Regulation of gene transfer and motility in the bacterium *Rhodobacter capsulatus* by the dinucleotide cyclic dimeric guanosine monophosphate (c-di-GMP)

by

© Purvikalyan Pallegar, M.Sc.,

A Thesis submitted to the

School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biology

Faculty of Science

Memorial University of Newfoundland

St. John's, Newfoundland and Labrador, Canada

September 2020

Abstract

Gene transfer agents (GTAs) are bacteriophage-like particles produced by several bacterial and archaeal lineages that contain small pieces of the producing cells' genomes that can be transferred to other cells in a process similar to transduction. One well-studied GTA is RcGTA, produced by the α-proteobacterium *Rhodobacter capsulatus*. RcGTA gene expression is regulated by several cellular regulatory systems, including the CckA-ChpT-CtrA phosphorelay. A previous study on CtrA, a DNA-binding response regulator protein, showed that transcription of multiple other regulator-encoding genes was affected by its loss. These included genes encoding proteins predicted to be involved in the synthesis and hydrolysis of the second messenger bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP). In this thesis, I investigated these genes and found that disruption of four of them (*rcc00620*, *rcc00645*, *rcc02629*, *rcc02857*) affected RcGTA production and flagellar motility. I performed site-directed mutagenesis of key catalytic residues in the functional domains responsible for diguanylate cyclase (DGC) and phosphodiesterase (PDE) activities involved in synthesizing and hydrolysing cyclic-di-GMP, respectively, and analysed the activities of the wild type and mutant proteins. Through this work, I demonstrated that c-di-GMP acts to inhibit RcGTA production and motility. I subsequently demonstrated that the enzymatic activity of one of the proteins, Rcc00620, is regulated through the phosphorylation status of its REC domain, which is controlled by a cognate histidine kinase protein, Rcc00621. In this system, the phosphorylated form of Rcc00620 is active as a PDE enzyme and stimulates gene transfer and motility. Interestingly, the *rcc00620*/*rcc00621* genes were acquired via horizontal gene transfer from a distantly related α-proteobacterium. I also explored a potential

mechanistic link between c-di-GMP and RcGTA regulation via the CckA-ChpT-CtrA phosphorelay and confirmed that c-di-GMP binds to CckA and this shifts its activity from kinase to phosphatase and thereby affects CtrA phosphorylation. All these results add c-di-GMP signaling to the collection of cellular regulatory systems controlling gene transfer in this bacterium and add gene transfer to the diverse list of activities affected by c-di-GMP in bacteria.

Acknowledgements

I take this privilege to acknowledge my indebtedness to my supervisor Dr. Andrew S. Lang for his active supervision, decisive comments, worthwhile suggestions, and encouragement all through this study. His immense support and motivation had made me to overcome difficult times and allowed me to make good progress during this research study. I am sincerely grateful to my co-supervisor Dr. Lourdes Peña-Castillo for her kind co-operation, encouragement, and continuous support during all the years of my research. I also express my deep gratitude to my graduate committee advisor Dr. Dawn Bignell for her guidance and support during the progress of the study.

I thank all the current and former lab members of Dr. Lang whose company and support helped to maintain a very friendly and healthy environment to work and my special thanks to Dr. Marta Canuti, post doctoral fellow in the lab for always being there to discuss and exchange of ideas. I thank and greatly appreciate all my co-authors whom I specially mentioned in co-authorship statement. I also express my special thanks to Dr. Sherri Christian for her invaluable suggestions on protein work and allowing me to use their lab.

My research for doctoral thesis has been supported by graduate student fellowships from the Memorial University School of Graduate Studies (SGS) and the Natural Science and Engineering Research Council (NSERC). The research in Dr. Langs's laboratory and Dr. Castillo's laboratory are supported by grants from NSERC. Finally, I would like to thank my family for always being there for me.

Table of contents

List of tables

List of figures

List of symbols, nomenclature or abbreviations

Abbreviation	Meaning
%	percent
µ	micro
α	alpha
Δ	delta
ANOVA	Analysis of variance
ADP	Adenosine di phosphate
ATP	Adenosine tri phosphate
BLAST	Basic Local Alignment Search Tool
bp	basepair
bps	basepairs
c-di-GMP	cyclic dimeric guanosine monophosphate
DNA	deoxyribonucleic acid
dsDNA	double stranded deoxyribonucleic acid
EDTA	ethylenediaminetetraceticacid
FDG	flourescein di-β-D-galactopyranoside
GTA	gene transfer agent
HGT	horizontal gene transfer
HSD	Honestly significant difference
IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilobases
kbp	kilobase pairs

LB	Luria-Bertani
Ni-NTA	Nickle nitrilotriacetic acid
nts	nucleotides
ORF	open reading frame
ORI	origin of replication
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
RBS	ribosome binding site
RCV	*Rhodobacter capsulatus* medium V
RcGTA	*Rhodobacter capsulatus* gene transfer agent
RNA	ribonucleic acid
rpm	revolutions per minute
s	seconds
SDS	Sodium dodecyl lsulphate
SDM	site directed mutagenesis
TM	transmembrane
YPS	yeast extract/peptone/salts medium

List of appendices

CHAPTER 1- Introduction and overview

1.1 Horizontal gene transfer

Gene transfer refers to the movement of genetic information from one cell to another. Horizontal gene transfer (also known as lateral gene transfer) is the movement of genetic information between genetically distinct cells, and this process plays an important role in bacterial evolution. It is believed that horizontal gene transfer (HGT) is the main force behind the acquisition of new, adaptive traits and genetic variation across bacterial strains (Raz and Tannenbaum 2010). It is a common and pervasive evolutionary mechanism occurring in prokaryotes (Gogarten 2003; Soucy, Huang, and Gogarten 2015) and evidence of HGT has also been reported for eukaryotes (Boto 2014; Soucy, Huang, and Gogarten 2015). Using a wide range of techniques to detect HGT events occurring in genomes, several studies have reported that an average of 12% of genes in bacterial genomes have been subjected to HGT (Nakamura et al. 2004; Zhaxybayeva et al. 2006), and other studies suggested that approximately 81% of genes in 181 sequenced prokaryotic genomes have been horizontally transferred in the long history of prokaryotic evolution (Dagan, Randrup, and Martin 2008).

The three traditional mechanisms by which bacterial DNA exchange occurs are transformation, conjugation and transduction, corresponding to uptake of free DNA, cell-cell transfer of plasmid DNA, and phage-mediated DNA transfer, respectively. The first-discovered method of HGT was transformation (Griffith 1928), where cells uptake naked DNA directly from the environment and it can be integrated into the genome. In conjugation, DNA transfer occurs because of cell-to-cell contact between a donor cell and

a recipient cell (Tatum and Lederberg 1947). Transduction is a process in which the genetic material from one bacterial cell is carried to another bacterium by a bacteriophage (Zinder and Lederberg 1952).

In addition to these mechanisms, novel processes are continually being identified such as gene transfer via nanotubes (Dubey and Yehuda 2011) and membrane vesicles (Mashburn and Whiteley 2006; Chiura et al. 2011). Another type of genetic exchange process, which has some commonalities with transduction, is mediated through bacteriophage-like particles called gene transfer agents (GTAs) (Lang, Zhaxybayeva, and Beatty 2012).

1.2 Bacteriophages

Viruses that infect bacteria are known as bacteriophages or phages. They are the most abundant biological entities on the planet, with a global phage count estimate of ~10^{31} (Suttle 2007), and therefore serve as an enormous nutrient and genetic reservoir (Wilhelm and Suttle 1999). They depend on the cellular machinery of bacteria to replicate and produce their progeny. It is believed that phages and their hosts have co-evolved by competing in a powerful "arms race", which has involved the development of novel defense mechanisms by bacteria and the strategic methods of phages to overcome these and successfully infect the cells.

The genetic makeup of phages can be single-stranded (ss) or double-stranded (ds) RNA or DNA and can be linear or circular, segmented or contiguous. However, phages with dsDNA genomes are the most common. The sizes of phage genomes range widely, e.g. from the 3.5-kb ssRNA in phage MS2 that infects *Escherichia coli* to the 500-kb

dsDNA in the *Bacillus* phage G. Phages also display diversity in their structures, although tailed phages are the most common type. A typical tailed phage consists of an icosahedral or prolate capsid head, which contains the dsDNA genome, and a tail that varies in length and structure depending on the phage family. The International Committee on Taxonomy of Viruses (ICTV) classifies phages taxonomically based on their genome and morphology (Ackermann 2009; Lefkowitz et al. 2018). More than 95% of identified phages have dsDNA genomes, possess tails, and belong to the order *Caudovirales*. The three families in this order are the *Myoviridae*, which have long, rigid, contractile tails, *Podoviridae*, which have short, non-contractile tails, and *Siphoviridae*, which have long, non-contractile, flexible tails (Ackermann 2007).

As mentioned earlier, phages depend on bacteria to replicate and produce their progeny. Phage recognize specific receptors on the bacterial cell surface, which partly determines phage specificity for particular host cells. Attachment to the cell is referred to as adsorption and this results in conformational changes in the phage virion that facilitate the transfer of the phage genetic material into the host cell (Young 1992; Karlsson et al. 2003; Puck and Lee 1955; Rakhuba et al. 2010). The exact mechanism of getting the phage genome into the cell varies depending on the phage type.

Phages are differentiated as lytic or temperate based on their life cycle options. Lytic phages replicate within infected cells and are then released into the environment by lysis of the infected cell. Temperate phages, such as phage λ that infects *E. coli*, can enter the lytic replication pathway after infection, or enter into the lysogenic pathway where they exist in a non-replicating prophage state. This can involve integration of their genetic material into the host genome (e.g. λ) or being maintained as an independent replicon

(e.g. P1 (Lobocka et al. 2004) and N15 (Ravin 2011), which both infect *E. coli*). In the lysogenic sate the phage genome is passed to daughter cells during bacterial cell division. Prophages can subsequently be activated to enter the lytic pathway, with replication and release of progeny viruses. In addition to these two "life-cycles" there is pseudo-lysogeny, an unstable stage in which the phage genome fails to replicate or become established as a prophage. Instead, the phage genome remains as a non-integrated "preprophage" for an extended period of time. Pseudolysogeny occurs mostly under nutrient-deprived conditions (Feiner et al. 2015).

It is believed that microbial communities are greatly influenced by phages because they are involved in controlling cell numbers and physiological states (Hennes, Suttle, and Chan 1995; Winter et al. 2004). Both narrow host range (restricted to specific species) and broad host range (capable of infecting multiple species) phages exist and they play a key role in mediating HGT. The abundance of phage-derived genetic information found in bacterial genomes (Canchaya et al. 2003; Casjens 2003) and the number of bacterial genes found in viral metagenomic studies (Rosario and Breitbart 2011) show the active role of phages in the evolution of prokaryotic genomes and their effects on microbial diversity.

1.3 Phage-like gene transfer agents (GTAs) and their distribution

An unusual mechanism of gene transfer, where bacteriophage-like particles transfer random pieces of DNA from one cell to another, was first identified in the α-proteobacterium *Rhodobacter capsulatus* (Marrs 1974), at that time known as *Rhodopseudomonas capsulata*. The process was similar to generalized transduction

(Figure 1.1), with DNase-resistant genetic exchange of antibiotic resistance markers that was not dependent on cell-cell contact (Marrs 1974). However, the particle responsible for the gene transfer differed from the transducing phages known at that time so it was referred to as a "gene transfer agent" (Marrs 1974). Each *R. capsulatus* GTA (now

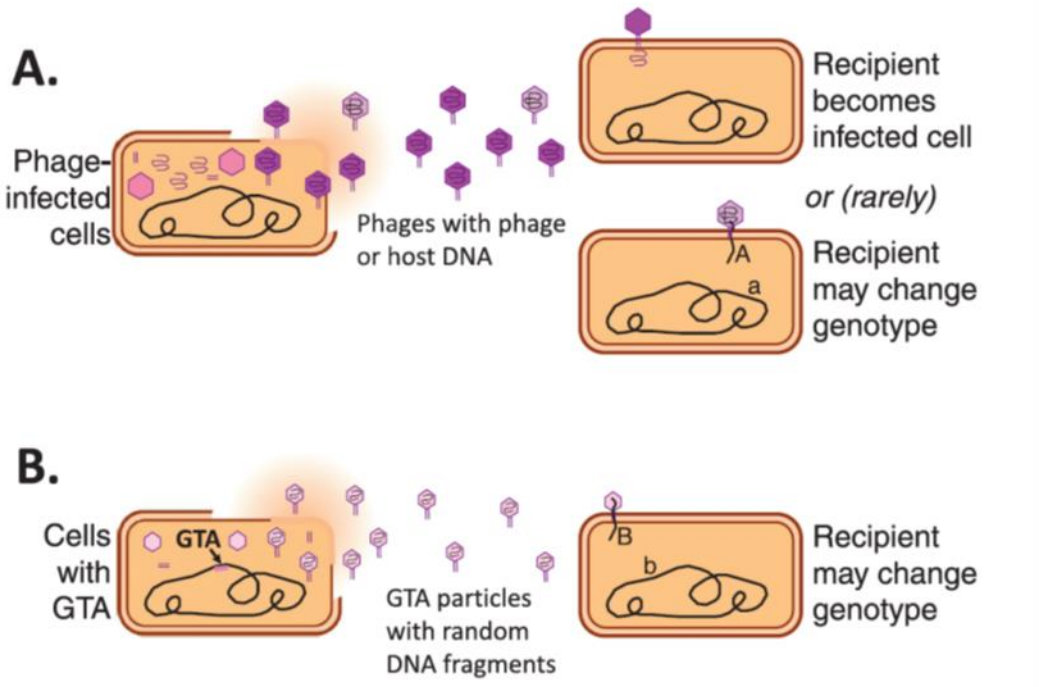


Figure 1.1. Comparison between a typical phage infection and GTA induction. A. Phage infection and transduction B. GTA production and gene transfer. Figure from (Redfield, Beatty, and Lang 2019).

known as RcGTA) particle is capable of packaging ~4 kb of dsDNA in its ~30 nm diameter capsid head and has a tail of ~50 nm in length (Yen, Hu, and Marrs 1979; Solioz and Marrs 1977). The ability to produce GTA and also to take up the DNA carried by GTA is found for most *R. capsulatus* strains, however there are some strains which can only either produce GTA or receive DNA from GTA, and some do neither (Marrs 1974;

Wall, Weaver, and Gest 1975). GTA functions similar to generalized transduction (Figure 1.2) but gene transfer events occur with much greater frequency, and all the genetic markers tested were readily transferred from donor cells to recipients (Yen, Hu, and Marrs 1979; Solioz, Yen, and Marrs 1975; Solioz and Marrs 1977). The main features which differentiate GTAs from bacteriophages are: 1) the DNA packaged within the capsid head is insufficient to encode the GTA structural proteins; 2) only the random parts of the producing cells' genome is packaged by GTA; and 3) GTA production is controlled by bacterial regulatory systems (Lang and Beatty 2007; Lang, Zhaxybayeva, and Beatty 2012).

Isolation of a GTA overproducer mutant strain that produces approximately three orders of magnitude more GTAs than wild type strains (Yen, Hu, and Marrs 1979) greatly aided GTA research. This made it possible to obtain an electron micrograph of purified GTA particles and to visualize the 4-kb packaged DNA directly on an agarose gel (Yen, Hu, and Marrs 1979).

Various prokaryotic species were subsequently shown to be producing GTAs. The δ-proteobacterium *Desulfovibrio desulfuricans* produces a GTA named Dd1 that packages approximately 14-kb fragments of dsDNA (Rapp and Wall 1987). The spirochete *Brachyspira hyodysenteriae* produces a GTA named VSH-1 that packages approximately 7.5 kb of dsDNA (Humphrey et al. 1997). The methanogenic archaeon *Methanococcus voltae* produces VTA (Voltae Transfer Agent) that packages approximately 4 kb of dsDNA (Eiserling et al. 1999). Lastly, the α-proteobacterium *Bartonella henselae* produces BaGTA that packages approximately 14 kb of dsDNA

(Québatte and Dehio 2019). There are structural differences among these different GTAs and it is now known that they are genetically unrelated and presumably originated from independent evolutionary events (Lang, Zhaxybayeva, and Beatty 2012; Lang, Westbye, and Beatty 2017). These discoveries of genetically unrelated GTAs arising independently in different lineages suggest they might have a beneficial role for the producing organisms (Lang, Westbye, and Beatty 2017; Stanton 2007).

1.4 *Rhodobacter capsulatus* and RcGTA

Rhodobacter capsulatus is a purple non-sulphur bacterium that is classified within the order *Rhodobacterales* of the class α-proteobacteria. Purple non-sulfur bacteria are typically found in stagnant or eutrophic aquatic environments and can grow under both aerobic and anaerobic conditions using various sources of organic carbon and terminal electron acceptors (Madigan and Jung 2009; Wall, Weaver, and Gest 1975). *R. capuslatus* is a very versatile organism and can grow phototrophically with either CO_2 or organic carbon as a carbon source, or in darkness by aerobic or anaerobic respiration, fermentation or chemolithotrophy (Pemberton, Horne, and McEwan 1998; Tabita and Govindjee 1995; Dubbs and Tabita 2004). Because of this metabolic diversity, *R. capsulatus* has been extensively used as a model organism for studying different bacterial physiological processes, especially anoxygenic photosynthesis (Imhoff 2006; Madigan 2006; Haselkorn et al. 2001; Pemberton, Horne, and McEwan 1998). *R. capsulatus* has also become a model organism for studying GTAs. The genome of *R. capsulatus* consists of a single 3.7-Mb chromosome and a 134-kb circular plasmid, both having relatively high GC content (66%) (Strnad et al. 2010; Haselkorn et al. 2001). In total, there are 3685 open reading frames (ORFs) with a coding density of 91%. Throughout the genome, there

are several large regions containing functionally related genes, such as the 45-kb photosynthesis gene cluster. It also contains 8 clustered regularly interspaced short palindromic repeat (CRISPR) regions and more than 200 phage-related genes (Strnad et al. 2010).

Although RcGTA was discovered in the early 1970s, it was only in 2000 that the genetic basis of its production started to be deciphered (Lang and Beatty 2000). Screening of a transposon insertion library for mutants that had lost RcGTA production identified three insertions within an approximately 14-kb gene cluster where several genes had homology with phage structural genes (Lang and Beatty 2000). This gene cluster was surrounded by cellular "housekeeping"-type genes and none of the genes had homology with phage genes associated with replication or lysis. Instead, the gene cluster appeared to represent a phage-like head-tail gene module. In addition to this main GTA structural cluster, there are two additional loci that encode proteins associated with RcGTA particles in the form of head spikes and tail fibers that are involved in attachment to recipient cells (Westbye et al 2016, Hynes et al 2016) and another locus encoding holin and lysin proteins for cell lysis and RcGTA release (Hynes et al 2012, Westbye et al 2013) (Figure 1.3).

Phylogenetic studies and advancements in bacterial genomics allowed a description of the evolutionary relationships between species carrying complete or incomplete versions of similar GTA gene clusters (Lang and Beatty 2007; Lang, Taylor, and Beatty 2002). It was found that the GTA-encoding genes are widespread among α-proteobacteria

and are especially well conserved within the order *Rhodobacterales* (Biers et al. 2008; Fu et al. 2010; Lang and Beatty 2007).

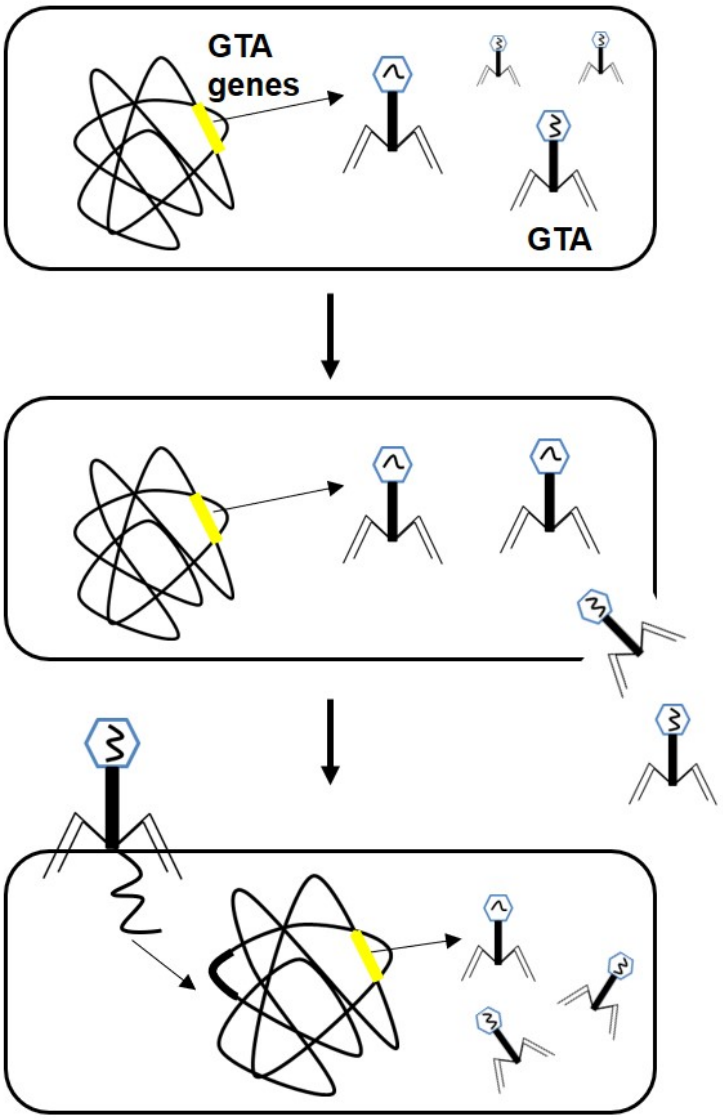


Figure 1.2. Diagrammatic representation of gene transfer agent production and release in *R. capsulatus*. All of the genes responsible for GTA production are present within the producing cell's genome (represented in yellow). Each GTA particle packages random

fragments of host cell DNA (black) and is released by the lysis of cell. Once released, the GTA particles may transfer the DNA to a recipient cell and it can then recombine into the recipient's genome.

The *Roseobacter* lineage in *Rhodobacterales* constitutes >25% of marine prokaryotic communities (Fu et al. 2010). Several of these bacteria have been found to make GTAs (Biers et al. 2008; Tomasch et al. 2018). The presence of GTA-encoding genes among a large number of bacterial species suggests that GTAs might play an important role in gene transfer in nature and evolution of bacterial genomes.

1.5 Regulation of RcGTA gene expression

It was observed during early work on RcGTA that its production varied according to growth conditions and culture growth phase (Solioz 1975; Solioz, Yen, and Marrs 1975), but molecular details on why this variation occurred were not determined until much later. The original transposon mutant screen that identified the RcGTA structural gene cluster also led to the discovery that transcription of these genes was dependent on a cellular two-component system (TCS) involving the histidine kinase CckA and response regulator CtrA (Lang and Beatty 2000). This TCS is now known to be a histidyl-aspartyl phosphorelay that also involves the intermediate histidine phosphotransferase ChpT (Mercer et al 2012). Additional cellular regulators and regulatory systems were also subsequently discovered that affect production and release of RcGTA. These include quorum sensing via N-acyl-homoserine lactone (AHL) synthase (GtaI) and LuxR-like AHL receptor (GtaR) proteins (Leung et al. 2012; Schaefer et al. 2002), a partner switching phosphorelay involving the RbaVWY proteins (Mercer and Lang 2014), and the SOS regulator protein LexA (Kuchinski et al. 2016).

Recent studies on a phage-derived regulatory gene, *gafA*, showed it to be essential for the production of RcGTA (Hynes et al. 2016) and it acts as a transcription factor by binding directly to the GTA promoter (Fogg 2019). Loss of the neighbouring gene, *rcc01866*, also caused impaired GTA production (Hynes et al. 2016). It has become clear that the regulation of RcGTA production and release is quite complex and linked with various aspects of *R. capsulatus* physiology, and many aspects of this still requires more investigation to be fully understood.

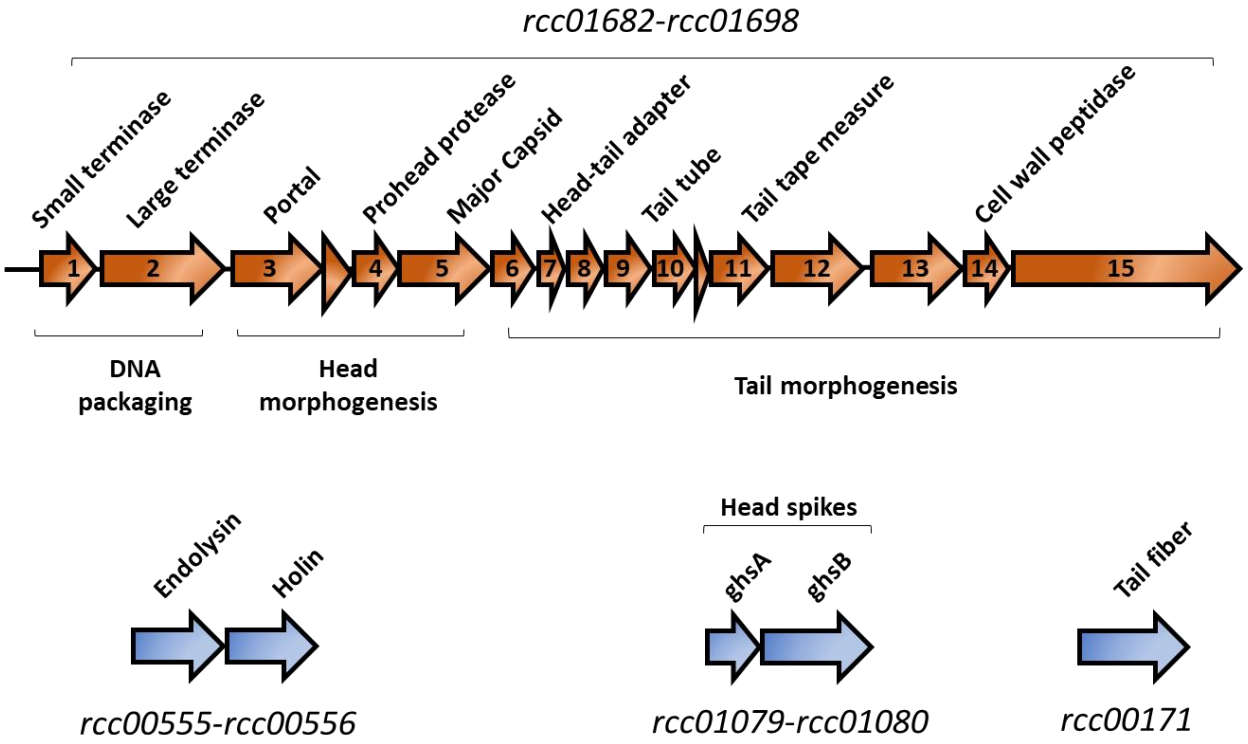


Figure 1.3. RcGTA genes in *Rhodobacter capsulatus*. The ORFs/genes along with known or predicted encoded protein functions are indicated. The major structural cluster genes are in orange and additional genes involved in RcGTA production (lysis, head spikes and tail fiber) located at different loci are in blue. Not to scale.

1.6 Two-component signalling systems

In bacteria, two component systems (TCSs) perceive diverse signals and convert them into appropriate cellular responses. These systems are widely distributed and integrated into a broad range of cellular signalling circuits. They are considered a major signal transducing system in prokaryotes and are also found in some eukaryotes (Jung et al. 2012; West and Stock 2001). A typical TCS is composed of sensory histidine kinase (HK) and response regulator (RR) proteins working to couple the stimulus-response functions (Shiro and Yamada 2008). These systems function by transfer of a phosphate group between the HK and RR, either directly or with the involvement of other proteins (Jung et al. 2012).

1.6.1 Histidine kinase (HK) proteins

In response to a specific stimulus, HK proteins, which are active as homodimers, autophosphorylate at a conserved histidine residue and then transfer the phosphoryl group to an aspartate residue in the cognate RR. In most cases, phosphorylation of the RR activates its output domain to deliver a cellular response. Sometimes, an intermediate protein called a histdine phosphotransferase (Hpt) is required to mediate the transfer, and such systems are known as two-component phosphorelays. In both TCS and phosphorelays, a phosphoryl group is transferred from the transmitter domain of the HK to a conserved aspartate in the receiver (REC) domain of the RR for signal transduction (Bhate et al. 2015; Casino, Rubio, and Marina 2010; Jung et al. 2012).

There is some variability and diversity in domain architectures among HKs, but a minimal HK consists of a sensory domain and a transmitter domain. The kinase activity for class-I HKs resides in two conserved domains, the C-terminal catalytic and ATP

binding (CA) domain and an N-terminal dimerization and histidine phosphotransfer (DHp) domain (together referred to as HK-type ATPase catalytic-HATPase_C domain) and these form the transmitter domain. ATP binds to the CA domain and the phosphoryl group of the ATP is transferred to a conserved histidine residue within DHp domain. The DHp domain is formed by two large antiparallel α-helices from each peptide in the homodimer and creates a four-helical bundle. This bundle provides the dimerization interface and it contains the conserved catalytic histidine residue. In some HKs, the membrane-distal part of the DHp domain forms an interface directly with the RR protein (Tomomori et al. 1999; Bhate et al. 2015). The CA domain binds ATP and has ATPase activity. For autophosphorylation to occur, a conformational change must happen in the CA domain so that it gets closer to the acceptor histidine residue in the DHp domain (Yamada et al. 2009). Each CA domain binds to a single ATP molecule and forms a tight loop known as the "ATP lid", and it has been found that at least one of the CA domains contains ATP or ADP at any given time (Bhate et al. 2015). Depending upon the linker between the two α-helices of the DHp domain, phosphorylation can occur in two different ways, either on the histidine residue of the DHp domain in the same polypeptide (*cis*-autophosphorylation) or in the dimer partner (*trans*-autophosphorylation) (Ashenberg, Keating, and Laub 2013; Podgornaia et al. 2013). Once the histidine residue gets phosphorylated, the aspartate of the receiver domain attacks His~P and results in the formation of Asp~P. This phosphorylation reaction occurs much faster (20-100 min^{-1}) than the autophosphorylation reaction (0.1-5 min^{-1}) (Bhate et al. 2015; Casino, Rubio, and Marina 2010).

Class-II HKs are not as common and are structurally different than class I-HKs. In this type, the histidine is part of a histidine phosphotransfer (Hpt) domain that is not located adjacent to the CA domain (Jung et al. 2012). Some HKs are “hybrid” and contain both a transmitter and a REC domain and perform intramolecular phosphotransfer. The phosphate is then transferred to the histidine residue of an additional HPt protein and finally passed to an aspartate residue on the RR.

HKs can also act as phosphatase enzymes and dephosphorylate the cognate RR (Dutta and Inouye 1996; Huynh, Noriega, and Stewart 2010), but there is no evidence of the phosphoryl group from the REC domain transferring back to the DHp domain to indicate that phosphatase activity is just the reversal of its kinase activity (Yamada et al. 2009). However, studies have shown that the dephosphorylation reaction is catalyzed by the DHp domain and does not require the histidine residue but requires the specific stretch of amino acid sequence next to the acceptor histidine residue (Dutta and Inouye 1996; Huynh, Noriega, and Stewart 2010; Chen et al. 2009). The signals that cause HKs to switch from kinase to phosphatase activity are still unclear and poorly understood. Recent studies in *Caulobacter crescentus* showed that binding of the secondary messenger bis-(3’-5’)-cyclic dimeric guanosine monophosphate (c-di-GMP) switches the HK CckA from kinase to phosphatase activity. Based on some *in vitro* studies, it has been shown that ADP might be acting as an activator for the phosphatase activity, but still needs to be confirmed *in vivo* (Lori et al. 2015).

1.6.2 Response regulator (RR) proteins

In most cases, RR proteins contain an output domain in addition to the REC domain. Phosphorylation of the REC domain by the cognate HK or Hpt protein regulates

the activity of the output domain. Output domain functions vary but helix-turn-helix DNA binding and GGDEF or EAL c-di-GMP-related enzymatic domains are some of the most common (Perry, Koteva, and Wright 2011; Galperin 2010; Krell et al. 2010). Indeed, more than 6% of RR proteins with enzymatic output domains are involved in c-di-GMP, indicating TCS represent a common mechanism for controlling c-di-GMP levels. There are also single domain RR proteins that lack an output domain that can modulate downstream targets by protein-protein interactions (Hecht and Newton 1995; Jenal and Galperin 2009). The CheY protein involved in chemotaxis is one of the best-studied examples of a single domain RR (Stock et al. 1989; Silversmith and Bourret 1999).

Phosphorylation of RRs promotes dimerization and makes them active (Gao and Stock 2010). The architecture of the REC domain consists of a central five-stranded β-sheet surrounded by a total of five α-helices. The C-terminal end of the central β-sheet harbours the conserved aspartate residue that gets phosphorylated. Phosphorylation of the aspartate residue results in a conformational change in the domain that triggers the formation of a homodimerization interface (Bourret 2010; Gao and Stock 2010). REC domains are believed to exist in equilibrium between two conformational states (active and inactive) and studies have also shown that RRs can exhibit autophosphorylation and autodephosphoryaltion activity (*in vitro*) when incubated with an appropriate phosphodonor, such as acetyl phosphate, in addition to being phosphorylated or dephosphorylated by a cognate HK or Hpt protein (Bourret 2010).

Overall, the diversity of stimuli and responses of TCSs is showcased by the number of different HK input and RR output domains. In addition, some HKs have multiple phosphotransfer targets and some RRs can be phosphorylated by multiple

kinases. This many-to-one and one-to-many pattern for TCSs differs from instances where RRs are accidentally phosphorylated by a non-cognate HK, which is considered to be very rare (Laub and Goulian 2007).

1.7 The CckA-ChpT-CtrA phosphorelay

The phosphorelay system involving the HK CckA, Hpt ChpT, and RR CtrA is widely conserved in α-proteobacteria (Brilli et al. 2010), and the functioning of this system is best understood from research with the developmental microbiology model bacterium *Caulobacter crescentus*, where it was also first discovered as an essential regulator of the cell cycle (Quon, Marczynski, and Shapiro 1996). CckA is a hybrid HK protein located at the cytoplasmic membrane and contains both HK and REC domains (Jacobs et al. 1999). Upon receiving an unknown signal, CckA autophosphorylates its HK domain and transfers the phosphate group from the histidine residue to the conserved aspartate residue in its REC domain. This allows phosphorylation of the cognate histidine on ChpT, which can then phosphorylate CtrA's REC domain aspartate residue (Biondi et al. 2006; Chen et al. 2009). All three of these proteins are required for *C. crescentus* viability. Phosphorylated CtrA (CtrA~P) binds directly to the chromosomal origin of replication and inhibits initiation of DNA replication (Quon et al. 1998). It also binds in the upstream regulatory regions of many genes and acts as a transcription factor. The CtrA-regulated genes in *C. crescentus* include genes involved in flagellar motility and cell division (Laub et al. 2000; 2002).

In *C. crescentus*, the kinase activity of CckA is regulated in at least two different ways. One is through the activity of the membrane protein DivL, which stimulates

CckA's kinase activity. The primary activity of CckA is as a phosphatase, which is observed in the absence of DivL and also when *cckA* is over-expressed (Chen et al. 2009; Tsokos, Perchuk, and Laub 2011). But the ability of DivL to stimulate CckA kinase activity is inhibited by phosphorylated DivK, whose phosphorylation state is dependent on PleC and DivJ (Childers et al. 2014; Tsokos and Laub 2012). The second is through the secondary messenger signaling molecule c-di-GMP. C-di-GMP binds to CckA and inhibits its kinase activity and stimulates its phosphatase activity (Lori et al. 2015).

Disruptions of the *R. capsulatus cckA, chpT* and *ctrA* genes affect the production and release of RcGTA particles and flagellar motility (Mercer et al. 2012; Lang and Beatty 2000; Lang and Beatty 2002). The *R. capsulatus* genes are not essential for viability as in *C. crescentus*, but site-directed mutational analysis of the *R. capsulatus* CckA protein showed that an increase in CtrA phosphorylation results in cell filamentation (Westbye et al. 2013), suggesting the possible involvement of this phosphorelay system in cell division in this bacterium also. Although there are some commonalities between the *R. capsulatus* and *C. crescentus* systems, there are also clearly differences. For example, homologs of DivK, PleC and DivJ are absent in *R. capsulatus.*

A common theme for CtrA function in α-proteobacteria is the regulation of genes involved in flagellar motility (Mercer et al. 2010). This was first observed in *C. crescentus* (Leclerc, Wang, and Ely 1998) and subsequently shown in other organisms such as *R. capsulatus* (Lang and Beatty 2002) and *Ruegeria* (previously known as *Silicibacter*) (Miller and Belas 2006). Transcriptomic and proteomic analyses of wild type

and *ctrA* mutant strains of *R. capsulatus* showed that more than 225 genes were dysregulated by the loss of CtrA (Mercer et al. 2010). This included all 73 predicted flagellar motility-associated genes and genes predicted to be involved in pilus synthesis and gas vesicle production. This gene set also included more than 20 genes encoding proteins predicted to be involved in signal transduction pathways or transcriptional regulation, and among these were nine predicted to be involved in signaling via the second messenger c-di-GMP. This suggested there might be a connection between c-di-GMP signaling and RcGTA production.

1.8 Cyclic dimeric guanosine monophosphate (c-di-GMP)

1.8.1 History

In 1987, researchers discovered a novel nucleotide molecule they initially thought was acting as an allosteric activator of cellulose biosynthesis in the fruit-degrading bacterium *Gluconacetobacter xylinus* (Ross et al. 1987). They determined its molecular structure as bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP). They subsequently found that c-di-GMP also affects cellulose synthesis in another α-proteobacterium, *Agrobacterium tumefaciens* (Amikam and Benziman 1989), showing that c-di-GMP was not specific for *G. xylinus* but had wider involvement in bacterial physiology.

The same group identified and sequenced the genes responsible for the metabolism of c-di-GMP (Tal 1998). These genes encode diguanylate cyclase (DGC) enzymes, which catalyse the synthesis reaction, and phosphodiesterase (PDE) enzymes, which break down c-di-GMP. Eventually, sequence analysis of several genes encoding these enzymes revealed conservation of a multi-domain architecture, with the Per-Arnt-

Sym (PAS) domain and GGDEF (named after conserved Gly-Gly-Asp-Glu-Phe amino acid sequence) and EAL (named after conserved Glu-Ala-Leu amino acid sequence) motifs, which has turned out to be the most common architecture in c-di-GMP-metabolizing proteins (Römling, Galperin, and Gomelsky 2013). The GGDEF motif is associated with all known DGC domains and its function was first defined in a response regulator protein, PleD, in *C. crescentus* (Hecht and Newton 1995). The EAL motif is associated with PDE function and its activity was first defined in a transcriptional repressor protein, BvgR, in *Bordetella pertussis* (Merkel, Barros, and Stibitz 1998). Studies involving the heterologous expression of genes containing only the EAL domain resulted in the decrease of c-di-GMP levels and also affected biofilm formation phenotypes (Simm et al. 2004; Tischler and Camilli 2004). These results indirectly supported the role of EAL domain in c-di-GMP degradation. Shortly after, biochemical evidence of EAL domain-associated PDE activity was shown in *Yersinia pestis* where the EAL domain protein HmsP was shown to hydrolyze the non-specific phosphodiesterase substrate bis-p-nitrophenyl phosphate (Bobrov, Kirillina, and Perry 2005). Similar evidence was also found using the EAL domain proteins YahA and DosP from *E. coli* (Schmidt, Ryjenkov, and Gomelsky 2005).

These findings revealed that c-di-GMP metabolism genes were common among bacteria and that this molecule regulated functions beyond cellulose synthesis. Now, after many more years of research, c-di-GMP is recognised as one of the most common bacterial second messengers, and it has been shown to regulate bacterial physiology in many ways (Römling, Galperin, and Gomelsky 2013; Jenal, Reinders, and Lori 2017).

1.8.2 Synthesis and degradation of c-di-GMP

In response to internal and environmental signals, the cellular levels of c-di-GMP are regulated by the actions of two opposing enzyme families: DGCs and PDEs. These two types of enzymes represent the largest known families of signalling proteins in bacteria and are found in all major bacterial phyla. The DGC enzymes catalyse the synthesis reaction by combining two GTP molecules (Chan et al. 2004; A J Schmidt, Ryjenkov, and Gomelsky 2005; Hickman, Tifrea, and Harwood 2005), whereas PDEs catalyse the degradation of cyclic-di-GMP into linear intermediate pGpG or two molecules of GMP. DGCs are characterised by the presence of a GGDEF motif as its functional domain, while PDEs contain either EAL or HD-GYP motifs (Figure 1.4). These proteins are found as single-domain enzymes or as dual-domain proteins that can have one or both of the domains active, as discussed in more detail below

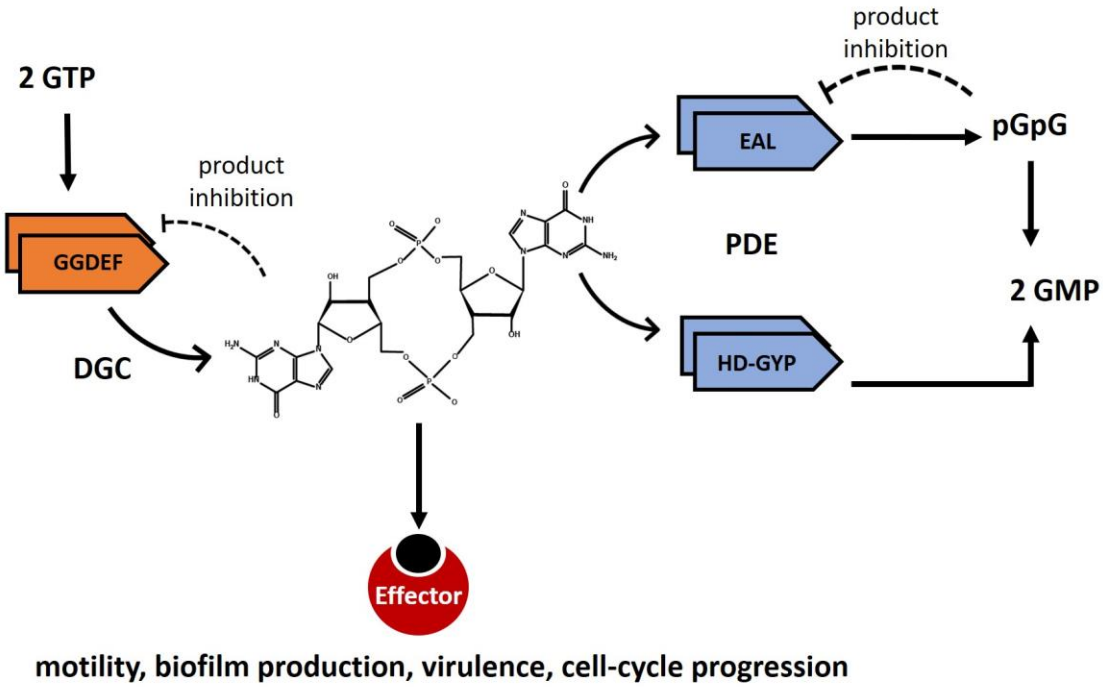


Figure 1.4. Schematic representation of synthesis and degradation of c-di-GMP. Two GTP molecules are used to synthesize c-di-GMP, catalysed by diguanylate cyclase (DGC) enzymes with GGDEF domains. C-di-GMP is hydrolysed into 2 GMP molecules or pGpG by phosphodiesterase (PDE) enzymes having HD-GYP or EAL domains, respectively. Dashed lines indicate feedback inhibition of these enzymes by their respective products. Examples of well-studied physiological process affected by c-di-GMP levels are given.

1.8.3 Diguanylate cyclases (DGCs): GGDEF motif-containing proteins

Structural and biochemical studies performed over the course of a decade on PleD, a DGC from *C. crescentus,* were fundamental for understanding the mechanistic aspects of DGC functioning (Hecht and Newton 1995; Chan et al. 2004). It was found that DGCs show similarity with adenylate cyclases in their structural fold (Pei and Grishin 2001). Each GTP molecule involved in synthesis of c-di-GMP is arranged in an antiparallel manner with the catalytic GGDEF domain to facilitate their condensation into c-di-GMP (Schirmer 2016). The first two glycine residues of the motif participate in GTP binding and the glutamate helps in metal ion coordination. The third amino acid residue of the signature motif, usually aspartate but occasionally a glutamate, also plays an important role in metal coordination and is crucial for DGC activity (Chan et al. 2004; Wassmann et al. 2007). Analysis of the GGDEF domain structures in PleD from *C. crescentus* and WspR from *Pseudomonas aeruginosa* by X-ray crystallography (Wassmann et al. 2007; De et al. 2009) indicated the presence of two sites: an active site (A-site) with the GGDEF motif and an allosteric or inhibitory site (I-site) with an RXXD motif (Christen et al. 2006; Schirmer and Jenal 2009). These two proteins also possess N-

terminal RR REC domains that facilitate dimerization upon phosphorylation (Paul et al. 2007; De et al. 2009). In PleD and WspR, the I-site acts as a receptor site for c-di-GMP to mediate feedback inhibition (Chan et al. 2004; Christen et al. 2006), with binding of c-di-GMP leading to immobilisation of the GGDEF domains into non-catalytic states. In this case, c-di-GMP acts as a non-competitive inhibitor of cyclase activity to maintain its levels at a defined concentration. However, this feature is not conserved in all DGCs. An alternative mechanism of DGC regulation was observed for the DgcZ protein from *E. coli*. DgcZ is a constitutive dimer and does not have an I-site. Instead, it contains an N-terminal chemoreceptor zinc-binding (CZB) domain in addition to the catalytic GGDEF domain, and this CZB domain functions in allosteric regulation of DgcZ activity (Zähringer et al. 2013). The binding of zinc to the CZB domain leads to a conformation change that affects the dimerization of the catalytic GGDEF site and turns it into an inactive form, whereas in the absence of zinc, the GGDEF residues remain in the proper positions as required for dimerization and synthesis of c-di-GMP.

1.8.4 Phosphodiesterases (PDEs): EAL or HD-GYP motif-containing proteins

Phosphodiesterase enzymes that carry out c-di-GMP hydrolysis contain either EAL or HD-GYP motifs in their PDE domains. EAL-containing PDEs degrade c-di-GMP into the linear 5'-phosphoguanylyl-(3'-5')-guanosine (pGpG) dinucleotide (Ross et al. 1986), whereas PDEs with HD-GYP motifs hydrolyse c-di-GMP into two GMP molecules (Ryan et al. 2006). Recent studies on EAL-containing PDEs demonstrated that the hydrolysis product pGpG is later processed to GMP by nanoRNA-degrading oligoribonuclease (Orn) (Orr et al. 2015; Cohen et al. 2015). Biochemical studies and structural analysis of PDEs have shown that the glutamic acid (E) in the EAL motif is

crucial for catalysis. The active site activates a water molecule that attacks and results in breakage of the P-O bond of c-di-GMP to generate a linear molecule. The reaction is affected by divalent cations, with Mg^{2+} or Mn^{2+} enhancing and Ca^{2+} and Zn^{2+} inhibiting PDE activity (Schmidt, Ryjenkov, and Gomelsky 2005).

Unlike DGCs, which are active as homodimers, EAL-containing proteins can retain PDE activity as monomers (Schmidt, Ryjenkov, and Gomelsky 2005). However, the EAL domain proteins characterized so far act as dimers or higher order oligomers (Tarutina, Ryjenkov, and Gomelsky 2006; Barends et al. 2009; Rao et al. 2008; Tchigvintsev et al. 2010). It also appears that the dimeric state is critical for PDE activation by environmental stimuli (Bai et al. 2012) and the dimeric state is thought to be the most probable functional form for hydrolysing c-di-GMP *in vivo* (Romling, Galperin, and Gomelsky 2013).

