative control (Bottom panel). Flth stands for flowthrough. The images shown are representatives of three independent experiments (n = 3) with the same outcomes.

Fig. 6. Domains 2, 3 and 4 of gp66 are necessary for gp63.1 binding.

A. Design of gp66 deletion mutants for gp63.1-gp66 interaction assay shown in panel B. The bar at the bottom of the figure shows the scale in amino acids. The color

code corresponds to putative domain architecture of gp66. The gray column denotes a region of gp66 that is necessary for gp63.1-gp66 complex formation. It comprises domains 2 and 3 (residues 138-294) that are predicted to be homologous to T4 gp10 domains D2 and D3

B. Gp63.1-gp66 interaction assay. The Coomassie-stained SDS-PAGE shows specifically bound fractions from an IMAC column loaded with N-terminally His-and SlyD-tagged gpS66 deletion mutants and tagless gp63.1. The expression enhancer tag SlyD and free gp63.1ht are also shown as negative controls. Gp66 mutant naming nomenclature is as in panel A. Molecular weight markers (kDa) are indicated on the right.

Fig. 7. Gp63.1 and gp66 form a two-branched host recognition complex.

A. A schematic diagram of gp63.1-gp66 host recognition complex position in a phage G7C virion. Gp66 consists of a particle binding domain (1) shown in red, two T4 gp10-like domains (2 and 3), two domains homologues to two N-terminal domains of gp63.1 (4 and 5) shown in purple, and putative host binding C-terminal part with predicted ß-helical topology. The N-terminal domain of gp63.1 (1) binds to T4 gp10-like domain 2 of gp66.

B. A diagram of the interacting region of G7C tailspikes that is conserved in receptor-binding proteins in a number of unrelated phages. The conserved core consists of two N-terminal domains of gp63.1 (denoted BD – branch domain) and five domains of gp66 – particle binding domain (1), two T4 gp10-like domains (2 and 3) and domains 4 and 5 (BD-like).

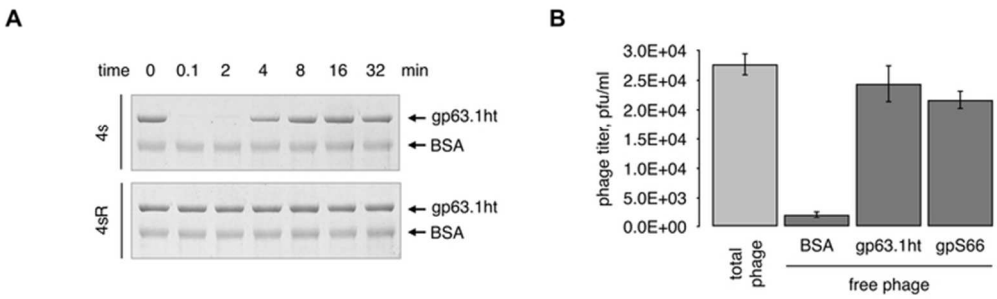


Fig. 1

64x23mm (300 x 300 DPI)

Fig 2

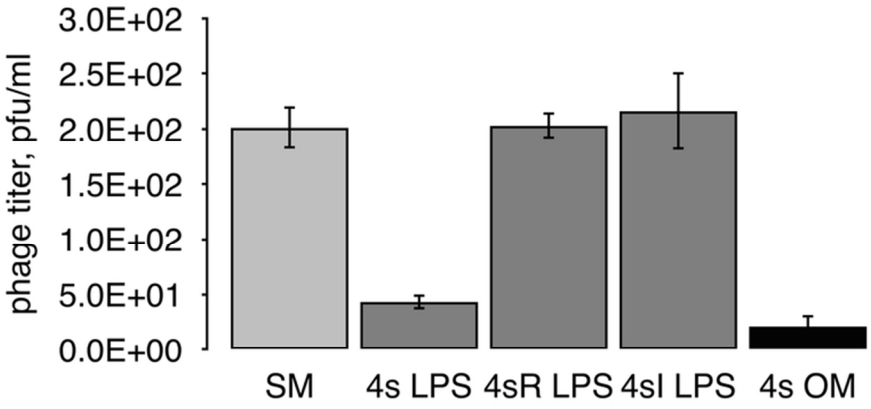


Fig. 2

74x45mm (300 x 300 DPI)

