



HAL
open science

Structural Study of the *Cobetia marina* Bacteriophage 1 (Carin-1) by Cryo-EM

Alessio D'acapito, Thomas Roret, Eleftherios Zarkadas, Pierre-Yves Mocaër, Florian Lelchat, Anne-Claire Baudoux, Guy Schoehn, Emmanuelle Neumann

► **To cite this version:**

Alessio D'acapito, Thomas Roret, Eleftherios Zarkadas, Pierre-Yves Mocaër, Florian Lelchat, et al.. Structural Study of the *Cobetia marina* Bacteriophage 1 (Carin-1) by Cryo-EM. *Journal of Virology*, 2023, 97 (4), pp.00248-23. 10.1128/jvi.00248-23 . hal-04256769

HAL Id: hal-04256769

<https://hal.science/hal-04256769>








Submitted on 24 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Structural Study of the *Cobetia marina* Bacteriophage 1 (Carin-1) by Cryo-EM

 Alessio d'Acapito,^a  Thomas Roret,^b  Eleftherios Zarkadas,^c Pierre-Yves Mocaër,^d  Florian Lelchat,^e  Anne-Claire Baudoux,^d
 Guy Schoehn,^a  Emmanuelle Neumann^a

^aUniv. Grenoble Alpes, CEA, CNRS, IBS, Grenoble, France

^bStation Biologique de Roscoff (SBR), CNRS FR2424, Sorbonne Université, Roscoff, France

^cUniv Grenoble Alpes, CNRS, CEA, EMBL, ISBG, Grenoble, France

^dSorbonne Université, CNRS, UMR7144, Station Biologique de Roscoff, Roscoff, France

^eLeo viridis, Groupe Tacthys, Plouzané, France

ABSTRACT Most of studied bacteriophages (phages) are terrestrial viruses. However, marine phages are shown to be highly involved in all levels of oceanic regulation. They are, however, still largely overlooked by the scientific community. By inducing cell lysis on half of the bacterial population daily, their role and influence on the bacterial biomass and evolution, as well as their impact in the global biogeochemical cycles, is undeniable. *Cobetia marina virus 1* (Carin-1) is a member of the *Podoviridae* family infecting the γ -*protoobacteria* *C. marina*. Here, we present the almost complete, nearly-atomic resolution structure of Carin-1 comprising capsid, portal, and tail machineries at 3.5 Å, 3.8 Å and 3.9 Å, respectively, determined by cryo-electron microscopy (cryo-EM). Our experimental results, combined with AlphaFold2 (AF), allowed us to obtain the nearly-atomic structure of Carin-1 by fitting and refining the AF atomic models in the high resolution cryo-EM map, skipping the bottleneck of *de-novo* manual building and speeding up the structure determination process. Our structural results highlighted the T7-like nature of Carin1, as well as several novel structural features like the presence of short spikes on the capsid, reminiscent those described for *Rhodobacter capsulatus* gene transfer agent (RcGTA). This is, to our knowledge, the first time such assembly is described for a bacteriophage, shedding light into the common evolution and shared mechanisms between gene transfer agents and phages. This first full structure determined for a marine podophage allowed to propose an infection mechanism different than the one proposed for the archetypal podophage T7.

IMPORTANCE Oceans play a central role in the carbon cycle on Earth and on the climate regulation (half of the planet's CO₂ is absorbed by phytoplankton photosynthesis in the oceans and just as much O₂ is liberated). The understanding of the biochemical equilibriums of marine biology represents a major goal for our future. By lysing half of the bacterial population every day, marine bacteriophages are key actors of these equilibriums. Despite their importance, these marine phages have, so far, only been studied a little and, in particular, structural insights are currently lacking, even though they are fundamental for the understanding of the molecular mechanisms of their mode of infection. The structures described in our manuscript allow us to propose an infection mechanism that differs from the one proposed for the terrestrial T7 virus, and might also allow us to, in the future, better understand the way bacteriophages shape the global ecosystem.

KEYWORDS bacteriophages, virus structure, cryo-EM, marine podovirus, Alphafold, virion structure

Viruses are the most abundant and widespread form of life on earth. Bacteriophages (also called phages) are viruses that infect bacteria. They are classified into orders

Editor Rebecca Ellis Dutch, University of Kentucky College of Medicine

Copyright © 2023 American Society for Microbiology. All Rights Reserved.

Address correspondence to Guy Schoehn, guy.schoehn@ibs.fr, or Emmanuelle Neumann, emmanuelle.neumann@ibs.fr.

The authors declare no conflict of interest.

Received 16 February 2023

Accepted 17 February 2023

Published 21 March 2023

based on their genome type (RNA or DNA) and on their morphology. Most known phages belong to the order of the Caudovirales (1). They are characterized by an icosahedral capsid that encloses the genome (double stranded DNA [dsDNA]) and a tail that varies in morphology depending on the family they belong to: *Myoviridae* having a long straight contractile tail, *Siphoviridae* characterized by their long flexible tail, and *Podoviridae* with a short tail (1). The host recognition and infection are generally carried out by the distal section of the tail, called the tip. Some phages tail tips are equipped with polysaccharide depolymerases that specifically degrade the capsule surrounding their host and allow them to reach their membrane for infection (2, 3).

The ocean is the largest reservoir of viruses on Earth, with estimates of 1 to 100 billion viruses mL^{-1} in the surface waters, and 10 to 1000-fold higher concentrations in sediments. They are shown to play a major role in the control of the marine bacterial population, and, thus, the whole biogeochemical equilibrium of oceans (4) even though, today, their genetic diversity remains an unexplored reservoir (5). Despite this, marine phages are largely overlooked by the scientific community compared to their terrestrial counterparts.

Cobetia marina virus 1 (Carin-1) is a lytic marine podophage that infects the γ -*proteobacteria* *C. marina* DSM 4741 (6). This bacterium is known for its ability to produce exopolysaccharides, either attached to the cell (as a capsule) or released in ambient water (as dissolved organic matter) (7, 8). Previous biochemical studies assessed the presence and the activity of a structural polysaccharide depolymerase enzyme (Dpo36) bound to the podophage Carin-1, that infects its bacterial host (3). It is likely that Carin-1 uses its polysaccharide depolymerase to degrade the host capsule in the aim to access the bacterial membrane receptor. The detection of virus-bound depolymerizing activity also points to unsuspected interactions between Carin-1 (and more globally marine phages) and dissolved organic matter. To date, detailed structural insights and, especially the Dpo36 location on Carin-1, are lacking.

T7, the *Podoviridae* prototype, has been extensively studied at the structural level by cryo-electron microscopy (cryo-EM) comprising functional studies on the infection mechanism and infectivity (9–12), making it one of the most studied podoviruses. Here, we present the nearly complete atomic structure of the Carin-1 phage comprising capsid, portal, and tail machinery at 3.5 Å, 3.8 Å, and 3.9 Å, respectively, obtained by cryo-EM single particle analysis. The models revealed a T7-like tail conformation, as well as the presence of new features as a capsid decoration protein complex for which similar architecture is, to our best knowledge, only described for gene transfer agents (GTA) (13). Moreover, based on available X-ray crystallography data and biochemical studies (personal communications), we localized the depolymerase Dpo36 on the tail fibers of the phage, suggesting an infection mechanism different from the one proposed for the archetypal podophage T7 (12).

RESULTS

Carin-1 overall structure. Mature Carin-1 phages were produced and purified as described in the materials and methods section. They were first imaged by negative staining (14) to check the quality of the sample. Cryo-EM images showed a good homogeneity of the sample with only a small proportion of the phage capsids being empty of DNA or damaged (Fig. S1a). Carin-1 exhibits the classical shape of a podophage: an icosahedral head with one vertex occupied by a short tail (Fig. 1a, b, and d) (1). The capsid has a diameter of 700 Å and a shell thickness of 50 Å; the tail is 180 Å long and 150 Å large. Six flexible fibers, visible at low contours level and associated with the connector hub, also decorate the tail. The capsid is full of DNA organized as concentric layers separated by 24 Å (Fig. 1d). Due to the size of the particle but also to the different symmetries present in the particle, the “divide and conquer” method was applied. Each phage was split *in silico* into 3 parts: the capsid, the tail, and the portal-connector assembly. By combining the different parts, a nearly complete atomic model

