# Identification of the first gene transfer agent (GTA) small terminase in *Rhodobacter capsulatus*, its role in GTA production and packaging of DNA

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

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

Running Title: Identification of the first GTA TerS

Keywords: Horizontal Gene Transfer, Bacterial Evolution, Gene Transfer Agent, *Rhodobacter*, Antimicrobial Resistance, Terminase, DNA Packaging, Bacteriophage, Transduction, Virus

Abstract = 122 words, Text = 6,631 words

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

*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-like genes are present throughout the alpha-proteobacteria and appear to have co-evolved with the host species, indicative of vertical inheritance (17, 18). Beyond the alpha-proteobacteria, functional GTAs have since been discovered experimentally in diverse prokaryotes, including animal pathogens of the *Brachyspira* genus (Spirochete) (19), *Desulfovibrio* spp. (delta-proteobacteria) (20) and the Archaeon *Methanococcus voltae* (21, 22). Each of these disparate GTAs was identified by chance during the study of phage-like particles or unusual levels of gene transfer. However, it is extremely difficult to systematically identify GTAs by bioinformatics alone because they are functionally analogous but genetically divergent from each other and their genes strongly resemble remnant bacteriophages. The difficulty of rapidly identifying GTAs is perhaps the major obstacle for expanding the breadth of research carried out on GTA producers.

The packaging of random bacterial DNA by GTAs is a fundamentally different behaviour to that of bacteriophages and other viruses (23). The primary aim of a phage is to distribute their own genes. Phages first replicate their genome, usually as a multi-copy concatamer. There is no evidence that GTAs possess any DNA replication genes or that the packaged DNA has been replicated, instead GTAs appear to contain the uncopied genome of the producing bacterium. For viruses, in all known cases the volume of the capsid is enough to contain the whole viral genome, however this is not the case for GTAs (7). An individual GTA virion is too small to package the genes required for its own synthesis, for example each RcGTA transfers only ~4 kb of DNA but the 14.5 kb core gene cluster plus several ectopic loci are required for mature GTA production. To achieve packaging specificity, dsDNA phages usually use initiation sites at a specific location in the phage genome that are recognized by the packaging machinery. Packaging initiation sites generate specificity with a defined DNA sequence, e.g. *cos*/*pac* sites (23–25), or with favourable topological features, e.g. conformational selection of intrinsically bent DNA by SPP1-like phages

(26). So far no evidence that GTAs target discrete packaging start sites has been presented and no conserved sequences or topologies have been implicated as *cos/pac* equivalents, all of which suggests that packaging initiation is indeed random.

Bacteriophages with a dsDNA genome use sophisticated molecular machinery, known as the terminase, to specifically recognize replicated phage DNA and to drive it into a preformed capsid (27). The capsid itself is essentially a passive receptacle and it is the terminase that provides DNA selectivity, enzymatic activity and motive force required to fill the capsid. The terminase is a complex of two oligomeric small and large terminase proteins, TerS and TerL, which are both indispensable for proper phage function and DNA packaging (27). TerL possesses the enzymatic activities required for DNA packaging: it has a C-terminal nuclease domain that cleaves the target DNA to produce a free end available for packaging and an N-terminal ATPase domain that translocates the DNA into a preformed capsid. Unlike TerL, TerS has no enzymatic activity and instead carries out a regulatory role being responsible for recognition of the phage genome's packaging initiation site, recruitment of TerL and modulation of TerL enzymatic activities (26, 28, 29).

In general, large terminase genes are sufficiently well conserved to allow confident identification by sequence identity alone, partly owing to the presence of the Walker ATP-interacting motifs, and thus most GTAs have an annotated *terL* gene. Small terminases, however, are smaller with little primary sequence conservation, which makes them far more challenging to identify *in silico*. No small terminase has been identified for any GTA to date. Given the role of terminase proteins in phage biology, it is highly likely that the GTA terminase plays a defining role in the packaging of random DNA. In this study, we definitively identify the small terminase of the model *R. capsulatus* GTA, demonstrate and localize its interaction with the large terminase and investigate its role in RcGTA production. Our characterization of the RcGTA TerS also allows us to speculate on the physical requirements of a GTA-type small terminase and to identify candidate small terminases in other GTA-producing species

## Results

**Characterization of RcGTA *g1* (*rcc01682*).** A gene encoding a TerL homologue (*rcc01683/*RcGTA *g2*) is readily identifiable within the *Rhodobacter capsulatus* SB1003 core RcGTA gene cluster (Fig. 1A). The RcGTA TerL has regions of strong homology with large terminases from several well-studied phages, including the presence of characteristic nuclease and Walker ATPase motifs (Fig. 2) (30, 31). Small terminases are far more difficult to predict and consequently no GTA small terminases have ever been identified. Most characterized phage TerS proteins have a modular structure: the N-terminal region comprises the helix-turn-helix DNA-binding domain, the central region contains a coiled-coil oligomerization domain and the C-terminus contains the TerL interaction segment (32). Such domain organization is a well conserved feature of TerS, despite the lack of sequence conservation.

In phage genomes, the small and large terminase genes are often co-localized and so the core RcGTA gene cluster was examined for genes that could encode a small terminase. The RcGTA gene cluster contains 17 predicted genes, of which at least six have been shown to be essential for GTA production (16). The first gene of the cluster, *rcc01682* (referred to hereafter as *g1* and the protein as gp1), is also thought to be essential for RcGTA activity (33), but no in-depth characterization has been carried out and no function has so far been assigned. The *g1* ORF is 324 bases and is located immediately upstream of the large terminase. The gp1 protein sequence was submitted to the JPRED4 protein secondary structure prediction server (34), which predicted an almost entirely helical structure (Fig. 1B). Subsequent analysis using the COILS server (35) (MTIDK matrix, all window sizes) indicated that of the three distinct α-helices, the first two are likely to form a coiled-coil (Fig. 1C) reminiscent of a phage TerS oligomerization domain. A more detailed structural prediction using the RaptorX structure prediction server (36) indicated similarity to the phage T4-like small terminase from *Aeromonas* phage 44RR (Fig. 1D & E). The 44RR TerS crystal structure (PDB: 3TXS) failed to resolve the N/C-terminal segments of residues 1-24 and 114-154 due to conformational variability, however, an N-terminal helix-turn-helix DNA-binding motif was predicted from the primary sequence (32). RcGTA gp1 begins with the coiled-coil domain and appears to lack a DNA-binding domain (Fig. 1D) at the N-terminus. No helix-turn-helix motif was detected by the Gym2.0 and NPS@ servers (37–39).

To confirm that *g1* is essential for RcGTA activity, a deletion mutant was produced in the GTA hyperproducer strain *R. capsulatus* DE442. Loss of the *g1* gene prevented all detectable gene transfer activity (Fig. 3A). Complementation with full length *g1* expressed ectopically from its own promoter effectively restored gene transfer to wild-type frequencies (Fig. 3A). Complementation was also attempted using *g1* constructs that lacked the sequence encoding either the first or third α-helical regions; in both cases gene transfer frequencies were indistinguishable from the uncomplemented DE442 Δ*g1* mutant (Fig. 3A).

It has previously been shown that the DE442 RcGTA hyperproducer packages sufficient genomic DNA into GTA particles to allow detection of a distinct 4 kb band in total DNA preparations (40). Given the predicted headful packaging mechanism used by RcGTA (8, 27), production of 4 kb DNA fragments can only occur if DNA is successfully packaged into the capsid. This property can be exploited to examine mutations that affect DNA packaging *in vivo*, independent of the release of infective GTA particles. Deletion of *g1* prevents any detectable accumulation of intracellular RcGTA 4 kb DNA fragments and *in trans* complementation restores DNA packaging (Fig. 3B). DNA contained within extracellular GTA particles is protected from enzymatic degradation. Isolation of DNase-insensitive DNA from the supernatant of DE442 wild-type and Δ*g1* strains yielded detectable RcGTA DNA for the wild-type only (Fig. 3C). As phage

TerS are responsible for binding to target DNA and stimulating the various enzymatic activities of the TerL that are required for DNA packaging, our data are entirely consistent with *g1* encoding the RcGTA small terminase.

**The C-terminus of RcGTA gp1 interacts with the ATPase domain of TerL.** In bacteriophage, the only protein that the small terminase is known to interact with is the large terminase. Indeed, the small terminase not only recognizes the bacteriophage DNA, but also recruits the large terminase and initiates the process of DNA packaging. Using the bacterial-2-hybrid assay, RcGTA gp1 was translationally coupled to the T25 domain of the *Bordetella pertussis* adenylate cyclase enzyme and RcGTA TerL was coupled to the adenylate cyclase T18 domain. Interaction between the two proteins brings together the two adenylate cyclase domains leading to cAMP production and subsequently β-galactosidase (41). In this assay, a distinct interaction can be seen between gp1 and gp2 (Fig. 3D). Truncation of gp1 to remove helix 1 had no appreciable effect on interaction with the large terminase, however, loss of helix 3 led to complete loss of interaction (Fig. 3D). Quantification of the results with a colorimetric β-galactosidase assay showed no significant difference between the helix 3 deletion and the no insert negative control, whereas the helix 1 deletion was indistinguishable from full length gp1 (Fig. 3E).

The RcGTA large terminase has clear homology with large terminase proteins of several well-studied phages (Fig. 2). The N-terminus of the protein contains the ATPase domain with conserved Walker A and B motifs (Fig. 2A), while the C-terminus contains the nuclease domain including three conserved nuclease motifs (Fig. 2B) (30). In the well-studied T4-like phages, it is the ATPase domain that directly interacts with the small terminase (42). To test whether the ATPase domain of the RcGTA large terminase is also responsible for interaction with gp1, translational fusions were made of each of the two domains with the adenylate cyclase T18 domain.

In a bacterial-2-hybrid assay, the TerL nuclease domain (V253-L455) had no significant interaction with RcGTA gp1 but the ATPase domain (L27-V258) produced a signal indistinguishable from full-length TerL (Fig. 3D & E).

