



HAL
open science

CryoEM Structure and Assembly Mechanism of a Bacterial Virus Genome Gatekeeper

Igor Orlov, Stephane Roche, Sandrine Brasilès, Natalya Lukoyanova, Marie-Christine Vaney, Paulo Tavares, Elena V. Orlova

► **To cite this version:**

Igor Orlov, Stephane Roche, Sandrine Brasilès, Natalya Lukoyanova, Marie-Christine Vaney, et al.. CryoEM Structure and Assembly Mechanism of a Bacterial Virus Genome Gatekeeper. Nature Communications, In press, 13 (7283), 10.1038/s41467-022-34999-8 . hal-03857097v1

HAL Id: hal-03857097

<https://hal.science/hal-03857097v1>

Submitted on 17 Nov 2022 (v1), last revised 26 Sep 2023 (v2)

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.



Distributed under a Creative Commons Attribution 4.0 International License

1 **CryoEM Structure and Assembly Mechanism**
2 **of a Bacterial Virus Genome Gatekeeper**

3
4 Igor Orlov^{1,2,*}, Stéphane Roche^{3,*}, Sandrine Brasiles³, Natalya Lukoyanova⁴, Marie-
5 Christine Vaney⁵, Paulo Tavares^{1,3}, Elena V Orlova^{1,4}

6
7 Affiliations

8
9 ¹Centre for Integrative Biology (CBI), Department of Integrated Structural Biology,
10 IGBMC, Université de Strasbourg, 67404 Illkirch, France;

11 ²Current address: Laboratory of Cellular and Structural Biology, The Rockefeller
12 University, 1230 York Avenue New York, NY 10065(212) 327-8000;

13 ³Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell
14 (I2BC), 91198, Gif-sur-Yvette, France;

15 ⁴Institute of Structural and Molecular Biology, Department of Biological Sciences,
16 Birkbeck College, Malet Street, London, WC1E 7HX, UK;

17 ⁵Institut Pasteur, Université de Paris, CNRS UMR3569, Unité de Virologie
18 Structurale, 75015 Paris, France

19
20 * Equal contribution

21
22 ¶ Corresponding authors:

23
24 Elena V. Orlova, e-mail: e.orlova@bbk.ac.uk

25 Paulo Tavares, e-mail: paulo.tavares@i2bc.paris-saclay.fr

26
27

28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

Abstract

Numerous viruses package their dsDNA genome into preformed capsids through a portal gatekeeper that is subsequently closed. We report the structure of the DNA gatekeeper complex of bacteriophage SPP1 (gp6₁₂gp15₁₂gp16₆) in the post-DNA packaging state at 2.7 Å resolution obtained by single particle cryo-electron microscopy. Comparison of the native SPP1 complex with assembly-naïve structures of individual components uncovered the complex program of conformational changes leading to its assembly. After DNA packaging, gp15 binds via its C-terminus to the gp6 oligomer positioning gp15 subunits for oligomerization. Gp15 refolds its inner loops creating an intersubunit β-barrel that establishes different types of contacts with six gp16 subunits. Gp16 binding and oligomerization is accompanied by folding of helices that close the portal channel to keep the viral genome inside the capsid. This mechanism of assembly has broad functional and evolutionary implications for viruses of the prokaryotic tailed viruses-herpesviruses lineage.

Key words:

bacteriophage, SPP1, virus structure, cryo-electron microscopy, single particle analysis, portal protein, protein-protein interaction, macromolecular assembly mechanism

52 Introduction

53

54 Many double-stranded DNA (dsDNA) viruses have a specialized portal complex at
55 one 5-fold vertex of their icosahedral capsid through which dsDNA enters and exits
56 the virion. Tailed prokaryotic viruses and herpes viruses form a viral capsid
57 precursor, termed procapsid, prior to DNA packaging. The procapsid contains a built-
58 in portal protein complex¹⁻⁵ which serves as a docking platform for the terminase.
59 The complex of the portal and terminase constitutes the viral genome packaging
60 motor that translocates dsDNA into the procapsid. Tightly packed DNA (>400
61 mg/mL) exerts a pressure on the capsid that can reach ~60 atm^{6,7}. Dissociation of
62 the terminase from the motor is followed by closure of the portal vertex to prevent
63 leakage of the packed viral genome³. Sealing of the portal channel is achieved by
64 binding head completion proteins to assemble a complex named connector in tailed
65 prokaryotic viruses, which consists of stacked cyclical oligomers creating a channel
66 for DNA traffic^{3,8-10}. Opening of the closed connector is necessary for DNA entry in
67 the tail tube and its subsequent delivery to the host cell at the beginning of infection.

68 CryoEM and X-ray structures were reported for a number of portal proteins^{5,8,11-14},
69 for the complex of P22 portal with the adaptor protein¹⁵, and for some isolated
70 proteins that participate in this process^{3,16-19}. CryoEM studies provided more
71 structural information on the portal complex assembled after DNA packaging¹⁰ and
72 when the tail is attached to the portal vertex^{8,20-23}. Nonetheless, the molecular
73 mechanism behind assembly of the DNA gatekeeper at the post-viral genome
74 packaging state is still unknown.

75 In bacteriophage SPP1, the gp15 and gp16 head completion proteins bind
76 sequentially to the dodecameric portal protein gp6 after DNA packaging^{9,10,19}. The
77 gp15 adaptor protein extends the portal tunnel that is fastened afterwards by binding
78 the gp16 stopper protein. Here we report a cryoEM structure of the 902 kDa
79 bacteriophage SPP1 connector in the post-DNA packaging state at a resolution of
80 2.7 Å. The *de novo* tracing of its 30 polypeptide chains within the cryoEM map
81 allowed us to determine an atomic model of the entire connector complex
82 (gp6₁₂gp15₁₂gp16₆). Structural comparison of the connector protein components with
83 their assembly-naïve states in solution infers a sequential path of conformational
84 changes and different types of interactions engaged during the assembly of the viral

85 DNA gatekeeper. This structural study reveals how symmetry transition takes place
86 within the connector to match tail symmetry.

87 88 **Results**

89
90 **Connector overall structure.** The 902 kDa SPP1 connector complex was purified
91 from disrupted tailless DNA-filled capsids⁹. The connector structure was determined
92 by single particle analysis cryoEM at a resolution of 2.7 Å (at 0.5 threshold of FSC)
93 (Fig 1, see Online Methods and Supplementary Table 1). This resolution allowed
94 unambiguous *de novo* tracing of the polypeptide chains of gp6, gp15 and gp16 within
95 the cryoEM map of the complex (Extended Data Figs. 1 and 2). We have found that
96 the connector comprises 12 subunits of both gp6 and gp15, but only 6 copies of
97 gp16 (Fig. 1) in contrast to previous reports^{10,19,20}. Gp15 and gp16 are monomeric
98 before they assemble in the connector²⁴, showing that connector assembling
99 involves gp6-gp15, gp15-gp15, gp15-gp16, and gp16-gp16 interaction events.

100 **The gp6 portal protein fold at the post-DNA packaging state.** The polypeptide
101 chain of gp6 was traced in the cryoEM density map from residue 28 to 470 out of
102 503 residues (Fig. 2a). The overall gp6 fold in the connector is similar to the fold of
103 gp6 in the 13-mer assembly-naïve form (PDB 2JES)¹². Gp6 encompasses four major
104 domains: the crown, wing, stem, and clip (Fig. 2a,b). It also features a tunnel loop
105 that protrudes towards the centre of the portal central tunnel and a β-hairpin at the
106 outer surface of the stem domain (Fig. 2a,b).