Another class of PDEs that are structurally and catalytically different from the EAL-containing enzymes have conserved HD-GYP motifs in their PDE domains. As mentioned earlier, these type of PDEs hydrolyse c-di-GMP into two GMP molecules without any intermediate steps (Bellini et al. 2014). HD-GYP is a subset of larger HD family displaying hydrolytic activity (Ryan et al. 2006). The c-di-GMP-specific PDE activity of the HD-GYP domain was originally predicted due to its frequent occurrence in proteins that also possessed DGC domains, similar to the tandem GGDEF-EAL motif-containing proteins (Galperin et al. 1999; Galperin, Nikolskaya, and Koonin 2001). Several HD-GYP-containing PDEs were identified in different bacteria such as *Xanthomonas* (Ryan et al. 2006), *Pseudomonas* (Ryan et al. 2009) and *Borrelia* (Sultan et al. 2011), but there were no clear mechanistic or biochemical insights about these proteins

until more recently. The first crystal structure of an active HD-GYP-containing PDE with its c-di-GMP substrate revealed the presence of a novel tri-nuclear catalytic binding site involving Fe^{2+} or Mn^{2+} (Bellini et al. 2014).

1.8.5 Tandem GGDEF and EAL or HD-GYP proteins

The first DGCs and PDEs identified in *G. xylinus* showed the presence of both GGDEF and EAL domains in a single protein (Tal et al. 1998; Chang et al. 2001). Genomic studies have revealed that such multidomain proteins are very common in bacteria, with approximately one-third of all GGDEF domains and two-thirds of all EAL domains found in this organization (Seshasayee, Fraser, and Luscombe 2010). The presence of these antagonistically functioning domains in the same protein raises an enzymatic conundrum about what determines the overall activity of the protein. This is elaborated further below.

1.8.5.1 Bifunctional enzymes and their regulation

One possible scenario for tandem GGDEF/EAL proteins is that both domains are enzymatically active, but they are differentially regulated with only one domain functioning at a given time depending upon external or internal signals. Such potentially bifunctional proteins have been well studied and include proteins that are the REC domain-containing RR part of TCSs that are regulated by HK proteins. Even though the number of tandem GGDEF/EAL proteins is large, only a few cases of truly bifunctional proteins have been described.

The *Rhodobacter sphaeroides* BphG1 is one of the examples of a bifunctional protein with tandemly arranged GGDEF-EAL domains. BphG1 is a bacteriophytochrome

protein with a photosensory module of PAS-GAF-PHY domains linked to GGDEF-EAL domains. Despite the presence of the photosensory module that is sensitive to light, this protein showed constitutive PDE activity irrespective of light (Tarutina, Ryjenkov, and Gomelsky 2006; Romling, Galperin, and Gomelsky 2013). However, when a truncated version of the protein lacking the C-terminal EAL domain was expressed in *E. coli* it displayed strong DGC activity in response to light (Tarutina, Ryjenkov, and Gomelsky 2006). In *Legionella pneumophilla*, the Lp10329 protein contains active DGC and PDE domains, with the activities regulated based upon the phosphorylation state of its REC domain (Levet-Paulo et al. 2011). The cognate histidine kinase, Lp10330, phosphorylates the conserved aspartate residue of Lp10329, which results in lower DGC activity of Lp10329 without affecting the PDE activity.

Protein-protein interactions can also modulate DGC and PDE activities. In *Vibrio parahaemolyticus*, the *scrC* gene which belongs to the *scrABC* operon, regulates the switch between motile and sessile cell forms (Boles and McCarter 2002). ScrC contains functional DGC and PDE domains linked to an N-terminal periplasmic sensor domain (Ferreira et al. 2008). When *scrC* is expressed alone, the protein shows DGC activity, whereas in the presence of ScrA and ScrB its activity is switched to PDE. At high cell densities, binding of ScrB to ScrC triggers the switch of its activity from DGC to PDE (Trimble and McCarter 2011).

1.8.5.2 Active and inactive domains

Another explanation for the enzymatic conundrum in some c-di-GMP signaling proteins is that one of the two domains is enzymatically inactive. However, the

enzymatically inactive domain can display other functions that affect the function of the other domain (Römling, Galperin, and Gomelsky 2013). The well-studied *C. crescentus* protein CC3396 is one of the best examples for a c-di-GMP protein carrying one active and one inactive domain. The DGC A-site in this protein contains an altered motif (GEDEF) and is not enzymatically active. However, this site displays very high affinity for GTP and binding of GTP at this inactive A-site leads to activation of the C-terminal EAL domain and increased c-di-GMP hydrolysis (Christen et al. 2005). The FimX protein from *P. aeruginosa* has similar properties, with an inactive DGC domain that acts as a receptor for GTP and activates the EAL domain (Kazmierczak, Lebron, and Murray 2006). Identification of an inactive EAL domain that binds to c-di-GMP and activates a GGDEF domain has not yet been reported, but studies on DgcA1 from *G. xylinus* showed that deletion of its enzymatically inactive EAL domain eliminated the protein's DGC activity, suggesting the potential importance of such degenerate EAL domains in the overall protein activity (Römling, Galperin, and Gomelsky 2013). Evaluation of EAL domain sequences from different bacterial genomes suggested that approximately 85% are enzymatically active (Seshasayee, Fraser, and Luscombe 2010).

1.8.6 Regulation of DGC and PDE activities by additional signalling domains

In most c-di-GMP signalling proteins found so far, the GGDEF, EAL and HD-GYP domains are linked with other N-terminal domains that are presumably involved in regulation of the enzymatic activities of the protein. These presumptive regulatory domains are often associated with the cytoplasmic membrane and can contain periplasmic loops that interact with small ligands or other periplasmic proteins. These domains include PAS, GAF, REC, CHASE (cyclases/histidine kinases associated sensory

extracellular), HAMP (present in histidine kinases, adenylate cyclases, methyl accepting proteins and phosphatases), the light sensing BLUF (blue light using flavin adenine dinucleotide) and various bacteriophytochrome domains. These N-terminal domains modulate the DGC and PDE enzyme activities in response respect to the specific stimuli they receive, but for most of these proteins the sensory input signals involved have not been identified. The most common domain architectures known are the combinations of REC, PAS, and/or GAF domains with GGDEF, EAL or HD-GYP domains (Römling, Galperin, and Gomelsky 2013). Indeed, the first-identified DGC/PDE of *G. xylinus* showed the association of GGDEF and EAL domains with an oxygen-sensing PAS domain (Tal et al. 1998; Chang et al. 2001). The PAS and GAF domains are involved in sensing various signals such as oxygen (O_2), nitric oxide (NO), carbon monoxide (CO), light, and quorum sensing molecules, and employ various bound ligands such as heme, flavin nucleotides, and different chromophores (Henry and Crosson 2010; Ho 2000; Deng et al. 2012). As discussed above, REC domains are regulated through their phosphorylation by cognate HKs, which are responsible for sensing the ultimate signal that controls the RR's output functionality.

In *P. aeruginosa*, two proteins, MucR and NbdA, each of which contain an N-terminal MHYT (methionine, histidine, tyrosine) domain followed by GGDEF and EAL domains, showed NO-induced PDE activity and promoted biofilm dispersal (Li et al. 2013). MHYT is a transmembrane domain predicted to possess the sensory function for diatomic gases like O_2, CO or NO through protein-bound copper ions. Addition of NO in this bacterium lead to the transcriptional activation of *nbdA* and thereby activated PDE

activity. MucR displayed both DGC and PDE activity *in vitro* but increased PDE activity with NO exposure (Li et al. 2013).

There have been multiple reports showing the importance of light-sensing (ranging from blue to far red) domains such as BLUF and LOV (light-oxygen-voltage sensing) controlling GGDEF/EAL domain proteins. In *Synechocystis* sp., the Cph2 protein consists of a GAF-GAF-GGDEF-EAL-CBCR (cyanobacteriochrome)-GGDEF multi-domain architecture and responds to blue light via the CBCR domain, which activates the DGC by stimulating the C-terminal GGDEF domain. The EAL domain's PDE activity is predicted to be controlled by the N-terminal GGDEF or GAF domains (Savakis et al. 2012). In other proteins, a novel membrane-integrated sensory domain containing a CSS (Cys-Ser-Ser amino acids) motif that is redox-regulated controls the activity of EAL domains (Hengge et al. 2016; Herbst et al. 2018). In *E. coli* there are five PDEs with CSS-EAL combinations that are all inactive due to the formation of a disulphide bond between the cysteine residues of the CSS motif. Mutations of the CSS motif to ASS resulted in high PDE activity in these proteins (Herbst et al. 2018).

Even though we have gained a large amount of knowledge on the structures and functions of DGCs and PDEs, testing and validating the physiological roles of individual enzymes under laboratory conditions can be a difficult task. The success rate for genetic studies to reveal a clear phenotype relevant for c-di-GMP signalling proteins is low. One main reason for this is difficulty with identification of specific signals that activate these enzymes. One good example for this was the study of PDEs in *E. coli*. Among 13 PDEs encoded in the genome, only one PDE protein, PdeH, was shown to hydrolyze c-di-GMP and act as a key regulator of motility (Reinders et al. 2016; Schmidt, Ryjenkov, and

Gomelsky 2005). However, gain-of-function mutants showed that other PDEs are active and substituted for the loss of *pdeH* in the stimulation of motility. This suggested that the specific input signals to activate these other PDEs in this organism were simply missing under the laboratory conditions used in the experiments (Reinders et al. 2016).

It is also important to note that DGCs and PDEs are not only working to maintain the homeostasis of c-di-GMP, but some DGCs and PDEs can also interact directly with effector molecules to participate in downstream signalling cascades (e.g. via protein-protein interactions) and thereby control various cellular processes (Tal et al. 1998; Lindenberg et al. 2013). In this case, these proteins are also acting as c-di-GMP sensors and controlling the activity of the interacting proteins. This activity is not limited to the so called "degenerate proteins" that have lost their catalytic abilities and act as receptors but is also observed in some active DGCs and PDEs. For example, PdeR (formerly known as YciR) in *E. coli* is an active PDE whose primary role is not to hydrolyse c-di-GMP but to sense it and affect the transcription of a downstream cascade of genes involved in amyloid curli fiber production via protein-protein interactions (Lindenberg et al. 2013).

1.9 Research goals and questions addressed

The initial hypothesis that c-di-GMP signaling might affect RcGTA production in *R. capsulatus* was based on the fact that CtrA is required for RcGTA production, at the level of RcGTA gene transcription, and that loss of CtrA affected the expression of multiple genes encoding predicted c-di-GMP signalling proteins (Mercer et al. 2010). Therefore, my primary research goal was to determine if some or all of these predicted c-di-GMP signalling genes that are dysregulated by the loss of CtrA are involved in regulating RcGTA production. I also evaluated the role of these genes in flagellar

motility. Furthermore, I evaluated the enzymatic activities of the relevant c-di-GMP signalling proteins. Related to this, I also investigated the effects of changing the intracellular concentrations of c-di-GMP on RcGTA production and motility. The results from this work are presented in Chapter 2. Revelations on the enzymatic activities of one of these proteins, Rcc00620, led me to investigate its involvement as part of a potential TCS, which is presented in Chapter 3. In the model bacterium *C. crescentus*, where the CtrA phosphorelay has been extensively studied, it was shown that c-di-GMP binds to the histidine kinase CckA and modulates its function and thereby the phosphorylation status of CtrA (Lori et al. 2015). I have investigated whether c-di-GMP also binds to CckA in *R. capsulatus* and thereby in turn also affects the CckA-ChpT-CtrA phosphorelay and regulation of RcGTA production in *R. capsulatus*. This work is presented in Chapter 4.

1.10 References

Ackermann, H. -W. (2007). 5500 Phages examined in the electron microscope.*Arch. Virol.* 152, 227–243. doi.org/10.1007/s00705-006-0849-1.

Ackermann, H. -W. (2009). Phage classification and characterization. *Methods. Mol. Biol*. 501, 127–140. doi.org/10.1007/978-1-60327-164-6_13.

Amikam, D., and Benziman, M. (1989). Cyclic diguanylic acid and cellulose synthesis in *Agrobacterium tumefaciens*. *J. Bacteriol*. 171, 6649-6655. doi.org/10.1128/jb.171.12.6649-6655.1989.

Ashenberg, O., Keating, A. E., and Laub, M. T. (2013). Helix bundle loops determine whether histidine kinases autophosphorylate in cis or in trans. *J. Mol. Biol*. 425, 1198-1209. doi.org/10.1016/j.jmb.2013.01.011.

Bai, Y., Yang, J., Zhou, X., Ding, X., Eisele, L. E., and Bai, G. (2012). *Mycobacterium tuberculosis* Rv3586 (DacA) is a diadenylate cyclase that converts ATP or ADP into c-di-amp. *PLoS One*. 7, e35206. doi.org/10.1371/journal.pone.0035206.

Barends, T. R. M., Hartmann, E., Griese, J. J., Beitlich, T., Kirienko, N. V., Ryjenkov, D. A., Reinstein, J., Shoeman, R. L., Gomelsky, M., and Schlichting, I. (2009). Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase. *Nature*. 459, 1015-1018. doi.org/10.1038/nature07966.

Bellini, D, Caly, D. L., Mccarthy, Y., Bumann, M., An, S. Q., Dow, J. M., Ryan, R. P., and Walsh, M. A. (2014). Crystal structure of an HD-GYP domain cyclic-di-gmp phosphodiesterase reveals an enzyme with a novel trinuclear catalytic iron centre. *Mol. Microbiol*. 91, 26-38. doi.org/10.1111/mmi.12447.

Bhate, M. A. P., Molnar, K. A. S., Goulian, M., and Degrado, W. F. (2015). Signal transduction in histidine kinases: Insights from new structures. *Structure*. 23, 981-

984. doi.org/10.1016/j.str.2015.04.002.
Biers, E. J., Wang, K., Pennington, C., Belas, R., Chen, F., and Moran, M. A. (2008). Occurrence and expression of gene transfer agent genes in marine bacterioplankton. *Appl. Environ. Microbiol*. 74, 2933-2939. doi.org/10.1128/AEM.02129-07.
Biondi, E. G., Reisinger, S. J., Skerker, J. M., Arif, M., Perchuk, B. S., Ryan, K. R., et al. (2006). Regulation of the bacterial cell cycle by an integrated genetic circuit. *Nature* 444, 899–904. doi:10.1038/nature05321.
Bobrov, A. G., Kirillina, O., and Perry, R. D. (2005). The phosphodiesterase activity of the HmsP EAL domain is required for negative regulation of biofilm formation in *Yersinia pestis*. *FEMS Microbiol. Lett*. 247, 123-130. doi.org/10.1016/j.femsle.2005.04.036.
Boles, B. R., and McCarter, L. L. (2002). *Vibrio parahaemolyticus* ScrABC, a novel operon affecting swarming and capsular polysaccharide regulation. *J. Bacteriol*. 184, 5946-5954. doi.org/10.1128/JB.184.21.5946-5954.2002.
Boto, L. (2014). Horizontal gene transfer in the acquisition of novel traits by metazoans. *Proc.Biol. Sci.* 281, 1777. doi.org/10.1098/rspb.2013.2450.
Bourret, R. B. (2010). Receiver domain structure and function in response regulator proteins. *Curr. Opin. Microbio*. 13, 142-149. doi.org/10.1016/j.mib.2010.01.015.
Brilli, M., Fondi, M., Fani, R., Mengoni, A., Ferri, L., Bazzicalupo, M., et al. (2010). The diversity and evolution of cell cycle regulation in alpha-proteobacteria: a comparative genomic analysis. *BMC Syst. Biol.* 4, 52. doi:10.1186/1752-0509-4-52.
Canchaya, C., Proux, C., Fournous, G., Bruttin, A., and Brussow, H. (2003). Prophage Genomics. *Microbiol. Mol. Biol. Rev.* 67, 238–276. doi.mmbr.asm.org/cgi/content/abstract/67/2/238.
Casino, P., Rubio, V., and Marina, A. (2010). The mechanism of signal transduction by two-component systems. *Curr. Opin. Struct. Biol*. 20, 763-771. doi.org/10.1016/j.sbi.2010.09.010.
Casjens, S. (2003). Prophages and bacterial genomics: What have we learned so far? *Mol. Microbiol.* 49, 277–300. doi/10.1046/j.1365-2958.2003.03580.x/abs.
Chan, C., Paul, R., Samoray, D., Amiot, N. C., Giese, B., Jenal, U., et al. (2004). Structural basis of activity and allosteric control of diguanylate cyclase. *Proc. Natl. Acad. Sci. U. S. A.* 101, 17084–17089. doi:10.1073/pnas.0406134101.
Chang, A. L., Tuckerman, J. R., Gonzalez, G., Mayer, R., Weinhouse, H., Volman, G., et al. (2001). Phosphodiesterase A1, a regulator of cellulose synthesis in *Acetobacter xylinum*, is a heme-based sensor. *Biochemistry* 40, 3420–3426. doi:10.1021/bi0100236.
Chen, Y. E., Tsokos, C. G., Biondi, E. G., Perchuk, B. S., and Laub, M. T. (2009). Dynamics of two phosphorelays controlling cell cycle progression in *Caulobacter crescentus*. *J. Bacteriol*. 191, 7417-7429. doi.org/10.1128/JB.00992-09.
Childers, W. S., Xu, Q., Mann, T. H., Mathews, I. I., Blair, A. J., Deacon, A. M., and Shapiro, L. (2014). Cell fate regulation governed by a repurposed bacterial histidine kinase. *PLoS Bio*. 12, 10. doi.org/10.1371/journal.pbio.1001979.
Chiura, H. X., Kogure, K., Hagemann, S., Ellinger, S., and Velimirov, B. (2011). Evidence for particle-induced horizontal gene transfer and serial transduction between bacteria. *FEMS Microbiol. Ecol.* 76, 576–91. doi.org/10.1111/j.1574-6941.2011.01077.x.
Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P., et al. (2006).

Allosteric control of cyclic di-GMP signaling. *J. Biol. Chem.* 281, 32015–32024. doi:10.1074/jbc.M603589200.

Christen, M., Christen, B., Folcher, M., Schauerte, A., and Jenal, U. (2005). Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. *J. Biol. Chem.* 280, 30829–30837. doi:10.1074/jbc.M504429200.

Cohen, D., Mechold, U., Nevenzal, H., Yarmiyhu, Y., Randall, T. E., Bay, D. C., Rich, J. D, et al. (2015). Oligoribonuclease is a central feature of cyclic diguanylate signaling in *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U. S. A.* 36, 11359-11364. doi.org/10.1073/pnas.1421450112.

Dagan, T., Randrup, Y. A., and Martin, W. (2008). Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc. Natl. Acad. Sci. U. S. A.* 105, 10039-10044. doi.org/10.1073/pnas.0800679105.

De, N. Navarro, M. V. A. S., Raghavan, R. V., and Sondermann, H. (2009). Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. *J. Mol.Biol.* 393, 619–633. doi.org/10.1016/j.jmb.2009.08.030.

Deng, Y., Schmid, N., Wang, C., Wang, J., Pessi, G., Wu, D., Lee, J., et al. (2012). Cis-2-dodecenoic acid receptor RpfR links quorum-sensing signal perception with regulation of virulence through cyclic dimeric guanosine monophosphate turnover. *Proc. Natl. Acad. Sci. U. S. A.* 109, 15479-15484. . doi.org/10.1073/pnas.1205037109.

Dubbs, J. M., and Tabita, F. R. (2004). Regulators of nonsulfur purple phototrophic bacteria and the interactive control of CO2 assimilation, nitrogen fixation, hydrogen metabolism and energy generation. *FEMS Microbiol. Rev.* 28, 353-376. doi.org/10.1016/j.femsre.2004.01.002.

Dubey, G. P., andYehuda, S. B. (2011). Intercellular nanotubes mediate bacterial communication. *Cell.* 144, 590–600. doi.org/S0092-8674(11)00016-X..

Dutta, R., and Inouye, M. (1996). Reverse phosphotransfer from OmpR to EnvZ in a kinase-/phosphatase+ mutant of envz (EnvZ·N347D), a bifunctional signal transducer of *Escherichia coli*. *J. Biol. Chem.* 271, 1424-1429. doi.org/10.1074/jbc.271.3.1424.

Eiserling, F., Pushkin, A., Gingery, M., and Bertani, G. (1999). Bacteriophage-like particles associated with the gene transfer agent of *Methanococcus voltae* ps. *J. Gen. Virol.* 80, 3305–3308. doi:10.1099/0022-1317-80-12-3305

Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I., and Herskovits, A. A. (2015). A new perspective on lysogeny: Prophages as active regulatory switches of bacteria. *Nat. Rev. Microbiol*. 13, 641-650. doi.org/10.1038/nrmicro3527.

Ferreira, R. B.R., Antunes, L. C. M., Greenberg, E. P., and McCarter, L. L. (2008). *Vibrio parahaemolyticus* ScrC modulates cyclic dimeric gmp regulation of gene expression relevant to growth on surfaces. *J. Bacteriol*. 190, 851-860. doi.org/10.1128/JB.01462-07.

Fogg, P. C. M. (2019). Identification and characterization of a direct activator of a gene transfer agent. *Nat. Commun.* 10, 595. doi:10.1038/s41467-019-08526-1.

Fu, Y., MacLeod, D. M., Rivkin, R. B., Chen, F., Buchan, A., and Lang, A. S. (2010). High diversity of *Rhodobacterales* in the subarctic north atlantic ocean and gene transfer agent protein expression in isolated strains. *Aquat. Microb. Ecol.* 59, 283–293. doi:10.3354/ame01398.

Galperin, M. Y., Natale, D. A., Aravind, L., and Koonin, E. V. (1999). A specialized

version of the HD hydrolase domain implicated in signal transduction. *J. Mol. Microbiol. Biotechnol*. 1, 303-305.
Galperin, M. Y. (2010). Diversity of structure and function of response regulator output domains. *Curr. Opin. Microbiol*. 150, 9. doi.org/10.1016/j.mib.2010.01.005.
Galperin, M. Y., Nikolskaya, A. N., and Koonin, E. V. (2001). Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiol. Lett*. 203, 11-21. doi.org/10.1016/S0378-1097(01)00326-3.
Gao, R., and Stock, A. M. (2010). Molecular strategies for phosphorylation-mediated regulation of response regulator activity. *Curr. Opin.Microbiol*. 13, 160-167. doi.org/10.1016/j.mib.2009.12.009.
Gogarten, J. P. (2003). Gene transfer: gene swapping craze reaches eukaryotes. *Curr. Biol*. 13, R53-R54. doi.org/10.1016/S0960-9822(02)01426-4.
Griffith, F. (1928). The significance of *Pneumococcal* types. *J. Hyg.* 27, 113–159. doi.org/10.1017/S0022172400031879.
Haselkorn, R., Lapidus, A., Kogan, Y., Vlcek, C., Paces, J., Paces, V., Ulbrich, P., et al. (2001). The *Rhodobacter capsulatus* genome. *Photosynth. Res.* 70, 43–52. doi.org/10.1023/A:1013883807771.
Hecht, G. B., and Newton, A., (1995). Identification of a novel response regulator required for the swarmer-to- stalked-cell transition in *Caulobacter crescentus*. *J. Bacteriol*. 177, 6223-6229. doi.org/10.1128/jb.177.21.6223-6229.1995.
Hengge, R., Galperin, M. Y., Ghigo, J. M., Gomelsky, M., Green, J., Hughes, K. T., Jenal, U., and Landini, P. (2016). Systematic nomenclature for GGDEF and EAL domain containing cyclic di-gmp turnover proteins of *Escherichia coli*. *J. Bacteriol*. 198, 7-11. doi.org/10.1128/JB.00424-15.
Hennes, K. P., Suttle, C. A., and Chan, A. M. (1995). Fluorescently labled virus probes show that natural virus populations can control the structure of marine microbial communities. *Appl. Environ. Microbiol.* 61,3623–27.
Henry, J. T., and Crosson, S. (2010). Ligand-binding PAS domains in a genomic, cellular, and structural context. *Ann. Rev. Microbiol*. 65, 261-286. doi.org/10.1146/annurev-micro-121809-151631.
Herbst, S., Lorkowski, M., Sarenko, O., Nguyen, T. K. L., Jaenicke, T., and Hengge, R. (2018). Transmembrane redox control and proteolysis of PdeC, a novel type of c-di-gmp phosphodiesterase. *EMBO J.*. 37, e97825. doi.org/10.15252/embj.201797825.
Hickman, J. W., Tifrea, D. F,. and Harwood, C. S. (2005). A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. *Proc. Natl. Acad. Sci. U. S. A.* 102, 14422-14427. doi.org/10.1073/pnas.0507170102.
Ho, Y.-S. J. (2000). Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic gmp receptor. *EMBO J.* 19, 5288-5299.. doi.org/10.1093/emboj/19.20.5288.
Humphrey, S. B., Stanton, T. B., Jensen, N. S., and Zuerner, R. L., (1997). Purification and characterization of VSH-1, a generalized transducing bacteriophage of *Serpulina hyodysenteriae*. *J. Bacteriol.* 179, 323–329. doi.org/10.1128/jb.179.2.323-329.1997.
Huynh, T. N., Noriega, C. E., and Stewart, V. (2010). Conserved mechanism for sensor phosphatase control of two-component signaling revealed in the nitrate sensor NarX. *Proc. Natl. Acad. Sci. U. S. A.* 107, 21140-21145. doi.org/10.1073/pnas.1013081107.
Hynes, A. P., Shakya, M., Mercer, R. G., Grüll, M. P., Bown, L., Davidson, F., et al. (2016). Functional and evolutionary characterization of a gene transfer agent's

multilocus “genome.” *Mol. Biol. Evol.* 33, 2530–2543. doi:10.1093/molbev/msw125.

Imhoff, J. F. (2006). Taxonomy and physiology of phototrophic purple bacteria and green sulfur bacteria. *Anoxy. Photosyn. Bact.*. 2, 1-15. doi.org/10.1007/0-306-47954-0_1.

Jacobs, C., Domian, I. J., Maddock, J. R., and Shapiro, L. (1999). Cell cycle-dependent polar localization of an essential bacterial histidine kinase that controls DNA replication and cell division. *Cell* 97, 111–120. doi:10.1016/S0092-8674(00)80719-9.

Jenal, U., and Galperin, M. Y. (2009). Single domain response regulators: Molecular switches with emerging roles in cell organization and dynamics. *Curr. Opin.Microbiol*. 12, 152-160. doi.org/10.1016/j.mib.2009.01.010.

Jenal, U., Reinders, A., and Lori, C. (2017). Cyclic di-GMP: second messenger extraordinaire. *Nat. Rev. Microbiol.* 15, 271–284. doi:10.1038/nrmicro.2016.190. doi.org/10.1038/nrmicro.2016.190.

Jung, K., Fried, L., Behr, S., and Heermann, R. (2012). Histidine kinases and response regulators in networks. *Curr. Opin.Microbiol*. 15, 118-124. doi.org/10.1016/j.mib.2011.11.009.

Karlsson, F., Borrebaeck, C. A. K., Nilsson, N., and Malmborg-Hager, A. C.(2003). The mechanism of bacterial infection by filamentous phages involves molecular interactions between TolA and phage protein 3 domains. *J. Bacteriol*. 185, 2628-2634. doi.org/10.1128/JB.185.8.2628-2634.2003.

Kazmierczak, B. I., Lebron, M. B., and Murray, T. S. (2006). Analysis of FimX, a phosphodiesterase that governs twitching motility in *Pseudomonas aeruginosa*. *Mol. Microbiol*. 60, 1026-1043. doi.org/10.1111/j.1365-2958.2006.05156.x.

Krell, T., Lacal, J., Busch, A., Jiménez, H. S., Guazzaroni, M. E., and Ramos, J. L.(2010). Bacterial sensor kinases: Diversity in the recognition of environmental signals. *Ann. Rev. Microbiol*. 64, 539-559. doi.org/10.1146/annurev.micro.112408.134054.

Kuchinski, K. S., Brimacombe, C. A., Westbye, A. B., Ding, H., and Beatty, T. J. (2016). The SOS response master regulator LexA regulates the gene transfer agent of *Rhodobacter capsulatus* and represses transcription of the signal transduction protein CckA. *J. Bacteriol.* 198, 1137–1148. doi:10.1128/JB.00839-15.

Lang, A. S., and Beatty, J. T. (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. doi:10.1073/pnas.97.2.859.

Lang, A. S., and Beatty, J. T. (2002). A bacterial signal transduction system controls genetic exchange and motility. *J. Bacteriol.* 184, 913–918. doi:10.1128/jb.184.4.913-918.2002.

Lang, A. S., Taylor, T. A., and Beatty, J. T. (2002). Evolutionary implications of phylogenetic analyses of the gene transfer agent (GTA) of *Rhodobacter capsulatus*. *J. Mol. Evol.* 55, 534–43. doi:10.1007/s00239-002-2348-7.

Lang, A. S., and Beatty, J. T. (2007). Importance of widespread gene transfer agent genes in α-proteobacteria. *Trends Microbiol.* 15, 54–62. doi:10.1016/j.tim.2006.12.001.

Lang, A. S., Westbye, A. B., and Beatty, J. T. (2017). The distribution, evolution, and roles of gene transfer agents in prokaryotic genetic exchange. *Annu. Rev. Virol.* 4, 87–104. doi:10.1146/annurev-virology-101416-041624.

Lang, A. S., Zhaxybayeva, O., and Beatty, J. T. (2012). Gene transfer agents: Phage-like elements of genetic exchange. *Nat. Rev. Microbiol.* 10, 472-482.

doi.org/10.1038/nrmicro2802.
Laub, M. T., Chen, S. L., Shapiro, L., and McAdams, H. H. (2002). Genes directly controlled by CtrA, a master regulator of the *Caulobacter* cell cycle. *Proc. Natl. Acad. Sci. U. S. A.* 99, 4632–4637. doi.org/10.1073/pnas.062065699.
Laub, M. T., and Goulian, M. (2007). Specificity in two-component signal transduction pathways. *Ann. Rev.Genet*. 41, 125-45. doi.org/10.1146/annurev.genet.41.042007.170548.
Laub, M. T., McAdams, H. H., Feldblyum, T., Fraser, C. M., and Shapiro, L. (2000). Global analysis of the genetic network controlling a bacterial cell cycle. *Science.* 290, 2144–2148. doi.org/10.1126/science.290.5499.2144.
Leclerc, G., Wang, S. P., and Ely, B. (1998). A new class of *Caulobacter crescentus* flagellar genes. *J. Bacteriol*. 180, 5010-5019. doi.org/10.1128/jb.180.19.5010-5019.1998.
Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., and Smith. D. B., (2018). Virus taxonomy: The database of the international committee on taxonomy of viruses (ICTV). *Nucleic Acids Res*. 46, 708-717. doi.org/10.1093/nar/gkx932.
Leung, M. M., Brimacombe, C. A., Spiegelman, G. B., and Beatty, J. T. (2012). The GtaR protein negatively regulates transcription of the *gtaRI* operon and modulates gene transfer agent (RcGTA) expression in *Rhodobacter capsulatus*. *Mol. Microbiol.* 83, 759–774. doi:10.1111/j.1365-2958.2011.07963.x.
Levet-Paulo, M., Lazzaroni, J. C., Gilbert, C., Atlan, D., Doublet, P., and Vianney, A. (2011). The atypical two-component sensor kinase Lpl0330 from *Legionella pneumophila* controls the bifunctional diguanylate cyclase-phosphodiesterase Lpl0329 to modulate bis-(3'-5')-cyclic dimeric GMP synthesis. *J. Biol. Chem.* 286, 31136–31144. doi:10.1074/jbc.M111.231340.
Li, Yi., Heine, S., Entian, M., Sauer, K., and Dinkel, N. F. (2013). NO-induced biofilm dispersion in *Pseudomonas aeruginos*a is mediated by an MHYT domain-coupled phosphodiesterase. *J. Bacteriol*. 195, 3531-3542. doi.org/10.1128/JB.01156-12.
Lindenberg, S., Klauck, G., Pesavento, C., Klauck, E., and Hengge, R. (2013). The EAL domain protein ycir acts as a trigger enzyme in a c-di-gmp signalling cascade in *E. coli* biofilm control. *EMBO J.* 32, 2001-2014. doi.org/10.1038/emboj.2013.120.
Lobocka, M. B., Rose, D. J., Plunkett, G., Rusin, M., Samojedny, A., Lehnherr, H., Yarmolinsky, M. B., and Blattner, F. R. (2004). Genome of bacteriophage P1. *J. Bacteriol*. 186, 7032-7068. doi.org/10.1128/JB.186.21.7032-7068.2004.
Lori, C., Ozaki, S., Steiner, S., Böhm, R., Abel, S., Dubey, B. N., et al. (2015). Cyclic di-GMP acts as a cell cycle oscillator to drive chromosome replication. *Nature* 523, 236–239. doi:10.1038/nature14473.
Madigan, M. T, and Jung, D. O. (2009). An overview of purple bacteria: Systematics, physiology, and habitats. *Purple Phototrophic Bacteria.* 2, 1–15.
Madigan, M. T. (2006). Microbiology of nitrogen fixation by anoxygenic photosynthetic bacteria. *Anoxygenic Photosynthetic Bacteria*. 2, 915-928. doi.org/10.1007/0-306-47954-0_42.
Marrs, B. (1974). Genetic recombination in *Rhodopseudomonas capsulata*. *Proc. Natl. Acad. Sci. U. S. A.* 71, 971–973. doi:10.1073/pnas.71.3.971.
Mashburn, W., Lauren, M., and Whiteley, M. (2006). Special delivery: Vesicle trafficking in prokaryotes. *Mol. Microbiol.* 61, 839–46. doi.org/10.1111/j.1365-

2958.2006.05272.x.
Mercer, R. G., Callister, S. J., Lipton, M. S., Pasa-Tolic, L., Strnad, H., Paces, V., et al. (2010). Loss of the response regulator CtrA causes pleiotropic effects on gene expression but does not affect growth phase regulation in *Rhodobacter capsulatus*. *J. Bacteriol.* 192, 2701–2710. doi:10.1128/JB.00160-10.
Mercer, R. G., Quinlan, M., Rose, A. R., Noll, S., Beatty, J. T., and Lang, A. S. (2012). Regulatory systems controlling motility and gene transfer agent production and release in *Rhodobacter capsulatus*. *FEMS Microbiol. Lett.* 331, 53–62. doi:10.1111/j.1574-6968.2012.02553.x.
Mercer, R. G., and Lang, A. S. (2014). Identification of a predicted partner-switching system that affects production of the gene transfer agent RcGTA and stationary phase viability in *Rhodobacter capsulatus*. *BMC Microbiol.* 14, 71. doi:10.1186/1471-2180-14-71.
Merkel, T. J., Barros, C., and Stibitz, S. (1998). Characterization of the bvgR locus of *Bordetella pertussis*. *J. Bacteriol*. 180, 1682-1690.
Miller, T. R., and Belas, R. (2006). Motility is involved in *Silicibacter* sp. TM1040 interaction with dinoflagellates. *Environ. Microbiol.* 8, 1648–1659. doi:10.1111/j.1462-2920.2006.01071.x.
Nakamura, Y., Itoh, T., Matsuda, H., and Gojobori, T. (2004). Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat. Genet.* 36, 760–66. doi:10.1038/ng1381
Orr, M. W., Donaldson, G. P., Severin, G. B., Wang, J., Sintim, H. O., Waters, C. M., and Lee, V. T. (2015). Oligoribonuclease is the primary degradative enzyme for pGpG in *Pseudomonas aeruginosa* that is required for cyclic-di-gmp turnover. *Proc. Natl. Acad. Sci. U. S. A.* 112, 5048-5057. doi.org/10.1073/pnas.1507245112.
Paul, R., Abel, S., Wassmann, P., Beck, A., Heerklotz, H., and Jenal, U. (2007). Activation of the diguanylate cyclase PleD by phosphorylation-mediated dimerization.. *J. Biol. Chem*. 282, 29170-29177. doi.org/10.1074/jbc.M704702200.
Pei, J.,, and Grishin, N. V. (2001). GGDEF Domain is homologous to adenylyl cyclase. *Proteins.* 42, 210-216. doi.org/10.1002/1097-0134(20010201)42:2<210::AID-PROT80>3.0.CO;2-8.
Pemberton, J. M., Horne, I. M., and McEwan, A. G. (1998). Regulation of photosynthetic gene expression in purple bacteria. *Microbiol*. 144, 267-278. doi.org/10.1099/00221287-144-2-267.
Perry, J., Koteva, K., and Wright,G. (2011). Receptor domains of two-component signal transduction systems. *Mol. BioSyst*. 7, 1388-1398. doi.org/10.1039/c0mb00329h.
Podgornaia, A. I., Casino, P., Marina, A., and Laub, M. T., (2013). Structural basis of a rationally rewired protein-protein interface critical to bacterial signaling. *Structure*. 21, 1636-1647. doi.org/10.1016/j.str.2013.07.005.
Puck, T. T., and Lee, H. H. (1955). Mechanism of cell wall penetration by viruses. ii. demonstration of cyclic permeability change accompanying virus infection of *Escherichia coli* B cells. *J. Exp. Med*. 101, 155-175. doi:10.1084/jem.101.2.151.
Québatte, M., and Dehio, C. (2019). Bartonella gene transfer agent: Evolution, function, and proposed role in host adaptation. *Cell. Microbiol*. 21, e13068. doi.org/10.1111/cmi.13068.
Quon, K. C., Yang, B., Domian, I. J., Shapiro, L., and Marczynski, G. T.(1998). Negative control of bacterial DNA replication by a cell cycle protein that binds at the

chromosome origin. *Proc. Nat. Acad. Sci. U. S. A.* 95, 20–125. doi:10.1073/pnas.95.1.120

Quon, K. C., Marczynski, G. T., and Shapiro, L. (1996). Cell cycle control by an essential bacterial two-component signal transduction protein. *Cell* 84, 83–93. doi:10.1016/S0092-8674(00)80995-2.

Rakhuba, D. V., Kolomiets, E. I., Dey, E. S., and Novik, G. I. (2010). Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Pol. J. Microbiol*. 59, 145-155.

Rao, F., Yang, Y., Qi, Y., and Liang, Z. X. (2008). Catalytic mechanism of cyclic di-gmp-specific phosphodiesterase: A study of the eal domain-containing RocR from *Pseudomonas aeruginosa*. *J. Bacteriol*. 190, 3622-3631. doi.org/10.1128/JB.00165-08.

Rapp, B. J., and Wall, J. D. (1987). Genetic transfer in *Desulfovibrio desulfuricans*. *Proc. Natl. Acad. Sci. U. S. A.* 84, 9128–9130. doi:10.1073/pnas.84.24.9128.

Ravin, N. V. (2011). N15: The linear phage-plasmid. *Plasmid*. 65, 102-109. doi.org/10.1016/j.plasmid.2010.12.004.

Raz, Y., and Tannenbaum, E. (2010). The influence of horizontal gene transfer on the mean fitness of unicellular populations in static environments. *Genetics* 185, 327–337. doi:10.1534/genetics.109.113613.

Redfield, R. J., Beatty, J. T., and Lang, A. S. (2019). Gene transfer agents. *Encycl Microbiol*. 2, 370-377. doi.org/10.1016/B978-0-12-809633-8.20784.

Reinders, A., Hee, C. S., Ozaki, S., Mazur, A., Boehm, A., Schirmer, T., and Jenal, U. (2016). Expression and genetic activation of cyclic di-gmp-specific phosphodiesterases in *Escherichia coli*. *J. Bacteriol*. 198, 448-462. doi.org/10.1128/jb.00604-15.

Römling, U., Galperin, M. Y., and Gomelsky, M. (2013). Cyclic di-GMP: The first 25 years of a universal bacterial second messenger. *Microbiol. Mol. Biol. Rev.* 77, 1–52. doi:10.1128/MMBR.00043-12.

Rosario, K., and Breitbart, M. (2011). Exploring the viral world through metagenomics. *Curr. Opin. Virol.* 1, 289–297.

Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Ohana, P. W., Mayer, R., Braun, S., et al. (1987). Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature*. 325, 279-281. doi.org/10.1038/325279a0.

Ross, P., Aloni, Y., Weinhouse, H., Michaeli, D., Ohana, P. W., Mayer, R., and Benziman, M. (1986). Control of cellulose synthesis *Acetobacter xylinum*. A unique guanyl oligonucleotide is the immediate activator of the cellulose synthase. *Carb. Res*. 149, 101-117. doi.org/10.1016/S0008-6215(00)90372-0.

Ryan, R. P., Fouhy, Y., Lucey, J. F., Crossman, L. C., Spiro, S., He, Y.-W., et al. (2006). Cell-cell signaling in *Xanthomonas campestris* involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. *Proc. Natl. Acad. Sci. U. S. A.* 103, 6712–6717. doi:10.1073/pnas.0600345103.

Ryan, R. P., Lucey, J., O'Donovan, K., McCarthy, Y., Yang, L., Nielsen, T. T., and Dow, J. M.(2009). HD-GYP domain proteins regulate biofilm formation and virulence in *Pseudomonas aeruginosa*. *Environ. Microbiol*. 11, 1126-1136. doi.org/10.1111/j.1462-2920.2008.01842.x.

Savakis, P., Causmaecker, S. D., Angerer, V., Ruppert, U., Anders, K., Essen, L. O., and Wilde, A. (2012). Light-induced alteration of c-di-gmp level controls motility of

Synechocystis sp. PCC 6803. *Mol. Microbiol*. 85, 239-252. doi.org/10.1111/j.1365-2958.2012.08106.x.

Schaefer, A. L., Taylor, T. A., Beatty, J. T., and Greenberg, E. P. (2002). Long-chain acyl-homoserine lactone quorum-sensing regulation of *Rhodobacter capsulatus* gene transfer agent production. *J. Bacteriol.* 184, 6515–21. doi:10.1128/JB.184.23.6515-6521.2002.

Schirmer, T. (2016). C-di-gmp synthesis: Structural aspects of evolution, catalysis and regulation. *J. Mol. Biol*. 428, 3683-3701. doi.org/10.1016/j.jmb.2016.07.023.

Schirmer, T., and Jenal, U. (2009). Structural and mechanistic determinants of c-di-GMP signalling. *Nat. Rev. Microbiol.* 7, 724–735. doi:10.1038/nrmicro2203.

Schmidt, A. J., Ryjenkov, D. A., and Gomelsky, M. (2005). The ubiquitous protein domain EAL is a c-di-gmp-specific phosphodiesterase: Enzymatically active and inactive EAL domains. *J. Bacteriol.* 187, 4774–4781. doi.org/10.1128/JB.187.14.4774-4781.2005.

Schmidt, A. J., Ryjenkov, D. A., and Gomelsky, M. (2005). The ubiquitous protein domain EAL is a c-di-GMP-specific phosphodiesterase: Enzymatically active and inactive EAL domains. *J. Bacteriol.* 187, 4774–4781. doi:10.1128/JB.187.14.4774-4781.2005.

Seshasayee, A. S.N., Fraser, G. M., and Luscombe, N. M. (2010). Comparative genomics of cyclic-di-gmp signalling in bacteria: Post-translational regulation and catalytic activity. *Nucleic Acids Res.* 38, 5970-5981. doi.org/10.1093/nar/gkq382.

Shiro, Y., and Yamada, S. (2008). Structural basis of the signal transduction in the two-component system. *Adv. Exp. Med. Biol*. 631, 22-39. doi.org/10.1007/978-0-387-78885-2_3.

Silversmith, R. E., and Bourret, R. B. (1999). Throwing the switch in bacterial chemotaxis. *Trends Microbiol*. 7, 16-22. doi.org/10.1016/S0966-842X(98)01409-7.

Simm, R., Morr, M., Kader, A., Nimtz, M., and Römling, U. (2004). GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessibility to motility. *Mol. Microbiol.* 53, 1123–1134. doi:10.1111/j.1365-2958.2004.04206.x.

Solioz, M. (1975). The gene transfer agent of *Rhodopseudomonas capsulata*. PhD Thesis. Saint Louis University.

Solioz, M., Yen, H. C., and Marrs, B. (1975). Release and uptake of gene transfer agent by *Rhodopseudomonas capsulata*. *J. Bacteriol.*123, 651–657.

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

Soucy, S. M., Huang, J., and Gogarten, J. P. (2015). Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* 16, 472–482. doi.org/10.1038/nrg3962.

Stanton, T. B. (2007). Prophage-like gene transfer agents–Novel mechanisms of gene exchange for *Methanococcus*, *Desulfovibrio*, *Brachyspira*, and *Rhodobacter* species. *Anaerobe* 13, 43–49. doi:10.1016/j.anaerobe.2007.03.004.

Stock, A. M., Mottonen, J. M., Stock, J. B., and Schutt, C. E. (1989). Three-dimensional structure of CheY, the response regulator of bacterial chemotaxis. *Nature*. 337, 745-749. doi.org/10.1038/337745a0.

Strnad, H., Lapidus, A., Paces, J., Ulbrich, P., Vlcek, C., Paces, V., et al. (2010). Complete genome sequence of the photosynthetic purple nonsulfur bacterium *Rhodobacter capsulatus* SB 1003. *J. Bacteriol.* 192, 3545–3546.

doi:10.1128/JB.00366-10.
Sultan, S. Z., Pitzer, J. E., Boquoi, T., Hobbs, G., Miller, M. R., and Motaleb, M. A. (2011). Analysis of the HD-GYP domain cyclic dimeric gmp phosphodiesterase reveals a role in motility and the enzootic life cycle of *Borrelia burgdorferi*. *Infect. Immun*. 79, 3273-3283. doi.org/10.1128/iai.05153-11.
Suttle, C. A. (2007). Marine viruses - Major players in the global ecosystem. *Nat. Rev. Microbiol.* 5, 801–812. doi.org/10.1038/nrmicro1750.
Tabita, F. R., and Govindjee. (1995). The biochemistry and metabloic regulation of carbon metabolism and Co2 fixation in purple bacteria. *Anoxygenic Photosynthetic Bacteria*, 2,885–914. doi.org/10.1007/0-306-47954-0_41.
Tal, R., Gelfand, D. H., Calhoon, R. D., Ben, B. A., Benziman, M., Wong, H. C. (1998). Cyclic di-guanylate metabolic enzymes. *US. Patent*.1998.
Tal, R., Wong, H. C., Calhoon, R., Gelfand, D., Fear, A. L., Volman, G., Mayer, R., et al. (1998). Three Cdg operons control cellular turnover of cyclic di-gmp in *Acetobacter xylinum*: Genetic organization and occurrence of conserved domains in isoenzymes. *J. Bacteriol.* 180, 4416–4425.
Tarutina, M., Ryjenkov, D. A., and Gomelsky, M. (2006). An unorthodox bacteriophytochrome from *Rhodobacter sphaeroides* involved in turnover of the second messenger c-di-GMP. *J. Biol. Chem.* 281, 34751–34758. doi:10.1074/jbc.M604819200.
Tatum, E. L., and Lederberg, J. (1947). Gene recombination in the bacterium *Escherichia coli*. *J. Bacteriol.* 53, 673–684.
Tchigvintsev, A., Xu, X., Singer, A., Chang, C., Brown, G., Proudfoot, M., Cui, H., et al. (2010). Structural insight into the mechanism of c-di-gmp hydrolysis by EAL domain phosphodiesterases. *J. Mol. Biol.* 402, 524-538. doi.org/10.1016/j.jmb.2010.07.050.
Tischler, A. D., and Camilli, A. (2004). Cyclic diguanylate (c-di-gmp) regulates *Vibrio cholerae* biofilm formation. *Mol. Microbiol*. 53, 857-869. doi.org/10.1111/j.1365-2958.2004.04155.x.
Tomasch, J., Wang, H., Hall, A. T. K., Patzelt, D., Preusse, M., Petersen, J., et al. (2018). Packaging of *Dinoroseobacter shibae* DNA into gene transfer agent particles is not random. *Genome Biol. Evol.* 10, 359–369. doi:10.1093/gbe/evy005.
Tomomori, C., Tanaka, T., Dutta, R., Park, H., Saha, S. K., Zhu, Y., Ishima, Y. R., et al. (1999). Solution structure of the homodimeric core domain of *Escherichia coli* histidine kinase EnvZ. *Nat. Struct.Biol*. 6, 729-734. doi.org/10.1038/11495.
Trimble, M. J., and McCarter, L. L. (2011). Bis-(3'-5')-cyclic dimeric gmp-linked quorum sensing controls swarming in *Vibrio parahaemolyticus*. *Proc. Natl. Acad. Sci. U. S. A.* 108, 18079–18084. doi.org/10.1073/pnas.1113790108.
Tsokos, C. G., Barrett, S. P., and Laub, M. T. (2011). A dynamic complex of signaling proteins uses polar localization to regulate cell-fate asymmetry in *Caulobacter crescentus*. *Dev. Cell.* 20, 329–341. doi.org/10.1016/j. devcel.2011.01.007.
Tsokos, C. G., and Laub, M. T. (2012). Polarity and cell fate asymmetry in *Caulobacter crescentus*. *Curr. Opin. Microbiol.* 15, 744-50. doi:10.1016/j.mib.2012.10.011.
Wall, J. D., Weaver, P. F., and Gest, H. (1975). Gene transfer agents, bacteriophages, and bacteriocins of *Rhodopseudomonas capsulata*. *Arch. Microbiol.* 105, 217–224. doi:10.1007/BF00447140.
Wall, J. D., Weaver, P. F., and Gest, H. (1975). Gene transfer agents, bacteriophages, and

bacteriocins of *Rhodopseudomonas capsulata*. *Arch. Microbiol*. 105, 217-224. doi.org/10.1007/BF00447140.

