1 **Identification of the first gene transfer agent (GTA) small terminase in *Rhodobacter***

2 ***capsulatus*, its role in GTA production and packaging of DNA**

3

4 Sherlock, D.[a], Leong, J.X. [a*], and Fogg, P.C.M. [a#]

5

6 [a] University of York, Biology Department, Wentworth Way, York, United Kingdom. YO10 5DD

7

8 Running Title: Identification of the first GTA TerS

9

10 Keywords: Horizontal Gene Transfer, Bacterial Evolution, Gene Transfer Agent, *Rhodobacter*,

11 Antimicrobial Resistance, Terminase, DNA Packaging, Bacteriophage, Transduction, Virus

12

13 Abstract = 122 words, Text = 6,631 words

14

15 #Address correspondence to Paul Fogg, paul.fogg@york.ac.uk

16

17 *Present address: The Centre for Organismal Studies (COS), Universität Heidelberg, Germany

Abstract

Genetic exchange mediated by viruses of bacteria (bacteriophages) is the primary driver of rapid

bacterial evolution. The priority of viruses is usually to propagate themselves. Most bacteriophages

use the small terminase protein to identify their own genome and direct its inclusion into phage

capsids. Gene transfer agents (GTAs) are descended from bacteriophages but they instead package

fragments of the entire bacterial genome without preference for their own genes. GTAs don't

selectively target specific DNA and no GTA small terminases are known. Here, we identified the

small terminase from the model *Rhodobacter capsulatus* GTA, which then allowed prediction of

analogues in other species. We examined the role of the small terminase in GTA production and

propose a structural basis for random DNA packaging.

Importance

Random transfer of and any and all genes between bacteria could be influential in spread of virulence or antimicrobial resistance genes. Discovery of the true prevalence of GTAs in sequenced genomes is hampered by their apparent similarity to bacteriophages. Our data allowed the prediction of small terminases in diverse GTA producer species and defining the characteristics of a "GTA-type" terminase could be an important step toward novel GTA identification. Importantly, the GTA small terminase shares many features with its phage counterpart. We propose that the GTA terminase complex could become a streamlined model system to answer fundamental questions about dsDNA packaging by viruses that have not been forthcoming to date.

Introduction

Viral transduction by bacteriophages is generally accepted to be the dominant mechanism for the rapid exchange of genes between bacteria. Viruses are the most abundant organisms in the environment; it is estimated that there are $>10^{30}$ viruses in the oceans alone and the majority of these are viruses of bacteria (1). The impact of bacteriophages is massive, from their crucial role in biogeochemical cycling in the oceans to the ubiquitous crAss phages that are intimately associated with >98% of tested human gut microbiomes (2, 3).

True viruses are essentially selfish – they use host resources to replicate their own genome and package it into the viral protein shell before the progeny move on to infect a new host. Host DNA can also be packaged by bacteriophages but this occurrence is usually incidental (4–6). By contrast, Gene Transfer Agents (GTAs) are small virus-like particles that exclusively package and transfer random fragments of their host bacterium's DNA to recipient bacteria (7, 8), with no preference for the propagation of their own genes. There are no known restrictions on the DNA that can be packaged into GTA particles and, consequently, any gene may be transferred by GTAs (8–10). An eye opening study of antibiotic gene transfer by GTAs in *in situ* marine microcosms, detected extraordinary transfer frequencies that were orders of magnitude greater than more established mechanisms (11).

GTAs were first discovered in the alpha-proteobacterium *Rhodobacter capsulatus*, which remains the model organism for study of GTAs today (12, 13). The *R. capsulatus* GTA (RcGTA) is encoded by a 14.5 kb core gene cluster that encodes a phage T4-like large terminase and most of the RcGTA structural proteins (portal, capsid, various tail proteins and glycoside hydrolases) required for RcGTA production (14). Recently, ectopic loci encoding tail fibres, head spikes and putative maturation proteins have also been identified (15, 16). Homologous clusters of RcGTA-

60 like genes are present throughout the alpha-proteobacteria and appear to have co-evolved with the

61 host species, indicative of vertical inheritance (17, 18). Beyond the alpha-proteobacteria,

62 functional GTAs have since been discovered experimentally in diverse prokaryotes, including

63 animal pathogens of the *Brachyspira* genus (Spirochete) (19), *Desulfovibrio* spp. (delta-

64 proteobacteria) (20) and the Archaeon *Methanococcus voltae* (21, 22). Each of these disparate

65 GTAs was identified by chance during the study of phage-like particles or unusual levels of gene

66 transfer. However, it is extremely difficult to systematically identify GTAs by bioinformatics alone

67 because they are functionally analogous but genetically divergent from each other and their genes

68 strongly resemble remnant bacteriophages. The difficulty of rapidly identifying GTAs is perhaps

69 the major obstacle for expanding the breadth of research carried out on GTA producers.

70 The packaging of random bacterial DNA by GTAs is a fundamentally different behaviour to

71 that of bacteriophages and other viruses (23). The primary aim of a phage is to distribute their own

72 genes. Phages first replicate their genome, usually as a multi-copy concatamer. There is no

73 evidence that GTAs possess any DNA replication genes or that the packaged DNA has been

74 replicated, instead GTAs appear to contain the uncopied genome of the producing bacterium. For

75 viruses, in all known cases the volume of the capsid is enough to contain the whole viral genome,

76 however this is not the case for GTAs (7). An individual GTA virion is too small to package the

77 genes required for its own synthesis, for example each RcGTA transfers only ~4 kb of DNA but

78 the 14.5 kb core gene cluster plus several ectopic loci are required for mature GTA production. To

79 achieve packaging specificity, dsDNA phages usually use initiation sites at a specific location in

80 the phage genome that are recognized by the packaging machinery. Packaging initiation sites

81 generate specificity with a defined DNA sequence, e.g. *cos/pac* sites (23–25), or with favourable

82 topological features, e.g. conformational selection of intrinsically bent DNA by SPP1-like phages

83 (26). So far no evidence that GTAs target discrete packaging start sites has been presented and no

84 conserved sequences or topologies have been implicated as *cos/pac* equivalents, all of which

85 suggests that packaging initiation is indeed random.

86 Bacteriophages with a dsDNA genome use sophisticated molecular machinery, known as the

87 terminase, to specifically recognize replicated phage DNA and to drive it into a preformed capsid

88 (27). The capsid itself is essentially a passive receptacle and it is the terminase that provides DNA

89 selectivity, enzymatic activity and motive force required to fill the capsid. The terminase is a

90 complex of two oligomeric small and large terminase proteins, TerS and TerL, which are both

91 indispensable for proper phage function and DNA packaging (27). TerL possesses the enzymatic

92 activities required for DNA packaging: it has a C-terminal nuclease domain that cleaves the target

93 DNA to produce a free end available for packaging and an N-terminal ATPase domain that

94 translocates the DNA into a preformed capsid. Unlike TerL, TerS has no enzymatic activity and

95 instead carries out a regulatory role being responsible for recognition of the phage genome's

96 packaging initiation site, recruitment of TerL and modulation of TerL enzymatic activities (26, 28,

97 29).

98 In general, large terminase genes are sufficiently well conserved to allow confident

99 identification by sequence identity alone, partly owing to the presence of the Walker ATP-

100 interacting motifs, and thus most GTAs have an annotated *terL* gene. Small terminases, however,

101 are smaller with little primary sequence conservation, which makes them far more challenging to

102 identify *in silico*. No small terminase has been identified for any GTA to date. Given the role of

103 terminase proteins in phage biology, it is highly likely that the GTA terminase plays a defining

104 role in the packaging of random DNA. In this study, we definitively identify the small terminase

105 of the model *R. capsulatus* GTA, demonstrate and localize its interaction with the large terminase

106   and investigate its role in RcGTA production. Our characterization of the RcGTA TerS also allows

107   us to speculate on the physical requirements of a GTA-type small terminase and to identify

108   candidate small terminases in other GTA-producing species

109   **Results**

110   **Characterization of RcGTA *g1* (*rcc01682*).** A gene encoding a TerL homologue

111   (*rcc01683/*RcGTA *g2*) is readily identifiable within the *Rhodobacter capsulatus* SB1003 core

112   RcGTA gene cluster (Fig. 1A). The RcGTA TerL has regions of strong homology with large

113   terminases from several well-studied phages, including the presence of characteristic nuclease and

114   Walker ATPase motifs (Fig. 2) (30, 31). Small terminases are far more difficult to predict and

115   consequently no GTA small terminases have ever been identified. Most characterized phage TerS

116   proteins have a modular structure: the N-terminal region comprises the helix-turn-helix DNA-

117   binding domain, the central region contains a coiled-coil oligomerization domain and the C-

118   terminus contains the TerL interaction segment (32). Such domain organization is a well conserved

119   feature of TerS, despite the lack of sequence conservation.

120   In phage genomes, the small and large terminase genes are often co-localized and so the

121   core RcGTA gene cluster was examined for genes that could encode a small terminase. The

122   RcGTA gene cluster contains 17 predicted genes, of which at least six have been shown to be

123   essential for GTA production (16). The first gene of the cluster, *rcc01682* (referred to hereafter as

124   *g1* and the protein as gp1), is also thought to be essential for RcGTA activity (33), but no in-depth

125   characterization has been carried out and no function has so far been assigned. The *g1* ORF is 324

126   bases and is located immediately upstream of the large terminase. The gp1 protein sequence was

127   submitted to the JPRED4 protein secondary structure prediction server (34), which predicted an

128   almost entirely helical structure (Fig. 1B). Subsequent analysis using the COILS server (35)

129   (MTIDK matrix, all window sizes) indicated that of the three distinct α-helices, the first two are

130   likely to form a coiled-coil (Fig. 1C) reminiscent of a phage TerS oligomerization domain. A more

131   detailed structural prediction using the RaptorX structure prediction server (36) indicated

8

132 similarity to the phage T4-like small terminase from *Aeromonas* phage 44RR (Fig. 1D & E). The

133 44RR TerS crystal structure (PDB: 3TXS) failed to resolve the N/C-terminal segments of residues

134 1-24 and 114-154 due to conformational variability, however, an N-terminal helix-turn-helix

135 DNA-binding motif was predicted from the primary sequence (32). RcGTA gp1 begins with the

136 coiled-coil domain and appears to lack a DNA-binding domain (Fig. 1D) at the N-terminus. No

137 helix-turn-helix motif was detected by the Gym2.0 and NPS@ servers (37–39).

138 To confirm that *g1* is essential for RcGTA activity, a deletion mutant was produced in the

139 GTA hyperproducer strain *R. capsulatus* DE442. Loss of the *g1* gene prevented all detectable gene

140 transfer activity (Fig. 3A). Complementation with full length *g1* expressed ectopically from its

141 own promoter effectively restored gene transfer to wild-type frequencies (Fig. 3A).

142 Complementation was also attempted using *g1* constructs that lacked the sequence encoding either

143 the first or third α-helical regions; in both cases gene transfer frequencies were indistinguishable

144 from the uncomplemented DE442 Δ*g1* mutant (Fig. 3A).

145 It has previously been shown that the DE442 RcGTA hyperproducer packages sufficient

146 genomic DNA into GTA particles to allow detection of a distinct 4 kb band in total DNA

147 preparations (40). Given the predicted headful packaging mechanism used by RcGTA (8, 27),

148 production of 4 kb DNA fragments can only occur if DNA is successfully packaged into the capsid.

