# Atomic Model of the Tailed Lactococcal Phage TP901-1 as Predicted by AlphaFold2: Revelations and Limitations

Jennifer Mahony , Adeline Goulet , Douwe Van Sinderen , Christian CAMBILLAU *

Posted Date: 13 November 2023

doi: 10.20944/preprints202311.0819.v1

Keywords: bacteriophage; virion structure; Lactococcus; structural biology; Alphafold; P335; TP901-1

*Article*

# Atomic Model of the Tailed Lactococcal Phage TP901-1 as Predicted by AlphaFold2: Revelations and Limitations

**Jennifer Mahony [1], Adeline Goulet [2], Douwe van Sinderen [1]* and Christian Cambillau [1,2]***

[1] School of Microbiology & APC Microbiome Ireland, University College Cork, Cork, Ireland
[2] Laboratoire d'Ingénierie des Systèmes Macromoléculaires (LISM), Institut de Microbiologie, Bioénergies et Biotechnologie (IMM), Aix-Marseille Université – CNRS, UMR 7255, Marseille, France
* Correspondence: CC : ccambillau@imm.cnrs.fr. DvS : d.vansinderen@ucc.ie

**Abstract:** Bacteria are engaged in a constant battle against preying viruses, called bacteriophages (or phages). These remarkable nano-machines pack and store their genomes in a capsid and inject it into the cytoplasm of their bacterial prey following specific adhesion to the host cell surface. Tailed phages possessing dsDNA genomes are the most abundant phages in the bacterial virosphere, particularly those with long, non-contractile tails. All tailed phages possess a nano-device at their tail tip that specifically recognizes and adheres to a suitable host cell surface receptor, being proteinaceous and/or saccharidic. Adhesion devices of tailed phages infecting Gram-positive bacteria are highly diverse and, for the majority, remain poorly understood. Their long, flexible and multi-domain encompassing tail limits experimental approaches to determine their complete structure. We have previously shown that the recently developed protein structure prediction program AlphaFold2 can overcome this limitation by predicting the structures of phage adhesion devices with confidence. Here, we extend this approach and employ AlphaFold2 to determine the structure of a complete phage, the lactococcal P335 phage TP901-1. Herein we report the structures of its capsid and neck, its extended tail, and the complete adhesion device, the baseplate, which was previously partially determined by X-ray crystallography.

**Keywords:** bacteriophage; virion structure; *Lactococcus*; structural biology; Alphafold; P335; TP901-1

## 1. Introduction

Bacteriophages are the most abundant biological entity on Earth with an estimated $10^{31}$ particles being present on our planet. Among these, tailed bacteriophages with double-stranded DNA genomes are the dominantly observed and reported in literature [1]. Until recently, tailed phages were classified within the Caudovirales order and described based on their morphology, possessing either a short (*Podoviridae*) or long tail that is contractile (*Myoviridae*) or non-contractile (*Siphoviridae*). Given the extent of their genetic, host species and ecological diversity, the taxonomy of phages has recently been re-established to reflect these evolutionarily-relevant factors and to remove reliance on morphology-based assignments [2]. The current viral class Caudoviricetes embraces 14 families assigned to one of four orders, while a further 33 families have been identified that have yet to be assigned to an order. Irrespective of the (sequence-based) taxonomy of phages, their morphology has a significant bearing on the types of interactions they will establish with their respective hosts [3, 4].

Phages with long, non-contractile tails (formerly called the *Siphoviridae*) are the most abundant in nature and several of these have become models to study phage-host interactions including the *Lactococcus* phages TP901-1 and p2, *Streptococcus thermophilus* phage STP1, *Bacillus subtilis* phage SPP1 and *Escherichia coli* phages lambda and T5 [5-10]. These phages are genetically distinct; however, in many cases, functional modules within their genomes are syntenic [11]. Furthermore, despite sequence disparity between phages of the same or distinct bacterial host species, the structure of their components is often conserved [4, 12-15]. The temperate lactococcal P335 phage TP901-1 is

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0819.v1



particularly well characterized with respect to its 3-dimensional structure and the interactions with its host [5, 16-18]. Furthermore, it has been employed as a model to understand the phenomenon of lysogeny and the factors that underpin the lytic-lysogenic lifestyle decision process [19-21]. TP901-1 structural module encompasses 22 genes which are responsible for the biosynthesis of the phage head, tail, baseplate and DNA packaging machinery (Figure 1).

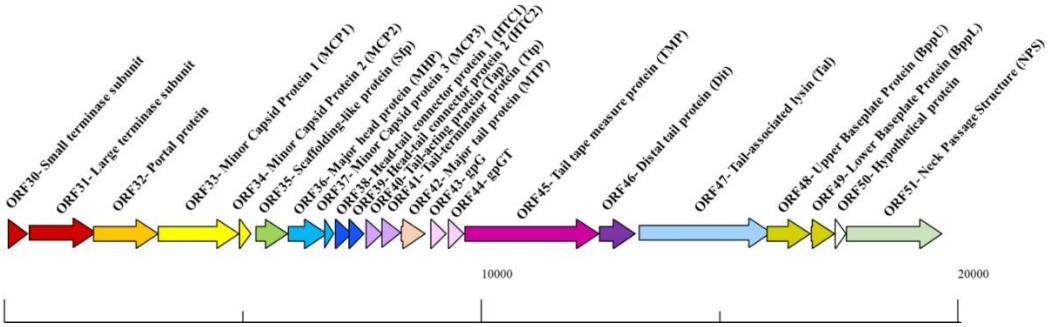


**Figure 1. Schematic representation of the structural module of TP901-1**. The functions are indicated above the arrows and the scale bar is presented at the base of the schematic in base pairs (bp). This figure and associated functions are based on those described in reference [17].

TP901-1 possesses an isometric capsid and a long non-contractile tail of approximately 135 nm [22]. At the distal end of its tail, TP901-1 presents a large, multi-protein, adhesion device termed a baseplate, which is comprised of the tail tape measure protein (TMP), distal tail protein (Dit), tail-associated lysin (Tal), and upper (BppU) and lower (BppL) baseplate proteins [5, 23]. The BppL component is the *bona fide* receptor binding protein (RBP) and is present as 18 trimers in the whole baseplate structure, which likely recognize and bind to cell wall polysaccharides presented on the surface of the host lactococcal cell. Its exquisite host specificity is based on heterologous expression of *Lactococcus lactis* genetic region encoding host-specific glycosyltransferases associated with cell wall polysaccharide biosynthesis [24].

The adhesion device of TP901-1 is in an infection-ready conformation and does not require activation by divalent cations such as calcium as is the case for certain tailed phages including the lactococcal phage p2 [25, 5]. Mutational analyses of TP901-1 genes coding for the capsid and tail proteins has provided insights into the assembly of the mature virion and the function of several previously uncharacterized gene products [17]. Furthermore, detailed genetic and structural analysis of the tail and baseplate proteins have rendered TP901-1 one of the best characterized phages capable of infecting Gram-positive bacteria [5, 22, 26].

The recent development of AlphaFold has significantly enhanced and advanced our capabilities to predict individual and multi-protein structures [27-29] and has transformed the field of structural biology. The ever-increasing number of phage genome and protein sequences and protein structures in public databases reinforces the need of reliable and rapid methods for the determination of (multi-component) protein structures and associated protein functionality. In the present study, we present predicted structures of TP901-1 capsid and neck, its extended tail, and its complete baseplate structure, which was previously partially determined by X-ray crystallography. AlphaFold predicted reliable structures of large parts of the phage in most cases, yet was unsuccessful in a small number of cases in which chaperone-aided folding is probably the reason for such structure prediction failure.

## 2. Materials and Methods

We performed predictions using either a Colab notebook running AlphaFold v2.3.1 (https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb) or HPC resources from GENCI-IDRIS running AlphaFold v2.3.1. The pLDDT values of predicted structure, stored in the pdb file as B-factors, as well as the PAE were plotted and are shown in Supplementary Figure S1. The final predicted protein or domain structures were submitted to the Dali server [30] to identify the closest structural homologs in the PDB. Visual representations of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0819.v1



structures were prepared with ChimeraX [31]. Coot [32, 33] was used to assemble predictions and visually analyze the predictions.

## 3. Results

### *3.1. The TP901-1 capsid*

Phage capsids are robust containers that carry and protect the viral genome packaged within its internal cavity [33]. Previously, the phage TP901-1 capsid structure was determined by negative staining Electron Microscopy (nsEM) and single particle analysis applying icosahedral symmetry at 15-Å resolution [22]. The mature capsid (660 Å wide) assembles 60 hexamers (hexons) and 11 pentamers (pentons) of the ORF36 major capsid protein (MCP) with a T=7 symmetry (EMD-2133). The 12th penton is replaced by a dodecamer of the portal protein. Here, we predicted the structures of the MCP hexamer and pentamer with confident scores, except for the N-terminus residues 1 to 15 (Figure 2). In both assemblies, the capsid MCP monomer contains classical domains of phage capsid protein including from the N-terminus to the C-terminus the β-hairpin "E-loop", the "P-domain" β-sheet, and a central "A-domain" formed of β-strands 4-6 and α-helices 3-5 (Figure 2A,C). While the A-domains form compact structures in the hexon and penton, the looser E-loops cover an adjacent monomer (Figure 2B,D).

