Analysis of key *Pseudomonas fluorescens* genes controlling rhizosphere colonisation and carbon metabolism

Rowena Ka Ying Fung B.Sc. (Hons) MRes

(University of Birmingham) (University of Nottingham)

Thesis submitted to The University of East Anglia

for the degree of Master of Philosophy

June 2018

Abstract

The rhizosphere is the complex soil environment that is affected by the root systems of plants, and contains microorganisms such as pathogens, plant growth promoting rhizobacteria. Many biotic interactions occur in the rhizosphere, including those between plants, pathogens and plant growth promoting rhizobacteria. An example of a plant growth promoting rhizobacteria, *Pseudomonas fluorescens* SBW25, strongly colonises the roots of plants.

Two uncharacterised proteins, RccR and RccA, a transcriptional regulator from the RpiR family and a putative c-di-GMP protein respectively, were studied in this project. RpiR regulators are associated with carbon metabolism, while c-di-GMP is a signalling molecule. Both RccR and RccA were shown to be important for *P. fluorescens* SBW25 in colonisation of the wheat rhizosphere. For successful colonisation to happen, bacterial motility plays an important role. Upon further investigation of this phenotype, RccR and RccA were shown to be functionally linked. Biochemical analysis on RccR established that RccR negatively regulates its own activity, as well as binding to seven additional DNA sites within the *P. fluorescens* SBW25 genome. The genes that RccR regulates are involved in carbon metabolism; *glcB, pntAA* and *PFLU3817* are part of the glyoxylate shunt; *gap* and *pckA* are involved in gluconeogenesis; and *aceE* is part of the pyruvate metabolism pathway.

Using a mariner transposon screen to identify other genes within the SBW25 genome which link *rccR* and *rccA,* links to a second RpiR transcriptional regulator; HexR, and very similar protein to RccR, were discovered. Work has begun on biochemical analysis to make comparisons between these two highly similar proteins.

Acknowledgements

Firstly, and foremost, I must say thank you to all of my friends and family that have given me the support and encouragement throughout these years, especially through the more challenging times. I am extremely grateful to my family for all the support given to me throughout my life, but in particular with my decision to return to study, moving to Norwich and further from home again and seemingly too far away. Jacqui Fischer and Siobhan Hoare, we've been friends since AHS, and as we've grown old together you've always been on the end of the phone for a long chat and words of encouragement. Alice Cowie, Ellie Marshall and Jeannette Shipman life would just not be the same without you three and a time for all of us to enjoy our meet-ups and forget the stresses of life. Finally, our amazing research assistant, and good friend, Dr Richard Little, thank you not only for all your help and technical assistance, but also for always being there for me, both whilst I was in the laboratory and since.

Another thank you must be given to my supervisor Dr Jacob Malone for giving me the opportunity to work in his laboratory. A big thank you must be said to the Erasmus and short-term students Stefan Paulusch, Eleni Vikeli and Libby Humphries who have all contributed to parts of Chapter 5.

Finally, I would like to thank everyone who helped me when ill; all of your support, encouragement and belief in me have helped me fight through the hard times. Without all of you, this thesis would not be here today, thank you all very much!

Abbreviations

σ	Sigma
β	Beta
Δ	Delta
α	Alpha
λ	Lambda
$^{-1}$	Per
%	Percentage
°C	Degrees centigrade
n	Nano (10^{-9})
µ	Micro (10^{-6})
m	Milli (10^{-3})
g	Grams
L	Litre
M	Molar
w/v	Weight/volume
V	Volts
UV	Ultra violet
OD	Optimal density

CFU	Colony forming units
rpm	Rotations per minute
psi	Pounds per square inch
bp	Single nucleotide base pair
kb	1,000 base pairs
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
mRNA	Messenger ribonucleic acid
dNTP	Deoxynucleotide triphosphates
GTP	Guanosine triphosphate
GMP	Guanosine monophosphate
c-di-GMP	Cyclic di-guanosine monophosphate
DGC	Diguanylate cyclase
PDE	Phosphodiesterase
PGPR	Plant growth promoting rhizobacteria
ED	Entner Doudoroff
TCA	Tricarboxylic acid
EMP	Embden-Meyerhof-Parnas
SIS	Sugar isomerase
HTH	Helix turn helix

Experimental techniques

PCR	Polymerase chain reaction
RT PCR	Reverse transcriptase polymerase chain reaction
EMSA	Electrophoretic mobility shift assay
CoIP	Co-immunoprecipitation
ChIP seq	Chromatin immunoprecipitation sequencing
IVET	in vivo expression technology
SDS PADE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Media and Solutions

LB	Lysogeny broth
KB	King's B
MS	Murashige and Skoog
M9	Minimal 9 media
TSS	Transformation and storage solution
TBS	Tris Buffered Saline
TMT	Tween, Milk powder and Tris buffered saline
TBE	Tris Borate EDTA
PBS	Phosphate buffer saline
TE	Tris EDTA

BSA	Bovine serum albumin
Tris-HCl	Tris hydrochloride
IPTG	Isopropyl β-D-1-thiogalactopyranoside
X-GAL	5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside
KDPG	2-keto-3-deoxy-6-phosphogluconate
ONGP	ortho-nitrophenyl-β-galactoside
poly [d(I-C)]	Poly-deoxy-inosinic-deoxy-cytidylic acid

Table of Contents

Chapter 1 – Introduction

Chapter 2 – Methods

Chapter 4 – The Characterisation of RccR

Chapter 5 – The Comparison of RccR and HexR

Chapter 6 – Discussion

References

Appendix

List of Figures

List of Tables

Chapter 1 – Introduction

1.1. The rhizosphere

The term rhizosphere was originally made by Hiltner in 1904 where it was defined as the soil layer of microorganisms that are influenced by the root system (Berg and Smalla, 2009); (Lugtenberg and Kamilova, 2009). The rhizosphere is formed of many different components, such as the roots of plants, bacteria, oomycetes and fungi, and finally insects and other small invertebrates. Due to all the numerous components present in the rhizosphere it is an extremely complex environment and a recent study identified over 33,300 bacterial and archaeal operational taxonomic units in the rhizosphere microbiome for the soil of the fungal pathogen *Rhizoctonia solani* in six different soil types (Mendes et al. 2011). Within the disease suppressive soils, Mendes and colleagues mainly discovered Pseudomonadaceae, Burkholderiaceae, Xanthomonadales and Lactobacillaceae. All of the components of the plant-pathogen-commensal soil system interact with one another, and is summarised in Figure 1.1.

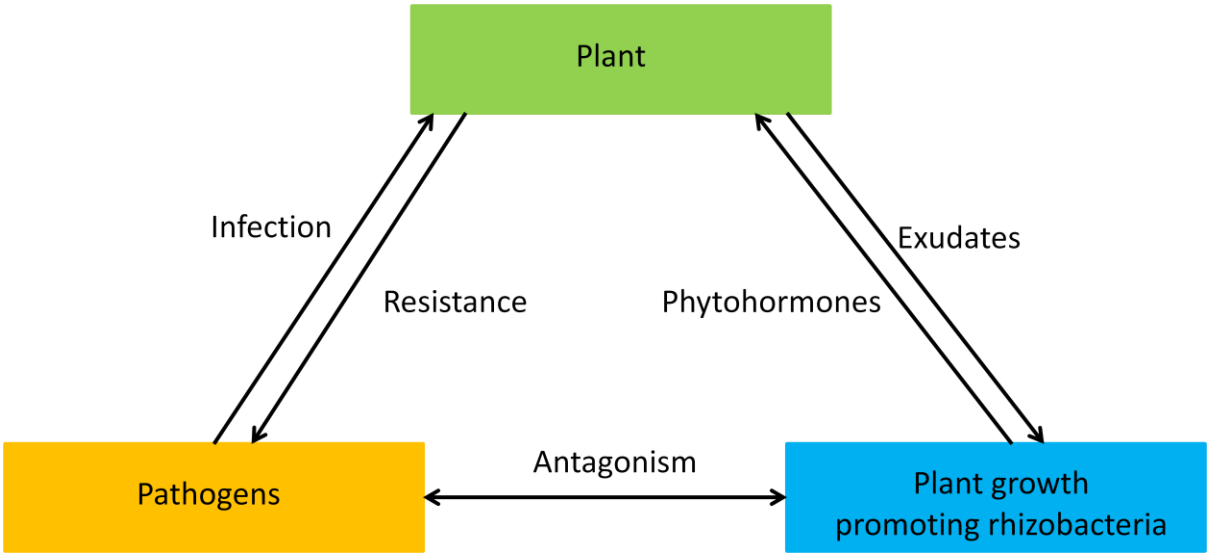

Figure 1.1. A summary of the interactions in the soil between plants, pathogens and plant growth promoting rhizobacteria. In the soil, there are many components, shown here are the main ones being plant growth promoting rhizobacteria, pathogens and plants, all of which interact. Plant growth promoting rhizobacteria have an antagonistic relationship with the pathogens in the soil and secrete phytohormones which affect the plant whilst using the

exudates from the plants as nutrients. The pathogens infect plants, whilst the plants have a resistance mechanism against pathogens.

Examples of interactions in the soil are pathogens interacting with plants via virulence factors, whilst some plant growth promoting rhizobacteria (PGPR), can prevent some damage from being caused. For more information about PGPR, see Section 1.2.2. Other bacteria in the rhizosphere may be neutral and have no effect on the plant or other bacteria present in the rhizosphere. As the soil in every environment is different, the rhizosphere of each is also different. This is due to the biotic and abiotic conditions, including the numerous components which make up the rhizosphere and how these vary in each environment, whilst the pH of the soil varies due to the composition of the root exudates that are present (Hartmann et al. 2009). Within the rhizosphere there is a substantially higher concentration of microbes present than is found in the surrounding soil (Hiltner, 1904). The nutrients which the plant exudes via the roots are utilised by the microbes in the rhizosphere as nutrients (Lugtenberg et al 2001). Some examples of the organic compounds which are secreted by the plants roots include amino acids, fatty acids, sugars and vitamins (Hartmann et al. 2009); (Lugtenberg and Kamilova, 2009). The formation of root exudates varies from species to species. For example, the oil radish exudate is rich in organic acids, while the pea exudate is rich in sugars (Jacoby, et al. 2017). Due to the wide-ranging exudates that are produced from plant to plant, the microorganisms and the numbers that are present in the rhizosphere also vary (Berg and Smalla, 2009). In a study examining the stability and succession of the rhizosphere microbiota, it was shown that plants, regardless of the species, are mainly responsible for the differences in rhizosphere microbiota. However, the soil composition also influences the stability of the rhizosphere (Tkacz et al. 2015).

1.1.2. Plant growth promoting rhizobacteria

By the late 1970s and early 1980s, some soil bacteria had been found to have plant growth promoting effects (Burr et al. 1973); (Teintze et al. 1981). Plant growth promoting rhizobacteria (PGPR) are described as bacteria which help to promote the health and development of the plant as well as, or preventing the growth of pathogens. PGPRs are able to colonise the roots of plants and some are able to colonise the inner root tissues (Vacheron et al. 2013). PGPRs modify the architecture of the roots leading to the production of

phytohormones and other signals enhancing lateral root branching and the development of root hairs (Vacheron et al. 2013). Many different bacterial communities exist in the rhizosphere and these are further complicated by the relationships that some species of PGPRs have with different plants. For example, *Pseudomonas fluorescens* interacts with wheat and flax plants (Landa et al. 2006), whilst crop management has an effect on rhizosphere populations as shown by (Mavrodi et al. 2012). Due to the complex nature of the rhizosphere, and the large number of different bacterial species present, including PGPRs and pathogens, there is a lot of competition for nutrients. For this reason, PGPRs produce various compounds in order to help their survival (Nguyen et al. 2016). Some of these compounds that the PGPRs produce are useful as biocontrol agents.

Pseudomonas species produce a number of different compounds in order to aid their survival in the soil, for example antimicrobials, bacteriocins and toxins. *Pseudomonas* species produce a number of antimicrobials, including siderophores to prevent the uptake of ions by pathogens (Mauchline and Malone, 2017) and the antifungal compounds pyoluteorin and pyrrolnitrin (Kirner et al. 1998); (Nowak-Thompson et al. 1999). Different *Pseudomonas* species produce various bacteriocins which are proteinaceous toxins that are able to destroy bacteria that are closely related to those who produced the toxin (Loper et al. 2012). Some strains of *Pseudomaonas* produce toxins in gene clusters encoding the toxins such as Mcf or the relaxed Fit toxin. The Mcf toxin has been shown to affect entomopathogenic nematodes using the *Photorhabdus* lineage (Ruffner et al. 2015). On the other hand, the Fit toxin is present in *Pseudomonas chlororaphis* and has an effect against the African cotton leafworm *Spodoptera littoralis* (Ruffner et al. 2013).

1.2. *Pseudomonas fluorescens*

Pseudomonas fluorescens are Gram negative and part of the gamma proteobacteria class of bacteria. They are rod shaped and motile with polar flagella. The typical *P. fluorescens* genome has approximately 6,000 genes, however, an analysis of numerous annotated *Pseudomonas* genomes identified significant genomic variability, with only about 20% of the genome of an individual bacterium comprising of the ‘core’ *Pseudomonas* genes found in all isolates (Garrido-Sanz et al. 2016). *P. fluorescens* can be found in a range of environments such as in fresh water, seawater and sediment. As well as these environments, *P. fluorescens*

is also found, and has adapted to survive in the soil where it is able to effectively colonise the roots of plants (O'Sullivan and O'Gara, 1992). By colonising the roots, the bacteria promote the growth of the plant as well as reducing plant disease (Haas and Defago 2005). In addition to this, *P. fluorescens* is found abundantly on the surface of leaves and contributes to the turnover of organic matter (O'Sullivan and O'Gara, 1992). In 1989, the strain SBW25 was isolated from the surface of a leaf of a sugar beet plant from University Farm, Wytham, Oxford, UK (Rainey and Bailey, 1996). This strain has since been characterised considerably, and is a model organism for both molecular microbiology (Lui et al. 2017); (Campilongo et al. 2017) and microbial evolutionary studies (Cairns et al. 2017); (Koza et al. 2017).

1.2.1. *Pseudomonas fluorescens* as a biocontrol agent

There are some strains of *P. fluorescens* that have been shown to useful as biocontrol agents (Naseby et al. 2001). In the study by Naseby et al. it was shown that the two strains of *P. fluorescens* SBW25 and CHA0 were the best at preventing the fungal pathogen *Pythium ultimum* attacking in the pea rhizosphere. Even though *P. fluorescens* is a biocontrol agent and can be utilised as one, it does not protect the plant from the soil phytopathogens in a specific manner as it is non-specific in its ability to protect plants (Couillerot et al. 2009). For this reason, *P. fluorescens* is able to control a number of plant diseases and in a range of environments, for example the strain F113 can protect sugar beet and potato from *P. ultimum* and *Pectobacterium carotovorum* respectively (Moënne-Loccoz et al. 1998), while the strain SBW25 is able to protect the pea from *P. ultimum* (Sanguin et al. 2008), and finally the strain Pf-5 protects cucumber from *P. ultimum* and bluegrass from *Drechslera poae* and *Sclerotinia homoeocarpa* respectively (Loper et al. 2007), to name a few. It is estimated that up to 10% of all rhizosphere species have biocontrol properties (Couillerot et al. 2009). When *P. fluorescens* is used as a biocontrol agent, there are still pathogens present in the soil, however these are in a lower quantity and are unable to cause infection to the plants. In these circumstances, the soil is classified as a suppressive soil which is defined as "a soil in which the pathogen does not establish or persist, establishes but causes little damage, or establishes and causes disease for a while but thereafter the disease is less important even though the pathogen may persist in the soil" (Baker and Cook, 1974); (Kwak and Weller, 2013).

1.2.2. Rhizosphere colonisation

During the process of plant colonisation, there are several stages the bacteria have to complete before being successful. In *Pseudomonas* species, the first part of plant colonisation is the movement along the gradient of the root exudates in the rhizosphere by use of chemotaxis. In order to locate the nutrients from the root and for bacterial attachment to occur, the Type IV pili, flagella and biosurfactants are all important (Alsohim et al. 2014); (Lugtenberg et al. 2001). In the latter stages of colonisation, the formation of micro-colonies on the surface of the plant is followed by a biofilm (Chin-A-Woeng et al. 1997). To assist in plant colonisation, *Pseudomonas* species produce enzymes which have effects on the plant, such as auxin producing and degrading enzymes that affect plant growth (Loper et al. 2012). Rhiszosphere colonisation requires a highly complex network of regulation, which enables the colonising bacteria to effectively respond to the different environments that make up the soil-plant interface. This control is mediated by several different pathways, including transcriptional regulation and second messenger-mediated signalling.

1.3. Cyclic di-guanosine monophosphate

One of the most important second messenger pathways controlling colonisation of the rhizosphere environment is **c**yclic **di**-**g**uanosine **m**ono**p**hosphate (c-di-GMP) signalling. C-di-GMP is a ubiquitous second messenger which is predominately found in the bacterial kingdom. C-di-GMP was first described by Benziman and colleagues in 1987 whilst studying the biosynthesis of cellulose in *Acetobacter xylinum*, (later renamed *Gluconacetobacter xylinus* (Ross et al. 1987)). In this study, Ross et al. identified that a cellulose synthase activator was present, but requires a multi component regulatory system as well as GTP to be produced in order for it to function. Later, it was discovered that this activator was c-di-GMP. Weinhouse and co-workers later identified that the BscA1 component of the cellulose activator was binding to c-di-GMP (Weinhouse et al. 1997).

Once again, using *A. xylinum*, three separate operons were identified each of which was involved in the metabolism of c-di-GMP (Tal et al. 1998). Each operon contained a pair of paralogs named *dgc* and *pde*, coding for a GGDEF and an EAL protein respectively, with the

dgc domain upstream of the *pde* domain. At this point in time, Tal et al. were unable to establish which domain was responsible for the synthesis or degradation of c-di-GMP.

1.3.1. The synthesis of c-di-GMP

For the cellular levels of c-di-GMP to be maintained, the synthesis and degradation of c-di-GMP is constantly occurring. The synthesis of c-di-GMP takes place by diguanylate cyclase (DGC) enzymes and requires two molecules of GTP, in addition to Mg^{2+} ions as a cofactor (Römling et al. 2013). Figure 1.2 shows how c-di-GMP is synthesised. When c-di-GMP is formed, the two separate GMP moieties are bound head-to-tail. The DGC activity is found in the GGDEF domain, and this was biochemically proven for the first time in 2004 for the PleD protein in *Caulobacter crescentus* (Paul et al. 2004). The GGDEF domain is named after the five amino acids which make up a conserved section of its active site; glycine, glycine, aspartic acid, glutamic acid and phenylalanine. Despite the GGDEF domain originally being named after these five amino acids, it has since been shown that the DGC activity is also present if the domain consists of either (S/G)G(D/E)EF. In 2007, using WspR from *P. fluorescens*, it was shown that the third residue could be either an aspartic acid or glutamic acid, a D or an E, in order to be functional (Malone et al. 2007), whilst the DGC activity in the SGDEF domain was shown in the protein ECA3270 from *Pectobacterium atrosepticum* (Pérez-Mendoza et al. 2011). Although enzmatically active DGCs contain the conserved amino acids, the overall number of amino acids in each protein varies, however, on average each protein containing this conserved domain has on average 180 amino acids (Galperin, 2004).

1.3.2. The degradation of c-di-GMP

C-di-GMP is degraded by specific phosphodiesterase (PDE) enzymes. Two different PDE enzymes are required for the degradation of c-di-GMP due to GMP being formed via the linear intermediate 5′-phosphoguanylyl-(3′,5′)-guanosine (pGpG). The first step of breaking c-di-GMP down into pGpG occurs by PDE-A enzymes. The PDE-A activity was first biochemically identified in the EAL domain (Christen et al. 2005); (Schmidt et al. 2005).

This intermediate is degraded by the second PDE enzyme, PDE-B, in a hydrolysis reaction in order to produce the end products of two GMP molecules (Schmidt et al. 2005). In 2015, two studies in *Pseudomonas aeruginosa* identified that the oligoribonuclease, Orn, is responsible for degrading pGpG into GMP and therefore the likely candidate for the PDE-B enzyme in *P. aeruginosa* (Cohen et al. 2015); (Orr et al. 2015). Although Orn was described in *P. aeruginosa*, there are homologues present throughout the bacterial kingdom (Cohen et al. 2015). The different stages of degradation of c-di-GMP are shown in Figure 1.2. In order for an EAL domain to be active, Mg^{2+} or Mn^{2+} ions are necessary; however, if Ca^{2+} or Zn^{2+} ions are present the EAL domain will be inactive (Tchigvintsev et al. 2010); (Schmidt et al. 2005). Just like the GGDEF domain, the EAL domain is named after three conserved amino acids found in the active site; glutamic acid, alanine and leucine. Once again, the number of amino acids in each protein with an EAL domain present differs from protein to protein, but the average protein has approximately 250 amino acids (Galperin, 2004).

Due to some bacterial species not encoding an EAL domain but having a GGDEF domain present, some questions were raised as to how c-di-GMP was being degraded in these species. For this reason, another domain was believed to have either PDE-A and/or PDE-B activity. Bioinformatic analysis identified a new protein domain, the HD GYP domain, as potentially being involved in the degradation of c-di-GMP (Galperin et al. 1999). Biochemical evidence for HD GYP phosphodiesterase activity was not published until 2013, when a HD GYP from *Vibrio cholerae* was shown to contains a non-heme diiron-carboxylate active site, of which only the reduced form is active (Miner et al. 2013). Even though it was 2013 until the biochemical evidence for HD GYP phosphodiesterase activity, there were publications suggesting that the HD GYP domain had phosphodiesterase activity, such as in *V. cholera* controlling biofilm formations (Hammer and Bassler, 2009) and in *P. aeruginosa* regulating biofilm formation and virulence (Ryan et al. 2009). In 2014, it was established that HD GYP phosphodiesterase activity degrades c-di-GMP directly into two GMP molecules using a novel trinuclear catalytic iron centre (Bellini et al. 2014).

1.3.3. C-di-GMP receptors

There are numerous c-di-GMP receptors in bacterial cells, however due to the lack of a single consensus sequence for these receptors, it is hard to identify them bioinformatically. More

information on the receptors of c-di-GMP can be found in the many review articles that have been published, such (Jenal et al. 2017); (Ryan et al. 2012) and (Hengge, 2009). A brief description of the main classes of receptors is given below.

In 2006, it was hypothesised using bioinformatics that the PilZ domain was required for c-di-GMP binding (Amikam and Galperin, 2006) and this was later confirmed experimentally using the PilZ domain containing protein YcgR from *Escherichia coli* (Ryjenkov et al. 2006). A new regulator of c-di-GMP was discovered in 2008: the GEMM domain (Sudarsan et al. 2008). The GEMM domain is a highly conserved mRNA domain known as a riboswitch. Riboswitches are regulatory fragments of mRNA, commonly found in the untranslated region of RNA, which bind to a small molecule therefore affecting the expression of genes from the mRNA itself and responding to the changes in the concentration of the target ligand (Winkler and Breaker, 2005). The GEMM domain was identified in the open reading frames immediately upstream of proteins with PDE or DGC activity.

Some GGDEF/EAL domains are not able synthesise or degrade c-di-GMP respectively, however these degenerate GGDEF/EAL domains may be able to bind to c-di-GMP. The first degenerate GGDEF protein that was shown to be able to bind to c-di-GMP was PopA from *C. crescentus* (Duerig et al. 2009). Also in 2009, LapD from *P. fluorescens* was shown to be a degenerate EAL protein which binds to c-di-GMP (Newell et al. 2009).

c-di-GMP also binds to AAA+ ATPases, the first of which to be described was FleQ in *P. aeruginosa* (Hickman and Harwood, 2008). These transcription factors are usually activated by phosphorylation, however in the case of FleQ, c-di-GMP interacts with the AAA+ ATPase domain. When bound to c-di-GMP, the structure of FleQ changes, altering the activity levels in ATPase which subsequently has downstream effects.

In addition, there are many other proteins which can bind to c-di-GMP. For example, in *Xanthomonas campestris*, the global regulator Clp which has a strong homology to Crp from *E. coli*, was hypothesised to bind to c-di-GMP (He et al. 2007). It was not until 2010 that experimental evidence confirmed that Clp binds to c-di-GMP inducing a conformation change stopping the interaction between Clp and the target DNA it was previously binding to (Tao et al. 2010). A second protein that has been shown to bind to c-di-GMP is the histidine kinase cell cycle kinase A, CckA, from *C. crescentus* (Lori et al. 2015). CckA binds to c-di-

GMP using its catalytic ATPase domains. In the presence of c-di-GMP, the phosphatase activity of CckA is stimulated whilst the default kinase activity is inhibited.

A summary of c-di-GMP synthesis, degradation and the receptors is shown in Figure 1.2.

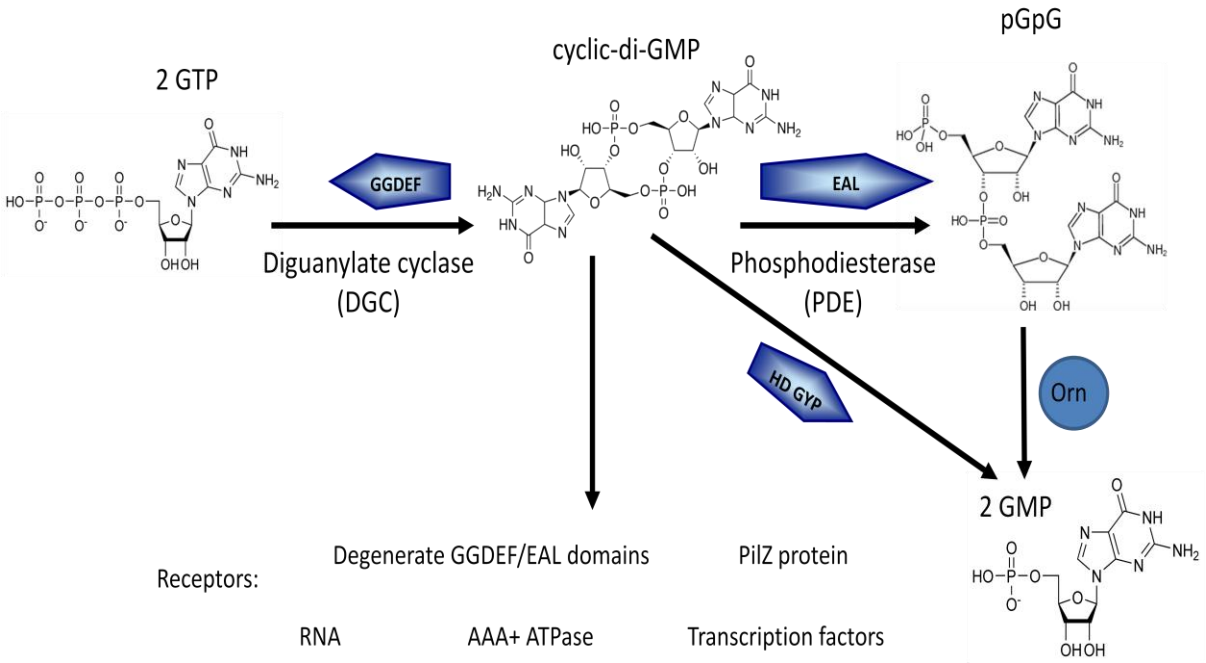

Figure 1.2. A summary of how c-di-GMP is synthesised, degraded and the receptors in the cell. C-di-GMP is synthesised from two molecules of GTP by diguanylate cyclase (DGC) enzymes found in the GGDEF domain, and degraded by specific phosphodiesterase (PDE) enzymes which are found in either EAL or HD GYP domains into the intermediate pGpG and this is further degraded into two molecules of GMP by the oligoribonuclease Orn. C-di-GMP binds to a number of receptors, including PilZ domains, transcription factors including FleQ, RNA in the form of riboswitches, and degenerate GGDEF domains.

1.3.4. C-di-GMP regulation

The majority of c-di-GMP proteins occur in multi domain signalling proteins suggesting that c-di-GMP is involved in sensing environmental signals. It is possible for multiple sensory input domains to be present, indicating that c-di-GMP may be involved in a complex response to environmental signals. In each case, the sensory domain is located at the N terminal of the GGDEF or EAL domain. A number of sensory domains have been associated

with c-di-GMP proteins, including PAS, GAF, HAMP and REC domains (Mills et al. 2011). PAS domains have been shown to bind to small molecules such as citrate, and also been shown to sense molecular oxygen. The GAF domain binds to cyclic mononucleotides and other small molecular weight effectors. The REC domain is usually part of a two component signal transduction system and activated by phosphorylation.

It has long been speculated that c-di-GMP is localised in the bacterial cell due to specific interactions between proteins, for example SadC and BifA in *P. aeruginosa* (Ribbe et al. 2017), and PleD and PopA in *C. crescentus* (Ozaki et al. 2014). However, in 2017 this was turned around following a study in *E. coli* K-12 and using all 29 DGC and PDE containing proteins. This study identified that c-di-GMP is organised into a hierarchical system, allowing multiple protein interactions to occur, containing a few master controllers and hubs (Sarenko et al. 2017).

1.3.5. The cellular effects of c-di-GMP

Since its discovery, c-di-GMP has been shown to have many different effects on bacterial cells, including the control of the production of virulence factors in animals and plants (Bhagirath et al. 2017); (Little et al. 2016) and cell cycle progression (Lori et al. 2015). C-di-GMP has also been implicated in the cellular functions of biofilm formation and the switch between motile and sessile cells; both of which are important for rhizosphere colonisation. When the concentration of c-di-GMP in the cell is low, bacteria are motile, whilst if the levels of c-di-GMP are high the bacteria are sessile and form biofilms. The link between c-di-GMP concentration and the motility of cells has been well documented in a range of bacteria, such as *P. aeruginosa*, *E. coli* and *Salmonella enterica* serovar Typhimurium (Simm et al. 2004).

As described in Section 1.3.1, motility and forming biofilms are important processes for plant colonising bacteria. C-di-GMP has been shown to be involved in these processes in many bacteria which colonise plants including the *flgZ* gene which is part of an operon encoding a PilZ domain protein which regulates swimming and biofilm formation in the *Pseudomonas* species (Martínez-Granero et al. 2014) and the Gac, Sad and Wsp pathways in *P. fluorescens* F113 which all independently contribute to swimming motility (Navazo et al. 2009). However, the system that has been studied the most is the Lap system from *P. fluorescens*

which controls biofilm formation. The LapD signalling protein contains both a GGDEF and EAL domain but lacks activity in these domains however c-di-GMP is able to bind to the EAL domain of LapD (Newell et al. 2009). In the presence of c-di-GMP, LapD interacts with the periplasmic protease LapG through its PAS domain (Boyd et al. 2012). When LapD interacts with LapG, it is unable to cleave the cell surface adhesin LapA therefore allowing a biofilm to form. On the other hand, in the absence of c-di-GMP, LapD undergoes a conformation change and loses its affinity for LapG which is then able to cleave the N terminal of LapA and release the adhesin making the structure of the biofilm destable (Chatterjee et al. 2014).

c-di-GMP is known to be involved in many systems, however further work is required to fully establish what role it has in the system; one such example is in *P. fluorescens* and the Rim system. The RimK protein is required to catalyse the addition of glutamate residues to the ribosomal modification enzyme RpsF, while proteomic changes link it to Hfq (Little et al. 2016). RimK is also controlled by c-di-GMP, which is interesting as the RimA protein is a PDE and may be regulating RimK by controlling the levels of c-di-GMP. The final protein, RimB, interacts with RimK, however the role this has in the system is still to be established. A model of how RimA and RimB interact with RimK has been proposed by Little et al. which is shown in Figure 1.3.

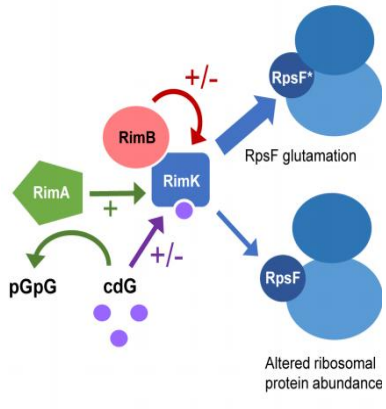

Figure 1.3. A model for how Rim system in *Pseudomonas fluorescens*. An increase in RimK activity also leads to an increase in RpsF glutamation, shown by the thickness of the arrow. RimK is further regulated by the two proteins RimA and RimB as well as the signalling molecule cyclic-di-GMP. cdG represents cyclic-di-GMP and +/- identifies that the nature of the protein/signalling molecule is currently undefined. Taken from (Little et al. 2016).

A second example where c-di-GMP is known to be important but investigations are still required, is in the plant pathogen *Erwinia amylovora* where c-di-GMP has been found to regulate the progression of disease (Edmunds et al. 2013). *E. amylovora* causes fire blight in plants of the family *Rosaceae*, which contains apples and pears. In the study by Edmunds and colleagues, it was established that three out of the five DGCs within the *E. amylovora* genome, EdcA, EdcC, and EdcE, are active. It was discovered that EdcC and EdcE are the two DGCs that regulate the processes involved in disease progression via motility, amylovoran production, biofilm production and the plant infection models that were used in the study.

1.4. Transcriptional regulators

In order to respond to the constantly changing external conditions and their internal metabolic environment, bacteria need to finely regulate the expression of their genes. There are various families of transcriptional regulators, each of which work in different ways. Each class of transcriptional regulator is classified by the sequence similarity of the binding motif or alternatively by the alignments of the amino acid sequences.

In general, transcriptional regulators are composed of two domains; the first senses the external or internal signals, whilst the second domain interacts with the DNA targets. In bacteria, the most common DNA binding domain is the helix turn helix domain (Ishihama, 2012). In order for each transcriptional factor to work differently, they bind to their ligand with high specificity which leads to the activation state (Brautaset et al. 2009).

Transcriptional regulators can act as either repressors or activators and in order to function they are usually controlled by a co-factor. Co-factors vary in size and can be any one of a number of molecules including small ions, sugars and peptides (van Hijum et al. 2009). Additional intangible signals, such as light (BLUF domain) have also been described as being involved in bacterial signal transduction (Gomelsky and Klug, 2002). Many transcriptional factors function as part of a kinase receiver pair and require protein-protein interactions with the signal input protein (Haldimann et al. 1996).

There are four ways in which repressors have been described to work. The first is by steric hindrance, where the repressor binds on or between the core promoter region preventing the RNA polymerase from binding, (Kohno et al. 1994). The second is repression by blocking of transcription elongation, which prevents the binding at the start codon and transcription from initiating (Schröder and Wagner, 2000). The third method is repression by DNA looping, whereby there are DNA binding site upstream and downstream of the promoter (Meinhardt et al. 2012). Finally, the last method is by the repression by the modulation of an activator, where the binding sites of the repressor and activator overlap, therefore only one of these transcriptional regulators can bind at any one time. (van Hijum et al. 2009).

Just like repressors, there are four different ways activators are known to work. Firstly, the transcriptional regulator binds upstream of the promoter and interacts with one of the core components of RNA polymerase. In the second method, the transcriptional regulator binds next to the promoter promoting the binding of a σ factor and transcription. The third method is by a DNA conformation change when the transcriptional factor binds to the promoter, which allows the σ factor to bind and transcription. The final method is activation by the modulation of a repressor (van Hijum et al. 2009).

1.4.1. The RpiR family of transcriptional regulators

The RpiR family of transcriptional regulators are characterised by their N terminal helix turn helix DNA binding domains and their phosphosugar binding C terminal sugar isomerase (SIS) domains (Bateman 1999). The organisation of the domains of RpiR transcriptional regulators can be seen in Figure 1.4.

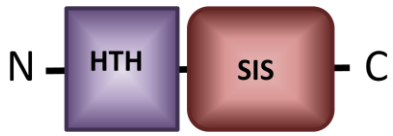

Figure 1.4. The domain organisation of RpiR transcriptional regulators. RpiR transcriptional regulators are composed of two domains; a N terminal helix turn helix DNA binding domain, represented by HTH, and a C terminal sugar binding domain annotated as SIS.

Despite RpiR first being identified in 1996, (Sørensen and Hove-Jensen, 1996), few RpiR transcriptional regulators have since been studied and surprisingly little is currently known about how this family of regulators function. To date, there has been no catalytic activity recorded for the SIS when linked to the helix turn helix domain from the same polypeptide (Daddaoua et al. 2009). RpiR transcriptional regulators have been associated with control of carbon metabolism. For example, RpiR in *Escherichia coli* is involved in ribose metabolism (Sørensen and Hove-Jensen, 1996), while MurR regulates the catabolism of N-Acetylmuramic acid, an amino sugar, in *E. coli* (Jaeger and Mayer, 2008). RpiR transcriptional regulators have been shown to be activators as well as repressors of gene transcription. For example, ClaR from *Lactococcus lactis* is an activator (Aleksandrzak-Piekarczyk et al. 2015), whilst in *Sinorhizobium meliloti*, IolR functions as a repressor (Kohler et al. 2011).