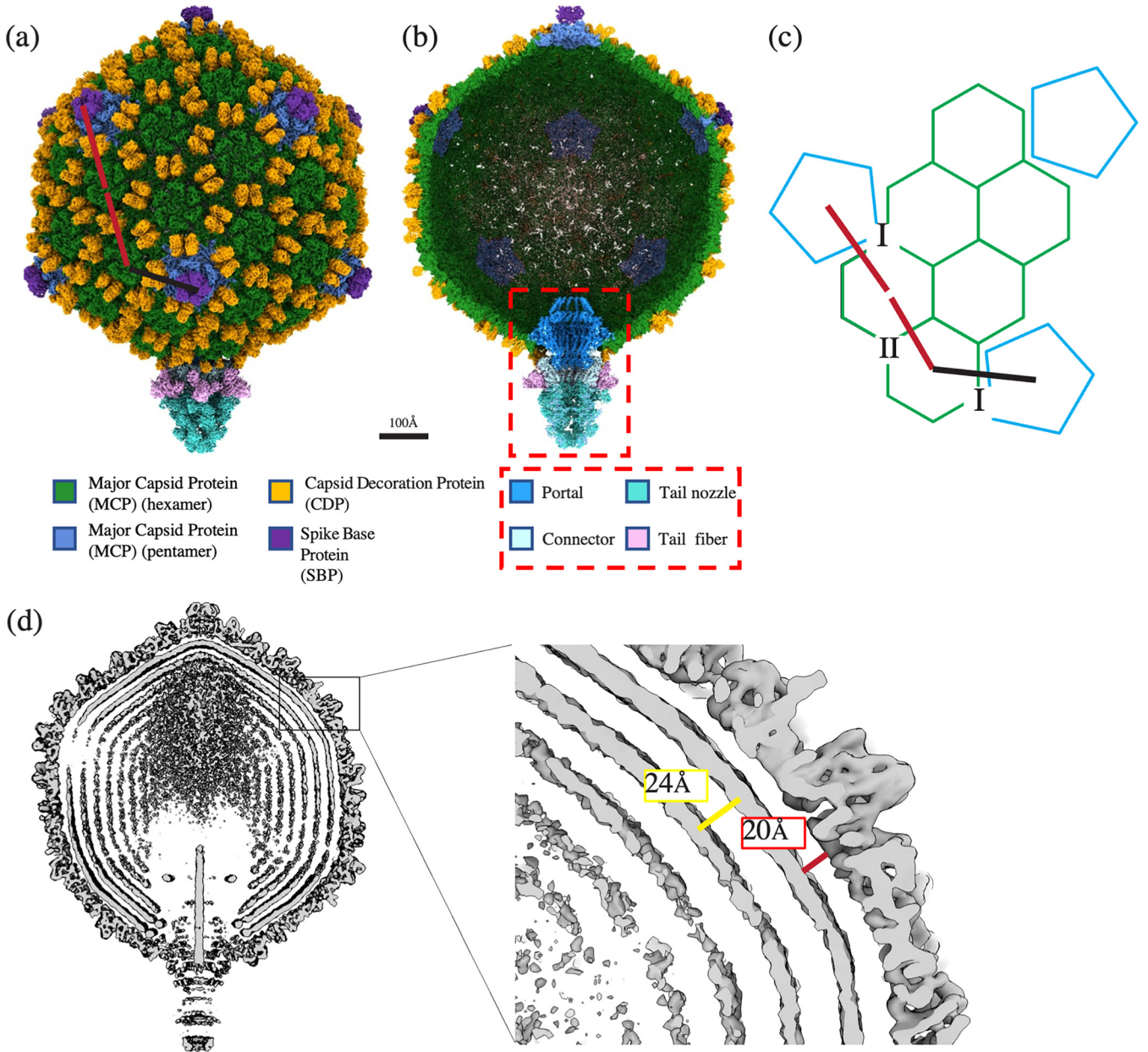


FIG 1 Three-dimensional structure of the Carin-1 bacteriophage. (a) External view of the isosurface representation of the Carin-1 three-dimensional structure. The T = 7,1 geometry is highlighted with h in red and k in black. (b) Internal view (the front half of the bacteriophage has been computationally removed and the DNA masked). For (a) and (b), the capsid (3.5 Å resolution) is represented in green (hexameric MCP), dark blue (pentameric MCP), yellow (CDP), and purple (SBP), and the portal (3.8 Å resolution) in blue, the tail assembly (3.9 Å resolution) comprising tail nozzle in blue, and connector in light blue. The resolved part of the tail fiber is in pink. (c) Schematic representation of the proteins layout. MPC hexamers are depicted in green and pentamers in blue, h and k are depicted as in (a). h = 2, k = 1. (d, left panel) Central section through the capsid density of a C1 reconstruction showing the capsid wall and the DNA arranging as concentric circles. (d, right panel) Detail showing the distances between the DNA-DNA layers (yellow bar) and DNA-capsid (red bar).

of the mature phage particle was obtained. It comprises the icosahedral capsid, the tail, the connector, and the portal. This model is based on the experimental cryo-EM Coulomb three-dimensional (3D) map combined with the generated AlphaFold2 models (15, 16) (Table S1).

Description of the icosahedral structure of the capsid. The capsid was reconstructed from 56,972 particles to 3.5 Å resolution (Fig. S1c) using Relion (17) by imposing icosahedral symmetry. It has a triangulation number T = 7,1 (Fig. 1a and c) and the shell is built by the major capsid protein (MCP) (Fig. 2b) that arranges both as hexamers on the facets and as pentamers on 11 of the vertices of the capsid (Fig. 1a, 2a). Compared

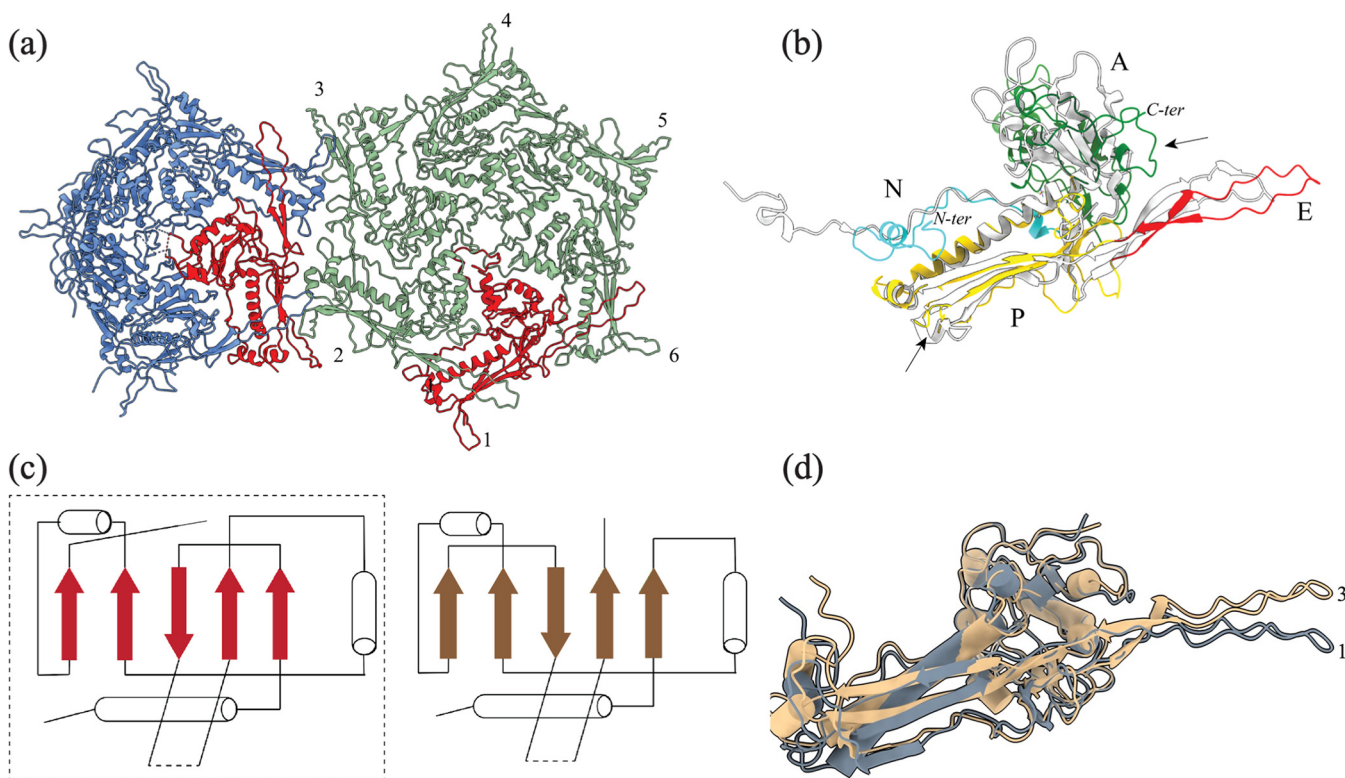


FIG 2 Structure of Carin-1 major capsid protein (MCP) seen from outside. (a) Top view of a pentameric (blue) and hexameric (green) assembly of the major capsid protein (MCP). One protein is depicted in red for clarity. Numbers correspond to individual subunits of the hexamer. (b) Ribbon diagram of Carin-1 (colored) and HK97 (PDB id 1OHG, light gray) MCP. The protein is divided into A-domain (green, containing the C-ter), P-domain (yellow), N-arm (light blue), and E-loop (red). Black arrows indicate, on the A-domain (left arrow), the 154 to 162 loop, and, on the P domain (right arrow), the short hairpin loop formed by residues 210 to 222, both involved in the contact with neighboring MCPs. (c) Comparison of the topologies of the A-domain of Carin-1 (red) and of the canonical HK97 (brown), showing differences in folding. (d) Superimposition of 2 MCPs (1 and 3, same numbering as in [a]) from the same hexamer shows differences in conformations.

to the podophage T7, the capsid is wider (Carin-1 diameter 700 Å versus 662 Å for T7 [9]) and is decorated with 3 extra types of proteins. On every local and icosahedral 2-fold symmetry axis, and sitting on the top of the MCP, one can find a globular shaped extra protein. This capsid decoration protein (CDP) is present as a dimer that bridges 2 capsomers (Fig. 1a and Fig. 3). The second extra complex is located on the top of the 11 MCP pentameric vertices (Fig. 4 and Fig. S2). This later complex is made of 2 different proteins, where one is well resolved, and the other shows great flexibility (less defined density) (Fig. S2).

Based on this global description, the architecture of Carin-1 capsid resembles more the one of the terrestrial podoviruses BPP-1 and Epsilon 15 than the one of T7 (18, 19).

Structure of the MCP (encoded by orf53). Orf53 encodes for the 346 amino acids long MCP (Table S2). MCP is organized either as hexamers making the facets of the capsid or as pentamers located on the 11 phage capsid vertices (Fig. 2). Based on the fit of the model obtained using AlphaFold2 (15, 16) into our map, the structure obtained by cryo-EM shows that its overall fold is a variation of the canonical HK97-fold also found in T7 (Fig. 2b and c) (9, 20, 21). Electron density was present in our experimental map for residues 3 to 344 in the hexamer, and for residues 3 to 295 and 308 to 344 for the pentamers. Overall, the MCP is organized in a very classical way, as in other bacteriophages, including T7 (9), into an N-terminal arm, Axial (A-), and Peripheral (P-) domains and an E-loop (Fig. 2b).

The A-domain is composed of a large β -sheet of 5 β -strands and 4 adjacent α -helices. However, several differences are visible when compared with the HK97 A-domain (20) (Fig. 2b and c): (i) the whole MCP A-domain is shifted toward the center of the hexamer by approximately 5 Å (Fig. 2b) when aligning the P-domains; (ii) the secondary structure