**RcGTA gp1 production is a prerequisite for tail attachment and efficient GTA capsid maturation.** As shown above, Δ*g1* mutants are unable to produce infective GTA particles or to package DNA, which indicates that RcGTA production has stalled early in the assembly process. To determine the developmental state of the stalled RcGTAs, we purified the RcGTA particles that were released by DE442 WT and Δ*g1* strains during lysis using nickel affinity purification. The RcGTA lysis genes (*rcc00555* and *rcc00556*) are located elsewhere in the *R. capsulatus* genome and should not be affected by the absence of a small terminase (8, 43). A plasmid containing the RcGTA capsid (*rcc01687*/RcGTA *g5*) with a C-terminal His6-tag was introduced into wild type DE442 and isogenic Δ*g1* strains. Timing of capsid expression was matched to GTA production by fusing the *g5* ORF directly to the previously characterized RcGTA promoter (40, 44). Incorporation of recombinant capsid monomers into nascent RcGTA particles allows affinity purification of the whole particles, as previously described (15). Concentrated samples were run on an SDS PAGE gel to qualitatively assess the relative protein content. Strong bands were evident in both samples at sizes consistent with the RcGTA capsid (post-translationally processed to 31.4 kDa (45)) and portal (42.8 kDa) proteins (Fig. 4). RcGTA$^{\mathrm{WT}}$, but not RcGTA$^{g1}$, also had several other visible bands (Fig. 4). RcGTA particles contain a distinctive 138.9 kDa putative tail fibre/host specificity protein (encoded by *rcc01698*/RcGTA *g15*) (45), and a band of this size was present only in the RcGTA$^{\mathrm{WT}}$ lane. The band was excised and positively identified as gp15 by MALDI-MS:MS (3 unique peptide hits, expect <0.05, total score 154).

Affinity purified RcGTA particles were submitted for shotgun liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis to determine the structural proteome of RcGTA$^{\mathrm{WT}}$ versus RcGTA$^{g1}$. In terms of number of peptides detected, both sample types yielded equivalent numbers for the RcGTA capsid and portal proteins (Fig. 5A). The GhsA and GhsB head spike proteins (encoded by *rcc01079* and *rcc01080,* respectively) were represented in both samples, however, 7 to 9-fold fewer GhsA/B peptides were detected in RcGTA$^{g1}$ (Fig. 5A). In contrast, peptide hits for the predicted RcGTA tail structures were almost completely absent in the RcGTA$^{g1}$ samples but abundant for RcGTA$^{\mathrm{WT}}$ (Fig. 5B). Transmission electron microscopy images corroborated the proteomic data. RcGTA$^{\mathrm{WT}}$ samples yielded intact GTA particles with clearly defined head spikes, portal apertures and dense staining of the heads, possibly indicative of tightly packaged DNA (Fig. 5C-E). RcGTA$^{g1}$ samples contained no evidence of tail structures, head spikes were present but at reduced frequency and portal structures were visible (Fig. 5F-H). Overall, RcGTA$^{g1}$ head structures appeared more prone to damage than wild-type, maturation was often incomplete and the contrast was poor - probably due to the absence of DNA (Fig. 5F-H). In agreement with data presented earlier (Fig. 3C), DNA extraction from affinity purified RcGTA$^{\mathrm{WT}}$ samples yielded characteristic 4 kb GTA DNA bands whereas no detectable DNA was recovered from RcGTA$^{g1}$ samples (Fig. 6A). Similar affinity chromatography using His6-tagged gp1 also allowed purification of RcGTA particles from culture supernatant. The overall concentration of RcGTA particles was much lower, presumably because the terminase complex dissociates after packaging is complete, but 4 kb GTA DNA was still recoverable (Fig. 6B). These data demonstrate a direct interaction between gp1 and the broader structural proteome for the first time, and support our hypothesis that gp1 is indeed the small terminase.

**RcGTA gp1 binds weakly to DNA.** A core role of phage small terminases is to recognize the phage genome and to target it for packaging into preformed capsids. RcGTAs don't package specific DNA but the large terminase still needs to be recruited to the host genomic DNA to initiate packaging, and it's plausible that this may be achieved via a non-specific affinity for DNA. In an electrophoretic motility shift assay (EMSA), we tested the ability of RcGTA gp1 to bind DNA *in vitro*. To obtain high concentration, soluble protein an N-terminal MBP-tag was used for gp1 purification. Purified gp1 exhibited low affinity for DNA with incomplete shifts occurring at micromolar concentrations - 87% of DNA substrate was bound at 40 µM protein concentration (Fig. 7). Six EMSA DNA substrates were used (351 to 2,944 bp PCR amplicons from distinct locations in the *R. capsulatus* genome, however, the identity of the DNA did not substantially affect the binding affinity. The size of the observed shift in DNA mobility was ~1500 bp or equivalent to 975 kDa, which is greater than would be expected for binding of a single protein monomer. The large reduction in mobility of the gp1-DNA complex indicates that gp1 could be binding as an oligomer (small terminases usually form characteristic ring structures), there could be multiple occupancy due to the lack of a specific binding site and/or the conformation of the DNA may have been altered.

**GTA small terminases can be predicted in other species.** Identification of the RcGTA small terminase allowed us to predict GTA *terS* genes in other alpha-proteobacterial species (Table 1), including two previously unannotated ORFs in *Parvularcula bermudensis* and *Dinoroseobacter shibae.* Interestingly, we were also able to predict small terminase genes in the distantly related delta-proteobacterium *Desulfovibrio desulfuricans* and the Archaeon *Methanoccus voltae* (Table 1). In each case the small terminase gene was immediately upstream of the cognate large terminase, the coding sequence for each small terminase was ~10-50% shorter than comparable phage counterparts (Table 1) and the predicted protein structures were almost entirely helical. Overall, the primary amino acid sequences of the various small terminases is poorly conserved, even between those found in closely related species (Fig. 8). However, for the Rhodobacterales GTA TerS proteins there is clear sequence similarity localized at the C-termini, specifically the third α-helix (Fig. 8). Conservation of this region supports our findings that the C-terminal helix is required for interaction with TerL, and that this interaction constrains TerS sequence divergence.

## Discussion

Gene Transfer Agents clearly share many structural and mechanistic features with bacteriophages, however, the most striking difference is that GTAs package and transfer random fragments of host DNA without any preference for their own genome. In bacteriophages, DNA packaging is carried out by the terminase complex, which is composed of multimeric small and large terminase subunits. Interestingly, the Enterobacteria phage T4 large terminase can promiscuously package heterologous linear DNA fragments into an empty phage head *in vitro* when TerS is absent, reminiscent of GTA-type DNA packaging, but the presence of TerS is essential for terminase activity *in vivo* (47). The large terminase has all the enzymatic capabilities required to package DNA i.e. a nuclease domain to create free DNA ends at the beginning/end of packaging and an ATPase domain to act as a motor to feed the DNA into the capsid (27, 31, 46). These data demonstrate that TerS is not strictly required for the process of packaging DNA into the capsid but is instead crucial for regulation (28). Depending on the particular phage, TerS forms an oligomeric ring consisting of 8 to 11 identical protein subunits, with the DNA binding domains arranged around the exterior surface. The TerS ring recognizes the packaging signal in the phage genome and has been proposed to wrap ~100 bp of DNA around the outside, along the circular surface formed by the DNA binding domains (48). TerS recruits TerL to make the initial DNA double strand break, but inhibits further DNA cleavage to prevent damage to the phage genome. The TerS/L complex docks to the phage head via the oligomeric portal protein, which contains a narrow aperture for the DNA to be fed through. TerS stimulates TerL ATP hydrolysis and translocation of the packaging complex along the phage genome. Once the genome has been tightly packaged into the capsid, TerL cleaves the DNA again to complete the process. The terminase disassociates, the portal aperture is plugged and tail assemblies are attached.

Given the role that small terminases play in phage DNA specificity it is likely that a comparable protein is responsible for random DNA packaging by GTAs, however, no GTA TerS proteins have so far been identified. Taken together, the molecular, genetic, proteomic and imaging data presented here all support the hypothesis that RcGTA gp1 is the small terminase. RcGTA gp1 is essential for RcGTA gene transfer and DNA packaging (Fig. 3A-C). RcGTA gp1 is predicted to have structural characteristics in common with phage TerS proteins, in particular a putative coiled-coil domain that is important for oligomerization in phage (Fig. 1) and a conserved C-terminal large terminase interaction domain (Fig. 3D-E & Fig. 8). Analysis of the *R. capsulatus* DE442 Δ*g1* mutant also allows us to postulate a model to describe RcGTA assembly. RcGTA capsid formation and incorporation of the portal aperture occurs independently of the terminase and DNA packaging. Proteomic analysis of the stalled RcGTA$^{g1}$ particles did not detect substantial presence of the large terminase protein, which suggests that either gp1 recruits TerL to the DNA first and the terminase hetero-complex then recruits the preformed capsids, or that the interaction between TerL and the portal is labile in the absence of TerS. Once the terminase-portal-capsid complex is assembled, headful DNA packaging can begin. In the absence of DNA packaging, efficient maturation of the RcGTA heads is impaired and RcGTA production stalls before the tail appendage is attached (Fig. 5).

A crucial difference between the RcGTA small terminase and its phage counterparts is the apparent lack of an N-terminal DNA-binding domain (Fig. 1). Previous work showed that deletion of the N-terminal region of bacteriophage SF6 and SPP1 TerS proteins led to a significant reduction in DNA binding affinity *in vitro* (49), but some binding was still retained. In addition, both T4 and P22 TerS proteins have N- and C-Terminal DNA binding activities, with non-specific DNA binding dependent upon a nine residue region in the P22 C-terminus, R143–K151 (48, 50). Here, we show that the RcGTA TerS protein can bind non-specifically to DNA at micromolar concentrations (Fig. 7). Absence of the specific DNA binding domain but retention of non-specific DNA binding could provide an explanation for random DNA packaging by GTAs. It is possible that the RcGTA TerS protein is also able to bind specific DNA sequences, however, this could not be tested because we have no evidence to suggest that this occurs *in vivo* and no GTA binding sites are currently known.