149 This property can be exploited to examine mutations that affect DNA packaging *in vivo*,

150 independent of the release of infective GTA particles. Deletion of *g1* prevents any detectable

151 accumulation of intracellular RcGTA 4 kb DNA fragments and *in trans* complementation restores

152 DNA packaging (Fig. 3B). DNA contained within extracellular GTA particles is protected from

153 enzymatic degradation. Isolation of DNase-insensitive DNA from the supernatant of DE442 wild-

154 type and Δ*g1* strains yielded detectable RcGTA DNA for the wild-type only (Fig. 3C). As phage

155   TerS are responsible for binding to target DNA and stimulating the various enzymatic activities of

156   the TerL that are required for DNA packaging, our data are entirely consistent with *g1* encoding

157   the RcGTA small terminase.

158   **The C-terminus of RcGTA gp1 interacts with the ATPase domain of TerL.** In bacteriophage,

159   the only protein that the small terminase is known to interact with is the large terminase. Indeed,

160   the small terminase not only recognizes the bacteriophage DNA, but also recruits the large

161   terminase and initiates the process of DNA packaging. Using the bacterial-2-hybrid assay, RcGTA

162   gp1 was translationally coupled to the T25 domain of the *Bordetella pertussis* adenylate cyclase

163   enzyme and RcGTA TerL was coupled to the adenylate cyclase T18 domain. Interaction between

164   the two proteins brings together the two adenylate cyclase domains leading to cAMP production

165   and subsequently β-galactosidase (41). In this assay, a distinct interaction can be seen between gp1

166   and gp2 (Fig. 3D). Truncation of gp1 to remove helix 1 had no appreciable effect on interaction

167   with the large terminase, however, loss of helix 3 led to complete loss of interaction (Fig. 3D).

168   Quantification of the results with a colorimetric β-galactosidase assay showed no significant

169   difference between the helix 3 deletion and the no insert negative control, whereas the helix 1

170   deletion was indistinguishable from full length gp1 (Fig. 3E).

171   The RcGTA large terminase has clear homology with large terminase proteins of several

172   well-studied phages (Fig. 2). The N-terminus of the protein contains the ATPase domain with

173   conserved Walker A and B motifs (Fig. 2A), while the C-terminus contains the nuclease domain

174   including three conserved nuclease motifs (Fig. 2B) (30). In the well-studied T4-like phages, it is

175   the ATPase domain that directly interacts with the small terminase (42). To test whether the

176   ATPase domain of the RcGTA large terminase is also responsible for interaction with gp1,

177   translational fusions were made of each of the two domains with the adenylate cyclase T18 domain.

178   In a bacterial-2-hybrid assay, the TerL nuclease domain (V253-L455) had no significant

179   interaction with RcGTA gp1 but the ATPase domain (L27-V258) produced a signal

180   indistinguishable from full-length TerL (Fig. 3D & E).

181   **RcGTA gp1 production is a prerequisite for tail attachment and efficient GTA capsid**

182   **maturation.** As shown above, Δ*g1* mutants are unable to produce infective GTA particles or to

183   package DNA, which indicates that RcGTA production has stalled early in the assembly process.

184   To determine the developmental state of the stalled RcGTAs, we purified the RcGTA particles

185   that were released by DE442 WT and Δ*g1* strains during lysis using nickel affinity purification.

186   The RcGTA lysis genes (*rcc00555* and *rcc00556*) are located elsewhere in the *R. capsulatus*

187   genome and should not be affected by the absence of a small terminase (8, 43). A plasmid

188   containing the RcGTA capsid (*rcc01687*/RcGTA *g5*) with a C-terminal His6-tag was introduced

189   into wild type DE442 and isogenic Δ*g1* strains. Timing of capsid expression was matched to GTA

190   production by fusing the *g5* ORF directly to the previously characterized RcGTA promoter (40,

191   44). Incorporation of recombinant capsid monomers into nascent RcGTA particles allows affinity

192   purification of the whole particles, as previously described (15). Concentrated samples were run

193   on an SDS PAGE gel to qualitatively assess the relative protein content. Strong bands were evident

194   in both samples at sizes consistent with the RcGTA capsid (post-translationally processed to 31.4

195   kDa (45)) and portal (42.8 kDa) proteins (Fig. 4). RcGTA^{WT}, but not RcGTA^{g1}, also had several

196   other visible bands (Fig. 4). RcGTA particles contain a distinctive 138.9 kDa putative tail

197   fibre/host specificity protein (encoded by *rcc01698*/RcGTA *g15*) (45), and a band of this size was

198   present only in the RcGTA^{WT} lane. The band was excised and positively identified as gp15 by

199   MALDI-MS:MS (3 unique peptide hits, expect <0.05, total score 154).

200   Affinity purified RcGTA particles were submitted for shotgun liquid chromatography-

201   tandem mass spectrometry (LC-MS/MS) analysis to determine the structural proteome of

202   RcGTA^{WT} versus RcGTA^{g1}. In terms of number of peptides detected, both sample types yielded

203   equivalent numbers for the RcGTA capsid and portal proteins (Fig. 5A). The GhsA and GhsB head

204   spike proteins (encoded by *rcc01079* and *rcc01080,* respectively) were represented in both

205   samples, however, 7 to 9-fold fewer GhsA/B peptides were detected in RcGTA^{g1} (Fig. 5A). In

206   contrast, peptide hits for the predicted RcGTA tail structures were almost completely absent in the

207   RcGTA^{g1} samples but abundant for RcGTA^{WT} (Fig. 5B). Transmission electron microscopy

208   images corroborated the proteomic data. RcGTA^{WT} samples yielded intact GTA particles with

209   clearly defined head spikes, portal apertures and dense staining of the heads, possibly indicative

210   of tightly packaged DNA (Fig. 5C-E). RcGTA^{g1} samples contained no evidence of tail structures,

211   head spikes were present but at reduced frequency and portal structures were visible (Fig. 5F-H).

212   Overall, RcGTA^{g1} head structures appeared more prone to damage than wild-type, maturation was

213   often incomplete and the contrast was poor - probably due to the absence of DNA (Fig. 5F-H). In

214   agreement with data presented earlier (Fig. 3C), DNA extraction from affinity purified RcGTA^{WT}

215   samples yielded characteristic 4 kb GTA DNA bands whereas no detectable DNA was recovered

216   from RcGTA^{g1} samples (Fig. 6A). Similar affinity chromatography using His6-tagged gp1 also

217   allowed purification of RcGTA particles from culture supernatant. The overall concentration of

218   RcGTA particles was much lower, presumably because the terminase complex dissociates after

219   packaging is complete, but 4 kb GTA DNA was still recoverable (Fig. 6B). These data demonstrate

220   a direct interaction between gp1 and the broader structural proteome for the first time, and support

221   our hypothesis that gp1 is indeed the small terminase.

222   **RcGTA gp1 binds weakly to DNA.** A core role of phage small terminases is to recognize the

223   phage genome and to target it for packaging into preformed capsids. RcGTAs don't package

224   specific DNA but the large terminase still needs to be recruited to the host genomic DNA to initiate

225   packaging, and it's plausible that this may be achieved via a non-specific affinity for DNA. In an

226   electrophoretic motility shift assay (EMSA), we tested the ability of RcGTA gp1 to bind DNA *in*

227   *vitro*. To obtain high concentration, soluble protein an N-terminal MBP-tag was used for gp1

228   purification. Purified gp1 exhibited low affinity for DNA with incomplete shifts occurring at

229   micromolar concentrations - 87% of DNA substrate was bound at 40 µM protein concentration

230   (Fig. 7). Six EMSA DNA substrates were used (351 to 2,944 bp PCR amplicons from distinct

231   locations in the *R. capsulatus* genome, however, the identity of the DNA did not substantially

232   affect the binding affinity. The size of the observed shift in DNA mobility was ~1500 bp or

233   equivalent to 975 kDa, which is greater than would be expected for binding of a single protein

234   monomer. The large reduction in mobility of the gp1-DNA complex indicates that gp1 could be

235   binding as an oligomer (small terminases usually form characteristic ring structures), there could

236   be multiple occupancy due to the lack of a specific binding site and/or the conformation of the

237   DNA may have been altered.

238   **GTA small terminases can be predicted in other species.** Identification of the RcGTA small

239   terminase allowed us to predict GTA *terS* genes in other alpha-proteobacterial species (Table 1),

240   including two previously unannotated ORFs in *Parvularcula bermudensis* and *Dinoroseobacter*

241   *shibae*. Interestingly, we were also able to predict small terminase genes in the distantly related

242   delta-proteobacterium *Desulfovibrio desulfuricans* and the Archaeon *Methanoccus voltae* (Table

243   1). In each case the small terminase gene was immediately upstream of the cognate large terminase,

244   the coding sequence for each small terminase was ~10-50% shorter than comparable phage

245   counterparts (Table 1) and the predicted protein structures were almost entirely helical. Overall,

246   the primary amino acid sequences of the various small terminases is poorly conserved, even

247   between those found in closely related species (Fig. 8). However, for the Rhodobacterales GTA

248   TerS proteins there is clear sequence similarity localized at the C-termini, specifically the third α-

249   helix (Fig. 8). Conservation of this region supports our findings that the C-terminal helix is

250   required for interaction with TerL, and that this interaction constrains TerS sequence divergence.

251   **Discussion**

252   Gene Transfer Agents clearly share many structural and mechanistic features with bacteriophages,

253   however, the most striking difference is that GTAs package and transfer random fragments of host

254   DNA without any preference for their own genome. In bacteriophages, DNA packaging is carried

255   out by the terminase complex, which is composed of multimeric small and large terminase

256   subunits. Interestingly, the Enterobacteria phage T4 large terminase can promiscuously package

257   heterologous linear DNA fragments into an empty phage head *in vitro* when TerS is absent,

258   reminiscent of GTA-type DNA packaging, but the presence of TerS is essential for terminase

259   activity *in vivo* (47). The large terminase has all the enzymatic capabilities required to package

260   DNA i.e. a nuclease domain to create free DNA ends at the beginning/end of packaging and an

261   ATPase domain to act as a motor to feed the DNA into the capsid (27, 31, 46). These data

262   demonstrate that TerS is not strictly required for the process of packaging DNA into the capsid but

263   is instead crucial for regulation (28). Depending on the particular phage, TerS forms an oligomeric

264   ring consisting of 8 to 11 identical protein subunits, with the DNA binding domains arranged

265   around the exterior surface. The TerS ring recognizes the packaging signal in the phage genome

266   and has been proposed to wrap ~100 bp of DNA around the outside, along the circular surface

267   formed by the DNA binding domains (48). TerS recruits TerL to make the initial DNA double

14

268 strand break, but inhibits further DNA cleavage to prevent damage to the phage genome. The

269 TerS/L complex docks to the phage head via the oligomeric portal protein, which contains a narrow

270 aperture for the DNA to be fed through. TerS stimulates TerL ATP hydrolysis and translocation

271 of the packaging complex along the phage genome. Once the genome has been tightly packaged

272 into the capsid, TerL cleaves the DNA again to complete the process. The terminase disassociates,

273 the portal aperture is plugged and tail assemblies are attached.

274   Given the role that small terminases play in phage DNA specificity it is likely that a comparable

275 protein is responsible for random DNA packaging by GTAs, however, no GTA TerS proteins have

276 so far been identified. Taken together, the molecular, genetic, proteomic and imaging data

277 presented here all support the hypothesis that RcGTA gp1 is the small terminase. RcGTA gp1 is

278 essential for RcGTA gene transfer and DNA packaging (Fig. 3A-C). RcGTA gp1 is predicted to

279 have structural characteristics in common with phage TerS proteins, in particular a putative coiled-

280 coil domain that is important for oligomerization in phage (Fig. 1) and a conserved C-terminal

281 large terminase interaction domain (Fig. 3D-E & Fig. 8). Analysis of the *R. capsulatus* DE442 Δ*g1*

282 mutant also allows us to postulate a model to describe RcGTA assembly. RcGTA capsid formation

283 and incorporation of the portal aperture occurs independently of the terminase and DNA

284 packaging. Proteomic analysis of the stalled RcGTA^{g1} particles did not detect substantial presence

285 of the large terminase protein, which suggests that either gp1 recruits TerL to the DNA first and

286 the terminase hetero-complex then recruits the preformed capsids, or that the interaction between

287 TerL and the portal is labile in the absence of TerS. Once the terminase-portal-capsid complex is

288 assembled, headful DNA packaging can begin. In the absence of DNA packaging, efficient

289 maturation of the RcGTA heads is impaired and RcGTA production stalls before the tail

290 appendage is attached (Fig. 5).