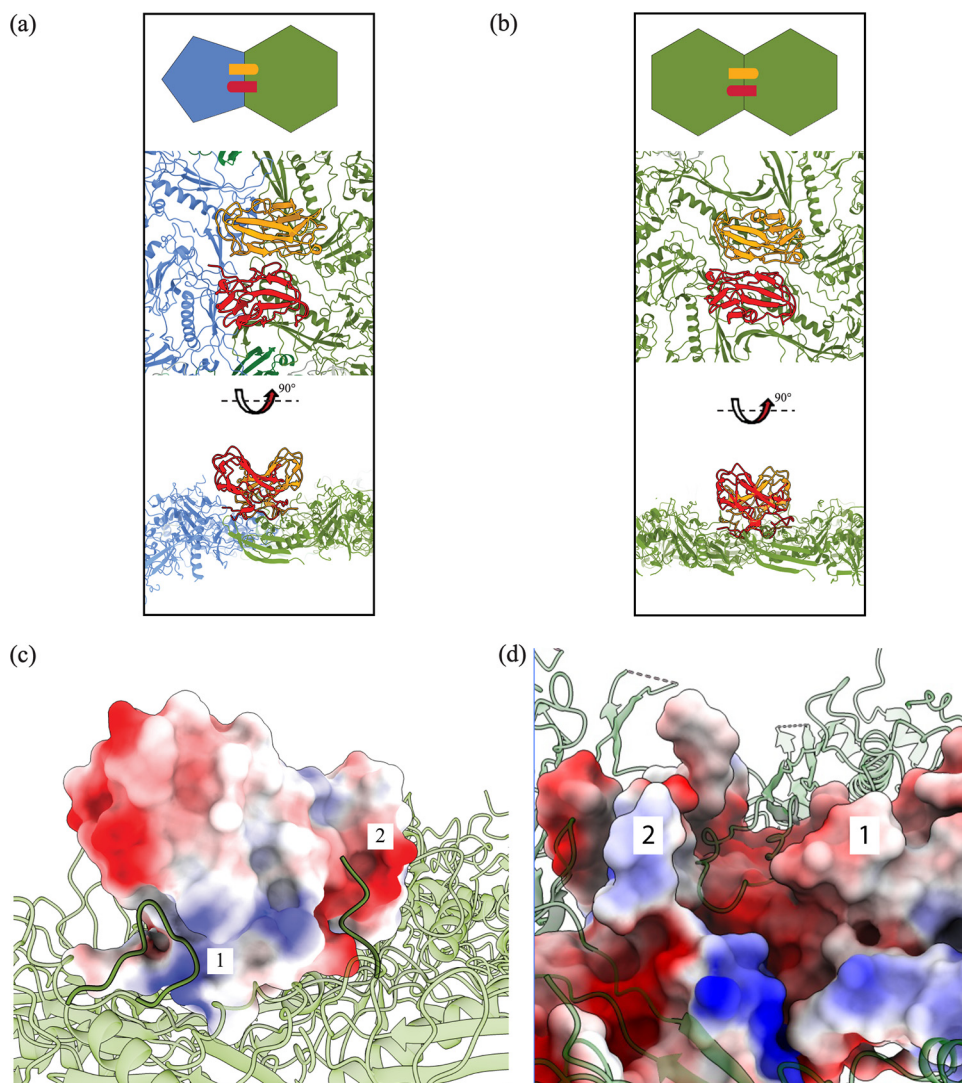


FIG 3 The capsid decoration protein dimer. (a) and (b) Arrangement of CDP over a pentamer and a hexamer of MCP (a) and 2 hexamers of MCP (b). For both representations, from top to bottom: schematic representation of the arrangement of CDPs over 2 capsomers (CDPs monomers are represented in orange and red for clarity. MCP pentamer is represented in blue, and the MCP hexamer in green); cartoon representation of the top view of the capsomers; side view of the capsomers after a 90° rotation. (c) Charge distribution on a CDP monomer which highlighted its interactions with 2 MCPs (same colors as [a] and [b]). (c, 1) Interaction with MCP1 P-domain. (c, 2) Interaction with MCP2 N-arm. (d) Charge distribution on a MCP1 P-domain. (d, 2) Interaction with MCP2 N-arm monomer which highlighted its interactions with a CDP monomer. Charge distribution in (c) and (d) are complementary.

elements are connected differently (Fig. 2c) within the A-domain: and (iii) the C-terminal part of the protein is extended compared to HK97 and lies on top of the β -sheet (Fig. 2b). Moreover, Carin-1 MCP A-domain also features a loop (residues 154 to 162) (Fig. 2b) that extends to contact the neighboring N-arm of the $n + 1$ subunit of the hexamer. As for the general architecture of the capsid, this alternative conformation of the A-domain compared to the canonical HK97 fold (20) is similar to the one described on the BPP-1 podophage (21).

The P-domain is composed of a long characteristic α -helix and 6 β -strands arranged as 1 extended antiparallel β -sheet as in HK97 (Fig. 2b). Residues 132 to 144 form a loop that extends over the β -strands and interacts with a CDP (see below). Residues 210 to 222 form a short hairpin loop that protrudes out of the capsid to reach (i) the P-domain of the subunit of another capsomer, (ii) the E-loop of the $n-1$ subunit of this capsomer, and (iii) the equivalent loop of the P-loop of the subunit on a third different

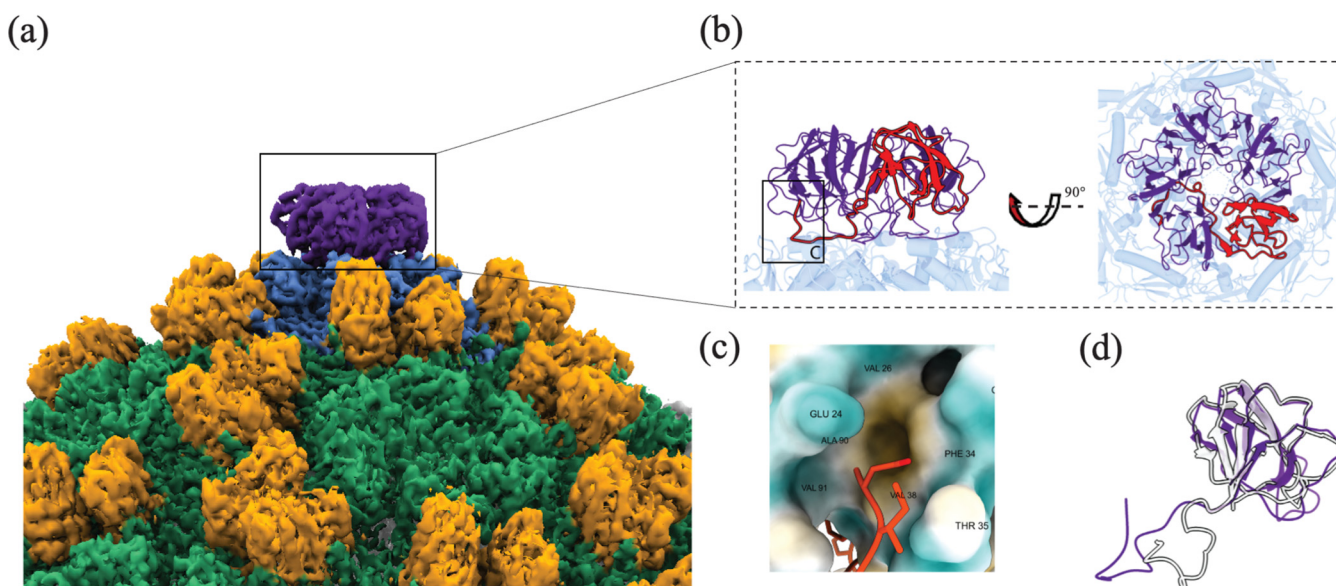


FIG 4 Structure of the spike base protein (SBP). (a) Location of the SBP in the Carin-1 capsid. Isosurface view of the capsid electron density. SBP is depicted in purple, MCP hexamers in green, MPC pentamers in blue, and CDP in orange. (b, left panel) Ribbon diagram of the side view of the spike base protein (SBP) showing the interaction between SBP (purple, 1 subunit in red), and the MPC (blue). (b, right panel) Top view of the SBP pentameric assembly. One subunit is colored in red for distinction. (c) Is the rectangle highlighted in (b) and depicts the interaction between the C-arm of 1 SBP subunit (orange) and the cavity formed by the next subunit (surface of hydrophobicity). Blue shows a hydrophilic surface and a yellow hydrophobic one. (c) is the rectangle highlighted in (b, left panel). (d) Superimposition of the structure of the SBP of Carin-1 (purple) and RcGTA (PDB id [6TSU](#), white) shown as ribbon diagrams.

capsomer (Fig. 2a). Overall, the hairpin loop is in contact with 2 other P-domains of 2 different capsomers locking the capsid structure.

The N-arm has 2 short α -helices and lies over the P-domain, as for HK97, but is bent by 90° so that the CDP can fit on it (see below and Fig. 3c and d).

The E-loop is composed by a characteristic 2-stranded β -sheet and a long hairpin loop. It does not differ from HK97 (Fig. 2b) (20). This loop extends over the P-loop of the $n + 1$ subunit of the hexamer and makes contact with 2 different CPs through the β -sheet and the tip of the loop.

The 6 quasi-equivalent subunits of a same hexamer differ slightly ($\text{RMSD}_{\text{max}} 3.0 \text{ \AA}$) with the most notable differences occurring at the level of the P-domain and E-loop. Depending on the subunit position, residues can be displaced up to 10 \AA at the level of the E-loop (Fig. 2d).

The pentameric conformation differs from the hexameric one by a notable bend of the E-loop (11.8 \AA) as it is the case for HK97 (20) or for T7 (9). The A-domain loop (aa 297 to 320) also adopts a different conformation by pointing toward the inside of the capsid. The intra- and inter-capsomer contacts of MCPs are mainly based on electrostatic interactions that allow quasi-equivalence with the T = 7,I capsid, as has been observed for other phages (22).

Structure of the CDP (coded by orf52). Elongated globular densities, corresponding to a decoration and/or stabilizing protein, were clearly visible on the external capsid surface from early reconstructions to the final one (Fig. 1a). Densities were remaining empty after the fitting of the MCP structure into the EM map. This is one of the main differences with the T7 capsid, as the latter is only made of 1 type of protein (9). AlphaFold2 (15, 16) was used to identify this new protein. We submitted several protein sequences derived from orfs adjacent to orf53 (the MCP) to the AlphaFold2 server, and found that the protein encoded by orf52 is the one forming the capsid decoration protein (CDP) (Table S2). Indeed, the CDP counts 130 amino acids and has a jellyroll fold built of 2 antiparallel β -sheets, each comprising 4 β -strands. On the capsid, CDPs can be found as dimers, 7 \AA apart from each other, and interacting through their upper-pointing β -sheets, forming a large β -stranded antiparallel β -sheet with a loop extruding on one end (Fig. 3a and b). The N- and C termini of each CDP are pointing toward the capsid

and interact directly with the P-domain of 2 MCPs from 2 different capsomers bridging them together. For each CDP dimer, the 2-fold symmetry axis is aligned with the radius of the capsid, the loop (residues 60 to 71) facing either inward in the capsomer or outward to the neighboring one. Each CDP clamps together the N-arm and the tip of the P-domain of a MCP with the P-domain 132 to 144 loop of an MCP of the adjacent capsomer, increasing the capsid stability. The N-terminus of a MCP is also anchored to residues 22 to 26 of the CDP (residues E6, N23, V24, I25, D26, and L27) through electrostatic interactions (Fig. 3c and d). In total, 1 CDP contacts 4 different MCPs from 2 adjacent capsomers, allowing them to robustly lock the capsomers into a solid capsid shell (Fig. 1a, and Fig. 3a and b). It can also play a role in cell binding like other phage decoration proteins (23). In this way, it can better resist the internal DNA pressure. Finally, Carin-1 CDP shows similarities with the BPP-1 phage's Bbp16 protein but lacks its elongated C-terminus (24).

Structure of the head spike assembly (spike base protein coded by orf50 and unidentified spike head protein). The tail-free vertices of the icosahedral capsid of Carin-1 are additionally decorated with turret-shaped spikes that we are calling head spike assembly (HSA) (Fig. 4a and b, and Fig. S2; feature not found in T7 [9], HK97 [20] and P22 [24, 25]). The "electron density" allows to distinguish 2 parts: the spike base protein (SBP) and a spike head protein (SHP) (Fig. S2b and c).

The part in contact with the MCP (SBP) is well defined as opposed to the apical region (SHP) (Fig. 4a, and Fig. S2c). The SHP was not modeled as it is too flexible to lead to a define density of high enough resolution, as shown by the two-dimensional (2D) classes (Fig. S2c). When using a higher contour level in the isosurface representation of the spike, a disk of density is appearing (Fig. S2b and c) confirming the flexibility of this protein.