1.4.2. The HexR transcriptional regulator

Prior to the start of this project, an RpiR transcriptional regulator had yet to be studied in *P. fluorescens*. However, one member of the RpiR family of transcriptional regulators, HexR, has been investigated in the closely related bacterial species *Pseudomonas putida* (Daddaoua et al. 2009) and *Pseudomonas syringae* (Mehmood et al. 2015), as well as in a second gamma proteobacteria species *Shewanella oneidensis* (Leyn et al. 2011). From these studies, it was identified that HexR binds to 2-keto-3-deoxy-6-phosphogluconate (KDPG), an intermediate of the Entner Doudoroff pathway. Two monomers of HexR were discovered to bind to the consensus sequence of 5'–TTGT$_{n=7–8}$ACAA–3' (Daddaoua et al. 2009). In the study by Daddaoua and co-workers, a three-dimensional homology model for both domains of HexR was produced as shown in Figure 1.5. This model identified that the two key amino acids in the DNA binding helix turn helix domain of HexR in *P. putida* which may interact with the DNA are arginine 54 and arginine 57. In addition to this, the model suggests that in the SIS domain the amino acids serine 140 and serine 184 are involved in recognising the effector.

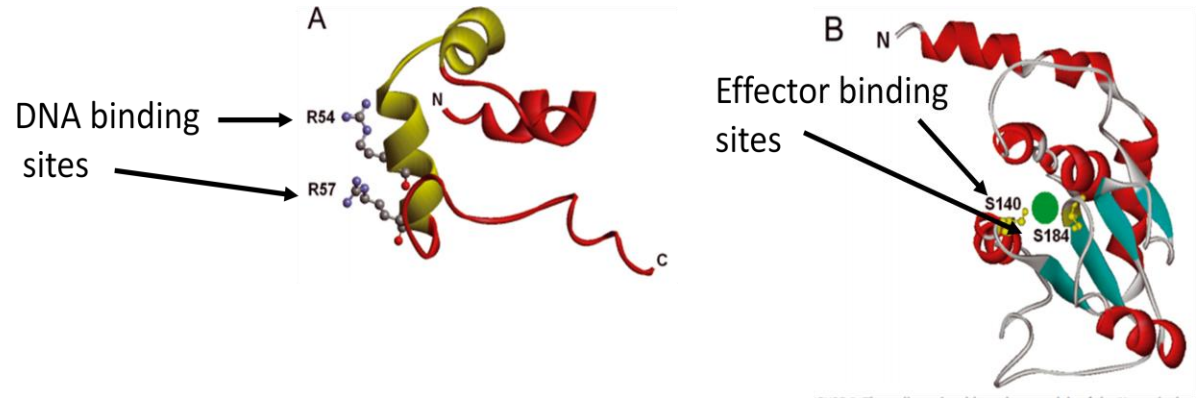

Figure 1.5. The three dimensional homology model of the N and C terminal domains of HexR. A) The N terminal DNA binding helix turn helix domain indicating the two arginine amino acids at 54 and 57 which are thought to interact with the DNA. B) The C terminal sugar isomerase domain with the amino acid serine highlighted at positions 140 and 184 which are believed to interact with the effector molecule shown in green. Taken from (Daddaoua et al. 2009).

1.5. The regulation of central carbon metabolism

Central carbon metabolism describes the key network of pathways and oxidation reactions of the main carbon sources in the cell (Papagianni, 2012). Central carbon metabolism pathways include the phosphotransferase system (PTS), glycolysis, gluconeogenesis, the pentose phosphate (PP) pathway, and the tricarboxylic acid (TCA) cycle and the glyoxylate shunt. Important pathways for this study are described in further detail below. In each pathway, there are key points which have been studied giving a better understanding to the respective pathway. In order for cells to efficiently regulate carbon metabolism, these metabolic pathways are interconnected.

1.5.1. Glycolysis

Glycolysis is the pathway responsible for the breakdown of glucose to pyruvate. However, in bacteria, due to the range of species, there is not one single pathway that is able to do this. The main glycolysis pathway is the Embden-Meyerhof-Parnas (EMP), however, as not all

bacteria have the key enzyme, 6-phosphofructokinase, required for the EMP pathway, an alternative is required for these. The alternative, the Entner Doudoroff (ED) pathway, was first described by Entner and Doudoroff in 1952 (Entner and Doudoroff, 1952). In this study, Entner and Doudoroff describe an alternative for the metabolism for glucose to the EMP pathway using the bacterium *Pseudomonas saccharophila*. The ED pathway is utilised by the *Pseudomonas* genus to metabolise glucose into the end products of glyceraldehyde-3-phosphate and pyruvate (del Castillo et al. 2008). Some bacteria have both the EMP and ED pathways, and if this is the case, the products from the ED pathway feed into the EMP pathway (Jojima and Inui, 2015). Both the EMP and ED pathways degrade glucose into the key components of phosphoenolpyruvate (PEP), pyruvate and acetyl-CoA which subsequently can be broken down in further reactions, for example, PEP is used in gluconeogenesis.

1.5.2. Gluconeogenesis

Gluconeogenesis is the reverse reaction of glycolysis which allows glucose synthesis to take from pyruvate via intermediate reactions. TCA cycle intermediates or carbon sources which enter the TCA cycle allow growth to occur via gluconeogenesis (Geddes and Oresnik, 2014). During gluconeogenesis essential polysaccharides, nucleotides, and amino acids are synthesised (diCenzo et al. 2017). Gluconeogenesis is ubiquitous among different species and found in plants, animals, bacteria and fungi. Three of the four enzymes in gluconeogenesis are reversible to those in glycolysis. The final enzyme, fructose-1,6-bisphosphatase is required to convert fructose-6-phosphate to glucose-6-phosphate (Sauer and Eikmanns, 2005).

1.5.3. The tricarboxylic acid cycle and the glyoxylate shunt

The tricarboxylic acid (TCA) cycle is also known as the Krebs cycle or the citric acid cycle. The TCA cycle is responsible for the complete oxidation of acetyl-CoA, which was produced in the earlier glycolysis reaction. In addition to acetyl-CoA, the TCA cycle can degrade fatty acids, ethanol or poly-β-hydroxybutyrate (Dunn et al. 2009). A complete cycle of the TCA

cycle produces two carbon dioxide molecules as well as other macromolecules the cell uses for diverse cellular processes.

The glyoxylate shunt is a unique anaplerotic variant of the TCA cycle and shares three of the five key enzymes with the TCA cycle (Kondrashov et al. 2006). Figure 1.6 shows where the glyoxylate shunt diverts from the TCA cycle.

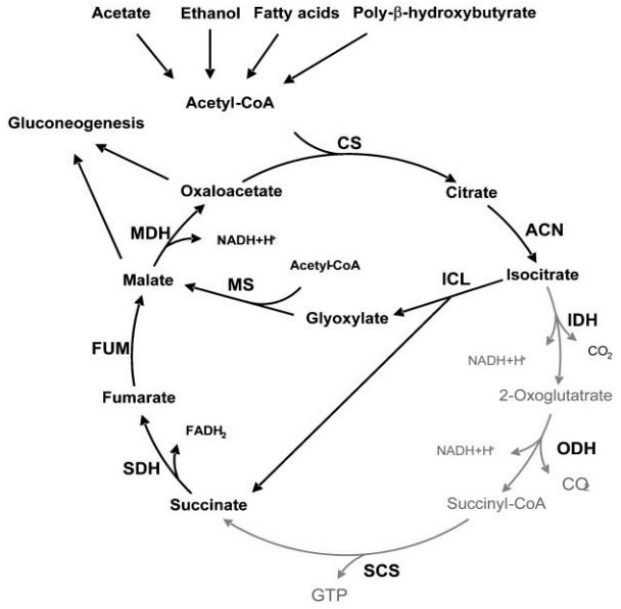

Figure 1.6. The enzymatic reactions of the TCA cycle and glyoxylate shunt. The TCA cycle is shown as a full circle, whilst the glyoxlate shunt is indicated by the dark arrows. The abbreviations for the enzymes: CS, citrate synthase; ACN, aconitase; IDH, isocitrate dehydrogenase; ODH, 2-oxoglutarate dehydrogenase; SCS, succinyl-CoA synthetase; SDH, succcinate deyhdrogenase; FUM, fumarase; MDH, malate dehydrogenase; MS, malate synthase; ICL, isocitrate lyase. Taken from (Dunn et al. 2009).

Once isocitrate is formed in the TCA cycle, this can either continue in the TCA cycle, or be diverted into the glyoxylate shunt where using the enzyme isocitrate lyase (ICL) it is converted into glyoxylate and succinate before the glyoxylate is broken down into malate by malate dehydrogenase. The glyoxylate shunt is found in many branches of life, including bacteria, fungi and plants (Kondrashov et al. 2006). If bacteria are grown on acetate, the glyoxylate shunt will have to be utilised to fully break down the acetate, as this is the only pathway that is able to do this (Kornberg, 1966). The end products of the TCA cycle and the glyoxylate shunt can be used for gluconeogenesis and other biosynthetic processes (Dunn et al. 2009).

1.6. The RccR transcriptional regulon

In 2017, RccR was described for the first time as a carbon regulator in *P. fluorescens* (Campilongo et al. 2017). RccR was identified to inversely control three carbon metabolism pathways; gluconeogenesis, the glyoxylate shunt and pyruvate metabolism. When pyruvate metabolism genes are regulated by RccR, the genes in gluconeogenesis and the glyoxylate shunt pathways are suppressed and vice versa. Depending on the carbon source that is available, it was shown by the authors that RccR regulates these genes accordingly and therefore switches to the appropriate pathway. The correct regulation of carbon metabolism was shown to be critical for effective rhizosphere colonisation by *P. fluorescens*. Figure 1.7 shows a model for RccR regulation and carbon metabolism which is the part of this study.

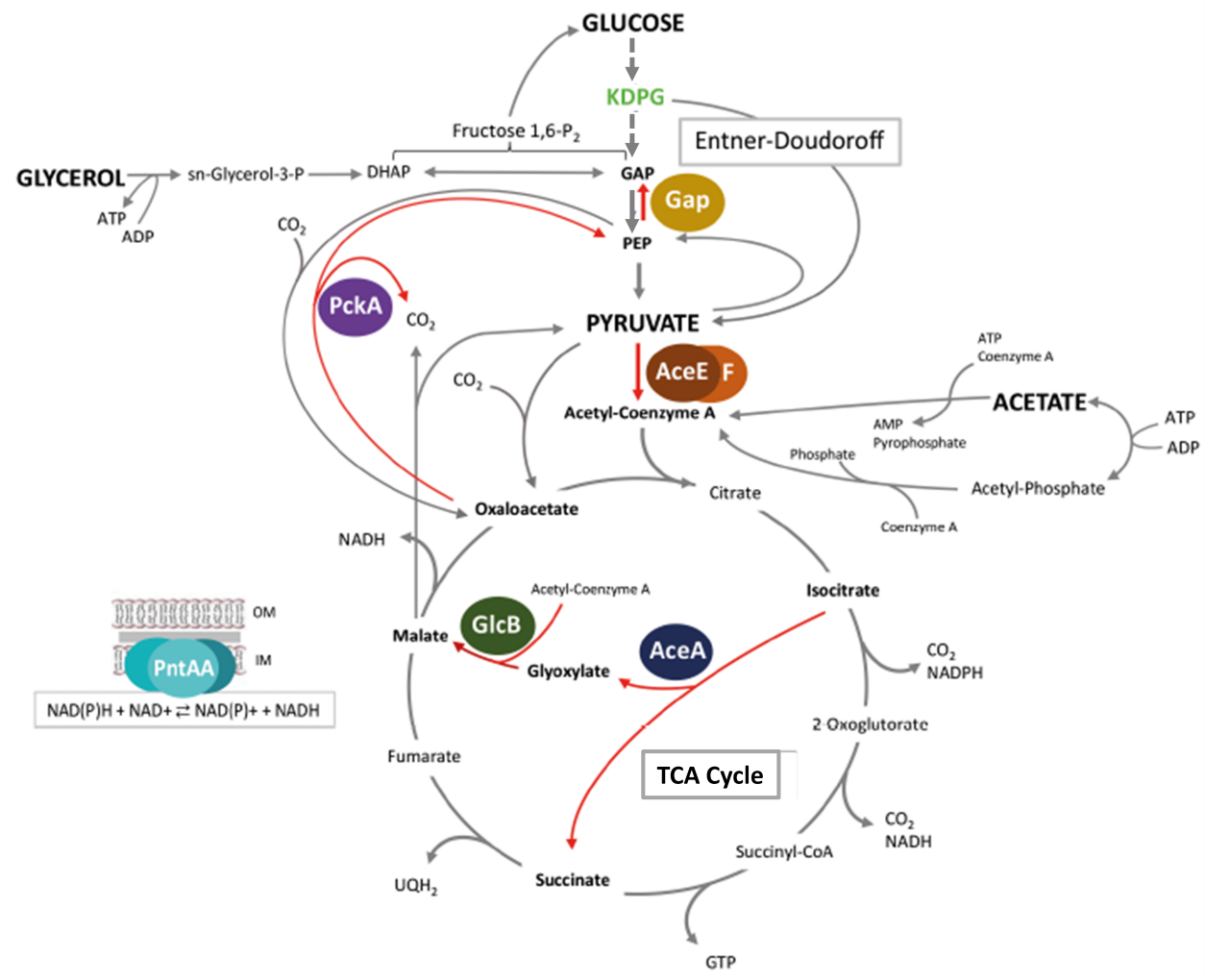

Figure 1.7. A model for RccR regulation in carbon metabolism. It is proposed that RccR regulates carbon metabolism from glucose, glycerol, pyruvate and acetate into the TCA cycle and the glyoxylate shunt. The seven protein targets of RccR are shown: PntAA/PFLU0112/B are subunits of the NAD(P) transhydrogenase membrane protein complex; PckA: phosphoenolpyruvate carboxykinase; AceE/F: pyruvate dehydrogenase subunits; Gap: glyceraldehyde-3-phosphate dehydrogenase; AceA: isocitrate lyase; GlcB: malate synthase G. The RccR regulated pathways are shown in red. Edited from (Campilongo et al. 2017).

1.7. Aim of the study

It is aimed to see if RccA, RccR and HexR from *P. fluorescens* have the ability to effect the colonisation of plant roots and allowing more effective growth. In addition to this, it is also hoped to discover the functions of these proteins. As HexR is involved in carbon metabolism in *P. putida* and *P. syringae*, this is also expected to be the case for HexR in *P. fluorescens*

and the newly identified protein RccR. This is the beginning of the first investigation which links the ability of *P. fluorescens* to effectively colonise plant roots with tight control of central carbon metabolism.

Chapter 2 – Methods

2.1. Strains and plasmids

The bacterial strains and plasmids used in this study are shown in Table 2.1 and Table 2.2 respectively.

Table 2.1. The bacterial strains used in this study.

Bacterial strain	Description	Reference
Pseudomonas fluorescens		
SBW25	Environmental *P. fluorescens* isolate	(Rainey and Bailey, 1996)
SBW25 Δ*rccR*	SBW25 with *rccR* (*PFLU6073*) deleted	This study
SBW25 Δ*rccA*	SBW25 with *rccA* (*PFLU6074*) deleted	This study
SBW25 Δ*rccA*Δ*rccR*	SBW25 with *rccA* and *rccR* (*PFLU6073 and PFLU6074*) deleted	This study
SBW25 Δ*hexR*	SBW25 with *hexR* (*PFLU4840*) deleted	This study
SBW25 Δ*0179*	SBW25 with *PFLU0179* deleted	This study
SBW25 Δ*rccA*Δ*0179*	SBW25 with *rccA* (*PFLU6074*) and *PFLU0179* deleted	This study
SBW25 p*hexRgabD*	SBW25 with the *hexR* binding site of *gabD* (*PFLU0180*) mutated from TTGT TCCGCCA ACAA to CTGT TCCGCCA GCAG	This study
SBW25 Δ*rccA* p*hexRgabD*	SBW25 with *rccA* (*PFLU6074*) deleted and the *hexR* binding site of *gabD* (*PFLU0180*) mutated from TTGT TCCGCCA ACAA to CTGT TCCGCCA GCAG	This study
SBW25-*lacZ*	SBW25 with the *lacZ* gene neutrally inserted between *PFLU1179* and *PFLU1180*	(Zhang and Rainey, 2007)

Bacterial strain	Description	Reference
Escherichia coli		
DH5α	*endA*1, *hsdR*17(r_K-m_K+), *supE*44, *recA*1, *gyrA* (Nal^r), *relA*1, Δ(*lacIZYA-argF*)U169, *deoR*, Φ80*dlacΔ(lacZ)M15*	(Woodcock et al. 1989)
BL21(D3) pLysS	SmR , K12 recF143 *lacIq lacZ*Δ.M15, *xylA*	Novagen
BW25113 pIJ790	BW25113 containing the helper plasmid pIJ790	(Gust et al. 2003)

Table 2.2. The plasmids used in this study.

Plasmid name	Description	Reference
pME3087	Tet^R, suicide vector; ColE1-replicon, IncP-1, Mob	(Voisard et al. 1994)
pTS1	pME3087 derivative containing a *sacB* counter-selection marker	(Scott et al. 2017)
pET42b(+)	Kc^R, purification vector, His_6-tag	Novagen
pET42b-*rccR*	pET42b (+) derivative with *rccR* as a *Hind*III - *Xba*I fragment	This study
pET42b-*hexR*	pET42b (+) derivative with *hexR* as a *NdeI* - *Xho*I fragment	This study
pME6032	Tet^R, P_K, 9.8 kb pVS1 derived shuttle vector	(Heeb et al. 2000)
pME6032-*rccR*	pME6032 derivative with *rccR* as a *Eco*RI - *Kpn*I fragment	This study
pME6032-*hexR*	pME6032 derivative with *hexR* as a *Eco*RI - *Sac*I fragment	This study
pGm6032	pME6032 derivative with *aacC1 as a Bam*HI -*Xba*I fragment	(Malone et al. 2010)
pGm6032-*rccR*	pGm derivative with *rccR* as a *Bam*HI - *Bam*HI fragment	This study
pIJ790	λ-RED (*gam*, *bet*, *exo*), *cat*, *araC*, *rep101ts* (CmlR)	(Gust et al. 2003)
pSUB11	Amplification vector for *flag*-FRT-Kan^R-FRT cassette	(Yu et al. 2000)
pFLP2	Amp^R, FRT cassette excision vector	(Hoang et al. 1998)
pALMAR3	Insertion vector for TetR Mariner transposon	(Malone et al. 2010)

2.2. The growth and selection of bacteria

The media, the buffers and solutions utilised throughout this study, and how to make them, are shown in Table 2.3. For all the media, solutions and buffers listed, 1L recipes are provided and are to be stored at room temperature, unless otherwise stated.

Table 2.3. The media, solutions and buffers used in this study and how to make them.

Media name	Ingredients (per litre)	
LB	10g Tryptone 10g NaCl Adjust to pH 7.0	5g Yeast extract (11g Agar)
KB	20g Peptone 1.5g K_2HPO_4 Adjust to pH 7.2	10ml Glycerol 1.5g $MgSO_4$ (15g Agar)
MS	20.61mM NH_4NO_3 3M $CaCl_2$ (anhydrous) 1.25mM KH_2PO_4 0.1mM $FeSO_4.7H_2O$ 0.08mM $MnSO_4.H_2O$ 5M KI 0.1M $CuSO_4$ (anhydrous) 0.09mM Sucrose 1.19M Thiamine.HCl Adjust to pH 5.7	18.79mM KNO_3 1.5M $MgSO_4$ (anhydrous) 0.1 mM Na_2EDTA 0.1mM H_3BO_3 0.03mM $ZnSO_4.H_2O$ 1.03M $Na_2MoO_4.2H_2O$ 0.11M $CoCl_2$ (anhydrous) 0.56mM i-Inositol 8g Agar
M9 minimal media (100ml)	10ml x10 M9 salts 100µl 100mM$CaCl_2$	200µl 1M $MgSo_4$
X2 TSS (100ml) Store at 4°C	0.8g Bacto-Tryptone 0.5g NaCl 10ml 1M $MgSO_4$ Adjust the pH to 6.5	0.5g Yeast extract 20g PEG8000 10ml DMSO Filter sterilise through a 0.2µm filter
EMSA buffer	100mM Tris HCL (pH7.5) 50mM DTT 10mM $MgCl_2$	10mM EDTA 50% Glycerol 100mM NaCl
Lysis buffer (for Chromatin Immunoprecipitation assay) Store at 4°C	10 mM Tris-HCl (pH8) 50 mM NaCl	4 $mgml^{-1}$ Lysozyme 1x Protease inhibitor
IP buffer (for Chromatin Immunoprecipitation assay) Store at 4°C	100 mM Tris-HCl (pH 8.0) 250 mM NaCl 1x Protease inhibitor	0.5% Triton X-100 0.1% SDS
Elution buffer (for Chromatin Immunoprecipitation assay) Store at 4°C	50 mM Tris-Cl (pH 7.6) 10 mM EDTA	1% SDS
Z buffer (for β galactosidase assay) Store at 4°C	0.06M Na_2HPO_4 0.01M KCl	0.09M $NaH_2PO.2H_2O$ 0.001M $MgSO_4.7H_2O$

Media name	Ingredients (per litre)			
Lysis buffer (for β galactosidase assay) Store at 4°C	100ml 50µl	Z buffer 10% SDS	0.27ml	β-mercaptoethanol
Nutrient solution Store at 4°C	1mM 800µM 35µM 0.8µM 0.3µM 18.4mM	$CaCl_2.2H_2O$ $MgSO_4$ H_3BO_3 $ZnCl_2$ $CuSO_4.5H_2O$ KH_2PO_4	100µM 10µM 9µM 0.5µM 6mM 20mM	KCl FeEDTA $MnCl_2.4H_2O$ $Na_2MoO_4.2H_2O$ KNO_3 Na_2HPO_4
'RNA Later' (1.5L)	935ml Stir until dissolved 25ml Adjust the pH to 5.2	Sterile water 1M Sodium citrate	700g 40ml	Ammonium sulphate 0.5M EDTA

All *P. fluorescens* strains were grown at 28°C. *E. coli* strains were grown at 37°C, unless otherwise stated. Colonies of both *P. fluorescens* and *E. coli* were grown on LB plates. Cultures of *P. fluorescens* were grown in LB media overnight before some of this was used to inoculate minimal media supplemented with a carbon source of interest, whilst cultures of *E. coli* were grown overnight in LB media.

The carbon sources, pyruvate, glucose and glycerol were used at a concentration of 0.4% w/v for growth experiments in minimal media, unless otherwise stated. 5% w/v sucrose was used to enable counter selection with pTS1 vectors.

To select for antibiotic resistance, antibiotics (Sigma) were used at the following final concentrations unless otherwise stated: tetracycline (tet) 12.5µgml^{-1}; kanamycin (kan) 50µgml^{-1}; gentamycin (gent) 25µgml^{-1}; carbenacillin (carb) 100µgml^{-1}; piperacillin (pip) 100µgml^{-1}; phosphomycin (phos) 100µgml^{-1}; chloramphenicol (chlor) 30µgml^{-1}.

IPTG and X-GAL at final concentrations of 200µgml^{-1} and 40µgml^{-1} respectively were used for β galactosidase blue/white selection, unless otherwise stated.

2.3. Molecular biology techniques

All DNA manipulation techniques were done according to the standard protocols with adjustments made when stated. Any kit that was utilised was done so according to the manufacturer's instructions.

2.3.1. Agarose gel electrophoresis

Gel electrophoresis was used to analyse DNA samples and separate the different sized DNA fragments. Depending on the size of DNA fragments, the agarose concentration was between 0.8% – 1.2% with ethidium bromide added. x6 gel loading dye (New England Biolabs) was added to the DNA sample followed by the loading of the agarose gel. A DNA marker of known molecular weight was also run alongside the samples, therefore allowing the identification of the sizes of the DNA samples. Gels were run at a constant voltage of 100V for one hour, or until the loading dye had nearly reached the end of the agarose gel, in x1 TBE, and were visualised using UV light.

2.3.2. Isolation of DNA, restriction digests and ligation reactions

Plasmid DNA was isolated from an overnight culture of bacterial cells using a NucleoSpin Plasmid Kit (Macherey-Nagel) and the plasmid DNA was eluted into TE buffer (5mM Tris pH8.0). DNA fragments that were required for downstream techniques were excised from agarose gels using a sterile scalpel whilst being visualised by UV light. The DNA was then extracted using the NucleoSpin Gel and PCR Clean-Up Kit (Macherey-Nagel) and eluted into TE buffer (5mM Tris pH8.0). To assist with downstream procedures, restriction digests using the DNA fragments, restriction endonucleases and buffers (New England Biolabs) were set up and incubated at 37°C for 2 hours. To prevent the re-ligation of vector DNA during ligations, the 3' phosphate group was removed using alkaline phosphatase (Roche) and the DNA incubated at 37°C for a further 20 minutes. The DNA was removed of impurities from the restriction digests and using agarose gel electrophoresis analysed to use a 1:3 ratio of vector to insert DNA in a ligation reaction. Each ligation reaction consisted of DNA

fragments with compatible ends and used T4 DNA Ligase with the associated buffer (New England Biolab) and incubated at 16°C overnight.

2.3.3. Transformation of plasmid DNA into *Escherichia coli*

Plasmid DNA was transformed into *E. coli* cells using either heat shock, electroporation or the transformation and storage solution methods as explained by (Sambrook and Russel 2001). When the heat shock method was used, 50µl of competent cells were used and shocked at 42°C for 20 seconds. In the case of electroporation, 50µl of competent cells were used and a 2500V pulse was passed through the cells. When using the transformation and storage solution method, an OD_{600} of 0.3 was reached before proceeding. In all cases, LB media was used for cells to recover for 1 hour at 37°C. For each reaction, a positive and negative control were included, using vector only DNA and no DNA respectively.

2.3.4. Transformation of plasmid DNA into *Pseudomonas fluorescens*

Prior to the transformation of plasmid DNA into *P. fluorescens*, competent cells were made. For each aliquot of *P. fluorescens* competent cells, 5ml of overnight culture was used. Cells were harvested and the pellet resuspended in 2ml of 10mM Hepes pH8.5. The wash step was repeated three times and the cells finally resuspended in 100µl of 10mM Hepes pH8.5. Transformation of plasmid DNA into the competent cells was done using electroporation as described above in Section 2.3.3. Cells were recovered in LB for 3 hours at 28°C and plated onto selective agar.

2.3.5. Polymerase chain reaction (PCR) for specific DNA amplification

Specific regions of DNA fragments were amplified by PCR. Oligonucleotide primers were obtained from Eurofins. All of the primers and their sequences used throughout this study are shown in Table 2.4.

Table 2.4. The primers used throughout this study.

Primer name	Sequence 5' → 3'	Additional notes
*rccR*upF	CGGGATCCTCAGCCTGGTGCGCCAAG	For the deletion of *rccR*. *Bam*HI restriction enzyme site added at the 5' end
*rccR*upR	CGTCTAGATTGCAACAGGTTCAAAG	For the deletion of *rccR*. *Xba*I restriction enzyme site added at the 5' end
*rccR*dnF	CGTCTAGAGCCCTCGACGACTAGC	For the deletion of *rccR*. *Xba*I restriction enzyme site added at the 5' end
*rccR*dnR	CGGAATTCCGGTCAACATGGAATTG	For the deletion of *rccR*. *Eco*RI restriction enzyme site added at the 5' end
*rccA*upF	CGGGATCCCAGCGCTGTTCGTCGAG	For the deletion of *rccA*. *Bam*HI restriction enzyme site added at the 5' end
*rccA*upR	CGTCTAGATCGACGGAGCATACACAAG	For the deletion of *rccA*. *Xba*I restriction enzyme site added at the 5' end
*rccA*dnF	CGTCTAGAAATGCAGATTTTCCTGAG	For the deletion of *rccA*. *Xba*I restriction enzyme site added at the 5' end
*rccA*dnR	CGGAATTCATGTGCGGGTCACTGTAG	For the deletion of *rccA*. *Eco*RI restriction enzyme site added at the 3' end
*rccR*outUp	CGGTGCACCAACTGCGTCAA	To confirm the deletion of *rccR*
*rccR*outDn	AACGCGTGCACTTCACCCAG	To confirm the deletion of *rccR*
*rccA*outUp	CCGTTGATAAACCATTGG	To confirm the deletion of *rccA*
*rccA*outDn	GGTGGGCTTCAAGGTCAC	To confirm the deletion of *rccA*
*rccR*OEup	TCGAATTCTTGAACCTGTTGCAACATATCG	For the over-expression of RccR. *Eco*RI restriction enzyme site added at the 5' end
*rccR*OEdn	AAGGTACCATACAGCGCTAGTCGTCG	For the over-expression of RccR. *Kpn*I restriction enzyme site added at the 5' end
*rccA*RTPCRup	CCAGGACCAACTGACGG	For the amplification of the predicted promoter sequence of *rccA*
*rccA*RTPCRdn	GCTGTGGCCATTGACC	For the amplification of the predicted promoter sequence of *rccA*
*rpoD*RTPCRup	CAACGAAGTAGACGAAAGCTG	For the amplification of the promoter sequence of *rpoD*
*rpoD*RTPCRdn	GACGGTTGATGTCCTTGATCTC	For the amplification of the promoter sequence of *rpoD*
*rccR*HisUp	GCAAGCTTAAGAGTCCTGTCCTTTGAACC	*Hind*III restriction enzyme site added at the 5' end
*rccR*HisDn	GATCTAGAGTGGTGGTGGTGGTGGTCGTCGAGGGCTTTGACG	*Xba*I restriction enzyme site, followed by the His_6 tag added
*rccR*EMSAup	CCAATGCAGGCCGAAG	For the amplification of the predicted promoter sequence of *rccR*
*rccR*EMSAdn	CCAAGGATGTAGCAAGC	For the amplification of the predicted promoter sequence of *rccR*
*rccR*BgalUp	CACGGATCCCCTGGGCGTGAAGCTGG	*Bam*HI restriction enzyme site added at the 5' end
*rccR*BgalDn	TACGGATCCCCAAGGATGTAGCAAGC	*Bam*HI restriction enzyme site added at the 5' end

Primer name	Sequence 5’ → 3’	Additional notes
6-FAM For	CTAAAACGACGGCCAGT	5' 6-FAM labelled primers for fluorescent EMSA experiments.
6-FAM Rev	CAGGAAACAGCTATGAC	5' 6-FAM labelled primers for fluorescent EMSA experiments.
*pntAA*ChIP*up*	CTAAAACGACGGCCAGTTTGCAGGCGAACTT	To confirm the binding of RccR to *pntAA* using EMSA
*pntAA*ChIPdn	CAGGAAACAGCTATGACATCAGTCGGTGGGAT	To confirm the binding of RccR to *pntAA* using EMSA
*pckA*ChIPup	CTAAAACGACGGCCAGTATGTTGCCCAA	To confirm the binding of RccR to *pckA* using EMSA
*pckA*ChIPdn	CAGGAAACAGCTATGACATGCACGTGGGAA	To confirm the binding of RccR to *pckA* using EMSA
*aceE*ChIPup	CTAAAACGACGGCCAGTTTGCTCCAGGGCGGAA	To confirm the binding of RccR to *aceE* using EMSA
*aceE*ChIPdn	CAGGAAACAGCTATGACTTGCATCGGCCAT	To confirm the binding of RccR to *aceE* using EMSA
*mfd*ChIPup	CTAAAACGACGGCCAGTTTGAGCTCCAA	To confirm the binding of RccR to *mfd* using EMSA
*mfd*ChIPdn	CAGGAAACAGCTATGACTTGCCCAGGTCTA	To confirm the binding of RccR to *mfd* using EMSA
2154ChIPup	CTAAAACGACGGCCAGTTAAGCCCCTTGA	To confirm the binding of RccR to *PFLU2154* using EMSA
2154ChIPdn	CAGGAAACAGCTATGACATGGCGACACCTT	To confirm the binding of RccR to *PFLU2154* using EMSA
3817ChIPup	CTAAAACGACGGCCAGTTACCCGCACTGAT	To confirm the binding of RccR to *PFLU3817* using EMSA
3817ChIPdn	CAGGAAACAGCTATGACAACGATGGCTCAA	To confirm the binding of RccR to *PFLU3817* using EMSA
*glcB*ChIPup	CTAAAACGACGGCCAGTAAGCCCTCTGGAA	To confirm the binding of RccR to *glcB* using EMSA
*glcB*ChIPdn	CAGGAAACAGCTATGACAAGTCTGCGGCTT	To confirm the binding of RccR to *glcB* using EMSA
*rccR*ChIPup	CTAAAACGACGGCCAGTAAGCAGCACGTGAT	To confirm the binding of RccR to *rccR* using EMSA
*rccR*ChIPdn	CAGGAAACAGCTATGACATCATCGCCACGTT	To confirm the binding of RccR to *rccR* using EMSA
ARB-1A	GGCCAGCGAGCTAACGAGACNNNNGTTGC	For identification of Tn insertion sites by arbitrary PCR.
ARB-1	GGCCAGCGAGCTAACGAGAC	Standard primer for arbitrary PCR identification of Tn insertion sites.
ALMAR3-PCR	CGCAAACCAACCCTTGGCAG	Amplification primer for arbitrary PCR for pALMAR3 genomic insertions
ALMAR3-seq	ACATATCCATCGCGTCCGCC	Sequencing primer for pALMAR3 genomic insertions after arbitrary PCR
*PFLU0179*UPF	CGGGATCCCTTCCTGTATGCCTCCAG	For deletion of *PFLU0179*, Tn-suppressor of Δ*rccA* motility defect. *Bam*HI restriction enzyme site added at the 5’ end

Primer name	Sequence 5’ → 3’	Additional notes
*PFLU0179*UPR	CGTCTAGAAAGGGTGGTGTCTGCTAAG	For deletion of *PFLU0179*, Tn-suppressor of Δ*rccA* motility defect. *Xba*I restriction enzyme site added at the 5’ end
*PFLU0179*DNF	CGTCTAGACAGGGCCAATTTGGAAAG	For deletion of *PFLU0179*, Tn-suppressor of Δ*rccA* motility defect. *Xba*I restriction enzyme site added at the 5’ end
*PFLU0179*DNR	CGGAATTCGATCGGTCAATGTGTGGTC	For deletion of *PFLU0179*, Tn-suppressor of Δ*rccA* motility defect. *Eco*RI restriction enzyme site added at the 5’ end
pHexRUpF	CGGGATCCGAAATCTCAAACGAATCC	For mutagenesis of HexR binding site upstream of *PFLU0180* (*gabD)* gene.
pHexRUpR	GATTCCCAGCTGTTCCGCCAGCAGGCCTTTATC	For mutagenesis of HexR binding site upstream of *PFLU0180* (*gabD)* gene.
pHexRDNF	GATAAAGGCCTGCTGGCGGAACAGCTGGGAATC	For mutagenesis of HexR binding site upstream of *PFLU0180* (*gabD)* gene.
pHexRDNR	CGGAATTCGTGAAGGACAATTTGCGCAC	For mutagenesis of HexR binding site upstream of *PFLU0180* (*gabD)* gene.
*hexR*upF	CGGGATCCCGGCTTTCATGAAGTCG	For the deletion of *hexR. Bam*HI restriction enzyme site added at the 5’ end
*hexR*upR	CGTCTAGAGCACGCGGTCCATTC	For the deletion of *hexR. Xba*I restriction enzyme site added at the 5’ end
*hexR*dnF	CGTTAGACGAGTTCAACTGAGCC	For the deletion of *hexR. Xba*I restriction enzyme site added at the 5’ end
*hexR*dnR	CGGAATTCGGCGTATTGGTCACG	For the deletion of *hexR. EcoR*I restriction enzyme site added at the 5’ end
*hexR*outUp	GCATCGTTGACCTTGCGC	To confirm the deletion of *hexR*
*hexR*outDn	GCAATTGCAGCTGTGCG	To confirm the deletion of *hexR*
*hexR*purifyF	GACATATGGACCGCGTGCGAAATC	*NdeI* restriction enzyme site added at the 5’ end
*hexR*purifyR	TCCTCGAGGTTGAACTCGTCGCCC	*XhoI* restriction enzyme site added at the 5’ end
*hexR*OEup	TCGAATTCATGGACCGCGTGCGAAATC	For the over-expression of HexR. *EcoR*I restriction enzyme site added at the 5’ end
*hexR*OEdn	AAGAGCTCCACAGCCTGGCTCAGTTG	For the over-expression of HexR. *Sac*I restriction enzyme site added at the 5’ end
pME6032M2Rev	CTAGTCCGAGGCCTCGAGATCTATCGATGCATGCCATGGTACCCATATGAATATCCTCCTTAG	For FLAG tagging genes inserted downstream of the *Kpn*I restriction site in pME6032
RccRM2For	CAGCCTGCGTCTGTCGCCCAAATCCGTCAAAGCCCTCGACGACGACTACAAAGACCATGACGG	For the FLAG tagging *rccR* in pME6032 vector with an inducible promoter.
HexRM2ChrFor	ATGGTACCGCCAGGCTGTGTGTTGG	To amplify HexR with a M2 tag incorporated in the chromosome.
HexRM2ChrRev	TAGGATCCCGTGCTGGTGATCGACC	To amplify HexR with a M2 tag incorporated in the chromosome.
HexRM2For	GCTTGAATGACAGCCGCTACCCGGTGGGCGACGAGTTCAACGACTACAAAGACCATGACGG	For the FLAG tagging of HexR in the pME6032 vector with an inducible promoter.