Wassmann, P., Chan, C., Paul, R., Beck, A., Heerklotz, H., Jenal, U., and Schirmer, T. (2007). Structure of BeF3-modified response regulator PleD: Implications for diguanylate cyclase activation, catalysis, and feedback inhibition. *Structure*. 15, 915-927. doi.org/10.1016/j.str.2007.06.016.

West, A. H., and Stock, A. M. (2001). Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem. Sci.* 26, 369-376. doi.org/10.1016/S0968-0004(01)01852-7.

Westbye, A. B., Leung, M. M., Florizone, S. M., Taylor, T. A., Johnson, J. A., Fogg, P. C., et al. (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. doi:10.1128/jb.00669-13.

Wilhelm, S. W., and Suttle, C. A. (1999). Viruses and nutrient cycles in the sea. *Bioscience.* 49, 781-788. doi.org/10.2307/1313569.

Winter, C., Smit, A., Herndl, G. J., and Weinbauer, M. G. (2004). Impact of virioplankton on archaeal and bacterial community richness as assessed in seawater batch cultures. *Appl.Environ.Microbiol*. 70, 804-813. doi.org/10.1128/AEM.70.2.804-813.2004.

Yamada, S., Sugimoto, H., Kobayashi, M., Ohno, A., Nakamura, H., and Shiro, Y. (2009). Structure of PAS-linked histidine kinase and the response regulator complex. *Structure*. 17, 1333-1344. doi.org/10.1016/j.str.2009.07.016.

Yen, H. C., Hu, N. T., and Marrs, B. L. (1979). Characterization of the gene transfer agent made by an overproducer mutant of *Rhodopseudomonas capsulata*. *J. Mol. Biol.* 131, 157–168. doi:10.1016/0022-2836(79)90071-8.

Young, R. (1992). Bacteriophage lysis: Mechanism and regulation. *Microbiol. Rev*. 56, 430-481.

Zähringer, F., Lacanna, E., Jenal, U., Schirmer, T., and Boehm, A. (2013). Structure and signaling mechanism of a zinc-sensory diguanylate cyclase. *Structure*. 21, 1149-1157. doi.org/10.1016/j.str.2013.04.026.

Zhaxybayeva, O., Gogarten, J. P., Charlebois, R. L., Doolittle, W. F., and Papke, R. T. (2006). Phylogenetic analyses of cyanobacterial genomes: Quantification of horizontal gene transfer events. *Genome Res.* 16, 1099–1108. doi:10.1101/gr.5322306

Zinder, N. D., and Lederberg, J. (1952). Genetic exchange in *Salmonella*. *J. Bacteriol.* 64, 679–699.

Co-authorship statement

Chapter 2 is a version of a manuscript published in the Journal of Bacteriology- (Pallegar Purvikalyan, Lourdes Peña-Castillo, Evan Langille, Mark Gomelsky, and Andrew S Lang. 2020. "Cyclic Di-GMP-Mediated Regulation of Gene Transfer and Motility in *Rhodobacter capsulatus*." *Journal of Bacteriology*. https://doi.org/10.1128/JB.00554-19). The basic concept of the project was designed by A.S. Lang. I prepared strains and performed all the experiments. Some of the initial mutant strains used in this study were made by C. Buckley. The HPLC analysis was performed by E. Langille. The manuscript was drafted and prepared by me and A.S. Lang with subsequent editorial input from the other co-authors.

Chapter 3 is a version of a manuscript published in in Journal of Molecular Biology– (Pallegar Purvikalyan, Marta Canuti, Lourdes Peña-Castillo, Evan Langille, and Andrew S Lang. 2020. "A two-component system acquired by horizontal gene transfer modulates gene transfer and motility via cyclic dimeric GMP". *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2020.07.001). Research in this chapter was proposed and designed by me and A.S. Lang. I prepared all the strains and carried out all the experiments. All the bioinformatic work including phylogenetic analysis was performed by M. Canuti. The HPLC analysis was performed by E. Langille. The manuscript was drafted and prepared by me, M. Canuti and A.S. Lang with subsequent editorial input from the other co-authors.

Chapter 4 is part of a manuscript in preparation for future submission. Research in this chapter was proposed and designed by me and A.S. Lang. I carried out all the

bioinformatic work, preparation of site directed mutant strains and c-di-GMP binding experiments. The other part of the future manuscript is prepared by the collaborators from the laboratory of Dr. J. T. Beatty, University of British Columbia, Vancouver.

CHAPTER 2- Cyclic-di-GMP-mediated regulation of gene transfer and motility in *Rhodobacter capsulatus*

2.1 Introduction

Gene transfer between cells plays an important role in bacterial evolution, with horizontal gene transfer (HGT) being the main force behind the acquisition of new, adaptive traits and genetic variation among bacterial strains (Raz and Tannenbaum 2010). In addition to the three canonical mechanisms by which bacterial DNA exchange occurs, i.e. transformation, conjugation and transduction, a different type of genetic exchange process is mediated through bacteriophage-like particles called gene transfer agents (GTAs). This gene transfer mechanism resembles the process of transduction, but GTAs are distinct from transducing bacteriophages (Lang and Beatty 2007; Stanton 2007). Similar to prophages, GTAs are encoded by genes within the producing organisms' genomes. However, GTAs are distinct from induced transducing prophages because all GTA particles contain only DNA from the cells' genomes. They also package less DNA than required to encode the particles, making them incapable of self-transmission (Stanton 2007; Lang and Beatty 2007).

Gene transfer agents are known to be produced by multiple bacteria and one archaeon (Lang, Westbye, and Beatty 2017; Tomasch et al. 2018). The first-discovered GTA (now known as RcGTA) is produced by *Rhodobacter capsulatus* (Marrs 1974), a purple non-sulfur α-proteobacterium that has been used as a model organism for various aspects of physiology such as anoxygenic photosynthesis (Strnad et al. 2010). Each RcGTA particle packages approximately 4 kb of double-stranded DNA (Yen, Hu, and

Marrs 1979) while the main gene cluster encoding the particles spans approximately 14 kb (Lang and Beatty 2000). Additional genes required for RcGTA production, function and release are located at distinct locations in the genome (Hynes et al. 2012; 2016; Westbye et al. 2016). Expression of the RcGTA genes is regulated by several cellular signaling systems as well as phage-related regulators (Fogg 2019; Lang, Westbye, and Beatty 2017). The cellular regulators include the CckA-ChpT-CtrA phosphorelay (Lang and Beatty 2000; Mercer et al. 2012), the GtaI-GtaR quorum sensing system (Schaefer et al. 2002; Leung et al. 2012), the Rba partner-switching phosphorelay (Mercer and Lang 2014), the SOS regulator LexA (Kuchinski et al. 2016), and the PAS domain protein DivL (Westbye et al. 2018).

The CtrA response regulator protein was first characterized in *Caulobacter crescentus* (Quon, Marczynski, and Shapiro 1996), where it acts as a master regulator of the cell cycle (Skerker and Laub 2004). Among all cellular RcGTA regulators identified to date, only loss of CtrA causes a complete loss of RcGTA production, which is caused by loss of transcription of most genes in the RcGTA gene cluster (Lang and Beatty 2000; Mercer et al. 2010). Loss of a phage-derived regulator (Hynes et al. 2016), which has been renamed *gafA* (Fogg 2019), also causes a complete loss of GTA production and this gene is also regulated by CtrA. Transcriptomic studies in *R. capsulatus* revealed that more than 225 genes are dysregulated in the absence of CtrA (Mercer et al. 2010), including more than 20 genes predicted to encode proteins involved in signal transduction or regulation of gene expression. This included proteins predicted to be involved in signaling via the second messenger bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP) based on the presence of conserved domains for c-di-GMP synthesis or degradation.

Cyclic di-GMP is an ubiquitous second messenger that controls various aspects of bacterial physiology (Romling, Galperin, and Gomelsky 2013; Jenal, Reinders, and Lori 2017). Cyclic di-GMP binds to a range of targets, including riboswitches and proteins, and affects diverse processes including motility, biofilm formation, virulence, and cell cycle progression. Inhibition of motility and promoting a sessile lifestyle and biofilm formation are the most widely conserved behaviors in bacteria in response to elevated levels of c-di-GMP. Two GTP molecules are used for the synthesis of c-di-GMP, catalyzed by diguanylate cyclase (DGC) enzymes that contain GGDEF motifs in their active sites (A-sites) (Paul et al. 2004; Chan et al. 2004; Ryjenkov et al. 2005). In addition to an A-site, many DGCs also carry an inhibitory site (I-site) motif, RxxD, which is involved in feedback inhibition (Christen et al. 2006; Schirmer and Jenal 2009). Cyclic di-GMP-specific phosphodiesterases (PDEs), characterized by EAL (Schmidt, Ryjenkov, and Gomelsky 2005; Christen et al. 2005; Tamayo, Tischler, and Camilli 2005) and HD-GYP (Ryan et al. 2006) domains, break down c-di-GMP into 5'-phosphoguanylyl-(3'-5')-guanosine (pGpG). Some proteins contain both GGDEF and EAL domains and can be bifunctional (Tarutina, Ryjenkov, and Gomelsky 2006; Levet-Paulo et al. 2011). It is also possible that only one domain is enzymatically active in such dual-domain proteins and enzymatically inactive domains can often bind former substrates, c-di-GMP (EAL) (Qi et al. 2011) or GTP (GGDEF) (Christen et al. 2005), and serve as regulatory sites (Wolfe and Visick 2010). The GGDEF and EAL domains are often present within proteins that contain additional periplasmic, membrane-embedded or cytoplasmic ligand-binding/signaling domains. These include the response regulator receiver (REC) domain and ligand-binding domains such as Per-ARNT-Sim (PAS) and cGMP-specific phosphodiesterases/adenylyl cyclases/FhlA (GAF) (Wolfe and Visick 2010).

Table 2.1. Properties of eight chromosomal c-di-GMP signaling genes whose transcript levels are affected by loss of CtrA.

Gene	Transcript fold-change in *ctrA* null mutant[a]	Protein accession number	Size of protein (aa)	C-di-GMP domains	Additional domains[b]
rcc00346	-7.6	ADE84111.1	514	GGDEF, EAL	
rcc00620	-14.0	ADE84385.1	610	GGDEF, EAL	REC
rcc00645	-7.7	ADE84410.1	1245	GGDEF, EAL	PAS
rcc02539	-8.1	ADE86269.1	641	GGDEF, EAL	
rcc02629	-8.1	ADE86359.1	353	GGDEF	
rcc02857	-12.5	ADE86586.1	1158	GGDEF, EAL	PAS
rcc03177	-19.5	ADE86901.1	280	EAL	
rcc03301	-4.5	ADE87025.1	1284	GGDEF, EAL	PAS

[a] From (Mercer et al. 2010)

[b] REC, response regulator receiver; PAS, Per-ARNT-Sim

The *R. capsulatus* genome (Strnad et al. 2010) carries 20 genes predicted to encode proteins containing GGDEF or EAL domains and the transcript levels of nine of these genes were significantly decreased in a *ctrA* null mutant (Mercer et al. 2010). Based on this observation we hypothesized that c-di-GMP signaling might affect the production of RcGTA. We have investigated the possible roles of the eight chromosomally encoded putative c-di-GMP signaling proteins from this group in *R. capsulatus* gene exchange (Table 2.1). We evaluated the potential enzymatic activities of four of these proteins that were implicated in RcGTA production via phenotypic assays in *Escherichia coli*. We also investigated the effects of changes in cellular levels of c-di-GMP on *R. capsulatus* gene

exchange. In addition, we investigated the roles of these four genes and c-di-GMP in *R. capsulatus* flagellar motility and conclude that elevated c-di-GMP levels inhibit RcGTA production and flagellar motility in this bacterium.

2.2 Materials and methods

2.2.1 Bacterial strains, media and culture conditions

All the experimental strains and plasmids used in this study are listed in Table 2.2 *R. capsulatus* was grown either anaerobically under photoheterotrophic conditions in complex YPS medium (Wall, Weaver, and Gest 1975) or aerobically in defined RCV medium (Beatty and Gest 1981) at 35°C. Appropriate antibiotics were used when required at the following concentrations: kanamycin (10 μg ml^{-1}), gentamycin (3 μg ml^{-1}), spectinomycin (10 μg ml^{-1}) and tetracycline (0.5 μg ml^{-1}). *E. coli* was grown at 37°C in LB medium supplemented with appropriate antibiotics when necessary: ampicillin (100 μg ml^{-1}), kanamycin (25 μg ml^{-1}), gentamycin (10 μg ml^{-1}), spectinomycin (50 μg ml^{-1}) and tetracycline (10 μg ml^{-1}).

Table 2.2. List of bacterial strains and plasmids used in this study.

Strains and plasmids	Description	Reference or source
***R. capsulatus* strains**		
SB1003	Genome-sequenced strain	(Strnad et al. 2010; Yen and Marrs 1976)
DW5	SB1003 Δ*puhA*	(Wong et al. 1996)

SB346	SB1003 with KIXX insertion in *rcc00346*	This study
SB620a	SB1003 with KIXX insertion in *rcc00620*	This study
SB620	SB1003 with 1068-bp deletion in *rcc00620* replaced by KIXX	This study
SB620.645	SB620 with 2469-bp deletion in *rcc00645* replaced by gentamycin resistance gene	This study
SB645	SB1003 with 2469-bp deletion in *rcc00645* replaced by KIXX	This study
SB645.2629.2857	SB2629 with 2469-bp deletion in *rcc00645* replaced by gentamycin resistance gene and 909-bp deletion in *rcc02857* replaced by spectinomycin resistance gene	This study
SB2539	SB1003 with KIXX insertion in *rcc02539*	This study
SB2629a	SB1003 with KIXX insertion in *rcc02629*	This study
SB2629	SB1003 with 541-bp deletion in *rcc02629* replaced by KIXX	This study

SB2857a	SB1003 with KIXX insertion in *rcc02857*	This study
SB2857	SB1003 with 909-bp deletion in *rcc02857* replaced by KIXX	This study
SB3177	SB1003 with KIXX insertion in *rcc03177*	This study
SB3301	SB1003 with KIXX insertion in *rcc03301*	This study
***E. coli* strains**		
MG1655	Wild type; motility indicator strain for DGC activity	(Blattner 1997)
MG1655 Δ*yhjH*	Motility indicator strain for PDE activity	(Ryjenkov et al. 2006; Simm et al. 2004)
BL21(DE3)	Curli fimbriae indicator strain for DGC activity	New England Biolabs; (Christen et al. 2006)
C600(pDPT51)	Plasmid-mobilizing strain	(Taylor et al. 1983)
S17-1	Plasmid-mobilizing strain	(Simon, Priefer, and Pühler 1983)
Plasmids		
pGEM-T Easy	TA PCR product cloning vector	Promega

pCM62	Broad host range vector; expression of genes in *E. coli* driven by *lac* promoter	(Marx and Lidstrom 2001)
pRR5C	Expression of genes in *R. capsulatus* driven by *puf* promoter	(Young, Reyes, and Beatty 1998)
p620	*rcc00620* and 440 bp of 5' sequence in *Kpn*I site of pCM62	This study
p620GGAAF	p620 with mutation in DGC domain	This study
p620AAL	p620 with mutation in PDE domain	This study
p620GGAAF/AAL	p620 with mutations in both DGC and PDE domains	This study
p645	*rcc00645* and 467 bp of 5' sequence in *Kpn*I site of pCM62	This study
p645GGAAF	p645 with mutation in DGC domain	This study
p645AVL	p645 with mutation in PDE domain	This study
p645GGAAF/AVL	p645 with mutations in both DGC and PDE domains	This study
p2629	*rcc02629* and 771 bp of 5' sequence in *Kpn*I site of pCM62	This study
p2629GGAAF	p2629 with mutation in DGC domain	This study
p2857	*rcc02857* and 99 bp of 5' sequence in *Kpn*I site of pCM62	This study
p2857GGAAF	p2857 with mutation in DGC domain	This study
p2857ATL	p2857 with mutation in PDE domain	This study

p2857GGAAF/AAL	p2857 with mutations in both DGC and PDE domains	This study
pDGC	Heterologous diguanylate cyclase gene from *R. sphaeroides* (RSP_3513) cloned into pCM62	This study
pRRDGC	Heterologous diguanylate cyclase gene from *R. sphaeroides* (RSP_3513) cloned into pRR5C	This study
pPDE	Heterologous phosphodiesterase gene from *G. xylinus* (*pdeA1*) cloned into pCM62	This study
pRRPDE	Heterologous phosphodiesterase gene from *G. xylinus* (*pdeA1*) cloned into pRR5C	This study
pX3	RcGTA *orfg3*'::'*lacZ* fusion	(Hynes et al. 2012)
pX3NP	RcGTA *orfg3*'::'*lacZ* fusion with no promoter	(Hynes et al. 2012)

2.2.2 Insertional mutagenesis, *trans*-complementation plasmids, and site-directed mutants

PCR amplifications were done using genomic DNA from *R. capsulatus* SB1003 as the template and the appropriate primers for each gene (Appendix 1, Table S2.1). The amplified products were cloned into pGEM-T Easy (Promega) according to the

manufacturer's guidelines. Gene disruptions were made by the insertion of the approximately 1.4-kb *Sma*I fragment of the kanamycin resistance-encoding KIXX fragment (Barany 1985) at specific restriction enzyme cut sites within the cloned PCR products, as detailed below.

The *rcc00346* open reading frame (ORF) was disrupted at the *BsaB*I site 319 bp from the start of the 1545-bp ORF; *rcc00620* at the *EcoR*I site 793 bp from the start of the 1833-bp ORF; *rcc00645* at the *Cla*I site 83 bp from the start of 3738-bp ORF (there is also a second *Cla*I site in this ORF, which results in a 2469-bp deletion); *rcc02539* at the *Stu*I site 469 bp from the start of the 1926-bp ORF; *rcc02540* at the *Hind*III site 300 bp from the start of the 2757-bp ORF; *rcc02629* at the *Msc*I site 517 bp from the start of the 1062-bp ORF; *rcc02857* at the *BamH*I site 422 bp from the start of the 3477-bp ORF; *rcc03177* at the *BstE*II site 102 bp from the start of the 843-bp ORF; and *rcc03301* at the *Stu*I site 526 bp from the start of the 3855-bp ORF. Gene disruptions were confirmed by restriction enzyme digestions and conjugated to *R. capsulatus* from *E. coli* C600 (pDPT51) (Taylor et al. 1983). RcGTA transfer of the disrupted genes into the chromosome of recipient SB1003 cells was then performed to generate *R. capsulatus* mutant strains (Hynes and Lang 2013). The resulting kanamycin-resistant strains were confirmed to contain only the disrupted versions of the genes by PCR using the original amplification primers.

Deletion mutants for the genes *rcc00620, rcc02629* and *rcc02857* were subsequently made by replacing portions of the ORFs with the KIXX fragment. For *rcc00620,* 1068 bp was deleted between two *Ava*I sites (deletion starts 144 bp into the ORF). For *rcc02629,* 541 bp was deleted between two *BamH*I sites (deletion starts 8 bp

into the ORF). For *rcc02857*, 909 bp was deleted between two *BsaB*I sites (deletion starts 2296 bp into the ORF). Chromosomal disruptions and subsequent confirmations were done as described above. For *rcc00645*, 2469 bp was already deleted between two *Cla*I sites when the original disruption was made. In addition to these individual mutants with KIXX insertions, additional mutants were made for *rcc00645* and *rcc002857* using different antibiotic resistance genes. A gentamycin resistance gene was inserted at the *ClaI* deletion site for *rcc00645* and a spectinomycin resistance gene was inserted into *rcc002857* at the *BsaB*I deletion site. These constructs were used to create double (*rcc00620* and *rcc00645*) and triple (*rcc00645*, *rcc02629* and *rcc02857*) mutants by RcGTA transfer into the chromosome of the appropriate recipient mutant strains, as described above.

Trans-complementation constructs were made using the plasmid pCM62 (Marx and Lidstrom 2001) as a vector. The structural genes and their upstream regulatory regions were amplified using gene-specific complementation primers (Appendix 1, Table S2.1). The amplified fragments were cloned into pCM62 as *Kpn*I fragments. The four knockout strains containing the empty plasmid were subsequently used as the reference strains.

To alter the c-di-GMP levels in *R. capsulatus* cells, plasmids carrying genes from other bacteria with known DGC (*dgcA* from *Rhodobacter sphaeroides*) (Ryjenkov et al. 2005) and PDE (*pdeA1* from *Gluconacetobacter xylinus*) (Chang et al. 2001) activities were constructed. The *dgcA* and *pdeA1* genes were amplified using gene-specific primers (Appendix 1, Table S2.1) and cloned into pRR5C (Young, Reyes, and Beatty 1998), which leads to transcription in *R. capsulatus* under the control of the *puf* promoter. These

genes were also cloned into pCM62 (Marx and Lidstrom 2001) to allow transcription in *E. coli* from the *lac* promoter.

To create point mutations in the predicted DGC and PDE domains, site-directed mutagenesis was performed using the QuickChange Lightning Site-Directed Mutagenesis kit (Agilent Technologies) as per the manufacturer's instructions. Briefly, the pCM62-based clones described above were used as templates for PCR with *PfuUltra* High-Fidelity DNA polymerase and gene-specific primers (Appendix 1,Table S2.1) designed to change the critical residues of the GGDEF (GGDEF to GGAAF) and EAL (EAL, EVL and ETL to AAL, AVL and ATL, respectively) domains. These substitutions were chosen as they were previously shown to disrupt the function of these domains (Newell et al. 2011; Kuchma et al. 2007). The methylated template DNA was then digested by incubation with *Dpn*I for 10 min at 37°C and the remaining DNA was transformed into *E. coli*. Mutations were confirmed by sequencing. These pCM62 constructs allow the genes to be transcribed from their native upstream sequences in *R. capsulatus* and from the plasmid's *lac* promoter in *E. coli*.

Plasmids were conjugated into *R. capsulatus* using *E. coli* S-17 (Simon, Priefer, and Pühler 1983).

2.2.3 Gene transfer bioassays

RcGTA-mediated gene transfer activity was measured as described (Hynes and Lang 2013), with quantification of the transfer of an essential photosynthesis gene, *puhA*, to a Δ*puhA* mutant strain, DW5 (Wong et al. 1996). Aerobically grown overnight cultures of test strains were normalized for density and used to inoculate anaerobic phototrophic cultures. These cultures were grown for approximately 48 hours, filtered using 0.45-µm

PVDF syringe filters and filtrates were assayed for RcGTA gene transfer activity. RcGTA activities for the mutant strains were measured as ratios relative to the parental strain SB1003 in at least three replicate experiments. For assaying the RcGTA production under aerobic conditions, the GTA donor cultures (SB1003 and SB645) were grown aerobically (shaken at 220 RPM) for approximately 48 hours at 35°C and then assayed for RcGTA gene transfer activity as described above. One-way analysis of variance (ANOVA) followed by Tukey HSD post-hoc analysis in R (Hesterberg, Chambers, and Hastie 1993) was used to identify statistically significant differences in RcGTA activities.

2.2.4 Quantification of c-di-GMP

The quantification of c-di-GMP levels in cells was done using a protocol adapted from Roy, Petrova, and Sauer (2016). Aerobically grown overnight cultures of the different *R. capsulatus* strains were normalized for cell density and used to inoculate anaerobic photoheterotrophic cultures. These cultures were grown for approximately 48 hours and 3 ml of each culture was removed and the cells pelleted by centrifugation at 4°C. The supernatant was discarded, and the pellet was washed twice with 1 ml of ice-cold phosphate-buffered saline (PBS). The cell pellet was then resuspended in 100 µl ice-cold PBS and incubated at 100°C for 5 minutes. Ice-cold 100% ethanol (186 µl, final concentration 65%) was added and the mixture was vortexed for 15 seconds followed by centrifugation at 4°C. The supernatant containing the extracted c-di-GMP was then collected and transferred to a new microfuge tube. This extraction procedure was repeated twice more for each cell pellet and the supernatants were pooled into a single tube. The pellets were saved for protein quantification. The combined supernatants were dried in a

centrifugal evaporator at 30°C for 4-5 hours. The resulting dried extracts were resuspended in 100 μl ultrapure water, briefly vortexed and then centrifuged at 16000 x *g* for 5 minutes. The supernatants were then transferred to 250-μl glass micro-inserts which were placed into high-performance liquid chromatography (HPLC) vials for analysis. C-di-GMP was detected using a Hewlett Packard 1050 HPLC system consisting of an autosampler, a quaternary pump and a multiple wavelength detector (Agilent Technologies). HPLC was performed with mobile phase parts A (10 mM ammonium acetate in H_2O) and B (10 mM ammonium acetate in methanol), with elution using a gradient of 5% B for the first 6 minutes, to 15% B at 11 minutes, 25% B at 25 minutes and 90% B at 17 minutes and a flow rate of 0.5 ml $minute^{-1}$. The backpressure of the system was 100 ± 5 bar. The injection volume was 20 μl and detection was at 253 nm. The runtime was 19 minutes with a post time of 11 minutes and the retention time of c-di-GMP was 13.0 ± 0.1 minutes. The "Chemstation" software (Agilent Technologies) was used to control the instrument and collect the data. Analytical separations were performed using a Luna 3μm C_{18} 100 x 4.6 mm column (Phenomonex). A standard curve was made using solutions of c-di-GMP (BIOLOG) in H_2O (80, 40, 20, 10, 5, 2.5, 1.25, and 0 μg l^{-1}). Standards were measured from triplicate injections. Culture samples were prepared in triplicate and each replicate was quantified from triplicate injections.

For protein quantification, the cell pellets from the c-di-GMP extractions were resuspended by adding 500 μl TE buffer and sonicated for a total of 2 minutes using 20-second bursts on ice. The protein concentration was determined using the Bradford assay (He 2011). Briefly, 60 μl of 10-fold diluted sample in TE buffer was added to a polystyrene cuvette containing 3 ml of Bradford reagent [0.005% (w/v) Coomassie

Brilliant Blue G-250, 8.5% (w/v) H_3PO_4]. The mixture was incubated for 5 minutes and the absorbance at 595 nm was measured. A standard curve was constructed using bovine serum albumin solution standards (25, 50, 100, 200, 500 and 1000 μg ml^{-1}). The concentration of c-di-GMP was then normalized to protein levels in the cells. One-way ANOVA followed by Tukey HSD post-hoc analysis in R was used to identify statistically significant differences in c-di-GMP levels.

2.2.5 Western blotting

Western blotting to detect the RcGTA major capsid protein (approximately 32 kDa) was carried out to quantify RcGTA protein levels within cells and released into the extracellular environment. This was performed on the same cultures that were used for gene transfer assays, as described (Mercer et al. 2012). Briefly, 0.5 ml of each culture was centrifuged at >14000 x *g* and a 0.2-ml sample was collected from the supernatant. The remaining supernatant was carefully removed from the cell pellet and 0.5 ml of TE buffer was added to resuspend the cells. Samples for SDS-PAGE were prepared by mixing 5 μl of cells and 10 μl of supernatant with SDS-PAGE sample loading buffer (NEB) and heating at 98°C for 5 minutes. Ten percent SDS-PAGE gels were used to separate the proteins followed by transfer onto nitrocellulose membranes by electroblotting in transfer buffer [48 mM Tris base, 39 mM glycine, 20% Methanol (v/v)]. The membranes were blocked with 5% (w/v) skim milk solution in TBST [20 mM Tris, 137 mM NaCl, 0.1% Tween-20 (v/v); pH 7.5] and incubated with the primary antibody, anti-*Rhodobacterales* GTA major capsid protein (Agrisera) (Fu et al. 2010), overnight at 4°C. After washing with TBST, membranes were incubated with secondary antibody, peroxidase-conjugated

anti-rabbit IgG (Santa Cruz Biotechnology), at room temperature for 1 hour. The SuperSignal West Femto Reagent Kit (Thermo Fisher Scientific) was used to detect the bands by chemiluminescence and images were captured using an Agilent ImageQuant LAS 4000 imaging system. Images were inverted and adjusted for brightness and contrast, and band intensities were quantified using ImageJ (Schneider, Rasband, and Eliceiri 2012).

2.2.6 β-galactosidase assays

A plasmid carrying an in-frame fusion of *orfg3* of the RcGTA structural gene cluster, along with upstream sequences including the cluster's promoter region, to *lacZ* (Hynes et al. 2012) was used to quantify RcGTA gene expression. A version lacking the promoter region was used as a negative control. The *R. capsulatus* strains were grown under the same conditions and for the same time as for RcGTA gene transfer activity assays and assayed for β-galactosidase activity as described (Miller 1992). Briefly, the cell density of each culture was measured via absorbance at 600 nm and 0.1 ml of each was centrifuged and resuspended in 1 ml of Z buffer [60 mM $Na_2HPO_4{\cdot}7H_2O$, 40 mM NaH_2PO_4, 1 mM $MgSO_4{\cdot}7H_2O$, 10 mM KCl, 50 mM β-mercaptoethanol; pH 7]. The cells were permeabilized by adding two drops of chloroform and one drop of 0.1% SDS followed by incubation at 28°C for 5 minutes. Ortho-nitrophenyl-β-galactoside was added (0.67 mg/ml final concentration) and the reactions were incubated at room temperature until visible yellow color developed and then stopped by addition of 1 M Na_2CO_3. Cell material was pelleted by centrifugation and the supernatants measured for absorbance at 420 nm. β-galactosidase activities were calculated in Miller units (Miller 1992).

2.2.7 Bioinformatic analyses

Protein sequence analyses, for identification of functional domains, were done using the SMART (Letunic 2004; Schultz et al. 1998) and Expasy-Prosite (de Castro et al. 2006) databases. Sequence alignments were done using Clustal Omega (Larkin et al. 2007).

2.2.8 *E. coli* motility assays

Swimming assays in *E. coli* MG1655 and *E.coli* MG1655 Δ*yhjH*, to detect DGC and PDE activities, respectively, were performed as described previously (Chen et al. 2014). The pCM62-based clones of *rcc00620, rcc00645*, *rcc02629,* and *rcc02857* genes, their respective site-directed mutants, and controls (*dgcA* from *R. sphaeroides* and *pdeA1* from *G. xylinus*) were transformed into the *E. coli* strains. Five µl of overnight cultures were inoculated onto semi-solid (0.25% agar) LB plates containing 0.25 mM IPTG and 0.5% NaCl and the plates were incubated at 37°C for 4-6 hours and photographed. The images were manipulated for brightness and contrast to help improve the visibility of the bacterial growth zones.

2.2.9 In-cell DGC activity assays

The same pCM62-based plasmids used for motility assays were also transformed into *E. coli* BL21(DE3) to perform Congo red binding assays as described previously (Chen et al. 2014). Briefly, 3 µl from overnight cultures for all the strains were streaked on LB agar containing 25 µg ml^{-1} Congo red and 0.1 mM IPTG and the plates were incubated at 28°C for 48 hours and photographed.

2.2.10 *R. capsulatus* motility assays

Aerobically grown overnight cultures were used to inoculate YPS agar (0.35%) stabs in test tubes that were incubated at 35°C under phototrophic conditions for 16-24 hrs. The tubes were photographed and the diameters of the zones of growth were measured using ImageJ. The stab motility assays were performed in three independent growth experiments. One-way ANOVA followed by Tukey HSD post-hoc analysis in R was used to identify statistically significant differences in the measured stab swim zones.

2.3 Results

2.3.1 Disruptions of four genes encoding predicted c-di-GMP signaling proteins affect RcGTA production

Insertional disruptions were made in the eight chromosomal genes predicted to encode c-di-GMP signaling proteins whose mRNA levels were affected by loss of CtrA (Table 1), which is a key regulator required for RcGTA gene expression. The strains with disruptions in four genes, *rcc00620, rcc00645, rcc02629* and *rcc02857*, showed appreciable differences in RcGTA production relative to the parental strain, whereas the other four gene disruptions did not (Appendix 1, Figure S2.1). Disruption of *rcc00620* decreased RcGTA production, whereas disruptions in *rcc00645, rcc02629* and *rcc02857* increased it.

Evaluation of the protein sequences for the four genes showed that Rcc00620, Rcc00645, and Rcc02857 contained both GGDEF and EAL domains, while Rcc02629 contained only a GGDEF domain (Figure 2.1A). Amino acid sequence analysis revealed that the GGDEF domains of all four proteins have all conserved residues required for

DGC activity (Figure 2.1B) (Romling, Galperin, and Gomelsky 2013). Similarly, the EAL domains of Rcc00620, Rcc00645 and Rcc02857 contain all conserved residues required for PDE activity (Figure 2.1C) (Romling, Galperin, and Gomelsky 2013). Therefore, Rcc02629 may possess DGC activity, while the three remaining proteins, Rcc00620, Rcc00645, and Rcc02857, may possess either DGC or PDE activity, or both activities.

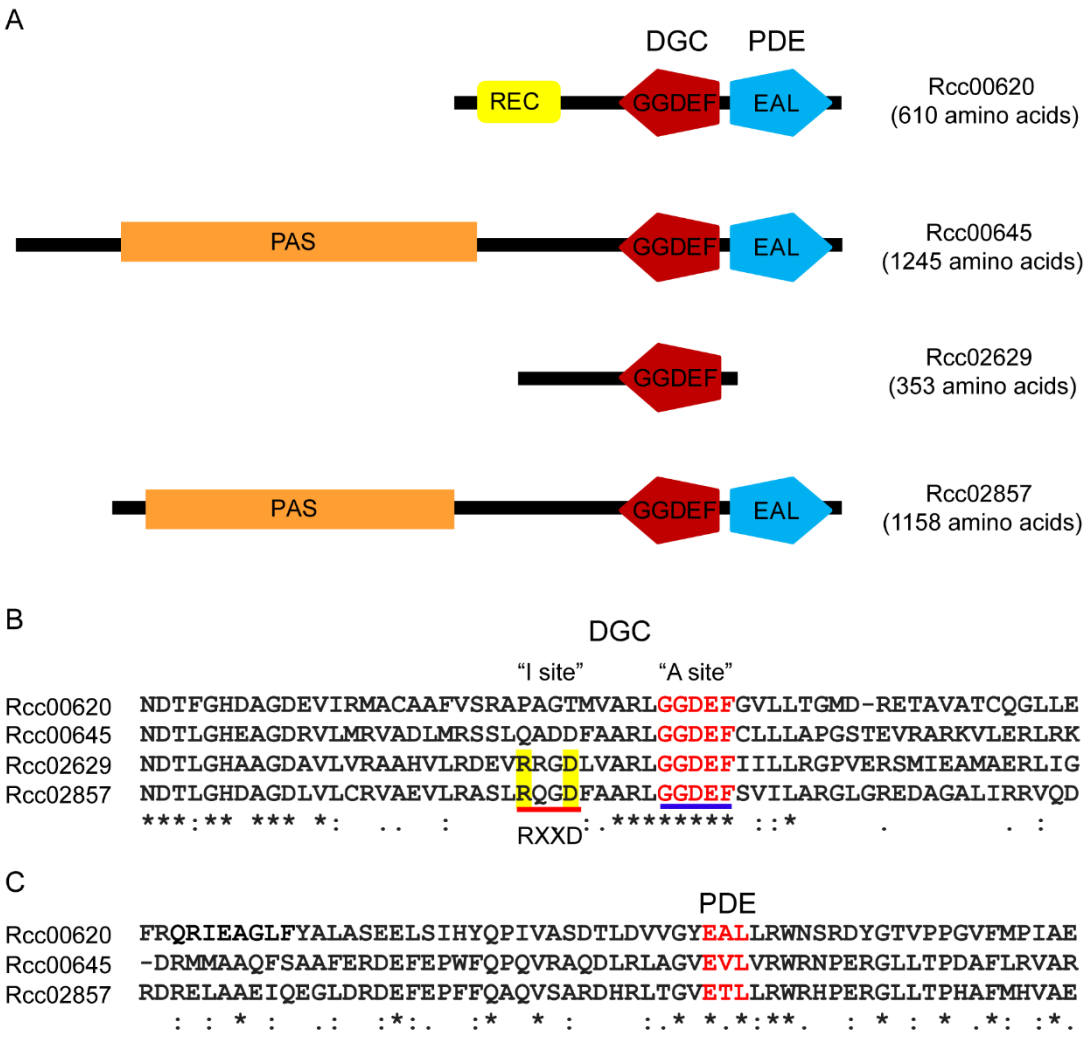


Figure 2.1. Predicted domains of four putative c-di-GMP signaling proteins that affect RcGTA gene transfer activity. A. Locations and organizations of predicted domains

identified in the four proteins. The domains are: REC, response regulator receiver; DGC, GGDEF (diguanylate cyclase); PDE, EAL (phosphodiesterase); PAS, Per-ARNT-Sim. B. Amino acid sequence alignments for the four proteins indicating A (active site: GGDEF, indicated by blue line) and I (inhibitory site: RXXD, indicated by red line) sites for the DGC domains. C. Amino acid sequence alignments for three of the proteins indicating the EAL motifs, which represent the PDE domains.

To eliminate the possibility that truncated proteins encoded by the disrupted versions of the *rcc00620, rcc02629* and *rcc02857* genes affected the results, we constructed deletion-insertion (as opposed to insertion only) mutations, where a large portion of each ORF was replaced with the KIXX fragment. The newly generated deletion-insertion mutants in *rcc00620, rcc02629* and *rcc02857* showed essentially the same phenotypes as the original insertion knockouts (Figure 2.2A). The original *rcc00645* disruption already featured a large deletion of the coding region, therefore obviating the need for a new mutant construction. *Trans*-complementation of all four mutants restored the RcGTA production to the parental strain levels (Figure 2.2A). Expression of plasmid-borne *rcc00645*, *rcc02629* and *rcc02857* in the parental strain reduced RcGTA production (Figure 2.2A). These results are consistent with the inhibitory role of Rcc00645, Rcc02629 and Rcc02857 in RcGTA production. Expression of *rcc00620* complemented the *rcc00620* mutation and increased RcGTA production in the parental strain, consistent with a stimulating role for Rcc00620 in RcGTA production. Quantification of gene expression for the RcGTA major structural gene cluster via a reporter fusion showed the changes in gene transfer activities in the four mutant strains matched with changes in RcGTA gene expression (Figure 2.3). Quantification of the

amounts of RcGTA major capsid protein within cells and released into the extracellular environment (Appendix 1, Figure S2.2) also indicated that the observed patterns in RcGTA activities in the mutant strains could be explained by changes in production and release of RcGTA for all four genes. Also, the relative RcGTA activities and capsid protein levels matched well when these genes were present in either mutant or parental strains. One exception was for *rcc02629,* where the decreased RcGTA activity observed when *rcc02629* was present on plasmid in SB1003 (Figure 2.2A) did not match with lower capsid protein levels (Appendix 1, Figure S2.2), although this strain also showed the largest standard deviation in this assay.

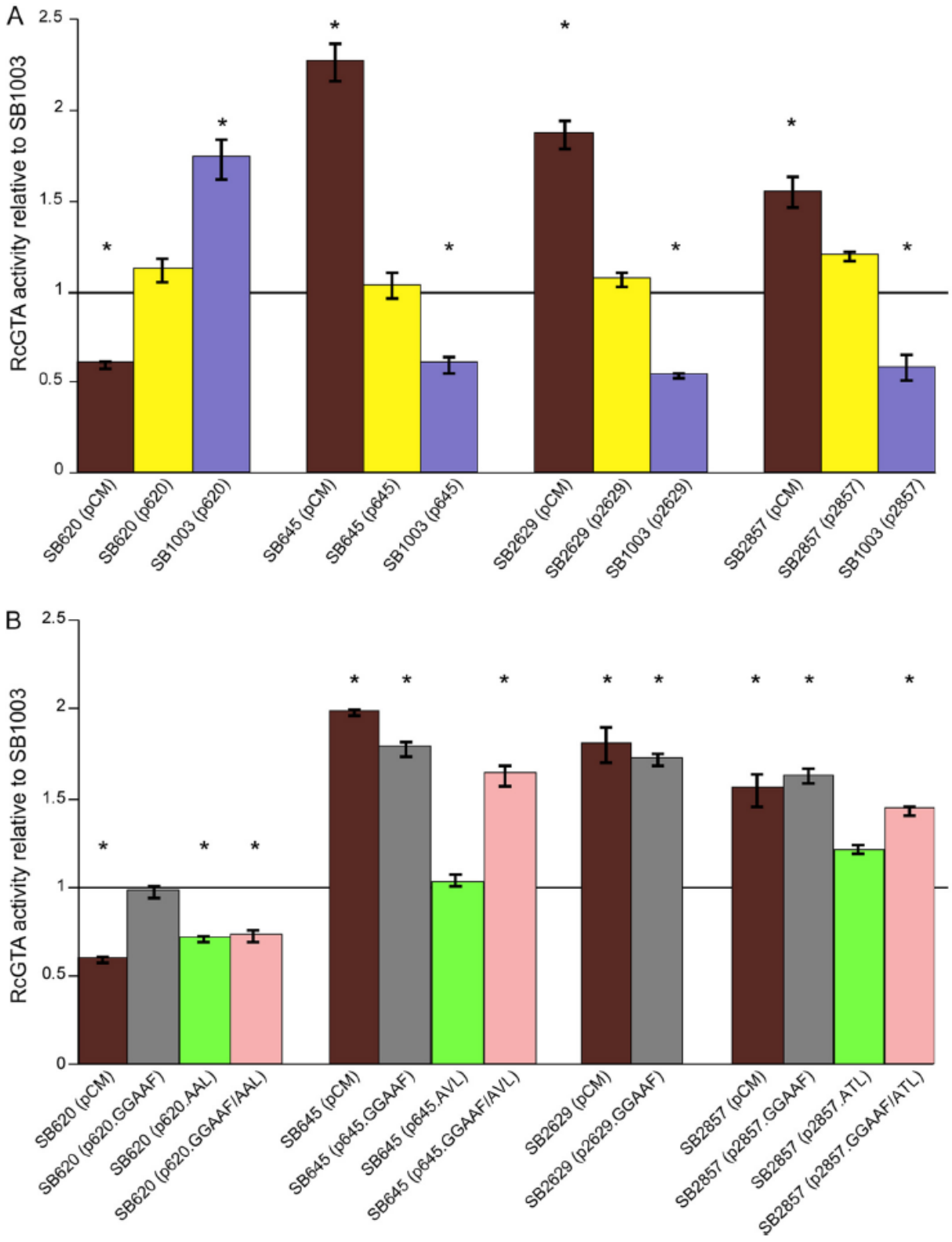

A
RcGTA activity relative to SB1003
2.5
2
1.5
1
0.5
0
SB620 (pCM)
SB620 (p620)
SB1003 (p620)
SB645 (pCM)
SB645 (p645)
SB1003 (p645)
SB2629 (pCM)
SB2629 (p2629)
SB1003 (p2629)
SB2857 (pCM)
SB2857 (p2857)
SB1003 (p2857)
B
RcGTA activity relative to SB1003
2.5
2
1.5
1
0.5
0
SB620 (pCM)
SB620 (p620.GGAAF)
SB620 (p620.AAL)
SB620 (p620.GGAAF/AAL)
SB645 (pCM)
SB645 (p645.GGAAF)
SB645 (p645.AVL)
SB645 (p645.GGAAF/AVL)
SB2629 (pCM)
SB2629 (p2629.GGAAF)
SB2857 (pCM)
SB2857 (p2857.GGAAF)
SB2857 (p2857.ATL)
SB2857 (p2857.GGAAF/ATL)

Figure 2.2. Effects of gene disruptions, *trans*-complementation, and site-directed mutagenesis of enzymatic domains on RcGTA gene transfer activity. The gene transfer activities for mutants and strains containing plasmid-borne copies of the genes (A) and strains containing site-directed mutant versions of the genes (B) are presented as averages from 3 replicates relative to the parental strain, SB1003, carrying the empty vector, pCM62. Bars represent the standard deviations and statistically significant differences ($p < 0.05$) compared to the control, identified using one-way ANOVA followed by Tukey HSD post-hoc analysis, are indicated by asterisks.

2.3.2 Analysis of GGDEF and EAL domains for the proteins affecting RcGTA production

To investigate which activities are involved in stimulating and inhibiting RcGTA production, we introduced site-directed mutations in the 'GGDEF' and 'EAL' motifs (Figure 2.1) of the proteins of all four proteins to generate GGAAF, AAL/AVL/ATL and GGAAF+AAL/AVL/ATL derivatives. The site-directed mutant derivatives were introduced into the respective knockout strains and RcGTA gene transfer activities were assayed. The GGAAF mutation in Rcc00620 had no effect on gene transfer activity, while mutation in the EAL domain (and both GGDEF and EAL domains) abolished the ability of Rcc00620 to complement the *rcc00620* mutation (Figure 2.2B). This observation suggests that Rcc00620 has PDE activity that stimulates RcGTA production. The GGAAF mutations in Rcc00645, Rcc02629, and Rcc02857 abolished their ability to complement their respective mutations, while mutations in the EAL domains in Rcc00645 and Rcc02629 had little to no effects on complementation (Figure 2.2B). These results

suggest that Rcc00645, Rcc02629, and Rcc02857 have DGC activities that inhibit RcGTA production in *R. capsulatus*.

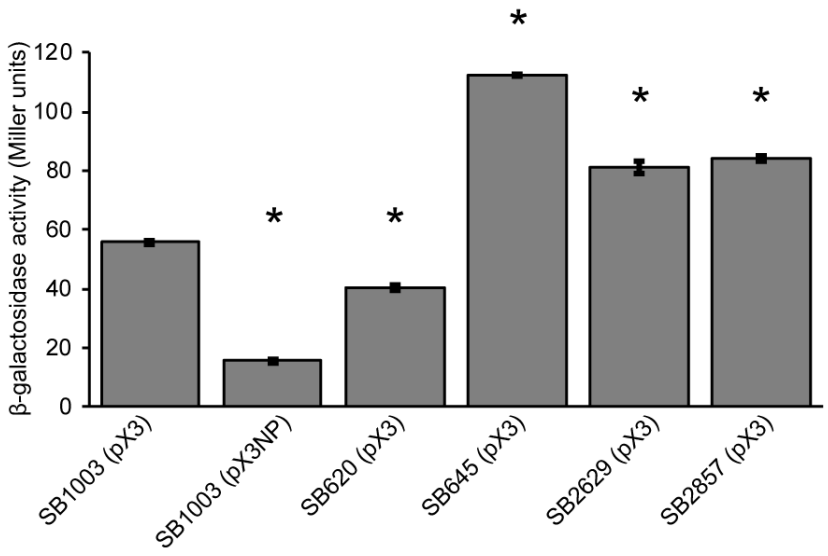


Figure 2.3. Effects of gene disruptions on RcGTA gene expression. β-galactosidase activities were measured for the indicated strains carrying the RcGTA *orfg3*'::'*lacZ* fusion construct, pX3. SB1003 (pX3NP) is the promoter-less negative control. The results are the average of three biological replicates with bars representing the standard deviations. Statistically significant differences ($p < 0.0001$) compared to the control, SB1003 (pX3), were identified using one-way ANOVA followed by Tukey HSD post-hoc analysis and are indicated by asterisks.