The well resolved electron density of the SBP shows a β -sheet-only based assembly. Orf50 was identified as previously described for the CDP (Table S2) to encode for the SBP and electron density for 87 out of 92 amino acids (5 to 91), as was present in the map. SBP has a jellyroll fold formed by 2 β -sheets of 3 and 4 β -strands, respectively (Fig. 4b and d). Two adjacent monomers interact through an elongated hook-like C-arm (residues 5 to 22) that locks monomer n with the $n + 1$ one (clockwise turn) from below (Fig. 4b), and accommodates into a hydrophobic cavity (Fig. 4c) present in the $n + 1$ monomer. SBP pentamers are held in place on the capsid vertex by electrostatic interactions. The loop (295 to 308) of the pentameric form of the MCP most probably plays a role in stabilizing the spike as the map shows considerable blurred densities at the interface between the 2 proteins.

To our best knowledge, it is the first time such an assembly is described for a bacteriophage. In the GTA (DNA-containing pseudo phage particles produced by bacteria for horizontal gene transfer), particles of the marine *Rhodobacter capsulatus* RcGTA (13), an architecture like the one presented here, is described and shown to be involved in the recognition of the polysaccharide capsule of the bacteria (13).

Even if we have not been able to model SHP, to try to identify it, we used a bioinformatics approach and compared the RcGTA known experimental data to our data. RcGTA SHP is known to contain polysaccharides binding sites (13) (Uniprot Accession id: [D5AR34](#)), but only the part interacting with the spike base protein (residues 2 to 10, PDB id [6TSU](#)) is structurally known (26).

Based on the similar sequence length and genome position, we propose that the Carin-1 SHP would be encoded by orf49. Moreover, sequence alignments show that 6 out of the 9 modeled RcGTA residues are the same, and 2 are of the same nature (M/A and I/L substitutions) (Fig. S2e).

To go further and structurally characterize the protein encoded by orf49, Alphafold2 models of both this protein (Fig. S2d) and RcGTA SHP (data not shown) were generated. The resulting structures share the same overall architecture consisting in 2 jellyroll like domains with a long N-terminal arm, confirming that orf49 could encode for the Carin-1 SHP. The accuracy of the predicted structure for the orf49 protein is not perfect as some portions of the structure have low confidences of prediction (Fig. S2d). Therefore, only

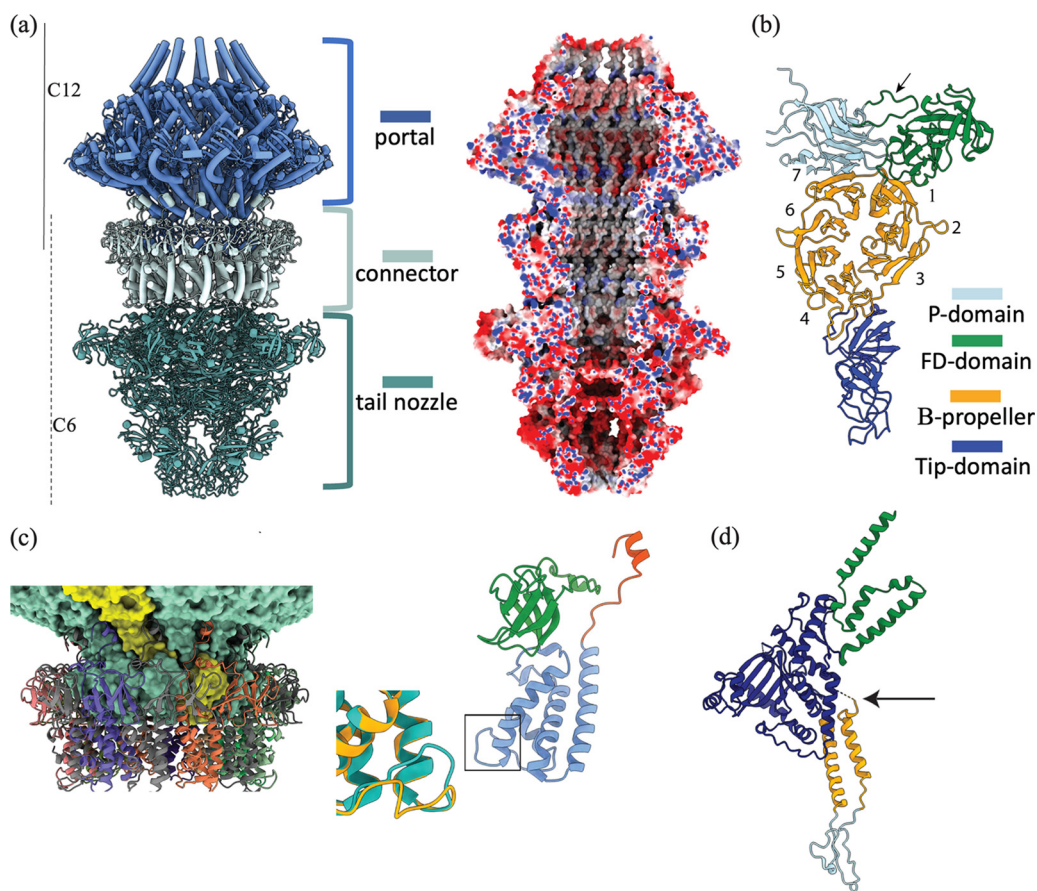


FIG 5 Structure of Carin-1 tail and portal assembly. (a, left panel) Ribbon diagram of the tail machinery including the portal (dark blue), the connector (light blue), and the tail nozzle (cyan). The symmetries of each part of the assembly are indicated. A, right panel) Same view as (a) but the front half of the complex has been removed and the surface colored by charges. (b) Ribbon representation of Carin-1 tail nozzle protein. The protein is divided into a P-domain (light blue), FD-domain (green), B-propeller (yellow, with numbered blades), and tip domain (blue). The black arrow indicates the 771 to 788 loop that acts as an interface with the tail fibers. (c, left panel) Assembly of the portal (depicted as blue surface, 1 subunit in yellow for clarity) and the connector (as ribbon, 1 color per subunit). (c, right panel) Ribbon diagrams of Carin-1 connector protein. Red, embracing helix; blue, helix bundle; green, fiber dock domain. The superimposition of the 2 possible conformations of the bundle loop are highlighted, either facing a tail fiber (blue) or not (orange). (d) Carin-1 portal protein. The protein comprises a crown domain (green), a wing domain (dark blue), the stem (yellow), and the clip (light blue). The structural localization of T7 channel valve is highlighted by the black arrow.

the portions where the backbone position can be considered as accurate (predicted local distance difference test IDDT above 50%) (Fig. S2d) were considered while drawing our conclusion (27).

Carin-1 orf49 encoded AlphaFold2 model was then submitted to the DALI server (28). The search relates orf49 distal domain to several peptidases, as well as the host interacting CBA120 phage tail fiber (Fig. S2e). These results are compatible with the experimental results on RCGTA spike head protein which is shown to contain oligosaccharide-interacting sites (13, 26). Carin-1 SHP (encoded by orf49) could, thus, be involved in the interaction with the host capsular polysaccharides.

Overall structure of the Carin-1 tail complex. The tail complex of Carin-1 is a tubular assembly composed of the portal, the connector, the tail nozzle, and the associated fibers. The complex is inserted into the capsid and located on one of its vertices (Fig. 1 a, b, and d). We solved the structure of the tail nozzle, the connector, the portal, and the proximal part of the tail fibers; encoded, respectively, by orf44, orf46, orf58, and orf36 (Fig. 5a and Table S2). A first map at 7 Å resolution was obtained by imposing a 6-fold symmetry on the entire complex where the dodecameric assemblies of the connector and portal were clearly visible. This map was then used for signal subtraction in order

to treat independently the portal (showing a 12-fold symmetry), and the tail and connector assemblies (showing a 6-fold symmetry). The whole tail assembly was reconstructed by imposing a 6-fold symmetry (for the nozzle, connector, and fibers, map at 3.9 Å resolution). The first 127 amino acids of the tail fibers are also visible in the tail reconstruction. By imposing 12-fold symmetry, the structure of the portal at 3.8 Å resolution was also obtained. The reconstruction clearly shows density above the portal that is most likely corresponding to the core proteins (i.e., the proteins that are ejected upon infection) but that were not solved (Fig. 1d) (11). All attempts at image analysis to obtain a 3D structure of the core protein failed, including the attempt to apply C8 symmetry as was done for T7 (10).

The interactions of the tail with the capsid are located at the level of the constriction between the connector and the portal assemblies (Fig. 1b and d), which generates a platform for interactions with adjacent MCPs. The portal wing domain sits on the inside of the capsid and the shell slots on the constriction. Due to the symmetry mismatch between the tail assembly (12-fold/6-fold) and the capsid (5-fold), there is an ambiguity about the exact interface between the two (6-fold symmetry averaged capsid for the tail reconstructions and 5-fold averaged tail for the capsid reconstruction). We don't have enough images to perform non-symmetrized reconstruction to reach significant resolution to bypass this ambiguity.

Structure of the tail nozzle (coded by orf44). The sequence coding for the tail nozzle was identified in the genome as described above (Table S2). The map obtained by signal subtraction allowed to fit in the density of all except the first 4 residues of the 830 amino acids long protein model obtained using AlphaFold2 (15, 16). The structure can be divided into 4 domains: a Tip domain (T-domain), a central β -propeller, a Fiber Dock domain (FD-domain), and a Platform domain (P-domain) on which the connector docks (Fig. 5b) (see below).

The central β -propeller is made of 7 blades all comprising the characteristic 4 anti-parallel β -sheets as described for T7 (10, 12). Blade 6 extends through the P-domain forming a long 8-stranded sheet. The propeller features, the so-called Velcro closure (29), on blade 7 (Fig. 5b and Fig. S3a), exhibit an extension between blades 2 and 3 to generate the T-domain.

As opposed to T7, the T-domain is connected to the β -propeller thanks to two 17 amino acid long loops (residues 158 to 175 and 317 to 334), and it comprises a 4 strands jellyroll β -barrel forming an "elbow" flanked by an additional 4-strands antiparallel β -sheet which resembles an OB-fold. The T-domain (Fig. 5b) is characterized by long distal loops between all secondary structure elements that could suggest a freedom for potential conformational change upon infection (10).

The P-domain (Fig. 5b) is facing the internal channel of the tail. It comprises the C- and N-terminal parts of the protein and its 2 β -sheets show a jellyroll- β -barrel-like fold. As described above, the P-domain interacts on one side with the β -propeller through β -sheet augmentation with the blade 7, increasing stability (Fig. 5b). On the opposite side, this domain acts as a platform on which the connector proteins will dock (Fig. 5a). One loop (residues 771 to 788) reaches the FD-domain and is involved in the interface with the tail fibers (see below) (Fig. 5b). The whole surface of the P-domain and the internal part of the T-domain are negatively charged, possibly to avoid interaction with the DNA during ejection (Fig. 5a).

The FD-domain points outward of the tail and acts as a scaffold on which the tail fibers dock. The domain is composed of 1 jellyroll β -barrel as described in T7, but also shows the insertion of another jellyroll β -barrel resembling OB-fold.

