Structure and Function of the Branched Receptor-Binding Complex of Bacteriophage CBA120

Michel Plattner[1,4,†], Mikhail M. Shneider[2], Nikolay P. Arbatsky[3], Alexander S. Shashkov[3], Alexander O. Chizhov[3], Sergey Nazarov[4], Nikolai S. Prokhorov[1], Nicholas M.I. Taylor[5], Sergey A. Buth[1], Michela Gambino[6], Yilmaz Emre Gencay[6,‡], Lone Brøndsted[6], Elizabeth M. Kutter[7], Yuriy A. Knirel[3] and Petr G. Leiman[1]

1 - Department of Biochemistry and Molecular Biology, *Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, 301 University Blvd, Galveston, TX 77555-0647, USA*
2 - Laboratory of Molecular Bioengineering, *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, 16/10 Miklukho-Maklaya St., 117997 Moscow, Russia*
3 - N. D. Zelinsky Institute of Organic Chemistry, *Russian Academy of Sciences, Moscow, Russia*
4 - École Polytechnique Fédérale de Lausanne, *Lausanne CH-1015, Switzerland*
5 - Structural Biology of Molecular Machines Group, *Protein Structure & Function Programme, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, Copenhagen 2200, Denmark*
6 - Department of Veterinary and Animal Sciences, *University of Copenhagen, Stigbøjlen 4, 1870 Frederiksberg C, Denmark*
7 - The Evergreen State College, *Olympia, WA 98505, USA*

Correspondence to Petr G. Leiman: *pgleiman@utmb.edu*
https://doi.org/10.1016/j.jmb.2019.07.022
Edited by M Gottesman

Abstract

Bacteriophages recognize their host cells with the help of tail fiber and tailspike proteins that bind, cleave, or modify certain structures on the cell surface. The spectrum of ligands to which the tail fibers and tailspikes can bind is the primary determinant of the host range. Bacteriophages with multiple tailspike/tail fibers are thought to have a wider host range than their less endowed relatives but the function of these proteins remains poorly understood. Here, we describe the structure, function, and substrate specificity of three tailspike proteins of bacteriophage CBA120—TSP2, TSP3 and TSP4 (*orf211* through *orf213*, respectively). We show that tailspikes TSP2, TSP3 and TSP4 are hydrolases that digest the O157, O77, and O78 *Escherichia coli* O-antigens, respectively. We demonstrate that recognition of the *E. coli* O157:H7 host by CBA120 involves binding to and digesting the O157 O-antigen by TSP2. We report the crystal structure of TSP2 in complex with a repeating unit of the O157 O-antigen. We demonstrate that according to the specificity of its tailspikes TSP2, TSP3, and TSP4, CBA120 can infect *E. coli* O157, O77, and O78, respectively. We also show that CBA120 infects *Salmonella enterica* serovar Minnesota, and this host range expansion is likely due to the function of TSP1. Finally, we describe the assembly pathway and the architecture of the TSP1–TSP2–TSP3–TSP4 branched complex in CBA120 and its related Vil-like phages.

Introduction

Infection of a bacterium by a bacteriophage begins with binding of the phage particle to the cell surface [1,2]. Most known bacteriophages utilize surface polysaccharides as receptors [3]. In gram-negative bacteria, the latter can be covalently linked to a lipid tail that forms the outer leaflet of the cell membrane (such structures are called lipopolysaccharides, or LPSs) or can be attached to the outer membrane or LPS non-covalently (these sugars are called capsular polysaccharides). The external moiety of LPS is called the O-specific polysaccharide (OPS) or O-antigen. It is a staggeringly diverse set of

carbohydrate structures that can be linear or branched, can be short or long, can contain common sugar monomers (e.g., glucose, galactose or *N*-acetylglucosamine), or more exotic sugars (e.g., β-D-*threo*-pentose or 6-deoxy-L-talose). These monomers can be linked in a great variety of ways [4,5].

Bacteriophages recognize cell surface polysaccharides with the help of tail fibers (TFs) or tailspike proteins (TSPs) [1,2,6–8]. TFs and TSPs emanate from the tail of the phage particle, a special organelle through which the phage genome and proteins are translocated from the phage capsid into the cell [9–12]. Thus, binding of TFs and TSPs to host cell surface structures sets the phage particle onto the path of irreversible attachment and genome injection [13]. Unlike TFs, TSPs are shorter and stockier and carry domains that are assumed to have enzymatic activity, although the actual activity is known for only a handful of TSPs [8,14–23]. Production of phage progeny depends on a multitude of factors, and the host cell can block the infection through a number of mechanisms after the phage particle has attached to the cell surface [24–29]. Nevertheless, the spectrum of ligands to which TFs and TSPs can bind is the primary determinant of bacteriophage's host range [30,31].

T4, P2, Mu, and other "model system" bacteriophages carry at most two types of TFs on their virions [32]. Members of the Viunalike genus to which phage CBA120 belongs are equipped with multiple TSPs and TFs [33]. These proteins are thought to form a branched structure in which each branch is formed by a single TSP or TF and the entire complex is attached to the baseplate via a fibrous protein [33]. The CBA120 TSP-TF cluster of genes contains a putative TF gene 215 (orf215), four TSP genes (orf210–orf213, further named TSP1–TSP4, respectively) that show similarity to each other and to other phage tailspikes [33,34], and a large TF or TSP gene (orf209) whose orthologs are found in large gram-positive SPO1-like bacteriophages where they participate in host cell binding [35,36]. Despite such an array of putative receptor-binding proteins, CBA120 is currently known to infect only *Escherichia coli* strains with O157 and O78 O-antigens [34].

The structure of TSP1 of CBA120 has been recently solved by x-ray crystallography [37]. Despite possessing a pectate lyase-like β-helix that is known to bind to and cleave polysaccharides [14], this protein did not show any enzymatic activity toward the O157:H7 *E. coli* O-antigen and did not bind to O157:H7 *E. coli* cells. TSP1 was reported to bind to non-O157:H7 *E. coli* strains ATCC 35218 (unknown serotype) and, somewhat surprisingly, to DH5α (K-12) [37] that CBA120 does not infect [34].

In this study, we describe x-ray crystal structures and enzymatic activities of TSP2, TSP3, and TSP4 of CBA120. We determined the structures of their O-antigen substrates and digestion products. We also report the crystal structure of TSP2 in complex with the repeating unit of the O157 O-antigen. We show that treatment of O157:H7 *E. coli* cells with TSP2 makes them resistant to CBA120 infection. Finally, we show that TSP1, TSP2, TSP3, and TSP4 form a branched hand-shaped complex *in vitro*. Its *in vitro* assembly is ordered as follows: a complex of TSP2 and TSP4 must form first, and either TSP1 or TSP3 can then bind to it independently from each other. Bioinformatics and the predicted structure of TSP complexes in other phages show that TSP4 is likely responsible for attachment of this complex to the CBA120 baseplate.

Results

CBA120 TSP2 and TSP4 contain a T4 gp10-like module

Bioinformatic analysis (Fig. 1) predicts the domain organization of the four TSP proteins of CBA120 with high confidence [38–40]. Upstream and adjacent to the common β-helical domain, which is located roughly in the middle of the protein, all four TSPs carry a module of two similar tandem domains (although there is only one copy of such domain in TSP2) first described in the structure of CBA120 TSP1 [37]. They will be referred to as TD1 and TD2 further in the text. TSP2 and TSP4 are larger than the other two proteins because they contain a phage T4 gp10-like module upstream of the TSP1-like domains. T4 gp10 is a multivalent, letter X-shaped protein that consists of four domains D1, D2, D3, and D4 (residues 1–155, 156–251, 252–395, and 396–602, respectively) that will be referred to as XD1, XD2, XD3, and XD4 further in the text [41]. T4 gp10 is responsible for attachment of the TF network containing gp11 and gp12 to the T4 baseplate. Gp11 and gp12 bind to domains XD2 and XD3 of gp10, respectively. Residues 1–160 of CBA120 TSP2 match XD2 and XD3 domains with a probability of 79% (*E*-value of 0.00016), and residues 82–165 of TSP4 match XD1, XD2, and XD3 with a probability of 98% (*E*-value of 1.3×10^{-9}) in HHpred analysis (Fig. 1).

Domain organization of CBA120 TSPs is typical for phage TSPs

To facilitate crystallization, the gp10-like and TD1-like domains of TSP2 and the gp10-like module of TSP4 were removed. The resulting mutants (called TSP2ΔN and TSP4ΔN, respectively), full-length TSP3, and their Se-methionine derivatives could then be crystallized. The single anomalous diffraction technique was used to obtain the phase information for the three crystal structures [42].

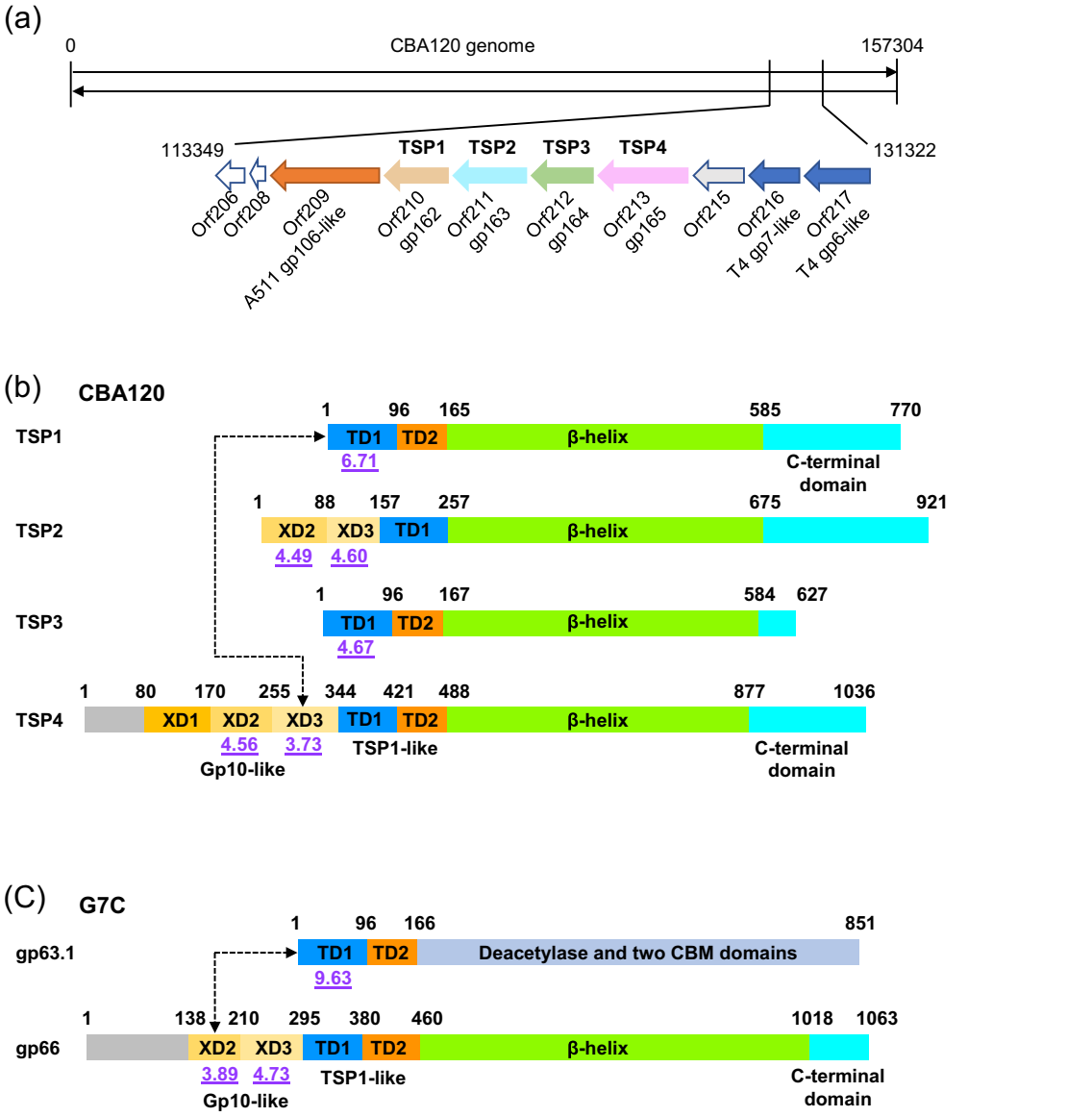


Fig. 1. Bioinformatic analysis of CBA120 TF/TSP genes and proteins. (A) The organization and location of TF/TSP genes in the CBA120 genome. (B) The domain architecture of TSP1, TSP2, TSP3, and TSP4 proteins. The two N-terminal tandem domains of TSP1 (TD1 and TD2) and their orthologs in all proteins are colored tangerine and cornflower blue. The lime rectangle is a β-helical domain. The C-terminal domain is colored cyan. The gp10-like domains XD1, XD2, and XD3 are colored in different shades of orange. The domain borders (in residue numbers) are shown above each protein. Calculated p*K*a's of TD1-like domains are shown below each protein with purple-colored underlined labels. The dashed arrow connects the most positively and most negatively charged domains that could interact based on complex formation experiments. (C) The domain architecture of two TSPs of the podophage G7C (gp63.1 and gp66). The color and labeling schemes as in panel B.