To test the cumulative effect of losses of the suspected DGC-encoding genes on GTA production, a *rcc00645.02629.02857* triple knockout mutant was constructed and tested for RcGTA activity. This strain did show an increase in RcGTA activity relative to the individual mutants (Appendix 1, Figure S2.3). Disruption of both *rcc00620* and

rcc00645 resulted in gene transfer activity comparable to the wild type strain (Appendix 1, Figure S2.3), indicating the loss of these two genes had compensatory effects.

2.3.3 Effects of changing intracellular c-di-GMP levels on RcGTA production

The above analyses supported the notion that the DGC activities of these *R. capsulatus* proteins are responsible for the reduction of RcGTA production, while PDE activity is responsible for stimulating RcGTA production. To test the effects of c-di-GMP levels on RcGTA production more directly, we expressed heterologous DGC- (*R. sphaeroides dgcA*) and PDE-encoding (*G. xylinus pde1*) genes in *R. capsulatus*. RcGTA gene transfer assays showed that expression of the PDE caused an approximately 50% increase in RcGTA production whereas expression of the DGC caused an approximately 40% decrease (Figure 2.4).

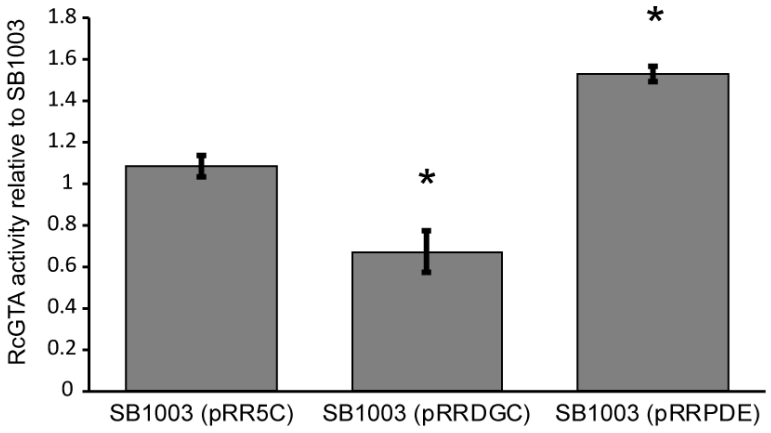


Figure 2.4. Expression of genes encoding known DGC and PDE enzymes affects RcGTA gene transfer activity. The gene transfer activities for *R. capsulatus* SB1003 carrying the expression plasmid pRR5C (pRR), pRR5C with *pdeA* from *G. xylinus* (pRRPDE), and pRR5C with *dgcA* from *R. sphaeroides* (pRRDGC) are presented as average ratios

relative to SB1003 carrying no exogenous plasmid. The data come from 3 replicates with bars showing the standard deviations and statistically significant differences ($p < 0.05$) compared to the control, identified using one-way ANOVA followed by Tukey HSD post-hoc analysis, are indicated by asterisks.

To confirm the disruptions or expression of the different genes were leading to changes in intracellular c-di-GMP levels, we quantified these in the different strains. The mutant strains and the parental strain with and without the heterologous DGC and PDE genes were cultured under the same conditions as used for the GTA bioassay experiments and subjected to c-di-GMP quantification. Expression of the heterologous PDE resulted in lower levels of c-di-GMP, as did loss of the *rcc00645, rcc02629* and *rcc02857* genes (Figure 2.5). Expression of the heterologous DGC increased the c-di-GMP levels, as did disruption of *rcc00620* (Figure 2.5) although the difference was not statistically significant for this strain.

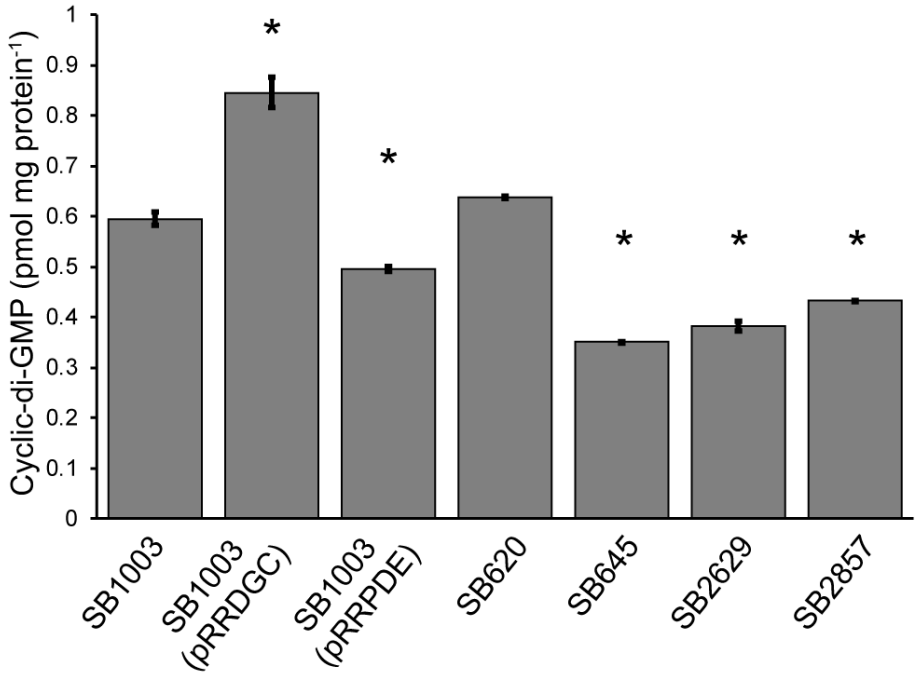


Figure 2.5. Quantification of intracellular c-di-GMP levels in *R. capsulatus* strains. C-di-GMP levels were measured by HPLC and normalized to protein content of the corresponding cell samples. The data come from 3 replicates with the bars representing the standard deviations and statistically significant differences ($p < 0.05$) compared to the control, identified using one-way ANOVA followed by Tukey HSD post-hoc analysis, are indicated by asterisks.

2.3.4 Assaying *R. capsulatus* proteins for DGC and PDE activities in *E. coli*

To gain a better understanding of the DGC or PDE activities of the *R. capsulatus* proteins, we expressed them in the *E. coli* c-di-GMP reporter strain MG1655, where expression of an active heterologous DGC that increases the intracellular c-di-GMP levels inhibits swimming motility on semi-solid agar (Chen et al. 2014). Expression of *rcc00620, rcc02629* and *rcc02857* in MG1655 decreased the swim zones compared to the

empty vector control (Figure 2.6A), indicative of their DGC activities in *E. coli*. Expression of *rcc00645* had a small inhibitory effect the swim zone. We also assessed a second c-di-GMP-dependent *E. coli* phenotype, fimbriae production in strain BL21(DE3), which can be detected by Congo red staining (Christen et al. 2006). The pattern was the same as for the motility assays, with expression of *rcc00620*, *rcc02629*, and *rcc02857* increasing Congo red staining compared to the vector control and expression of *rcc00645* having no effect (Figure 2.6B).

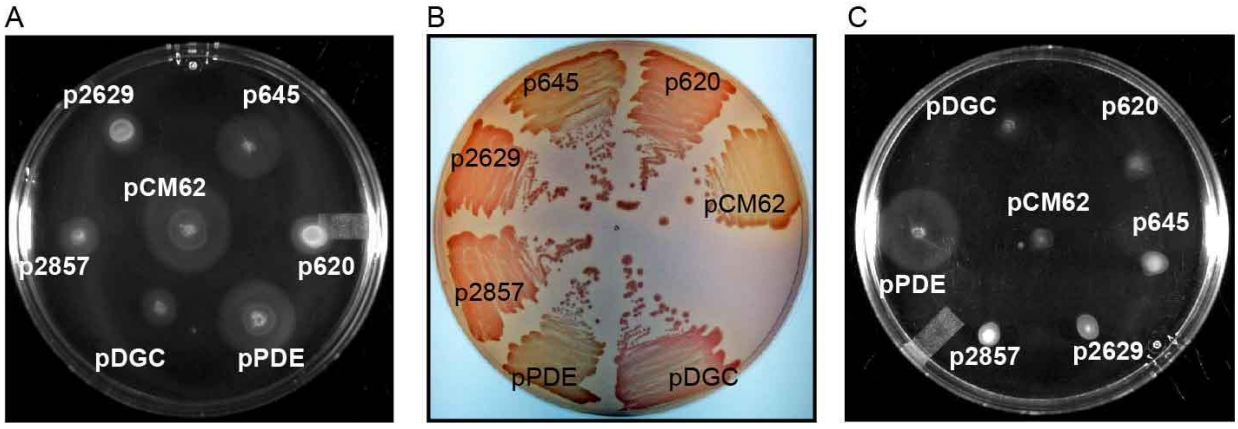


Figure 2.6. Evaluation of *R. capsulatus* proteins for potential DGC and PDE activities in *E. coli*. A. Motility of *E. coli* MG1655 on semi-solid agar (0.25%), which is reduced by DGC activity, when containing the indicated plasmids. B. Congo red binding by *E. coli* BL21(DE3), where DGC activity increases fimbriae production and Congo Red binding, when containing the indicated plasmids. C. Motility of *E. coli* MG1655 Δ*yhjH* on semi-solid media (0.25%), which is increased by PDE activity, when containing the indicated plasmids. In all experiments, transcription of the genes from the plasmid's *lac* promoter was induced with IPTG.

We next assessed the effects of the site-directed mutations in the GGDEF and EAL domains in these *E. coli* DGC assays. The GGAAF mutations in Rcc00620, Rcc02629, and Rcc02857 decreased or abolished their presumed DGC activities in the motility (Appendix 1, Figure S2.4A) and Congo red binding (Appendix 1, Figure S2.4B) assays, which is consistent with these three proteins possessing DGC activities in *E. coli*. The EAL domain mutations did not affect the activities of the Rcc00620 and Rcc02857 proteins (Appendix 1, Figure S2.4AB). The GGAAF mutation in Rcc00645 had no effect in either assay, but mutation of the EAL domain resulted in increased DGC activity in both assays (Appendix 1, Figure S2.4AB). These results are consistent with Rcc00645 having modest levels of both DGC activity and PDE activity, with the DGC activity only evident when the PDE activity is abolished.

The wild type and mutant genes were also tested in *E. coli* MG1655 Δ*yhjH*, where a drop in intracellular c-di-GMP levels caused by expression of an active PDE restores swimming motility on semi-solid agar (Ryjenkov et al. 2006). Expression of *rcc00645*, *rcc00620* and *rcc02857* had little or no observable effect on the swimming phenotype (Figure 2.6C). However, the GGAAF mutations in Rcc00620, Rcc00645 and Rcc02857 increased the PDE activities of all three, as indicated by the larger swim zones (Appendix 1, Figure S2.5). Mutation of the EAL domains in these GGAAF mutants reduced or abolished this evidence of PDE activity (Appendix 1, Figure S2.5). As expected, Rcc02629 did not exhibit PDE activity, nor did its GGAAF mutant version (Appendix 1, Figure S2.5).

Taken together, these assays indicated that, in *E. coli*, Rcc02629 acts as a DGC while Rcc00620, Rcc00645 and Rcc02857 possess both DGC and PDE activities.

2.3.5 Effect of aerobic versus anaerobic growth on Rcc00645 and RcGTA production

The Rcc00645 protein contains several PAS domains, one of which (aa 142-244) is predicted to bind heme, a common moiety involved in oxygen sensing. Since both its GGDEF and EAL domains appear to be enzymatically active based on our assays, it is possible that this bifunctional protein switches from acting as a DGC under anaerobic conditions, which were used for the RcGTA production assays, to a PDE under aerobic conditions, which were used for the *E. coli* motility assays. To test this hypothesis, we assayed GTA production by SB645 when grown under aerobic versus anaerobic conditions. Interestingly, there was no significant difference in gene transfer activity for this mutant compared to the parental strain under aerobic conditions, compared to the large increase observed in this strain under anaerobic conditions (Appendix 1, Figure S2.6). This suggests oxygen does affect Rcc00645 such that it is more active as a DGC under anaerobic conditions. However, there is no evidence from this result that the protein acts as a PDE under aerobic conditions in *R. capsulatus* as gene transfer activity was not lower in the mutant strain.

2.3.6 C-di-GMP and *R. capsulatus* flagellar motility

Intracellular c-di-GMP concentrations are commonly involved in regulating flagellar motility in bacteria (Romling, Galperin, and Gomelsky 2013). We therefore conducted flagellar motility assays to look for effects of the gene disruptions and expression of the heterologous PDE and DGC genes on this phenotype in *R. capsulatus*. The parental strain, SB1003, and its non-motile *ctrA* mutant derivative, SBRM1, served

as controls. Expression of the heterologous DGC caused a significant decrease in motility while the heterologous PDE significantly increased the swim diameter (Figure 2.7). Therefore, c-di-GMP inhibits flagellar motility in *R. capsulatus*. With this in mind, the phenotypes of the mutant strains (Figure 2.7) were all as predicted based on the results of the experiments presented above (e.g. Figure 2.4), although the difference was only statistically significant for SB645.

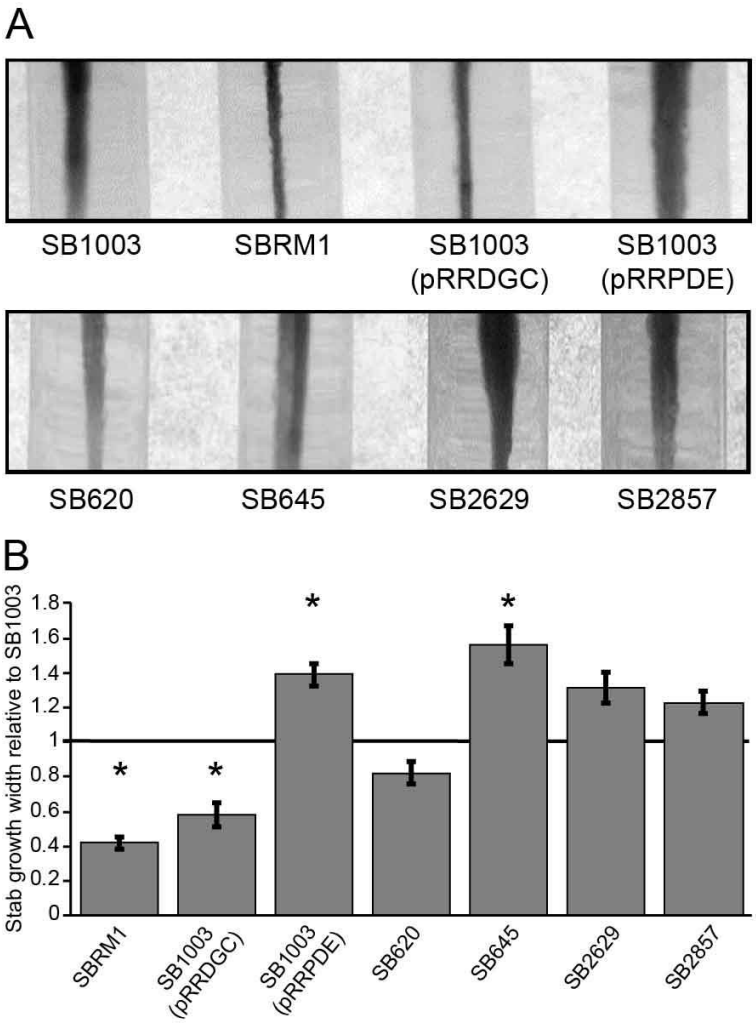


Figure 2.7. Effects of gene disruptions and alterations of c-di-GMP levels on *R. capsulatus* flagellar motility. A. Motility assays of the four mutant strains and the parental strain, SB1003, carrying the heterologous PDE and DGC expression plasmids. The

parental strain without any manipulation and its non-motile *ctrA* null mutant derivative, SBRM1, are included as reference strains. The distance of growth away from the center of the stab lines in 0.35% agar reflects the relative motility of the strain. B. The swimming diameters were measured from three replicate assays and plotted as relative to SB1003. The bars represent the standard deviation and statistically significant differences ($p < 0.05$) compared to the control, identified using one-way ANOVA followed by Tukey HSD post-hoc analysis, are indicated by asterisks.

2.4 Discussion

The genome of *R. capsulatus* is predicted to encode 2 EAL, 2 GGDEF and 16 tandem GGDEF/EAL domain-containing proteins (Strnad et al. 2010). Several of these genes were identified in a previous transcriptomic study as having significantly changed transcript levels in the absence of the response regulator CtrA (Mercer et al. 2010), which is absolutely required for the production of RcGTA (Lang and Beatty 2000). This led us to hypothesize that these proteins potentially involved in c-di-GMP signaling might affect RcGTA production. Indeed, disruptions of four of the genes were found to affect RcGTA gene transfer activity, as did increasing their copy number, which presumably increased their expression, in the parental strain (Figure 2.2; Table 2.3). Disruptions of *rcc00645*, *rcc02629* and *rcc02857* resulted in similar RcGTA phenotypes with increases in RcGTA production compared to the parental strain, and loss of all three of these genes showed a slightly larger effect than for any one individual gene (Appendix 1, Figure S2.3). Disruption of *rcc00620* lowers RcGTA production compared to the parental strain. Both reporter fusion assays and western blotting showed that loss of the four genes affected RcGTA gene expression (Figures 3 and Appendix 1, S2.2). Importantly, increasing the

intracellular concentration of c-di-GMP by expression of a heterologous DGC inhibited gene transfer activity, while decreasing the c-di-GMP levels by expression of a heterologous PDE stimulated gene transfer activity (Figure 2.4; Table 3). The intracellular c-di-GMP levels in the studied *R. capsulatus* strains matched the expectations based on the phenotypic experiments (Figure 2.5) and these results establish that c-di-GMP acts as an inhibitor of RcGTA production and that, with respect to effects on RcGTA, the Rcc00645, Rcc02629 and Rcc02857 proteins act as DGCs in *R. capsulatus* whereas Rcc00620 acts as a PDE. Furthermore, c-di-GMP levels and these specific signaling proteins are also implicated in regulating flagellar motility (Figure 2.7), further entwining the co-regulation of RcGTA production and motility in *R. capsulatus* (Lang and Beatty 2002; Mercer et al. 2012).

Table 2.3. Summary of phenotypes from *R. capsulatus* and *E. coli* assays.

Strain/plasmid	*R. capsulatus* phenotype (RcGTA activity)[a]	*E. coli* phenotype		
		Fimbriae production (Congo red binding)[b]	MG1655 motility[b]	MG1655 Δ*yhjH* motility[c]
SB620	↓			
SB1003 (p620)	↑			
SB645	↑			
SB1003 (p645)	↓			
SB2629	↑			
SB1003 (p2629)	↓			
SB2857	↑			

SB1003 (p2857)	↑			
SB620.645.	=			
SB645.2629.2857	↑			
SB1003 (pRRDGC)	↓			
SB1003 (pRRPDE)	↑			
pDGC		↑	↓	=
pPDE		=	=	↑
p620		↑	↓	=
p620GGAAF	=	=	=	↑
p620AAL	↓	↑	↓	=
p620GGAAF/AAL	↓	=	=	=
p645		=	=	=
p645GGAAF	↑	=	=	↑
p645AVL	=	↑	↓	=
p645GGAAF/AVL	↑	=	=	=
p2629		↑	↓	=
p2629GGAAF	↑	=	=	=
p2857		↑	↓	=
p2857GGAAF	↑	=	=	↑
p2857ATL	=	↑	↓	=
p2857GGAAF/ATL	↑	=	=	=

[a] From Figures 2, 4, S2.1 and S2.3

[b] From Figures 6 and S2.4

[c] From Figures 6 and S2.5

Disruptions of the different DGC-encoding genes affect gene transfer to different degrees (Figure 2.2A). This could be due to differences in the expression levels of the different genes or enzymatic activities of the encoded proteins or some combination of both these factors. Examination of the transcript levels for these genes (Mercer et al. 2010) from the same growth conditions as used for the experiments here indicated *rcc00645* has the highest transcript levels of the three genes, possibly explaining some of our results. However, *rcc02857* transcript levels are more than three-fold higher than those of *rcc02629*, yet loss of *rcc02629* had a bigger effect on gene transfer activity. Increasing the copy number for all three genes in the wild type background resulted in similar effects (Figure 2.2A) and the c-di-GMP levels in the three different mutants are quite similar (Figure 2.5). Therefore, there does not seem to be an easy explanation in terms of transcript levels or magnitude of effects on intracellular c-di-GMP levels for the differences in the observed effects for the different genes.

Bifunctional DGC-PDE proteins are fairly common (Tarutina, Ryjenkov, and Gomelsky 2006; Romling, Galperin, and Gomelsky 2013), and three of the proteins implicated in RcGTA production contain intact versions of both GGDEF and EAL domains while one contains only a GGDEF domain (Figure 2.1). Site-directed mutations in these domains (Figure 2.2B) validate the interpretations based on the null mutant (Figure 2.2A) and heterologous gene expression results (Figure 2.4). The functionality of these domains and the enzymatic activities of the proteins were further interrogated in *E. coli* c-di-GMP reporter strain assays. These confirmed that Rcc02629 possesses DGC activity and revealed that the other three proteins display both DGC and PDE activities

(Figures 6, 7, S2.4, and S2.5; Table 3). From the *E. coli* assays, for Rcc00620 and Rcc02857 the DGC activity dominates and PDE activity is only evident when the GGDEF domain is mutated, whereas for Rcc00645 the PDE activity dominates and the DGC activity is enhanced by mutation of the EAL domain. These results confirm the functionality of these domains, but the activities of the proteins in *E. coli* and how this relates to their activities in *R. capsulatus* needs to take into consideration that the three bifunctional proteins contain additional signaling/regulatory domains that likely regulate their enzymatic activities (Wolfe and Visick 2010). Indeed, a preliminary investigation of a possible effect of oxygen on the activity of Rcc00645, which contains a PAS domain predicted to bind heme, suggests this bifunctional protein shows higher DGC activity under anaerobic conditions (Appendix 1, Figure S2.6). It is possible that this differential activity of Rcc00645 with respect to RcGTA might be part of the explanation for the observation made more than 40 years ago that much less RcGTA production occurs when cultures are grown aerobically versus anaerobically (Solioz 1975). Rcc00620 displays DGC activity in *E. coli*, whereas it acts as a PDE in *R. capsulatus.* We speculate that the switch between DGC and PDE activities may depend on the phosphorylation status of its N-terminal REC domain (Levet-Paulo et al. 2011; Ryjenkov et al. 2005). The phosphorylation status may differ in *R. capsulatus* and *E. coli* due to the lack of a cognate kinase/phosphatase protein(s) in *E. coli*.

How c-di-GMP is connected to RcGTA production mechanistically is unknown at present. One possible link may involve CckA, the sensor kinase component of the CckA-ChpT-CtrA histidyl-aspartyl phosphorelay. This regulatory system has been extensively characterized in *C. crescentus* (Curtis and Brun 2010; Tsokos and Laub 2012), where it was also first discovered (Quon, Marczynski, and Shapiro 1996; Jacobs et al. 1999;

Biondi et al. 2006). It is almost universally conserved within the class α-proteobacteria (Brilli et al. 2010), although its functions vary among lineages within this group (Barnett et al. 2001; Bellefontaine et al. 2002; Curtis and Brun 2010; Kim, Heindl, and Fuqua 2013; Greene et al. 2012; Miller and Belas 2006; Mercer et al. 2010). Kinase and phosphatase activities of CckA are regulated directly by c-di-GMP in *C. crescentus* (Mann et al. 2016; Lori et al. 2015). Binding of c-di-GMP by CckA favors phosphatase activity, thus resulting in CtrA existing primarily in the non-phosphorylated state. The *C. crescentus* CckA c-di-GMP binding sites are conserved in the *R. capsulatus* protein sequence, suggesting that a similar regulation might also occur in this bacterium and the effects of disrupting or overexpressing the four genes studied here could therefore be mediated in part by CckA. The dysregulation of these genes in the absence of CtrA, including evidence for direct regulation of *rcc00645* by CtrA due to the presence of an upstream consensus binding site (Mercer et al. 2010), would then also suggest there is some sort of feedback loop at work, with these proteins potentially modulating CtrA phosphorylation state and CtrA directly and/or indirectly affecting transcript levels of these genes. A similar feedback situation may also exist in the α-proteobacterium *Dinoroseobacter shibae*, where CtrA also affects the transcript levels of c-di-GMP signaling genes (Koppenhöfer et al. 2019). Indeed, there are many commonalities in the interconnections of GTA-regulating systems (quorum sensing, LexA, etc.) with other aspects of biology (e.g. flagellar motility) in the two species.

The production and release of RcGTA particles is not as straightforward as once believed. Most of the particle structure is encoded in an approximately 14-kb gene cluster (Lang and Beatty 2000), but additional essential genes for the structure (Hynes et al.

2016) and particle release (Hynes et al. 2012) are encoded elsewhere. Genes encoding head spike proteins (Westbye et al. 2016), which are not essential but that improve the particles' binding to recipient cells (Westbye et al. 2016; Hynes et al. 2016), are also encoded elsewhere. Importantly, some of these genes are transcribed in the presence of unphosphorylated CtrA, such as the main structural cluster (Mercer et al. 2012), while others require phosphorylated CtrA and are not expressed in the absence of CckA or ChpT, such as the lysis and head spike genes (Westbye et al. 2013; Westbye et al. 2016). It is also important to note that CckA is required for proper function and release of RcGTA independent of CtrA and CtrA~P (Mercer et al. 2012) and therefore, if c-di-GMP is affecting CckA activity, some of the effects seen in this study could also be due to this CtrA-independent pathway.

A previous network-based gene co-expression analysis of *R. capsulatus* microarray transcriptomic data from 23 different strains and/or growth conditions sorted the *R. capsulatus* genes into 40 distinct modules (Peña-Castillo et al. 2014). The *rcc00645* gene falls within the same gene module as the RcGTA structural cluster (Lang and Beatty 2000), lysis (Hynes et al. 2012), head spike (Westbye et al. 2016; Hynes et al. 2016), putative tail spike (Hynes et al. 2016; Lang, Westbye, and Beatty 2017), and phage-related regulatory (Fogg 2019; Hynes et al. 2016) genes. The *rcc00620*, *rcc02629* and *rcc02857* genes are within a different gene module that also contains the RcGTA-regulatory partner-switching phosphorelay genes, *rbaVWY* (Mercer and Lang 2014). Therefore, in addition to the connections of these c-di-GMP signaling genes with respect to CtrA, RcGTA production and motility, they show linkages in regulation that are maintained through analysis of diverse transcriptomic datasets.

To the many bacterial processes known to be affected by c-di-GMP signaling (Romling, Galperin, and Gomelsky 2013; Jenal, Reinders, and Lori 2017) this study adds one more, gene transfer. Our experiments showed that elevated c-di-GMP concentrations inhibit RcGTA gene transfer and flagellar motility in *R. capsulatus*. Of the four proteins implicated in affecting RcGTA production and motility, three contain both GGDEF and EAL domains, all of which appear enzymatically functional. Regulation of the activities of these proteins, most likely involving their additional signaling domains (Figure 2.1), is therefore expected to be an important aspect of their functioning with respect to these behaviors. Indeed, oxygen appears to be an important factor for the activity of Rcc00645. Additional research is also required to evaluate the possible connection between c-di-GMP signaling uncovered here and the CckA-ChpT-CtrA regulatory pathway.

2.5 References

Barany, F. (1985). Single-stranded hexameric linkers: a system for in-phase insertion mutagenesis and protein engineering. *Gene* 37, 111–123. doi:10.1016/0378-1119(85)90263.

Barnett, M. J., Hung, D. Y., Reisenauer, A., Shapiro, L., and Long, S. R. (2001). A homolog of the CtrA cell cycle regulator is present and essential in *Sinorhizobium meliloti*. *J. Bacteriol.* 183, 3204–3210. doi:10.1128/JB.183.10.3204-3210.2001.

Beatty, J. T., and Gest, H. (1981). Generation of succinyl-coenzyme A in photosynthetic bacteria. *Arch. Microbiol.* 129, 335–340. doi:10.1007/BF00406457.

Bellefontaine, A.-F., Pierreux, C. E., Mertens, P., Vandenhaute, J., Letesson, J.-J., and De Bolle, X. (2002). Plasticity of a transcriptional regulation network among alpha-proteobacteria is supported by the identification of CtrA targets in *Brucella abortus*. *Mol. Microbiol.* 43, 945–960. doi:10.1046/j.1365-2958.2002.02777.x.

Biondi, E. G., Reisinger, S. J., Skerker, J. M., Arif, M., Perchuk, B. S., Ryan, K. R., et al. (2006). Regulation of the bacterial cell cycle by an integrated genetic circuit. *Nature* 444, 899–904. doi:10.1038/nature05321.

Blattner, F. R. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1462. doi:10.1126/science.277.5331.1453.

Brilli, M., Fondi, M., Fani, R., Mengoni, A., Ferri, L., Bazzicalupo, M., et al. (2010). The diversity and evolution of cell cycle regulation in alpha-proteobacteria: a comparative genomic analysis. *BMC Syst. Biol.* 4, 52. doi:10.1186/1752-0509-4-52.

Chan, C., Paul, R., Samoray, D., Amiot, N. C., Giese, B., Jenal, U., et al. (2004). Structural basis of activity and allosteric control of diguanylate cyclase. *Proc. Natl. Acad. Sci. U. S. A.* 101, 17084–17089. doi:10.1073/pnas.0406134101.

Chang, A. L., Tuckerman, J. R., Gonzalez, G., Mayer, R., Weinhouse, H., Volman, G., et al. (2001). Phosphodiesterase A1, a regulator of cellulose synthesis in *Acetobacter xylinum*, is a heme-based sensor. *Biochemistry* 40, 3420–3426. doi:10.1021/bi0100236.

Chen, L. H., Köseoğlu, V. K., Güvener, Z. T., Myers-Morales, T., Reed, J. M., D'Orazio, S. E. F., et al. (2014). Cyclic di-GMP-dependent signaling pathways in the pathogenic firmicute *Listeria monocytogenes*. *PLoS Pathog.* 10, e1004301. doi:10.1371/journal.ppat.1004301.

Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P., et al. (2006). Allosteric control of cyclic di-GMP signaling. *J. Biol. Chem.* 281, 32015–32024. doi:10.1074/jbc.M603589200.

Christen, M., Christen, B., Folcher, M., Schauerte, A., and Jenal, U. (2005). Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. *J. Biol. Chem.* 280, 30829–30837. doi:10.1074/jbc.M504429200.

Curtis, P. D., and Brun, Y. V. (2010). Getting in the Loop: regulation of development in *Caulobacter crescentus*. *Microbiol. Mol. Biol. Rev.* 74, 13–41. doi:10.1128/MMBR.00040-09.

De Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., et al. (2006). ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* 34, W362–365. doi:10.1093/nar/gkl124.

Fogg, P. C. M. (2019). Identification and characterization of a direct activator of a gene transfer agent. *Nat. Commun.* 10, 595. doi:10.1038/s41467-019-08526-1.

Fu, Y., MacLeod, D. M., Rivkin, R. B., Chen, F., Buchan, A., and Lang, A. S. (2010). High diversity of *Rhodobacterales* in the subarctic North Atlantic ocean and gene transfer agent protein expression in isolated strains. *Aquat. Microb. Ecol.* 59, 283–293. doi:10.3354/ame01398.

Greene, S. E., Brilli, M., Biondi, E. G., and Komeili, A. (2012). Analysis of the CtrA pathway in magnetospirillum reveals an ancestral role in motility in alphaproteobacteria. *J. Bacteriol.* 194, 2973–2986. doi:10.1128/JB.00170-12.

He, F. (2011). Bradford protein assay. *Bioprotocol* 1, e45. doi:10.21769/bioprotoc.45.

Hesterberg, T., Chambers, J. M., and Hastie, T. J. (1993). *Statistical Models in S*. doi:10.2307/1269676.

Hynes, A. P., and Lang, A. S. (2013). *Rhodobacter capsulatus* gene transfer agent (RcGTA) activity bioassays. *Bioprotocol* 3, e317. doi:10.1093/molbev/msw125.

Hynes, A. P., Mercer, R. G., Watton, D. E., Buckley, C. B., and Lang, A. S. (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. doi:10.1111/j.1365-2958.2012.08113.x.

Hynes, A. P., Shakya, M., Mercer, R. G., Grüll, M. P., Bown, L., Davidson, F., et al. (2016). Functional and evolutionary characterization of a gene transfer agent's multilocus "genome." *Mol. Biol. Evol.* 33, 2530–2543. doi:10.1093/molbev/msw125.

Jacobs, C., Domian, I. J., Maddock, J. R., and Shapiro, L. (1999). Cell cycle-dependent polar localization of an essential bacterial histidine kinase that controls DNA replication and cell division. *Cell* 97, 111–120. doi:10.1016/S0092-8674(00)80719-

9.
Jenal, U., Reinders, A., and Lori, C. (2017). Cyclic di-GMP: second messenger extraordinaire. *Nat. Rev. Microbiol.* 15, 271–284. doi:10.1038/nrmicro.2016.190.
Kim, J., Heindl, J. E., and Fuqua, C. (2013). Coordination of division and development influences complex multicellular behavior in *Agrobacterium tumefaciens*. *PLoS One* 8, e56682. doi:10.1371/journal.pone.0056682.
Koppenhöfer, S., Wang, H., Scharfe, M., Kaever, V., Wagner-Döbler, I., and Tomasch, J. (2019). Integrated transcriptional regulatory network of quorum sensing, replication control, and SOS response in *Dinoroseobacter shibae*. *Front. Microbiol.* 10, 803. doi:10.3389/fmicb.2019.00803.
Kuchinski, K. S., Brimacombe, C. A., Westbye, A. B., Ding, H., and Beatty, T. J. (2016). The SOS response master regulator LexA regulates the gene transfer agent of *Rhodobacter capsulatus* and represses transcription of the signal transduction protein CckA. *J. Bacteriol.* 198, 1137–1148. doi:10.1128/JB.00839-15.
Kuchma, S. L., Brothers, K. M., Merritt, J. H., Liberati, N. T., Ausubel, F. M., and O'Toole, G. A. (2007). BifA, a cyclic-di-GMP phosphodiesterase, inversely regulates biofilm formation and swarming motility by *Pseudomonas aeruginosa* PA14. *J. Bacteriol.* 189, 8165–8178. doi:10.1128/JB.00586-07.
Lang, A. S., and Beatty, J. T. (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. doi:10.1073/pnas.97.2.859.
Lang, A. S., and Beatty, J. T. (2002). A bacterial signal transduction system controls genetic exchange and motility. *J. Bacteriol.* 184, 913–918. doi:10.1128/jb.184.4.913-918.2002.
Lang, A. S., and Beatty, J. T. (2007). Importance of widespread gene transfer agent genes in α-proteobacteria. *Trends Microbiol.* 15, 54–62. doi:10.1016/j.tim.2006.12.001.
Lang, A. S., Westbye, A. B., and Beatty, J. T. (2017). The distribution, evolution, and roles of gene transfer agents in prokaryotic genetic exchange. *Annu. Rev. Virol.* 4, 87–104. doi:10.1146/annurev-virology-101416-041624.
Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi:10.1093/bioinformatics/btm404.
Letunic, I. (2004). SMART 4.0: towards genomic data integration. *Nucleic Acids Res.* 32, 142D–144. doi:10.1093/nar/gkh088.
Leung, M. M., Brimacombe, C. A., Spiegelman, G. B., and Beatty, J. T. (2012). The GtaR protein negatively regulates transcription of the *gtaRI* operon and modulates gene transfer agent (RcGTA) expression in *Rhodobacter capsulatus*. *Mol. Microbiol.* 83, 759–774. doi:10.1111/j.1365-2958.2011.07963.x.
Levet-Paulo, M., Lazzaroni, J. C., Gilbert, C., Atlan, D., Doublet, P., and Vianney, A. (2011). The atypical two-component sensor kinase Lpl0330 from *Legionella pneumophila* controls the bifunctional diguanylate cyclase-phosphodiesterase Lpl0329 to modulate bis-(3'-5')-cyclic dimeric GMP synthesis. *J. Biol. Chem.* 286, 31136–31144. doi:10.1074/jbc.M111.231340.
Lori, C., Ozaki, S., Steiner, S., Böhm, R., Abel, S., Dubey, B. N., et al. (2015). Cyclic di-GMP acts as a cell cycle oscillator to drive chromosome replication. *Nature* 523, 236–239. doi:10.1038/nature14473.
Mann, T. H., Seth Childers, W., Blair, J. A., Eckart, M. R., and Shapiro, L. (2016). A cell

cycle kinase with tandem sensory PAS domains integrates cell fate cues. *Nat. Commun.* 7, 11454. doi:10.1038/ncomms11454.
Marrs, B. (1974). Genetic recombination in *Rhodopseudomonas capsulata*. *Proc. Natl. Acad. Sci. U. S. A.* 71, 971–973. doi:10.1073/pnas.71.3.971.
Marx, C. J., and Lidstrom, M. E. (2001). Development of improved versatile broad-host-range vectors for use in methylotrophs and other gram-negative bacteria. *Microbiology* 147, 2065–2075. doi:10.1099/00221287-147-8-2065.
Mercer, R. G., Callister, S. J., Lipton, M. S., Pasa-Tolic, L., Strnad, H., Paces, V., et al. (2010). Loss of the response regulator CtrA causes pleiotropic effects on gene expression but does not affect growth phase regulation in *Rhodobacter capsulatus*. *J. Bacteriol.* 192, 2701–2710. doi:10.1128/JB.00160-10.
Mercer, R. G., and Lang, A. S. (2014). Identification of a predicted partner-switching system that affects production of the gene transfer agent RcGTA and stationary phase viability in *Rhodobacter capsulatus*. *BMC Microbiol.* 14, 71. doi:10.1186/1471-2180-14-71.
Mercer, R. G., Quinlan, M., Rose, A. R., Noll, S., Beatty, J. T., and Lang, A. S. (2012). Regulatory systems controlling motility and gene transfer agent production and release in *Rhodobacter capsulatus*. *FEMS Microbiol. Lett.* 331, 53–62. doi:10.1111/j.1574-6968.2012.02553.x.
Miller, T. R., and Belas, R. (2006). Motility is involved in *Silicibacter* sp. TM1040 interaction with dinoflagellates. *Environ. Microbiol.* 8, 1648–1659. doi:10.1111/j.1462-2920.2006.01071.x.
Newell, P. D., Yoshioka, S., Hvorecny, K. L., Monds, R. D., and O'Toole, G. A. (2011). Systematic analysis of diguanylate cyclases that promote biofilm formation by *Pseudomonas fluorescens* Pf0-1 . *J. Bacteriol.* 193, 4685–4698. doi:10.1128/JB.05483-11.
Paul, R., Weiser, S., Amiot, N. C., Chan, C., Schirmer, T., Giese, B., et al. (2004). Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev.* 18, 715–727. doi:10.1101/gad.289504.
Peña-Castillo, L., Mercer, R. G., Gurinovich, A., Callister, S. J., Wright, A. T., Westbye, A. B., et al. (2014). Gene co-expression network analysis in *Rhodobacter capsulatus* and application to comparative expression analysis of *Rhodobacter sphaeroides*. *BMC Genomics* 15, 730. doi:10.1186/1471-2164-15-730.
Qi, Y., Chuah, M. L. C., Dong, X., Xie, K., Luo, Z., Tang, K., et al. (2011). Binding of cyclic diguanylate in the non-catalytic EAL domain of FimX induces a long-range conformational change. *J. Biol. Chem.* 286, 2910–7. doi:10.1074/jbc.M110.196220.
Quon, K. C., Marczynski, G. T., and Shapiro, L. (1996). Cell cycle control by an essential bacterial two-component signal transduction protein. *Cell* 84, 83–93. doi:10.1016/S0092-8674(00)80995-2.
Raz, Y., and Tannenbaum, E. (2010). The influence of horizontal gene transfer on the mean fitness of unicellular populations in static environments. *Genetics* 185, 327–337. doi:10.1534/genetics.109.113613.
Römling, U., Galperin, M. Y., and Gomelsky, M. (2013). Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol. Mol. Biol. Rev.* 77, 1–52. doi:10.1128/MMBR.00043-12.
Roy, A., Petrova, O., and Sauer, K. (2016). Extraction and quantification of cyclic di-

GMP from *Pseudomonas aeruginosa*. *Bioprotocol* 3, e828. doi:10.21769/bioprotoc.828.

Ryan, R. P., Fouhy, Y., Lucey, J. F., Crossman, L. C., Spiro, S., He, Y.-W., et al. (2006). Cell-cell signaling in *Xanthomonas campestris* involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. *Proc. Natl. Acad. Sci. U. S. A.* 103, 6712–6717. doi:10.1073/pnas.0600345103.

Ryjenkov, D. A., Simm, R., Römling, U., and Gomelsky, M. (2006). The PilZ domain is a receptor for the second messenger c-di-GMP: the PilZ domain protein YcgR controls motility in enterobacteria. *J. Biol. Chem.* 281, 30310–30314. doi:10.1074/jbc.C600179200.

Ryjenkov, D. A., Tarutina, M., Moskvin, O. V., and Gomelsky, M. (2005). Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: Insights into biochemistry of the GGDEF protein domain. *J. Bacteriol.* 187, 1792–1798. doi:10.1128/JB.187.5.1792-1798.2005.

Schaefer, A. L., Taylor, T. A., Beatty, J. T., and Greenberg, E. P. (2002). Long-chain acyl-homoserine lactone quorum-sensing regulation of *Rhodobacter capsulatus* gene transfer agent production. *J. Bacteriol.* 184, 6515–21. doi:10.1128/JB.184.23.6515-6521.2002.

Schirmer, T., and Jenal, U. (2009). Structural and mechanistic determinants of c-di-GMP signalling. *Nat. Rev. Microbiol.* 7, 724–735. doi:10.1038/nrmicro2203.

Schmidt, A. J., Ryjenkov, D. A., and Gomelsky, M. (2005). The ubiquitous protein domain EAL is a c-di-GMP-specific phosphodiesterase: Enzymatically active and inactive EAL domains. *J. Bacteriol.* 187, 4774–4781. doi:10.1128/JB.187.14.4774-4781.2005.

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi:10.1038/nmeth.2089.

Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5857–5864. doi:10.1073/pnas.95.11.5857.

Simm, R., Morr, M., Kader, A., Nimtz, M., and Römling, U. (2004). GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessibility to motility. *Mol. Microbiol.* 53, 1123–1134. doi:10.1111/j.1365-2958.2004.04206.x.

Simon, R., Priefer, U., and Pühler, A. (1983). A broad host range mobilization system for *in vivo* genetic engineering: Transposon mutagenesis in Gram-negative bacteria. *Bio/Technology* 1, 784–791. doi:10.1038/nbt1183-784.

Skerker, J. M., and Laub, M. T. (2004). Cell-cycle progression and the generation of asymmetry in *Caulobacter crescentus*. *Nat. Rev. Microbiol.* 2, 325–337. doi:10.1038/nrmicro864.

Solioz, M. (1975). The gene transfer agent of *Rhodopseudomonas capsulata*. PhD Thesis. Saint Louis University.

Stanton, T. B. (2007). Prophage-like gene transfer agents–Novel mechanisms of gene exchange for *Methanococcus*, *Desulfovibrio*, *Brachyspira*, and *Rhodobacter* species. *Anaerobe* 13, 43–49. doi:10.1016/j.anaerobe.2007.03.004.

Strnad, H., Lapidus, A., Paces, J., Ulbrich, P., Vlcek, C., Paces, V., et al. (2010). Complete genome sequence of the photosynthetic purple nonsulfur bacterium *Rhodobacter capsulatus* SB 1003. *J. Bacteriol.* 192, 3545–3546. doi:10.1128/JB.00366-10.

Tamayo, R., Tischler, A. D., and Camilli, A. (2005). The EAL domain protein VieA is a cyclic diguanylate phosphodiesterase. *J. Biol. Chem.* 280, 33324–33330. doi:10.1074/jbc.M506500200.

Tarutina, M., Ryjenkov, D. A., and Gomelsky, M. (2006). An unorthodox bacteriophytochrome from *Rhodobacter sphaeroides* involved in turnover of the second messenger c-di-GMP. *J. Biol. Chem.* 281, 34751–34758. doi:10.1074/jbc.M604819200.

Taylor, D. P., Cohen, S. N., Clark, W. G., and Marrs, B. L. (1983). Alignment of genetic and restriction maps of the photosynthesis region of the *Rhodopseudomonas capsulata* chromosome by a conjugation-mediated marker rescue technique. *J. Bacteriol.* 154, 580–590.

Tomasch, J., Wang, H., Hall, A. T. K., Patzelt, D., Preusse, M., Petersen, J., et al. (2018). Packaging of *Dinoroseobacter shibae* DNA into gene transfer agent particles is not random. *Genome Biol. Evol.* 10, 359–369. doi:10.1093/gbe/evy005.

Tsokos, C. G., and Laub, M. T. (2012). Polarity and cell fate asymmetry in *Caulobacter crescentus*. *Curr. Opin. Microbiol.* 15, 744–50. doi:10.1016/j.mib.2012.10.011.

Wall, J. D., Weaver, P. F., and Gest, H. (1975). Gene transfer agents, bacteriophages, and bacteriocins of *Rhodopseudomonas capsulata*. *Arch. Microbiol.* 105, 217–224. doi:10.1007/BF00447140.

Westbye, A. B., Kater, L., Wiesmann, C., Ding, H., Yip, C. K., and Beatty, J. T. (2018). The protease ClpXP and the PAS-domain protein DivL regulate CtrA and gene transfer agent production in *Rhodobacter capsulatus*. *Appl. Environ. Microbiol.* 84, e00275-18. doi:10.1128/AEM.00275-18.

Westbye, A. B., Kuchinski, K., Yip, C. K., and Beatty, J. T. (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. doi:10.1016/j.jmb.2015.12.010.

Westbye, A. B., Leung, M. M., Florizone, S. M., Taylor, T. A., Johnson, J. A., Fogg, P. C., et al. (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. doi:10.1128/jb.00669-13.

Wolfe, A. J., and Visick, K. L. (2010). *The Second Messenger Cyclic Di-GMP*. ASM Press doi:10.1128/9781555816667.

Wong, D. K. H., Collins, W. J., Harmer, A., Lilburn, T. G., and Beatty, J. T. (1996). Directed mutagenesis of the *Rhodobacter capsulatus puhA* gene and Orf 214: Pleiotropic effects on photosynthetic reaction center and light-harvesting 1 complexes. *J. Bacteriol.* 178, 2334–2342. doi:10.1128/jb.178.8.2334-2342.1996.

Yen, H. C., Hu, N. T., and Marrs, B. L. (1979). Characterization of the gene transfer agent made by an overproducer mutant of *Rhodopseudomonas capsulata*. *J. Mol. Biol.* 131, 157–168. doi:10.1016/0022-2836(79)90071-8.

Yen, H. C., and Marrs, B. (1976). Map of genes for carotenoid and bacteriochlorophyll biosynthesis in *Rhodopseudomonas capsulata*. *J. Bacteriol.* 126, 619–629.

Young, C. S., Reyes, R. C., and Beatty, J. T. (1998). Genetic complementation and kinetic analyses of *Rhodobacter capsulatus* ORF1696 mutants indicate that the ORF1696 protein enhances assembly of the light-harvesting I complex. *J. Bacteriol.* 180, 1759–1765.