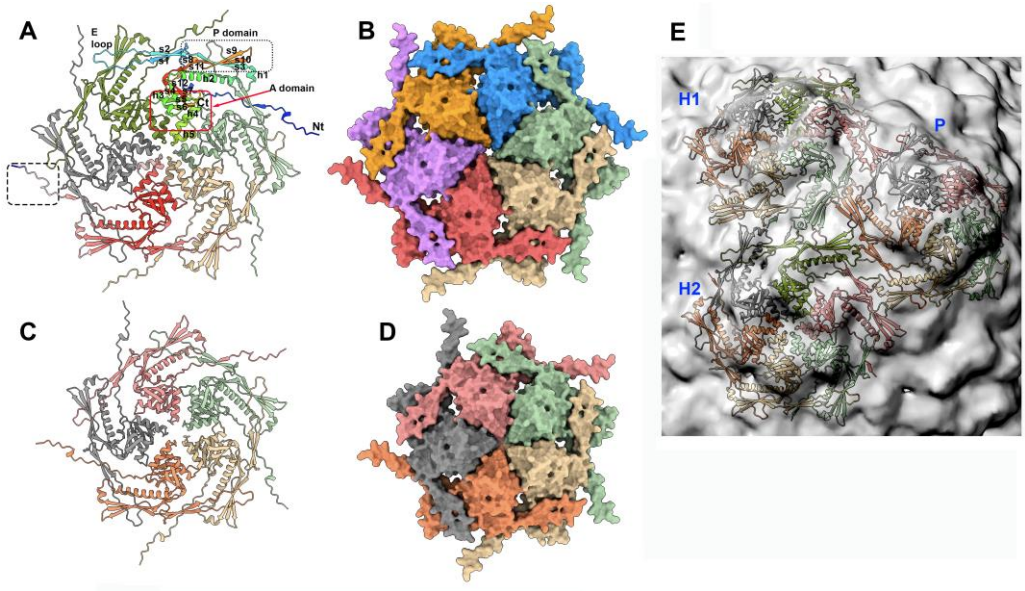


**Figure 2. Capsid's hexon and penton structures.** (**A**) Ribbon view of the predicted hexon structure. One monomer is rainbow colored and its domains are labeled. The monomer at the figure's bottom is colored according to pLDDT (quality of the prediction from blue to red). Note the low-confidence structure of the N-terminus. (**B**) Surface representation of the hexon (same orientation and scale as in (A)). (**C**) Ribbon view of the predicted penton structure. (**D**) Surface representation of the penton (same orientation and scale as in (C)). (**E**) Ribbon view of the structures of two hexons (H1 and H2) and one penton (P) fitted in the nsEM 3D reconstruction (EMD-2133).

In the entire capsid, each penton is surrounded by five hexons. Therefore, in order to analyse protein contacts at the hexon-hexon and hexon-penton interfaces, we attempted to predict the structures of two hexons and of one hexon with one penton by feeding AlphaFold with 12 and 11 MCP sequences, respectively. However, in both cases AlphaFold returned separated hexons and penton, thus limiting the possibility to examine the interfaces. In order to overcome this limitation and to evaluate the predictions in the context of the whole capsid, we fitted two predicted hexons and one penton (in the absence of residues 1-15) in the capsid nsEM 3D reconstruction (cross-correlation values of 0.81 (hexon) and 0.78 (penton) (Figure 2E). The TP901-1 capsid is structurally very close to

 doi:10.20944/preprints202311.0819.v1



that of the coliphage HK97 [35] as reported by Dali (PDB ID 1ohg; Z-score=18.1; rmsd=3.0 Å for 248 aligned residues of 280 residues in total).

*3.2. The procapsid scaffolding protein*

Procapsids (virion precursors without DNA)are assembled from the dodecameric portal by the addition of MCP hexons and pentons [36]. This process is promoted by scaffolding proteins that have been proposed to establish bridges between capsid MCPs [37, 38]. Little is known, however, on the structure of these scaffolding proteins alone and when they are attached to the MCPs. The procapsid structure of the *Staphylococcus aureus* phage 80α revealed the presence of a α-helix attached to each MCP in the internal face of the procapsid [37] (Supplementary Figure S2A,B). This short α-helix, which belongs to the C-terminal end of the scaffolding protein, is only a small portion of the phage 80α's full-length scaffolding protein (209 residues). The putative scaffolding protein of TP901-1 is slightly longer than that of 80α at its C-terminus (220 residues) (Supplementary Figure S2B). Therefore, we predicted their structures, which assemble dimers with good statistics (Supplementary Figure S2C). However, while the 80α's scaffolding protein shows an unstructured C-terminal end, the one of TP901-1 shows three well-folded C-terminal α-helices (Supplementary Figure S2d). We also predicted the structure of the complex between these C-terminal helices and a MCP hexon, and we found that each MCP is bound to this helical domain on the internal face of the hexon (Supplementary Figure S2E,F).

*3.3. Dodecameric portal and adaptor and the hexameric stopper.*

3.3.1. The dodecameric portal (ORF32)

Phage portals are dodecameric assemblies that replace a penton in the capsid structure [39-41]. They are bound to a dodecameric adaptor that is also connected to the hexameric stopper. The portal/adaptor/stopper complex is often referred to as the neck or the genome gatekeeper [42]. [41]. The TP901-1 portal dodecameric assembly comprises 5484 residues in total (12x457 residues), which is beyond the prediction capability of AlphaFold currently limited to ~4000 residues. Therefore, we used a multi-step approach to overcome this limit and obtain the structures of the portal dodecamer. First, we predicted the structure of a portal trimer to identify the regions of the portal protein that are not involved in self-assembly, essentially the N-terminal (1-64) and the C-terminal (380-457) ends, and define "short" versions of the portal protein (316 residues) and of its dodecamer (3,792 residues) for structure prediction. We confirmed that the predicted structures of the monomers in the "short" dodecamer superimposed perfectly on that of the predicted full-length portal. Then, a full-length dodecameric portal was formed by superimposing full-length monomers on each subunit of the "short" dodecamer with Coot [33].

Phage TP901-1 portal dodecamer has dimensions of ~160 Å (external diameter ) and ~110 Å (height excluding the C-terminus) with a central channel of ~30Å in diameter (Figure 3A; Figure 4A,B). As with other phage portal proteins, each monomer contains the domains named the wing, stem, clip and crown [39-42] (Figure 3B). The wing is the central core of the portal.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0819.v1



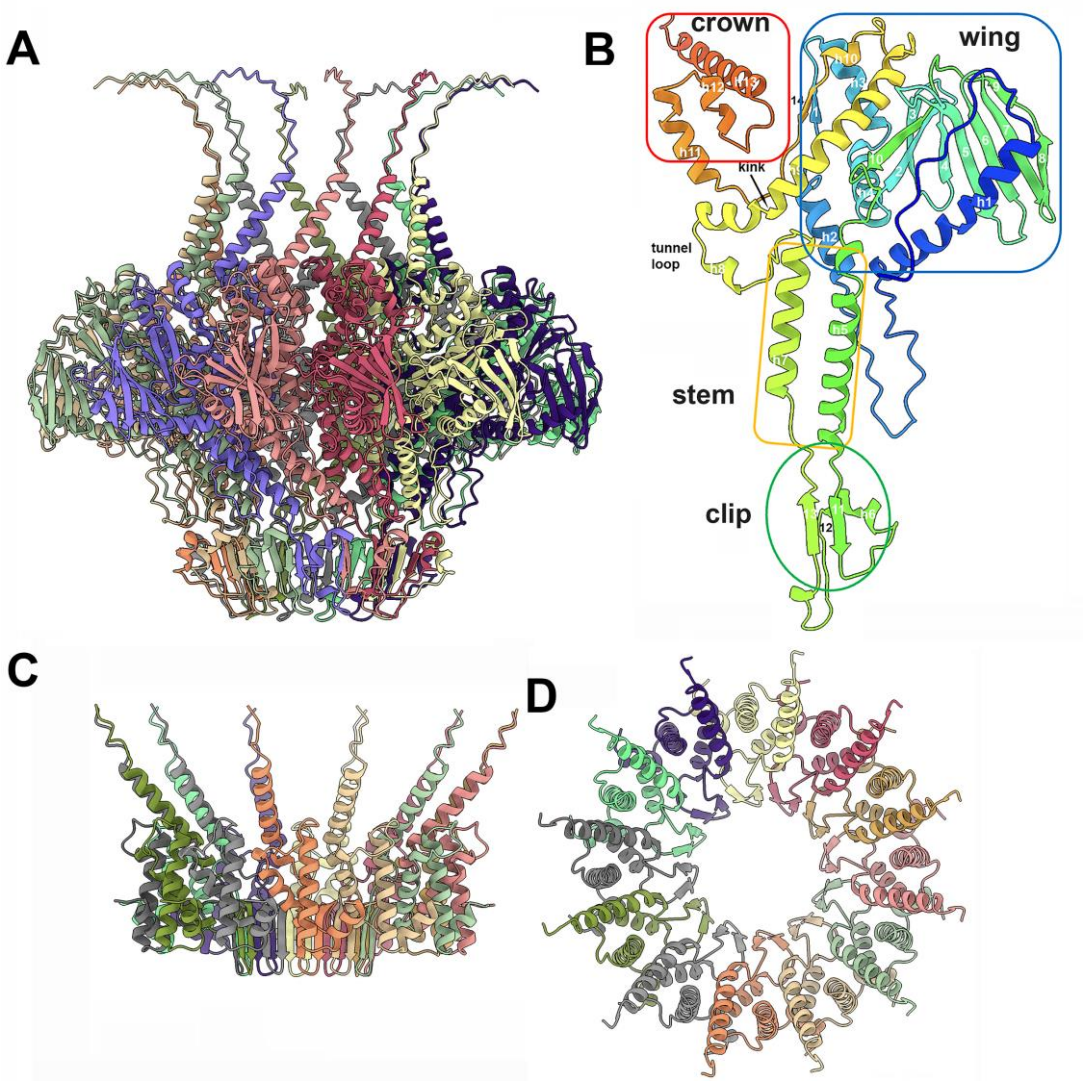


**Figure 3. The dodecameric portal and adaptor structures**. (**A**) The portal dodecamer has a size of 160 x 110 Å. (**B**) structure of the portal monomer with the different domains conserved in all portals. The C-terminal extension is poorly predicted and is disordered. (**C**) Lateral view of the stopper with the extended C-terminal α-helices and the central β-sheet. (**D**) same view rotated 90° relative to (C).