The structures of the three proteins are similar with the large β-helical domain as their main feature (Fig. 2). In agreement with the bioinformatic analysis, the N-terminal modules consisting of TD1 and TD2 domains of TSP1, TSP3, and TSP4ΔN have nearly identical Cα traces and can be superimposed with a root mean square deviation (RMSD) of 0.5–0.6 Å (depending on the pair) for the whole 160 residues in the alignment (Fig. 3). Unlike the other two proteins, TSP1 carries a large positively charged patch on its surface (Fig. 4) and contains a buried Zn ion on the threefold axis [37]. It is coordinated by the side chains of three symmetry-related His25 that emanate from the C-terminal part of the first α-helix of the

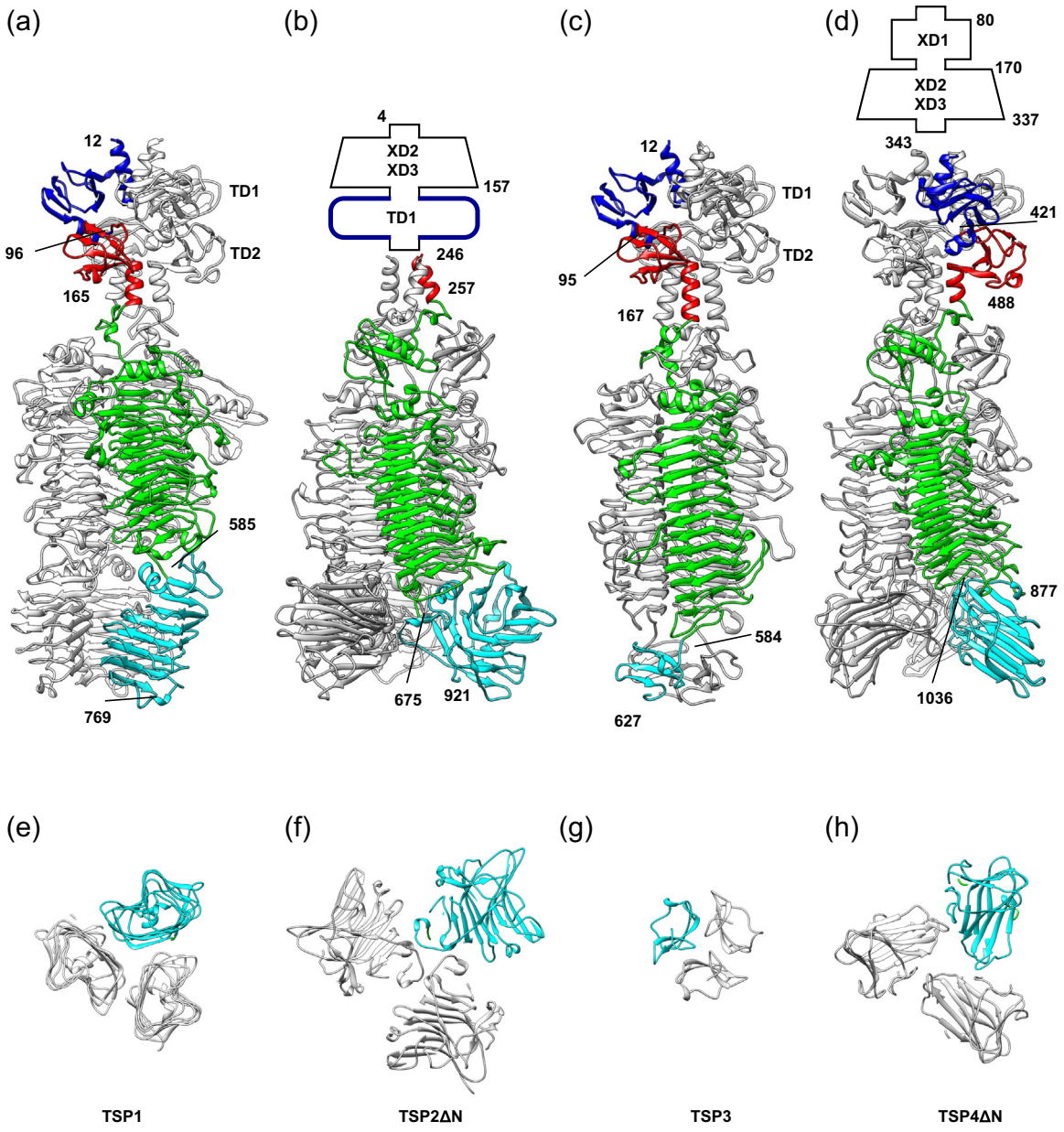


Fig. 2. Crystal structures of CBA120 TSP proteins TSP1–TSP4. Panels A, B, C, and D present side views of TSP1, TSP2ΔN, TSP3, and TSP4ΔN, respectively. Panels E, F, G, and H present a central slice (orthogonal to the 3-fold) through the C-terminal domain of the TSP protein shown in the panel immediately above. In each trimer, one chain is colored according to the domain organization of TSP1. TD1 and TD2 domains are colored blue and red, respectively. The catalytic β-helical domain is colored green, and the C-terminal domain is colored cyan. The missing N-terminal modules of TSP2 (B) and TSP4 (D) are represented in a schematic form according the bioinformatic prediction. XD1, XD2, and XD3 stand for T4 gp10-like domains (see Fig. 1). TD1 is tandem domain 1, a homolog of which is found at the N-terminus of TSP1.

TD1 domain. The corresponding residues in TSP3 and TSP4ΔN are Asn24 and Val351, respectively.

Poly- and oligosaccharide substrates bind to cavities and groves on the surfaces of other TSPs [15]. In all four CBA120 TSPs, such features are found on the interface between the neighboring polypeptide chains (Fig. 4). Furthermore, the active sites are usually particularly strongly negatively charged and the central cavity in all the four CBA120 TSPs displays this property. Hydrolases are known to utilize carboxylic groups of two aspartic or glutamic acid side chains for catalysis [43]. These groups are usually between 5 and 10 Å apart [43]. The interchain cavities of all four TSPs

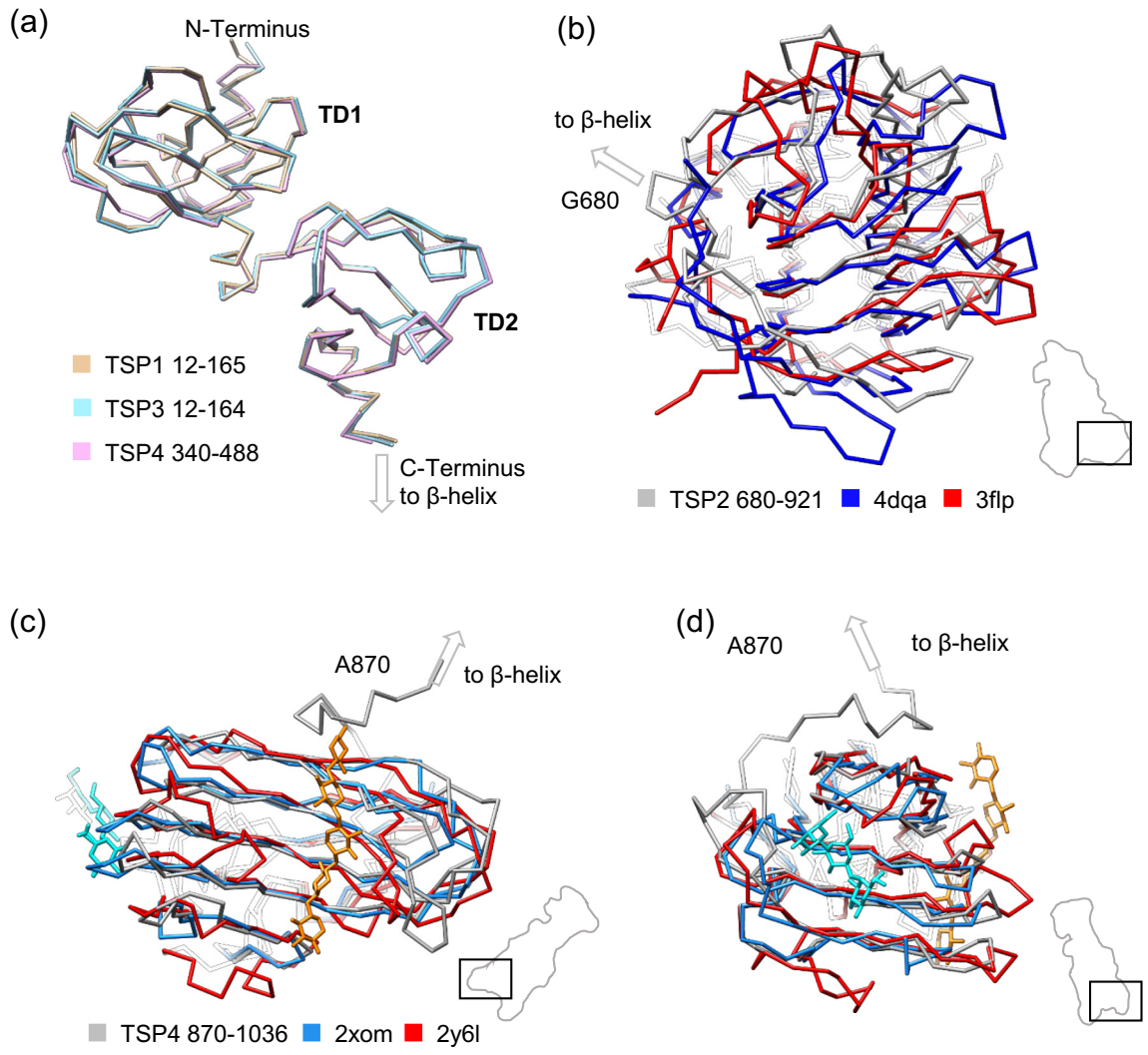


Fig. 3. Structure and similarity of the N- and C-terminal domains to other domains. (A) Superposition of the N-terminal domains TD1 and TD2 of TSP1, TSP3, and TSP4ΔN. These domains share a high sequence identity and superimpose with an RMSD of 0.5–0.6 Å (depending on the pair). (B) Superposition of the TSP2 C-terminal domain onto the putative carbohydrate binding protein (BACOVA_03559) from *B. ovatus* (PDB: 4dqa) and onto the heptameric SAP-like pentraxin from *L. polyphemus* (PDB: 3flp). The DALI *Z*-scores of the two superpositions are 13.4 and 13.1, respectively, whereas the RMSDs are the same at 3.5 Å. The structure-based sequence identities of the two superpositions are 13% and 9%, respectively. (C and D) Two views of the superposition of the TSP4 C-terminal domain onto the carbohydrate-binding modules CBM61 (PDB: 2xom) and CBM4–2 (PDB: 2y6l). The DALI *Z*-scores of two superpositions are 15.6 and 15.3, respectively, and the RMSDs are 2.3 and 2.1 Å. The CBM4–2 (orange-colored sugar) binding pocket is located near the interchain interface of TSP4. The substrate-binding pocket of CBM61 (cyan-colored sugar) is on the outer surface, facing away from the interchain interface. A CBM4–2-like binding pocket would lead the oligosaccharide directly into the active site cavity of TSP4.

contain such residues at appropriate distances, suggesting that all four TSPs could be hydrolases. Indeed, a crystal structure of CBA120 TSP3 has been recently reported and its active site is located in the interchain negatively charged cavity shown in Fig. 4C [44].

The C-terminal domains of the four TSP proteins are predominantly β-structural, but their folds are different (Fig. 2). In other TSP proteins, such domains were proposed to play a role in the trimerization and folding of the protein or in binding to surface polysaccharides [20,21]. In the four CBA120 TSP proteins, a large part of the C-terminal domain solvent-exposed surface is buried in the trimeric interface (Table 1). However, the role of these domains in trimerization is uncertain

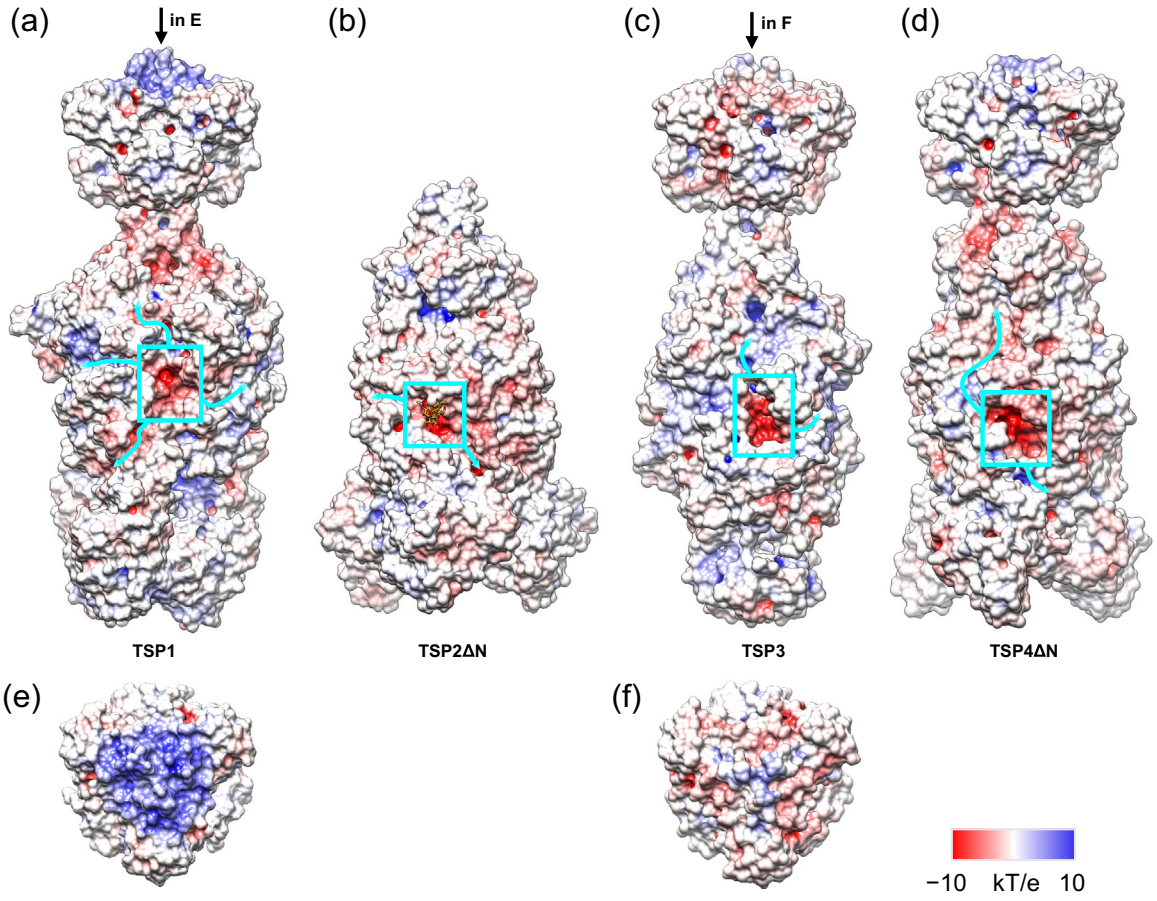


Fig. 4. The distribution of electrostatic potential on the molecular surface of CBA120 TSP proteins. (A–D) Side views of TSP1, TSP2ΔN, TSP3, and TSP4ΔN, respectively. All four TSPs feature a prominent negatively charged cavity (semitransparent cyan rectangle) on the interface of two adjacent chains. The cavity is part of a grove on the surface of the protein (semitransparent cyan lines extending from the cyan rectangle). (E and F) End-on views of the surfaces of TSP1 and TSP3 (their N-terminal domains) that interact with other TSPs in the TSP1–TSP2–TSP3–TSP4 complex.

because their predicted free energy of dissociation is small and negative (i.e., dissociation is a favorable process in solution) (Table 1) [45]. For comparison, the predicted free energy of dissociation of the P22 tailspike C-terminal domain (PDB code 1TSP, residues 545–666), which is known to be required for trimerization [46], has a large positive value of 115.7 kcal/mol.