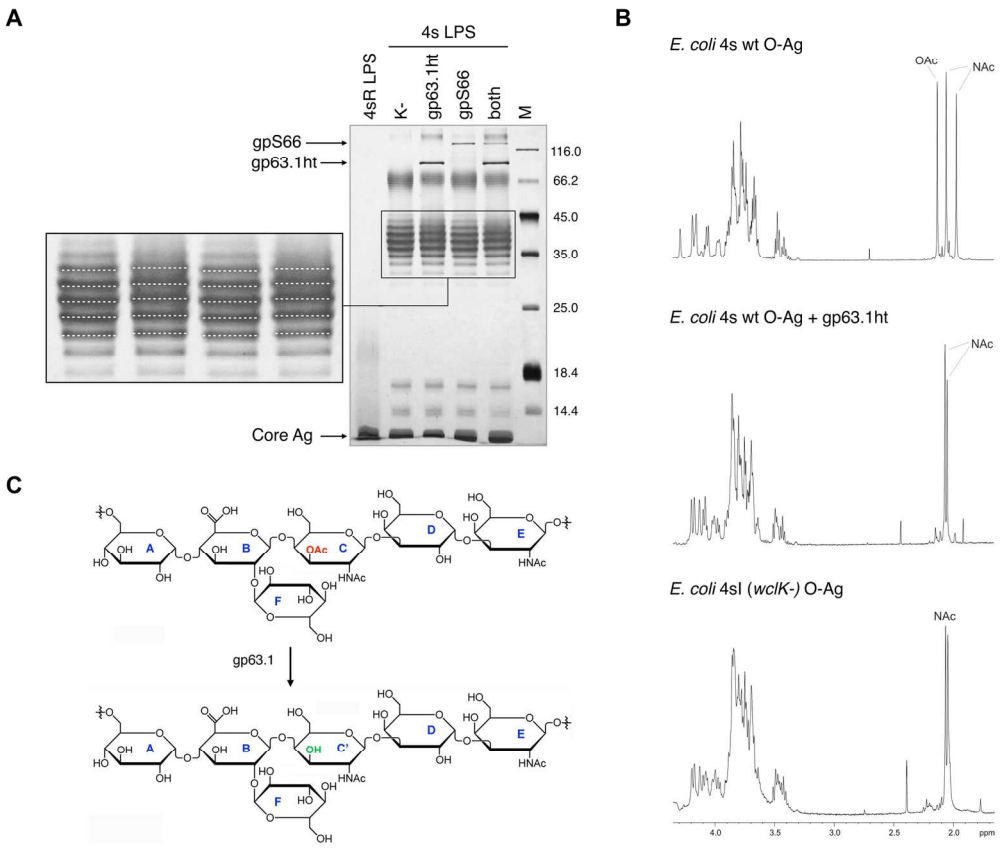


Fig. 3

159x141mm (300 x 300 DPI)