The wing domains form the central core of the portal assembly. The wing starts with an extended stretch (residues 1-18) followed by a long α-helix (residues 20-45), a loop (residues 46-67) and a second α-helix (residues 68-80). A first β-strand is followed by two α-helices (residues 81-121) and a β-hairpin (residues 122-142). A β-sheet of six antiparallel β-strands (residues 144-201) is followed by a β-strand antiparallel to the β-hairpin and a loop connected to the first α-helix (residues 223-246) of the stem. The clip follows with β-strand 11, a loop, α-helix 6, β-strand 12, another loop and β-strand 13. β-strands 11 and 13 form an antiparallel β-sheet. The β-strand 13 is connected to the second α-helix of the stem (residues 292-312), which is followed by the tunnel loop and a long, kinked α-helix (residues 326-363) conserved in phage portal proteins. After a short α-helix, the last β-strand 14 closes the wing by forming an antiparallel β-structure with β-strand 1, which is followed by an extended stretch connected to the three α-helices of the crown (385-425). It is noteworthy that the C-terminal end of the portal protein (residues 420-457) is predicted with low confidence. Interactions between portal monomers are strong, as they involve 4100 Å$^2$ of buried surface area, ~14% of the total surface area (29,000 Å$^2$). As mentioned above, these interactions involve most of the residues along the protein chain, with the exception of the N- and C-termini. The Dali server reported an impressively strong hit with the *B. subtilis* phage SPP1 portal protein (PDB ID: 3jes-U; Z-score=27.4; rmsd=2.6 Å;

 doi:10.20944/preprints202311.0819.v1



lali: 348/370; [39]) and a similarly good hit with the *E. coli* podophage T7 portal (PDB ID:6qwp-F; Z-score=20.7; rmsd=5.5 Å; lali: 352/486; [43]).

3.3.2. The dodecameric adaptor (ORF 38)

The predicted structure of the adaptor protein starts with an α-helix (H1), followed by two additional α-helices (H2, H3), an anti-parallel β-hairpin (S1, S2), and a long C-terminal α-helix (H4) (Figure 3B,C; Figure 4C). In the predicted adaptor dodecameric assembly, the twelve β-hairpins form a cylindrical β-sheet of 24 antiparallel β-strands. This cylinder, which is the conduit for DNA passage, has a diameter of ~28 Å. The C-terminal α-helix H4 (residues 85-96) is followed by an unstructured segment (residues 86-105) and the C-terminal β-strand (residues 105-110). The interaction between adjacent adaptor monomers is strong, as it buries a surface area of 1320 Å$^2$ of a total surface of 8900 Å$^2$ (~15 %). These intermolecular contacts involve the β-barrel as well as α-helices 3 sandwiched between α-helices 2 and 1 of an adjacent monomer and α-helices 2 from adjacent monomers. Submission to the Dali server reported rather poor hits with bacteriophage T7 tail protein gp12 (PDB ID:7ey7-T; Z-score=6.5; rmsd=4.7 Å; lali: 86/193) and with podophage T7 adaptor (PDB ID:2kz4; Z-score=9.1; rmsd=3.3 Å; lali: 94/112) and native GTA particle stopper (PDB ID:7box-T; Z-score=6.4; rmsd=4.8 Å; lali: 88/195). Therefore, the structure of the adaptor seems to be novel, in particular its 24-stranded β-barrel.

3.3.3. The portal/adaptor complex

Since the prediction of a portal/adaptor dodecamer was also beyond the possibilities of AF2, we predicted the structure of a complex between 12 clip domains of the portal protein (residues 246-288) and 12 adaptor proteins, which are known to interact in portal/adaptator assemblies. This clip/adaptor dodecamer was superimposed on the full-length portal dodecamer to obtain a complete portal/adaptor dodecamer.

The portal/adaptor interface involves a limited number of residues (Figure 4A,B). In particular, the C-terminal β-strand (residues 105-110) forms a 4-stranded β-sheet together with two adjacent portal monomers (Figure 4, inset 1), leading to a very modest interaction surface area of 710 Å$^2$ (~1.8 % of the portal total surface and ~7.8 of the adaptor total surface). This limited interaction results in a poor compactness at the portal/adaptor interface with "holes" in the adaptor structure.

3.3.4. The adaptor/stopper (ORF39 complex

We predicted the structure of six stopper proteins together with twelve adaptor proteins (1938 residues in total) in a single job submission. The adaptor structure in this complex is similar to that observed in the adaptor/portal complex, with the exception of its C-terminus which is predicted as a α-helix which poor statistics in the absence of the clip domain (Figure 4C). Each stopper monomer contains a β-sandwich of two β-sheets with four antiparallel β-strands each, a long β-hairpin and a α–helix. β-strands 5, 6, 10 and 9 form the first β-sheets, and β-strands 4, 1, 7 and 8 form the second one. The β-hairpin (β-strands 2 and 3) is an extension of β-strands 1 and 4. The unique α-helix is found between β-strands 5 and 6. The central cavity of the stopper hexameric assembly is partly obstructed by the six α-helices, while the β-hairpins are projected out of the central axis. The contacts between two adjacent stopper monomers involve the loops between the β-strands and the α-helix, as well as the α-helices themselves. Therefore, the interaction between adjacent monomers is weak with a buried surface area of ~400 Å$^2$ of the total 7580 Å$^2$ surface (5.2 %). Consequently, AlphaFold2 was unable to predict the stopper hexameric structure alone likely because of this weak surface area. The two best hits obtained with the Dali server were those of the NMR structure of a *Shigella flexneri* protein, which likely comes from a temperate phage (PDB ID: 2kz4; Z-score=9.1; rmsd=3.3 Å; lali: 94/112; unpublished) and of the native GTA particle stopper (PDB ID: 6te9; Z-score=7.6; rmsd=4.0 Å; lali: 84/110;[44]).

The adaptor-stopper interface is rather flat. Each stopper monomer interacts with three adaptor monomers with buried surfaces of 770 Å$^2$, 300 Å$^2$ and 260 Å$^2$ reaching a total interface area of 1330 Å$^2$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0819.v1



on a total surface area of 7580 Å² (17.5 %), a value indicating that the adaptor dodecamer and the stopper hexamer strongly interact with each other. (Figure 4A,D). Additionnal contacts involve the stopper's loops s4-s5 and s6-s7 with the adaptor's loop h1-h2, and the stopper's loop s9-s10 with the adjacent adaptor's loop h1-h2.

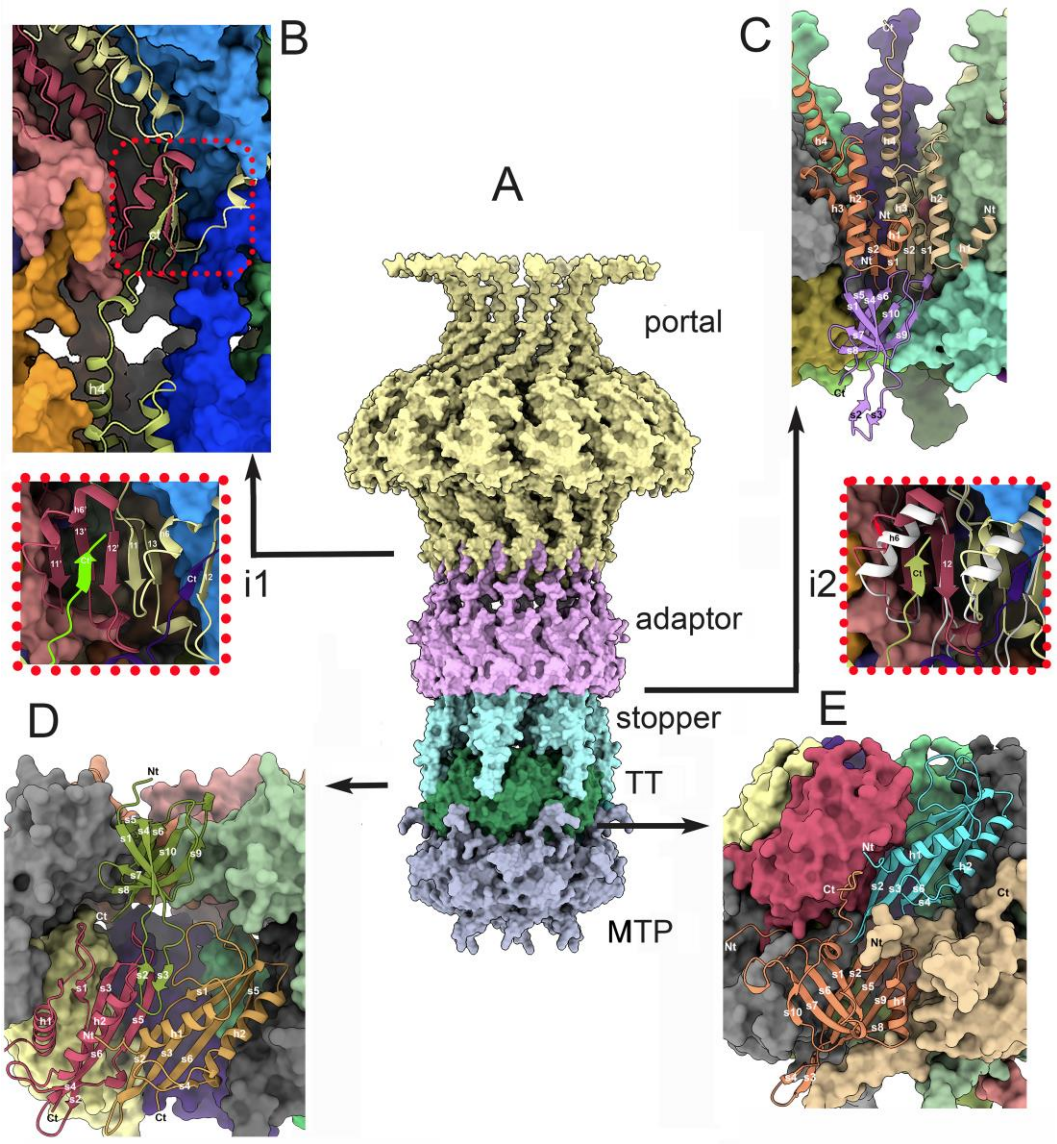


**Figure 4. Portal to major tail protein structure.** (**A**) Dodecameric portal (yellow) and adaptor (violet), hexameric stopper (light blue), tail terminator (TT, green) and major tail protein (MTP, grey). (**B**) The portal/adaptor interface (close-up view in **inset1, i1**) involves four β-strands: portal i s13 and s11 (anti-parallel), portal i+1 s12 and adaptor C-terminal β-strand stacks against α-helix 6 and portal i+1 α-strands s11 and s13. (**C**) Adaptor/stopper interface involves two adaptors (beige and orange) and one stopper (violet). Contacts originate from stopper's N-terminus and loops joining the α-strands of the three monomers. The insertion of the adaptor displaces α-helix 6 (**inset2, i2**). (**D**) Stopper /TT interface: The long stopper's hairpin s2-s3 (light green) is inserted between two TT monomers (red and beige). Other contacts involve the long TT's loops h1-s1and s3-s5. (**E**) The TT/MTP interface. The MTP i-1 N-terminus (beige) is inserted between MTP i (orange) and TT (light blue), and comprises the majority of TT/MTP contacts. The MTP C-terminus is inserted vertically between two TT monomers.