DALI search [47] for structures similar to these domains identified several sugar-binding proteins except for the C-terminal domain of TSP3. Remarkably, a domain with a similar fold is found upstream of the β-helix domain in the gp42 TSP of *Acinetobacter* phage vb_AbaP_AS12 (DALI *Z*-score of 3.7, RMSD of 1.9 Å with 40 Cα atoms in alignment, 10% sequence identity) and in the gp49 TSP of *Acinetobacter* phage Fri1 (DALI *Z*-score of 3.3, RMSD of 2.1 Å with 40 Cα atoms in alignment, 13% sequence identity). The C-terminal domain of TSP2 is a good match (DALI *Z*-score of 13.1 and an RMSD of 3.5 Å) to the SAP-like pentraxin from *Limulus polyphemus* (Protein Data Bank code 1qtj) [48] and to the putative carbohydrate binding protein BACOVA_03559 from *Bacteroides ovatus* ATCC 8483 (PDB code 4dqa) (Fig. 3B). The C-terminal domain of TSP4 is a very good match (*Z*-

Table 1. PISA analysis of C-terminal domains of TSP1, TSP2, TSP3, and TSP4

C-terminal domain (residue numbers)	Surface area (Å^2)	Buried area (Å^2)	Buried area (%)	ΔG^{diss} (kcal/mol)
TSP1 (585–769)	20,841.4	3335.0	16.0	−2.0
TSP2 (680–921)	26,957.6	2299.5	8.5	−2.0
TSP3 (576–627)	8665.2	2631.4	30.4	−1.3
TSP4 (870–1036)	20,523.5	2621.2	12.8	−9.1

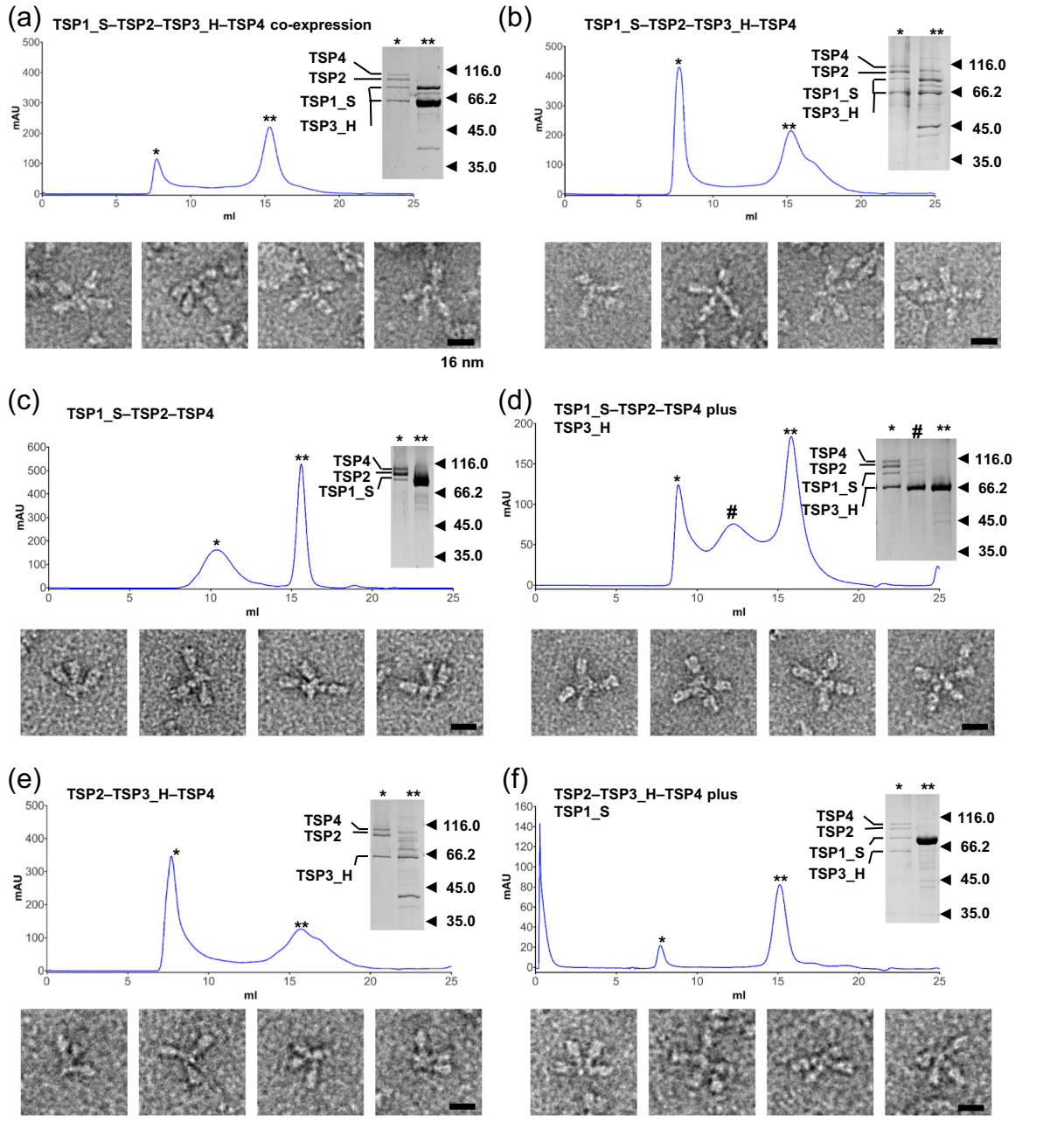


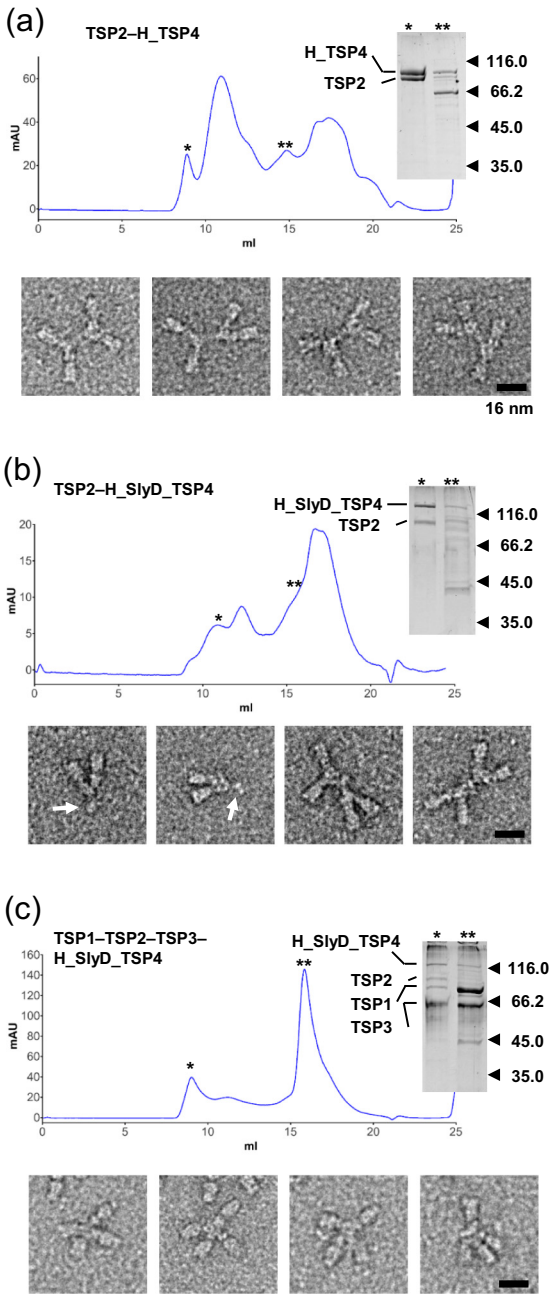
Fig. 5. Analysis of the assembly of the TSP1–TSP2–TSP3–TSP4 complex. Each panel shows a Superose 6 size-exclusion chromatogram and SDS-PAGE analysis of the two fractions labeled with a single (*) and double (**) asterisk. The composition of the high-molecular-weight, single-asterisk fraction is then analyzed with negative-staining electron microscopy and shown in four representative images. The scale bar is 16 nm. The numbers next to the SDS-PAGE correspond to the position of molecular weight standards. (A) Co-expression of C-terminally Strep-tagged TSP1 (TSP1_S), untagged TSP2, C-terminally His-tagged TSP3 (TSP_H), and untagged TSP4. (B) A complex of TSP1_S, TSP2, TSP3_H, and TSP4 can be purified from mixed lysates in which the four proteins were expressed separately. (C) TSP1_S, TSP2, and TSP4 form a complex. (D) Adding TSP3_H to the TSP1_S–TSP2–TSP4 results in a TSP1_S–TSP2–TSP3_H–TSP4 complex. (E) TSP2, TSP3_H, and TSP4 form a complex. (F) Adding TSP1_S to the TSP2–TSP3_H–TSP4 results in a TSP1_S–TSP2–TSP3_H–TSP4 complex.

score of 15.6 and an RMSD of 2.1 Å) to the carbohydrate binding domain (CBM) 61 and to CBM4–2 (Fig. 3C, 3D) [49]. Hence, the C-terminal domains of CBA120 TSP1, TSP2, and TSP4 are likely to participate in binding to cell surface oligo- and polysaccharides. The C-terminal domain of TSP3 can also

promote binding of TSP3 to cell surface sugars, but its exact function remains to be determined.

TSP1, TSP2, TSP3, and TSP4 form a complex and its assembly pathway is ordered

A branched, brush-like structure of unknown composition emanates from the baseplate of CBA120 [33,34]. We expressed TSP1, TSP2, TSP3, and TSP4 separately or as a contiguous cluster and found that the four proteins form a complex that could be purified to homogeneity by affinity and size exclusion chromatography (Fig. 5).



The four components of this complex are present in equimolar amounts. When imaged in an electron microscope, the complex resembled the structure emanating from the CBA120 baseplate. We, therefore, conclude that TSP1, TSP2, TSP3, and TSP4 are responsible for forming this branched, brush-like structure in the CBA120 phage particle.

To delineate the assembly pathway of the TSP1–TSP2–TSP3–TSP4 complex, we expressed and purified His- or Strep-tagged TSPs in various combinations. The following information was taken into account in the placement of affinity tags: (1) the C-termini of TSP2 and TSP4 could not be tagged as they were either partially or completely buried in the structure (Fig. 2); (2) unlike TSP2, TSP4 has a conserved N-terminal extension beyond the T4 gp10-like region, an organization that is similar to G7C gp66 TSP that is known to interact with the virion and thus was deemed to be amenable for tagging [8] (Fig. 1); and (3) the rest of the N-terminal domains were left unmodified because N-terminal domains or peptides attach TSPs and TFs to the virion or to each other in other systems [8,41,50,51].

Complex formation analysis of the two shorter TSPs (TSP1 and TSP3) initially gave a false positive because $TSP1^{Strep}$ bound to HisTrap resin despite lacking a His-tag and co-eluted from the resin together with $TSP3^{His}$. However, no high molecular complex could be detected in the subsequent size exclusion chromatography step after untagged TSP2 or TSP4 was added to this "complex" (Fig. S1). To the contrary, the two longer TSPs (TSP2 and TSP4) formed a real complex. Either His-tag or a

Fig. 6. TSP2 and TSP4 form a complex. Each panel shows a Superose 6 size-exclusion chromatogram and SDS-PAGE analysis of the two fractions labeled with a single (*) and double (**) asterisk. The composition of the high-molecular-weight, single-asterisk fraction is then analyzed with negative-staining electron microscopy and shown in four representative images. The scale bar is 16 nm. The numbers next to the SDS-PAGE correspond to the position of molecular weight standards. (A) TSP2 and N-terminally His-tagged TSP4 (H_TSP4) form a complex with an equimolar ratio of components. The complex contains two trimers of TSP2 and TSP4 each, which dimerize via their N-terminal domains. (B) Introduction of an additional domain (SlyD) to the N-terminus of TSP4 (H_SlyD–TSP4) interferes with the (TSP2–TSP4):(TSP2–TSP4) complex formation and a proper assembly intermediate TSP2–H_SlyD_TSP4 can be seen in EM. This complex carries an additional density (white arrow) likely corresponding to the SlyD domain. A (TSP2–H_SlyD_TSP4): (TSP2–H_SlyD_TSP4) complex can also be detected in this fraction. (C) The His-SlyD-tag does not interfere with the formation of a complex containing all four TSPs. A TSP1–TSP2–TSP3–H_SlyD_TSP4 complex can be purified by independent expression of the components and lysate mixing and then visualized in EM.

Table 2. Complexes of TSP proteins

Components of mixed lysates or co-expressed								Complex forms?	Figure panel
TSP1	$TSP1^{Strep}$	TSP2	$TSP3^{His}$	TSP3	TSP4	$^{His}TSP4$	$^{HisSlyD}TSP4$		
	✓	✓	✓		✓			Yes	5A, 5B, 5D, 5F
✓		✓		✓			✓	Yes	6C
	✓		✓					No	Not shown[a]
		✓				✓		Yes	6A
		✓					✓	Yes	6B
	✓	✓	✓					No	S1A
	✓		✓		✓			No	S1B
	✓	✓			✓			Yes	5C
		✓	✓		✓			Yes	5E

[a] See the subsequent step shown in Fig. S1A and S1B.

His-tag-SlyD expression/folding enhancer leader at the N-terminus of TSP4 (^{His}TSP4 or ^{HisSlyD}TSP4) could be used to purify a TSP2–TSP4 complex in which TSP2 was tagless (Fig. 6A and 6B).