In summary, RcGTA gp1 is the first GTA small terminase to be described to date. We hypothesize that GTA small terminases possess all of the regulatory abilities of phage small terminases but lack of an N-terminal DNA-binding domain abolishes DNA sequence specificity. Loss of the specific DNA binding region could allow non-specific binding of random DNA sequences, which is the defining characteristic of GTA-type TerS proteins. The greatest barrier to novel GTA identification and an understanding of their true prevalence in the environment, is the lack of an effective identification method. Based on the data gained from RcGTA, we were able to predict the small terminases from other known GTAs using gene size, neighbourhood and protein secondary structure prediction analyses. Confirmation and in-depth characterization of these proteins could allow us to pinpoint the defining characteristics of GTA-type terminases with a view to enhanced discovery of novel GTAs in existing genome datasets. Furthermore, we also anticipate that the smaller size and simpler organization of GTAs, compared to phages, will 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 against *E. coli*) + 10 µg ml$^{-1}$ kanamycin (plasmid selection). Plates were incubated o/n at 30°C then restreaked onto fresh agar to obtain single colonies.

**Gene Knock-Outs.** Knock-outs were created by RcGTA transfer. pCM66T plasmid constructs were created with a gentamicin resistance cassette flanked by 500-1000 bp of DNA from either side of the target gene. Assembly was achieved by a one-step, four component NEBuilder (NEB) reaction and transformation into NEB 10-beta cells. Deletion constructs were introduced into the RcGTA hyperproducer strain by conjugation and a standard GTA bio-assay was carried out to replace the intact chromosomal gene with the deleted version.

***Rhodobacter* Gene Transfer Assays.** In *Rhodobacter*, the assays were carried out essentially as defined by Leung and Beatty (2013) (54). RcGTA donor cultures were grown anaerobically with illumination in YPS for ~48 h and recipient cultures were grown aerobically in RCV for ~24 h. For overexpression experiments, donor cultures were first grown aerobically to stationary phase then anaerobically for 24 h. Cells were cleared from donor cultures by centrifugation and the supernatant filtered through a 0.45 µm syringe filter. Recipient cells were concentrated 3-fold by centrifugation at 5,000 x g for 5 min and resuspension in 1/3 volume G-Buffer (10 mM Tris-HCl (pH 7.8), 1 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM NaCl, 0.5 mg ml$^{-1}$ BSA). Reactions were carried out in polystyrene culture tubes (Starlab) containing 400 µl G-Buffer, 100 µl recipient cells and 100 µl filter donor supernatant, then incubated at 30°C for 1 h. 900 µl YPS was added to each tube and incubated for a further 3 h. Cells were harvested by centrifugation at 5,000 x g and plated on YPS + 100 µg ml$^{-1}$ rifampicin (for standard GTA assays) or 3 µg ml$^{-1}$ gentamicin (for gene knock-outs).

**DNA Purification.** To isolate total intracellular DNA, 1 ml samples of relevant bacterial cultures were taken for each nucleic acid purification replicate. Generally, sampling occurred during stationary phase but for overexpression experiments samples were taken 6 h and 24 h after transition to anaerobic growth. Total DNA was purified according to the Purification of Nucleic Acids by Extraction with Phenol:Chloroform protocol (53). To isolate extracellular DNA contained in RcGTA virions, *R. capsulatus* DE442 cultures (23 ml) were grown anaerobically with illumination in YPS for ~48 h at 30°C. Cells were cleared from the cultures by centrifugation at 15,000 x g for 10 min and the supernatant was filtered through a 0.45 µm syringe filter. RcGTAs were precipitated by addition of PEG8000 to a final concentration of 10% (w/v) and then incubated at 4°C for 1 h with continuous rolling. Precipitated RcGTAs were pelleted by centrifugation at 10,000 x g for 10 min. The pellet was resuspended in 500 µl G-Buffer. Bacterial DNA and RNA was removed by overnight incubation with Basemuncher nuclease (Expedeon) in the presence of 10 mM $MgCl_2$ at 30°C. Nuclease digestion was inhibited by addition of 50 mM EDTA. DNA was extracted with Phenol:Chloroform:Isoamyl Alcohol (25:24:1, pH 8.0) as previously described (53).

**Bacterial-two-hybrid (B2H) assays.** The procedure and the resources were as described in (41). Plasmids encoding T18 (pUT18C and derivatives) and the compatible plasmids encoding T25 (pKT25 and derivatives) were introduced pairwise into competent BTH101 by co-transformation. Selection was using LB agar containing 50 µg/ml kanamycin, 100 µg/ml ampicillin, 1 mM IPTG and 80 µg/ml X-Gal, and plates were incubated at 30°C for 24-48 h. The phenotype of BTH101 (*cya*-) can be complemented if the two domains of adenylate cyclase (T18 and T25) are brought into close proximity, and this can be achieved by fusing interacting protein partners to each domain. The readout for complementation of the *cya*- phenotype (indicating a positive interaction between the two fusion proteins) is the induction of *lac* (blue colonies on IPTG, XGal), whereas 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 $Na_2HPO_4$, 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 $Na_2HPO_4$, 40 mM $NaH_2PO_4$, 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 ($\Delta OD_{420}$/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

Tris-HCl (pH 8), 1 mM $MgCl_2$, 1 mM $CaCl_2$ and 1 mM NaCl), then incubated at room temperature for 1 h with agitation. Beads were applied to 25 ml gravity flow columns (Thermo-Fisher) and washed with 200 ml G* buffer supplemented with 40 mM imidazole. RcGTA particles were eluted using 5 ml G* buffer supplemented with 400 mM imidazole. Eluted RcGTAs were concentrated and imidazole depleted to <1 mM by iterative dilution and ultrafiltration using a 100 kDa Spin-X UF20 device (Corning).

**Electron Microscopy.** Affinity purified RcGTAs were directly applied to 200 mesh copper grids with a formvar/carbon support film and allowed to adsorb for four minutes. The grids were washed with three drops of deionised water and then negatively stained with uranyl acetate solution (55). Samples were analyzed on a Tecnai 12 BioTWIN G2 transmission electron microscope operating at 120kV, and images were captured using the Ceta camera (Thermo-Fisher).

**Protein Purification.** Protein overexpression was carried out in the NEB Express *E. coli* strain (NEB) containing the relevant T7 expression plasmid (Table 2). Expression from the T7 promoter was induced at mid-exponential growth phase with 0.2 mM IPTG at 20°C overnight. His6-tagged (56) and MBP-tagged (40) proteins were purified as described previously. All chromatography steps were carried out on an AKTA Prime instrument (GE Healthcare). Purified proteins were concentrated in a Spin-X UF Centrifugal Concentrator (Corning). Samples were stored at -80°C in binding buffer plus 50% glycerol.

**Electrophoretic motility shift assays (EMSA).** DNA substrates were prepared by PCR amplification with oligonucleotides indicated in Table 3 and cleaned with a Monarch DNA clean-up kit (NEB). 10 µl EMSA mixtures contained 100 ng of DNA, binding buffer based on reference (57) (25 mM HEPES, 50 mM K-glutamate, 1 mM dithiothreitol, 0.05% Triton X-100, 4% Glycerol, 1 µg poly dI:dC; pH 8.0) and purified protein at stated concentrations. Binding assays were carried out at room temperature for 30 min. Samples were run on a 0.8% agarose gel in 0.5X TBE at 80 V for 2 h at room temperature. Gels were stained with Sybr Safe (Invitrogen) and imaged on a GelDoc transilluminator (BioRad).

**Sample Preparation for Mass Spectrometry.** For MALDI-MS:MS protein identification, purified RcGTA samples were run on a TEO-Tricine 4-12% SDS Mini Gel (Expedeon) at 150 V for 45 minutes. Gels were stained with InstantBlue protein stain (Expedeon) for a 1 h before destaining with ultrapure water for 1 h. Protein bands of interest were excised. For shotgun LC-MS:MS, samples were run into a 7 cm NuPAGE Novex 10% Bis-Tris Gel (Life Technologies) at 200 V for 6 mins. Gels were stained with SafeBLUE protein stain (NBS biologicals) for 1 h before destaining with ultrapure water for 1 h.

In-gel tryptic digestion was performed after reduction with DTE and S-carbamidomethylation with iodoacetamide. Gel pieces were washed two times with 50% (v:v) aqueous acetonitrile containing 25 mM ammonium bicarbonate, then once with acetonitrile and dried in a vacuum concentrator for 20 min. Sequencing-grade, modified porcine trypsin (Promega) was dissolved in the 50 mM acetic acid supplied by the manufacturer, then diluted 5-fold by adding 25 mM ammonium bicarbonate to give a final trypsin concentration of 0.02 µg/µl. Gel pieces were rehydrated by adding 10 µl of trypsin solution, and after 5 min enough 25 mM ammonium bicarbonate solution was added to cover the gel pieces. Digests were incubated overnight at 37°C.

**MALDI-MS:MS.** A 1 µl aliquot of each peptide mixture was applied to a ground steel MALDI target plate, followed immediately by an equal volume of a freshly-prepared 5 mg/mL solution of 4-hydroxy-α-cyano-cinnamic acid (Sigma) in 50% aqueous (v:v) acetonitrile containing 0.1% , trifluoroacetic acid (v:v).

Positive-ion MALDI mass spectra were obtained using a Bruker ultraflex III in reflectron mode, equipped with a Nd:YAG smart beam laser.  MS spectra were acquired over a range of 800-4000 m/z.  Final mass spectra were externally calibrated against an adjacent spot containing 6 peptides (des-Arg1-Bradykinin, 904.681; Angiotensin I, 1296.685; Glu1-Fibrinopeptide B, 1750.677; ACTH (1-17 clip), 2093.086; ACTH (18-39 clip), 2465.198; ACTH (7-38 clip), 3657.929.). Monoisotopic masses were obtained using a SNAP averagine algorithm (C 4.9384, N 1.3577, O 1.4773, S 0.0417, H 7.7583) and a S/N threshold of 2.