CHAPTER 3- A two-component system acquired by horizontal gene transfer modulates gene transfer and motility via cyclic dimeric GMP

3.1 Introduction

Two-component systems (TCS) are widely used by bacteria to sense and respond to various internal and external stimuli (Stock, Robinson, and Goudreau 2000; Gao and Stock 2009; Zschiedrich, Keidel, and Szurmant 2016). A classic TCS comprises a sensor histidine kinase and a response regulator, proteins that are multidomain in nature, and cognate pairs are often encoded by neighbouring genes (Shiro and Yamada 2008). The chemistry of the TCS involves the sequential phosphorylation of two different amino acids that, for the majority, are a histidine within the kinase and an aspartic acid within the response regulator, and they are therefore also referred to as histidyl-aspartyl phosphorelay systems. In response to a specific signal, the kinase becomes activated and autophosphorylates on the histidine residue using ATP as the phosphodonor. The phosphate is subsequently transferred to the aspartic acid in the N-terminal receiver (REC) domain of the cognate response regulator, causing a conformational change that activates a C-terminal output/effector domain for the appropriate cellular response (Stock, Robinson, and Goudreau 2000). Different response regulators contain different types of output domains, such as helix-turn-helix DNA-binding, enzymatic, or protein/ligand-binding domains, and the activities of these domains are regulated based on the phosphorylation state of the REC domain (Perry, Koteva, and Wright 2011; Galperin 2010; Krell et al. 2010). Many response regulators trigger adaptive responses through the direct alteration of gene expression (Casino, Rubio, and Marina 2010), but >5% are predicted to be involved in cyclic dimeric guanosine monophosphate (c-di-GMP)

biosynthesis and/or degradation, indicating extensive TCS regulation of the cellular levels of this important signaling molecule (Romling, Galperin, and Gomelsky 2013).

C-di-GMP acts as a second messenger in bacteria, with changes in its levels affecting various processes in different organisms, such as motility, aggregation behaviors, and developmental transitions (Romling, Galperin, and Gomelsky 2013; Jenal, Reinders, and Lori 2017; Hengge 2009). Diguanylate cyclase (DGC) and phosphodiesterase (PDE) enzymes mediate the synthesis and degradation of c-di-GMP, respectively. Diguanylate cyclases contain a GGDEF motif in their active site, which catalyzes the cyclization of c-di-GMP from two GTP molecules (Paul et al. 2004). Phosphodiesterases contain EAL or HD-GYP motifs and break c-di-GMP into 5'-phosphoguanylyl-(3'-5')-guanosine (5'-pGpG) or two GMP, respectively (Simm et al. 2004; Tischler and Camilli 2004; Römling and Galperin 2017). Proteins containing both GGDEF and EAL domains are also common. In some of these proteins, one of the two domains is enzymatically inactive and serves as a regulatory site by binding to the GTP or c-di-GMP substrates (Hengge 2009). Others are bifunctional enzymes, such as BphG1 in *Rhodobacter sphaeroides* (Tarutina, Ryjenkov, and Gomelsky 2006), MSDGC-1 in *Mycobacterium smegmatis* (Kumar and Chatterji 2008), and ScrC in *Vibrio parahaemolyticus* (Ferreira et al. 2008), but it is still not completely clear if/how these two activities are reciprocally regulated and what determines the overall activity of such proteins. However, these proteins typically also contain additional sensory/signaling domains, such as Per-ARNT-Sim (PAS) or REC, that can affect the c-di-GMP-related domains in response to various stimuli (Romling, Galperin, and Gomelsky 2013).

Rhodobacter capsulatus is a purple non-sulfur photosynthetic bacterium that belongs to the class Alphaproteobacteria. It has been studied with respect to various

cellular functions (Strnad et al. 2010), especially phototrophy, and it is also known for being the first organism found to exchange DNA via a gene transfer agent (GTA) (Marrs 1974). GTAs, which are now known to be produced by multiple species of bacteria and one archaeon, resemble bacteriophages but they package small fragments of the producing cell's genome (4-14 kb, depending on the GTA) that is transferred to other cells (reviewed in (Lang, Beatty, and Rice 2017; Lang, Zhaxybayeva, and Beatty 2012; Lang and Beatty 2007; Stanton 2007)). Production of the *R. capsulatus* GTA (RcGTA) is affected by two regulators with clear evolutionary connections to phages (Hynes et al. 2016; Fogg 2019) as well as multiple cellular regulatory systems [reviewed in (Lang, Beatty, and Rice 2017)], including quorum sensing via the GtaI-GtaR proteins (Schaefer et al. 2002; Leung et al. 2012) and the CckA-ChpT-CtrA histidyl-aspartyl phosphorelay (Lang and Beatty 2000; Mercer et al. 2012). A transcriptomic study focused on CtrA showed that more than 20 genes encoding predicted signal transduction and transcription-regulating proteins were affected by the loss of this response regulator (Mercer et al. 2010). These included nine proteins predicted to be involved in c-di-GMP signaling, and four of these were subsequently shown to affect RcGTA production as well as flagellar motility by altering c-di-GMP levels (Pallegar et al. 2020). One of these four proteins, Rcc00620, possesses a response regulator REC domain and both GGDEF and EAL domains and was shown to act as a positive regulator of RcGTA production by acting as a PDE enzyme (Pallegar et al. 2020). However, this protein acted as a DGC in *Escherichia coli* (Pallegar et al. 2020). This differential activity in the two bacteria led us to speculate that the activity of one or both domains might be regulated by the phosphorylation status of the N-terminal REC domain and through the action of a cognate histidine kinase protein.

In this study, we used genetic manipulations, site-directed mutagenesis, and *in vitro* phosphorylation assays to determine if the c-di-GMP-modulating protein Rcc00620 acts as a response regulator in a TCS with a putative cognate histidine kinase, Rcc00621. We evaluated the role of the Rcc00620 REC domain and Rcc00621 in regulating the enzymatic activity of Rcc00620 in *R. capsulatus* by quantifying RcGTA production, cell motility, and c-di-GMP levels in relevant strains, and via *E. coli* c-di-GMP indicator assays. We also investigated the evolutionary history of the two genes and their protein motif conservation across alphaproteobacterial species. We show that these genes were horizontally acquired by an ancestral *Rhodobacter* from within the alphaproteobacterial order *Sphingomonadales* and encode a functional TCS where the c-di-GMP-related enzymatic activity of Rcc00620 is regulated through phosphorylation of its REC domain by the histidine kinase Rcc00621.

3.2 Methods and materials

3.2.1 Sequence and phylogenetic analyses

The SMART (Letunic 2004; Schultz et al. 1998) and Expasy-Prosite (de Castro et al. 2006) databases were used to identify functional domains in the protein sequences and TMpred (Hofmann and Stoffel 1993) and Phobius (Käll, Krogh, and Sonnhammer 2007) were used to predict transmembrane domains.

rcc00620 and *rcc00621* homologs were identified using the BLASTn (Altschul et al. 1990) online tool with adjusted scoring parameters (Match/Mismatch Scores: 1/-1; Gap Costs: Existence 2, Extension 1) and searches were performed within the "non-redundant" and "whole-genome shotgun contigs" (wgs) databases using the genomic sequence of *R. capsulatus* SB1003 that contains both genes (accession number

CP001312, nt 666094-669794) as query. Only hits that showed >89% continuous coverage were considered. For nucleotide sequences that lacked annotations, protein sequence predictions and annotations were performed in Geneious R11 (Biomatters). Accession numbers of sequences used in this study are available in the supplementary material (Appendix 2, Table S3.1, Figures S3.2, S3.3 and S3.4).

Phylogenetic analyses were performed using either 16S rDNA reference sequences obtained from the NCBI 16S RefSeq database (https://www.ncbi.nlm.nih.gov/refseq/targetedloci/16S_process/) or with sequences identified with BLASTn. Sequence alignments were produced with MAFFT (Katoh and Standley 2013), alignments were polished with TrimAl (Capella-Gutiérrez, Silla-Martínez, and Gabaldón 2009) through the online tool Phylemon (Sánchez et al. 2011), and trees were built with MEGA7 (S. Kumar, Stecher, and Tamura 2016) using the neighbor-joining (Saitou and Nei 1987) and maximum-likelihood (Felsenstein 1981) methods for large and small datasets, respectively. The best model for distance estimation, identified by a model test analysis, was used to calculate distances between sequences and a bootstrap analysis was performed to evaluate cluster robustness (Felsenstein 1985). Sequence identities (1–p-distance) were calculated with MEGA7.

3.2.2 Bacterial strains, media and growth conditions

All bacterial strains and plasmids used in this study are listed in Table 3.1. *R. capsulatus* was grown at 35 °C under either photoheterotrophic anaerobic conditions in YPS medium (Wall, Weaver, and Gest 1975) or aerobically in RCV medium (Beatty and Gest 1981), supplemented with appropriate antibiotics when required: kanamycin (10 μg ml^{-1}), gentamycin (3 μg ml^{-1}), and tetracycline (0.5 μg ml^{-1}). *E. coli* was grown in LB

medium supplemented with appropriate antibiotics when required: kanamycin (25 μg ml^{-1}), gentamycin (10 μg ml^{-1}), and tetracycline (10 μg ml^{-1}).

Table 3.1. List of bacterial strains and plasmids used in this study.

Strains and plasmids	Description	Reference or source
***R. capsulatus* strains**		
SB1003	Genome-sequenced strain	(Strnad et al. 2010; Yen and Marrs 1976)
DW5	SB1003 Δ*puhA*	(Wong et al. 1996)
SB620	SB1003 with 1068-bp deletion in *rcc00620* replaced by KIXX fragment	(Pallegar et al. 2020)
SB620.621	SB1003 with 3379-bp deletion in *rcc00620.rcc00621* replaced by spectinomycin resistance-encoding Omega fragment	This study
SB621	SB1003 with 2469-bp deletion in *rcc00621* replaced by spectinomycin resistance-encoding Omega fragment	This study
***E. coli* strains**		
NEB5α	Cloning strain	New England Biolabs
MG1655	Wild type; motility indicator strain for DGC activity	(Blattner et al. 1997)

MG1655 Δ*yhjH*	Motility indicator strain for PDE activity	(Ryjenkov et al. 2006; Simm et al. 2004)
BL21(DE3)	Curli fimbriae indicator strain for DGC activity	New England Biolabs; (Christen et al. 2006)
C600(pDPT51)	Plasmid-mobilizing strain	(Taylor et al. 1983)
S17-1	Plasmid-mobilizing strain	(Simon, Priefer, and Pühler 1983)
Plasmids		
pGEM-T Easy	TA PCR product cloning vector	Promega
pCM62	Broad host range vector; expression of genes in *E. coli* driven by *lac* promoter	(Marx and Lidstrom 2001)
pET-28a	Expression vector for purifying 6X-His-tagged proteins	Novagen (USA)
p620	*rcc00620* and 440 bp of 5' sequence in *Kpn*I site of pCM62	(Pallegar et al. 2020)
p620D86A	p620 with mutation in REC domain	This study
p620D86E	p620 with mutation in REC domain	This study
p621	*rcc00621* and 1049 bp of 5' sequence in *EcoR*I site of pCM62	This study
p621H361N	p621 with mutation in H-box of HK domain	This study

p620.621	*rcc00620.rcc00621* and 440 bp of 5' sequence in *Kpn*I site of pCM62	This study
pET-28a.620	Expression vector for C-terminal 6X-His-tagged Rcc00620	This study
pET-28a.621ΔTM	Expression vector for C-terminal 6X-His-tagged Rcc00621ΔTM	This study
pDGC	Heterologous diguanylate cyclase gene from *Rhodobacter sphaeroides* (RSP_3513) cloned into pCM62	(Pallegar et al. 2020)
pPDE	Heterologous phosphodiesterase gene from *Gluconacetobacter xylinus* (*pdeA1*) cloned into pCM62	(Pallegar et al. 2020)

3.2.3 Construction of gene disruptions and *trans*-complementation plasmids

The *rcc00621* gene was amplified by PCR using *R. capsulatus* SB1003 genomic DNA as template with corresponding primers (Appendix 2, Table S3.2). The amplified product was then cloned into pGEM-T Easy (Promega). Gene disruptions were made by insertion of the approximately 2-kb *Sma*I fragment of the spectinomycin resistance-encoding Omega fragment (Prentki and Krisch 1984) at specific restriction enzyme sites. *rcc00621* was disrupted by digestion with *Srf*I and *Msc*I, resulting in the deletion of 1222 bp from the gene. The double knockout mutant of both *rcc00620* and *rcc00621* was made by digestion with *Nru*I, resulting in a deletion of 3379 bp across the two genes. Mutation constructs were confirmed by restriction enzyme mapping and the disrupted genes were

transferred into the *R. capsulatus* SB1003 chromosome by RcGTA for generation of the mutant strains as described previously (Hynes and Lang 2013). Chromosomal gene disruptions were confirmed by PCR.

Trans-complementation of mutants was done with the relevant genes and their upstream regulatory regions, amplified using gene-specific primers (Appendix 2, Table S3.2), cloned into pCM62 (Marx and Lidstrom 2001). The amplified fragments were cloned into pCM62 using *Eco*RI for *rcc00621* and *Kpn*I for *rcc00620.rcc00621*. The knockout strains containing the empty vector were used as the respective reference strains.

3.2.4 Gene transfer bioassays and quantification of the RcGTA major capsid protein

A gene transfer bioassay quantifying the transfer of an essential photosynthesis gene, *puhA*, to a Δ*puhA* mutant strain (DW5) (Wong et al. 1996) was used to measure the production and release of RcGTA particles (Hynes and Lang 2013). Briefly, *R. capsulatus* strains were grown under photoheterotrophic conditions in YPS medium for approximately 48 hours, filtered using 0.45-µm PVDF syringe filters, and the filtrates assayed for RcGTA gene transfer activity. RcGTA activities were calculated as ratios relative to the parental wild type strain, SB1003, in three replicate experiments. Statistically significant differences in RcGTA gene transfer activities were identified using one-way analysis of variance (ANOVA) followed by Tukey HSD *post-hoc* analysis in R (Hesterberg, Chambers, and Hastie 1993). RcGTA major capsid protein levels were quantified within the cells and in the culture medium by performing western blots on samples from the same cultures that were used for gene transfer assays, as described previously (Pallegar et al. 2020).

3.2.5 Site-directed mutagenesis of Rcc00620 and Rcc00621

Substitutions in the REC domain of Rcc00620 and of the phosphoryl acceptor histidine in Rcc00621 were created by performing site-directed mutagenesis using the QuickChange Lightning Site-Directed Mutagenesis kit (Agilent Technologies) as per the manufacturer's instructions. Briefly, *rcc00620* cloned in pCM62 was used as a template for PCRs with *PfuUltra* High-Fidelity DNA polymerase and primers (Appendix 2, Table S3.2) designed to change the aspartic acid within the REC domain (position 86) to alanine (D86A) or glutamate (D86E) to mimic the permanently unphosphorylated or phosphorylated state, respectively (Scharf 2010). Similarly, *rcc00621* cloned in pCM62 was used as template for SDM PCRs using primers (Appendix 2, Table S3.2) to change the conserved histidine residue within the predicted H-box (position 361) to asparagine (H361N). The methylated template DNAs were then digested by incubation with *Dpn*I for 10 min at 37 °C and the remaining DNA was transformed into *E. coli*. Mutations were confirmed by sequencing and plasmids were subsequently conjugated into *R. capsulatus* using *E. coli* S-17 (Simon, Priefer, and Pühler 1983).

3.2.6 Expression and purification of recombinant proteins from *E. coli*

To create recombinant C-terminal 6X-His-tagged proteins, *rcc00620* and *rcc00621* were amplified using gene-specific primers (Appendix 2, Table S3.2) and cloned as *Nco*I/*Hind*III fragments into pET-28a (Novagen). *rcc00621* was amplified and cloned such that only the predicted soluble cytoplasmic portion of the protein would be present and the N-terminal region of the protein (283 amino acids) containing the two predicted trans-membrane domains was not included (Rcc00621ΔTM). The resulting plasmids were confirmed to be as expected by sequencing and transformed into *E. coli*

BL21(DE3) (New England Biolabs). For protein purification, overnight cultures were used to inoculate 200 ml of LB broth with appropriate antibiotics followed by incubation at 30 °C with shaking at 220 rpm. After one hour, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM to induce expression and the cultures were incubated at 30 °C for another 5-6 hours with shaking at 220 rpm. The cells were harvested by centrifugation (5000 x *g* for 10 minutes at 4 °C) and the pellets resuspended in lysis buffer [50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 0.1% Benzonase® nuclease (Qiagen), 1 mg ml^{-1} lysozyme; pH 8] and incubated on ice for 30 minutes. The suspension was then centrifuged at 10000 x *g* for 30 minutes at 4 °C. The supernatant was then collected and mixed 4:1 (v/v) with Ni-NTA agarose slurry (Qiagen) and incubated at 4 °C for 1 hour. After incubation, the mix was loaded into a polypropylene column and washed twice with wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole; pH 8) and the recombinant proteins were eluted in four 0.5-ml aliquots of elution buffer [50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole; pH 8]. The eluted proteins were analysed by SDS-PAGE followed by Coomassie Brilliant Blue staining. The appropriate eluted protein fractions were pooled and dialyzed against dialysis buffer (50 mM NaH_2PO_4, 300 mM NaCl; pH 8) and the final protein concentrations determined using the Bradford method (He 2011).

3.2.7 *In-vitro* phosphorylation assays

Phosphorylation assays were performed in phosphorylation reaction buffer [100 mM Tris-HCl, 50 mM KCl, 5 mM $MgCl_2$; pH 8]. Acetyl phosphate-dependent phosphorylation of Rcc00620 (5 µM) was tested in this buffer supplemented with 40 mM acetyl phosphate with incubation at 30 °C for 30 minutes. Phosphotransfer from Rcc00621 to Rcc00620 was tested by incubation of Rcc00621ΔTM (5 µM) in the reaction

buffer with 10 mM ATP at 30 °C for 30 minutes, followed by addition of Rcc00620 (5 µM). For both assays, aliquots were removed at T=0 and 60 minutes and 3X SDS-PAGE loading buffer was immediately added. Samples were boiled at 100 °C for 5 minutes and analyzed by electrophoresis on 8% PhosTag^{TM} (APExBIO) SDS-PAGE gels with Coomassie Brilliant Blue staining.

3.2.8 Assays for DGC and PDE enzymatic activities in *E. coli*

DGC and PDE enzymatic activities in *E. coli* were assayed by swimming and Congo red binding assays as described previously (Chen et al. 2014). Briefly, the pCM62-based clones of *rcc00620* and its site-directed mutants, *rcc00620.621*, and control DGC- and PDE-encoding genes (Table 1) were transformed into *E. coli* MG1655, MG1655 Δ*yhjH*, and BL21(DE3). The MG1655 strain is highly motile on semi-solid medium and a decrease in its swim zone due to the expression of a heterologous gene is taken as indication of DGC activity by the encoded protein (Chen et al. 2014). Similarly, the MG1655 Δ*yhjH* strain is non-motile on semi-solid medium and was used to test for evidence of PDE activity, which is indicated by an increase in swim zone diameter (Simm et al. 2004; Girgis et al. 2007). With the BL21(DE3) strain, an increase in the intracellular level of c-di-GMP causes an increase in fimbriae production that can be visualized by Congo red staining (Chen et al. 2014).

For swimming assays, 5 µl from each overnight culture was spotted on semisolid LB medium (0.25% agar) supplemented with 0.25 mM IPTG and 0.5% NaCl. The plates were incubated at 37 °C for 4 to 6 hours and photographed. Images were subsequently adjusted for brightness and contrast for better visibility. Congo red binding assays were done by streaking from overnight cultures onto LB agar plates containing 25 µg ml^{-1}

Congo red and 0.25 mM IPTG. The plates were incubated at 28 °C for 48 hours, photographed and adjusted for brightness and contrast for better visibility.

3.2.9 *R. capsulatus* motility assays

Aerobically grown overnight cultures were used to inoculate YPS agar (0.3%) stabs, which were then incubated at 35 °C under phototrophic conditions for 10-12 hours and photographed. The pictures were adjusted for brightness and contrast and the diameters of the zones of growth were measured using ImageJ (Schneider, Rasband, and Eliceiri 2012). Statistically significant differences in the swim zones were identified by one-way ANOVA followed by Tukey HSD *post-hoc* tests.

3.2.10 Quantification of c-di-GMP

The quantification of c-di-GMP levels in cells was performed as described previously (Pallegar et al. 2020). Briefly, aerobically grown overnight cultures of the different *R. capsulatus* strains were normalized for cell density and used to inoculate anaerobic photoheterotrophic cultures. These cultures were grown for approximately 48 hours and 3 ml of each culture was removed and the cells pelleted by centrifugation at 4 °C. The cells were boiled in phosphate-buffered saline (PBS) and subjected to three rounds of extraction with 65% ethanol. The three extraction supernatant fractions were pooled, dried and resuspended in ultrapure water for quantification of c-di-GMP by high-performance liquid chromatography (HPLC). Culture samples were prepared in triplicate and each replicate was quantified from triplicate injections. A standard curve was made using solutions of c-di-GMP (BIOLOG) that were measured from triplicate injections. The cell pellets were resuspended in 500 µl TE buffer followed by sonication on ice in 20-second bursts for a total of 2 minutes. The protein concentrations were determined

using the Bradford assay (He 2011). The concentration of c-di-GMP was then normalized to protein levels. One-way ANOVA followed by Tukey HSD *post-hoc* analysis was used to identify statistically significant differences in c-di-GMP levels.

3.3 Results

3.3.1 Rcc00620 and Rcc00621 functional domains

The previously studied Rcc00620 protein is encoded in a two-gene operon with a gene (*rcc00621*) predicted to encode a histidine kinase protein (Figure 3.1A). We performed a detailed analysis of the conserved domains in the two proteins (Figure 3.1A). Rcc00620 contains an N-terminal REC domain, a central GGDEF (DGC) domain, and a C-terminal EAL (PDE) domain. For Rcc00621, all typical histidine kinase domains were identified: the HAMP domain, the H-box that contains the histidine autophosphorylation site (Kim and Forst 2001; Grebe and Stock 1999), and the ATP-binding catalytic domain (HATPase_C), as well as a periplasmic CHASE4 sensory domain located between two predicted transmembrane segments. The HAMP domain is a cytoplasmic helical linker domain in histidine kinase and methyl-accepting chemotaxis proteins and the CHASE4 domain is commonly found in various classes of transmembrane receptors that are part of signal transduction pathways (Zhulin, Nikolskaya, and Galperin 2003).

Characteristic sequence motifs within the domains were identified and analyzed using full-length alignments obtained with all identified Rcc00620 and Rcc00621 homologs (N=205; a subset is shown in Figure 3.1B). The REC (including the phosphoryl acceptor aspartic acid, D), GGDEF, and EAL domains in the response regulator proteins were highly conserved among the homologs (Figure 3.1B). Within the histidine kinase proteins, the H-box contained a highly conserved sequence (HNxRNxLxP), which showed characteristics of both type I and type II kinases, as defined by Kim and Forst

(Kim and Forst 2001). Specifically, considering the histidine as position 1 in the H-box, the presence of a positively charged amino acid (arginine, R) at position 4 would be indicative of a type I kinase, but the lack of the type I-associated proline (P) at position 6 and the presence of an asparagine (N) at position 5 are typical of a type II kinase. Within the HTPase_C catalytic domain, we identified orthodox N-box (NLxxNA), D/G-box (D/G1-box: DxGxG), F-box (FxxG), and G-box (G4-box: GxGxG) motifs, and two potential additional G-boxes (G2-box, G3-box) (Kim and Forst 2001; Grebe and Stock 1999).

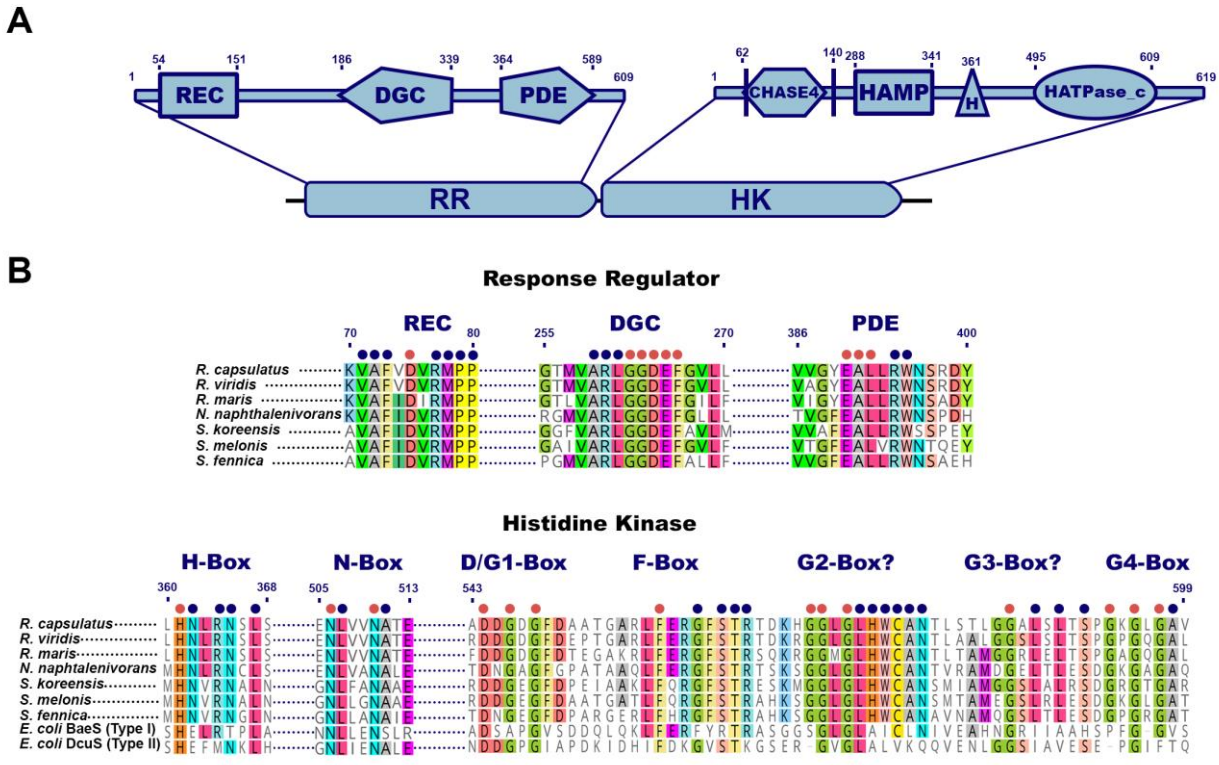


Figure 3.1. Domain architectures of the Rcc00620 and Rcc00621 proteins of *R. capsulatus* and their homologs in α-proteobacteria. A. The organization of the main domains in the Rcc00620 response regulator (RR) (REC: receiver; DGC: diguanylate cyclase; PDE: phosphodiesterase) and Rcc00621 histidine kinase (HK) (CHASE4:

sensory domain; HAMP: cytoplasmic helical linker domain; H: H-box; HATPase_c: ATP-binding catalytic domain) proteins. The numbers on the top indicate the amino acid positions. B. Alignments showing the conservation of functional domains in the proteins. All of the individual component of the HATPase_C domain are shown for the histidine kinases (the canonical N-Box, D/G1 Box, F-Box, G4-Box and two other putative G-Boxes indicated by "?"). The alignments include the three *Rhodobacter* species and the sole *Novosphingobium* species known to contain the genes, plus additional representative members of the order *Sphingomonadales.* The histidine kinase alignments also include prototypical type I and II (Kim and Forst 2001) representatives from *E. coli.* Amino acids conserved in >199 of the 205 total analyzed sequences (Appendix 2, Table S3.1) are indicated with a dot, with residues that characterize the motifs indicated with red dots.

3.3.2 Both *rcc00620* and *rcc00621* are required for normal RcGTA production

Since it is known from a previous study that loss of *rcc00620* results in decreased RcGTA gene transfer activity (Pallegar et al. 2020), we disrupted *rcc00621* and performed RcGTA gene transfer bioassays to determine if a similar phenotype resulted and potentially indicating the proteins encoded by these two genes work in conjunction to regulate RcGTA. Disruption of *rcc00621* showed a similar phenotype, with decreased RcGTA activity compared to the parental strain (Figure 3.2A). Similar to the individual mutants, a double knockout strain (SB620.621) also showed a decrease in RcGTA activity (Figure 3.2B). *Trans*-complementation of the SB621 and SB620.621 strains with plasmids carrying the respective genes reversed the effects of the mutations to wild type levels or higher (Figure 3.2AB). The presence of the plasmid-borne *rcc00621* and *rcc00620-621* in the parental strain also showed opposite effects compared to the

respective mutants, with increased RcGTA activity (Figure 3.2A,B). Quantification of the amounts of RcGTA major capsid protein, within the cells and released into the extracellular environment, matched the patterns observed for the RcGTA activities of the tested strains (Appendix 2, Figure S3.1).

3.3.3 Rcc00620 phosphorylation is required for its PDE activity

We wanted to determine if the enzymatic activity of Rcc00620 was regulated through its REC domain. To test this, we investigated the effects of different phosphorylation states of Rcc00620 on RcGTA production by creating point mutations at the predicted phosphorylation site in the REC domain (aspartic acid, D86). The D86 residue was changed to alanine (D86A) and glutamate (D86E) to mimic the unphosphorylated and phosphorylated states, respectively (Scharf 2010). Plasmids carrying these mutants were introduced into the SB620.621 double mutant strain and RcGTA activity was assayed (Figure 3.2C). The unphosphorylated state mimic

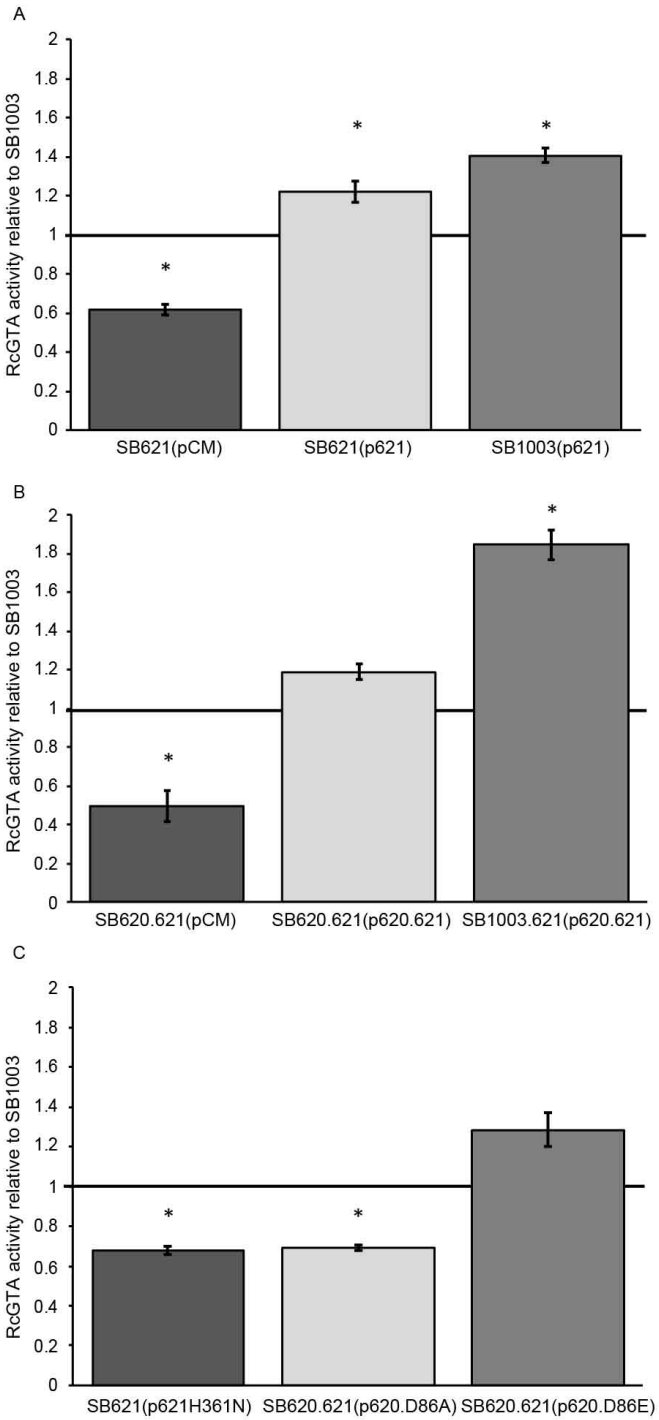


Figure 3.2. Effects of gene disruptions, *trans*-complementation, and site-directed mutagenesis of functional domains on RcGTA gene transfer activity. A gene transfer bioassay quantifying the transfer of *puhA*, an essential photosynthesis gene, to a Δ*puhA* mutant strain was used to measure the production and release of RcGTA particles. A. The

gene transfer activities of the *rcc00621* null mutant strain SB621(pCM), its complement SB621(p621), and the parental strain carrying *rcc00621* on a plasmid SB1003(p621). B. Gene transfer activities in the *rcc00620-621* double knockout null mutant SB620.621(pCM), its complement SB620.621(p620.621), and the parental strain carrying both genes on a plasmid SB1003(p620.621). C. *Trans*-complementation in the double mutant with the Rcc00620 REC domain site-directed mutants, SB620.621(p620.D86A), SB620.621(p620.D86E) and of the SB621 strain with the Rcc00621 kinase active site histidine site-directed mutant SB621(p621H361N). The data are the average values from 3 replicates, relative to the parental strain, SB1003 carrying the empty vector, pCM62. Error bars represent the standard deviations, and statistically significant differences ($p<0.05$) compared to the control, identified using one-way ANOVA followed by Tukey HSD *post-hoc* analysis, are indicated by asterisks.

resulted in a decrease in RcGTA activity similar to the knockout strain, whereas the phosphomimetic version reversed the effect of the loss of both genes (Figure 3.2C). These results indicate that it is the phosphorylated version of Rcc00620 that is active as a positive regulator of RcGTA production, presumably as a consequence of its PDE activity.

We also made a substitution at the predicted autophosphorylation site (histidine, H361) within the H-box of Rcc00621 and changed this residue to an asparagine (H361N) to prevent autophosphorylation. The presence of this mutated version in the SB621 mutant strain did not complement the reduction in RcGTA activity (Figure 3.2C), unlike the presence of the native *rcc00621*. This indicates that this predicted

autophosphorylation site in Rcc00621 is essential for its activity with respect to RcGTA regulation, which is presumably mediated via the phosphorylation of Rcc00620.

3.3.4 Rcc00621 phosphorylates Rcc00620 *in vitro*

To confirm that Rcc00620 can be phosphorylated on its REC domain, we performed an *in vitro* phosphorylation assay using acetyl phosphate as a phosphodonor (Lukat et al. 1992; Kinoshita-Kikuta et al. 2015). We tested if this occurred with purified Rcc00620 using PhosTag^{TM} SDS-PAGE, where phosphorylated proteins migrate more slowly than non-phosphorylated versions (Barbieri and Stock 2008). Incubation of Rcc00620 with acetyl phosphate for 60 mins resulted in the appearance of an additional, more slowly migrating band representing phosphorylated Rcc00620 (Figure 3.3). We then performed an *in vitro* phosphorylation assay using the purified Rcc00620 and Rcc00621ΔTM proteins to determine whether Rcc00621 can phosphorylate Rcc00620. Incubation of Rcc00621ΔTM with ATP followed by addition of Rcc00620 and continued incubation showed that Rcc00620 was phosphorylated (Figure 3.3).

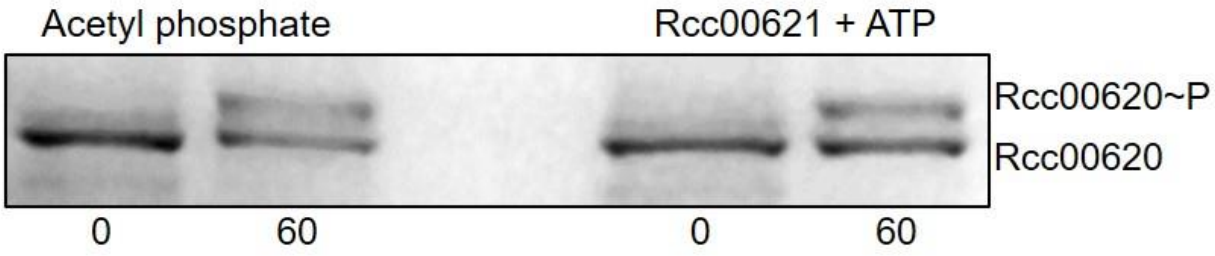


Figure 3.3. *In vitro* phosphorylation assays. A. The purified 6X-His-Rcc00620 protein was incubated at 30 °C with 40 mM acetyl phosphate with samples collected from the reaction mix at T=0 and 60 minutes. B. 6X-His-Rcc00621ΔTM was incubated with ATP for 30 minutes at 30 °C, followed by addition of 6X-His-Rcc00620 and incubation at 30 °C for 60 minutes with samples collected from this reaction mix at T=0 and 60 minutes.

Samples were analyzed by 8% PhosTag™ SDS-PAGE followed by Coomassie Brilliant Blue staining.

3.3.5 Assessment of Rcc00620-Rcc00621 activity via *Escherichia coli* phenotypic assays

In order to indirectly assess the potential DGC and PDE activities associated with the Rcc00620-621 TCS, we used three c-di-GMP-sensitive *E. coli* phenotypic assays to evaluate the role of the Rcc00620 REC domain's enzymatic activity. These assays are commonly used to indirectly assess the c-di-GMP-related enzymatic activities of proteins from different bacteria (Christen et al. 2006; Chen et al. 2014). C-di-GMP inhibits motility of *E. coli* MG1655 on semi-solid medium such that a reduction of the swim zone when the strain expresses a foreign gene is taken as evidence that the gene encodes an enzyme with DGC activity. DGC activity was also detected by performing a second assay, where an increase in c-di-GMP levels in *E. coli* BL21(DE3) leads to an increase in fimbriae production and increased Congo Red staining. Similarly, *E. coli* MG1655 Δ*yhjH*, which lacks a key PDE (Simm et al. 2004) and is non-motile on semi-solid medium (Girgis et al. 2007), was used to test for evidence of PDE activity as indicated by an increase in the swim zone diameter. The native and D86A mutant versions of Rcc00620 reduced the swim zone of

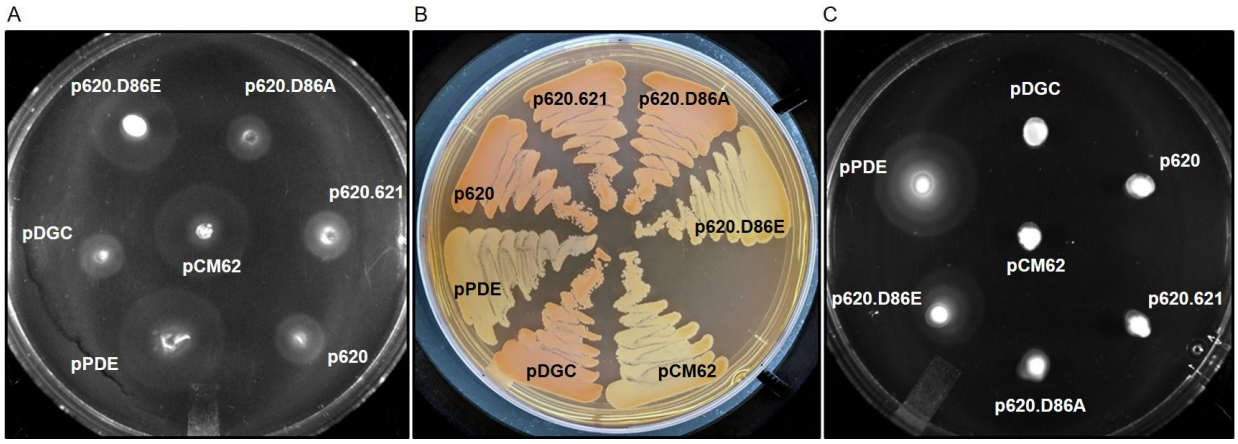


Figure 3.4. Evaluating enzymatic activities of *R. capsulatus* proteins in *E. coli*. A. Motility of *E. coli* MG1655 on semi-solid medium, which is reduced by DGC activity, when containing the indicated plasmids. B. Congo Red binding by *E. coli* BL21(DE3), where DGC activity increases fimbriae production and Congo red binding, when containing the indicated plasmids. C. Motility of *E. coli* MG1655 Δ*yhjH* on semi-solid agar, which is increased by PDE activity, when containing the indicated plasmids. In all experiments, the pCM62 plasmid vector is used and the transcription of the genes from the plasmid's *lac* promoter was induced with IPTG.

strain MG1655 (Figure 3.4A) and increased Congo red staining by strain BL21(DE3) (Figure 3.4B), indicating an increase in intracellular c-di-GMP levels that is presumably due to DGC activities of the proteins. In contrast, the phosphomimetic D86E version did not show any evidence of DGC activity in these assays (Figure 3.4A,B) but showed an increase in the swim zone in the MG1655 Δ*yhjH* motility assay (Figure 3.4C), indicating a decrease in c-di-GMP levels that is presumably due to PDE activity of this protein. The presence of both *rcc00620* and *rcc00621* produced the same results as when only

rcc00620 or its D86A mutant was present (Figure 3.4A,B,C), indicating the presence of Rcc00621 was not able to make Rcc00620 act as a PDE in *E. coli*.

3.3.6 The Rcc00620-621 TCS affects *R. capsulatus* flagellar motility and intracellular c-di-GMP levels

As it is known that increased intracellular c-di-GMP levels decrease *R. capsulatus* flagellar motility (Römling, Galperin, and Gomelsky 2013), we performed flagellar motility assays in tubes containing semi-solid agar and compared the swim zones of various mutants to evaluate the role of the Rcc00620-621 TCS in this behavior. The strains where *rcc00620*, *rcc00621*, or both genes were disrupted showed decreased motility compared to the parental strain (Figure 3.5), although the difference was not significant for SB621 ($p=0.08$). The Rcc00620 D86E mutant restored motility to the double knockout whereas the D86A mutant did not.

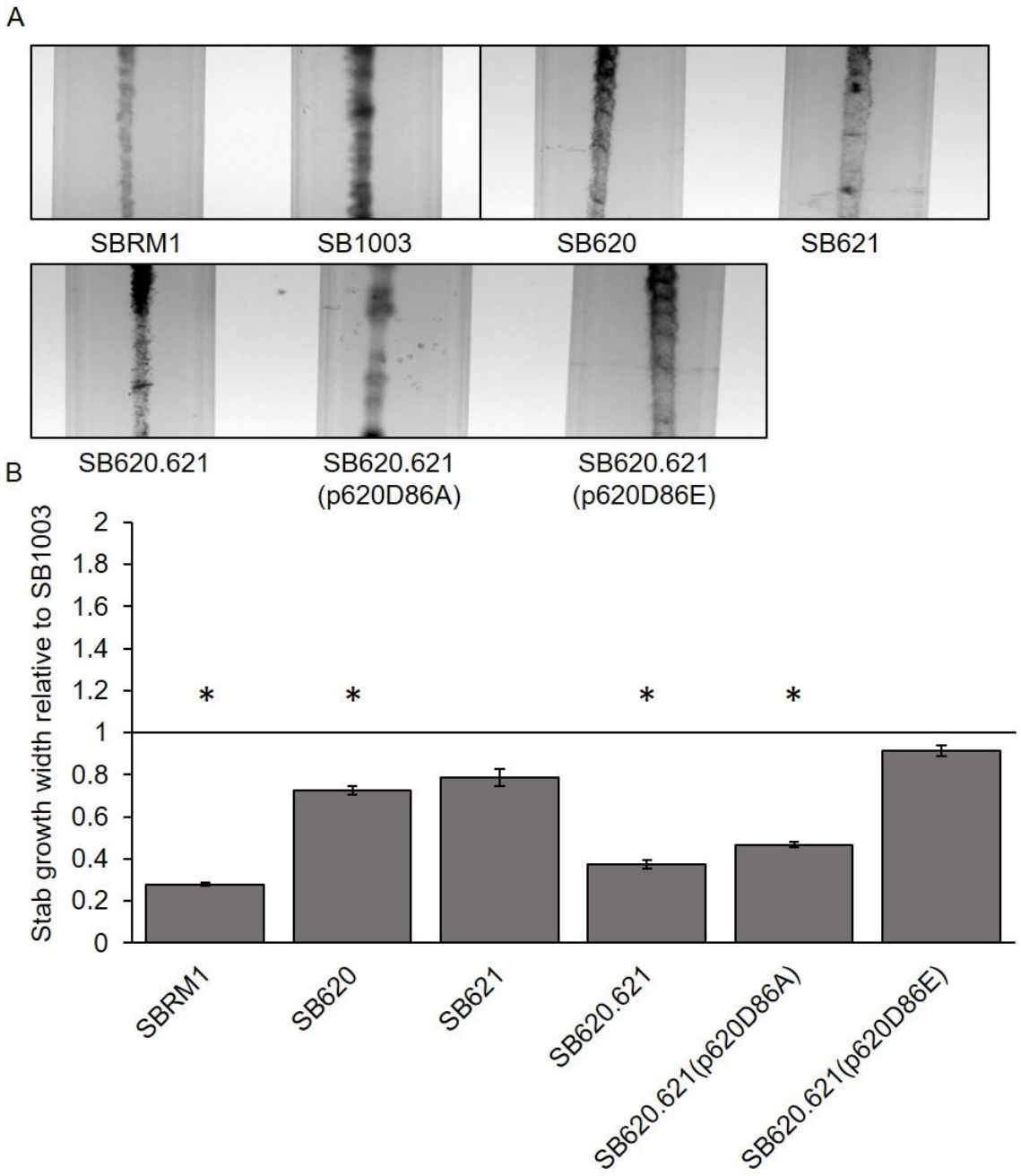


Figure 3.5. Role of the Rcc00620-Rcc00621 TCS in *R. capsulatus* flagellar motility. A. Representative motility assay results for the *rcc00620* null mutant strain (SB620), *rcc00621* null mutant strain (SB621), *rcc00620-621* double knockout null mutant (SB620.621), and *trans*-complementation in the double mutant with the Rcc00620 REC domain site-directed mutants SB620.621(p620.D86A) and SB620.621(p620.D86E). The

cells are inoculated into the tubes containing soft agar from a liquid culture and the subsequent distance of growth away from the center of the stab line shows the relative flagellar motility of the strain. The parental strain, SB1003, and its non-motile *ctrA* null mutant derivative, SBRM1, are included as references. B. The swimming diameters were measured from three replicate assays and plotted relative to SB1003. The bars represent the standard deviations and statistically significant differences ($p<0.05$) compared to the control, identified using one-way ANOVA followed by Tukey HSD *post-hoc* analysis, are indicated by asterisks.

We also evaluated the role of Rcc00620-621 TCS in contributing to the intracellular levels of c-di-GMP by quantifying c-di-GMP within the different strains. Loss of both genes resulted in significantly higher levels of c-di-GMP (Figure 3.6). Although the differences were not statistically significant, the trends for the D86A and D86E mutants in the double knockout strain (Figure 3.6) were as expected based on all previous results that indicated phosphorylation at this site causes the Rcc00620 protein to be active as a PDE.

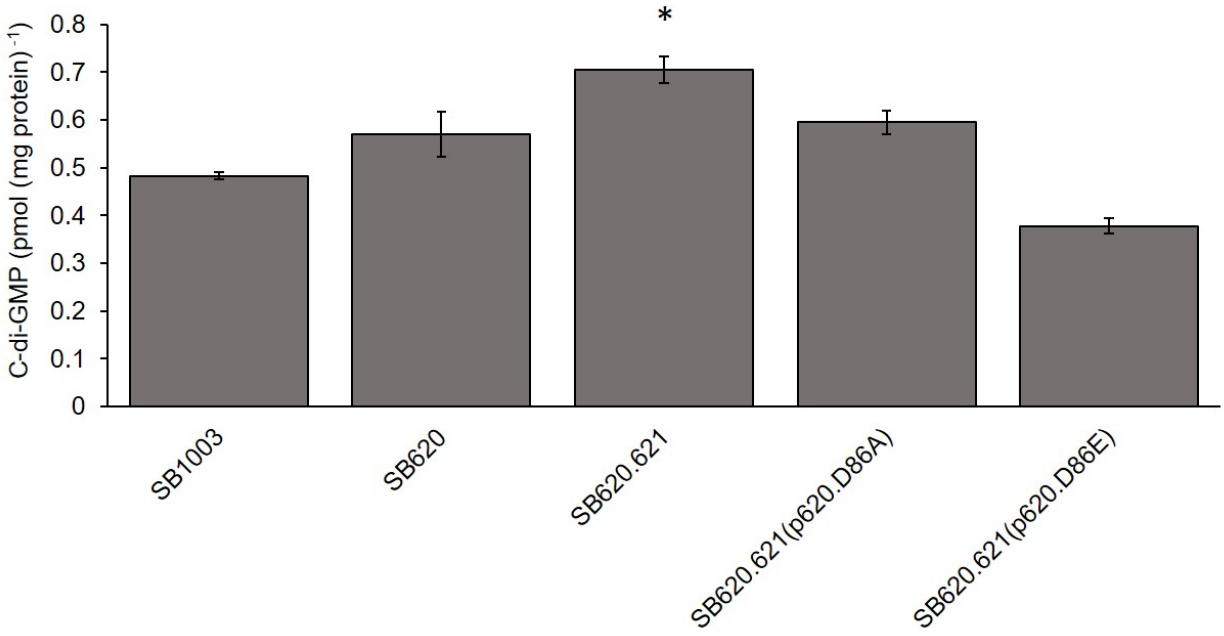


Figure 3.6. Quantification of intracellular c-di-GMP levels in *R. capsulatus* strains. C-di-GMP levels for *rcc00620* null mutant (SB620), *rcc00620-621* double knockout null mutant (SB620.621), and SB620.621 *trans*-complemented with the Rcc00620 REC domain site-directed mutants, SB620.621(p620.D86A) and SB620.621(p620.D86E) were measured by HPLC and normalized to the protein content of the cells. Quantification was performed from 3 replicates with the bars representing the standard deviations and the strain with a statistically significant difference ($p<0.05$) compared to the control, identified using one-way ANOVA followed by Tukey HSD *post-hoc* analysis, is indicated by an asterisk.

