



HAL
open science

Biogenesis of a Bacteriophage Long Non-Contractile Tail

Anait Seul, Sandrine Brasilès, Isabelle Petitpas, Rudi Lurz, Valérie Campanacci, Christian Cambillau, Frank Weise, Mohamed Zairi, Paulo Tavares, Isabelle Auzat

► **To cite this version:**

Anait Seul, Sandrine Brasilès, Isabelle Petitpas, Rudi Lurz, Valérie Campanacci, et al.. Biogenesis of a Bacteriophage Long Non-Contractile Tail. *Journal of Molecular Biology*, 2021, 433 (18), pp.167112. 10.1016/j.jmb.2021.167112 . hal-03312278

HAL Id: hal-03312278

<https://hal.science/hal-03312278v1>

Submitted on 16 Nov 2021

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.

Biogenesis of a Bacteriophage Long Non-Contractile Tail

Anait Seul¹, Sandrine Brasilès^{1,2}, Isabelle Petitpas¹, Rudi Lurz³, Valérie Campanacci^{4, #}, Christian Cambillau⁴, Frank Weise^{3, #}, Mohamed Zairi¹, Paulo Tavares^{1,2,*}, and Isabelle Auzat^{1,2,*}

1 - Unité de Virologie Moléculaire et Structurale, Centre de Recherche de Gif, CNRS UPR 3296 and IFR115, CNRS, Gif-sur-Yvette, France

2 - Institute for Integrative Biology of the Cell, Université Paris-Saclay, CEA, CNRS, 91198, Gif-sur-Yvette, France

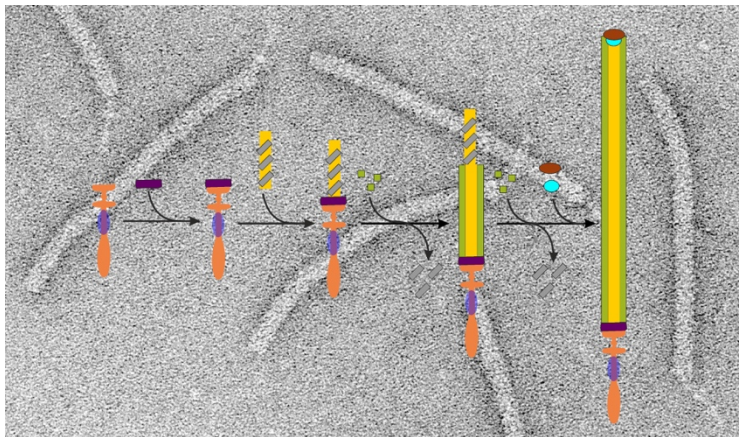
3 - Max Planck Institute for Molecular Genetics, D-14195 Berlin, Germany

4 - Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS and Universités d'Aix-Marseille I & II, Campus de Luminy, Marseille, France.

*Correspondence to: Paulo Tavares, paulo.tavares@i2bc.paris-saclay.fr (*P. Tavares*) or to Isabelle Auzat, isabelle.auzat@i2bc.paris-saclay.fr (*I. Auzat*). Tel. +33 1 6982 3860

#Present addresses: Valérie Campanacci, Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France; Frank Weise, Department of Molecular Biology, NMI, Reutlingen, Germany.

Graphical Abstract



Research Highlights

- Long bacteriophage tails are nanotube devices for delivery of viral DNA to the host cell cytoplasm
- Sequential program of protein interactions leading to assembly of phage SPP1 tail is uncovered
- Tail chaperones are essential to stabilize the TMP structure and for tail tube assembly
- Gp16.1, a highly conserved phage protein, acts during assembly of the tail interface for capsid binding
- SPP1 uses a minimal protein set to build long phage tails

Biogenesis of a Bacteriophage Long Non-Contractile Tail

**Anait Seul¹, Sandrine Brasilès^{1,2}, Isabelle Petitpas¹, Rudi Lurz³, Valérie Campanacci⁴,
#, Christian Cambillau⁴, Frank Weise^{3,#}, Mohamed Zairi¹, Paulo Tavares^{1,2,*}, and
Isabelle Auzat^{1,2,*}**

*1 - Unité de Virologie Moléculaire et Structurale, Centre de Recherche de Gif, CNRS
UPR 3296 and IFR115, CNRS, Gif-sur-Yvette, France*

*2 - Institute for Integrative Biology of the Cell, Université Paris-Saclay, CEA, CNRS,
91198, Gif-sur-Yvette, France*

3 - Max Planck Institute for Molecular Genetics, D-14195 Berlin, Germany

*4 - Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS and
Universités d'Aix-Marseille I & II, Campus de Luminy, Marseille, France.*

*Correspondence to: Paulo Tavares, paulo.tavares@i2bc.paris-saclay.fr (*P. Tavares*) or to
Isabelle Auzat, isabelle.auzat@i2bc.paris-saclay.fr (*I. Auzat*). Tel. +33 1 6982 3860

#Present addresses: Valérie Campanacci, Institute for Integrative Biology of the Cell
(I2BC), CEA, CNRS, Univ Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France;
Frank Weise, Department of Molecular Biology, NMI, Reutlingen, Germany.

2

3 **Abstract**

4 Siphoviruses are main killers of bacteria. They use a long non-contractile tail to recognize
5 the host cell and to deliver the genome from the viral capsid to the bacterial cytoplasm.
6 Here, we define the molecular organization of the *Bacillus subtilis* bacteriophage SPP1
7 ~6.8 MDa tail and uncover its biogenesis mechanisms. A complex between gp21 and the
8 tail distal protein (Dit) gp19.1 is assembled first to build the tail cap (gp19.1-gp21Nter)
9 connected by a flexible hinge to the tail fiber (gp21Cter). The tip of the gp21Cter fiber is
10 loosely associated to gp22. The cap provides a platform where tail tube proteins (TTPs)
11 initiate polymerization around the tape measure protein gp18 (TMP), a reaction dependent
12 on the non-structural chaperones gp17.5 and gp17.5* (TACs). Gp17.5 is essential for
13 stability of gp18 in the cell. Helical polymerization stops at a precise tube length followed
14 by binding of proteins gp16.1 and gp17 (THJP) to build the tail interface for attachment to
15 the capsid portal system. This finding uncovers the function of the extensively conserved
16 gp16.1-homologs in the assembly of long tails. All SPP1 tail components, apart from gp22,
17 share homology to conserved proteins whose coding genes' synteny is broadly maintained
18 in siphoviruses. They conceivably represent the minimal essential protein set necessary to
19 build functional long tails. Proteins homologous to SPP1 tail building blocks feature a
20 variety of add-on modules that diversify extensively the tail core structure, expanding its
21 capability to bind host cells and to deliver the viral genome to the bacterial cytoplasm.

22 **KEYWORDS** siphoviruses; macromolecular assembly; chaperone; tape measure; delivery
23 device

24

25 **Introduction**

26 Tail tubes are machines designed to deliver macromolecules to target cells. They were first
27 identified in bacterial viruses (bacteriophages or phages). The function of phage tails is to
28 recognize specifically host bacteria and to deliver a surgical attack of the cell envelope,
29 opening the way for viral genome transfer from the phage capsid to the bacterial cytoplasm
30 [1–9]. A diverse group of structurally and functionally related long contractile tails was
31 later found to be built by bacteria for delivery of protein effectors to other bacteria or
32 eukaryotic cells [10–15]. More than 95 % of the currently known bacterial viruses have a

33 tail device associated to an icosahedral capsid that shields the viral genome (order
34 *Caudovirales* [16]). This structural organization was retained throughout evolution for
35 efficient infection of Eubacteria and several Archaea [17]. Phage tails can be short (family
36 *Podoviridae*), long but non-contractile (family *Siphoviridae*) or long and contractile
37 (family *Myoviridae*). The tails of siphoviruses and myoviruses are composed of a long tube
38 and of a host adsorption apparatus. Their architectural components are built by structurally
39 conserved proteins. In addition, myoviruses have a distinctive sheath around the tail tube,
40 absent in siphoviruses, which contracts at the onset of infection. Tails are attached to the
41 unique portal vertex of the icosahedral capsid through which DNA enters and exits the
42 capsid [18]. The end of the tail distal from the capsid has the adsorption apparatus that is
43 responsible for interaction with the host cell surface. This structure varies significantly in
44 complexity from a single fiber to elaborated baseplates [2,19–24]. The sequential nature of
45 the siphoviruses' long tail assembly pathways was established in pioneer studies of phage
46 lambda and supported by more recent studies on lactococcal phages Tuc2009 [25] and
47 TP901-1 [26], but the underlying molecular assembly mechanisms remain largely
48 unknown.

49 Bacteriophage SPP1 is a well-characterized siphovirus that infects the Gram-positive
50 bacterium *Bacillus subtilis*. Its long non-contractile tail follows the conserved building plan
51 of long-tailed phages featuring an adsorption apparatus at the end distal from the capsid.
52 Adsorption of SPP1 to *B. subtilis* is mediated by the gp21 (Receptor Binding Protein
53 (RBP) of SPP1) carboxyl terminus (Cter) region (Receptor Binding Protein (RBP) of
54 SPP1) of the tail fiber [4]. The fiber is attached to the tail cap [19] that is composed of the
55 gp21 trimer amino terminus (Nter) and a gp19.1 hexamer (Distal Tail Protein (Dit) of
56 SPP1) [27,28]. This structure closes the 160 nm-long tail tube that is built by a helical
57 array of the Tail Tube Proteins (TTPs) gp17.1 and gp17.1* [19,29] surrounding a core
58 probably formed by the Tape Measure Protein (TMP) gp18. Purified gp17.1 has the
59 intrinsic property to polymerise into nanotubes of variable length that have a helical lattice
60 similar to the one found in phage tails [30,31]. The tail tube is tapered by the Tail-to-Head
61 Joining Protein (THJP) gp17 that creates the interface for attachment to the capsid
62 connector structure [32,33]. Although SPP1 tail protein components and their organization
63 are partially known, we still lack crucial information on the complete set of effectors
64 necessary for tail assembly and on the program of their sequential reactions to build the
65 complete structure. Here we describe the composition of the SPP1 tail, its architectural

66 organization, and the set of sequential reactions leading to its assembly. Bioinformatics and
67 comparative biology reveal that the SPP1 tail is representative of a minimal core to build a
68 long tail nano-device.

69

70 **Results**

71 **Morphology and composition of bacteriophage SPP1 tails**

72 In order to determine the SPP1 tail composition we established methods to obtain highly
73 purified structures (Figure 1A and Materials & Methods). Tails were purified from lysates
74 of the non-permissive *B. subtilis* strain YB886 infected with a conditional lethal mutant
75 defective for production of the major capsid protein gp13 (SPP1*sus31*, hereafter
76 abbreviated SPP1gp13⁻) [34]. A fraction of the tails was precipitated by 10 % (w/v)
77 polyethylene glycol 6000-8000 (pellet P2 in Figure 1A) while a second fraction required
78 addition of 1 M NaCl to be pelleted (P3) from the supernatant of S2 (Figure 1A). The
79 resuspended tail pellets were fractionated by anion exchange chromatography (AEC). Tails
80 from P2 and P3 exhibited a slightly different chromatographic behaviour: they elute
81 between ~230 and ~520 mM NaCl in the case of P2 and between ~280 and ~580 mM NaCl
82 in the case of P3 (Figure 1A). The procedure yielded highly pure tails in case of P3 (best
83 fraction P3f8) while elimination of minor contaminants from P2 (pool of P2f4 to P2f10 of
84 the AEC) required sedimentation through a 5-20 % (w/v) sucrose gradient (best fraction
85 P2f6S) (Figure 1A). The purified tails were identical to the ones found in phage particles.
86 They consisted of a 174 nm-long flexible tube terminating at the head proximal end in a
87 ring narrower than the tail tube diameter [29] and in a tail fiber on the other end (Figure
88 1B). The fiber (or tip) that is anchored at the tail cap [19,27] is a thin structure with a
89 central ellipsoid-like element and a broader, somewhat diffuse, extremity (Figure 1B).

90 Protein amino-terminal sequencing (gp19.1 and gp21), mass spectrometry (gp16.1,
91 gp17, gp17.1, gp17.1*, gp18, gp19.1 and gp21), and western blotting (gp17, gp17.1,
92 gp17.1*, gp18, gp19.1, gp21 and gp22) allowed to assign gp17.1, gp17.1*, gp18, gp19.1,
93 and gp21 to specific bands in the Coomassie Blue pattern of purified tails (Figure 1C,D,
94 Figure S1 and Table S1). Gp18 was detected in three independent bands by mass
95 spectrometry (Figure 1C, Figure S1, Table S1). The species with low electrophoretic
96 mobility (>175 kDa) corresponds probably to gp18 aggregates or SDS-resistant multimers

97 (gp18_{mult}). The apparent molecular masses of the two other bands are approximately 145
98 kDa (gp18_{145kDa}) and 73 kDa (gp18_{73kDa}). Independent mass spectrometry experiments
99 showed that both species have a good coverage of most of the gp18 sequence (900 residues
100 of the 1032 total protein sequence) (Figure S1 and Table S1). There is no straightforward
101 explanation for this behavior. It may be speculated that the faster migrating band
102 corresponds to differently truncated versions of gp18. Alternatively, the full-length gp18
103 adopts two conformations that bind SDS differently leading to an abnormal faster
104 migration of one of them (gp18_{73kDa}) as found for some proteins with transmembrane
105 segments [35]. An atypical electrophoretic behavior of gp18 can result of its basic
106 theoretical pI (10.4), its putative transmembrane segments (Figure S2A), and its extended
107 conformation inside the tail tube [19]. The three gp18 forms were recognized by a
108 polyclonal antibody raised against its N-terminal region (gp18₁₈₁₋₄₃₅) (Figure 1D) although
109 the band corresponding to the full-length protein gp18_{145kDa} was frequently less intense or
110 not detected (e.g. phage particles in Figure 1D). We have thus used the gp18_{mult} and
111 gp18_{73kDa} bands as the most faithful markers for wild type gp18 detection. Gp16.1 and
112 gp17 co-migrated in a single band in SDS-PAGE (Figure 1C).

113 Interestingly, gp22 was present in P3f8 tails but was found neither in P2f6S tails not in
114 phage particles (Figure 1D). This feature did not correlate with any detectable difference of
115 tail morphology (Figure 1B) but presence of gp22 could eventually account for the
116 different precipitation and chromatographic behaviour of P3f8 and P2f6S tails (Figure 1A).
117 Immunoelectron microscopy with anti-gp22 IgG showed no detectable labeling of the two
118 tail preparations, probably because the antibody stripped gp22 weakly bound to the P3f8
119 tails. This hypothesis was confirmed by the observation that when tails were pre-fixed with
120 glutaraldehyde, the tail fiber region of P3f8 tails was cross-linked by IgG through the
121 region between their ellipsoid element and the distal extremity (Figure 2A,B). Anti-rabbit
122 IgG colloidal gold labelling confirmed the specificity of gp22 immunolabeling (Figure
123 2C). Gp22 thus associates to the tail tip structure during SPP1 assembly but the interaction
124 is weak after release from infected cells leading to dissociation of gp22 from the tip
125 structure. To investigate if the association with gp22 is a widespread feature of gp21
126 homologous proteins we carried pBlast to identify homologs of gp22. The resulting 43
127 Blast hits identified proteins encoded immediately downstream of the gene coding for a
128 gp21 homolog (Figure S3A). Gp21-homologs co-encoded with gp22 present homology to
129 the complete SPP1 gp21 sequence (Figure S3B), consistent with an interaction of gp22

130 with the gp21 Cter fiber domain. Both gp22 and the sub-set of gp21 proteins are very
131 closely related to their homologs (Figure S3C,D). However, more than 100 genomes code
132 proteins highly homologous to gp21 (e-value < 2e-150) throughout their complete length
133 but no detectable gp22-homologs (Figure S3E). We conclude that the gp21-gp22
134 interaction, although conserved in a sub-set of phages, is not an essential feature for gp21
135 fiber function.

136

137 **Identification of SPP1 tail conditional lethal mutants**

138 The library of SPP1 conditional lethal mutants [36] was screened for phages unable to
139 produce tail structures or that produced DNA-filled capsids without tails in non-permissive
140 infections. Mutations were subsequently mapped to specific SPP1 genes using a plasmid
141 complementation assay. All putative tail genes were localized between the SPP1 capsid
142 and lysis gene clusters (Figure 3). Previously uncharacterized *sus* mutations were mapped
143 in genes *16.1* (*sus666*), *17.5/17.5** (*sus9*), *17.5** (*sus222*) and *18* (*sus76*) (Figure 3 and
144 Table S2). Thermo sensitive (*ts*) mutations impaired in multiplication of SPP1 at 46 °C
145 were assigned by complementation to genes *19.1* (*tsJ20*) and *21* (*tsJ5*, *tsJ15*, and *tsA58*)
146 (Figure 3 and Table S2). DNA sequencing showed that mutants in gene *21* carried multiple
147 mutations leading to several single amino acid substitutions in gp21 (Figure 3). However,
148 the three mutants were complemented by a plasmid coding for gp21 assigning the
149 temperature sensitivity phenotype to this protein. Other conditional lethal mutations
150 mapped in genes coding for tail proteins but that were not further characterized are
151 compiled in Figure 3 and in Table S2. All mutants are hereafter named according to the
152 gene they affect for simplicity (e.g. SPP1*sus666* is named SPP1*gp16.1*; in the case that
153 several mutants of a single gene are studied the original mutant name is included as, for
154 example, SPP1*gp21*⁻(*tsJ5*)).

155 The tail composition analysis combined with the mutants mapping and sequencing
156 identified essential genes for tail assembly (Tables S1 and S2). They all coded for SPP1
157 tail structural proteins with the exception of gp17.5 and gp17.5* (Figure 1D). Gp17.5* was
158 proposed to be generated by a -1 frameshift during translation of gene *17.5* at a TTTTTT
159 slippery sequence based on bioinformatics [29,37]. Mutation *sus222* in SPP1*gp17.5**
160 introduced a stop codon downstream of this putative frameshift position disrupting
161 specifically the gene *17.5** coding frame (Figure 3 and Table S2). This finding

162 demonstrated that gp17.5* is generated by a recoding event and that it is an essential
163 protein for SPP1 multiplication.

164 No conditional lethal mutations were identified in gene 22 and in open reading frames
165 *orf23*, *orf23.1*, and *orf24* localized between gene 21 and the lysis genes 24.1, 25 and 26
166 [38,39].

167

168 **Stability of SPP1 tail protein components**

169 Synthesis of SPP1 proteins required for tail assembly was detected 12-15 min (10 min in
170 case of gp17.5) after infection of *B. subtilis* at 37 °C and their amount increased until lysis
171 (Figure 4A). Gp17.1, gp17.5, gp19.1, gp21 and gp22 appear as single bands in western
172 blots. In the case of gp18, the two or three different bands found in purified phage particles
173 and tails (see above; Figure 1D) were detected specifically with the gp18 Nter antibody
174 also in extracts of infected bacteria (white arrowheads in Figure 4). Furthermore, those
175 bands were absent in infections with SPP1*gp18*⁻ (Figure 4B). This feature and the increase
176 of intensity of bands during infection were used as criteria to distinguish gp18 species from
177 several non-specific bands recognized by the antibody (Figure 4). In contrast to other tail
178 components that were stable upon freezing and thawing of extracts, the gp18 species were
179 labile, being detected best in freshly prepared extracts, and their ratios varied somehow
180 between preparations.