To test whether the attachment of the shorter TSPs to the TSP2–TSP4 complex is coordinated or independent, $TSP1^{Strep}$, TSP2, $TSP3^{His}$, and TSP4 were expressed separately. The $TSP1^{Strep}$, TSP2, and TSP4 cell lysates were combined, purified using a Strep-affinity resin and subjected to size exclusion chromatography. The three proteins formed a complex (Fig. 5C, Table 2). Addition of $TSP3^{His}$ (which was purified separately with the help of a HisTrap resin) to the $TSP1^{Strep}$–TSP2–TSP4 complex resulted in a complex containing all four of the proteins in equimolar amounts (Fig. 5D). The procedure was then repeated with first mixing TSP2, $TSP3^{His}$, and TSP4 lysates, purifying the TSP2–$TSP3^{His}$–TSP4 complex using a HisTrap resin and size exclusion chromatography (Fig. 5E, Table 2), and then adding $TSP1^{Strep}$, which was purified separately on a Strep-affinity resin. A complex of all four of the TSPs could again be detected (Fig. 5F). The results of the complex formation studies are summarized in Table 2 and Fig. 7A.

TSP2 degrades the O157 OPS, binds to *E. coli* O157:H7, and blocks CBA120 infection

To identify the role of each of the four CBA120 TSPs in binding and attachment to an O157 *E. coli* host, we subjected them to several assays. The ability of each of the four TSPs to cleave and digest O157 LPS that was purified from *E. coli* NCTC 12900 (O157:H7) was examined with the help of sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 8A). TSP2ΔN digested this LPS down to small oligosaccharides, whereas TSP4ΔN also showed some, albeit very weak, activity (Fig. 8A).

The four CBA120 TSPs were then tested for binding to live *E. coli* NCTC 12900. To eliminate possible interference from improperly folded or hydrophobic N-terminal domains, the TSP constructs that were previously employed in the crystallographic analysis (Fig. 2) were used. Of the four TSPs, TSP2ΔN was the only protein that showed any type of binding in this assay. This interaction was transient, and the protein dissociated from the cells within less than a minute (Fig. 8B, 8C). TSP1, TSP3, and TSP4ΔN did not bind (Fig. S2).

TSP1 was reported to bind to non-O157:H7 *E. coli* strains ATCC 35218 and DH5α [37], which CBA120 does not infect ([34] and our own data). This binding was detected with the help of the immunodot blot assay. However, the same study stated that no interaction between fluorescently labeled TSP1 and DH5α could be observed in fluorescence microscopy. We did not detect binding of TSP1 to our DH5α strain (Fig. S2D). It is possible that our assay and fluorescence microscopy are significantly less sensitive than the immunodot blot assay and cannot detect very short-lived, transient interactions.

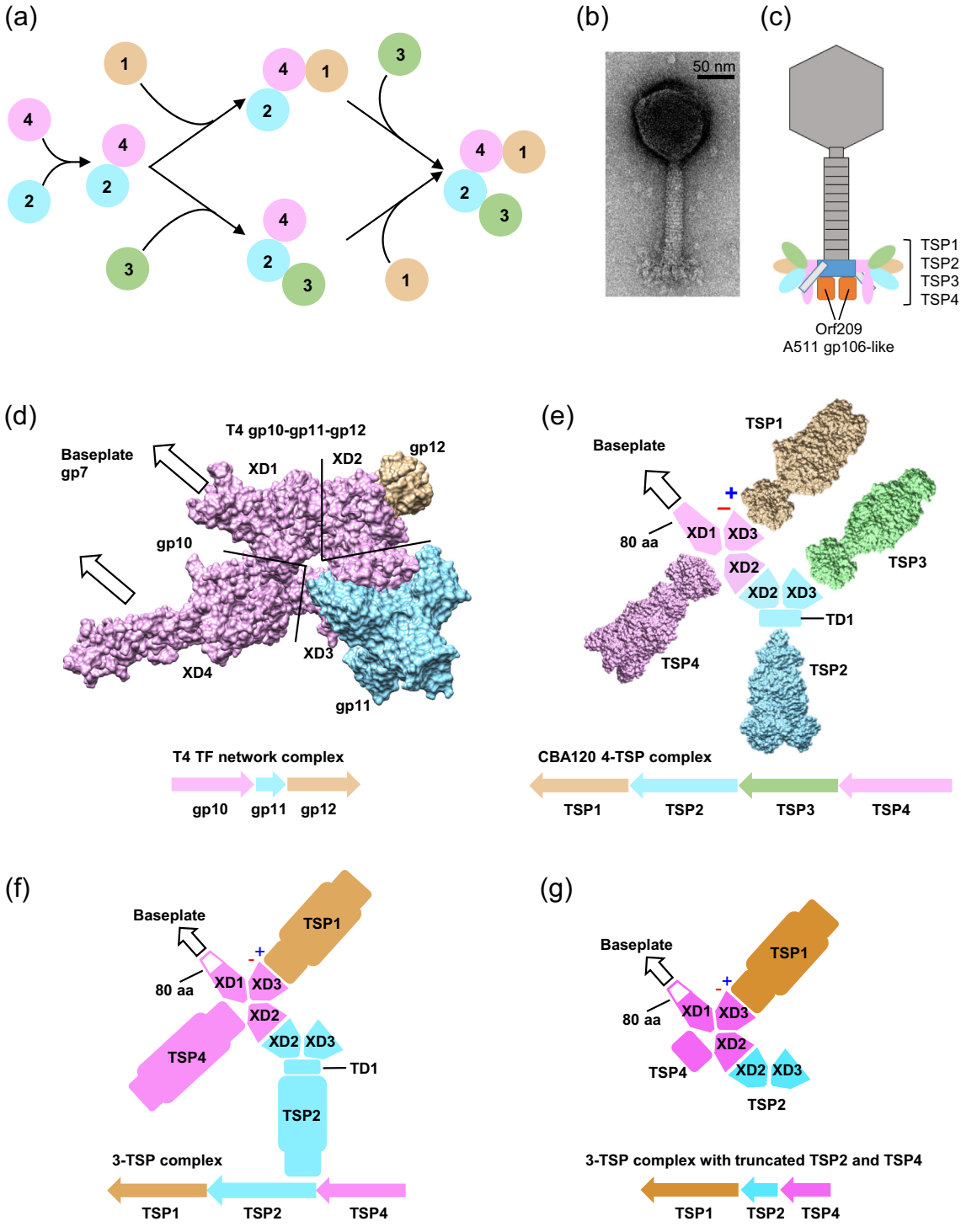
The effect of the modification of the cell surface by each of the four TSPs on the ability of CBA120 to infect its O157 host was evaluated with the help the soft agar overlay assay. *E. coli* NCTC 12900 cells (in soft agar) were mixed CBA120 and one of the following proteins—TSP2ΔN, TSP4ΔN, bovine serum albumin (BSA; negative control 1), or TSP3 (negative control 2) and plated on agar plates. The concentrations of the cells and the phage were chosen such that the uninhibited infection assay would yield about 200 plaques per plate (Fig. 8D, 8E). TSP2ΔN fully blocked the infection, BSA and TSP3 had no effect, and TSP4ΔN had a small but statistically significant inhibitory effect.

TSP2 binds the substrate in a cavity on the interface of two adjacent polypeptide chains

To identify the substrate-binding site on the surface of TSP2ΔN, it was co-crystallized with the O157 O-antigen. A well-resolved electron density of a ligand was located in a negatively charged pocket

on the interface of two adjacent chains (Fig. 9), consistent with the electrostatic potential distribution discussed earlier (Fig. 4). In the process of crystallization, the O-antigen was digested to the tetrasaccharide repeating unit found in the NMR analysis (see below, Oligosaccharide **1,** Fig. 10A). The atomic structure of the Glc-GalNAc-Rha4NAc-Fuc tetrasaccharide fit the electron density with a correlation coefficient of 0.93 (calculated in the real space [52]). The substrate-binding pocket of TSP2ΔN was not unusual and contained aromatic residues W539 and F483, hydrophobic L445, and polar Q426 (Fig. 9B), the types of residues often found in other sugar-binding sites [53,54].

The role of the two aspartic acids D506 and D571 located in the vicinity of the terminal fucose in the



catalysis was tested by mutagenesis. The D571A mutant was insoluble. The D571N mutant expressed in the soluble form but was so unstable that its activity could not be assayed. The D506A mutant expressed well and was soluble and stable. It bound to O157 *E. coli* similar to the wild-type protein. It cleaved the O157 O-antigen, and the pattern of the cleavage products on SDS-PAGE was identical to that of the wild type. Interestingly, point mutations in other residues of the substrate binding site—K536A, Q570A, and Y623F (Fig. 9B)—did not affect binding of TSP2 to O157 *E. coli* or cleavage of O157 O-antigen. Although these experiments failed to identify the catalytic site, they suggest that the actual substrate binding site spans more than one Glc-GalNAc-Rha4NAc-Fuc tetrasaccharide unit. Indeed, the cavity that binds the Glc-GalNAc-Rha4NAc-Fuc unit is a part of a large grove on the surface of the protein (Fig. 4B) that can accommodate the $(Glc\text{-}GalNAc\text{-}Rha4NAc\text{-}Fuc)_n$ polymer, the actual substrate of TSP2.

CBA120 infects *Salmonella enterica* serovar Minnesota, *E. coli* O77, and *E. coli* O78

Putative substrates of TSP1, TSP3, and TSP4 and, as a consequence, additional hosts of CBA120 were identified with the help of bioinformatic analysis that took into account a previous finding demonstrating that prophage TSP sequences correlate very well with the O-antigens of the cells that harbor them [55] (see Supplemental Information), This analysis suggested that TSP1 targets a surface polysaccharide of *S. enterica* serovar Minnesota or *Citrobacter freundii* and that the substrates of TSP3 and TSP4 are O77-like and O78 antigens, respectively.

Indeed, we found that CBA120 can infect *S. enterica* serovar Minnesota and *E. coli* O77 and O78 strains (Table 3). The titer of CBA120 on *S. enterica* serovar Minnesota and *E. coli* O78:H^- was similar to that on O157:H7 (NCTC 12900) to which it might have adapted due to the prior propagation. The plaques were of normal morphology. A slightly reduced titer with pinpoint plaques was observed on lawns of *E. coli* O77 strains (Table 3).

To further validate the derivation of the CBA120 host range from bioinformatic analysis, we examined the host range of phage S117 [56], a close relative of CBA120 with two very similar TSPs. Namely, S117 and CBA120 TSP1 sequences are 93% identical over their entire lengths (GenBank accession codes AXC40875.1 and YP_004957864.1, respectively), and the TSP2 proteins are 99% identical (GenBank codes AXC40876.1 and YP_004957865.1, respectively). We reasoned that these TSPs should allow S117 to infect *S. enterica* serovar Minnesota and *E. coli* O157. Indeed, S117 infected the two *S. enterica* serovar Minnesota strains listed in Table 3 with CBA120-like titers (2.00E+10 pfu/ml and 3.00E+10 pfu/ml on S.123286 and S.119883, respectively) and gave small clear plaques. The titer of S117 on *E. coli* O157 was 1.00E+9 pfu/ml or 50 fold lower than that of CBA120, and the plaques were pinpoint.

The other two TSPs of S117—TSP3 and TSP4 (GenBank codes AXC40877.1 and AXC40878.1, respectively)—are similar to their CBA120 counterparts (GenBank YP_004957866.1 and YP_004957867.1, respectively) only in their N-terminal domains. Namely, residues 1–165 of the TSP3 proteins of the two phages are 95% identical and residues 1–486 of the TSP4 proteins of CBA120 and S117 are 97% identical. The rest of the TSPs (the putative catalytic β-helical and C-terminal domains) are different. S117 did not infect any of the O77 or O78 strains tested.

TSP2, TSP3, and TSP4 cleave O157, O77 and O78 O-antigens, respectively

To fully understand how TSPs cleave O-antigens, we determined the structure of O-antigen digestion

Fig. 7. The assembly pathway and the structure of the TSP complex in CBA120 and other Viunalike viruses. (A) A schematic representation of the complex formation process. TSP2 and TSP4 have to interact first to become capable of binding of either of the two smaller TSPs. Binding of TSP1 or TSP3 to the TSP2–TSP4 complex is independent from each other and is not ordered. (B) An electron micrograph of CBA120. (C) A schematic representation of bacteriophage CBA120 derived from earlier EM images [33] and cryo-EM reconstruction (Sergey Nazarov's PhD Thesis, DOI: https://doi.org/10.5075/epfl-thesis-6507). (D) The structure of the gp10–gp11–gp12 complex in the molecular surface representation extracted from the atomic model of the T4 baseplate [41]. The four domains comprising gp10 are labeled XD1, XD2, XD3, and XD4. Only a small N-terminal part of the gp12 fiber (colored tan) is shown for clarity. (E) A model of the architecture of the CBA120 TSP1–TSP2–TSP3–TSP4 complex. TSP2 and TSP4 contain T4 gp10-like domains (XD1, XD2, XD3) that are either responsible for attachment of the entire gp10–gp11–gp12 complex to the baseplate (XD1) or for forming off-axis sites for binding trimeric gp11 and gp12 proteins (XD2 and XD3). The charge complementarity of the TSP1 N-terminal TD1 domain and the XD3 domain of TSP4 is shown with plus and minus symbols. The following viruses have a similar gene cluster and orthologous domains in their four tailspikes: Vil, SFP10, ΦSH19, phiSboM-AG3, Det7, ECML-4, Maynard, Marshall, vB_SalM_SJ2, vB_SalM_SJ3, and S117. TSP1 in vB_SalM_SJ3 consists of domains TD1 and TD2 only (the rest is absent). (F) The TSP cluster of PhaxI phage consists of three TSPs (TSP3 is missing). Each of the PhaxI TSPs contains a predicted enzymatic domain. (G) LIMEStone1 phage contains only one TSP (TSP1) with a putative enzymatic domain. TSP1 is attached to a complex comprising the N-terminal parts of TSP2 and TSP4. No gp10-like domains are predicted in TSP2.