107 The RMSD is small between the Cα positions in the wing domain (residues 29-54,
108 88-169, 178-208 and 369-420) (RMSD 0.7 Å) and in the portal stem (residues 55-62,
109 81-87, 256-280, 327-346) (RMSD 0.5 Å) (Fig. 2b). Noticeable structural differences
110 were found within the crown area, tunnel loops, β-hairpins, and clip domains (Fig.
111 2b). The tunnel loop forms a short helix in the connector complex that is absent in
112 the gp6 13-mer¹² (Fig. 2b). The low density of this area in the connector suggests
113 that this segment is flexible and can adopt other conformations. The most significant
114 difference between the two structures is observed in the gp6 clip domain that makes
115 the most extensive intersubunit bonding in the portal oligomer (Fig. 2c,d). The clip is
116 formed by an intersubunit triple-stranded β-sheet, composed of strands β12 and β14
117 from subunit *i* and β13 from subunit *i-1*, and helix α4 from subunit *i* (Fig. 2c,d). These

118 structural elements are found in both gp6 structures (12-mer and 13-mer) but exhibit
119 visible shifts between residues Gln282 at the end of helix α 3 and Pro325 at the
120 beginning of helix α 5 (Fig. 2b-d). Helix α 4 has a tilted orientation in the gp6 13-mer¹²
121 with its N-terminus Pro296 residue localized at the clip bottom (Fig. 2c; Extended
122 Data Fig. 3a). In the connector, Pro296 is shifted upwards by 7 Å bringing helix α 4 to
123 a more horizontal orientation closer to the gp6 β -hairpin of adjacent subunit $i+2$ (Fig.
124 2d). Changes in the clip domains of the connector gp6 oligomer lead to formation of
125 pockets where the C-termini of gp15 bind to the portal oligomer (Extended Data Fig.
126 3b,c). These pockets are absent in the assembly-naïve gp6 13-mer (Extended Data
127 Fig. 3a) explaining why it does not bind gp15²⁴.

128 **Mapping of vital DNA packaging mutations in the portal structure.** The structure
129 of gp6 in the post-DNA packaging state allows to rationalize the effect of mutations
130 previously shown to impair the SPP1 DNA packaging process^{25,26}. Since these
131 mutated gp6 proteins are incorporated in procapsids, this implies that they are folded
132 correctly and interact with the other procapsid components ensuring proper
133 procapsid assembly²⁵. We discerned several groups of residues critical for DNA
134 packaging (Fig. 2e).

135 A set of mutations that block DNA packaging was found in the gp6 crown. This
136 region forms the upper part of the tunnel channel through which DNA is translocated
137 into the capsid interior. Mutations affecting the crown structural organization involve
138 residues engaged in inter-subunit bonding (Ser428, Ile437, Phe449) and/or
139 destabilizing the crown hydrophobic core (Ile437Val, Ala443Thr, Phe449Leu) formed
140 by helices α 7 and α 8 (Fig. 2e, in green; Supplementary Table 2).

141 Another cluster of mutated residues affect either the gp6 tunnel loop structure
142 (Ser350, Glu352), or the loop interaction with α 5 (Val347, Lys373) and the portal
143 crown (Gly360). They highlight a role of the loops in the DNA packaging process
144 (Fig. 2e, in red; Supplementary Table 2). We previously showed that substitution
145 Glu352Gly leads to reduction of the terminase ATPase activity²⁷. This cross-talk
146 between tunnel loops and the terminase during DNA translocation is mediated,
147 conceivably, through helix α 5 and the gp6 clip region²⁸.

148 DNA packaging is also disrupted by mutation of Pro325 (Fig. 2e, in purple) located at
149 a hinge position between helix α 5 and the clip. The polypeptide chains of the gp6 13-

150 mer and of the gp6 connector 12-mer change their conformation at this point leading
151 to a different position of the clip in the two structures (Fig. 2b-e). Substitution
152 Pro325Leu disrupts inter-subunit bonding and might extend the length of helix $\alpha 5$,
153 affecting the correct positioning of the clip elements.

154 The next important set of mutations is found in loops anchoring the gp6 β -hairpin at
155 proximity to the wing and stem (Fig. 2b, in blue). We hypothesize that the mutations
156 in this hairpin change its position with respect of the clip domain (Fig. 2d, $\beta 1$ - $\beta 2$ in
157 blue), preventing formation of the clip pocket. The essential role of the clip pocket in
158 DNA packaging is revealed by the highly detrimental effect of amino acid
159 substitutions in Asn290, Gly293 and Glu294²⁵ of loop $\beta 12$ - $\alpha 4$ (Fig. 2e, in cyan;
160 Supplementary Table 2). These three residues are exposed to the portal exterior
161 (Fig. 2e). Asn290 is located at the tip of loop $\beta 12$ - $\alpha 4$, with its side chain engaged in
162 intrasubunit and intersubunit hydrogen bonds with the carbonyl of Asp314 and with
163 the nitrogen of Gly313, respectively (Extended Data Fig. 4). These interactions
164 mimic beta interactions and are essential for the correct positioning of loops $\beta 12$ - $\alpha 4$
165 and $\beta 13$ - $\beta 14$ within the clip domain. Substitution of Asn290 disrupts one or both
166 hydrogen bonds destabilizing these loops. Gly293 and Glu294 make no side chain
167 contacts suggesting that the phenotype of their substitution results from a defect in
168 the direct interaction with the terminase²⁹. Disruption of this interaction upon
169 disassembly of the packaging motor renders the gp6 clip pocket ready for gp15
170 binding. Sequential interaction of the terminase and gp15 with the same gp6 region
171 of the clip is further corroborated by finding that mutation Glu294Lys disrupts SPP1
172 DNA packaging while substitution of the same residue Glu294Gly impairs only the
173 subsequent assembly step of binding to gp15²⁶ (Supplementary Table 2; see below).
174 Collectively this structure-function analysis highlights structural elements lining the
175 portal channel and the clip as essential factors for phage DNA packaging.

176 ***The gp15 structure and gp6-gp15 interface.*** Termination of DNA packaging is
177 followed by departure of the terminase and attachment of gp15 to gp6, initiating
178 connector assembly. The core of gp15 comprises 4 helices (Fig. 1c, right panel). Its
179 C-terminus is directed towards gp6 while the hairpin $\beta 1$ - $\beta 2$ points towards gp16 (Fig.
180 1b). Gp15 interacts with the gp6 oligomer mostly via insertion of its C-terminus into
181 the gp6 clip pocket (Fig. 3a,b; Extended Data Fig. 3c). The gp15 C-terminus
182 establishes ten interchain bonds with the clip of gp6 chain *i* (Supplementary Table 3).

183 Gp15 His37 makes also a bond with clip residue Asp292 of gp6 chain $i+1$ (Fig. 3b)
184 while gp15 Arg102 establishes a hydrogen bond with the carbonyl of Asp68 from the
185 β -hairpin of gp6 chain $i+2$ (Fig. 3b; Supplementary Table 3). This interaction network
186 explains why mutation Glu294Gly in gp6, which disrupts its hydrogen bond with gp15
187 Met100 (Fig. 3b), or replacement of the five C-terminal residues (ArgLysMetAlaArg)
188 in gp15 with MetAlaGly, prevent gp15 stable binding to gp6^{20,26}. These interactions
189 are characterized by electrostatic potential complementarity between the gp15 C-
190 terminus and the gp6 pocket and their shape matching (Extended Data Fig. 5a).