291   A crucial difference between the RcGTA small terminase and its phage counterparts is the

292   apparent lack of an N-terminal DNA-binding domain (Fig. 1). Previous work showed that deletion

293   of the N-terminal region of bacteriophage SF6 and SPP1 TerS proteins led to a significant

294   reduction in DNA binding affinity *in vitro* (49), but some binding was still retained. In addition,

295   both T4 and P22 TerS proteins have N- and C-Terminal DNA binding activities, with non-specific

296   DNA binding dependent upon a nine residue region in the P22 C-terminus, R143–K151 (48, 50).

297   Here, we show that the RcGTA TerS protein can bind non-specifically to DNA at micromolar

298   concentrations (Fig. 7). Absence of the specific DNA binding domain but retention of non-specific

299   DNA binding could provide an explanation for random DNA packaging by GTAs. It is possible

300   that the RcGTA TerS protein is also able to bind specific DNA sequences, however, this could not

301   be tested because we have no evidence to suggest that this occurs *in vivo* and no GTA binding sites

302   are currently known.

303   In summary, RcGTA gp1 is the first GTA small terminase to be described to date. We

304   hypothesize that GTA small terminases possess all of the regulatory abilities of phage small

305   terminases but lack of an N-terminal DNA-binding domain abolishes DNA sequence specificity.

306   Loss of the specific DNA binding region could allow non-specific binding of random DNA

307   sequences, which is the defining characteristic of GTA-type TerS proteins. The greatest barrier to

308   novel GTA identification and an understanding of their true prevalence in the environment, is the

309   lack of an effective identification method. Based on the data gained from RcGTA, we were able

310   to predict the small terminases from other known GTAs using gene size, neighbourhood and

311   protein secondary structure prediction analyses. Confirmation and in-depth characterization of

312   these proteins could allow us to pinpoint the defining characteristics of GTA-type terminases with

313   a view to enhanced discovery of novel GTAs in existing genome datasets. Furthermore, we also

314   anticipate that the smaller size and simpler organization of GTAs, compared to phages, will

315   provide the opportunity to develop a superior model system for structural and mechanistic studies.

Materials and Methods

Bacterial Strains. Two wild-type *Rhodobacter* strains were used – rifampicin resistant SB1003 (ATCC BAA-309) and rifampicin sensitive B10 (51). The RcGTA overproducer strain DE442 is of uncertain provenance but has been used in a number of RcGTA publications (44, 52). The *E. coli* S17-1 strain, which contains chromosomally integrated *tra* genes, was used as a donor for all conjugations. NEB 10-beta Competent *E. coli* (New England Biolabs, NEB) were used for standard cloning and plasmid maintenance; T7 Express Competent *E. coli* (NEB) were used for overexpression of proteins for purification.

Cloning. All cloning reactions were carried out with either the In-Fusion Cloning Kit (CloneTech) or NEBuilder (NEB) to produce the constructs listed in Table 2. All oligonucleotides were obtained from IDT (Table 3) and designed with an optimal annealing temperature of 60°C when used with Q5 DNA Polymerase (NEB). In summary, destination plasmids were linearized using a single restriction enzyme (pCM66T (BamHI), pEHisTEV (NcoI), pKT25 (BamHI), pUT18C (BamHI)), or linearized by PCR (pETFPP_2 using primers CleF and CleR). Inserts were amplified using primers with 15 bp 5' overhangs that have complementary sequence to the DNA with which it is to be recombined.

Transformation. Plasmids were introduced into *E. coli* by standard heat shock transformation (53), and into *Rhodobacter* by conjugation. For conjugation, 1 ml aliquots of an *E. coli* S17-1 donor containing the plasmid of interest and the *Rhodobacter* recipient were centrifuged at 5,000 x g for 1 min, washed with 1 ml SM buffer, centrifuged again and resuspended in 100 µl SM buffer. 10 µl of concentrated donor and recipient cells were mixed and spotted onto YPS agar or spotted individually as negative controls. Plates were incubated o/n at 30°C. Spots were scraped, suspended in 100 µl YPS broth and plated on YPS + 100 µg ml^{-1} rifampicin (counter-selection

18

339   against *E. coli*) + 10 µg ml^{-1} kanamycin (plasmid selection). Plates were incubated o/n at 30°C

340   then restreaked onto fresh agar to obtain single colonies.

341   **Gene Knock-Outs.** Knock-outs were created by RcGTA transfer. pCM66T plasmid constructs

342   were created with a gentamicin resistance cassette flanked by 500-1000 bp of DNA from either

343   side of the target gene. Assembly was achieved by a one-step, four component NEBuilder (NEB)

344   reaction and transformation into NEB 10-beta cells. Deletion constructs were introduced into the

345   RcGTA hyperproducer strain by conjugation and a standard GTA bio-assay was carried out to

346   replace the intact chromosomal gene with the deleted version.

347   ***Rhodobacter* Gene Transfer Assays.** In *Rhodobacter*, the assays were carried out essentially as

348   defined by Leung and Beatty (2013) (54). RcGTA donor cultures were grown anaerobically with

349   illumination in YPS for ~48 h and recipient cultures were grown aerobically in RCV for ~24 h.

350   For overexpression experiments, donor cultures were first grown aerobically to stationary phase

351   then anaerobically for 24 h. Cells were cleared from donor cultures by centrifugation and the

352   supernatant filtered through a 0.45 µm syringe filter. Recipient cells were concentrated 3-fold by

353   centrifugation at 5,000 x g for 5 min and resuspension in 1/3 volume G-Buffer (10 mM Tris-HCl

354   (pH 7.8), 1 mM MgCl$_2$, 1 mM CaCl$_2$, 1 mM NaCl, 0.5 mg ml^{-1} BSA). Reactions were carried out

355   in polystyrene culture tubes (Starlab) containing 400 µl G-Buffer, 100 µl recipient cells and 100

356   µl filter donor supernatant, then incubated at 30°C for 1 h. 900 µl YPS was added to each tube and

357   incubated for a further 3 h. Cells were harvested by centrifugation at 5,000 x g and plated on YPS

358   + 100 µg ml^{-1} rifampicin (for standard GTA assays) or 3 µg ml^{-1} gentamicin (for gene knock-outs).

359   **DNA Purification.** To isolate total intracellular DNA, 1 ml samples of relevant bacterial cultures

360   were taken for each nucleic acid purification replicate. Generally, sampling occurred during

361   stationary phase but for overexpression experiments samples were taken 6 h and 24 h after

362 transition to anaerobic growth. Total DNA was purified according to the Purification of Nucleic

363 Acids by Extraction with Phenol:Chloroform protocol (53). To isolate extracellular DNA

364 contained in RcGTA virions, *R. capsulatus* DE442 cultures (23 ml) were grown anaerobically with

365 illumination in YPS for ~48 h at 30°C. Cells were cleared from the cultures by centrifugation at

366 15,000 x g for 10 min and the supernatant was filtered through a 0.45 µm syringe filter. RcGTAs

367 were precipitated by addition of PEG8000 to a final concentration of 10% (w/v) and then incubated

368 at 4°C for 1 h with continuous rolling. Precipitated RcGTAs were pelleted by centrifugation at

369 10,000 x g for 10 min. The pellet was resuspended in 500 µl G-Buffer. Bacterial DNA and RNA

370 was removed by overnight incubation with Basemuncher nuclease (Expedeon) in the presence of

371 10 mM $MgCl_2$ at 30°C. Nuclease digestion was inhibited by addition of 50 mM EDTA. DNA was

372 extracted with Phenol:Chloroform:Isoamyl Alcohol (25:24:1, pH 8.0) as previously described

373 (53).

374 **Bacterial-two-hybrid (B2H) assays.** The procedure and the resources were as described in (41).

375 Plasmids encoding T18 (pUT18C and derivatives) and the compatible plasmids encoding T25

376 (pKT25 and derivatives) were introduced pairwise into competent BTH101 by co-transformation.

377 Selection was using LB agar containing 50 µg/ml kanamycin, 100 µg/ml ampicillin, 1 mM IPTG

378 and 80 µg/ml X-Gal, and plates were incubated at 30°C for 24-48 h. The phenotype of BTH101

379 (*cya-*) can be complemented if the two domains of adenylate cyclase (T18 and T25) are brought

380 into close proximity, and this can be achieved by fusing interacting protein partners to each

381 domain. The readout for complementation of the *cya-* phenotype (indicating a positive interaction

382 between the two fusion proteins) is the induction of *lac* (blue colonies on IPTG, XGal), whereas

383 no induction (white colonies) indicates no fusion protein interaction.

Assay of β-galactosidase activity. Colonies obtained from the B2H plasmids introduced into BH101 were spotted onto selective agar. The confluent spots were used to inoculate 200 μl aliquots of LB supplemented with 50 μg/ml kanamycin, 100 μg/ml ampicillin and 1 mM IPTG in a 96-well plate. Plates were covered and incubated for 16 h at 30°C with agitation. Absorbance (OD$_{600}$) readings were taken using a plate reader. In a second 96-well plate, 80 μl aliquots of permeabilization solution (100 mM Na$_2$HPO$_4$, 20 mM KCl, 2 mM MgSO$_4$, 0.06% (w/v) CTAB, 0.04% (w/v) sodium deoxycholate, 0.0054% (v/v) TCEP) were prepared. 20 μl aliquots from each well of the cultured bacteria were added to the corresponding wells of the plate containing the permeabilization solution and the mixtures incubated at room temperature for 15 min. 25 μl of the permeabilized samples were then added to 150 μl of substrate solution (60 mM Na$_2$HPO$_4$, 40 mM NaH$_2$PO$_4$, 1 mg/ml ONPG and 0.0027% (v/v) TCEP) that had been placed in a third 96-well plate. Absorbance (OD$_{420}$) readings were taken in the plate reader at 10 minute intervals over 60 min at 30°C. The maximum 2-point slope was calculated (ΔOD$_{420}$/min/ml).

Affinity purification of RcGTA particles. Purification of RcGTA particles was carried out as previously described with minor modifications (15, 45). Plasmids pCMF142 or pCMF173 (Table 2) were conjugated into the RcGTA overproducer strain *R. capsulatus* DE442 and an isogenic RcGTA *g1* deletion strain. pCMF142 and pCMF173 use the RcGTA promoter to express the RcGTA major capsid protein or gp1, respectively, with a hexa-histidine purification tag incorporated at the C-terminus. 100 ml cultures of DE442 and DE442 Δ*g1* were grown in YPS medium to stationary phase at 30°C, anaerobically with constant illumination. The cultures were cleared by centrifugation at 15,000 x g for 10 min followed by syringe filtration through a 0.45 μm pore filter. Tris – HCl (pH 8) was added to a final concentration of 10 mM. Each filtrate was mixed with 3 ml Amintra Ni-Agarose beads (Expedeon) pre-equilibrated with G* buffer (10 mM

407   Tris-HCl (pH 8), 1 mM MgCl$_2$, 1 mM CaCl$_2$ and 1 mM NaCl), then incubated at room temperature

408   for 1 h with agitation. Beads were applied to 25 ml gravity flow columns (Thermo-Fisher) and

409   washed with 200 ml G* buffer supplemented with 40 mM imidazole. RcGTA particles were eluted

410   using 5 ml G* buffer supplemented with 400 mM imidazole. Eluted RcGTAs were concentrated

411   and imidazole depleted to <1 mM by iterative dilution and ultrafiltration using a 100 kDa Spin-X

412   UF20 device (Corning).

413   **Electron Microscopy.** Affinity purified RcGTAs were directly applied to 200 mesh copper grids

414   with a formvar/carbon support film and allowed to adsorb for four minutes. The grids were washed

415   with three drops of deionised water and then negatively stained with uranyl acetate solution (55).

