The role of post-transcriptional regulators in pathogenesis and secondary metabolite production in *Serratia* sp. ATCC 39006.

Nabil M. Wilf

Department of Biochemistry

St. John's College, University of Cambridge

March 2011

This dissertation is submitted for the degree of Doctor of Philosophy

The role of post-transcriptional regulators in pathogenesis and secondary metabolite production in *Serratia* sp. ATCC 39006.

Nabil M. Wilf

Summary

Serratia sp. ATCC 39006 (S39006) is a Gram-negative bacterium that is virulent in plant (potato) and animal (*Caenorhabditis elegans*) models. It produces two secondary metabolite antibiotics, prodigiosin and a carbapenem, and the plant cell wall degrading exoenzymes, pectate lyase and cellulase. A complex regulatory network controls production of prodigiosin, including a quorum sensing (QS) system, and the role of post-transcriptional regulation was investigated. It was hypothesized that Hfq-dependent small regulatory RNAs (sRNAs) might also play a role. Hfq is an RNA chaperone involved in post-transcriptional regulation that plays a key role in stress response and virulence in other bacterial species. An S39006 Δ*hfq* mutant was constructed and in the mutants production of prodigiosin and carbapenem was abolished, while production of the QS molecule, butanoyl homoserine lactone (BHL), was unaffected. Using transcriptional fusions, it was found that Hfq regulated the QS response regulators, SmaR and CarR. Additionally, exoenzyme production and swimming motility were decreased in the Δ*hfq* mutant, and virulence was attenuated in potato and *C. elegans*. It was also shown that the phenotype of an *hfq* mutant is independent of its role in regulating the stationary phase sigma factor, *rpoS*. In order to define the complete regulon of Hfq and identify relevant potential sRNAs, deep sequencing of strand-specific cDNAs (RNA-seq) was used to analyse the whole transcriptome of S39006 WT and the Δ*hfq* mutant. The regulon of another post-transcriptional regulator, RsmA, also involved in regulating prodigiosin production, was investigated by performing RNA-seq on an *rsmA* mutant. Moreover, global changes in the proteome of the *hfq* mutant was analysed using an LC-MS/MS approach with isobaric tags for relative and absolute quantification (iTRAQ). This study confirms a role for Hfq in pathogenesis and the regulation of antibiotic production in S39006, and begins to provide a systems-level understanding of Hfq and RsmA regulation using a combination of transcriptomics and proteomics.

Declaration

This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text.

No part of this dissertation has been submitted for any other degree or qualification. The length of this thesis does not exceed the limit set by the Biology degree committee.

This research project was performed in the laboratory of Professor George Salmond, Department of Biochemistry at the University of Cambridge, except for specific elements of the work that were performed elsewhere as described herein, during the period January 2008 to January 2011.

Nabil M. Wilf
March 2011

To my grandfather,
Muhammad Said Fikri,
And my parents.

"We should take care not to make the intellect our god;
it has, of course, powerful muscles, but no personality."
-Albert Einstein

"The third Tajallí is concerning arts, crafts and sciences. Knowledge is as wings to man's life, and a ladder for his ascent. Its acquisition is incumbent upon everyone. The knowledge of such sciences, however, should be acquired as can profit the peoples of the earth. Great indeed is the claim of scientists and craftsmen on the peoples of the world. In truth, knowledge is a veritable treasure for man, and a source of glory, of bounty, of joy, of exaltation, of cheer and gladness unto him."
-Bahá'u'lláh

Acknowledgements

I give Thee thanks, O my Lord, for having enabled me to travel to this land and engage in the study of sciences. Noble hast Thou created humanity, O my God. I implore Thee to assist me and to confirm my efforts to rise unto that for which I wast created, so to serve mankind and carry forward an ever-advancing civilization. O Thou the Glory of the Most Glorious! Yá Bahá'u'l-Abhá!

I would like to thank my supervisor Professor George Salmond for his enthusiasm, guidance, and the opportunity to work on this project. I would like to especially thank Dr. Neil Williamson and Dr. Josh Ramsay who provided invaluable support and assistance. I thank Simon Poulter for providing strain SP21 for experimental use. I thank Nicola Darling who identified the *rpoS* mutant in a transposon mutagenesis screen, which was investigated in this work. I would also like to thank all other members of the Salmond, Welch, and Hong groups for advice, useful discussions, and providing an enjoyable atmosphere in which to work.

I thank Prof. Ben Luisi for critical reading of the manuscript on the S39006 *hfq* mutant and Kasia Bandyra for providing purified *E. coli* Hfq protein for experimental use.

I thank the team at the Cambridge Centre for Proteomics led by Dr. Kathryn Lilley for assistance with proteomics experiments. I thank Renata Feret for guiding me in performing 2D-DiGE, Svenja Hester for analyzing iTRAQ-labelled protein samples, and Dr. Laurent Gatto for analysis of iTRAQ results.

I thank the members of the Sanger Institute, especially Prof. Gordon Dougan, head of the Microbial Pathogenesis team, for giving me the opportunity to apply RNA-seq using the latest high-throughput sequencing technologies as part of this study. I thank Dr. Robert Kingsley for working with me throughout my time at the Sanger Institute to guide the direction of the project and provide invaluable support and assistance. I thank Heidi Hauser and Nicholas Croucher for mapping of RNA-seq reads, and I especially thank Dr. Adam Reid for performing and explaining the DEseq analysis of the RNA-seq data and preparing the MA plots.

For assistance with experiments with *C. elegans*, I thank Dr. Leo Kurz (Marseille, France) for explaining the use of worm strain DH26 and Jakob Møller-Jensen (University of Southern Denmark) for discussions about the use of fluorescing bacteria to visualize colonization of the worm gut.

I especially thank my previous undergraduate supervisor and still mentor, Prof. Roger Wartell, for introducing me to Hfq many years ago in the summer of 2003 after my first year of undergraduate studies. It was a fateful encounter the wisdom of which only time revealed. I thank Prof. Wartell for continuing to provide support and helpful discussions during the course of my doctoral studies. I also thank Taylor Updegrove for providing advice with Hfq-RNA gel-shift experiments.

I thank the fellows and staff of St. John's College for providing a collegial and social environment throughout my course and funding travel expenses to conferences. I thank my tutor, Dr. Frank Salmon, for helpful and pleasant conversations.

I thank the Society for General Microbiology for funding me to attend their conferences during my course. I am especially grateful and honoured by the award of the Young Microbiologist of the Year from the SGM, which was received on 7 September 2010 at the SGM Autumn Meeting in Nottingham, UK.

I would like to thank the Gates Cambridge Trust for generously funding tuition and living expenses that provided the opportunity to pursue my course at Cambridge. It was an honour to have been present during the visit on 12 June 2009 of Bill and Melinda Gates, my personal benefactors, and hear them speak of their philanthropic causes.

A final word of appreciation goes to my friends and all the friends in the Bahá'í community who became as family and generously poured their love and support throughout my time here. You were always of good cheer and exceptional hospitality, and your friendship carried me through any difficulties in my work. May the Best-Beloved of the worlds continue to shower you with His bounties and favors.

Table of contents

List of Figures

List of Tables

Abbreviations

°C	degrees Celsius
AHL	acylated homoserine lactone
Ab	antibiotic
aa	amino acids
Ap	Ampicillin
ATCC	American Tissue Culture Collection
BHL	*N*-butanoyl-L-homoserine lactone
bp	base pairs
cAMP	cyclic adenosine 3',5'-monophosphate
Car	1-carbapen-2-em-3-carboxylic acid
Cro	*Citrobacter rodentium* ICC 168
c-di-GMP	bis-(3′-5′)-cyclic dimeric guanosine monophosphate
Cm	Chloramphenicol
CRP	cAMP receptor protein
Dda	*Dickeya dadantii* 3937
DMDC	Dimethyl dicarbonate
DNA	deoxyribonucleic acid
g	grams
GFP	green fluorescent protein
HHL	*N*-hexanoyl-L-homoserine lactone
HK	histidine kinase
HTH	helix-turn-helix
IL-2	Interleukin 2
kb	kilobase pairs
Km	Kanamycin
L	litres
LB	Luria broth
LBA	Luria broth agar
m	milli (10^{-3})

M	molar (mol/L)
MAP	2-methyl-3-amylpyrrole
MBC	4-methoxy-2,2'bipyrrole-5-carboxyaldehyde
MW	molecular weight
ml	millilitres
MM	minimal media
mRNA	messenger RNA
nm	nanometers
ncRNA	non-coding RNAs
OHHL	*N*-3-(oxohexanoyl)-L-homoserine lactone
ONPG	*o*-nitrophenyl-b-D-galactopyranoside
ORF	open reading frame
PAI	pathogenicity island
Pat	*Pectobacterium atroseptica* SCRI 1043
Pcc	*Pectobacterium carotovorum* subsp. *carotovorum*
PCR	polymerase chain reaction
Pi	inorganic phosphate
pI	isoelectric point
Prodigiosin	2-methyl-3-pentyl-6-methoxyprodigiosin
qRT-PCR	quantitative reverse transcription PCR
QS	quorum sensing
RACE	rapid amplification of cDNA ends
RBS	ribosome binding site
RNA	ribonucleic acid
RND	resistance-nodulation-cell-division
rpm	revolutions per minute
RR	response regulator
SNP	single-nucleotide polymorphism
sRNA	small regulatory RNA
S39006	*Serratia* sp. ATCC 39006
ss	single-stranded

Sma	*Serratia marcescens*
Sp	Spectinomycin
Spr	*Serratia proteamaculans*
TSS	transcription start site
T2SS	Type II secretion system
T4P	Type IV pili
UTR	untranslated region
WT	wild type S39006
WTSI	Wellcome Trust Sanger Institute (Hinxton, UK)
µ	micro (10^{-6})
µl	microlitres

Publications

Chapter 4 of this work has been adapted into an article which has been accepted for publication in the journal *Environmental Microbiology*.

Additional work from this study will be adapted into articles for submission to scientific journals.

Chapter 1. Introduction

1.1 Secondary metabolism

The exploitation of naturally occurring microbial products for medicinal use first began with the advent of penicillin by its discoverer, Alexander Fleming, who in 1929 published his historic observation that *Penicillium notatum* killed his bacterial culture of *Staphylococcus aureus* (Demain, 2006). This led the way in the hunt for other antibiotics the use of which has had a revolutionary impact on modern day society by treating devastating infections, lowering infant and child mortality, and extending life spans. Antibiotics are examples of naturally occurring secondary metabolites whose tremendous diversity hidden within nature is being uncovered and studied for the benefit of civilization. Secondary metabolism is a multifaceted activity that refers to the set of chemical pathways in a cell that lead to the production of secondary metabolites whose role is not essential or primary for normal growth, development, and reproduction of the organism. Secondary metabolites are diverse in function and composition but are defined as small, low molecular weight molecules dispensable for normal cell activity (Fox and Howlett, 2008). Yet, these metabolic 'add-on' molecules are thought to confer a competitive advantage to the fitness of the organism in evolutionary terms (Vining, 1992). They function "(i) as competitive weapons used against other bacteria, fungi, amoebae, plants, insects, and large animals; (ii) as metal transporting agents; (iii) as agents of symbiosis between microbes and plants, nematodes, insects, and higher animals; (iv) as sexual hormones; and (v) as differentiation effectors" (Demain and Adrio, 2008).

In addition to antibiotics, examples of secondary metabolites include pigments, toxins, inhibitors, polyketides, and a whole range of other molecules. By 2002, over 22,000 bioactive compounds had been discovered from microbes and documented in the scientific and patent literature and many continue to find their way into the clinic as antibacterial, antifungal, antiparasitic, anticancer and immunosuppressive agents (Bérdy, 2005). One such class of molecules being developed into chemotherapeutic

drugs are the prodiginines, which are produced by various bacteria, including some *Serratia* spp.

1.2 *Serratia* sp. ATCC 39006 and its secondary metabolites

The Gram-negative bacteria of the genus *Serratia* cause a range of opportunistic human, plant, and insect diseases. Most human infections are nosocomial and result from the well-studied pathogen *Serratia marcescens*. Clinical isolates of this species are increasingly problematic as it develops resistance to traditional antibiotics (Młynarczyk et al., 2007). A characteristic feature of this species is its ability to produce the red-pigmented molecule prodigiosin. Much of the biosynthetic pathway and regulatory network has been uncovered and described in the literature from work done on the strain *Serratia* sp. ATCC 39006 (S39006). Originally, this strain, which is an environmental isolate, was thought to be an example of *Serratia marcescens*, but later it was determined to be a taxonomically ill-defined strain (Slater et al., 2003). In addition to prodigiosin, S39006 produces the plant cell wall degrading exoenzymes pectate lyase and cellulase and a carbapenem (1-carbapen-2-em-3-carboxylic acid; Car), which is a class of β-lactam antibiotics with broad-spectrum activity (Coulthurst et al., 2005b).

The red phenotype of S39006 and the ease by which prodigiosin, carbapenem, and its exoenzymes are easily assayed make this strain a genetically tractable system in which to study the regulation of prodigiosin, which is an example of a potentially clinically useful, model secondary metabolite (Thomson et al., 2000). Furthermore, the genome of S39006 was recently sequenced toward the end of this study and this has aided research on this strain. As a result of these factors, and building on recent research advances in S39006, the regulatory network controlling the production of secondary metabolites and virulence factors was investigated in this study.

1.2.1 Prodigiosin

1.2.1.1 Structure and Biosynthesis

Deemed a model secondary metabolite, prodigiosin (2-methyl-3-pentyl-6-methoxyprodigiosin) belongs to the prodiginine family of red-pigmented tripyrrole molecules. It confers the distinctive red phenotype to colonies of *Serratia* spp. and is

also produced by certain Gram-positive and Gram-negative bacteria isolated from marine and terrestrial environments, including *Streptomyces* spp., *Hahella* spp., and *Vibrio* spp. (Williamson et al., 2007). Prodigiosin is produced in the late exponential and stationary phases of growth when nutrients are limiting. The pathways involved in its biosynthesis have been determined in a previous study from this lab. The enzymes involved in production are encoded by the *pig* operon, consisting of 15 biosynthetic genes (*pigA-O*) which are transcribed as a polycistronic mRNA (Williamson et al., 2005). The biosynthesis proceeds by a bifurcated pathway culminating in the enzymatic condensation by PigC of the terminal products of the two pathways, 2-methyl-3-n-amyl-pyrrole (MAP) and 4-methoxy-2,2′-bipyrrole-5-carbaldehyde (MBC), to form prodigiosin (Figure 1.1). The *pig* cluster and its metabolic pathways are similar to other gene clusters involved in the production of related prodiginine compounds, such as undecylprodigiosin in *Streptomyces coelicolor* A3(2) and cycloprodigiosin in *Vibrio gazogenes* (Williamson et al., 2006).

A feature of the enzymology is the relaxed substrate specificity of the enzyme, PigC, which catalyzes the final condensation step in the production of prodigiosin. It is capable of incorporating similar non-native substrates to produce prodiginine analogues. Obatoclax, a proposed drug in clinical trials, was synthesized in the lab in this way using analogues of MAP and MBC (Chawrai et al., 2008). Thus, there is biotechnological potential to generate novel prodiginines with enhanced therapeutic activity or reduced toxicity from a wild type or mutated PigC enzyme.

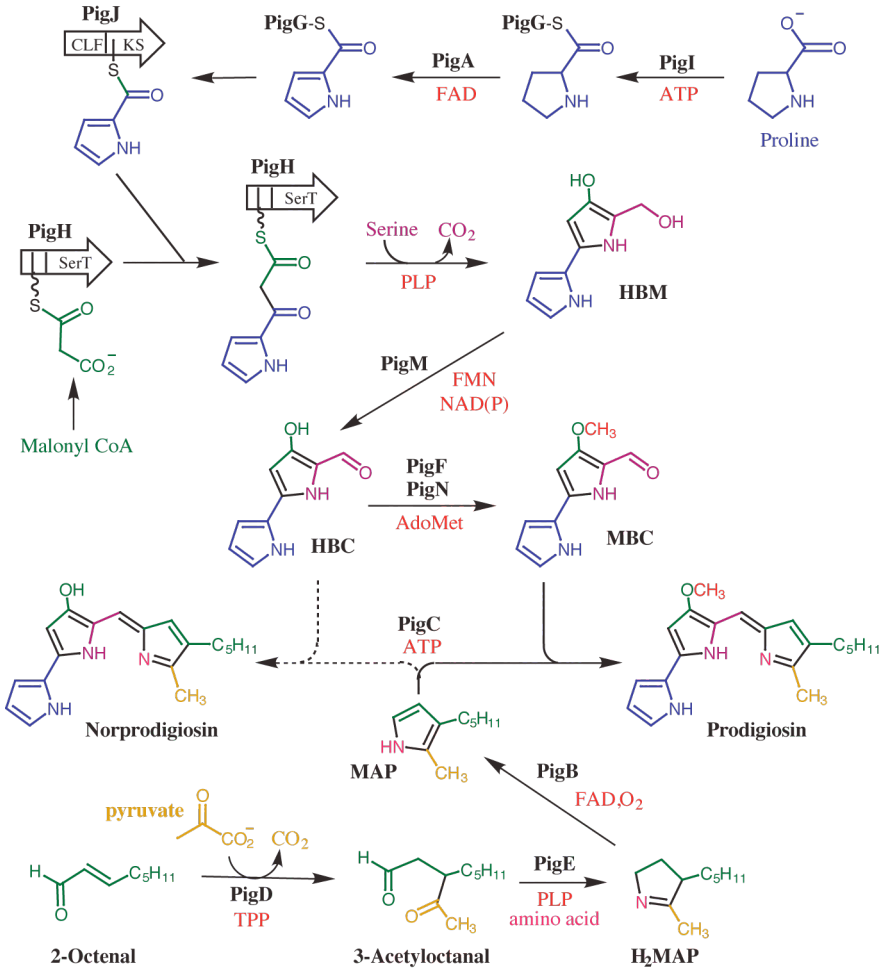


Figure 1.1. The biosynthetic pathway of prodigiosin.