For each spot the ten strongest precursors, with a S/N greater than 30, were selected for MS/MS fragmentation. Fragmentation was performed in LIFT mode without the introduction of a collision gas. The default calibration was used for MS/MS spectra, which were baseline-subtracted and smoothed (Savitsky-Golay, width 0.15 m/z, cycles 4); monoisotopic peak detection used a SNAP averagine algorithm (C 4.9384, N 1.3577, O 1.4773, S 0.0417, H 7.7583) with a minimum S/N of 6.  Bruker flexAnalysis software (version 3.3) was used to perform spectral processing and peak list generation.

**Shotgun LC-MS:MS.** Peptides were extracted by washing three times with 50% (v/v) aqueous acetonitrile containing 0.1% trifluoroacetic acid (v/v), before being dried down in a vacuum concentrator and reconstituting in aqueous 0.1% trifluoroacetic acid (v/v). Samples were loaded onto a nanoAcquity UPLC system (Waters) equipped with a nanoAcquity Symmetry C18, 5 µm trap (180 µm x 20 mm Waters) and a nanoAcquity HSS T3 1.8 µm C18 capillary column (75 □m x 250 mm, Waters). The trap wash solvent was 0.1% (v/v) aqueous formic acid and the trapping flow rate was 10 µl/min. The trap was washed for 5 min before switching flow to the capillary column. Separation used a gradient elution of two solvents (solvent A: aqueous 0.1% (v/v) formic acid; solvent B: acetonitrile containing 0.1% (v/v) formic acid). The capillary column flow rate was 350 nl/min and the column temperature was 60°C. The gradient profile was linear 2-35% B over 20 mins. All runs then proceeded to wash with 95% solvent B for 2.5 min. The column was returned to initial conditions and re-equilibrated for 25 min before subsequent injections.

The nanoLC system was interfaced with a maXis HD LC-MS/MS system (Bruker Daltonics) with CaptiveSpray ionisation source (Bruker Daltonics). Positive ESI-MS and MS/MS spectra were acquired using AutoMSMS mode. Instrument control, data acquisition and processing were performed using Compass 1.7 software (microTOF control, Hystar and DataAnalysis, Bruker Daltonics). Instrument settings were: ion spray voltage: 1,450 V, dry gas: 3 l/min, dry gas temperature 150°C, ion acquisition range: m/z 150-2,000, MS spectra rate: 2 Hz, MS/MS spectra rate: 1 Hz at 2,500 cts to 10 Hz at 250,000 cts, cycle time: 3 s, quadrupole low mass: 300 m/z, collision RF: 1,400 Vpp, transfer time 120 ms. The collision energy and isolation width settings were automatically calculated using the AutoMSMS fragmentation table, absolute threshold 200 counts, preferred charge states: 2 – 4, singly charged ions excluded. A single MS/MS spectrum was acquired for each precursor and former target ions were excluded for 0.8 min unless the precursor intensity increased fourfold.

**Bioinformatics.** Tandem mass spectral data were submitted to database searching against the unrestricted NCBInr database (version 20190131, 187087713 sequences; 68237485887 residues) using a locally-running copy of the Mascot program (Matrix Science Ltd., version 2.5.1), through the Bruker ProteinScape interface (version 2.1). Search criteria specified: Enzyme, Trypsin; Fixed modifications, Carbamidomethyl (C); Variable modifications, Oxidation (M); Peptide tolerance, 150 ppm; MS/MS tolerance, 0.75 Da; Instrument, MALDI-TOF-TOF. Results were filtered to accept only peptides with an expect score of 0.05 or lower.

Helix turn helix predictions were carried out using NPS@ (37, 38) and Gym2.0 (39) using the default settings. Protein secondary structure and coiled coil predictions were made using JPRED4 (34) and COILS (35), respectively. Protein 3D structures were predicted using RaptorX (36). Molecular graphics/analyses were performed with the UCSF Chimera package v1.13 (58). Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311). COBALT was used for protein sequence alignments and PROMALS3D for alignment of predicted protein structure (59, 60). Jalview was used to visualize alignments (61). FIJI software was used image analysis (62). Figure graphics were produced using CorelDraw 2018. Statistical analysis was carried out using Sigmaplot software version 13 (Systat Software Inc.) and, for each use, the test parameters are indicated in the text and/or figure legends.

**Acknowledgements:** This work was wholly supported by a Wellcome Trust/Royal Society Sir Henry Dale Fellowship Grant (109363/Z/15/Z). We would like to acknowledge the contributions of the University of York Technology Facility to this work. Electron microscopy images were produced by the Imaging and Cytometry lab and proteomics were carried out by the York Centre of Excellence in Mass Spectrometry, which was created thanks to a major capital investment through Science City York, supported by Yorkshire Forward with funds from the Northern Way Initiative, and subsequent support from EPSRC (EP/K039660/1; EP/M028127/1). We would also like to thank Prof. Fred Antson for critical reading of the manuscript.

**Author Contributions:** Conceptualization, P.C.M.F; Methodology, D.S., J.X.L. and P.C.M.F.; Investigation, D.S., J.X.L. and P.C.M.F.; Writing, P.C.M.F.; Visualization, P.C.M.F., Funding Acquisition, P.C.M.F.; Resources, P.C.M.F.; Supervision, D.S. and P.C.M.F.

**Declaration of Interests:** The authors declare no competing interests.

## References

1. Suttle CA. 2007. Marine viruses--major players in the global ecosystem. Nat Rev Microbiol **5**:801–812.

2. Shkoporov AN, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP, Hill C. 2018. ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. Nat Commun **9**:4781.

3. Breitbart M, Bonnain C, Malki K, Sawaya NA. 2018. Phage puppet masters of the marine microbial realm. Nature Microbiology **3**:754–766.

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

5. Sato K, Campbell A. 1970. Specialized transduction of galactose by lambda phage from a deletion lysogen. Virology **41**:474–487.

6. Chen J, Quiles-Puchalt N, Chiang YN, Bacigalupe R, Fillol-Salom A, Chee MSJ, Fitzgerald JR, Penadés JR. 2018. Genome hypermobility by lateral transduction. Science **362**:207–212.

7. Lang AS, Zhaxybayeva O, Beatty JT. 2012. Gene transfer agents: phage-like elements of genetic exchange. Nat Rev Microbiol **10**:472–482.

8. Hynes AP, Mercer RG, Watton DE, Buckley CB, Lang AS. 2012. DNA packaging bias and differential expression of gene transfer agent genes within a population during production and release of the *Rhodobacter capsulatus* gene transfer agent, RcGTA. Mol Microbiol **85**:314–325.

9. Tomasch J, Wang H, Hall ATK, Patzelt D, Preusse M, Petersen J, Brinkmann H, Bunk B, Bhuju S, Jarek M, Geffers R, Lang AS, Wagner-Döbler I. 2018. Packaging of *Dinoroseobacter shibae* DNA into Gene Transfer Agent particles is not random. Genome Biol Evol **10**:359–369.

10. Berglund EC, Frank AC, Calteau A, Vinnere Pettersson O, Granberg F, Eriksson A-S, Näslund K, Holmberg M, Lindroos H, Andersson SGE. 2009. Run-off replication of host-adaptability genes is

associated with gene transfer agents in the genome of mouse-infecting *Bartonella grahamii*. PLoS Genet **5**:e1000546.

11. McDaniel LD, Young E, Delaney J, Ruhnau F, Ritchie KB, Paul JH. 2010. High frequency of horizontal gene transfer in the oceans. Science **330**:50.

12. Solioz M, Marrs B. 1977. The gene transfer agent of *Rhodopseudomonas capsulata*. Purification and characterization of its nucleic acid. Arch Biochem Biophys **181**:300–307.

13. Marrs B. 1974. Genetic recombination in *Rhodopseudomonas capsulata*. Proc Natl Acad Sci U S A **71**:971–973.

14. Lang AS, Beatty JT. 2000. Genetic analysis of a bacterial genetic exchange element: the gene transfer agent of *Rhodobacter capsulatus*. Proc Natl Acad Sci U S A **97**:859–864.

15. Westbye AB, Kuchinski K, Yip CK, Beatty JT. 2016. The Gene Transfer Agent RcGTA contains head spikes needed for binding to the *Rhodobacter capsulatus* polysaccharide cell capsule. J Mol Biol **428**:477–491.

16. Hynes AP, Shakya M, Mercer RG, Grüll MP, Bown L, Davidson F, Steffen E, Matchem H, Peach ME, Berger T, Grebe K, Zhaxybayeva O, Lang AS. 2016. Functional and evolutionary characterization of a gene transfer agent's multilocus "genome". Mol Biol Evol **33**:2530–2543.

17. Shakya M, Soucy SM, Zhaxybayeva O. 2017. Insights into origin and evolution of α-proteobacterial gene transfer agents. Virus evolution **3**:vex036.

18. Lang AS, Beatty JT. 2006. Importance of wide spread gene transfer agent genes in alpha-proteobacteria. Trends in microbiology **15**:54–62.

19. Motro Y, La T, Bellgard MI, Dunn DS, Phillips ND, Hampson DJ. 2009. Identification of genes associated with prophage-like gene transfer agents in the pathogenic intestinal spirochaetes *Brachyspira hyodysenteriae, Brachyspira pilosicoli* and *Brachyspira intermedia*. Vet Microbiol **134**:340–345.

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

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

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

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

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

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

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

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

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

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

30. Ponchon L, Boulanger P, Labesse G, Letellier L. 2006. The endonuclease domain of bacteriophage terminases belongs to the resolvase/integrase/ribonuclease H superfamily: a bioinformatics analysis validated by a functional study on bacteriophage T5. J Biol Chem **281**:5829–5836.