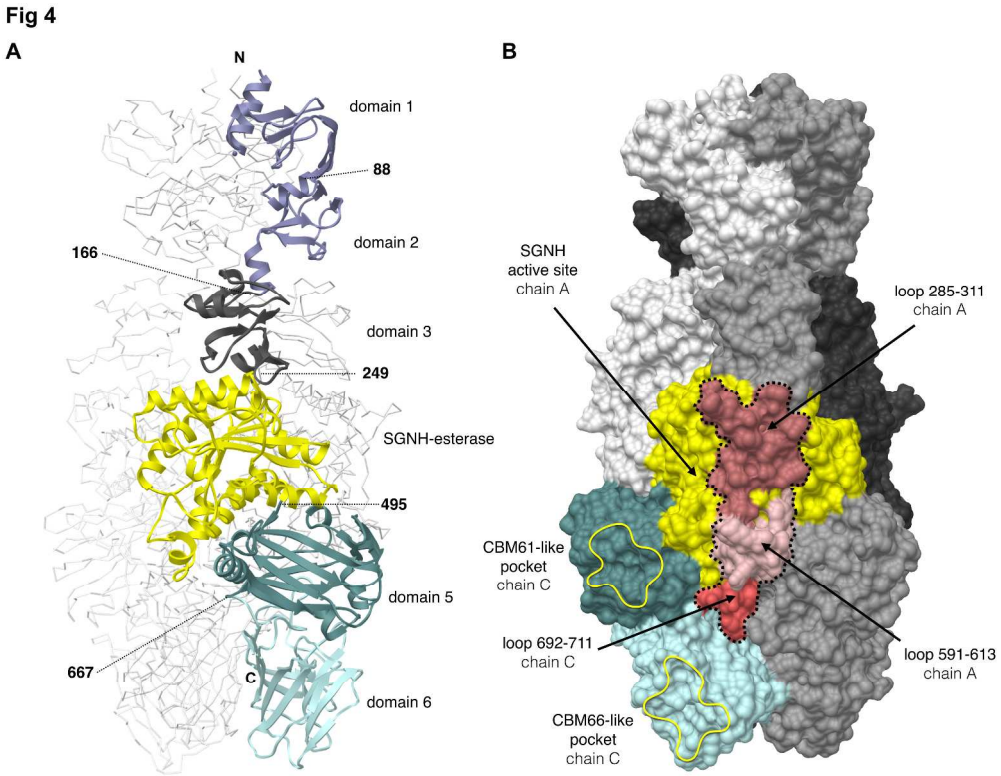


Fig. 4

1478x1145mm (72 x 72 DPI)

Fig 5

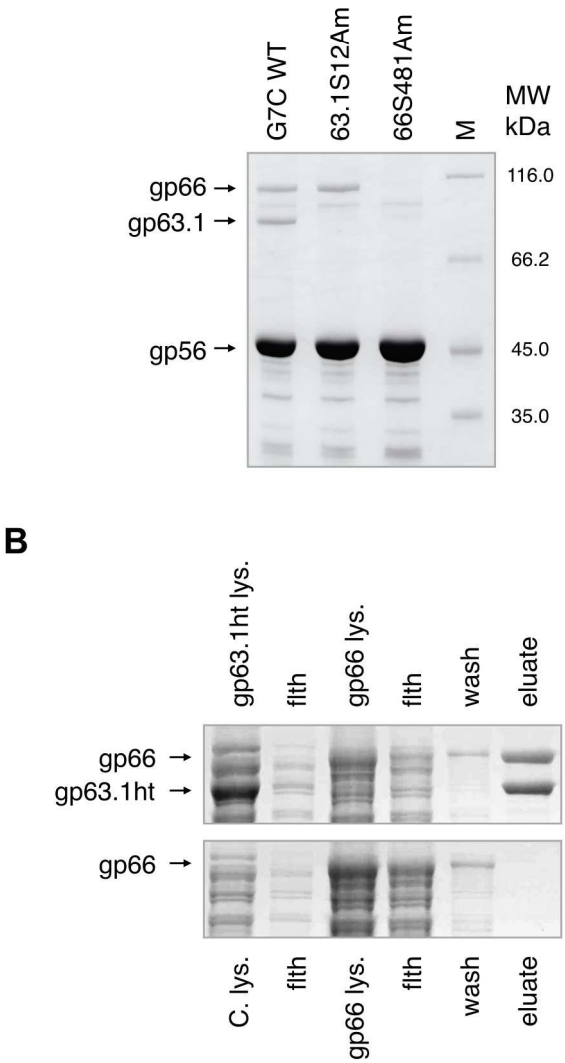


Fig. 5

250x520mm (300 x 300 DPI)