181 The presence of different tail proteins was analyzed in cells infected with mutants
182 defective for the production of functional forms of each individual tail protein. Non-
183 permissive infections with suppressor-sensitive mutants showed that the corresponding
184 gene products were absent in infected cells (Figure 4B). Gp17.1, gp17.5, gp19.1, gp21 and
185 gp22 were detected as single, stable bands, independently of the presence of other tail
186 proteins (Figure 4B,C). Gp21 was present in non-permissive infections of SPP1*gp21*⁻
187 (*tsJ15*), SPP1*gp21*⁻(*tsA58*) and SPP1*gp21*⁻(*tsJ5*) (Figure 4C) revealing that the mutant
188 proteins were produced at non-permissive temperatures but were not-functional. Analysis
189 of SPP1*gp19.1*⁻ that codes gp19.1_{S244N} revealed that the mutant protein appears less stable
190 in non-permissive infections (Figure 4C). The amino acid substitution S244N disrupts
191 hydrogen bonding of S244 with Y175 and Q242 in the galectin domain of gp19.1 (Figure
192 S4A-D) [27] that holds the gp21Nter in the SPP1 cap [28], providing a plausible

193 explanation for gp19.1_{S244N} lower stability and non-functionality at high temperature.
194 Gp17.1 of SPP1*gp19.1*⁻ exhibited reproducibly a reduced electrophoretic mobility (Figure
195 4C) due to presence of an additional mutation (Table S2). The resulting gp17.1_{I135M} amino
196 acid substitution did not impair tail assembly because the temperature sensitive phenotype
197 of this phage was complemented by gene *19.1* expressed from a plasmid.

198 Contrary to other tail components, detection of the mature gp18 pattern of bands found
199 in assembled tails required the production of several other tail proteins. No gp18 was
200 detected in absence of SPP1*gp17.5*⁻, while an additional gp18 species with an apparent
201 mass of ~83 kDa (gp18_{83kDa}) was found when gp17.1, gp17.5*, gp19.1 or gp21 were absent
202 (diamonds in Figures 4B,C and 5). In contrast, infections with SPP1*gp13*⁻, which is
203 defective in capsid assembly but forms normal tails [34], yielded a gp18 band pattern
204 identical to the one found in cells infected with wild type phages and in SPP1 virions.
205 Collectively, these results show that gp18 relies on interactions with gp17.5 for stability
206 and on assembly of the tail tube to achieve the protection pattern observed in infectious
207 virions.

208

209 **Tail proteins complexes and definition of the SPP1 tail assembly pathway**

210 In order to define the network of interactions leading to tail assembly, we produced
211 extracts of cells infected with mutants defective for production of each tail protein. Stable
212 tail protein complexes formed at different steps of the assembly pathway were enriched by
213 centrifugation through a sucrose cushion and then sedimented through a 10-30 % (w/v)
214 glycerol gradient. Different running times followed by fractionation and composition
215 analyses of both sedimentation steps were tested to obtain complete tail structures or its
216 assembly intermediates (Figure S5).

217 Under conditions in which complete tails are assembled (SPP1*gp13*⁻ infection; Figure
218 1), all tail proteins peaked as one fast sedimenting species corresponding to complete tails
219 (Figure 6A, top panel). These structures were rather pure and morphologically similar to
220 the tails in Figure 1B. High purity was achieved, when necessary, by an additional anion
221 exchange step as in Figure 1A. When the major tail proteins gp17.1/gp17.1*, the tape
222 measure protein gp18, or the non-structural proteins gp17.5/gp17.5* were absent, gp19.1
223 and gp21 co-sedimented as a ~19 S form (Figure 6A). The sedimentation coefficient was

224 estimated by comparison to the behavior of standards of known sedimentation coefficient
225 run in parallel (Figure 6B, bottom panel). Silver staining showed that the corresponding
226 fractions were significantly contaminated by other proteins (not shown) hampering
227 electron microscopy observation of the putative gp19.1-gp21 complexes. Slower
228 sedimenting species corresponding to full length gp21 (~11 S) and a shorter gp21 form (8.5
229 S) were detected. They corresponded to species with apparent molecular masses of ~142
230 kDa and ~81 kDa in SDS-PAGE, respectively. The shorter gp21 form, detectable with
231 antibodies raised against both the protein Nter and Cter regions, was not found in extracts
232 of infected cells (Figure 4), showing that it was generated during the partial purification
233 procedure. The two gp21 forms co-sedimented partially in a sample from non-permissive
234 infection with SPP1*gp19.1*⁻ that codes for the non-functional mutant gp19.1_{S244N} (Figure
235 6B). In contrast, gp19.1 was not found in the upper fractions of the glycerol gradient of a
236 sample from infection with SPP1*gp21*⁻(*tsJ5*) (Figure 6B), probably because gp19.1 was not
237 pelleted in the sucrose cushion step that precedes the glycerol gradient (Figure S5).

238 Assembly of tails with wild type length require presence of gp21, gp19.1, gp18, gp17.5,
239 gp17.5*, and the TTP gp17.1 (the minor TTP gp17.1* species is not essential for phage
240 assembly [29]) (Figure 6). Analysis of the composition of purified tails identified two other
241 tail structural proteins, gp17 and gp16.1 (Figure 1C and Table S1). We have previously
242 shown that gp17 is the SPP1 tail-to-head joining protein found at the tail end distal from
243 the adsorption apparatus [33]. In order to investigate the role and topology of gp16.1 in the
244 tail structure, we used phage SPP1*gp16.1*⁻ (Figure 3). During infection under non-
245 permissive conditions tail proteins, including gp17, are produced but not gp16.1 (Figure
246 7A). Tails purified from those lysates had a sedimentation behavior similar to wild type
247 tails (Figures S6 and 6A, respectively). They lacked gp16.1, as anticipated, but also gp17
248 (Figure 7B). Electron microscopy observation showed that the tails had a normal
249 adsorption apparatus and tube length (Figure 7C). However, they did not exhibit the gp17
250 ring that caps the tail tube in wild type tails (white arrows in Figure 7C) [33]. Therefore,
251 gp16.1 is a structural component required for normal tail construction. It binds to the
252 assembled tail tube extremity preceding attachment of gp17 that forms the head-to-tail
253 joining ring found in isolated tails.

254

255 **Plasticity of the tail components building plans**

256 We then carried out homology searches of the SPP1 tail proteins to investigate the
257 landscape of their diversity and modular organization. Amino acid sequence homology was
258 found predominantly to prophage-encoded proteins. Homologous proteins were detected in
259 Firmicutes, mostly in phages and prophages of the *Bacillus* genus. Their genes synteny is
260 highly conserved, following the arrangement found in the genome of SPP1 (Figure 3)
261 [29,40,41]. A previous study showed that the major TTP of SPP1 gp17.1 can be decorated
262 with a Cter FN3 domain that is exposed to the tube outside [29]. The alignment of the TMP
263 gp18 homologous proteins reveals a pattern of homology throughout the complete gp18
264 sequence punctuated with internal deletions and insertions (Figure S2B) which likely play
265 a role to determine tail length as shown experimentally for phages λ [42], T4 [43], and
266 TP901-1 [44]. Most gp18-homologs have N- and C-termini longer than gp18 which might
267 further extend the tail tube length and/or play a role in specific interactions with proteins at
268 the tail tube ends.

269 Alignment of the 253 amino acids-long gp19.1 Dit amino acid sequence with its
270 homologs revealed a modular organization (Figure S4E). Its Nter, the most conserved
271 sequence, has a TTP-like fold in the gp19.1 hexamer (Figure S4A,B) that extends the tail
272 tube [28]. The gp19.1 Cter galectin domain sequence is less conserved albeit homology is
273 detected (Figure S4E). This domain irradiates outwards from the tail tube hexamer-like
274 structure (Figure S4A,B) to clamp, together with the Nter ring, the gp21 Nter trimer in the
275 tail cap structure [28]. A clear disruption of homology (Figure S4E) is found in a loop
276 protruding to the outside of the galectin domain (Figure S4A). This marks an increase in
277 sequence variability and length of gp19.1 homologs (Figure S4E). HHPred predicts that
278 their Nter has a gp19.1 TTP-like fold, as expected, but also that the far Cter end is
279 structurally related to the Cter of gp19.1 galectin domain. The two regions are separated in
280 numerous cases by either an internal module of unknown function, by a carbohydrate
281 binding module previously reported for “evolved Dits” of phages infecting other Gram-
282 positive bacteria (PDB 6LY8) [45,46], or by combination of several modules. This is
283 exemplified by HHPred structural homology analysis of the 783 amino acids-long Dit
284 WP_077616833.1 that distinguishes three modules integrated in the galectin domain: a
285 module of unknown function, a module with several structural homologs including
286 bacterial cell envelope binders (PDB 6LVV, 3FLP), and the carbohydrate binding module
287 of “evolved Dits” (PDB 6LY8). Collectively, these results highlight the remarkable
288 structural plasticity of gp19.1 homologous proteins.

289 Gp21 homologs exhibit high sequence conservation on their Nter module (Figure
290 S3B,E). In SPP1, gp21₁₋₅₂₂ was shown to form a trimer that closes the tail cap [28].
291 HHpred reveals also structural relatedness of gp21₁₋₃₉₉ to the structure of trimeric proteins
292 from phage tails like a *Listeria monocytogenes* EGD-e prophage tail component (PDB
293 3GS9), *Staphylococcus* phage 80 α (PDB 6V8I) [47] and the *Escherichia coli* myophage
294 Mu tail baseplate (PDB 1WRU) but also to tail-related cellular structures like pyocin R2
295 (PDB 6U5H), the baseplate of *Photobacterium* PVC system (PDB 6J0M), or the
296 anti-feeding secretion system of *Serratia entomophila* (PDB 6RBK). The gp21 Nter cap
297 module is thus a highly conserved trimeric building block found in tails of siphovirus and
298 myophages of evolutionarily distant hosts that is also used by different tail tube-like
299 cellular weapons [48]. These are frequently individual proteins. They can close reversibly
300 the tail tube end [9,20,28] or extend the tube that is plugged by the tail cell puncturing
301 device (PDB 1K28 [49], 1WRU). Their common function appears to be acting as a Tail
302 Adaptor Protein (TAP) that builds the 6 to 3-fold symmetry transition at the tail end [9].

303 The gp21 Nter is followed by a sequence of weak structural homology to glycoside
304 hydrolases consistent with a role of gp21 in cell wall degradation (Tal function) although
305 this activity was not showed experimentally. The Cter of gp21 is significantly more diverse
306 in sequence and no significant structure-based homology hits were found with HHpred.
307 The gp21 closest homologs align through their complete length (Figure S3E) and their
308 genes are frequently linked to genes coding gp22-homologs (Figure S3A,B). More distant
309 gp21-homologs show homology within a shorter Cter region and then have an unrelated
310 sequence at their Cter end or alternate segments of homology with unrelated sequences
311 (Figure S3E). The less related gp21-homologs align only through their Nter with a clear
312 homology breakpoint around residue 522 of gp21 (Figure S3E), in excellent correlation
313 with the domain that closes the tail cap. Importantly, all these proteins have a Cter domain
314 which can be longer than 1000 residues, irrespectively of the length of homology with
315 gp21. We conclude that gp21-homologs represent a group of proteins that build long tail
316 fibers covalently attached to the tail cap trimeric module that closes reversibly the tail tube.
317 We propose that they are an all-in-one surgical device that binds the bacterial receptor,
318 most likely perforates the cell envelope, and opens to deliver the TMP and viral DNA to
319 the host cell. They differ from shorter trimers (<650 residues-long) whose complete
320 structure is related to the gp21Nter (individual TAP proteins). Those are found in
321 siphoviruses (e.g. phages p2 ORF16, *L. monocytogenes* EGD-e prophage gp18,

322 staphylococcal phage 80 α gp59), and in myoviruses (e.g. phages T4 gp27, Mu gp44)
323 featuring complex baseplates with numerous RBPs hooked in the Dit protein hub and/or
324 other proteins that form an axial fiber extending from the gp21 Nter-like cap region
325 [9,20,22,47,49].

326

327 **Discussion**

328 In this work we established the bacteriophage SPP1 tail composition and characterized
329 its assembly intermediates to obtain a model of the complete set of sequential protein-
330 protein interaction program leading to the functional tail (Figure 8). The SPP1 tail has a
331 simple architecture when compared to other siphoviruses representing the minimal core of
332 essential elements to build a functional tail. The two SPP1 tail originalities are its transient
333 gp22, found in a limited number of phages with SPP1 gp21 homologs (Figure S3), and the
334 optional FN3 domain of the TTP gp17.1* [29].

335 Gp21 is a multi-functional trimeric protein whose Nter closes reversibly the tail cap
336 (TAP function) [28] and the Cter has a receptor-binding activity (RBP) [4]. These
337 functions are associated to two distinct domains connected by a flexible hinge underneath
338 the cap that allows free rotation and tilting of the fiber RBP relative to the tail tube (Figure
339 2A,B) for three-dimensional search of receptor YueB in the *B. subtilis* surface. Like SPP1,
340 the tail central fiber is the essential tool for host recognition and receptor binding in phages
341 lambda [50] and T5 [9,51] which are the other well-characterized siphoviruses that use a
342 bacterial protein receptor for host recognition. In these phages the tail cap plug and the
343 RBP are different proteins while SPP1 gp21 typifies a single-protein minimal system to
344 carry out both functions (TAP/RBP). This modular organization is highlighted by sequence
345 alignment of gp21 homologous proteins (Figure S3E). Higher divergence of their RBP
346 Cter module is consistent with increased plasticity to recognize different bacterial receptors
347 while other homologous proteins lack the Cter module probably because the RBP function
348 is accomplished by a different phage protein. A gp21 ~11 S species that is stably formed in
349 absence of the cap gp19.1 (Dit) protein was the first precursor of tail assembly detected in
350 our study (Figure 6B). This species is likely a trimer, as found in viral particles [19], but its
351 elongated shape hampered molecular mass confirmation from the sedimentation data. The
352 gp21₁₋₅₂₂ Nter cap plug could be produced soluble in heterologous systems [28] but full-
353 length gp21 and its Cter were insoluble in spite of extensive optimisation work (25; our

354 observations). It is possible that correct folding and trimerization of the gp21 Cter fiber
355 domain occurs at low gp21 concentration to prevent aggregation, or that it requires a
356 dedicated, yet unidentified, chaperone during infection. The transient binding of gp22 to
357 the fiber (Figure 2) renders it a potential candidate for this role. Although all gp22
358 homologous proteins are encoded adjacently to genes coding full-length gp21 homologs,
359 they are found only on a minority of the large number of gp21 homologous proteins
360 presently identified (Figure S3). We thus favour the hypothesis that gp22 does not play a
361 critical role in tail fiber assembly but might have another function like shielding the tail
362 fiber from adsorbing to cell debris when infected cells lyse.

363 When gp21 and gp19.1 are produced during SPP1 infection, they form the 19 *S* cap-tail
364 fiber complex (Figure 6). The direct interaction between gp19.1 and the gp21₁₋₅₂₂ Nter
365 purified proteins was previously demonstrated to build an SPP1 tail cap-like structure [28]
366 similar to the one found in viral particles [19]. The 19 *S* complex accumulates in bacteria
367 infected with phages that do not produce the major tail tube protein gp17.1, the tape
368 measure gp18 or its non-structural chaperones gp17.5 and gp17.5* (Figure 6). Thus, in the
369 case of SPP1, no stable intermediate is formed between the gp19.1-gp21 complex and tail
370 tube assembly effectors in contrast to phage lambda, where a complex of the tail fiber-cap
371 with the TMP gpH and possibly the chaperone gpG was detected [52,53]. We hypothesize
372 that such complex is also assembled to initiate SPP1 tail tube formation (bracketed in
373 Figure 8) but is unstable when the TTP does not polymerize around the TMP scaffold to
374 build the tail tube. Anchoring of the TMP C-terminus trimer on the Dit central cavity of
375 phage 80α [47] further supports the notion that this interaction is central to initiate tube
376 assembly.

377 Detection of SPP1 TMP gp18 in infected cells depends strictly on the presence of the
378 non-structural protein gp17.5 (Figures 4B,C and 5). Gp17.5 chaperones most likely gp18
379 folding and/or shields gp18 from protease attack, thus maintaining it in an extended
380 conformation competent for tail tube assembly before protection inside the tube. In
381 mutants defective for assembly of the cap-tail fiber complex (gp19.1 and gp21) or of the
382 tail tube (gp17.1 and gp17.1*), there is a gp18_{83kDa} species that is reproducibly found in
383 infected cells but is absent if tail tube construction is achieved. We conclude that p17.5 is
384 necessary for gp18_{83kDa} stability. Gp18_{83kDa} could be a precursor of gp18_{73kDa} during
385 assembly as the latter species is very poorly detected when the tail tube is not assembled

386 (Figure 4). However, the low amount of gp18 in infected bacteria and the lability of its
387 forms hampered investigation of this hypothesis. Gp17.5*, whose first 112 residues are
388 similar to gp17.5 (Figure 4A) [29], is not necessary to chaperone gp18_{83kDa} (Figure 5) but
389 is essential for tail tube assembly. This is consistent with a role of the gp17.5* Nter to
390 cross-talk with the gp17.5-gp18 complex and its unique Cter to have a different function.
391 The Cter could act as a tail cap-gp18-gp17.5 initiator complex that primes gp17.1 helical
392 polymerization for building-up the tail tube, as proposed for phage lambda [54,55]. There
393 are, however, noteworthy differences between the two phage systems, such as the simpler
394 two-protein cap-tail fiber system that primes tail tube assembly and the requirement of
395 gp17.5 for SPP1 TMP gp18 stability in the infected cell. In contrast, lambda full-length
396 gpH is stably produced, yet in a non-functional form [55], and associates to the tail tip
397 when the chaperone gpG is not present in the cell [52,53].

398 Another major finding of our work was the assignment of SPP1 gp16.1 to a specific role
399 at the stage of tail assembly termination. Bioinformatics showed that gp16.1 homologous
400 proteins encoded by a gene adjacent to the THJP gene are widespread in siphoviruses and
401 myoviruses consistent with a conserved function [41]. Mutants of phages lambda (gpZ)
402 [52], P2 (gpS) [56], Mu (gpG) [57], and TP901-1 (Tap_{TP901-1}) [26] related their gp16.1
403 homologs to phage infectivity or to tail to capsid joining. Furthermore, their relative of
404 phage T5 is a tail component (T5 (p143) [51]). However, their precise function in phage
405 assembly remained unknown. We have now filled this gap showing that gp16.1 is a
406 component of the tail that is incorporated on its structure after tube assembly. The TMP
407 gp18 remains stable at these assembly steps between the halt of tail tube polymerization
408 and the final viral particle state (Figure 7B). Gp16.1 thus acts on the assembled tail tube
409 and its presence correlates with recruitment of gp17 to form a ring at the extremity of the
410 tail (arrow in Figure 7C). This ring is a hexameric structure that binds to the capsid
411 connector gp16 protein [58], finalizing assembly of the infective virion [33].

412 Collectively, our work provides a comprehensive model of the complete SPP1 tail
413 assembly pathway and elects it as a minimal model system for tail build-up. Its essential
414 eight core protein components are gp21 and gp19.1 constructing the cap-tail fiber complex,
415 the tail tube gp18 and gp17.1 building blocks assisted by non-structural chaperones gp17.5
416 and gp17.5*, and the tail-capsid joint effectors gp16.1 and gp17. Functional equivalents of
417 these proteins are found in the vast majority of long-tailed phages (Figures S2-S4).