3.3.5. The collar and the neck passage proteins

A low resolution (20 Å) nsEM structure of the *TP901-1* neck was reported previously [22]. We used ChimeraX to fit the predicted portal/adaptor/stopper-associated atomic model in this 3D

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0819.v1



reconstruction (EMD-2227). While the overall fit is satisfying for the three modules (CC=081), a ring-shaped density *around* the portal/adaptor interface cannot be interpreted with our model (Figure 5A,B). This density positions exactly on the "holes" of the adaptor surface. Interestingly, a similar ring-shaped density has previously been ascribed to a "collar" formed by the N-terminal domains of six neck passage proteins (*NPS)* [46]. *NPS* are elongated proteins, anchored to the neck, terminated by carbohydrate binding modules. The NPS of TP901-1 *(ORF51)* were first reported by Vegge et al. [45]. However, to our knowledge, nothing is known about the precise molecular attachment of the NPS to the neck. The closest example of a structure resembling the callar is represented by the tail fibers, FibU and FibL of the *Staphylococcus aureus* siphophage 80 α [47] or the BppU N-terminal domain of phage TP901-1 [5]. In *S. aureus* phage 80α the six trimeric fibers (18 monomers) are attached to the phage tail by a ring of 12 N-terminal domains plus six N-terminal domains located above the ring [47].

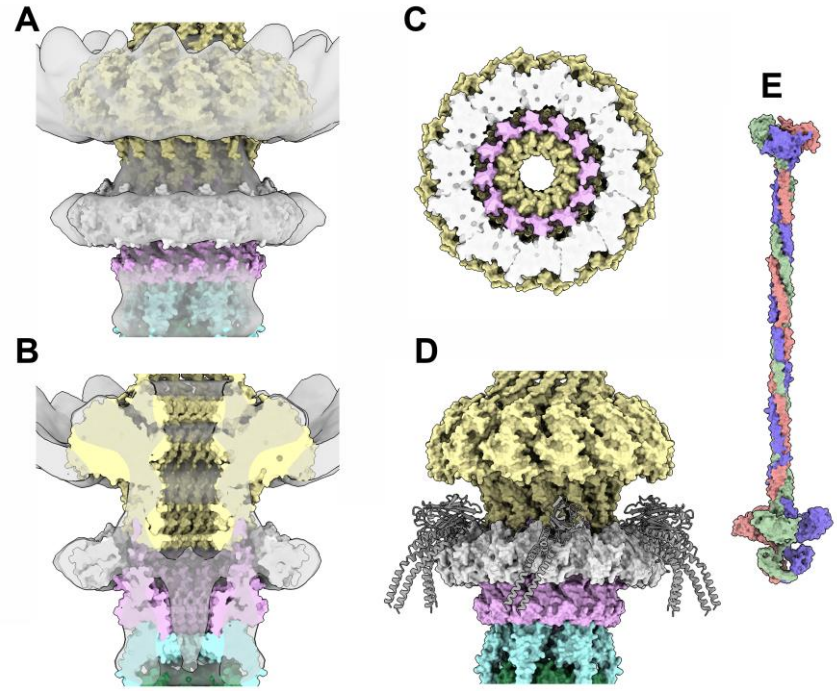


**Figure 5. The fit of the neck in the nsEM density map.** (**A**) The fit of the neck (portal 12-mer, yellow; adaptor 12-mer, pink; stopper 6-mer light blue; NPS N-termini 12-mer, white) in the nsEM map at 20.0 Å resolution. (**B**) Same view as in (A), but slabbed at half diameter. The NPS N-termini are located just below the portal/adaptor junction. (**C**) View at 90° from (A). The NPS N-termini cover the cavities observed at the portal/ adaptor interface. (**D**) The phage 80α fiber structure was superimposed onto the NPS N-termini to illustrate a possible structure of the NPS. (**E**) Surface view of the structural prediction of the NPS trimer with a total length of 48 nm.

Our reasoning was that this arrangement might also apply to NPS. AF2 predicted a ring of 12 N-terminal domains of NPS (residues 1-125) that yielded excellent statistics and fitted well in the nsEM map (Figure 5a-c). We then attempted to predict a complex between the adaptor and 12 N-terminal domains of NPS, which unfortunately failed. Nonetheless, fitting the FibU ring on top of the dodecameric NPS N-terminal domain ring might provide insights as to how the NPS is projected out of the tail axis (Figure 5D).

### *3.4. The neck/tail junction and the tail*

Structures of several protein complexes were predicted for the neck/tail junction and the tail: i) the complex formed by the adaptor (minus its C-terminal helix), stopper, tail terminator (TT) and MTP (12/6/6/6 mer); ii) the adaptor/stopper complex (12/6 mer); iii) the stopper/TT complex (6/6 mer); iv) the TT/MTP complex (6/6 mer); and v) the MTP/MTP complex (6/6 mer).

 doi:10.20944/preprints202311.0819.v1



3.4.1. The stopper/tail terminator (TT; ORF41) complex

We predicted the structure of the complex between a stopper hexamer and a TT hexamer. The TT is composed of a 6-stranded β-sheet (s1-s6) covered by two α-helices (h1-h2) inserted between the β-strands s3 and s4 (Fig 4d). In the TT hexamer, the six β-sheets form a cylindrical β-barrel maintained by inter-monomer contacts between β-strands s2 and s5. The TT/TT interactions cover 840 Å$^2$ on a total surface of 7700 Å$^2$ (~11%). A Dali search of the most closely related structures in the PDB reported a small number of relevant hits, among which the most pertinent was the retrieved structural similarity hit with the tail terminator of phage GTA (PDB ID 6te9; Z-score=14.0; rmsd=2.4 Å; lali: 122/134;[43]).

In the stopper/TT complex, the stopper β-hairpins have a conformation different from that observed in the absence of the TT. While the stopper β-sandwich structure and the stopper/stopper interactions are maintained, the β-hairpins move by ~20° to embrace the TT hexameric ring (Supplementary Figure 3, Supplementary Movie 1). Each stopper β-hairpin stacks against the TT's β-strand s5 of monomer i, and is sandwiched between the α-helix h2 of the monomer i and the α-helix h1 of the monomer i+1. Additional contacts involve the TT's long loops h1-s1 and s5-s6 and the stopper's C-terminus and β-sandwich loops. Each stopper monomer interacts with two TT monomers that cover 400 and 330 Å$^2$ (730 Å$^2$ in total, ~10% of the total surface).

3.4.2. The tail terminator (TT) / major tail protein (MTP; ORF42) complex

We predicted the structure of the complex between a TT hexamer and a MTP hexamer. The TT hexamer structure is similar to that observed in the absence of the MTPs. The MTP structure displays a sandwich of two anti-parallel β-sheets (Figure 4E). The five β-strands of the external β-sheet have a connectivity s1, s6, s7, s10, and the four β-strands of the internal β-sheet have a connectivity s2, s5, s9, s8. An α-helix between β-strands s1 and s2 and a long β-hairpin s3,s4 complete the structure. The internal β-sheets in the MTP hexamer assemble as a tight anti-parallel β-barrel of 24 β-strands maintained by interactions between the β-strands s2 and s8. Additional inter-monomer contacts involve the N-terminus that sneaks within the adjacent monomer. The β-hairpin establishes strong contact with the adjacent MTP β-sandwich and the β-hairpin that strongly interact with the adjacent MTP β-sandwich and β-hairpin. The α-helix is also in contact with the loops of the external β-sheet. The buried surface of these inter MTP monomer interactions involves 1340 Å$^2$ out of 12,200 Å$^2$ in total, ~11%). Dali search identified several hits in the PDB, among which MTP of phage SPP1 was the best (PDB ID:6yeg; Z=15.2; rmsd=2.3 Å; lali: 157/172;[48]). The N-terminus of the MTP establishes contacts with adjacent MTP and TT moieties, while the C-terminus contacts the closest TT and the adjacent one (Figure 4E). The TT s2,s3 loop extends to contact the MTP's internal β-sheet. Overall, each TT interacts with three MTPs: MTP i covers 770 Å$^2$ of its surface and MTPs i-1 and i+1 380 Å$^2$ and 110 Å$^2$, respectively, for a total of 1,260 Å$^2$ (~16% of TT surface).