3.3.7 The *rcc00620* and *rcc00621* genes were acquired by horizontal gene transfer

To evaluate the importance of this Rcc00620-621 TCS throughout the evolution of the class Alphaproteobacteria, we performed a BLAST search using the genomic region of *R. capsulatus* SB1003 containing both *rcc00620* and *rcc00621* (accession number CP001312, nt 666094-669794) to identify genomes containing contiguous homologs of both genes. Within the 82 complete or almost complete *Rhodobacter* genomes, homologs were found in all sequenced strains belonging to three species, *R. capsulatus* (N=13), *R. maris* (N=1), and *R. viridis* (N=1), but not in the other 10 species and 28 strains without species designations for which the complete genomes were available (Table S3.1). Since most of the complete genome projects were identified within the whole-genome shotgun contigs database, we tested whether the undetectability of homologs was the result of low sequencing coverage by evaluating the presence of 6 other genetic markers in these genomes. Most genomes contained all or most of the investigated markers, which included *rpoB* (DNA dependent RNA polymerase subunit beta), *gyrB* (DNA gyrase subunit B), and four ORFs flanking *620/621* in *R. capsulatus* (Figure 3.7A; see below) (Table S1). This strengthened the conclusion that these genes were not present in the genomes of most species within the genus *Rhodobacter*.

While the genes were positioned at the same genomic location in *R. capsulatus* and *R. viridis*, between genes for an ATP-binding cassette (ABC) transporter protein and a molybdopterin biosynthesis protein (MoeA), they were located at a different position in *R. maris*, between genes for glycerate 2-kinase (GckA) and conserved hypothetical proteins (Figure 3.7A).

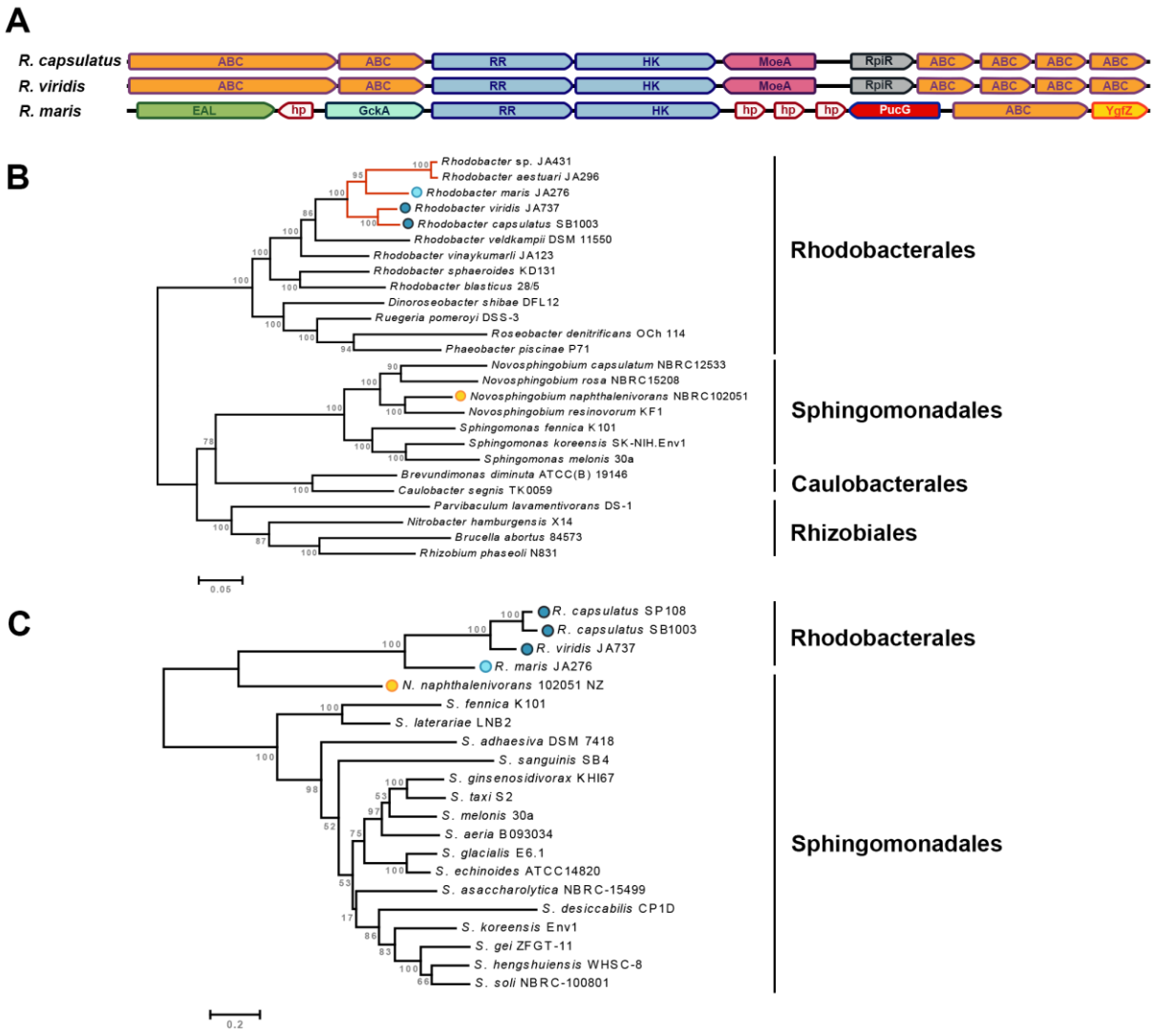


Figure 3.7. Evolutionary history of *rcc00620* and *rcc00621*. A. Genomic contexts of the genes within *Rhodobacter* spp. The *rcc00620.621* genes are indicated in blue, while the various neighboring ORFs are indicated in different colors corresponding to the predicted/annotated encoded proteins (ABC: ATP-binding cassette transporter proteins; MoeA: molybdopterin biosynthesis protein; RpiR: RpiR family transcriptional regulator; EAL: EAL domain-containing protein; GckA: glycerate 2-kinase; PucG: alanine-glyoxylate aminotransferase family protein; YgfZ: folate-binding protein; hp: hypothetical protein). B. Maximum-likelihood phylogenetic analysis (GTR +G +I model)

of representative members of the class Alphaproteobacteria based on *rpoB*, which encodes the DNA-dependent RNA polymerase subunit beta. Branches of the *Rhodobacter* clade derived from the ancestor that was likely the receiver of the genes are colored in orange. C. Maximum-likelihood phylogenetic analysis (JTT +G model) of the concatenated response regulator and histidine kinase protein sequences of representative members of the Alphaproteobacteria that contain the genes. In panels B and C, the *Rhodobacter* spp. that contain the genes are indicated by blue dots (the shade of blue corresponds to the different genomic localizations, as shown in panel A), while *Novosphingobium naphthalenivorans*, which contains the most closely related homologs of the *Rhodobacter* genes, is indicated by a yellow dot.

This suggested that these genes could be a movable element that was either gained or lost during the evolution of *Rhodobacter* species since they diverged from a common ancestor. To explore this possibility, the relationships among *Rhodobacter* strains were studied using three genetic markers commonly used for phylogenetic inference in bacteria (16S rDNA, *rpoB*, and *gyrB*; Appendix 2, Figure S3.2) (Ogier et al. 2019; Liu et al. 2012). In all cases, *R. capsulatus*, *R. viridis* and *R. maris* were included within one highly supported clade, suggesting common ancestry (summarized in Figure 3.7B). Matching what we observed for genomic structure (Figure 3.7A), *R. capsulatus* and *R. viridis* were phylogenetically more closely related, while *R. maris* was in a different sub-clade with two other *Rhodobacter* species, *R. aestuarii* and another yet-unnamed strain (*Rhodobacter* sp. JA431), in which the TCS genes were not found. There were three additional *Rhodobacter* species for which only 16S sequences were available that were also included in this clade (Appendix 2, Figure S3.2). Two were close relatives of *R.*

viridis (*R. sediminis* and *R. azollae*) and one of *R. maris* (*R. lacus*) and it is possible they also possess these genes, but genomic sequence data are not currently available for these species.

To identify whether other bacteria within the class Alphaproteobacteria contained these genes, a further BLASTn analysis was performed. This led to the identification of homologs in 190 sequenced strains, all belonging to the order Sphingomonadales (accession numbers provided in Appendix 2, Table S3.1). Phylogenetic analyses performed with all identified homologs revealed that the closest relatives to the *Rhodobacter* proteins were those found in *Novosphingobium naphthalenivorans*, a naphthalene-degrading bacterium (Zuzuki and Hiraishi 2007) (Figure 3.7C; Appendix 2, Figures S3.3 and S3.4). Interestingly, among the 95 *Novosphingobium* complete genome projects that contained a complete *rpoB*, only this species contained these genes.

As shown in the phylogenetic analysis of *rpoB* sequences from representative members of the class Alphaproteobacteria (Figure 3.7B), the two genera *Rhodobacter* (order Rhodobacterales) and *Novosphingobium* (order Sphingomonadales) are clearly not monophyletic, but their versions of these TCS proteins are closely related (Figure 3.7C). Indeed, on average, the Rcc00620 and Rcc00621 homologs from within the genus *Rhodobacter* were 54.3% and 39.5% identical to those of *N. naphthalenivorans* and 48.1% and 32.8% identical to those of other Sphingomonadales, respectively. Interestingly, the *N. naphthalenivorans* protein sequences were only as similar to those of other Sphingomonadales (52.6% and 33%, respectively) as they were to the ones from *Rhodobacter*.

3.4 Discussion

3.4.1 Rcc00620 and Rcc00621 form a two-component system involved in the regulation of c-di-GMP levels, RcGTA production, and flagellar motility

In *R. capsulatus*, Rcc00620 and Rcc00621 contain all of the conserved domains and residues required to act as TCS response regulator and histidine kinase, respectively (Figure 3.1). These domains in both proteins have been well conserved throughout evolution, suggesting this is an important regulatory system in multiple species and genera within the class Alphaproteobacteria. The disruption of *rcc00621* and both *rcc00620-621* resulted in decreased gene transfer activity (Figure 3.2), with the changes in gene transfer activity accompanied by corresponding changes in production and release of the RcGTA capsid protein (Appendix 2, Figure S3.1), as reported previously for *rcc00620* (Pallegar et al. 2020). Therefore, these two genes are both involved in positively regulating RcGTA production. Mutations of the predicted phosphoryl receptor residue in the REC domain of Rcc00620, made to mimic the different phosphorylation states, showed this site and its phosphorylation status modulate the enzymatic activity of Rcc00620. The REC-GGDEF-EAL domain architecture is one of the most commonly found in c-di-GMP-metabolizing proteins (Romling, Galperin, and Gomelsky 2013), and approximately 8% of all GGDEF/EAL proteins also contain a REC domain (Seshasayee, Fraser, and Luscombe 2010). Our data show that phosphorylation of Rcc00620 makes it act as a PDE in *R. capsulatus*, which thereby positively affects RcGTA production and motility (Figures 3.2, 3.5 and 3.6). Surprisingly, none of the assays show any evidence that Rcc00620 acts as a DGC in *R. capsulatus* in the unphosphorylated form as the D86A mutant gave phenotypes undistinguishable from the null mutant (Figures 3.2, 3.5 and 3.6). This contrasts with the results from *E. coli*, where the unphosphorylated protein

appears to function as DGC because the assays showed it resulted in lower c-di-GMP levels in the cells (Figure 3.4). It is possible that under different conditions, factors other than its phosphorylation state could cause Rcc00620 to act as a DGC. Although the D86E version did act as a PDE in *E. coli*, the presence of both *rcc00620* and *rcc00621* only yielded DGC activity in this bacterium. This suggests the Rcc00621 histidine kinase is not active for Rcc00620 phosphorylation in *E. coli*, at least under the conditions used in these experiments. This is could be due to lack of the appropriate stimulus or other factor(s) required to activate its kinase activity. Unfortunately, there does not appear to be any information available about the signal sensed for any other CHASE4 domain and so this represents a need for future research. It is also possible that another factor, such as improper folding, insolubility, or a lack of insertion into the cytoplasmic membrane, prevented Rcc00621 from being active in *E. coli*.

As mentioned above, there are many known examples of RR proteins possessing DGC and PDE domains. However, to the best of our knowledge, only one other RR protein containing both DGC and PDE domains has been characterized with respect to its regulation (Levet-Paulo et al. 2011). In that system, phosphorylation of the REC domain causes a switch from DGC to PDE activity (Levet-Paulo et al. 2011). This is similar to our Rcc00620-621 TCS, where the Rcc00621-mediated phosphorylation of the Rcc00620 REC domain results in PDE activity and a reduction of intracellular c-di-GMP levels.

We were able to identify a typical H-box for Rcc00621, which contains the conserved histidine residue that is the site of autophosphorylation (Figure 3.1). Substitution of this residue (H361N) prevented *trans*-complementation of the null mutant (Figure 3.2) and incubation of the purified native protein with Rcc00620 and ATP

resulted in phosphorylation of Rcc00620 (Figure 3.3). Overall, the phenotypic and *in vitro* data validate our initial hypothesis about the two proteins acting as a TCS.

C-di-GMP plays a vital role in bacterial signaling and controls diverse physiological processes by binding to various targets (Romling, Galperin, and Gomelsky 2013). Besides RcGTA production, flagellar motility is also regulated by c-di-GMP levels in *R. capsulatus* (Pallegar et al. 2020), and we show here that both members of this Rcc00620-621 TCS affect this process (Figure 3.5), as expected from previous results with Rcc00620. Similarly, the trends for the intracellular levels of c-di-GMP (Figure 3.6) matched expectations based on the other experiments.

3.4.2 The *rcc00620* and *rcc00621* genes were acquired by horizontal gene transfer

While homologs of the *R. capsulatus rcc00620/621* genes were absent from most *Rhodobacter* species, they were identified within the genomes of two additional closely related species (*R. viridis* and *R. maris*; Figure 3.7A). Interestingly, these three bacteria are part of a clade that also includes two other species, *R. aestuarii* and another unnamed strain or species, that seem to be lacking the genes (Figure 3.7B) and three other species (*R. sediminis*, *R. azollae*, and *R. lacus*; Appendix 2, Figure S3.2) for which sufficient genomic sequence data were not available to conclude whether they contain homologs or not. Since most species within the order Rhodobacterales lack these genes, we can hypothesize that they were acquired during the evolution of *Rhodobacter* by an ancestor of the species within the one clade and subsequently lost by *R. aestuarii* and related species. However, the genomic sequences available for *R. aestuarii* and the related species were identified in the whole-genome shotgun contigs database and it is possible that genomic regions containing these genes were not sequenced or not included in the available scaffolds.

Within the class Alphaproteobacteria we found Rcc00620 and Rcc00621 homologs encoded in the genomes of 190 bacteria within the order Sphingomonadales, most of which belonged to the genus *Sphingomonas*. They were also found in one of the 95 complete genomes available from bacteria in the genus *Novosphingobium* within the Sphingomonadales, and our data indicate that the genes found in this species (*N. naphthalenivorans*) are the closest known relatives to that of *Rhodobacter* (Figure 3.7B). Since these two genera are clearly not monophyletic (Figure 3.7C), as they belong to different orders, the fact that their genes are monophyletic reflects an unexpected evolutionary link. The genes were found in multiple *Sphinogmonas* spp. but only once in *Novosphingobium* and the genes found in *N. naphthalenivorans* were phylogenetically located between those from *Rhodobacter* and those from the other Sphingonomadales. We postulate therefore that the genes could have been acquired at a certain point by an ancestral *Rhodobacter* strain from an ancestral *Novosphingobium,* which in turn had acquired them from an ancestral *Sphingomonas,* or that they both acquired them in a similar time-frame from a similar source, and the genes then diverged over time. Interestingly, the genomic locations in which the genes are found varies considerably among members of both the *Rhodobacter* (Figure 3.7) and *Sphingomonas* (data not shown) genera, suggesting these genes move around together as lineages diverge. Unfortunately, only a limited number of genomes from members of the genus *Rhodobacter* closely related to *R. capsulatus* are available and further complete genome sequencing efforts are required to fully elucidate this aspect. Although further studies will be required to experimentally confirm whether GTA-mediated horizontal gene transfer can occur between these distantly related bacteria, it is possible that the transfer of the genes was GTA-mediated because members of the order *Sphingomonadales* possess the

genes to produce GTAs (Shakya, Soucy, and Zhaxybayeva 2017; Lang and Beatty 2007; Lang, Zhaxybayeva, and Beatty 2012; Biers et al. 2008).

This TCS is just one of multiple regulatory systems that contribute to the control of RcGTA production (Lang, Westbye, and Beatty 2017). Some of these other systems, such as quorum sensing and the CckA-ChpT-CtrA phosphorelay, also regulate GTA production in another bacterium, *Dinoroseobacter shibae* (Koppenhöfer et al. 2019; Tomasch et al. 2018; Wang et al. 2014). *D. shibae* lacks this particular TCS, but c-di-GMP does affect expression of its GTA genes (Koppenhöfer et al. 2019).

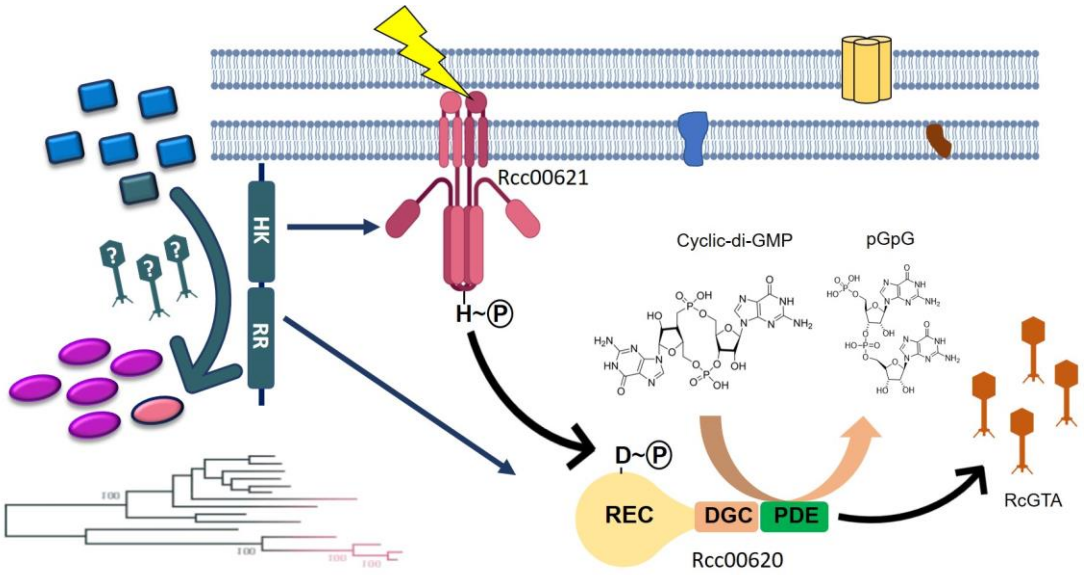


Figure 3.8. Proposed model for the Rcc00620-621 TCS controlling c-di-GMP synthesis in *R. capsulatus*. The HK Rcc00621 gets activated upon sensing an unknown stimulus and undergoes autophosphorylation at a conserved histidine residue and transfers the phosphoryl group to a conserved aspartate residue of its cognate RR protein, Rcc00620. The phosphorylated form of Rcc00620 acts as PDE and promotes c-di-GMP hydrolysis. The decrease in c-di-GMP levels promotes RcGTA production and motility. The TCS-encoding genes are only present in a few phylogenetically related *Rhodobacter*

species and they appear to have been acquired horizontally by an ancestral *Rhodobacter* from a member of the *Sphingomonadales*.

3.5 Concluding remarks

In this study, we delineated the role of a TCS involved in the regulation of c-di-GMP levels in *R. capsulatus*, and the consequent effects on gene transfer and flagellar motility. The Rcc00620 protein possesses a REC-DGC-PDE multidomain architecture, which is one of the most common domain architectures among c-di-GMP signaling proteins. Here, by assaying RcGTA production, motility and c-di-GMP levels in *R. capsulatus* and using c-di-GMP indicator assays in *E. coli*, we show that the enzymatic activity of Rcc00620 is modulated by phosphorylation of its REC domain by the cognate histidine kinase Rcc00621. Very few potentially bifunctional proteins have been characterized to date so this study adds important new information about how such proteins can be regulated, in this case via phosphorylation as part of a TCS. Intriguingly, this TCS that ultimately affects gene transfer activity in *R. capsulatus* seems to have been horizontally acquired from another genus and order within the class Alphaproteobacteria. Bacteria in this other genus also have the genetic capacity to produce GTAs, although such production has not been documented to our knowledge. What processes this TCS regulates in other organisms (e.g. GTA gene expression and/or motility?) and to which stimuli it responds are questions that future studies will have to answer to help clarify the role of this conserved TCS in different lineages. A graphical summary of the findings for this chapter is shown above (Figure 3.8).

3.6 References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi.org/10.1016/S0022-2836(05)80360-2.

Barbieri, C. M., and Stock, A. M. (2008). Universally applicable methods for monitoring response regulator aspartate phosphorylation both in vitro and in vivo using Phos-Tag-based reagents. *Anal. Biochem.* 376, 73-82. doi.org/10.1016/j.ab.2008.02.004.

Beatty, J. T., and Gest, H. (1981). Generation of succinyl-coenzyme A in photosynthetic bacteria. *Arch. Microbiol.* 129, 335–340. doi:10.1007/BF00406457.

Biers, E. J., Wang, K., Pennington, C., Belas, R., Chen, F., and Moran, M. A. (2008). Occurrence and expression of gene transfer agent genes in marine bacterioplankton. *Appl. Environ. Microbiol*. 74, 2933-2939. doi.org/10.1128/AEM.02129-07.

Blattner, F. R. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1462. doi:10.1126/science.277.5331.1453.

Capella,G. S., Martínez, J. M. S., and Gabaldón, T. (2009). TrimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 25, 1972-1975. doi.org/10.1093/bioinformatics/btp348.

Casino, P., Rubio, V., and Marina, A. (2010). The mechanism of signal transduction by two-component systems. *Curr. Opin. Struct. Biol*. 20, 763-771. doi.org/10.1016/j.sbi.2010.09.010.

De Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., et al. (2006). ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* 34, W362–365. doi:10.1093/nar/gkl124.

Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P., et al. (2006). Allosteric control of cyclic di-GMP signaling. *J. Biol. Chem.* 281, 32015–32024. doi:10.1074/jbc.M603589200.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol.* Evol. 17, 368-376.. doi.org/10.1007/BF01734359.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*. 39, 783-791. doi.org/10.2307/2408678.

Ferreira, R. B.R., Antunes, L. C. M., Greenberg, E. P., and McCarter, L. L. (2008). *Vibrio parahaemolyticus* ScrC modulates cyclic dimeric gmp regulation of gene expression relevant to growth on surfaces.. *J Bacteriol*. 190, 851-860. doi.org/10.1128/JB.01462-07.

Fogg, P. C. M. (2019). Identification and characterization of a direct activator of a gene transfer agent. *Nat. Commun.* 10, 595. doi:10.1038/s41467-019-08526-1.

Galperin, M. Y. (2010). Diversity of structure and function of response regulator output domains. *Curr. Opin. Microbiol*. 150, 9. doi.org/10.1016/j.mib.2010.01.005.

Gao, R., andStock, A. M. (2009). Biological insights from structures of two-component proteins. *Ann. Rev. Microbiol*. 63, 133-154. doi.org/10.1146/annurev.micro.091208.073214.

Grebe, T. W., and Stock, J. B. (1999). The histidine protein kinase superfamily. *Adv. Microb. Physiol*. 41, 139-227 doi.org/10.1016/s0065-2911(08)60167-8.

He, F. (2011). Bradford protein assay. *Bioprotocol* . doi.org/10.21769/bioprotoc.45.

Hengge, R. (2009). Principles of c-di-gmp signalling in bacteria. *Nat. Rev. Microbiol*. 7, 263-273. doi.org/10.1038/nrmicro2109.

Hesterberg, T., Chambers, J. M., and Hastie, T. J. (1993). *Statistical Models in S*. doi:10.2307/1269676.
Hofmann, K., and Stoffel, W.(1993). TMbase: A database of membrane spanning protein segments. *Biol. Chem*. 374, 166. doi.org/10.1515/bchm3.1993.374.1-6.143.
Hynes, A. P., and Lang, A. S. (2013). *Rhodobacter capsulatus* gene transfer agent (RcGTA) activity bioassays. *Bioprotocol* 3, e317. doi:10.1093/molbev/msw125.
Hynes, A. P., Shakya, M., Mercer, R. G., Grüll, M. P., Bown, L., Davidson, F., et al. (2016). Functional and evolutionary characterization of a gene transfer agent's multilocus "genome." *Mol. Biol. Evol.* 33, 2530–2543. doi:10.1093/molbev/msw125.
Jenal, U., Reinders, A., and Lori, C. (2017). Cyclic di-GMP: second messenger extraordinaire. *Nat. Rev. Microbiol.* 15, 271–284. doi:10.1038/nrmicro.2016.190. doi.org/10.1038/nrmicro.2016.190.
Käll, L., Krogh, A., and Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction-the phobius web server. *Nucleic Acids Res*. 35, W429-432. doi.org/10.1093/nar/gkm256.
Katoh, K., and Standley, D. M.(2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772-780. doi.org/10.1093/molbev/mst010.
Kim, D. J.,, and Forst, S. (2001). Genomic analysis of the histidine kinase family in bacteria and archaea. *Microbiology*. 147, 1197-1212. doi.org/10.1099/00221287-147-5-1197.
Kinoshita, K.,, Kinoshita, E. E., Eguchi, Y., Yanagihara, S., Edahiro, K., Inoue, Y., Taniguchi, M et al. (2015). Functional characterization of the receiver domain for phosphorelay control in hybrid sensor kinases. *PloS One*. 7, e0132598. doi.org/10.1371/journal.pone.0132598.
Krell, T., Lacal, J., Busch, A., Jiménez, H. S., Guazzaroni, M. E., and Ramos, J. L.(2010). Bacterial sensor kinases: Diversity in the recognition of environmental signals. *Ann. Rev.Microbiol*. 64, 539-559. doi.org/10.1146/annurev.micro.112408.134054.
Kumar, M., and Chatterji, D. (2008). Cyclic-di-gmp: A second messenger required for long-term survival, but not for biofilm formation, in *Mycobacterium smegmatis*. *Microbiology*. 154, 2942-2955. doi.org/10.1099/mic.0.2008/017806-0.
Kumar, S., Stecher, G., and Tamura, K (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Bio. Evol.* 33, 1870-1874. doi.org/10.1093/molbev/msw054.
Lang, A. S., and Beatty, J. T. (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. doi:10.1073/pnas.97.2.859
Lang, A. S., Beatty, J. T., and Rice, P. A. (2017). Mobile genetic elements and horizontal gene transfer in prokaryotes. *Curr. Opin. Microbiol.* 38, 5-7. doi.org/10.1016/j.mib.2017.09.018.
Lang, A. S., and Beatty, J. T. (2007). Importance of widespread gene transfer agent genes in α-proteobacteria. *Trends Microbiol.* 15, 54–62. doi:10.1016/j.tim.2006.12.001
Lang, A. S., Zhaxybayeva, O., and Beatty, J. T. (2012). Gene transfer agents: Phage-like elements of genetic exchange. *Nat. Rev. Microbiol*. 10, 472-482. doi.org/10.1038/nrmicro2802.
Letunic, I. (2004). SMART 4.0: towards genomic data integration. *Nucleic Acids Res.* 32,

142D–144. doi:10.1093/nar/gkh088.
Leung, M. M., Brimacombe, C. A., Spiegelman, G. B., and Beatty, J. T. (2012). The GtaR protein negatively regulates transcription of the *gtaRI* operon and modulates gene transfer agent (RcGTA) expression in *Rhodobacter capsulatus*. *Mol. Microbiol.* 83, 759–774. doi:10.1111/j.1365-2958.2011.07963.x.
Liu, W.,, Li, L., Khan, A. Md., and Zhu, F. (2012). Popular molecular markers in bacteria. *Mol. Gen. Microbiol. Virusol.*3, 14-17. doi.org/10.3103/S0891416812030056.
Lukat, G. S., McCleary, W. R., Stock, A. M., and Stock, A. M. (1992). Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors. *Proc. Natl. Acad. Sci. U.S.A.* 89, 718-722. doi.org/10.1073/pnas.89.2.718.
Marrs, B. (1974). Genetic recombination in *Rhodopseudomonas capsulata*. *Proc. Natl. Acad. Sci. U. S. A.* 71, 971–973. doi:10.1073/pnas.71.3.971.
Marx, C. J., and Lidstrom, M. E. (2001). Development of improved versatile broad-host-range vectors for use in methylotrophs and other gram-negative bacteria. *Microbiology* 147, 2065–2075. doi:10.1099/00221287-147-8-2065.
Mercer, R. G., Callister, S. J., Lipton, M. S., Pasa-Tolic, L., Strnad, H., Paces, V., et al. (2010). Loss of the response regulator CtrA causes pleiotropic effects on gene expression but does not affect growth phase regulation in *Rhodobacter capsulatus*. *J. Bacteriol.* 192, 2701–2710. doi:10.1128/JB.00160-10.
Mercer, R. G., Quinlan, M., Rose, A. R., Noll, S., Beatty, J. T., and Lang, A. S. (2012). Regulatory systems controlling motility and gene transfer agent production and release in *Rhodobacter capsulatus*. *FEMS Microbiol. Lett.* 331, 53–62. doi:10.1111/j.1574-6968.2012.02553.x.
Ogier, J. C.,, Pagès, S., Galan, M., Barret, M., and Gaudriault, S. (2019). RpoB, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing. *BMC Microbiol*. 19, 171. doi.org/10.1186/s12866-019-1546-z.
Pallegar, P., Castillo, L. P., Langille, E., Gomelsky, M., and Lang, A. S. (2020). Cyclic di-gmp-mediated regulation of gene transfer and motility in *Rhodobacter capsulatus*. *J Bacteriol.* 202, e00554-19. doi.org/10.1128/JB.00554-19.
Paul, R., Weiser, S., Amiot, N. C., Chan, C., Schirmer, T., Giese, B., et al. (2004). Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev.* 18, 715–727. doi:10.1101/gad.289504.
Perry, J., Koteva, K., and Wright,G. (2011). Receptor domains of two-component signal transduction systems. *Mol. BioSyst.* 7, 1388-1398. doi.org/10.1039/c0mb00329h.
Prentki, P., and Krisch,H. M. (1984). In vitro insertional mutagenesis with a selectable DNA fragment. *Gene.* 29,303–313.
Römling, U., Galperin, M. Y., and Gomelsky, M. (2013). Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol. Mol. Biol. Rev.* 77, 1–52. doi:10.1128/MMBR.00043-12.
Römling, U., and Galperin, M. Y. (2017). Discovery of the second messenger cyclic di-gmp. *Meth. Mol. Biol*. 1657, 1-8. doi.org/10.1007/978-1-4939-7240-1_1.
Ryjenkov, D. A., Simm, R., Römling, U., and Gomelsky, M. (2006). The PilZ domain is a receptor for the second messenger c-di-GMP: the PilZ domain protein YcgR controls motility in enterobacteria. *J. Biol. Chem.* 281, 30310–30314. doi:10.1074/jbc.C600179200.

Saitou, N, and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
Sánchez, R., Serra, F., Tárraga, J., Medina, I., Carbonell, J., Pulido, L., María, A. D., et al. (2011). Phylemon 2.0: A suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Res.* 39, W470-W474. doi.org/10.1093/nar/gkr408.
Schaefer, A. L., Taylor, T. A., Beatty, J. T., and Greenberg, E. P. (2002). Long-chain acyl-homoserine lactone quorum-sensing regulation of *Rhodobacter capsulatus* gene transfer agent production. *J. Bacteriol.* 184, 6515–21. doi:10.1128/JB.184.23.6515-6521.2002.
Scharf, B. E. (2010). Summary of useful methods for two-component system research. *Curr. Opin. Microbiol*. 13, 246-252. doi.org/10.1016/j.mib.2010.01.006.
Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi:10.1038/nmeth.2089.
Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5857–5864. doi:10.1073/pnas.95.11.5857.
Seshasayee, A. S.N., Fraser, G. M., and Luscombe, N. M. (2010). Comparative genomics of cyclic-di-gmp signalling in bacteria: Post-translational regulation and catalytic activity. *Nucleic Acids Res.* 38, 5970-5981. doi.org/10.1093/nar/gkq382.
Shakya, M., Soucy, S. M., and Zhaxybayeva, O. (2017). Insights into origin and evolution of α-proteobacterial gene transfer agents. *Virus Evol.* 3, vex036. doi.org/10.1093/ve/vex036.
Shiro, Y., and Yamada, S. (2008). Structural basis of the signal transduction in the two-component system. *Adv. Exp. Med. Biol*. 631, 22-39. doi.org/10.1007/978-0-387-78885-2_3.
Simm, R., Morr, M., Kader, A., Nimtz, M., and Römling, U. (2004). GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessibility to motility. *Mol. Microbiol.* 53, 1123–1134. doi:10.1111/j.1365-2958.2004.04206.x.
Simon, R., Priefer, U., and Pühler, A. (1983). A broad host range mobilization system for *in vivo* genetic engineering: Transposon mutagenesis in Gram-negative bacteria. *Bio/Technology* 1, 784–791. doi:10.1038/nbt1183-784.
Stanton, T. B. (2007). Prophage-like gene transfer agents–Novel mechanisms of gene exchange for *Methanococcus*, *Desulfovibrio*, *Brachyspira*, and *Rhodobacter* species. *Anaerobe* 13, 43–49. doi:10.1016/j.anaerobe.2007.03.004.
Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000). Two-component signal transduction. *Ann. Rev. Biochem.* 69,183–215.
Strnad, H., Lapidus, A., Paces, J., Ulbrich, P., Vlcek, C., Paces, V., et al. (2010). Complete genome sequence of the photosynthetic purple nonsulfur bacterium *Rhodobacter capsulatus* SB 1003. *J. Bacteriol.* 192, 3545–3546. doi:10.1128/JB.00366-10.
Tarutina, M., Ryjenkov, D. A., and Gomelsky, M. (2006). An unorthodox bacteriophytochrome from *Rhodobacter sphaeroides* involved in turnover of the second messenger c-di-GMP. *J. Biol. Chem.* 281, 34751–34758. doi:10.1074/jbc.M604819200.
Taylor, D. P., Cohen, S. N., Clark, W. G., and Marrs, B. L. (1983). Alignment of genetic and restriction maps of the photosynthesis region of the *Rhodopseudomonas*

capsulata chromosome by a conjugation-mediated marker rescue technique. *J. Bacteriol.* 154, 580–590.
Tischler, A. D., and Camilli, A. (2004). Cyclic diguanylate (c-di-gmp) regulates *Vibrio cholerae* biofilm formation. *Mol. Microbiol*. 53, 857-869. doi.org/10.1111/j.1365-2958.2004.04155.x.
Wall, J. D., Weaver, P. F., and Gest, H. (1975). Gene transfer agents, bacteriophages, and bacteriocins of *Rhodopseudomonas capsulata*. *Arch. Microbiol.* 105, 217–224. doi:10.1007/BF00447140.
Wong, D. K. H., Collins, W. J., Harmer, A., Lilburn, T. G., and Beatty, J. T. (1996). Directed mutagenesis of the *Rhodobacter capsulatus puhA* gene and Orf 214: Pleiotropic effects on photosynthetic reaction center and light-harvesting 1 complexes. *J. Bacteriol.* 178, 2334–2342. doi:10.1128/jb.178.8.2334-2342.1996.
Yen, H. C., and Marrs, B. (1976). Map of genes for carotenoid and bacteriochlorophyll biosynthesis in *Rhodopseudomonas capsulata*. *J. Bacteriol.* 126, 619–629.
Zhulin, I. B., Nikolskaya, A. N., and Galperin, M. Y. (2003). Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in bacteria and archaea. *J. Bacteriol.* 185, 285–294. doi.jb.asm.org/cgi/content/abstract/185/1/285.
Zschiedrich, C. P., Keidel, V., and Szurmant, H. (2016). Molecular mechanisms of two-component signal transduction. *J. Mol.Biol*. 428, 3752-3775. doi.org/10.1016/j.jmb.2016.08.003.
Zuzuki, S., and Hiraishi, A. (2007). *Novosphingobium naphthalenivorans* Sp. Nov., a naphthalene-degrading bacterium isolated from polychlorinated-dioxin-contaminated environments. *J. Gen. Appl. Microbiol*. 53, 221-228. doi.org/10.2323/jgam.53.221.

CHAPTER 4- Investigating c-di-GMP binding to CckA and its effects on the CckA-ChpT-CtrA phosphorelay system in *Rhodobacter capsulatus*

4.1 Introduction

Bacteria are very versatile and can be found in diverse environmental niches. Many bacteria are able to adapt to changes in their environment and employ specific signal transduction mechanisms to adapt their intracellular environment to an appropriate state in response to these changes. These signal transduction systems most often function through regulatory proteins that can sense environmental stimuli and trigger a particular response. For example, some proteins bind to particular metabolites or ions and this increases the affinity of the protein for specific DNA sequences, resulting in changes in gene expression. Two-component systems (TCSs) are specialized signal transduction systems found in bacteria that couple stimulus-response functions (Stock, Robinson, and Goudreau 2000; Gao and Stock 2009). Multiple TCSs are typically found within each bacterium, with each particular TCS specialized to respond to a specific stimulus, such as pH, temperature, osmotic pressure, or specific chemical compounds, and trigger the appropriate cellular response. Stimuli can be environmental or intracellular.

A classic TCS comprises a sensor histidine kinase (HK) protein and a response regulator (RR) protein. The transfer of a phosphoryl group from a conserved histidine residue in the transmitter domain of the HK to a conserved aspartate residue in the receiver (REC) domain of the RR protein forms the basis of this signalling system (Stock, Robinson, and Goudreau 2000) (Figure 4.1A). Phosphorylation of the RR alters the activity of its output domain. Some TCSs are more complex and include additional phosphotransfer steps. These are known as histidyl-aspartyl phosphorelays or multicomponent systems. These phosphorelays involve a “hybrid” HK that contains a C-

terminal REC domain, which is an intermediary phosphoacceptor between the transmitter domain and an additional histidine phosphotransferase (HPT) protein that phosphorylates the cognate RR (Gao and Stock 2009) (Figure 4.1B). Hybrid HKs are fairly common and examples in *Escherichia coli* include ArcB, BarA, EvgS and RcsC (Mizuno 1997).

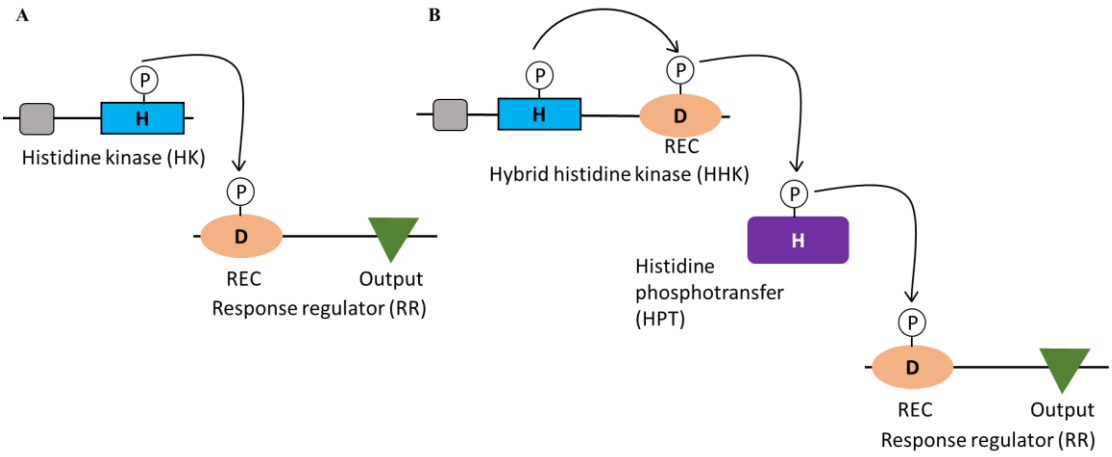


Figure 4.1. Schematic representation of two-component signalling systems. A. A classic TCS involving HK and RR proteins. In response to a stimulus received at the sensory domain (grey), a conserved histidine residue in the transmitter domain (blue) of the HK autophosphorylates and then transfers the phosphate to a conserved aspartate residue in the receiver domain (REC, beige) of the RR. This affects the activity of the RR output domain (green). B. A histidyl-aspartyl phosphorelay system involving hybrid HK and intermediate HPT (purple) proteins.

A phosphorelay system involving the CckA, ChpT and CtrA proteins (hybrid HK, HPT and RR, respectively) is widely conserved in the class *Alphaproteobacteria* (Brilli et al. 2010). It has been best characterized in *Caulobacter crescentus* where it functions to regulate cell cycle and developmental processes (Jacobs et al. 1999; Laub et al. 2002;

Chen et al. 2009; Lori et al. 2015). CckA contains two N-terminal Per-ARNT-Sim (PAS) sensory domains and, in response to an unknown signal, it causes the phosphorylation of CtrA via ChpT (Jacobs et al. 1999; Biondi et al. 2006), as described above for this type of phosphorelay. In *C. crescentus*, CtrA acts as a "master regulator" and controls at least 25% of total cell cycle regulated genes (Skerker and Laub 2004). It acts at more than 100 promoters, thereby directly regulating the expression of genes involved in diverse cellular processes including flagellar motility and chemotaxis (Laub et al. 2002).

Like other HKs, CckA also acts as a phosphatase and can cause the dephosphorylation of CtrA. In *C. crescentus*, the activity of CckA is controlled in part by another sensor kinase protein, DivL, which stimulates the kinase activity of CckA. In the absence of DivL, or when over-expressed, CckA primarily acts as a phosphatase (Chen et al. 2009; Tsokos, Perchuk, and Laub 2011). The ability of DivL to stimulate CckA kinase activity is inhibited by phosphorylated DivK, whose phosphorylation state is dependent on PleC and DivJ (Childers et al. 2014; Tsokos and Laub 2012). Interestingly, it was recently discovered that CckA's activity is also controlled by the dinucleotide molecule cyclic-dimeric-guanosine monophosphate (c-di-GMP) (Lori et al. 2015; Dubey et al. 2016; Kaczmarczyk et al. 2020). Increased levels of c-di-GMP during the G1-S transition phase cause CckA to switch from kinase to phosphatase mode. C-di-GMP binds to CckA and activates its phosphatase activity, which results in the dephosphorylation of CtrA upon entry into S-phase (Lori et al. 2015). CckA ligand affinity assays also showed that c-di-GMP binding to CckA is seen only in the presence of adenosine diphosphate (ADP) and these two molecules appear to synergistically stabilize CckA phosphatase activity (Dubey et al. 2016).

As mentioned above, homologs of these proteins are widespread in the *Alphaproteobacteria*, but only a few of these have been studied. However, these studies have shown that the regulatory functions of this system vary among species (Mercer et al. 2010; Greene et al. 2012; Francez-Charlot, Kaczmarczyk, and Vorholt 2015; Bird and MacKrell 2011; Bellefontaine et al. 2002). For example, while the *C. crescentus* CckA-ChpT-CtrA system regulates cell cycle processes and is essential for viability, the homologs found in *Rhodobacter capsulatus* are not essential (Lang and Beatty 2000; Mercer et al. 2012). In this bacterium, the CckA-ChpT-CtrA phosphorelay regulates flagellar motility (Lang and Beatty 2002; Mercer et al. 2012), which appears to be the most commonly shared trait among alphaproteobacteria, and the production of its gene transfer agent, RcGTA (Lang and Beatty 2000). A DivL homolog was also identified in *R. capsulatus* and, although it is truncated compared to the *C. crescentus* protein, it was shown to affect CtrA phosphorylation by stimulating the kinase activity of CckA (Westbye et al. 2018). Loss of DivL corresponded with reduced CckA kinase activity and increased phosphatase activity, thereby affecting the expression of genes activated by CtrA~P. However, homologs of DivK, PleC and DivJ that control the activity of DivL in *C. crescentus* are absent in *R. capsulatus.*

In the previous chapters of this thesis (Chapters 2 and 3), I showed that the second messenger c-di-GMP affects flagellar motility and RcGTA production in *R. capsulatus*. In this chapter, I have attempted to link the c-di-GMP effects with the CckA-ChpT-CtrA phosphorelay. I investigated the binding of c-di-GMP to CckA *in vitro* and also tested the effects of ADP on this interaction. Key amino acids in CckA, chosen on the basis of their conservation with identical residues that affect the activity of the *C. crescentus* protein, were mutated and the effects of these mutations on c-di-GMP binding were also tested.

These CckA constructs were shared with J.T. Beatty's research group (at University of British Columbia) for them to evaluate the functionality of these proteins via *in vitro* phosphotransfer experiments as part of a larger collaborative study. Additionally, it is known that allosteric or inhibitory site (I-site) RXXD motifs that are present in diguanylate cyclase enzymes can act as receptors for c-di-GMP to carry out feedback inhibition (Christen et al. 2006; Schirmer and Jenal 2009). Therefore, I also mutated I-site motifs in CckA to test their potential roles in c-di-GMP binding.

4.2 Materials and methods

4.2.1 Bacterial strains and growth conditions

All the plasmids used in this study are listed in Table 4.1. *E. coli* strains NEB5α and BL21(DE3) (New England Biolabs) were used for cloning and protein overexpression, respectively. *E. coli* strains were grown at 37 °C or 30 °C for protein expression in LB medium supplemented with kanamycin (25 μg ml^{-1}) when necessary.

Table 4.1. List of plasmids used in this study.

Strains and plasmids	Description	Reference or source
Plasmids		
pGEM-T Easy	TA PCR product cloning vector	Promega
pET28-a	Expression vector for expressing 6X-His-tagged proteins	Novagen
pCckAΔTM	Expression vector for C-terminal 6X-His-tagged CckA without transmembrane region	This study

pG395T	pCckAΔTM with mutation at position 395 aa residue to substitute glycine to threonine	This study
pG396E	pCckAΔTM with mutation at position 396 aa residue to substitute glycine to glutamate	This study
pH399A	pCckAΔTM with mutation at position 399 aa residue to substitute histidine to alanine	This study
pV443P	pCckAΔTM with mutation at position 443 aa residue to substitute valine to proline	This study
pY589D	pCckAΔTM with mutation at position 589 aa residue to substitute tyrosine to aspartate	This study
pI-1	pCckAΔTM with mutation at 1st predicted I-site motif starting at position 166	This study
pI-2	pCckAΔTM with mutation at 2nd predicted I-site motif starting at position 212	This study

pI-3	pCckAΔTM with mutation at 3rd predicted I-site motif starting at position 317	This study
pI-4	pCckAΔTM with mutation at 4th predicted I-site motif starting at position 341	This study
pI-5	pCckAΔTM with mutation at 5th predicted I-site motif starting at position 458	This study

4.2.2 Protein domain analysis

Protein sequence analyses for identification of functional domains were done using the SMART (Letunic 2004; Schultz et al. 1998) and Expasy-Prosite (de Castro et al. 2006) databases. Sequence alignments were done using ClustalW Omega (Larkin et al. 2007). Transmembrane regions were predicted using TMHMM Server 2.0 (Krogh et al. 2001) and the TM Pred database (Hofmann and Stoffel 1993).