418 Remarkably, all components of the fiber and tube can fuse with additional domains (e.g.
419 the insertions in the galectin domain of gp19.1-homologs [Figure S4], additions to the
420 gp17.1 Cter [29], and the elaboration of gp21 Cter-homologs [Figure S3]) or recruit other
421 protein elements (e.g. the hub role of gp19.1-homologs to build elaborated baseplates
422 [59]). Such plasticity delivers a large panoply of functional add-ons that diversify the
423 phage particle capabilities to adsorb to surfaces (48 and references therein), recognize
424 different host receptors [15,45,46,59], to attack the cell envelope [61], and to initiate host
425 cell takeover [62]. These large combinatorial possibilities offer phages with immense
426 opportunities to change and adapt. The SPP1 tail offers a gadget-free system, apart from
427 the dispensable gp17.1* Cter FN3 domain [29] and possibly gp22 (Figure 2), to
428 characterize the limited number of essential sequential reactions to build a long flexible
429 tail. The molecular principles driving the process apply very likely to the general building
430 of tail tube-like delivery devices.

431

432 **Materials & Methods**

433 **Bacterial and phage strains**

434 *E. coli* DH5 α , was used for cloning procedures. Bacteriophages SPP1 wild type [63], *sus*
435 and *ts* mutants were from the MPIMG collection [36]. The permissive strains *B. subtilis*
436 HA101B [64] was used to amplify all *sus* mutants with the exception of
437 SPP1gp17.5/gp17.5* (mutant SPP1*sus9* [34]) that was amplified in *B. subtilis* BG295
438 [65]. Mutants *ts* were multiplied in strain YB886 [66] at 30°C. Non-permissive infections
439 were carried in strain YB886 for *sus* mutants and in the same strain at 46°C for *ts* mutants.

440

441 **Cloning procedures**

442 Genes *16.1* to *17.1* were amplified from SPP1 phage DNA by PCR using primers 60102BZ
443 and 60102AZ (Table S3). The PCR fragment cleaved with EcoRI-PstI was cloned into
444 vector pBluescript pSK- (Stratagene) cut with the same enzymes generating pSK17. A
445 BamHI-MspI fragment carrying gene *16.1* was excised from pSK17 and cloned in plasmid
446 pRSETA (Invitrogen) cleaved with BamHI-PvuII to generate pBT361.

447 Genes *17.5* to *22* were amplified from SPP1 phage DNA using primers 17.5CS1 and

448 23NCS1 (Table S3) in a ExpandTM Long Template PCR System (Roche) reaction. The
449 PCR fragment cleaved with BamHI-PstI was cloned into plasmid pHP13 [67] cut with the
450 same enzymes to generate pPT220. Plasmid pPT220 was then cleaved with EcoRI and re-
451 ligated to obtain pPT221 that codes genes 21 and 22.

452 Genes 17.5 to 17.5* were amplified from SPP1 phage DNA by PCR using primers
453 gene 17.5 CS and gene 17.5* NCS. The PCR fragment was cleaved with BamHI-PstI and
454 cloned into plasmid pRSETA cleaved with the same enzymes to generate pPT118. Genes
455 17.5 to 17.5* of pPT118 were then amplified by PCR using primers prset1 and prset2
456 (Table S3). The PCR fragment cleaved with SmaI-PstI was cloned into SmaI-PstI sites
457 under the control of promoter PN25/0 in a pHP13-derived shuttle vector [68] generating
458 pPT200. This plasmid codes gp17.5 and gp17.5* fused to the pRSETA Nter tag featuring a
459 hexahistidine sequence (Invitrogen). Plasmid pPT200 was used as template for site-
460 directed mutagenesis with the QuikChange kit (Stratagene) using primers gene 17.5 to
461 17.5* CS and gene 17.5 to 17.5* NCS (Table S3) to insert a C nucleotide in the putative
462 slippery sequence (TTTTTT) proposed to generate the -1 frameshift for translation of
463 gp17.5* [29]. The resulting plasmid pPT209 codes tagged gp17.5*. Site-directed
464 mutagenesis of pPT200 with primers gene 17.5 only CS 2mut and gene 17.5 only NCS
465 2mut (Table S3) was used to degenerate the genes 17.5-17.5* putative slippery sequence
466 for production of gp17.5 only. A by-product clone (pPT212) carrying deletion of gene 17.5
467 first phenylalanine codon and a nucleotide substitution changing the second phenylalanine
468 codon to leucine within the slippery sequence (TTTTTT→CTT) proved to be the most
469 effective way to disrupt the -1 frameshift. Plasmid pPT212 leads to exclusive
470 overproduction of tagged gp17.5 carrying mutation F₁₁₁F→L₁₁₁.

471 Overproduction of several gp18 polypeptides was attempted due to the known
472 difficulty for stable production of the tape measure protein. The criteria to design
473 polypeptides for production were based on sequence homology conservation and to avoid
474 (i) putative transmembrane segments, (ii) breaking putative α -helices, (iii) predicted
475 unstructured regions, and (iv) low sequence complexity regions. Fragments coding gp18₁₈₁₋
476 ₄₃₅ (primers gene 18 181 CS and gene 18 435 NCS (Table S3)), gp18₃₄₇₋₄₃₅ (primers gene
477 18 347 CS and gene 18 435 NCS), and gp18₆₇₂₋₉₁₉ (primers gene 18Cter CS and gene 18
478 919 NCS) were generated by PCR. They were then cleaved with BamHI-PstI and cloned
479 into pRSETA (Invitrogen) cut with the same enzymes. The resulting plasmids pPT122,

480 pPT124 and pPT123, respectively, encode the gp18 polypeptides with a pRSETA Nter tag
481 containing a hexahistidine sequence.

482 A 756 bp fragment covering the gene *19.1* sequence lacking the initiation and stop
483 codons (position from 15365 to 16120 of the SPP1 nucleotide sequence; access code
484 X97918.3) was amplified by PCR from SPP1*wt* DNA with two flanking sequences coding
485 for 3'-PstI and 5'-AgeI restriction sites. Oligonucleotides F-gp19.1 and R-gp19.1 (Table
486 S3) were used as forward and reverse primer, respectively. The purified PCR product was
487 totally sequenced at GATC Biotech (Konstanz, Germany), cleaved with PstI and AgeI, and
488 cloned into the pIA2 shuttle vector [33] digested with the same restriction enzymes. The
489 resulting plasmid was named pIA29 (6His-gp19.1).

490 Genes *23* and *23.1* were amplified by PCR and cloned by Gateway recombination into
491 the pETG-20A vector using an established strategy [69,70]. The resulting plasmids
492 encoded His₆-tagged thioredoxin followed by a tobacco etch virus (TEV) protease cleavage
493 site fused to the Nter of gp23 (this work) and gp23.1 [70].

494 A PCR fragment bearing gene *24* was amplified with oligonucleotides F-gp24 and R-
495 gp24 (Table S3) followed by cloning in the pIA2 vector [33] as described above for gene
496 *19.1*.

497

498 **Mapping and sequencing of SPP1 conditional lethal mutations**

499 The SPP1 conditional lethal mutants collection [36] was screened against individual clones
500 of the tail genes ORFeome in *B. subtilis* using a complementation spot assay in 96-well
501 plates. The screening led to the identification of *sus* mutants in genes *17.5*, *17.5**, and *18*.
502 An alternative large scale mapping approach used plasmids covering clusters of SPP1 late
503 genes (pPT25 [29], pPT220 and pPT221) followed by fine mapping to individual genes.
504 This approach led to identification of a supressor-sensitive mutant in gene *16.1* and of *ts*
505 mutants in genes *17.5*, *19.1* and *21*. Identification of mutations was carried out by DNA
506 sequencing of the gene complementing the conditional lethal mutant. The observation that
507 the gene *19.1*-defective mutant SPP1*ts20* codes for gp17.1 with a retarded electrophoretic
508 mobility (Figure 4C) prompted also sequencing of its gene *17.1* which carries a missense
509 mutation in addition to the one in gene *19.1* (Figure 3).

510

511 **Purification of SPP1 phage tails**

512 Growth of the bacterial host strain, lysis and enzymatic treatment of the lysate were carried
513 out at 37 °C and the following steps at 4 °C, unless stated otherwise. Figure 1A illustrates
514 the method. *B. subtilis* strain YB886 was grown in 1 L of LB medium to an OD₆₀₀ of 0.8,
515 then supplemented with 10 mM CaCl₂ and infected with SPP1*gp13*⁻ mutant phages at an
516 input multiplicity (i.m.) of 0.5. At 2 h post-infection, Roche complete protease inhibitors
517 (without EDTA) were added, and the lysate was treated with 200 µg mL⁻¹ of lysozyme, 5
518 µg mL⁻¹ of DNase and 80 µg mL⁻¹ of RNase for 15 min. After overnight centrifugation at
519 12785 g, the supernatant was stirred with 10 % w/v of PEG 6000-8000 for 1 h, and
520 centrifuged for 1 h at 11325 g. The pellet P2 was resuspended in a total volume of 14 mL
521 of buffer TBT (100 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM MgCl₂) and dialysed
522 against 100 mM Tris-HCl pH 7.5, 10 mM MgCl₂. The P2 dialysate was centrifuged for
523 15 min at 13776 g and the supernatant was consecutively filtered through a 0.45 µm, and a
524 0.20 µm syringe filter (Sartorius). The sample was applied to a Ressource™Q anion
525 exchange column (GE Healthcare) and eluted with consecutive gradients of 0 M to 0.35 M
526 NaCl (6 column volumes (CV)) and 0.35M to 1M NaCl (0.5 CV) in 100 mM Tris pH 7.5,
527 10 mM MgCl₂ at room temperature and at a flow rate of 6 mL min⁻¹. Tail containing
528 fractions identified by SDS-PAGE and Coomassie Blue staining were pooled and
529 concentrated in a Vivaspin 20 device (MWCO 10,000, Sartorius), the buffer being
530 concomitantly exchanged for TBT. Aliquots of 200 µL each were loaded onto 5 to 20 %
531 w/v sucrose gradients prepared in TBT, which were spun in an SW41 rotor (Beckman) at
532 35000 rpm, 4°C, for 4 h and 15 min. Fractions of 1 mL each were collected starting from
533 the bottom of each tube, and analyzed by SDS-PAGE and Coomassie staining.

534 The supernatant S2 of the PEG precipitation step (Figure 1A) was precipitated by
535 addition of NaCl to 1 M over a period of 20 min. After stirring for another 1 h, the sample
536 was centrifuged for 1 h at 11325 g. The pellet P3 was resuspended in a volume of 14 mL of
537 TBT and dialysed against 100 mM Tris-HCl pH 7.5, 10 mM MgCl₂. The dialysate was
538 filtered through a 0.20 µm syringe filter and subjected to anion exchange chromatography,
539 as described for the resuspended P2 sample. Fractions containing purified SPP1*sus31* tails
540 were stored at 4 °C.

541 A lower scale phage tail purification starting from 130 mL culture infected either with
542 SPP1*gp13*⁻ or with SPP1*gp16.1*⁻ used the method developed for partial purification of tail

543 complexes from *B. subtilis* lysates followed by sedimentation of the semi-purified tail
544 protein complexes through a 10 % to 30 % w/v glycerol gradient prepared in TBT buffer,
545 as detailed in the sections with those titles below. Tail-enriched fractions were pooled,
546 dialysed against 100 mM Tris-HCl pH 7.5, 10 mM MgCl₂ and subjected to an anion
547 exchange chromatography step under the conditions described above to obtain pure tails.

548

549 **Partial purification of tail protein complexes from *B. subtilis* lysates**

550 Complexes of tail proteins were enriched and partially purified from lysates of *B. subtilis*
551 YB886 infected with the SPP1_{sus} or SPP1_{ts} mutants of interest (Table S2). During growth,
552 infection and lysis, cultures were maintained at 37 °C when working with *sus* mutants, and
553 at 46 °C for the *ts* mutants. A 130 mL LB culture of the host strain, grown to OD₆₀₀ = 0.8
554 and supplemented with 10 mM CaCl₂, was split for parallel infection of 40 mL cultures,
555 each with the SPP1 mutant(s) under analysis, at an i.m. of 5 pfu cfu⁻¹. At 37 °C, lysis was
556 allowed to proceed for 1 h after infection. The lysates were supplemented with Roche
557 complete protease inhibitors and treated with 200 µg mL⁻¹ of lysozyme, 10 µg mL⁻¹
558 DNase and 100 µg mL⁻¹ RNase for 15 min. At 46 °C, the time spans were reduced to 48
559 min post-infection and 12 min for enzymatic treatment. All the following steps were
560 carried out on ice or at 4 °C. Lysates were pre-cleared by centrifugation for 15 min at
561 39,281 g, splitted in four aliquots of ~10 mL that were overlayed on top of 3 mL cushions
562 of 20 % w/v sucrose in buffer TBT, and run at 32,000 rpm in a SW41 rotor (Beckman).
563 The sedimentation time was 15 h or prolonged to 19 h in case of *ts* mutant lysates and of
564 the *sus76* lysate run in parallel with those. After centrifugation, three fractions of
565 approximately 8 mL, 2.1 mL and 1.6 mL were collected by pipetting, starting from the top
566 of each tube, leaving the bottom 0.6 to 0.8 mL of the sucrose cushion in each tube. Pellets
567 were resuspended in this remaining liquid, and material from the four tubes was pooled for
568 each mutant.

569

570 **Sedimentation of semi-purified tail protein complexes**

571 The pooled samples from the sucrose cushions were concentrated to 200 µL each, and the
572 buffer was concomitantly exchanged for sucrose-free TBT using Vivaspin 20 devices
573 (MWCO 10,000). Samples were loaded onto 10 % to 30 % w/v glycerol gradients prepared

574 in buffer TBT and centrifuged at 35000 rpm, 4 °C in an SW41 rotor. Sedimentation times
575 were 7 h in the case of SPP1*gp13*⁻ and SPP1*gp16.1*⁻, 12 h for the other SPP1*sus* mutants
576 and 15 h for the *ts* mutants and the SPP1*gp18*⁻ control sedimented in parallel with those
577 (Figure 6). Twelve fractions of 1 mL each were collected with a syringe and peristaltic
578 pump, starting from the bottom of each tube. Pellets were resuspended in the roughly 0.8
579 mL remaining in each tube after fractionation.

580

581 **Mass spectrometry of tail proteins**

582 Samples were separated on SDS-PAGE gels and stained with Coomassie Blue. Excised
583 bands were digested with trypsin and peptides were analysed by MALDI-TOF (instrument:
584 Voyager DE-STR, Applied Biosystems) and ESI-MS/MS (nanoAcquity /Q-TOF Premier,
585 Waters). Detected peptides were searched against a database containing known and
586 putative gene products of SPP1 and against the nrNCBI database, using the MASCOT
587 (www.matrixscience.com [71]) and FindPept (<http://expasy.org/tools/findpept-ref.html>
588 [72]) softwares. All mass spectrometry analyses were performed by D. Cornu at the
589 Service d'Identification et de Caractérisation des Protéines par Spectrométrie de masse
590 (SICaPS) d'IMAGIF (CNRS, Gif-sur-Yvette).

591

592 **N-terminal sequencing of tail proteins**

593 Samples were separated on 16 % SDS-PAGE gels and transferred as described above into
594 a PVDF membrane, which had been briefly pre-incubated with methanol. After the
595 transfer, the membrane was abundantly rinsed with water and Coomassie stained (0.1 %
596 w/v Coomassie Blue, 40 % v/v ethanol, 1 % v/v acetic acid). After several steps of
597 destaining with 50 % v/v ethanol, the membrane was dried on Whatman paper before
598 cutting out the bands of interest. Amino-terminus sequencing was carried out by J.
599 d'Alayer at the Plateforme d'Analyse et de Microséquençage des Protéines (Institut
600 Pasteur, Paris), using a Procise ABI 494 sequencer (Applied Biosystems).

601

602 **Identification of tail components by western blotting**

603 Aliquots of fractions P3f8 and P2f6S (Figure 1) were concentrated 1.7-fold and 5-fold,
604 respectively. 10 μ L each of the concentrated samples and of disrupted phage particles
605 (2×10^{12} pfu mL^{-1}) were mixed with 5 x sample buffer (45 % v/v glycerol, 0.023 % w/v
606 bromophenol blue, 14 % w/v SDS, 0.18 M Tris-HCl pH 6.8, 1.27 M β -mercaptoethanol)
607 and heated to 100 $^{\circ}\text{C}$ for 5 min. Multiples of these three samples were separated on 15 %
608 SDS-PAGE gels and transferred onto nitrocellulose membranes at 100 V during 2 h 45 min
609 in western blot transfer buffer. Membranes were Ponceau stained, then blocked with 5 %
610 low fat milk powder in PBS with 0.1% Tween 20 (PBS-T) buffer for 1 h. After washing
611 twice for 10 min each with PBS-T, the membrane was cut into stripes, which were
612 incubated for 40 min with a suitable dilution of the respective primary antibodies in PBS-T
613 / 5 % milk. After washing as before, the membranes were incubated with a 10,000-fold
614 dilution of the secondary antibody (goat anti-rabbit, peroxidase-coupled) in PBS-T for 30
615 min. After 6 final washes of 5 min in PBS-T, proteins were detected using ECL western
616 Blotting Reagents (GE Healthcare) and medical X-ray films. DNA-filled capsids and
617 phage particles were disrupted as described [73] to be used as references in Coomassie
618 Blue-stained gels and western blot.

619

620 **Detection of tail proteins in extracts of infected cells**

621 Bacterial cultures of *B. subtilis* YB886 were grown at 37 $^{\circ}\text{C}$ for infection with SPP1 $_{wt}$ or
622 SPP1 $_{sus}$ mutants, and at 46 $^{\circ}\text{C}$ for non-permissive infections with SPP1 $_{ts}$ mutants. For
623 direct comparison, SPP1 $_{wt}$ and SPP1 $_{gp17.1/gp17.1^*}$ control infections were also carried
624 out at 46 $^{\circ}\text{C}$. Cultures were grown in LB medium to an OD_{600} of 0.8 and supplemented
625 with 10 mM CaCl_2 . An aliquot of 2 mL of the non-infected culture was centrifuged for 2.5
626 min at 13,200 rpm and room temperature in an eppendorf 5415D centrifuge. The pellet was
627 briefly re-spun to completely remove all liquid by pipetting and frozen at -20 $^{\circ}\text{C}$.
628 Meanwhile, the culture was infected with SPP1 $_{wt}$ or with the mutant phage of interest at an
629 i.m. of 3 pfu cfu^{-1} . For kinetics of tail protein production, aliquots of 2 mL were taken at
630 various time points after infection and treated as described for the non-infected sample. For
631 comparison of protein production in different mutants, samples were taken shortly before
632 lysis, i.e. 25 min after infection at 37 $^{\circ}\text{C}$ and 20 min after infection at 46 $^{\circ}\text{C}$. Frozen pellets
633 were transferred to ice and resuspended in 100 μ L of resuspension buffer (50 mM Tris-
634 HCl, pH 8, 50 mM glucose, 1 mM EDTA, 1 mg mL^{-1} lysozyme freshly prepared, and

635 protease inhibitors (Roche)), followed by 5 min of incubation at room temperature.
636 Samples were then mixed with 100 μ L each of ice-cold lysis buffer (150 mM NaCl, 1%
637 NP-40, 50 mM Tris-Cl pH 7.5, 5 mM MgCl₂, 1 μ g mL⁻¹ DNase, 100 μ g mL⁻¹ RNase, and
638 protease inhibitors (Roche)), incubated on ice for 30 min and centrifuged for 30 min at
639 14000 g and 4 °C to remove debris. Extracts were either stored at -80 °C or used directly
640 for SDS-PAGE and western Blot, which were done as described for the purified tail
641 samples.

642

643 **Proteins production, purification and antibody production**

644 One liter of BL21 (DE3) (pLysS) (pBT361) was grown in Luria-Bertani broth (LB)
645 supplemented with ampicillin (100 μ g/mL) and chloramphenicol (30 μ g/mL) at 37 °C to
646 an OD_{560 nm} of 0.8. The culture was induced with 5 mM IPTG and harvested 2 h later. The
647 cell pellet was lysed in K-phosphate buffer pH 7.8, 500 mM KCl, 6 M guanidine
648 hydrochloride and denatured tagged gp16.1 was purified by Ni-NTA affinity
649 chromatography following the protocols established for protein (pC)CAT [74].