31. Kanamaru S, Kondabagil K, Rossmann MG, Rao VB. 2004. The functional domains of bacteriophage t4 terminase. J Biol Chem **279**:40795–40801.

32. Sun S, Gao S, Kondabagil K, Xiang Y, Rossmann MG, Rao VB. 2012. Structure and function of the small terminase component of the DNA packaging machine in T4-like bacteriophages. Proc Natl Acad Sci U S A **109**:817–822.

33. Leung MM-Y. 2010. PhD Thesis. University of British Columbia. CtrA and GtaR : two systems that regulate the Gene Transfer Agent in *Rhodobacter capsulatus*.

34. Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015. JPred4: a protein secondary structure prediction server. Nucleic Acids Res **43**:W389–94.

35. Lupas A, Van Dyke M, Stock J. 1991. Predicting coiled coils from protein sequences. Science **252**:1162–1164.

36. Källberg M, Margaryan G, Wang S, Ma J, Xu J. 2014. RaptorX server: a resource for template-based protein structure modeling. Methods Mol Biol **1137**:17–27.

37. Combet C, Blanchet C, Geourjon C, Deléage G. 2000. NPS@: network protein sequence analysis. Trends Biochem Sci **25**:147–150.

38. Dodd IB, Egan JB. 1990. Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res **18**:5019–5026.

39. Narasimhan G, Bu C, Gao Y, Wang X, Xu N, Mathee K. 2002. Mining protein sequences for motifs. J Comput Biol **9**:707–720.

40. Fogg PCM. 2019. Identification and characterization of a direct activator of a gene transfer agent. Nat Commun **10**:595.

41. Karimova G, Pidoux J, Ullmann A, Ladant D. 1998. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proc Natl Acad Sci U S A **95**:5752–5756.

42. Gao S, Rao VB. 2011. Specificity of interactions among the DNA-packaging machine components of T4-related bacteriophages. J Biol Chem **286**:3944–3956.

43. Westbye AB, Leung MM, Florizone SM, Taylor TA, Johnson JA, Fogg PC, Beatty JT. 2013. Phosphate concentration and the putative sensor kinase protein CckA modulate cell lysis and release of the *Rhodobacter capsulatus* gene transfer agent. J Bacteriol **195**:5025–5040.

44. Fogg PCM, Westbye AB, Beatty JT. 2012. One for all or all for one: heterogeneous expression and host cell lysis are key to gene transfer agent activity in *Rhodobacter capsulatus*. PLoS ONE **7**:e43772.

45. Chen F, Spano A, Goodman BE, Blasier KR, Sabat A, Jeffery E, Norris A, Shabanowitz J, Hunt DF, Lebedev N. 2009. Proteomic analysis and identification of the structural and regulatory proteins of the *Rhodobacter capsulatus* gene transfer agent. J Proteome Res **8**:967–973.

46. Kondabagil KR, Zhang Z, Rao VB. 2006. The DNA translocating ATPase of bacteriophage T4 packaging motor. J Mol Biol **363**:786–799.

47. Zhang Z, Kottadiel VI, Vafabakhsh R, Dai L, Chemla YR, Ha T, Rao VB. 2011. A promiscuous DNA packaging machine from bacteriophage T4. PLoS Biol **9**:e1000592.

48. Gao S, Zhang L, Rao VB. 2016. Exclusion of small terminase mediated DNA threading models for genome packaging in bacteriophage T4. Nucleic Acids Res **44**:4425–4439.

49. Büttner CR, Chechik M, Ortiz-Lombardía M, Smits C, Ebong I-O, Chechik V, Jeschke G, Dykeman E, Benini S, Robinson CV, Alonso JC, Antson AA. 2012. Structural basis for DNA recognition and loading into a viral packaging motor. Proc Natl Acad Sci U S A **109**:811–816.

50. Nemecek D, Lander GC, Johnson JE, Casjens SR, Thomas GJ. 2008. Assembly architecture and DNA binding of the bacteriophage P22 terminase small subunit. J Mol Biol **383**:494–501.

51. Wall JD, Weaver PF, Gest H. 1975. Gene transfer agents, bacteriophages, and bacteriocins of *Rhodopseudomonas capsulata*. Arch Microbiol **105**:217–224.

52. Ding H, Moksa MM, Hirst M, Beatty JT. 2014. Draft genome sequences of six *Rhodobacter capsulatus Strains*, YW1, YW2, B6, Y262, R121, and DE442. Genome Announc **2**.

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

54. Leung M, Beatty J. 2013. *Rhodobacter capsulatus* Gene Transfer Agent transduction assay. Bio-protocol **3(4)**: e334. DOI: 10.21769/BioProtoc.334.

55. Booth DS, Avila-Sakar A, Cheng Y. 2011. Visualizing proteins and macromolecular complexes by negative stain EM: from grid preparation to image acquisition. J Vis Exp. **22**(58): 3227.

56. Fogg PCM, Younger E, Fernando BD, Khaleel T, Stark WM, Smith MCM. 2018. Recombination directionality factor gp3 binds ϕC31 integrase via the zinc domain, potentially affecting the trajectory of the coiled-coil motif. Nucleic Acids Res **46**:1308–1320.

57. Wiethaus J, Schubert B, Pfänder Y, Narberhaus F, Masepohl B. 2008. The GntR-like regulator TauR activates expression of taurine utilization genes in *Rhodobacter capsulatus*. J Bacteriol **190**:487–493.

58. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem **25**:1605–1612.

59. Papadopoulos JS, Agarwala R. 2007. COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics **23**:1073–1079.

60. Pei J, Kim B-H, Grishin NV. 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res **36**:2295–2300.

61. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics **25**:1189–1191.

62. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: an open-source platform for biological-image analysis. Nat Methods **9**:676–682.

63. Liu H, Naismith JH. 2009. A simple and efficient expression and purification system using two newly constructed vectors. Protein Expr Purif **63**:102–111.

64. Fogg MJ, Wilkinson AJ. 2008. Higher-throughput approaches to crystallization and crystal 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 [**Δ*g1***], *g1* deletion complemented with full length *g1* [**Δ*g1*(*g1*)**], *g1* deletion complemented with *g1* lacking helix 1 [**Δ*g1*(Δh1)**] and *g1* deletion complemented with *g1* lacking helix 3 [**Δ*g1*(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 [**Δ*g1***] and a complemented *g1* deletion

[**Δ*g1*(*g1*)**]. Bioline Hyperladder 1 kb DNA ladder [**M**] and purified RcGTA DNA [**GTA**] is shown for size comparison. The location of genomic DNA, RcGTA DNA and the 4 kb DNA ladder band are annotated. **C.** Agarose gel of DNA isolated from purified RcGTA particles released by DE442 wild-type and *g1* deletion strains. **D.** Interaction between RcGTA gp1 and the large terminase (TerL). 10 µl spots of individual bacterial-2-hybrid assay transformations. Blue/green indicates a positive interaction and white indicates no interaction. Reactions shown are as follows: "gp1" = gp1 vs TerL, "-ve" = no insert control, "gp1Δh1" = gp1 helix 1 deletion vs TerL, "gp1Δh3" = gp1 helix 3 deletion vs TerL, "Nuc" = gp1 vs TerL nuclease domain, "ATP" = gp1 vs TerL ATPase domain. **E.** Histogram showing quantification of the interactions shown in panel A by β-galactosidase assay. Statistical significance is shown above the chart (ANOVA, n=3, ns=not significant, *** p<0.01).

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

SDS PAGE gel of affinity purified RcGTAs produced by DE442 wild-type [**WT**] and *g1* deletion strains [**Δg1**]. Expedeon Tri-Color marker is included for size comparison [**M**], with approximate molecular weights annotated to the left of the gel. Bands predicted to contain the RcGTA portal and capsid proteins are annotated, as well the tail fibre (**GTA TF**) that was confirmed by MALDI mass spectrometry.

**Figure 5. Gp1 is a prerequisite for RcGTA assembly. A & B.** Structural proteome of RcGTA particles. LC-MS:MS analysis of affinity purified RcGTA particles produced by DE442 wild-type [**WT**] and *g1* deletion [**Δ*g1***] strains. RcGTA head proteins and tail proteins are shown separately in panels Band A and B, respectively. **D-H.** Transmission electron micrographs of RcGTA particles. Images in panels D-G were taken at 68,000x magnification and the scale bar in panel C represents 50 nm. The panel H image was taken at 49,000x magnification and the scale bar represents 100 nm. Black arrow heads indicate head spikes and white arrows indicate portal apertures.

**Figure 6. DNA content of affinity purified RcGTA particles.** Agarose gels of DNA extracted directly from chimeric his6-tagged RcGTAs are shown. His-tags were incorporated into nascent RcGTA particles by ectopic expression of his6-tagged gp5 (**A**) or gp1 (**B**) proteins. Tagged RcGTAs were purified from *R.*

*capsulatus* DE442 culture supernatants using nickel agarose affinity chromatography. The genotype of the producer cells is indicated directly above each gel – wild-type (**WT**) or RcGTA *g1* gene knock-out (**Δg1**). DNA marker hyperladder 1 kb is included for reference (**M**); the locations of the 4 kb reference band and GTA DNA are annotated.

**Figure 7. RcGTA gp1 *in vitro* DNA binding. A.** Representative agarose gel (0.8% w/v) showing the stated concentrations of gp1 protein binding to DNA in an electrophoretic mobility shift assay (EMSA). The locations of unbound and shifted DNA are annotated. Substrate DNA in the assay shown is a 1.4 kbp PCR amplification of an arbitrarily chosen region flanking the *rcc01398* gene from *R. capsulatus* (amplified using *rcc01398* F & R primers, Table 3). Bioline hyperladder 1 kb DNA marker is shown for size comparison [**M**]. **B.** Quantification of EMSAs by band intensity analysis. Data is show is the average of two EMSAs carried out independently in time and with different DNA substrates (*rcc01397* and *rcc01398* genes). Individual data points are plotted as well as the mean line.