191 **Gp15 conformational changes.** Assembly-naïve gp15 (PDB 2KBZ) is a monomer
192 in solution composed of an α -helical core, flexible loop $\alpha1$ - $\alpha2$, a large unstructured
193 loop between helices $\alpha2$ and $\alpha3$, and the flexible C terminus, as shown by NMR
194 structures¹⁹. The α -helical core $\alpha1$ - $\alpha3$ maintains its conformation after assembly of
195 gp15 in the connector but $\alpha0$, loop $\alpha2$ - $\alpha3$, and the C-terminus undergo major
196 conformational changes (Fig. 3c, in green, blue and orange, respectively). The gp15
197 C-terminus interacts with gp6 (Fig. 3a,b; Extended Data Figure 3c). The N-terminus
198 and $\alpha0$ undergo 145° rotation to establish contacts with helices $\alpha1$ and $\alpha2$ (Fig. 3c).
199 Helix $\alpha0$ makes also links with the adjacent subunit through residues Arg5 and Arg8
200 forming salt bridges with Glu23 of $\alpha1$ from subunit $j-1$ (Fig. 3d). The most dramatic
201 conformational change takes place by folding of loop $\alpha2$ - $\alpha3$ into the $\beta1$ - $\beta2$ hairpin
202 that forms the gp15 24-stranded intermolecular β -barrel (Figs. 1b,c and 3c).

203 **Gp15 intersubunit interactions within the connector.** The surface of interaction
204 between a single gp15 subunit and gp6 in the connector is only 713 Å² and the
205 calculated dissociation energy is -1.1 kcal/mol as assessed by EBI-PISA³⁰
206 (Supplementary Table 4). This indicates that the single gp15 monomer does not
207 bind strongly to the gp6 ring³¹. In contrast, adjacent gp15 subunits have a large
208 interaction surface of 1220 Å² with 22 intersubunit hydrogen bonds (Supplementary
209 Table 3), accounting for a dissociation energy of -12.4 kcal/mol (Supplementary
210 Table 4). The initial attachment of gp15 monomers to gp6 is thus followed by gp15
211 intersubunit interactions that ensure stable association of gp15 to the gp6 oligomer.

212 The gp15-gp15 interface comprises lateral contacts between helix $\alpha1$ of subunit j and
213 helices $\alpha0$, $\alpha2$ and $\alpha3$ of the adjacent subunit $j-1$ (Supplementary Table 3).
214 Furthermore, helix $\alpha0$ binds helix $\alpha1$ of adjacent subunits at the connector periphery

215 through a network of salt bridges and hydrogen bonds (Fig. 3d). The gp15 β 1- β 2
216 hairpins of neighbour subunits form an intersubunit 24-stranded β -barrel distal from
217 the portal (Figs. 1b,c) that outlines the 31 Å-wide central tunnel of gp15 (Fig. 1c).
218 Within each gp15 subunit the β -hairpin is connected to its α -helical core via loops α 2-
219 β 1 and β 2- α 3 (Figs. 3c,e). The carboxylate group of Glu85 in loop β 2- α 3 of subunit j
220 establishes five hydrogen bonds with the main chain and side chains of Ser88 and
221 Thr89 of subunit $j+1$ (Fig. 3e), stabilising the overall positioning of the β -barrel
222 strands.

223 **Gp16 structure and the gp15-gp16 interface.** The bottom part of the connector is
224 formed by six gp16 subunits. Each of them comprises a β -barrel core (β 3 to β 6), the
225 tunnel α -helix α 1 and the β 1- β 2 loop that protrudes below the connector and
226 presumably interacts with the tail (Fig. 1d).

227 The interface between the gp15 dodecamer and the gp16 hexamer is characterized
228 as well by electrostatic complementarity (Extended Data Fig. 5b). Most of the gp15-
229 gp16 interactions involve the bottom rim of the gp15 24-strand β -barrel (Figs. 1b and
230 3f,g). Each gp16 subunit makes contacts with four gp15 subunits (Fig. 3f;
231 Supplementary Table 3). Remarkably, residues Thr77 of three different gp15
232 subunits interact with the same gp16 subunit (Fig. 3f). Thr77 of gp15 chain $j+1$
233 makes a hydrogen bond with gp16 Ser91 while Thr77 of gp15 chain j establishes
234 hydrogen bonds with residues Tyr61 and Arg98 of gp16. Finally, Thr77 of gp15 of
235 chain $j-1$ is involved in two hydrogen bonds with gp16 Gln39. Additionally, Arg73 of
236 gp15 chain j makes a salt bridge with Glu95 of gp16 (Fig. 3f). Disruption of this
237 interaction by mutation Arg73Glu in gp15 was previously shown to prevent stable
238 binding of gp16 to gp15²⁰. A second type of interactions between gp16 and gp15
239 takes place by insertion of the gp16 N-terminus in a crease between the gp15 β -
240 barrel and loops α 2- β 1 of gp15 _{j} and α 0- α 1 of gp15 _{$j+1$} (Fig. 3g). This network of
241 interactions establishes the transition of 12 to 6-fold symmetry within the connector.
242 Overall, the interaction between each subunit of gp16 and the gp15 oligomer has a
243 surface of 910 Å² and dissociation energy of -13.6 kcal/mol indicating on strong
244 bonding between gp15 and gp16 oligomer

245 **Gp16 conformational changes and portal tunnel closure.** The gp16 subunit in the
246 connector has a well-defined six-stranded β -barrel core (Fig. 1d) like the assembly-

247 naïve gp16 (PDB 2KCA) monomer in solution. The most dramatic change within the
248 gp16 structure takes place in flexible loop $\beta 2'$ - $\beta 3$ (residues 43-51) that folds into the
249 tunnel α -helix (Fig. 4a). The α -helices of six gp16 subunits in the connector are hold
250 together by 5 intersubunit hydrogen bonds between adjacent subunits. Gln43 makes
251 four of them and the fifth is between Tyr47-Glu45 (Fig. 4b). The tunnel helices form a
252 constriction in the portal channel with a diameter of ~ 12 Å: between side chains of
253 opposite gp16 Gln51 residues (Fig. 1d). This is the only region of the connector
254 tunnel narrower than the 20 Å diameter of a DNA double-helix, locking DNA inside
255 the capsid after its encapsidation. The interaction between gp16 subunits is notably
256 weak with a surface of 513 Å² and a calculated dissociation energy of -1.6 kcal/mol
257 (Supplementary Table 4), suggesting that the gp16 hexamer is held primarily by
258 bonds with gp15 subunits. The N-terminus residues 12-30, forming loop $\beta 1$ - $\beta 2$,
259 protrude below the connector and presumably interact with the tail (Fig. 1d). Tail
260 binding to gp16 might further stabilise its structure in the infectious tailed phage²⁰.

261 **Discussion and conclusions**

262 The 2.7 Å resolution structure of the SPP1 gp6₁₂gp15₁₂gp16₆ complex reported here
263 uncovers the detailed molecular architecture of the complete viral DNA gatekeepers
264 from tailed prokaryotic viruses. Comparison with the structure of its non-assembled
265 components reveals the molecular events leading to assembly of the gatekeeper
266 (Fig. 5) that have extensive functional and evolutionary implications.