650 One liter of *E. coli* BL21DE3 (pLysS) (pPT212) that overproduces gp17.5 was grown
651 in Luria-Bertani broth (LB) supplemented with ampicillin (100 μ g/mL) and
652 chloramphenicol (30 μ g/mL) at 30 °C to an OD_{560 nm} of 0.6 and induced with 1 mM IPTG.
653 After 3 h of growth at 30 °C, bacteria were sedimented at 10,000 g for 30 min, pellets were
654 resuspended in 30 mL LB, transferred to a 50 mL Falcon tube, re-sedimented, and the cell
655 pellet was stored frozen at -80°C. The pellet was thawed on ice and resuspended in ice-
656 cold 50 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 25 mM imidazole supplemented with a
657 mixture of antiproteases (Complete EDTA-free, Roche). Bacteria were lysed by 4 cycles of
658 3 min sonication with pauses of 1 min (amplitude 60, 30–40 W in a Vibra Cell sonicator)
659 and centrifuged at 17,000 g for 1 h at 4° C. The supernatant containing soluble gp17.5 was
660 filtered through a 0.22 μ m filter and applied to a 5 mL HisTrapTM column coupled to an
661 Akta system (GE Healthcare) equilibrated with 50 mM Hepes-NaOH pH 7.5, 500 mM
662 NaCl, 25 mM imidazole. After washing with 25 mL of the same buffer, the column was
663 eluted with a 20 mL linear gradient to 50 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 500
664 mM imidazole. Fractions of the gp17.5 peak were collected in tubes containing EDTA to
665 reach a final concentration of 1 mM, pooled, applied to a SuperdexTM 75 26/60 size-
666 exclusion chromatography column equilibrated with 50 mM Hepes-NaOH pH 7.5, 500

667 mM NaCl, 1 mM EDTA and run with the same buffer. Eluted samples containing gp17.5
668 were pooled, supplemented with a cocktail of antiproteases (Complete EDTA-free, Roche),
669 and stored in aliquots at -80 °C.

670 Gp17.5* was produced in *E. coli* BL21DE3 (pLysS) (pPT209) and purified as
671 described above for gp17.5 with the exception that no size exclusion chromatography step
672 was included in the purification procedure.

673 *E. coli* BL21DE3 (pLysS) transformed with pRSETA plasmid derivatives coding
674 different gp18 polypeptides were grown in Luria-Bertani broth (LB) supplemented with
675 ampicillin (100 µg/mL) and chloramphenicol (30 µg/mL) at 37 °C to an OD_{560 nm} of 0.6
676 followed by induction with 1 mM IPTG. After 2 h growth at 37°C, a 1 mL of culture was
677 concentrated 10-fold by centrifugation and 5 µL were loaded on SDS-PAGE gels for
678 Coomassie Blue staining or western blot developed with anti-hexahistidine antibodies and
679 polyclonal serum generated against SPP1 viral particles. Detectable gp18 protein
680 production was found only in *E. coli* BL21DE3 (pLysS) (pPT122) producing gp18₁₈₁₋₄₃₅.
681 This strain was grown at 37°C in 1.5 L of LB broth supplemented with ampicillin (100
682 µg/mL) and chloramphenicol (30 µg/mL) to an OD₆₀₀ of 0.8 and induced with 1 mM
683 IPTG. After 3 h of growth at 37°C bacteria were sedimented at 10,000 g for 30 min,
684 resuspended in 30 mL LB, transferred to 50 mL Falcon tubes, re-sedimented, and the cell
685 pellet was stored frozen at -80° C. The pellet was thawed on ice and resuspended in ice-
686 cold 50 mM Hepes-KOH pH 8.0, 500 mM KCl, 10 mM imidazole supplemented with a
687 mixture of antiproteases (Complete EDTA-free, Roche). Bacteria were lysed by 5 cycles of
688 3 min sonication with pauses of 1 min (amplitude 60, 30–40 W in a Vibra Cell sonicator)
689 and centrifuged at 17,000 g for 45 min at 4°C. The pellet containing insoluble gp18₁₈₁₋₄₃₅
690 was resuspended in 50 mM Hepes-KOH pH 8.0, 500 mM KCl, 10 mM imidazole, 6 M
691 guanidinium hydrochloride and stirred at room temperature for 1 h to solubilize gp18₁₈₁₋₄₃₅.
692 The suspension was centrifuged at 17,000 g for 45 min at 20°C and the supernatant filtered
693 sequentially through 0.45 and 0.22 µm filters. The sample was applied to a 1 mL
694 HisTrap™ column coupled to an Akta system (GE Healthcare) equilibrated with 50 mM
695 Hepes-KOH pH 8.0, 500 mM KCl, 10 mM imidazole, 6 M guanidinium hydrochloride and
696 washed with 10 mL of the same buffer at 16° C. The buffer was exchanged with 10 mL in
697 50 mM Hepes-KOH pH 8.0, 500 mM KCl, 10 mM imidazole, 8 M urea and the protein
698 was eluted with a 20 mL linear gradient to 50 mM Hepes-KOH pH 8.0, 500 mM KCl, 500

699 mM imidazole, 8 M urea. Fractions containing gp18₁₈₁₋₄₃₅ were pooled, concentrated by
700 ultrafiltration in a Vivaspin device (10 kDa cut-off) and dialysed sequentially against 50
701 mM Hepes-KOH pH 8.0, 500 mM KCl, 6 M urea and against the same solution with 5 M
702 urea. Aliquots of denatured gp18₁₈₁₋₄₃₅ were stored at -80°C.

703 *E. coli* JS218 freshly transformed with plasmid pIA29 was grown at 37°C in LB
704 supplemented with erythromycin (30 µg/mL) and chloramphenicol (10 µg/mL) to an OD₆₀₀
705 of 0.35-0.40. Cultures were transferred to 27°C and induced with 2 mM IPTG when the
706 OD₆₀₀ reached 0.50-0.55 and incubation was continued overnight. Pelleted bacteria were
707 resuspended in cold lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Triton
708 X100, 10 mM MgSO₄) supplemented with a mixture of antiproteases (Complete EDTA-
709 free, Roche) and disrupted by performing three bursts of 1 min (amplitude 60, pulse 3, 30–
710 40 W) in a sonicator (Vibra Cell) with pauses of 1 min. The crude extract was centrifuged
711 at 4°C twice for 20 minutes at 27000 x g. Imidazole was then added to a final
712 concentration of 20 mM and recombinant hexahistidine gp19.1 protein was purified by
713 metal-chelating chromatography using a HiTrap column (GE Healthcare) following the
714 supplier conditions (NaH₂PO₄ 20 mM pH 7.4, NaCl 500 mM). The column was washed
715 with increasing concentrations of imidazole. Gp19.1 was eluted at 100 mM imidazole,
716 concentrated and further purified by size exclusion chromatography on a Superdex 75
717 column (GE Healthcare) run in 100 mM NaCl, 20 mM Tris-HCl, pH 7.5.

718 Gp22, gp23 and gp23.1 fused to a TEV-cleavable hexahistidine-thioredoxin tag were
719 overproduced in *E. coli* and purified by nickel-affinity chromatography, TEV proteolysis
720 to remove the tag, and size exclusion chromatography as previously described [27,69,70].

721 Gp24 was produced from pIA39 following the protocol used for gp19.1 with some
722 modifications. Induction was performed at 37°C instead of 27°C for 3 hours with 2 mM
723 IPTG when the OD₆₀₀ reached 0.6. After an identical treatment of the bacterial pellet, the
724 crude extract was directly loaded on a HiTrap column, without addition of imidazole.
725 Gp24 was eluted at 500 mM, concentrated and loaded on a Superdex 75 column (GE
726 Healthcare) run in 100 mM NaCl, 20 mM Tris-HCl, pH 7.5.

727 Polyclonal antibodies were produced by immunisation of rabbits with purified soluble
728 (gp17.5, gp17.5*, gp19.1, gp22, gp23, gp23.1 and gp24) and denatured (gp16.1 and
729 fragment gp18₁₈₁₋₄₃₅) proteins as described [75] at the INAF platform (Gif-sur-Yvette,
730 France) and at FZB Biotechnik GmbH (Berlin, Germany). Sera against gp17, gp17.1Nter,

731 gp21Nter and gp21Cter were previously described [4,29,33]. The IgG fraction of anti-
732 gp22 serum was purified by affinity chromatography in a Hi-Trap Protein A HP (GE
733 Healthcare) as described by the suppliers.

734

735 **Electron microscopy and immunolabeling**

736 Tail samples were negatively stained with 2 % uranyl acetate before imaging by electron
737 microscopy [76].

738 Tails P2f6 and P3f8 in TBT buffer were split in two identical samples of 25 μ L. One
739 set of samples was fixed with glutaraldehyde (Sigma-Aldrich) at a final concentration of
740 0.2 % for 10 min at room temperature. The reaction was quenched with 0.1 M
741 glycylglycine for 30 min at room temperature. Control samples were processed identically
742 with the exception that the fixative was replaced by an identical volume of water. All
743 samples were then labeled with purified anti-gp22 IgG overnight at room temperature. The
744 volume was adjusted to 60 μ L and loaded on a microspin column filled with 400 μ L
745 Sephacryl 500 slurry equilibrated with buffer TBT. The column was centrifuged for 5 min
746 at 800 g to recover the tails and immune-complexes in the flow through while unbound
747 IgG was retained in the size-exclusion matrix. An aliquot of each sample was used for
748 negative staining with 2% uranyl acetate and visualized by electron microscopy. The
749 remaining sample was incubated overnight with 1 μ L of goat anti-rabbit IgG complexed
750 with 5 nm colloidal gold (BritishBioCell) and prepared for electron microscopy
751 observation as above.

752

753 **Bioinformatics**

754 SPP1 tail sequence-related proteins were detected by pBlast [77] (e-value ≤ 0.0001) and
755 filtered through CD-Hit [78] to identify proteins with less than 80 % identity through 80%
756 of their amino acid sequence relative to the shortest sequence. Sequences were then aligned
757 using Clustal Omega set to default parameters [79]. Residues conservation was color-
758 coded using the UGENE program default function [80]. Gp21 and gp22 pBlast results were
759 used unfiltered for the analysis in Figure S3. Neighbourhood of their coding genes was
760 investigated by manual inspection of their coding genomes. Guide trees were computed
761 from pBlast sequence alignments using Clustal Omega and displayed with UGENE [80].

762 Homology search combined with structure prediction was carried out using HHpred [81].
763 Prediction of transmembrane segments was carried using TMHMM.2 set to default
764 parameters [82].

765 Protein structures were analysed and displayed with Pymol.

766

767 **CRedit authorship contribution statement**

768 **Anait Seul:** Writing - original draft, Conceptualization, Methodology - tails and tail sub-structures
769 purification methods/tails characterization/tail proteins fate during infection, Investigation, Visualization.
770 **Sandrine Brasilès:** Methodology - protein purification methods, Resources, Investigation. **Isabelle**
771 **Petitpas:** Methodology - protein purification methods, Resources, Investigation. **Rudi Lurz:** Investigation,
772 Visualization. **Valérie Campanacci:** Methodology - protein purification methods, Resources. **Christian**
773 **Cambillau:** Writing - review & editing, Supervision, Project administration, Funding acquisition. **Frank**
774 **Weise:** Writing - review & editing, Methodology - protein purification methods, Resources, Investigation.
775 **Mohamed Zairi:** Investigation, Formal analysis, Visualization. **Paulo Tavares:** Writing - original draft,
776 Conceptualization, Resources, Investigation, Visualization, Supervision, Project administration, Funding
777 acquisition. **Isabelle Auzat:** Writing - original draft, Conceptualization, Methodology - protein purification
778 methods/tail proteins fate during infection, Resources, Investigation, Visualization, Supervision, Project
779 administration.
780

781

782 **Data and Materials Availability Statement**

783 Experimental data of this manuscript have been contained within this manuscript or are
784 available upon request from P.T. (paulo.tavares@i2bc.paris-saclay.fr) or I.A.
785 (isabelle.auzat@i2bc.paris-saclay.fr). Biological materials are available upon request from
786 P.T or I.A.

787

788 **Acknowledgments**

789 We thank Marie-Christine Vaney (Institut Pasteur, Paris) for Figure S4A-D and useful
790 discussions. David Cornu (SICaPS, I2BC) is acknowledged for mass spectrometry
791 measurements. This work has benefited from the facilities and expertise of the I2BC
792 proteomic platform (Proteomic-Gif, SICaPS), supported by IBiSA, Ile de France Region,
793 Plan Cancer, CNRS and Paris-Saclay University. We thank Jacques d'Alayer (Plate-forme
794 d'Analyse et de Microséquençage des Protéines, Institut Pasteur, Paris) for proteins N-
795 terminal sequencing.

796

797 **Funding**

798 The work was supported by institutional funding from CNRS and by ANR (grant
799 Siphophages to C.C. and P.T.).

800

801

802 **Declaration of Competing Interest**

803 The authors declare that they have no known competing financial interests or personal
804 relationships that could have appeared to influence the work reported in this paper.

805

806

807 **Appendix A. Supplemental data**

808 Supplemental data to this article can be found online at xxx.

809 **Table S1**, Fingerprint of SPP1 proteins studied in this work, PDF file, 0.1 MB

810 **Table S2**, Protein coding genes studied in this work and their conditional lethal mutants.

811 See Figure 4 for a graphical representation, PDF file, 0.1 MB

812 **Table S3**, Primers used in this work, PDF file, 0.1 MB

813 **Figure S1**, Mass spectrometry of tryptic peptides from SPP1 tail proteins. PDF file, 0.2
814 MB

815 **Figure S2**, SPP1 gp18 putative transmembrane segments and sequence alignment with
816 homologous proteins. PNG file, 2.2 MB

817 **Figure S3**, SPP1 gp22 homologs are co-encoded with gp21 homologs in a sub-set of
818 genomes that code for gp21 homologs. PNG file, 8.5 MB

819 **Figure S4**, SPP1 gp19.1 domain structure and its homologous proteins. PNG file, 4 MB

820 **Figure S5**, Partial purification of tail protein complexes. PNG file, 0.4 MB

821 **Figure S6**, Sedimentation of SPP1 tails produced during non-permissive infection with
822 SPP1*gp16.1*. PDF file, 0.4 MB

823

824

825

826

References

- 827 [1] C.A. Roessner, D.K. Struck, G.M. Ihler, Injection of DNA into liposomes by
828 bacteriophage λ , *J. Biol. Chem.* 258 (1983) 643–648.
- 829 [2] J. Böhm, O. Lambert, A.S. Frangakis, L. Letellier, W. Baumeister, J.L. Rigaud, FhuA-
830 mediated phage genome transfer into liposomes, *Curr. Biol.* 11 (2001) 1168–1175.
831 [https://doi.org/10.1016/S0960-9822\(01\)00349-9](https://doi.org/10.1016/S0960-9822(01)00349-9).
- 832 [3] C. São-José, S. Lhuillier, R. Lurz, R. Melki, J. Lepault, M.A. Santos, P. Tavares, The
833 ectodomain of the viral receptor YueB forms a fiber that triggers ejection of
834 bacteriophage SPP1 DNA, *J. Biol. Chem.* 281 (2006) 11464–11470.
835 <https://doi.org/10.1074/jbc.M513625200>.
- 836 [4] I. Vinga, C. Baptista, I. Auzat, I. Petipas, R. Lurz, P. Tavares, M.A. Santos, C. São-José,
837 Role of bacteriophage SPP1 tail spike protein gp21 on host cell receptor binding and
838 trigger of phage DNA ejection, *Mol. Microbiol.* 83 (2012) 289–303.
839 <https://doi.org/10.1111/j.1365-2958.2011.07931.x>.
- 840 [5] A.R. Davidson, L. Cardarelli, L.G. Pell, D.R. Radford, K.L. Maxwell, Long
841 Noncontractile Tail Machines of Bacteriophages, *Adv Exp Med Biol.* 726 (2012) 115–
842 142. https://doi.org/10.1007/978-1-4614-0980-9_6.
- 843 [6] B. Hu, W. Margolin, I.J. Molineux, J. Liu, The Bacteriophage T7 Virion Undergoes
844 Extensive Structural Remodeling During Infection, *Science* (80-.). 339 (2013) 576–579.
845 <https://doi.org/10.1126/science.1231887>.
- 846 [7] B. Hu, W. Margolin, I.J. Molineux, J. Liu, Structural remodeling of bacteriophage T4
847 and host membranes during infection initiation, *Proc. Natl. Acad. Sci.* 112 (2015)
848 E4919–E4928. <https://doi.org/10.1073/pnas.1501064112>.
- 849 [8] C.-A. Arnaud, G. Effantin, C. Vivès, S. Engilberge, M. Bacia, P. Boulanger, E. Girard,
850 G. Schoehn, C. Breyton, Bacteriophage T5 tail tube structure suggests a trigger
851 mechanism for Siphoviridae DNA ejection, *Nat. Commun.* 8 (2017) 1953.
852 <https://doi.org/10.1038/s41467-017-02049-3>.
- 853 [9] R. Linares, C.-A. Arnaud, S. Degroux, G. Schoehn, C. Breyton, Structure, function and
854 assembly of the long, flexible tail of siphophages, *Curr. Opin. Virol.* 45 (2020) 34–42.
855 <https://doi.org/10.1016/j.coviro.2020.06.010>.
- 856 [10] P.G. Leiman, M. Basler, U.A. Ramagopal, J.B. Bonanno, J.M. Sauder, S. Pukatzki, S.K.
857 Burley, S.C. Almo, J.J. Mekalanos, Type VI secretion apparatus and phage tail-
858 associated protein complexes share a common evolutionary origin, *Proc. Natl. Acad. Sci.*
859 106 (2009) 4154–4159. <https://doi.org/10.1073/pnas.0813360106>.
- 860 [11] L.G. Pell, V. Kanelis, L.W. Donaldson, P. Lynne Howell, A.R. Davidson, The phage
861 major tail protein structure reveals a common evolution for long-tailed phages and the
862 type VI bacterial secretion system, *Proc. Natl. Acad. Sci.* 106 (2009) 4160–4165.
863 <https://doi.org/10.1073/pnas.0900044106>.
- 864 [12] G. Bönemann, A. Pietrosiuk, A. Mogk, Tubules and donuts: a type VI secretion story,
865 *Mol. Microbiol.* 76 (2010) 815–821. <https://doi.org/10.1111/j.1365-2958.2010.07171.x>.
- 866 [13] A. Zoued, E. Durand, Y.R. Brunet, S. Spinelli, B. Douzi, M. Guzzo, N. Flaugnatti, P.
867 Legrand, L. Journet, R. Fronzes, T. Mignot, C. Cambillau, E. Cascales, Priming and