Diagram of the prodigiosin biosynthetic pathway taken from Williamson *et al.* (2005). Prodigiosin synthesis proceeds via a bifurcated pathway culminating in the enzymatic condensation by PigC of the substrates MAP (2-methyl-3-n-amyl-pyrrole) and MBC (4-methoxy-2,2'-bipyrrole-5-carbaldehyde). Norprodigiosn can also be produced from the condensation of HBC (4-hydroxy-2,2'-bipyrrole-5-carbaldehyde) and MAP which is indicated by the dashed arrows. HBC is converted from HBM (4-hydroxy-2,2'-bipyrrole-5-methanol) by the oxidoreductase PigM. The tripyrrole ring is the characteristic feature of prodiginine compounds.

1.2.1.2 Therapeutic Properties

Bacterial prodiginines have various therapeutic properties such as antibacterial, antifungal, antimalarial, anticancer, and immunosuppressive activities (Williamson et al., 2007). Currently, the immunosuppressive and anticancer activities are the most promising features in the drug development of prodiginines. Prodigiosin effectively suppressed T-cells by blocking IL-2 dependent proliferation and, in mice, suppressed

graft-versus-host disease. Additionally, prodigiosin has been shown to be a proapoptotic anticancer agent active against numerous cell lines, including lung, colon, kidney, and breast cancers, with little cytotoxicity to noncancerous cells. A number of prodiginine or prodiginine-like compounds feature in preclinical and clinical trials. Aida Pharmaceuticals (China) is developing prodigiosin into a treatment for pancreatic cancer, and Obatoclax (see pg. 3), developed by Gemin X pharmaceuticals (Canada), is in Phase II clinical trials for treatment against multiple solid tumors and hematologic malignancies (Williamson et al., 2007). The search continues for synthetic prodigines with increased therapeutic activity and decreased cytotoxicity .

1.2.1.3 Physiological Role of Prodigiosin

The exact physiological role of prodigiosin in *Serratia* spp. is still unknown and is debated in the literature. The majority of strains isolated in a clinical setting and grown at 37°C do not produce prodigiosin. Only a small subset produce it when grown at lower temperatures such as 30°C, whereas strains isolated from the environment are often pigmented. In addition to temperature, several other physiological factors impinge on the production of prodigiosin, including salt concentration, light, and pH (Williamson et al., 2006). It has been suggested that, because the surfactant produced by S39006 is able to solubilize and disperse prodigiosin, it may act as an antimicrobial weapon in competition against other organisms, aiding in colonization by *Serratia*. Evidence for this notion is that swarming activity is co-regulated with the biosynthesis of prodigiosin (Williamson et al., 2008). In this manner, *Serratia* may use prodigiosin to gain a competitive advantage in environmental niches.

1.2.2 Carbapenem

The second notable secondary metabolite produced by S39006 is 1-carbapen-2-em-3-carboxylic acid (Car), a carbapenem antibiotic and member of the β-lactam class of antibiotics. The S39006 *car* gene cluster was sequenced and analysed in previous studies from this lab and was shown to be conserved with the equivalent cluster in the plant pathogen *Pectobacterium carotovorum* (formerly *Erwinia carotovora* subsp *carotovora*) (Coulthurst et al., 2005b). The operon consists of eight genes; *carA-E* encoding the proteins for carbapenem production with *carF* and *carG* conferring

intrinsic resistance to the antibiotic by an unknown mechanism. The function of *carH* is unknown but mutants defective in *carH* still synthesize normal levels of carbapenem. The regulation of carbapenem and prodigiosin production is highly coordinated in S39006 (Thomson et al., 2000; Fineran et al., 2005b).

1.3 Genetic Regulation of Prodigiosin Production in S39006

Subsequent to the detailed study of the *pig* cluster and the enzymatic pathways involved in biosynthesis of prodigiosin, continuing work focused on uncovering the mechanism for the genetic regulation of prodigiosin production. When S39006 is grown in liquid culture, prodigiosin production increases from late exponential to stationary phases of growth. Because prodigiosin is pigmented and colonies of S39006 are pink on LB plates, the technique of random transposon mutagenesis was used in facile assays to identify genes involved in the regulation of prodigiosin production. Transposon mutant colonies with altered prodigiosin phenotypes were screened to identify the corresponding gene containing the transposon insertion. Several research reports using this technique progressively revealed a highly complex gene regulatory network coordinating various environmental inputs, including population cell density and the levels of phosphate and gluconate, into an integrated circuit controlling the regulation of prodigiosin production (summarized in Figure 1.2 and Table 1.1) (Gristwood et al., 2008). Over 20 regulatory inputs to this network have been characterized to date, with further research, including the work from this study, continuing to uncover additional regulatory factors.

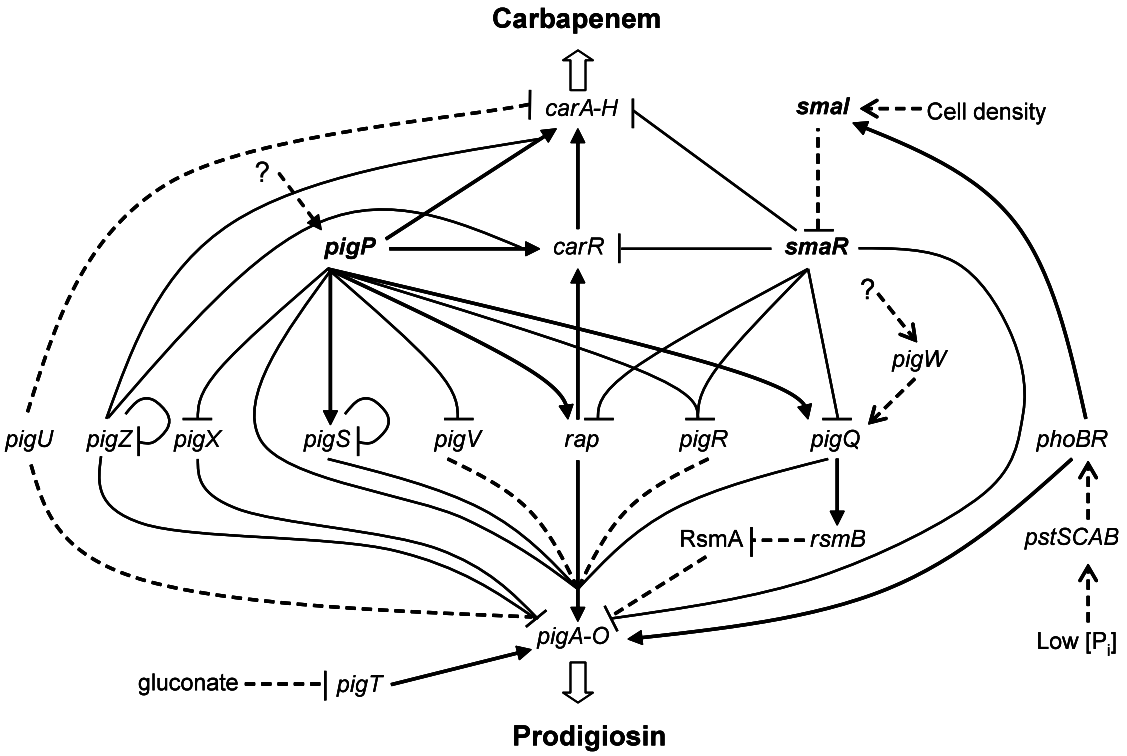


Figure 1.2. A proposed hierarchical model of the regulatory network governing prodigiosin and carbapenem production.

The diagram taken from Fineran *et al.* (2006) depicts the integration of various circuits involved in the regulation of prodigiosin and carbapenem production. The three known major inputs into the network include cell density (sensed via the SmaI/R QS circuit), phosphate levels (sensed via the *phoBR/pstSCAB* circuit), and gluconate levels (sensed via *pigT*). Unknown environmental signals are proposed to regulate *pigP* and *pigW*. *rap, pigR*, and *pigQ* represent points of integration between the QS and *pigP* circuits. Transcriptional activation is represented by bold arrowheads and solid lines, and repression is depicted with flat arrowheads. Dashed lines represent non-transcriptional interactions such as the sensing of environmental factors (e.g. low Pi, or other unknown factors), post-transcriptional effects (e.g. rsmAB regulation), or instances where the mode of regulation has not yet been determined (e.g. pigU, pigV, and pigR).

Table 1.1. Regulators of prodigiosin production.

Gene	Description	Reference
1. pigP	Transcriptional regulator	(Fineran et al., 2005b), (Gristwood et al., 2011)
2. pigQ	Phosphorelay two-component system, transcriptional regulator	(Fineran et al., 2005b)
3. pigR	Adenylate cyclase	(Fineran et al., 2005b)
4. pigS	Transcriptional regulator, SmtB/ArsR family	(Fineran et al., 2005b), (Gristwood et al., 2011)
5. pigT	Transcriptional regulator, GntR-type	(Fineran et al., 2005a)
6. pigU	Transcriptional regulator, LysR family	(Fineran et al., 2005b)
7. pigV	Predicted inner membrane regulatory protein	(Fineran et al., 2005b)
8. pigW	Phosphorelay two-component system, sensor kinase	(Fineran et al., 2005b)
9. pigX	GGDEF/EAL domain protein	(Williamson et al., 2008)
10. pigZ	Transcriptional repressor, TetR/AcrR family	(Gristwood et al., 2008)
11. phoBR	Two-component system for sensing inorganic phosphate (Pi)	(Slater et al., 2003), (Gristwood et al., 2009)
12. pstSCAB-phoU	ATP-dependent Pi uptake transporter	(Slater et al., 2003), (Gristwood et al., 2009)
13. rap	Transcriptional regulator, SlyA/MarR family	(Thomson et al., 1997), (Slater et al., 2003), (Fineran et al., 2005b)
14. rsmA	RNA-binding post-transcriptional regulator	(Williamson et al., 2008)
15. rsmB	Small RNA antagonist of RsmA	(Williamson et al., 2008)
16. smaI	*N*-acyl homoserine lactone synthase	(Thomson et al., 2000), (Crow, 2001)
17. smaR	QS LuxR family transcriptional regulator	(Thomson et al., 2000), (Crow, 2001)

1.3.1 Quorum Sensing

Quorum sensing (QS) is a cell-cell communication mechanism by which bacteria sense the population cell density in their environment via detection of a diffusible signaling molecule, resulting in synchronization of gene expression and coordination of behaviors in the population (Ng and Bassler, 2009). Different types of signaling molecules or "autoinducers" have been identified that are used by various bacterial species for QS. Gram-positive bacteria synthesize and detect modified oligopeptides, whereas *N*-acyl homoserine lactones (AHLs) are a major class of freely diffusible signaling molecule used by many Gram-negative proteobacteria. QS was first observed in the bioluminescent marine bacterium *Vibrio fischeri* the genetic basis of which is encoded by the LuxI/LuxR regulatory system, a paradigmatic example of QS in Gram-negative bacteria (Figure 1.3). LuxI is the synthase for *N*-(3-oxohexanoyl)-homoserine lactone (OHHL). OHHL is freely diffusible across the cell membrane and increases in concentration with increasing cell density of the population in a confined space. At a threshold concentration, OHHL binds to the LuxR protein, a DNA-binding transcriptional activator, and this complex activates transcription of the luciferase *luxICDABE* operon responsible for bioluminesence and increases expression of *luxI*, resulting in higher levels of the OHHL signal and thereby creating an autoinduction positive feedback loop (Ng and Bassler, 2009).

In S39006, secondary metabolite and exoenzyme production was found to be under the control of QS. S39006 possesses a LuxIR-type QS system encoded by the homologous *smaIR* locus (**s**econdary **m**etabolite **a**ctivator) which is transcribed convergently. However, SmaR is a member of a family of atypical LuxR-type regulators that are active only in the absence of cognate AHLs and whose activity is to inhibit expression of target genes (Tsai and Winans, 2010). SmaI directs the synthesis of two signalling molecules, *N*-butanoyl-L-homoserine lactone (BHL) and *N*-hexanoyl-L-homoserine lactone (HHL), with BHL being the major product (Thomson et al., 2000). At high cell density, SmaR-mediated repression of the *pig* and *car* gene clusters and *carR* is de-repressed by rising BHL/HHL levels, resulting in the biosynthesis of prodigiosin and carbapenem. CarR is a LuxR homologue that activates, in an AHL-independent manner, the transcription of the *car* operon (Poulter et al., 2011). Figure

1.4 summarizes the mechanism of QS control of secondary metabolite production in S39006. However, a *smaR* mutant produces only slightly higher levels of carbapenem, and prodigiosin production is unaffected, indicating that the QS system is superimposed on underlying regulatory mechanisms required to activate secondary metabolite production (Slater et al., 2003).

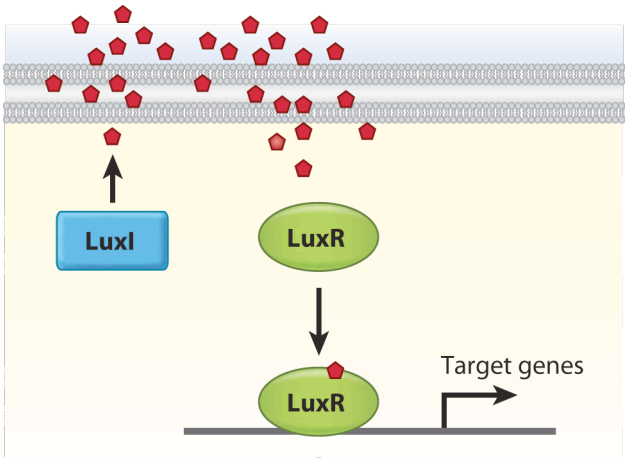


Figure 1.3. Model of a LuxIR-type QS system.

The diagram taken from Ng and Bassler (2009) depicts a paradigmatic Gram-negative LuxIR-type QS system. LuxI synthesizes AHL autoinducers depicted by red pentagons which is freely diffusible across the membrane. At a threshold concentration, LuxR binds AHL and recognizes a consensus binding site (lux box) upstream of target genes activating their expression.

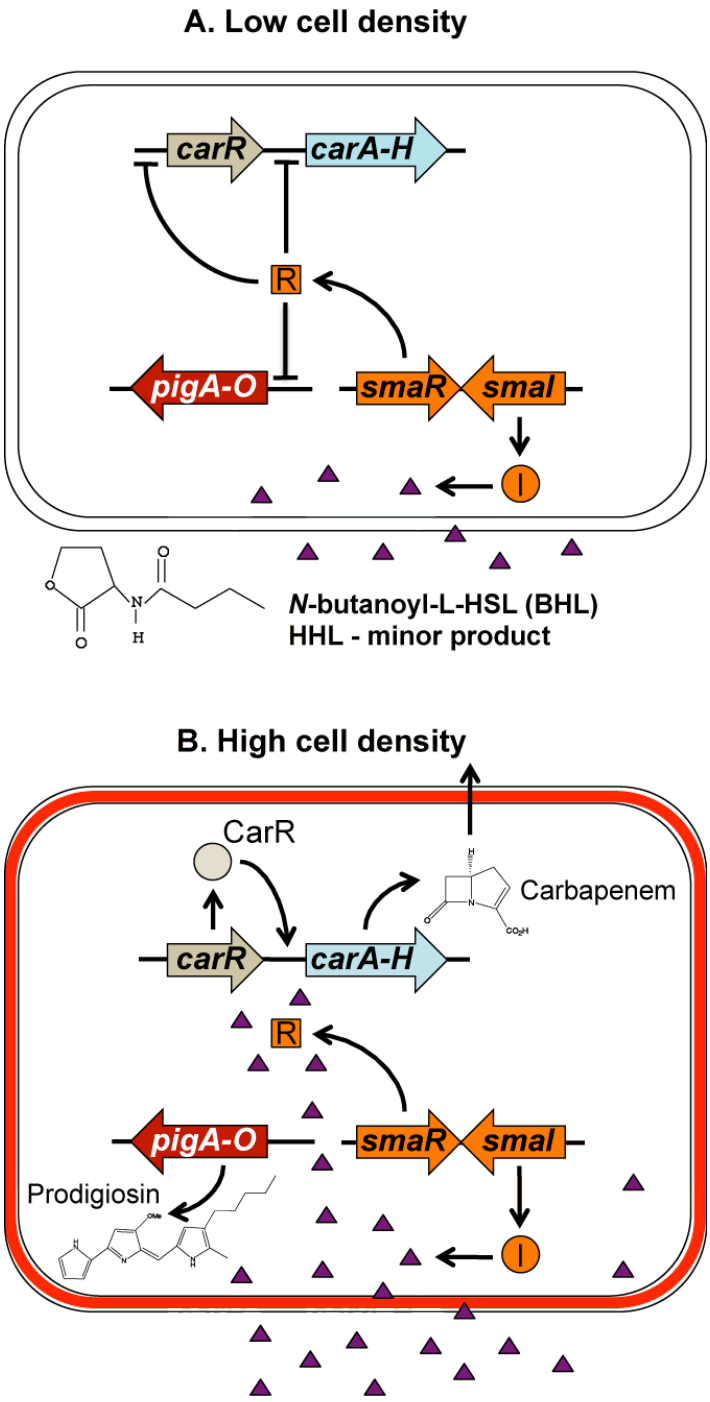


Figure 1.4. Model of the QS regulation of secondary metabolite production in S39006.