267 The connector structure shows the conformation of the portal protein gp6 in its post-
268 DNA packaging state. Locations of mutations that arrest DNA packaging before the
269 endonucleolytic cleavage that terminates the process^{25,26} identify clusters of residues
270 in several structural elements of gp6 (Fig. 2e). Two clusters were found in the crown
271 and in loops that line the portal tunnel. Other mutations affect structural elements
272 that build the clip pocket exposed for binding of the terminase (Fig. 2e and Extended
273 Data Fig. 3b). The previous finding that motion of stem helix $\alpha 5$ is necessary for DNA
274 packaging²⁸ and localisation of mutations in the gp6 structure suggest that the tunnel
275 loops act jointly with the terminase during DNA packaging^{27,29}. Their cross-talk
276 through helix $\alpha 5$ might support DNA translocation, driven primarily by the terminase
277 pumping activity²⁹, and a ratchet mechanism in which tunnel loops act to prevent
278 sliding of DNA out of the phage capsid as it reaches high concentration inside the

279 capsid. The previously proposed ratchet function of the tunnel loops^{8,32,33}, is
280 conceivably coordinated with the DNA packaging motor activity. This retention
281 mechanism operating during genome packaging is transient. Keeping the DNA
282 inside the capsid requires binding of gp15 and gp16 to the portal system upon
283 terminase departure¹⁰. The process is coordinated by sequential interaction of the
284 terminase and gp15 with a common binding interface within a pocket in the gp6 clip
285 (Figs 2e and 3a,b).

286 Comparison of the NMR structures of SPP1 assembly-naïve gp15 and gp16¹⁹ with
287 their conformations within the connector highlights the structural rearrangements
288 occurring during assembly of the DNA gatekeeper (Figs. 3c, 4a and 5). The analysis
289 of gp15 and gp16 structural homologues (Extended Data Figs. 6-8) expands further
290 our understanding of their conformational landscape. Gp15-like proteins have a
291 conserved α -helical core but differ in the conformations of loop $\alpha 0$ - $\alpha 1$, loop $\alpha 2$ - $\alpha 3$
292 and/or of the C-terminus. These variations inform on structural changes on the
293 pathway from the monomer to the dodecameric state (Extended Data Fig. 6). The C-
294 terminus of SPP1 gp15 has a defined conformation only if it is bound to the portal
295 protein. That corroborates with their role in anchoring the gp15-like adaptor proteins
296 to the portal clip (Fig. 5a). During connector assembly, gp15 helix $\alpha 0$ changes its
297 position to a more horizontal orientation (Figs. 3c and 5a,b). Interestingly, in the non-
298 assembled monomer of the gp15 homologue YqbG helix $\alpha 0$ adopts a position,
299 stabilized by intrachain bonds, that is close to the orientation found in the gp15
300 connector state (Extended Data Fig. 6b). In SPP1 such conformation requires
301 intersubunit bonding of $\alpha 0$ that bridges helices $\alpha 1$ of neighbour subunits (Fig. 3d).
302 Disruption of these interactions in mutant gp15 Arg5Glu Arg8Glu (Fig. 3d) allows for
303 formation of the gp15-g16 complex in phage capsids but prevents closure of the
304 connector tunnel²⁰. Re-positioning of $\alpha 0$ is thus an essential assembly step (step **a** to
305 **b** in Fig. 5; Movie 1).

306 The NMR structures of gp15 and YqbG monomers^{19,34} both show highly mobile
307 loops $\alpha 2$ - $\alpha 3$ (Extended Data Fig. 6b). Its α -helical core has intersubunit interactions
308 similar to those between SPP1 gp15 subunits within the connector. The homologue
309 gp6 protein from siphophage HK97 forms a 13-mer ring in assembly-naïve
310 conditions¹⁸ and the loop $\alpha 2$ - $\alpha 3$ of HK97gp6¹⁸ has an antiparallel β -like interchain
311 interaction (Extended Data Figs. 6b and 7a). This loop is located in a position similar

312 to loop $\alpha 2$ - $\alpha 3$ from SPP1 gp15 that folds to build a β -sheet within the connector
313 (Extended Data Figs. 6b and 7b). Although gp6 of HK97 forms most likely a 12-mer
314 when it assembles in the phage particle¹⁸, the 13-mer structure suggests that its
315 oligomerization leads to formation of β -like interactions between the $\alpha 2$ - $\alpha 3$ loops of
316 adjacent subunits. We hypothesize that they represent a precursor state of an inter-
317 subunit β -barrel in HK97, as was found in SPP1 gp15. Rigid-body fitting of gp15
318 NMR monomers into the SPP1 connector exposes multiple clashes between loops
319 $\alpha 2$ - $\alpha 3$ of adjacent subunits, dictating conformational changes in their fold. These
320 changes result conceivably in formation of the β -barrel during gp15 oligomerization
321 (**a** to **c** in Fig. 5; Movie 1). Collectively, analysis of SPP1 gp15 homologous protein
322 structures reveal different conformations that are necessary for achieving the
323 functional gp15 assembled state.

324 The gp15 β -barrel creates a platform for gp16 attachment to the SPP1 portal
325 complex (Fig. 5d). Connector stoppers, typified by gp16, are characterised by a
326 common β -barrel core³ (Extended Data Fig. 8). The monomers of phage λ gpFII¹⁷
327 and XkdH have a flexible loop equivalent to the SPP1 gp16 monomer $\beta 2'$ - $\beta 3$ loop.
328 This loop is folded into a α -helix in the monomer of the putative stopper SF1141
329 (Extended Data Fig. 8b). Loop $\beta 2'$ - $\beta 3$ of SPP1 gp16 folds also into a α -helix after
330 binding of gp16 subunits to the gp15 ring in SPP1 (step **f** to **h** in Fig. 5; Movie 1).
331 Five inter-helices contacts stabilize the helical conformation occluding the portal
332 channel (Fig. 4b) while not imposing a strong barrier for disruption of this bonding
333 that is necessary for subsequent channel opening. The weak gp16 intersubunit
334 bonding (Supplementary Table 4) indicates that its interaction with gp15 plays a
335 major role for endorsed positioning of the six gp16 subunits in the connector. This
336 corroborates with mutations in gp15 that impair specifically portal tunnel closure by
337 gp16²⁰ (see above). The 12 to 6-fold symmetry transition within the gp15-gp16
338 complex enables a direct match of the connector with the hexameric structure of the
339 phage pre-assembled long tail.

340 The portal vertex complex represents an essential DNA traffic gate conserved
341 among all viruses of the prokaryotic tailed viruses-herpesviruses lineage³⁵. The
342 portal protein component is conserved in the lineage. In contrast, the effectors acting
343 after DNA packaging diverged as viruses evolved to infect different cellular hosts
344 (Fig. 6). Eukaryotic herpesviruses deviated first by extension of the portal protein clip

345 with “tentacle” α -helices protruding towards a cap that closes the portal system^{3,4}
346 (Fig. 6). Clip domains, like in SPP1 gp6, are one of the most conserved structural
347 elements in the portal protein from tailed bacteriophages^{3,14}. These viruses have also
348 gp15-like proteins that form dodecamers extending the portal DNA tunnel. However,
349 the β -barrel region of SPP1 gp15 has changed significantly in the homologous
350 proteins from phages with a short tail (Fig. 6). Phages P68 and ϕ 29 have a β -barrel
351 longer than $>100 \text{ \AA}$ forming the short tail DNA conduit^{23,37} (Extended Data Fig.7c,d)
352 while T7⁸ and possibly P22¹⁵ have no equivalent to the SPP1 gp15 β -barrel (Fig. 6).
353 This evolutionary divergence is further marked by the absence of gp16 homologous
354 proteins in phages with short tails, with the possible exception of a domain from the
355 phage ϕ 29 knob protein (Fig. 6 and Extended Data Fig. 8). In contrast, gp16-like
356 proteins are conserved among phages with long tails³⁸. The molecular architecture of
357 the gp15-gp16 interface accomplishes the connector-to-tail symmetry transition and
358 matches the connector tunnel-forming gp15 structure with the tail tube-like fold of
359 gp16³⁹. This β -barrel fold is the common feature of some components of long tail
360 tubes^{39,40}. The gp16-like proteins evolved to establish three main functions: firstly the
361 interaction with gp15-like proteins, secondly to build a temporary retention system of
362 the viral DNA, and thirdly to bind to the tail-to-head joining protein at the tail tube
363 end^{3,20,21,41,42}.