416   Samples were analyzed on a Tecnai 12 BioTWIN G2 transmission electron microscope operating

417   at 120kV, and images were captured using the Ceta camera (Thermo-Fisher).

418   **Protein Purification.** Protein overexpression was carried out in the NEB Express *E. coli* strain

419   (NEB) containing the relevant T7 expression plasmid (Table 2). Expression from the T7 promoter

420   was induced at mid-exponential growth phase with 0.2 mM IPTG at 20°C overnight. His6-tagged

421   (56) and MBP-tagged (40) proteins were purified as described previously. All chromatography

422   steps were carried out on an AKTA Prime instrument (GE Healthcare). Purified proteins were

423   concentrated in a Spin-X UF Centrifugal Concentrator (Corning). Samples were stored at -80°C

424   in binding buffer plus 50% glycerol.

425   **Electrophoretic motility shift assays (EMSA).** DNA substrates were prepared by PCR

426   amplification with oligonucleotides indicated in Table 3 and cleaned with a Monarch DNA clean-

427   up kit (NEB). 10 µl EMSA mixtures contained 100 ng of DNA, binding buffer based on reference

428   (57) (25 mM HEPES, 50 mM K-glutamate, 1 mM dithiothreitol, 0.05% Triton X-100, 4%

429   Glycerol, 1 µg poly dI:dC; pH 8.0) and purified protein at stated concentrations. Binding assays

430   were carried out at room temperature for 30 min. Samples were run on a 0.8% agarose gel in 0.5X

431   TBE at 80 V for 2 h at room temperature. Gels were stained with Sybr Safe (Invitrogen) and

432   imaged on a GelDoc transilluminator (BioRad).

433   **Sample Preparation for Mass Spectrometry.** For MALDI-MS:MS protein identification,

434   purified RcGTA samples were run on a TEO-Tricine 4-12% SDS Mini Gel (Expedeon) at 150 V

435   for 45 minutes. Gels were stained with InstantBlue protein stain (Expedeon) for a 1 h before

436   destaining with ultrapure water for 1 h. Protein bands of interest were excised. For shotgun LC-

437   MS:MS, samples were run into a 7 cm NuPAGE Novex 10% Bis-Tris Gel (Life Technologies) at

438   200 V for 6 mins. Gels were stained with SafeBLUE protein stain (NBS biologicals) for 1 h before

439   destaining with ultrapure water for 1 h.

440   In-gel   tryptic   digestion   was   performed   after   reduction   with   DTE   and   S-

441   carbamidomethylation with iodoacetamide. Gel pieces were washed two times with 50% (v:v)

442   aqueous acetonitrile containing 25 mM ammonium bicarbonate, then once with acetonitrile and

443   dried in a vacuum concentrator for 20 min. Sequencing-grade, modified porcine trypsin (Promega)

444   was dissolved in the 50 mM acetic acid supplied by the manufacturer, then diluted 5-fold by adding

445   25 mM ammonium bicarbonate to give a final trypsin concentration of 0.02 µg/µl. Gel pieces

446   were rehydrated by adding 10 µl of trypsin solution, and after 5 min enough 25 mM ammonium

447   bicarbonate solution was added to cover the gel pieces. Digests were incubated overnight at 37°C.

448   **MALDI-MS:MS.** A 1 µl aliquot of each peptide mixture was applied to a ground steel MALDI

449   target plate, followed immediately by an equal volume of a freshly-prepared 5 mg/mL solution of

450   4-hydroxy-α-cyano-cinnamic acid (Sigma) in 50% aqueous (v:v) acetonitrile containing 0.1% ,

451   trifluoroacetic acid (v:v).

452   Positive-ion MALDI mass spectra were obtained using a Bruker ultraflex III in reflectron

453   mode, equipped with a Nd:YAG smart beam laser. MS spectra were acquired over a range of 800-

454   4000 m/z. Final mass spectra were externally calibrated against an adjacent spot containing 6

455   peptides (des-Arg1-Bradykinin, 904.681; Angiotensin I, 1296.685; Glu1-Fibrinopeptide B,

456   1750.677; ACTH (1-17 clip), 2093.086; ACTH (18-39 clip), 2465.198; ACTH (7-38 clip),

457   3657.929.). Monoisotopic masses were obtained using a SNAP averagine algorithm (C 4.9384, N

458   1.3577, O 1.4773, S 0.0417, H 7.7583) and a S/N threshold of 2.

459   For each spot the ten strongest precursors, with a S/N greater than 30, were selected for

460   MS/MS fragmentation. Fragmentation was performed in LIFT mode without the introduction of a

461   collision gas. The default calibration was used for MS/MS spectra, which were baseline-subtracted

462   and smoothed (Savitsky-Golay, width 0.15 m/z, cycles 4); monoisotopic peak detection used a

463   SNAP averagine algorithm (C 4.9384, N 1.3577, O 1.4773, S 0.0417, H 7.7583) with a minimum

464   S/N of 6. Bruker flexAnalysis software (version 3.3) was used to perform spectral processing and

465   peak list generation.

466   **Shotgun LC-MS:MS.** Peptides were extracted by washing three times with 50% (v/v) aqueous

467   acetonitrile containing 0.1% trifluoroacetic acid (v/v), before being dried down in a vacuum

468   concentrator and reconstituting in aqueous 0.1% trifluoroacetic acid (v/v). Samples were loaded

469   onto a nanoAcquity UPLC system (Waters) equipped with a nanoAcquity Symmetry C18, 5 µm

470   trap (180 µm x 20 mm Waters) and a nanoAcquity HSS T3 1.8 µm C18 capillary column (75 □m

471   x 250 mm, Waters). The trap wash solvent was 0.1% (v/v) aqueous formic acid and the trapping

472   flow rate was 10 µl/min. The trap was washed for 5 min before switching flow to the capillary

473   column. Separation used a gradient elution of two solvents (solvent A: aqueous 0.1% (v/v) formic

474   acid; solvent B: acetonitrile containing 0.1% (v/v) formic acid). The capillary column flow rate

475 was 350 nl/min and the column temperature was 60°C. The gradient profile was linear 2-35% B

476 over 20 mins. All runs then proceeded to wash with 95% solvent B for 2.5 min. The column was

477 returned to initial conditions and re-equilibrated for 25 min before subsequent injections.

478 The nanoLC system was interfaced with a maXis HD LC-MS/MS system (Bruker

479 Daltonics) with CaptiveSpray ionisation source (Bruker Daltonics). Positive ESI-MS and MS/MS

480 spectra were acquired using AutoMSMS mode. Instrument control, data acquisition and processing

481 were performed using Compass 1.7 software (microTOF control, Hystar and DataAnalysis, Bruker

482 Daltonics). Instrument settings were: ion spray voltage: 1,450 V, dry gas: 3 l/min, dry gas

483 temperature 150°C, ion acquisition range: m/z 150-2,000, MS spectra rate: 2 Hz, MS/MS spectra

484 rate: 1 Hz at 2,500 cts to 10 Hz at 250,000 cts, cycle time: 3 s, quadrupole low mass: 300 m/z,

485 collision RF: 1,400 Vpp, transfer time 120 ms. The collision energy and isolation width settings

486 were automatically calculated using the AutoMSMS fragmentation table, absolute threshold 200

487 counts, preferred charge states: 2 – 4, singly charged ions excluded. A single MS/MS spectrum

488 was acquired for each precursor and former target ions were excluded for 0.8 min unless the

489 precursor intensity increased fourfold.

490 **Bioinformatics.** Tandem mass spectral data were submitted to database searching against the

491 unrestricted NCBInr database (version 20190131, 187087713 sequences; 68237485887 residues)

492 using a locally-running copy of the Mascot program (Matrix Science Ltd., version 2.5.1), through

493 the Bruker ProteinScape interface (version 2.1). Search criteria specified: Enzyme, Trypsin; Fixed

494 modifications, Carbamidomethyl (C); Variable modifications, Oxidation (M); Peptide tolerance,

495 150 ppm; MS/MS tolerance, 0.75 Da; Instrument, MALDI-TOF-TOF. Results were filtered to

496 accept only peptides with an expect score of 0.05 or lower.

497   Helix turn helix predictions were carried out using NPS@ (37, 38) and Gym2.0 (39) using

498   the default settings. Protein secondary structure and coiled coil predictions were made using

499   JPRED4 (34) and COILS (35), respectively. Protein 3D structures were predicted using RaptorX

500   (36). Molecular graphics/analyses were performed with the UCSF Chimera package v1.13 (58).

501   Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the

502   University of California, San Francisco (supported by NIGMS P41-GM103311). COBALT was

503   used for protein sequence alignments and PROMALS3D for alignment of predicted protein

504   structure (59, 60). Jalview was used to visualize alignments (61). FIJI software was used image

505   analysis (62). Figure graphics were produced using CorelDraw 2018. Statistical analysis was

506   carried out using Sigmaplot software version 13 (Systat Software Inc.) and, for each use, the test

507   parameters are indicated in the text and/or figure legends.

520   **References**

521   1.   Suttle CA. 2007. Marine viruses--major players in the global ecosystem. Nat Rev Microbiol **5**:801–

522   812.

523   2.   Shkoporov AN, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP, Hill C. 2018.

524   ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects

525   *Bacteroides intestinalis*. Nat Commun **9**:4781.

526   3.   Breitbart M, Bonnain C, Malki K, Sawaya NA. 2018. Phage puppet masters of the marine

527   microbial realm. Nature Microbiology **3**:754–766.

528   4.   Thierauf A, Perez G, Maloy AS. 2009. Generalized transduction. Methods Mol Biol **501**:267–286.

529   5.   Sato K, Campbell A. 1970. Specialized transduction of galactose by lambda phage from a deletion

530   lysogen. Virology **41**:474–487.

531   6.   Chen J, Quiles-Puchalt N, Chiang YN, Bacigalupe R, Fillol-Salom A, Chee MSJ, Fitzgerald JR,

532   Penadés JR. 2018. Genome hypermobility by lateral transduction. Science **362**:207–212.

533   7.   Lang AS, Zhaxybayeva O, Beatty JT. 2012. Gene transfer agents: phage-like elements of genetic

534   exchange. Nat Rev Microbiol **10**:472–482.

535   8.   Hynes AP, Mercer RG, Watton DE, Buckley CB, Lang AS. 2012. DNA packaging bias and

536   differential expression of gene transfer agent genes within a population during production and

537   release of the *Rhodobacter capsulatus* gene transfer agent, RcGTA. Mol Microbiol **85**:314–325.

538   9.   Tomasch J, Wang H, Hall ATK, Patzelt D, Preusse M, Petersen J, Brinkmann H, Bunk B, Bhuju S,

539   Jarek M, Geffers R, Lang AS, Wagner-Döbler I. 2018. Packaging of *Dinoroseobacter shibae* DNA

540   into Gene Transfer Agent particles is not random. Genome Biol Evol **10**:359–369.