3.4.3. The MTP/MTP rings

Employing AlphaFold2 we predicted three stacked rings, each consisting of MTP hexamers, representing a total of 3,042 residues. We then added a fourth hexamer with Coot [33] leading to four stacked rings (Figure 6A). These four MTP rings fit into the 20.0 Å resolution nsEM 3D reconstruction (EMD-2228). However, the map had to be flipped to get the best result (CC=0.79) Figure 6B). The six MTPs' β-hairpins of a hexamer form a platform onto which the lower hexamer is stacked (Figure 6C). The contacts are established between the β-hairpin and the N-terminus of adjacent monomers and between the C-terminus of a lower monomer and the external β-sheet of an upper monomer (Figure 6D). Each MTP of one ring contacts four MTPs of the other ring: the largest interaction is established with the nearest monomer i (835 Å$^2$), the other interactions of 485, 105 and 45 Å$^2$ are established with monomers i+1, i+2 and i-1, respectively, with a total surface of 1,500 Å$^2$ (~12.4% of total surface).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0819.v1



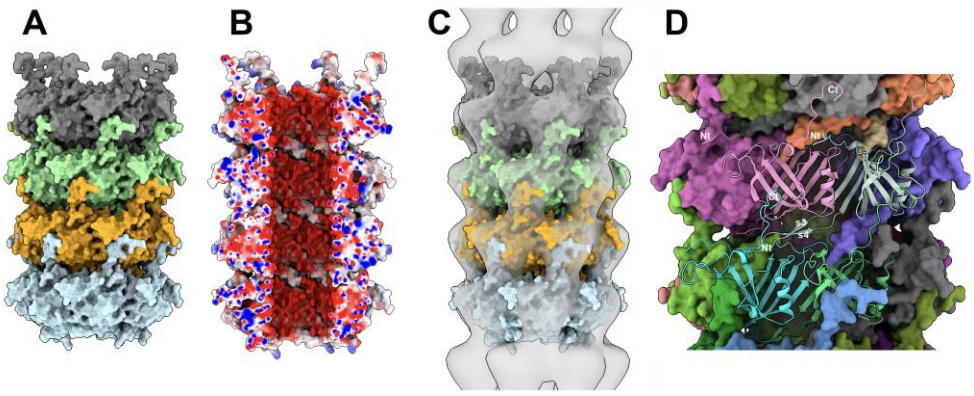


**Figure 6. The tail tube.** (**A**) View of four MTP rings forming a section of the tail tube (Top of the view is the direction of capsid). Note the insertion of each monomer C-terminus inside a monomer cavity in the above ring. (**B**) Same orientation as in (A) but slabbed at half-diameter and colored according to electrostatics. (**C**) Fit of the MTPs in the nsEM map at 20Å resolution (EMD-2228). (**D**) MTP/MTP interactions between stacked rings. Interactions involve the N- and C-termini and the β-hairpin with strands s3 and s4.

*3.5. The baseplate*

The baseplate of TP901-1 comprises four proteins: the Dit, Tal, BppU and RBP (ORFs46,47, 48, and 49, respectively) [5, 11, 49, 50]. The Dit hexamer and the Tal trimer form the central part of the baseplate and extend the last MTP ring of the tail. Moreover, the tape measure protein (TMP; ORF45) has been proposed to fill the central channel of the tail, Dit and Tal [26]. The crystal structure of TP901-1 baseplate, without the Tal, was previously reported [5]. In this complex, six BppU trimers are attached to the Dit hexamer, each trimer projecting three RBP trimers at the baseplate periphery. Our goal here was to assess the MTP/Dit interface and the Tal/Dit interface, and to model a complete baseplate structure. To this end, we predicted the structure of a complex of MTP (x6) , Dit (x3) and Tal (x3) structural domain (residues 1-380) (3,672 residuesin total). We also predicted the structure of the full-length Tal trimer and superimposed it on the MTP/Dit/Tal-Nter predicted structure using Coot [33] (Figure 7A). Lastly, we predicted the structure of the trimeric BppU, the trimericRBP, and their complex, and compared them to the X-ray structures.

3.5.1. The MTP/Dit interface

The predicted structure of the Dit hexamer superimposes well on the Dit crystal structure (r.m.s.d.=0.93 Å) [5]. Each Dit monomer comprises an N-terminal domain (residues 1-145) forming a β-sandwich with an α-helix and a β-hairpin, and a C-terminal domain (residues 146-255) folded as a galectin-like β-sandwich, also found in other Dit proteins [25, 51] (Figure 7B). The six N-terminal domains form the Dit central core, with a Dit/Dit interface burying a surface of 1,400 Å$^2$, while the six C-terminal domains extend out of the ring and do not interact with each other. Each Dit monomer projects a β-hairpin (strands s3, s4) towards the neighboring monomer, thereby ensuring the cohesion of the hexamer, and forming a large part of the interface between the MTP and Dit rings. The Dit N-terminal segment (residues 1-12) extends towards the MTP covering its external β-sheet (Figure 7B). Each MTP monomer contacts three Dit monomers with buried surfaces of 910 Å$^2$, 520 Å$^2$ and 160 Å$^2$ for monomers i, i+1 and i+2, respectively, accounting for a total surface of 1590 Å$^2$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0819.v1



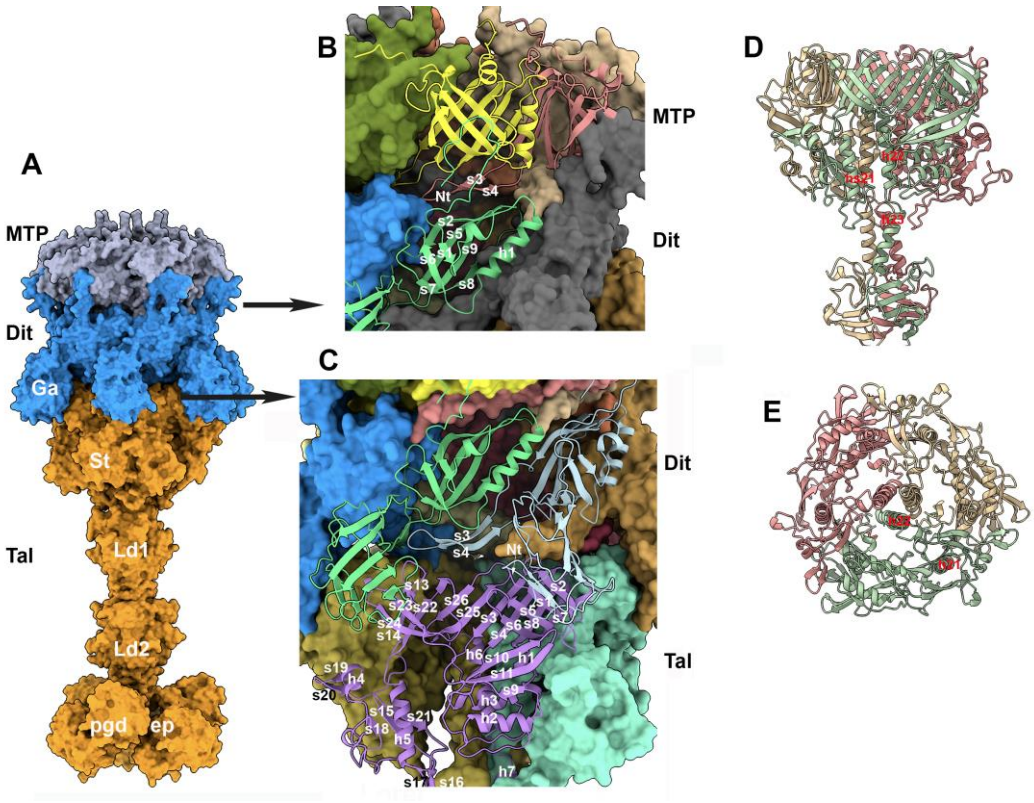


**Figure 7. The distal MTP and the central baseplate**. (**A**) Surface view of the distal hexameric MTP (grey), the hexameric Dit (blue) and the trimeric Tal (orange). Ga: Dit's galectin domain; St: Tal's conserved structural domain; Ld1 and Ld2: linker domains 1 and 2; pgd and endopeptidase (ep) domains. (**B**) Ribbon view of the MTP/Dit interface. (**C**) Ribbon view of the Dit/Tal interface. (**D**) Ribbon side-view of the Tal trimer N-terminus (residues 1-484). (**E**) Same view rotated by 90° within the phage main axis.

### 3.5.2. The trimeric Tal and the Dit/Tal interface

The Dit/Tal interface has a 6-fold to 3-fold symmetry mismatch reported for all Dit/Tal structures in tailed phages [25, 52, 53]. Each of the two Tal sub-domains faces one Dit thereby assembling a pseudo-6-fold symmetrical Tal ring (for details, see [52]). Two other domains form the lower part of the Tal that were shown to undergo conformational changes opening the Tal for dsDNA ejection [25]. Here, the TP901-1 Tal trimer is maintained in a semi-open position as it accommodates three h22 helices in its center (Figure 7D,E). They are followed by three h23 helices abutting to structural domains of 3 β-sheets made up of 5 anti-parallel β-strands each (Figure 7D). This β-domain is is followed by a second β-stranded structural domain and the two enzymatic domains already described elsewhere [46, 54]. The two Tal monomer domains in contact with Dit form a β-sandwich, the internal part formed by 8 β-strands. The Tal trimer is therefore tightly assembled through a 24 β-strands continuous belt (Figure 7C,D). The two Tal monomer (i) domains located below the belt do not interact together, each interacting with the i-1 and i+1 Tal monomer (Figure 7C). The interactions between the Dit ring and the Tal trimer are again mediated by a hairpin of the Dit (s3,s4), but also by the interaction of Dit's galectin domains with the two upper domains of the Tal (Figure 7a,c). Each Tal interacts with three Dit monomers: a major interaction with a Dit (1,240 Å$^2$), and two other interactions with flanking molecules of 360 Å$^2$ and 510 Å$^2$, respectively. The total of 2,110 Å$^2$, accounts for 10% of the total Tal surface area.