The S39006 QS system is encoded by the *smaIR* operon. SmaI directs synthesis of the QS signalling molecules, BHL/HHL. At low cell density, SmaR represses the expression of the *pig* and *car* clusters and *carR*. At high cell density, the increasing concentration of BHL/HHL induces derepression of SmaR-regulated target genes allowing production of carbapenem and

prodigiosin. Activation is depicted by pointed arrowheads and repression by flat arrowheads. Dashed lines indicate putative regulation by Hfq.

1.3.2 Two-Component Signal Transduction Systems

Some of the known regulators of prodigiosin production identified consist of two-component signal transduction systems, a well-studied common regulatory mechanism found in bacteria (Stock et al., 2000). The system normally consists of a membrane-associated histidine kinase (HK) and a response regulator (RR) that is often a transcriptional activator in its active state. The HK senses an environmental stimulus and in response, autophosphorylates a conserved histidine residue. The phosphoryl group is then transferred to the RR on a conserved aspartate residue, activating the output domain of the RR, such as a DNA binding domain.

Alternatively, the HK could be involved in a multi-step phosphorelay system whereby the phosphate group is transferred initially to an additional RR domain on the HK itself before it is transferred to a cytoplasmic RR that activates transcription of target genes.

1.3.2.1 Phosphate Availability

S39006 regulates the production of prodigiosin and carbapenem in response to the availability of inorganic phosphate (Pi). Pi is sensed by an inner membrane-associated high affinity transporter (PstSCAB-PhoU) which activates the PhoBR two-component system. Mutants defective in the *pstSCABphoU* operon are hyper-pigmented (10-fold over wild type) and produce 2-fold more carbapenem (Slater et al., 2003). In low Pi, the PstSCABPhoU transporter induces the HK, PhoR, to activate (by phosphorylation) PhoB, which contributes towards transcriptional activation of *smaI*, *pigA*, and another Pig regulator, *rap*. Increased expression of Rap in response to the phosphate response and higher levels of BHL/HHL further activates *pigA* and *carA* production. Thus, the Pst transporter and PhoBR two-component system regulate the biosynthesis of prodigiosin and carbapenem via QS-dependent and independent pathways (Slater et al., 2003; Gristwood et al., 2009). The role of phosphate sensing in the regulatory network of prodigiosin is depicted in the summary diagram Figure 1.2.

1.3.2.2 PigQW Two-Component System

In addition to the Pst/PhoBR system, a second two-component phosphorelay system encoded by *pigQ* and *pigW*, was found to be a part of the prodigiosin regulatory network. PigQW are homologues of the *E. coli* GacA and GacS two-component system which has been identified in a number of Gram-negative bacteria and has been to shown to have an important role in the production of secondary metabolites and extracellular enzymes involved in pathogenicity (Heeb and Haas, 2001). GacS/PigW is the membrane-associated component which is autophosphorylated in response to an unknown environmental stimulus. GacS/PigW then activates the RR, GacA/PigQ, via phosphotransfer which then upregulates the transcription of small RNAs of the *E. coli* Csr system, or the equivalent S39006 Rsm (**r**egulator of **s**econdary **m**etabolism) system. The small RNA CsrB/RsmB sequesters the RNA-binding protein (CsrA/RsmA) from repressing the translation of target mRNA transcripts, resulting in the upregulation of prodigiosin production (Lapouge et al., 2008). A model of the Csr system is represented in Figure 1.5.

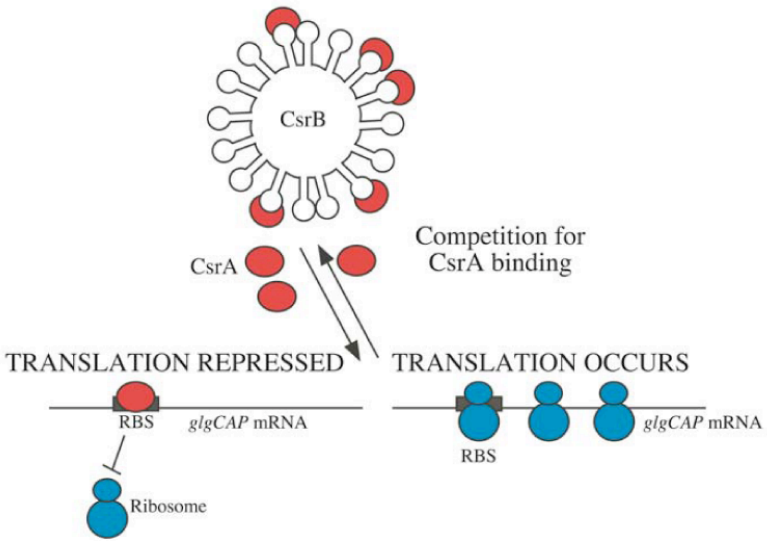


Figure 1.5. Model of action of the Csr system.

Model of CsrAB action from Majdalani *et al.* (2005). CsrA is an RNA-binding protein that inhibits the translation of target mRNAs, e.g. *glgCAP* mRNA, by binding to the ribosome binding site (RBS). The activation of GacS and GacA leads to the expression of the small RNA, CsrB, which titrates CsrA away from target mRNAs. Ribosomes are then free to bind to and translate *glgCAP* mRNA.

1.3.3 Gluconate repression of prodigiosin production

Prodigiosin production was found to be dependent upon levels of gluconate (Fineran et al., 2005a). The *pigT* gene encodes a GntR homologue from *E. coli* which is a transcriptional regulator and represses, in the absence of gluconate, the *gntI* group of genes involved in gluconate utilization (Izu et al., 1997). PigT was shown to directly activate transcription from the *pigA* promoter. The addition of gluconate deactivates PigT, resulting in a reduction of prodigiosin production. However, the purpose for gluconate-mediate repression of prodigiosin production is unknown.

1.3.4 PigP, a master regulator of prodigiosin production

PigP is transcriptional regulator containing an N-terminal helix-turn-helix (HTH) DNA binding domain. It was uncovered as a major regulatory protein or master regulator which activates prodigiosin production by altering the expression of six other Pig regulators (*pigQ, pigR, pigS, pigV, pigX* and *rap*) and the *pig* cluster itself (Fineran et al., 2005b). Three of these genes (*pigQ, pigR,* and *rap*) are jointly controlled by PigP and the QS circuit and therefore represent a 'node' where multiple environmental or physiological signals are integrated to regulate secondary metabolism (Williamson et al., 2006).

1.3.5 Additional regulators of prodigiosin production

Several additional regulators are involved in the control of prodigiosin production. Rap (**r**egulator of **a**ntibiotic and **p**igment) contains a winged HTH DNA-binding domain and controls prodigiosin at the level of transcription of the biosynthetic cluster. It is a member of the MarR family of transcriptional regulators and mutants of *rap* produce no detectable prodigiosin or carbapenem (Thomson et al., 1997). PigR is a predicted class IV adenylate cyclase which is presumed to activate prodigiosin production through cyclic adenosine 3',5'-monophosphate (cAMP). It is thought that cAMP complexes with the cAMP receptor protein (CRP) which then binds to a putative binding site for the cAMP-CRP complex in the promoter region of the *pig* cluster (Fineran, 2006).

Another intracellular messenger thought to be involved in prodigiosin production is bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP). Pig X is a

predicted inner-membrane signalling protein and genetic evidence indicates that it might function as a cyclic diguanosine monophosphate (c-di-GMP) phosphodiesterase (Fineran et al., 2007). *pigX* mutants show elevated levels of prodigiosin although carbapenem is unaffected. PigV is also predicted to be an inner membrane regulatory protein with similarity to YgfX in *E. coli* but the mechanism by which it activates prodigiosin production is unknown (Williamson et al., 2006).

PigS contains a predicted DNA binding domain and is thought to be a part of the ArsR/SmtB-family of regulators which is involved in the control of stress resistance to high levels of heavy metal ions. PigU is of the LysR-type of transcriptional regulators and directly represses pigment production (Williamson et al., 2006). PigZ is a transcriptional repressor of the TetR family. In the absence of PigZ, a putative resistance-nodulation-cell-division (RND) transporter is expressed which leads to increased prodigiosin levels (Gristwood et al., 2008). The network of regulators is summarized in diagram Figure 1.2.

1.3.6 Pigment regulatory factor

A search for additional genetic factors with a regulatory role in pigment production uncovered an unknown putative **p**igment **r**egulatory **f**actor (*prf*) gene the mutation of which results in a hyper-pigmented phenotype. Mutants defective in *prf* were picked out in a general screen for hyper-pigmented mutants in 39006, and in two screens searching for bypass mutations that restore pigment production and suppress the effect of *pigP* and *rsmB* mutants (L. Woodley, unpublished). The phenotypes and role of *prf* was investigated in this study (see Chapter 3).

1.4 Small RNA regulation by Hfq

The involvement of the Rsm system of post-transcriptional regulation by small RNAs in the control of prodigiosin production contributed to the hypothesis that members from the larger family of Hfq-dependent trans-acting sRNAs may likewise play a role in the regulatory network. The RNA-binding protein Hfq has been shown to have a central role in other bacterial species in post-transcriptional regulation via modulation of the binding of small regulatory RNAs (sRNAs) to their target messenger

RNAs (mRNAs). The effect of an *hfq* deletion in S39006 on secondary metabolism and pathogenesis was explored in this study (see Chapter 4).

1.4.1 Hfq structure

Hfq was first identified as a host factor protein required for the replication of the RNA phage Qβ (Franze de Fernandez et al., 1968). Its role as a global post-transcriptional regulator through its function in sRNA regulation highlights it as a key protein through which bacteria alter gene expression in order to adapt to changes in the environment. Hfq is structurally and evolutionarily homologous to the eukaryotic Sm and Lsm proteins that are involved in pre-mRNA splicing, mRNA stability, and translation, and Hfq is highly conserved across numerous bacterial species (Sun et al., 2002). Hfq forms a hexamer in a doughnut-like ring structure based on crystallization and electron microscopy studies (Sauter et al., 2003). It binds to RNA strongly, particularly AU-rich single-stranded RNA, with a preference for binding next to a stem-loop region (Brescia et al., 2003).

1.4.2 The role of Hfq in post-transcriptional regulation

The importance of Hfq as a global regulator of gene expression is underpinned by the pronounced pleiotropic phenotypes of *E. coli hfq* null mutants (Tsui et al., 1994). The properties of these mutants matched defects caused by mutations in *rpoS*, which encodes a stationary phase sigma factor. RpoS is a global regulator of several genes expressed during stationary phase or various stress situations in bacteria. Hfq was shown to be necessary for the translation of *rpoS* mRNA in *E. coli* and *S. typhimurium* (Brown and Elliott, 1996; Muffler et al., 1996).

Hfq upregulates *rpoS* via the action of two sRNAs, DsrA and RprA (Gottesman, 2004). The inhibition of RpoS mRNA translation is due to the secondary structure of the 5’ untranslated region (UTR) which folds into a hairpin that sequesters the ribosome binding site (Cunning et al., 1998). The sRNAs, DsrA and RprA, permit *rpoS* translation by binding to complementary sequences in the 5’ UTR (untranslated region) altering the inhibitory secondary structure and allowing access to the ribosome binding site (Majdalani et al., 1998; Lease and Woodson, 2004). The effect of a loss of *rpoS* function

in S39006 on secondary metabolism and pathogenesis was explored in this study (see Chapter 5).

Although Hfq and dependent small RNAs positively regulate the translation of *rpoS*, the majority of small RNAs negatively regulate their target mRNAs by short, imperfect binding to the 5' region (Papenfort and Vogel, 2010). Additionally, a single small RNA can regulate multiple targets such as the sRNA, RybB, which regulates multiple outer membrane protein mRNAs in *Salmonella* (Vogel, 2009). Moreover, high-throughput sequencing of a cDNA library of RNAs that co-immunoprecipitated with Hfq from *Salmonella typhimurium* identified 727 mRNAs and 64 sRNAs which suggests about a 1:7 ratio of sRNA to target mRNAs (Sittka et al., 2008).

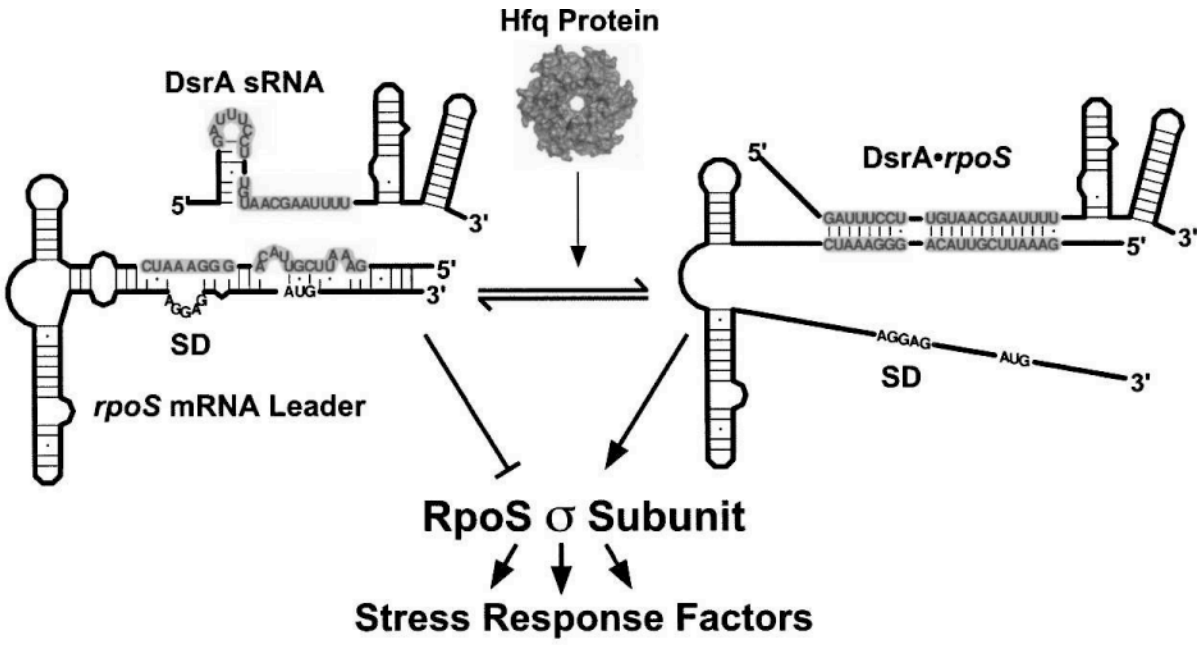


Figure 1.6. Model of Hfq-dependent small RNA regulation of *rpoS* mRNA.

Model of Hfq-mediated small RNA regulation of *rpoS* mRNA from Soper and Woodson (2008). *rpoS* mRNA contains an inhibitory secondary structure that occludes the ribosome binding site (RBS), represented by the Shine-Dalgarno sequence (SD) in the diagram. Hfq facilitates the pairing of the highlighted complementary regions of DsrA sRNA and *rpoS* mRNA, exposing the RBS to allow ribosomes to bind and translate the mRNA. Translational expression or activation is represented by solid arrowheads and repression is depicted with flat arrowheads.

1.4.3 The role of Hfq in pathogenesis

A growing body of research over the past decade has shown that Hfq plays an essential role in pathogenesis in some Gram-negative and Gram-positive bacteria (Table 1.2). Transcriptomic and proteomic approaches suggest that Hfq may regulate up to 20% of all genes in some organisms including secretion systems that play a role in pathogenesis (Chao and Vogel, 2010). For the *S. typhimurium hfq* mutant, significant reductions were observed in secretion of several Type III secretion system (T3SS) effector proteins encoded by the *Salmonella* pathogenicity island 1 (SPI-1) which facilitates host invasion or intracellular survival (Sittka et al., 2007). Deep sequencing analysis of Hfq-associated RNAs identified *hilD*, the master transcriptional regulator of SPI-1, as a target of Hfq (Sittka et al., 2008). Conversely, other studies demonstrated that *hfq* deletion mutants of enterohaemorrhagic and enteropathogenic *E. coli*, *V. cholerae*, and *P. aeruginosa* showed increased transcript abundance for both effector and structural genes of the T3SS (Hansen and Kaper, 2009; Shakhnovich et al., 2009).

The S39006 genome contains a cluster of predicted T3SS *hrp* genes which are highly related to T3SS genes from various plant pathogens including *Pectobacterium atrosepticum* SCRI1043, the causative agent of soft rot and blackleg potato diseases (Bell et al., 2004). The virulence factors regulated by S39006 Hfq were explored in this study (see Chapter 6).

Table 1.2. A summary of *hfq mutants* attenuated for virulence in pathogenic bacteria.