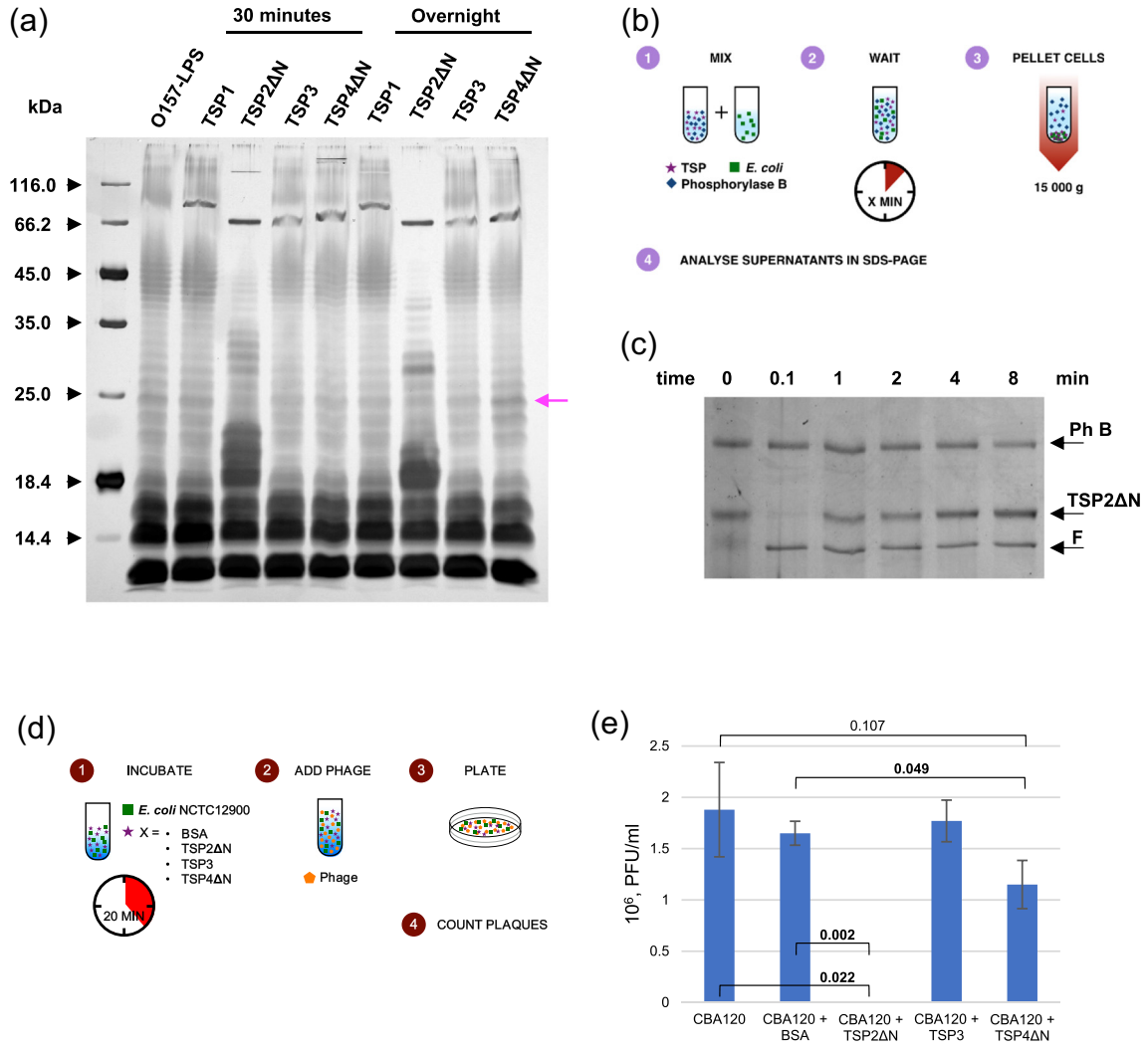


Fig. 8. Interaction of TSP proteins with the purified O157 LPS and the cell surface of *E. coli*. O157:H7. (A) Cleavage of O157 LPS by CBA120 TSP proteins. The cleavage was stopped by the addition of the Laemmli SDS buffer upon incubation with the corresponding TSP protein for 30 min and overnight and assayed by silver-stained SDS PAGE. The magenta arrow points to possible products of O157 O-antigen degradation by TSP4ΔN. (B) Schematic of the TSP-cell binding assay. (C) Binding of TSP2ΔN to *E. coli* O157:H7 assayed as described in panel B and visualized with Coomassie-stained SDS-PAGE. The binding is transient and TSP2 is released into the solution within 1 min. Phosphorylase B (the arrow labeled "Ph B") was used as a negative-binding and gel-loading control. The bacterial flagellin (the arrow labeled "F") detached from the cell surface during the assay. A control experiment without any added proteins also resulted in in the release of the flagellin and OmpC (confirmed by mass spectrometry). (D) A schematic of the assay that tests inhibition of CBA120 infection by TSPs. (E) Titers of CBA120 in the presence of TSP2ΔN, TSP3, TSP4ΔN, and bovine serum albumin (BSA) measured using a soft agar overlay assay. The error bars represent the standard deviation of plaque count in three (n = 3) independent experiments. TSP2ΔN completely inhibits plaque formation. TSP4ΔN shows a statistically significant reduction of plaque formation. The *p*-values are calculated for the Student's two-tail *t*-test with unequal variance.

reaction products by two-dimensional Nuclear Magnetic Resonance (NMR) spectroscopy and high-resolution electrospray ionization mass spectrometry (HR-ESI-MS). The O157 O-antigen was purified from *E. coli* NCTC 12900 (O157:H7), the same strain that was previously used in all other assays and crystallographic analysis. O77 and O78 antigens were available to us, and their structures were determined in earlier studies [57–60]. The activity of TSP1 toward its *S. enterica* serovar Minnesota O-antigen substrate was not analyzed.

NMR showed that TSP2, TSP3, and TSP4 are hydrolases supporting the assumption that their active sites are likely to be negatively charged (Fig. 10). TSP2ΔN cleaved the O157 O-antigen down to its repeating unit—a Glc-GalNAc-Rha4NAc-Fuc tetrasaccharide (called Oligosaccharide **1**) with an L-fucose residue at its reducing end. Details of the

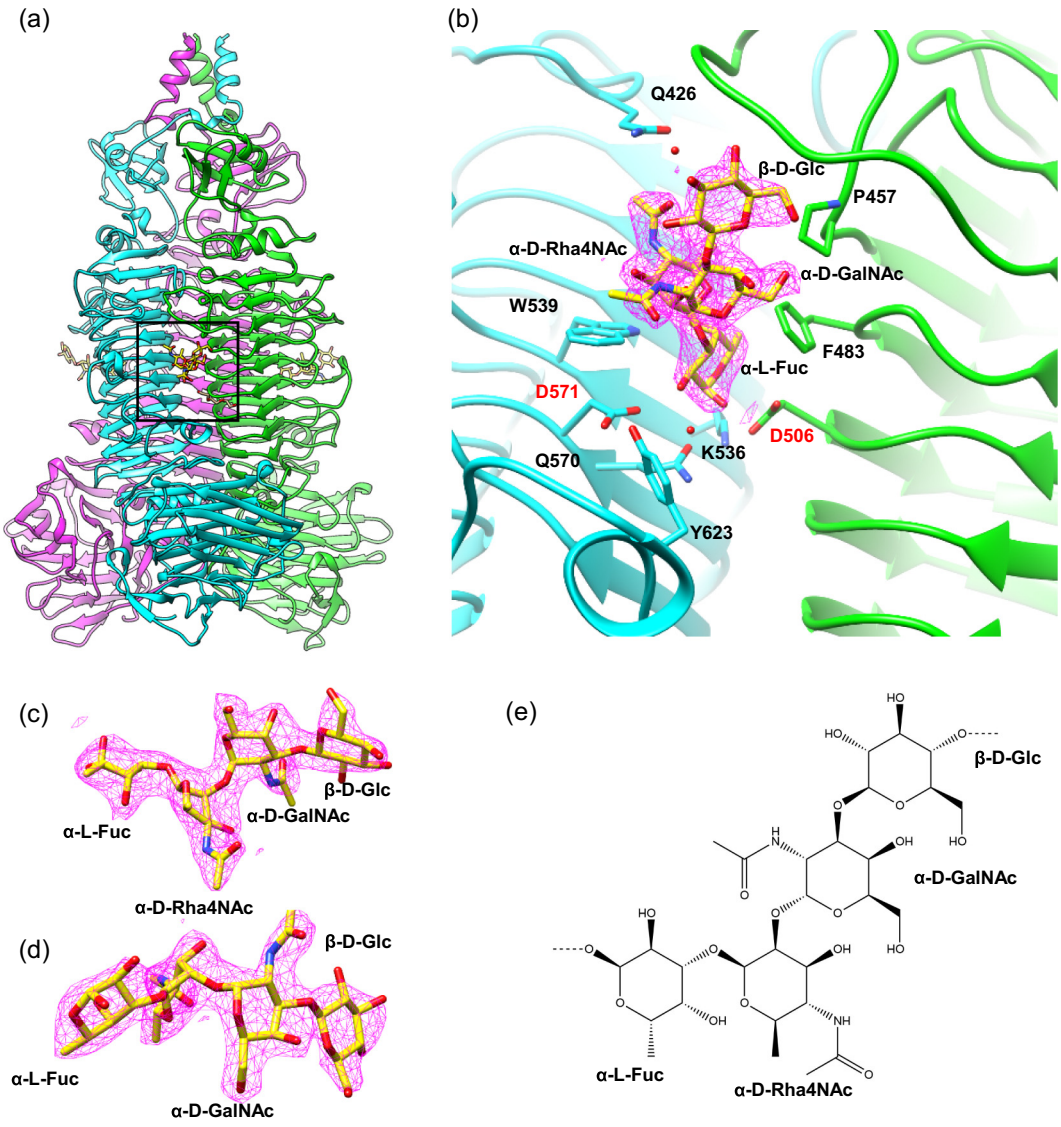


Fig. 9. Structure of TSP2ΔN with a repeating tetrasaccharide unit of the O157 O-antigen bound to it. (A) The location of the binding site in the TSP2ΔN structure. (B) A detailed view of the active site. Side chains of residues that participate in substrate binding are shown in the stick representation. The putative catalytic residues are labeled with red labels. The contour level of the $2F_o - F_c$ electron density map of the O157 O-antigen repeating tetrasaccharide unit is one standard deviation above the mean. (C and D) The structure and electron density map (contoured at in panel B) of the O157 O-antigen repeating tetrasaccharide unit are shown in two different orientations. (E) The chemical structure of the bound O157 O-antigen repeating tetrasaccharide unit.

NMR and MS analyses are given in the Supplemental Information. TSP3 cleaved the O77 O-antigen down to a dimer of the repeating tetrasaccharide unit, with a mannose at the reducing end (Oligosaccharide **4**, Fig. 10B). TSP4 cleaved the O78 O-antigen down to its tetrasaccharide repeating unit (Oligosaccharide **5**) with a GlcNAc residue at the reducing end (Fig. 10C). Interestingly, TSP4ΔN also cleaved the O157 O-antigen down to a mixture of oligomers (including the monomer) of Oligosaccharide **1** (Oligosaccharides **2** and **3**). The relative ratio of the tetrasaccharide monomer/dimer/trimer/higher oligomers was ~2/1/2/5. This was in agreement with its inhibitory function during infection and its weak digestion activity that could be detected in the SDS-PAGE analysis (Fig. 8).

The structure of Oligosaccharides **1** through **5** (Fig. 10) was confirmed by negative ion mode HR ESI MS (Supplemental Information). For TSP2 and the O157 substrate, a $[M\text{-}H]^-$ ion peak was detected with an *m*/*z* value of 715.28. It matched the calculated molecular mass of Oligosaccharide **1** (716.28 Da). For the TSP3 and O77 substrate, an $[M\text{-}H]^-$ ion peak at *m*/*z* of 1395.48 matched the

(a)

→4)-β-D-Glc*p*-(1→3)-α-D-Gal*p*NAc-(1→2)-α-D-Rha*p*4NAc-(1→3)-α-L-Fuc*p*-(1→		O157 OPS
β-D-Glc*p*-(1→3)-α-D-Gal*p*NAc-(1→2)-α-D-Rha*p*4NAc-(1→3)-L-Fuc	**1**	O157 OPS + TSP2ΔN O157 OPS + TSP4ΔN
β-D-Glc*p*-(1→3)-α-D-Gal*p*NAc-(1-[→2)-α-D-Rha*p*4NAc-(1→3)-α-L-Fuc*p*-(1→ -4)-β-D-Glc*p*-(1→3)-α-D-Gal*p*NAc-(1-]$_n$→2)-α-D-Rha*p*4NAc-(1→3)-L-Fuc	**2, 3**	O157 OPS + TSP4ΔN

(b)

→2)-β-D-Man*p*-(1→3)-α-D-Glc*p*NAc-(1→6)-α-D-Man*p*-(1→2)-α-D-Man*p*-(1→		O77 OPS
β-D-Man*p*-(1→3)-α-D-Glc*p*NAc-(1→6)-α-D-Man*p*-(1→2)-α-D-Man*p*-(1→ -2)-β-D-Man*p*-(1→3)-α-D-Glc*p*NAc-(1→6)-α-D-Man*p*-(1→2)-D-Man	**4**	O77 OPS + TSP3

(c)

→4)-β-D-Glc*p*NAc-(1→4)-β-D-Man*p*-(1→4)-α-D-Man*p*-(1→3)-β-D-Glc*p*NAc-(1→		O78 OPS
β-D-Glc*p*NAc-(1→4)-β-D-Man*p*-(1→4)-α-D-Man*p*-(1→3)-D-GlcNAc	**5**	O78 OPS + TSP4ΔN

Fig. 10. Structures of the *E. coli* O157, O77, and O78 O-antigens and the products of digestion of these polysaccharides by CBA120 TSPs. (A) O157 OPS, the intact O157 O-antigen. Oligosaccharide **1** (the repeating unit of O157 O-antigen) is the product of its digestion by TSP2ΔN. TSP4ΔN digested the O-antigen down to Oligosaccharide **1**, Oligosaccharide **2** (n = 1, a dimer of the repeating unit), Oligosaccharide **3** (n = 2, a trimer of the repeating unit), and larger oligosaccharides. (B) The O77 OPS and the product of its digestion by TSP3 (Oligosaccharide **4**, a dimer of the repeating unit). (C) The O78 OPS and the product of its digestion by TSP4ΔN (Oligosaccharide **5**).

calculated molecular mass of Oligosaccharide **4** (1396.49 Da), For TSP4 and the O78 substrate, an $[M-H]^-$ ion peak at *m*/*z* of 747.27 matched the calculated molecular mass of Oligosaccharide **5** (748.27 Da). For TSP4 and O157 substrate, $[M-H]^-$ ion peaks at *m*/*z* of 1413.55 and 2111.82 corresponded to a dimeric and trimeric form of Oligosaccharide **1**–Oligosaccharide **2** (1414.56 Da) and Oligosaccharide **3** (2112.83 Da), as expected.