### 3.5.3. The Dit/BppU and BppU/RBP complexes

The assembly of the hexameric Dit with the 18 BppU is currently beyond the capacity for AF2 prediction. We therefore tried different predictions involving the hexameric Dit and three or six BppU

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2023 doi:10.20944/preprints202311.0819.v1



monomers, without success. However, like for the NPS -neck complex (see above), we were able to successfully predict the structure of the hexameric Dit with a dodecamer of the BppU N-terminal domain (residues 1-140) (Figure 8A,B) that superimposes well on the TP901-1 baseplate crystal structure. We also successfully successfully retrieved the predicted structure of three BppU with three RBP trimers. While the structure of the N-terminus (1-170) of BppU is not assembled as in the X-ray structure, due to the absence of the Dit hexamer, the following helical domain and C-terminal RBP-attachment domains are folded as in the X-ray structure. Furthermore, the attachment of the RBP trimer to its BppU receptor is very close to that observed in the X-ray structure (PDB: 4v96) (Figure 8C) the BppU/RBPcomplex fits well in the nsEM map (EMD-1793) (Figure 8E). Even at the level of side chains, three BppU hydrophobic residues (Ile219, Phe226 and Phe232) that were shown in the X-ray structure to play a crucial role in the BppU/RBP interaction are observed in the same conformations [5] (Figure 8D). Finally, the interactions of the TMP within the virion could not be deciphered. The TMP was shown to participate to the initial step of the tail tube formation by assembling to the Dit/Tal/BppU/RBP complex and being assisted by specific chaperones [23]. Moreover, it was shown that the TMP forms a trimer within the phage tail tube, down to the Tal trimer [10, 47, 55], indicating that the TMP should interact with the TT, MTP and Dit hexamers, as well as with the Tal trimer. We therefore tried to predict the interaction of (i) a trimer of the TMP N-terminus with a TT hexamer, (ii) a trimer of the TMP middle region with a MTP hexamer, (iii) a trimer of the TMP the C-terminus with the Dit hexamer/Tal trimer, all without any success. In all cases, a trimer of TMP could not be predicted. This may be due to the fact that the TMP assembly is driven by chaperones that form transient structural states ? as reported previously [23].

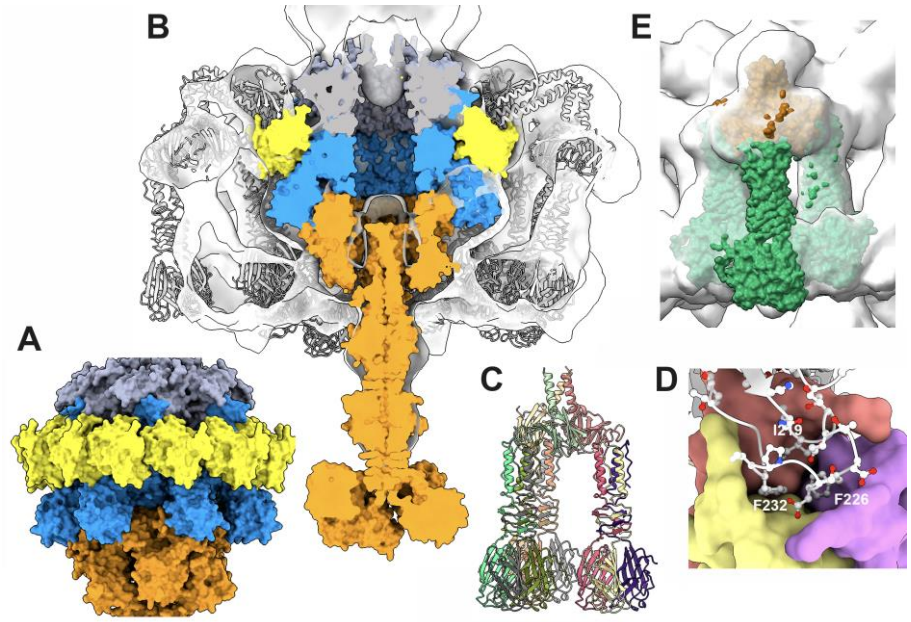


**Figure 8. The peripheral baseplate**. (**A**) Surface view of the dodecameric BppU N-terminal domain (yellow) complexed to the hexameric Dit (blue); (MTP: grey; Tal: orange). (**B**) Slabbed view of the nsEM map (EMD-1793; 25 Å resolution) of the complete baseplate with the distal hexameric MTP (grey), the hexameric Dit (blue), the trimeric Tal (orange) and the dodecameric BppU N-terminal domain (yellow) fitted inside. The grey ribbon inside the map represents the X-ray structure (PDB id 4v96). (**C**) The AF2 predicted structure of the "tripod" formed of the trimeric BppU C-terminal domain holding three RBP trimers. (**D**) Close-up of the BppU/RBP interface with the hydrophobic residues Ile 219, Phe 226 and Phe 232 also identified in the X-ray structure. (**E**) Surface view of the "tripod" fit in the nsEM map.

## 4. Discussion

The virion structure of the tailed phage that we present here is one of the most complete to date. For such prediction-based structural work, the validity of the predicted structures is an important

 doi:10.20944/preprints202311.0819.v1



issue. AlphaFold2 has been recognized to provide crystal-grade structures provided that the pLDDT values are sufficiently high, i.e. above 80% [27-29, 56]. Here, we checked the structure validity at three levels. First, internally to AF2, we accepted structures that meet the AF2 internal validation score given by the pLDDT values, i.e. when the pLDDT values were above 70% in the folded regions. Although loops are often below this threshold, they do not significantly affect the overall fold. Moreover, we also examined the PEA plots that give confidence scores of protein-protein interactions as well as verified that the pLDDT values in the protein complex were at least equal or higher than those of the non-complexed structure. Second, we used an external validity criterion by assessing the agreement between the predicted assemblies and the experimental TP901-1 nsEM 3D reconstructions in the 15-25 Å-resolution range. We also use the available crystal structures to evaluate the capabilities and limitations of AF2 predictions. As a third validation approach, we checked with the Dali server [30] whether the predicted folds were already reported in the PDB, without being an obligate criterion as structures with new fold may be generated (see below). Additionally, we used the PISA server [57] to analyse the quality of the interactions within our assemblies.

In our predicted structure of the TP901-1virion, the hexon and penton of the capsid were well predicted, except for the N-termini as they may be involved in interactions with other hexons or pentons. Predictions of hexon/hexon or hexon/penton interactions were unsuccessful: the hexon and penton were predicted as separate units without any interactions. The procapsid assembly requires a so-called scaffolding protein (SP), which acts as a chaperone to ensure correct positioning of the subunits [58, 59]. Structures of full or partial SP have been reported (PDB ID 1tx9, 2gp8, 1no4, 8dt0), and they are helical proteins as seen in phage the phi29, where it forms a dimer of two long α-helices [58]. In the procapsid of staphylococcal phage 80α, the SP C-terminal α-helix was observed in contact with each of the capsid component [60]. In phage TP901-1, the predicted structure of he expression product of *orf35*, which is located immediately upstream of the MCP-encoding gene, displays a pattern of two long helices and some shorter ones, making it a good candidate for being an SP.

The predicted dodecameric portal exhibits all the characteristics of genuine phage portals previously reported [43, 61, 62]. However, the TP901-1 portal has a unique feature: its clip domain contains three β-strands originating from two different portals and a fourth β-strand coming from the adaptor. The TP901-1 adaptor also presents a unique feature with its 24-stranded β-barrel forming its internal channel. Interestingly, the not compact and rather loose interface between the portal and adaptor is compensated by the presence of a ring of six NPS trimers whose N-terminal domains cover the adaptor ring. The hexameric stopper of TP901-1 is similar to that of other phages and exhibits long β-hairpins that embrace the tail tube upon complexation.

The tail and baseplate components, including the TT, MTP, and Dit, share structural features that have previously been reported [15, 63]. The superimposition of the monomers reveals that the three components possess a β-hairpin that interacts with the other subunits of an hexameric ring and also provides a stacking platform between hexameric rings. The MTP and Dit also possess an extended N-terminus that inserts into a crevice of an above hexamer in the MTP/MTP and Dit/MTP interactions. While platforms of β-hairpins were also observed in the MTP stacking of the tailed phages 80α [47] and T5 [55, 64], the N-terminal lock was only observed in 80α for which a β-helix helps the interaction.

The structure of the TP901-1 baseplate has previously been reported, though without its Tal component [5]. Here, the full-length Tal completes the X-ray structure. The central channel of the Tal N-terminal structural domain (1-380) is filled by three $\alpha$-helices that link this domain to the functional C-terminal domain. Surprisingly, this central channel is filled in phage 80 by the C-terminus of the three TMPs contained in the tail tube [47]. However, the position of the three helices in TP901-1 Tal is logical considering the Tal topology. Indeed, these helices and the C-terminal domain should dramatically rearrange to allow the TMP exit and DNA ejection during the initial stages of infection. Such a rearrangement has recently been reported for phage T5. Upon contact of the tail tip with T5's receptor, the membrane protein FhuA, the Tal-like protein pb3 that obstructs the tail exit channel opens and rotates on the tail side, thus allowing the TMP (pb2) to insert into the membrane [10, 55].

We postulate that a similar mechanism is operating in the case of phage TP901-1 upon baseplate/cell wall polysaccharide binding.

The prediction of the RBP structures and of their interaction with BppU was excellent, as were the prediction and localization of 12 out of the 18 BppU N-terminal domains that form a ring similar to the NPS ring or to the tail fiber rings observed in phage 80α [47]. However, the six remaining BppU N-terminal domains and the BppU topology were out of reach due to the fact that the complete structure is well over the residue limit. Lastly, all attempts to model the TMP in complex with the MTP, Dit, and tail chaperone were unsuccessful.

To conclude, we have found that AF2 has impressive prediction capabilities, provided that the ensemble used for prediction involves less than a residue limit of around 4000. Most of our prediction failures were primarily linked to this limitation. However, as evidenced by the capsid and parts of the baseplate, the limitations may not be due to this factor only. Despite this, we illustrate here that AF2's predictions provide an avenue for further investigation into phage structure. This is particularly important for phages that remain poorly characterized at the structural level, such as those infecting human pathogens and Mycobacteria [65].

**Supplementary Materials:** The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Figure S1: title; Table S1: title; Video S1: title.

**Author Contributions:** C.C. and D.v.S. led the study. C.C. performed the AlphaFold2 predictions and analysis. C.C. and D.v.S. supervised and examined the results of the experiments. The initial concept manuscript was written by C.C., J.M. A.G., and D.v.S.. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Science Foundation Ireland (SFI; https://www.sfi.ie/), under grant numbers SFI/12/RC/2273-P2 (DvS) and 20/FFP-P/8664 (JM).