541   10.   Berglund EC, Frank AC, Calteau A, Vinnere Pettersson O, Granberg F, Eriksson A-S, Näslund K,

542   Holmberg M, Lindroos H, Andersson SGE. 2009. Run-off replication of host-adaptability genes is

543   associated with gene transfer agents in the genome of mouse-infecting *Bartonella grahamii*. PLoS

544   Genet **5**:e1000546.

545   11.   McDaniel LD, Young E, Delaney J, Ruhnau F, Ritchie KB, Paul JH. 2010. High frequency of

546   horizontal gene transfer in the oceans. Science **330**:50.

547   12.   Solioz M, Marrs B. 1977. The gene transfer agent of *Rhodopseudomonas capsulata*. Purification

548   and characterization of its nucleic acid. Arch Biochem Biophys **181**:300–307.

549   13.   Marrs B. 1974. Genetic recombination in *Rhodopseudomonas capsulata*. Proc Natl Acad Sci U S A

550   **71**:971–973.

551   14.   Lang AS, Beatty JT. 2000. Genetic analysis of a bacterial genetic exchange element: the gene

552   transfer agent of *Rhodobacter capsulatus*. Proc Natl Acad Sci U S A **97**:859–864.

553   15.   Westbye AB, Kuchinski K, Yip CK, Beatty JT. 2016. The Gene Transfer Agent RcGTA contains

554   head spikes needed for binding to the *Rhodobacter capsulatus* polysaccharide cell capsule. J Mol

555   Biol **428**:477–491.

556   16.   Hynes AP, Shakya M, Mercer RG, Grüll MP, Bown L, Davidson F, Steffen E, Matchem H, Peach

557   ME, Berger T, Grebe K, Zhaxybayeva O, Lang AS. 2016. Functional and evolutionary

558   characterization of a gene transfer agent's multilocus "genome". Mol Biol Evol **33**:2530–2543.

559   17.   Shakya M, Soucy SM, Zhaxybayeva O. 2017. Insights into origin and evolution of α-

560   proteobacterial gene transfer agents. Virus evolution **3**:vex036.

561   18.   Lang AS, Beatty JT. 2006. Importance of wide spread gene transfer agent genes in alpha-

562   proteobacteria. Trends in microbiology **15**:54–62.

563   19.   Motro Y, La T, Bellgard MI, Dunn DS, Phillips ND, Hampson DJ. 2009. Identification of genes

564   associated with prophage-like gene transfer agents in the pathogenic intestinal spirochaetes

565   *Brachyspira hyodysenteriae, Brachyspira pilosicoli* and *Brachyspira intermedia*. Vet Microbiol

566   **134**:340–345.

29

567   20.   Rapp BJ, Wall JD. 1987. Genetic transfer in *Desulfovibrio desulfuricans*. Proc Natl Acad Sci U S A
568   **84**:9128–9130.

569   21.   Bertani G. 1999. Transduction-like gene transfer in the methanogen *Methanococcus voltae*. J
570   Bacteriol **181**:2992–3002.

571   22.   Eiserling F, Pushkin A, Gingery M, Bertani G. 1999. Bacteriophage-like particles associated with
572   the gene transfer agent of *Methanococcus voltae* PS. J Gen Virol **80 (Pt 12)**:3305–3308.

573   23.   Rao VB, Black LW. 2013. DNA Packaging in Bacteriophage T4 - Madame Curie Bioscience
574   Database - Austin (TX): Landes Bioscience.

575   24.   Nichols BP, Donelson JE. 1978. 178-Nucleotide sequence surrounding the *cos* site of
576   bacteriophage lambda DNA. J Virol **26**:429–434.

577   25.   Sternberg N, Coulby J. 1990. Cleavage of the bacteriophage P1 packaging site (*pac*) is regulated by
578   adenine methylation. Proc Natl Acad Sci U S A **87**:8070–8074.

579   26.   Greive SJ, Fung HKH, Chechik M, Jenkins HT, Weitzel SE, Aguiar PM, Brentnall AS, Glousieau
580   M, Gladyshev GV, Potts JR, Antson AA. 2016. DNA recognition for virus assembly through
581   multiple sequence-independent interactions with a helix-turn-helix motif. Nucleic Acids Res
582   **44**:776–789.

583   27.   Black LW. 2015. Old, new, and widely true: The bacteriophage T4 DNA packaging mechanism.
584   Virology **479-480**:650–656.

585   28.   Al-Zahrani AS, Kondabagil K, Gao S, Kelly N, Ghosh-Kumar M, Rao VB. 2009. The small
586   terminase, gp16, of bacteriophage T4 is a regulator of the DNA packaging motor. J Biol Chem
587   **284**:24490–24500.

588   29.   Leavitt JC, Gilcrease EB, Wilson K, Casjens SR. 2013. Function and horizontal transfer of the
589   small terminase subunit of the tailed bacteriophage Sf6 DNA packaging nanomotor. Virology
590   **440**:117–133.

591   30.   Ponchon L, Boulanger P, Labesse G, Letellier L. 2006. The endonuclease domain of bacteriophage

592   terminases belongs to the resolvase/integrase/ribonuclease H superfamily: a bioinformatics analysis

593   validated by a functional study on bacteriophage T5. J Biol Chem **281**:5829–5836.

594   31.   Kanamaru S, Kondabagil K, Rossmann MG, Rao VB. 2004. The functional domains of

595   bacteriophage t4 terminase. J Biol Chem **279**:40795–40801.

596   32.   Sun S, Gao S, Kondabagil K, Xiang Y, Rossmann MG, Rao VB. 2012. Structure and function of

597   the small terminase component of the DNA packaging machine in T4-like bacteriophages. Proc

598   Natl Acad Sci U S A **109**:817–822.

599   33.   Leung MM-Y. 2010. PhD Thesis. University of British Columbia. CtrA and GtaR : two systems

600   that regulate the Gene Transfer Agent in *Rhodobacter capsulatus*.

601   34.   Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015. JPred4: a protein secondary structure prediction

602   server. Nucleic Acids Res **43**:W389–94.

603   35.   Lupas A, Van Dyke M, Stock J. 1991. Predicting coiled coils from protein sequences. Science

604   **252**:1162–1164.

605   36.   Källberg M, Margaryan G, Wang S, Ma J, Xu J. 2014. RaptorX server: a resource for template-

606   based protein structure modeling. Methods Mol Biol **1137**:17–27.

607   37.   Combet C, Blanchet C, Geourjon C, Deléage G. 2000. NPS@: network protein sequence analysis.

608   Trends Biochem Sci **25**:147–150.

609   38.   Dodd IB, Egan JB. 1990. Improved detection of helix-turn-helix DNA-binding motifs in protein

610   sequences. Nucleic Acids Res **18**:5019–5026.

611   39.   Narasimhan G, Bu C, Gao Y, Wang X, Xu N, Mathee K. 2002. Mining protein sequences for

612   motifs. J Comput Biol **9**:707–720.

613   40.   Fogg PCM. 2019. Identification and characterization of a direct activator of a gene transfer agent.

614   Nat Commun **10**:595.

615   41.   Karimova G, Pidoux J, Ullmann A, Ladant D. 1998. A bacterial two-hybrid system based on a

616   reconstituted signal transduction pathway. Proc Natl Acad Sci U S A **95**:5752–5756.

617   42.   Gao S, Rao VB. 2011. Specificity of interactions among the DNA-packaging machine components

618   of T4-related bacteriophages. J Biol Chem **286**:3944–3956.

619   43.   Westbye AB, Leung MM, Florizone SM, Taylor TA, Johnson JA, Fogg PC, Beatty JT. 2013.

620   Phosphate concentration and the putative sensor kinase protein CckA modulate cell lysis and

621   release of the *Rhodobacter capsulatus* gene transfer agent. J Bacteriol **195**:5025–5040.

622   44.   Fogg PCM, Westbye AB, Beatty JT. 2012. One for all or all for one: heterogeneous expression and

623   host cell lysis are key to gene transfer agent activity in *Rhodobacter capsulatus*. PLoS ONE

624   **7**:e43772.

625   45.   Chen F, Spano A, Goodman BE, Blasier KR, Sabat A, Jeffery E, Norris A, Shabanowitz J, Hunt

626   DF, Lebedev N. 2009. Proteomic analysis and identification of the structural and regulatory

627   proteins of the *Rhodobacter capsulatus* gene transfer agent. J Proteome Res **8**:967–973.

628   46.   Kondabagil KR, Zhang Z, Rao VB. 2006. The DNA translocating ATPase of bacteriophage T4

629   packaging motor. J Mol Biol **363**:786–799.

630   47.   Zhang Z, Kottadiel VI, Vafabakhsh R, Dai L, Chemla YR, Ha T, Rao VB. 2011. A promiscuous

631   DNA packaging machine from bacteriophage T4. PLoS Biol **9**:e1000592.

632   48.   Gao S, Zhang L, Rao VB. 2016. Exclusion of small terminase mediated DNA threading models for

633   genome packaging in bacteriophage T4. Nucleic Acids Res **44**:4425–4439.

634   49.   Büttner CR, Chechik M, Ortiz-Lombardía M, Smits C, Ebong I-O, Chechik V, Jeschke G,

635   Dykeman E, Benini S, Robinson CV, Alonso JC, Antson AA. 2012. Structural basis for DNA

636   recognition and loading into a viral packaging motor. Proc Natl Acad Sci U S A **109**:811–816.

637   50.   Nemecek D, Lander GC, Johnson JE, Casjens SR, Thomas GJ. 2008. Assembly architecture and

638   DNA binding of the bacteriophage P22 terminase small subunit. J Mol Biol **383**:494–501.

639   51.   Wall JD, Weaver PF, Gest H. 1975. Gene transfer agents, bacteriophages, and bacteriocins of

640   *Rhodopseudomonas capsulata*. Arch Microbiol **105**:217–224.

641   52.   Ding H, Moksa MM, Hirst M, Beatty JT. 2014. Draft genome sequences of six *Rhodobacter*

642   *capsulatus Strains*, YW1, YW2, B6, Y262, R121, and DE442. Genome Announc **2**.

643   53.   Maniatis T, Fritsch EF, Sambrook J. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring

644   Harbor laboratory press.

645   54.   Leung M, Beatty J. 2013. *Rhodobacter capsulatus* Gene Transfer Agent transduction assay. Bio-

646   protocol **3(4)**: e334. DOI: 10.21769/BioProtoc.334.

647   55.   Booth DS, Avila-Sakar A, Cheng Y. 2011. Visualizing proteins and macromolecular complexes by

648   negative stain EM: from grid preparation to image acquisition. J Vis Exp. **22**(58): 3227.

649   56.   Fogg PCM, Younger E, Fernando BD, Khaleel T, Stark WM, Smith MCM. 2018. Recombination

650   directionality factor gp3 binds φC31 integrase via the zinc domain, potentially affecting the

651   trajectory of the coiled-coil motif. Nucleic Acids Res **46**:1308–1320.

652   57.   Wiethaus J, Schubert B, Pfänder Y, Narberhaus F, Masepohl B. 2008. The GntR-like regulator

653   TauR activates expression of taurine utilization genes in *Rhodobacter capsulatus*. J Bacteriol

654   **190**:487–493.

655   58.   Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004.

656   UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem

657   **25**:1605–1612.

658   59.   Papadopoulos JS, Agarwala R. 2007. COBALT: constraint-based alignment tool for multiple

659   protein sequences. Bioinformatics **23**:1073–1079.

660   60.   Pei J, Kim B-H, Grishin NV. 2008. PROMALS3D: a tool for multiple protein sequence and

661   structure alignments. Nucleic Acids Res **36**:2295–2300.

662   61.   Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2--a

663   multiple sequence alignment editor and analysis workbench. Bioinformatics **25**:1189–1191.