Discussion

Inter-chain substrate-binding sites are common in TSPs

TSP2 binds its substrate and has its putative active site at the interface between two polypeptide chains. The putative substrate binding site in the other three CBA120 TSPs is also in a grove between the subunits (Figs. 2 and 4). Phage TSPs are likely to have evolved from monomeric β-helical proteins in which the active site is in a crevice on the face of the β-helix [15]. Many of the phage tailspikes studied earlier (e.g., in phage P22, Det7, or HK620) had the active site on the face of the β-helix [14,16,18]. Our work and more recent studies of Sf6 and ΦAB6 TSPs [17,22] show that an interchain active site is not uncommon. In these TSPs, the active site "migrated" from the intra-chain to an inter-chain one [15]. TSPs that possess two different active sites on the same molecule (intra- and inter-chain) could exist. Such a TSP would be able to recognize and digest or modify at least two different substrates.

T4 gp10 module is key to forming the branched TF/TSP structure

Beta-helical domains are widespread in phage TSPs and their occurrence in CBA120 is not surprising. More interesting is the repeat utilization of the phage T4 gp10-like module in the TSP2 and TSP4 tailspike proteins of CBA120 and in gp66 of phage G7C (Fig. 1) [8]. Similar to the T4 baseplate structure, where the gp10 module mediates off-axis attachment of two trimeric proteins gp11 and gp12 to the baseplate (Fig. 7D), this module is likely to be responsible for side or off-axis attachment of CBA120 and G7C TSPs giving the complex a branched appearance.

G7C gp66 contains two gp10-like domains, XD2 and XD3, but most likely only one of them is utilized because gp66 forms a 1:1 complex with gp63.1. In CBA120 TSPs, the situation is more complicated. Both longer CBA120 TSPs (TSP2 and TSP4) contain XD2 and XD3 domains each of which can potentially attach a trimeric tailspike (Figs. 1 and 7E).

Table 3. CBA120 host range agrees with experimentally confirmed specificity of its TSPs

Species	Database code or strain ID	Full serotype	Group	Titer (pfu/ml)	Plaque morphology (size and turbidity) and notes	TSP/O-antigen specificity
S. enterica Minnesota	JEO1794[a] S.123286			4.00E+10	Normal clear	TSP1/*S. enterica* O-antigen (putative)
S. enterica Minnesota	JEO1874 S.119883			4.00E+10	Normal clear	
E. coli	NCTC 12900[b]	O157:H7		5.00E+10	Normal clear	TSP2 (TSP4)/O157
E. coli	C344–80[c]	O77:H45	ETEC[d]	1.00E+08	Pinpoint	TSP3/O77
E. coli	C348–80	O77:H45	ETEC	1.00E+08	Pinpoint	
E. coli	C1106–98	O77:H9	A/EEC[e]	–	No single plaques (clear lysis spot in up to 10-fold dilutions)	
E. coli	C252–02	O77:H18	A/EEC	1.00E+08	Pinpoint	
E. coli	C710–09	O77:H34	EAggEC[f]	1.00E+08	Pinpoint	
E. coli	JEO632	O78[g]		3.00E+09	Small	TSP4/O78
E. coli	C493–88	O78:H10	ETEC	–	No single plaques (clear lysis spot in up to 1000-fold dilutions)	
E. coli	C497–88	O78:H10	ETEC	8.00E+08	Normal very turbid	
E. coli	C2209–99	O78:H$^-$	ETEC	3.00E+10	Normal clear	
E. coli	C882–10	O78:H2	EAggEC	5.00E+09	Small	
E. coli	C843–12	O78:H11	ETEC	–	No single plaques (clear lysis spot in up to 1000-fold dilutions)	

[a] Codes starting with JEO designate strains from John E. Olsen's database (Department of Veterinary and Animal Sciences, University of Copenhagen).
[b] NTCT stands for The National Collection of Type Cultures (Public Health England).
[c] Codes starting with C designate strains from Flemming Scheutz's database (Department of Bacteria, Parasites and Fungi, Statens Serum Institute).
[d] Enterotoxigenic *E. coli.*
[e] Attaching and effacing *E. coli.*
[f] Enteroaggregative *E. coli.*
[g] The H component of the serotype is unknown.

However, we showed that neither of them interacts with either of the shorter tailspikes, so no G7C-like tailspike pair of a (long TSP)-(short TSP) is formed. Instead, TSP2 and TSP4 must interact first to become competent for binding of any of the short TSPs (Table 2). The off-axis binding site in both longer TSPs is clearly dependent on the presence of the other long TSP. A chain or domain swapping event takes place within XD2 and XD3 in T4 gp10 [41]. It is possible that some form of interdigitation or chain swapping occurs when TSP2 and TSP4 interact to form the off-axis binding sites for TSP1 and TSP3. This type of interaction might make the assembly sequential and could reduce the amount of incorrectly assembled intermediates.

Interestingly, a gp10-like module is likely utilized in the TSP complex of K1E, K1–5, and SP6-like phages [31,61–64]. These phages carry two TSPs that are attached to two different sites on the adaptor protein gp37 [51]. HHpred suggests that the C-terminal part of gp37 (residues 160–316) has a structure similar to domains XD2 and XD3 [39]. Despite a low sequence identity (14%) and only 77.3% probability (as calculated by HHpred), the spatial arrangement of the attachment sites and the overall structure of the adaptor-tailspike complex of these phages matches those of the gp10–gp11–gp12 complex supporting the tentative bioinformatic prediction.

TSP1 likely interacts with the gp10_D3 domain of TSP4

The N-terminal TD1 domains of CBA120 TSP1 and G7C gp63.1 display similar positively charged patches on their surfaces that are most likely involved in complex formation (off-axis attachment to the longer TSPs) (Fig. 4) [8]. The calculated p*K*a values [65] of CBA120 TSP1 and G7C gp63.1 are 6.71 and 9.63, respectively (Fig. 1B and C). Their most likely partners are the XD3 domain of CBA120 TSP4 with a p*K*a of 3.73 and the XD2 domain of gp66 of G7C with a p*K*a of 3.89, respectively (Fig. 1B and C). Furthermore, the positively charged N-terminal domains of CBA120 TSP1 and G7C gp63.1 contain a buried Zn ion situated on the threefold axis [37]. Conservation of the surface properties and the buried Zn ion in the TSPs of two unrelated phages is rather remarkable, considering that neither the charge nor the Zn ion are conserved in CBA120 TSP3 despite the latter being more similar to TSP1 (70%/82% sequence identity/similarity) than gp63.1 is to TSP1 (70%/78% sequence identity/similarity) in terms of their amino acid sequence within this region (the first 166 residues). It is therefore possible that the charge and the buried Zn ion are a recurring theme in the assembly of branched TSPs in other G7C-like and Viunalike phages.

TSP4 attaches the TSP1–TSP2–TSP3–TSP4 complex to the baseplate

Bioinformatics and the experimental work described above make it possible to explain the organization of TSPs in CBA120 and other Viunalike phages (Fig. 7E, F, and G). The TSP gene cluster of CBA120 where each of the four TSPs contains a predicted enzymatic domain is common among CBA120 relatives (Fig. 7E). However, there are a number of phages that contain only three or even one TSP with a predicted enzymatic domain (e.g., LIMEstone 1 with its only 'full-fledged' TSP1) (Fig. 7G). TSP3 is missing in several phages, suggesting that it is dispensable for the assembly and function of the TSP complex (Fig. 7F). The N-terminal part of TSP2 appears to be universally conserved, even in phages with a single putatively enzymatically active TSP, suggesting that in the CBA120-like TSP complex, the TSP2–TSP4 interaction is required to form an off-axis binding site for TSP1 attachment.

The most conserved part of the TSP cluster is the N-terminal part of TSP4, which contains gp10-like domains XD1, XD2, and XD3, and an additional 80-residue extension upstream of the XD1 domain (Figs. 1 and 7E). The latter is likely to be responsible for the attachment of the entire cluster to the baseplate (Fig. 7D and E). TSP4 can be considered to be a direct ortholog of T4 gp10 in which the XD4 domain is missing in some phages or is replaced with an enzymatically active module (e.g., with a β-helix and C-terminal sugar-binding domain as in CBA120) (Fig. 7D and E). This supposition makes the organization of the TSP1–TSP2–TSP3–TSP4 complex in Viunalike phages analogous to that of the gp10–gp11–gp12 TF network in T4 because in the T4 baseplate gp12 and gp11 interact with domains XD2 and XD3 of gp10 and the whole complex is attached to the baseplate via the N-terminal domain XD1 of gp10 (Fig. 7E).

Materials and Methods

Cloning

Clontech HiFi DNA polymerase and the primer pairs listed in Table S1 were used to amplify TSP genes directly from a pure CBA120 phage sample (without DNA purification). The amplified fragments were cloned into the commercial (Novagen) and homemade vectors (pTSL, Genbank: KU314761 [41] and pEEVa2, a pTSL derivative, Genbank: MK533800), depending on the requirement of the experiment. To increase the yield of protein for crystallization, TSP1, TSP2ΔN, and TSP4ΔN carried a folding and expression enhancer SlyD tag at their N-termini. The tag was separated from the protein by a tobacco etch virus (TEV) protease cleavage site.

Protein expression and purification for crystallization

The TSPs were all expressed in *E. coli* B834 and purified to homogeneity prior to crystallization according to the same purification protocol. Briefly, the LB medium was inoculated with an overnight culture or fresh cells from an LB-agar plate. The culture was grown in an orbital shaker-incubator at 37 °C and 200 rpm until its optical density at 600 nm reached a value of about 0.6 to 0.8. The culture was cooled on ice prior to addition of isopropyl β-d-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM to induce protein expression. The culture was then incubated in an orbital shaker-incubator at 18 °C and 200 rpm overnight. The cells were harvested by centrifugation (15 min at 5000*g*) and resuspended in lysis buffer [20 mM Tris–HCl (pH 8.0), 150 mM NaCl] or in buffer IMAC A (which had the same composition but also contained 20 mM imidazole), depending on whether the sample was intended for metal-affinity purification. The cell suspension was lysed by sonication (the temperature was maintained below 10 °C during the lysis), and the cell debris was separated from the soluble fraction by centrifugation (35,000*g* for 20 min).

The overexpressed proteins were affinity purified using either 5 ml or 10 ml HisTrap HP columns (GE Healthcare) or a 5-ml StrepTrap HP column (GE Healthcare) depending on the expression volume and affinity tag. The proteins were eluted from the HisTrap resin with a step gradient of imidazole in Buffer IMAC A. The final concentration of imidazole was 250 mM. The loading and equilibration buffer for the StrepTrap chromatography was 50 mM Tris–HCl (pH 8.0), 150 mM NaCl, and the elution buffer contained 5 mM desthiobiothin. The eluted protein was dialyzed against a 5 mM Tris–HCl (pH 8.0) buffer overnight at 4 °C. This step was different for proteins that carried expression tags that needed to be removed (e.g., for crystallization of TSP1, TSP2ΔN, and TSP4ΔN). The TEV protease was added directly to the dialysis bag (the TEV:protein ratio was 1:50), and the mixture was dialyzed overnight at room temperature in the following buffer: 1 mM DTT, 0.5 mM EDTA, and 5 mM Tris–HCl (pH 8.0). The proteins were further purified by ion-exchange chromatography (MonoQ 10/100 GL; GE Healthcare) with a linearly increasing concentration of NaCl [Buffer A: 20 mM Tris–HCl (pH 8.0); Buffer B: 1 M NaCl, 20 mM Tris–HCl (pH 8.0)]. The final purification step was size exclusion chromatography (Superdex 200 16/60 HiLoad; GE Healthcare) in 10 mM Tris–HCl (pH 8.0) and 150 mM NaCl buffer. For crystallization, the proteins were concentrated to 10–20 mg/ml in the gel filtration buffer and stored at 4 °C.

Selenomethionine proteins were expressed by following a standard SelenoMet Medium (Molecular dimensions Limited) protein production protocol and purified following the procedure developed for the native protein.

Protein crystallization and data collection

For each protein, more than 1000 unique crystallization conditions (Jena Biosciences) were screened using Mosquito robot (TTP Labtech) and MRC2 plates. Individual hits were further optimized by a hanging drop method in 24-well plates. TSP2ΔN–O157 O-antigen complex was obtained by co-crystallization of TSP2ΔN and highly purified O157 LPS. TSP2ΔN was concentrated to high concentration (37 mg/ml) and incubated with 1% O157 LPS at 37 °C for 60 min prior to drop setting. Crystallization conditions are listed in Table S2.

Diffraction data were collected at the PXI and PXII beamlines at the Swiss Light Source (SLS PSI, Villigen, Switzerland, Table S2). The data sets were integrated with XDS [66,67]. The crystallographic phases were calculated from the single anomalous diffraction (SAD) experiment [42] using Shelx [68] and HKL2MAP [69] or using molecular replacement in Phaser [70]. The electron density maps were improved by Parrot, and an initial model was build either by Arp/Warp [71] or Buccaneer [72]. Structures were refined in the reciprocal space using Refmac5 [73] from the CCP4 suite [74] or Phenix_-Refine [75] from the PHENIX suite [76] for later refinement steps in iteration with real space refinement in COOT [77].

Isolation of O157 LPS (O-antigen plus the lipid tail)

A hot aqueous phenol extraction procedure was used to isolate the LPSs [78,79]. An overnight cell culture was pelleted by centrifugation (4500 *g*, 15 min at 20 °C), resuspended in 1/25th of the initial cell culture volume of lysis buffer [50 mM Tris–HCl (pH 8.0), 100 mM NaCl] and lysed by ultrasonic sonication. Lysozyme, DNase I, and RNase A were added to this mixture to standard recommended concentrations, and the sample was incubated at 37 °C for 60 min. Proteinase K was then added, and the sample was incubated for 60 min at 65 °C first and then overnight at 37 °C. The sample was mixed with two volumes of aqueous phenol, vigorously stirred, and incubated at 70 °C for 30 min. The sample was cooled down, transferred into a dialysis bag (Visking Dialysis Tubing, Type 36/32, Serva) without separating it into water-soluble and phenol fractions, and dialyzed against tap water until the phenol smell was gone. The insoluble fraction was separated by centrifugation (15,000 *g*, 20 min at 4 °C). The soluble fraction was combined with 10 volumes of 96% ethanol, NaCl was added to a final concentration of 0.5 M, and the mixture was incubated at −20 °C overnight. The LPS was pelleted by centrifugation at 15,000*g* for 20 min at 4 °C and dissolved in water, and the ethanol–NaCl precipitation procedure was repeated once again with the addition of 10 volumes of ethanol and NaCl to a final concentration 0.5 M. The precipitated LPS was dissolved in 1 ml of distilled water and dialyzed against distilled water overnight. The purified LPS was lyophilized.