Fig 6

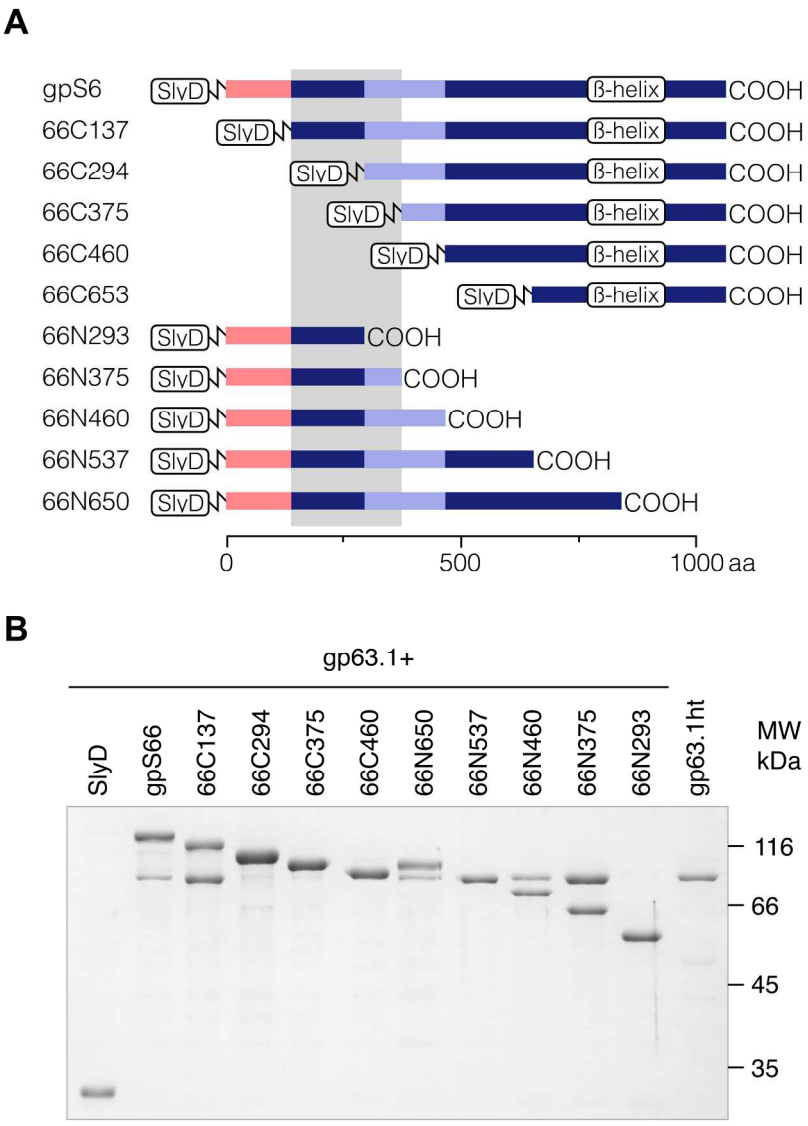


Fig. 6

176x260mm (300 x 300 DPI)