- 868 polymerization of a bacterial contractile tail structure, *Nature*. 531 (2016) 59–63.
869 <https://doi.org/10.1038/nature17182>.
- 870 [14] D. Böck, J.M. Medeiros, H.-F. Tsao, T. Penz, G.L. Weiss, K. Aistleitner, M. Horn, M.
871 Pilhofer, In situ architecture, function, and evolution of a contractile injection system,
872 *Science* (80-.). 357 (2017) 713–717. <https://doi.org/10.1126/science.aan7904>.
- 873 [15] A. Goulet, S. Spinelli, J. Mahony, C. Cambillau, Conserved and Diverse Traits of
874 Adhesion Devices from Siphoviridae Recognizing Proteinaceous or Saccharidic
875 Receptors, *Viruses*. 12 (2020) 512. <https://doi.org/10.3390/v12050512>.
- 876 [16] H. Ackermann, Phage Classification and Characterization, in: *Methods Mol. Biol.*, 2009:
877 pp. 127–140. https://doi.org/10.1007/978-1-60327-164-6_13.
- 878 [17] M. Krupovic, V. Cvirkaite-Krupovic, J. Iranzo, D. Prangishvili, E. V. Koonin, Viruses of
879 archaea: Structural, functional, environmental and evolutionary genomics, *Virus Res*.
880 244 (2018) 181–193. <https://doi.org/10.1016/j.virusres.2017.11.025>.
- 881 [18] P. Tavares, The Bacteriophage Head-to-Tail Interface, in: J.R. Harris, D. Bhella (Eds.),
882 *Subcell. Biochem.*, Springer, 2018: pp. 305–328. [https://doi.org/10.1007/978-981-10-](https://doi.org/10.1007/978-981-10-8456-0_14)
883 [8456-0_14](https://doi.org/10.1007/978-981-10-8456-0_14).
- 884 [19] C. Plisson, H.E. White, I. Auzat, A. Zafarani, C. São-José, S. Lhuillier, P. Tavares, E. V.
885 Orlova, Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection, *EMBO J*.
886 26 (2007) 3720–3728. <https://doi.org/10.1038/sj.emboj.7601786>.
- 887 [20] G. Sciara, C. Bebeacua, P. Bron, D. Tremblay, M. Ortiz-Lombardia, J. Lichiere, M. van
888 Heel, V. Campanacci, S. Moineau, C. Cambillau, Structure of lactococcal phage p2
889 baseplate and its mechanism of activation, *Proc. Natl. Acad. Sci.* 107 (2010) 6852–6857.
890 <https://doi.org/10.1073/pnas.1000232107>.
- 891 [21] C. Bebeacua, P. Bron, L. Lai, C.S. Vegge, L. Brøndsted, S. Spinelli, V. Campanacci, D.
892 Veesler, M. van Heel, C. Cambillau, Structure and Molecular Assignment of
893 Lactococcal Phage TP901-1 Baseplate, *J. Biol. Chem.* 285 (2010) 39079–39086.
894 <https://doi.org/10.1074/jbc.M110.175646>.
- 895 [22] D. Veesler, C. Cambillau, A Common Evolutionary Origin for Tailed-Bacteriophage
896 Functional Modules and Bacterial Machineries, *Microbiol. Mol. Biol. Rev.* 75 (2011)
897 423–433. <https://doi.org/10.1128/MMBR.00014-11>.
- 898 [23] D. Veesler, S. Spinelli, J. Mahony, J. Lichiere, S. Blangy, G. Bricogne, P. Legrand, M.
899 Ortiz-Lombardia, V. Campanacci, D. van Sinderen, C. Cambillau, Structure of the phage
900 TP901-1 1.8 MDa baseplate suggests an alternative host adhesion mechanism, *Proc.*
901 *Natl. Acad. Sci.* 109 (2012) 8954–8958. <https://doi.org/10.1073/pnas.1200966109>.
- 902 [24] N.M.I. Taylor, N.S. Prokhorov, R.C. Guerrero-Ferreira, M.M. Shneider, C. Browning,
903 K.N. Goldie, H. Stahlberg, P.G. Leiman, Structure of the T4 baseplate and its function in
904 triggering sheath contraction, *Nature*. 533 (2016) 346–352.
905 <https://doi.org/10.1038/nature17971>.
- 906 [25] S. Mc Grath, H. Neve, J.F.M.L. Seegers, R. Eijlander, C.S. Vegge, L. Brøndsted, K.J.
907 Heller, G.F. Fitzgerald, F.K. Vogensen, D. van Sinderen, Anatomy of a Lactococcal
908 Phage Tail, *J. Bacteriol.* 188 (2006) 3972–3982. <https://doi.org/10.1128/JB.00024-06>.
- 909 [26] S.R. Stockdale, B. Collins, S. Spinelli, F.P. Douillard, J. Mahony, C. Cambillau, D. van
910 Sinderen, Structure and Assembly of TP901-1 Virion Unveiled by Mutagenesis, *PLoS*

- 911 One. 10 (2015) e0131676. <https://doi.org/10.1371/journal.pone.0131676>.
- 912 [27] D. Veessler, G. Robin, J. Lichière, I. Auzat, P. Tavares, P. Bron, V. Campanacci, C.
913 Cambillau, Crystal structure of bacteriophage SPP1 distal tail protein (gp19.1): A
914 baseplate hub paradigm in gram-positive infecting phages, *J. Biol. Chem.* 285 (2010)
915 36666–36673. <https://doi.org/10.1074/jbc.M110.157529>.
- 916 [28] A. Goulet, J. Lai-Kee-Him, D. Veessler, I. Auzat, G. Robin, D.A. Shepherd, A.E.
917 Ashcroft, E. Richard, J. Lichière, P. Tavares, C. Cambillau, P. Bron, The Opening of the
918 SPP1 Bacteriophage Tail, a Prevalent Mechanism in Gram-positive-infecting
919 Siphophages, *J. Biol. Chem.* 286 (2011) 25397–25405.
920 <https://doi.org/10.1074/jbc.M111.243360>.
- 921 [29] I. Auzat, A. Dröge, F. Weise, R. Lurz, P. Tavares, Origin and function of the two major
922 tail proteins of bacteriophage SPP1, *Mol. Microbiol.* 70 (2008) 557–569.
923 <https://doi.org/10.1111/j.1365-2958.2008.06435.x>.
- 924 [30] C. Langlois, S. Ramboarina, A. Cukkemane, I. Auzat, B. Chagot, B. Gilquin, A.
925 Ignatiou, I. Petitpas, E. Kasotakis, M. Paternostre, H.E. White, E. V. Orlova, M. Baldus,
926 P. Tavares, S. Zinn-Justin, Bacteriophage SPP1 tail tube protein self-assembles into β -
927 structure-rich tubes, *J. Biol. Chem.* 290 (2015) 3836–3849.
928 <https://doi.org/10.1074/jbc.M114.613166>.
- 929 [31] M. Zinke, K.A.A. Sachowsky, C. Öster, S. Zinn-Justin, R. Ravelli, G.F. Schröder, M.
930 Habeck, A. Lange, Architecture of the flexible tail tube of bacteriophage SPP1, *Nat.*
931 *Commun.* 11 (2020) 5759. <https://doi.org/10.1038/s41467-020-19611-1>.
- 932 [32] B. Chagot, I. Auzat, M. Gallopin, I. Petitpas, B. Gilquin, P. Tavares, S. Zinn-Justin,
933 Solution structure of gp17 from the Siphoviridae bacteriophage SPP1: Insights into its
934 role in virion assembly, *Proteins Struct. Funct. Bioinforma.* 80 (2012) 319–326.
935 <https://doi.org/10.1002/prot.23191>.
- 936 [33] I. Auzat, I. Petitpas, R. Lurz, F. Weise, P. Tavares, A touch of glue to complete
937 bacteriophage assembly: The tail-to-head joining protein (THJP) family, *Mol. Microbiol.*
938 91 (2014) 1164–1178. <https://doi.org/10.1111/mmi.12526>.
- 939 [34] B. Becker, N. De la Fuente, M. Gassel, D. Günther, P. Tavares, R. Lurz, T.A. Trautner,
940 J.C. Alonso, Head morphogenesis genes of the *Bacillus subtilis* bacteriophage SPP1, *J.*
941 *Mol. Biol.* 268 (1997) 822–839. <https://doi.org/10.1006/jmbi.1997.0997>.
- 942 [35] A. Rath, M. Glibowicka, V.G. Nadeau, G. Chen, C.M. Deber, Detergent binding
943 explains anomalous SDS-PAGE migration of membrane proteins, *Proc. Natl. Acad. Sci.*
944 106 (2009) 1760–1765. <https://doi.org/10.1073/pnas.0813167106>.
- 945 [36] B. Behrens, G. Lüder, M. Behncke, T.A. Trautner, A.T. Ganesan, The genome of *B.*
946 *subtilis* phage SPP1, *Mol. Gen. Genet. MGG.* 175 (1979) 351–357.
947 <https://doi.org/10.1007/BF00397235>.
- 948 [37] J. Xu, R.W. Hendrix, R.L. Duda, Conserved Translational Frameshift in dsDNA
949 Bacteriophage Tail Assembly Genes, *Mol. Cell.* 16 (2004) 11–21.
950 <https://doi.org/10.1016/j.molcel.2004.09.006>.
- 951 [38] S. Fernandes, C. São-José, Probing the function of the two holin-like proteins of
952 bacteriophage SPP1, *Virology.* 500 (2017) 184–189.
953 <https://doi.org/10.1016/j.virol.2016.10.030>.

- 954 [39] L.M. Godinho, M.E.S. Fadel, C. Monniot, L. Jakutyte, I. Auzat, A. Labarde, K. Djacem,
955 L. Oliveira, R. Carballido-Lopez, S. Ayora, P. Tavares, The revisited genome of bacillus
956 subtilis bacteriophage SPP1, *Viruses*. 10 (2018). <https://doi.org/10.3390/v10120705>.
- 957 [40] D. Veessler, G. Robin, J. Lichière, I. Auzat, P. Tavares, P. Bron, V. Campanacci, C.
958 Cambillau, Crystal Structure of Bacteriophage SPP1 Distal Tail Protein (gp19.1), *J. Biol.*
959 *Chem.* 285 (2010) 36666–36673. <https://doi.org/10.1074/jbc.M110.157529>.
- 960 [41] A. Lopes, P. Tavares, M.-A. Petit, R. Guérois, S. Zinn-Justin, Automated classification
961 of tailed bacteriophages according to their neck organization, *BMC Genomics*. 15 (2014)
962 1027. <https://doi.org/10.1186/1471-2164-15-1027>.
- 963 [42] I. Katsura, Determination of bacteriophage λ tail length by a protein ruler, *Nature*. 327
964 (1987) 73–75. <https://doi.org/10.1038/327073a0>.
- 965 [43] N.K. Abuladze, M. Gingery, J. Tsai, F.A. Eiserling, Tail Length Determination in
966 Bacteriophage T4, *Virology*. 199 (1994) 301–310.
967 <https://doi.org/10.1006/viro.1994.1128>.
- 968 [44] J. Mahony, M. Alqarni, S. Stockdale, S. Spinelli, M. Feyereisen, C. Cambillau, D. van
969 Sinderen, Functional and structural dissection of the tape measure protein of lactococcal
970 phage TP901-1, *Sci. Rep.* 6 (2016) 36667. <https://doi.org/10.1038/srep36667>.
- 971 [45] S. Hayes, R. Vincentelli, J. Mahony, A. Nauta, L. Ramond, G.A. Lugli, M. Ventura, D.
972 van Sinderen, C. Cambillau, Functional carbohydrate binding modules identified in
973 evolved dits from siphophages infecting various Gram-positive bacteria, *Mol. Microbiol.*
974 110 (2018) 777–795. <https://doi.org/10.1111/mmi.14124>.
- 975 [46] M.-E. Dieterle, S. Spinelli, I. Sadovskaya, M. Piuri, C. Cambillau, Evolved distal tail
976 carbohydrate binding modules of *L. actobacillus* phage J-1: a novel type of anti-receptor
977 widespread among lactic acid bacteria phages, *Mol. Microbiol.* 104 (2017) 608–620.
978 <https://doi.org/10.1111/mmi.13649>.
- 979 [47] J.L. Kizziah, K.A. Manning, A.D. Dearborn, T. Dokland, Structure of the host cell
980 recognition and penetration machinery of a *Staphylococcus aureus* bacteriophage, *PLOS*
981 *Pathog.* 16 (2020) e1008314. <https://doi.org/10.1371/journal.ppat.1008314>.
- 982 [48] E. Cascales, Microbiology: And Amoebophilus Invented the Machine Gun!, *Curr. Biol.*
983 27 (2017) R1170–R1173. <https://doi.org/10.1016/j.cub.2017.09.025>.
- 984 [49] S. Kanamaru, P.G. Leiman, V.A. Kostyuchenko, P.R. Chipman, V. V. Mesyanzhinov, F.
985 Arisaka, M.G. Rossmann, Structure of the cell-puncturing device of bacteriophage T4,
986 *Nature*. 415 (2002) 553–557. <https://doi.org/10.1038/415553a>.
- 987 [50] R. Hendrix, R. Duda, Bacteriophage lambda PaPa: not the mother of all lambda phages,
988 *Science* (80-.). 258 (1992) 1145–1148. <https://doi.org/10.1126/science.1439823>.
- 989 [51] Y. Zivanovic, F. Confalonieri, L. Ponchon, R. Lurz, M. Chami, A. Flayhan, M.
990 Renouard, A. Huet, P. Decottignies, A.R. Davidson, C. Breyton, P. Boulanger, Insights
991 into Bacteriophage T5 Structure from Analysis of Its Morphogenesis Genes and Protein
992 Components, *J. Virol.* 88 (2014) 1162–1174. <https://doi.org/10.1128/JVI.02262-13>.
- 993 [52] I. Katsura, P.W. Kühl, Morphogenesis of the tail of bacteriophage lambda, *J. Mol. Biol.*
994 91 (1975) 257–273. [https://doi.org/10.1016/0022-2836\(75\)90379-4](https://doi.org/10.1016/0022-2836(75)90379-4).
- 995 [53] L.-C. Tsui, R.W. Hendrix, Proteolytic processing of phage λ tail protein gpH: timing of
996 the cleavage, *Virology*. 125 (1983) 257–264. <https://doi.org/10.1016/0042->

- 997 6822(83)90199-X.
- 998 [54] J. Xu, R.W. Hendrix, R.L. Duda, A Balanced Ratio of Proteins from Gene G and
999 Frameshift-Extended Gene GT Is Required for Phage Lambda Tail Assembly, *J. Mol.*
1000 *Biol.* 425 (2013) 3476–3487. <https://doi.org/10.1016/j.jmb.2013.07.002>.
- 1001 [55] J. Xu, R.W. Hendrix, R.L. Duda, Chaperone–Protein Interactions That Mediate
1002 Assembly of the Bacteriophage Lambda Tail to the Correct Length, *J. Mol. Biol.* 426
1003 (2014) 1004–1018. <https://doi.org/10.1016/j.jmb.2013.06.040>.
- 1004 [56] J.A. Lengyel, R.N. Goldstein, M. Marsh, R. Calendar, Structure of the bacteriophage P2
1005 tail, *Virology.* 62 (1974) 161–174. [https://doi.org/10.1016/0042-6822\(74\)90312-2](https://doi.org/10.1016/0042-6822(74)90312-2).
- 1006 [57] F.J. Grundy, M.M. Howe, Morphogenetic structures present in lysates of amber mutants
1007 of bacteriophage Mu, *Virology.* 143 (1985) 485–504. [https://doi.org/10.1016/0042-](https://doi.org/10.1016/0042-6822(85)90388-5)
1008 [6822\(85\)90388-5](https://doi.org/10.1016/0042-6822(85)90388-5).
- 1009 [58] Y. Chaban, R. Lurz, S. Brasilès, C. Cornilleau, M. Karreman, S. Zinn-Justin, P. Tavares,
1010 E. V. Orlova, Structural rearrangements in the phage head-to-tail interface during
1011 assembly and infection, *Proc. Natl. Acad. Sci. U. S. A.* 112 (2015) 7009–7014.
1012 <https://doi.org/10.1073/pnas.1504039112>.
- 1013 [59] S. Spinelli, D. Veessler, C. Bebeacua, C. Cambillau, Structures and host-adhesion
1014 mechanisms of lactococcal siphophages, *Front. Microbiol.* 5 (2014).
1015 <https://doi.org/10.3389/fmicb.2014.00003>.
- 1016 [60] J.S. Fraser, K.L. Maxwell, A.R. Davidson, Immunoglobulin-like domains on
1017 bacteriophage: weapons of modest damage?, *Curr. Opin. Microbiol.* 10 (2007) 382–387.
1018 <https://doi.org/10.1016/j.mib.2007.05.018>.
- 1019 [61] P. Boulanger, P. Jacquot, L. Plançon, M. Chami, A. Engel, C. Parquet, C. Herbeuval, L.
1020 Letellier, Phage T5 Straight Tail Fiber Is a Multifunctional Protein Acting as a Tape
1021 Measure and Carrying Fusogenic and Muralytic Activities, *J. Biol. Chem.* 283 (2008)
1022 13556–13564. <https://doi.org/10.1074/jbc.M800052200>.
- 1023 [62] M. Piuri, G.F. Hatfull, A peptidoglycan hydrolase motif within the mycobacteriophage
1024 TM4 tape measure protein promotes efficient infection of stationary phase cells, *Mol.*
1025 *Microbiol.* 62 (2006) 1569–1585. <https://doi.org/10.1111/j.1365-2958.2006.05473.x>.
- 1026 [63] S. Riva, M. Polsinelli, A. Falaschi, A new phage of *Bacillus subtilis* with infectious
1027 DNA having separable strands, *J. Mol. Biol.* 35 (1968) 347–356.
1028 [https://doi.org/10.1016/S0022-2836\(68\)80029-4](https://doi.org/10.1016/S0022-2836(68)80029-4).
- 1029 [64] S. Okubo, T. Yanagida, Isolation of a suppressor mutant in *Bacillus subtilis.*, *J.*
1030 *Bacteriol.* 95 (1968) 1187–1188. <https://doi.org/10.1128/JB.95.3.1187-1188.1968>.
- 1031 [65] A. Dröge, P. Tavares, In vitro Packaging of DNA of the *Bacillus subtilis* bacteriophage
1032 SPP1 1 Edited by J. Karn, *J. Mol. Biol.* 296 (2000) 103–115.
1033 <https://doi.org/10.1006/jmbi.1999.3449>.
- 1034 [66] R.E. Yasbin, P.I. Fields, B.J. Andersen, Properties of *Bacillus subtilis* 168 derivatives
1035 freed of their natural prophages, *Gene.* 12 (1980) 155–159. [https://doi.org/10.1016/0378-](https://doi.org/10.1016/0378-1119(80)90026-8)
1036 [1119\(80\)90026-8](https://doi.org/10.1016/0378-1119(80)90026-8).
- 1037 [67] P. Haima, S. Bron, G. Venema, The effect of restriction on shotgun cloning and plasmid
1038 stability in *Bacillus subtilis* Marburg, *Mol. Gen. Genet. MGG.* 209 (1987) 335–342.
1039 <https://doi.org/10.1007/BF00329663>.