664   62.   Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden

665   C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona

666   A. 2012. Fiji: an open-source platform for biological-image analysis. Nat Methods **9**:676–682.

667   63.   Liu H, Naismith JH. 2009. A simple and efficient expression and purification system using two

668   newly constructed vectors. Protein Expr Purif **63**:102–111.

669   64.   Fogg MJ, Wilkinson AJ. 2008. Higher-throughput approaches to crystallization and crystal

670   structure determination. Biochem Soc Trans **36**:771–775.

Legends

Figure 1. Location and structure of the RcGTA small terminase, gp1. A. Schematic of the RcGTA

core gene cluster. Genes are shown as arrows with *R. capsulatus* SB1003 gene designations below and

known or predicted protein functions above. Arrows are coloured according to type - DNA packaging

(black), head associated (red) and tail associated (cyan). **B.** Amino acid sequence of RcGTA gp1 with the

predicted secondary structure indicated. Boxes represent α-helices and lines are disordered. The

boundaries of helix 1 (Δh1) and helix 3 (Δh3) truncations used in this study are shown as lines beneath

the sequence with the new terminal amino acids annotated. **C.** RcGTA gp1 coiled coil prediction using

COILS. The three window sizes for the prediction are annotated on the graph and colour coded. **D.** 3D

structural prediction of RcGTA gp1 using RaptorX and visualized using UCSF Chimera. Terminal amino

acids are annotated. **E.** Crystal structure of *Aeromonas* phage 44RR TerS, visualized with UCSF

Chimera.

Figure 2. Conserved functional motifs of the RcGTA large terminase protein, gp2. Alignments of the

N-terminal ATPase domains (**A**) and C-terminal nuclease domains (**B**) of RcGTA gp2 (ADE85428),

phage T4 gp17 (AAD42422), phage T5 TerL (AAS77194), phage T7 gp19 (AAP33962) and phage SPP1

gp2 (CAA39537). Alignment were made using COBALT and visualized using Jalview. Amino acid

similarity is indicated using the Clustal colour scheme. Amino acid position numbers in the full-length

protein are shown at the beginning and end of each row. The location of the Walker A, Walker B, motif

III/ATP-coupling and nuclease motifs are underlined and annotated (30).

Figure 3. The role of gp1 in RcGTA production. A. Histogram showing the results of a gene transfer

assay using the following donor strains - *R. capsulatus* DE442 wild-type [**WT**], *g1* deletion [**Δg1**], *g1*

deletion complemented with full length *g1* [**Δg1(g1)**], *g1* deletion complemented with *g1* lacking helix 1

[**Δg1(Δh1)**] and *g1* deletion complemented with *g1* lacking helix 3 [**Δg1(Δh3)**]. Statistical significance is

shown above the chart (ANOVA, n=3, ns=not significant, *** p<0.01). **B.** Agarose gel of total DNA

isolated from *R. capsulatus* DE442 wild-type [**WT**], *g1* deletion [**Δg1**] and a complemented *g1* deletion

696   [**Δ*g1*(*g1*)**]. Bioline Hyperladder 1 kb DNA ladder [**M**] and purified RcGTA DNA [**GTA**] is shown for

697   size comparison. The location of genomic DNA, RcGTA DNA and the 4 kb DNA ladder band are

698   annotated. **C.** Agarose gel of DNA isolated from purified RcGTA particles released by DE442 wild-type

699   and *g1* deletion strains. **D.** Interaction between RcGTA gp1 and the large terminase (TerL). 10 µl spots of

700   individual bacterial-2-hybrid assay transformations. Blue/green indicates a positive interaction and white

701   indicates no interaction. Reactions shown are as follows: "gp1" = gp1 vs TerL, "-ve" = no insert control,

702   "gp1Δh1" = gp1 helix 1 deletion vs TerL, "gp1Δh3" = gp1 helix 3 deletion vs TerL, "Nuc" = gp1 vs TerL

703   nuclease domain, "ATP" = gp1 vs TerL ATPase domain. **E.** Histogram showing quantification of the

704   interactions shown in panel A by β-galactosidase assay. Statistical significance is shown above the chart

705   (ANOVA, n=3, ns=not significant, *** $p < 0.01$).

706   **Figure 4. Comparison of the major structural proteins in wild-type RcGTA vs a gp1 knock-out.**

707   SDS PAGE gel of affinity purified RcGTAs produced by DE442 wild-type [**WT**] and *g1* deletion strains

708   [**Δg1**]. Expedeon Tri-Color marker is included for size comparison [**M**], with approximate molecular

709   weights annotated to the left of the gel. Bands predicted to contain the RcGTA portal and capsid proteins

710   are annotated, as well the tail fibre (**GTA TF**) that was confirmed by MALDI mass spectrometry.

711   **Figure 5. Gp1 is a prerequisite for RcGTA assembly. A & B.** Structural proteome of RcGTA particles.

712   LC-MS:MS analysis of affinity purified RcGTA particles produced by DE442 wild-type [**WT**] and *g1*

713   deletion [**Δg1**] strains. RcGTA head proteins and tail proteins are shown separately in panels Band A and

714   B, respectively. **D-H.** Transmission electron micrographs of RcGTA particles. Images in panels D-G were

715   taken at 68,000x magnification and the scale bar in panel C represents 50 nm. The panel H image was taken

716   at 49,000x magnification and the scale bar represents 100 nm. Black arrow heads indicate head spikes and

717   white arrows indicate portal apertures.

718   **Figure 6. DNA content of affinity purified RcGTA particles.** Agarose gels of DNA extracted directly

719   from chimeric his6-tagged RcGTAs are shown. His-tags were incorporated into nascent RcGTA particles

720   by ectopic expression of his6-tagged gp5 (**A**) or gp1 (**B**) proteins. Tagged RcGTAs were purified from *R.*

721   *capsulatus* DE442 culture supernatants using nickel agarose affinity chromatography. The genotype of the

722   producer cells is indicated directly above each gel – wild-type (**WT**) or RcGTA *g1* gene knock-out (**Δg1**).

723   DNA marker hyperladder 1 kb is included for reference (**M**); the locations of the 4 kb reference band and

724   GTA DNA are annotated.

725   **Figure 7. RcGTA gp1 *in vitro* DNA binding. A.** Representative agarose gel (0.8% w/v) showing the

726   stated concentrations of gp1 protein binding to DNA in an electrophoretic mobility shift assay (EMSA).

727   The locations of unbound and shifted DNA are annotated. Substrate DNA in the assay shown is a 1.4 kbp

728   PCR amplification of an arbitrarily chosen region flanking the *rcc01398* gene from *R. capsulatus*

729   (amplified using *rcc01398* F & R primers, Table 3). Bioline hyperladder 1 kb DNA marker is shown for

730   size comparison [**M**]. **B.** Quantification of EMSAs by band intensity analysis. Data is show is the average

731   of two EMSAs carried out independently in time and with different DNA substrates (*rcc01397* and

732   *rcc01398* genes). Individual data points are plotted as well as the mean line.

733   **Figure 8. GTA TerS protein alignment.** COBALT multiple sequence alignment of putative GTA small

734   terminases from Rhodobacterales species *Dinoroseobacter shibae* (DsGTA), *Oceanicola granulosus*

735   (OgGTA), *Rhodobacter capsulatus* (RcGTA) and *Ruegeria pomeroyi* (RpGTA). Intensity of colour for

736   each amino acid is based on percentage identity. Predicted secondary structure is indicated below the

737   alignment with "h" indicating helical.

738 **Table 1. Predicted small terminases from known GTAs**

Host Species	Gene Name	Protein Accession	Size (kDa)	Size (aa)
Rhodobacter capsulatus	*rcc001682*	ADE85427	11.5	107
Oceanicola granulosus	*OG2516_RS04255*	EAR49554	12.9	114
Ruegeria pomeroyi	*SPO2267*	AAV95531	12.6	114
Parvularcula bermudensis	n/a	CP002156* (1595455-1595796)	13.1	114
Oceanicaulis alexandrrii	*OA2633_14800*	EAP88801	10.1	92
Methanococcus voltae	*Mvol_0412*	ADI36072	14.6	125
Desulfovibrio desulfuricans	*Ddes_0720*	ACL48628	13.5	125
Bartonella grahamii	*Bgr_16770*	WP_041581600	11.9	107
Bartonella australis	*BAnh1_10950*	AGF74963	11.7	109
Dinoroseobacter shibae	n/a	CP000830* (2306070..2306432)	12.9	121
Aeromonas **phage 44RR**	*gene 16*	NP_932507	17.3	154
Enterobacteria phage T4	*gene 16*	NP_049775	18.4	164
Bacillus **phage SF6**	*gene 1*	CAK29441	16.0	145
Bacillus **phage SPP1**	*gene 1*	CAA39536	20.8	184

739

740 Grey rows are well characterized phage small terminases included for comparison

741 * Where no gene/protein has previously been annotated the accession number for the bacterial genome is

742 provided with the nucleotide position of the new ORF indicated in brackets.

743 **Table 2. Plasmids used in this study.**

Name	Description	Reference
pCM66T	pCM66T was a gift from Mary Lidstrom	Addgene plasmid #
	Broad host range vector; ColE1, OriV, IncP/traJ, Kanamycin^R	74738
pUT18C	Bacterial two hybrid vector	(41)
pKT25	Bacterial two hybrid vector	(41)
pEHisTEV	Expression vector; T7 promoter, His6 tag, TEV cleavage site, Kanamycin^R	(63)
pETFPP_22	Expression vector; T7 promoter, His6/MBP tags, 3c cleavage site, Kanamycin^R	(64)
pCMF170	RcGTA promoter fused to RcGTA *g1* in pCM66T	This Study
pJXL1	RcGTA promoter fused to RcGTA *g1*Δh1 in pCM66T	This Study
pJXL2	RcGTA promoter fused to RcGTA *g1*Δh3 in pCM66T	This Study
pCMF143	T25 fused to RcGTA *g1* in pKT25	This Study
pJXL3	T25 fused to RcGTA *g1*Δh1 in pKT25	This Study
pJXL4	T25 fused to RcGTA *g1*Δh3 in pKT25	This Study
pCMF144	T18 fused to RcGTA *g2* in pUT18C	This Study
pCMF238	T18 fused to RcGTA *g2* ATPase domain in pUT18C	This Study
pCMF239	T18 fused to RcGTA *g2* nuclease domain in pUT18C	This Study
pCMF153	His6-RcGTA *g1* in pEHisTEV	This Study
pCMF166	His6-MBP-RcGTA *g1* in pETFPP_22	This Study
pCMF142	RcGTA promoter fused to RcGTA *g5*-His6 in pCM66T	This Study
pCMF173	RcGTA promoter fused to RcGTA *g1*-His6 in pCM66T	This Study
pCMF172	RcGTA *g1* flanking DNA interrupted with Gentamycin^R in pCM66T	This Study

744

Table 3. Oligonucleotides used in this study.