**Figure 8. GTA TerS protein alignment.** COBALT multiple sequence alignment of putative GTA small terminases from Rhodobacterales species *Dinoroseobacter shibae* (DsGTA), *Oceanicola granulosus* (OgGTA), *Rhodobacter capsulatus* (RcGTA) and *Ruegeria pomeroyi* (RpGTA). Intensity of colour for each amino acid is based on percentage identity. Predicted secondary structure is indicated below the alignment with “h” indicating helical.

**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*<br>(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*<br>(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 |

Grey rows are well characterized phage small terminases included for comparison

* Where no gene/protein has previously been annotated the accession number for the bacterial genome is provided with the nucleotide position of the new ORF indicated in brackets.

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

| Name | Description | Reference |
|---|---|---|
| **pCM66T** | pCM66T was a gift from Mary Lidstrom Broad host range vector; ColE1, OriV, IncP/traJ, Kanamycin$^R$ | Addgene plasmid # 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 |

**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** | CGACTCTAGAGGATCTGGGGGGGCTTGGGAACAAT |
| **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 |

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

**Figure 1**

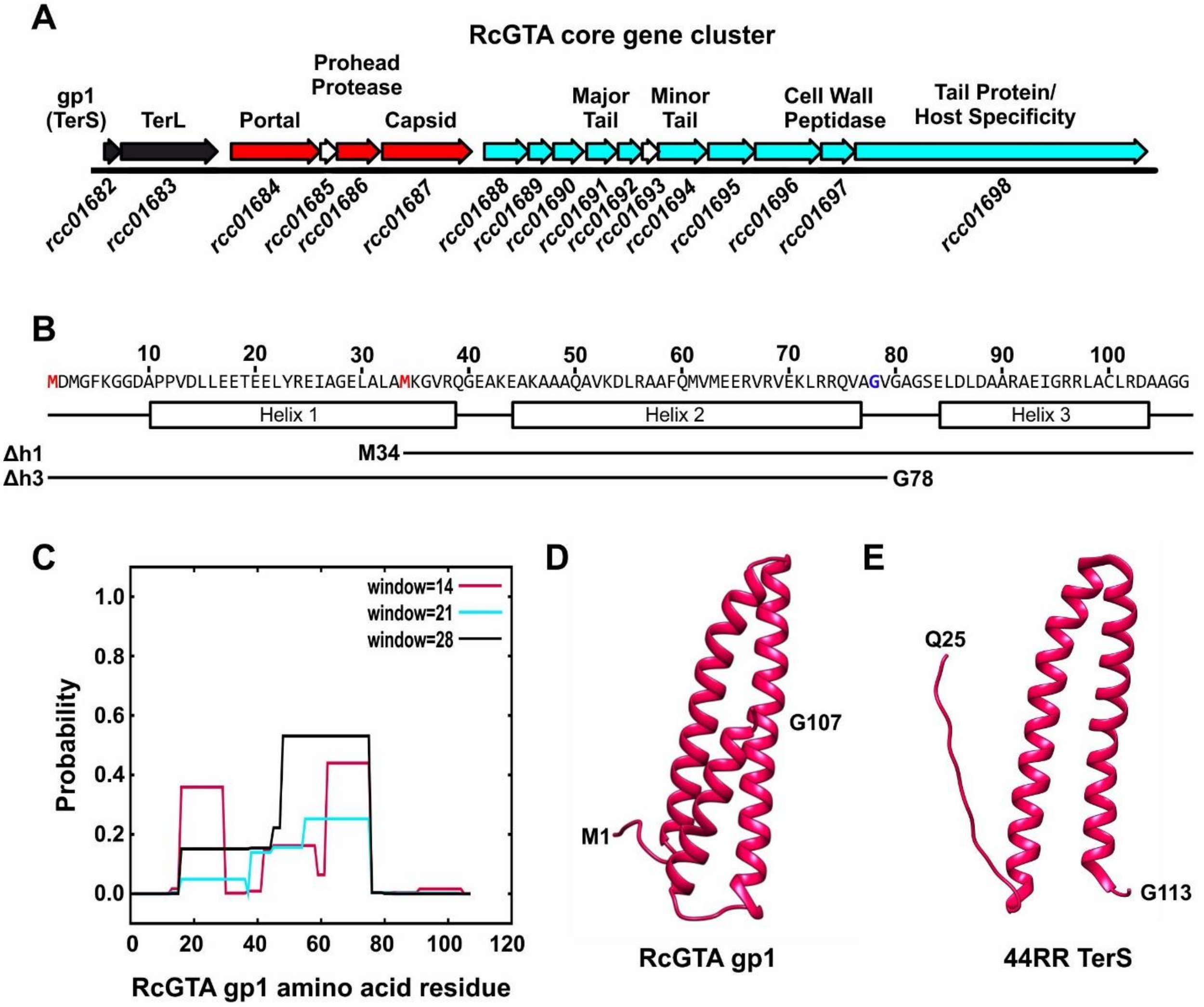

A
RcGTA core gene cluster
gp1
(TerS)
TerL
Portal
Prohead
Protease
Capsid
Major
Tail
Minor
Tail
Cell Wall
Peptidase
Tail Protein/
Host Specificity
rcc01682
rcc01683
rcc01684
rcc01685
rcc01686
rcc01687
rcc01688
rcc01689
rcc01690
rcc01691
rcc01692
rcc01693
rcc01694
rcc01695
rcc01696
rcc01697
rcc01698
B
10
20
30
40
50
60
70
80
90
100
MDMGFKGGDAPPVDLLEETEELYREIAGELALAMKGVRQGEAKEAKAAAQAVKDLRAAFQMVMEERVRVEKLRRQVAGVGAGSELDLDAARAEIGRRLACLRDAAGG
Helix 1
Helix 2
Helix 3
Δh1
M34
Δh3
G78
C
window=14
window=21
window=28
Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
120
RcGTA gp1 amino acid residue
D
M1
G107
RcGTA gp1
E
Q25
G113
44RR TerS