Bacterial Species	Study
Brucella abortus	(Robertson and Roop, 1999)
Pseudomonas aeruginosa (PAO1)	(Robertson and Roop, 1999)
Listeria monocytogenes	(Christiansen et al., 2004)
Vibrio cholerae	(Ding et al., 2004)
Legionella pneumophila	(McNealy et al., 2005)
Shigella flexneri	(Sharma and Payne, 2006)
Salmonella typhimurium	(Sittka et al., 2007)
Uropathogenic *Escherichia coli* (UPEC)	(Kulesus et al., 2008)

Yersinia pestis	(Geng et al., 2009)
Francisella tularensis	(Meibom et al., 2009)
Neisseria meningitidis	(Fantappiè et al., 2009)
Enterohaemorrhagic *E. coli* (EPEC)	(Hansen and Kaper, 2009; Shakhnovich et al., 2009)
Burkholderia cepacia	(Sousa et al., 2010)
Yersinia pseudotuberculosis	(Schiano et al., 2010)
Borrelia burgdorferi	(Lybecker et al., 2010)

1.5 Mapping bacterial transcriptomes with RNA-seq

With the advent of recent high-throughput sequencing technologies it is now possible to identify small and other non-coding RNA elements from whole transcriptome profiles generated by sequencing a library of cDNAs reverse transcribed from extracted RNA in a method referred to as RNA-seq (Sorek and Cossart, 2009). Deep sequencing of transcriptomes is revolutionizing the understanding and complexity of bacterial transcriptomes. The many advantages and new insights afforded by RNA-seq over previous transcriptomic methods, chiefly microarrays, include:

- Discovery of new genes coding for small proteins;
- Discovery of non-coding RNA (ncRNA) elements such as,
    - Trans-acting small RNAs,
    - Riboswitches, and
    - *cis*-antisense RNAs; and
- Definition of transcription start sites (TSS), 5' and 3' untranslated regions (UTR), and operon maps.
- Quantitation of gene expression by measuring relative transcript abundance

Additionally, RNA-seq can be adapted to analyze a subset of the transcriptome such as Hfq-associated RNAs (Sittka et al., 2008).

One of the challenges with RNA-seq is to prepare strand-specific cDNA libraries in order to preserve information regarding the original coding strand. cDNAs must be amplified into double-stranded cDNA in order to prepare the sample for sequencing on

the different platforms currently available including the Illumina Genome Analyzer, the Roche 454, and the ABI SOLiD. Different methods have been employed to generate strand-specific libraries which can be mostly divided into two main classes, one that ligates distinct adaptors to mark the 5' and 3' ends of the RNA trancript, and the second that relies on marking one strand by chemical modification (Levin et al., 2010). Another method includes sequencing of single-stranded cDNAs which relies on the formation of inter- and intra-molecular interactions between cDNAs to prime the ligation of sequencing adapters (Croucher et al., 2009). The S39006 transcriptome was sequenced using this latter method in order to identify ncRNA elements and quantify changes in gene expression in *hfq* and *rsmA* mutants (see Chapter 6).

1.6 Investigating host-pathogen interactions

Research in the area of host-pathogen interactions seeks to elucidate the mechanism by which a pathogen attacks a host, and the defence response of the host to the pathogen. The chance of success for each organism to overcome the other in this antagonistic relationship is the result of a long evolutionary "arms race" between pathogen and host (Tan et al., 1999a). Pathogens that evolve new virulence factors and develop new means to evade host recognition are counteracted by advances in host defence responses such as modified receptors to recognize new strains of pathogens or secreted virulence factors. The margin of "victory" for pathogen or host can be a delicate balance in which the tipping point of success for either side depends on a complex cascade of both intracellular and extracellular signalling. In order to tease apart the complex interactions underlying the pathogenesis of microbes and the defence response of infected hosts, several *in vitro* and *in vivo* infection models have been developed to identify the genetic factors involved (Pradel and Ewbank, 2004). A host organism is most likely to share defence mechanisms with its closest relatives. But studies are increasingly showing that some mechanisms of pathogen attack and host defence are widely conserved, indicating that research using any model species, whether vertebrate, invertebrate, or plant, may shed light on fundamental mechanisms of pathogenesis that are universal (Aballay and Ausubel, 2002; Pradel and Ewbank, 2004). Some of these universal mechanisms involve adherence and/or invasion by pathogens of host cells and secretion of effectors, as well as conserved signalling

pathways of innate immunity in host cells to recognize pathogens and respond with antimicrobial peptides.

1.6.1 *C. elegans* is a model host for animal pathogenesis

In order to dissect host-pathogen interactions at a molecular level, the soil nematode, *Caenorhabditis elegans*, has become a popular choice as a model host organism among researchers for its practicality and the avoidance of ethical considerations. It offers a compromise between genetic complexity and tractability with its small (97 Mb) fully sequenced genome, in comparison to other model hosts. Additional advantages include its short 2-3 week life span and its highly conserved biological processes with mammals (Aballay and Ausubel, 2002). Work by the Ausubel laboratory over a decade ago using the human pathogen *Pseudomonas aeruginosa* strain 14 (PA14) established *C. elegans* as "a model organism to study innate immunity and identify virulence factors of infectious bacteria" (Mahajan-Miklos et al., 1999; Tan et al., 1999a; Tan et al., 1999b; Zhang, 2008). As a host model it possesses the advantages of allowing relatively high-throughput screening of transposon mutant libraries of bacteria to identify virulence factors, which is rapid, inexpensive, and much more practical compared to mammalian hosts (Gravato-Nobre and Hodgkin, 2005).

C. elegans can live for up to two and a half weeks at 25°C sustained by the normal laboratory feeder strain *E. coli* OP50 grown on nematode growth medium (NGM) (Couillault and Ewbank, 2002). *C. elegans* is a filter feeder that ingests bacterial cells suspended in liquid through the mouth, concentrating food in the pharynx and returning liquid out through the mouth (L. Riddle, 1997). Bacteria are transported through the pharynx by peristaltic action. The base of the pharynx is referred to as the terminal bulb and contains the grinder which consists of three plates with teeth-like ridges. As the muscles of the pharynx contract, the plates rotate against one another disrupting or grinding the bacteria before passing through the pharyngeal-intestinal valve into the intestine. Defects in grinding lead to the ingestion of whole bacteria which proliferate in the intestines, distending the lumen and causing sickness (Labrousse et al., 2000). Pathogenic bacteria have been observed to get past the grinder of normal worms and cause intestinal infections. However, the basis of the ability to do this is unknown.

In order to evaluate whether a bacterium is pathogenic to *C. elegans*, worms are typically grown until the fourth larval stage (L4) and then transferred to plates cultured with the bacterium of interest. A decrease in the life span of worms fed on the bacterium (compared to OP50) provides an indication of virulence. Surviving worms are counted each day or hour depending on the speed of killing and are considered dead when they fail to respond to touch. Worms are transferred daily to fresh plates in order to avoid confusion with progeny when counting. A temperature sensitive strain of *C. elegans* that is sterile at 25°C but reproduces at 15°C can also be used for this purpose. In order to ensure that the bacterium tested is an adequate nutritional source for *C. elegans*, the worms are fed heat-killed bacteria to evaluate if the survival rate is similar for OP50. There is a correlation between edibility of bacteria and cell size whereby, as cell size increases, edibility decreases (Avery and Shtonda, 2003).

1.6.2 The different modes of *Pseudomonas aeruginosa* killing of *C. elegans*

P. aeruginosa is an opportunistic pathogen that infects immunocompromised individuals and is a leading cause of lung infections in cystic fibrosis patients. It also displays an ability to infect a broad host range of both animals and plants. *P. aeruginosa* kills *C. elegans* by four known mechanisms: slow or fast-killing, lethal paralysis, and red death, depending on the strain of *P. aeruginosa* and growth media indicating that "the mode and extent of killing depends on a variety genetic and environmental factors" (Aballay and Ausubel, 2002).

The Ausubel laboratory initially described two forms of killing of *C. elegans* by *P. aeruginosa* strain 14: fast versus slow killing. PA14 grown on a slightly modified version of NGM (a relatively low-osmolarity medium) killed *C. elegans* over the course of several days (2-3 dy; slow-killing) by an infection-like process in which PA14 accumulated in worm intestines (Tan et al., 1999a). Microscopy of worms fed after 48 hours on PA14 expressing green fluorescent protein (GFP) showed the intestinal lumen full of PA14/GFP compared to the control *E. coli* DH5α/GFP. PA14 grown on the relatively high-osmolarity peptone-glucose medium (PG) killed *C. elegans* in a matter of hours (4-24 hr; fast-killing) by the secretion of low-molecular-weight toxins, with phenazines identified as one of the mediators of killing (Mahajan-Miklos et al., 1999). The toxin-mediated killing mechanism was confirmed by feeding worms heat-killed or antibiotic-

killed PA14 which killed as efficiently as live PA14, whereas heat-killed PA14 did not kill worms under slow-killing conditions (Tan et al., 1999a). Additionally, fast-killing was also still observed in *C. elegans* that had not come into contact with PA14 bacterial cells but which were cultured on a filter covering the PG plate and removed prior to seeding with worms, allowing for the diffusion of toxins into the media killing the worms. Another toxin-mediated mode of killing referred to as lethal paralysis was observed for the more commonly studied strain of *P. aeruginosa*, PAO1, grown on rich brain-heart infusion medium. This kills worms after 4 hours by inducing neuromuscular paralysis through the generation of neurotoxins, including hydrogen cyanide (Darby et al., 1999; Aballay and Ausubel, 2002). The most recently described killing mechanism is referred to as red death for the visible red colour in dead worms. Worms were fed on PAO1 grown on low phosphate NGM which induces virulence systems through phosphate, quorum sensing, and iron signalling, and kills 70% of worms after 50 hours (Zaborin et al., 2009). A summary of the different modes of *P. aeruginosa* is provided in Table 1.3.

1.6.3 *Serratia marcescens* killing of *C. elegans*

The pathogenicity of other bacteria has also been investigated using *C. elegans* as a model host. Over the past decade the lab of Jonathan Ewbank has studied the host-pathogen interaction between the Gram-negative bacterium, *Serratia marcescens* strain Db 11 (Db11), and *C. elegans*. They determined that Db11 kills *C. elegans* following the establishment of an infection in the intestinal lumen (Mallo et al., 2002). However, unlike *Pseudomonas aeruginosa*, additional toxin-mediated killing mechanisms have not been described for *S. marcescens*. Feeding worms heat-killed Db11 and heat-killed PA14 elucidated the toxin-independent killing mechanism. Worms fed on heat-killed Db11 did not appear sick and survived as long as that of worms fed on *E. coli* OP50. Additionally, after as a little as 2 hours of contact, worms fed on Db11 expressing GFP were shown to have intact Db11-GFP in the intestinal lumen. After 3 days of feeding on Db11, intestinal cells were damaged with the lumen distended and full of intact bacteria (Kurz et al., 2003).

Table 1.3. Summary of killing modes of *C. elegans*.

Organism	***C. elegans* pathology**	**Reference**
Pseudomonas aeruginosa strain PAO1	Toxin- hydrogen cyanide (lethal paralysis)	(Darby et al., 1999)
Pseudomonas aeruginosa strain PAO1	Diminished life span[1] (red death)	(Zaborin et al., 2009)
Pseudomonas aeruginosa strain PA14	Toxin- phenazine (fast-killing)	(Mahajan-Miklos et al., 1999)
Pseudomonas aeruginosa strain PA14	Intestinal infection (slow-killing)	(Tan et al., 1999a)
Serratia marcescens strain Db 11	Intestinal infection	(Kurz et al., 2003)

[1]"Diminished life span" indicates that the killing mode has not been determined in more detail.

1.6.4 Innate immunity of *C. elegans*

All animal organisms are exposed to a wide variety of bacteria throughout their lifetimes and have evolved immune responses to defend against infectious agents. For vertebrates this involves the complex and interconnected systems of innate and adaptive immunity. The more specialized adaptive immune system, which includes T-cells and B-cells, has the capability to recognize and attack certain pathogens and "remember them" in the course of future infections to generate immunity.

The evolutionarily older innate immune response defends the host from bacteria in a non-specific manner using genetically hard-wired mechanisms. Cells of the innate immune system in vertebrates include leukocytes such as macrophages and neutrophils which identify and eliminate pathogens by engulfing them. Despite not possessing an adaptive immune response, invertebrates rely on mechanisms of innate immunity to defend against pathogens. The realization that there are features of innate immunity that are evolutionary conserved and universal among vertebrates, invertebrates, and plants is regarded as a remarkable development in modern immunology (Gravato-Nobre and Hodgkin, 2005). Although the innate immune system of *C. elegans* differs from vertebrates and is simpler than the popular invertebrate model, *Drosophila*

melanogaster, studies on the defence response of *C. elegans* against pathogenic bacteria have yielded important insights in how animals identify and defend against infection. This has the increased the use of *C. elegans* as a model organism outside the fields of developmental biology and neurobiology for which it first gained popularity (Gravato-Nobre and Hodgkin, 2005). The simplicity of *C. elegans* as an animal model organism allows for the study of immune defence response in the context of the whole organism at a molecular level that is not yet feasible in more complex model organisms (Shapira and Tan, 2008). The genetic tractability of *C. elegans* allows the quick screening of various random mutations to identify single genes important for protection against infection and to elucidate signalling pathways involved in host defences. Studies of host-pathogen interactions using the model host *C. elegans* to study the human pathogens, *P. aeruginosa* and *S. marcescens* among others, have demonstrated the conservation of signalling pathways related to innate immunity in animals.

Before a pathogen even enters *C. elegans*, the first line of defence for the worm is to distinguish food from foe and avoid pathogens in the environment. It has been observed that *C. elegans* will avoid the lawns of the pathogens *P. aeruginosa* (Tan et al., 1999a), *S. enterica* (Aballay et al., 2000), and *S. marcescens* (Pujol et al., 2001) in addition to others. Experimental evidence suggests that the behaviour of lawn avoidance may be mediated by a Toll receptor since it was observed that a *C. elegans tol-1* mutant is defective in its avoidance of *S. marcescens* Db10 and 11 (Pujol et al., 2001). Serrawettin W, a lipopeptide biosurfactant, was identified as the chemical repellent produced by *S. marcescens* and perceived by *C. elegans* through the olfactory sensory neurons AWB, among possibly other neurons (Pradel et al., 2007). In addition to its role in pathogen avoidance behaviour, the role of TOL-1 in innate immune responses was confirmed by demonstrating that a *C. elegans tol-1* mutant is killed by significant invasion of the pharynx by *S. enterica*. It was shown that TOL-1 is required for the correct expression of ABF-2, which is a defensin-like molecule expressed in the pharynx (Tenor and Aballay, 2008). The discovery of the role of pathogen recognition through the Toll-like receptors provides increasing evidence that this aspect of innate immunity may be universally shared by animals.

1.7 Aims of this study

The aim of this study is to further the understanding of the regulatory network governing secondary metabolism and virulence in *Serratia* 39006. This study investigated the role of the post-transcriptional regulator and RNA-binding protein, Hfq, and the sigma factor, RpoS, in secondary metabolism and pathogenesis. The link between Hfq and QS regulation was explored as well as the function of the sRNA, RprA. A combination of transcriptomics (RNA-seq) and proteomics were used to provide a systems-level understanding of Hfq and RsmA regulation and to identify similarities and differences in the regulons of these two global post-transcriptional regulators.

Chapter 2. Materials and Methods

2.1 Media, reagents and solutions

The antibiotics and supplements used in this study are listed in Table 2.1 and media and standard solutions are listed in Table 2.2. Solutions, where appropriate, were sterilised by either autoclaving at 121°C for 20 min or filtration through 0.22 μM filters (Millipore).

2.2 Bacterial strains, plasmids and culture conditions

Bacterial strains, phage, and plasmids used in this study are described, with a brief description of their construction, in Table 2.3. *Serratia* strains were grown at 30°C and *Escherichia coli* strains at 37°C in Luria broth and on LB agar plates. When required antibiotics and supplements were used at concentrations shown in Table 2.1. Overnight cultures were grown by inoculating 5 ml LB in 25 ml universal tubes with a single colony and growing overnight at the appropriate temperature on a rotating wheel. For growth curves, 25 ml or 50 ml LB media in 250 ml conical flasks were inoculated with cells from an overnight culture at a starting OD of 0.02, and grown in a shaking water bath at 200 rpm. Cells from overnight cultures were pelleted by centrifugation and resuspended in 250 μl LB prior to use. Bacterial growth or optical density was measured in a Unicam Heλios spectrophotometer at 600 nm.

Table 2.1. Antibiotics and supplements used in this study.

Chemical	Abbreviation	Stock concentration (mg/ml)[1]	Final concentration (μg/ml)
Antibiotic			
Ampicillan	Ap	100 in water	100
Chloramphenicol	Cm	25 in EtOH	25
Erythromycin	Em	200 in EtOH	200
Kanamycin	Km	50 in water	50
Spectinomycin	Sp	50 in water	50
Streptomycin	Sm	100 in water	100

Tetracyline	Tc	10 in MeOH	10
Supplement			
Diaminopimelate	DAP	60 mM = 11.7 mg/ml in water	300 μM
5-bromo-4-chloro-3-indolyl-D-galactopyranoside	X-gal	20 mg/ml in dimethylformamide	40 μg/ml
Isopropyl-β-D-thiogalactopyranoside	IPTG	100 mM = 24 mg/ml in water	0.1 mM

[1]All solutions stored at -20°C except for DAP which was stored at 4°C.

Table 2.2. Solutions used in this study.