| Name	Sequence (5'-3')
pGTA F*[1]	CGACTCTAGAGGATCGATTGTCGATCAGATCAC
pGTA R*[1]	GCTGACCATCGCCAGGGCCAGTTCC
g1 (66T) R	CGGTACCCGGGGATCTCAACCTCCTGCGGCGTC
pGTA g1 Δh1 inv F	CAAGACATGAAAGGGGTTCGCCAG
pGTA g1 Δh1 inv R	CCCTTTCATGTCTTGCGTGACCCG
pGTA g1 Δh3 R	CGGTACCCGGGGATCCTAACCGGCAACTTGTCTGC
T25-g1 F	CGACTCTAGAGGATCTGAAAGGGGTTCGCCAG
T25-g1 R	AGGTACCCGGGGATCTCAACCTCCTGCGGCGTC
T25-g1 Δh1 F	CGACTCTAGAGGATCTGGACATGGGGTTCAAG
T25-g1 Δh3 R	AGGTACCCGGGGATCCTAACCGGCAACTTGTCTGC
T18C-g2 F	CGACTCTAGAGGATCTGGGGGGGGCTTGGGAACAAT
T18C-g2 R	CGGTACCCGGGGATCTCAAAGCCCGCGCACCTG
T18C-g2 Nuc F	CGACTCTAGAGGATCGTATGGTTCTGCTGGAGGATGTC
T18C-g2 ATP R	CGGTACCCGGGGATCCTAGACATCCTCCAGCAGAAC
H6-g1 F*[2]	TTTCAGGGCGCCATGGACATGGGGTTCAAG
H6-g1 R*[2]	CCGATATCAGCCATGTCAACCTCCTGCGGCGTC
MBP-g1 F	TCCAGGGACCAGCAATGGACATGGGGTTCAAG
MBP-g1 R	TGAGGAGAAGGCGCGGTCAACCTCCTGCGGCGTC
g1-H6 R	CGGTACCCGGGGATCTCAATGGTGATGGTGATGGTGACCTCCTGCGGCGTCGCG
g5-H6 F	CTGGCGATGGTCAGCATGAAGACCGAGACCAAG
g5-H6 R	CGGTACCCGGGGATCTTAGTGATGGTGATGGTGATGCGAGGCGGCAAACTTCAAC
g1 UP R	GGGAATCAGGGGATCCTGGCGAACCCCTTTCAT
g1 DOWN F	AACAATTCGTTCAAGAGACAAGTTGCCGGTGTCG
g1 DOWN R	CGGTACCCGGGGATCGTCCAAATACGCCCTTGCG
Gent F	GATCCCCTGATTCCCTTTGT
Gent R	CTTGAACGAATTGTTAGG

rcc01397 **F**[*3]	CGACTCTAGAGGATCCCAGCGCGTAGATCGACG
rcc01397 **R**[*3]	CGGTACCCGGGGATCGCGATTGCCAACATCGCC
rcc01398 **F**[*4]	CGACTCTAGAGGATCCGCTTTCGCCTGCGCCTGC
rcc01398 **R**[*4]	CGGTACCCGGGGATCCTCGGCATGGATCCAGTGC
gafA **F**[*5]	CGACTCTAGAGGATCAGGAAGCCCTTGCCATAGG
gafA **R**[*5]	CGGTACCCGGGGATCGCGAAGCTGGAGTTCAACC
rcc00555 **F**[*6]	TAATCGCGGCCTCGAATCGTCATCGACCTGAAGGC
rcc00555 **R**[*6]	ATTTTGAGACACAACCGAAATCAGGTTAACGATCC

746   * Primers used to generate EMSA substrate DNA, superscript numbers 1-6 indicate primer pairs

747

Figure 1
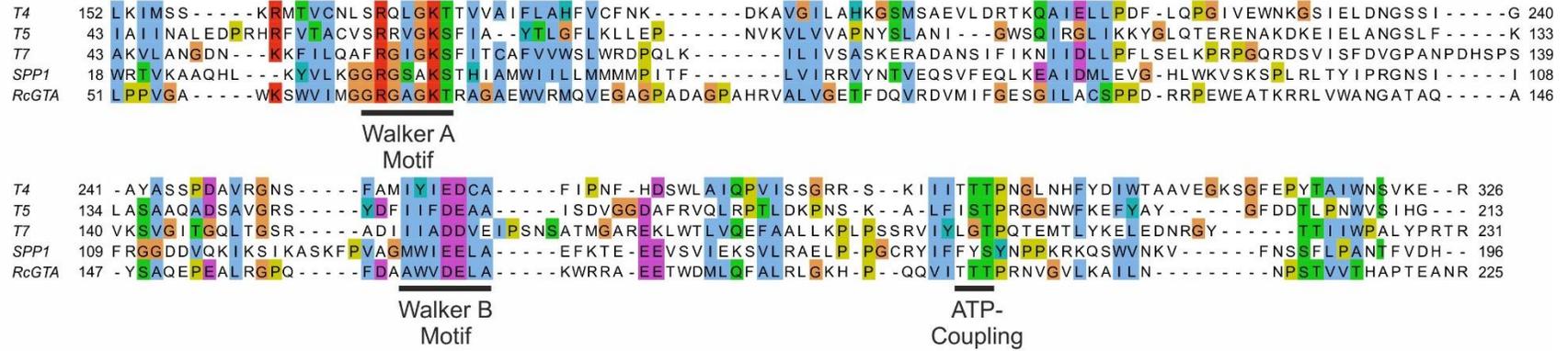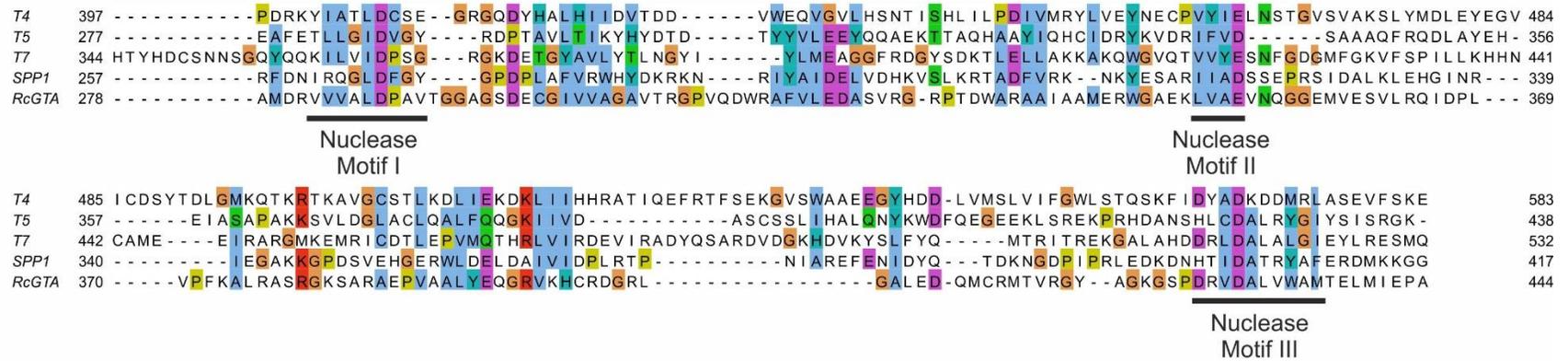
Figure 2