4.2.3 Expression and purification of recombinant proteins from *E. coli*

To create a recombinant C-terminal 6X-histidine tagged protein, the *cckA* gene (*rcc01749*) excluding the region encoding the N-terminal transmembrane portion was amplified using gene-specific primers (Appendix 3, Table S4.1) and cloned as a *Nco*I/*Hind*III fragment into the pET28-a vector. The resulting plasmid was sequence-confirmed and transformed into *E. coli* BL21 (DE3). It was also used for mutagenesis (described below). For protein expression, overnight cultures were used to inoculate 200 ml of LB broth with kanamycin and incubated at 30 °C with shaking at 220 rpm. After one hour, isopropyl-β-D-thiogalactopyranoside (IPTG) was added at a final concentration

of 1 mM to induce gene expression and cells were incubated for a further 5-6 hours. Cells were harvested by centrifugation (5000 X *g* for 10 minutes at 4 °C) and cell pellets were resuspended in lysis buffer [50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 0.1% (v/v), Benzonase® nuclease (Qiagen), 1 mg ml^{-1} lysozyme (w/v); pH 8] and incubated on ice for 30 minutes. The suspended mix was centrifuged at 10000 X *g* for 30 minutes at 4 °C. The supernatant was collected and mixed 4:1 (v/v) with Ni-NTA agarose slurry (Qiagen) and incubated at 4 °C for 1 hour. After incubation, the mix was loaded into a polypropylene column and washed twice with wash buffer [50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole; pH 8.0] and the recombinant proteins were eluted in four 0.5-ml aliquots of elution buffer [50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole; pH 8.0]. The eluted proteins were analysed by SDS-PAGE and Coomassie Brilliant Blue staining. The eluted protein fractions were pooled as appropriate, dialyzed against dialysis buffer [50 mM NaH_2PO_4, 300 mM NaCl; pH 8] and quantified using a Bradford protein assay (He 2011).

4.2.4 Amino acid substitutions

The above-mentioned CckAΔTM construct in the pET28-a vector was used as a template for performing site-directed mutagenesis (SDM) using the QuickChange Lightning SDM kit (Agilent Technologies) as per the manufacturer's instructions. Key amino acid residues in the CckA protein that affect kinase and phosphatase activities and in predicted I-site motifs that might act as c-di-GMP binding sites were chosen and mutated to test their potential involvement in c-di-GMP binding. Mutagenesis PCRs were done using *PfuUltra* High-Fidelity DNA polymerase and site-specific primers ((Appendix 3, Table S4.1) designed to create point mutations to change glycine 395 to threonine (G395T), glycine 396 to glutamate (G396E), histidine 399 to alanine (H399A), valine 443

to proline (V443P), tyrosine 589 to aspartate (Y589D) and five predicted I-site motifs, RXXD, to GXXG. The methylated template DNAs were then digested by incubation with *Dpn*I for 10 minutes at 37 °C and the remaining DNA was transformed into *E. coli*. Plasmids carrying mutations were confirmed by sequencing and later transformed into *E. coli* BL21 (DE3) for protein expression and purification.

4.2.5 C-di-GMP binding assays

In vitro pull-down assays using streptavidin beads and biotinylated c-di-GMP were performed to assay c-di-GMP binding by CckA as described previously (Chambers and Sauer 2017), with minor modifications. Briefly, a 20-µl reaction mix containing the purified CckAΔTM protein (10 µM), with or without 1 µl of biotinylated c-di-GMP (200 pmoles), 2 µl of 10X reaction buffer (100 mM Tris, 1M KCl, 10 mM DTT; pH 7.5), 2 mM EDTA, and with or without 5 mM ADP was incubated at room temperature for 1 hour. The reaction mixture was then mixed with streptavidin beads (blocked with 5% skim milk) in 250 µl TBST [20 mM Tris, 137 mM NaCl, 0.1% Tween-20 (v/v); pH 7.5] and incubated at room temperature on a vortex mixer shaking at 1400 rpm for 1 hour. The beads were collected using a magnetic strand and the supernatant was removed. The beads were washed 4-5 times by adding 1 ml of TBST and incubating on the vortex mixer for 15 minutes at room temperature. The beads were resuspended in 15 µl of sterile distilled water and mixed with 5 µl of 3X SDS-PAGE loading buffer.

4.2.6 Western blotting

The streptavidin bead suspensions were analysed to detect and quantify the His-tagged proteins by western blotting. The samples were heated at 98 °C for 5 minutes, followed by collection of the beads on a magnetic stand, and the supernatant was collected and run on a 10% SDS-PAGE gel. The proteins were transferred onto

nitrocellulose membranes by electroblotting in transfer buffer [48 mM Tris base, 39 mM glycine, 20% methanol (v/v)]. The membranes were blocked with 5% (w/v) skim milk solution in TBST and incubated with the primary antibody, anti-HisTag protein (Thermo Fisher Scientific), overnight at 4 °C. After washing with TBST, membranes were incubated with secondary antibody, peroxidase-conjugated anti-rabbit IgG (Santa Cruz Biotechnology), at room temperature for 1 hour. The SuperSignal West Femto Reagent Kit (Thermo Fisher Scientific) was used to detect the bands by chemiluminescence and images were captured using an Agilent ImageQuant LAS 4000 imaging system. Images were inverted and adjusted for brightness and contrast, and band intensities were quantified using ImageJ (Schneider, Rasband, and Eliceiri 2012).

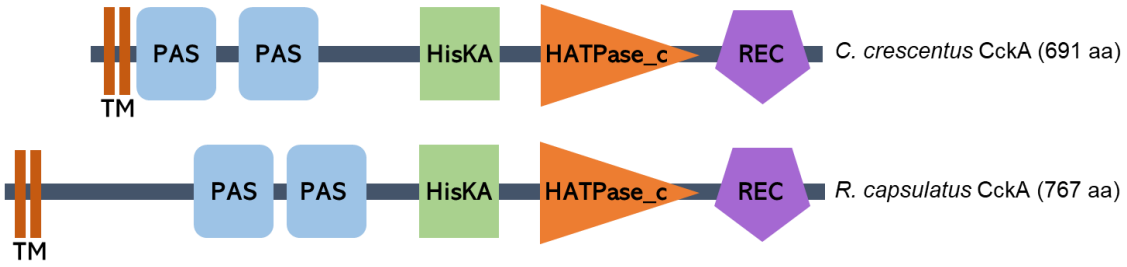
4.3 Results

4.3.1 Comparison of the *R. capsulatus* and *C. crescentus* CckA proteins

The *C. cresenctus cckA* homolog found in *R. capsulatus* encodes a 767 amino acid (aa) protein that shares 44% overall identity with *C. crescentus* CckA. The functional domain organisations of the *R. capsulatus* and *C. crescentus* proteins showed the same pattern (Figure 4.2A). Both proteins contain two predicted transmembrane regions (aa 10-32 and 39-58 of the *R. capsulatus* protein) and 2 PAS domains (aa 174-239 and 266-328 of the *R. capsulatus* protein) in their N-terminal regions and a HisKA domain (aa 389-455), an HATPase_c domain (aa 496-619), and a response regulator receiver (REC) domain (aa 644-767) in their C-terminal regions (Figure 4.2A). There is a high level of sequence conservation in the C-terminal regions of the proteins (aa 323-767 of the *R. capsulatus* protein) and the HisKA, HATPase_c and REC domains share 67%, 50% and 44% identity, respectively (Figure 4.2B), while the N-terminal regions have much lower identity. Key aa residues identified in *C. crescentus* CckA that are the sites of

phosphorylation, H322 and D623, were conserved in the *R. capsulatus* protein (H399 and D696, respectively) (Figure 4.2B).

A



B

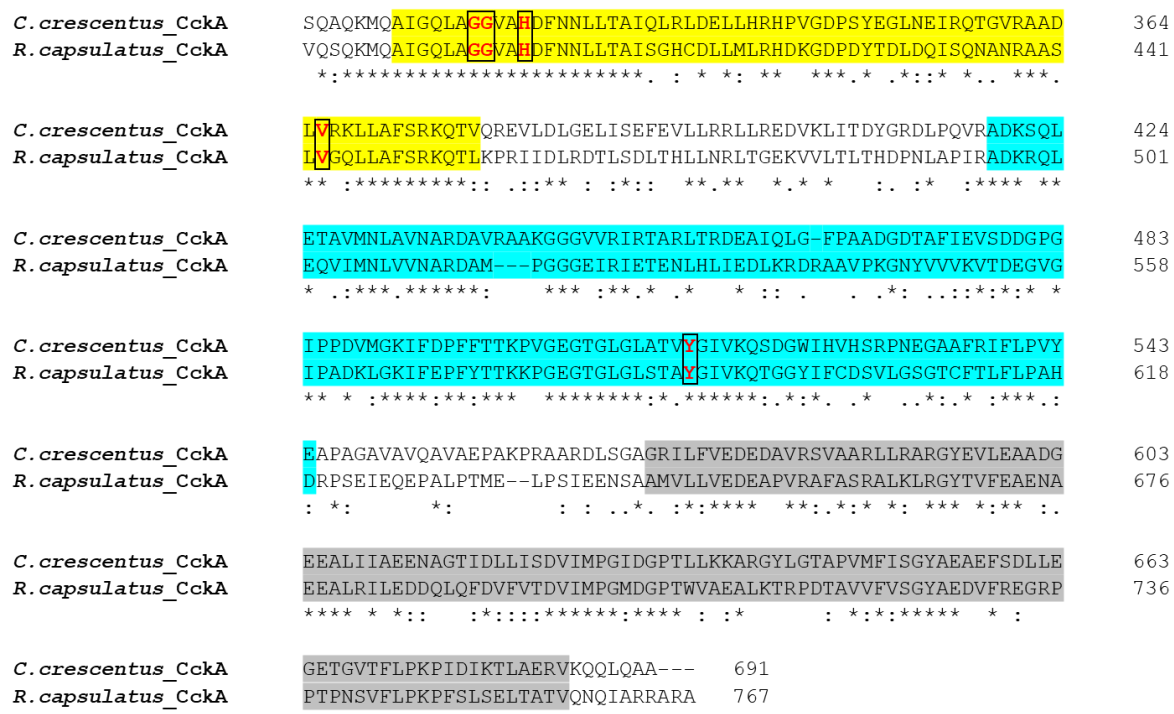


Figure 4.2. Comparison of the *R. capsulatus* and *C. crescentus* CckA proteins. A. Locations and organizations of predicted domains: TM, transmembrane; PAS, Per-ARNT-Sim; HisKA, histidine kinase; HATPase_c, ATP-binding catalytic domain; REC, response regulator receiver. B. Amino acid sequence alignments for *R. capsulatus* and *C. crescentus* CckA proteins indicating the HisKA (highlighted in yellow), HATPase_c (highlighted in blue), and REC (highlighted in grey) domains. Key amino acid residues mutated in this study are in red and boxed, except for the I-site motifs that do not fall within this region.

4.3.2 C-di-GMP binds to *R. capsulatus* CckA *in vitro*

Binding assays showed that the *R. capsulatus* CckAΔTM protein binds to c-di-GMP and that addition of ADP increased the binding (Figure 4.3). Quantification of the

western blot band intensities showed that the binding in the absence of ADP was significantly higher than the negative control (CckA + beads in the absence of biotinylated c-di-GMP) and binding in the presence of ADP was significantly higher than without (Figure 4.3).

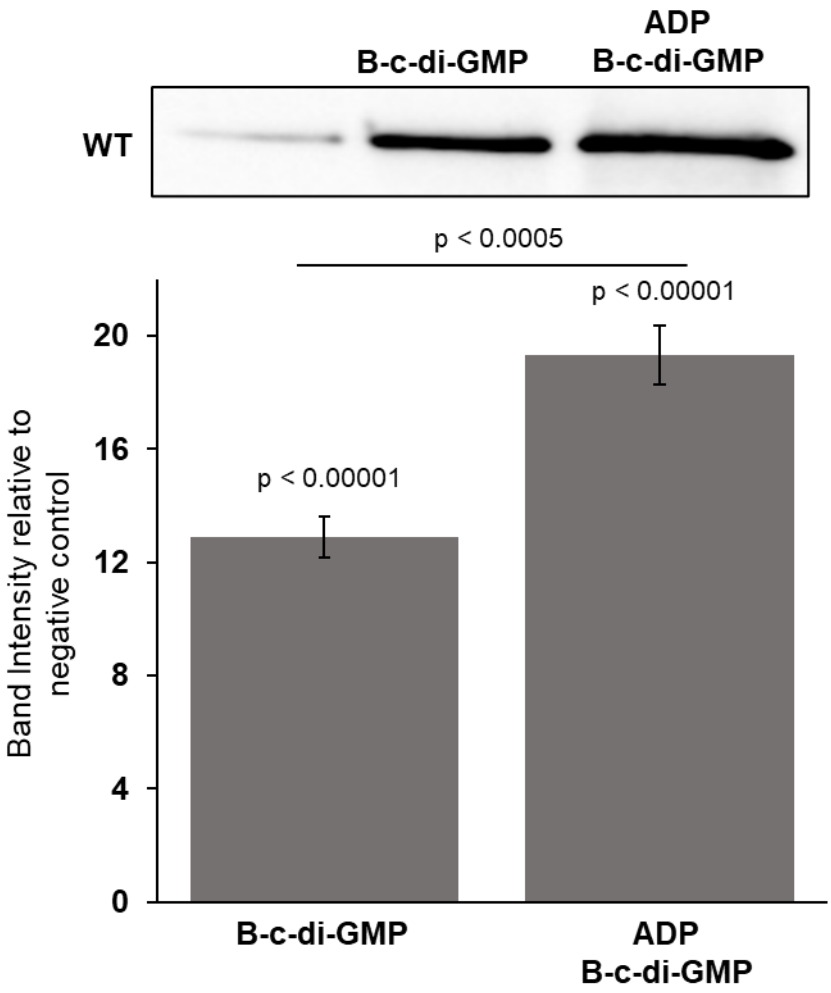


Figure 4.3. C-di-GMP binding assays with CckAΔTM. The protein was incubated with the molecules indicated above the lanes (biotinylated c-di-GMP, (B-c-di-GMP); adenosine diphosphate, (ADP)) at room temperature for 1 hour followed by addition of streptavidin beads. The eluted samples were analyzed by western blotting using anti-His-tag primary antibody. The band intensities are presented as an average of three replicates

relative to negative control (no B-c-di-GMP or ADP). The p-value above each bar represents the difference compared to the negative control while the p-value above the line represents the difference between the results with and without ADP (one-way analysis of variance followed by Tukey's HSD post-hoc test).

4.3.3 The Y589D mutation eliminates c-di-GMP binding

The tyrosine at position 589 in the *R. capsulatus* CckA corresponds with that at position 514 in *C. crescentus*, which was shown to be required for binding to c-di-GMP. The Y589D substitution showed a decrease in c-di-GMP binding, regardless of the presence or absence of ADP (Figure 4.4).

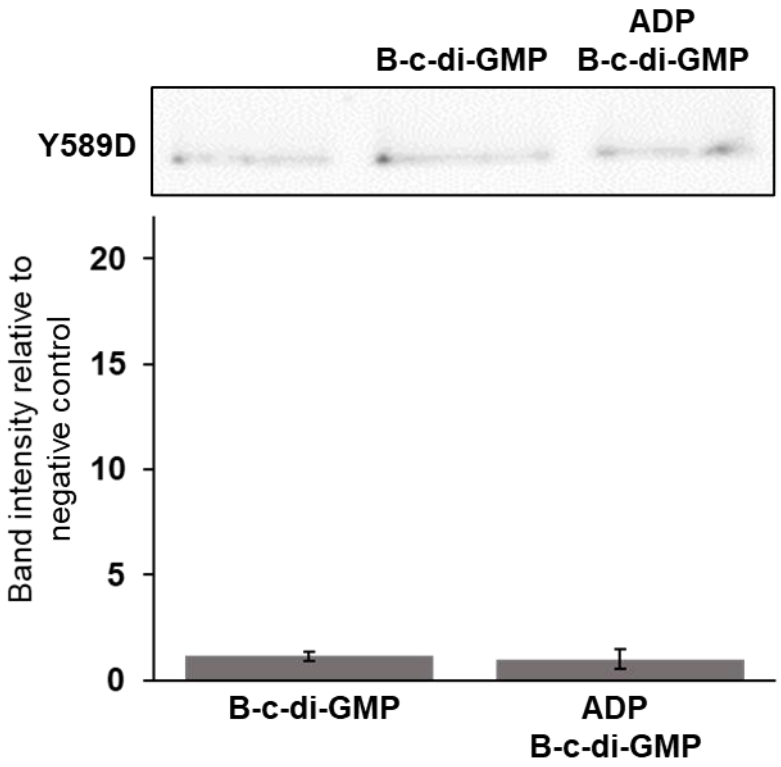


Figure 4.4. C-di-GMP binding assays with the Y589D version of CckAΔTM. The protein was incubated with the molecules indicated above the lanes (biotinylated c-di-GMP, B-c-di-GMP; adenosine diphosphate, ADP) at room temperature for 1 hour followed by

addition of streptavidin beads. The eluted samples were analyzed by western blotting using anti-His-tag primary antibody. The band intensities are presented as an average of three replicates relative to negative control (no B-c-di-GMP or ADP). No significant difference was identified compared to the negative control for either experimental sample (one-way analysis of variance followed by Tukey's HSD post-hoc test).

4.3.4 Effects of other mutations on c-di-GMP binding

In addition to the wild type and Y589D CckAΔTM proteins, I also performed binding assays on the other mutants that were constructed for the purpose of *in vitro* phosphotransferase assays. These were mutations at key aa residues corresponding to those that are involved in the normal functioning of CckA kinase and phosphatase activities in *C. crescentus*, made to test the conservation of the function of these residues in the two distantly related bacteria. My preliminary experiments did not provide any indication that c-di-GMP binding was affected by these substitutions (Figure 4.5).

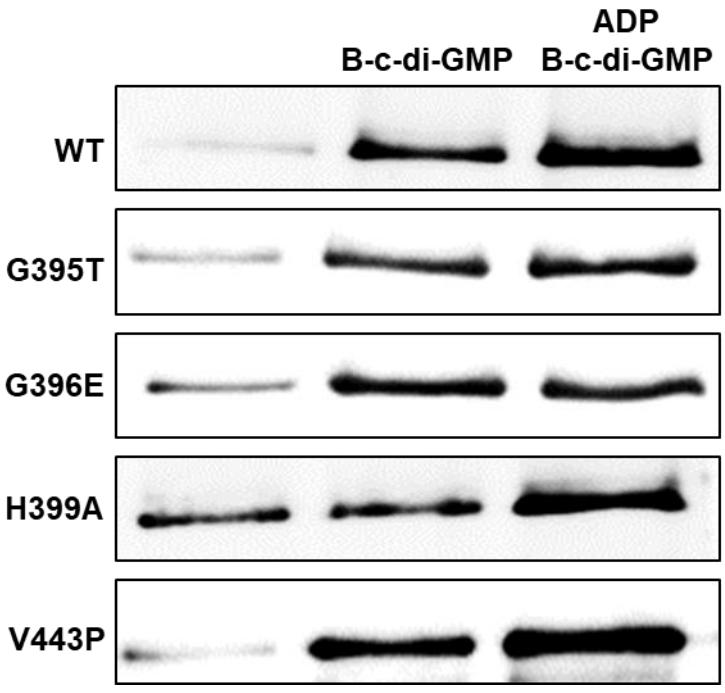


Figure.4.5. Effect of additional point mutations on c-di-GMP binding. Key amino acid residues predicted to be involved in kinase and phosphatase activities were selected and mutated. The different proteins were incubated with the molecules indicated above the lanes (biotinylated c-di-GMP, B-c-di-GMP; adenosine diphosphate, ADP) at room temperature for 1 hour followed by addition of streptavidin beads. The eluted samples were analyzed by western blotting using anti-His-tag primary antibody to detect the His-tagged protein.

4.3.5 Effects of I-site motif mutations on c-di-GMP binding

The *R. capsulatus* CckA sequence contains 5 I-site motifs (RXXD), which are sites involved in c-di-GMP binding in c-di-GMP signaling proteins. However, none of the mutations created in these motifs showed any observable difference for c-di-GMP binding compared the wild type protein in my initial experiments (Figure 4.6).

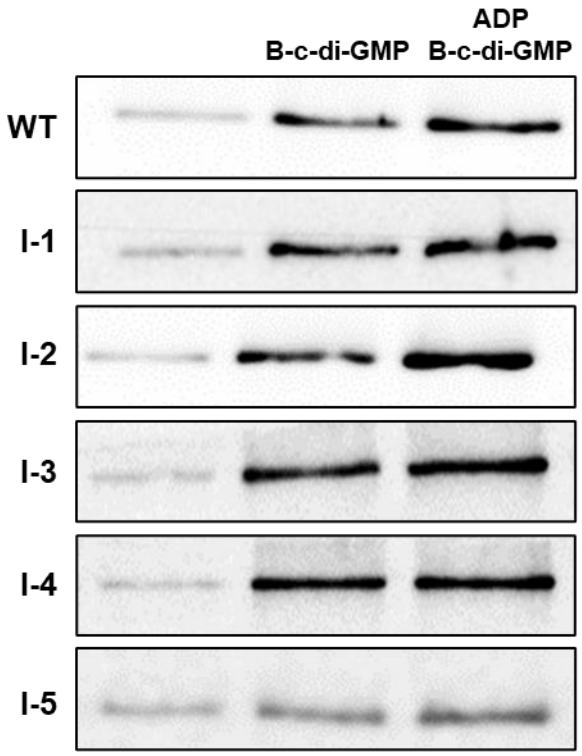


Figure 4.6. Effects of I-site mutations on c-di-GMP binding. The 5 I-site (RXXD) motifs were mutated (to GXXG) and each protein was assayed for c-di-GMP binding. The proteins were incubated with the molecules indicated above the lanes (biotinylated c-di-GMP, B-c-di-GMP; adenosine diphosphate, ADP) at room temperature for 1 hour followed by addition of streptavidin beads. The eluted samples were analyzed by western blotting using anti-His-tag primary antibody to detect the His-tagged protein.

4.4 Discussion

In a previous chapter, I showed that c-di-GMP levels affect RcGTA production, but how c-di-GMP is connected to RcGTA production mechanistically was unknown. Recent findings in another bacterium, *C. crescentus*, showed that c-di-GMP binds to CckA to modulate its function and thereby regulates the CckA-ChpT-CtrA phosphorelay. Based on these findings, we hypothesized that c-di-GMP likely binds to the *R. capsulatus* CckA and thereby regulates RcGTA production in *R. capsulatus*. The c-di-GMP binding assays performed using the purified CckAΔTM protein confirmed that c-di-GMP binds to *R. capsulatus* CckA *in vitro* (Figure 4.3). The Y589D mutation resulted in loss of c-di-GMP binding (Figure 4.4). For the Y589D mutation, the tyrosine residue at position 589 corresponds to the residue found in *C. crescentus* (Y514) and *A. tumefaciens* (Y674) known to be involved in the formation of hydrogen-bonding with guanosine in c-di-GMP (Lori et al. 2015; Dubey et al. 2016). In both species, these residues are crucial for c-d-GMP binding and mutations of those residues also showed diminished phosphatase activity (Lori et al. 2015).

I also tested the other CckA mutants I constructed for phosphotransfer experiments, which contain substitutions predicted to affect kinase (G395T, G396E &

H399A) and phosphatase (V443P) activities. None of these appeared to be affected for c-di-GMP binding (Figure 4.5) as observed for the Y589D protein (Figure 4.4). However, I was not able to perform enough replicates and the assays with these mutants need to be repeated to obtain properly quantitative data. An X-ray crystallographic structural study of the *C. crescentus* CckA revealed that the amino acid residues responsible for binding of both c-di-GMP and ATP are mostly confined to the catalytic region (CA domain) of the protein, although several residues in the dimerization histidine phosphotransfer (DHp) domain also help with this function (Dubey et al. 2016). These additional mutations created are mostly confined to the DHp region of the protein except for V443P, where it lies near the linker that connects DHp and CA domain and therefore it will be interesting to obtain quantitative data for these mutants to see if there are any intermediate effects on c-di-GMP binding.

The I-site motif (RXXD) present in diguanylate cyclase enzymes is involved in feedback inhibition by acting as a receptor for c-di-GMP (Christen et al. 2006; Schirmer and Jenal 2009). I hypothesized that the five I-site motifs present in the *R. capsulatus* CckA protein might be acting as binding sites for c-di-GMP. Although I was not able to complete all the replicates for this experiment and the data presented must be considered preliminary, none of the mutants created to inactivate those five motifs showed any obvious effect on c-di-GMP binding in the assays I performed (Figure 4.6), suggesting that not all RXXD motifs act as binding sites for c-di-GMP and that there are other amino acids involved in the functioning of these sites.

In *C. crescentus*, it was initially reported that c-di-GMP binds CckA only in the presence of ADP (Dubey et al. 2016). However, the results in my experiments show that ADP is not required for binding to occur, although the addition of ADP does increase the

binding (Figure 4.3). These results correspond with more recent findings from *C. crescentus* where the c-di-GMP inhibition of kinase activity in the presence of ADP was mainly seen as an additive effect rather than cooperative (Mann and Shapiro 2018). Interestingly, the *in vitro* phosphotransfer assays performed to test the effect of ADP on c-di-GMP-mediated phosphatase stimulation did not show a difference in the presence or absence of ADP (Calderon 2020).

Previous genetic manipulations and the resulting phenotypic effects on flagellar motility and RcGTA production support the notion that CckA, ChpT and CtrA function as a phosphorelay in *R. capsulatus* (Lang and Beatty 2000; Lang and Beatty 2002; Mercer et al. 2012). However, direct biochemical evidence for this has not been provided. Therefore, the construct for the His-tagged CckAΔTM protein that I made was sent to collaborators (at the University of British Columbia) for *in vitro* phosphotransfer assays. The individual protein components were purified and used for *in vitro* phosphorylation and dephosphorylation assays in the presence and absence of c-di-GMP, which allowed the evaluation of CckA- and ChpT-dependent CtrA phosphorylation and the potential role of c-di-GMP in CtrA dephosphorylation (Calderon 2020). In the absence of c-di-GMP, those assays confirmed the occurrence of autophosphorylation of CckA and that phosphotransfer between CckA and CtrA was mediated by ChpT. Addition of c-di-GMP resulted in CckA-dependent dephosphorylation of CtrA, suggesting that c-di-GMP is acting as a stimulator for CckA phosphatase activity in *R. capsulatus*, analogous to the *C. crescentus* system.

The G395T, G396E, H399A, V443P and Y589D mutants that I created were also tested for their kinase and phosphatase activities *in vitro* (Calderon 2020). The Y589D

mutant showed normal kinase activity and reduced phosphatase activity, corresponding with no c-di-GMP binding observed in c-di-GMP binding assays (Figure 4.4). The G395T and G396E mutations were predicted to affect the kinase activity based on the *C. crescentus* model and both displayed high kinase activity and reduced phosphatase activity. The H399A mutation showed a decrease in kinase activity but displayed normal phosphatase activity, also matching expectations based on the analogous *C. crescentus* H322A mutant. However, the V443P mutant displayed neither kinase nor phosphatase activity, which does not match the results observed in the analogous V366P *C. crescentus* mutant that showed increased kinase activity (Chen et al. 2009).

In *C. crescentus*, the levels of c-di-GMP are regulated by the diguanylate cyclase PleD and mutation of *pleD* directly impacts the CtrA phosphorelay via c-di-GMP (Abel et al. 2013; Paul et al. 2004). The *R. capsulatus* genome contains 20 genes encoding proteins with domains associated with the synthesis and/or degradation of c-di-GMP. Eight of these genes were identified as having lower transcript levels in a *ctrA* null mutant strain (Mercer et al. 2010), suggesting possible interactions between the CtrA phosphorelay system and signaling via c-di-GMP. Indeed, I subsequently showed that four of these genes affect RcGTA gene expression and production and flagellar motility (Chapter 2), which are two processes also regulated by CtrA. I found that c-di-GMP inhibits RcGTA production and flagellar motility in *R. capsulatus*. The results presented in this chapter strengthen the initial findings and suggest that the link between c-di-GMP and these processes is likely, at least in part, due to switching CckA from kinase to phosphatase mode and thereby affecting the phosphorylation of CtrA.

4.5 References

Abel, S., Bucher, T., Nicollier, M., Hug, I., Kaever, V., Wiesch, P. A. Z., and Jenal, U. (2013). Bi-modal distribution of the second messenger c-di-gmp controls cell fate and asymmetry during the *Caulobacter* cell cycle. *PLoS Genet*. 9, e1003744. doi.org/10.1371/journal.pgen.1003744.

Bellefontaine, A.-F., Pierreux, C. E., Mertens, P., Vandenhaute, J., Letesson, J.-J., and De Bolle, X. (2002). Plasticity of a transcriptional regulation network among alpha-proteobacteria is supported by the identification of CtrA targets in *Brucella abortus*. *Mol. Microbiol.* 43, 945–960. doi:10.1046/j.1365-2958.2002.02777.x.

Biondi, E. G., Reisinger, S. J., Skerker, J. M., Arif, M., Perchuk, B. S., Ryan, K. R., et al. (2006). Regulation of the bacterial cell cycle by an integrated genetic circuit. *Nature* 444, 899–904. doi:10.1038/nature05321.

Bird, T. H.,, and MacKrell, A. (2011). A CtrA homolog affects swarming motility and encystment in *Rhodospirillum centenum*. *Arch. Microbiol.* 193, 451–59. doi.org/10.1007/s00203-011-0676-y.

Brilli, M., Fondi, M., Fani, R., Mengoni, A., Ferri, L., Bazzicalupo, M., et al. (2010). The diversity and evolution of cell cycle regulation in alpha-proteobacteria: a comparative genomic analysis. *BMC Syst. Biol.* 4, 52. doi:10.1186/1752-0509-4-52.

Calderon, F. R., (2020). Characterization of a signal transduction phosphorelay controlling *Rhodobacter capsulatus* gene transfer agent (RcGTA) gene expression. Master's Thesis. University of British Columbia

De Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., et al. (2006). ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* 34, W362–365. doi:10.1093/nar/gkl124.

Chambers, J.R., and Sauer, K. (2017). Detection of cyclic di-gmp binding proteins utilizing a biotinylated cyclic di-gmp pull-down assay. *Met. Mol. Bio*. 1657, 317-329. doi.org/10.1007/978-1-4939-7240-1_25.

Chen, Y. E., Tsokos, C. G., Biondi, E. G., Perchuk, B. S., and Laub, M. T. (2009). Dynamics of two phosphorelays controlling cell cycle progression in *Caulobacter crescentus*. *J. Bacteriol*. 191, 7417-7429. doi.org/10.1128/JB.00992-09.

Childers, W. S., Xu, Q., Mann, T. H., Mathews, I. I., Blair, A. J., Deacon, A. M., and Shapiro, L. (2014). Cell fate regulation governed by a repurposed bacterial histidine kinase. *PLoS Bio*. 12, 10. doi.org/10.1371/journal.pbio.1001979.

Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P., et al. (2006). Allosteric control of cyclic di-GMP signaling. *J. Biol. Chem.* 281, 32015–32024. doi:10.1074/jbc.M603589200.

Dubey, B. N., Lori, S. O., Geoffrey, F., Ivan, P. M., Urs, J., and Tilman, S., (2016). Cyclic di-GMP mediates a histidine kinase/phosphatase switch by noncovalent domain cross-linking. *Sci. Adv.* 2, 9. doi.org/10.1126/sciadv.1600823.

Francez, C. A., Kaczmarczyk, A., andVorholt, J. A (2015). The branched CcsA/CckA-ChpT-CtrA phosphorelay of *Sphingomonas melonis* controls motility and biofilm formation. *Mol. Microbiol*. 97, 47-63. doi.org/10.1111/mmi.13011.

Gao, R., andStock, A. M. (2009). Biological insights from structures of two-component proteins. *Ann. Rev. Microbiol*. 63, 133-154. doi.org/10.1146/annurev.micro.091208.073214.

Greene, S. E., Brilli, M., Biondi, E. G., and Komeili, A. (2012). Analysis of the CtrA

pathway in magnetospirillum reveals an ancestral role in motility in alphaproteobacteria. *J. Bacteriol.* 194, 2973–2986. doi:10.1128/JB.00170-12.
He, F. (2011). Bradford protein assay. *Bioprotocol* 1, e45. doi:10.21769/bioprotoc.45.
Hofmann, K., and Stoffel, W.(1993). TMbase: A database of membrane spanning protein segments. *Biol. Chem*. 374, 166. doi.org/10.1515/bchm3.1993.374.1-6.143.
Jacobs, C., Domian, I. J., Maddock, J. R., and Shapiro, L. (1999). Cell cycle-dependent polar localization of an essential bacterial histidine kinase that controls DNA replication and cell division. *Cell* 97, 111–120. doi:10.1016/S0092-8674(00)80719-9.
Andreas, K., Antje M. Hempel, Christoph von Arx, Raphael Böhm, Badri N. Dubey, B. N., Jutta Nesper, J., Tilman Schirmer, T., Sebastian Hiller, S., and Urs Jenal, U. (2020). Precise timing of transcription by c-di-gmp coordinates cell cycle and morphogenesis in *Caulobacter*.11, 816. *Nat. Comm*. doi.org/10.1038/s41467-020-14585-6.
Krogh, A., Larsson, B., Von, H. G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes. *J. Mol. Biol.*. 305, 567-580. doi.org/10.1006/jmbi.2000.4315.
Lang, A. S., and Beatty, J. T. (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. doi:10.1073/pnas.97.2.859.
Lang, A. S., and Beatty, J. T. (2002). A bacterial signal transduction system controls genetic exchange and motility. *J. Bacteriol.* 184, 913–918. doi:10.1128/jb.184.4.913-918.2002.
Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi:10.1093/bioinformatics/btm404.
Laub, M. T., Chen, S. L., Shapiro, L., and McAdams, H. H. (2002). Genes directly controlled by CtrA, a master regulator of the *Caulobacter* cell cycle. *Proc. Natl. Acad. Sci. U. S. A.* 99, 4632–4637. doi.org/10.1073/pnas.062065699.
Letunic, I. (2004). SMART 4.0: towards genomic data integration. *Nucleic Acids Res.* 32, 142D–144. doi:10.1093/nar/gkh088.
Lori, C., Ozaki, S., Steiner, S., Böhm, R., Abel, S., Dubey, B. N., et al. (2015). Cyclic di-GMP acts as a cell cycle oscillator to drive chromosome replication. *Nature* 523, 236–239. doi:10.1038/nature14473.
Mann, T. H., and Shapiro, L. (2018). Integration of cell cycle signals by multi-PAS domain kinases. *Proc. Natl. Acad. Sci. U. S. A.* 115, 7166–7173. doi.org/10.1073/pnas.1808543115.
Mercer, R. G., Callister, S. J., Lipton, M. S., Pasa-Tolic, L., Strnad, H., Paces, V., et al. (2010). Loss of the response regulator CtrA causes pleiotropic effects on gene expression but does not affect growth phase regulation in *Rhodobacter capsulatus*. *J. Bacteriol.* 192, 2701–2710. doi:10.1128/JB.00160-10.
Mercer, R. G., Quinlan, M., Rose, A. R., Noll, S., Beatty, J. T., and Lang, A. S. (2012). Regulatory systems controlling motility and gene transfer agent production and release in *Rhodobacter capsulatus*. *FEMS Microbiol. Lett.* 331, 53–62. doi:10.1111/j.1574-6968.2012.02553.x.
Mizuno, T.. (1997). Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli*. *DNA Research*. 4, 161-168.

doi.org/10.1093/dnares/4.2.161.
Paul, R., Weiser, S., Amiot, N. C., Chan, C., Schirmer, T., Giese, B., et al. (2004). Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev.* 18, 715–727. doi:10.1101/gad.289504.
Schirmer, T., and Jenal, U. (2009). Structural and mechanistic determinants of c-di-GMP signalling. *Nat. Rev. Microbiol.* 7, 724–735. doi:10.1038/nrmicro2203.
Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi:10.1038/nmeth.2089.
Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5857–5864. doi:10.1073/pnas.95.11.5857.
Skerker, J. M., and Laub, M. T. (2004). Cell-cycle progression and the generation of asymmetry in *Caulobacter crescentus*. *Nat. Rev. Microbiol.* 2, 325–337. doi:10.1038/nrmicro864
Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000). Two-component signal transduction. *Ann. Rev. Biochem.* 69,183–215.
Tsokos, C. G., Barrett, S. P., and Laub, M. T. (2011). A dynamic complex of signaling proteins uses polar localization to regulate cell-fate asymmetry in *Caulobacter crescentus*. *Dev. Cell.* 20, 329–341. doi.org/10.1016/j.devcel.2011.01.007.
Tsokos, C. G., and Laub, M. T. (2012). Polarity and cell fate asymmetry in *Caulobacter crescentus*. *Curr. Opin. Microbiol.* 15, 744–50. doi:10.1016/j.mib.2012.10.011.
Westbye, A. B., Kater, L., Wiesmann, C., Ding, H., Yip, C. K., and Beatty, J. T. (2018). The protease ClpXP and the PAS-domain protein DivL regulate CtrA and gene transfer agent production in *Rhodobacter capsulatus*. *Appl. Environ. Microbiol.* 84, e00275-18. doi:10.1128/AEM.00275-18.

CHAPTER 5- Summary and future directions

Over the past two decades, c-di-GMP-mediated regulation in bacteria has become a widely studied topic. It is now clear that this universal secondary messenger molecule controls a plethora of cellular activities in bacteria and the list keeps growing (Romling, Galperin, and Gomelsky 2013). The results presented in this thesis now add gene transfer to this growing list. In *R. capsulatus*, gene transfer between cells is mediated by bacteriophage-like particles called gene transfer agents (GTAs) (Marrs 1974; Yen, Hu, and Marrs 1979). The regulation of production of this GTA, called RcGTA, is complex as multiple regulatory systems affect RcGTA gene expression and release. These include the CckA-ChpT-CtrA phosphorelay (Lang and Beatty 2000; Mercer et al. 2012), the GtaI-GtaR- quorum sensing system (Leung et al. 2012; Schaefer et al. 2002) and additional regulators that either directly or indirectly affect the production of GTA (Mercer and Lang 2014; Hynes et al. 2016; Fogg 2019; Kuchinski et al. 2016; Westbye et al. 2018). Among all of the regulators that were identified, CtrA was particularly important because it remained for a long time the only regulator whose absence resulted in a complete loss of RcGTA production. It was subsequently determined that this is because CtrA is required for the transcription of another regulatory gene (Hynes et al. 2016) that acts at the RcGTA structural gene cluster promoter (Fogg 2019). CtrA also has other regulatory roles in *R. capsulatus* and its loss affects the transcript levels of more than 200 genes (Mercer et al. 2010). As part of the investigation to unravel and understand the complex downstream signalling pathways connecting CtrA and RcGTA production, I studied the eight chromosomally encoded c-di-GMP signalling genes that were identified in a transcriptomic study as having lowered transcripts in the absence of CtrA (Mercer et al.

2010). In my research I established a definitive link between c-di-GMP and RcGTA production in *R. capsulatus*, identified specific mechanisms that control the functioning of c-di-GMP signaling proteins in this bacterium, and began to explore a mechanistic link between c-di-GMP and RcGTA regulation via the histidine kinase CckA.

The first and major part of my thesis involved investigating the roles of genes encoding putative c-di-GMP signaling proteins in RcGTA production. Among the eight putative c-di-GMP signalling genes that were investigated by genetic approaches, I found four genes (*rcc00620*, *rcc00645*, *rcc02629*, and *rcc02857*) that were affecting RcGTA production. Mutations of the genes had differing effects on RcGTA production, but the issue of whether the proteins might be responsible for increasing or lowering c-di-GMP levels in cells was complicated by the fact that three of them have both DGC and PDE domains, which are involved in the synthesis and degradation of c-di-GMP, respectively. The fourth protein, Rcc02629, only has a DGC domain. Expression of heterologous genes encoding established DGC and PDE enzymes, site-directed mutagenesis of key catalytic residues in the *R. capsulatus* proteins, quantification of c-di-GMP levels in various experimental strains, and phenotypic assays in *E. coli* strains expressing the *R. capsulatus* genes allowed me to determine the enzymatic activities of the different proteins, with respect to RcGTA regulation, and that c-di-GMP inhibits RcGTA production. For the putatively bi-functional proteins, two act as DGCs and inhibit RcGTA production whereas the third acts as a PDE and positively regulating RcGTA production. For Rcc00645, which has a specific type of PAS domain predicted to be involved in binding to heme, I found that the availability of oxygen affects its activity. Furthermore, these c-di-GMP signalling genes also affected flagellar motility and I showed that increased c-di-

GMP levels decreases motility in *R. capsulatus*, which has been commonly observed in other bacteria (Romling, Galperin, and Gomelsky 2013).

In Chapter 3, I characterized a predicted two-component signalling system (TCS), where the histidine kinase (HK) protein Rcc00621 phosphorylates the cognate response regulator (RR) protein Rcc00620, which switches this putatively bifunctional protein to act as a PDE. This was done using a combination of site-directed mutagenesis of key catalytic residues in the *R. capsulatus* proteins, quantification of c-di-GMP levels in various experimental strains, *in vitro* phosphotransfer assays with purified recombinant proteins, and phenotypic assays in *E. coli* strains expressing the *R. capsulatus* genes. We derived the evolutionary history of the genes encoding the TCS proteins using phylogenetic analysis and documented key motif conservation across alphaproteobacterial species. This showed that these genes are only present in a few phylogenetically related *Rhodobacter* species and that they have likely been acquired horizontally by an ancestral *Rhodobacter* from a bacterium in the *Sphingomonadales*, a different order within the class *Alphaproteobacteria*. Members of this order also possess GTA gene clusters, raising the possibility that the HGT event was mediated by a GTA. In addition to the effects of this TCS on RcGTA production, I also showed that it is regulating motility in *R. capsulatus* via c-di-GMP.

The CckA-ChpT-CtrA phosphorelay system is widespread in alphaproteobacteria and best-studied in *C. crescentus* (Brilli et al. 2010). It is now known from work in *C. crescentus* that the binding of c-di-GMP to CckA causes it to switch from kinase to phosphatase activity and this thereby modulates the phosphorylation state of CtrA (Lori et al. 2015; Dubey et al. 2016). In the last chapter I confirmed that c-di-GMP binds to CckA

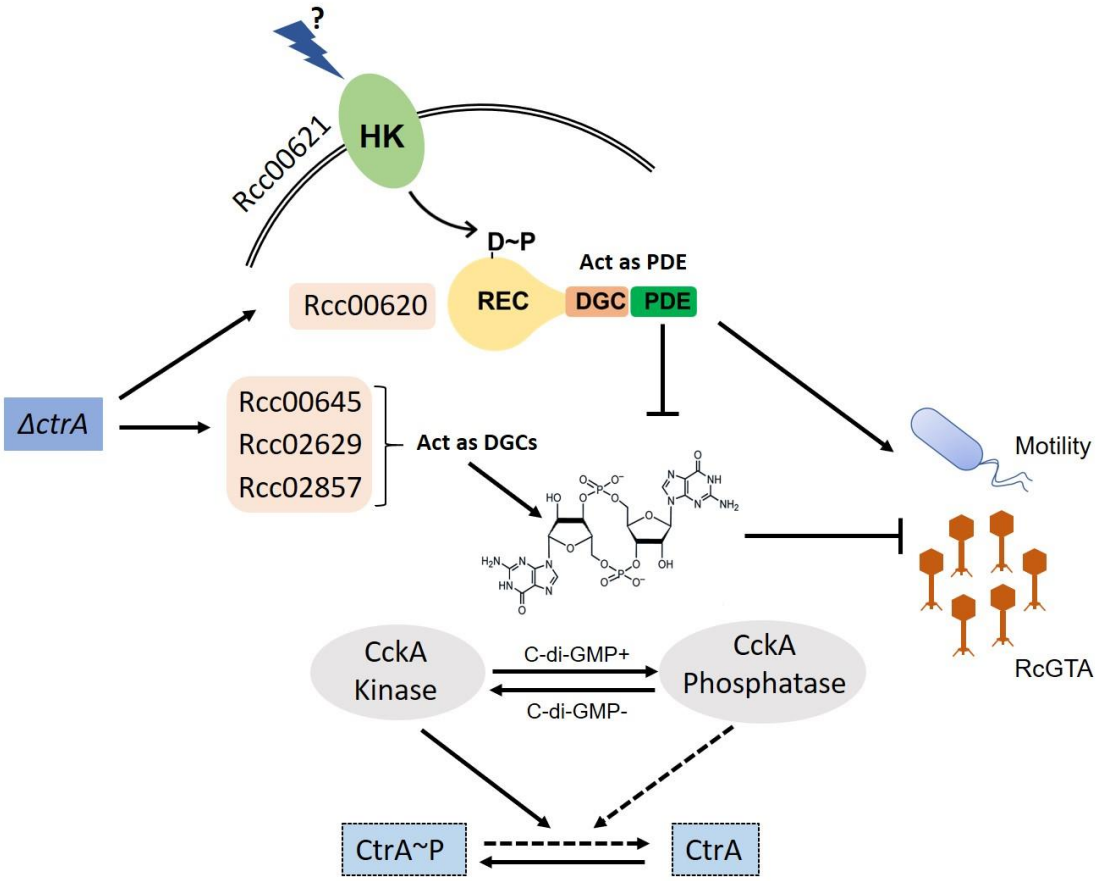


Figure 5.1. Overall model for the proposed c-di-GMP-mediated regulation of motility and RcGTA production in *R. capsulatus*. The response regulator CtrA affects the transcription of four genes encoding predicted c-di-GMP signalling proteins: Rcc00620, Rcc00645, Rcc02629 and Rcc02857. Three of the proteins act as DGCs and synthesize c-di-GMP and thereby act as negative regulators of RcGTA production and motility. Rcc00620 act as PDE and promotes c-di-GMP hydrolysis, RcGTA production and motility. Rcc00620 is regulated by the histidine kinase protein Rcc00621, which gets activated upon sensing an unknown stimulus and undergoes autophosphorylation at a conserved histidine residue and transfers the phosphoryl group to a conserved aspartate residue of Rcc00620. The levels of c-di-GMP affect the CckA-ChpT-CtrA phosphorelay,

where c-di-GMP binds to CckA and modulates its kinase versus phosphatase activities and leads to the dephosphorylation of CtrA.

in *R. capsulatus*. Constructs that I made were also used by another research group for collaborative work to study *in vitro* phosphotransfer among the CckA, ChpT and CtrA proteins. These assays confirmed the occurrence of CckA to ChpT to CtrA phosphotransfers, and also showed that c-di-GMP stimulates CtrA dephosphorylation (Calderon 2020). Assays performed using the CckA mutants I created in this study showed that the Y589D mutant showed decreased phosphatase activity and normal kinase activity and confirmed the Y589 residue is crucial for c-di-GMP binding. I also tested the possible role of CckA I-site motifs in c-di-GMP binding because of their role in this activity in other proteins, however my preliminary data did not support this as a function for these sites in CckA.