Solution	Components (per L)
Media	
Cel Agar	10 g carboxymethyl cellulose 16 g Bacto agar 25 ml 20% (w/v) Bacto Yeast Extract 4 ml 50% (w/v) glycerol 20 ml 50× phosphate 10 ml 10% (w/v) $(NH_4)_2SO_4$ 1 ml 1 M $MgSO_4$
Luria Broth (LB)	10 g Bacto tryptone 5 g Bacto yeast extract 5 g NaCl
Minimal Medium	20 ml 50× phosphate 10 ml 10% (w/v) $(NH_4)_2SO_4$ 0.41 ml 1 M $MgSO_4$ 10 ml 20% carbon source

Solution	Components (per L)
NGM agar	3 g NaCl 2.5 g peptone 20 g agar 972 ml H_2O [5 mg/ml cholesterol in EtOH] [1 ml 1 M $MgSO_4$] [25 ml 1 M phosphate buffer (pH 6.0)] [1 ml 1 M $CaCl_2$]
Pel Agar	16 g Bacto agar 5 ml 20% (w/v) Bacto Yeast Extract 10 ml 10% (w/v) $(NH_4)_2SO_4$ 1 ml 1 M $MgSO_4$ 10 ml 50% (w/v) glycerol 250 ml 2% polygalacturonic acid 20 ml 50× phosphate
Pel Phosphate Buffer	15 g Na_2HPO_4 0.7 g $NaH_2PO_4.H_2O$ (pH 8.0)
50× Phosphate	350 g K_2HPO_4 100 g KH_2PO_4 (pH 6.9-7.1)
Phosphate buffered saline (PBS)	1 PBS tablets (Gibco) dissolved in 100 ml H_2O to provide: 8 g/L NaCl 0.2 g/L KCl 1.15 g/L Na_2HPO_4 0.2 g/L KH_2PO_4 pH 7.3
Tryptone Swarm Agar (TSA)	10 g Bacto tryptone 5 g NaCl 3 g Bacto agar

Solution	Components (per L)
lacZ gene expression assay	
Z buffer	8.52 g Na_2HPO_4 anhydrous 6.24 g $NaH_2PO_4.2H_2O$ 0.75 g KCl 0.25 g $MgSO_4.7H_2O$ 0.7 ml β-mercaptoethanol (add just before use) (pH 7.0)
Phage work	
Phage Buffer	10 mM Tris Base 10 mM $MgSO_4$ 0.01% (w/v) gelatine (pH 7.4)
DNA work	
6× DNA loading buffer	5 mg/ml orange G 50% glycerol
Solution A	9.9 ml 1 M $MnCl_2$ 49.5 ml 1 M $CaCl_2$ 198 ml 50 mM MES 742.6 ml H_2O
Solution A + 15% glycerol	10 ml 1 M $MnCl_2$ 50.1 ml 1 M $CaCl_2$ 200.5 ml 50 mM MES 300.6 ml 50 % glycerol 438.6 ml H_2O
RNA work	
Ammonium acetate 5M	385.4 g NH_4OAc 900 ml H_2O

Solution	Components (per L)
Binding buffer	20 mM Tris pH 8.0, 100 mM NH_4Cl 50 mM NaCl 50 mM KCl DMDC-treated H_2O
Loading dye	0.25% w/v bromophenol blue 30% glycerol DMDC-treated H_2O
Lysis buffer for RNA extraction using hot-phenol	10 mM sodium acetate pH 5.2 0.15 M sucrose 2% SDS DMDC-treated H_2O pH 4.8
<u>Gel electrophoresis</u>	
Agarose gel mix	1.0% agarose in TBE 500 ng Ethidium Bromide per ml gel
Native polyacrylamide gel (5%, 10 ml)	7.27 ml H_2O 1.66 ml 30% 29:1 acrylamide/bisacrylamide 1.0 ml 5× TBE buffer 0.07 ml 10% ammonium persulphate 0.01 ml TEMED
50× TAE buffer	242 g Tris base 57.1 ml glacial acetic acid 100 ml 0.5 M EDTA (pH 8.0)
5× TBE buffer	450 mM Tris-Cl pH 8.0 450 mM boric acid 10 mM EDTA pH 8.0

Solution	Components (per L)
1D SDS-PAGE	
Coomassie stain	10% Methanol 7% acetic acid 0.25 g Coomassie R-250 dye per 100 ml
Coomassie de-stain	10% MeOH 7% acetic acid
Resolving gel mix (10% gel, 10 ml)	4.0 ml H_2O 3.3 ml 30% acrylamide/bisacrylamide 2.5 ml 1.5 M TrisHCl (pH 8.8) 0.1 ml 10% SDS 0.1 ml 10% ammonium persulphate 0.004 ml TEMED
Stacking gel mix (10 ml)	6.8 ml H_2O 1.7 ml 30% acrylamide/bisacrylamide 1.25 ml 1 M TrisHCl (pH 6.8) 0.1 ml 10% SDS 0.1 ml 10% ammonium persulphate 0.01 ml TEMED
SDS-PAGE running buffer	25 mM Tris 0.2 M Glycine 0.1% SDS
4× SDS sample buffer	40 mM TrisHCl (pH 6.8) 40% glycerol 4 mM EDTA 2.5 % SDS 0.2 mg/ml bromophenol blue

[a] Items in brackets are added as sterile solutions after autoclaving.

Table 2.3. Bacterial strains, phage, and plasmids used in this study.

Strain/plasmid	Genotype/phenotype	Reference
Escherichia coli		
BW20767	RP4-2-*tet*:Mu-1*kan*::Tn7 integrant *leu-63*::IS*10 recA1 creC510 hsdR17 endA1 zbf-5 uidA*(Δ*Mlu*I):*pir*$^{+}$ *thi*	(Metcalf et al., 1996)
CC118 *λpir*	*araD, Δ(ara, leu), ΔlacZ74, phoA20, galK, thi-1, rspE, rpoB, argE, recA1, λpir*	(Herrero et al., 1990)
DH5α	F^{-}, Φ80*ΔdlacZM15, Δ(lacZYA-argF)U169, endA1, recA1, hsdR17 (rκ^{-}mκ^{+}), deoR, thi-1, supE44, λ^{-}, gyrA96, relA1*	Gibco/BRL
ESS	β-lactam super sensitive indicator strain	(Bainton et al., 1992)
β2163	(F^{-}) RP4-2-Tc::Mu Δ*dapA*::(*erm-pir*) [Km^{R} Em^{R}]	(Demarre et al., 2005)
Serratia		
ATCC 39006	Wild type (Car^{+}, Pig^{+})	(Bycroft et al., 1987)
LacA (WT)	Lac^{-} derivative of ATCC 39006, made by EMS mutagenesis	(Thomson et al., 2000)
Mutants derived from LacA		
HSPIG67	*pigP*::mini-Tn*5lacZ1*, Km^{R}; random transposon mutant	(Fineran et al., 2005b)
ISTSO4	*smaI*::mini-Tn*5*Sm/Sp, *pigX*::miniTn*5lacZ1*, Sp^{R}, Km^{R}	(Fineran et al., 2005b)
MCP2L	*pigA*::mini-Tn*5lacZ1*, Km^{R}; insertion is 44 bp after the start codon	(Crow, 2001; Slater et al., 2003)
MCA54	*carA*::mini-Tn*5lacZ1*, Km^{R}; insertion is 93 bp after the start codon	(Thomson et al., 2000; Crow, 2001)
MCR14	*carR*::mini-Tn*5lacZ1*, Km^{R}	(Slater et al., 2003)
NMW1	*prf*::Tn-KRCPN1*lacZ1*, Km^{R}	This study
NMW2	*prf*::Tn-KRCPN1*lacZ1*, Km^{R}	This study

Strain/plasmid	Genotype/phenotype	Reference
NMW3	*prf*::Tn-KRCPN1*lacZ1*, Km^R	This study
NMW4	*prf*::Tn-KRCPN1*lacZ1*, Km^R	This study
NMW5	*prf*::Tn-KRCPN1*lacZ1*, Km^R	This study
NMW6	*prf*::Tn-KRCPN1*lacZ1*, Km^R	This study
NMW7	*rsmA*::Tn-KRCPN1*lacZ1*, Km^R	(N. Williamson, in preparation)
NMW8	*Δhfq*	This study
NMW9	*hfq::cat*	This study
NMW10	*hfq::cat, rsmA::*Tn-KRCPN1*lacZ1,* Cm^R, Km^R	This study
NMW11	*vfmE*::*cat*, *prf*::Tn-KRCPN1*lacZ1*, Cm^R, Km^R	This study
NMW12	*hfq*::*cat*, *prf*::Tn-KRCPN1*lacZ1*, Cm^R, Km^R	This study
NMW14	*hfq*::*cat*, *pigA*::mini-Tn*5lacZ1*, Cm^R, Km^R	This study
NMW15	*hfq*::*cat*, *smaR*::*lacZ1*, Cm^R, Km^R	This study
NMW16	*hfq*::*cat*, *carA*::mini-Tn*5lacZ1*, Cm^R, Km^R	This study
NMW17	*hfq*::*cat*, *carR*::mini-Tn*5lacZ1*, Cm^R, Km^R	This study
NMW18	*hfq*::*cat*, *pigW*::mini-Tn*5*Sm/Sp, Cm^R, Sp^R	This study
NMW19	*hfq*::*cat*, *pigQ*::mini-Tn*5*Sm/Sp, Cm^R, Sp^R	This study
NMW20	*hfq*::*cat*, *pigP*::mini-Tn*5lacZ1*, Cm^R, Km^R	This study
NMW21	*hfq*::*cat*, *rap*::mini-Tn*5lacZ1*, Cm^R, Km^R	This study
NMW22	*hfq*::*cat*, *pstS*::mini-Tn*5*Sm/Sp, Cm^R, Sp^R	This study
NMW23	*hfq*::*cat*, *pigX*::mini-Tn*5*Sm/Sp, Cm^R, Sp^R	This study
NMW24	*hfq*::*cat*, *smaI*::mini-Tn*5*Sm/Sp, Cm^R, Sp^R	This study
NMW25	*hfq*::*cat*, *smaI*::mini-Tn*5lacZ1*, *smaR*::cat, Cm^R, Km^R	This study
NMW25	*rpoS*::Tn-*uidA, cat,* Cm^R	This study
NMW26	*rpoS*::Tn-*uidA, cat, Δhfq*, Cm^R	This study
NMW27	*rpoS*::Tn-*uidA, cat, pigQ*::mini-Tn*5*Sm/Sp, Cm^R, Sp^R	This study
NMW28	*rpoS*::Tn-*uidA, cat, pigW*::mini-Tn*5*Sm/Sp, Cm^R, Sp^R	This study
NMW29	*rpoS*::Tn-*uidA, cat, pigP::*mini-Tn*5lacZ1,* Cm^R, Km^R	This study
NMW30	*rpoS*::Tn-*uidA, cat, pigA::*mini-Tn*5lacZ1,* Cm^R, Km^R	This study

Strain/plasmid	Genotype/phenotype	Reference
NMW31	*rpoS*::Tn-*uidA, cat, carA::*mini-Tn*5lacZ1,* Cm^R, Km^R; pink phenotype	This study
NMW32	*rpoS*::Tn-*uidA, cat, carA::*mini-Tn*5lacZ1,* Cm^R, Km^R; red phenotype	This study
NMW33	*rpoS*::Tn-*uidA, cat, rsmA::*mini-Tn*5lacZ1,* Cm^R, Km^R	This study
NWA19	*lacA*, Δ*pigC*	This study
NWX21	*flhC*::Tn-DS1028, *pigX*::mini-Tn*5*Sm/Sp, *carA*::mini-Tn*5lacZ1,* Cm^R, Sp^R, Km^R	(Williamson et al., 2008)
PIG17S	*pigQ*::mini-Tn*5*Sm/Sp, Sp^R; derivative of HSPIG17, constructed by transposon exchange	(Fineran et al., 2005b)
PIG62S	*pigW*::mini-Tn*5*Sm/Sp, Sp^R; derivative of HSPIG62, constructed by transposon exchange	(Fineran et al., 2005b)
RAPL	*rap*::mini-Tn*5lacZ1*, Km^R; derivative of RAPS, constructed by transposon exchange	(Fineran et al., 2005b)
ROP2	*pstS::*mini-Tn*5*Sm/Sp, Sp^R	(Slater et al., 2003)
ROP4S	*pigX*::mini-Tn*5*Sm/Sp, Sp^R; derivative of ROP4, constructed by transposon exchange	(Fineran et al., 2005b)
SP21	*smaR*::*lacZ*, Δ*pigC*, Km^R	This study
VFME	*hfq*::*cat*, Cm^R	(N. Williamson, unpublished)
Phage		
ϕOT8	*Serratia* generalized transducing phage	(Evans et al., 2010)
Vector		
pBlueScript II KS+	Cloning vector, ColE1 replicon, Ap^R	Stratagene
pFAJamp	pFAJ1700 (Dombrecht et al., 2001) modified to remove the Tetracycline resistance cassette. Digested with BseRI and re-ligated.	(J.P. Ramsay, unpublished)
pGFP	pQE80oriT carrying GFP amplified from pGreenTIR (Miller and Lindow, 1997)	(J.P. Ramsay unpublished)

Strain/plasmid	Genotype/phenotype	Reference
pKNG101	Marker exchange suicide vector, *sacBR*, mobRK2, oriR6K, Sm^R	(Kaniga et al., 1991)
pKRCPN1	Delivery plasmid for transposon KRCPN1 which contains *lacZ1*, Km^R	(K. Robertson unpublished)
pQE-80L	Expression vector for native or N-terminal hexahistidine proteins, Ap^R	Qiagen
pQE80oriT	pQE80L modified to contain the oriT site	(J.P. Ramsay, unpublished)
pNB1	S39006 *hfq* including up and downstream regions cloned into pBluescript	This study
pNB2	S39006 Δ*hfq* including up and downstream regions cloned into pBluescript	This study
pNB3	S39006 *hfq::cat* including up and downstream regions cloned into pBluescript	This study
pNB5	*hfq* deletion construct from pNB2 subcloned into pKNG101	This study
pNB6	*hfq* deletion construct from pNB3 subcloned into pKNG101	This study
pNB8	*prf* promoter fragment 1 cloned into pRW50	This study
pNB9	*prf* promoter fragment 2 cloned into pRW50	This study
pNB10	*prf* promoter fragment 3 cloned into pRW50	This study
pNB11	*prf* promoter fragment 4 cloned into pRW50	This study
pNB12	*prf* promoter fragment 5 cloned into pRW50	This study
pNB18	S39006 *hfq* with its native promoters cloned into pFAJ1700	This study
pNB22	*C. rodentium hfq* with its native promoters cloned into pFAJ1700	This study
pNB35	*E. coli hfq* with its native promoters cloned into pFAJ1700	This study
pNB36	S39006 *hfq* cloned under the T5 promoter in pQE80oriT	This study
pNB37	*C. rodentium hfq* cloned under the T5 promoter in pQE80oriT	This study
pNB38	*E. coli hfq* cloned under the T5 promoter in pQE80oriT	This study
pNB39	S39006 *prf* gene cloned into pQE80oriT	This study

Strain/plasmid	Genotype/phenotype	Reference
pNB42	S39006 *rpoS* with its native promoter cloned into pFAJamp	This study
pNB43	S39006 *rpoS* cloned under the T5 promoter in pQE80oriT	This study
pNB44	*E.coli rpoS* gene cloned under the T5 promoter in pQE80oriT	This study
pNB45	*E.coli rpoS* gene with its native promoter cloned into pFAJamp	This study
pNB46	S39006 or1385 UTR+ORF cloned into pQE80oriT	This study
pNB47	S39006 or1385 UTR cloned into pQE80oriT	This study
pNB48	S39006 or1385 ORF cloned into pQE80oriT	This study
pNB49	S39006 *rprA* cloned into pQE80oriT	This study
pNB50	*E. coli dsrA* cloned into pQE80oriT	This study
pSP58	1 kb fragment containing *smaR* flanking regions in *Spe*I-*Apa*I sites of pBlueScript II KS$^+$, Ap^R	This study
pSP59	*lacZ*-Km^R fragment from pKRCPN1 in *Nhe*I-*Sph*I sites of pSP58, Ap^R, Km^R	This study
pSP60	*Spe*I-*Apa*I fragment from pSP59 in pKNG101, Sp^R, Km^R	This study

2.3 ϕOT8 transduction

2.3.1 Preparation of ϕOT8 lysates

The generalised transducing phage ϕOT8 was used to transduce markers between strains of *Serratia*. 100 μl from a dilution series of a ϕOT8 lysate, usually 10^0-10^{-5}, was added to 4 ml of 0.35% molten LBA and 100 μl of a S39006 overnight culture in order to prepare a lysate of ϕOT8. The mixture was poured onto the surface of an LBA plate, allowed to set and incubated overnight at 30°C. The 0.35% agar containing ϕOT8 was harvested from the plate showing just less than confluent lysis and transferred to a glass universal. The plate was rinsed with 4 ml phage buffer and added to the universal. $NaHCO_3$-treated chloroform (500 μl) was added and the universal was vortexed for 1 min to break up the agar. The ϕOT8 lysate was centrifuged at 2,219 g for 20 min at 4°C

to pellet the agar. The ϕOT8 lysate was decanted into a glass universal, 50 µl of chloroform was added to maintain sterility, and stored at 4°C. In order to determine phage concentration of the lysate, a ϕOT8 dilution series was made and 20 µl from each dilution was spotted onto a top lawn of S39006 and incubated overnight at 30°C. ϕOT8 plaques were counted and the concentration of the lysate expressed as plaque forming units per ml (pfu/ml).

2.3.2 Transduction of S39006 with ϕOT8

Double mutants of S39006 were constructed by infection of a recipient mutant strain by a ϕOT8 lysate grown on the donor mutant strain. A 5 ml culture of the S39006 recipient strain was grown overnight and then pelleted by centrifugation at 2,219 g for 10 min and was resuspended in 1 ml LB. 100 µl of the ϕOT8 lysate was added and incubated for 1 hr at 30°C. The mixture was pelleted by centrifugation at 2,219 g for 10 min and the cells were resuspended in 10 ml LB. The cells were washed twice more with 10 ml LB to reduce the number of ϕOT8 pfu/ml. The transductants were plated onto LBA plates containing the appropriate antibiotic(s) and incubated at 30°C for 48 hr. The newly created strains were streaked out twice more onto LBA plates to eliminate the possibility of ϕOT8 contamination and the appropriate antibiotic resistance was checked before experimental use.