Two adjacent monomers of the tail nozzle are interacting together on 3 distinct points located on the P-domain, on the β -propeller, and on the T-domain as described for T7 (Fig. 5b). The P-P-domains interface is a flat surface, and interactions are driven by complementary electrostatic charges, as for the T-domain. The linker between blade 3 and blade 4 of the β -propeller is a negatively charged loop (residues 382 to 389) that protrudes and docks into the positively charged cavity of the β -propeller located in the next monomer, locking the 2 monomers in place. This leaves the "elbow" and

the higher part of the β -propeller free for possible conformational changes upon infection, as for T7 (12) (Fig. S3a).

Structure of the connector protein (encoded by orf46). The connector proteins bridge between the portal and the tail (Fig. 5a and S4c). The protein is encoded by orf46 (Table S2), and the amino acid 2 to 207 structural model (the protein is 222 aa long) generated from AlphaFold2 (15, 16) was placed into the density. The map showed that it assembles as a dodecameric ring of 125 Å diameter with a 6-fold symmetry. Indeed, 2 successive monomers are either facing a tail fiber or the inter-tail-fiber space. This results in different conformations of loop 14 to 29 (Fig. 5c). Except this loop, the 2 monomers of the asymmetric unit share the same fold (RMSD 1.6 Å) composed of a bundle of 4 α -helices, a fiber docking domain comprising a jellyroll of twice 3 antiparallel β -strands, and a C-terminal embracing helix (Fig. 5c) which is also very similar to T7 (10).

The C-terminal arm ends in an α -helix that docks in between 2 monomers of the portal (Fig. 5c). As all the other tail proteins, the connector shows a complementary distribution of the charges on its surface at the subunit interface that enhances stability. The dodecameric ring of the portal is inserted into the ring of the connector (Fig. 5c). Each monomer is stabilized by the C-terminal arm of the portal monomer that locks the hook of the portal clip domain. On the “lower” part, the α -helix bundle of the connector is attached to the P-domain of the tail nozzle that forms a flat surface in the complex. The N-terminal part of the tail nozzle protein acts as a pin that locks the connector ring into place; all interactions described above are stabilized by an extensive network of electrostatic interactions.

Structure of the portal protein (encoded by orf58). The complex (170 Å in diameter and 140 Å height) is formed by 12 copies of the protein coded by orf58 (Fig. 5d and Table S2). The central channel ranges from 68 to 42 Å in width. As described for other phage portals (12, 24), it comprises wing, crown, stem, and clip domains (Fig. 5d). It is also sharing great resemblance with T7's gp8 (10). The resolution of the 12-fold symmetry map is heterogenous ranging from 3.7 Å to 6 Å depending on the domain. Residues 4 to 218; 231 to 368, and 399 to 523 were docked in the map with great confidence from an AlphaFold2 model (15, 16). The wing is the core of the structure and has a triangular shape that points outward of the tail and sits on the inner side of the MCP shell. It comprises 8 α -helices with 4 arranged as an α -bundle and a sandwich of 2 β -sheets of 5 and 3 β -strands, respectively (Fig. 5d). Different from T7, we observe the presence of 2 extra helices at the upper tip of the triangle. In the dodecameric form, the wing domains line straight one against the other in a bookshelf-like fashion and are held together by the complementary distribution of charges on both wings. The stability is further increased by the interlocking of loops 163 to 169 and 263 to 273 that protrude from the β -sandwich into a cavity formed by the β -sandwich of the n-1 monomer (clockwise rotation observed from the capsid). The stem is composed of 2 α -helices which form the characteristic double layer of 24 helices when assembled in the entire mature portal. The clip domain is located at the bottom of the assembly. It is responsible for the interaction with the neighboring subunits through 2 inter subunit β -sheets as well as with the connector proteins through its hook-like shape (Fig. 5c and d) (see below). At the level of the crown, the map was sensibly noisier, and the resolution was limited to 5 Å. However, 4 α -helices were clearly visible and the C- α chain was easy to fit into the density up to amino acid 523 (of 556). The C-terminus of the protein arranges as a long α -helix, as in T7 (12). Interestingly, as T7's dodecamer shows a screw like assembly of the C termini that wraps around the DNA, Carin-1 helices lay straight parallel to the DNA which possibly suggests a difference in DNA packaging or release dynamics (Fig. S4b) (10, 12). The amino acids ranging from 368 to 399 were not modeled into the density due to high noise, but the chain is clearly pointing toward the DNA. Thanks to structural alignments of Carin-1 portal protein with T7's gp8, one can suggest that the unmodelled density could correspond to the channel valve (Fig. 5d, and Fig. S1d and Fig. S4b) (10).

Interaction between the tail machinery and the DNA. The tube-like density of Carin-1's dsDNA was clearly visible on the 2D classes. It extends clearly through the

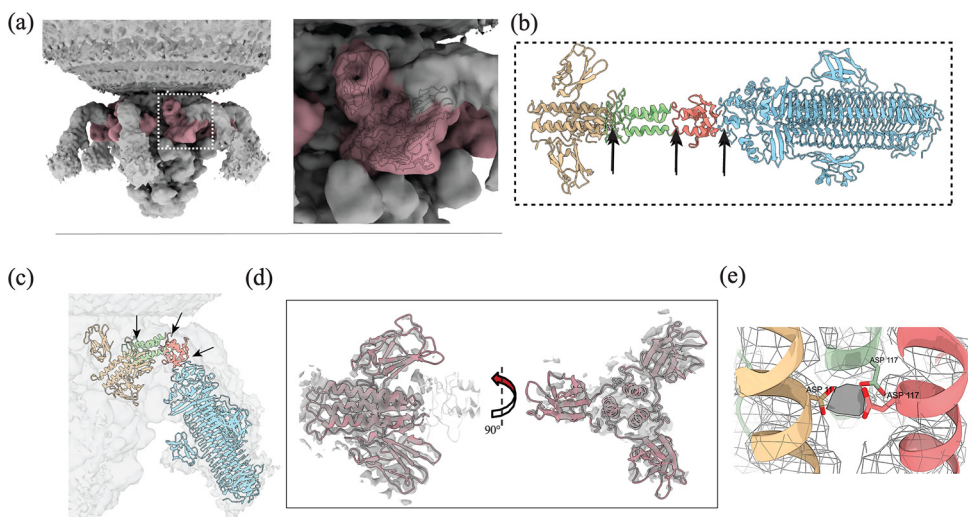


FIG 6 Dpo36 is located at the level of the tail fibers. (a, left panel) Overall 7.2 Å 3D reconstruction of the tail assembly shown as an isosurface. The tail fibers are detached from the capsid. The part of the density that allowed to reconstruct the N-term of Dpo36 is colored in pink. (a, right panel) Detailed view of the dotted square zone highlighted on (a, left panel), with the N-term of Dpo36 docked into the EM density. (b) Ribbon diagram of the crystal structure of Dpo36 (personal communication) colored by segment. (c) Ribbon diagram of the crystal structure of Dpo36 fitted by segments in the medium reconstruction of the tail fibers at 7.2 Å, same coloring as (b). (b and c) Black arrows indicate the hinges where the structure bends to get its characteristic shape. (d) Ribbon diagram of 2 orientations of Dpo36 (the red part is built in the cryo-EM map and the white part from the crystal structure) docked in the cryo-EM reconstruction of the tail fiber. (e) Detail of the interaction between the 3 helices of the N-term of Dpo36 depicted as ribbon (1 color per subunit). The electron density of the amino acid chain is shown as mesh, and the putative cation as plane surface.

core and portal assemblies (Fig. 1d). The electron density shows that the portal protein locks the DNA at the level of the putative channel valve and the uppermost part of the crown domain of the portal. In our structure (coming from phages filled with DNA) the tail nozzle holds and secures the dsDNA thanks to 2 negatively charged closing gates (aa 458 to 468 and 411 to 423, the gates are 5 and 9 Å wide, respectively) located in the internal channel of the nozzle (Fig. 1b and d, and Fig. 5a, and Fig. S3d). Both gates are composed of loops that extrude from β -sheets and are held together by a tight net of intra- and intermolecular hydrogen bondings and hydrophobic interactions, ensuring a hermetic closing of the tail. The opening of the channel must involve a large rearrangement of the whole tail nozzle in order to allow DNA ejection. This strategy possibly resembles the one from T7 as a similar architecture was described (10).

Structure and function of the tail fibers (encoded by orf36). Carin-1 exhibits 6 flexible tail fibers. As opposed to T7 (12), they are all detached from the capsid (Fig. 1a and b, and Fig. 6a). They extend from the tail nearly horizontally for 70 Å before bending on a 75° hinge (aa 127 to 170 on the crystal structure) and extending straight for 140 Å (Fig. 6a and c).

Due to the intrinsic flexibility of the tail fibers, our EM map only allowed to build the atomic model of the N-terminal's first 127 amino acids of the trimer of the tail fibers from the AlphaFold2 model (15, 16) that we generated from the sequence of the Dpo36 protein (encoded by orf36) (Table S2) (Fig. 6a). We fitted, in each subunit, 2 domains: one 7-stranded β -sheet immunoglobulin fold domain, and a straight α -helix (Fig. 6d). The interaction between the monomers occurs at the level of the helices that arrange as a bundle held by hydrophobic interactions. This assembly could be reinforced by the putative presence of a cation locked between the helices by the 3 asparagine residues (aa 117) as the map clearly shows a spherical density at this level (Fig. 6e).

The trimer sits on the 771 to 788 loop of the tail nozzle FD-domain, with tryptophane 780 of the FD that inserts in the hydrophobic pocket made by the α -bundle. Remaining interactions occurs between the immunoglobulin fold of each monomer

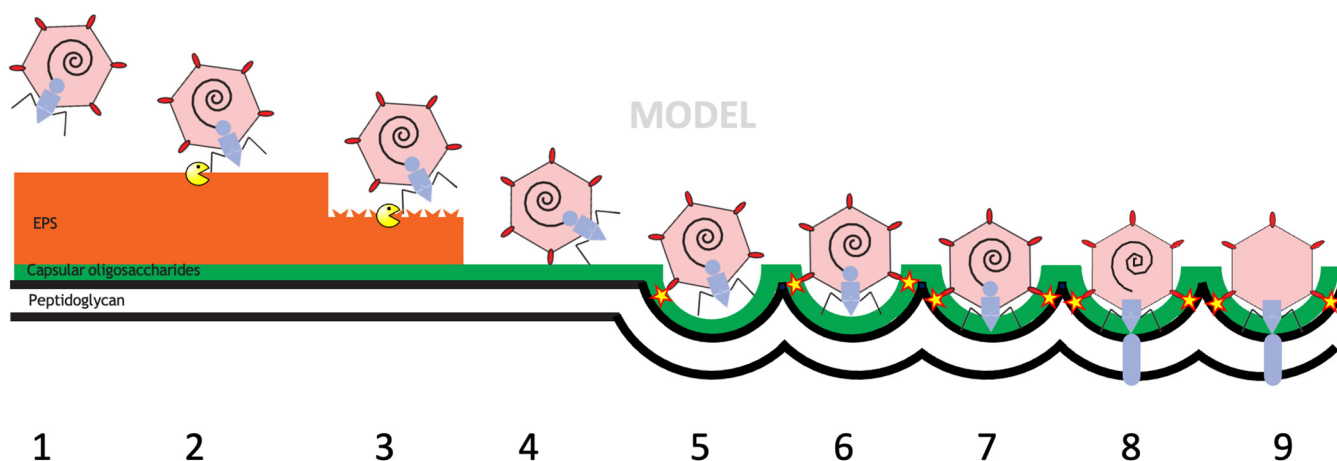


FIG 7 Proposed model for the infection mechanism of Carin-1. (1) Free particle. (2 and 3) Carin-1 degrades specifically (yellow « pacmac » *Cobetia Marina*'s EPS thanks to Dpo36). (4 and 5) Carin-1 could recognize (star) the capsular oligosaccharides of the host thanks to the head spike assemblies. (6) The phage reorients to infect the host and penetrate into the capsular oligosaccharides (7). (8 and 9) Ejection of the core proteins and perforation of the double membrane to inject the DNA in the cytoplasm of the host.

and either (i) the distal part of the FD-domain or (ii) the adjacent connector protein. The third outward pointing monomer does not interact with the tail (Fig. 6a).