- 1040 [68] A. Isidro, A.O. Henriques, P. Tavares, The portal protein plays essential roles at different
1041 steps of the SPP1 DNA packaging process, *Virology*. 322 (2004) 253–263.
1042 <https://doi.org/10.1016/j.virol.2004.02.012>.
- 1043 [69] D. Veesler, S. Blangy, S. Spinelli, P. Tavares, V. Campanacci, C. Cambillau, Crystal
1044 structure of *Bacillus subtilis* SPP1 phage gp22 shares fold similarity with a domain of
1045 lactococcal phage p2 RBP, *Protein Sci.* 19 (2010) 1439–1443.
1046 <https://doi.org/10.1002/pro.416>.
- 1047 [70] D. Veesler, S. Blangy, J. Lichière, M. Ortiz-Lombardía, P. Tavares, V. Campanacci, C.
1048 Cambillau, Crystal structure of *Bacillus subtilis* SPP1 phage gp23.1, a putative
1049 chaperone, *Protein Sci.* 19 (2010) 1812–1816. <https://doi.org/10.1002/pro.464>.
- 1050 [71] D.N. Perkins, D.J.C. Pappin, D.M. Creasy, J.S. Cottrell, Probability-based protein
1051 identification by searching sequence databases using mass spectrometry data,
1052 *Electrophoresis*. 20 (1999) 3551–3567. [https://doi.org/10.1002/\(SICI\)1522-
1053 2683\(19991201\)20:18<3551::AID-ELPS3551>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2).
- 1054 [72] A. Gattiker, W. V. Bienvenut, A. Bairoch, E. Gasteiger, FindPept, a tool to identify
1055 unmatched masses in peptide mass fingerprinting protein identification, *Proteomics*. 2
1056 (2002) 1435–1444. [https://doi.org/10.1002/1615-9861\(200210\)2:10<1435::AID-
1057 PROT1435>3.0.CO;2-9](https://doi.org/10.1002/1615-9861(200210)2:10<1435::AID-PROT1435>3.0.CO;2-9).
- 1058 [73] I. Vinga, A. Dröge, A.C. Stiege, R. Lurz, M.A. Santos, R. Daugelavičius, P. Tavares,
1059 The minor capsid protein gp7 of bacteriophage SPP1 is required for efficient infection of
1060 *Bacillus subtilis*, *Mol. Microbiol.* 61 (2006) 1609–1621. [https://doi.org/10.1111/j.1365-
1061 2958.2006.05327.x](https://doi.org/10.1111/j.1365-2958.2006.05327.x).
- 1062 [74] A. Isidro, M.A. Santos, A.O. Henriques, P. Tavares, The high-resolution functional map
1063 of bacteriophage SPP1 portal protein, *Mol. Microbiol.* 51 (2004) 949–962.
1064 <https://doi.org/10.1046/j.1365-2958.2003.03880.x>.
- 1065 [75] E. Harlow, D. Lane, *Antibodies: a Laboratory Manual*, Cold Spring Harbor Laboratory
1066 Press, 1988.
- 1067 [76] A.C. Steven, B.L. Trus, J.V. Maizel, M. Unser, D.A.D. Parry, J.S. Wall, J.F. Hainfeld,
1068 F.W. Studier, Molecular substructure of a viral receptor-recognition protein, *J. Mol.*
1069 *Biol.* 200 (1988) 351–365. [https://doi.org/10.1016/0022-2836\(88\)90246-X](https://doi.org/10.1016/0022-2836(88)90246-X).
- 1070 [77] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment
1071 search tool, *J. Mol. Biol.* 215 (1990) 403–410. [https://doi.org/10.1016/S0022-
1072 2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- 1073 [78] L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: accelerated for clustering the next-
1074 generation sequencing data, *Bioinformatics*. 28 (2012) 3150–3152.
1075 <https://doi.org/10.1093/bioinformatics/bts565>.
- 1076 [79] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H.
1077 McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, Fast, scalable
1078 generation of high-quality protein multiple sequence alignments using Clustal Omega,
1079 *Mol. Syst. Biol.* 7 (2011) 539. <https://doi.org/10.1038/msb.2011.75>.
- 1080 [80] K. Okonechnikov, O. Golosova, M. Fursov, Unipro UGENE: a unified bioinformatics
1081 toolkit, *Bioinformatics*. 28 (2012) 1166–1167.
1082 <https://doi.org/10.1093/bioinformatics/bts091>.

- 1083 [81] L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J.
1084 Söding, A.N. Lupas, V. Alva, A Completely Reimplemented MPI Bioinformatics
1085 Toolkit with a New HHpred Server at its Core, *J. Mol. Biol.* 430 (2018) 2237–2243.
1086 <https://doi.org/10.1016/j.jmb.2017.12.007>.
- 1087 [82] A. Krogh, B. Larsson, G. von Heijne, E.L.. Sonnhammer, Predicting transmembrane
1088 protein topology with a hidden markov model: application to complete genomes¹¹Edited
1089 by F. Cohen, *J. Mol. Biol.* 305 (2001) 567–580. <https://doi.org/10.1006/jmbi.2000.4315>.
- 1090 [83] B. Behrens, G. Lüder, M. Behncke, T.A. Trautner, A.T. Ganesan, The genome of *B.*
1091 *subtilis* phage SPP1 - Physical arrangement of phage genes, *MGG Mol. Gen. Genet.* 175
1092 (1979) 351–357. <https://doi.org/10.1007/BF00397235>.
- 1093
- 1094

1095

1096 **Figure Legends**

1097 **Figure 1** Purification of SPP1 phage tails and identification of their protein components.

1098 (A) Purification of SPP1 tails from *B. subtilis* YB886 infected with SPP1 $gp13^-$, using
1099 differential precipitation, anion exchange chromatography (AEC) and zonal centrifugation.

1100 The lanes of the Coomassie-stained SDS-PAGE gels are aligned with the corresponding
1101 fractions in the AEC chromatograms. For pellet P2, AEC fractions pooled for further
1102 purification by zonal centrifugation are shaded. Purified tails peak in P2 sucrose gradient
1103 fraction 6 (P2f6S) and in P3 AEC fraction 8 (P2f8), respectively (both shaded). (B)

1104 Electron micrographs of negatively stained phage tails from fractions P2f6S and P3f8.
1105 Morphological elements of the tail and average measurements of their length and diameter.

1106 ($n=20$) are depicted in the scheme on the right. (C) Assignment of bands detected on a
1107 Coomassie gel (15 % (w/v) SDS-PAGE) of a P3f8 sample, based on mass spectrometry

1108 (MS) and N-terminal (N-ter) sequencing (Table S1) as labeled. The portal protein gp6 and
1109 the major capsid protein gp13, present only in the phage particle control, are identified

1110 based on their apparent molecular mass. (D) Coomassie-stained SDS-PAGE gel
1111 (15 % (w/v)) of 10-fold concentrated samples P2f6S and P3f8 and western blots of the

1112 same samples, re-diluted 2-fold (P2f6S) and 6-fold (P3f8), respectively. Disrupted phage
1113 particles are run as controls. Note that anti-gp17.1N detects both major tail proteins,

1114 gp17.1 ($M_{r, app} \sim 21$ kDa) and gp17.1* (~ 29 kDa) [29] while anti-gp18N recognizes three
1115 different species of gp18.

1116

1117 **Figure 2** Immunolabeling of gp22 in P3f8 tails. Overview (A) and detail (B) of P3f8 tails

1118 cross-linked by purified anti-gp22 IgG through their tip region. The cartoons are visual
1119 aids to identify the position of IgG relative to the tail structure whose elements are color-

1120 coded as in Figure 3. Red lines in (A) show the relative position of the tail fiber and tail
1121 tube central symmetry axis to highlight the free-tilting joint between the cap and the fiber.

1122 Immuno-labelling required pre-fixation of the P3f8 tails with glutaraldehyde. (C) Labelling
1123 of the anti-gp22 IgG position with goat IgG anti-rabbit IgG complexed with colloidal gold.

1124

1125 **Figure 3** Schematic representation of the SPP1 virion's structural organization and of the
1126 genetic map of the two SPP1 tail gene operons. The region between nucleotides 7803 and
1127 20825 of the SPP1 genome (NCBI accession: X97918.3 [39]) corresponds to the 3' portion
1128 of the gene *II* operon (promoter *PL4*) and to the operon controlled by promoter *PL5*
1129 (figure bottom). It comprises open reading frames *I3* to *24.1*. Coding regions for structural
1130 proteins previously identified [19,33] or characterized in this work are depicted as colored
1131 arrows. Genes coding for unassigned (gp14, gp23, gp23.1, and gp24) and non-structural
1132 proteins (gp17.5 and gp17.5*) are represented in grey. The arrangement of protein
1133 components within the SPP1 virion established in this and in previous work [19,29,33,58]
1134 is schematized on top of the figure. Proteins are represented in the same colors as the
1135 corresponding coding genes below. Their abbreviation name according to function is
1136 displayed on top: THJP – Tail to Head Joining Protein, TCP – Tail Completion Protein;
1137 TTP – Tail Tube Protein; TMP – Tape Measure Protein; Dit – Distal Tail protein; TAP –
1138 Tail Adaptor Protein; RBP – Receptor Binding Protein. The tail tube is depicted cut-open
1139 to show the TMP gp18 inside the tail tube central channel. Gp22 is rendered semi-
1140 transparent to indicate its transient association (Figure 1D) to the tail fiber region identified
1141 in Figure 2. Conditional lethal mutants that disrupt individual protein function investigated
1142 in this study are shown in the figure middle labelled with their original names [83]. They
1143 were renamed in the manuscript text according to the gene they knock-out for simplicity.
1144 SPP1*sus31* (named here SPP1*gp13*⁻) [34] and SPP1*sus45* (named SPP1*gp17.1*⁻) [29] were
1145 previously described while all other mutants were mapped by complementation and
1146 sequenced in this study (Table S2). Designations of suppressor sensitive (*sus*) mutants are
1147 shown in blue-framed boxes, featuring also the position of the stop codon preventing the
1148 production of the respective protein. Temperature sensitive (*ts*) mutants and their mutation
1149 sites are framed in green. The black bars at the bottom of the figure identify the coding
1150 region of each tail gene that was expressed to purify proteins for raising polyclonal rabbit
1151 antibodies. The lowercase “d” indicates that denatured polypeptides were used for
1152 immunization. Anti-gp17.1 [29] and anti-gp21 [4] antibodies were previously described.

1153

1154 **Figure 4** Production and stability of SPP1 tail proteins in infected cells. (A-C) Production
1155 of SPP1 proteins in cells infected with SPP1 wild type (kinetics in (A)), suppressor
1156 sensitive (B) and temperature sensitive (C) mutants defective in tail genes (Figure 3).

1157 Infections of the non-permissive host *B. subtilis* YB886 were carried out at 37°C (A,B) and
1158 46°C (C). Bacteria infected with mutants defective in the proteins shown above the gel
1159 lanes were harvested at the time indicated. Tail proteins in cell extracts were detected by
1160 western blot. The three gp18-specific bands detectable in SPP1*wt* infections are marked by
1161 arrowheads and the gp18 band observed only after infection with SPP1*gp17.1*⁻ (B,C) and
1162 SPP1*ts* mutants (C) with diamonds to distinguish them from nonspecific background. The
1163 signal of the capsid portal protein gp6 was used to compare the level of overall phage
1164 protein production in infections with different mutants. * - the capsids run in this lane carry
1165 a gp6 form with a hexahistidine tag at its Cter leading to its lower electrophoretic mobility
1166 when compared to wild type gp6.

1167

1168 **Figure 5** Function of tail chaperones gp17.5 and gp17.5* on the TMP gp18 stability. The
1169 experimental conditions and symbols in the western blots are as in Figure 4.

1170

1171 **Figure 6** Identification of SPP1 tail protein complexes from bacteria infected with
1172 suppressor sensitive and temperature sensitive mutants (Figure 3). (A) Structures partially
1173 purified from lysates of the non-permissive strain *B. subtilis* YB886 infected with SPP1*sus*
1174 mutant phages defective in the production of the major capsid protein gp13, TTPs
1175 gp17.1/gp17.1*, tail assembly chaperones gp17.5/gp17.5*, and the TMP gp18. Structures
1176 concentrated through a sucrose cushion were applied to glycerol gradients (10 % to 30 %
1177 (w/v)) and spun at 35000 rpm in a SW41 rotor (Figure S5), for 7 h in the case of
1178 SPP1*gp13*⁻ infection and 12 h for the others. (B) Structures produced in infections a 46°C
1179 by SPP1*ts* mutants defective in the production of gp19.1 and gp21. Gradients of
1180 concentrated structures were spun in parallel with a SPP1*gp18*⁻ infection sample and a
1181 mixture of proteins with known sedimentation coefficients (bottom). The sedimentation
1182 conditions were as above except that the run was for 15 h. Tail proteins were detected by
1183 western blot. Polyclonal antibodies raised against SPP1 particles, which give a strong
1184 reaction with the Cter of gp17.1* [29], were used to detect the two SPP1 major tail
1185 proteins. Calibration proteins were detected by Coomassie staining (bottom).

1186

1187 **Figure 7** Function of gp16.1 in tail assembly. (A) Western blot of extracts of the non-
1188 permissive B subtilis strain YB886 infected with SPP1*gp13*⁻ and SPP1*gp16.1*⁻. Tail
1189 proteins gp16.1, gp17 and gp19.1 were detected with specific antibodies. (B) Composition
1190 of tails produced in non-permissive infections with SPP1*gp13*⁻ (tails) and SPP1*gp16.1*⁻
1191 (fraction 4 of the gradient in Figure S6). Western blots were developed with antibodies
1192 against tail structural proteins and the capsid portal protein gp6 (control). Phage particles
1193 and capsids are used as controls. (C) EM of negatively stained purified tails. White arrows
1194 indicate a narrow ring on the SPP1*gp13*⁻ tail region that binds to tails. This ring is absent in
1195 SPP1*gp16.1*⁻ tails.

1196

1197 **Figure 8** Assembly pathway model for the SPP1 phage tail. The protein complexes formed
1198 sequentially and mutants blocking individual assembly steps are depicted according to the
1199 data of this study. Brackets indicate a putative labile complex of gp18 bound to the cap
1200 gp19.1-gp21 complex. “?” indicates gp22 and gp17.5* putative interaction steps during
1201 assembly. See text for details.

1202

1203

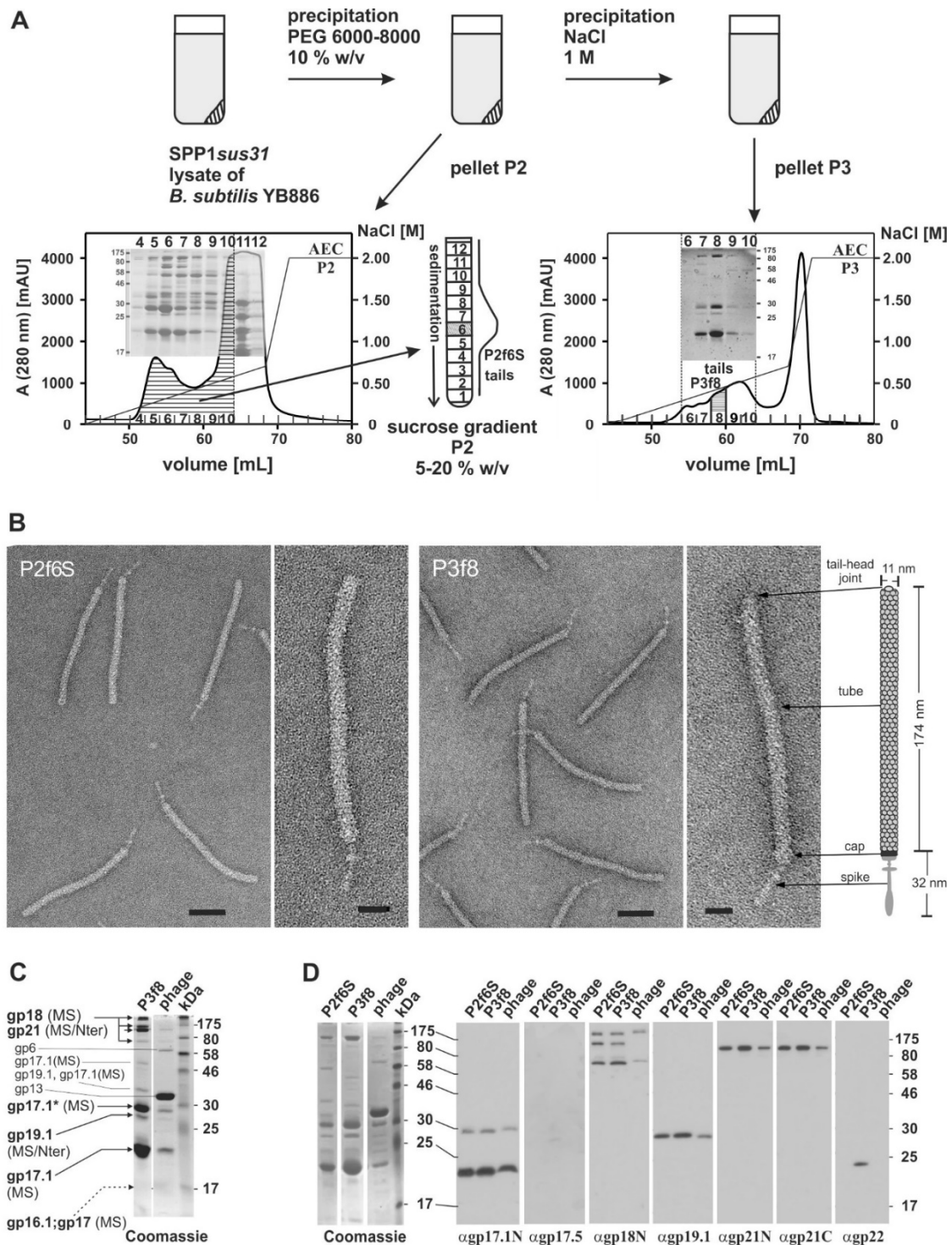
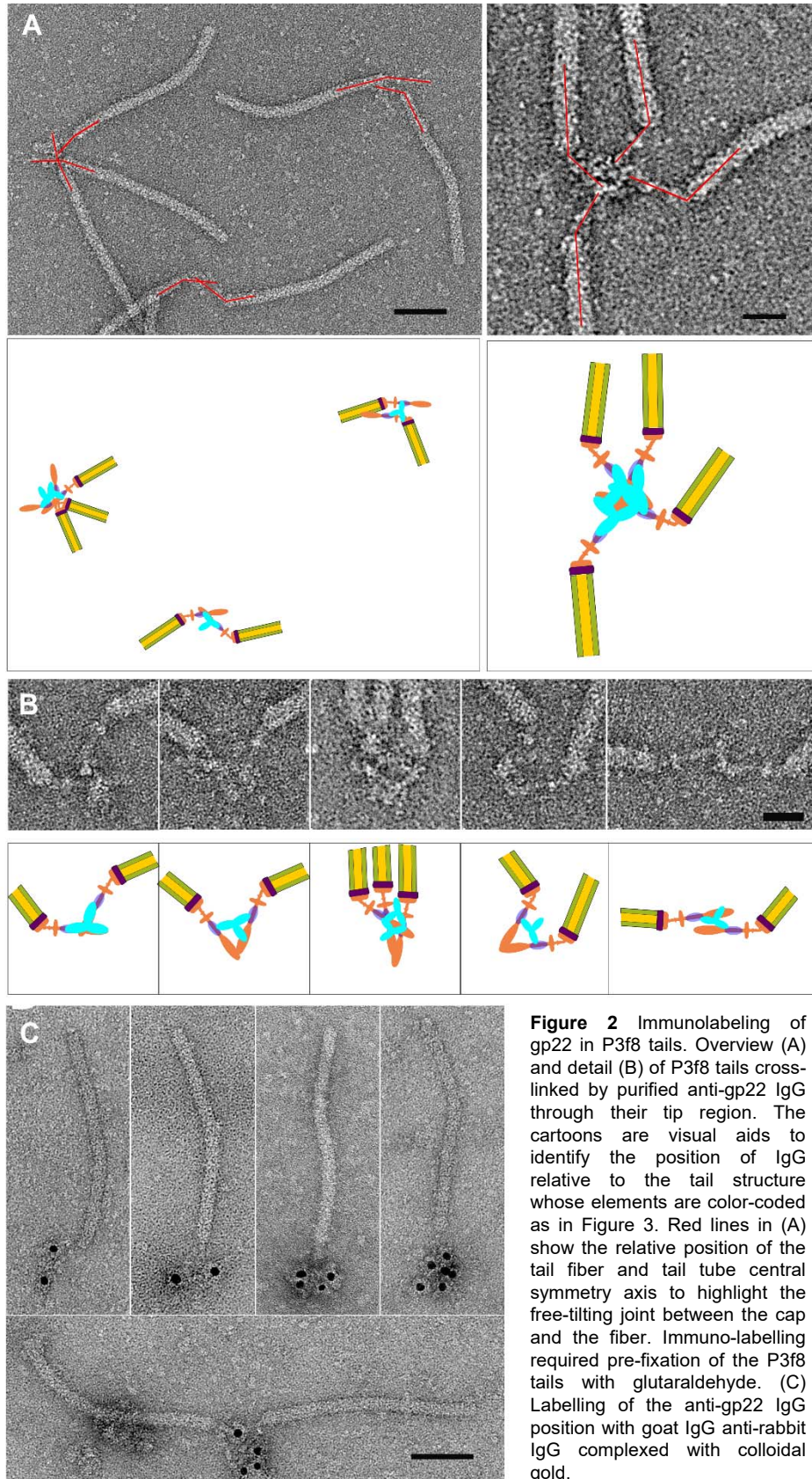


Figure 1 Purification of SPP1 phage tails and identification of their protein components. (A) Purification of SPP1 tails from *B. subtilis* YB886 infected with SPP1 $gp13$, using differential precipitation, anion exchange chromatography (AEC) and zonal centrifugation. The lanes of the Coomassie-stained SDS-PAGE gels are aligned with the corresponding fractions in the AEC chromatograms. For pellet P2, AEC fractions pooled for further purification by zonal centrifugation are shaded. Purified tails peak in P2 sucrose gradient fraction 6 (P2f6S) and in P3 AEC fraction 8 (P3f8), respectively (both shaded). (B) Electron micrographs of negatively stained phage tails from fractions P2f6S and P3f8. Morphological elements of the tail and average measurements of their length and diameter. ($n=20$) are depicted in the scheme on the right. (C) Assignment of bands detected on a Coomassie gel (15 % (w/v) SDS-PAGE) of a P3f8 sample, based on mass spectrometry (MS) and N-terminal (N-ter) sequencing (Table S1) as labeled. The portal protein gp6 and the major capsid protein gp13, present only in the phage particle control, are identified based on their apparent molecular mass. (D) Coomassie-stained SDS-PAGE gel (15 % (w/v)) of 10-fold concentrated samples P2f6S and P3f8 and western blots of the same samples, re-diluted 2-fold (P2f6S) and 6-fold (P3f8), respectively. Disrupted phage particles are run as controls. Note that anti-gp17.1N detects both major tail proteins, gp17.1 ($M_{r, app} \sim 21$ kDa) and gp17.1* (~ 29 kDa) [29] while anti-gp18N recognizes three different species of gp18.