Depending on the DNA polymerase being used and the expected size of the DNA fragment, the temperature and length of time at each stage of the cycle varied. *Taq* DNA polymerase amplifies DNA at 1kb per minute, while Phusion DNA polymerase has an extension time of 1kb per 30 seconds as summarised in Table 2.5.

Table 2.5. The cycling conditions used in PCR.

Stage	Temperature	Duration	Number of cycles
1	95°C 98°C	5 minutes (*Taq*) 2 minutes (Phusion)	1
2	95°C	30 seconds	25
	50°C - 72°C	45 seconds	
	72°C	1 minute per kb (*Taq*)	
		30 seconds per kb (Phusion)	
3	72°C	5 minutes	1

The choice of DNA polymerase was dependent on the downstream procedures the DNA fragment would be used for. As *Taq* DNA polymerase (New England Biolabs) has no proof reading ability, this polymerase was utilised to screen for potential mutants. On the other hand, Phusion DNA polymerase (New England Biolabs) was used when the resulting DNA fragment was required for downstream DNA manipulation techniques as this polymerase has a high fidelity.

2.3.6. DNA sequencing

DNA sequencing, as described by (Slatko et al. 2001), was used to confirm the sequence of a construct in a plasmid, or the absence of a gene in a mutant. The pre-mix Big Dye Terminator (version 3.1) with the accompanying x5 Sequencing Buffer (Life Technologies Ltd.) and the required primer was used at a concentration of 10mM in each sequencing reaction. Amplification of the DNA fragment was carried out in a thermo cycler using cycles of denaturation, annealing and elongation of the DNA, as shown in Table 2.6.

Table 2.6. The cycling conditions used in a DNA sequencing reaction.

Stage	Temperature	Duration	Number of cycles
1	96°C	1 minute	1
2	96°C	10 seconds	25 cycles
	50°C	5 seconds	
	60°C	4 minutes	
3	4°C	10 minutes	1

Reactions were sequenced by Eurofins and the chromatograms analysed to identify any errors in base calling and confirm the regions of sequence that could be used with confidence for downstream processes.

2.3.7. Reverse transcriptase polymerase chain reaction (RT PCR)

For each RT PCR, the One Step RT PCR kit (Qiagen) was used. Prior to each reaction the RNA required was thawed on ice and immediately placed back in ice. For each condition, a control pair of primers annealing to the *rpoD* gene, in addition to a 'no template' control, was used. For all reactions, the final concentration of RNA utilised was 0.25ngμl^{-1} which was diluted using RNase free water. The denaturation of the RNA followed by the amplification of DNA was carried out in a thermo cycler using cycles of denaturation, annealing and elongation of the DNA, shown in Table 2.7.

Table 2.7. The cycling conditions used in a RT PCR reaction.

Stage	Temperature	Duration	Number of cycles
1	50°C	30 minutes	1
2	95°C	15 minutes	1
3	94°C	30 seconds	35 cycles
	60°C	30 seconds	
	72°C	1 minute	
4	72°C	10 minutes	1

The products were separated on a 1.8% agarose gel. 4μl of gel loading dye was mixed with 6μl of PCR product and 9μl of this mixture was loaded into the gel.

2.3.8. Making markerless gene deletions in *Pseudomonas fluorescens*

Gene deletions in *P. fluorescens* were created using an adaptation of the method described by (Voisard et al. 1994). Briefly, 500bp of upstream and downstream flanking DNA was amplified and ligated into pME3087 between *Eco*RI and *Bam*HI. The resulting vectors were transformed into *P. fluorescens* with single crossovers selected using tetracycline and re-streaked. Cultures of single colonies were grown overnight in LB, diluted 1 in 100 into fresh media and grown for 2 hours before 5µlml^{-1} of tetracycline was added to inhibit the growth of those cells that had lost the tetracycline cassette. After an additional hour of growth, cells were harvested and the pellet resuspended in fresh LB containing 5µlml^{-1} tetracycline plus phosphomycin and piperacillin to kill any remaining growing bacteria. The cultures were returned to grow for a further 4 – 6 hours before being washed with LB and a serial dilution made and plated on LB agar. Single colonies were selected and patched onto LB with/out tetracycline and those that were tetracycline sensitive then underwent colony PCR screening testing for gene deletion/modification. A summary of the molecular work involved in creating the mutants is provided in Figure 2.1.

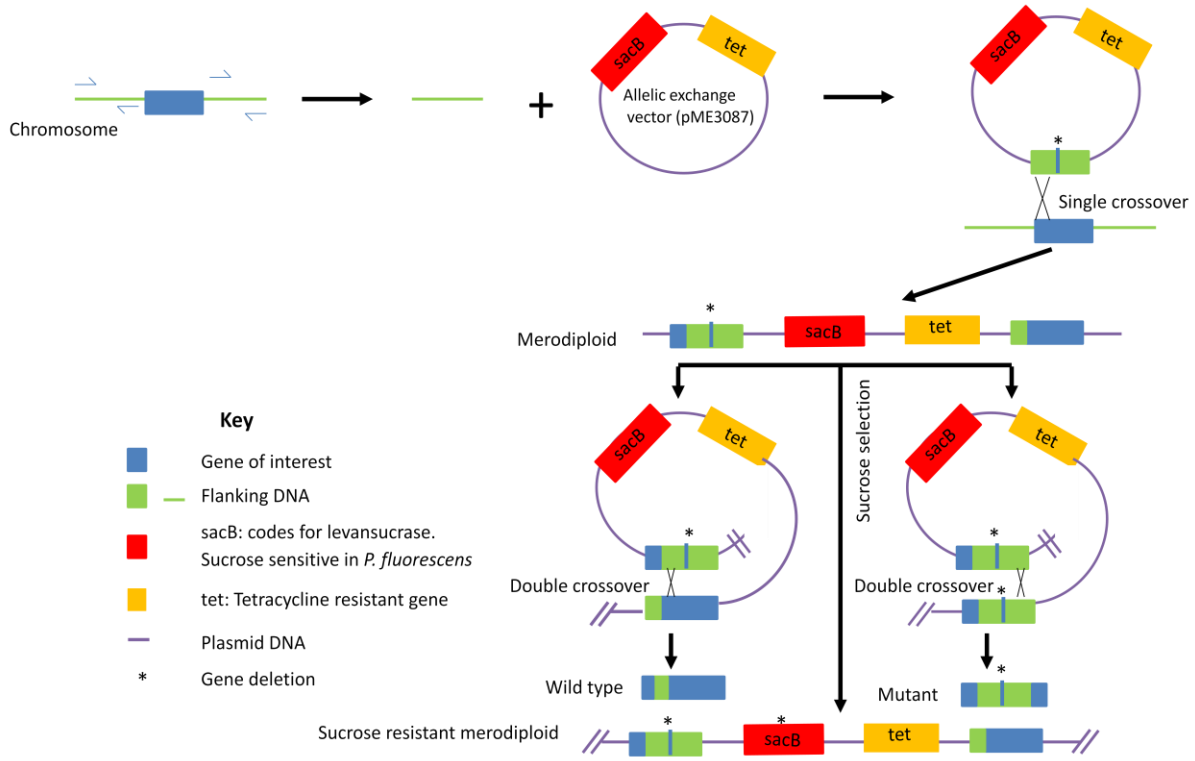

Figure 2.1. An overview of the molecular work involved to create markerless *P. fluorescens* deletions. Part of this summary has been adapted from (Hmelo et al. 2015).

2.3.9. The over-expression of genes in *Pseudomonas fluorescens*

The plasmid pME6032 was used to over-express the gene of interest. An explanation of the method is given by (Heeb et al. 2000). Each gene of interest was amplified and ligated into the pME6032 plasmid between *Eco*RI and *Kpn*I, unless the primers, found in Table 2.4, state otherwise.

2.3.10 C terminal FLAG epitope tagging by λ red recombination

Using the pME6032 over-expression plasmids of the genes, the addition of a C terminal FLAG epitope tag to the gene was carried out as explained by (Yu et al. 2000). Firstly, the gene of interest and the FLAG epitope were amplified from the relevant pME6032 over-expression plasmid and the pSUB11 plasmid respectively. Both of these fragments were transformed into the temperature sensitive λ red *E. coli* BW251123 pIJ790 and grown at 28°C. The pIJ790 plasmid was removed using a second transformation and the cells grown at 37°C. The pME6032 plasmid was removed and the sequence of the FLAG epitope was confirmed. Following this confirmation, the pME6032 plasmid was transformed into *E. coli* DH5α followed by the pFLP2 in order to remove the downstream kanamycin gene as explained by (Hoang et al. 1998). A number of colonies were picked and grown on both LB with tetracycline ± kanamycin, and only those in the absence of kanamycin were of interest. Finally, the pFLP2 plasmid was removed by the addition of sucrose onto the LB and tetracycline plates. The pME6032-gene-FLAG plasmid was removed for further use. A summary of the method used is shown in Figure 2.2.

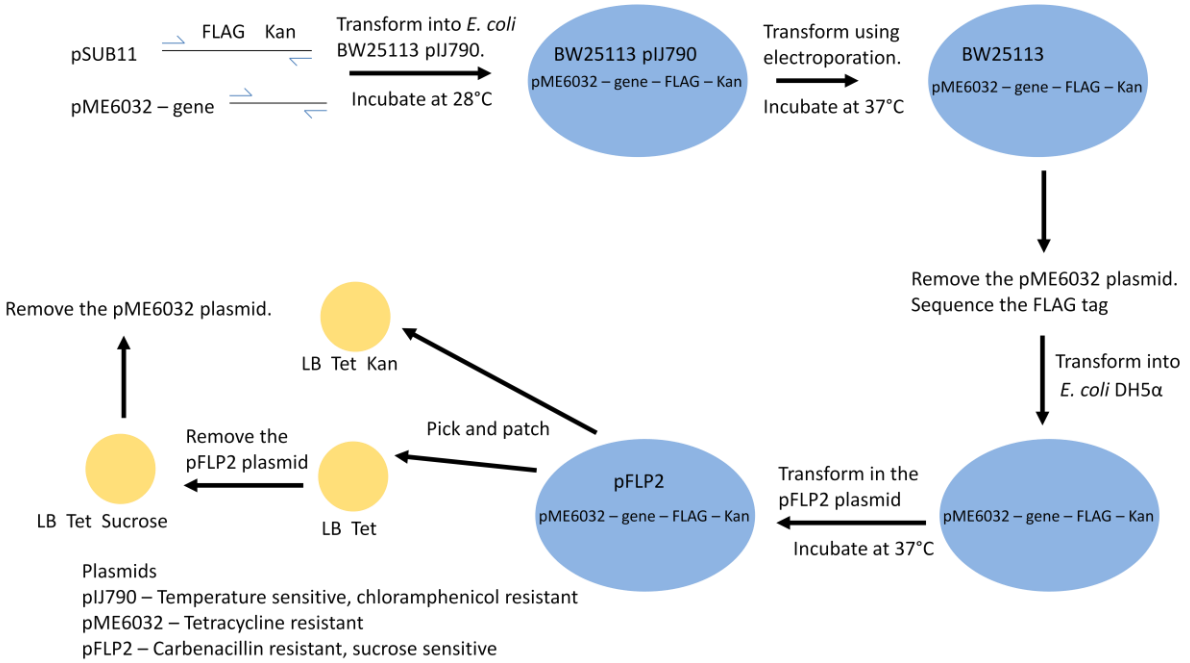

Figure 2.2. Diagrammatic explanation of the C terminal FLAG epitope tagging by recombination method.

2.3.11. Incorporation of the FLAG tag into the SBW25 chromosome

This utilised previously constructed strains made by C terminal FLAG epitope tagging be recombination method as described in section 2.3.10 and the two step allelic exchange method explained by (Hmelo et al. 2015). Figure 2.3 provides a summary of the stages involved in this procedure. In Stage 1, the gene of interest with the C terminal FLAG tag was removed from the pME6032 plasmid using the appropriate restriction enzymes, and the downstream chromosomal gene was amplified. A triple ligation resulted in this new construct. Stage 2 involved inserting this new construct into the allelic exchange vector pTS1 at the relevant restriction enzyme sites. Finally in Stage 3, when transformed into SBW25, a single crossover event could occur allowing the two step allelic exchange to continue. To identify if the FLAG epitope had been incorporated into the SBW25 chromosome successfully, PCR and Western blot analysis was used.

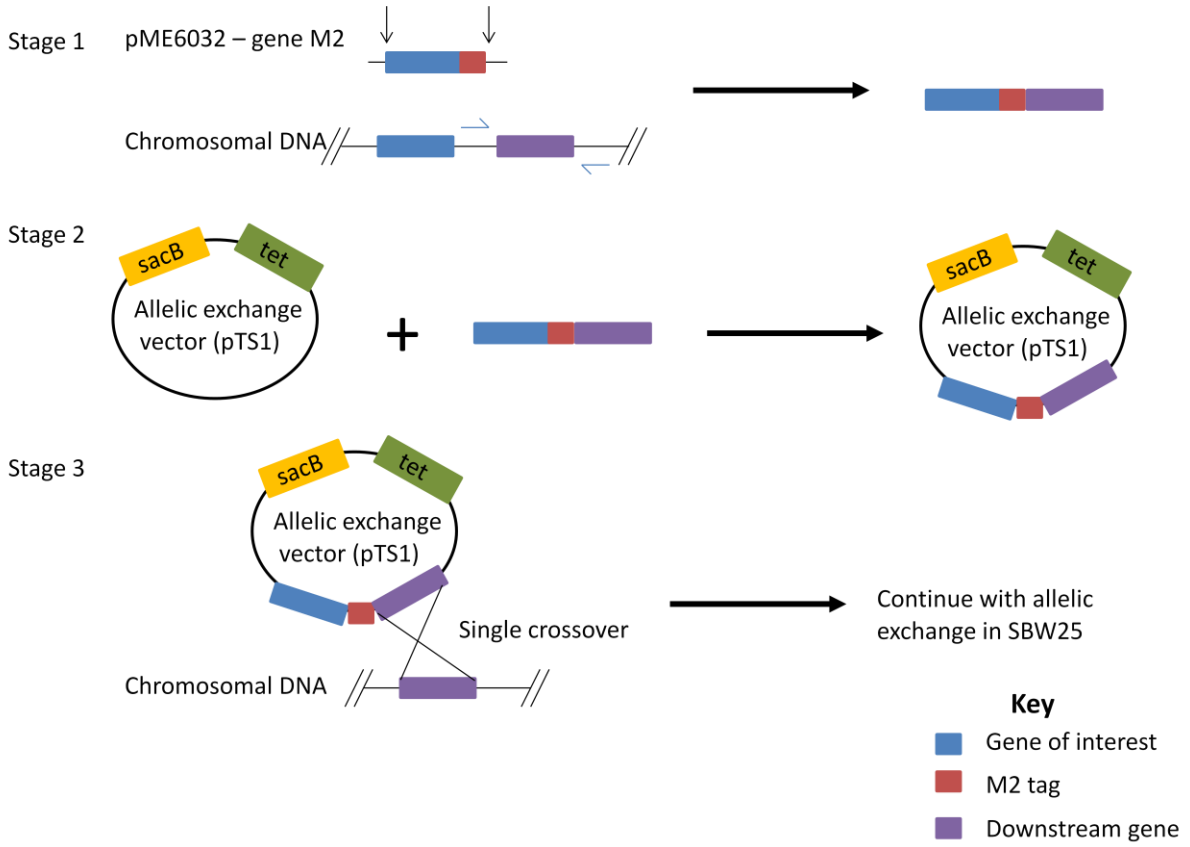

Figure 2.3. A schematic diagram of the FLAG tag was incorporated into the SBW25 chromosome.

2.3.12. Mariner transposon screen

The pALMAR3 plasmid was transformed into the Δ*rccA* mutant strain and plated onto selected agar to produce a library of thousands of individual random transposon insertion mutants. Single colonies were picked at random, grown in 96 well plates containing LB overnight before swarming motility was observed using 0.5% KB agar. Isolates that reverted the Δ*rccA* mutant phenotype back to the wild type phenotype (Figure 2.4) were re-grown in LB and a second swarming motility assay using M9 minimal media supplemented with pyruvate was done. For more information on the swarming motility assay see Section 2.5.4. Samples that still reverted the *rccA* mutant phenotype back to the wild type on the minimal media plate were of interest. The transposon insertion sequence was amplified from the target colony by semi-random PCR in each case, sequenced and identified using a BLAST search. Figure 2.4 provides a brief summary of how the transposon screen was done.

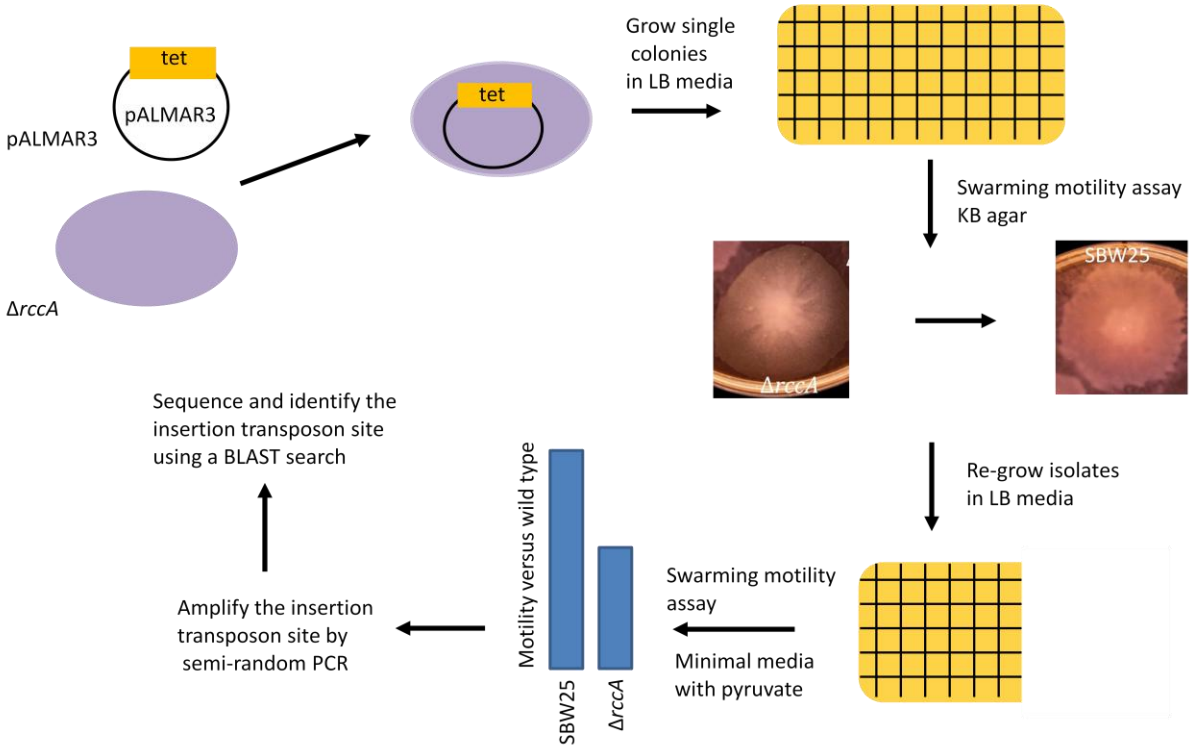

Figure 2.4. A summary of how the mariner transposon screen was done.

2.4. Biochemical assays

2.4.1. Protein purification

For the purification of His_6 tagged proteins, the over-expression vector pET42b and *Escherichia coli* BL21 (DE3) pLysS cells were used. The full length gene of interest was amplified and cloned in-frame into pET42b replacing the start and stop codons with the restriction enzymes being utilised in the cloning procedure. To ensure the resulting protein fusion was inducible, a small scale induction was done by using an overnight LB culture of cells and a 1 in 50 dilution to inoculate 50ml of LB with the appropriate antibiotics. Cells were allowed to grow at 37°C until an OD_{600} of 0.6 was reached and the gene of interest was induced using IPTG at a final concentration of 100µM whilst a 5ml sample was removed as non-induced control, and both samples were replaced at 37°C for a further 2 hours. Samples were analysed by SDS-PAGE to identify if the protein was inducible. In preparation for purification, large scale protein induction was carried out in a similar way to the small scale induction as above, however, 750µl of cell culture was used to inoculate a 50ml LB culture

and incubated overnight before 7.5ml of this was used to inoculate 500ml LB in a 2L conical flask, each time with the appropriate antibiotics present. Cultures were grown until an OD_{600} of 0.6 was reached, induced using IPTG at a concentration of 100µM and a non-induced control was also prepared. After 2 hours of induction, the cells were harvested by centrifugation at 4°C and the pellet resuspended in 30ml of Equilibration buffer (this varied depending on the protein; see individual protein purification results for the buffer used). The cells were lysed using a French press set at 1000psi and repeated once. The soluble protein was retrieved by centrifugation at 4°C for 20 minutes. Protein purification was done using NTA-Ni chromatography, using 1ml HiTrap Chelating columns (GE Healthcare) equilibrated with the Equilibration buffer and loaded with the soluble protein. The sample was removed from the column using a gradient of 0% - 100% of Wash buffer, and 1ml fractions of this collected. 20µl of each fraction was removed and prepared for SDS-PAGE analysis, while the remainder of the fraction was stored in a final concentration of 2% glycerol and snap frozen using liquid nitrogen.

Identification of fractions containing the purified protein was done by SDS-PAGE. Each fraction, along with non-induced, induced samples, broken cells, the soluble cell and the pellet samples were analysed by SDS-PAGE. All of the additional samples were taken at the relevant points in the procedure.

2.4.2. SDS – PAGE

SDS-PAGE analysis was used to analyse and differentiate the sizes of proteins. 12.5% acrylamide gels, were utilised unless otherwise stated. Each protein sample was mixed with x2 SDS Buffer and denatured at 99°C for 5 minutes before being loaded onto the gel in addition to a protein markers of known molecular masses (Expedeon). Gels were run at a constant voltage of 200V for approximately 1.5 hours or until the dye had reached the bottom of the glass plates in x1 Running Buffer. Gels were visualised using InstantBlue solution (Expedeon) and a light box.

2.4.3. Western blot analysis

Western blot analysis was used to identify specific proteins using antibodies to target and visualise these proteins. The proteins are separated using an acrylamide gel followed by the transfer of the proteins from the gel to a membrane by electrophoretic transfer with the membrane between the gel and the positive electrode. Non-specific binding of proteins to the membrane is prevented by the use of the blocking solution. A summary of the technique, theory and some trouble-shooting is given by (Mahmood and Yang, 2012). Differences to the method used were: 12.5% acrylamide gels; polyvinylidene difluoride membranes (Millipore); a blocking solution of Tween, Milk powder and TBS (TMT); the primary antibody was diluted in TMT using a final dilution of 1 in 2,000; an anti rabbit secondary antibody at a dilution of 1 in 10,000; detection of proteins used the enhanced chemiluminescence Prime Western Blotting Detection Reagent (GE Healthcare).

2.4.4. Electrophoretic mobility shift assay (EMSAs)

Two types of electrophoretic mobility shift assay (EMSAs) were done throughout this study; the first with ethidium bromide being used to stain the DNA and the second fluorescent EMSAs. In each case, the promoter region, or predicted promoter region, was amplified from the *P. fluorescens* SBW25 chromosome and the protein of interest was used. When ethidium bromide was used to stain the DNA, the method as explained by (Schumacher et al. 2017) was used. When fluorescent EMSAs were used the method described by (Crack et al. 2015) was utilised.

2.4.5. Chromatin immunoprecipitation sequencing (ChIP seq) assay

Chromatin immunoprecipitation sequencing (ChIP seq) assays were done using a similar protocol described by (Kuras and Struhl, 1999), however some adaptations were made. Following bacterial growth, cultures were incubated with 1% formaldehyde at room temperature for 20 minutes followed by the addition of 125mM glycine for 5 minutes to stop the crosslinking reaction. The cells were harvested and washed four times using cold PBS before being lysed in 1ml of lysis buffer. Following this, the DNA of the cell extracts was

fragmented to an average of 500bp using sonication. 50µl of this extract was removed for 'total DNA preparation'. For the immunoprecipitation of RccR crosslinked DNA, 1ml of the extract was immunoprecipitated with 10µl of polyclonal anti-RccR antibody at 4˚C for 4 hours. Once incubation with ProteinA affinity gel (Sigma) at 4˚C for an hour had occurred, the beads were washed twice with IP buffer and lastly TE buffer. The immunoprecipitated material was eluted using 100µl of elution buffer. To reverse the crosslinking of immunoprecipitated and 'total DNA' samples were incubated at 65°C overnight. Following Proteinase K (Roche) treatment, the immunoprecipitate and 'total DNA' samples were extracted using phenol-chloroform and purified using a QiaQuick kit (Qiagen). Illumina TruSeq ChIP-seq libraries were produced using these samples and selected to approximately 200 – 300bp before being sequenced on a single Illumina HiSeq 2000/2500 lane in High Output mode using 100bp single-end reads. A summary of this procedure can be seen in Figure 2.5.

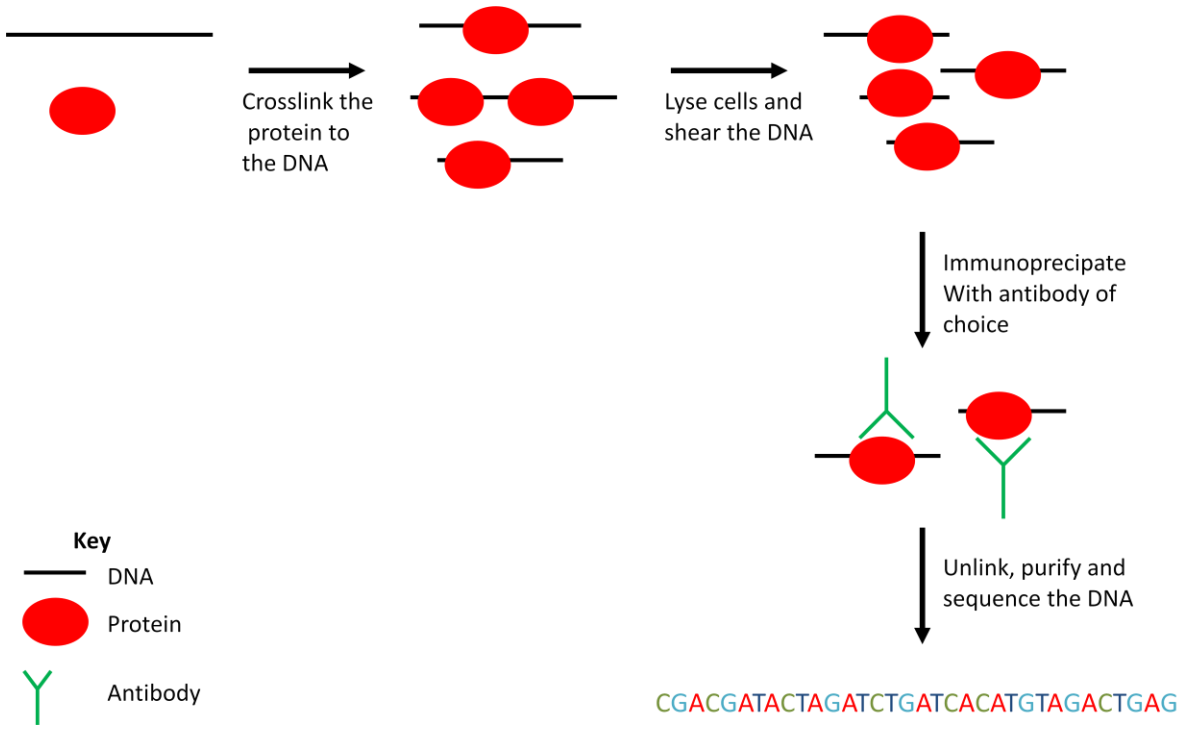

Figure 2.5. Diagram of the laboratory stages of the chromatin immunoprecipitation sequencing assay.

Following the sequencing of libraries, reads were provided in fastq files and aligned to the *P. fluorescens* SBW25 genome using the bowtie2 software resulting in single SAM (.sam) files for each fastq file. Further analysis was done using a combination of Perl scripts dependent on the BioPerl toolkit and R scripts. For each SAM file, the coverage (the number of reads

mapping) to each nucleotide of the SBW25 genome was calculated and saved in files referred to as coverage files. Using these coverage files, a local enrichment was calculated using a moving window of 51 nucleotides and moving in steps of 21 nucleotides (the sum of coverage at each nucleotide position in the 51 nucleotide window divided by the sum of coverage at each nucleotide position in a 4,001 nucleotide window centred around the 51 nucleotide window), resulting in an enrichment value for every 25 nucleotides in the SBW25 genome. Any enrichment value of less than 1.5 was removed followed by the fitting of a negative binomial distribution to the data using the fitdistr function of the MASS package in R resulting in the size and mu parameters of the binomial distribution. The values of size and mu parameters were used to calculate the p values for every enrichment ratio using the pnbinom function of R. Using the p.adjust function of R, the p values were adjusted for multiple testing using the Benjamini and Hochberg method resulting in tables with three columns titled: Genomic position; Enrichment ratio and Adjusted p value. Using the Scaffold programme, it was identified where in the genome the protein bound to the DNA in the SBW25 wild type strain and not the mutant strain being tested. Once these positions had been identified, the gene/s which the protein was binding to could be identified. An overview of this is shown in Figure 2.6.

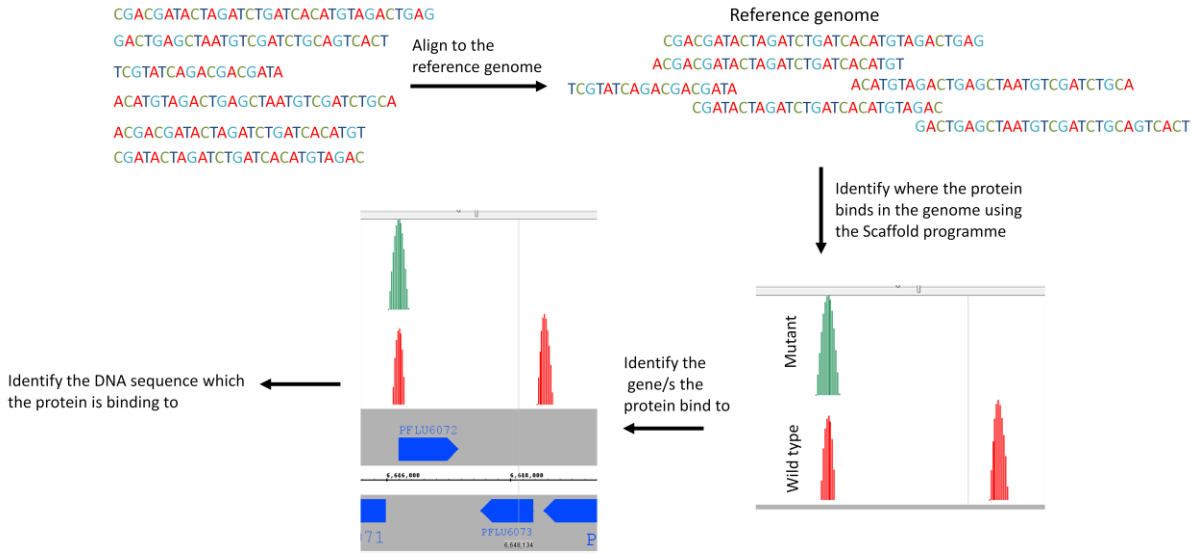

Figure 2.6. Simple diagram of how the raw sequencing data occurred following the chromatin immunoprecipitation sequencing analysis.

2.4.6. β galactosidase assay

The predicted *rccR* promoter sequence was amplified and ligated into the pGm6032 plasmid at the *Bam*HI site. Potential transformants were PCR screened and sequenced ensuring the promoter DNA fragment was inserted in the correct orientation. The final plasmid construct was transformed into *P. fluorescens* SBW25 and *P. fluorescens* SBW25 Δ*rccR*.

The β galactosidase assay was done at 30°C using overnight cultures of minimal media supplemented with either pyruvate or glucose. The cells were broken open using Lysis buffer and chloroform, then the β galactosidase activity was measured using ONGP as a substrate and the reaction stopped using 1M sodium carbonate. Notes were taken of the OD_{600} of the starting culture, the OD_{420} and OD_{550} of the final reaction and the length of time the reaction took. Each sample was duplicated and the pGm6032 plasmid was used to give a baseline reading.

The following formula was used to calculate the β galactosidase activity of each sample:

$$\frac{1000 \times (OD_{420} - (1.75 \times OD_{550}))}{\text{Time taken (minutes)} \times \text{Volume of sample added (ml)} \times OD_{600}}$$

2.5. Additional assays

2.5.1. Root colonisation assay

The root colonisation assay is a competition experiment and requires both the mutant being tested and the SBW25-*lacZ* strain. A minimum of eight plants were utilised for each *P. fluorescens* mutant. The full method for the root colonisation assay has explained by (Campilongo et al. 2017). In summary, paragon wheat seeds were sterilised using 70% ethanol and 5% hypochlorite and left to germinate in the dark for 72 hours on 0.8% MS agar. Individual seeds were transferred to sterile 50ml tubes containing medium grain vermiculite with Nutrient solution and grown in a controlled environment, at 25°C with a 16 hour light cycle, for seven days. After seven days of growth, the seeds were ready to be inoculated with bacteria: overnight cultures were grown in minimal media supplemented with pyruvate and a

serially diluted using phosphate buffer. In total, 1 x 10^3 CFU of both mutant and SBW25-*lacZ* were used to inoculate the plants. The plants were grown for a further seven days, when the shoots were removed and 20ml PBS added to assist with the suspension of the bacteria by vortexing. A serial dilution was created and plated onto IPTG, X-GAL and carbenicillin plates. The SBW-*lacZ* strain formed blue colonies while the mutant created white colonies. The ratio of blue:white colonies in each case was calculated and used to infer the relative competitive colonisation ability of the mutant strain.

2.5.2. Growth assay

The bacterial growth of cells was measured using a microplate spectrophotometer and a similar method described by (Campilongo et al. 2017). A total growth volume of 200µl per well was used and 10µl of cell culture at an OD_{600} of 0.05 was used to initiate growth.

2.5.3. RNA extraction of *Pseudomonas fluorescens* from liquid culture

The method used to extract the RNA from *P. fluorescens* in liquid culture is explained by (Campilongo et al. 2017) with a few amendments made. For each strain, three 5ml LB overnight cultures were used to inoculate three 50ml M9 pyruvate cultures using a 1 in 100 dilution and these were grown for 24 hours with shaking. RNA quantification was carried out using the Qubit RNA assay and the Qubit Fluorometer (Life Technologies).

2.5.4. Swarming motility assays

Swarming motility assays were done late in the afternoon and incubated overnight at room temperature.

The overnight LB cell cultures and dilutions of 1 in 50 and 1 in 100 were used to inoculate minimal media supplemented with pyruvate and grown for 6 hours. The OD_{600} of each set of dilutions was measured and the most consistent set of cultures used for the assay by normalising all the OD_{600} measurements to that of the lowest using LB media. Each

swarming motility plate had 25ml of agar present, made to an overall consistency of 0.3% agar and each plate was dried evenly for exactly 1 hour in a flow hood. 2µl of the normalised LB cultures were spotted onto the swarming plates. Up to four spots were placed onto each plate. On each plate, different strains were present to ensure overall consistency, and at least three replicates were done for each strain in each assay.

Two different sets of plates were made; a nutrient rich media (KB, with NaCl and Congo Red dye) or minimal media supplemented a carbon source. For KB agar plates, the phenotype of the swarms were photographed, whilst for the minimal media plates the diameters of the swarms were measured, following 16 and 20 hours of incubation.

Chapter 3 – RccR and RccA are Crucial in Plant Colonisation

3.1. Introduction

Previous studies, such as (Matilla et al. 2007) and (Gao et al. 2014), have shown that cyclic-di-GMP (c-di-GMP) may have an important role in the rhizosphere of plants. The early years of the Malone laboratory have for this reason, been interested in understanding how c-di-GMP may be a significant factor in the rhizosphere. In 2009, (Silby et al. 2009), used an in vivo expression technology (IVET) study to identify 146 specifically up-regulated genes in the *Pseudomonas fluorescens* SBW25 sugar beet rhizosphere. Of the 146 genes identified, seven of these were of relevance to the Malone laboratory due to their potential ability to either produce or degrade c-di-GMP. Figure 3.1 shows the seven genes that were of interest to the Malone laboratory. As seen in Figure 3.1, the domain structure of each protein varies; some only contain a GGDEF or an EAL domain, whilst others have both GGDEF and EAL domains and finally in addition to the GGDEF and/or EAL domain of the protein some also include sensory and/or transmembrane domains.