A more complete structure (residues 6 to 663 over 828) of the tail fiber trimer was previously solved by X-ray crystallography in an extended conformation. The homotrimer is characterized by a 3- α -helix bundle flanked by 3 immunoglobulin folds on the N-term and a long β -helix also flanked by 3 immunoglobulin folds on the C-term (Fig. 6b). We were able to dock the crystal structure in the low resolution EM map, highlighting 3 hinges that give the assembly of its characteristic shape (Fig. 6b and c). Interestingly, the AlphaFold2 model (15, 16) fit the experimental EM map better than the crystal structure (Table S1). Indeed, the angle between the α -helix domain and the immunoglobulin domain is narrower in the *in-phago* EM model (and in the AlphaFold2 prediction) when compared to the isolated protein crystal structure, which shows a more open conformation. This is probably due to the contact between the different proteins in the crystal but could also be due to a conformation of the isolated protein in contrast to that integrated in the phage assembly.

Furthermore, biochemical studies characterized a polysaccharide depolymerase activity located within the C-terminal part/domain of Dpo36 (personal communication). Our structural results allow to determine the location of the depolymerase in the phage architecture, situating it close to the outer tips of tail fibers. Altogether, these results show and assess the key role of the tail fibers on the predation and infection mechanism and specificity, as described for other bacteriophages (30).

DISCUSSION

Overall, we solved the nearly complete structure of the marine podophage Carin-1. Our structural results allow us to propose an infection mechanism, which differs in some aspects from the one proposed for T7 (10, 12).

As *C. marina* cells are embedded in the L₆ exopolysaccharide (EPS) matrix (7), the polysaccharide depolymerase activity of the tail fibers might allow the phage to degrade and specifically pass the barrier of its host's EPS. The characteristic detached-from-the-capsid conformation makes the tail fibers available for polysaccharide recognition and degradation. It is documented that the tail fibers are the primary host recognition and attachment site for most of known phages (30, 31). As Carin-1 tail fibers are only shown to bear a depolymerase activity, Carin-1 specificity is driven by the described high substrate specificity of the tail fibers (3, 6, 32) but seems to lack the classical host-attachment site vastly described for phage tail fibers (33). We suggest that Carin-1 developed a different host recognition strategy as other known phages, driven by the HSAs that decorate its capsid (Fig. 7) (13).

In *Rotobacter capsulatus* GTA (RcGTA), structures sharing (i) the same structural localization as Carin-1's HSA, (ii) the same structural characteristics (RMSD of the SBPs = 0.095 Å, SHP are long and flexible as in RcGTA) (Fig. 4d) were shown to be involved in the infection mechanism by acting as a starting point for host capsule recognition (13). As biofilms can be populated by different living entities and also supported by bioinformatics results (Fig. S2), we suggest, that the HSAs of Carin-1 are involved in the specific recognition of the host capsule in the context of the biofilm, allowing a productive infection. This differs from the canonical T7 infection mechanism as T7's host recognition is driven by the tail spikes that do not present any depolymerase activity (10). This is just a model and supplemental experimental data are needed to confirm this hypothesis.

To cope with the apparent lack of attachment sites on the tail fibers, we propose that another level of host specificity could be added thanks to the OB-folds present on the tail nozzle assembly (absent from T7 whose attachment is driven by the tail fibers) (Fig. S3a) (34) but present in some podophage decoration protein.

As OB-folds are shown to recognize and bind oligosaccharides (34), the structures decorating the tail nozzle could therefore be a putative additional host-binding site, and is possibly involved after HSA recognition in similar mechanisms than described for RcGTA, that first recognizes the capsular polysaccharides with its HSAs and then binds the bacterial envelope with the tail machinery (13). These results add another layer of structural homology between GTAs and bacteriophages, highlighting evolutionary linkage between both. Moreover, with *R. capsulatus* being a marine bacterium, recognition devices, such as HSAs, could be a common feature to marine phages and phage-like particles.

As for T7, the release of the phage DNA is triggered by the recognition of lipopolysaccharides in the surface of *E. coli* (10), the similarity in architecture of the portal, connector, and tail nozzle proteins suggests that the DNA release of Carin-1 is highly like the one proposed for T7 (10, 12, 10).

Overall, our results are, to our knowledge, the first high resolution structures of a complete marine bacteriophage, the fourth full mature structure of a Gram-negative infecting podovirus and the second T7-like podovirus structure (9, 12, 24, 25, 35). This structure is showing great structure similarity with known podophages despite the low sequence homology. This also allowed us to give further insights into the biology of marine bacteriophages and the evolutionary relation between GTAs and phages.

All the models built for this work were first generated with the AlphaFold2 artificial intelligence algorithm (15, 16), showing great results (Table S1). In organisms, such as bacteriophages, where structural homology is high across species, artificial intelligence can remove the throughput bottleneck of atomic modeling and focusing efforts only on model refinement. Moreover, the barrier of sequence homology can be overtaken in cases like bacteriophage where sequence homology is low, but structure homology is high, thus facilitating the identification of proteins in unannotated genomes and allowing kind of visual annotation of genomes.

Overall, the advances of AlphaFold2 allowed to drastically decrease the time spent on repetitive tasks such as model building and sequence assignment, increasing the throughput of the cryo-EM pipeline, and showing once again the incredible potential that artificial intelligence has to offer to the field to understand biological processes by determining macromolecular assembly structure.

MATERIALS AND METHODS

Bacterial host culture. The bacterial host *C. marina* DSM 4741 was cryo-conserved at -80°C in a 17% glycerol solution. Cultures were reactivated on a solid autoclaved ZoBell medium (5 g peptone, 1 g yeast extract, 15 g agar diluted in 1 L of 0.2 μm filtered seawater). One day prior to the virus production, a single colony was picked from the solid medium and transferred into autoclaved ZoBell Broth (5 g Peptone, 1 g Yeast extract in 1 L of 0.2 μm filtered seawater). The culture was incubated overnight at room temperature (approximately 20°C) under vigorous agitation (180 rpm) on a rotary table.

Viral culture. The viral isolate (Carin-1) that infects the marine bacteria *C. marina* DSM 4741 was isolated by plaque assay from coastal Atlantic marine waters (6). Stock solutions of viral culture was kept at 4°C in the dark. Prior to the production for cryo-EM analysis, the infectivity of the viral suspension was

assessed by spot test. Therefore, an aliquot the stock viral suspension was serially diluted in SM buffer (100 mM NaCl, 8 mM MgSO₄·7H₂O, 50 mM Tris-HCl 1M pH 7.5 in MilliQ water), and a 5 μL drop of each dilution was spotted on a host lawn obtained by plating a 1:4 mixture of host culture in molten agar (Zobell Broth completed with 0.6% [wt/vol] noble agar) onto a Zobell agar plate (5 g Peptone, 1 g Yeast extract, 15 g Agar in 1 L of 0.2 μm filtered seawater). The plate was incubated overnight at room temperature, and translucent spots were enumerated to compute the titer of the viral suspension.

Bacteriophage production and purification. A volume of 300 mL of host culture growing in ZoBell medium was infected at a multiplicity of infection on 0,1 based on the spot assay conducted previously. The mixture was incubated at room temperature under low agitation (60 rpm) on a rotary table for 18 h. The infection was stopped by centrifugation at 7000 g for 30 min (Eppendorf 5810R) to remove non-infected host cells. The supernatant was filtered through 0.2 μm PES filters (Filtration unit Stericup GP Millipore) to collect the virions. The viral suspension was concentrated by tangential filtration using a Vivaflow 200 cartridge (Sartorius Stedim) made of polyethylsulfone (PES) with a 30 kDa cutoff. The 300 mL suspension was concentrated down to 25 mL, the membrane was flushed back with another 25 mL of 30 kDa filtered seawater to improve the recovery yield leading to a final sample volume of 50 mL. The 50 mL-sample was further concentrated using a Vivaspin20 centrifugal concentrators (Sartorius Stedim) made of PES with a cutoff 30 kDa. Repeated runs of centrifugation (6000 g, 15 min) were performed until the sample reached a final volume of 1 to 2 mL. The concentrator was then incubated overnight at 4°C in order to maximize virus recovery. Finally, the viral concentrate was purified on a linear 10 to 40% sucrose gradient by ultracentrifugation using a SW41Ti rotor at 134074g for 45 min, 4°C (Optima XPN-80, Beckman). The band was extracted with a syringe (1 mL) and washed 3 times with 15 mL of SM buffer in a 30 kDa PES vivaspin column to remove traces of sucrose. The purified viral suspension was also kept overnight at 4°C to improve virus recovery.

Preparation of the cryo-EM grids and image acquisition. Phages were first imaged by negative-stained using Uranyl Acetate to check the state, the homogeneity, and the concentration of the sample (14). Because the concentration was too low for cryo-EM, the sample was concentrated by pelleting it using centrifugation at 15,000 rpm for 15 min, and resuspending it in 15 μL of phosphate-buffered saline (PBS) prior to vitrification and cryo-EM data collection. A total of 3.5 μL of Carin-1 samples were deposited on negatively glow discharged (45s; 25 mA) Quantifoil R1.2/1.3 EM grids and plunge-frozen in liquid nitrogen-cooled ethane using a Vitrobot (ThermoFisher) (100% humidity, Temperature 20°C; Blot force 0; blotting time 5s; wait time 5s; drain time 0.5s). The grids were transferred either to a ThermoFisher Scientific Glacios electron microscope or to a FEI Krios G3 electron microscope. A total of 981 micrographs (40 frames movies, total dose 40 electrons/Å²) were collected on the Glacios electron microscope on a Gatan K2 summit direct electron detector, and 8,630 micrographs (30 frames movies, total dose 30 electrons/Å²) on the Krios electron microscope equipped with a Quantum energy filter coupled to a Gatan K3 direct electron detector (European Synchrotron Radiation Facility) (36). The data collection was performed using SerialEM 3.8 software (Glacios) and EPU (Krios) with a defocus range of $-0.8 \mu\text{m}/-2.5 \mu\text{m}$ (Glacios) or $-0.7 \mu\text{m}/-4 \mu\text{m}$ (Krios). A nominal magnification of x22k (Glacios, K2 super resolution) or x64k (Krios) was used resulting in a pixel size of 0.94 Å (Glacios) or 1.35 Å (Krios).