2.4 Recombinant DNA techniques

All molecular biological techniques were performed by standard methods (Sambrook et al., 1989), unless stated otherwise. Oligonucleotide primers were obtained from Sigma-Genosys and are listed in Table 2.4 and Table 2.5.

Table 2.4. Oligonucleotide primers used in this study.

Name	Sequence (5'-3')[a]	RE Site	Notes
InVitroRNA Pro	GAAATTAATACGACTCACTA TAGGGAGACCA		Universal primer with T7 promoter for in vitro transcription; blue is T7 sequence and red is +1 base of transcript
KR31	AAAAACTTTCAGTGCCGCC		pKRCPN1 R out LHS

Name	Sequence (5'-3')[a]	RE Site	Notes
KR36	CTCCTTCATTACAGAAACGGC		pKRCPN1 F out RHS
KR37	GGTTAATTGGTTGTAACACTGGC		pKRCPN1 F out RHS
KR88	AAAATAGGCGTATCACGAGGC		pQE-80L F primer for MCS; 397 bp product
KR89	AGCTCCTGAAAATCTCGC		pQE-80L R primer for MCS
LER1	GCGTGCATAATAAGCCCTACAC		mini-Tn*5*Sm/Sp F out RHS
oNMW1	AATGCCGCACTGGTATC		Sequencing primer for 3' end of *argG* in Serratia
oNMW4-F	CCGACGGCSTCVGGIAAAAC		*hfq* locus identitification and sequencing
oNMW4-R	TCTTTYTCWATTGCCTGAAGCTG		*hfq* locus identitification and sequencing
oNMW5	CGAGACCTACCCTCCGTTAC		sp1 for 5' RACE of *prf*
oNMW6	TTTG^TCGACCCTACCGTGTTTAGTAACGATTTC	*Sal*I	sp2 for 5' RACE of *prf*; upstream of oNMW5
oNMW7	TGTGTACAATTGGCTGATTCTC		Primer walking of *hfq* locus
oNMW8	GCAATTTCTTCCGCTTTACC		Primer walking *hfq* locus
oNMW9	TTAAGG^ATCCCTATTGCAGCAATGAACACTC	*Bam*HI	sp3 for 5' RACE of *prf*; upstream of NMW6
oNMW10	CACCATCTTCTTTCATCGCAGC		For sequencing 3' end of *argG*
oNMW11F	TTAAG^TCGACATCAGTGCCGCTGGTCGTATTCCT	*Sal*I	Primer used for construction of NMW8 and NMW9
oNMW11R	GCAGTGTAGG^ATCCAGTGTGGCAA	*Bam*HI	Primer used for construction of NMW8 and NMW9
oNMW12	CGGTAATGAGCCTTATTCA^GATCTTAGCCATTCTATATTTTCC	*Bgl*II	Primer used for construction of NMW8 and NMW9
oNMW13	GAAAATATAGAATGGCTAA^GATCTGAATAAGGCTCATTACC	*Bgl*II	Primer used for construction of NMW8 and NMW9
oNMW19	GCGACTACATTTATGGCTGGCAG		For sequencing 3' end of *argG*
oNMW20	TTAAA^AGCTTTGTTTAGTAACGATTTCCAGC	*Hin*dIII	pRW50 *prf* primer assay RHS respective to *prf* start codon
oNMW21	TTAAG^AATTCATGCTGTTTTCAGCAATTCAG	*Eco*RI	pRW50 *prf* primer assay LHS respective to *prf* start codon
oNMW22	TTAAG^AATTCTTTTTTTAATTTTGAGTGTTCATTG	*Eco*RI	pRW50 *prf* primer assay LHS
oNMW23	TTAAG^AATTCGGATTTTATCTTACTGATTTTATTTG	*Eco*RI	pRW50 *prf* primer assay LHS

Name	Sequence (5'-3')[a]	RE Site	Notes
oNMW24	TTAAG^AATTCCAATATGCC CTTTTTATTGG	*Eco*RI	pRW50 *prf* primer assay LHS
oNMW25	TTAAG^AATTCCCGGCCGTT ATACTG	*Eco*RI	pRW50 *prf* primer assay LHS
oNMW31-F	CTGATCGATATCCGTGACC		Primer to confirm by PCR the marker exchange to create *hfq* mutant
oNMW31-R	ACCAGATCATGCGGTAACTG		Primer to confirm by PCR the marker exchange to create *hfq* mutant
oNMW36-F	TTAAGG^ATCCCCGACGGCS TCVGGIAAAAC	*Bam*HI	Amplification of *hfq* for cloning into pFAJamp vector to generate pNB36
oNMW36-R	TTAACTGCA^GCCTCTGCCA TGGTAAACC	*Pst*I	Amplification of *hfq* for cloning into pFAJamp vector to generate pNB36
oNMW38-F	TTAAGG^ATCCCCGACGGCC TCAGGCAAAAC	*Bam*HI	*C. rodentium hfq* F primer for cloning
oNMW41	TTAACTGCA^GATACCCTGG CTCCCCGTG	*Pst*I	*C. rodentium hfq* R primer for cloning
oNMW53-F	TTAAA^AGCTTCCGACGGCC TCCGGTAAAAC	*Hin*dIII	*E. coli* K12 *hfq* F primer for cloning
oNMW53-R	TTAAG^AATTCATCGCTGGC TCCCCGTG	*Eco*RI	*E. coli* K12 *hfq* R primer for cloning
oNMW56-F	GATG^AATTCATTAAAGAGG AGAAATTAACTATGGCTAAG GGGCAATC	*Eco*RI	F primer for cloning *hfq* genes from S39006, *C. rodentium*, and *E. coli* into pQE80oriT; rbs in blue
oNMW56-R	GATA^AGCTTTTATTCAGCG TCATCGCTATC	*Hin*dIII	R primer for cloning S39006 *hfq* into pQE80oriT to make pNB36
oNMW57	GATA^AGCTTTTATTCAGTT TCTTCGCTGTCC	*Hin*dIII	R primer for cloning *C. rodentium hfq* into pQE80oriT to make pNB37
oNMW58	GATA^AGCTTTTATTCGGTT TCTTCGCTGTC	*Hin*dIII	R primer for cloning *E. coli hfq* into pQE80oriT to make pNB40
oNMW69-F	ATAGGGAGACCAGAAGTCAA TTAACCGTTGTAGCA		F primer for in vitro transcription of *pigA* transcript; 250 bp; blue overlaps with InVitroRNAPro primer
oNMW69-R	GTATCAAGCGCCCCAG		R primer for in vitro transcription of *pigA* transcript; 250 bp

Name	Sequence (5'-3')[a]	RE Site	Notes
oNMW70-F	ATAGGGAGACCATTGCGATATGTTTTTATCAAGG		F primer for in vitro transcription of *smaR* transcript; 301 bp transcript; blue overlaps with InVitroRNAPro primer
oNMW70-R	CCTGAGCTGATCATGATATTTTC		R primer for in vitro transcription of *smaR* transcript; 301 bp transcript
oNMW71-F	ATAGGGAGACCAGCCGTGCCCTATGACAA		F primer for in vitro transcription of *smaI* transcript; 266 bp transcript; blue overlaps with InVitroRNAPro primer
oNMW71-R	GTTTCTATAAACCGCAAACTGC		R primer for in vitro transcription of *smaI* transcript; 266 bp transcript
oNMW72-F	ATAGGGAGACCAATTAGAGTCATGATGTTACATATCAATAAT		F primer for in vitro transcription of *carR* transcript; 339 bp transcript; blue overlaps with InVitroRNAPro primer
oNMW72-R	GAGATCGTTATCATCCCAGG		R primer for in vitro transcription of *carR* transcript; 339 bp transcript
oNMW73-F	ATAGGGAGACCAAGAATTAAAAATCGATCTCTTCGT		F primer for in vitro transcription of *carA* transcript; 252 bp transcript; blue overlaps with InVitroRNAPro primer
oNMW73-R	TCCGGCTACCGTTTTCAG		R primer for in vitro transcription of *carA* transcript; 252 bp transcript
oNMW74-F	ATAGGGAGACCATGGGTAGCACCGGAACC		F primer for in vitro transcription of S39006 *rpoS* transcript (250 bp) and *E. coli rpoS* transcript (260 bp); blue overlaps with InVitroRNAPro primer
oNMW74-R	CTCTTCTTCTGTCAGCGCT		R primer for in vitro transcription of S39006 *rpoS* transcript; 250 bp transcript
oNMW79	GTTCCTGTTCTACTAAGGCCTT		R primer for in vitro transcription of *E.coli rpoS* transcript; 260 bp transcript

Name	Sequence (5'-3')[a]	RE Site	Notes
oNMW82-F	GATG^AATTCCTTCGCCAAG CTCTTCC	*Eco*RI	F primer for cloning S39006 *rpoS* into pFAJamp (1879 bp, pNB42)
oNMW82-R	GATA^AGCTTTTATTCACGA AACAGCTCTTCA	*Hin*dIII	R primer for cloning S39006 *rpoS* into pFAJamp (1879 bp, pNB42) and pQE80oriT (993 bp, pNB43)
oNMW83	GATG^AATTCATTAAAGAGG AGAAATTAACTATGAGCCAA AGTACGCTGA	*Eco*RI	F primer for cloning S39006 *rpoS* into pQE80oriT (993 bp, pNB43); rbs in blue
oNMW84-F	ATG^AATTCATTAAAGAGGA GAAATTAACTATGAGTCAGA ATACGCTGAAA	*Eco*RI	F primer for cloning *E. coli* *rpoS* into pQE80oriT (993 bp, pNB44); rbs in blue
oNMW84-R	GATA^AGCTTTTACTCGCGG AACAGCG	*Hin*dIII	R primer for cloning *E. coli* *rpoS* into pQE80oriT (993 bp, pNB44)
oNMW85	GATG^AATTCCAGCCAGTAG CTCAGCAG	*Eco*RI	F primer for cloning *E. coli* *rpoS* into pFAJamp (1928 bp, pNB45)
oNMW91-F	ATG^AATTCATTAAAGAGGA GAAATTAACTACAGATGAGC TCGAATAGTCG	*Eco*RI	F primer for cloning S39006 or1385 UTR + ORF into pQE80oriT (1105 bp, pNB46); rbs in blue
oNMW91-R	GATA^AGCTTTCAGCGCCCA CAAAAG	*Hin*dIII	R primer for cloning S39006 or1385 UTR + ORF into pQE80oriT (1105 bp, pNB46)
oNMW92	GATA^AGCTTCTTTTTGAGA AGCGCATACA	*Hin*dIII	R primer for cloning S39006 or1385 UTR into pQE80oriT (415 bp, pNB47) with oNMW91-F as F primer
oNMW93	GATG^AATTCATTAAAGAGG AGAAATTAACTATGTGGTCT TCTGACGCAAC	*Eco*RI	F primer for cloning S39006 or1385 ORF (885 bp, pNB48) with oNMW91-R as R primer; rbs in blue
oNMW94-F	GATC^AATTGCCGGCGCTAT TGTTAACGTA	*Mfe*I	F primer for cloning S39006 RprA into pQE80oriT (pNB49); 111 bp amplicon
oNMW94-R	GATG^GATCCATAAAAAAAA GCCCATCTCAGGG	*Bam*HI	R primer for cloning S39006 RprA into pQE80oriT (pNB49); 111 bp amplicon
oNMW95-F	GATC^AATTGAACACATCAG ATTTCCTGGTG	*Mfe*I	F primer for cloning S39006 DsrA into pQE80oriT (pNB50); 94 bp amplicon

Name	Sequence (5'-3')[a]	RE Site	Notes
oNMW95-R	GATG^GATCCAATAAAAAAA TCCCGACCCTGA	*Bam*HI	R primer for cloning S39006 DsrA into pQE80oriT (pNB50); 94 bp amplicon
NW128	GATAATAAGCGGATGAATGG CAG		pDS1028-*uidA*-*cat* F out RHS
NW198	GAAGGCTGGGTGGCTAATAA G		*rsmA* F primer
NW221	ACA^GATCTCGGCTATTTAA CGACCC	*Bgl*II	Amplifcation of *cat* gene from pACYC184; F primer
NW222	TAA^GATCTCCGGAAGATCA CTTCGCAG	*Bgl*II	Amplifcation of *cat* gene from pACYC184; R primer
PF5	CTGACTGCAGGCATCATACA TACG		*pigQ* R primer
PF7	GCTAAAGCTTTATGAACCTT TTG		*pigQ* F primer
PF39	CAGCACCTTCACGTAGTTGC		*pigW* R primer
PF46	GCAAGGACGAGAATTTCCC		pRW50 sequening primer (faces *Eco*RI site)
PF47	CTGGCATTTGGTCAGCG		pRW50 sequening primer (faces *Hin*dIII site)
PF55	GCATGAATTCGATCATTTAC CGTCAC		*pigW* F primer
PF100	CCCAGTCACGACGTTG		pKRCPN1 R out LHS
PF106	GACCACACGTCGACTAGTGC N_{10}AGAG		Random primed PCR Primer 1
PF107	GACCACACGTCGACTAGTGC N_{10}ACGCC		Random primed PCR Primer 2
PF108	GACCACACGTCGACTAGTGC N_{10}GATAC		Random primed PCR Primer 3
PF109	GACCACACGTCGACTAGTGC		Random primed PCR Adapter primer
PFAJ1700-F	AGGTTTCCCGACTGGAAAGC		pFAJ1700 F primer for MCS
PFAJ1700-R	GGACAACTGCTACGCAGATG TC		pFAJ1700 R primer for MCS
PKNG101A	CAACTTAACGTAAAAACAAC TTCAGA		pKNG101 primer *Sal*I/*Bam*HI end of MCS
PKNG101B	TACACTTCCGCTCAGGTCCT TGTCCT		pKNG101 primer *Spe*I end of MCS
PQEF2	TCGTCTTCACCTCGAGAAAT C		pQE-80L F primer for MCS
PQER	GTCATTACTGGATCTATCAA CAGG		pQE-80L R primer for MCS
SP2	GCCTTGATGTTACCCGAGAG C		mini-Tn*5*Sm/Sp R out of LHS

Name	Sequence (5'-3')[a]	RE Site	Notes
SP203	ACATACTAGTACCATTATTA CGAATAGTCAATAATC		Primer used to construct strain SP21
SP204	TTGGCATTGGTAACACCTAG GCATGCTTTGCTAGCATTAG ACACCCAATTATCTCCTT		Primer used to construct strain SP21
SP205	ACATGCTAGCAAAGCATGCC TAGGTGTTACCAATGCCAA		Primer used to construct strain SP21
SP206	ACATGGGCCCTAGAAACAAG ATTCCCTAACATG		Primer used to construct strain SP21
SP209	ACATGCTAGCAGCAGCGATT GATGCAGGAG		Primer used to construct strain SP21
SP210	ACATGCATGCATGCTCTGCC AGTGTTACAAC		Primer used to construct strain SP21
T3	ATTAACCCTCACTAAAGGGA		pUC or pBluescript sequencing primer
T7	TAATACGACTCACTATAGGG		pUC or pBluescript sequencing primer

[a] The symbol "^" indicates the point of digestion by the restriction enzyme in the sequence

Table 2.5. Oligonucleotide primers used for qRT-PCR.

Gene	Sequence (5'-3')	Name
16S rRNA	GGCCTTCGGGTTGTAAAGTC	3916SF
16S rRNA	GCTTTACGCCCAGTCATTC	3916SR
gyrB	CGTCTGCGTGAGTTGTCATTC	oNMW28-F
gyrB	CGGTGTCTTGTTACGGTTGAG	oNMW28-R
pigA	GGATGCCATTCAAACTTTTGG	oNMW76-F
pigA	TTGGTACCGGAGAAAATTCGA	oNMW76-R
pigB	TGATGATGGCTGTCTTTGTG	oNMW27-F
pigB	GTGGTCGGGAAGGTTCTC	oNMW27-R
carA	GGAGAATTCTGCTTCTTGATTG	qcarAf
carA	CTGACAAACCCGATCTTCAC	qcarAr
carR	CCATTTTCCTGGGATGATAAC	qcarRf
carR	TAGTCAGGATCGCCAGATTG	qcarRr
smaI	TGCCTGTGGATGAGCATAAC	oNMW75-F
smaI	TCCCTGACGCAGAATGATAG	oNMW75-R
smaR	GACAAATCCTGCGATGATGA	oNMW65-F
smaR	CCATGCTTGCCCAGTAAAGT	oNMW65-R
luxS	TGGCGTGGAAGTTGTTGATA	oNMW66-F

Gene	Sequence (5'-3')	Name
luxS	CATTGAGCTCCGGGATTTTA	oNMW66-R
rpoS	AGGATTCCGCTTTTCAACCT	oNMW61-F
rpoS	CTATCGAGCTGTTCGGCAAT	oNMW61-R
rsmA	GCGAAACCCTCATGATTG	qrsmAf
rsmA	CAATAAGATGTTGGCTGAGACTTC	qrsmAr
rhlA	GGTGAGTCACTGGAACATAAT	RHLAF
rhlA	ACAAACCACCCCATGAAAT	RHLAR
flhC	CGAAAGTGAAACGCAATTAAG	FLHCF
flhC	TACCGCCTCAACGCTACTAC	FLHCR

2.4.1 DNA separation, purification, digestion and ligation

Plasmid DNA was purified using the QIAprep spin mini-prep kit (Qiagen) and genomic DNA was purified using the DNeasy kit (Qiagen), both according to the manufacturer's instructions. DNA concentration was quantified using a NanoDrop ND-100 spectrophotometer (NanoDrop Technologies, Wilmington, DE). DNA was eluted in H_2O and plasmid DNA stored at –20°C and genomic DNA at 4°C. DNA was purified from agarose gels, PCR amplifications, or digestions using the QIAGEN QIAquick or QIAEX II gel extraction kits depending on the size of DNA fragment purified.