Degradation of the O157 LPS with TSPs for SDS-PAGE analysis

The following procedure was used to perform degradation of the O157 LPS by TSPs and to analyze the degradation products with the help of SDS-PAGE. Lyophilized O157 LPS was first dissolved in water to obtain a solution with a concentration of 20 mg/ml. The reaction mixture had the following composition: (1) the buffer used in protein size exclusion chromatography [10 mM Tris–HCl (pH 8.0), 150 mM NaCl], (2) the lyophilized LPS at 2 mg/ml and (3) the protein at 5 μg/ml. The mixture was incubated at 37 °C for 30 min or overnight. The reaction was stopped by the addition of Laemmli buffer that was followed by incubation at 95 °C for 5 min. LPS (15 μg) was loaded onto the SDS-PAGE. The digestion products were visualized by silver staining (Thermo Fisher Scientific) of 15% SDS-PAGE.

Purification of O157 O-antigen and digestion with TSPs for NMR experiments

The O-antigen was separated from the lipid by boiling LPS samples (~100 mg each) with aqueous 2% acetic acid (4 mL, 100 °C, 2 h). The precipitated fraction was removed by centrifugation (12,000 *g*, 20 min). The supernatant was fractionated by size exclusion chromatography on Sephadex G-50 gel (GE Healthcare) in aqueous 0.1% acetic acid. The elution was monitored with a UV detector (Waters, USA) at a wavelength of 206 nm, and the high-molecular-mass species was selected for further analysis. Fractions containing O157 O-antigen were pooled together and lyophilized.

Purified O-antigen samples were dissolved in 50 ml Tris–HCl (pH 8.0) to a final concentration of 1 mg/ml. A TSP protein was added to this solution to a final concentration of 0.1 mg/ml, and the mixture was incubated overnight at 37 °C. The protein was denatured by incubation at 95 °C for 10 min and cooled to room temperature. The proteins were further digested with proteinase K (100 μg/ml) at 65 °C for 60 min. The O-antigen degradation products were fractionated by size exclusion chromatography on Sephadex G-25 gel in aqueous 0.1% acetic acid.

NMR spectroscopy

Lyophilized O-antigen samples were dissolved in 99.9% D_2O, re-lyophilized, and dissolved in 99.95% D_2O. The NMR analysis was performed using a

Bruker Avance II 600 MHz spectrometer. The O157 spectra were recorded at 30 °C, whereas the O77 and O78 were analyzed at 40 °C. Sodium 3-trimethylsilylpropanoate-2,2,3,3-d_4 (δ_H 0, δ_C −1.6) was used for calibration. Two-dimensional $^1H,^1H$ correlation spectroscopy (COSY), $^1H,^1H$ total correlation spectroscopy (TOCSY), and $^1H,^{13}C$ heteronuclear single-quantum coherence (HSQC) spectroscopy experiments were performed using standard Bruker software. The Bruker TopSpin 2.1 program was used to acquire and process the NMR data. A spin-lock time of 100 ms and a mixing time of 150 ms were used in the TOCSY and two-dimensional Rotating-frame nuclear Overhauser effect spectroscopy (ROESY) experiments, respectively. A 60-ms delay was used for evolution of long-range couplings to optimize $^1H,^{13}C$ heteronuclear multiple-bond correlation (HMBC) experiments for coupling constant $J_{H,C}$ 8 Hz.

HR ESI mass spectrometry

HR ESI mass spectrometry was performed in the negative ion mode using a maXis instrument (Bruker Daltonics). Oligosaccharide samples (~0.05 mg/ml) were dissolved in a 1:1 (v/v) water–acetonitrile mixture and injected with a syringe at a flow rate of 3 µl/min. The capillary entrance voltage was set at 3 kV and the shield voltage at −500 V. Nitrogen was used as the drying gas, and the interface temperature was set at 180 °C. The internal calibration was performed with ESI Calibrant Solution (Agilent).

TSP binding assay

NCTC 12900 O157:H7 *E. coli* were grown to an optical density of about 1.0 at 600 nm. A 6-ml fraction of this cell culture was pelleted by centrifugation (5000*g*, 5 min at 4 °C) and resuspended in 200 µl of 20 mM Tris–HCl (pH 8.0) and 150 mM NaCl buffer, which contained a TSP and rabbit muscle phosphorylase P (the loading and negative binding control protein, Sigma) each at a concentration of 0.05 mg/ml. The mixture was moved to 37 °C, and at different time points, the cells were pelleted down by centrifugation (20,000*g* for 30 s at room temperature) and a 30 µl sample of supernatant was drawn (Fig. 3). These samples were analyzed by 8% SDS-PAGE and Coomassie stain. A 6% gel was used for TSP3.

CBA120-TSP infection inhibition assay

Bacteriophage CBA120 was produced and purified according to the published procedure [34]. The phage titer was determined by a dilution series and a titer of 2.5×10^6 PFU/ml was chosen for the competition experiment. *E. coli* O157:H7 (NCTC 12900) were grown at 37 °C to an OD600 nm of 0.3 and placed on ice to slow down further growth. A TSP or BSA at 0.5 mg/ml [in 20 mM Tris–HCl (pH 8.0), 150 mM NaCl] was added to 100 µl of this cell suspension to a final concentration of 0.5 mg/ml, and the mixture was incubated at 37C for 20 min. This suspension (the entire volume) was then mixed with 10 µl of phage stock and 5 ml of top agar (that was kept at 42 °C). This mixture was then poured onto an agar plate and incubated overnight at 37 °C. The number of plaques was recorded. The experiment was performed in triplicate, and a Student's *t*-test was then used to evaluate the statistical outcome (Fig. 8D and E).

TSP complex formation and negative staining electron microscopy

For complex formation studies, the TSPs were expressed individually or co-expressed in B21 star cells (Thermo Scientific). The protein expression was carried out in 50 ml of LB as described above. The cells were harvested by centrifugation (5000*g* for 10 min at 4 °C). The cell pellet was resuspended in 5 ml of Buffer W [50 mM Tris–HCl (pH 8.0), 150 mM NaCl] and lysed by ultrasonic sonication. In case of co-expression, TSP complexes were purified by His- or Strep-tag affinity chromatography (1 ml HisTrap HP or StrepTrap HP column, GE Healthcare) followed by size exclusion chromatography (Superose 6 10/300 GL GE Healthcare). In case of separate expression of TSPs, cleared cell lysates of individually expressed TSPs were mixed in equal ratios (v/v) and dialyzed overnight into Buffer W at room temperature. The mixture was clarified by centrifugation, and TSP complexes were purified by affinity and size exclusion chromatography. In case of sequential complex formation, the lysate of a new TSP was mixed with the purified subcomplex, the solution was incubated overnight, and the new complex was then purified by affinity and size exclusion chromatography.

The concentration of the purified complexes was adjusted to approximately 0.02 mg/ml and applied to 200 mesh copper grids with carbon support film (Electron Microscopy Sciences). The sample was stained with freshly made 2% uranyl acetate and imaged on a JEOL 1400 microscope.

Host range studies

Phage infectivity against the *S. enterica* serovar Minnesota and *E. coli* O77 and O78 strains listed in Table 3 was determined using a procedure similar to the one published previously [56]. Namely, 10-fold serial dilutions (up to 10^{-7}) of the phage stocks in SM buffer [0.1 M NaCl, 8 mM $MgSO_4.7H_2O$, 50 mM Tris–HCl (pH 7.5)] were prepared. An overnight culture of strains (100 µl) grown in LB medium at

37 °C was mixed with 4 ml of molten overlay agar (LB medium with 0.6% agar), spread on 9 cm LA (LB with 1.2% agar) plates, and spotted with three drops (10 μl) of each phage dilution previously prepared. Following incubation overnight at 37 °C, plaques were counted and plaque-forming units per milliliter (pfu/ml) were calculated for each strain. Two independent replicates of the experiments were performed.

Accession numbers

Protein structures and associated diffraction data have been deposited to the Protein Data Bank under the following accession numbers: TSP2ΔN, 5W6P; TSP2ΔN with O157, 5W6S; TSP3, 5W6F; TSP4ΔN, 5W6H.

Acknowledgments

Purification of the O-antigen digestion products, NMR, and HR ESI MS analyses were supported by the Russian Science Foundation (Project No. 14-14-01042). The project was supported by the Swiss National Foundation Grant 310030_166383 and UTMB startup grant to P.G.L. We thank Drs. Osnat Herzberg and Daniel Nelson of the Institute for Bioscience and Biotechnology Research (University of Maryland) for sharing their data on substrate diversity of TSP3 that is not described in this report. We thank Flemming Scheutz (Department of Bacteria, Parasites and Fungi, The International Collaborating Centre for Reference and Research on Escherichia and Klebsiella, Statens Serum Institute, Copenhagen, Denmark) and John E. Olsen (Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark) for providing bacterial strains used in the determination of phage host range (Table 3).

Author Contributions: M.P. cloned and purified TSP1, TSP2, TSP3, and TSP4, and crystallized TSP2ΔN, TSP2ΔN–O157 complex, TSP3, and TSP4ΔN. M.P. also purified O-antigen from O157 cells, performed complex formation analysis and electron microscopy imaging, solved and refined all crystal structures, and did all phage-related experiments. S.A.B. helped M.P. with protein purification and structure determination of TSP3. M.M.S. cloned TSP2ΔN, TSP3, and TSP4ΔN. S.N. and N.M.I.T. contributed to structural and bioinformatic analysis of the CBA120 particle. N.S.P. advised M.P. on the purification of O157 LPS. N. P.A., A.S.S., A.O.C., and Y.A.K. performed purification, NMR spectroscopy, and mass spectrometry of the O-antigen digestion products. M.P. wrote the first draft of the paper, which was then rewritten by P.G.L. and edited by Y.A.K. M.G., Y.E.G., and L. B. analyzed the host range of CBA120 and S117 using bacterial strains typed by Flemming Scheutz and John E. Olsen. E.K. isolated and extensively characterized CBA120 and determined its unusual very narrow host range on the ECOR collection. All authors read and commented on the manuscript.

Conflict of Interest: The authors declare no conflict of interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2019.07.022.

Received 1 March 2019;
Received in revised form 10 July 2019;
Accepted 10 July 2019
Available online 17 July 2019

Keywords:
bacteriophage attachment;
host-cell recognition;
x-ray crystallography;
NMR;
bacterial surface polysaccharides

Current affiliation: M. Plattner, Institute of Medical Microbiology, University of Zurich, Gloriastrasse 30/32, 8006 Zurich, Switzerland.

‡Current affiliation: Y.E. Gencay, SNIPR Biome, Lersø Parkallé 44, 2100 Copenhagen, Denmark.

Abbreviations:
TSP, tailspike protein; TF, tail fiber; LPS, lipopolysaccharide; OPS, O-specific polysaccharide or O-antigen; gp, gene product; orf, open reading frame; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; HR-ESI-MS, high-resolution electrospray ionization mass spectrometry; RMSD, root mean square deviation; NMR, nuclear magnetic resonance.

References

[1] J.H. Wilson, R.B. Luftig, W.B. Wood, Interaction of bacteriophage T4 tail fiber components with a lipopolysaccharide fraction from *Escherichia coli*, J. Mol. Biol. 51 (1970) 423–434.

[2] B. Hu, W. Margolin, I.J. Molineux, J. Liu, Structural remodeling of bacteriophage T4 and host membranes during infection initiation, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) E4919–E4928.

[3] J. Bertozzi Silva, Z. Storms, D. Sauvageau, Host receptors for bacteriophage adsorption, FEMS Microbiol. Lett. 363 (2016).