43
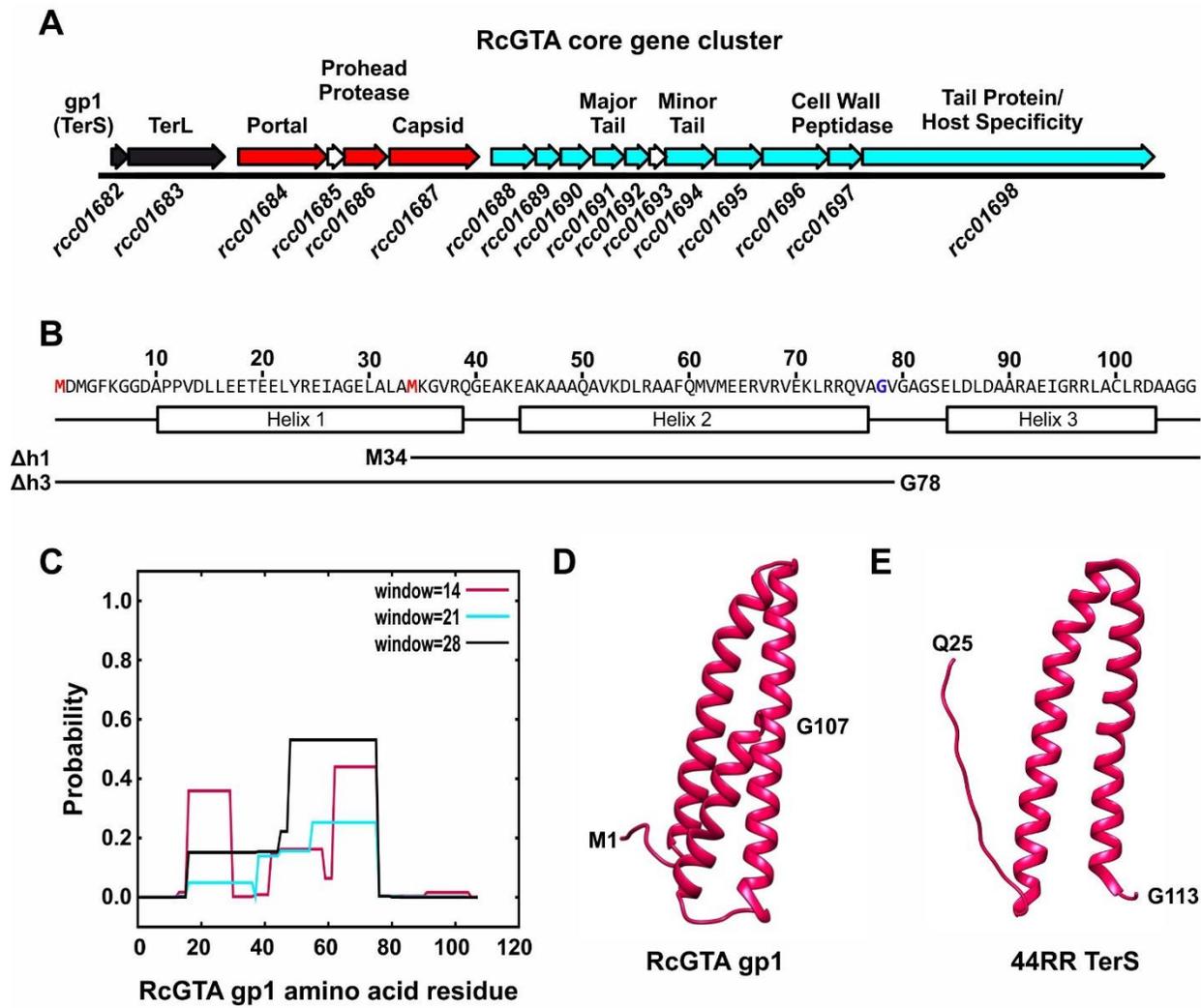
Figure 3
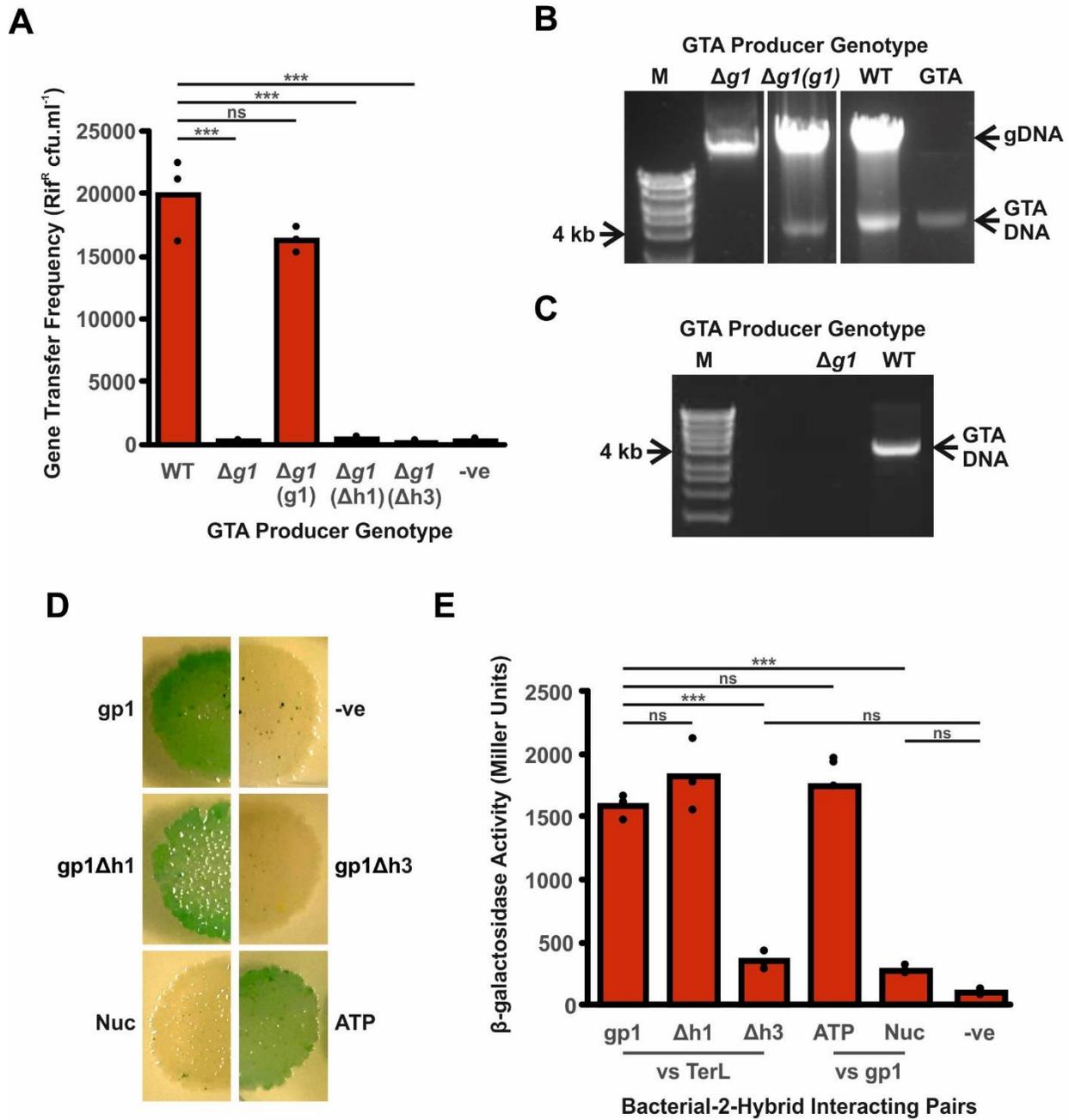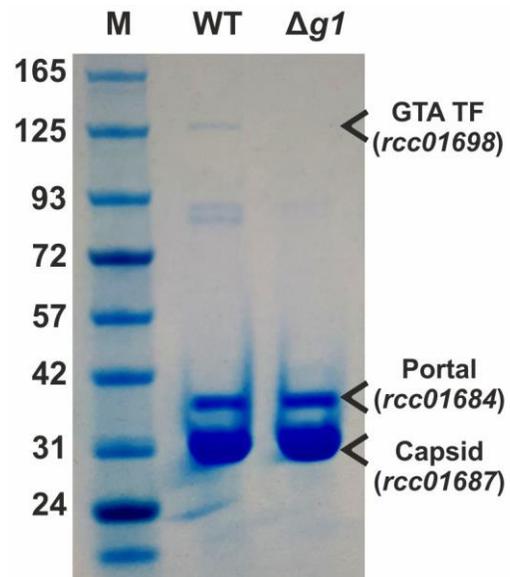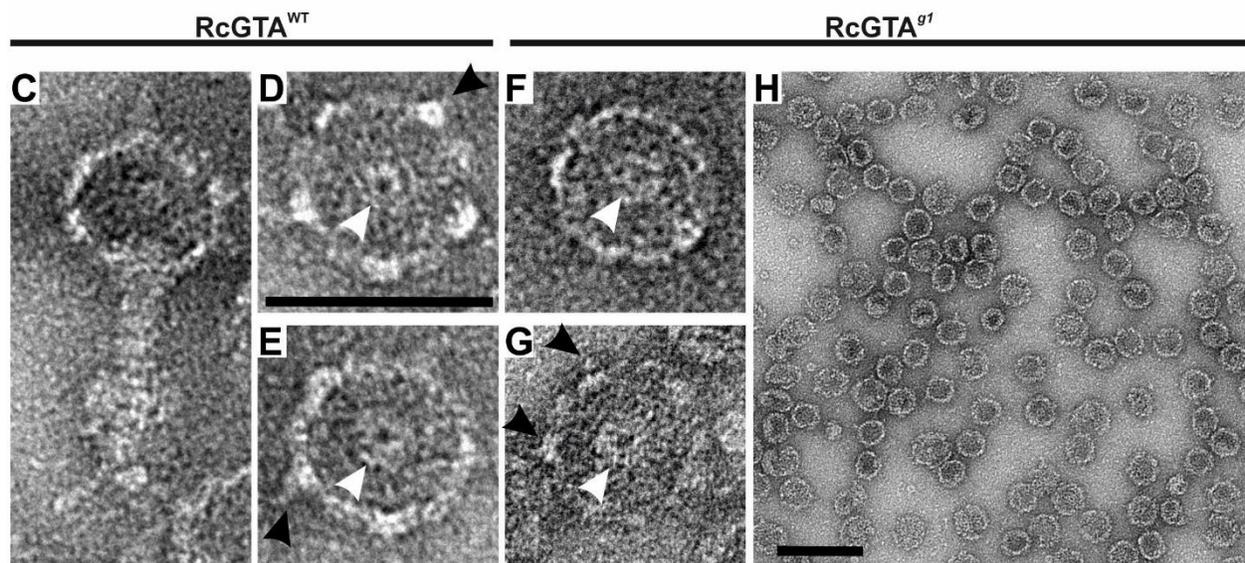
759   **Figure 4**

760
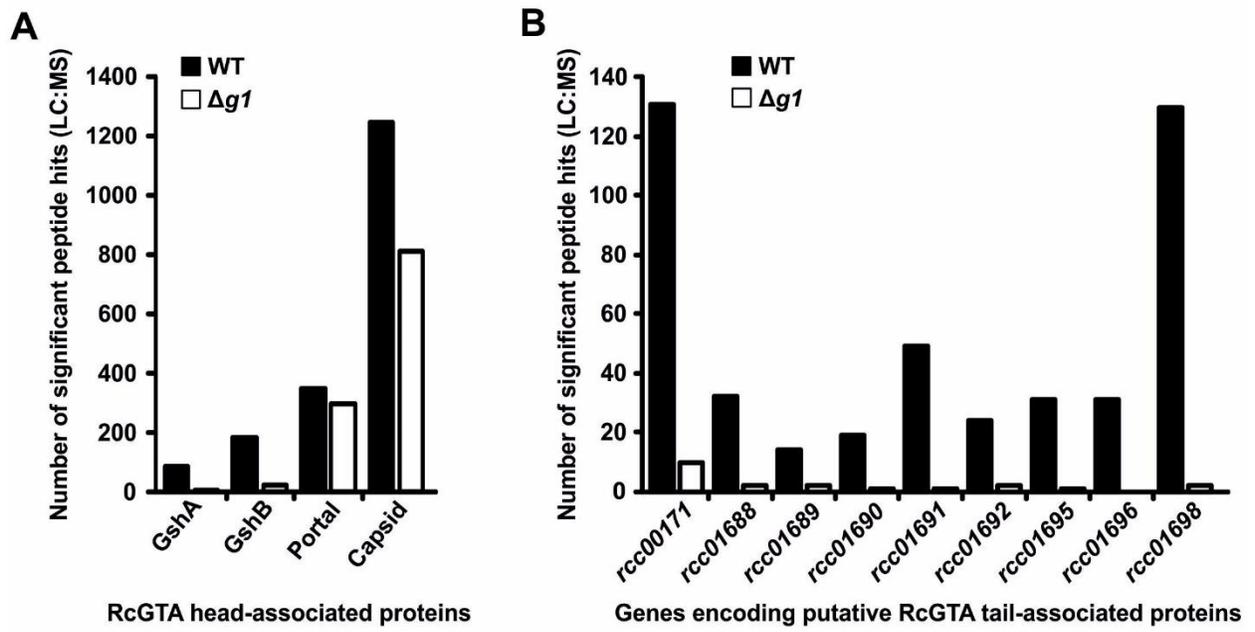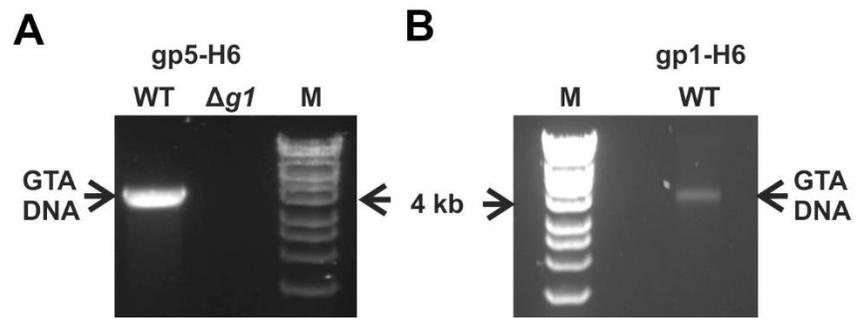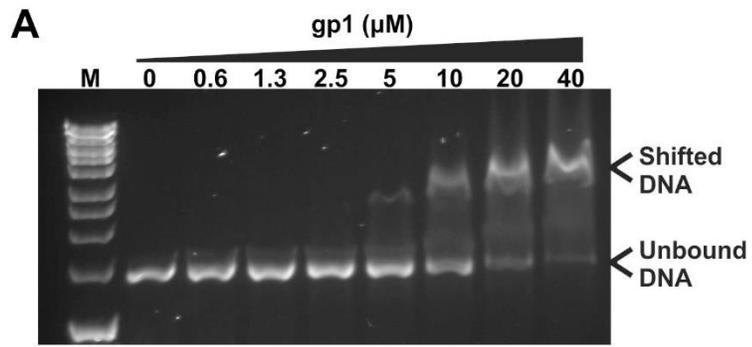

761

762

763   **Figure 5**

764

765

766

767   **Figure 6**

768

769

770

771   **Figure 7**

772

773

774

775 **Figure 8**

776

		10	20	30	40	50	60	70	
DsGTA	1	MIG-DTGTGACQPG---------SFLDAATDQVVYLRNCVQCAIVRVEELARTGTPNDTSAGEFRKLLKEL	61
OgGTA	1	MSR-------PEPERDPDREAWDLLIHEREQLLRATGETL----AEMVERLRGGEG--GDFRKMVSKAGDV	58
RcGTA	1	MDMGFKG-GDAPPV---------DLLEETEELYREIAGEL----ALAMKGVRQGEA--KEAKAAAQAVKDL	55
RpGTA	1	MTL-------ITPE---------ERISRTAELLQSLENSIRDLRNAAEDLQKRIRA--GEDGDLAGYGKQM	53
Structure		hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh hhhhhhhhhhh

		80	90	100	110	120	130	
DsGTA	62	REISG---IALREESRLAEQLAKENGGL-HAG-AYDLVAARAEIGRRLADLRTARSHPDVSGEPE	124
OgGTA	59	EFALR---KMIEIREKY-DDWHAKRGGELTGN-RFDADDARADIGRKLDRLRDAGGAGGVS----	117
RcGTA	56	RAAFQ---MVMEERVRV-EKLRRQVAGVGAGS-ELDLDAARAEIGRRLACLRDAAGG--------	110
RpGTA	54	GQAASLIRECQKVEASFAEQVRREAGIA-QGGYALDLDRARSEIGCRLARLRKCCREGAVSE---	117
777 | Structure | | hhhhh   hhhhhhhhhhhhhhhhhhh   hhhhhhhhhhhhhhhhhhhhhh |

49