Image analysis. The data analysis was performed with the software Relion 3.1.3 (17). Image processing of the Glacios data set allowed to obtain low to medium resolution models for the capsid and the whole tail machinery that were directly used as references for 3D refinement of the Krios data set. The 2 data sets were analyzed in the same way, unless stated otherwise. Gain normalized movies were imported, dose weighted, and motion corrected using Relion's own implementation, and CTF parameters were determined for each micrograph with Ctfind 4.1 (37), for both data sets. A ranking of the best images, based on maximum resolution, defocusing, figure of merit and ice quality, as well as manual cleaning, allowed us to select 933 micrographs from the Glacios data set and 6,615 from the Krios one.

Image analysis of the capsid. Manual picking of approximately 200 viral capsid particles allowed to calculate 2D classes with representatives of 3 representative orientations of the icosahedral capsid: 5-, 3- and 2- fold symmetries. These were used for automated picking on a small subset of the micrographs for parameter optimization prior to picking on all micrographs resulting in a total of 6,398 particles for the Glacios data set and 78,546 for the Krios one. Due to the large size of the capsid, the particles were extracted in 800 × 800 pixel² boxes and binned 4 times (i.e., 200 × 200 pix²) for the processing steps. Particles were extensively cleaned by iterative 2D classification to reach a final set of 5,795 capsid particles for the Glacios data set, and 56,972 for the Krios one (Table S3). From the Glacios data set, an initial 3D model of the capsid was generated with Relion without imposing any symmetry and used for 3D refinement by imposing the icosahedral symmetry. The final 6.7 Å resolution Glacios model was used for mask creation and subsequent masked 3D refinement on the Krios data set using a box size of 800² pix² and a pixel size of 1.35 Å/pixel, leading to a 4.4 Å resolution map. After further postprocessing (B-factor: -108 Å²), and polishing (anisotropic magnification estimation, per particle CTF estimation, defocus, and astigmatism fitting and beam tilt estimation) the final map showed a resolution of 3.5 Å as estimated by Fourier Shell Correlation (gold-standard FSC used in Relion of 0.143 [17]).

Image analysis of the tail, portal, and connector. As template-based automatic picking was not successful to select only the tails, the parameters were optimized to pick all peripheral 5-fold axes of all visible capsids in all micrographs, thus comprising all the visible tails. Iterative 2D classification led to the cleanup of the data set and resulted in 2 sets of particles: the head spikes (processed but didn't lead to any improved reconstruction compared to the one performed on the entire capsid) and the tails. The circular mask used for the tail comprised the whole nozzle and part of the capsid. The 336 particles picked in the Glacios data set were used to generate an *ab initio* model that was further 3D refined by imposing the C6 symmetry to a 22.3 Å resolution reconstruction. As for the capsid, the Glacios tail 3D

final model was used as an initial model for 3D refinement using the 11,395 Krios tail particles (box size 500×500 pixel² and 1.35 Å/pixel). The 6-fold symmetrized 3D reconstruction of the whole tail machinery, including the dodecameric assembly of the portal and the connector, exhibit a 7.2 Å resolution.

Starting from the 2D classes of the tail, the particles were reextracted and recentered on the tail nozzle. The 7.2 Å resolution map was used to create a mask for signal subtraction. Recentered particles were subtracted and re-boxed in 300×300 pixel², and the mask aligned and applied to the reconstruction in Chimera (38). The model obtained using the C6-symmetry-imposed 3D refinement on the subtracted particles reached a resolution of 4.5 Å, and was further postprocessed (B-factor: -80.0 Å²) to a resolution of 3.9 Å according to the FSC (0.143 criterion implemented in Relion [17]).

In a second step, particles were also recentered on the portal and on the connector, and reextracted as before. The same protocol as for the tail nozzle was followed except that a 12-fold symmetry was applied for the 3D refinements. The portal and connector assembly maps reached a resolution of 4.3 Å before postprocessing and 3.8 Å after postprocessing (B-factor: -70 Å²) as estimated by gold-standard FSC (0.143 criterion implemented in Relion [17]).

Bioinformatics, open reading frame (orf) identification, and figure preparation. The genome of Carin-1 was entirely sequenced and annotated by collaborators (personal communication). A BLAST (39) search of all orfs helped to identify candidates for the MCP, tail nozzle, portal, and connector proteins. These sequences were submitted to AlphaFold2 (15, 16) via the ColabFold notebook (available at <https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>), and the monomer models rigidly fitted into the respective maps to identify the corresponding one. We identified the MCP, CDP, tail nozzle connector, and portal proteins with this strategy. For the spike base protein (SBP), the corresponding density was isolated from the capsid reconstruction using the Chimera "Volume Eraser" tool and segmented in putative monomers using Segger inside Chimera (37). One monomer was then submitted to Phenix_map_to_model (40) to build a polyA chain in the density using default parameters. Based on (i) the fact that all structural proteins of the phage genome are clustered, (ii) the length of the segment built by Phenix, and (iii) the apparent predominant β -sheet based fold of the protein visible in the reconstruction, the 6 candidates orfs (orf50, 51, 52, 56, 59, and 60) were submitted to the Jpred4 server (41) for further sorting. The resulting candidates were then modeled using AlphaFold2 (15, 16) and fitted into the EM density as described above.

All RMSDs were calculated, and all figures were prepared in ChimeraX (42).

General model building. All models were built the same way unless stated otherwise. Each raw AlphaFold2 model (15, 16) was first rigidly fitted into the EM density. In a second step, a flexible fit by interactive molecular dynamics using ISOLDE in ChimeraX (43) was performed to fit the main chain. The amino acids that did not belong to any resolved density were removed at this stage.

The geometry of each monomer model was then improved by several iterations of manual rebuilding using Coot (44) and of automated refinement using Real_space_refinement in Phenix (45) (using default parameters at the maximal resolution of the map). The models were then assembled as homomultimers to remove clashes between subunits. Given that the maps are at too low resolution to allow per-atom placement, particular attention was given to build models with the best possible stereochemistry (total number of bond angles and length outliers lower than 4σ), and with as close as possible to no Ramachandran or rotamer outliers. Moreover, all the clashes between side chains under 1.5 Å were manually corrected. Once the models reached satisfying refinement statistics, the entire asymmetric unit was assembled, and the inter-protein clashes further refined by iterations of manual checks in Coot and real space refinement in Phenix. One complete asymmetric unit surrounded by 2 neighbors asymmetric units were finally refined the same way as described above to remove inter-asymmetric unit clashes. Toward the end of each refinement stage, structures were submitted to Molprobit (46) to pinpoint the issues of the structure. Finally, the model of each asymmetric unit was submitted to the PDB validation server (available at <https://validate-rcsb-2.wwpdb.org/>) for a final quality check.

Model building of the capsid. The AlphaFold2 model (15, 16) of the MCP was docked in the map unambiguously, thanks to the P-domain α -helix that served as landmark. Due to hexamer pseudo-symmetry, each unit was individually refined as described above. For the pentameric form of the MCP, one of the refined hexamer models was used and flexibly fitted to accommodate it in the density and refined as described above. The 7 CPs composing 1 asymmetric unit were individually refined as well because of the slightly different conformation of the MCPs in the underlying hexamer/pentamer interface. The SBP model was built and refined as described above.

Model building of the portal. The AlphaFold2 model (15, 16) of the portal was built as described above except for the α -helical crown domain that was first built in the lower resolution density region using Coot's "place helix here" tool. The different polyA chains alone were linked together, the correct different amino acids were built, and then real space refined in Phenix at 5 Å resolution. The refined crown domain was then connected to the rest of the molecule and refined as described above.

Model building of the tail complex. The connector protein model was first fitted and refined in the portal-connector complex map because the reconstruction at this level is of better quality than in the 6-fold symmetrized tail one. This PDB model was then refined, as described above, in the tail map in order to discriminate between the 2 possible conformations of loop 14 to 29. The tail nozzle protein was refined as described above. The first 150 aa of the Dpo36 trimer were modeled with the AlphaFold2 multimer ColabFold notebook (available at <https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb>). The model of the trimer was then fitted into the density, trimmed, and refined as described above.

Data availability. Cryo-EM density maps of the capsid, the tail and the connector/portal, and the associated atomic coordinate have been respectively deposited in the PDB and in the Electron

Microscopy Data Bank (EMDB) under the following accession codes: PDB ID [8CJZ](#), [EMD-16687](#); PDB ID [8CK1](#), [EMD-16689](#); PDB ID [8CK0](#), [EMD-16688](#).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 2 MB.

ACKNOWLEDGMENTS

We thank Leandro Estrozi, Grégory Effantin, Dominique Housset, and Romain Linares for help with the processing; Aymeric Peuch for maintaining our cluster; and Pierre Legrand, Serena Sirigu, and Mirjam Czjzek for providing the pdb of Dpo36, and for reading and comments on the manuscript. This research used the EM facility at the Grenoble Instruct-ERIC Center (ISBG; UAR 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB). IBS platform access was supported by FRISBI (ANR-10-INBS-05-02) and GRAL, a project of the University Grenoble Alpes graduate school (Écoles Universitaires de Recherche), CBH-EUR-GS (ANR-17-EURE-0003).

The IBS electron microscope facility is supported by the Auvergne-Rhône-Alpes Region, the Fondation pour la Recherche Médicale (FRM), the Fonds FEDER, and the GIS-Infrastructures en Biologie Santé et Agronomie (IBiSA).

We acknowledge the provision of BAG experimental time from the CM01 facility at the ESRF Eazhisai Kandiah for the CM01 data collection. This research was partly funded by the Agence Nationale de la Recherche, grant numbers ANR-21-CE11-0023 to G.S. and E.N., and ANR-15-CE01-0009-01 to A.-C.B.

IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG, CEA).

We declare no conflict of interest.

G.S. and A.-C.B. conceived the project. A.-C.B. prepared the Carin-1 sample. G.S. optimized cryo-grid preparation. G.S., E.Z., and Eazhisai Kandiah (E.S.R.F.) recorded the cryo-EM data. A.D.A. processed the data under the supervision of E.N. A.D.A., E.N., and G.S. analyzed and interpreted the structure. A.D.A., E.N., and G.S. wrote the paper with contributions from A.-C.B. and all other authors. All authors contributed to the editing of the manuscript.