364 The architecture of the connector and its assembly mechanism reveal the
365 evolutionary process of the tailed prokaryotic viruses-herpesviruses lineage (Fig. 6).
366 Lineage branching begins with one of their common building blocks, a portal protein,
367 to assemble viral machines that infect eukaryotic cells or use a tail device to attack
368 prokaryotes. In tailed prokaryotic viruses there are gp15-like proteins that bind to the
369 portal protein to form short tails or connect the gp16-like proteins. The latter
370 interaction creates an interface for attaching long tail tubes assembled in an
371 independent pathway. Our connector structure and comparative structural biology
372 study proposes a molecular mechanism at work at different stages of phage
373 connector assembly in phages with long tails and reveals its divergence breakpoints
374 during viral lineage evolution.

375 **Online content**

376 Online Methods, Nature Research reporting summaries, Extended Data Figs. 1 to 8
377 and Supplementary Tables 1 to 4, Movie 1, acknowledgements, peer review
378 information.

379 **Author contributions** E.V.O. and P.T. designed and supervised the research. S.B.
380 purified SPP1 DNA-filled capsids and connectors. N.L. prepared the EM grids, I.O.,
381 E.V.O. analysed the EM data and did structural analysis. I.O., E.V.O. S.R. and M-
382 C.V. analysed the map and performed modelling. P.T., E.V.O., and S.R. wrote the
383 manuscript; I.O., E.V.O. S.R. and M.-C.V. have prepared the figures. All authors
384 participated in the discussions of the results, figures, and critical reading of the
385 manuscript. All authors approved the final version of the manuscript.

386 **Data Availability Statement:** The connector map and atomic models of gp6, gp15,
387 and gp16 are deposited to the EMD database under accession codes EMD-14509
388 and PDB entry 7Z4W.

389 **Acknowledgments:** We thank Dr D. Houldershaw and Y. Goudetsidis for computer
390 support at Birkbeck throughout the duration of the project. Cryo-EM samples were
391 prepared and optimised at the ISMB EM facility at Birkbeck College, with financial
392 support from Wellcome Trust (079605/2/06/2). We acknowledge Diamond Light
393 Source for access and support of the cryo-EM facilities at the UK National electron
394 Bio-Imaging Centre (eBIC, under proposal EM14704) funded by the Wellcome Trust,
395 the Medical Research Council UK and the Biotechnology and Biological Sciences
396 Research Council. We grateful to Dr Y, Chaban for his assistance with the data
397 collection at the eBIC. We thank Dr A. Isidro (VMS, CNRS, Gif-sur-Yvette, France)
398 for a sample of HSV-1 C-capsids and Dr R. Lurz (MPIMG, Berlin, Germany) for
399 kindly providing the micrographs shown in Fig. 6. Part of this work was supported by
400 institutional funding from CNRS.

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

402 **References**

- 403 1. Bazinet, C. & King, J. The DNA Translocating Vertex of DSDNA
404 Bacteriophage. *Annu. Rev. Microbiol.* **39**, 109–129 (1985).
- 405 2. Newcomb, W. W. *et al.* The UL6 Gene Product Forms the Portal for Entry of
406 DNA into the Herpes Simplex Virus Capsid. *J. Virol.* **75**, 10923–10932 (2001).

- 407 3. Tavares, P. The Bacteriophage Head-to-Tail Interface. in *Subcellular*
408 *Biochemistry* (eds. Harris, J. R. & Bhella, D.) vol. 88 305–328 (Springer, 2018).
- 409 4. Liu, Y.-T., Jih, J., Dai, X., Bi, G.-Q. & Zhou, Z. H. Cryo-EM structures of herpes
410 simplex virus type 1 portal vertex and packaged genome. *Nature* **570**, 257–
411 261 (2019).
- 412 5. Dedeo, C. L., Cingolani, G. & Teschke, C. M. Portal Protein: The Orchestrator
413 of Capsid Assembly for the dsDNA Tailed Bacteriophages and Herpesviruses.
414 *Annu. Rev. Virol.* **6**, 141–160 (2019).
- 415 6. Smith, D. E. *et al.* The bacteriophage ϕ 29 portal motor can package DNA
416 against a large internal force. *Nature* **413**, 748–752 (2001).
- 417 7. Rao, V. B. & Feiss, M. Mechanisms of DNA Packaging by Large Double-
418 Stranded DNA Viruses. *Annu. Rev. Virol.* **2**, 351–378 (2015).
- 419 8. Cuervo, A. *et al.* Structures of T7 bacteriophage portal and tail suggest a viral
420 DNA retention and ejection mechanism. *Nat. Commun.* **10**, 3746 (2019).
- 421 9. Lurz, R. *et al.* Structural organisation of the head-to-tail Interface of a bacterial
422 virus. *J. Mol. Biol.* **310**, 1027–1037 (2001).
- 423 10. Orlova, E. V. *et al.* Structure of a viral DNA gatekeeper at 10 Å resolution by
424 cryo-electron microscopy. *EMBO J.* **22**, (2003).
- 425 11. Simpson, A. A. *et al.* Structure of the bacteriophage ϕ 29 DNA packaging
426 motor. *Nature* **408**, 745–750 (2000).
- 427 12. Lebedev, A. A. *et al.* Structural framework for DNA translocation via the viral
428 portal protein. *EMBO J.* **26**, (2007).
- 429 13. Sun, L. *et al.* Cryo-EM structure of the bacteriophage T4 portal protein
430 assembly at near-atomic resolution. *Nat. Commun.* **6**, 7548 (2015).
- 431 14. Javed, A. *et al.* Cryo-EM Structures of Two Bacteriophage Portal Proteins
432 Provide Insights for Antimicrobial Phage Engineering. *Viruses* **13**, 2532 (2021).
- 433 15. Olia, A. S., Prevelige, P. E., Johnson, J. E. & Cingolani, G. Three-dimensional
434 structure of a viral genome-delivery portal vertex. *Nat. Struct. Mol. Biol.* **18**,
435 597–603 (2011).
- 436 16. Maxwell, K. L. *et al.* The solution structure of bacteriophage λ protein W, a
437 small morphogenetic protein possessing a novel fold¹¹Edited by P. E. Wright.
438 *J. Mol. Biol.* **308**, 9–14 (2001).
- 439 17. Maxwell, K. L., Yee, A. A., Arrowsmith, C. H., Gold, M. & Davidson, A. R. The
440 Solution Structure of the Bacteriophage λ Head–Tail Joining Protein, gpFII. *J.*
441 *Mol. Biol.* **318**, 1395–1404 (2002).
- 442 18. Cardarelli, L. *et al.* The Crystal Structure of Bacteriophage HK97 gp6: Defining
443 a Large Family of Head–Tail Connector Proteins. *J. Mol. Biol.* **395**, 754–768
444 (2010).
- 445 19. Lhuillier, S. *et al.* Structure of bacteriophage SPP1 head-to-tail connection
446 reveals mechanism for viral DNA gating. *Proc. Natl. Acad. Sci. U. S. A.* **106**,
447 8507–8512 (2009).