**Figure 2**

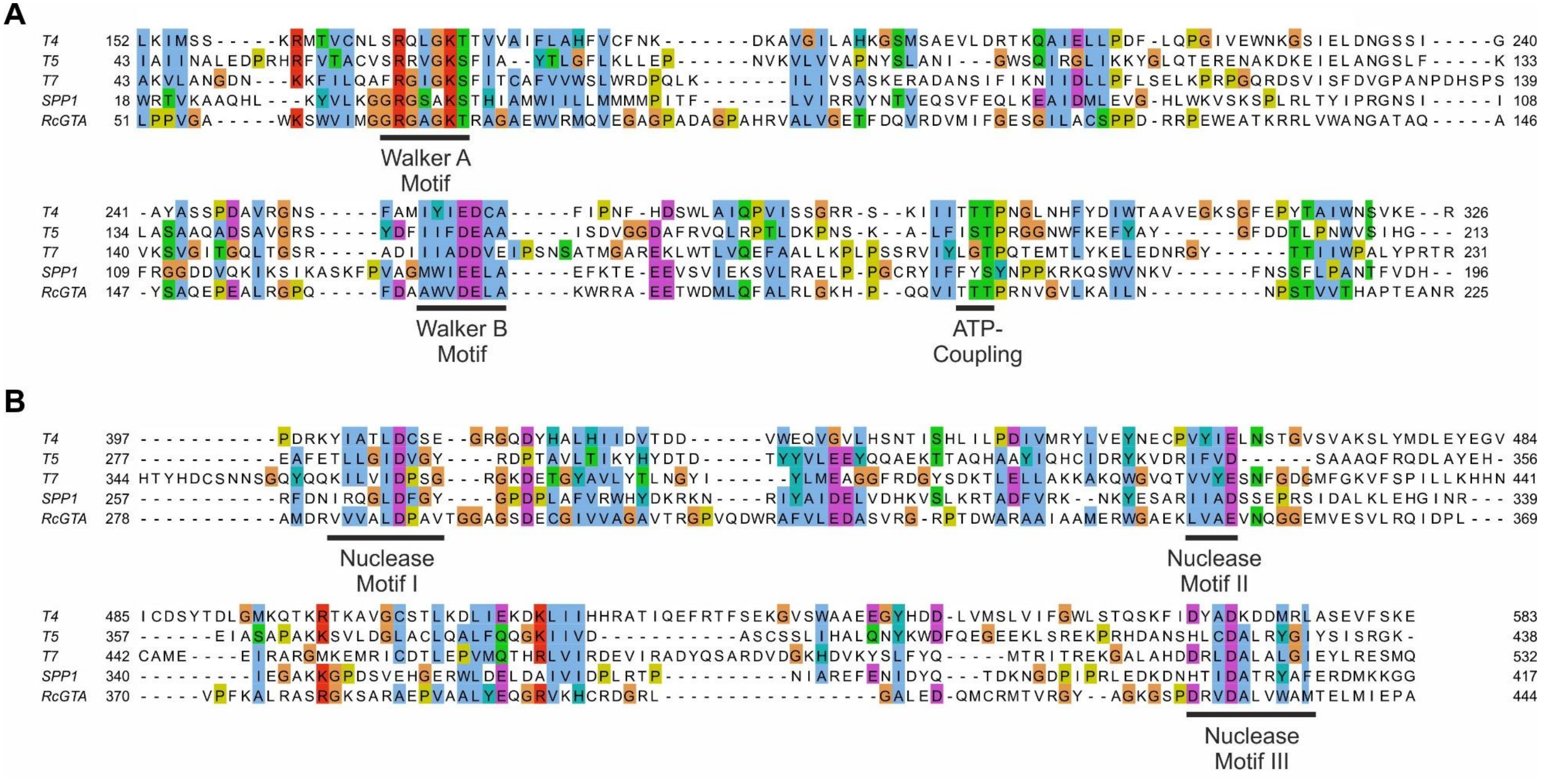

A
T4 152 LKIMSS-----KRMTVCNLSRQLGKTTVVAIFLAHFVCFNK-------DKAVGILAHKGSMSAEVLDRTKQAIELLPDF-LQPGIVEWNKGSIELDNGSSI-----G 240
T5 43 IAIINALEDPRHRFVTACVSRRVGKSFIA--YTLGFLKLLEP------NVKVLVVAPNYSLANI---GWSQIRGLIKKYGLQTERENAKDKEIELANGSLF-----K 133
T7 43 AKVLANGDN---KKFILQAFRGIGKSFITCAFVVWSLWRDPQLK-------ILIVSASKERADANSIFIKNIIDLLPFLSELKPRPGQRDSVISFDVGPANPDHSPS 139
SPP1 18 WRTVKAAQHL---KYVLKGGRGSAKSTHIAMWIILLMMMMPITF-------LVIRRVYNTVEQSVFEQLKEAIDMLEVG-HLWKVSKSPLRLTYIPRGNSI-----I 108
RcGTA 51 LPPVGA-----WKSWVIMGGRGAGKTRAGAEWVRMQVEGAGPADAGPAHRVALVGETFDQVRDVMIFGESGILACSPPD-RRPEWEATKRRLVWANGATAQ-----A 146
Walker A Motif
T4 241 -AYASSPDAVRGNS-----FAMIYIEDCA-----FIPNF-HDSWLAIQPVISSGRR-S--KIIITTTPNGLNHFYDIWTAAVEGKSGFEPYTAIWNSVKE--R 326
T5 134 LASAAQADSAVGRS-----YDFIIFDEAA-----ISDVGGDAFRVQLRPTLDKPNS-K--A-LFISTPRGGNWFKEFYAY------GFDDTLPNWVSIHG--- 213
T7 140 VKSVGITGQLTGSR-----ADIIIADDVEIPSNSATMGAREKLWTLVQEFAALLKPLPSSRVIYLGTPQTEMTLYKELEDNRGY------TTIIWPALYPRTR 231
SPP1 109 FRGGDDVQKIKSIKASKFPVAGMWIEELA-----EFKTE-EEVSVIEKSVLRAELP-PGCRYIFFYSYNPPKRKQSWVNKV------FNSSFLPANTFVDH-- 196
RcGTA 147 -YSAQEPEALRGPQ-----FDAAWVDELA-----KWRRA-EETWDMLQFALRLGKH-P--QQVITTTPRNVGVLKAILN---------NPSTVVTHAPTEANR 225
Walker B Motif
ATP-Coupling
B
T4 397 -----------PDRKYIATLDCSE--GRGQDYHALHIIDVTDD------VWEQVGVLHSNTISHLILPDIVMRYLVEYNECPVYIELNSTGVSVAKSLYMDLEYEGV 484
T5 277 -----------EAFETLLGIDVGY----RDPTAVLTIKYHYDTD-----TYYVLEEYQQAEKTTAQHAAYIQHCIDRYKVDRIFVD------SAAAQFRQDLAYEH- 356
T7 344 HTYHDCSNNSGQYQQKILVIDPSG---RGKDETGYAVLYTLNGYI------YLMEAGGFRDGYSDKTLELLAKKAKQWGVQTVVYESNFGDGMFGKVFSPILLKHHN 441
SPP1 257 -----------RFDNIRQGLDFGY----GPDPLAFVRWHYDKRKN----RIYAIDELVDHKVSLKRTADFVRK--NKYESARIIADSSEPRSIDALKLEHGINR--- 339
RcGTA 278 -----------AMDRVVVALDPAVTGGAGSDECGIVVAGAVTRGPVQDWRAFVLEDASVRG-RPTDWARAAIAAMERWGAEKLVAEVNQGGEMVESVLRQIDPL--- 369
Nuclease Motif I
Nuclease Motif II
T4 485 ICDSYTDLGMKQTKRTKAVGCSTLKDLIEKDKLIIHHRATIQEFRTFSEKGVSWAAEEGYHDD-LVMSLVIFGWLSTQSKFIDYADKDDMRLASEVFSKE 583
T5 357 ------EIASAPAKKSVLDGLACLQALFQQGKIIVD-----------ASCSSLIHALQNYKWDFQEGEEKLSREKPRHDANSHLCDALRYGIYSISRGK- 438
T7 442 CAME----EIRARGMKEMRICDTLEPVMQTHRLVIRDEVIRADYQSARDVDGKHDVKYSLFYQ-----MTRITREKGALAHDDRLDALALGIEYLRESMQ 532
SPP1 340 ---------IEGAKKGPDSVEHGERWLDELDAIVIDPLRTP----------NIAREFENIDYQ---TDKNGDPIPRLEDKDNHTIDATRYAFERDMKKGG 417
RcGTA 370 -----VPFKALRASRGKSARAEPVAALYEQGRVKHCRDGRL-----------------GALED-QMCRMTVRGY--AGKGSPDRVDALVWAMTELMIEPA 444
Nuclease Motif III

**Figure 3**

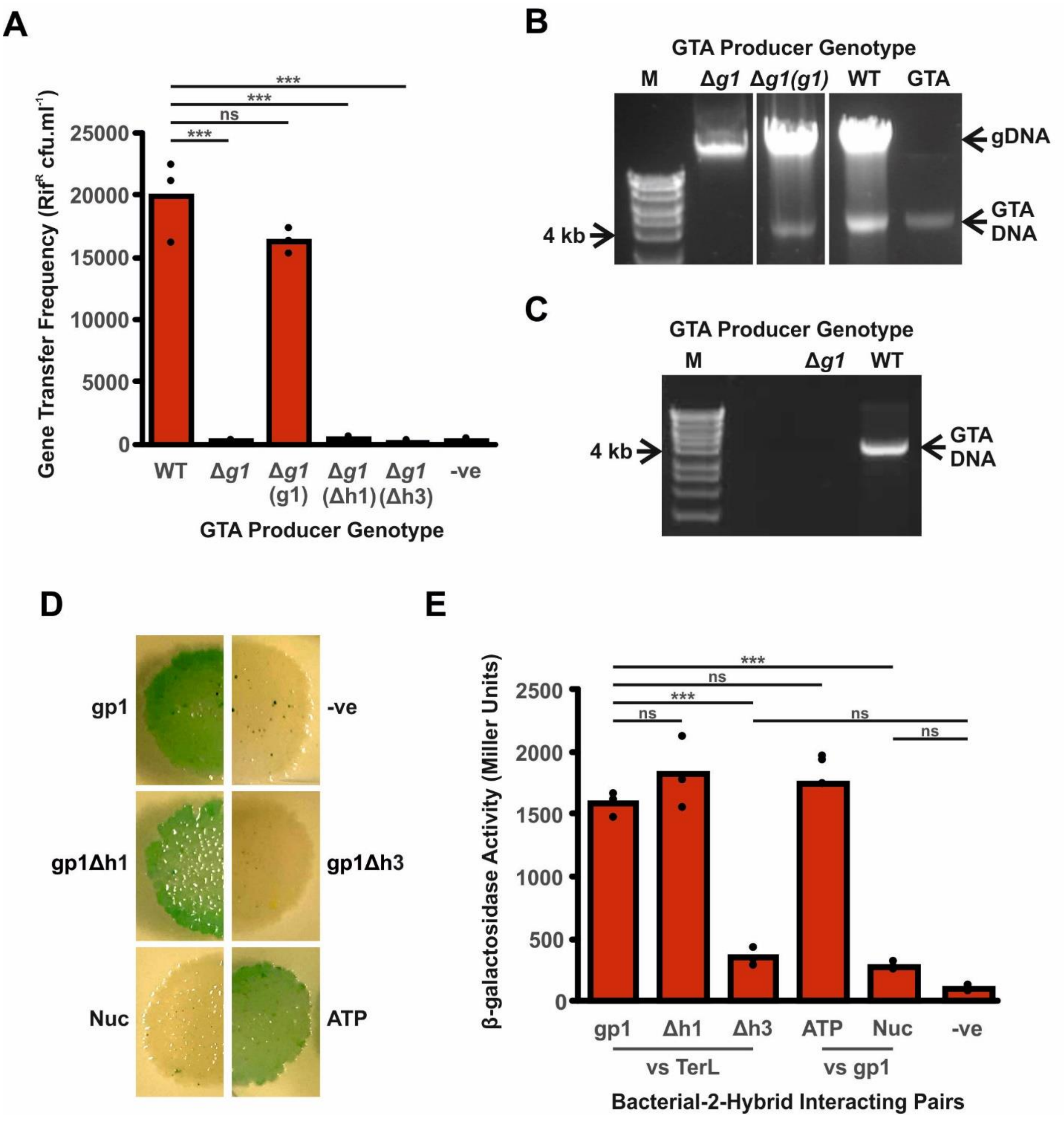

A
Gene Transfer Frequency (Rif$^R$ cfu.ml$^{-1}$)
25000
20000
15000
10000
5000
0
***
***
ns
***
WT
Δg1
Δg1 (g1)
Δg1 (Δh1)
Δg1 (Δh3)
-ve
GTA Producer Genotype
B
GTA Producer Genotype
M
Δg1
Δg1(g1)
WT
GTA
gDNA
GTA DNA
4 kb
C
GTA Producer Genotype
M
Δg1
WT
4 kb
GTA DNA
D
gp1
-ve
gp1Δh1
gp1Δh3
Nuc
ATP
E
β-galactosidase Activity (Miller Units)
2500
2000
1500
1000
500
0
***
ns
***
ns
ns
ns
gp1
Δh1
Δh3
ATP
Nuc
-ve
vs TerL
vs gp1
Bacterial-2-Hybrid Interacting Pairs