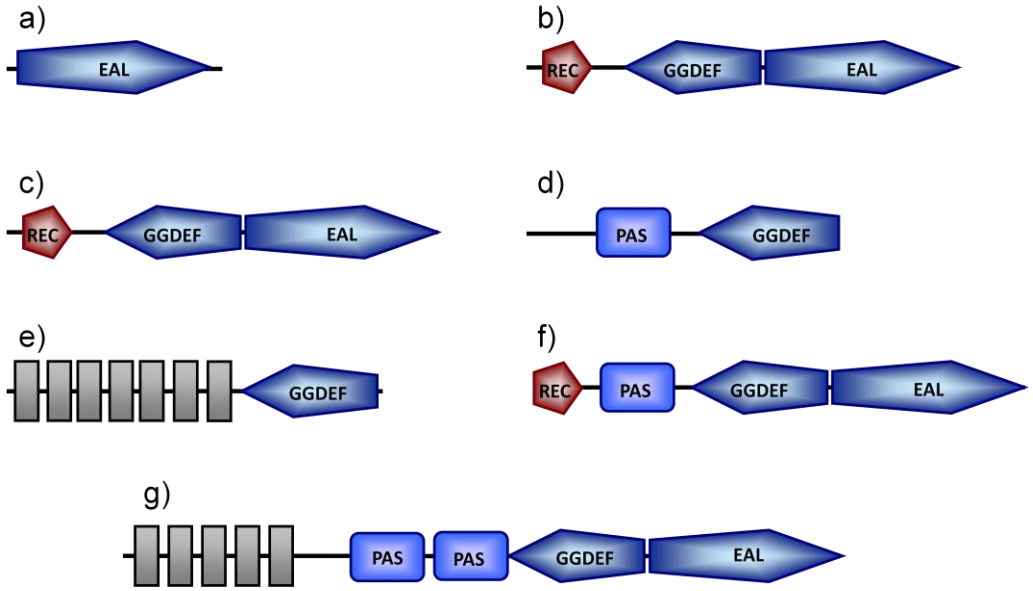

Figure 3.1. The seven up-regulated genes identified in *P. fluorescens* of interest to the Malone laboratory. The genes are a) *PFLU0261*, b) *PFLU1114*, c) *PFLU4858*, d) *PFLU5127*, e) *PFLU5608*, f) *PFLU5698* and e) *PFLU6074*. The GGDEF and EAL domains synthesise and degrade c-di-GMP respectively, while the REC and PAS domains are two different types of sensory domains. The grey boxes represent transmembrane domains.

Two of these seven genes, *PFLU4858* and *PFLU5698*, have already been studied in *Pseudomonas aeruginosa* (Kuchma et al. 2007) and *Pseudomonas putida* (Matilla et al. 2011) respectively. PFLU4858 is an orthologue of BifA which has mainly been studied in *P. aeruginosa* where it was shown to inversely regulate biofilm formation and the swarming motility of cells (Kuchma et al. 2007). In addition to *P. aeruginosa*, BifA has also been investigated in *P. putida* and *Pseudomonas syringae* where it has been shown to also regulate the formation of biofilms (Jiménez-Fernández et al. 2015) and to be involved in the virulence of bacteria in the *P. syringae* complex (Aragón et al. 2015) respectively. *PFLU5698* is an orthologue of *rup4959* which has been shown to be important in the *P. putida* maize rhizosphere (Matilla et al. 2011). At the beginning of this study, the remaining five genes had not been studied in *P. fluorescens* or any other organism. However, further work has now been done on two of five genes, *PFLU0261* and *PFLU6074*. More information on *PFLU0261*, now known as RimA, can be found in Section 1.4.5, and *PFLU6074* will be discussed in this chapter.

In order to decide which of the remaining uncharacterised five proteins to study further preliminary experiments, using mutants of each gene, relating to c-di-GMP, for example the ability to produce a biofilm, the motility of each strain and the ability to bind to Congo Red dye, were done. Due to the results of the preliminary experiments and in addition to the unusual structure of the protein, this will be discussed further in Section 3.3.1, it was decided to study *PFLU6074*. *PFLU6074* was later re-named *rccA* (for rhizosphere colonisation controller), and some of the preliminary experiments were investigated further.

Although there have been previous studies using *P. fluorescens* SBW25 (Silby et al. 2009), the wheat rhizosphere, root colonisation and the potential link of c-di-GMP (Matilla et al. 2007), no study has combined all of these together. This chapter will focus on two genes: *rccA* and the immediately downstream *rccR* from *P. fluorescens* to understand how they affect root colonisation in the wheat rhizosphere.

3.2. Aims

During this chapter, it is aimed to:

- create non-polar *rccA*, *rccR* and *rccAR* mutants in *P. fluorescens* SBW25.
- test the *rccA*, *rccR* and *rccAR* mutants and identify if they have effects on root colonisation in the *P. fluorescens* SBW25 wheat rhizosphere.
- use swarming motility tests to identify if *rccA* and *rccR* are functionally linked.

3.3. Results

3.3.1. Computational analysis of *rccR* and *rccA*

In order to understand more about *rccA*, the *Pseudomonas* Genome DB website (Winsor et al. 2006) was utilised to examine the gene and the surrounding area more closely. From this, it was found that a second gene, *PFLU6073* and since re-named *rccR*, was only approximately 160bp downstream from *rccA*, raising the possibly of the two genes potentially being in an operon. The genetic organisation of *rccR* and *rccA* can be seen in Figure 3.2.

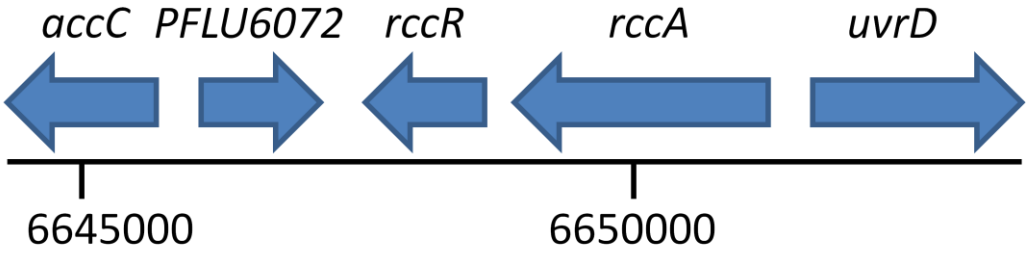

Figure 3.2. The position and organisation of *rccR* and *rccA* in the *P. fluorescens* chromosome. The direction of the arrowhead specifies the direction in which transcription of the gene takes place. The numbers underneath the black line indicates where in the chromosome the genes can be found, with the numbers representing the number of base pairs from the origin. The surrounding genes *accC*, *PFLU6072* and *uvrD* encode for pyruvate carboxylase subunit A, a LysR transcriptional regulator and a DNA-dependent helicase II respectively.

Before laboratory work began on *rccR* and *rccA*, computational analysis of the two genes and the proteins were done. The non-coding upstream DNA sequences of each gene were analysed using the promoter predicting programme on the Berkeley Drosophila Genome Project website, http://www.fruitfly.org/seq_tools/promoter.html, to identify if any potential promoter sequences were present. In addition to this, the amino acid sequences of the encoded proteins were examined to identify if they were members of any families of known proteins using the Pfam website, http://pfam.xfam.org/ (Finn et al. 2016).

From the computational analysis undertaken, it was discovered that RccA is a putative c-di-GMP protein containing both a GGDEF and an EAL domain. In addition to these GGDEF and EAL domains, RccA also has two sensory domains known as PAS domains, and five transmembrane domains, for more information about the importance of sensory domains for c-di-GMP proteins, see Section 1.4.4. Due to these transmembrane domains, it can be concluded that RccA is sitting in the membrane of the cell and could potentially be used to send signals to either side of the membrane. The organisation of the domains of RccA are shown in Figure 3.3.

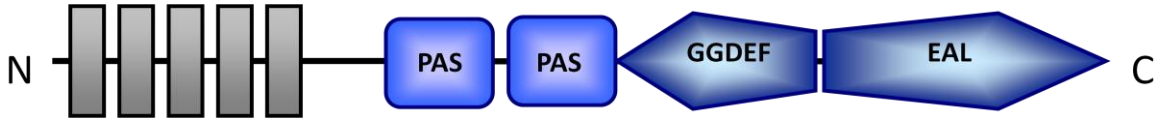

Figure 3.3. The domain organisation of RccA. The five grey rectangles represent the transmembrane domains, while the two sensory PAS domains and the GGDEF and EAL domains are represented by blocks with their names inside them.

Figure 3.4 shows the amino acid sequence of RccA and where the conserved GGDEF and EAL sequences are which signify their domains. In the case of the GGDEF domain, in RccA it is actually composed of a SGDEF sequence, but it still remains to be seen if this will have an effect on the function of the protein. From this analysis, it can be concluded that RccA is a transmembrane protein, but as of yet the function is unknown. It is however likely to be related to c-di-GMP due to the two sensory domains and the GGDEF and EAL domains that are present.

MCMLRRRFGLTIIRTLPMTLSTDLFGPSAAPAQVLRKHYATEMAVERTRLLYQGSLLPTLLMLVNGLVCAWLLWN
PKQYLLDSIWLVWLLALVAMRVIQVAAFDSAMPSRQAQPVWRRMFMLGSAVSGLTLATAAIALVPADSFMQQAWV
FGLIGAATLSASVAYAVSVPAFLSFALPCLVPSIIYLFWSGDPQQRGWGVLGLILLASLSLVAWQVNRLIQRGLL
RRFQNQALIEHLQQAQQRSEQLNQELVREVEQRRQVEQELREAQVGLQDRVAQRSQELDAASLALNKSEARLAMA
LQASELGLWDWNLQTDEVHHTQLKELFGLEPEYVTAMLGHLKPRLHPEDLPLLKRALVEHLKGRSEDYQVEYRVR
HGDGHWVWIEDRGRAVERAPSGRVTRMLGTRRDISASKALEEQQRLASTVFEAASEGIVILDPDYVLIAVNQAFS
RVTGFETDDMIGRNVVELPSSRDARRHFPVIRQALLSHGTWQGELVETRKNGELYPQWLQLNVVRDVRGKVSHIV
GFFADLSARRESEERMRYLTHYDELTGLANRSLFRERLREAHQRVRQGGRSLALLHINLDRFKLLNDSLGHEVAD
QLLQKMARRLINALPEADTIARLSGDEFAVLFDAYGNLSSLARVATRLLAKLRVPVTVEGHELVVSASMGVSLLP
DNAREISALVSQSNMAMQHAKHLGGNNFQFYTDSLQASTLERLQLENHLRKAIEERQLSVFYQPKLCLATGKLNA
AEALIRWEHPQWGMVPPGDFIGLAEETGLIVPLGEFVLREACWQACEWQRQGLAPIRVSVNLSVHQLRQGKLVSL
VRQVLEETGLDPQYLELELTESQLLDSVEHIIATFQQLRDLGVKLAIDDFGTGYSSLSYLKRIPVDYVKIDQTFI
RGLGQGREDAAITRAIIAMAHGLALKVVAEGVEDQQQLDFLRGERCDEVQGYLISRPMQAEGLADLLRKNADFPE

Figure 3.4. The amino acid sequence of RccA. The five transmembrane domains are shown in orange, while the two individual PAS domains of RccA are highlighted in green. The GGDEF domain is highlighted by red and the conserved sequence making this a GGDEF domain, in this case SGDEF, is shown in red. In contrast, the EAL domain is shown in blue with the conserved EAL sequence also highlighted in blue.

From the DNA sequence analysis, three potential promoter sequences were identified which can be seen in Figure 3.5. All these sequences are predicted and it is not currently clear which sequence, if any, is utilised and what binds to it, therefore experimental work is required to determine this.

5′GCCCGCAACAGGCGGTGCGCCAGGCCGTGGAAGGTGCCTACCCACATGCCGGCCGGGCTGATGCCCATCAACT
GCTCGATGCGGTGGCGCATCTCGGCAGCGGCCTTGTTGGTGAAGGTCACCGACAGGATGGAATGCGGGGACGCG
TTCTCGACCTGGATCAACCAGGCGATACGGTGCACCAGCACTCGGGTTTTACCGGAGCCAGCACCGGCCAGGACC
AACTGACGGCCAACGGGGGCGGCTACGGCCTGGCGTTGGGCATCGTTGAGGGAGTTCAGCAAAAGAGAGAGATC
ATCGCGCATCGGCGCATTCTATGGCCCACGGGGGAGTGGGGCAAATACCAATGCCGGGTGGTCGACGGAAAACA
ATCAGGGTGATGACCGGTTGGTCAATGGCCACAGCCCGCGCAGTGTCTCGCTCAAGCCGTCTGGCCCCGAAGTTT
GCGGCGCATGAGCATGCCGTTTTTATGACGTAGAGCAGTTTGGCCCAAGCGTTTGC 3′

Figure 3.5. The three predicted promoter sequences of *rccA*. The 500bp upstream DNA sequence of *rccA* with the three predicted promoter sequences highlighted in red. The predicted transcription start site is shown by the larger letter in each case.

Using the amino acid sequence, it was shown that RccR is composed of two domains; a DNA binding domain and a sugar binding domain. The composition of RccR can be seen in Figure 3.6. From this analysis, it is predicted that RccR belongs to the RpiR family of transcriptional regulators, where the C terminal sugar binding domain is further characterised as a sugar isomerase (SIS) domain (Bateman, 1999).

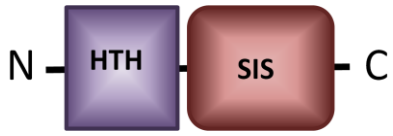

Figure 3.6. The domain organisation of RccR. RccR consists of two different domains; a N terminal helix turn helix DNA binding domain, represented by HTH, and a C terminal sugar binding domain annotated as SIS.

When the DNA sequence of *rccR* was analysed, one single potential promoter sequence was identified as shown in Figure 3.7. From this data, it is unknown if this is the promoter sequence for *rccR*, and if it is what binds to it. There is a high possibility that RccR binds to a type of sugar due to the SIS domain that is present in RccR itself. Once again though, experimental evidence will be required to answer both of these questions.

5′ACTGCGCGACCTGGGCGTGAAGCTGGCCATCGACGATTTCGGCACCGGTTATTCGTCCCTCAGCTACCTCAAG
CGTATCCCCGTGGACTACGTGAAGATCGACCAGACCTTTATTCGTGGACTCGGGCAGGGGCGTGAAGACGCGGCC
ATCACCCGCGCAATTATCGCCATGGCCCATGGGCTGGCACTCAAGGTGGTGGCAGAAGGTGTGGAAGACCAGCAG
CAGCTGGATTTCCTGCGCGGTGAGCGGTGTGACGAAGTGCAGGGCTATTTGATCAGCCGGCCAATGCAGGCCGAA
GGGTTGGCAGATTTGTTACGGAAAAATGCAGATTTTC**CTGAGTGATGCCAAGGTCGCCCCTGTGGCTACAACTTG**
GCT**GCCCTTAAA**TTCAAAGGCGGGCAGGGCGATTTAGTGACGCGTCGGACGGGCAAAACGGCAATTCTTGTAGT
ATAACTACAAGCTTGCTACATCCTTGGCCCTGCCCATAACAAGAGTCCTGTCCT 3′

Figure 3.7. The predicted promoter sequence of RccR. The 500bp upstream DNA sequence of *rccR* with the predicted promoter sequence highlighted in red. The transcription start site is shown by the larger letter.

3.3.2. The creation of non-polar *rccR* and *rccA* mutants in *Pseudomonas fluorescens* SBW25

In order to start analysing *rccR* and *rccA*, a non-polar mutant of each gene was created in *P. fluorescens* SBW25, in addition to a double mutant, *rccAR*. It was decided to create non-polar mutants in order that the downstream genes were not affected by the deletion of the previous gene. The two step allelic exchange method was utilised as this creates a mutant which is identical to the wild type SBW25 strain only without the deleted gene. By using the suicide vector, in this case pME3087, the first crossover can take place by selecting for tetracycline resistant bacteria. To ensure that the suicide vector has been removed and the second crossover has taken place, tetracycline sensitive cells are screened for. Full details of the two step allelic exchange method are described in Section 2.4.9.

Using Phusion polymerase and the primers *rccR*upF and *rccR*upR a 500bp fragment of the upstream *rccR* DNA was amplified. The downstream 500bp of *rccR* were amplified using Phusion polymerase and the primers *rccR*dnF and *rccR*dnR. In order to amplify the 500bp of upstream genomic SBW25 DNA of *rccA* primers *rccA*upF and *rccA*upR were used. The 500bp downstream DNA of *rccA* was amplified using primers *rccA*dnF and *rccA*dnR. Finally, the DNA fragments required for the *rccAR* mutant were created using the primers *rccR*upF and *rccR*upR for the upstream DNA and primers *rccA*dnF and *rccA*dnR for the downstream DNA. The sequences for the primers used to create Δ*rccR*, Δ*rccA* and Δ*rccAR* are listed in Table 2.4. The final chromosomal structures of the *rccA*, *rccR* and *rccAR* mutants are shown in Figure 3.8b), Figure 3.8c) and Figure 3.8d) respectively.

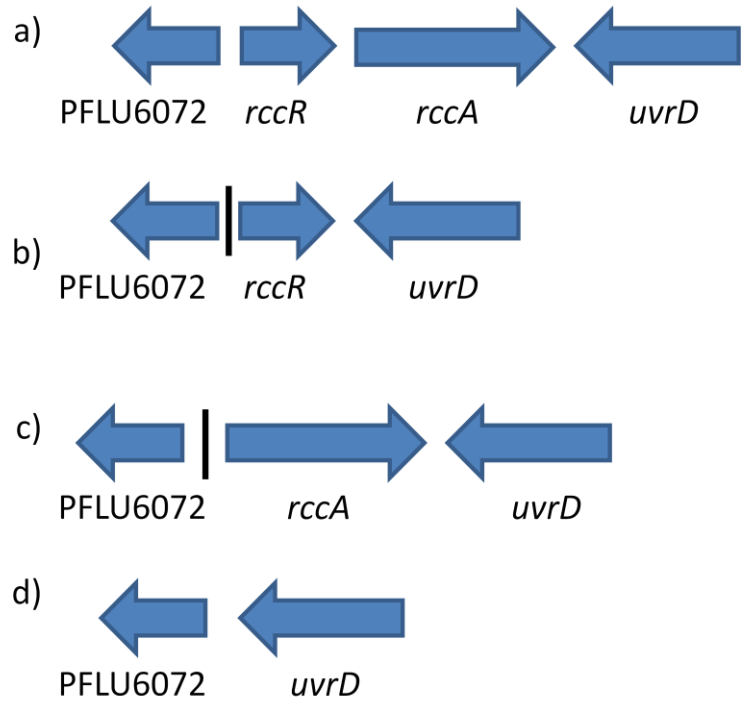

Figure 3.8. The chromosomal structure of the *rccA*, *rccR* and *rccAR* mutants. The chromosomal structure of a) wild type, b) Δ*rccA*, c) Δ*rccR* and d) Δ*rccAR*. The adjacent genes to *rccR* and *rccA* are also shown and the orientation of each gene is shown by the direction of the blue arrow. The black lines indicate where genes have been deleted.

3.3.3. Δ*rccR*, Δ*rccA* and Δ*rccAR* have wheat root colonisation deficiencies

To identify if Δ*rccR*, Δ*rccA* or Δ*rccAR* have any effect on the colonisation of wheat seedlings, a root colonisation assay was set up. This assay utilises a strain of SBW25-*lacZ* bacteria which competes with the mutant strain of bacteria to colonise the wheat seedlings. When samples of *P. fluorescens* are taken from the root environment, the SBW25-*lacZ* bacteria are identified due to blue selection on IPTG and X-GAL agar plates. On the other hand, the mutant strain of bacteria being tested does not have the *lacZ* gene present and therefore present as white colonies on the same agar plates. In order to identify if there was a difference between the wild type and the mutant strain being tested, the ratio of the two was calculated in each case.

To confirm that the SBW25-*lacZ* strain being utilised in this experiment does not have an effect on the colonisation of the wheat seedlings, a control experiment using the SBW25 wild

type strain instead of a mutant was done by Dr Richard Little. The results of this can be seen in Figure 3.9 and shows that the SBW25-*lacZ* strain does not affect the ability of the bacteria to colonise the roots of the wheat seedlings. As this is the case, it can be concluded that this strain acts like the wild type SBW25 and can be safely used in place of the SBW25 strain. When the Δ*rccR*, Δ*rccA* and Δ*rccAR* strains were tested to see how their root colonisation efficiencies compared to the SBW25 wild type, it could be seen that all three strains were less efficient than wild type. This result is shown in Figure 3.9. Even though Δ*rccR*, Δ*rccA* and Δ*rccAR* all have root colonisation deficiencies, the reasons for this are not yet clear and further investigations are required.

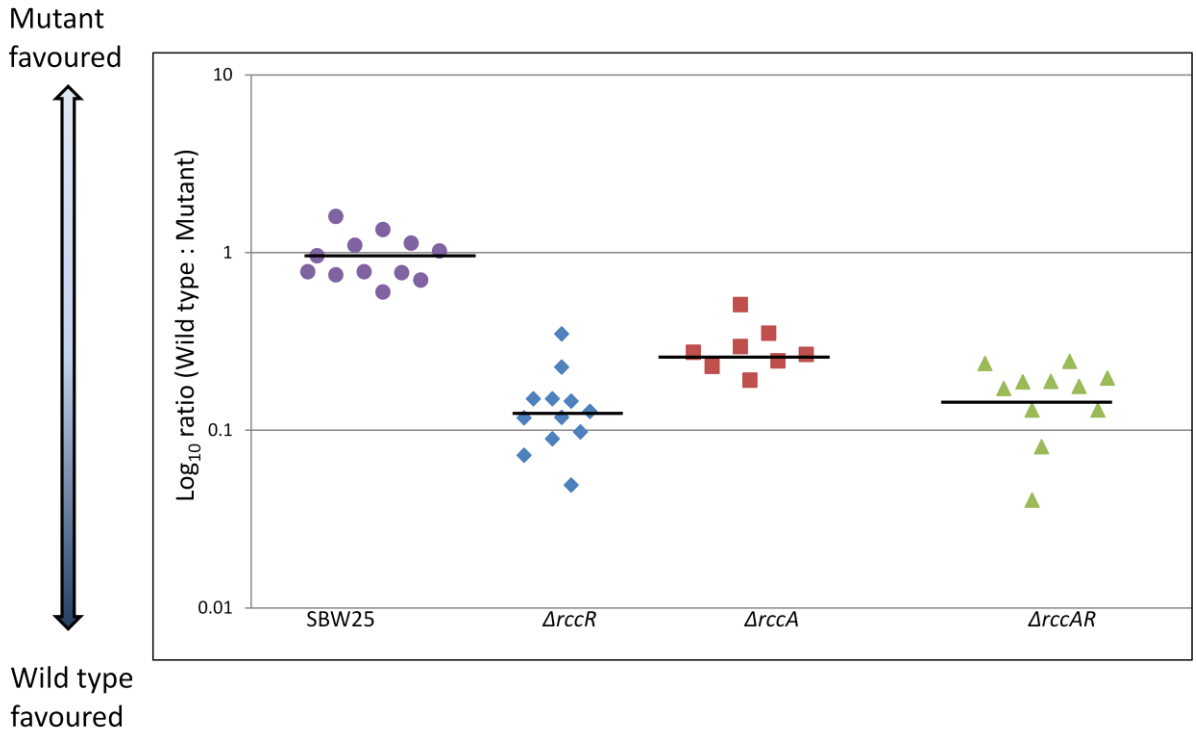

Figure 3.9. Δ*rccR*, Δ*rccA* and Δ*rccAR* all have wheat root colonisation deficiencies compared to wild type. Each dot represents an individual wheat seedling and the median root colonisation efficiency by the strain is shown by the solid black line. In each strain, there is a statistically significant difference in the colonisation efficiency compared to the wild type SBW25 when using the Mann Whitney U test and a p value < 0.05.

3.3.4. Δ*rccR* and Δ*rccAR* do not have growth defects, but Δ*rccA* does when grown in pyruvate

To confirm that the root colonisation deficiencies seen by the Δ*rccR*, Δ*rccA* and Δ*rccAR* were all due to root colonisation, growth curves of the mutants were done to ensure that there were no growth defects present. The growth curves were done in a defined minimal media to better understand any differences in phenotypes that may occur. As well as this, two different carbon sources were tested; pyruvate and glucose. The results of the growth curve, as shown in Figure 3.10a, identify that Δ*rccR*, Δ*rccA* and Δ*rccAR* are able to grow in minimal media supplemented with glucose just like the SBW25 wild type strain is. This is also the case for Δ*rccR* and Δ*rccAR* when grown in the addition of pyruvate as a carbon source, however, Δ*rccA* appears to have a small growth defect compared to SBW25, which is shown in Figure 3.10b. These results indicate that Δ*rccR* and Δ*rccAR* do not have a growth defect when grown in minimal media with the addition of either pyruvate or glucose. On the other hand, Δ*rccA* only has a small growth deficiency when grown in minimal media supplemented with pyruvate.

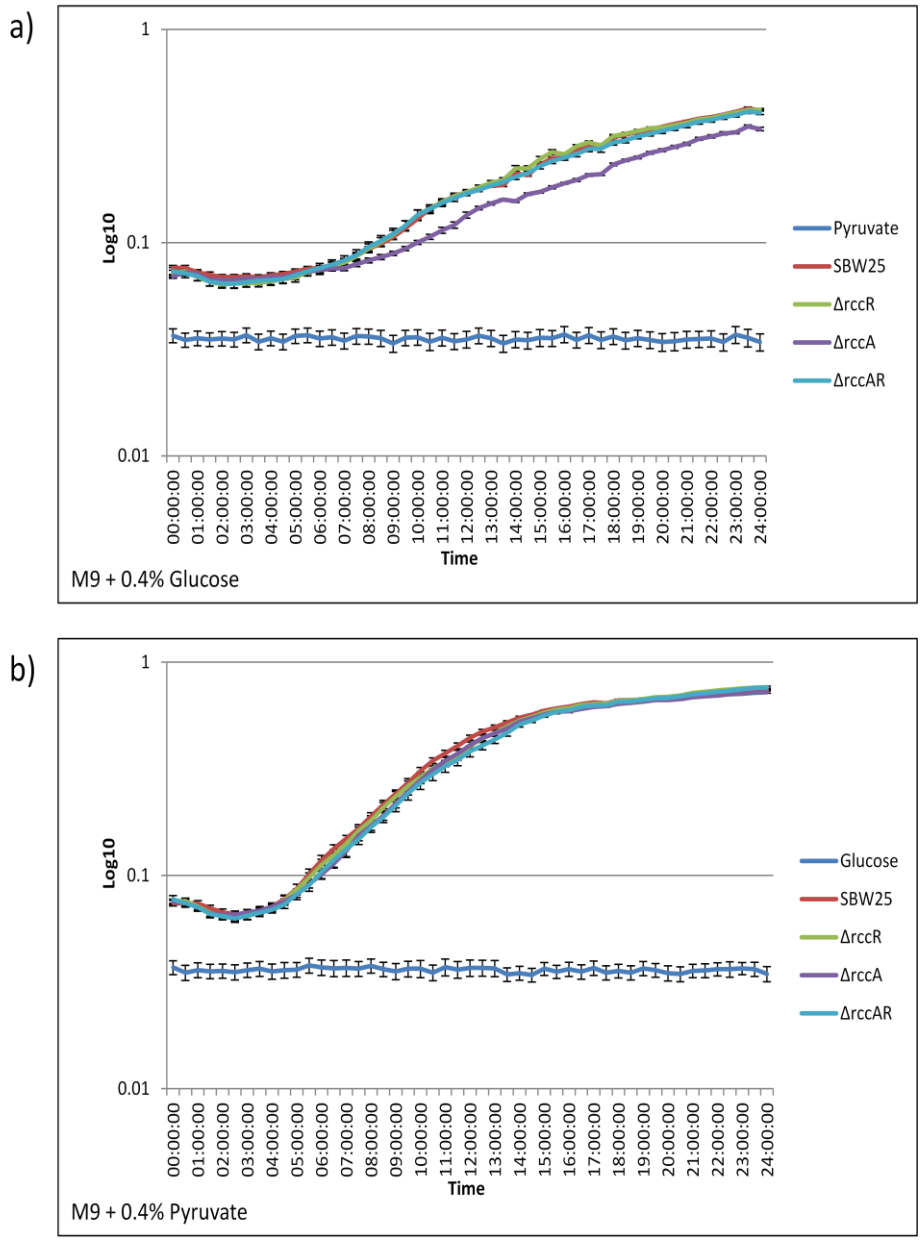

Figure 3.10. Growth curve of SBW25, Δ*rccR*, Δ*rccA* and Δ*rccAR* when grown in minimal media supplemented with either glucose or pyruvate. a) Strains grown in addition to glucose. b) Strains grown with pyruvate as the carbon source. When grown in glucose, all the strains grew at approximately the same rate. When grown in pyruvate SWB25, Δ*rccR* and Δ*rccAR* all have approximately identical growth rates, but Δ*rccA* has a slower rate of growth. The standard deviation of the six samples used are shown for each time point.

3.3.5. *rccR* and *rccA* are functionally linked

As there are a number of phenotypes that enable bacteria to effectively colonise the roots (Turnbull et al. 2001), one of these was studied further to see if this would have an effect on *rccR* and *rccA*. The motility of bacteria is not only important to colonise roots, but the levels of c-di-GMP in the cell affect the motility of the bacteria (Jones et al. 2015); (Römling et al. 2013); (Hengge, 2009). Due to there being a variety of motility assays, the swarming motility assay was utilised due to this assay using semi solid agar and this being the closest replicate to the soil environment; the swimming motility assay is done using a liquid media whilst twitching motility uses solid agar and looks at the pili of bacteria instead of the flagella.

The first type of swarming motility assay done with the mutant strains was using 0.3% KB agar supplemented with 1% NaCl and Congo Red dye. After overnight incubation at room temperature, the *rccA* mutant has a unique irregular morphology compared to the wild type SBW25 strain and the *rccR* mutant, as seen in Figure 3.11. However, when a *rccAR* mutant is tested in this swarming motility assay the wild type phenotype is restored again, which can also be seen in Figure 3.11. This data suggests that the swarming motility is dependent on *rccA*.

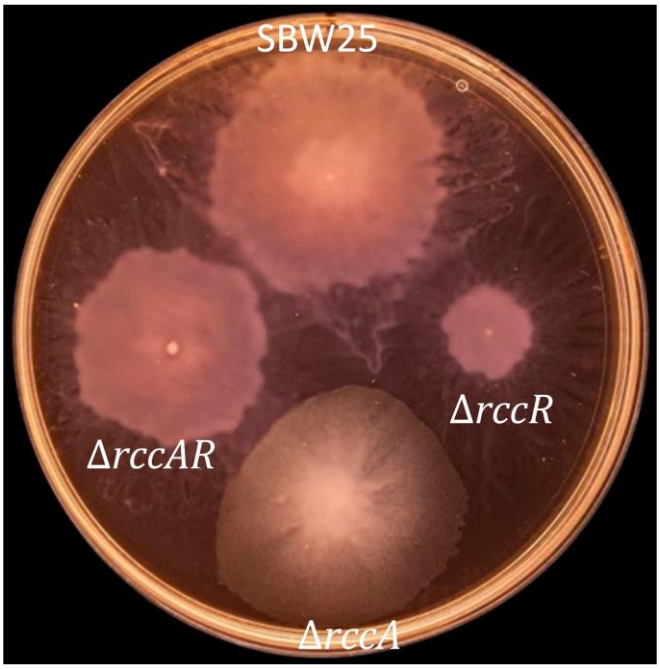

Figure 3.11. The swarming motility of SBW25, Δ*rccR*, Δ*rccA* and Δ*rccAR* on KB agar with NaCl and Congo Red Dye. The swarming morphology of Δ*rccA* is unique and irregular compared to SBW25 and all of the other strains tested.

Before the second type of motility assay could be done, the *rccR* gene was required to be inserted into the pME6032 over-expression plasmid. To begin with, *rccR* was amplified from SBW25 chromosomal DNA using the primers *rccR*OEup and *rccR*OEdn and Phusion polymerase in a PCR. The sequences of the primers can be found in Table 2.4. This amplified fragment was inserted into the pME6032 plasmid using the restriction enzyme sites *Eco*RI and *Kpn*I. Following the sequencing of the pME6032-*rccR* construct, it was now possible to test the swarming motility of *rccR* when over-expressed using the pME6032 plasmid in SBW25 wild type and the Δ*rccAR* background strains. In order to replicate the conditions as previously used, the agar utilised in this swarming motility was M9 minimal media supplemented with pyruvate. As no distinct phenotype was present on this media using these strains, the size of the swarms were measured after 16 hours and 20 hours and compared to the wild type control as shown in Figure 3.12.

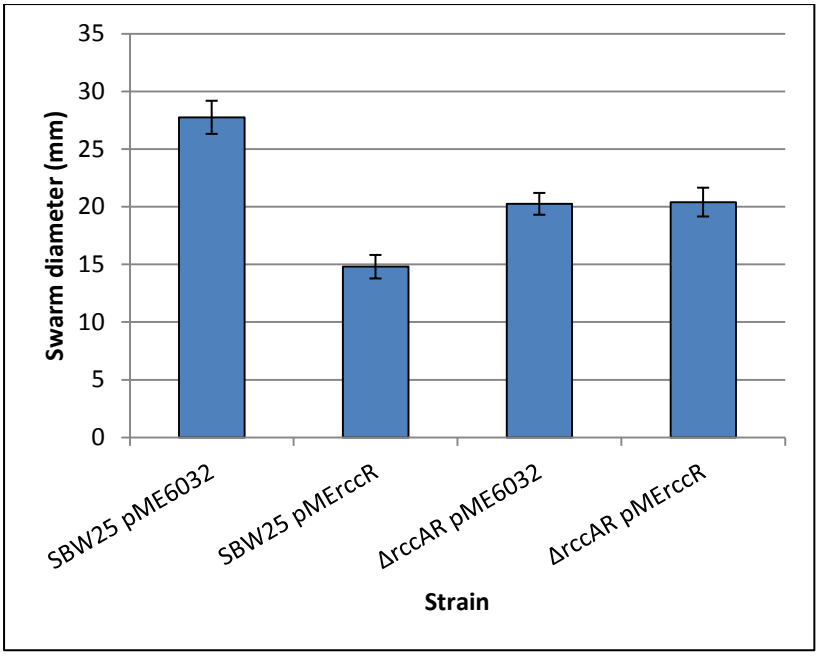

Figure 3.12. The swarming motility of an over-expressed *rccR* when in the SBW25 or the Δ*rccAR* background. When *rccR* is over-expressed in SBW25, there is a decrease in the diameter of the swarm seen compared to the SBW25 control. However, when this is done in the Δ*rccAR* background, the swarm diameters of the Δ*rccAR* control and the over-expression *rccR* are approximately the same. The student paired sample t-test was used to compare the over-expression of *rccR* to control in the relevant background at a p value of 0.05. Only in the SBW25 background were the results statistically significant.

When Δ*rccR* is over-expressed in the SBW25 wild type background, there is a reduced swarm diameter compared to the wild type control. However, when this is repeated in the Δ*rccAR* background, there is very little difference seen, if any, in the swarm diameters between the Δ*rccAR* control and the Δ*rccR* over-expression strain. This result again suggests that the effects of RccR and RccA on swarming motility are linked; in this case the *rccR* swarming motility phenotype is dependent on *rccA*.

The final swarming motility assay done was using the mutant strains on minimal media which was supplemented with different carbon sources. The minimal media M9 was utilised to replicate the root environment as much as possible, while different carbon sources were used to identify if it would make a difference. The carbon sources of choice were pyruvate, glycerol and glucose. The size of each swarm was measured after 16 hours and 20 hours and compared to the wild type control. The results are shown in Figure 3.13.

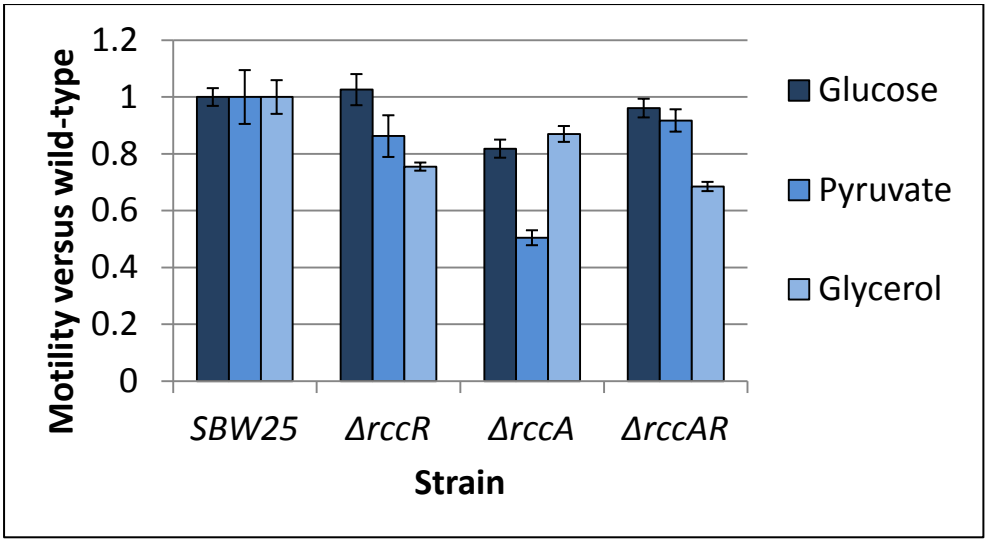

Figure 3.13. The swarming motility of SBW25, Δ*rccR*, Δ*rccA* and Δ*rccAR* when grown on different carbon sources. When grown in these carbon sources, Δ*rccR* and Δ*rccAR* have the largest to smallest swarm sizes when grown in glucose, pyruvate and glycerol respectively. Δ*rccA* appears to be different and has the largest swarm size when grown in glycerol followed by glucose and a very small swarm size when grown in pyruvate compared to Δ*rccR* and Δ*rccAR*. The p value < 0.05 using the student paired sample t-test, where each strain is compared to the wild type SBW25 in the same carbon source. The data for Δ*rccR* is only statistically significant in the presence of glycerol. All of the data for Δ*rccA* is statistically significant. Finally, when grown in either glucose or pyruvate, the data for Δ*rccAR* were both statistically significant.