- 448 20. Chaban, Y. *et al.* Structural rearrangements in the phage head-to-tail interface
449 during assembly and infection. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7009–7014
450 (2015).
- 451 21. Bárdy, P. *et al.* Structure and mechanism of DNA delivery of a gene transfer
452 agent. *Nat. Commun.* **11**, 3034 (2020).
- 453 22. Leiman, P. G., Chipman, P. R., Kostyuchenko, V. A., Mesyanzhinov, V. V. &
454 Rossmann, M. G. Three-Dimensional Rearrangement of Proteins in the Tail of
455 Bacteriophage T4 on Infection of Its Host. *Cell* **118**, 419–429 (2004).
- 456 23. Xu, J., Wang, D., Gui, M. & Xiang, Y. Structural assembly of the tailed
457 bacteriophage ϕ 29. *Nat. Commun.* **10**, 2366 (2019).
- 458 24. Lhuillier, S. Etude biochimique et structurale de l'éjection de l'ADN de SPP1.
459 (Université de Paris-Sud, France, 2006).
- 460 25. Isidro, A., Santos, M. A., Henriques, A. O. & Tavares, P. The high-resolution
461 functional map of bacteriophage SPP1 portal protein. *Mol. Microbiol.* **51**, 949–
462 962 (2004).
- 463 26. Isidro, A., Henriques, A. O. & Tavares, P. The portal protein plays essential
464 roles at different steps of the SPP1 DNA packaging process. *Virology* **322**,
465 253–263 (2004).
- 466 27. Oliveira, L., Henriques, A. O. & Tavares, P. Modulation of the viral ATPase
467 activity by the portal protein correlates with DNA packaging efficiency. *J. Biol.*
468 *Chem.* **281**, 21914–21923 (2006).
- 469 28. Cuervo, A., Vaney, M.-C., Antson, A. A., Tavares, P. & Oliveira, L. Structural
470 rearrangements between portal protein subunits are essential for viral DNA
471 translocation. *J. Biol. Chem.* **282**, (2007).
- 472 29. Oliveira, L., Cuervo, A. & Tavares, P. Direct Interaction of the Bacteriophage
473 SPP1 Packaging ATPase with the Portal Protein. *J. Biol. Chem.* **285**, 7366–
474 7373 (2010).
- 475 30. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from
476 Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).
- 477 31. Chen, J., Sawyer, N. & Regan, L. Protein-protein interactions: General trends
478 in the relationship between binding affinity and interfacial buried surface area.
479 *Protein Sci.* **22**, 510–515 (2013).
- 480 32. Grimes, S., Ma, S., Gao, J., Atz, R. & Jardine, P. J. Role of ϕ 29 Connector
481 Channel Loops in Late-Stage DNA Packaging. *J. Mol. Biol.* **410**, 50–59 (2011).
- 482 33. Bayfield, O. W., Steven, A. C. & Antson, A. A. Cryo-EM structure in situ
483 reveals a molecular switch that safeguards virus against genome loss. *Elife* **9**,
484 (2020).
- 485 34. Liu, G. *et al.* NMR structure of protein yqbG from *Bacillus subtilis* reveals a
486 novel α -helical protein fold. *Proteins Struct. Funct. Bioinforma.* **62**, 288–291
487 (2005).
- 488 35. Abrescia, N. G. A., Bamford, D. H., Grimes, J. M. & Stuart, D. I. Structure

- 489 Unifies the Viral Universe. *Annu. Rev. Biochem.* **81**, 795–822 (2012).
- 490 36. Li, Z., Pang, J., Dong, L. & Yu, X. Structural basis for genome packaging,
491 retention, and ejection in human cytomegalovirus. *Nat. Commun.* **12**, 4538
492 (2021)..
- 493 37. Hrebík, D. *et al.* Structure and genome ejection mechanism of *Staphylococcus*
494 *aureus* phage P68. *Sci. Adv.* **5**, (2019).
- 495 38. Lopes, A., Tavares, P., Petit, M.-A., Guérois, R. & Zinn-Justin, S. Automated
496 classification of tailed bacteriophages according to their neck organization.
497 *BMC Genomics* **15**, 1027 (2014).
- 498 39. Cardarelli, L. *et al.* Phages have adapted the same protein fold to fulfill multiple
499 functions in virion assembly. *Proc. Natl. Acad. Sci.* **107**, 14384–14389 (2010).
- 500 40. Veessler, D. & Cambillau, C. A Common Evolutionary Origin for Tailed-
501 Bacteriophage Functional Modules and Bacterial Machineries. *Microbiol. Mol.*
502 *Biol. Rev.* **75**, 423–433 (2011).
- 503 41. Auzat, I., Petitpas, I., Lurz, R., Weise, F. & Tavares, P. A touch of glue to
504 complete bacteriophage assembly: The tail-to-head joining protein (THJP)
505 family. *Mol. Microbiol.* **91**, 1164–1178 (2014).
- 506 42. Seul, A. *et al.* Biogenesis of a Bacteriophage Long Non-Contractile Tail. *J. Mol.*
507 *Biol.* **433**, 167112 (2021).
- 508

509 **Figure legends**

510 **Fig. 1 | Structure of the SPP1 connector (gp6₁₂gp15₁₂gp16₆) complex.** **a**,
511 CryoEM map of the SPP1 connector. One subunit of each gp6, gp15 and gp16
512 oligomer is shown in blue, magenta, and green, respectively. **b**, Atomic model of the
513 connector. For clarity, only the front half of the oligomer is shown. Four subunits of
514 gp6 are displayed in purple, cyan, blue and black. Four subunits of gp15 are in
515 orange, red, magenta, and brown. Three copies of gp16 are in variations of green.
516 Subunit numbers with their colour codes indicated on the right are used consistently
517 throughout the manuscript. **c**, gp15 dodecamer viewed from the capsid side (left
518 panel). Secondary structure elements are labelled on one subunit (right panel). **d**,
519 gp16 hexamer viewed from the capsid side (left panel). Secondary structure
520 elements are labelled on one subunit (right panel).

521

522 **Fig. 2 | Comparison of SPP1 portal protein gp6 structures and localization of**
523 **mutations impairing viral DNA packaging.** **a**, Secondary structural elements of the
524 gp6 subunit in the SPP1 connector. Domains are colour coded: the crown domain
525 (residues 421-470) is in green, the wing domain (residues 28-54, 88-255 and 369-
526 420) in orange, the tunnel loop with the tilted region of helix $\alpha 6$ (residues 347-368) in
527 red, the stem domain (residues 55-62, 81-87, 256-280 and 327-346) in magenta, the
528 gp6 β -hairpin (residues 63-80) in blue and the clip domain (residues 281-326) in
529 cyan. **b**, Superposition of the gp6 subunit from the connector (12-mer) and the
530 subunit of the gp6 assembly-naïve 13-mer (shown in grey). **c**, Clip domain of the gp6
531 13-mer (PDB 2JES¹²). **d**, Clip domain of gp6 in the connector complex. Residues
532 307-312 of gp6, chain *i-1*, are shown in purple, 281-326 of chain *i* in cyan, residues
533 291-294 of chain *i+1* in blue and residues 64-76 of chain *i+2* in dark blue. The clip is
534 in the same orientation as in **b**. 7Å distance indicates the shift of $\alpha 4$ in the connector
535 relative to $\alpha 4$ in the gp6 13-mer (see **c**). **e**, Localization of mutations impairing SPP1
536 DNA packaging with residues coloured according to their mutation group (see
537 Supplementary Table 2). The connector gp6 subunit is rotated 180° relative to **a** and
538 **b**.