DNA molecules were separated by 1% agarose gel electrophoresis using 1× TAE running buffer and DNA stained with 500 ng/ml of ethidium bromide in accordance with standard procedures (Sambrook et al., 1989). A 6× loading buffer was added to samples prior to loading in the gel and a linear DNA ladder (New England Biolabs) was included in every gel. Gels were visualised using a Gene Genius Bio-Imaging System (Syngene, Synoptics Ltd.).

DNA was digested with restriction endonuclease enzymes according to the supplied instructions. 1 μg of plasmid vector was digested in preparation for use in ligation reactions. Digested DNA for use in ligation reactions was analyzed and separated using 1% agarose gel electrophoresis and the appropriate band was cut out and the DNA purified. For DNA ligations, the concentration of vector and insert DNA was determined using the NanoDrop and a 1:4 molar ratio of vector to insert was used to set up the ligation reactions. The ligation reactions (20-30 μl) were performed with

T4 DNA ligase (NEB) and incubated overnight at room temperature. Moles of DNA were approximated using the equations:

Moles DNA = vol µL DNA × conc DNA ng/µL × (1 mole/ M.W. DNA)

[1]Molecular Weight of dsDNA = (# nucleotides × 607.4) + 157.9

Restriction enzyme digestion (scalable from 20-50 µl) with an example of a 20 µl reaction volume:

Reagents:	**Volume (µl)**
H_2O	x[1]
RE Buffer 10×	2
10× BSA	2
DNA (~1 µg of plasmid)	x[1]
RE 1	1
RE 2	1
Total	20

Ligation reaction (scalable from 20-50 µl) with an example of a 20 µl reaction volume:

Reagents:	**Volume (µl)**
H_2O	x[1]
Insert	x[1]
Vector	x[1]
LigaseBuffer 10×	2
T4 Ligase	1
Total	20

[1]Volume of component is adjustable as appropriate with regard to the total volume of the reaction.

2.4.2 Polymerase chain reaction (PCR)

PCRs were performed using either Taq polymerase (NEB or Sigma), when low fidelity was required, or Phusion high-fidelity polymerase (Finnzymes). Reactions were performed in a MJ Research PTC-200, Peltier Thermal Cycler, DNA Engine, or Veriti Thermal Cycler.

Annealing temperatures and extension times were adjusted according to the requirements of the specific reaction and polymerase enzyme. In general, the annealing

[1] From technical resources, Ambion

temperature was set to 3°C higher than the melting temperature of the primers when using Phusion, versus 5°C lower than the melting temperature when using Taq. Phusion is optimised for fast polymerisation. Example reaction mixtures and the accompanying PCR program are shown below for each enzyme. The total volume was adjusted, scaling each component accordingly.

Colony PCR was used to screen colonies by placing a small sample of cells picked with a pipette tip in 50 µl of water and heating to 99°C for 10 minutes to liberate DNA. Cellular debris was pelleted by centrifugation for 10 min in a tabletop centrifuge for 10 minutes at maximum speed and 2 µl of the liquid was added as template DNA to a 10 µl PCR reaction.

Taq reaction conditions (scalable from 10-20 µl):

Reagents:	**Volume (µl)**
H_2O	6.7
ThermoPol Buffer 10×	1
dNTPs (10 mM)	0.2
F primer (10 µM)	0.5
R primer (10 µM)	0.5
Template DNA	1
Taq polymerase	0.1
Total	10

Stage	**Temp (°C)**	**Time**	**Cycles**
1. Initial denaturation	95	2 min	1×
2. Denaturation	95	30 s	
3. Annealing	55	30 s	30×
4. Extension	72	1 min/kb	
5. Final extension	72	7 min	1×
6. End	12	Forever	Forever

Phusion Taq reaction conditions (scalable from 20-50 µl):

Reagents:	**Volume (µl)**
H_2O	12.4
HF buffer 5×	4
dNTPs (10 mM)	0.4
F primer (10 µM)	1
R primer (10 µM)	1
Template DNA	1
Phusion Taq	0.2
Total	20

Stage	**Temp (°C)**	**Time**	**Cycles**
1. Initial denaturation	98	2 min	1×
2. Denaturation	98	10 s	25×
3. Annealing	(+3°C of primer with lowest Tm)	15 s	
4. Extension	72	15s/kb for plasmid DNA or 30s/kb for gDNA template	
5. Final extension	72	5 min	1×
6. End	12	Forever	Forever

2.4.3 Random-primed PCR

Random primed PCR was used to amplify sequences from a defined into an undefined region of DNA. For example, random primed PCR was used in this study to amplify across a transposon insertion site where the transposon sequence was defined but the insertion site in the chromosome unknown. For transposon, TnKRCPN1, primer oKR36 was used as the transposon specific primer facing out of the right hand side. PCR was performed using a random primer mix (PF106, PF107 and PF108) and the transposon specific primer on crude DNA extracts from colonies. A second PCR was performed using an aliquot of DNA from the first PCR with the nested transposon specific primer, KR37, and the adapter primer (PF109) which overlapped the 5' end of the random primers. The resulting mix of PCR fragments were purified and sequenced with the nested gene specific primer, KR37. Sequenced products allowed for the identification of transposon insertion site in the chromosome (see Chapter 3.2.1).

2.4.4 Quantitative PCR

Quantitative reverse transcription PCR (qRT-PCR) was used to measure mRNA transcript levels, and hence changes in expression. Oligonucleotides used as primers for quantitative PCR are listed in Table 2.5 and were designed to amplify products between 100-200 bp. The synthesis of cDNA and quantitative PCR was performed as described (Williamson et al., 2008). cDNAs were synthesized from total RNA by reverse transcription using random hexamers. Minus reverse transcriptase (RT) controls were performed as a control for DNA contamination. A 5 µl aliquot of template cDNA was used in a 25 µl reaction containing 1× SYBR Green PCR Master Mix (Applied Biosystems) and 10 pmol of the appropriate primers. PCR amplification was performed using an ABI PRISM 7300 real-time PCR system. Amplification was performed with 40 cycles of 95°C (15 s) and 60°C (1 min). Fluorescence data were processed using SDS software (ABI) to produce threshold cycle (Ct) values for each sample. Relative gene expression was obtained using 16S rRNA as the control with mRNA/16S rRNA = 1 in the WT. Experiments were performed in triplicate.

2.4.5 5'RACE

5' RACE (rapid amplification of cDNA ends) was performed using the Roche 5'/3' RACE 2nd generation kit, following the manufacturer's specifications. The following specific primers (SP1-3) were used for mapping the transcriptional starts of *prf* (NW210, oNMW5 and oNMW6). Specific primers oNMW5, oNMW6, and oNMW9 were used in the second attempt of 5'RACE of *prf*. The PCR products were separated by 1% agarose gel electrophoresis and visualized using a UV imager. Dominant DNA bands were cut out, purified, and sequenced.

2.4.6 DNA sequencing and sequence analysis

DNA sequencing of plasmid and amplicon DNA was performed at the DNA Sequencing Facility, Department of Biochemistry, University of Cambridge, on an ABI 3730xl DNA analyser. Nucleotide sequence data was analysed using Vector NTI v. 10.0 (Invitrogen) and/or BLAST (Altschul et al., 1997).

2.5 Mutagenesis of S39006

2.5.1 Random transposon mutagenesis of S39006

Random transposon mutagenesis of PIG17S was performed by conjugation with *E. coli* BW20767 carrying plasmid pKRCPN1. 15 μL of an overnight (O/N) culture of PIG17S and the *E. coli* donor strain were mixed and spotted onto an LBA plate and incubated overnight at 30°C. Five more mixtures were also spotted onto the LBA plate. Following conjugation, the mating mixture was scraped off the LBA plate and resuspended in 1 ml of LB and 100 μl aliquots of a 10^{-2} dilution were plated on LBA plates with appropriate antibiotic selection. Plates were incubated for 4-5 days at 30°C and mutant colonies were identified for further analysis using random primed PCR to identify the transposon insertion site (see Chapter 3.2.1). Negative controls of PIG17S and *E. coli* prior to conjugation were also spotted onto the LBA plates.

2.5.2 Marker exchange mutagenesis of *hfq* and *smaR*

Orthologues of the *miaA* and *hflX* genes located, respectively, up- and downstream of *hfq* were aligned using ClustalW. Highly conserved sequences were selected to design degenerate primers, oNMW4-F/R, to amplify by PCR the *hfq* gene from S39006 genomic DNA. oNMW4-F is 1,002 bp upstream of the *hfq* translation start site and oNMW4-R is 1,268 bp downstream of the stop site. The PCR product was cloned into pBluescript to generate plasmid pNB1, and the insert region was sequenced by primer walking. To create the Δ*hfq* and *hfq::cat* mutants, 771 bp upstream and 805 bp downstream of *hfq* was amplified by PCR using primer pairs, oNMW11-F/oNMW12, and oNMW13/oNMW-11R, and ligated in a triple ligation reaction into *Bam*HI-*Sal*I digested pBluescript to create pNB2. oNMW11-F and oNMW11-R contained, respectively, *Bam*HI and *Sal*I restriction sites overhanging on the 5' end, and oNMW12 and oNMW13 were designed to include the first nine nucleotides of the *hfq* coding sequence followed by a *Bgl*II restriction site and a stop codon. A 980 bp fragment containing the *cat* gene encoding resistance to chloramphenicol was amplified by PCR from pACYC184 using primers NW221 and NW222 containing *Bgl*II restriction sites overhanging on the 5' end. The *cat* gene was ligated into the *Bgl*II site in pNB2 to generate pNB3. The Δ*hfq* and *hfq::cat* constructs were excised from pNB2 and pNB3 by

digestion with *Bam*HI and *Sal*I and subsequently cloned into the respective restriction sites into the marker-exchange plasmid pKNG101 to generate pNB5 and pNB6. Allelic exchange using pNB5 and pNB6 was performed using a similar protocol to Kaniga *et al.* (1991). The resulting Δ*hfq* and *hfq::cat* mutants were confirmed by DNA sequencing, and the *hfq::cat* mutation was transduced into a clean S39006 *lacA* genetic background using phage ϕOT8 with Cm^R selection.

The *smaR*::*lacZ* transcriptional fusion was created in a *pigC* mutant of *Serratia* 39006 (NWA19) by Simon Poulter using the allelic exchange plasmid pKNG101, to create strain SP21. The two 500 bp regions flanking the *smaR* gene (including the first nine nucleotides of *smaR*) were PCR amplified from *Serratia* genomic DNA using primers SP203 and SP204 (upstream) and SP205 and SP206 (downstream). The resulting SP203/SP204 and SP205/SP206 PCR products were mixed in equal quantities and used as template in an overlap extension PCR reaction with primers SP203 and SP206, fusing the *smaR* flanking regions and incorporating internal *Nhe*I and *Sph*I restriction sites. This PCR product was then digested with *Spe*I and *Apa*I, and cloned into pBlueScript KS II^+ to create plasmid pSP58. A 4.8 kb fragment containing the *lacZ* gene and a kanamycin resistance cassette was then PCR amplified from plasmid pKRCPN1 using primers SP209 and SP210, digested with *Nhe*I and *Sph*I, and cloned into *Nhe*I-*Sph*I digested pSP58, to create plasmid pSP59. The *Spe*I-*Apa*I fragment from plasmid pSP59 was cloned into the marker-exchange plasmid pKNG101 to create plasmid pSP60. Allelic exchange using pSP60 was performed using a similar protocol to Kaniga *et al.* (1991). A resulting *smaR*::*lacZ* mutant was confirmed by DNA sequencing, and the mutation was transduced into a clean NWA19 genetic background using phage ϕOT8.

2.6 Plasmid constructs of *hfq* for complementation

To complement the *hfq* mutants, the native S39006 *hfq* gene (or its orthologues) was cloned either with its native promoter into pFAJamp or under control of the T5 promoter in pQE80oriT. Because *hfq* was shown to be under the control of three independent promoters within the intergenic region and coding sequence of *miaA* in *E. coli* (Tsui et al., 1996), the coding sequence of *hfq* and a 1 kb region upstream of the

start codon was PCR-amplified from genomic DNA of S39006 (oNMW36-F/R), *C. rodentium* (oNMW38-F, oNMW41), and *E. coli* (oNMW53-F/R), generating plasmids pNB18, pNB22, and pNB35, respectively. The coding sequence of the same *hfq* orthologues were amplified by PCR from genomic DNA of S39006 (oNMW56-F/R), *C. rodentium* (oNMW56-F, oNMW57), and *E. coli* (oNMW56-F, oNMW58), generating plasmids pNB36, pNB37, and pNB38, respectively.

2.7 Total RNA extraction

For qPCR of the *prf* mutant (see Chapter 3.2.5), total RNA was extracted from S39006 grown to early stationary phase using the QIAGEN RNeasy mini kit according to the manufacturer's instructions. Genomic DNA was removed using DNase I (Qiagen) on-column digestion during the extraction procedure. A further round of DNase digestion was performed using TURBO DNase (Ambion), a high efficiency DNase, to completely remove genomic DNA. PCR amplification for 16S rRNA with primers 3916SF/R (Table 2.5) using template from a negative control reverse transcription reaction (i.e. no RT enzyme added) showed that genomic DNA was removed to below PCR-detectable levels.

For RNA-seq of S39006, the hot phenol method was employed to extract total RNA (Mattatall and Sanderson, 1996) used for synthesis of cDNA libraries which were subjected to high-throughput sequencing. This method is recommended by the group of Jörg Vogel which has applied it to perform RNA-seq on many bacterial species (C. Sharma, personal communication) (Sharma et al., 2010). Most commercially available standard column-purification kits do not guarantee isolation of RNA species less than 150-200 nt whereas phenol:chloroform extraction will harvest total RNA including small RNAs. Three replicates of three strains [WT, Δ*hfq* (NMW8), and *rsmA*::Tn (NMW7)] were grown at the same time for RNA extraction. RNA was extracted from cell cultures (2 ml) taken at early stationary phase (10 h, OD_{600} = 2.0 for WT) growth in LB at 30°C and mixed with 4mL RNA Later (Ambion). Cells were resuspended in 3 ml of lysis buffer (10 mM sodium acetate, 0.15 M sucrose, pH 4.8, 2% sodium dodecyl sulfate). To this was added 3 ml phenol (water saturated, pH 4.3) heated to 65°C. The mixture was vortexed, incubated at 65°C for 5 min, then incubated at 0°C for 5 min, followed by centrifugation at 5,000 g for 30 min at 4°C. The aqueous phase (top layer) was

transferred to a new tube and the phenol extraction repeated followed by a final extraction with 3 ml chloroform. The RNA was precipitated overnight in 2.5 volumes of ethanol at 20°C and centrifuged. The RNA pellet was washed with 75% ethanol and resuspended in 100 µl H_2O. TURBO DNase digestion was followed by the addition of 10 µl 10× DNase reaction buffer and 1 µl TURBO DNase (Ambion) and the reaction incubated at 37°C for 30 min. RNA was extracted with an equal volume of chloroform, precipitated with 2.5 volumes of ethanol, dissolved in 50 µl H_2O and stored at -70°C. PCR amplification of 16S rRNA before and after DNase digestion confirmed the removal of genomic DNA to below PCR-detectable levels. RNA was quantified using a Nanodrop ND1000 (NanoDrop Technologies) and the quality evaluated using an Agilent Bioanalyser (Agilent Technologies).

2.8 Transcriptomic studies with RNA-seq

2.8.1 cDNA synthesis

cDNA libraries were prepared for RNA-seq using the the Illumina sequencing platform as described previously but without depletion of 23S and 16S rRNA (Croucher et al., 2009). This method employs the synthesis of single-stranded (ss) cDNAs followed by library preparation for sequencing on the Illumina platform which has been shown to generate strand-specific sequence reads. 1 µg RNA was reverse transcribed using random primers (GE Healthcare) and Superscript III (Invitrogen) at 45°C for two hours. +RT and –RT reactions were assembled as controls. The RNA was hydrolysed by the addition of 1.5 µl 1M NaOH and heat denatured by incubation at 70°C for 20 min, followed neutralisation with 1.5 µl of 1M HCl. Second strand synthesis was omitted in order to retain strand specific sequence information. The +RT reactions were purified using an Illustra AutoSeq G50 spin column (GE Healthcare) and cDNAs delivered to Dr. Robert Kingsley at the Wellcome Trust Sanger Institute (WTSI) for submission to the Illumina sequencing team.

The transcribed genes of 16S rRNA and *pigB* were amplified by PCR with a maximum amplicon of 200 bp as a positive control for reverse transcription. 1 µl from the +RT reaction was used as template. Whereas absence of PCR product using –RT reaction as template confirmed the absence of contaminating genomic DNA. qPCR was

performed to measure relative transcript levels of genes known to be differentially expressed in the *hfq* and *rsmA* mutant samples prepared for RNA-seq (Figure 6.6). Measurement of differential expression by qPCR provided further confirmation that genomic DNA was sufficiently removed from RNA.

cDNA synthesis reaction conditions:

Reagents	**Volume µl**	**Incubation step**
Total RNA (1 µg)	12.73	
Random oligos (5 µg)	1.67	
RNaseOUT	1.0	
Total	15.4	70°C for 10 min followed by 5 min on ice
5× SSIII buffer	6.0	
0.1M DTT	3.0	
dNTP mix (25 mM of C,G,A,T)	0.6	
Nuclease-free water	2.85	
Actinomycin D (1.2 mg/ml)	0.15	
Supercript III RT	2.0	
Total	30.0[a]	42°C for 2 hours
1M NaOH	1.5	
Total	31.5	70°C for 20 min
1M HCl	1.5	
Total	33.0	Reaction applied to Illustra AutoSeq G50 spin column (GE)

[a] Components for the second part of the reaction were prepared in a 10× master mix and aliquoted to each of the nine samples.