When grown in the presence of pyruvate, the *rccR* and the *rccAR* mutants both have a reduced swarming motility compared to wild type; however, the swarming motility of the *rccA* mutant is reduced by half. Whilst in the presence of glycerol, compared to the wild type strain the swarming motility of all the mutants are reduced once again, this time however by approximately 20%, 15% and 30% respectively. Finally, when glucose was used as the carbon source, it was only the *rccA* mutant which had a reduced swarming motility by approximately 20%. These results suggest that *rccR* and *rccA* respond differently to carbon sources when measured by swarming motility.

3.3.6. The carbon source does not affect the mRNA levels of *rccR* or *rccA*

Finally, to understand more about the relationship between *rccR* and *rccA*, the transcription of the two genes was studied using reverse transcriptase PCR, (RT PCR). In addition to *rccR* and *rccA*, a third gene, the housekeeping sigma factor *rpoD*, (Shimada et al. 2014) was used as a control. The primers used in the RT PCR are listed in Table 2.4. As all the previous assays were done using minimal media, it was also used in this experiment. Due to the differences seen when different carbon sources were used in the swarming motility assays, it was again decided to see if these would make a difference when studying the transcripts.

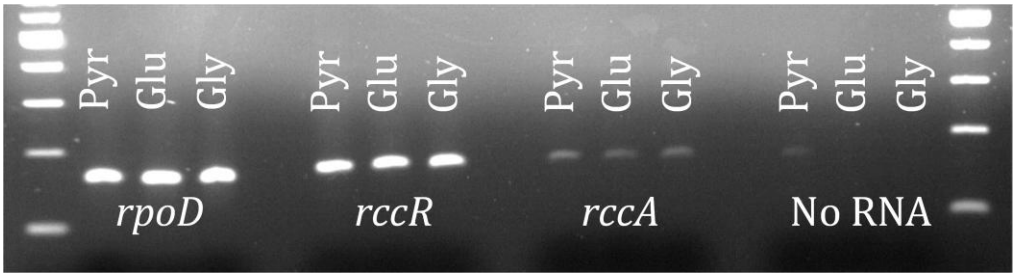

Figure 3.14. The reverse transcriptase PCR of *rccR* and *rccA* when grown in minimal media with either pyruvate, glucose or glycerol as a carbon source. *rpoD* is known to be a constitutive promoter and was used as a control gene, and no RNA in the negative controls. Both *rccR* and *rccA* appear to have approximately the same level of transcript in each of the media, however, the levels seen in *rccA* appear to be at a lower concentration than those of *rccR*. The following abbreviations have been used; Pyr for pyruvate; Glu for glucose and Gly for glycerol.

The results of the RT PCR are shown in Figure 3.14 and show that the control gene *rpoD*, a RNA polymerase σ^{70} has identical transcript levels when supplemented with either pyruvate, glycerol or glucose. This result was also seen for the transcripts of *rccR* and *rccA*, however, the overall levels of the *rccA* transcript appears to be less intense in all media (Figure 3.14). The decrease in the intensity of the *rccA* transcripts compared to *rccR*, as seen in Figure 3.14, could mean that there is a lower concentration of *rccA* present compared to *rccR*, although this will need to be verified. These results indicate that *rccR* and *rccA* do not appear to interact at the transcriptional level. In addition to this, the expression of *rccR* and *rccA* does not appear to be affected by the carbon source that is present in the media.

3.4. Conclusions

From the work done so far, it has been shown that *rccR* and *rccA* both play an important role in the colonisation of the *P. fluorescens* SBW25 wheat rhizosphere. This is in agreement with the results that *rccA* is up-regulated in the *P. fluorescens* SBW25 sugar beet rhizosphere (Silby et al. 2009). As these studies have been done using different crops, wheat in this study and sugar beet in the Silby et al. study, these results suggest that *rccA* is important in the general colonisation of the rhizospheres of different plants.

The data shown in this chapter suggest that the functions *rccR* and *rccA* are linked in some way and that the swarming phenotype of an *rccA* is dependent on *rccR*. It can be speculated that c-di-GMP is involved in this link, however from the current data this can not be proven. At the moment, RccA is a putative c-di-GMP protein and this could potentially mean that c-di-GMP binds to RccR. For example, Hickman and Harwood have previously demonstrated that the transcription factor FleQ binds to c-di-GMP in addition to activating the expression of the biogenesis of flagella genes as it was previously known to do (Hickman and Harwood, 2008). Although RccR and FleQ belong to different families of transcription factors, it is still possible that c-di-GMP is able to bind to both of these. One of the current problems in the c-di-GMP field has been the inability to predict what the second messenger binds to, as there is no consensus binding sequence and a number of different proteins have been discovered.

Finally, the swarming assays identified that *rccA* and *rccR* have different effects when grown on different carbon sources. It is a possibility that one, or even both, of the PAS domains present in *rccA* is stimulated by pyruvate, or another carbon source, as previous research has shown that PAS domains respond to a number of different environmental signals (Vogt and Schippers, 2015); (Becker et al. 2011). As well as this, RccR is known to be a member of the RpiR transcriptional regulators, which have a sugar binding domain (Bateman, 1999). Therefore, it is possible that the sugar binding domain of RccR is either responding to the differing carbon sources or alternatively breaking down the products from the carbon sources that were utilised in the swarming motility assays (Daddaoua et al. 2009).

Even though the carbon sources had different effects on *rccA* and *rccR* and their ability to swarm, the level of mRNA in *rccA* and *rccR* does not appear to vary when in each of the different carbon sources. Although the earlier data in this chapter indicates that there is a link between *rccA* and *rccR*, this link does not appear to be at the transcriptional level. It has also

been shown that *rccR* is auto-repressed in different carbon sources, due to the fact that there are two different binding sites present in its promoter region (Campilongo et al. 2017). Despite this being the case, it is still possible that RccA and RccR interact after translation has taken place, or even allosterically (Christen et al. 2006). In the subsequent chapters, this intriguing connection between RccR and RccA will be investigated in more detail.

Chapter 4 – The Characterisation of RccR

4.1. Introduction

In this chapter, the RccR protein will be characterised further. It was decided to focus on RccR for two main reasons following the analysis that was done in Chapter 3. Firstly, the deletion of *rccR* had a strong effect on the ability of *Pseudomonas fluorescens* to colonise the roots of wheat seedlings. Secondly, all of the swarming motility assays that were utilised indicated that *rccR* could be involved in the suppression of the *rccA* phenotype. Although RccR is part of the previously studied RpiR family of transcriptional regulators, and is similar to the Enter Doudoroff (ED) pathway regulator HexR (Daddaoua et al. 2009), it has never been studied before and will therefore be further characterised in this chapter. For further information on the comparison of RccR and HexR from *P. fluorescens* SBW25, including an amino acid sequence alignment, see Chapter 5.

4.2. Aims

During this chapter, it is proposed to:

- over-express and purify RccR in *Escherichia coli* using a His_6 tag at the C terminus.
- use the purified RccR protein to produce a polyclonal antibody.
- identify if RccR auto-regulates its own transcription.
- identify and confirm the genes that RccR binds to within the *P. fluorescens* SBW25 genome.

4.5. Results

4.3.1 Over-expression and purification of RccR with a C terminus His_6 tag in *Escherichia coli*

To understand more about the function of RccR, it was required to over-express and purify the protein for biochemical analysis. In order to do this, the over-expression vector pET42b

was utilised, in which a His_6 tag was added to the C terminus of *rccR* enabling the protein to be purified from *Escherichia coli*. Firstly, primers *rccR*HisUp and *rccR*HisDn were utilised in a PCR with Phusion polymerase and genomic DNA to amplify *rccR*. During the PCR, the His_6 tag was incorporated to the end of the gene using primer *rccR*HisDn. Table 2.4 shows the full sequence of the primers that were utilised in the PCR. Figure 4.1 shows the cloning gel of the pET42b-*rccR* construct following the restriction digest of four individual clones.

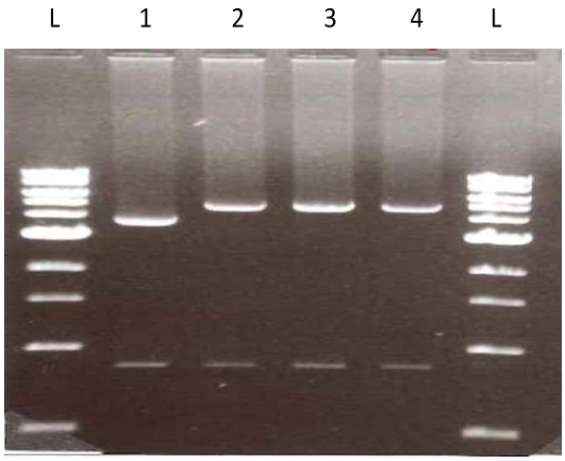

Figure 4.1. The cloning gel of the pET42b-*rccR* construct. Following a restriction digest using the enzymes *Nde*I and *Xho*I, the pET42b vector and the *rccR* fragment can be seen at approximately 5kb and 900bp respectively. Where L stands for ladder and 1 – 4 represent different samples. Sample 1 was rejected due to the abnormal vector size.

Following the vector construction, in order to confirm that *rccR* was able to be over-expressed, a small-scale induction of RccR was attempted using *E. coli* BL21 (DE3) pLysS cells. Following the induction of RccR, a SDS-PAGE analysis was utilised to see if there was a difference in the induced and non-induced samples. Figure 4.2 shows that RccR can be induced on a small-scale therefore the next step was to scale up the expression of RccR and attempt to purify RccR. During purification, fractions of the RccR protein eluted from the column were collected in order to test which ones contained the purified RccR protein. A small sample from each fraction was then used in an SDS-PAGE in order to identify which fraction contained only RccR, and the results of this can be seen in Figure 4.3.

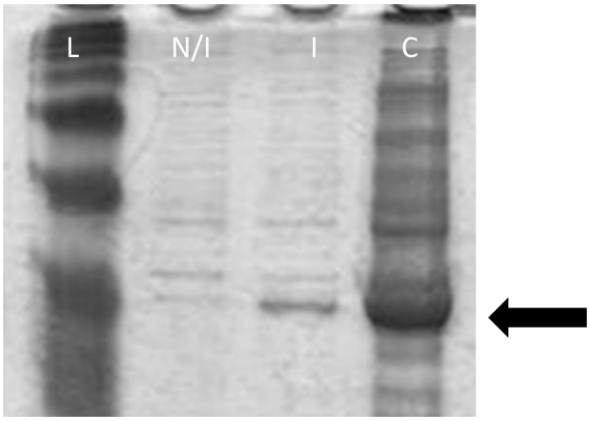

Figure 4.2. The small scale induction of RccR. There is a difference in the induced and non-induced samples, therefore when induced, RccR is over-expressed. Where L stands for ladder, N/I for non-induced sample, I for induced sample, N/I for non-induced sample and C concentrated sample. The black arrow indicates the position of the RccR-His_6 band.

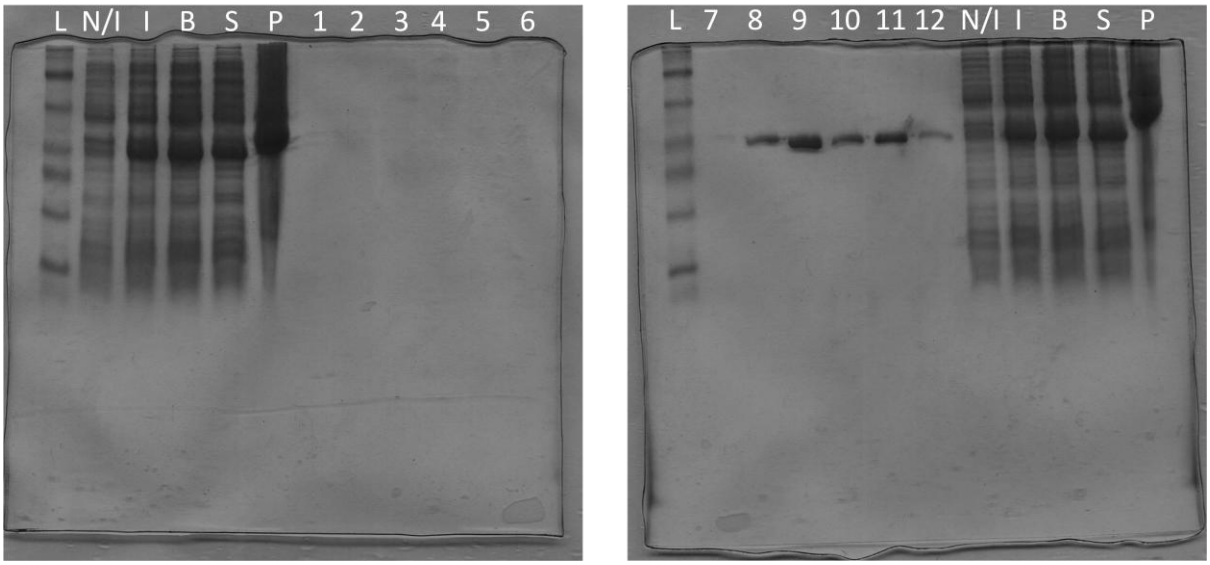

Figure 4.3. The SDS-PAGE identifying the purified RccR protein. Two SDS-PAGE gels were required for the number of fractions obtained, however, only fractions 8 – 12 contain RccR. The letters L, N/I, I, B, S and P stand for ladder, non induced, induced, broken cells, soluble protein and pellet respectively. While the numbers represent the fraction number taken from the column.

Following the purification of RccR, and in order to enable future experiments, a polyclonal antibody for RccR was raised in rabbits, using the company Dundee Life Sciences. In preparation for the production of the polyclonal antibody, 500ml of a 4.14mgml^{-1} protein solution of RccR was sent to Dundee Life Sciences. To ensure that the polyclonal antibody

received from Dundee Life Sciences was immunogenic for RccR, Western blot analysis was utilised with this RccR antibody as the primary antibody. Figure 4.4 shows the final Western blot where the strains with the RccR protein present have a band present. This indicates that the antibody has a high affinity to detect RccR *in vivo*. It can also be seen in Figure 4.4 that a second band is present, which indicates that this antibody is not specific to RccR. In lane 5, the Δ*rccAR* strain is present, however a band can be seen, which is unexpected. However, this result could be explained by the similarity of RccR and HexR and how both of these proteins interact with RccA; see Chapter 3 and Chapter 5 for more information respectively.

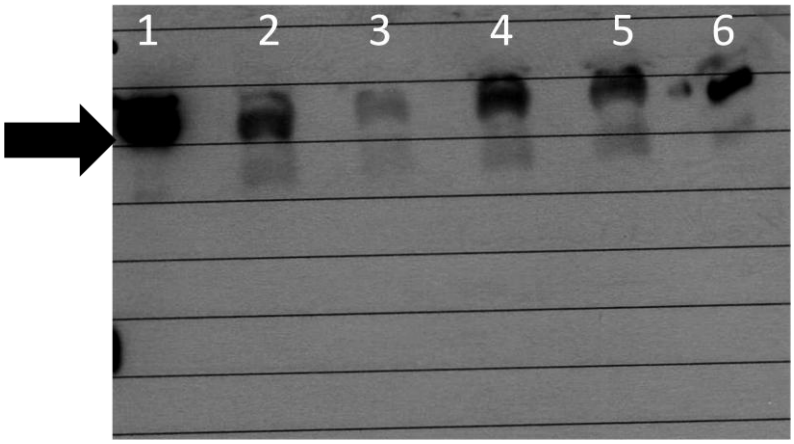

Figure 4.4. Western blot confirmation of the RccR polyclonal antibody. The strains used for the Western blot analysis are as follows and are shown in lanes 1 – 6 respectively. 1) 1 in 100 dilution of purified RccR protein; 2) SBW25; 3) Δ*rccR*; 4) Δ*rccA*; 5) Δ*rccAR* and 6) pME6032-*rccR* over-expression. The black arrow indicates where the RccR protein is, which is not present in the Δ*rccR* strain.

4.3.2. RccR binds to the predicted promoter sequence upstream of *rccR*

In Section 3.3.1, it was shown using bioinformatic analysis that *rccR* is predicted to have a single promoter. In order to identify if RccR does bind to this predicted promoter region, electrophoretic mobility shift assays (EMSAs) were utilised where ethidium bromide was used to stain the DNA in question. The primers *rccR*EMSAup and *rccR*EMSAdn, the full sequence of which can be found in Table 2.4, were utilised in a Phusion PCR to amplify the predicted promoter region of *rccR* of approximately 200bp from the SBW25 chromosome. In the EMSA, the 100bp ladder was used as a DNA marker and a negative control of bovine

serum albumin (BSA) was used as BSA is known not to bind to DNA. In each reaction, an increasing amount of the purified RccR protein was incubated with the same amount of the predicted *rccR* promoter DNA. The results of the EMSA can be seen in Figure 4.5 and show that RccR does bind to the predicted *rccR* promoter sequence. It can therefore be concluded that RccR is potentially able to auto-regulate. It can also be argued that where there is a RccR-DNA complex present, a second RccR-DNA complex can be seen too (Figure 4.5). This second RccR-DNA complex is more noticeable in the second and third lanes where the RccR-DNA complexes are present. This raises the possibility that RccR binds to DNA as a dimer when at the correct concentrations, which could be plausible as HexR, another member of the RpiR family is also thought to bind as dimer (Daddaoua et al. 2009).

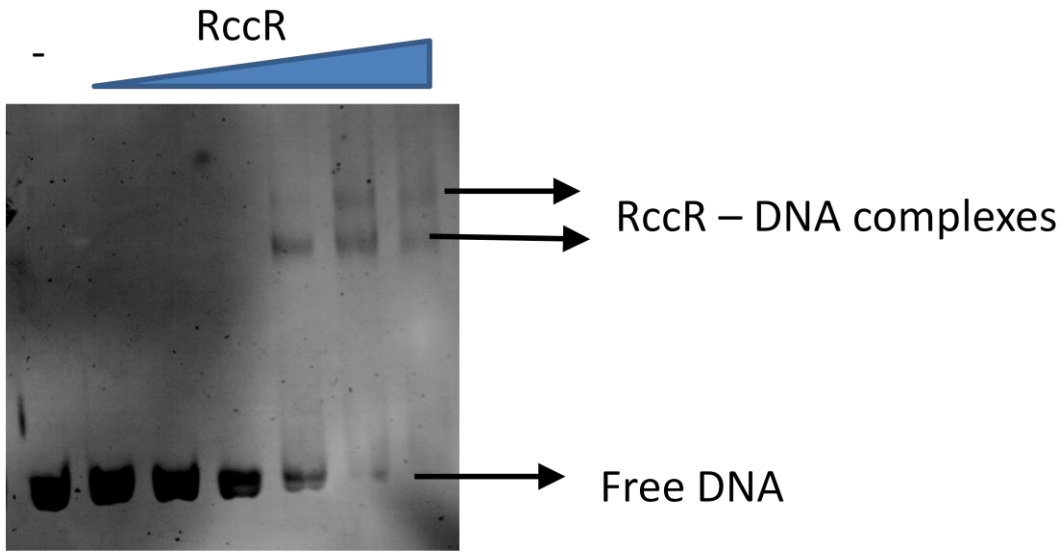

Figure 4.5. EMSA confirming that RccR binds to the promoter region of *rccR* DNA. The concentration of *rccR* promoter DNA is 30ngµl^{-1}, while the concentration of RccR ranges from 0.1mM to 7.5mM. As the concentration of RccR increases a RccR-DNA complex is formed, with a second RccR-DNA complex seen. The – indicates the negative control of BSA.

4.3.3. RccR negatively regulates its own expression

After the EMSA confirmation that RccR binds to its own predicted promoter region, there is still the outstanding question of whether RccR positively or negatively regulates its own expression. In order to answer this, *lacZ* transcriptional reporter fusions were made to *rccR*.

The SBW25 *rccR* promoter region sequence of 465bp was amplified using the primers *rccR*BgalUp and *rccR*BgalDn and Phusion polymerase. The full sequence of the primers can be found in Table 2.4. This amplified product was cloned into the multiple cloning site of the pGm6032 plasmid using the Bam*HI* restriction enzyme site. Figure 4.6 shows a cloning gel of pGm6032-p*rccR-lacZ* following a Taq polymerase screen to identify if the *rccR* promoter sequence was inserted into the pGM6032 plasmid.

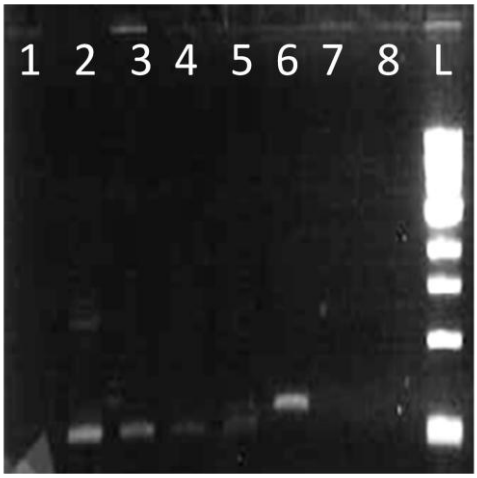

Figure 4.6. The cloning gel of pGm6032-p*rccR-lacZ*. A cloning gel showing the results of a Taq polymerase screen to identify if the *rccR* promoter sequence was inserted into the pGM6032-*lacZ* plasmid. A product of approximately 500bp was expected as seen in lanes 2 – 5. Different samples are shown in each lane and the ladder is in the lane with the L.

To confirm that the *rccR* promoter sequence was in the pGm6032 plasmid in the correct orientation, the construct was sequenced before experiments proceeded. Following this, β galactosidase assays were utilised in the SBW25 wild type background, as well as the Δ*rccR* background, to establish if RccR is a positive or negative auto-regulator. In both cases, the empty vector was used as a control. As the swarming motility assays in Section 3.3.5, identified that different carbon sources resulted in differing effects, minimal media was utilised in this assay with two different carbon sources; pyruvate and glucose. Figure 4.7 shows the results of the β galactosidase assay when the *rccR* promoter was in the SBW25 or the Δ*rccR* background and either pyruvate or glucose was available as a carbon source. Although in each case there is a small amount of β galactosidase activity from the controls, it is extremely low compared to the samples. Importantly, no difference can be seen in the β galactosidase activity when grown in pyruvate or glucose. Therefore, if carbon sources do play a part in the regulation of *rccR*, it is unlikely to be at the level of transcription. Even

though the carbon source does not matter, the different background the *rccR* promoter is placed in does, and this is also summarised in Figure 4.7. It can be seen in Figure 4.7 that when in the Δ*rccR* background, the pGm-*rccR* has an increase in β galactosidase activity compared to when in the SBW25 wild type background. This result indicates that RccR is repressing its own transcription, and a simple model as shown in Figure 4.8 can be made.

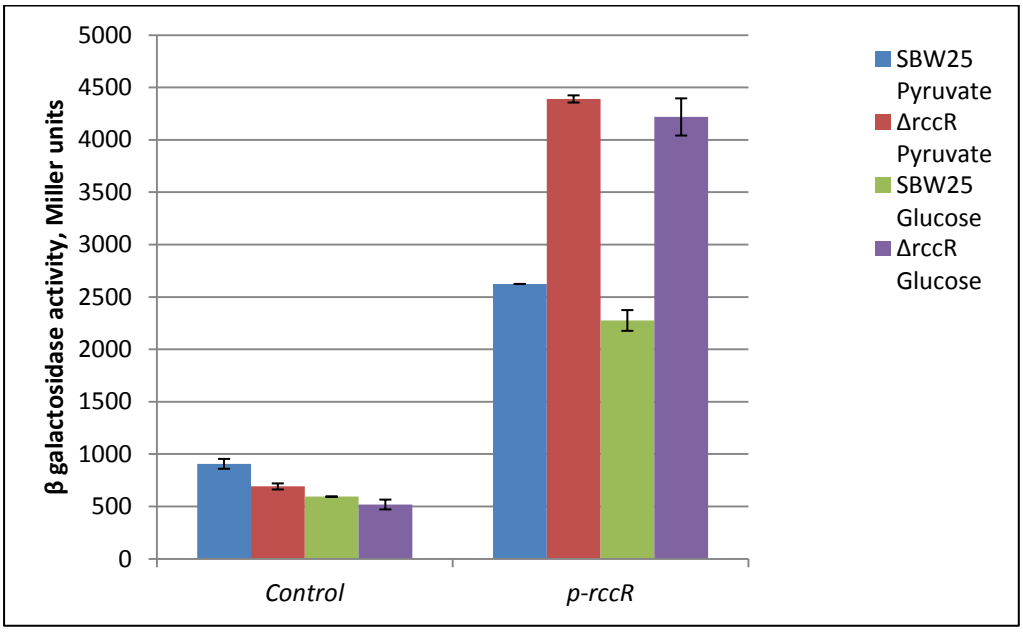

Figure 4.7. The β galactosidase assay when grown in pyruvate or glucose. The empty pGm vector was used as the control sample. The error bars show the standard deviation of the samples. In each case, the empty vector was compared to the vector with the *rccR* promoter region inserted into it in the same background and carbon source, and each result is statistically significant. (p value < 0.05 using the student paired sample t-test).

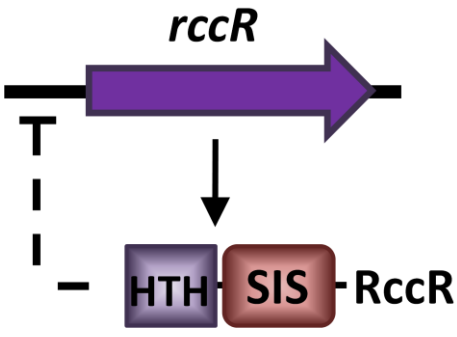

Figure 4.8. A simple repression model of how *rccR* could be working. Following the transcription of *rccR*, the translation of RccR will inhibit any further *rccR* transcription.

4.3.4. The mRNA levels of *rccR* peak during log phase

In order to determine when the transcription of *rccR* occurs in the cell cycle, cells were grown in either minimal media supplemented with pyruvate or glycerol. Samples were taken at timepoints during the lag, log and stationary phases of the cell cycle, and a real time PCR was done using the primers *rccR*EMSAup and *rccR*EMSAdn. As a control, the *rpoD* gene, a σ^{70} factor, was tested using the primers *rpoD*RTPCRup and *rpoD*RTPCRdn. The sequences of all the primers can be found in Table 2.4.

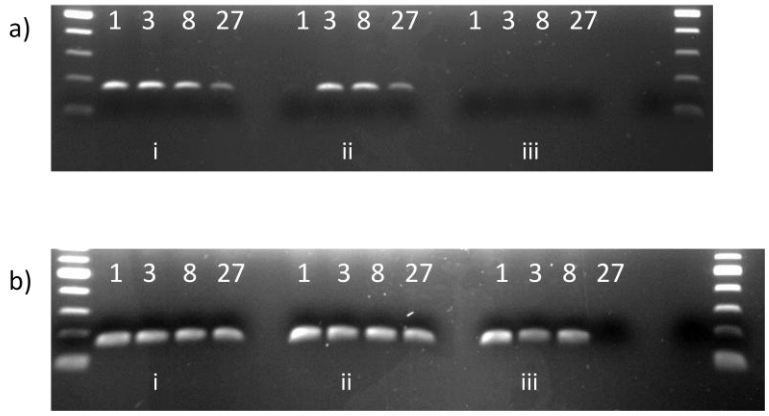

Figure 4.9. The real time PCR of *rccR* and *rpoD* when grown in minimal media with either pyruvate or glycerol. a) The *rccR* real time PCR and b) the *rpoD* real time PCR. The numbers represent the time points in hours when the samples were taken. i, ii and iii indicate that the samples were SBW25 grown in minimal media with pyruvate, SBW25 grown in minimal media with glycerol and finally Δ*rccR* and grown in minimal media with pyruvate respectively.

The results of both real time PCRs can be seen in Figure 4.9. When *rccR* is grown in minimal media with pyruvate, it is expressed until and including the log phase of growth, but during stationary phase appears to decrease. However, if grown in minimal media with the addition of glycerol, *rccR* does not appear to be expressed at all in the lag phase of growth. Expression of *rccR* peaks once again during log phase before declining in the stationary phase. These results indicate that *rccR* expression peaks during log phase of growth when grown in minimal media with either pyruvate or glycerol as the additional carbon source. Although this may be the case, it can not be concluded without further experiments to quantify the amount of RNA transcribed by *rccR* during each of these phases of growth.

4.3.5. RccR binds to the promoter regions of other metabolic genes

In order to identify any other DNA regions of the SBW25 chromosome that RccR binds to, a chromatin immunoprecipitation sequencing (ChIP seq) experiment was done using the polyclonal antibody of RccR. Wild type SBW25 and Δ*rccR* cells were grown in minimal media, to replicate the conditions of the other experiments done. From the results of the real time PCR in Section 4.3.4, SBW25 and Δ*rccR* cells were grown to mid-log phase of growth for cells grown in minimal media with glycerol and minimal media with pyruvate respectively. Following the immunoprecipitation with protein-A sepharose, and to identify if the ChIP seq was successful, a limited PCR was done to amplify a promoter that the protein in question, in this case RccR, binds to. As RccR was only known to bind to *rccR*, a limited PCR using the primers *rccR*EMSAup and *rccR*EMSAdn and the samples as template DNA, to identify when a *rccR* product could be seen on the agarose gel. Figure 4.10 shows RccR binds to *rccR* in all of the samples where the total DNA is present. However, in the Co-IP samples, it is only in the SBW25 wild type samples and not the not in the Δ*rccR* samples where RccR binds to *rccR* after 20 cycles in the PCR. In addition to this, Figure 4.10 also shows that the *rccR* region has been enriched in the SBW25 wild type Co-IP samples, as seen in lanes 5 and 7, compared to the SBW25 genomic DNA in lane 9. Finally, on the agarose gel in Figure 4.10, no difference can be seen between the samples grown in glycerol or pyruvate. The ChIP seq samples were subsequently given to The Genome Analysis Centre (TGAC) for quality control, library preparation and sequencing.

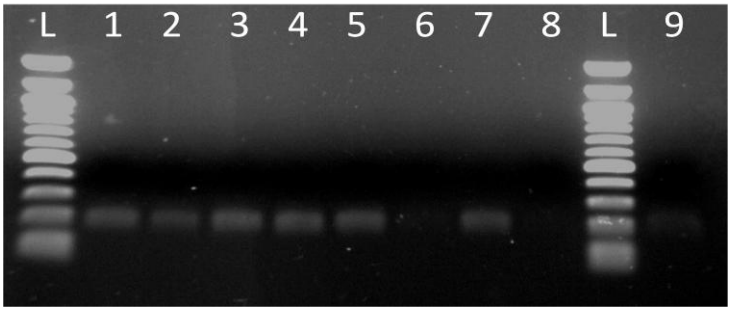

Figure 4.10. Limited PCR of the RccR ChIP seq samples. In lanes 1 – 4 the total DNA samples are present and in the order of SBW25 M9 Pyruvate, Δ*rccR* M9 Pyruvate, SBW25 Glycerol and Δ*rccR* M9 Glycerol. In lanes 5 – 8 the Co-IP samples are present and in the same order as lanes 1 – 4. In lane 9 SBW25 genomic DNA is present, and L is loaded with ladder.

Following the preparation and sequencing of the samples by TGAC, the analysis of the raw data was done by Govind Chandra. Differences in the SBW25 and the Δ*rccR* samples were identified and a total of eight DNA targets were discovered, including the already known *rccR* gene itself. The other targets can be categorised into metabolic genes, as shown in Table 4.1.

Table 4.1. The DNA targets identified by ChIP seq analysis that RccR binds to.

Gene name	Description	Additional information
pntAA	NAD(P) transhydrogenase subunit alpha 1	
pckA	Phosphoenolpyruvate carboxykinase	
aceE	Pyruvate dehydrogenase subunit E1	Binding in the intergenic region between these genes
glnD	Bifunctional glutamine-synthetase adenylyltransferase/de adenylyltransferase	
Mfd	Transcription-repair coupling factor	Binding in the intergenic region between these genes
Gap	Glyceraldehyde-3-phosphate dehydrogenase	
PFLU2154	Hypothetical protein	
PFLU3817	Isocitrate lyase	
PFLU5622	Hypothetical protein	Binding in the intergenic region between these genes
glcB	Malate synthase G	
rccR	Member of the RpiR family of transcriptional regulators	

Three of these genes, *glcB*, *pntAA* and *PFLU3817* an isocitrate lyase are involved in the glyoxylate shunt pathway. A further two of the genes, *gap* and *pckA*, are part of the gluconeogenesis pathway. Although no difference was seen between the samples grown in pyruvate and those grown in glycerol, one of the genes identified, *aceE*, is associated with the pyruvate metabolic pathway.

4.3.6. EMSA confirmation of the new seven ChIP seq DNA targets

In order to confirm that RccR binds to DNA sequences the ChIP seq experiment identified, EMSAs were done using fluorescent tags to identify the DNA in question. Phusion polymerase and specific primers, identified in Table 2.4, were used to amplify each of the seven ChIP seq DNA targets from the SBW25 chromosome, followed by a second nested

PCR using the primers 6-FAM For and 6-FAM Rev and Taq polymerase to create the fluorescent tags.

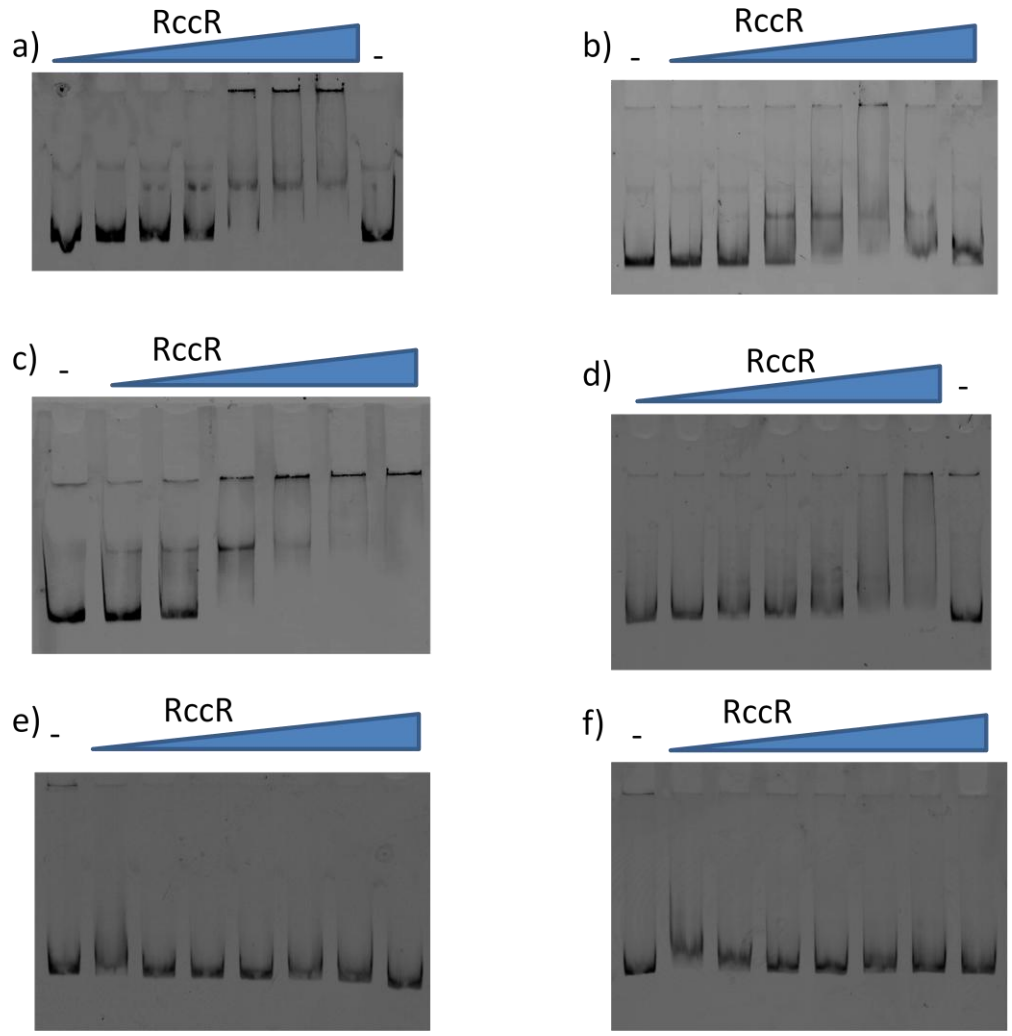

Figure 4.11. The EMSA of the RccR ChIP seq DNA targets results using fluorescent tags. An increase concentration of RccR was used in each EMSA with a negative control of BSA as shown by the – in each figure. The promoter DNA from *pckA*, *glcB* and *aceE* was used in a), b) and c) and a RccR-DNA complex can be seen. In d) the promoter DNA from the hypothetical protein PFLU2154 was used and a protein shift can be seen, however a clear RccR-DNA complex can not be seen. Finally, in e) and f) the promoter DNA from *mfd* and isocitrate lyase was used and no DNA shift can be seen.