**Data Availability Statement:** Coordinates files have been deposited at ModelArchive (modelarchive.org) under the deposition identifiers XXX….

**Acknowledgments:** This work was performed in part using HPC resources from GENCI-IDRIS (Grant 2023-AD010714075). We acknowledge UCSF ChimeraX for molecular graphics that is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Ackermann, H. W. "Bacteriophage Electron Microscopy." *Adv Virus Res* 82 (2012): 1-32.
2. Turner, D., A. N. Shkoporov, C. Lood, A. D. Millard, B. E. Dutilh, P. Alfenas-Zerbini, L. J. van Zyl, R. K. Aziz, H. M. Oksanen, M. M. Poranen, A. M. Kropinski, J. Barylski, J. R. Brister, N. Chanisvili, R. A. Edwards, F. Enault, A. Gillis, P. Knezevic, M. Krupovic, I. Kurtboke, A. Kushkina, R. Lavigne, S. Lehman, M. Lobocka, C. Moraru, A. Moreno Switt, V. Morozova, J. Nakavuma, A. Reyes Munoz, J. Rumnieks, B. L. Sarkar, M. B. Sullivan, J. Uchiyama, J. Wittmann, T. Yigang, and E. M. Adriaenssens. "Abolishment of Morphology-Based Taxa and Change to Binomial Species Names: 2022 Taxonomy Update of the Ictv Bacterial Viruses Subcommittee." *Arch Virol* 168, no. 2 (2023): 74.
3. Comeau, A. M., D. Tremblay, S. Moineau, T. Rattei, A. I. Kushkina, F. I. Tovkach, H. M. Krisch, and H. W. Ackermann. "Phage Morphology Recapitulates Phylogeny: The Comparative Genomics of a New Group of Myoviruses." *PLoS One* 7, no. 7 (2012): e40102.
4. Goulet, A., S. Spinelli, J. Mahony, and C. Cambillau. "Conserved and Diverse Traits of Adhesion Devices from Siphoviridae Recognizing Proteinaceous or Saccharidic Receptors." *Viruses* 12, no. 5 (2020).
5. Veesler, D., S. Spinelli, J. Mahony, J. Lichiere, S. Blangy, G. Bricogne, P. Legrand, M. Ortiz-Lombardia, V. Campanacci, D. van Sinderen, and C. Cambillau. "Structure of the Phage Tp901-1 1.8 Mda Baseplate Suggests an Alternative Host Adhesion Mechanism." *Proc Natl Acad Sci U S A* 109, no. 23 (2012): 8954-8.
6. Lavelle, K., Goulet, A., McDonnel, B., Spinelli, S., van Sinderen, D., Mahony, J. and Cambillau, C. "Revisiting the Host Adhesion Determinants of Streptococcus Thermophilus Siphophages." *Microb. Biotech.* 13, no. 6 (2020): 1765-79.

7. Bebeacua, C., D. Tremblay, C. Farenc, M. P. Chapot-Chartier, I. Sadovskaya, M. van Heel, D. Veesler, S. Moineau, and C. Cambillau. "Structure, Adsorption to Host, and Infection Mechanism of Virulent Lactococcal Phage P2." *J Virol* 87, no. 22 (2013): 12302-12.
8. Baptista, C., M. A. Santos, and C. Sao-Jose. "Phage Spp1 Reversible Adsorption to Bacillus Subtilis Cell Wall Teichoic Acids Accelerates Virus Recognition of Membrane Receptor Yueb." *J Bacteriol* 190, no. 14 (2008): 4989-96.
9. Xu, J., R. W. Hendrix, and R. L. Duda. "Chaperone-Protein Interactions That Mediate Assembly of the Bacteriophage Lambda Tail to the Correct Length." *J Mol Biol* 426, no. 5 (2014): 1004-18.
10. Degroux, S., G. Effantin, R. Linares, G. Schoehn, and C. Breyton. "Deciphering Bacteriophage T5 Host Recognition Mechanism and Infection Trigger." *J Virol* 97, no. 3 (2023): e0158422.
11. Mahony, J., S. R. Stockdale, B. Collins, S. Spinelli, F. P. Douillard, C. Cambillau, and D. van Sinderen. "Lactococcus Lactis Phage Tp901-1 as a Model for Siphoviridae Virion Assembly." *Bacteriophage* 6, no. 1 (2016): e1123795.
12. Fokine, A., and M. G. Rossmann. "Molecular Architecture of Tailed Double-Stranded DNA Phages." *Bacteriophage* 4, no. 1 (2014): e28281.
13. Spinelli, S., C. Bebeacua, I. Orlov, D. Tremblay, B. P. Klaholz, S. Moineau, and C. Cambillau. "Cryo-Electron Microscopy Structure of Lactococcal Siphophage 1358 Virion." *J Virol* 88, no. 16 (2014): 8900-10.
14. Goulet, A., R. Joos, K. Lavelle, D. Van Sinderen, J. Mahony, and C. Cambillau. "A Structural Discovery Journey of Streptococcal Phages Adhesion Devices by Alphafold2." *Front Mol Biosci* 9 (2022): 960325.
15. Veesler, D., and C. Cambillau. "A Common Evolutionary Origin for Tailed-Bacteriophage Functional Modules and Bacterial Machineries." *Microbiol Mol Biol Rev* 75, no. 3 (2011): 423-33, first page of table of contents.
16. Ruiz-Cruz, S., A. Erazo Garzon, P. Kelleher, F. Bottacini, S. O. Breum, H. Neve, K. J. Heller, F. K. Vogensen, S. Palussiere, P. Courtin, M. P. Chapot-Chartier, E. Vinogradov, I. Sadovskaya, J. Mahony, and D. van Sinderen. "Host Genetic Requirements for DNA Release of Lactococcal Phage Tp901-1." *Microb Biotechnol* 15, no. 12 (2022): 2875-89.
17. Stockdale, S. R., B. Collins, S. Spinelli, F. P. Douillard, J. Mahony, C. Cambillau, and D. van Sinderen. "Structure and Assembly of Tp901-1 Virion Unveiled by Mutagenesis." *PLoS One* 10, no. 7 (2015): e0131676.
18. Ostergaard Breum, S., H. Neve, K. J. Heller, and F. K. Vogensen. "Temperate Phages Tp901-1 and Philc3, Belonging to the P335 Species, Apparently Use Different Pathways for DNA Injection in Lactococcus Lactis Subsp. Cremoris 3107." *FEMS Microbiol Lett* 276, no. 2 (2007): 156-64.
19. Rasmussen, K. K., A. Palencia, A. K. Varming, H. El-Wali, E. Boeri Erba, M. Blackledge, K. Hammer, T. Herrmann, M. Kilstrup, L. Lo Leggio, and M. R. Jensen. "Revealing the Mechanism of Repressor Inactivation During Switching of a Temperate Bacteriophage." *Proc Natl Acad Sci U S A* 117, no. 34 (2020): 20576-85.
20. Pedersen, M., J. T. Neergaard, J. Cassias, K. K. Rasmussen, L. Lo Leggio, K. Sneppen, K. Hammer, and M. Kilstrup. "Repression of the Lysogenic P(R) Promoter in Bacteriophage Tp901-1 through Binding of a Ci-Mor Complex to a Composite O(M)-O(R) Operator." *Sci Rep* 10, no. 1 (2020): 8659.
21. Varming, A. K., K. K. Rasmussen, Z. Zong, P. W. Thulstrup, M. Kilstrup, and L. Lo Leggio. "Flexible Linker Modulates the Binding Affinity of the Tp901-1 Ci Phage Repressor to DNA." *FEBS J* 289, no. 4 (2022): 1135-48.
22. Bebeacua, C., L. Lai, C. S. Vegge, L. Brondsted, M. van Heel, D. Veesler, and C. Cambillau. "Visualizing a Complete Siphoviridae Member by Single-Particle Electron Microscopy: The Structure of Lactococcal Phage Tp901-1." *J Virol* 87, no. 2 (2013): 1061-8.
23. Vegge, C. S., L. Brondsted, H. Neve, S. Mc Grath, D. van Sinderen, and F. K. Vogensen. "Structural Characterization and Assembly of the Distal Tail Structure of the Temperate Lactococcal Bacteriophage Tp901-1." *J Bacteriol* 187, no. 12 (2005): 4187-97.
24. Ainsworth, S., I. Sadovskaya, E. Vinogradov, P. Courtin, Y. Guerardel, J. Mahony, T. Grard, C. Cambillau, M. P. Chapot-Chartier, and D. van Sinderen. "Differences in Lactococcal Cell Wall Polysaccharide Structure Are Major Determining Factors in Bacteriophage Sensitivity." *mBio* 5, no. 3 (2014): e00880-14.
25. Sciara, G., C. Bebeacua, P. Bron, D. Tremblay, M. Ortiz-Lombardia, J. Lichiere, M. van Heel, V. Campanacci, S. Moineau, and C. Cambillau. "Structure of Lactococcal Phage P2 Baseplate and Its Mechanism of Activation." *Proc Natl Acad Sci U S A* 107, no. 15 (2010): 6852-7.
26. Mahony, J., M. Alqarni, S. Stockdale, S. Spinelli, M. Feyereisen, C. Cambillau, and D. V. Sinderen. "Functional and Structural Dissection of the Tape Measure Protein of Lactococcal Phage Tp901-1." *Sci Rep* 6 (2016): 36667.
27. Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli,