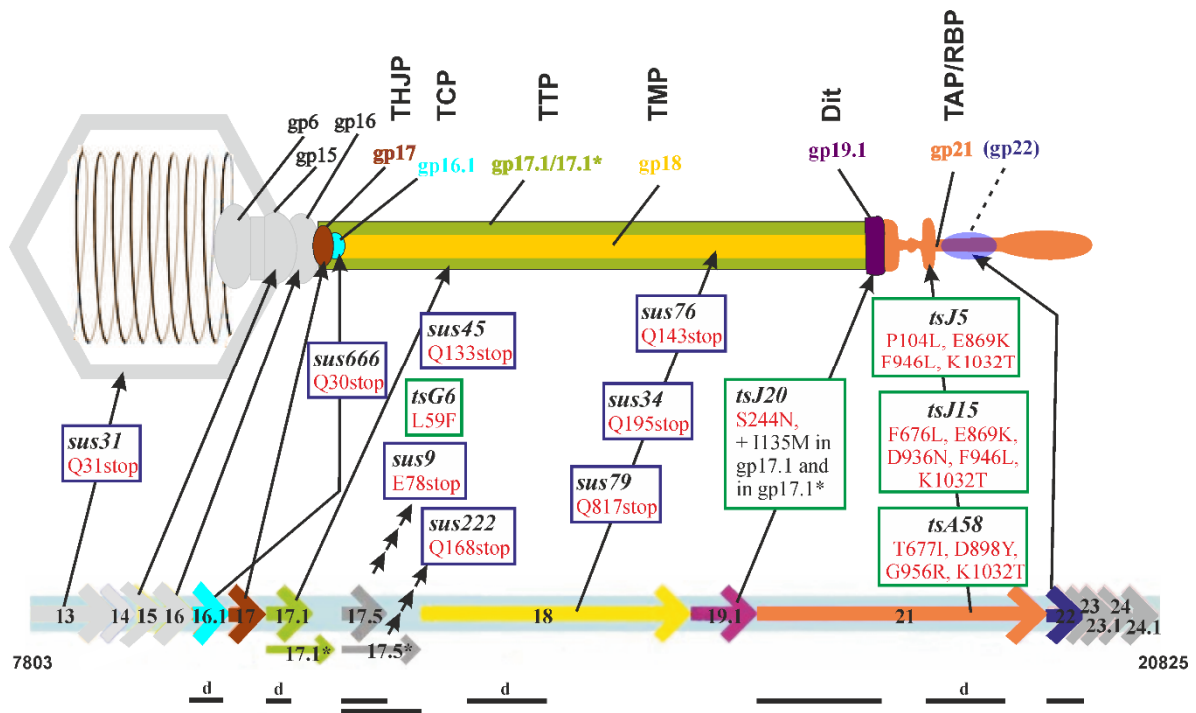


Figure 3 Schematic representation of the SPP1 virion's structural organization and of the genetic map of the two SPP1 tail gene operons. The region between nucleotides 7803 and 20825 of the SPP1 genome (NCBI accession: X97918.3 [39]) corresponds to the 3' portion of the gene 11 operon (promotor *PL4*) and to the operon controlled by promotor *PL5* (figure bottom). It comprises open reading frames 13 to 24.1. Coding regions for structural proteins previously identified [19,33] or characterized in this work are depicted as colored arrows. Genes coding for unassigned (gp14, gp23, gp23.1, and gp24) and non-structural proteins (gp17.5 and gp17.5*) are represented in grey. The arrangement of protein components within the SPP1 virion established in this and in previous work [19,29,33,58] is schematized on top of the figure. Proteins are represented in the same colors as the corresponding coding genes below. Their abbreviation name according to function is displayed on top: THJP – Tail to Head Joining Protein, TCP – Tail Completion Protein; TTP – Tail Tube Protein; TMP – Tape Measure Protein; Dit – Distal Tail protein; TAP – Tail Adaptor Protein; RBP – Receptor Binding Protein. The tail tube is depicted cut-open to show the TMP gp18 inside the tail tube central channel. Gp22 is rendered semi-transparent to indicate its transient association (Figure 1D) to the tail fiber region identified in Figure 2. Conditional lethal mutants that disrupt individual protein function investigated in this study are shown in the figure middle labelled with their original names [83]. They were renamed in the manuscript text according to the gene they knock-out for simplicity. SPP1*sus31* (named here SPP1*gp13*) [34] and SPP1*sus45* (named SPP1*gp17.1*) [29] were previously described while all other mutants were mapped by complementation and sequenced in this study (Table S2). Designations of suppressor sensitive (*sus*) mutants are shown in blue-framed boxes, featuring also the position of the stop codon preventing the production of the respective protein. Temperature sensitive (*ts*) mutants and their mutation sites are framed in green. The black bars at the bottom of the figure identify the coding region of each tail gene that was expressed to purify proteins for raising polyclonal rabbit antibodies. The lowercase “d” indicates that denatured polypeptides were used for immunization. Anti-gp17.1 [29] and anti-gp21 [4] antibodies were previously described.

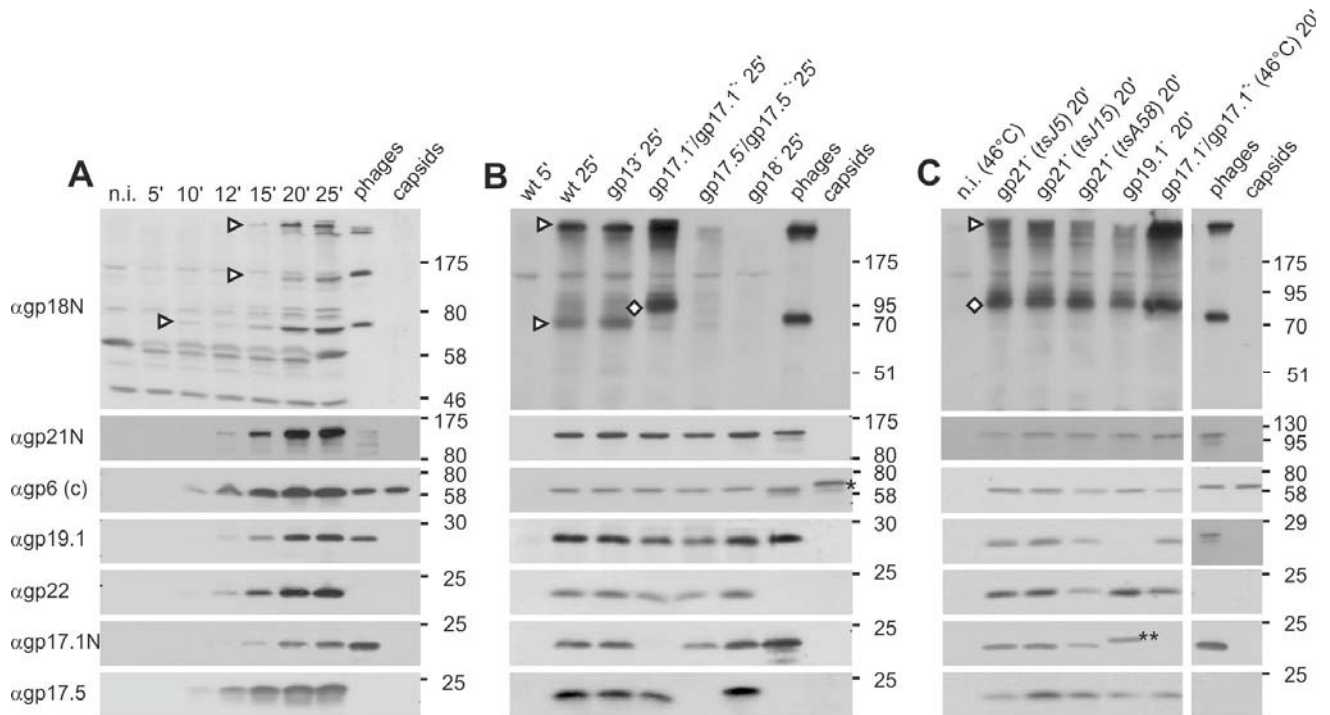


Figure 4 Production and stability of SPP1 tail proteins in infected cells. (A-C) Production of SPP1 proteins in cells infected with SPP1 wild type (kinetics in (A)), suppressor sensitive (B) and temperature sensitive (C) mutants defective in tail genes (Figure 3). Infections of the non-permissive host *B. subtilis* YB886 were carried out at 37°C (A,B) and 46°C (C). Bacteria infected with mutants defective in the proteins shown above the gel lanes were harvested at the time indicated. Tail proteins in cell extracts were detected by western blot. The three gp18-specific bands detectable in SPP1*wt* infections are marked by arrowheads and the gp18 band observed only after infection with SPP1*gp17.1* (B,C) and SPP1*ts* mutants (C) with diamonds to distinguish them from nonspecific background. The signal of the capsid portal protein gp6 was used to compare the level of overall phage protein production in infections with different mutants. * - the capsids run in this lane carry a gp6 form with a hexahistidine tag at its Cter leading to its lower electrophoretic mobility when compared to wild type gp6.

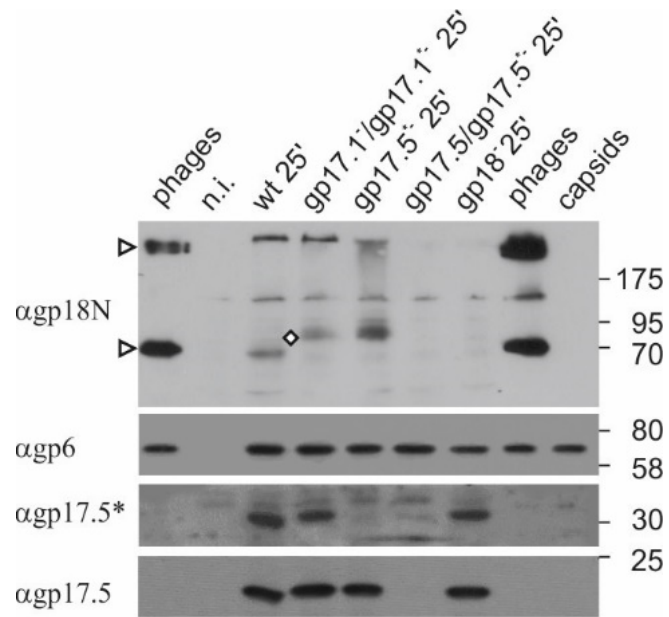


Figure 5 Function of tail chaperones gp17.5 and gp17.5* on the TMP gp18 stability. The experimental conditions and symbols in the western blots are as in Figure 4.

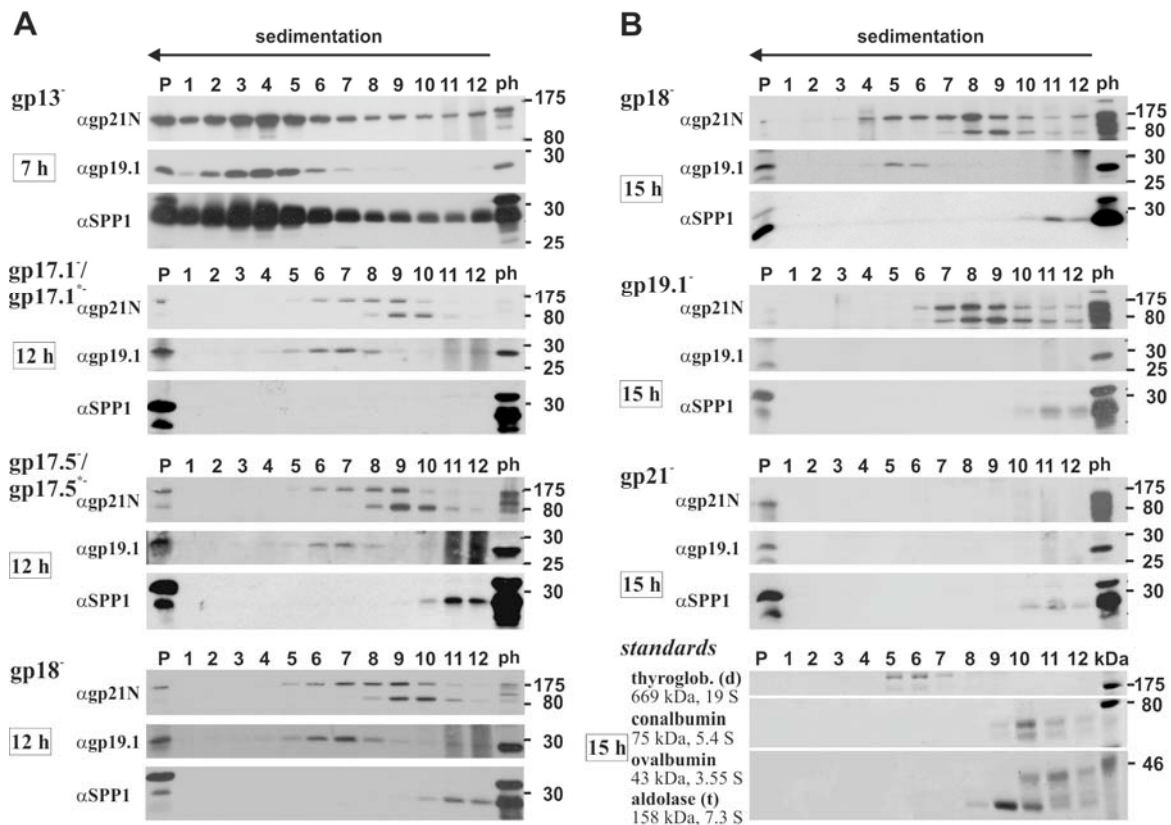


Figure 6 Identification of SPP1 tail protein complexes from bacteria infected with supressor sensitive and temperature sensitive mutants (Figure 3). (A) Structures partially purified from lysates of the non-permissive strain *B. subtilis* YB886 infected with SPP1^{sus} mutant phages defective in the production of the major capsid protein gp13, TTPs gp17.1/gp17.1^{*}, tail assembly chaperones gp17.5/gp17.5^{*}, and the TMP gp18. Structures concentrated through a sucrose cushion were applied to glycerol gradients (10 % to 30 % (w/v)) and spun at 35000 rpm in a SW41 rotor (Figure S5), for 7 h in the case of SPP1^{gp13⁻} infection and 12 h for the others. (B) Structures produced in infections a 46°C by SPP1^{ts} mutants defective in the production of gp19.1 and gp21. Gradients of concentrated structures were spun in parallel with a SPP1^{gp18⁻} infection sample and a mixture of proteins with known sedimentation coefficients (bottom). The sedimentation conditions were as above except that the run was for 15 h. Tail proteins were detected by western blot. Polyclonal antibodies raised against SPP1 particles, which give a strong reaction with the Cter of gp17.1^{*} [29], were used to detect the two SPP1 major tail proteins. Calibration proteins were detected by Coomassie staining (bottom).

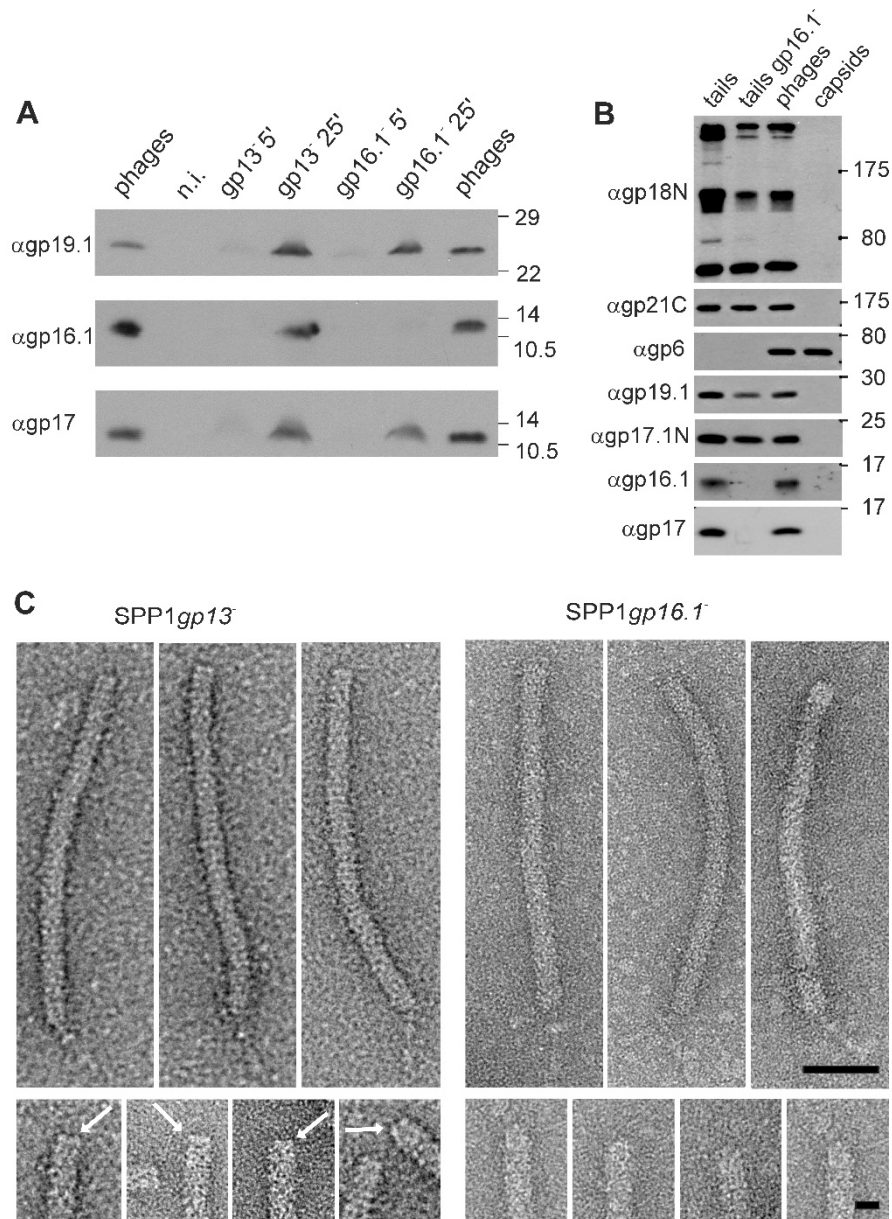


Figure 7 Function of gp16.1 in tail assembly. (A) Western blot of extracts of the non-permissive *B subtilis* strain YB886 infected with SPP1*gp13⁻* and SPP1*gp16.1⁻*. Tail proteins gp16.1, gp17 and gp19.1 were detected with specific antibodies. (B) Composition of tails produced in non-permissive infections with SPP1*gp13⁻* (tails) and SPP1*gp16.1⁻* (fraction 4 of the gradient in Figure S6). Western blots were developed with antibodies against tail structural proteins and the capsid portal protein gp6 (control). Phage particles and capsids are used as controls. (C) EM of negatively stained purified tails. White arrows indicate a narrow ring on the SPP1*gp13⁻* tail region that binds to tails. This ring is absent in SPP1*gp16.1⁻* tails.