The results of six of these EMSAs are shown in Figure 4.11. Four of the six ChIP seq targets bind to RccR and a RccR-DNA complex can be seen which are shown in Figure 4.11a), b), c) and d) which have *pckA*, *glcB*, *aceE* and the hypothetical protein PFLU2154 respectively as

their DNA targets. It can also be seen in Figures 4.11a) and b) that a second RccR-DNA complex is present which would require RccR to bind to these DNA targets as a dimer. On the other hand, the remaining two DNA targets, *mfd* and isocitrate lyase, did not bind to RccR, when a confirmation EMSA was done, as seen in Figure 4.11e) and f). From the EMSAs, it can be concluded that RccR only binds to four of the seven new ChIP seq targets; *pckA*, *glcB*, *aceE* and *PFLU2154*. In addition to binding to RccR, it was also observed that RccR was binding to *pckA* and *glcB* as a dimer. From these preliminary results, the remaining three DNA targets did not identify that RccR was able to bind to *mfd* or isocitrate lyase**.**

4.3.7. Predicted RccR consensus sequence

To determine if there was a consensus sequence for the binding of RccR, the DNA sequences from the eight ChIP seq targets were utilised in a MEME analysis (Bailey and Elkan 1994). Even though three results were provided, only the first gave a palindromic sequence, and this can be seen in Figure 4.12a. The results suggest that the predicted consensus sequence could be ATGTAGTG$_{n=12}$CACTACAT.

a)

b)

Name	Strand	Start	*p*-value	Sites
6. Isocitrate	+	541	7.55e-14	TAGACTAAAG ATGTAGTGCTTGAAAAAAAGCACTACAA AACATTTGAC
5. PFLU2154	-	367	3.99e-13	ACAGCAATTT ATGTAGTGAGCAAAAATAATCACTACAT AACCGTTGAC
7. glcB	+	193	1.48e-11	GCAAGTCATC ATGTAGTATGCCCAACCGTGCACTACAT AATGAAATCT
4. gap	-	315	7.64e-11	AACCTTCCAC ATGTAGGGGCTGAAAAAAGGGGCTACAT TATGCCGCCC
1. pntAA	+	158	8.32e-11	TCAAGGGATT ATGACGGGTGACGAAAAATTAACTACAA GTCACCCTGT
2. pckA	+	1375	1.87e-10	TCCCACGTGC ATGTAGGGTCCGCCGAAGACCACTACCT GG
8. rccR	-	135	9.01e-10	ACGGCAATTC TTGTAGTATAACTACAAGCTTGCTACAT CCTTGGCCCT
3. aceE	-	222	6.67e-9	AACTTCTACC TTTTCTTGTTCGCAAAAGATCAACAATA ACGATTATGC

Figure 4.12. The predicted consensus sequence for RccR following MEME analysis. a) The graphical results from the MEME analysis where the two boxes indicate the palindromic sequence where RccR is expected to bind. b) The individual sequences of the ChIP seq DNA targets that give the results for the sequence seen above.

In Figure 4.12b, it can be seen that only four of the eight ChIP seq DNA targets fully match this sequence; *pntAA*, *pckA*, *rccR* and *aceE* do not appear to have this sequence as their

consensus sequence as they have the following sequences respectively: ATGAC$_{n=12}$TACAA, ATGTA$_{n=12}$TACCT, TTGTA$_{n=12}$TACAT and TTTTC$_{n=12}$CAATA.

The other two results from the MEME analysis are shown in Figure 4.13. These results did not identify clear palindromic sequences, and for that reason are unlikely to be the consensus sequence for RccR.

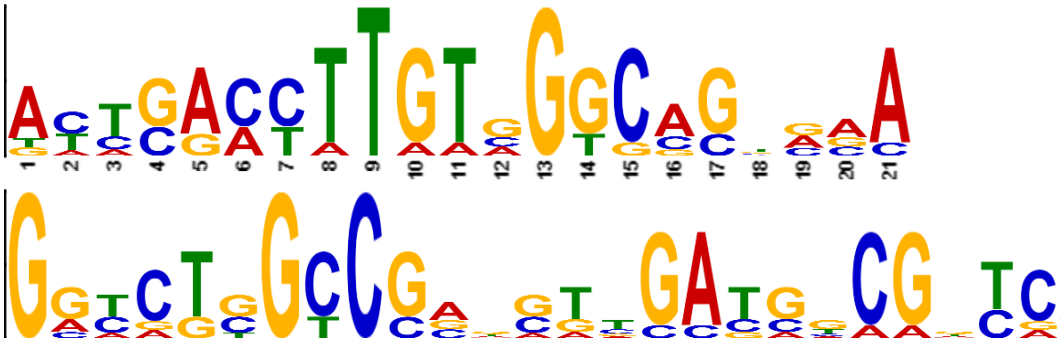

Figure 4.13. The other potential RccR consensus sequences.

From these results, it is likely that the first sequence, ATGTAGTG$_{n=13}$CACTACAT, will be the consensus sequence for RccR, however, further experimental analysis will need to be done before this can be established.

4.6. Conclusions

Even though a RccR polyclonal antibody has been produced in this study, it is not completely specific to RccR. As shown in the next chapter, the amino acid sequence of RccR is very similar to HexR, another member of the RpiR transcriptional regulator family. Therefore, it is highly probable that the RccR polyclonal antibody that has been produced is able to bind to other members of the RpiR family of transcriptional regulators within the SBW25 genome. In addition to this, the RccR polyclonal antibody appeared to bind in the Δ*rccAR* strain which was unexpected: this may be due to *rccA* and *hexR* being linked, as shown in Chapter 5. Despite this being the case, the RccR antibody produced was still sufficiently specific enough to be used in a ChIP seq experiment. A successful ChIP seq experiment was conducted, as shown in a confirmatory PCR to the promoter region of *rccR*, allowing the RccR binding regions within the SBW25 chromosome to be identified.

It was shown by EMSA analysis that RccR binds to the promoter region of *rccR*. Not only was a single RccR-DNA complex seen, but potentially, a second RccR-DNA complex was also present, suggesting that RccR could bind to its target promoters as a dimer. This has also been shown before for HexR; another RpiR regulator in *Pseudomonas putida* by Daddaoua and colleagues in 2009. Following this discovery, the results from the β galactosidase assay show that RccR is a repressor, which some RpiR regulators are known to be whilst others are activators (Daddaoua et al. 2009); (Sørensen and Hove-Jensen, 1996); (Yamamoto et al. 2001).

The results from the ChIP seq experiment identified eight DNA targets, including the *rccR* gene itself. All of the DNA targets appear to be involved in metabolism, mainly the glyoxylate shunt and gluconeogenesis pathway or the pyruvate metabolic pathway, suggesting that RccR has a role in metabolism like other RpiR transcriptional regulators (Leyn et al. 2011). It was later shown that all of the RccR ChIP seq DNA targets were the same as those identified in this study (Campilongo et al. 2017). Although it was shown in the Campilongo study that RccR regulates all the DNA targets, this was only done indirectly by quantitative real time PCR, and the direct RccR-DNA binding to three promoters, *aceA*, *aceE* and *rccR*, was tested. On the other hand, I was able to demonstrate the direct binding of several other promoter targets. Of the additional seven DNA targets, apart from *rccR*, only *pckA*, *glcB* and *aceE* were found to bind to RccR via EMSA analysis and in two of these cases, *pckA* and *glcB* can potentially bind as dimers. From the EMSA analysis in this study, the remaining four DNA targets did not confirm RccR binding, however, using quantitative real time PCR analysis on all seven DNA targets Campilongo et al. verified them.

Finally, from the ChIP seq analysis and the eight DNA targets discovered, a potential consensus sequence for RccR was found using MEME analysis. The first, and most likely, consensus sequence identified by MEME analysis was shown to be correct by Campilongo et al. in 2017.

Following the discovery of the functional link between *rccR* and *rccA* in Chapter 3, the next chapter will focus on establishing genes within the *P. fluorescens* SBW25 genome linking *rccR* and *rccA* and investigating these further.

Chapter 5 – The Comparison of RccR and HexR in *Pseudomonas fluorescens* in SBW25

5.1. Introduction

Previously in Chapter 3, it was identified that *rccR* and *rccA* are functionally linked. When swarming motility was monitored on KB agar with NaCl and Congo Red dye, a unique and atypical phenotype was seen for the Δ*rccA* strain, but this was recovered to a wild type phenotype in the Δ*rccAR* strain. In a second swarming motility test, the plasmid pME6032 over-expressing the *rccR* gene was tested in the SBW25 or Δ*rccAR* background. This result showed that when compared to the empty plasmid control, there was only a reduction in the swarm diameter in the SBW25 background. Both of these results indicate that the effects of RccA and RccR on swarming motility are linked. Therefore, in this chapter, additional experimental analysis, beginning with a mariner transposon screen, will be done in order to determine how *rccR* and *rccA* are linked. Any genes of interest that are discovered from the mariner transposon screen will be studied in more detail.

RccR is 74.3% similar to another transcriptional regulator; HexR (Daddaoua et al. 2009). Both RccR and HexR belong to the RpiR family of transcriptional regulators. HexR has previously been studied in *Pseudomonas putida* and a second gamma proteobacterial species *Shewanella oneidensis* (Daddaoua et al. 2009); (Leyn et al. 2011). However, it has not been examined in *P. fluorescens*. Further information about the RpiR family of transcriptional regulators and HexR can be found in Sections 1.5.1 and 1.5.2 respectively. As RccR and HexR are similar at the amino acid level, a comparison of the two proteins will be conducted in *P. fluorescens* SBW25, with further analysis on HexR in this chapter.

Some of the experimental work in this chapter has been done by Erasmus students, with experimental design and guidance given by myself. Any practical work done by another person is clearly mentioned in each case.

5.2. Aims

Throughout this chapter, the aims are to:

- use a mariner transposon screen to identify other genes which link *rccR* and *rccA*.
- create a *hexR* mutant in *P. fluorescens* SBW25.
- confirm if the *hexR* mutant has an effect on root colonisation in the *P. fluorescens* SBW25 wheat rhizosphere.
- prepare for additional chromatin immunoprecipitation sequencing (ChIP seq) experiments by creating chromosomal *hexR*-M2 and *rccR*-M2 tags
- over-express and purify HexR in *Escherichia coli* using a His_6 C terminal tag in preparation for a ChIP seq experiment and subsequent electrophoretic mobility shift assays (EMSAs).

5.3. Results

5.3.1. A mariner transposon screen identifies other genes linked to *rccA*

To identify other genes within the *P. fluorescens* SBW25 genome which also recover the *rccA* swarming motility phenotype, a mariner transposon screen was conducted by the Erasmus student Stefan Paulusch. The mariner family of transposable elements are diverse and widespread throughout the animal kingdom (Lampe et al. 1999). Mariner transposons encode for a transposase which is part of the superfamily of transposases and retroviral integrases and is distinguished by the conserved D,D35E motif in the catalytic domain of the protein (Lampe et al. 1999). The unique swarming motility phenotype of the Δ*rccA* strain, described in Section 3.3.5 was examined in this transposon screen. To begin with, the pALMAR3 mariner transposon was transformed into the Δ*rccA* strain resulting in the transposon inserting into the chromosomes of individual cells at random sites. Approximately 3,500 transformant colonies were picked, grown and tested to identify if the Δ*rccA* swarming motility was restored to wild type on KB agar supplemented with NaCl and Congo Red dye as seen in Figure 3.12 in Section 3.3.5. A selection of the screened transformants can be seen following the swarming assay using KB agar supplemented with NaCl and Congo Red dye in Figure 5.1. Figure 5.1 shows that some of the transposons insertions restore Δ*rccA* motility to

wild type as is required in this screen, while others have a similar motility phenotype to Δ*rccA* and finally some motility phenotypes can not be confirmed and need to be repeated.

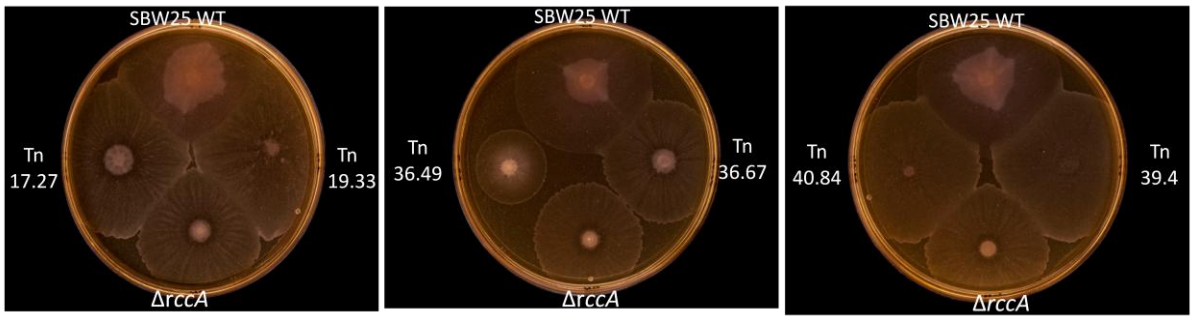

Figure 5.1. Swarming motility of a selection of mariner transposon inserts using KB agar supplemented with NaCl and Congo Red dye. Each plate contains a SBW25 wild type and a Δ*rccA* control as well as two transposon inserts. Any insert which restored the *rccA* swarming motility to wild type was of interest. Inserts Tn36.49, Tn39.4 and Tn40.84 all restore swarming motility to wild type, while inserts Tn17.27, Tn19.33 and Tn36.67 have a similar motility to Δ*rccA*.

If a particular mutant colony restored the *rccA* swarming motility phenotype, the sample was re-grown and a second motility test was done on minimal media with pyruvate, as explained in Section 2.7.3. If both of these tests were positive for phenotypic recovery, the locations of transposon insertions were identified by semi-random PCR and sequencing before the BLAST database was utilised to identify the transposon site. The sites that were identified are shown in Table 5.1.

These genes can be split into three categories; 1) metabolism genes, 2) signal transduction genes, and 3) hypothetical or unknown genes. Six out of the thirteen genes can be placed in the metabolism category; two, *PFLU0920* and *PFLU4766*, in carbon metabolism, while another two, *glnD* and *PFLU1514*, in nitrogen metabolism. *PFLU0920* encodes for a glycerate dehydrogenase, however it is also known as hydroxypyruvate reductase due to the reversibility of the enzyme and is found in eukaryotes, including plants (Muñoz-Bertomeu et al. 2013); (Ho et al. 1999) as well as prokaryotes (But et al. 2017). Glycerate dehydrogenase is mainly associated with the serine cycle. *PFLU4766* encodes for an acetyl-CoA synthetase, which has been shown in other bacteria, for example, *Pseudomonas putida* U, *Escherichia coli* and *Methanosaeta thermophila*, to convert acetate to acetyl-CoA via the intermediate acetyl-AMP (Arias-Barrau et al. 2006); (Renilla et al. 2012); (Valgepea et al. 2010); (Berger et al. 2012).

Table 5.1. The genes whose disruption recovers the *rccA* swarming motility phenotype following a mariner transposon screen.

PFLU number	Gene name	Gene function	Additional comments
PFLU1268	PII uridylyl-transferase (*glnD*)	Nitrogen metabolism	
PFLU0179	Putative phage integrase		PFLU0180 (*gabD*) could be *hexR* regulated
PFLU0179/PFLU5963	Putative phage integrase/Hypothetical protein		Two independent transposon insertions
PFLU1255	Glycerol-3-phosphate acyltransferase		
PFLU0920	Glycerate dehydrogenase	Carbon metabolism	PFLU0918 could be *hexR* regulated
PFLU4766	Acetyl-CoA synthetase	Carbon metabolism	Two independent transposon insertions
PFLU1512-14	PFLU1514 putative glutamine synthetase	Nitrogen metabolism	
PFLU1765	Cytochrome C biogenesis protein		
PFLU0087	Putative two-component system sensor kinase	Signal transduction	
PFLU5329	Putative cyclic-di-GMP signalling protein	Signal transduction	
PFLUr4	23S ribosomal RNA		
PFLU0426	Hypothetical protein	Metabolism	
PFLU0156	Hypothetical protein		

glnD and *PFLU1514* encode for a PII uridylyl-transferase and a putative glutamine synthetase respectively. *glnD* has been shown to sense low nitrogen levels in cells, leading to the bacterial nitrogen stress response (Yurgel et al. 2013). An orthologue of *PFLU1514* is *glnA1* in *Mycobacterium bovis*, and this has been shown to respond to the availability of nitrogen (Tripathi et al. 2013).

The two signal transduction genes identified in the transposon screen, *PFLU5329* and *PFLU0087*, encode for a putative c-di-GMP protein and a two component sensor kinase respectively, and relate directly to RccA. Firstly, RccA is also a putative c-di-GMP protein (see Chapter 3 for more information). Orthologues of *PFLU5329* indicate that this is MorA and although it is highly conserved within the *Pseudomonas* genus, its role in the development of the flagella and biofilm formation varies between species (Choy et al. 2004).

Secondly, c-di-GMP signalling is often part of a two component system. This is the case for the orthologues of the second gene, *PFLU0087*, which encode for KinB; an alginate biosynthesis sensor protein (Damron et al. 2009).

Even though one of the genes identified, *PFLU0426*, is only a hypothetical protein in *P. fluorescens*, orthologues of this gene suggest that it is either a FAD binding protein or a flavin dependent dehydrogenase. This would indicate that like other genes from the transposon screen, PFLU0426 is also involved in metabolism.

5.3.2. Mutating the HexR consensus binding site of *gabD* restores the swarming motility phenotype of *rccA*

As seen in Table 5.1, one of the genes discovered from the mariner transposon screen was the putative phage integrase gene, *PFLU0179*. Downstream of the putative phage integrase gene is a *hexR* binding site according to the consensus sequence from *P. putida* (Daddaoua et al. 2009), and further upstream from this is the *gabD* gene. *gabD* encodes for succinate-semialdehyde dehydrogenase I, indicating that this is involved in carbon metabolism. The *hexR* binding site is in the regulatory region of *gabD*; the genomic region under investigation is shown in Figure 5.2.

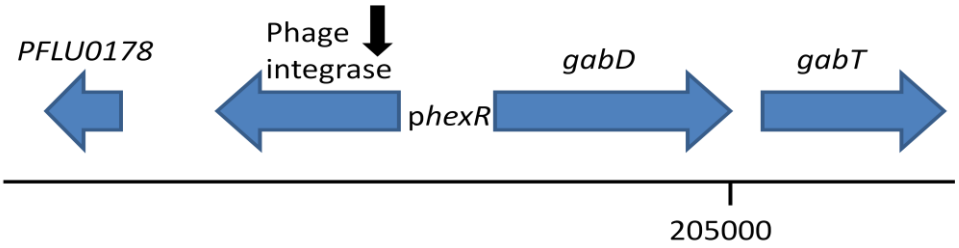

Figure 5.2. The position and organisation of *gabD* and the putative phase integrase gene, *PFLU0179*, in the *P. fluorescens* chromosome. The direction of the arrowhead specifies the direction in which transcription of the gene takes place. The binding site of *hexR* is in the regulatory region of *gabD*. The black arrow above the phase integrase gene indicates where the transposon inserted in the chromosome. The surrounding genes, *PFLU0178* and *gabT* encode for a hypothetical protein and 4-aminobutyrate aminotransferase respectively. The numbers underneath the black line indicate where in the chromosome the genes can be found, with the numbers representing the number of base pairs from the origin.

The practical work done in this section was completed by two short-term visiting students; Eleni Vikeli and Libby Humphries. In order to understand if it is the putative phage integrase gene or *gabD* regulation via the downstream *hexR* binding site that is responsible for the restoration of swarming motility seen in the mariner transposon screen, two different mutants were made using the two step allelic exchange method: a non-polar mutant of the phage integrase gene, and a point mutant in the downstream *hexR* binding site using site specific mutagenesis. The mutant with a different swarming motility in the SBW25 wild type and the Δ*rccA* background will indicate that this genetic perturbation is responsible for the restoration of swarming motility seen in the mariner transposon screen.

Firstly, the non-polar mutant of the phage integrase gene was created by amplifying the 500bp of upstream DNA and 500bp of downstream SBW25 genomic DNA using primers *PFLU0179*UPF and *PFLU0179*UPR and primers *PFLU0179*DNF and *PFLU0179*DNR respectively with Phusion polymerase in two different PCRs. The sequences of the primers can be found in Table 2.4. The insert was used to create a deletion construct in pME3087 which was used to delete the *PFLU0179* gene in the SBW25 and Δ*rccA* backgrounds by allelic exchange. Next, a similar construct was created to mutate the downstream *hexR* binding site from TTGT TCCGCCA ACAA to CTGT TCCGCCA GCAG, where the underlined bases are those which have been mutated. In addition to this, the *hexR* binding site was also altered in the same way in the Δ*rccA* background.

Following the creation of these mutants, the swarming motility of each was tested using minimal media supplemented with pyruvate, as in Section 3.3.5. Each mutant was tested in the wild type SBW25 background as well as the Δ*rccA* background to identify if like *rccR*, the *hexR* binding site is linked to swarming motility or if it is the putative phase integrase gene. When the mutants are tested in the SBW25 wild type background and compared to the wild type, there is no difference in the swarming motility, which can be seen in Figure 5.3. However, when the mutants are tested in the Δ*rccA* background and compared to Δ*rccA* wild type, there was a difference in the swarming motility as seen in Figure 5.3. Figure 5.3 also shows that the putative phage integrase gene mutant has the same swarming motility when compared to the *rccA* mutant in which it was tested in. This result suggests that it is therefore not this gene that is restoring the *rccR* swarming phenotype that is seen in the mariner transposon screen. It can also be seen in Figure 5.3 that the *hexR* binding site mutant increases swarming motility when compared to the *rccA* mutant it was tested in. The

swarming motility seen by the *hexR* binding site mutant in the *rccA* background not only increased but was restored to the SBW25 wild type swarming motility. This result suggests that it is the loss of *gabD* regulation by *hexR* that is responsible for the restoration of *rccA* swarming motility.

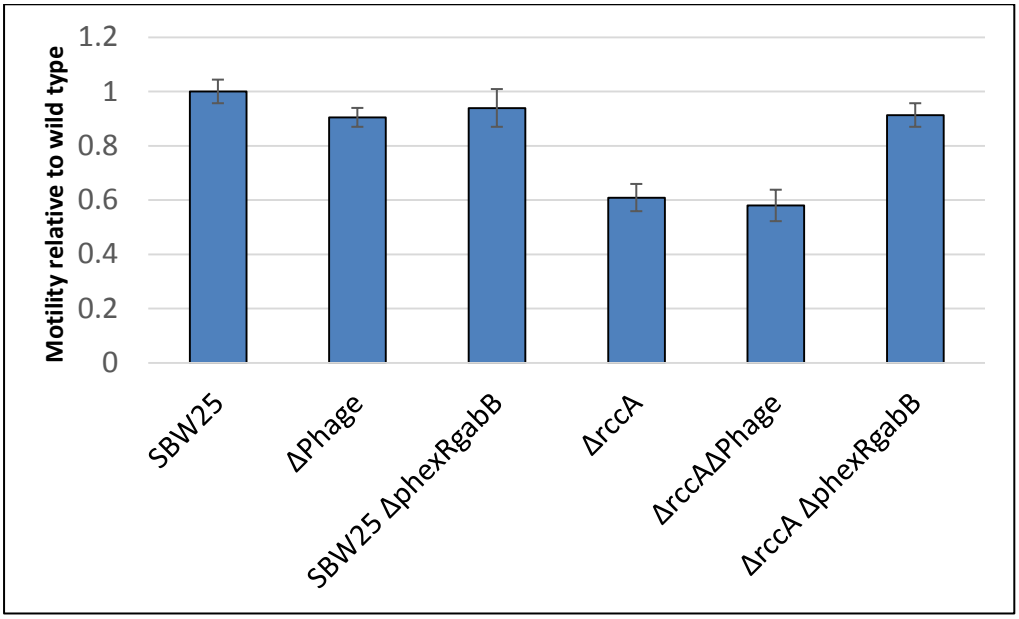

Figure 5.3. The swarming motility of the deleted phage integrase gene and mutated *hexR* binding site of *gabD*. When in the SBW25 wild type background, there is no difference in the swarming motility between the two different mutants. However, when the same mutants are tested in the *rccA* background, the deleted phage integrase gene has approximately the same swarming motility as *rccA*, but the mutated *hexR* binding site of *gabD* strain has recovered swarming motility. Using the student t test and a p value < 0.05 to compare each mutant to its respective background strain, there is only one statistically significant result: the mutated *hexR* binding site of *gabD* in the Δ*rccA* background.

5.3.3. Computational analysis of *hexR*

Following the mariner transposon screen to discover other genes in the SBW25 genome that recover the Δ*rccA* phenotype and the identification that some of these genes were regulated by the HexR transcriptional regulator, computational analysis of the *hexR* gene in the SBW25 genome was done using similar analysis as in Chapter 3. Using the upstream DNA of *hexR*, a single potential promoter sequence was identified for *hexR* and this can be seen in Figure 5.4.

5′GTCATCGGCTTTCATGAAGTCGACGTGAACATAGGTCAGGCGCGCCAGGAAACGCTGGGCGATGGTTTCGTCC
AGTTCCTTGCCGACGTATTTGCGCAGTGCTTGTTCGATGTGCGCCAGGTGCTGTTGCTCGGAACCGGCTTCCCGG
GCCAGGGCCAGGATGCGCGTGTCGTCGTGCAGGAGCCCCGCGCCATCAAGTTGGTAGAGGGCAGGAAATAACTTG
CGCAGCGCCAGATCACCCAAGGCGCCAAACAGGGCAAAGGTGCAGGGTTCTACGGTTATCGAAGGCATGATGTTT
GTTCTTTTATCAAGTTAAGCTACAAATACCTTTTTTCAAGGCATCACTCAAGGAAAAATGTAGTAA**T**AACCACA
ACATTTTCCCGAAATACGCATTCCGAGTGGTGGTGTTCAGCTACCCTCAGTAGGATAGGCCACCGCAAAAGACCA
CTTATCGATTGGTTACCACCCTAATTTGCATCGTCGCCCGAAGGAAAGACTGA 3′

Figure 5.4. The predicted promoter sequence of *hexR*. The 500bp upstream DNA sequence of *rccA* with the predicted promoter sequence highlighted in red. The predicted transcription start site is shown by the larger letter.

The amino acid sequence of HexR showed that it is predicted to be composed of two domains; a helix turn helix domain, and a sugar isomerase (SIS) domain as seen in Figure 5.5.

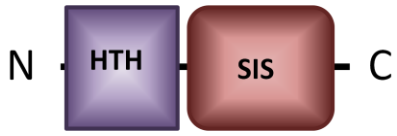

Figure 5.5. The domain organisation of HexR. HexR consists of a N terminal helix turn helix DNA binding domain and a C terminal sugar binding domain. These are annotated by HTH and SIS respectively.

This domain organisation is identical to RccR and confirms that like RccR, HexR belongs to the RpiR family of transcriptional regulators. Due to both HexR and RccR being members of the RpiR family, an alignment of the amino acid sequences from the *P. fluorescens* SBW25 genome was done and this can be seen in Figure 5.6. From the alignment, it was shown that HexR and RccR are 41% identical and 74.3% similar at the amino acid level.

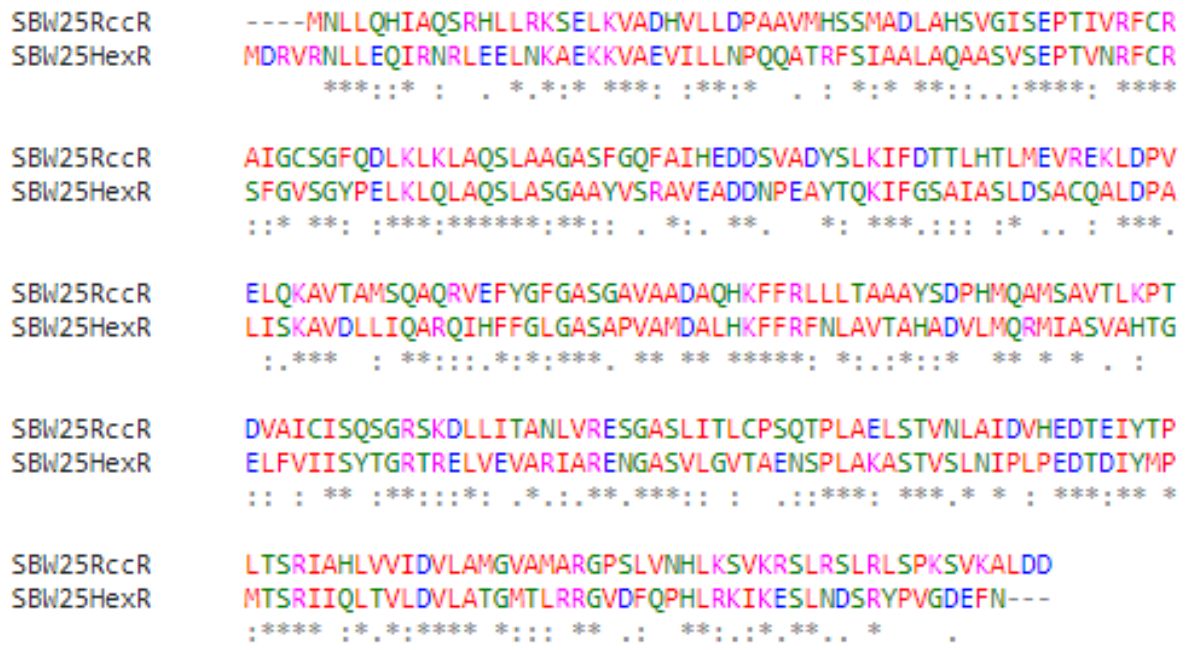

Figure 5.6. An amino acid alignment of RccR and HexR from *P. fluorescens*. A sequence alignment was done using the Clustal Omega program with the amino acid sequences of RccR and HexR from *P. fluorescens*. The * indicates amino acids that are 100% identical, while the : and . represent amino acids with similar and weakly similar properties respectively.

5.3.4. The creation of the *hexR* mutant

For further analysis to take place on *hexR* in *P. fluorescens*, similar to the analysis that was done on *rccR*, a *hexR* mutant was made using the two step allelic exchange method. The SBW25 genomic DNA and Phusion polymerase were utilised in two separate PCRs, with primers *hexR*upF and *hexR*upR in the first PCR and primers *hexR*dnF and *hexR*dnR in the second PCR, to amplify the 500bp of upstream and downstream DNA respectively. The sequences of the primers utilised to create the *hexR* mutant can be found in Table 2.4. Figure 5.7 shows the pME3087 product and the two PCR products that were used to create the deletion construction for allelic exchange to take place in the SBW25 background.

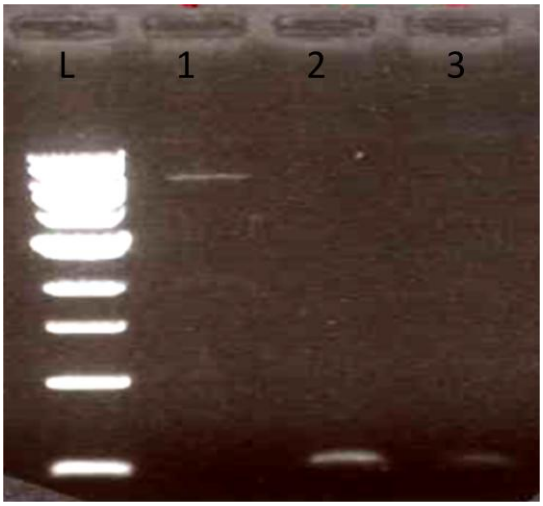

Figure 5.7. The ligation gel for pME3087 and *hexR*. The vector pME3087 and the PCR products of the 500bp upstream and downstream flanking DNA of *hexR* are present on an agarose gel following a restriction digestion in preparation for a ligation reaction. Where L represents the ladder and in lane 1 pME3087 (*Bam*HI/*Eco*RI); lane 2 *hexR* upstream (*Bam*HI/*Xba*I); lane 3 *hexR* downstream (*Xba*I/*Eco*RI).

Figure 5.8 shows the cloning gel which identifies that a *hexR* mutant has been created.

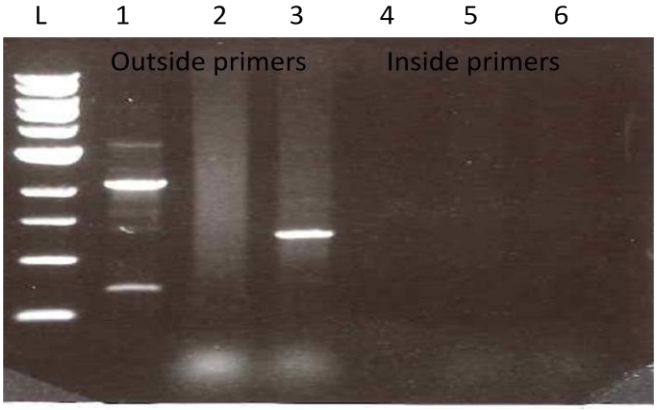

Figure 5.8. Cloning gel confirming the creation of the *hexR* mutant. Primers outside and inside of the deleted region of *hexR* were used to confirm if a mutant had been created. A 1kb product is expected for a *hexR* mutant compared to a 2.2kb product for the SBW25 wild type when using the outside primers, however, no products are expected when using the inside primers. In lanes 1 – 3 are the samples using the outside primers and in the order of SBW25 wild type, *hexR* mutant sample 1 and *hexR* mutant sample 2. In lanes 4 – 6 are the samples using the inside primers and in the same order as lanes 1 – 3. L represents the ladder.

From Figure 5.8 it can be seen that only sample 2 is a *hexR* mutant. The organisation of *hexR* and the final construct of Δ*hexR* are shown in Figure 5.9.

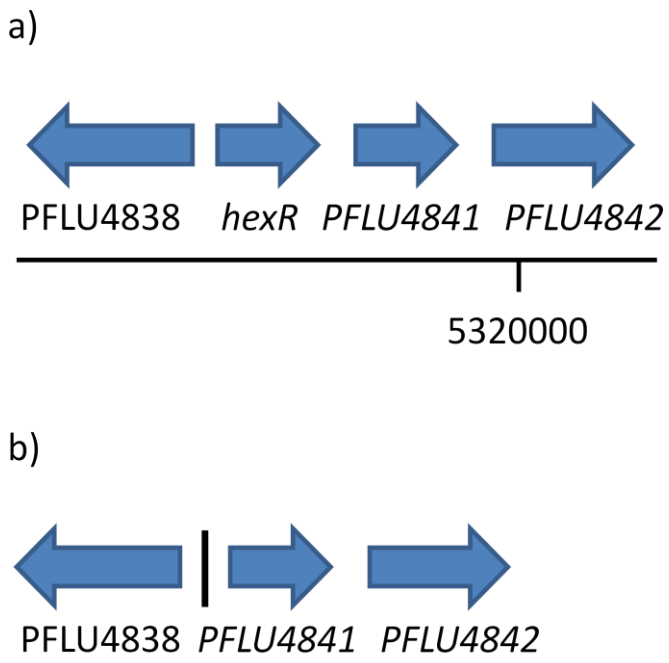

Figure 5.9. The position and organisation of *hexR* in *P. fluorescens* and the final construct of the *hexR* mutant. The direction of the blue arrowheads specifies the direction the transcription of the gene takes place. a) Shows the position and organisation of *hexR* and the surrounding genes, with the number underneath the black line indicating where in the chromosome the genes can be found, with the numbers representing the number of base pairs from the origin. The chromosomal structure of the *hexR* mutant is shown in b). The black line indicates where the *hexR* gene should be.

The creation of this *hexR* mutant will now allow further analysis of whether *hexR* has an effect on the ability to colonise the roots of the plants, as well as the effects it has on swarming motility. These two processes are crucial to investigate as RccR was shown in Chapter 3 to have important roles in controlling both phenotypes.

5.3.5. Initial experiments indicate that Δ*hexR* does not have a wheat colonisation defect

To identify if the deletion of the *hexR* gene has an effect during the colonisation of wheat seedlings, like Δ*rccR* and Δ*rccA* do as seen in Chapter 3, a root colonisation assay was set up. In this assay, the wild type SBW25 and the Δ*hexR* bacteria compete to colonise the wheat seedlings. As the SWB25 strain utilised in this assay is marked with the *lacZ* gene, it can be distinguished from the mutant strain by the use IPTG and X-GAL agar plates, where the wild type-*lacZ* and mutant strains form blue and white colonies respectively. Unlike Δ*rccR*, Figure 5.10 shows that Δ*hexR* does not have an effect on colonisation compared to the wild type SBW25. This data therefore indicates that HexR does not appear to play an important role in plant colonisation.