Fig 7

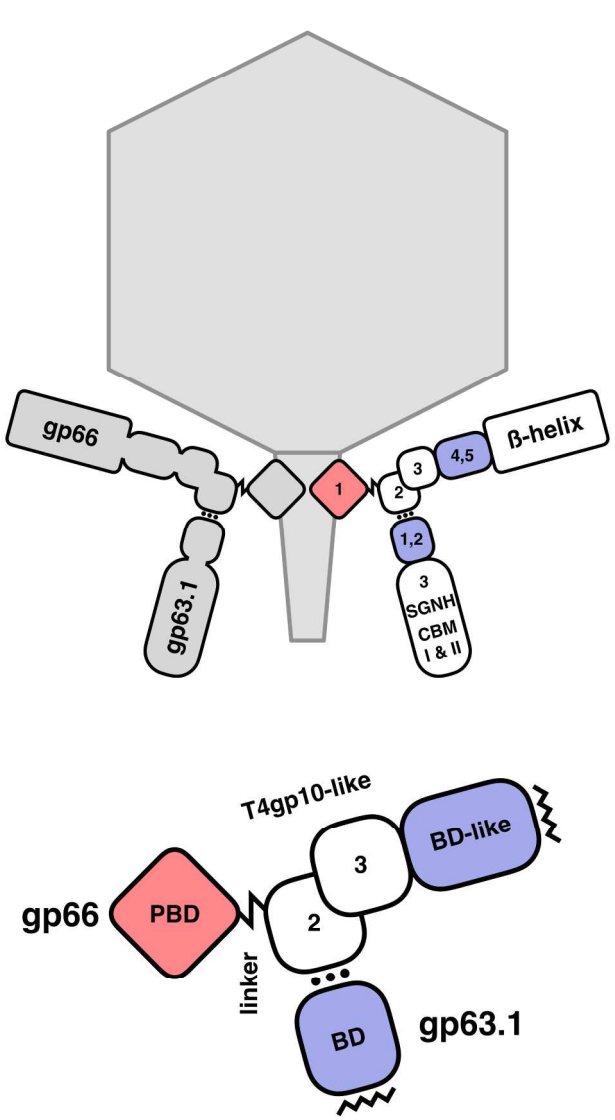
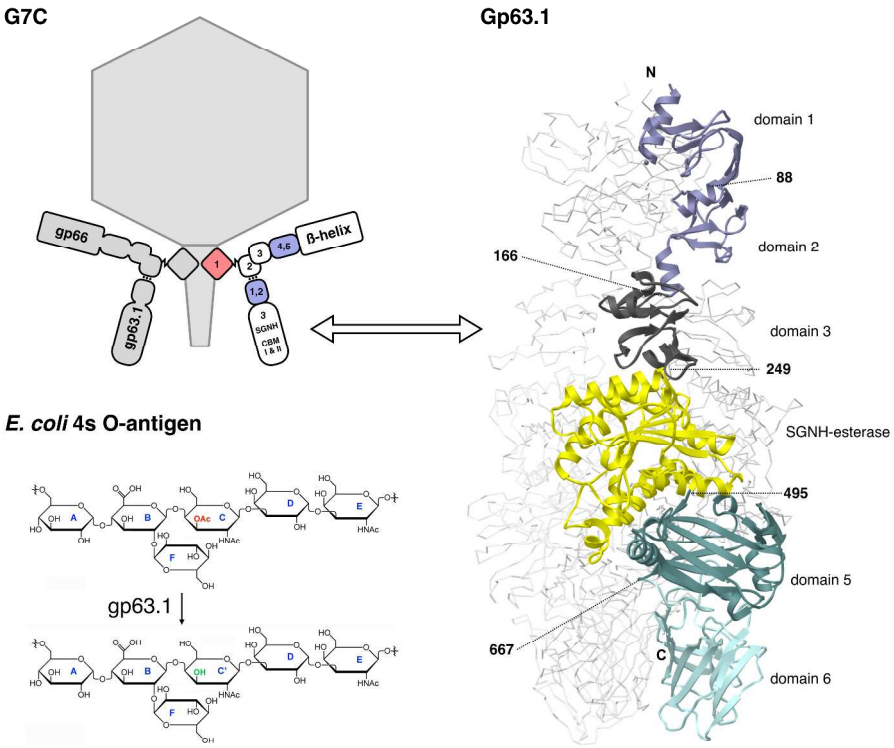


Fig. 7

204x347mm (300 x 300 DPI)

N4-like phage G7C recognizes its *E. coli* 4s host cell with the help of gp63.1 and gp66 tailspike proteins. Gp66 is attached to the phage particle directly whereas gp63.1 binds to gp66. Gp63.1 is an esterase that deacetylates the *E. coli* 4s O-antigen while leaving its polysaccharide backbone intact. This minor modification of the O-antigen is absolutely essential for G7C infection.



A schematic of phage G7C with its receptor-binding tailspike proteins, a crystal structure of the G7C gp63.1 tailspikes and its mode of action on the O-antigen substrate.

1587x1411mm (72 x 72 DPI)