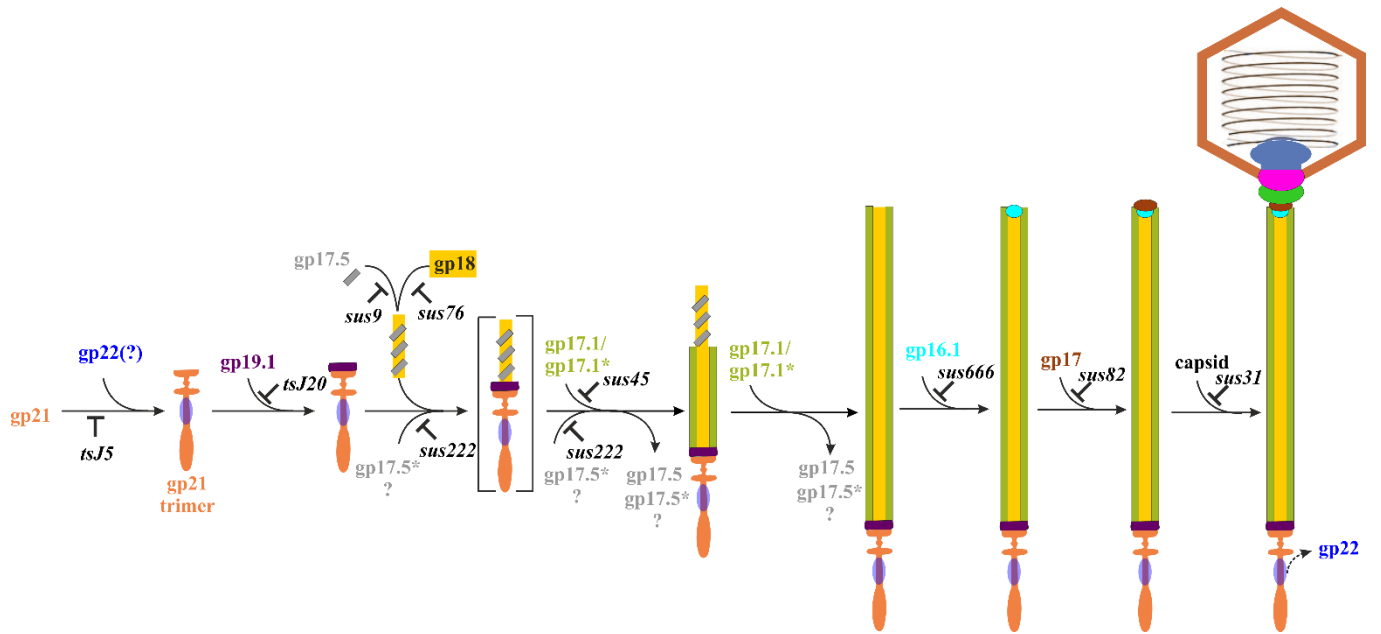


Figure 8 Assembly pathway model for the SPP1 phage tail. The protein complexes formed sequentially and mutants blocking individual assembly steps are depicted according to the data of this study. Brackets indicate a putative labile complex of gp18 bound to the cap gp19.1-gp21 complex. "?" indicates gp22 and gp17.5* putative interaction steps during assembly. See text for details.

Table S1 Fingerprint of SPP1 proteins studied in this work

protein	M _{r, th} [kDa] ^a	M _{r, app} [kDa] ^b	Number of amino acids	Sequence coverage ^c			N-terminal sequence ^d	pI _{calc} ^e	protein detection ^f and references to previous work ^g
				set 1 (MALDI)	set 2a (LC-MS/MS)	set 2b (MALDI)			
gp16.1	15.9	17	141	---	19-105 (36 %)	5-105 (40 %)	---	9.18	tail component (this work)
gp17	15.0	17	134	---	13-121 (59 %)	13-121 (59 %)	---	5.34	tail component [1,2]
gp17.1	19.2	21	177	---	1-177 (54 %)	---	---	4.31	tail component [3–5]
gp17.1*	28.2	29	264	---	1-209 (47 %)	---	---	4.36	tail component [3,4]
gp17.5	20.2	18	173	---	---	---	---	5.13	nonstructural protein (this work) [3,4]
gp17.5*	34.0	30	297	---	---	---	---	5.45	nonstructural protein (this work) [3,4]
gp18	110.9	>175 145 73	1032	6-916 (47 %) --- 6-1010 (44 %)	6-1010 (39 %) 6-902 (48 %) 6-902 (34 %)	6-909 (49 %) 6-909 (41 %) 21-909 (14 %)	--- --- ---	10.45	tail component; the gp18 species enriched in extracts of, e.g., <i>sus45</i> infected cells migrates at about 83 kDa (Fig. 4B,C) (this work) [3]
gp19.1	28.6	27	253	1-251 (80 %)	1-251 (55%)	---	MNIYDI	4.72	tail component (this work) [3,6,7]
gp21	123.6	138	1108	65-1107 (55 %)	8-1094 (53 %)	65-956 (47 %)	SKNIWIM	5.06	tail component (this work) [3,7,8]
gp22	16.7	23	146	1-146 (75 %)	---	---	---	4.32	tail component (this work) [9]
gp23	6.1	---	54	---	---	---	---	5.32	not detected in infected cells (this work)
gp23.1	5.8	---	51	---	---	---	---	3.73	not detected in infected cells (this work) [10]
gp24	8.4	---	76	---	---	---	---	4.41	not detected in infected cells (this work)

^amolecular mass calculated from the protein amino acid sequence using the ExPASy ProtParam tool

^bmolecular mass estimated from the migration behaviour in SDS-PA gels. For the gp18 species and gp21 the estimation was made from 10 % w/v Coomassie stained gels (or, more accurately, from western blots) and for other proteins from 15 % w/v Coomassie gel

^cMass spectrometry of SPP1 tail structural proteins. “set 1”: set of 5 bands cut from two SDS-PA gels (15 % w/v and 20 % w/v) and analysed by peptide mapping using MALDI-TOF. “set 2”: set of 13 bands cut from two SDS-PA gels (15 % w/v both) and analysed by peptide mapping using (a) LC-MS/MS (all bands analysed) and (b) MALDI-TOF (only selected bands from set 2).

^dNter sequence determined by Edman degradation.

^eisoelectric points estimated from the sequences using the ExPASy ProtParam tool.

^fwestern blot detection of tail proteins in SPP1 phage particles, tails and in extracts of infected cells.

^greferences:

- [1] B. Chagot, I. Auzat, M. Gallopin, I. Petitpas, B. Gilquin, P. Tavares, S. Zinn-Justin, Solution structure of gp17 from the Siphoviridae bacteriophage SPP1: Insights into its role in virion assembly, *Proteins Struct. Funct. Bioinforma.* 80 (2012) 319–326. <https://doi.org/10.1002/prot.23191>.
- [2] I. Auzat, I. Petitpas, R. Lurz, F. Weise, P. Tavares, A touch of glue to complete bacteriophage assembly: The tail-to-head joining protein (THJP) family, *Mol. Microbiol.* 91 (2014) 1164–1178. <https://doi.org/10.1111/mmi.12526>.
- [3] C. Plisson, H.E. White, I. Auzat, A. Zafarani, C. São-José, S. Lhuillier, P. Tavares, E. V. Orlova, Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection, *EMBO J.* 26 (2007) 3720–3728. <https://doi.org/10.1038/sj.emboj.7601786>.
- [4] I. Auzat, A. Dröge, F. Weise, R. Lurz, P. Tavares, Origin and function of the two major tail proteins of bacteriophage SPP1, *Mol. Microbiol.* 70 (2008) 557–569. <https://doi.org/10.1111/j.1365-2958.2008.06435.x>.
- [5] C. Langlois, S. Ramboarina, A. Cukkemane, I. Auzat, B. Chagot, B. Gilquin, A. Ignatiou, I. Petitpas, E. Kasotakis, M. Paternostre, H.E. White, E. V. Orlova, M. Baldus, P. Tavares, S. Zinn-Justin, Bacteriophage SPP1 tail tube protein self-assembles into β -structure-rich tubes, *J. Biol. Chem.* 290 (2015) 3836–3849. <https://doi.org/10.1074/jbc.M114.613166>.
- [6] D. Veessler, G. Robin, J. Lichière, I. Auzat, P. Tavares, P. Bron, V. Campanacci, C. Cambillau, Crystal Structure of Bacteriophage SPP1 Distal Tail Protein (gp19.1), *J. Biol. Chem.* 285 (2010) 36666–36673. <https://doi.org/10.1074/jbc.M110.157529>.
- [7] A. Goulet, J. Lai-Kee-Him, D. Veessler, I. Auzat, G. Robin, D.A. Shepherd, A.E. Ashcroft, E. Richard, J. Lichière, P. Tavares, C. Cambillau, P. Bron, The Opening of the SPP1 Bacteriophage Tail, a Prevalent Mechanism in Gram-positive-infecting Siphophages, *J. Biol. Chem.* 286 (2011) 25397–25405. <https://doi.org/10.1074/jbc.M111.243360>.
- [8] I. Vinga, C. Baptista, I. Auzat, I. Petitpas, R. Lurz, P. Tavares, M.A. Santos, C. São-José, Role of bacteriophage SPP1 tail spike protein gp21 on host cell receptor binding and trigger of phage DNA ejection, *Mol. Microbiol.* 83 (2012) 289–303. <https://doi.org/10.1111/j.1365-2958.2011.07931.x>.
- [9] D. Veessler, S. Blangy, S. Spinelli, P. Tavares, V. Campanacci, C. Cambillau, Crystal structure of Bacillus subtilis SPP1 phage gp22 shares fold similarity with a domain of lactococcal phage p2 RBP, *Protein Sci.* 19 (2010) 1439–1443. <https://doi.org/10.1002/pro.416>.
- [10] D. Veessler, S. Blangy, J. Lichière, M. Ortiz-Lombardía, P. Tavares, V. Campanacci, C. Cambillau, Crystal structure of Bacillus subtilis SPP1 phage gp23.1, a putative chaperone, *Protein Sci.* 19 (2010) 1812–1816. <https://doi.org/10.1002/pro.464>.

Table S2 Protein coding genes studied in this work and their conditional lethal mutants. See Figure 3 for a graphical representation.

gene product ^a	coding region ^b	<i>sus</i> mutants ^c	nucleotide substitution (position)	stop codon	<i>ts</i> mutants ^c	nucleotide substitution	amino acid substitution	reference (mapping / sequencing) ^e
gp13	7803-8777	<i>sus31</i>	CAG→TAG (7893)	Q31 <i>stop</i>	---			[1]
gp16.1	9644-10069	<i>sus666</i>	CAA→TAA (7922)	Q30 <i>stop</i>	---			this work
gp17	10066-10470	<i>sus82</i>	CAA→TAA (19399)	Q111 <i>stop</i>	---			[2]
gp17.1	10484-11017	<i>sus45</i>	CAA→TAA (10880)	Q133 <i>stop</i>	---			[3]
gp17.1*	10484-11279	<i>sus45</i>	CAA→TAA (10880)	Q133 <i>stop</i>	---			[3]
gp17.5	11363-11884	<i>sus9</i>	GAG→TAG (11594)	E78 <i>stop</i>	<i>tsG6</i>	CTT→TTT (11537)	L59F	this work
gp17.5*	11363-12255	<i>sus9</i> <i>sus222</i>	GAG→TAG (11594) CAA→TAA (11863)	E78 <i>stop</i> Q168 <i>stop</i>	<i>tsG6</i>	CTT→TTT (11537)	L59F	this work
gp18	12267-15365	<i>sus76</i> <i>sus34</i> <i>sus79</i>	CAA→TAA (12693) CAA→TAA (12849) CAA→TAA (14715)	Q143 <i>stop</i> Q195 <i>stop</i> Q817 <i>stop</i>	---			this work
gp19.1	15362-16123	---			<i>tsJ20</i>	AGT→AAT, (16092) (and ATT→ATG (10889)) ^d	S244N (and I135M in <i>gp17.1</i> and <i>gp 17.1*</i>) ^d	this work

			<i>tsJ5</i>	CCT→CTT (16447), GAA→AAA (18741), TTC→TTA (18974), AAA→ACA (19231)	P104L, E869K, F946L, K1032T	
gp21	16137-19463	---	<i>tsJ15</i>	TTT→CTT (18162), GAA→AAA (18741), GAC→AAC (18942), TTC→TTA (18974), AAA→ACA (19231)	F676L, E869K, D936N, F946L, K1032T	this work
			<i>tsA5</i>	ACA→ATA (18166), GAT→TAT (18828), GGG→AGG (19002), AAA→ACA (19231)	T677I, D898Y, G956R, K1032T	
gp22	19476-19916	---	---			
gp23	19932-20096	---	---			
gp23.1	20089-20244	---	---			
gp24	20237-20467	---	---			

^agene products shown to be either SPP1 structural, tail proteins or otherwise involved in tail morphogenesis are displayed in bold

^bcoordinates in the SPP1 nucleotide sequence, NCBI access code X97918.3 [4]

^cSPP1 *sus* and *ts* mutants are from our collection [5]. Mutants in bold were functionally characterized in this study.

^dSPP1 *tsJ20* carries a *ts* mutation in gene *19.I*, as assessed by complementation assays, and a missense mutation in genes *17.I/17.I** that leads to amino acid change I135M in gp17.1/gp17.1*

^ereferences:

- [1] B. Becker, N. De la Fuente, M. Gassel, D. Günther, P. Tavares, R. Lurz, T.A. Trautner, J.C. Alonso, Head morphogenesis genes of the Bacillus subtilis bacteriophage SPP1, *J. Mol. Biol.* 268 (1997) 822–839. <https://doi.org/10.1006/jmbi.1997.0997>.
- [2] I. Auzat, I. Petitpas, R. Lurz, F. Weise, P. Tavares, A touch of glue to complete bacteriophage assembly: The tail-to-head joining protein (THJP) family, *Mol. Microbiol.* 91 (2014) 1164–1178. <https://doi.org/10.1111/mmi.12526>.
- [3] I. Auzat, A. Dröge, F. Weise, R. Lurz, P. Tavares, Origin and function of the two major tail proteins of bacteriophage SPP1, *Mol. Microbiol.* 70 (2008) 557–569. <https://doi.org/10.1111/j.1365-2958.2008.06435.x>.
- [4] L.M. Godinho, M.E.S. Fadel, C. Monniot, L. Jakutyte, I. Auzat, A. Labarde, K. Djacem, L. Oliveira, R. Carballido-Lopez, S. Ayora, P. Tavares, The revisited genome of bacillus subtilis bacteriophage SPP1, *Viruses.* 10 (2018). <https://doi.org/10.3390/v10120705>.
- [5] B. Behrens, G. Lüder, M. Behncke, T.A. Trautner, A.T. Ganesan, The genome of B. subtilis phage SPP1 - Physical arrangement of phage genes, *MGG Mol. Gen. Genet.* 175 (1979) 351–357. <https://doi.org/10.1007/BF00397235>.

Table S3 Primers used in this work

Primer	Sequence ^a	Restriction sites
60102BZ	<u>GGGAATTCAAAGGAGAAGGATCCATGTCGGTTAGAATCGACCC</u>	EcoRI, BamHI
60102AZ	<u>CCGGTACCCTGCAGCATGTGTTTTACCCCTCTC</u>	KpnI, PstI
17.5CS1	<u>CCTCCAGGATCCGTCATCGATGAGTGTAATAAAAATGAGAGGGG</u>	BamHI, ClaI
23NCS1	<u>CCGGTACCCTGCAGCCGGGTTGATAATTTCCGG</u>	PstI
gene 17.5 CS	<u>CGGCCCGGGAAAGGAGAAGGATCCATGAGAGGGGTAAAACACATG</u>	SmaI, BamHI
gene 17.5* NCS	<u>GAACTGCAGCTATTTAGGTGTGAATTTTAGTCC</u>	PstI
prset1	<u>CGGCCCGGGAAAGGAGAAATACATATGCGGGGTTCTC</u>	SmaI, NdeI
prset2	<u>ATACAAGCTTCGATTCCATGGTACC</u>	HindIII
gene 17.5 to 17.5* CS	<u>GCATTGGATAACTCAGGTTTCTTCCAAACGAGCTGTGCG</u>	
gene 17.5 to 17.5* NCS	<u>CGCACAGCTCGTTTGGAAGAAACCTGAGTTATCCAATGC</u>	
gene 17.5 only CS 2mut	<u>GCATTGGATAACTCAGGATTCTTCAAACGAGCTGTGCG</u>	
gene 17.5 only NCS 2mut	<u>CGCACAGCTCGTTTGAAGAATCCTGAGTTATCCAATGC</u>	
gene 18Cter CS	<u>GAAGGATCCGACAAAATAAAAGAATATACGG</u>	BamHI
gene 18Cter NCS	<u>GAACTGCAGCTATCGTCTTTTCCTCTTCC</u>	PstI

gene 18 181 CS	<i>GAAGGATCC</i> <u>GAGGACTTCAA</u> AAGCAATTGG	BamHI
gene 18 347 CS	<i>GAAGGATCC</i> <u>GCGGATATGATGACATCTTTAG</u>	BamHI
gene 18 435 NCS	TGAACTCGAGCTGCAGCT <u>ACTTACCGATTTC</u> CGGGTGC	XhoI, PstI
gene 18 919 NCS	TGAACTCGAGCTGCAGCT <u>ACTTGATTGTTTT</u> CGGAATGGC	XhoI, PstI
F-gp19.1	ATGCTGCAG <u>AATATATATGACATTCTG</u>	PstI
R-gp19.1	TCAACCGGT <u>TAAAAACTTTTCTGTAAA</u>	AgeI
F-gp24	ATGCTGCAG <u>ACTAAGCTAACTGAAAAT</u>	PstI
R-gp24	TCAACCGGT <u>CTCCGGTTGTGGTTCTCC</u>	AgeI

^aPrimers sequence identical to SPP1 is underlined. Point mutations are non-underlined within SPP1 sequences and a nucleotide insertion is in red, non-underlined. Translation ribosome binding sites (complementary to *B. subtilis* 16 S rRNA 3'), start and stop codons are in bold type. Restriction sites are in italics and listed in the column on the right.