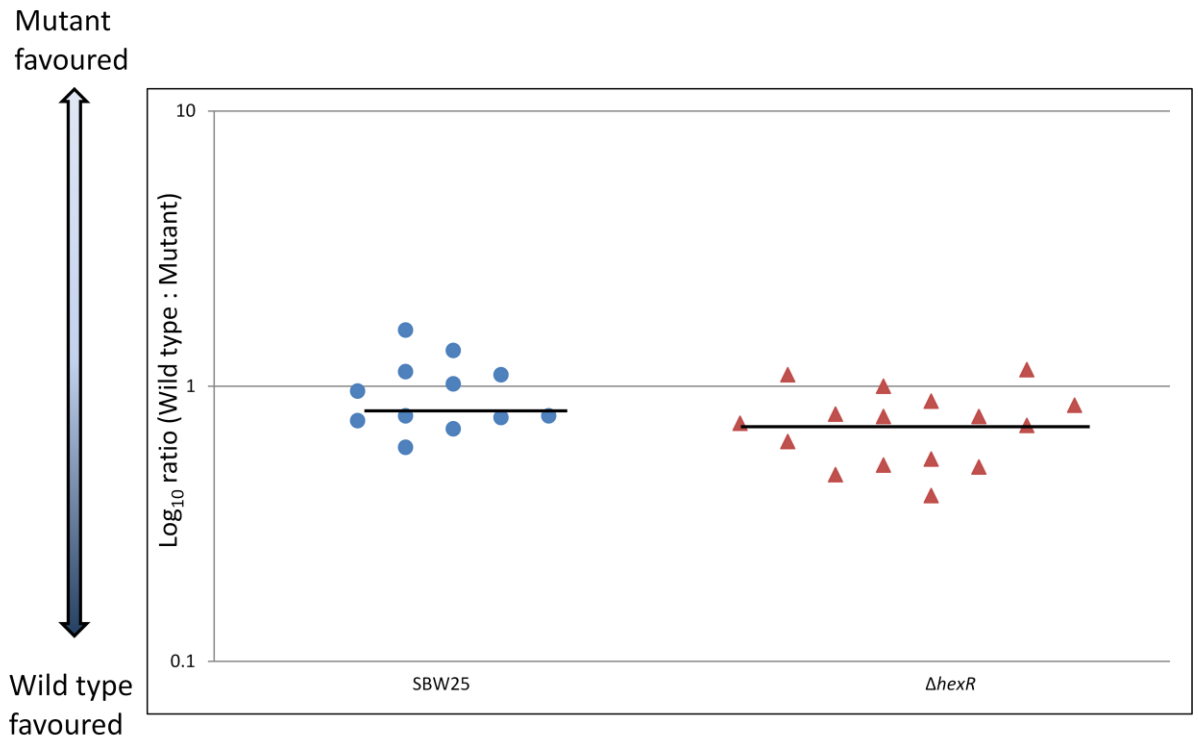

Figure 5.10. Δ*hexR* does not have a root colonisation deficiency compared to SBW25 wild type. Each dot represents an individual wheat seedling and the median root colonisation efficiency of each strain is shown by the solid black line. There is no statistical significance between the colonisation efficiencies of SBW25 wild type and Δ*hexR* when using the Mann Whitney U test (p value of 0.7).

5.3.6. Over-expression and purification of HexR with a C terminal His_6 tag

Even though HexR has been studied in the gamma proteobacteria species *Shewanella oneidensis* and the closely related *P. putida*, there have been no studies in *P. fluorescens*. Therefore, to understand more about HexR and its biochemistry it was over-expressed from the pET42b vector in *E. coli* and purified using a His_6 tag at the C terminus of the protein. To begin with, a PCR reaction using the primers *hexR*purifyF and *hexR*purifyR, was done to amplify *hexR* from the SBW25 chromosome. The full sequence of *hexR*purifyF and *hexR*purifyR can be found in Table 2.4. The amplified *hexR* fragment was incorporated upstream of the His_6 tag in the pET42b vector using the restriction enzymes designed within the primers used to amplify the fragment. Following the construction of the pET42b-*hexR* vector, a small scale induction of HexR was done using *E. coli* BL21 (DE3) pLysS cells to identify if the over-expression of *hexR* was possible. The results of the small scale induction can be seen in Figure 5.11 and indicate that this is possible, leading the way for a larger scale expression of HexR and to the eventual purification of the protein.

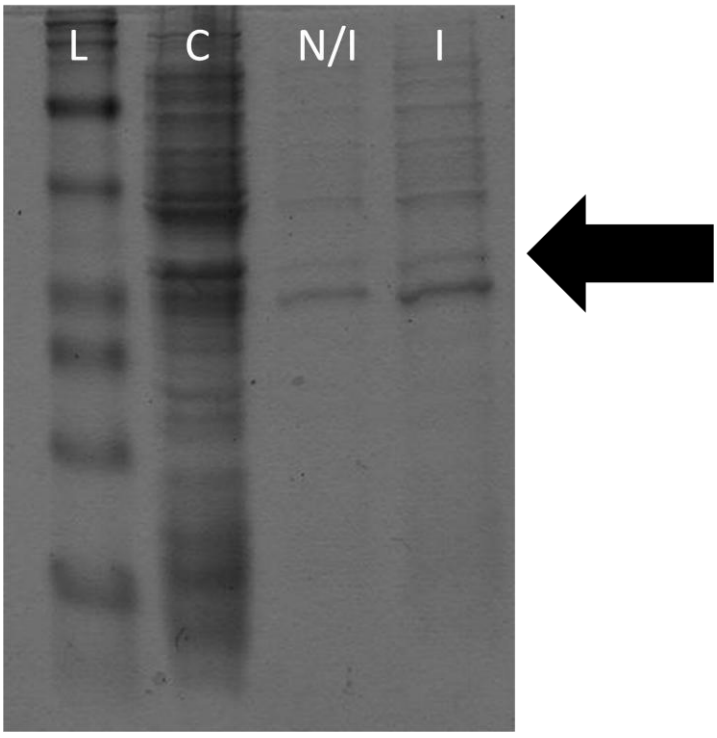

Figure 5.11. The small scale induction of HexR. There is a difference in the induced and non-induced samples, therefore when induced, HexR is over-expressed. Where L indicates the protein ladder, C concentrated sample, N/I non-induced sample and I indicates the induced sample. The black arrow indicates the position of the HexR-His_6 band.

Using the His_6 tag and a nickel affinity column, HexR was purified by fast protein liquid chromatography. The column was washed of impurities using the equilibration buffer made of 50mM Tris-HCl, 300mM NaCl, 1mM dithiothreitol and 10mM imidazole at pH7.9 before HexR was eluted using a gradient of the elution buffer composed of 50mM Tris-HCl, 300mM NaCl, 1mM dithiothreitol and 10mM imidazole at pH7.9.

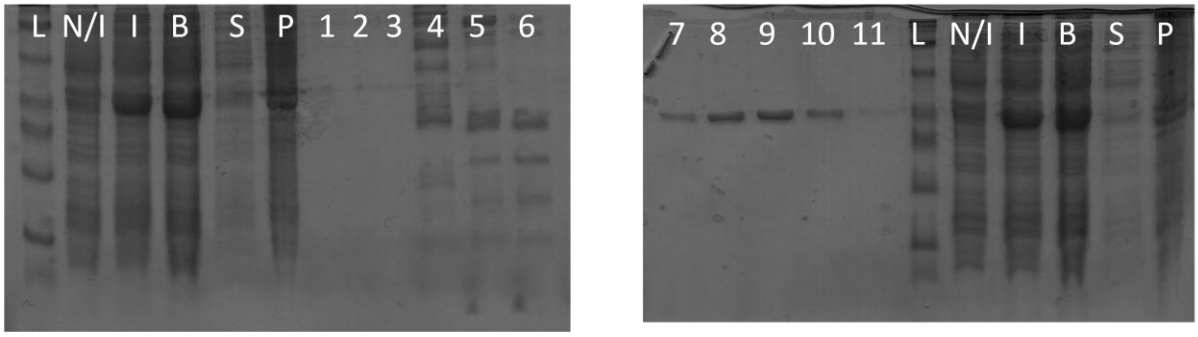

Figure 5.12. The SDS-PAGE identifying the purified HexR protein. Two SDS-PAGE gels were required for the number of fractions obtained, however, only fractions 7 – 10 contain HexR. The letters L, N/I, I, B, S and P stand for ladder, non induced, induced, broken cells, soluble protein and pellet respectively. While the numbers represent the fraction number taken from the column.

Fractions of the eluted protein were collected and tested using SDS-PAGE to identify those that only contained HexR. The resulting SDS-PAGE can be seen in Figure 5.12 and shows that HexR has been purified, therefore allowing further analysis to be done using this protein in the future.

5.3.7. The preparation of *hexR* and *rccR* flag tagged strains to enable ChIP seq analysis

In order to identify the DNA sequences that HexR binds to in *P. fluorescens* chromosome, a ChIP seq experiment is required. The results from this ChIP seq will not only provide answers to the DNA sequences HexR binds to, but it will also allow more comparisons to take place between HexR and RccR. Due to not having a polyclonal antibody for HexR in SBW25, a construct was produced enabling a flag epitope tag to be incorporated onto the end of the *hexR* gene within the chromosome, allowing the M2 antibody to be utilised in the ChIP

seq experiment. In addition to the *hexR* gene, the flag epitope tag was integrated at the end of the *rccR* gene in a second gene manipulation vector to allow a comparative ChIP seq experiment to take place. For a full description of how the chromosomal M2 vector tags were prepared see Section 2.3.11. Firstly, the gene of interest was removed from the over-expression plasmid, pME6032, if a strain was available, otherwise this was prepared. At the same time, the flag epitope tag was amplified from the pSUB11 vector and the resulting two products were ligated and transformed into *E. coli*, resulting in the gene-flag tag. Once the sequence of the flag epitope tag was confirmed to be correct and in-frame, the following stages could be done. Approximately 500bp of the downstream chromosomal DNA was amplified in a PCR using Phusion polymerase and specific primers, for *rccR* and *hexR* were used respectively. All the sequences of the primers used for this piece of work can be found in Table 2.4. This 500bp of downstream DNA was required for the first crossover event to occur, as explained in Chapter 3, and therefore it was added downstream of the flag epitope of the gene-flag construct. In addition to this, the suicide vector, pTS1, was prepared and the gene-flag-chromosomal DNA construct was cloned into the multiple cloning site to create the allelic exchange vector. Once this new gene-flag-chromosomal DNA construct was made, the allelic exchange method was able to be used to add the flag epitope into the *P. fluorescens* chromosome at the required position.

Even though these two strains were created, it was not clear if the M2 tag that was incorporated into the chromosome was neutral, and there was not enough time to do further experiments to establish this. Therefore, before these strains can be utilised in a ChIP seq experiment, this will need to be clarified.

5.4. Conclusions

In the mariner transposon screen for Δ*rccA* swarming phenotype recovery, it was discovered that the genes identified could be split into two categories; metabolic and signalling genes. The metabolic genes or orthologues of these genes that were identified in the transposon screen indicate that these were involved in carbon and nitrogen metabolism. For example, *PFLU4766* encodes an acetyl-CoA synthetase and is an orthologue of *acsA*. In *Escherichia coli*, once *acsA* metabolises acetate into acetyl-CoA via the intermediate acetyl-AMP, the acetyl-CoA is metabolised further by the tricarboxylic acid cycle or the glyoxylate shunt

(Renilla et al. 2012); (Valgepea et al. 2010). In *Vibrio vulnificus*, when acetate is the only carbon source available, *acsA* is required for growth (Kim et al. 2015). All of these studies have shown that AcsA expression is induced by acetate however it is repressed by glucose.

The two signalling genes which were identified in the mariner transposon screen relate to RccA, which is a putative c-di-GMP protein. PFLU5329 is also a putative c-di-GMP protein, while PFLU0087 is expected to be two-component system sensor kinase. In other *Pseudomonas* species, orthologues of PFLU5329 are annotated as MorA, which is involved in controlling the biogenesis of flagella. This may have been identified in the screen as the unique swarming motility phenotype of the Δ*rccA* strain was examined as part of the screen. Even though MorA is conserved among *Pseudomonas* species, there are varying results for its function in each species (Choy et al. 2004). Orthologues of *PFLU0087* indicate that this encodes the alginate biosynthesis sensor protein KinB and is also part of a two component system (Damron et al. 2009). *PFLU0087* may have been identified in the mariner transposon screen as alginate is important in the formation of the wrinkly colony spreader seen in *Pseudomonas aeruginosa* (Hay et al. 2009). Although KinB does not regulate c-di-GMP metabolism, there are many other two component systems that do, for example VxrA and VxrB in *Vibrio cholerae* (Teschler et al. 2017), RavA and RavR in *Xanthomonas campestris* (Tao et al. 2014), GacA and GacS in *P. putida* (Martínez-Gil et al. 2014) to name a few.

Unfortunately, *rccR* was not identified as one of the transposon insertion sites during the mariner transposon screen. Although it was expected that *rccR* would appear in this transposon screen from the previous results from Chapter 3, it was not totally surprising from the number of mutants compared to the number of genes in the genome; approximately 3,500 mutants to approximately 6,000 genes in the whole genome. In addition to this, due to the number of stages in this screen, it was difficult as well as time consuming to do, therefore a note of caution must be taken when planning, analysing and interpreting the results from these types of screens. Even though the genes that were picked up do look promising, it is likely that not every single gene that restores the *rccA* swarming motility phenotype was identified, as seen by the lack of *rccR* in the screen. It may be that *rccA* is sensing the changes in the environment, such as the differing levels of organic substances in the cell and responding accordingly. If this is the case, the genes identified in the mariner transposon screen would therefore be the ones which recovered the metabolic imbalance imposed by the *rccA* mutant used in the screen.

Previously in this study, it was shown that *rccA* and *rccR* were functionally linked, and later established that RccR is a regulator of the glyoxylate shunt and gluconeogenesis pathways by Campilongo et al. in 2017. By combining the mariner transposon screen and swarming motility assays, it was identified that the loss of *gabD* regulation by *hexR* was responsible for the restoration of *rccA* swarming motility. This observation suggests that either a) *rccA* and *gabD* may also be linked or b) RccR and GabD have similar functions with respect to the metabolic balance of SBW25.

It was shown in this study that the *hexR* mutant does not have an effect on the colonisation of the roots in the wheat rhizosphere. However, this is disputed by Campilongo et al. who found that *hexR* is important for plant colonisation. This would suggest that as I only managed to do this assay once, it was not a replicable result. Alternatively, there may have been a problem with the many reagents that are utilised in the assay when I did the experiment the one time, more than likely contamination.

Throughout this chapter, many useful components for future analysis into HexR and its relationship with, if there is one, RccR and/or RccA have been constructed. For example, it has been demonstrated that it is possible to over-express and purify HexR from *P. fluorescens* in *E. coli*. As well as this, *rccR*-flag and *hexR*-flag chromosomal genetic constructs have been made in order for a ChIP seq experiment to done to compare the genes the two proteins regulate.

Chapter 6 – Discussion

6.1. RccR and RccA have important roles in plant rhizosphere colonisation

During this study, it was identified that *rccR* and *rccA* are both important for plant rhizosphere colonisation, which is in agreement with the findings from a study by Campilongo et al. in 2017. As well as these studies examining the wheat rhizosphere and the bacterium *P. fluorescens* SBW25, *rccA* has been identified as an important gene by Silby et al. in 2009 when an in vivo expression technology (IVET) study was used to identify the up-regulated genes in the *P. fluorescens* SBW25 sugar beet rhizosphere. In 2007, a study into the *Pseudomonas putida* rhizosphere identified that a cyclic-di-GMP (c-di-GMP) gene, *rup4959*, was important for the colonisation of plants (Matilla et al. 2007). Even though *rup4959* is not an orthologue of *rccA*, the findings by Matilla and co-workers identified that c-di-GMP signalling could be involved in the colonisation process. A second study, this time using *P. fluorescens* F113, the c-di-GMP related genes *sadB*, *wspR* and *gacS* and the leguminous plant alfalfa identified that these three genes were all required for the colonisation of the alfalfa root tips (Barahona et al. 2010).

As well as studying the effect of *rccR* and *rccA* on the ability of SBW25 to colonise plants, a *hexR* mutant was also investigated, but showed no effect on plant colonisation in *P. fluorescens* SBW25. This result however differs from that seen by Campilongo et al. who also studied *hexR* in *P. fluorescens* SBW25, and showed that a *hexR* deletion has an effect on the colonisation of plants and is therefore an important gene in this process (Campilongo et al. 2017). Even though RccR and HexR are similar, the functions of each protein in *P. fluorescens* is now known and was described by Campilongo et al.. *hexR* is known to regulate the Entner Doudoroff (ED) pathway in *Pseudomonas aeruginosa* (Hager et al. 2000) and *P. putida* (Daddaoua et al. 2009). In the study by Campilongo et al., HexR in *P. fluorescens* was confirmed to regulate the *edd*, *gap-1* and *zwf* genes. Like in other species, it was identified that HexR in *P. fluorescens* functions as previously described and it was also shown in the presence of glycerol, pyruvate or acetate that the *hexR* mutant has a different growth profile to the wild type, where in these conditions the ligand for *hexR* is unable to accumulate.

6.2. ChIP seq results

Following further analysis of the RccR protein, it was shown that like other transcriptional regulators RccR is able to negatively regulate its own expression. Other RpiR family proteins have been shown to be the subject of autoregulation, for example, RpiR in *Escherichia coli* is also a repressor (Sørensen and Hove-Jensen, 1996), whilst GlvR in *Bacillus subtilis* is an activator (Yamamoto et al. 2001). From computational analysis, a single promoter for *rccR* was identified and this DNA region was used in electrophoretic mobility shift assays (EMSAs) to show that RccR is able to bind to the upstream region of the *rccR* gene. From the EMSA results, it could be seen that RccR is able to form two RccR-DNA complexes suggesting that binding may occur as a dimer in some circumstances. This is not the first time that a RpiR transcriptional regulator has been proposed to bind as a dimer; in *Pseudomonas putida*, HexR has also been identified as binding as a dimer (Daddaoua et al. 2009).

From the chromatin immunoprecipitation sequencing (ChIP seq) analysis on RccR, it was identified that RccR binds to eight different DNA sites within the SBW25 chromosome. However, in this study, only half of these DNA sites were confirmed by EMSA analysis. Subsequently, Campilongo et al. were able to confirm all eight DNA binding sites of RccR by using quantitative real time PCR (Campilongo et al. 2017). In the EMSA experiments, it could be seen that RccR formed two RccR-DNA complexes when in the presence of *pckA* or *glcB* DNA, once again suggesting that RccR may be able to bind as a dimer in some circumstances, just like it has been shown in *P. putida* with HexR (Daddaoua et al. 2009). Not only was RccR shown to bind as a dimer, it was seen to bind as a monomer in the case of *aceE*, and finally in the EMSAs with the promoter DNA from *PFLU2154* a protein shift was seen but no RccR-DNA complex. These three different types of binding suggest that RccR may have different modes of binding to its DNA targets. Later, it was shown by Campilongo et al. that one of the RccR DNA targets, *aceE*, has a particularly complex binding pattern with two similar RccR binding sites separated by 68bp. With two RccR binding sites in the upstream region of *aceE*, it may be expected that RccR binds to *aceE* as a dimer, however this was not the case and further analysis by Campilongo et al. suggests that the DNA bends in order for both binding sites to bind to RccR.

Three of the eight ChIP seq DNA targets, *pckA*, *aceEF* and *gap*, have previously been identified as having a role in the central carbon metabolism of *Proteobacteria* and being regulated by *hexR* (Leyn et al. 2011). In addition to these three genes, malate synthase and

isocitrate lyase encoded by *aceBA*, were also present on the list of genes regulated by *hexR* and having a role in carbon metabolism in *Proteobacteria* by Leyn and colleagues. *aceBA* is interesting, as the ChIP seq analysis identified *glcB* and *PFLU3817* which encoded malate synthase G and isocitrate lyase respectively, although these were two different DNA targets.

Using the results from the ChIP seq analysis on RccR and inputting the DNA sequences of all eight targets into MEME analysis, a single consensus sequence was identified. This consensus sequence, ATGTAGTG$_{n=12}$CACTACAT, was found amongst six of the eight ChIP seq DNA targets. Although this was the case, three of the four DNA targets only had a few DNA base changes. Due to time limitations, this could not be confirmed, but once again a similar consensus sequence was later confirmed by Campilongo et al. in 2017. The consensus sequence which was identified in this study only has a linker region of 12bp, however in the study by Campilongo et al. the linker region is made up of 15bp. As well as this difference, this consensus sequence was also shown in *aceE* and manually found in *pckA*. Even though *pckA* has the same consensus sequence, it has a linker region of 16bp making it an extra base pair longer which may explain why it was not discovered in the MEME analysis.

In addition to this consensus sequence, Campilongo et al. also identified a second consensus sequence which was very similar to this (TGTAGTG$_{n=3}$CACTACA), but this was only found upstream of *rccR* and *aceE*, where two were found separated by 68bp. This observation, with the others seen in this and the Campilongo study, indicate that RccR is able to regulate *aceE* differently depending on the conditions. Both of these consensus sequences were confirmed in their study by ReDCaT surface plasmon resonance assays, however, only weak binding of RccR to *pckA* was seen therefore suggesting that binding to the *pckA* promoter may not be physiologically relevant.

6.3. The roles of RccR and HexR

As both RccR and HexR are members of the RpiR family of transcriptional regulators, it was assumed that they are both involved in carbon metabolism due to the C terminal SIS domain that is characteristic of these regulators. Of those investigated so far, RpiR and MurR in *Escherichia coli* have been found to bind to allose 6-phosphate and N-acetylmuramic acid 6-phosphate respectively (Sørensen and Hove-Jensen, 1996); (Jaeger and Mayer, 2008), while

in *Haemophilus influenza* SiaR binds to glucosamine 6-phosphate (Johnston et al. 2010). Due to the expectation that RccR and HexR are both involved in carbon metabolism, the majority of the experiments in this study have been done using a variety of media containing defined carbon sources, mainly pyruvate, glucose or glycerol. Previous studies by (del Castillo et al. 2008) into the Entner Doudoroff (ED) pathway have found that HexR is important in this pathway as well as being involved in glucose catabolism. Since this study, it has been confirmed that RccR controls cellular responses to a number of difference carbon sources including pyruvate and acetate (Campilongo et al. 2017).

Although I did not have enough time to fully investigate the function of RccR, it was always expected that RccR and HexR would have similar or related functions due to their sequence similarity. One of the key functions of RccR that was established in this study was that it was a negative regulator of its own transcription; this however, is different to HexR which does not regulate its own transcription (del Castillo et al. 2008). As well as this observation, it was shown from the ChIP seq results that RccR binds to metabolic genes from the glyoxylate shunt, gluconeogenesis pathway and the pyruvate metabolic pathway. The glyoxylate shunt was first described by Kornberg and colleagues and is similar to the tricarboxylic acid (TCA) cycle, with six out of the eight reactions present: the first two reactions differ, allowing growth on acetate as the main carbon source (Kornberg and Krebs, 1957); (Kornberg and Madsen, 1958). The glyoxylate shunt is present in bacteria, plants and fungi (Kondrashov et al. 2006); (Dunn et al. 2009). At isocitrate, the carbon is converted into glyoxylate and succinate by the enzyme isocitrate lyase. Finally, a second enzyme; malate synthase completes the pathway by producing malate from glyoxylate and acetyl-CoA (Kondrashov et al. 2006). Both an isocitrate lyase, *PFLU3817*, and malate synthase, *glcB*, were identified in the ChIP seq analysis for RccR, confirming that RccR binds upstream of genes from the glyoxylate shunt. The gluconeogenesis pathway is also essential for growth on acetate (Phue et al. 2005). After further experiments by Campilongo et al. it was established that RccR is a regulator of the glyoxylate shunt and gluconeogenesis pathways. As well as this, the ligand which RccR binds to was also identified; just like HexR in *P. putida*, (Daddaoua et al. 2009) it appears that RccR binds to the ED pathway intermediate 2-keto-3-deoxy-6-phosphogluconate (KDPG) (Campilongo et al. 2017). To summarise, both of these proteins have some important similarities, such as their ligand and structure, however, they also are unique regulators which have different binding sites, modes of actions and cellular roles.

Although this study was unable to assign a definite function for HexR in *P. fluorescens*, previous work on HexR in *P. putida* (del Castillo et al. 2008); (Daddaoua et al. 2009) and in *Pseudomonas aeruginosa* (Hager et al. 2000) have established that HexR in these organisms is involved in the ED pathway. In 2009, the ligand and the consensus sequence for HexR was identified by Daddaoua and co-workers as the ED pathway intermediate KDPG and the sequence of $TTGT_{n=7–8}ACAA$. It is partly for this reason that the studies of HexR in *P. fluorescens* began here and it is also expected to be part of the ED pathway. Additional work done by Campilongo et al. was able to confirm that HexR in is indeed part of the ED pathway in *P. fluorescens*, binds to the same ligand, KDPG, and finally has the same consensus sequence, $TTGT_{n=7–8}ACAA$, as in *P. putida*.

6.4. Mariner transposon screen results

From the mariner transposon screen, which was used to identify genes within the SBW25 genome that might link *rccR* and *rccA*, further genes were actually discovered that assisted with the identification and assignment of regulatory function to *rccR*. Some of genes were involved in the metabolism of carbon, such as *PFLU0920* and *PFLU1514*, which encode a glycerate dehydrogenase and a putative glutamine synthetase, and have been shown to encode similar genes in various bacteria (Leyn et al. 2011). For example, although not identical to *PFLU0920*, the *aceEF* genes encode for a pyruvate dehydrogenase in other *Pseudomonas* species as well as *Oceanospirillales*. Both *PFLU0920* and the *aceEF* genes are dehydrogenases, however, the sugar to be broken down is different in each case.

Six of the thirteen genes identified in the screen are involved in metabolism. One of these genes was *PFLU4766* which encodes an acetyl-CoA synthetase and is an orthologue of *acsA*. Previous studies have identified that acetate is degraded to acetyl-CoA via the intermediate acetyl-AMP by *acsA*. In *Escherichia coli*, the tricarboxylic acid (TCA) cycle or the glyoxylate shunt then metabolises the acetyl-CoA further (Renilla et al. 2012); (Valgepea et al. 2010). It is now known that RccR is involved in the glyoxylate shunt (Campilongo et al. 2017).

A further two metabolic genes were identified however these are involved in nitrogen metabolism; *glnD* and *PFLU1514* encoding for a PII uridylyl-transferase and a putative

glutamine synthetase respectively. In order for the nitrogen stress response to be activated in *Sinorhizobium meliloti*, the low levels of nitrogen are sensed by *glnD* (Yurgel et al. 2013). Orthologues of *PFLU1514*, for example *glnA* and *glnN* in the cyanobacterium *Synechocystis*, and *glnA1* in *Mycobacterium bovis*, have been shown to respond to the availability of nitrogen in the cells (Reyes et al. 1997); (Tripathi et al. 2013). Even though some of the genes identified in the transposon screen are involved in nitrogen metabolism, these result in carbon molecules which are used in numerous pathways.

Following the mariner transposon screen, a swarming motility assay confirmed that the regulation of *gabD* was responsible for the restoration of a Δ*rccA* swarming motility phenotype. This result suggests that *gabD* and *rccA* are linked, possibly in a similar way to how *rccA* and *rccR* are, although this is still not known. Alternatively, it could be that *gabD* and *rccR* have similar functions. *gabD* encodes for succinate-semialdehyde dehydrogenase I, which is important for the metabolism of nitrogen when its availability is limited and degrades succinate-semialdehyde into succinate (Schneider et al. 2002). This succinate can then be metabolised as part of the TCA cycle. This is similar to RccR which is now known to bind to genes from the glyoxylate shunt and the gluconeogenesis pathway (Campilongo et al. 2017). Therefore, both GabD and RccR have an impact on the levels of carbon in the cell, and it may be that in both cases RccA senses these changes and responds accordingly.

6.5. What role does *rccA* have?

From the work in this study and afterwards by Campilongo et al., the function of RccA has still not been discovered. However, at the beginning of this study, it was shown by the swarming motility assays that there was a functional link between *rccR* and *rccA*. Due to further interest in RccR and the mariner transposon screen providing data about *hexR* regulation when identifying genes linked to *rccA* and this being followed, the link between *rccR* and *rccA* was never fully established.

RccA is expected to be involved in c-di-GMP signalling. In this study, it has been shown that *rccA* is important in the colonisation of wheat roots in *P. fluorescens*. Another study using *P. fluorescens* SBW25 and wheat identified that the ribosomal modifying protein, RimK, binds to the c-di-GMP protein RimA and both individually affect the ability of *P. fluorescens* to

colonise wheat (Little et al. 2016). A third study using *P. fluorescens* SBW25 identified the *wss* operon as being important in the rhizosphere (Gal et al. 2003), and the *wss* operon is known to be regulated by the Wsp pathway and includes the c-di-GMP protein WspR (Spiers et al. 2003). In addition to this, the c-di-GMP proteins GacS, SadB and WspR in *P. fluorescens* F113 were shown to have an effect on the colonisation of alfalfa roots (Barahona et al. 2010). In 2007, it was shown that c-di-GMP may have a role in the colonisation of maize by *Pseudomonas putida* (Matilla et al. 2007). Due to the range of cellular functions that c-di-GMP is involved in, it is not surprising that it has since been found to be involved in plant colonisation as suggested by Matilla and colleagues. Previous studies have identified that c-di-GMP is involved in a number of cellular functions, for some recent reviews see Jenal et al. 2017; Sadiq et al. 2017 and Hengge, 2016. For example, it is known that c-di-GMP is known to affect the motility of species, for example *bifA* in *Pseudomonas aeruginosa* (Kuchma et al. 2007) and *phoB* in *Vibrio cholera* (Pratt et al. 2009).

The results from the mariner transposon screen in this study identified that signal transduction genes are involved in restoring the Δ*rccA* motility back to wild type. One of these genes, *PFLU5329*, is the orthologue of MorA. MorA is highly conserved among the *Pseudomonas* species however its function in *P. putida* and *P. aeruginosa* were shown to be different; in *P. putida*, MorA is important in the timing for flagella biogenesis as well as motility and biofilm formation, whereas in *P. aeruginosa* only biofilm formation is affected by MorA (Choy et al. 2004). Although MorA appears to function differently in each species, the difference in motility seen in *P. putida* could explain why *PFLU5329* was identified in the mariner transposon screen. In addition to this, this screen also identified *hexR* regulated loci therefore some *rccA* genes may also be associated with *hexR*. If RccA is involved in c-di-GMP signalling and even though the function of RccR is now known, it is still possible that RccR is able to bind to c-di-GMP as well. Previously, it has been shown by Hickman and Harwood that the transcriptional regulator FleQ functions as previously known beforehand as the activator of flagella biogenesis genes, but also binds to c-di-GMP (Hickman and Harwood, 2008).

As well as the c-di-GMP related domains, RccA also has two sensory PAS domains. It is possible that RccA uses these domains to sense changes in the environment, leading to further c-di-GMP mediated downstream events for the cell. As *rccR* is upstream of *rccA*, and the function of *rccR* has been identified (Campilongo et al. 2017), it is possible that *rccA* is

detecting changes in the carbon availability for the cell while *rccR* reacts to these changes. When the swarming motility assay done with differing carbon sources, varying results were seen on each of the carbon sources used; this may have been due to the PAS domains sensing the changes in the environment and this should be investigated further. Previous work has shown the importance of PAS domains in sensing environmental changes (Petrova and Sauer 2012). Consistent with this observation, a number of c-di-GMP related proteins have sensory domains as well as GGDEF and/or EAL domains (Mills et al. 2011).

Following the identification and insight into three metabolic genes as well as a further three identified from the mariner transposon screen, it could be proposed that RccA is involved in sensing the changes of substances within the cell. During the screen and in the absence of *rccA*, the metabolic genes identified may have altered their expression to restore the balance back to a wild type level.

Even though neither the GGDEF nor the EAL domain of RccA have as of yet been shown to synthesise or degrade c-di-GMP, this can not be ruled out. It has already been identified in Chapter 3 that in RccA the conserved GGDEF amino acids are actually composed of SGDEF. This however does not affect the ability of the diguanylate cyclase as it was shown that a SGDEF is active in *Pectobacterium atrosepticum* (Pérez -Mendoza et al. 2011)

Following the identification of the function of RccA, it will be important to identify how RccA is linked to RccR and HexR, as this study has shown it is linked to both of these proteins and there is some evidence that all three proteins are subject to compensatory regulation in vivo. For example, according to the Western blot analysis used to confirm the RccR polyclonal antibody, the Δ*rccAR* strain was still able to bind to the antibody, despite the deletion of *rccR* in this mutant. This could have been due to an interaction between HexR and RccA, and in the absence of RccA more HexR was expressed. If this was the case, the close similarity between RccR and HexR meant that cross reaction was seen, resulting in a detectable band in the Δ*rccAR* background. As well as this, the mariner transposon screen identified *hexR* loci and one of these was confirmed as being linked to *rccA* when the *hexR* mutated loci of *gabD* was responsible for the restoration of *rccA* swarming motility.

6.6. Applications

The work from this study has been continued and published, which is shown in Appendix A. The ability of *P. fluorescens* to colonise plants is of importance due to *P. fluorescens* being a PGPR and in this study both *rccR* and *rccA* were shown to have an effect on the ability to colonise plants, and later *hexR* by Campilongo et al. The discovery of *rccR* being involved in carbon metabolism is important as the environment in which *P. fluorescens* are naturally found varies, especially in the soil where plant colonisation will occur.

6.7. Future work

My focus was on understanding more about RccR and HexR. However, now that their functions have been determined in the publication in 2017 by Campilongo et al. (see Appendix A), it is still unknown how RccA fits into the picture, more work needs to be done to better understand this. For example:

- mutate the individual PAS domains and see the effects these have especially sensing the changes in carbon sources.
- determine if both or either the GGDEF or EAL domain of RccA is active and therefore synthesises or degrades c-di-GMP respectively.
- over-express and purify RccA in preparation for further biochemical analysis.
- use different protein-protein interaction techniques to discover other proteins that RccA interacts with. Including co-immunoprecipitation and the identification of the proteins through mass spectrometry. It would be interesting to see if these are c-di-GMP related like RccA itself is predicted to be, or involved in carbon metabolism like RccR and HexR.