539

540

541 **Fig. 3 | Intermolecular interfaces within the SPP1 connector.** **a**, gp6-gp15
542 interface. The C-terminus of the gp15 subunit (in magenta) interacts with the clip
543 pocket formed by three gp6 subunits displayed in cyan (*i*), blue (*i+1*) and dark blue
544 (*i+2*) as in Fig. 1b. View from the outside of the connector. **b**, Zoomed in view of the
545 gp6-gp15 interactions of the outlined area in (a). Interchain hydrogen bonds and salt
546 bridges between gp6 and gp15 subunits are displayed as dashed lines in black.
547 **c**, Comparison of the assembly-naïve gp15 monomer (left) (PDB 2KBZ¹⁹) and the
548 gp15 subunit in the connector (centre) superposed through their α -helical core
549 (magenta). Helix α_0 is shown in green, loop α_2 - α_3 is in blue, and the C-terminus is in
550 orange. The overlay of the gp15 subunit from the connector with the NMR monomer
551 (in grey) is shown in the right panel. The green arrow indicates rotation of helix α_0 on
552 $\sim 145^\circ$. **d**, Intersubunit interaction of gp15 α_0 with helix α_1 of adjacent subunit viewed
553 from outside of the connector. **e**, Gp15 Glu85 intersubunit bonding viewed from the
554 connector tunnel. **f**, View from the outer surface of the β -barrel shows the interface
555 contacts between the gp15 β -barrel bottom rim and gp16. Dashed lines indicate
556 hydrogen bonds and salt bridges between one gp16 subunit and three gp15
557 subunits. **g**, Side view of the β -barrel showing the interaction between gp15 and the
558 gp16 N-terminus.

559

560 **Fig. 4 | Conformational changes of gp16.** **a**, Comparison of the NMR structure of
561 the assembly-naïve gp16 monomer (left panel, PDB 2KCA¹⁹) with one subunit of
562 gp16 in the connector (central panel). The N-terminus is shown in dark blue
563 (residues 1-7) and loop β_2' - β_3 is in red (residues 40-56). The right panel shows the
564 overlay of two gp16 conformations aligned through their β -sheet core, the NMR
565 structure is indicated in grey. **b**, Tunnel helices of two adjacent subunits of the gp16
566 hexamer viewed from the capsid side. Black dashed lines show hydrogen bonds.
567 The position of the 6-fold symmetry axis is shown by a black hexamer.

568

569 **Fig. 5 | Model of connector assembly.** **a**, Anchoring of the gp15 monomer in the
570 gp6 clip. Twenty models of the gp15 monomer NMR structures¹⁹ are superposed to
571 highlight stable and flexible structural elements before assembly (left panel). The N-
572 terminus α_0 and loop α_2 - α_3 (shown in green and blue, respectively) undergo

573 conformational changes during assembly. **a-c**, Binding to gp6 of adjacent gp15
574 subunits leads to clashes between gp15 loops $\alpha 2$ - $\alpha 3$ inducing changes and folding
575 of the $\beta 1$ - $\beta 2$ hairpin (bent blue arrow in **a** indicates the change occurring in **b** upon
576 oligomerization), building the gp15 intersubunit β -barrel. Helices $\alpha 0$ reposition during
577 oligomerization (bent green arrow in **a** shows the change occurring in **b**), to bridge
578 helices $\alpha 1$ of adjacent subunits. Views from the tail side in the bottom part of the
579 figure show only gp15 subunits for clarity. **d**, Binding of gp16 to the gp15 ring.
580 Superposed twenty models of the gp16 monomer NMR structure¹⁹ are shown on the
581 left panel. The gp16 C-terminus and loop $\beta 2'$ - $\beta 3$, that change conformation during
582 assembly, are rendered in dark blue and red, respectively. **e-h**, Folding of gp16
583 loops $\beta 2'$ - $\beta 3$ to tunnel helices stabilized by inter-helices interactions following
584 oligomerization of gp16 in the connector. Views are shown from the phage tail side.

585

586 **Fig. 6 | Evolutionary divergence of the prokaryotic tailed viruses-herpesviruses**
587 **lineage at the portal vertex.** The portal protein (in grey-blue) is conserved in all
588 viruses of the lineage. Addition of a turret domain (in blue) to the portal in
589 herpesviruses marks their divergence from prokaryote viruses. The adaptor protein
590 (light magenta) that binds to the portal clip of prokaryotic viruses has various
591 structural features in the region distal from the portal. They define two evolutionary
592 division points for viruses with short tails, typified by phages T7 and $\phi 29$, from the
593 phages with long tails. The adaptor proteins of phages with long tails encompass a
594 β -barrel for binding stopper proteins (in green). The stopper builds the docking
595 interface for pre-assembled long tails. Domains of the herpes cytomegalovirus
596 (HCMV) portal protein and of tailed phage adaptors that define divergence points in
597 the lineage evolutionary trajectory are shown in dark blue and magenta, respectively.
598 The structures presented are the SPP1 connector (gp6₁₂gp15₁₂gp16₆; PDB 7ZW4)
599 (this work), HCMV pUL104₁₂ (PDB 7ETM)³⁶, T7 gp8₁₂gp11₁₂ (PDB 6R21)⁸, and $\phi 29$
600 gp10₁₂gp11₁₂ (PDB 6QZF)²³. Representative images of a herpesvirus capsid and the
601 mature phages in negative stain (2% uranyl acetate) are shown close to the
602 corresponding connectors. The scale bar corresponds to 50 nm.

603

604

Online Methods

605

606 **Connector complex purification.** The SPP1 connector complex purification was previously
607 described⁹. Briefly, SPP1 tailless particles were produced by infection of the non-permissive
608 host *B. subtilis* YB886 with phage SPP1^{sus9}⁴³ and purified by isopycnic centrifugation in a
609 discontinuous CsCl gradient⁴⁴. Tailless particles were disrupted with 50 mM EDTA for 30 min
610 at 55°C. Then, MgCl₂ was added to a final concentration of 100 mM and free DNA was
611 digested overnight with 50 U Benzonase in an oven at 37°C. Connector complexes were
612 then purified by sedimentation through a 10-30% (w/v) glycerol gradient in TBT buffer (100
613 mM Tris-Cl, pH 7.5, 100 mM NaCl, 10 mM MgCl₂) run at 35,000 rpm in a SW41 rotor
614 (Beckman Coulter) for 3 h at 4°C. Fractions containing connectors were identified by the
615 presence of gp6, gp15 and gp16 in SDS-PAGE and western blots⁹. They were then pooled
616 followed by concentration and buffer exchanged to 0.5x TBT in a Vivaspin microconcentrator
617 with a cut-off of 100 kDa.

618 **Cryo-EM grid preparation and data acquisition.** 3 µL of suspension of purified
619 gp6gp15gp16 connector complexes at 0.64 mg/mL were applied to a freshly glow-
620 discharged (using PELCO Easiglow, Ted Pella, USA) C-flat grid (Protochips, USA; 2/2 400
621 mesh) and blotted for 5 seconds before plunge-freezing the grids and vitrified using a
622 Vitrobot Mark IV (ThermoFisher™) at 96% humidity, and 4 °C. Data were collected on a
623 Titan Krios microscope (ThermoFisher™, the eBIC Diamond light source facility, Harwell,
624 Oxfordshire, UK) operating at 300 kV with the specimen maintained at liquid nitrogen
625 temperatures. Images were recorded using a Falcon III camera (ThermoFisher™) in
626 integrating mode that enables faster data acquisition. Data collection was done using the
627 EPU software (ThermoFisher™) within a defocus range of -1.2 to -2.5 µm. Movies (40
628 frames per movie) were collected with a dose of 1.12 e⁻/Å² per frame at the specimen plane
629 over a 1 s exposure. The calibrated pixel size was 0.69 Å at the specimen level.