[4] L. Wang, Q. Wang, P.R. Reeves, The variation of O antigens in gram-negative bacteria, Subcell Biochem. 53 (2010) 123–152.
[5] I. Lerouge, J. Vanderleyden, O-antigen structural variation: mechanisms and possible roles in animal/plant–microbe interactions, FEMS Microbiol. Rev. 26 (2002) 17–47.
[6] S.A. Buth, M.M. Shneider, D. Scholl, P.G. Leiman, Structure and analysis of R1 and R2 pyocin receptor-binding fibers, Viruses. 10 (2018).
[7] Y. Xiang, P.G. Leiman, L. Li, S. Grimes, D.L. Anderson, M.G. Rossmann, Crystallographic insights into the autocatalytic assembly mechanism of a bacteriophage tail spike, Mol. Cell 34 (2009) 375–386.
[8] N.S. Prokhorov, C. Riccio, E.L. Zdorovenko, M.M. Shneider, C. Browning, Y.A. Knirel, et al., Function of bacteriophage G7C esterase tailspike in host cell adsorption, Mol. Microbiol. 105 (2017) 385–398.
[9] P.G. Leiman, F. Arisaka, M.J. van Raaij, V.A. Kostyuchenko, A.A. Aksyuk, S. Kanamaru, et al., Morphogenesis of the T4 tail and tail fibers, Virol. J. 7 (2010) 355.
[10] A.L. Davidson, L. Cardarelli, L.G. Pell, D.R. Radford, K.L. Maxwell, Long noncontractile tail machines of bacteriophages, in: M.G. Rossmann, V.B. Rao (Eds.), Viral Molecular Machines, Springer, New York, 2012.
[11] S.R. Casjens, I.J. Molineux, Short noncontractile tail machines: adsorption and DNA delivery by podoviruses, in: M. G. Rossmann, V.B. Rao (Eds.), Viral Molecular Machines, Springer, New York, 2012.
[12] P.G. Leiman, M.M. Shneider, Contractile tail machines of bacteriophages, Adv. Exp. Med. Biol. 726 (2012) 93–114.
[13] D. Andres, C. Hanke, U. Baxa, A. Seul, S. Barbirz, R. Seckler, Tailspike interactions with lipopolysaccharide effect DNA ejection from phage P22 particles in vitro, J. Biol. Chem. 285 (2010) 36768–36775.
[14] S. Steinbacher, U. Baxa, S. Miller, A. Weintraub, R. Seckler, R. Huber, Crystal structure of phage P22 tailspike protein complexed with *Salmonella* sp. O-antigen receptors, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 10584–10588.
[15] P.G. Leiman, I.J. Molineux, Evolution of a new enzyme activity from the same motif fold, Mol. Microbiol. 69 (2008) 287–290.
[16] M. Walter, C. Fiedler, R. Grassl, M. Biebl, R. Rachel, X.L. Hermo-Parrado, et al., Structure of the receptor-binding protein of bacteriophage det7: a podoviral tail spike in a myovirus, J. Virol. 82 (2008) 2265–2273.
[17] J.J. Muller, S. Barbirz, K. Heinle, A. Freiberg, R. Seckler, U. Heinemann, An intersubunit active site between supercoiled parallel beta helices in the trimeric tailspike endorhamnosidase of *Shigella flexneri* phage Sf6, Structure. 16 (2008) 766–775.
[18] S. Barbirz, J.J. Muller, C. Uetrecht, A.J. Clark, U. Heinemann, R. Seckler, Crystal structure of *Escherichia coli* phage HK620 tailspike: podoviral tailspike endoglycosidase modules are evolutionarily related, Mol. Microbiol. 69 (2008) 303–316.
[19] J.E. Thompson, M. Pourhossein, A. Waterhouse, T. Hudson, M. Goldrick, J.P. Derrick, et al., The K5 lyase KflA combines a viral tail spike structure with a bacterial polysaccharide lyase mechanism, J. Biol. Chem. 285 (2010) 23963–23969.
[20] D. Schwarzer, C. Browning, K. Stummeyer, A. Oberbeck, M. Muhlenhoff, R. Gerardy-Schahn, et al., Structure and biochemical characterization of bacteriophage phi92 endosialidase, Virology. 477 (2015) 133–143.
[21] K. Stummeyer, A. Dickmanns, M. Muhlenhoff, R. Gerardy-Schahn, R. Ficner, Crystal structure of the polysialic acid-degrading endosialidase of bacteriophage K1F, Nat. Struct. Mol. Biol. 12 (2005) 90–96.
[22] M.J. Lai, K.C. Chang, S.W. Huang, C.H. Luo, P.Y. Chiou, C. C. Wu, et al., The tail associated protein of *Acinetobacter baumannii* phage PhiAB6 is the host specificity determinant possessing exopolysaccharide depolymerase activity, PLoS One 11 (2016), e0153361.
[23] D. Andres, Y. Roske, C. Doering, U. Heinemann, R. Seckler, S. Barbirz, Tail morphology controls DNA release in two Salmonella phages with one lipopolysaccharide receptor recognition system, Mol. Microbiol. 83 (2012) 1244–1253.
[24] J.E. Samson, A.H. Magadan, M. Sabri, S. Moineau, Revenge of the phages: defeating bacterial defences, Nat. Rev. Microbiol. 11 (2013) 675–687.
[25] S.J. Labrie, J.E. Samson, S. Moineau, Bacteriophage resistance mechanisms, Nat. Rev. Microbiol. 8 (2010) 317–327.
[26] G. Gasiunas, T. Sinkunas, V. Siksnys, Molecular mechanisms of CRISPR-mediated microbial immunity, Cell. Mol. Life Sci. 71 (2014) 449–465.
[27] R. Barrangou, J. van der Oost, Bacteriophage exclusion, a new defense system, EMBO J. 34 (2015) 134–135.
[28] T. Goldfarb, H. Sberro, E. Weinstock, O. Cohen, S. Doron, Y. Charpak-Amikam, et al., BREX is a novel phage resistance system widespread in microbial genomes, EMBO J. 34 (2015) 169–183.
[29] E. Harrison, A.J. Wood, C. Dytham, J.W. Pitchford, J. Truman, A. Spiers, et al., Bacteriophages limit the existence conditions for conjugative plasmids, MBio. 6 (2015), e00586.
[30] S.R. Williams, D. Gebhart, D.W. Martin, D. Scholl, Retargeting R-type pyocins to generate novel bactericidal protein complexes, Appl. Environ. Microbiol. 74 (2008) 3868–3876.
[31] D. Gebhart, S.R. Williams, D. Scholl, Bacteriophage SP6 encodes a second tailspike protein that recognizes *Salmonella enterica* serogroups C2 and C3, Virology. 507 (2017) 263–266.
[32] N.M.I. Taylor, M.J. van Raaij, P.G. Leiman, Contractile injection systems of bacteriophages and related systems, Mol. Microbiol. 108 (2018) 6–15.
[33] E.M. Adriaenssens, H.W. Ackermann, H. Anany, B. Blasdel, I.F. Connerton, D. Goulding, et al., A suggested new bacteriophage genus: "Viunalikevirus", Arch. Virol. 157 (2012) 2035–2046.
[34] E.M. Kutter, K. Skutt-Kakaria, B. Blasdel, A. El-Shibiny, A. Castano, D. Bryan, et al., Characterization of a Vil-like phage specific to *Escherichia coli* O157:H7, Virol. J. 8 (2011) 430.
[35] J. Novacek, M. Siborova, M. Benesik, R. Pantucek, J. Doskar, P. Plevka, Structure and genome release of Twort-like Myoviridae phage with a double-layered baseplate, Proc. Natl. Acad. Sci. U. S. A. 113 (2016) 9351–9356.
[36] R.C. Guerrero-Ferreira, M. Hupfeld, S. Nazarov, N.M. Taylor, M.M. Shneider, J.M. Obbineni, et al., Structure and transformation of bacteriophage A511 baseplate and tail upon infection of *Listeria* cells, EMBO J. 38 (2019).
[37] C. Chen, P. Bales, J. Greenfield, R.D. Heselpoth, D.C. Nelson, O. Herzberg, Crystal structure of ORF210 from *E. coli* O157:H1 phage CBA120 (TSP1), a putative tailspike protein, PLoS One 9 (2014), e93156.
[38] J. Soding, A. Biegert, A.N. Lupas, The HHpred interactive server for protein homology detection and structure prediction, Nucleic Acids Res. 33 (2005) W244–W248.
[39] V. Alva, S.Z. Nam, J. Soding, A.N. Lupas, The MPI bioinformatics toolkit as an integrative platform for advanced

protein sequence and structure analysis, Nucleic Acids Res. 44 (2016) W410–W415.
[40] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[41] N.M. Taylor, N.S. Prokhorov, R.C. Guerrero-Ferreira, M.M. Shneider, C. Browning, K.N. Goldie, et al., Structure of the T4 baseplate and its function in triggering sheath contraction, Nature. 533 (2016) 346–352.
[42] B.C. Wang, Resolution of phase ambiguity in macromolecular crystallography, Methods Enzymol. 115 (1985) 90–112.
[43] G. Davies, B. Henrissat, Structures and mechanisms of glycosyl hydrolases, Structure. 3 (1995) 853–859.
[44] Greenfield J, Shang X, Luo H, Zhou Y, Heselpoth RD, Nelson DC, et al. Structure and tailspike glycosidase machinery of ORF212 from *E. coli* O157:H7 phage CBA120 (TSP3). Scientific reports. 2019;9:7349.
[45] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, J. Mol. Biol. 372 (2007) 774–797.
[46] M.J. Gage, A.S. Robinson, C-terminal hydrophobic interactions play a critical role in oligomeric assembly of the P22 tailspike trimer, Protein Sci. 12 (2003) 2732–2747.
[47] L. Holm, L.M. Laakso, Dali server update, Nucleic Acids Res. 44 (2016) W351–W355.
[48] A.K. Shrive, A.M. Metcalfe, J.R. Cartwright, T.J. Greenhough, C-reactive protein and SAP-like pentraxin are both present in *Limulus polyphemus* haemolymph: crystal structure of Limulus SAP, J. Mol. Biol. 290 (1999) 997–1008.
[49] C.A. Consortium, Ten years of CAZypedia: a living encyclopedia of carbohydrate-active enzymes, Glycobiology. 28 (2018) 3–8.
[50] G.C. Lander, R. Khayat, R. Li, P.E. Prevelige, C.S. Potter, B. Carragher, et al., The P22 tail machine at subnanometer resolution reveals the architecture of an infection conduit, Structure. 17 (2009) 789–799.
[51] P.G. Leiman, A.J. Battisti, V.D. Bowman, K. Stummeyer, M. Muhlenhoff, R. Gerardy-Schahn, et al., The structures of bacteriophages K1E and K1-5 explain processive degradation of polysaccharide capsules and evolution of new host specificities, J. Mol. Biol. 371 (2007) 836–849.
[52] I.J. Tickle, Statistical quality indicators for electron-density maps, Acta Crystallogr D Biol Crystallogr. 68 (2012) 454–467.
[53] P.J. Simpson, H. Xie, D.N. Bolam, H.J. Gilbert, M.P. Williamson, The structural basis for the ligand specificity of family 2 carbohydrate-binding modules, J. Biol. Chem. 275 (2000) 41137–41142.
[54] L. von Schantz, M. Hakansson, D.T. Logan, B. Walse, J. Osterlin, E. Nordberg-Karlsson, et al., Structural basis for carbohydrate-binding specificity—a comparative assessment of two engineered carbohydrate-binding modules, Glycobiology. 22 (2012) 948–961.
[55] S.R. Casjens, J.H. Grose, Contributions of P2- and P22-like prophages to understanding the enormous diversity and abundance of tailed bacteriophages, Virology. 496 (2016) 255–276.
[56] Y.E. Gencay, M. Gambino, T.F. Prussing, L. Brondsted, The genera of bacteriophages and their receptors are the major determinants of host range, Environ Microbiol 21 (6) (2019) 2095–2111.
[57] Knirel YA, Valvano MA. Bacterial Lipopolysaccharides. Structure, Chemical Synthesis, Biogenesis and Interaction With Host Cells. 1st ed. Vienna: Springer-Verlag Wien; 2011. p. XI, 440.
[58] B. Liu, A. Furevi, A.V. Perepelov, X. Guo, H. Cao, Q. Wang, et al., Structure and genetics of *Escherichia coli* O antigens, FEMS Microbiol Rev (2019) (in press).
[59] H. Yildirim, A. Weintraub, G. Widmalm, Structural studies of the O-polysaccharide from the *Escherichia coli* O77 lipopolysaccharide, Carbohydr. Res. 333 (2001) 179–183.
[60] P.E. Jansson, B. Lindberg, G. Widmalm, K. Leontein, Structural studies of the *Escherichia coli* O78 O-antigen polysaccharide, Carbohydr. Res. 165 (1987) 87–92.
[61] D. Scholl, S. Rogers, S. Adhya, C.R. Merril, Bacteriophage K1-5 encodes two different tail fiber proteins, allowing it to infect and replicate on both K1 and K5 strains of *Escherichia coli*, J. Virol. 75 (2001) 2509–2515.
[62] A.T. Dobbins, M. George Jr., D.A. Basham, M.E. Ford, J.M. Houtz, M.L. Pedulla, et al., Complete genomic sequence of the virulent *Salmonella* bacteriophage SP6, J. Bacteriol. 186 (2004) 1933–1944.
[63] D. Scholl, S. Adhya, C.R. Merril, Bacteriophage SP6 is closely related to phages K1-5, K5, and K1E but encodes a tail protein very similar to that of the distantly related P22, J. Bacteriol. 184 (2002) 2833–2836.
[64] K. Stummeyer, D. Schwarzer, H. Claus, U. Vogel, R. Gerardy-Schahn, M. Muhlenhoff, Evolution of bacteriophages infecting encapsulated bacteria: lessons from *Escherichia coli* K1-specific phages, Mol. Microbiol. 60 (2006) 1123–1135.
[65] M.R. Wilkins, E. Gasteiger, A. Bairoch, J.C. Sanchez, K.L. Williams, R.D. Appel, et al., Protein identification and analysis tools in the ExPASy server, Methods Mol. Biol. 112 (1999) 531–552.
[66] W. Kabsch, Integration, scaling, space-group assignment and post-refinement, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 133–144.
[67] W. Kabsch, XDS, Acta crystallographica Section D Biological crystallography. 66 (2010) 125–132.
[68] G.M. Sheldrick, Experimental phasing with SHELXC/D/E: combining chain tracing with density modification, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 479–485.
[69] T. Pape, T.R. Schneider, HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs, J. Appl. Crystallogr. 37 (2004) 843–844.
[70] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser crystallographic software, J. Appl. Crystallogr. 40 (2007) 658–674.
[71] G. Langer, S.X. Cohen, V.S. Lamzin, A. Perrakis, Automated macromolecular model building for x-ray crystallography using ARP/wARP version 7, Nat. Protoc. 3 (2008) 1171–1179.
[72] K. Cowtan, The Buccaneer software for automated model building. 1. Tracing protein chains, Acta Crystallogr D Biol Crystallogr. 62 (2006) 1002–1011.
[73] G.N. Murshudov, P. Skubak, A.A. Lebedev, N.S. Pannu, R.A. Steiner, R.A. Nicholls, et al., REFMAC5 for the refinement of macromolecular crystal structures, Acta Crystallogr D Biol Crystallogr. 67 (2011) 355–367.
[74] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, et al., Overview of the CCP4 suite and current developments, Acta Crystallogr. D Biol. Crystallogr. 67 (2011) 235–242.
[75] P.D. Adams, P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, et al., PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta Crystallogr D Biol Crystallogr. 66 (2010) 213–221.
[76] P.D. Adams, P.V. Afonine, G. Bunkoczi, V.B. Chen, N. Echols, J.J. Headd, et al., The Phenix software for automated

determination of macromolecular structures, Methods. 55 (2011) 94–106.
[77] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr D Biol Crystallogr. 66 (2010) 486–501.
[78] O. Westphal, K. Jann, Bacterial lipopolysaccharides extraction with phenol–water and further applications of the procedure, Methods in Carbohydrate Chemistry. (1965) 83–91.
[79] N.P. Arbatsky, M. Wang, A.S. Shashkov, A.O. Chizhov, L. Feng, Y.A. Knirel, et al., Structure of the O-polysaccharide of *Cronobacter sakazakii* O2 with a randomly O-acetylated L-rhamnose residue, Carbohydr. Res. 345 (2010) 2090–2094.