References

Aleksandrzak-Piekarczyk T, Stasiak-Różańska L, Cieśla J, Bardowski J. (2015) ClaR – a novel key regulator of cellobiose and lactose metabolism in *Lactococcus lactis* IL1403. **Appl Microbiol Biotechnol 99(1):337-47**

Alsohim AS, Taylor TB, Barrett GA, Gallie J, Zhang XX, Altamirano-Junqueira AE, Johnson LJ, Rainey PB, Jackson RW. (2014) The biosurfactant viscosin produced by *Pseudomonas fluorescens* SBW25 aids spreading motility and plant growth promotion. **Environ Microbiol 16(7):2267-81**

Amikam D and Galperin MY (2006) PilZ domain is part of the bacterial c-di-GMP binding protein. **Bioinformatics 22(1):3-6**

Aragón IM, Pérez-Mendoza D, Gallegos MT, Ramos C. (2015) The c-di-GMP phosphodiesterase BifA is involved in the virulence of bacteria from the *Pseudomonas syringae* complex. **Mol Plant Pathol 16(6):604-15**

Arias-Barrau E, Olivera ER, Sandoval A, Naharro G, Luengo JM. (2006) Acetyl-CoA synthetase from *Pseudomonas putida* U is the only acyl-CoA activating enzyme induced by acetate in this bacterium. **FEMS Microbiol Lett 260(1):36-46**

Bailey TL and Elkan C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. **Proc Int Conf Intell Syst Mol Biol 2:28-36**

Baker KF and Cook RJ. (1974) Biological Control of Plant Pathogens. Freeman, San Francisco, CA

Barahona E, Navazo A, Yousef-Coronado F, Aguirre de Cárcer D, Martínez-Granero F, Espinosa-Urgel M, Martín M, Rivilla R. (2010) Efficient rhizosphere colonization by *Pseudomonas fluorescens* F113 mutants unable to form biofilms on abiotic surfaces. **Environ Microbiol 12(12):3185-95**

Bateman A. (1999) The SIS domain: a phosphosugar-binding domain. **Trends Biochem Sci 24(3):94-5**

Becker DF, Zhu W, Moxley MA. (2011) Flavin redox switching of protein functions. **Antioxid Redox Signal 14(6):1079-91**

Bellini D, Caly DL, McCarthy Y, Bumann M, An SQ, Dow JM, Ryan RP, Walsh MA. (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(1):26-38**

Berg G and Smalla K. (2009) Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. **FEMS Microbiol Ecol 68(1):1-13**

Berger S, Welte C, Deppenmeier U. (2012) Acetate activation in *Methanosaeta thermophila*: characterization of the key enzymes pyrophosphatase and acetyl-CoA synthetase. **Archaea 2012:315153**

Bhagirath AY, Somayajula D, Li Y, Duan K. (2017) CmpX affects virulence in *Pseudomonas aeruginosa* through the Gac/Rsm signaling pathway and by modulating c-di-GMP levels. **J Membr Biol**

Boyd CD, Chatterjee D, Sondermann H, O'Toole GA. (2012) LapG, required for modulating biofilm formation by *Pseudomonas fluorescens* Pf0-1, is a calcium-dependent protease. **J Bacteriol 194(16):4406-14**

Brautaset T, Lale R, Valla S. (2009) Positively regulated bacterial expression systems. **Microb Biotechnol 2(1):15-30**

Burr TJ, Schroth MN, Suslow T. (1973) Increased potato yields by treatment of seedpieces with specific strains of *Pseudomonas fluorescens* and *P. putida*. **Phytopathology 68:1377-1383**

But SY, Egorova SV, Khmelenina VN, Trotsenko YA. (2017) Biochemical properties and phylogeny of hydroxypyruvate reductases from methanotrophic bacteria with different C1-assimilation pathways. **Biochemistry (Mosc) 82(11):1295-1303**

Campilongo R, Fung RKY, Little RH, Grenga L, Trampari E, Pepe S, Chandra G, Stevenson CEM, Roncarati D, Malone JG. (2017) One ligand, two regulators and three binding sites: How KDPG controls primary carbon metabolism in *Pseudomonas*. **PLoS Genet 13(6):e1006839**

Cairns J, Frickel J, Jalasvuori M, Hiltunen T, Becks L. (2017) Genomic evolution of bacterial populations under coselection by antibiotics and phage. **Mol Ecol 26(7):1848-1859**

Chatterjee D, Cooley RB, Boyd CD, Mehl RA, O'Toole GA, Sondermann H. (2014) Mechanistic insight into the conserved allosteric regulation of periplasmic proteolysis by the signaling molecule cyclic-di-GMP. **Elife 3:e03650**

Chin-A-Woeng TFC, de Priester W, van der Bij A, Lugtenberg BJJ. (1997) Description of the colonization of a gnotobiotic tomato rhizosphere by *Pseudomonas fluorescens* biocontrol strain WCS365, using scanning electron microscopy. **Mol Plant Microbe Interact 10(1):79-86**

Choy WK, Zhou L, Syn CK, Zhang LH, Swarup S. (2004) MorA defines a new class of regulators affecting flagellar development and biofilm formation in diverse *Pseudomonas* species. **J Bacteriol 186(21):7221-8**

Christen M, Christen B, Folcher M, Schauerte A, Jenal U. (2005) Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. **J Biol Chem 280(35):30829-37**

Christen B, Christen M, Paul R, Schmid F, Folcher M, Jenoe P, Meuwly M, Jenal U. (2006) Allosteric control of cyclic di-GMP signaling. **J Biol Chem 281(42):32015-24**

Cohen D, Mechold U, Nevenzal H, Yarmiyhu Y, Randall TE, Bay DC, Rich JD, Parsek MR, Kaever V, Harrison JJ, Banin E. (2015) Oligoribonuclease is a central feature of cyclic diguanylate signaling in *Pseudomonas aeruginosa*. **Proc Natl Acad Sci USA 112(36):11359-64**

Couillerot O, Prigent-Combaret C, Caballero-Mellado J, Moënne-Loccoz Y. (2009) *Pseudomonas fluorescens* and closely-related fluorescent *pseudomonads* as biocontrol agents of soil-borne phytopathogens. **Lett Appl Microbiol 48(5):505-12.**

Crack JC, Munnoch J, Dodd EL, Knowles F, Al Bassam MM, Kamali S, Holland AA, Cramer SP, Hamilton CJ, Johnson MK, Thomson AJ, Hutchings MI, Le Brun NE. (2015) NsrR from *Streptomyces coelicolor* is a nitric oxide-sensing [4Fe-4S] cluster protein with a specialized regulatory function. **J Biol Chem 290(20):12689-704**

Daddaoua A, Krell T, Ramos JL. (2009) Regulation of glucose metabolism in *Pseudomonas*: the phosphorylative branch and Entner-Doudoroff enzymes are regulated by a repressor containing a sugar isomerase domain. **J Biol Chem 284(32):21360-8**

Damron FH, Qiu D, Yu HD. (2009) The *Pseudomonas aeruginosa* sensor kinase KinB negatively controls alginate production through AlgW-dependent MucA proteolysis. **J Bacteriol 191(7):2285-95**

del Castillo T, Duque E, Ramos JL. (2008) A set of activators and repressors control peripheral glucose pathways in *Pseudomonas putida* to yield a common central intermediate. **J Bacteriol 190(7):2331-9**

diCenzo GC, Muhammed Z, Østerås M, O'Brien SA, Finan TM. (2017) A key regulator of the glycolytic and gluconeogenic central metabolic pathways in *Sinorhizobium meliloti*. **Genetics 207(3):961-974**

Duerig A, Abel S, Folcher M, Nicollier M, Schwede T, Amiot N, Giese B, Jenal U. (2009) Second messenger-mediated spatiotemporal control of protein degradation regulates bacterial cell cycle progression. **Genes Dev 23(1):93-104**

Dunn MF, Ramírez-Trujillo JA, Hernández-Lucas I. (2009) Major roles of isocitrate lyase and malate synthase in bacterial and fungal pathogenesis. **Microbiology 155(Pt10):3166-75**

Edmunds AC, Castiblanco LF, Sundin GW, Waters CM. (2013) Cyclic di-GMP modulates the disease progression of *Erwinia amylovora*. **J Bacteriol 195(10):2155-65**

Entner N and Doudoroff M. (1952) Glucose and gluconic acid oxidation of *Pseudomonas saccharophila*. **J Biol Chem 196(2):853-62**

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. (2016) The Pfam protein families database: towards a more sustainable future. **Nucleic Acids Res 44(D1):D279-85**

Gal M, Preston GM, Massey RC, Spiers AJ, Rainey PB. (2003) Genes encoding a cellulosic polymer contribute toward the ecological success of *Pseudomonas fluorescens* SBW25 on plant surfaces. **Mol Ecol 12(11):3109-21**

Galperin MY, Natale DA, Aravind L, Koonin EV. (1999) A specialized version of the HD hydrolase domain implicated in signal transduction. **J Mol Microbiol Biotechnol 1(2):303-5**

Galperin MY. (2004) Bacterial signal transduction network in a genomic perspective. **Environ Microbiol 6(6):552-67**

Gao S, Romdhane SB, Beullens S, Kaever V, Lambrichts I, Fauvart M, Michiels J. (2014) Genomic analysis of cyclic-di-GMP-related genes in rhizobial type strains and functional analysis in *Rhizobium etli*. **Appl Microbiol Biotechnol 98(10):4589-602**

Garrido-Sanz D, Meier-Kolthoff JP, Göker M, Martín M, Rivilla R, Redondo-Nieto M. (2016) Genomic and genetic diversity within the *Pseudomonas fluorescens* complex. **PLoS One 11(2):e0150183**

Geddes BA and Oresnik IJ (2014) Physiology, genetics, and biochemistry of carbon metabolism in the alphaproteobacterium *Sinorhizobium meliloti*. **Can J Microbiol 60(8):491-507**

Gomelsky M and Klug G. (2002) BLUF: a novel FAD-binding domain involved in sensory transduction in microorganisms. **Trends Biochem Sci 27(10):497-500**

Gust B, Challis GL, Fowler K, Kieser T, Chater KF. (2003) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. **Proc Natl Acad Sci USA 100(4):1541-6**

Haas D and Defago G. (2005) Biological control of soil-borne pathogens by fluorescent *pseudomonads*. **Nat Rev Microbiol 3(4):307-19**

Hager PW, Calfee MW, Phibbs PV. (2000) The *Pseudomonas aeruginosa devB*/SOL homolog, *pgl*, is a member of the *hex* regulon and encodes 6-phosphogluconolactonase. **J Bacteriol 182(14):3934-41**

Haldimann A, Prahalad MK, Fisher SL, Kim SK, Walsh CT, Wanner BL. (1996) Altered recognition mutants of the response regulator PhoB: a new genetic strategy for studying protein-protein interactions. **Proc Natl Acad Sci USA 93(25):14361-6**

Hammer BK and Bassler BL. (2009) Distinct sensory pathways in *Vibrio cholerae* El Tor and classical biotypes modulate cyclic dimeric GMP levels to control biofilm formation. **J Bacteriol 191(1):169-77**

Hartmann A, Schmid M, van Tuinen D, Berg G. (2009) Plant-driven selection of microbes. **Plant Soil 321:235–257**

Hay ID, Remminghorst U, Rehm BH. (2009) MucR, a novel membrane-associated regulator of alginate biosynthesis in *Pseudomonas aeruginosa*. **Appl Environ Microbiol 75(4):1110-20**

He YW, Ng AY, Xu M, Lin K, Wang LH, Dong YH, Zhang LH. (2007) *Xanthomonas campestris* cell-cell communication involves a putative nucleotide receptor protein Clp and a hierarchical signalling network. **Mol Microbiol 64(2):281-92**

Heeb S, Itoh Y, Nishijyo T, Schnider U, Keel C, Wade J, Walsh U, O'Gara F, Haas D. (2000) Small, stable shuttle vectors based on the minimal pVS1 replicon for use in gram-negative, plant-associated bacteria. **Mol Plant Microbe Interact 13(2):232-7**

Hengge R. (2009) Principles of c-di-GMP signalling in bacteria. **Nat Rev Microbiol 7(4):263-73**

Hengge R. (2016) Trigger phosphodiesterases as a novel class of c-di-GMP effector proteins. **Philos Trans R Soc Lond B Biol Sci 371(1707)**

Hickman JW and Harwood CS. (2008) Identification of FleQ from *Pseudomonas aeruginosa* as a c-di-GMP-responsive transcription factor. **Mol Microbiol 69(2):376-89**

Hiltner L. (1904) Über neuere erfahrungen und probleme auf dem gebiete der bodenbakteriologie unter besonderer berücksichtigung der gründüngung und der brache. **Arb DLG 98: 59–78**

Hmelo LR, Borlee BR, Almblad H, Love ME, Randall TE, Tseng BS, Lin C, Irie Y, Storek KM, Yang JJ, Siehnel RJ, Howell PL, Singh PK, Tolker-Nielsen T, Parsek MR, Schweizer HP, Harrison JJ. (2015) Precision-engineering the *Pseudomonas aeruginosa* genome with two-step allelic exchange. **Nat Protoc 10(11):1820-41**

Ho CL, Noji M, Saito M, Saito K. (1999) Regulation of serine biosynthesis in *Arabidopsis*. Crucial role of plastidic 3-phosphoglycerate dehydrogenase in non-photosynthetic tissues. **J Biol Chem 274(1):397-402**

Hoang TT, Karkhoff-Schweizer RR, Kutchma AJ, Schweizer HP. (1998) A broad-host-range Flp-FRT recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked *Pseudomonas aeruginosa* mutants. **Gene 212(1):77-86**

Ishihama A. (2012) Prokaryotic genome regulation: a revolutionary paradigm. **Proc Jpn Acad Ser B Phys Biol Sci 88(9):485-508**

Jacoby R, Peukert M, Succurro A, Koprivova A, Kopriva S. (2017) The role of soil microorganisms in plant mineral nutrition – current knowledge and future directions. **Front Plant Sci 8:1617**

Jaeger T and Mayer C. (2008) The transcriptional factors MurR and catabolite activator protein regulate N-acetylmuramic acid catabolism in *Escherichia coli*. **J Bacteriol 190(20):6598-608**

Jenal U, Reinders A, Lori C. (2017) Cyclic di-GMP: second messenger extraordinaire. **Nat Rev Microbiol 15(5):271-284**

Jiménez-Fernández A, López-Sánchez A, Calero P, Govantes F. (2015) The c-di-GMP phosphodiesterase BifA regulates biofilm development in *Pseudomonas putida*. **Environ Microbiol Rep 7(1):78-84**

Johnston JW, Shamsulddin H, Miller AF, Apicella MA. (2010) Sialic acid transport and catabolism are cooperatively regulated by SiaR and CRP in nontypeable *Haemophilus influenzae*. **BMC Microbiol 10:240**

Jojima T and Inui M. (2015) Engineering the glycolytic pathway: A potential approach for improvement of biocatalyst performance. **Bioengineered 6(6):328-34**

Jones CJ, Utada A, Davis KR, Thongsomboon W, Zamorano Sanchez D, Banakar V, Cegelski L, Wong GC, Yildiz FH. (2015) C-di-GMP regulates motile to sessile transition by modulating MshA pili biogenesis and near-surface motility behavior in *Vibrio cholerae*. **PLoS Pathog 11(10):e1005068**

Kim MJ, Kim J, Lee HY, Noh HJ, Lee KH, Park SJ. (2015) Role of AcsR in expression of the acetyl-CoA synthetase gene in *Vibrio vulnificus*. **BMC Microbiol 15:86**

Kirner S, Hammer PE, Hill DS, Altmann A, Fischer I, Weislo LJ, Lanahan M, van Pée KH, Ligon JM. (1998) Functions encoded by pyrrolnitrin biosynthetic genes from *Pseudomonas fluorescens*. **J Bacteriol 180(7):1939-43**

Kohno K, Yasuzawa K, Hirose M, Kano Y, Goshima N, Tanaka H, Imamoto F. (1994) Autoregulation of transcription of the *hupA* gene in *Escherichia coli*: evidence for sterichindrance of the functional promoter domains induced by HU. **J Biochem 115(6):1113-8**

Kohler PR, Choong EL, Rossbach S. (2011) The RpiR-like repressor IolR regulates inositol catabolism in *Sinorhizobium meliloti*. **J Bacteriol 193(19):5155-63**

Kondrashov FA, Koonin EV, Morgunov IG, Finogenova TV, Kondrashova MN. (2006) Evolution of glyoxylate cycle enzymes in Metazoa: evidence of multiple horizontal transfer events and pseudogene formation. **Biol Direct 1:31**

Kornberg HL and Krebs HA. (1957) Synthesis of cell constituents from C2-units by a modified tricarboxylic acid cycle. **Nature 179(4568):988-91**

Kornberg HL and Madsen NB. (1958) The metabolism of C_2 compounds in microorganisms. 3. Synthesis of malate from acetate via the glyoxylate cycle. **J Biochem 68(3):549-57**

Kornberg HL (1966) The role and control of the glyoxylate cycle in *Escherichia coli*. **J Biochem 99(1):1-11**

Koza A, Kusmierska A, McLaughlin K, Moshynets O, Spiers AJ. (2017) Adaptive radiation of *Pseudomonas fluorescens* SBW25 in experimental microcosms provides an understanding of the evolutionary ecology and molecular biology of A-L interface biofilm. **FEMS Microbiol Lett 364(12)**

Kuchma SL, Brothers KM, Merritt JH, Liberati NT, Ausubel FM, O'Toole GA. (2007) BifA, a cyclic-di-GMP phosphodiesterase, inversely regulates biofilm formation and swarming motility by *Pseudomonas aeruginosa* PA14. **J Bacteriol 189(22):8165-78**

Kuras L and Struhl K. (1999) Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme. **Nature 399(6736):609-13**

Kwak YS and Weller DM. (2013) Take-all of wheat and natural disease suppression: A Review. **J Plant Pathol 29(2):125-35**

Lampe DJ, Akerley BJ, Rubin EJ, Mekalanos JJ, Robertson HM. (1999) Hyperactive transposase mutants of the Himar1 mariner transposon. **Proc Natl Acad Sci U S A. 96(20):11428-33.**

Landa BB, Mavrodi OV, Schroeder KL, Allende-Molar R, Weller DM. (2006) Enrichment and genotypic diversity of *phlD*-containing fluorescent *Pseudomonas* spp. in two soils after a century of wheat and flax monoculture. **FEMS Microbiol Ecol 55(3):351-68**

Leyn SA, Li X, Zheng Q, Novichkov PS, Reed S, Romine MF, Fredrickson JK, Yang C, Osterman AL, Rodionov DA. (2011) Control of proteobacterial central carbon metabolism by the HexR transcriptional regulator: a case study in *Shewanella oneidensis*. **J Biol Chem 286(41):35782-94**

Little RH, Grenga L, Saalbach G, Howat AM, Pfeilmeier S, Trampari E, Malone JG. (2016) Adaptive remodeling of the bacterial proteome by specific ribosomal modification regulates *Pseudomonas* infection and niche colonisation. **PLoS Genet 12(2):e1005837**

Liu Y, Gokhale CS, Rainey PB, Zhang XX. (2017) Unravelling the complexity and redundancy of carbon catabolic repression in *Pseudomonas fluorescens* SBW25. **Mol Microbiol 105(4):589-605**

Loper JE, Kobayashi DY, Paulsen IT. (2007) The genomic sequence of *Pseudomonas fluorescens* Pf-5: Insights into biological control. **Phytopathology 97(2):233-8**

Loper JE, Hassan KA, Mavrodi DV, Davis EW 2nd, Lim CK, Shaffer BT, Elbourne LD, Stockwell VO, Hartney SL, Breakwell K, Henkels MD, Tetu SG, Rangel LI, Kidarsa TA, Wilson NL, van de Mortel JE, Song C, Blumhagen R, Radune D, Hostetler JB, Brinkac LM, Durkin AS, Kluepfel DA, Wechter WP, Anderson AJ, Kim YC, Pierson LS 3rd, Pierson EA, Lindow SE, Kobayashi DY, Raaijmakers JM, Weller DM, Thomashow LS, Allen AE, Paulsen IT. (2012) Comparative genomics of plant-associated *Pseudomonas* spp.: insights into diversity and inheritance of traits involved in multitrophic interactions. **PLoS Genet 8(7):e1002784**

Lori C, Ozaki S, Steiner S, Böhm R, Abel S, Dubey BN, Schirmer T, Hiller S, Jenal U. (2015) Cyclic di-GMP acts as a cell cycle oscillator to drive chromosome replication. **Nature 523(7559):236-9**

Lugtenberg BJ, Dekkers L, Bloemberg GV. (2001) Molecular determinants of rhizosphere colonization by *Pseudomonas*. **Annu Rev Phytopathol 39:461-90**

Lugtenberg B and Kamilova F. (2009) Plant-growth-promoting rhizobacteria. **Annu Rev Microbiol 63:541-56.**

Mahmood T, Yang PC. (2012) Western Blot: Technique, Theory, and Trouble Shooting **N Am J Med Sci 4(9):429-34**

Malone JG, Williams R, Christen M, Jenal U, Spiers AJ, Rainey PB. (2007) The structure-function relationship of WspR, a *Pseudomonas fluorescens* response regulator with a GGDEF output domain. **Microbiology 153(Pt 4):980-94**

Malone JG, Jaeger T, Spangler C, Ritz D, Spang A, Arrieumerlou C, Kaever V, Landmann R, Jenal U. (2010) YfiBNR mediates cyclic di-GMP dependent small colony variant formation and persistence in *Pseudomonas aeruginosa*. **PLoS Pathog 6(3):e1000804**

Martínez-Granero F, Navazo A, Barahona E, Redondo-Nieto M, González de Heredia E, Baena I, Martín-Martín I, Rivilla R, Martín M. (2014) Identification of *flgZ* as a flagellar gene encoding a PilZ domain protein that regulates swimming motility and biofilm formation in *Pseudomonas*. **PLoS One 9(2):e87608**

Martínez-Gil M, Ramos-González MI, Espinosa-Urgel M. (2014) Roles of cyclic di-GMP and the Gac system in transcriptional control of the genes coding for the *Pseudomonas putida* adhesins LapA and LapF. **J Bacteriol 196(8):1484-95**

Matilla MA, Espinosa-Urgel M, Rodríguez-Herva JJ, Ramos JL, Ramos-González MI. (2007) Genomic analysis reveals the major driving forces of bacterial life in the rhizosphere. **Genome Biol 8(9):R179**

Matilla MA, Travieso ML, Ramos JL, Ramos-González MI. (2011) Cyclic diguanylate turnover mediated by the sole GGDEF/EAL response regulator in *Pseudomonas putida*: its role in the rhizosphere and an analysis of its target processes. **Environ Microbiol 13(7):1745-66**

Mauchline TH and Malone JG. (2017) Life in earth - the root microbiome to the rescue? **Curr Opin Microbiol 37:23-28**

Mavrodi OV, Mavrodi DV, Parejko JA, Thomashow LS, Weller DM. (2012) Irrigation differentially impacts populations of indigenous antibiotic-producing *pseudomonas* spp. in the rhizosphere of wheat. **Appl Environ Microbiol 78(9):3214-20**

Mehmood A, Abdallah K, Khandekar S, Zhurina D, Srivastava A, Al-Karablieh N, Alfaro-Espinoza G, Pletzer D, Ullrich MS. (2015) Expression of extra-cellular levansucrase in *Pseudomonas syringae* is controlled by the in planta fitness-promoting metabolic repressor HexR. **BMC Microbiol 15:48**

Meinhardt S, Manley MW Jr, Becker NA, Hessman JA, Maher LJ 3rd, Swint-Kruse L. (2012) Novel insights from hybrid LacI/GalR proteins: family-wide functional attributes and biologically significant variation in transcription repression. **Nucleic Acids Res 40(21):11139-54**

Mendes R, Kruijt M, de Bruijn I, Dekkers E, van der Voort M, Schneider JHM, Piceno YM, DeSantis TZ, Andersen GL, Bakker PA, Raaijmakers JM. (2011) Deciphering the rhizosphere microbiome for disease-suppressive bacteria. **Science 332(6033):1097-100**

Mills E, Pultz IS, Kulasekara HD, Miller SI. (2011) The bacterial second messenger c-di-GMP: mechanisms of signalling. **Cell Microbiol 13(8):1122-9**

Miner KD, Klose KE, Kurtz DM Jr. (2013) An HD-GYP cyclic di-guanosine monophosphate phosphodiesterase with a non-heme diiron-carboxylate active site. **Biochemistry 52(32):5329-31**

Moënne-Loccoz, Y, Powell J, Higgins P, McCarthy J, O'Gara F. (1998) An investigation of the impact of biocontrol *Pseudomonas fluorescens* F113 on the growth of sugarbeet and the performance of subsequent clover-*Rhizobium* symbiosis. **Appl Soil Ecol (7):225–237**

Muñoz-Bertomeu J, Anoman A, Flores-Tornero M, Toujani W, Rosa-Téllez S, Fernie AR, Roje S, Segura J, Ros R. (2013) The essential role of the phosphorylated pathway of serine biosynthesis in *Arabidopsis*. **Plant Signal Behav 8(11):e27104**

Naseby DC, Way JA, Bainton NJ, Lynch JM. (2001) Biocontrol of *Pythium* in the pea rhizosphere by antifungal metabolite producing and non-producing *Pseudomonas* strains. **J Appl Microbiol 90(3):421-9**

Navazo A, Barahona E, Redondo-Nieto M, Martínez-Granero F, Rivilla R, Martín M. (2009) Three independent signalling pathways repress motility in *Pseudomonas fluorescens* F113. **Microb Biotechnol 2(4):489-98**

Newell PD, Monds RD, O'Toole GA. (2009) LapD is a bis-(3',5')-cyclic dimeric GMP-binding protein that regulates surface attachment by *Pseudomonas fluorescens* Pf0-1. **Proc Natl Acad Sci USA 106(9):3461-6**

Nguyen DD, Melnik AV, Koyama N, Lu X, Schorn M, Fang J, Aguinaldo K, Lincecum TL Jr, Ghequire MG, Carrion VJ, Cheng TL, Duggan BM, Malone JG, Mauchline TH, Sanchez LM, Kilpatrick AM, Raaijmakers JM, Mot R, Moore BS, Medema MH, Dorrestein PC. (2016) Indexing the *Pseudomonas* specialized metabolome enabled the discovery of poaeamide B and the bananamides. **Nat Microbiol 2:16197**

Nowak-Thompson B, Chaney N, Wing JS, Gould SJ, Loper JE. (1999) Characterization of the pyoluteorin biosynthetic gene cluster of *Pseudomonas fluorescens* Pf-5. **J Bacteriol 181(7):2166-74**

Orr MW, Donaldson GP, Severin GB, Wang J, Sintim HO, Waters CM, Lee VT. (2015) Oligoribonuclease is the primary degradative enzyme for pGpG in *Pseudomonas aeruginosa* that is required for cyclic-di-GMP turnover. **Proc Natl Acad Sci USA 112(36):e5048-57**

O'Sullivan DJ and O'Gara F. (1992) Traits of fluorescent *Pseudomonas* spp. involved in suppression of plant root pathogens. **Microbiol Rev 56(4):662-76**

Ozaki S, Schalch-Moser A, Zumthor L, Manfredi P, Ebbensgaard A, Schirmer T, Jenal U. (2014) Activation and polar sequestration of PopA, a c-di-GMP effector protein involved in *Caulobacter crescentus* cell cycle control. **Mol Microbiol 94(3):580-94**

Papagianni M. (2012) Recent advances in engineering the central carbon metabolism of industrially important bacteria. **Microb Cell Fact 11:50**

Paul R, Weiser S, Amiot NC, Chan C, Schirmer T, Giese B, Jenal U. (2004) Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. **Genes Dev 18(6):715-27**

Pérez-Mendoza D, Coulthurst SJ, Humphris S, Campbell E, Welch M, Toth IK, Salmond GP (2011) A multi-repeat adhesin of the phytopathogen, *Pectobacterium atrosepticum*, is secreted by a Type I pathway and is subject to complex regulation involving a non-canonical diguanylate cyclase. **Mol Microbiol 82(3):719-33**

Petrova OE and Sauer K. (2012) PAS domain residues and prosthetic group involved in BdlA-dependent dispersion response by *Pseudomonas aeruginosa* biofilms. **J Bacteriol 194(21):5817-28**

Phue JN, Noronha SB, Hattacharyya R, Wolfe AJ, Shiloach J. (2005) Glucose metabolism at high density growth of *E. coli* B and *E. coli* K: differences in metabolic pathways are responsible for efficient glucose utilization in *E. coli* B as determined by microarrays and Northern blot analyses. **Biotechnol Bioeng 90(7):805-20**

Pratt JT, McDonough E, Camilli A. (2009) PhoB regulates motility, biofilms, and cyclic di-GMP in *Vibrio cholerae*. **J Bacteriol 191(21):6632-42**

Rainey PB and Bailey MJ, (1996) Physical and genetic map of the *Pseudomonas fluorescens* SBW25 chromosome. **Mol Microbiol 19(3):521-33**

Renilla S, Bernal V, Fuhrer T, Castaño-Cerezo S, Pastor JM, Iborra JL, Sauer U, Cánovas M. (2012) Acetate scavenging activity in *Escherichia coli*: interplay of acetyl-CoA synthetase and the PEP-glyoxylate cycle in chemostat cultures. **Appl Microbiol Biotechnol 93(5):2109-24**

Reyes JC, Muro-Pastor MI, Florencio FJ. (1997) Transcription of glutamine synthetase genes (*glnA* and *glnN*) from the cyanobacterium *Synechocystis* sp. strain PCC 6803 is differently regulated in response to nitrogen availability. **J Bacteriol 179(8):2678-89**

Ribbe J, Baker AE, Euler S, O'Toole GA, Maier B. (2017) Role of cyclic di-GMP and exopolysaccharide in Type IV pilus dynamics. **J Bacteriol 199(8)**

Römling U, Galperin MY, Gomelsky M. (2013) Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. **Microbiol Mol Biol Rev 77(1):1-52**

Ross P, Weinhouse H, Aloni Y, Michaeli D, Weinberger-Ohana P, Mayer R, Braun S, de Vroom E, van der Marel GA, van Boom JH, Benziman M. (1987) Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. **Nature 325(6101):279-81**

Ruffner B, Péchy-Tarr M, Ryffel F, Hoegger P, Obrist C, Rindlisbacher A, Keel C, Maurhofer M. (2013) Oral insecticidal activity of plant-associated *pseudomonads*. **Environ Microbiol 15(3):751-63**

Ruffner B, Péchy-Tarr M, Höfte M, Bloemberg G, Grunder J, Keel C, Maurhofer M. (2015) Evolutionary patchwork of an insecticidal toxin shared between plant-associated *pseudomonads* and the insect pathogens *Photorhabdus* and *Xenorhabdus*. **BMC Genomics 16:609**

Ryan RP, Lucey J, O'Donovan K, McCarthy Y, Yang L, Tolker-Nielsen T, Dow JM. (2009) HD-GYP domain proteins regulate biofilm formation and virulence in *Pseudomonas aeruginosa*. **Environ Microbiol 11(5):1126-36**

Ryan RP, Tolker-Nielsen T, Dow JM. (2012) When the PilZ don't work: effectors for cyclic di-GMP action in bacteria. **Trends Microbiol 20(5):235-42**

Ryjenkov DA, Simm R, Römling U, 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(41):30310-4**

Sadiq FA, Flint S, Li Y, Liu T, Lei Y, Sakandar HA, He G. (2017) New mechanistic insights into the motile-to-sessile switch in various bacteria with particular emphasis on *Bacillus subtilis* and *Pseudomonas aeruginosa*: a review. **Biofouling 33(4):306-326**

Sambrook J, Russel DW. (2001) Molecular cloning. A laboratory manual. 3rd ed. ed: **Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.**

Sanguin H, Kroneinsen L, Gazengel K, Kyselkova,M, Remenant B, Prigent-Combaret C, Grundmann GL, SarniguetA et al. (2008) Development of a 16S rRNA microarray approach for the monitoring of rhizosphere *Pseudomonas* populations associated with the decline of take-all disease of wheat. **Soil Biol Biochem 40:1028–1039**

Sarenko O, Klauck G, Wilke FM, Pfiffer V, Richter AM, Herbst S, Kaever V, Hengge R. (2017) More than enzymes that make or break cyclic di-GMP-local signaling in the interactome of GGDEF/EAL domain proteins of *Escherichia coli*. **MBio 8(5)**

Sauer U and Eikmanns BJ. (2005) The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. **FEMS Microbiol Rev 29(4):765-94**

Schmidt AJ, Ryjenkov DA, Gomelsky M. (2005) The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains. **J Bacteriol 187(14):4774-81**

Schneider BL, Ruback S, Kiupakis AK, Kasbarian H, Pybus C, Reitzer L. (2002) The *Escherichia coli gabDTPC* operon: specific gamma-aminobutyrate catabolism and nonspecific induction. **J Bacteriol 184(24):6976-86**

Schröder O and Wagner R. (2000) The bacterial DNA-binding protein H-NS represses ribosomal RNA transcription by trapping RNA polymerase in the initiation complex. **J Mol Biol 298(5):737-48**

Schumacher MA, Zeng W, Findlay KC, Buttner MJ, Brennan RG, Tschowri N. (2017) The *Streptomyces* master regulator BldD binds c-di-GMP sequentially to create a functional BldD2-(c-di-GMP)4 complex. **Nucleic Acids Res 45(11):6923-6933**

Scott TA, Heine D, Qin Z, Wilkinson B. (2017) An L-threonine transaldolase is required for L-threo-β-hydroxy-α-amino acid assembly during obafluorin biosynthesis. **Nat Commun. 8:15935**

Shimada T, Yamazaki Y, Tanaka K, Ishihama A. (2014) The whole set of constitutive promoters recognized by RNA polymerase RpoD holoenzyme of *Escherichia coli*. **PLoS One 9(3):e90447**

Silby MW, Cerdeño-Tárraga AM, Vernikos GS, Giddens SR, Jackson RW, Preston GM, Zhang XX, Moon CD, Gehrig SM, Godfrey SA, Knight CG, Malone JG, Robinson Z, Spiers AJ, Harris S, Challis GL, Yaxley AM, Harris D, Seeger K, Murphy L, Rutter S, Squares R, Quail MA, Saunders E, Mavromatis K, Brettin TS, Bentley SD, Hothersall J, Stephens E, Thomas CM, Parkhill J, Levy SB, Rainey PB, Thomson NR. (2009) Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. **Genome Biol 10(5):R51**

Simm R, Morr M, Kader A, Nimtz M, Römling U. (2004) GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. **Mol Microbiol 53(4):1123-34**

Slatko BE, Albright LM, Tabor S, Ju J. (2001) DNA sequencing by the dideoxy method. **Curr Protoc Mol Biol. May;Chapter 7 doi: 10.1002/0471142727.mb0704as47**

Sørensen KI and Hove-Jensen B. (1996) Ribose catabolism of *Escherichia coli*: characterization of the *rpiB* gene encoding ribose phosphate isomerase B and of the *rpiR* gene, which is involved in regulation of *rpiB* expression. **J Bacteriol 178(4):1003-11**

Spiers AJ, Bohannon J, Gehrig SM, Rainey PB. (2003) Biofilm formation at the air-liquid interface by the *Pseudomonas fluorescens* SBW25 wrinkly spreader requires an acetylated form of cellulose. **Mol Microbiol 50(1):15-27**

Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, Link KH, Breaker RR. (2008) Riboswitches in eubacteria sense the second messenger cyclic di-GMP. **Science 321(5887):411-3**

Tal R, Wong HC, Calhoon R, Gelfand D, Fear AL, Volman G, Mayer R, Ross P, Amikam D, Weinhouse H, Cohen A, Sapir S, Ohana P, Benziman M. (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(17):4416-25**

Tao F, He YW, Wu DH, Swarup S, Zhang LH. (2010) The cyclic nucleotide monophosphate domain of *Xanthomonas campestris* global regulator Clp defines a new class of cyclic di-GMP effectors. **J Bacteriol 192(4):1020-9**

Tao J, Li C, Luo C, He C. (2014) RavA/RavR two-component system regulates *Xanthomonas campestris* pathogenesis and c-di-GMP turnover. **FEMS Microbiol Lett 358(1):81-90**

Tchigvintsev A, Xu X, Singer A, Chang C, Brown G, Proudfoot M, Cui H, Flick R, Anderson WF, Joachimiak A, Galperin MY, Savchenko A, Yakunin AF. (2010) Structural insight into the mechanism of c-di-GMP hydrolysis by EAL domain phosphodiesterases. **J Mol Biol 402(3):524-38**

Teintze M, Hossain MB, Barnes CL, Leong J, van der Helm D. (1981) Structure of ferric pseudobactin, a siderophore from a plant growth promoting *Pseudomonas*. **Biochemistry 20(22):6446-57**

Teschler JK, Cheng AT, Yildiz FH. (2017) The two-component signal transduction system VxrAB positively regulates *Vibrio cholera* biofilm formation. **J Bacteriol 199(18)**

Tkacz A, Cheema J, Chandra G, Grant A, Poole PS. (2015) Stability and succession of the rhizosphere microbiota depends upon plant type and soil composition. **J ISME 9(11):2349-59**

Tripathi D, Chandra H, Bhatnagar R. (2013) Poly-L-glutamate/glutamine synthesis in the cell wall of *Mycobacterium bovis* is regulated in response to nitrogen availability. **BMC Microbiol 13:226**

Turnbull GA, Morgan JA, Whipps JM, Saunders JR. (2001) The role of bacterial motility in the survival and spread of *Pseudomonas fluorescens* in soil and in the attachment and colonisation of wheat roots. **FEMS Microbiol Ecol 36(1):21-31**

Vacheron J, Desbrosses G, Bouffaud ML, Touraine B, Moënne-Loccoz Y, Muller D, Legendre L, Wisniewski-Dyé F, Prigent-Combaret C. (2013) Plant growth-promoting rhizobacteria and root system functioning. **Front Plant Sci 4:356**

Valgepea K, Adamberg K, Nahku R, Lahtvee PJ, Arike L, Vilu R. (2010) Systems biology approach reveals that overflow metabolism of acetate in *Escherichia coli* is triggered by carbon catabolite repression of acetyl-CoA synthetase. **BMC Syst Biol 4:166**

van Hijum SA, Medema MH, Kuipers OP. (2009) Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. **Microbiol Mol Biol Rev 73(3):481-509**

Vogt JH and Schippers JH. (2015) Setting the PAS, the role of circadian PAS domain proteins during environmental adaptation in plants. **Front Plant Sci 6:513**

Voisard C, Bull CT, Keel C, Laville J, Maurhofer M. (1994) Biocontrol of root diseases by *Pseudomonas fluorescens* CHA0: current concepts and experimental approaches. **O'Gara F, Dowling DN, Boesten B (eds) Molecular Ecology of Rhizosphere Microorganisms 1994:67-89.**

Weinhouse H, Sapir S, Amikam D, Shilo Y, Volman G, Ohana P, Benziman M. (1997) c-di-GMP-binding protein, a new factor regulating cellulose synthesis in *Acetobacter xylinum*. **FEBS Lett 416(2):207-11**

Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FS. (2016) Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. **Nucleic Acids Res 44(D1):D646-53**

Woodcock DM, Crowther PJ, Doherty J, Jefferson S, DeCruz E, Noyer-Weidner M, Smith SS, Michael MZ, Graham MW. (1989) Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. **Nucleic Acids Res 17(9):3469-78**

Winkler WC, Breaker RR. (2005) Regulation of bacterial gene expression by riboswitches. **Annu Rev Microbiol. 59:487-517**

Yamamoto H, Serizawa M, Thompson J, Sekiguchi J. (2001) Regulation of the *glv* operon in *Bacillus subtilis*: YfiA (GlvR) is a positive regulator of the operon that is repressed through CcpA and cre. **J Bacteriol 183(17):5110-21**

Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL. (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. **Proc Natl Acad Sci USA 97(11):5978-83**

Yurgel SN, Rice J, Kahn ML. (2013) Transcriptome analysis of the role of GlnD/GlnBK in nitrogen stress adaptation by *Sinorhizobium meliloti* Rm1021. **PLoS One 8(3):e58028**

Zhang XX and Rainey PB. (2007) Construction and validation of a neutrally-marked strain of *Pseudomonas fluorescens* SBW25. **J Microbiol Methods 71(1):78-81**

Appendix A

This appendix provides the publication of the work from this study and the addition work afterwards. Campilongo et al. PLoS Genetics 2017