REFERENCES

- Ackermann HW. 1998. Tailed bacteriophages: the order caudovirales. *Adv Virus Res* 51:135–201. [https://doi.org/10.1016/s0065-3527\(08\)60785-x](https://doi.org/10.1016/s0065-3527(08)60785-x).
- Yan J, Mao J, Xie J. 2014. Bacteriophage polysaccharide depolymerases and biomedical applications. *BioDrugs* 28:265–274. <https://doi.org/10.1007/s40259-013-0081-y>.
- Mocaër P-Y. 2019. From gene to ecosystem : an integrative study of polysaccharide depolymerases bound to marine viruses. *Ecosystems*:1–206.
- Fuhrman JA. 1999. Marine viruses and their biogeochemical and ecological effects. *Nature* 399:541–548. <https://doi.org/10.1038/21119>.
- Hurwitz BL, Sullivan MB. 2013. The Pacific Ocean Virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology *PLoS One* 8:e57355. <https://doi.org/10.1371/journal.pone.0057355>.
- Lelchat F, Mocaer PY, Ojima T, Michel G, Sarthou G, Bucciarelli E, Cérantola S, Colliac-Jouault S, Boisset C, Baudoux A-C. 2019. Viral degradation of marine bacterial exopolysaccharides. *FEMS Microbiol Ecol* 95. <https://doi.org/10.1093/femsec/fiz079>.
- Lelchat F, Cérantola S, Brandily C, Colliac-Jouault S, Baudoux A-C, Ojima T, Boisset C. 2015. The marine bacteria *Cobetia marina* DSMZ 4741 synthesizes an unexpected K-antigen-like exopolysaccharide. *Carbohydr Polym* 124:347–356. <https://doi.org/10.1016/j.carbpol.2015.02.038>.
- Maréchal J-P, Culioli G, Hellio C, Thomas-Guyon H, Callow ME, Clare AS, Ortalo-Magné A. 2004. Seasonal variation in antifouling activity of crude extracts of the brown alga *Bifurcaria bifurcata* (Cystoseiraceae) against cyprids of *Balanus amphitrite* and the marine bacteria *Cobetia marina* and *Pseudoalteromonas haloplanktis*. *J Exp Mar Biol Ecol* 313:47–62. <https://doi.org/10.1016/j.jembe.2004.07.016>.
- Guo F, Liu Z, Fang P-A, Zhang Q, Wright ET, Wu W, Zhang C, Vago F, Ren Y, Jakana J, Chiu W, Serwer P, Jiang W. 2014. Capsid expansion mechanism of bacteriophage T7 revealed by multistate atomic models derived from cryo-EM reconstructions. *Proc Natl Acad Sci U S A* 111:E4606–E4614. <https://doi.org/10.1073/pnas.1407020111>.
- Chen W, Xiao H, Wang L, Wang X, Tan Z, Han Z, Li X, Yang F, Liu Z, Song J, Liu H, Cheng L. 2021. Structural changes in bacteriophage T7 upon receptor-induced genome ejection. *Proc Natl Acad Sci U S A* 118:e2102003118. <https://doi.org/10.1073/pnas.2102003118>.
- Pérez-Ruiz M, Pulido-Cid M, Luque-Ortega JR, Valpuesta JM, Cuervo A, Carrascosa JL. 2021. Assisted assembly of bacteriophage T7 core components for genome translocation across the bacterial envelope. *Proc Natl Acad Sci U S A* 118:e2026719118. <https://doi.org/10.1073/pnas.2026719118>.
- Hu B, Margolin W, Molineux IJ, Liu J. 2013. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. *Science* 339:576–579. <https://doi.org/10.1126/science.1231887>.
- Westbye AB, Kuchinski K, Yip CK, Beatty JT. 2016. The gene transfer agent RcGTA contains head spikes needed for binding to the *Rhodobacter capsulatus* polysaccharide cell capsule. *J Mol Biol* 428:477–491. <https://doi.org/10.1016/j.jmb.2015.12.010>.
- Neumann E, Kawasaki T, Effantin G, Estrozi LF, Chatchawankanphanich O, Yamada T, Schoehn G. 2020. 3D structure of three jumbo phage heads. *J Gen Virol* 101:1219–1226. <https://doi.org/10.1099/jgv.0.001487>.
- Cramer P. 2021. AlphaFold2 and the future of structural biology. *Nat Struct Mol Biol* 28:704–705. <https://doi.org/10.1038/s41594-021-00650-1>.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstern S, Silver D, Vinyals O, Senior AW,

- Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
17. Scheres SHW. 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* 180:519–530. <https://doi.org/10.1016/j.jsb.2012.09.006>.
 18. Dai W, Hodes A, Hui WH, Gingery M, Miller JF, Zhou ZH. 2010. Three-dimensional structure of tropism-switching *Bordetella* bacteriophage. *Proc Natl Acad Sci U S A* 107:4347–4352. <https://doi.org/10.1073/pnas.0915008107>.
 19. Jiang W, Baker ML, Jakana J, Weigele PR, King J, Chiu W. 2008. Backbone structure of the infectious ϵ 15 virus capsid revealed by electron cryomicroscopy. *Nature* 451:1130–1134. <https://doi.org/10.1038/nature06665>.
 20. Helgstrand C, Wikoff WR, Duda RL, Hendrix RW, Johnson JE, Liljas L. 2003. The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 Å resolution. *J Mol Biol* 334:885–899. <https://doi.org/10.1016/j.jmb.2003.09.035>.
 21. Zhang X, Guo H, Jin L, Czornyj E, Hodes A, Hui WH, Nieh AW, Miller JF, Zhou ZH. 2013. A new topology of the HK97-like fold revealed in *Bordetella* bacteriophage by cryoEM at 3.5 Å resolution. *Elife* 2:e01299. <https://doi.org/10.7554/eLife.01299>.
 22. Hendrix RW, Johnson JE. 2012. Bacteriophage HK97 capsid assembly and Maturation. In Rossmann M, Rao V (ed). *Viral molecular machines. Advances in experimental medicine and biology*, vol 726:351–63. Springer, Boston, MA.
 23. Dedeo CL, Teschke CM, Alexandrescu A. 2020. Keeping it together: structures, functions, and applications of viral decoration proteins. *Viruses* 12:1163. <https://doi.org/10.3390/v12101163>.
 24. McNulty R, Cardone G, Gilcrease EB, Baker TS, Casjens SR, Johnson JE. 2018. Cryo-EM elucidation of the structure of bacteriophage P22 virions after genome release. *Biophys J* 114:1295–1301. <https://doi.org/10.1016/j.bpj.2018.01.026>.
 25. Lander GC, Tang L, Casjens SR, Gilcrease EB, Prevelige P, Poliakov A, Potter CS, Carragher B, Johnson JE. 2006. The structure of an infectious P22 virion shows the signal for headful DNA packaging. *Science* 312:1791–1795. <https://doi.org/10.1126/science.1127981>.
 26. Bárdy P, Füzik T, Hrebík D, Pantůček R, Beatty JT, Plevka P. 2020. Structure and mechanism of DNA delivery of a gene transfer agent. *Nat Commun* 11. <https://doi.org/10.1038/s41467-020-16669-9>.
 27. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Židek A, Bridgland A, Cowie A, Meyer C, Laydon A, Velankar S, Kleywegt GJ, Bateman A, Evans R, Pritzel A, Figurnov M, Ronneberger O, Bates R, Kohl SAA, Potapenko A, Ballard AJ, Romera-Paredes B, Nikolov S, Jain R, Clancy E, Reiman D, Petersen S, Senior AW, Kavukcuoglu K, Birney E, Kohli P, Jumper J, Hassabis D. 2021. Highly accurate protein structure prediction for the human proteome. *Nature* 596:590–596. <https://doi.org/10.1038/s41586-021-03828-1>.
 28. Holm L. 2022. Dali server: structural unification of protein families. *Nucleic Acids Res* 50:W210–W215. <https://doi.org/10.1093/nar/gkac387>.
 29. Fülöp V, Jones DT. 1999. β propellers: structural rigidity and functional diversity. *Current Opinion in Structural Biology* 9:715–721. [https://doi.org/10.1016/S0959-440X\(99\)00035-4](https://doi.org/10.1016/S0959-440X(99)00035-4).
 30. Knecht LE, Veljkovic M, Fieseler L. 2020. Diversity and function of phage encoded depolymerases. *Front Microbiol* 10. <https://doi.org/10.3389/fmicb.2019.02949>.
 31. Yehl K, Lemire S, Yang AC, Ando H, Mimeo M, Torres MDT, de la Fuente-Nunez C, Lu TK. 2019. Engineering phage host-range and suppressing bacterial resistance through phage tail fiber mutagenesis. *Cell* 179:459–469. <https://doi.org/10.1016/j.cell.2019.09.015>.
 32. Broecker NK, Barbirz S. 2017. Not a barrier but a key: how bacteriophages exploit host's O-antigen as an essential receptor to initiate infection. *Mol Microbiol* 105:353–357. <https://doi.org/10.1111/mmi.13729>.
 33. Mouroso JT, Awe A, Guo W, Batra H, Ganesh H, Wu X, Zhu J. 2022. "Understanding bacteriophage tail fiber interaction with host surface receptor: the key 'blueprint' for reprogramming phage host range. *Int J Molecular Sciences* 23. <https://doi.org/10.3390/ijms232012146>.
 34. Flynn RL, Zou L. 2010. Oligonucleotide/oligosaccharide-binding fold proteins: a growing family of genome guardians. *Critical Rev in Biochemistry and Molecular Biology* 45:266–275. <https://doi.org/10.3109/10409238.2010.488216>.
 35. Li F, Hou CFD, Yang R, Whitehead R, Teschke CM, Cingolani G. 2022. High-resolution cryo-EM structure of the *Shigella* virus Sf6 genome delivery tail machine. *Sci Adv* 8:eadc9641. <https://doi.org/10.1126/sciadv.adc9641>.
 36. Kandiah E, Giraud T, de Maria Antolinos A, Dobias F, Effantin G, Flot D, Hons M, Schoehn G, Susini J, Svensson O, Leonard GA, Mueller-Dieckmann C. 2019. CM01: a facility for cryo-electron microscopy at the European Synchrotron. *Acta Crystallogr D Struct Biol* 75:528–535. <https://doi.org/10.1107/S2059798319006880>.
 37. Rohou A, Grigorieff N. 2015. CTFIND4: fast and accurate defocus estimation from electron micrographs. *J Struct Biol* 192:216–221. <https://doi.org/10.1016/j.jsb.2015.08.008>.
 38. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF chimera - a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612. <https://doi.org/10.1002/jcc.20084>.
 39. Altschup SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Bio* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
 40. Pintilie GD, Zhang J, Goddard TD, Chiu W, Gossard DC. 2010. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J Struct Biol* 170:427–438. <https://doi.org/10.1016/j.jsb.2010.03.007>.
 41. Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* 43:W389–W394. <https://doi.org/10.1093/nar/gkv332>.
 42. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. 2021. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci* 30:70–82. <https://doi.org/10.1002/pro.3943>.
 43. Croll TI. 2018. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* 74:519–530. <https://doi.org/10.1107/S2059798318002425>.
 44. Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132. <https://doi.org/10.1107/S0907444904019158>.
 45. Afonine PV, Poon BK, Read RJ, Sobolev OV, Terwilliger TC, Urzhumtsev A, Adams PD. 2018. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* 74:531–544. <https://doi.org/10.1107/S2059798318006551>.
 46. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, Jain S, Lewis SM, Arendall WB, Snoeyink J, Adams PD, Lovell SC, Richardson JS, Richardson DC. 2018. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci* 27:293–315. <https://doi.org/10.1002/pro.3330>.