and D. Hassabis. "Highly Accurate Protein Structure Prediction with Alphafold." *Nature* 596, no. 7873 (2021): 583-89.
28. Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. "Applying and Improving Alphafold at Casp14." *Proteins* 89, no. 12 (2021): 1711-21.
29. Jumper, J., and D. Hassabis. "Protein Structure Predictions to Atomic Accuracy with Alphafold." *Nat Methods* 19, no. 1 (2022): 11-12.
30. Holm, L., S. Kaariainen, P. Rosenstrom, and A. Schenkel. "Searching Protein Structure Databases with Dalilite V.3." *Bioinformatics* 24, no. 23 (2008): 2780-1.
31. Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. "Ucsf Chimera--a Visualization System for Exploratory Research and Analysis." *J Comput Chem* 25, no. 13 (2004): 1605-12.
32. Emsley, P., and K. Cowtan. "Coot: Model-Building Tools for Molecular Graphics." *Acta Crystallogr D Biol Crystallogr* 60, no. Pt 12 Pt 1 (2004): 2126-32.
33. Emsley, P., B. Lohkamp, W. G. Scott, and K. Cowtan. "Features and Development of Coot." *Acta Crystallogr D Biol Crystallogr* 66, no. Pt 4 (2010): 486-501.
34. Morais, M. C., K. H. Choi, J. S. Koti, P. R. Chipman, D. L. Anderson, and M. G. Rossmann. "Conservation of the Capsid Structure in Tailed Dsdna Bacteriophages: The Pseudoatomic Structure of Phi29." *Mol Cell* 18, no. 2 (2005): 149-59.
35. Helgstrand, C., W. R. Wikoff, R. L. Duda, R. W. Hendrix, J. E. Johnson, and L. Liljas. "The Refined Structure of a Protein Catenane: The Hk97 Bacteriophage Capsid at 3.44 a Resolution." *J Mol Biol* 334, no. 5 (2003): 885-99.
36. Rao, V. B., A. Fokine, Q. Fang, and Q. Shao. "Bacteriophage T4 Head: Structure, Assembly, and Genome Packaging." *Viruses* 15, no. 2 (2023).
37. Kizziah, J. L., K. A. Manning, A. D. Dearborn, E. A. Wall, L. Klenow, R. L. L. Hill, M. S. Spilman, S. M. Stagg, G. E. Christie, and T. Dokland. "Cleavage and Structural Transitions During Maturation of Staphylococcus Aureus Bacteriophage 80alpha and Sapi1 Capsids." *Viruses* 9, no. 12 (2017).
38. Ignatiou, A., S. Brasiles, M. El Sadek Fadel, J. Burger, T. Mielke, M. Topf, P. Tavares, and E. V. Orlova. "Structural Transitions During the Scaffolding-Driven Assembly of a Viral Capsid." *Nat Commun* 10, no. 1 (2019): 4840.
39. Lebedev, A. A., M. H. Krause, A. L. Isidro, A. A. Vagin, E. V. Orlova, J. Turner, E. J. Dodson, P. Tavares, and A. A. Antson. "Structural Framework for DNA Translocation Via the Viral Portal Protein." *Embo J* 26, no. 7 (2007): 1984-94.
40. Rao, V. B., A. Fokine, and Q. Fang. "The Remarkable Viral Portal Vertex: Structure and a Plausible Model for Mechanism." *Curr Opin Virol* 51 (2021): 65-73.
41. Feiss, M., and V. B. Rao. "The Bacteriophage DNA Packaging Machine." *Adv Exp Med Biol* 726 (2012): 489-509.
42. Orlov, I., S. Roche, S. Brasiles, N. Lukoyanova, M. C. Vaney, P. Tavares, and E. V. Orlova. "Cryoem Structure and Assembly Mechanism of a Bacterial Virus Genome Gatekeeper." *Nat Commun* 13, no. 1 (2022): 7283.
43. Cuervo, A., M. Fabrega-Ferrer, C. Machon, J. J. Conesa, F. J. Fernandez, R. Perez-Luque, M. Perez-Ruiz, J. Pous, M. C. Vega, J. L. Carrascosa, and M. Coll. "Structures of T7 Bacteriophage Portal and Tail Suggest a Viral DNA Retention and Ejection Mechanism." *Nat Commun* 10, no. 1 (2019): 3746.
44. Bardy, P., T. Fuzik, D. Hrebik, R. Pantucek, J. Thomas Beatty, and P. Plevka. "Structure and Mechanism of DNA Delivery of a Gene Transfer Agent." *Nat Commun* 11, no. 1 (2020): 3034.
45. Vegge, C. S., H. Neve, L. Brondsted, K. J. Heller, and F. K. Vogensen. "Analysis of the Collar-Whisker Structure of Temperate Lactococcal Bacteriophage Tp901-1." *Appl Environ Microbiol* 72, no. 10 (2006): 6815-8.
46. Goulet, A., J. Mahony, C. Cambillau, and D. van Sinderen. "Exploring Structural Diversity among Adhesion Devices Encoded by Lactococcal P335 Phages with Alphafold2." *Microorganisms* 10, no. 11 (2022).
47. Kizziah, J.L., Manning, K.A., Dearborn,A.D., Dokland, T. . "Structure of the Host Cell Recognition and Penetration Machinery of a Staphylococcus Aureus Bacteriophage." *PLOS-Pathogens* (2020).
48. Zinke, M., K. A. A. Sachowsky, C. Oster, S. Zinn-Justin, R. Ravelli, G. F. Schroder, M. Habeck, and A. Lange. "Architecture of the Flexible Tail Tube of Bacteriophage Spp1." *Nat Commun* 11, no. 1 (2020): 5759.
49. Labrie, S. J., J. Josephsen, H. Neve, F. K. Vogensen, and S. Moineau. "Morphology, Genome Sequence, and Structural Proteome of Type Phage P335 from Lactococcus Lactis." *Appl Environ Microbiol* 74, no. 15 (2008): 4636-44.
50. Mahony, J., Oliveira, J., Collins, B., Haanemaaijer, L., Lugli, G.A., Neve, H., Ventura, M., Kouwen, T.R., Cambillau, C., van Sinderen, D. . "Genetic and Functional Characterisation of the Lactococcal P335 Phage-Host Interactions." *BMC Genomics* submitted (2017).

51. Veesler, D., G. Robin, J. Lichiere, I. Auzat, P. Tavares, P. Bron, V. Campanacci, and C. Cambillau. "Crystal Structure of Bacteriophage Spp1 Distal Tail Protein (Gp19.1): A Baseplate Hub Paradigm in Gram-Positive Infecting Phages." *J Biol Chem* 285, no. 47 (2010): 36666-73.
52. Kanamaru, S., P. G. Leiman, V. A. Kostyuchenko, P. R. Chipman, V. V. Mesyanzhinov, F. Arisaka, and M. G. Rossmann. "Structure of the Cell-Puncturing Device of Bacteriophage T4." *Nature* 415, no. 6871 (2002): 553-7.
53. Taylor, N. M., N. S. Prokhorov, R. C. Guerrero-Ferreira, M. M. Shneider, C. Browning, K. N. Goldie, H. Stahlberg, and P. G. Leiman. "Structure of the T4 Baseplate and Its Function in Triggering Sheath Contraction." *Nature* 533, no. 7603 (2016): 346-52.
54. Stockdale, S. R., J. Mahony, P. Courtin, M. P. Chapot-Chartier, J. P. van Pijkeren, R. A. Britton, H. Neve, K. J. Heller, B. Aideh, F. K. Vogensen, and D. van Sinderen. "The Lactococcal Phages Tuc2009 and Tp901-1 Incorporate Two Alternate Forms of Their Tail Fiber into Their Virions for Infection Specialization." *J Biol Chem* 288, no. 8 (2013): 5581-90.
55. Linares, R., C. A. Arnaud, G. Effantin, C. Darnault, N. H. Epalle, E. Boeri Erba, G. Schoehn, and C. Breyton. "Structural Basis of Bacteriophage T5 Infection Trigger and E. Coli Cell Wall Perforation." *Sci Adv* 9, no. 12 (2023): eade9674.
56. Varadi, M., S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Zidek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, and S. Velankar. "Alphafold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models." *Nucleic Acids Res* 50, no. D1 (2022): D439-D44.
57. Krissinel, E., and K. Henrick. "Inference of Macromolecular Assemblies from Crystalline State." *J Mol Biol* 372, no. 3 (2007): 774-97.
58. Morais, M. C., S. Kanamaru, M. O. Badasso, J. S. Koti, B. A. Owen, C. T. McMurray, D. L. Anderson, and M. G. Rossmann. "Bacteriophage Phi29 Scaffolding Protein Gp7 before and after Prohead Assembly." *Nat Struct Biol* 10, no. 7 (2003): 572-6.
59. Li, S., P. Roy, A. Travesset, and R. Zandi. "Why Large Icosahedral Viruses Need Scaffolding Proteins." *Proc Natl Acad Sci U S A* 115, no. 43 (2018): 10971-76.
60. Dearborn, A. D., E. A. Wall, J. L. Kizziah, L. Klenow, L. K. Parker, K. A. Manning, M. S. Spilman, J. M. Spear, G. E. Christie, and T. Dokland. "Competing Scaffolding Proteins Determine Capsid Size During Mobilization of Staphylococcus Aureus Pathogenicity Islands." *Elife* 6 (2017).
61. Orlova, E. V., P. Dube, E. Beckmann, F. Zemlin, R. Lurz, T. A. Trautner, P. Tavares, and M. van Heel. "Structure of the 13-Fold Symmetric Portal Protein of Bacteriophage Spp1." *Nat Struct Biol* 6, no. 9 (1999): 842-6.
62. Cuervo, A., M. C. Vaney, A. A. Antson, P. Tavares, and L. Oliveira. "Structural Rearrangements between Portal Protein Subunits Are Essential for Viral DNA Translocation." *J Biol Chem* 282, no. 26 (2007): 18907-13.
63. Linares, R., C. A. Arnaud, S. Degroux, G. Schoehn, and C. Breyton. "Structure, Function and Assembly of the Long, Flexible Tail of Siphophages." *Curr Opin Virol* 45 (2020): 34-42.
64. Arnaud, C. A., G. Effantin, C. Vives, S. Engilberge, M. Bacia, P. Boulanger, E. Girard, G. Schoehn, and C. Breyton. "Bacteriophage T5 Tail Tube Structure Suggests a Trigger Mechanism for Siphoviridae DNA Ejection." *Nat Commun* 8, no. 1 (2017): 1953.
65. Cambillau, C., and A. Goulet. "Exploring Host-Binding Machineries of Mycobacteriophages with Alphafold2." *J Virol* 97, no. 3 (2023): e0179322.