**Figure 4**

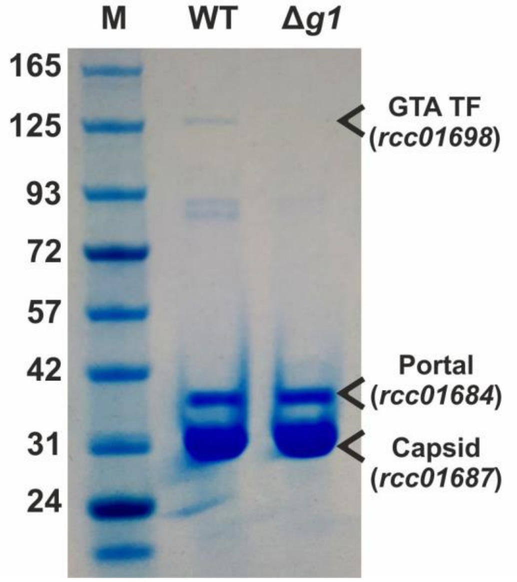

M
WT
Δ*g1*
165
125
93
72
57
42
31
24
GTA TF
(*rcc01698*)
Portal
(*rcc01684*)
Capsid
(*rcc01687*)

**Figure 5**

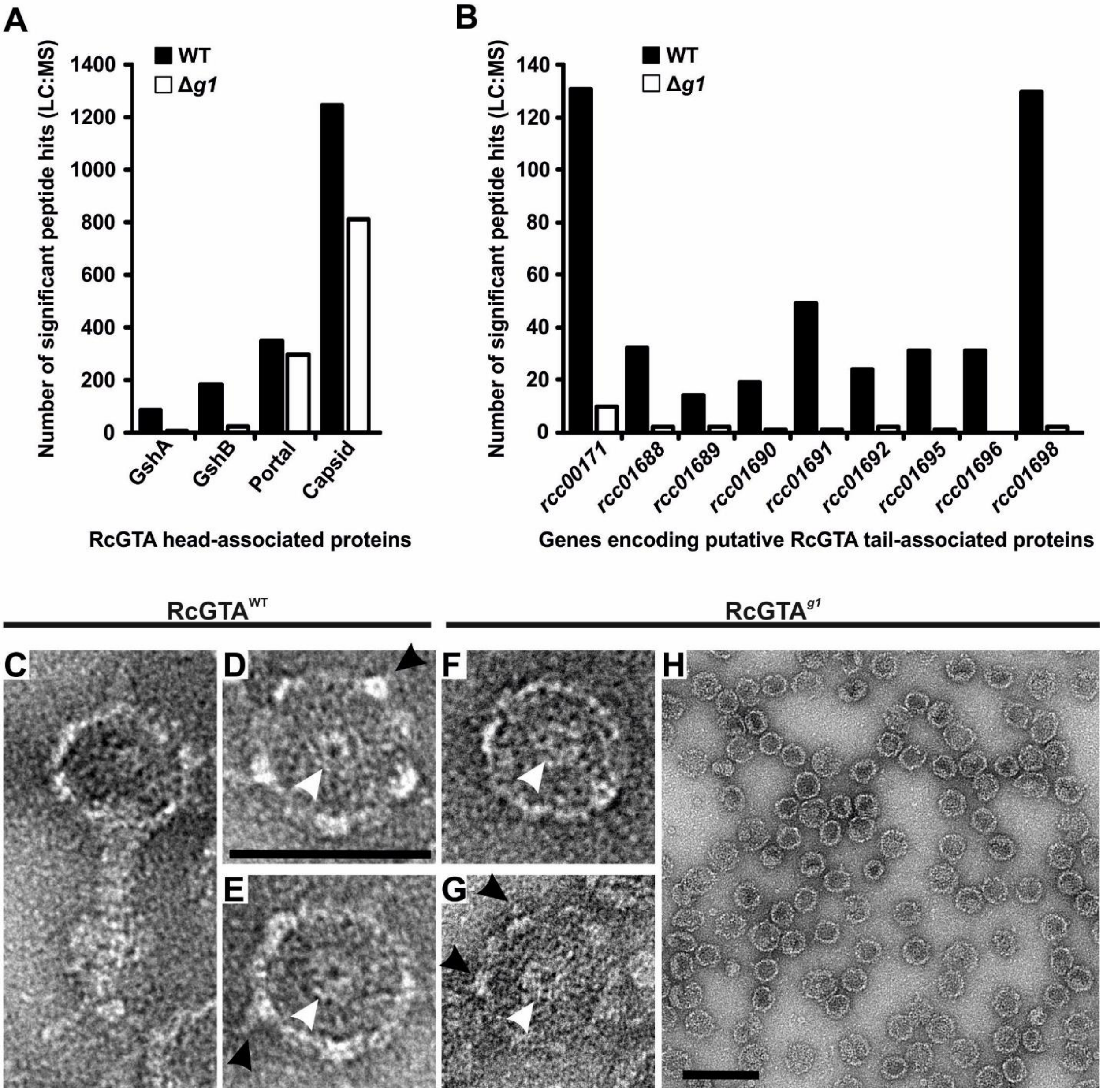

A
WT
Δg1
Number of significant peptide hits (LC:MS)
1400
1200
1000
800
600
400
200
0
GshA
GshB
Portal
Capsid
RcGTA head-associated proteins
B
WT
Δg1
Number of significant peptide hits (LC:MS)
140
120
100
80
60
40
20
0
rcc00171
rcc01688
rcc01689
rcc01690
rcc01691
rcc01692
rcc01695
rcc01696
rcc01698
Genes encoding putative RcGTA tail-associated proteins
RcGTA^WT
RcGTA^g1
C
D
E
F
G
H

**Figure 6**

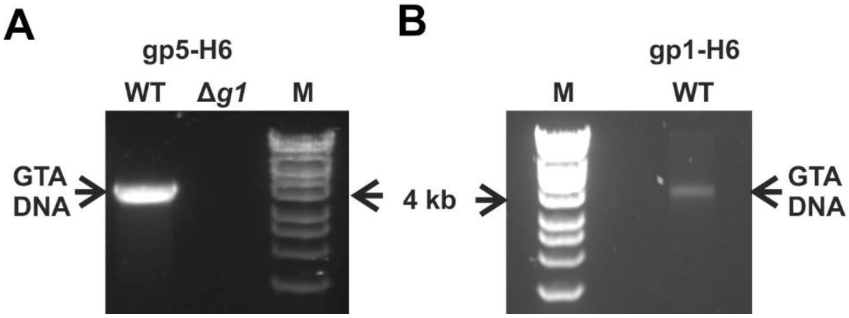


**Figure 7**

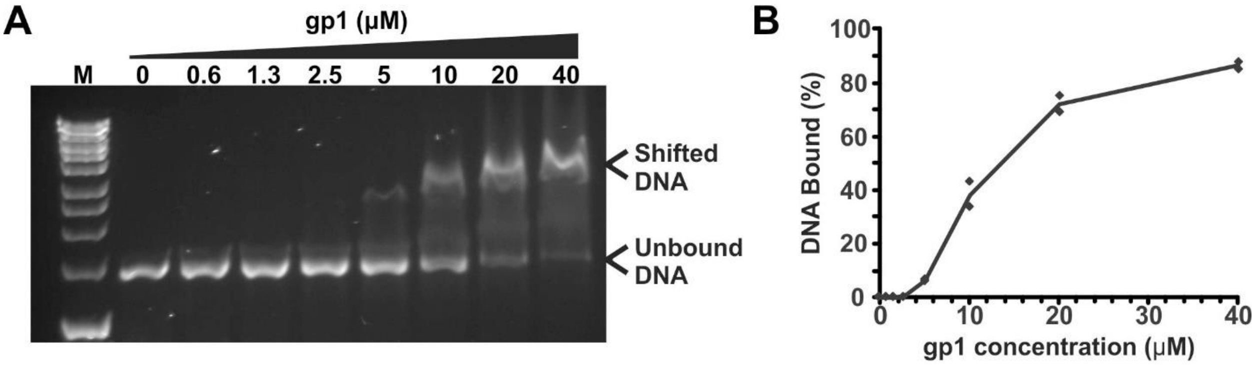
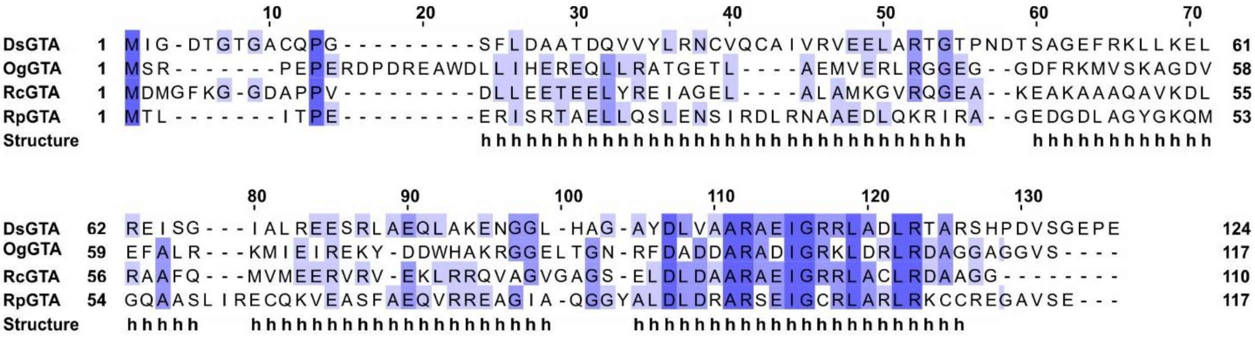

A
gp1 (μM)
M
0
0.6
1.3
2.5
5
10
20
40
Shifted DNA
Unbound DNA
B
100
80
60
40
20
0
DNA Bound (%)
0
10
20
30
40
gp1 concentration (μM)

Figure 8



```
                    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                              hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh    hhhhhhhhhhhh

                   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
Structure   hhhhh   hhhhhhhhhhhhhhhhhhhh     hhhhhhhhhhhhhhhhhhhhhh
```