Overall, the findings from my thesis provide some new insights towards understanding the connection of various regulatory systems and RcGTA production and motility in *R. capsulatus*. However, further research will still need to be done to address some additional questions related to these findings. Are there other c-di-GMP-binding proteins besides CckA that are directly or indirectly involved in RcGTA and motility regulation? Another key question is about the stimuli that are controlling these signaling proteins. I found that oxygen is likely a key regulator of one of the DGC enzymes (Rcc00645), but what activates the Rcc00621 HK protein and triggers the reduction of c-di-GMP levels through the activity of Rcc00620? How did a regulatory unit acquired by horizontal gene transfer come to regulate gene transfer activity in the new “host” organism? There is a clear connection between c-di-GMP and the functioning of the

CckA-ChpT-CtrA phosphorelay, but it is not fully clear yet if this is the only way that c-di-GMP might be connected to regulation of the affected cellular behaviors. Among the various important bacterial activities such as motility, biofilm formation, and cell cycle processes, I now added gene transfer as a new behavior affected by c-di-GMP.

5.1 References

Brilli, M., Fondi, M., Fani, R., Mengoni, A., Ferri, L., Bazzicalupo, M., et al. (2010). The diversity and evolution of cell cycle regulation in alpha-proteobacteria: A comparative genomic analysis. *BMC Syst. Biol.* 4, 52. doi:10.1186/1752-0509-4-52.

Calderon, F. R., (2020). Characterization of a signal transduction phosphorelay controlling *Rhodobacter capsulatus* gene transfer agent (RcGTA) gene expression. Master's Thesis. University of British Columbia.

Dubey, B. N., Lori, S. O., Geoffrey, F., Ivan, P. M., Urs, J., and Tilman, S., (2016). Cyclic di-GMP mediates a histidine kinase/phosphatase switch by noncovalent domain cross-linking. *Sci. Adv.* 2, 9. doi.org/10.1126/sciadv.1600823.

Fogg, P. C. M. (2019). Identification and characterization of a direct activator of a gene transfer agent. *Nat. Commun.* 10, 595. doi:10.1038/s41467-019-08526-1.

Hynes, A. P., Shakya, M., Mercer, R. G., Grüll, M. P., Bown, L., Davidson, F., et al. (2016). Functional and evolutionary characterization of a gene transfer agent's multilocus "genome." *Mol. Biol. Evol.* 33, 2530–2543. doi:10.1093/molbev/msw125.

Kuchinski, K. S., Brimacombe, C. A., Westbye, A. B., Ding, H., and Beatty, T. J. (2016). The SOS response master regulator LexA regulates the gene transfer agent of *Rhodobacter capsulatus* and represses transcription of the signal transduction protein CckA. *J. Bacteriol.* 198, 1137–1148. doi:10.1128/JB.00839-15.

Lang, A. S., and Beatty, J. T. (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. doi:10.1073/pnas.97.2.859.

Leung, M. M., Brimacombe, C. A., Spiegelman, G. B., and Beatty, J. T. (2012). The GtaR protein negatively regulates transcription of the *gtaRI* operon and modulates gene transfer agent (RcGTA) expression in *Rhodobacter capsulatus*. *Mol. Microbiol.* 83, 759–774. doi:10.1111/j.1365-2958.2011.07963.x.

Lori, C., Ozaki, S., Steiner, S., Böhm, R., Abel, S., Dubey, B. N., et al. (2015). Cyclic di-GMP acts as a cell cycle oscillator to drive chromosome replication. *Nature* 523, 236–239. doi:10.1038/nature14473.

Marrs, B. (1974). Genetic recombination in *Rhodopseudomonas capsulata*. *Proc. Natl. Acad. Sci. U. S. A.* 71, 971–973. doi:10.1073/pnas.71.3.971.

Mercer, R. G., Callister, S. J., Lipton, M. S., Pasa-Tolic, L., Strnad, H., Paces, V., et al. (2010). Loss of the response regulator CtrA causes pleiotropic effects on gene expression but does not affect growth phase regulation in *Rhodobacter capsulatus*. *J. Bacteriol.* 192, 2701–2710. doi:10.1128/JB.00160-10.

Mercer, R. G., and Lang, A. S. (2014). Identification of a predicted partner-switching

system that affects production of the gene transfer agent RcGTA and stationary phase viability in *Rhodobacter capsulatus*. *BMC Microbiol.* 14, 71. doi:10.1186/1471-2180-14-71.

Mercer, R. G., Quinlan, M., Rose, A. R., Noll, S., Beatty, J. T., and Lang, A. S. (2012). Regulatory systems controlling motility and gene transfer agent production and release in *Rhodobacter capsulatus*. *FEMS Microbiol. Lett.* 331, 53–62. doi:10.1111/j.1574-6968.2012.02553.x.

Römling, U., Galperin, M. Y., and Gomelsky, M. (2013). Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol. Mol. Biol. Rev.* 77, 1–52. doi:10.1128/MMBR.00043-12.

Schaefer, A. L., Terumi, T. A., Beatty, J. T., and Greenberg, E. P. (2002). Long-chain acyl-homoserine lactone quorum-sensing regulation of *Rhodobacter capsulatus* gene transfer agent production. *J. Bacteriol.* 184, 6515–6521. doi.org/10.1128/JB.184.23.6515-6521.2002.

Westbye, A. B., Kater, L., Wiesmann, C., Ding, H., Yip, C. K., and Beatty, J. T. (2018). The protease ClpXP and the PAS-domain protein DivL regulate CtrA and gene transfer agent production in *Rhodobacter capsulatus*. *Appl. Environ. Microbiol.* 84, e00275-18. doi:10.1128/AEM.00275-18.

Yen, H. C., Hu, N. T., and Marrs, B. L. (1979). Characterization of the gene transfer agent made by an overproducer mutant of *Rhodopseudomonas capsulata*. *J. Mol. Biol.* 131, 157–168. doi:10.1016/0022-2836(79)90071-8.

Appendix 1: Supplementary figures and tables for chapter 2

Table S2.1. List of primers used in this study.

Primer	Sequence (5'-3')[a]
346-F	CGCGCGGTGTCCGGAATGA
346-R	CCGTCCTGCCCGACTTCCTGA
620-F	CGGCGATCTATGTGGTGGCGG
620-R	TTGCAGCGCCATCGTGTCGAA
645-F	TAGGTACCTTTCCAAGGCGACGAC
645-R	TAGGTACCTTTGTGTTCGCTCTCTTGC
2539-F	TTCCATGCCGAAATAGGCCGC
2539-R	GGCGCCGTCGTCGATCTGAAT
2540-F	AAAGGCGCGGTGCAGCATGAA
2540-R	ACCGCGGAAGAACGTGAAGCC
2629-F	ATGCCCCGGCCGCGCCAGCA
2629-R	CCTTGGCCCCGAGGGGTTCG
2857-F	CACAGGATGCGGAACACCCCG
2857-R	CGAACCAGCACACCGAGGCAT
3177-F	CAGCCGCACCGACAGATAGCC
3177-R	CCGCATCGCCAAGCTGGATGT
3301-F	GGCCCGATGGGCTTTGCCAT
3301-R	CCCGAGGCCGAGGAATGGGA
620-CompF	TAGGTACCGGCTCGCTTTGGCT
620-CompR	TAGGTACCCAAGCACCACCGCA
645-CompF	TAGGTACCTTTCCAAGGCGACGAC
645-CompR	TAGGTACCTTTGTGTTCGCTCTCTTGC
2629-CompF	TAGGTACCGACAGGGACACGGTCTCG
2629-CompR	TAGGTACCCCAGAAGGCCGATG
2857-CompF	TAGGTACCCGAAATCGTCTCGGT
2857-CompR	TAGGTACCGCAATGTGCTGACCT
DGC-F	AGGAGGGCCATGATGCAGGACTGCGAGAAACT

DGC-R	TGGTCCGCAGAAGGCG
PDE-F	AGGAGGACGACGATGCCCGACATCACAGCCCTC
PDE-R	TGGCTCCATCTATCCCGTAGCA
620-GGAAF-F	CGGCTGGGCGGCGCCGCATTCGGCGTGCTG
620-GGAAF-R	CAGCACGCCGAATGCGGCGCCGCCCAGCCG
645-GGAAF-F	CGGCTTGGCGGCGCCGCATTCTGCCTGCTG
645-GGAAF-R	CAGCAGGCAGAATGCGGCGCCGCCAAGCCG
2629-GGAAF-F	CGGCTGGGCGGCGCCGCATTCATCATCCTG
2629-GGAAF-R	CAGGATGATGAATGCGGCGCCGCCCAGCCG
2857-GGAAF-F	CGTCTTGGCGGCGCCGCATTCTCGGTGATC
2857-GGAAF-R	GATCACCGAGAATGCGGCGCCGCCAAGACG
620-AAL-F	GTGGTGGGTTATGCGGCGCTCTTGCGCTGG
620-AAL-R	CCAGCGCAAGAGCGCCGCATAACCCACCAC
645-AVL-F	CTGGCCGGGGTCGCGGTTCTGGTACGCTGG
645-AVL-R	CCAGCGTACCAGAACCGCGACCCCGGCCAG
2857-ATL-F	CTGACCGGGGTCGCGACGCTGCTGCGCTGG
2857-ATL-R	CCAGCGCAGCAGCGTCGCGACCCCGGTCAG

[a] Underlined sequences indicate restriction sites or ribosome binding sites added for cloning or protein translation purposes, respectively.

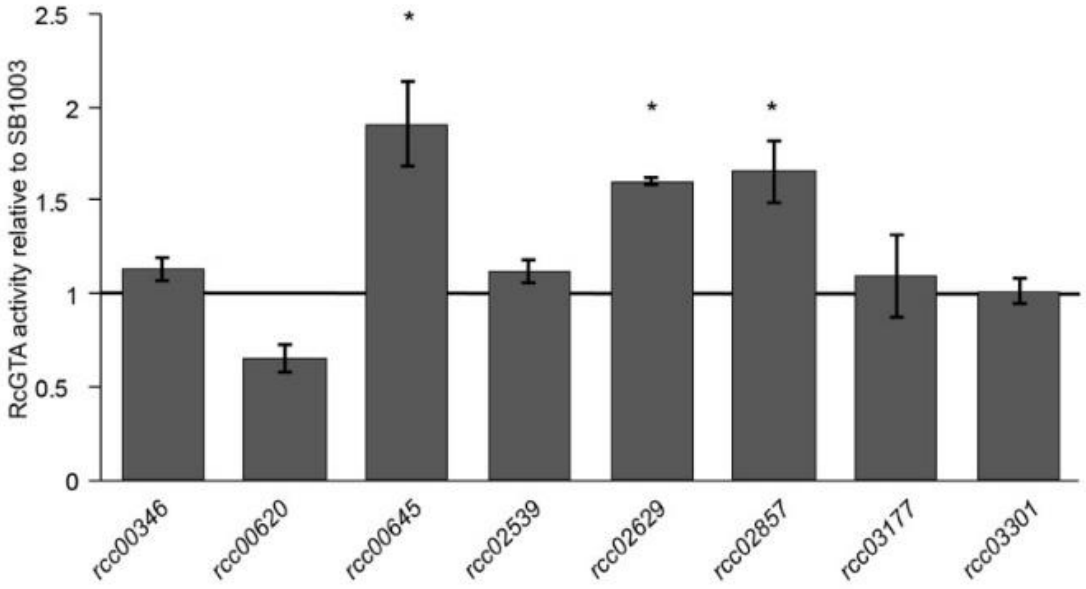


Figure S2.1. Effects of disruptions of genes encoding putative c-di-GMP signaling proteins on RcGTA gene transfer activity. The gene transfer activities for eight predicted cyclic-di-GMP gene mutants are presented as averages from three replicates relative to the parental strain, SB1003, and the bars represent the standard deviations. Statistically significant differences ($p < 0.05$) compared to the control, identified using one-way analysis of variance (ANOVA) followed by Tukey HSD post-hoc analysis, are indicated by asterisks.

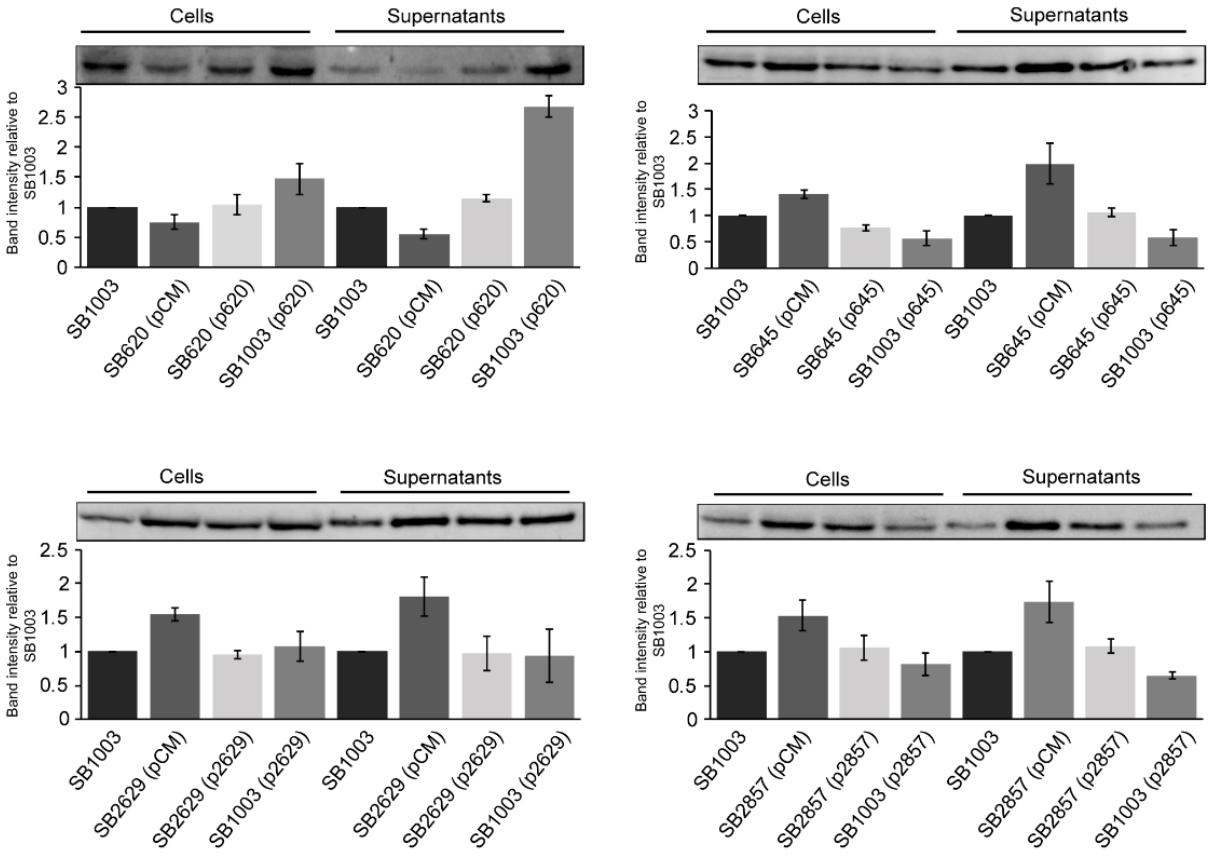


Figure S2.2. Western blots for the RcGTA major capsid protein in cells and culture supernatants. Blots were performed on all replicate gene transfer bioassay cultures (Figure 2A) and one representative set of blots is shown for each. Graphs represent the average band intensities from the replicate blots relative to the parental strain, SB1003, calculated using ImageJ, and bars represent the standard deviations.

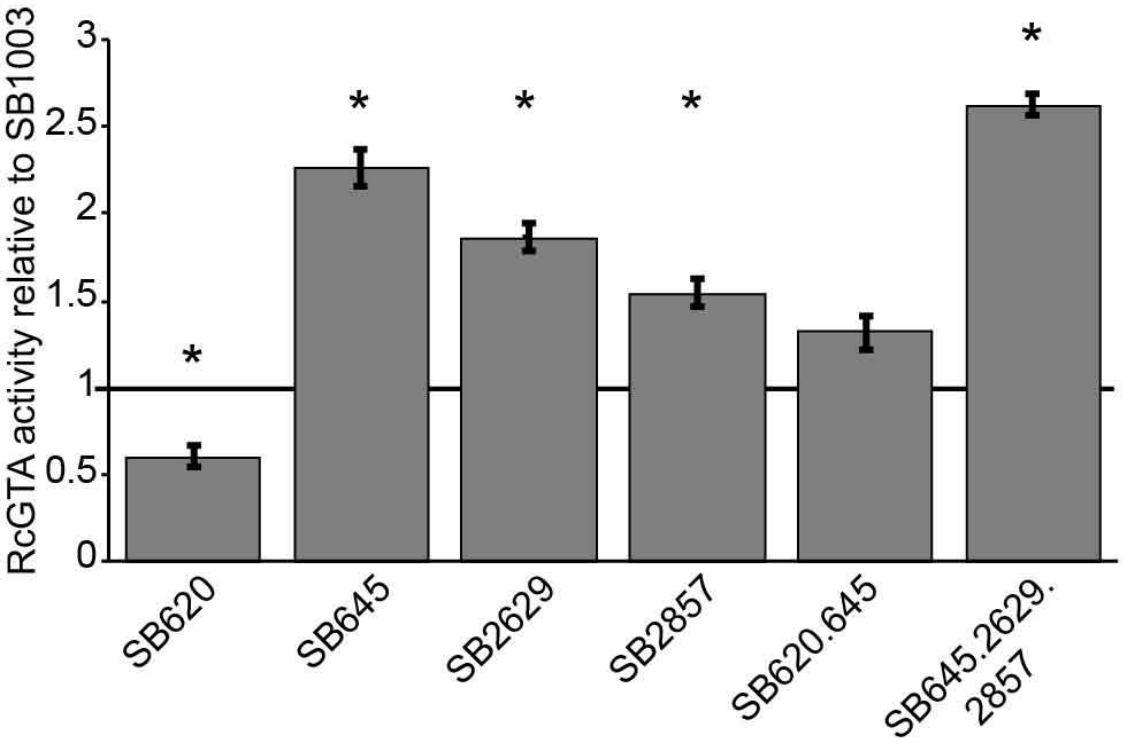


Figure S2.3. Effects of multiple gene disruptions on RcGTA gene transfer activity. The gene transfer activities for a double mutant (SB620.645) and a triple mutant (SB645.2629.2857) are compared to each of the single gene mutants (data from Figure 2A). The data are presented as averages from three replicates relative to the parental strain, SB1003, and the bars represent the standard deviations. Statistically significant differences ($p < 0.05$) compared to the control, identified using one-way ANOVA followed by Tukey HSD post-hoc analysis, are indicated by asterisks.

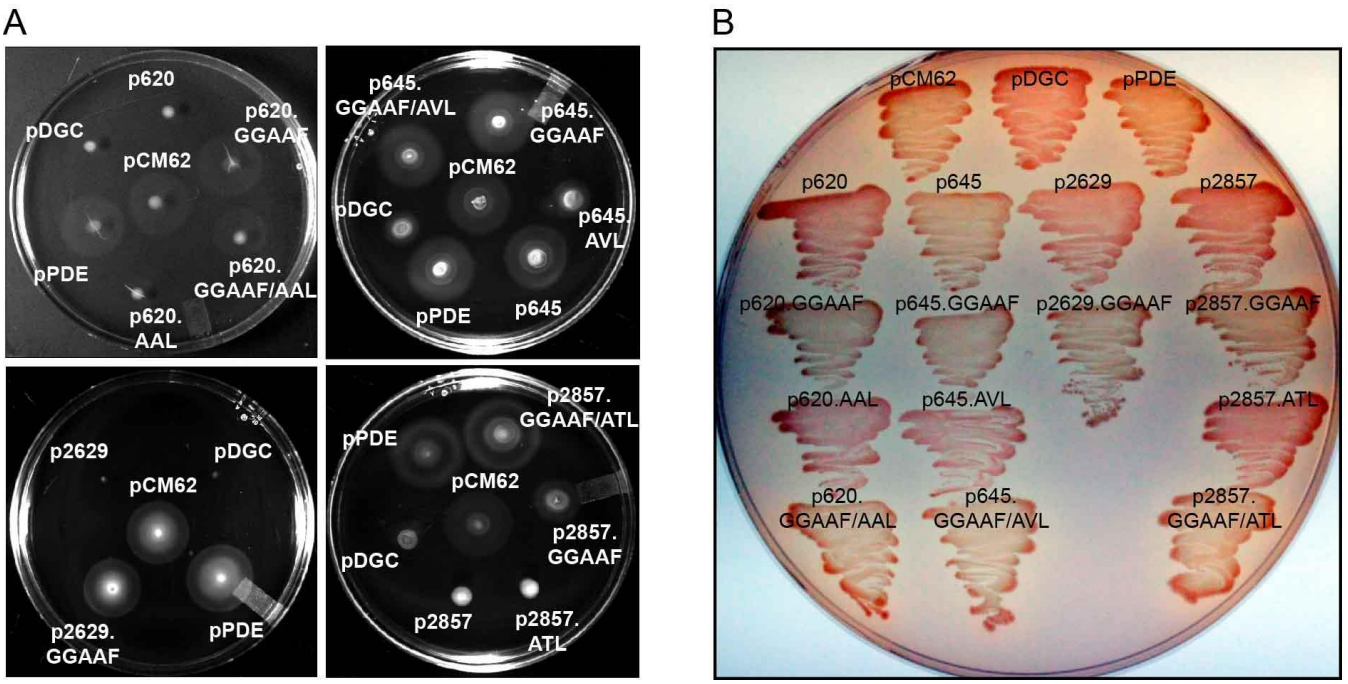


Figure S2.4. Effect of GGDEF and EAL domain mutations on DGC activity in *E. coli*. A. Motility of *E. coli* MG1655 on semi-solid agar (0.25%), which is reduced by DGC activity, when containing the indicated plasmids. Transcription from the plasmid's *lac* promoter was induced with IPTG. B. Congo red binding by *E. coli* BL21(DE3), where DGC activity increases fimbriae production and Congo red binding, when containing the indicated plasmids. Transcription from the plasmid's *lac* promoter was induced with IPTG.

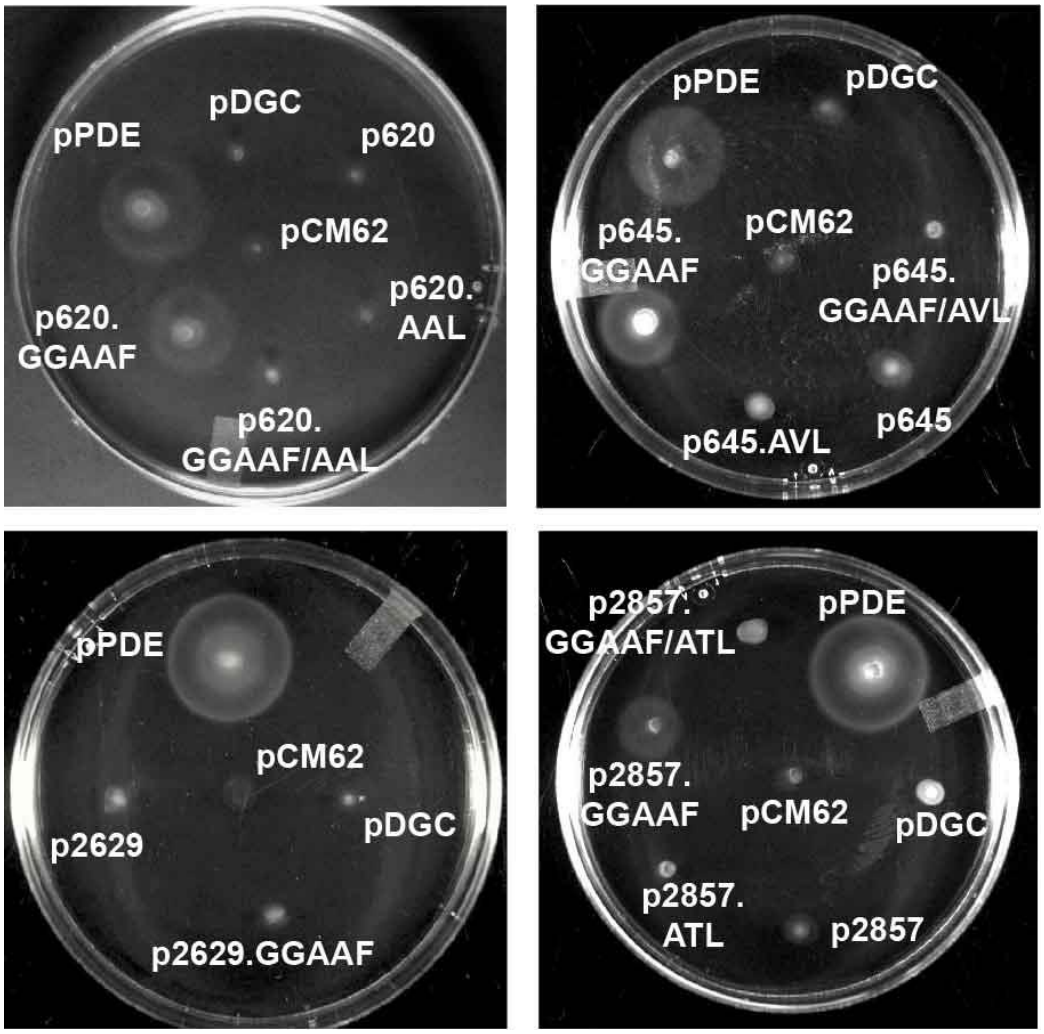


Figure S2.5. Evaluation of *R. capsulatus* proteins with point mutations in GGDEF and EAL domains for potential PDE activities in *E. coli*. Restoration of motility to *E. coli* MG1655 Δ*yhjH* on semi-solid media (0.25%) indicates PDE activity. Transcription from the plasmid's *lac* promoter was induced with IPTG.

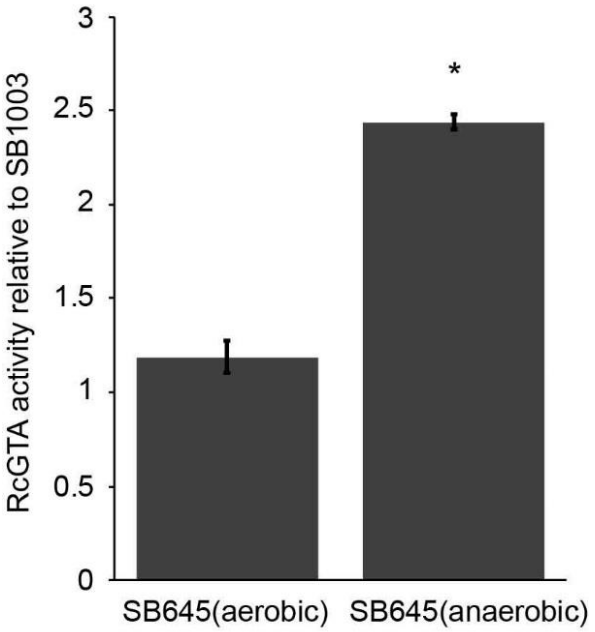


Figure S2.6. Effect of oxygen on gene transfer activity in strain SB645. The *R. capsulatus* strains SB1003 and SB645 were grown under both aerobic heterotrophic and anaerobic photoheterotrophic conditions and assayed for RcGTA production. Data for SB645 are presented as averages from three replicates relative to the parental strain and the bars represent the standard deviations. The asterisk indicates a statistically significant difference ($p < 0.05$) compared to the control, identified using one-way ANOVA followed by Tukey HSD post-hoc analysis.

Appendix 2: Supplementary figures and tables for chapter 3

Table S3.1. *rcc00620*/*rcc00621* accession numbers. (xls.file)

Table S.3.2. List of primers used in this study.

Primer	Sequence (5'-3')[a]
620-F	CGGCGATCTATGTGGTGGCGG
620-R	TTGCAGCGCCATCGTGTCGAA
620-KpnI-F	TAGGTACCGGCTCGCTTTGGCT
620-KpnI-R	TAGGTACCCAAGCACCACCGCA
621-KpnI-R	TAGGTACCCGGCGTCACGGGACACGATG
620-D86A-F	AAGGTCGCCTTCGTCGCTGTGCGGATGCCG
620-D86A-R	CGGCATCCGCACAGCGACGAAGGCGACCTT
620-D86E-F	AAGGTCGCCTTCGTCGAGGTGCGGATGCCG
620-D86E-R	CGGCATCCGCACCTCGACGAAGGCGACCTT
621-H361N-F	ACCGGGCTTTTGAACAATCTGCGCAACAGT
621-H361N-R	ACTGTTGCGCAGATTGTTCAAAAGCCCGGT
621-H326N-F	ATCGGCGCCCTGAACCGGGATTTCAACGAG
621-H326N-R	CTCGTTGAAATCCCGGTTCAGGGCGCCGAT
pET_620-F	TATACCATGGATGACCATTCTCGTCATTGATGAC
pET_620-R	TATAAGCTTGCCCGCGGCACTGCGGCCGGGTTC
pET_621-F	TATACCATGGCGGCGGCTGTCGACGCAGGT
pET_621-R	TATAAGCTTGCCCGCCGGGCGCAACC
621_ML-F	CTTGCGCTGCTGGGCCTGCTGGCGGCG
621_ML-F1	CCCGGGACCGGGGCGCCCGACCCCGAT
620_ML-F	GCAGTTTCATCGAGGGGATGGACCGC

[a] Underlined sequences indicate restriction sites added for cloning purposes.

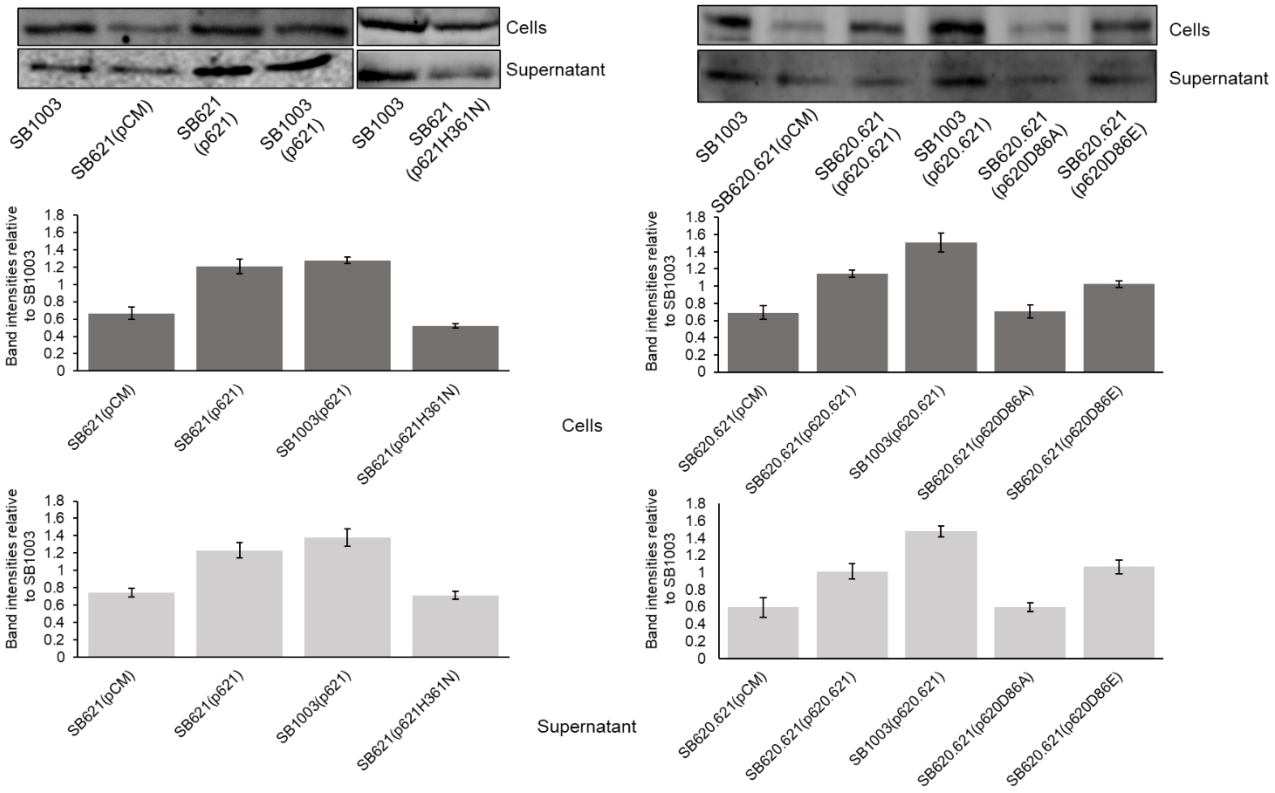


Figure S3.1. Western blotting quantification of the RcGTA major capsid protein. Blots were perfomed on all replicate gene transfer bioassay cultures and one representative set of blots is shown for each strain. Graphs show the average band intensities from the replicate blots relative to parental strain, SB1003, calculated using ImajeJ. The bars represent standard deviations.

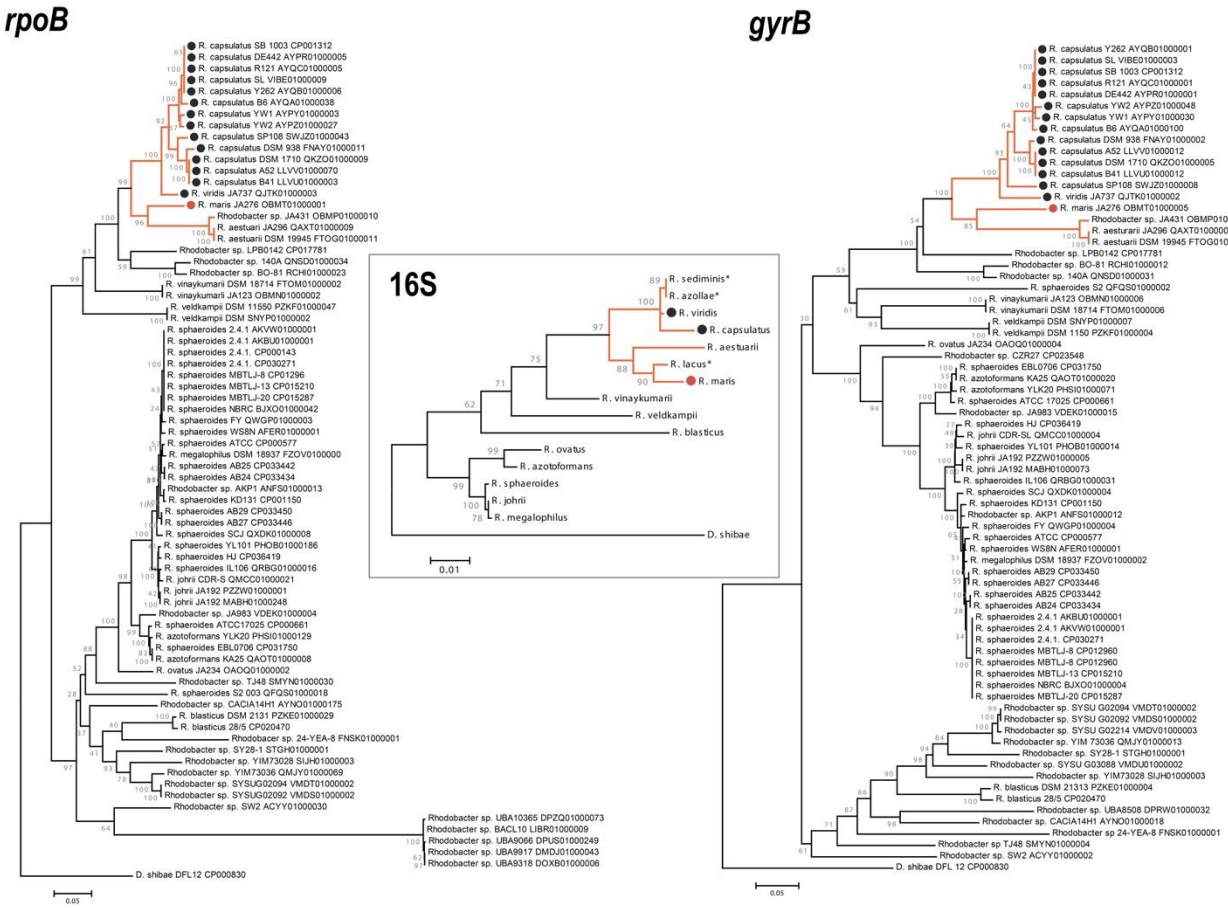


Figure S3.2. Phylogenetic relationships among members of the genus *Rhodobacter*. Phylogenetic trees based on the RNA polymerase beta subunit gene (*rpoB*), 16S rDNA, and DNA gyrase B gene (*gyrB*) were obtained with the maximum-likelihood method with MEGA 7. The General Time Reversible (*rpoB* and *gyrB*) and Tamura 3-parameters (16S) models, identified as the best fitting models after a model test analysis, were used to estimate genetic distances between sequences, a discrete Gamma distribution was used to model evolutionary rate differences among sites, and the rate variation model allowed for some sites to be evolutionarily invariable. The outcome of the bootstrap analysis is shown next to the nodes. Strains in which *rcc00620-621* homologs were identified are labelled with a dot whose color corresponds to the genomic organizations surrounding these genes

(Figure 7 in the main text). The clade including descendants of the receiver of the two genes is highlighted in orange.

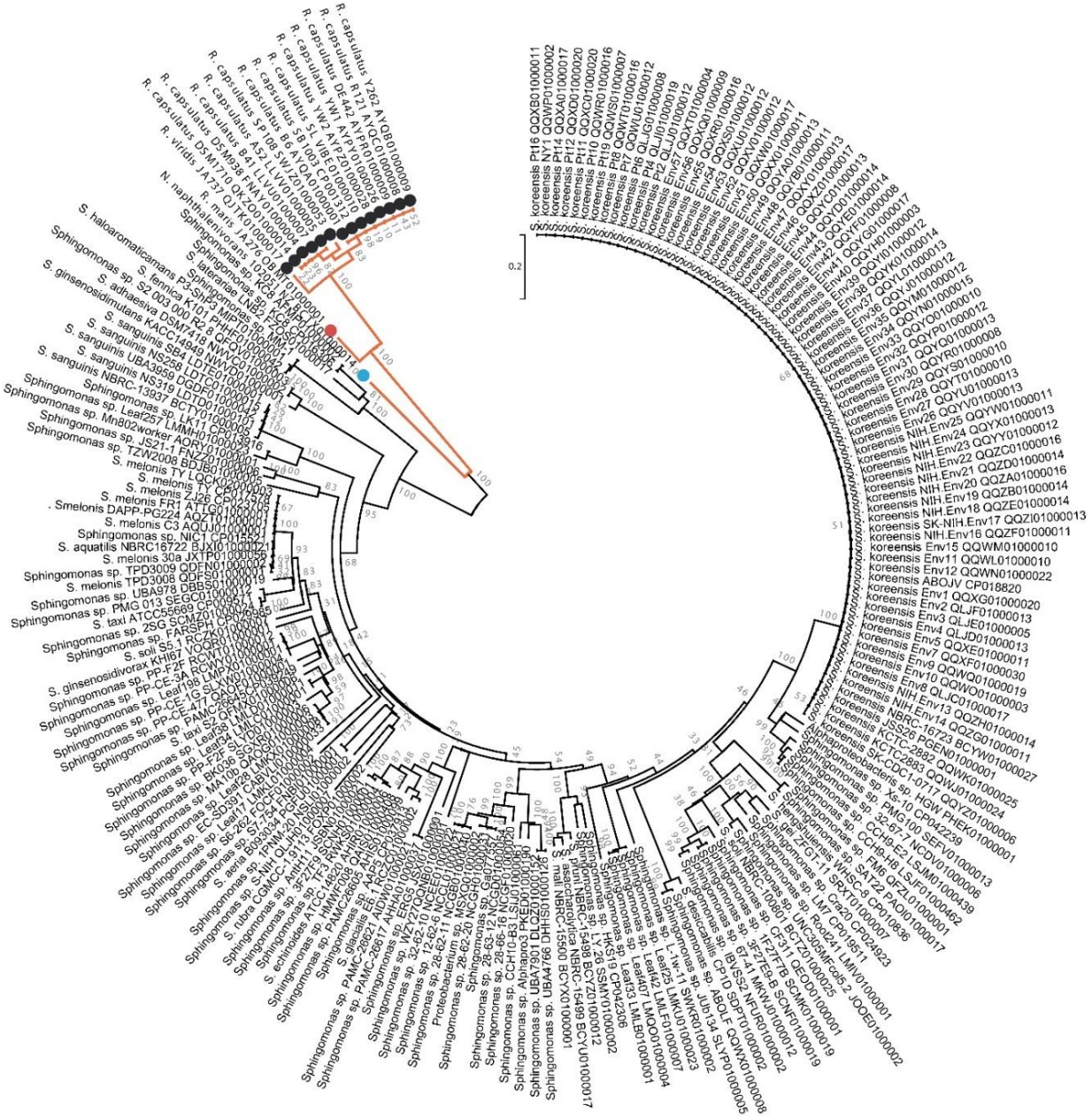


Figure S3.3. Phylogenetic analysis of the Rcc00620 protein homologs found in Alphaproteobacteria. The phylogenetic trees was built with MEGA 7 with the neghbor-joining method using the JTT model , identified as the best fitting model after a model test analysis. A discrete Gamma distribution was used to model evolutionary rate

differences among sites (+G=0.6) and the outcome of the bootstrap analysis is shown next to the nodes. *Rhodobacter* strains in which the two genes were identified are labelled with a black or red dot whose color corresponds to the genomic organizations (Figure 7 in the main text), while the *Novosphingobium* strain is indicated by a blue dot. The clade including *Rhodobacter* and *Novosphingobium* strains is highlighted in orange.

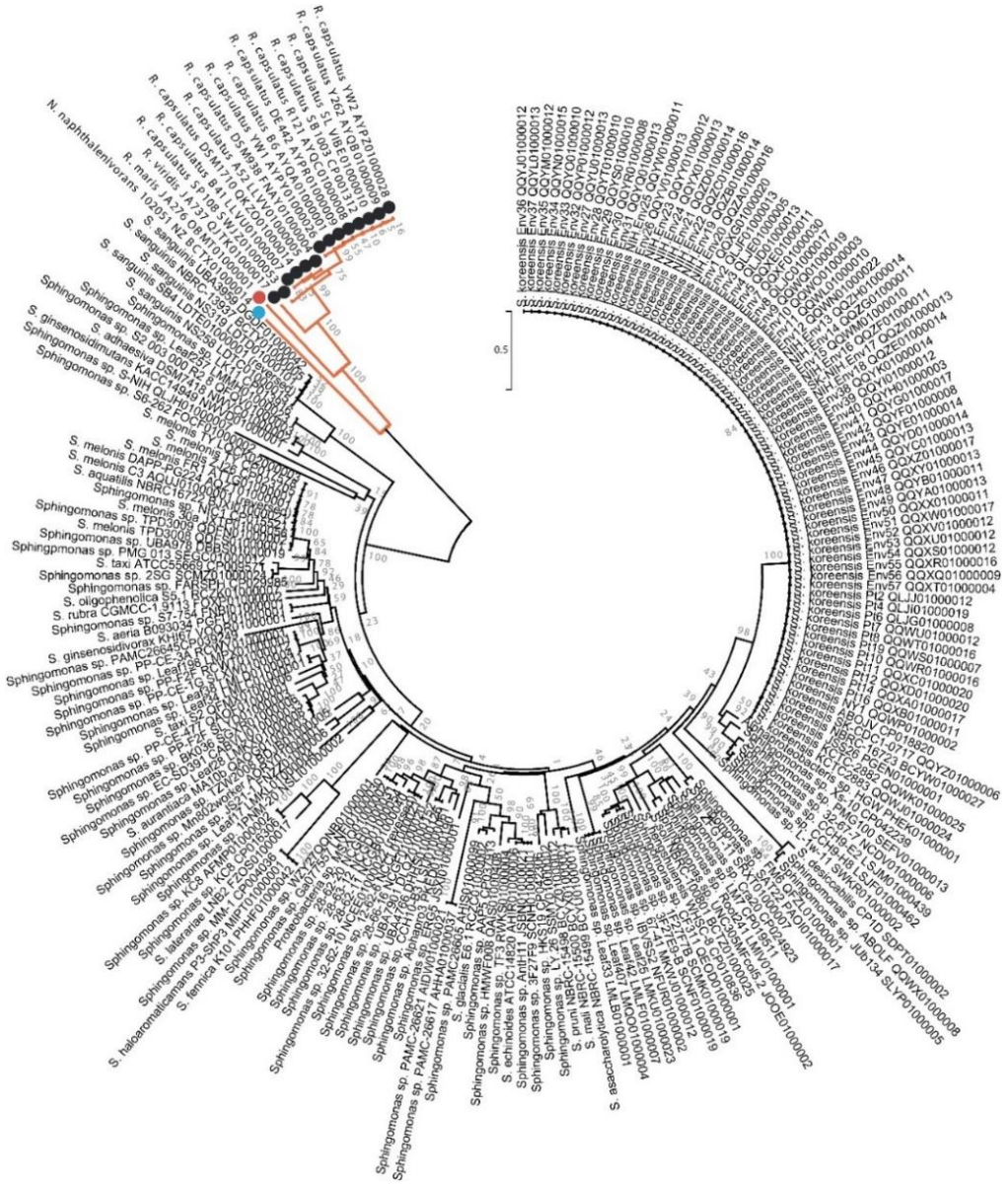


Figure S3.4. Phylogenetic analysis of the Rcc0621 homologs found in Alphaproteobacteria. The phylogenetic trees was built with MEGA 7 with the neghbor-joining method using the JTT model, identified as the best fitting model after a model test analysis. A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G=0.75) and the outcome of the bootstrap analysis is shown next to the nodes. *Rhodobacter* strains in which the two genes were identified are labelled with a black or red dot whose color corresponds to the genomic organizations (Figure 7 in the

main text), while the *Novosphingobium* strain is indicated by a blue dot. The clade including *Rhodobacter* and *Novosphingobium* strains is highlighted in orange.

Appendix 3: Supplementary tables for chapter 4

Table S4.1. List of primers used in this study

Primer	Sequence (5'-3')
CckA_ΔTM-F	TATACCATGGCGGGGGCGATTTCCGCC
CckA_ΔTM-R	TATAAGCTTGGCCCGCGCCCGGCGG
CckA_G395T-F	GGCCAGCTTGCGACCGGGGTTGCGCATGATTTC
CckA_G395T-R	GAAATCATGCGCAACCCCGGTCGCAAGCTGGCC
CckA_G396E-F	GGCCAGCTTGCGGGCGAGGTTGCGCATGATTTC
CckA_G396E-R	GAAATCATGCGCAACCTCGCCCGCAAGCTGGCC
CckA_H399A-F	GGCGGGGTTGCGGCTGATTTCAACAACTTG
CckA_H399A-R	CAAGTTGTTGAAATCAGCCGCAACCCCGCC
CckA_V443P-F	GCGGCCTCGCTGCCCGGGCAGCTTCTGGCG
CckA_V443P-R	CGCCAGAAGCTGCCCGGGCAGCGAGGCCGC
CckA_Y589D-F	CTGGGGCTCTCGACCGCCGACGGGATCGTCAAG
CckA_Y589D-R	CTTGACGATCCCGTCGGCGGTCGAGAGCCCCAG
CckA_1st_Isite-F	TTCCTGTGGGGGCTGGAGCACATGGCGGAA
CckA_1st_Isite-R	TTCCGCCATGTGCTCCAGCCCCCACAGGAA
CckA_2nd_Isite-F	GAGCGTGTCGGCACGCTCGGCCGGATCTTC
CckA_2nd_Isite-R	GAAGATCCGGCCGAGCGTGCCGACACGCTC
CckA_3rd_Isite-F	GGGCTGGGCGGTCCGGTGCATGACTGGGTT
CckA_3rd_Isite-F	AACCCAGTCATGCACCGGACCGCCCAGCCC
CckA_4th_Isite-F	GTGCTGCGGGCGGGGCGCGGCGGCCGCGAGGTC
CckA_4th_Isite-R	GACCTCGCGGCCGCCGCGCCCCGCCCGCAGCAC
CckA_5th_Isite-F	CTGAAGCCGGGGATCATCGGTCTGCGCGAC
CckA_5th_Isite-R	GTCGCGCAGACCGATGATCCCCGGCTTCAG
CckA_ML_FP1	GGCGTCGAAATCGGGCACGATC
CckA_ML_FP2	GCAGCTGAAGACGCTGGAAGGCC