630 **Image Processing.** 3,876 movies were aligned using MotionCorr²⁴⁵. CTFFIND4⁴⁶ was used
631 to determine defocus values (Extended Data Fig. 1). Micrographs were screened manually
632 to assess CTF quality and selected based on the presence of high-resolution Thon rings at
633 least to 4 Å and beyond for further processing. A set of 10 randomly selected micrographs
634 were used for manual picking of particles. Five 2D classes corresponding to the side views
635 were then used for automated particle picking from the entire data set. About 520,000
636 particle images were picked from ~3,500 micrographs using Relion 2.0⁴⁷. Correction of the
637 contrast transfer function was done by phase flipping during the following processing.

638 Initial steps of processing were done in Relion 2.0. The data set of particle images was
639 subjected to 3 rounds of 2D classification to remove atypical images that included some ice

640 contamination, particles that were overlapping or attached to the frame boundaries. The
641 best classes representing mostly side views (~10,000 images) were selected firstly for the
642 initial alignments and reconstruction. The portal protein gp6 map (EMD-1021)¹⁰ low pass
643 filtered to 20 Å, that revealed symmetry C12, was used as an initial model. Then we have
644 performed a 3D classification using 5 classes. No symmetry was applied during these steps
645 of analysis. Two classes were not well defined, two classes were rather similar with apparent
646 C12 symmetry and the last one had mirrored structure with approximate 12-fold symmetry.

647 The two structures with the correct handedness, assessed by fitting the main domains of
648 one subunit from the gp6 13 mer X-ray structure (PDB code 2JES)¹², were averaged and
649 used as a starting model to obtain the structure of the connector complex using the entire
650 data set of 520,000 particles. Several iterations of 3D refinement alternated with 3D
651 classification resulted in a structure at ~3.5 Å resolution.

652 In the next steps of refinement, we used a subset of movie frames from 2 to 21 after their
653 alignment to eliminate frames with high dose of radiation and re-extracted particle images.
654 These images and the model obtained in Relion 2.0 were exported in CryoSPARC⁴⁸ for
655 further processing. After three rounds of 2D classification in CryoSPARC, 401,807 particle
656 images were selected for the next two rounds of refinement. The structure of the
657 gp6gp15gp16 complex with imposed symmetry C12 indicated that density of the region
658 corresponding to the gp16 protein was nearly twice lower when compared to the densities of
659 gp6 and gp15. This finding indicated that protein components within the connector have
660 different symmetries. Therefore, we reduced the symmetry to investigate the distribution of
661 densities corresponding to each protein component. Utilisation of 3-fold symmetry revealed
662 six subunits in the region of gp16 implying that the gp16 ring has C6 symmetry. This
663 symmetry was used at the final refinement of the complex structure. The electron densities
664 corresponding to each of the three proteins were within the same range in the structure
665 obtained.

666 The final structure of the complex was obtained at a resolution of 2.4 Å as estimated by the
667 gold-standard Fourier shell correlation (FSC) at a 0.143 threshold and 2.7 Å at a 0.5
668 threshold. Local resolution variations in the reconstruction was estimated using ResMap⁴⁹
669 (Extended Data Fig. 2).

670 **Polypeptide chains tracing.** Interpretation of the cryoEM map was done using Chimera⁵⁰,
671 COOT⁵¹ and PHENIX⁵². The initial atomic model building for gp6 was done by rigid body
672 fitting of the stem domain (helices α 3 and α 5, residues 256-280 and 327-368 from the X-ray
673 structure of a gp6 13-mer ring, PDB code 2JES¹²). The remaining amino acids of the
674 sequence were fitted *de novo* by manual tracing using COOT with extensive adjustments of

675 the model to the cryoEM densities. That step was followed by real space refinement of the
676 atomic model against the experimental map in PHENIX, including restraints for rotamers, c-
677 beta, and Ramachandran constraints. Note that the short β -strand β 10 (residues 215-217) is
678 only found in chains A, C, E, G, I, and K due to poor local electron density in the gp6 wing.
679 The atomic models of gp15 and gp16 were build *de novo* in the cryoEM map. The amino
680 acid sequences were traced manually with COOT and real space refinement with PHENIX in
681 the cryoEM density map.

682 Refined atomic models of each protein in the connector complex were obtained by several
683 rounds of rebuilding in COOT and real space refinement with rotamer, c-beta, and
684 Ramachandran restraints in PHENIX. Manual inspection of the residues in the complete gp6,
685 gp15 and gp16 polypeptide chains tracing was done using COOT to correct the orientation
686 of rotamers and to trace regions of lower electron density quality (gp6 residues 170-176 and
687 217-241). Protein residues of the final atomic model show well-refined geometrical
688 parameters: most favoured regions, 91.71%; additionally, allowed regions, 8.18%; and
689 0.10% of outliers in Ramachandran plots (Supplementary Table 1). Inter-subunit bonding,
690 interaction interfaces and subunits dissociation energy in the connector complex were
691 computed using PISA³⁰ (Supplementary Tables 3 and 4). The protein interfaces electrostatic
692 potential was calculated and displayed using Pymol⁵³.

693 **3D homology search.** Structure homology searches were carried with DALI⁵⁴ using
694 individual subunits of the connector (this work) and the NMR structures of gp15 and gp16¹⁹.
695 Residue conservation of structurally aligned homologous proteins was analysed and
696 displayed in the linear protein sequence with DALI⁵⁴. Pairwise structure comparison between
697 domains of gp6 in the connector (this work) and in the assembly-naïve gp16 (PDB 2JES¹²)
698 were carried with the DALI server⁵⁴ using default settings. The PDB coordinates of the
699 corresponding domains were provided as input. The C α coordinates of the polypeptide chain
700 in the domains were used for calculation of the RMSD.

701 **Structure display.** Figures were prepared using Pymol⁵³ and Chimera⁵⁰ software.

702 **Reporting Summary.** Further information on research design is available in the Nature
703 Research Reporting Summary linked to this article.

704 **References**

- 705 43. Becker, B. *et al.* Head morphogenesis genes of the Bacillus subtilis
706 bacteriophage SPP1. *J. Mol. Biol.* **268**, 822–839 (1997).
- 707 44. Jakutyte, L. *et al.* Bacteriophage infection in rod-shaped gram-positive
708 bacteria: Evidence for a preferential polar route for phage SPP1 entry in

- 709 Bacillus subtilis. *J. Bacteriol.* **193**, (2011).
- 710 45. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion
711 for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
- 712 46. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation
713 from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
- 714 47. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-
715 EM structure determination with parallelisation using GPUs in RELION-2. *Elife*
716 **5**, (2016).
- 717 48. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC:
718 algorithms for rapid unsupervised cryo-EM structure determination. *Nat.*
719 *Methods* **14**, 290–296 (2017).
- 720 49. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution
721 of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
- 722 50. Pettersen, E. F. *et al.* UCSF Chimera?A visualization system for exploratory
723 research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- 724 51. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics.
725 *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
- 726 52. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and
727 crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 531–544 (2018).
- 728 53. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline
729 State. *J. Mol. Biol.* **372**, 774–797 (2007).
- 730 54. The PyMOL Molecular Graphics System, Version 2.2.3, Schrödinger, LLC.
- 731 55. Holm, L. DALI and the persistence of protein shape. *Protein Sci.* **29**, 128–140
732 (2020).
- 733
- 734