2.8.2 Library preparation and Illumina sequencing

Synthesized cDNAs were used for constructing paired-end sequencing libraries for the Illumina GAII platform by the WTSI sequencing team as described previously (Croucher et al., 2009). cDNA samples were first sheared by nebulization (35 psi, 6 min). Duplexes were then blunt ended through an end repair reaction using large Klenow fragment, T4 polynucleotide kinase and T4 polymerase. A single 3′ adenosine moiety was added to the cDNA using Klenow exo− and dATP. Illumina adapters, containing primer sites for flow cell surface annealing, amplification and sequencing, were ligated onto the repaired ends of the cDNA. Gel electrophoresis was used to select for DNA constructs 200–250 bp in size, which were subsequently amplified by 18 cycles of PCR with Phusion polymerase. The libraries were denatured with sodium hydroxide

and diluted to 3.5 pM in hybridization buffer for loading onto a single lane of an Illumina GAII flow cell. Cluster formation, primer hybridization and sequencing reactions were performed according to the manufacturer's recommended protocol (Bentley et al., 2008).

2.8.3 Read mapping, visualization, and differential expression analysis

Sequence reads were mapped to the S39006 genome using the MAQ software by Nicholas Croucher (WTSI). Mapped reads were viewed in BAM view in the Artemis genome browser (Carver et al., 2010). However, due to unknown technical problems, strand-specific information was limited to non-proper paired reads (see Chapter 6.2.3). In order to identify interesting non-coding RNAs, the BAM files for the three WT replicates were cumulatively loaded into BAM view, the filter for hiding non-proper paired reads was applied and the transcriptional profile of the entire genome was scanned by eye. The DEseq package in the R software environment was used to calculate digital gene expression from the RNA-seq data and analyze differential gene expression between the WT and the mutants of *hfq* and *rsmA*. The DEseq package models count data using a negative binomial distribution (Anders and Huber, 2010), and analysis with DEseq was performed by Adam Reid (WTSI). *P* values were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) correction.

2.9 Hfq-RNA gel shift assay

2.9.1 RNA preparation for *in vitro* binding

DNA templates preceded with the T7 promoter sequence were amplified by PCR from genomic DNA using primers listed in Table 2. The 5' UTR of *pigA*, *carA*, *carR*, *smaR*, and *smaI* were amplified from the predetermined transcription start site (TSS) through a portion of the coding sequence. The position of the TSS and 3' end of the *in vitro* transcript relative to the translation start site is *pigA* (-44, +206), *carA* (-42, +210), *carR* (-108, +231), *smaR* (-182, +119), and *smaI* (-51, +215) and the oligonucleotides used for amplification from genomic DNA are, respectively, oNMW69-F/R, oNMW73-F/R, oNMW72-F/R, oNMW70-F/R, and oNMW71-F/R. A region of the 5' UTR and coding

sequence of S39006 *rpoS* (-148, +102) similar to the same region of *E. coli rpoS* RNA (rpoS-S) used in a previous study was amplified using oNMW74-F/R (Updegrove et al., 2008). In order to include the T7 promoter sequence at the 5' end, the PCR products were amplified in a second round of PCR using the oligonucleotide InVitroRNAPro (carrying the T7 promoter sequence) as the forward primer and the reverse primer for each respective primer pair. The oligonucleotides of the original forward primers (F) included additional sequence at the 5' end to hybridize with oligonucleotide InVitroRNAPro. PCR products were purified by spin column purification (Promega) and 5 µl analyzed by agarose gel electrophoresis, showing a single dominant band.

PCR products were then used as template for *in vitro* transcription using T7 RNA polymerase (Fermentas) followed by DNase I digestion (Qiagen) according to manufacturers' protocol. RNA was purified by ammonium acetate precipitation, dissolved in water, and quantified using a Nanodrop ND1000 spectrophotometer. RNA transcripts gave one dominant band when resolved using a 5% polyacrylamide gel.

2.9.2 Preparation of Hfq

Isolation and purification of *E. coli* Hfq was performed as described by Worrall *et al.* (2008) with modifications by Kasia Bandyra in the lab of Ben Luisi (Dept of Biochemistry, Cambridge). The lysate from centrifugation steps was kept at 22°C to minimize precipitation. Obtained protein was fractionated by gel filtration in buffer composed of 50 mM Tris/HCl (pH 8.0), 100 mM NaCl and 100 mM KCl. The protein was provided by K. Bandyra.

2.9.3 Electrophoretic gel mobility-shift assay

Binding reactions between *E. coli* Hfq and *in vitro* transcribed RNA were prepared as described (Updegrove et al., 2008). A 15 µl volume reaction was prepared by combining: 5 µl of 30 nM RNA (10 nM final concentration), 7.5 µl of Hfq protein–to give an appropriate Hfq concentration–or reaction buffer (20 mM Tris–HCl (pH 8.0), 100 mM NH_4Cl, 50 mM NaCl, 50 mM KCl), and 2.5 µl of loading buffer (0.25% bromophenol blue, 30% glycerol). All reagents were in reaction buffer so that final conditions were 20 mM Tris–HCl (pH 8.0), 100 mM NH_4Cl, 50 mM NaCl, 50 mM KCl, 5% glycerol. Reactions were incubated for at least ten minutes at room temperature and

RNA/protein complexes were resolved by electrophoresis on native 5% (w/v) (acrylamide 29:1 (w/w) /bisacrylamide) gels in 0.5× TBE at 100 V for 2 hours. Gels were stained using 1× SYBR Gold (Invitrogen) for 45 min followed by visualization of nucleic acid bands using a UV imager. Gel bands were quantified using GeneTools (Syngene).

2.10 Proteomic studies with iTRAQ

In order to correlate transcriptomic and proteomic changes in the *hfq* mutant, triplicate cultures of WT and *hfq* were grown under the same conditions as samples subjected to RNA-seq. Cells from 25 ml cultures at early stationary phase (10 h, OD_{600} = 2.0 for WT) growth in LB at 30°C were pelleted by centrifugation. Cells were washed with 1 ml phosphate-buffered saline and the cell pellet resuspended in 1.01 ml CHAPS lysis buffer (8 M urea, 4% CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propane-sulphonate), 5 mM magnesium acetate, 10 mM Tris–HCl, pH 8.0) containing 1× protease inhibitor (protease inhibitor cocktail set I, Calbiochem). Cells were sonicated in 10 second bursts (10 seconds on and 20 seconds off) for 5 cycles while on ice to prevent the samples from overheating. The sonication ensured the complete lysis of cells and sheared the genomic DNA to reduce viscosity. Samples were centrifuged to pellet cell debris and insoluble material, and the supernatant was transferred to a new Eppendorf tube. The centrifugation was repeated a second time to ensure removal of insoluble material and soluble protein was stored at -20°C.

The protein concentration of the soluble fraction was determined using the Bio-Rad DC protein assay as described by the manufacturer. A preliminary 2D-DiGE analysis of replicates WT-1 and hfq-1 was performed with the assistance of Renata Feret in the Cambridge Centre for Proteomics (CCP) as described previously in order to observe the level of global changes in protein expression (Coulthurst et al., 2008). Two-dimensional gel electrophoresis separation was performed using a linear pH 3–10 separation in the first dimension and 12% SDS-PAGE in the second dimension. Images were obtained using a Typhoon 9400 (GE Healthcare) and detection of protein spots and altered expression was performed using the DeCyder software package (GE Healthcare). The

result showed significant changes in protein expression with an estimated ~10% of the protein spots showing more than a 2-fold change.

2.10.1 Protein labelling with iTRAQ

Protein samples were precipitated, digested into peptides, and labelled with isobaric tags for relative or absolute quantitation (iTRAQ 4-plex) as previously described (Castrillo et al., 2007). Chilled acetone (1.5 ml) was added to 250 µl of soluble protein sample. The tube was inverted three times and left at -20°C overnight. Precipitated proteins were pelleted by a 10 minute centrifugation at 3,000 rpm, and resuspended in iTRAQ labeling buffer (8 M urea, 2% Triton X100, 0.1% SDS and 25 mM triethyl ammonium bicarbonate (TEAB) pH 8.5). The detergent-compatible BCA protein assay (Pierce, Rockford, IL) was used to determine protein concentration.

iTRAQ labelling was performed using 100 µg protein from each sample or a pooled sample comprising equal amounts of each sample. The pooled protein sample served as an internal reference allowing for a comparison of relative quantitation. First, samples were reduced by the addition of TCEP (4 mM Tris(2-carboxyethyl) phosphine, 20°C, 1 h) and cysteines blocked by the addition of MMTS (8 mM methyl methanethiosulfonate, 20°C 10 minutes). Samples were diluted with 50 mM TEAB (pH 8.5) such that the final urea concentration was below 1 M, digested with trypsin overnight at 37°C (Promega; 2.5 µg added at 0 and 1 h) and lyophilized using a speed vacuum centrifuge.

Each lyophilized sample was resuspended in 100 µl labeling buffer (0.25 M TEAB, 75% ethanol), added to one unit of the corresponding iTRAQ reagent and incubated for 1 h at 20°C. Residual reagent was quenched by adding 100 µl water and incubated for a further 15 minutes at 20°C. The samples belonging to the iTRAQ comparisons were then pooled and lyophilized. Pool A consisted of samples WT-1, hfq-1, WT-2, and the internal pooled sample which were labelled with iTRAQ tags 114, 115, 116, and 117, respectively. Pool B consisted of samples hfq-2, the internal pooled sample, hfq-3, and WT-3 which were labelled with iTRAQ tags 114, 115, 116, and 117, respectively. Sample and pooled replicates were labelled with different tags to take into account the slight differences between the performance of the iTRAQ reporter ions in the mass spectrometer. Lyophilized samples were delivered to Svenja Hester (CCP) who

performed pre-fractionation of the peptide samples using reverse-phase chromotagraphy followed by LC-MS/MS analysis using the Q-Tof Premier tandem mass spectromoter.

2.10.2 Reverse-phase chromatography

After the peptides had been labelled, the next step was to apply reverse-phase chromatography to pre-fractionate peptides according to their property of hydrophobicity at high pH. The lyophilized samples were resuspended in 50 µl ammonium formate (20 mM, pH 10.0). The sample was loaded onto a C18 UPLC column (bridged ethyl hybrid, Waters, 2.1 × 150 mm column dimension, 1.7 µm particle size, 130 Å pore size). The peptides were separated and eluted using a 50 min linear gradient of 0-70% ACN at a flow rate of 0.25 ml/min. Eluted peptides were monitored with a diode array detector with a wavelength range of 200-400 nm and were manually collected in fractions at 2 min intervals. 30 fractions were collected per sample and reduced to dryness using a speed vacuum centrifuge. Fractions 8-26 for sample A and fractions 8-24 for sample B were analyzed by liquid chromatography followed by tandem mass spectrometry (LC-MS/MS). Fractions 25-26 for sample B were omitted because of very low peptide abundance.

2.10.3 LC-MS/MS analysis

Peptide data collection was performed using a Q-Tof Premier tandem mass spectrometer (Waters) as described previously (Karp et al., 2010). Samples were resuspended in 30 µl 0.1% formic acid. Peptides were separated by a second dimension of reverse-phase chromatography at low pH using a nanoAcquity HPLC pump (Waters) and a Waters Atlantis C18 analytical column (nanoAcquity) (Symmetry C18, 3 µM particle size, 75 µM i.d. × 100mm). Peptides (1 µl) were injected using an auto sampler into a 10 µl sample loop and then loaded onto a pre-column (Waters Symmetry C18, 5 µM particle size, 180 µM i.d x 20mm) with 0.1% formic acid at a flow rate of 5 µL/min for 4 minutes. Peptides were concentrated onto the pre-column and residual wash was directed to waste. After this period, the six port valve was switched to allow elution of the peptides at a flow rate of 300 nl/min from the pre-column onto the analytical column. The gradient employed was a 100 min linear gradient of 2-55% ACN. In MS

mode, data were collected from *m/z* 400-1600. The instrument was run in data dependent acquisition mode which means that following each MS scan, generally, the three most intense peptide ions were selected for MS/MS analysis. MS/MS data were collected between *m/z* 50-2000. Briefly, peptides were selected for MS/MS based on charge-state recognition (2+ and 3+). MS/MS data were collected for a total of 3 seconds, before the instrument was switched back to MS mode. Peptides which were selected for MS/MS analysis were excluded from re-selection for MS/MS for 300 s.

2.10.4 MS data analysis and protein quantification

MS data analysis was performed at the CCP. The QTof Premier raw files were converted into three file formats: 1) centroided pkl (peak list) format using ProteinLynx Global Server (PLGS) V2.4, 2) mzML format in order to obtain charge state information, and 3) mzXML format which is a universal MS file format created using Masswolf. The information contained in these three file formats was compiled to create an MGF (Mascot Generic Format) file. The MGF file was searched against the S39006 proteome database using the Mascot search algorithm to yield matches of MS/MS spectra to peptides. The S39006 proteome database was generated from the predicted ORFs in the reference genome (see Chapter 6.2.1 for assembly of the reference genome). Within Mascot, the modifications used were as follows: fixed, iTRAQ 4-plex (K), iTRAQ 4-plex (N-term), methylthio modification (C); variable, oxidation (M), iTRAQ (Y). The MS tolerance was 1 Da, and the MS/MS tolerance was 0.8 Da. Each MGF file was also searched against a decoy proteome database to estimate the false discovery rate.

The search results outputted by Mascot were re-analyzed by Mascot percolator, which is a machine learning algorithm that yields peptide/spectrum matches with a more robust false discovery rate (Brosch et al., 2009). The output of Mascot percolator and the corresponding mzXML file were inputted into iSPY (P. Charles, manuscript in preparation), which matches peptide identifications from the MS/MS spectra with reporter ion peak areas to determine the relative abundance of the peptides. Results were filtered to apply a false discovery rate of 1%.

2.11 Preparation and analysis of flagellar protein extracts

Flagellar protein extracts were prepared from S39006 grown to early stationary phase and analyzed by 1D SDS-PAGE. An aliquot of 25 ml cultures was taken and normalized to an OD = 2.0 in 5 ml. Cells were pelleted by centrifugation for 15 minutes at 4000 rpm (4°C) and resuspended in 5 ml PBS pH 7.4. The flagella were sheared from cells on ice using a T8 disperser/homogenizer (setting 3) for 1 minute. The cells were pelleted by centrifugation and protein in the supernatant was precipitated by the addition of 0.55 ml 100% trichloroacetic acid solution to a final concentration of 10% (w/v). Precipitated protein was pelleted by centrifugation at 12500 rpm (>17,000 rcf) for 30 minutes at 4°C. The pellet was washed with ice cold acetone (1ml) and stored at -20°C.

Protein pellets were resuspended in SDS-PAGE loading buffer (0.05 ml) and 0.02 ml analyzed by 10% 1D SDS-PAGE. A gel of 1 mm thickness, containing 10 lanes, was made and run using the Mini-PROTEAN Tetra cell system (Bio-Rad). The 1D SDS-PAGE gel was run at 80 V, 20 min, to run the samples through the stacking gel, followed by 120 V, 1.5 hr, to separate protein bands. The gel was stained by shaking with Coomassie stain overnight, prior to de-staining with Coomassie de-stain solution overnight. The gel was imaged using a Gene Genius Bio-Imaging System (Syngene, Synoptics Ltd.) and placed in 10% MeOH to decrease contamination. The gel was sent to Len Packman at the Protein & Nucleic Acid Chemistry Facility (Dept of Biochemistry) for identification of protein bands by MALDI-TOF analysis.

2.12 Phenotypic assays

2.12.1 Prodigiosin assay

Assays for prodigiosin were performed as described previously (Slater et al., 2003). One ml samples from bacterial cultures were pelleted by centrifugation at 13,000 rpm for 10 min in a benchtop microcentrifuge and the cell pellets stored at -20°C. The pellets were resuspended in 1 ml of acidified ethanol (4% 1 M HCl in ethanol) by pipetting and then vortexed for 5 sec. The samples were then pelleted by centrifugation at 13,000 rpm for 10 min in a benchtop micro-centrifuge and 950 µl of supernatant was transferred to a 1 ml cuvette. Pig levels were measured at A_{534} in a

Unicam Heλios spectrophotometer with acidified ethanol as a blank. Relative prodigiosin concentration per cell was expressed as (A_{534}/ml/OD_{600}) × 50.

2.12.2 Carbapenem and *N*-AHL plate assay

Assays for carbapenem were performed as described previously (Slater et al., 2003). Detection of *N*-AHLs was performed using the *Serratia* LIS bioassay as described previously (Thomson et al., 2000), but using strain ISTSO4 as the sensor strain. One ml samples from bacterial cultures were pelleted by centrifugation at 13,000 rpm for 10 minutes in a benchtop microcentrifuge. The supernatants (800 μl) were filter-sterilized with a 0.22 μM filter (Millipore) and stored at 4°C until assayed the same day. Bioassay plates were prepared by adding 100 μl of the appropriate sensor strain to 100 ml of 0.75% molten LBA, which was poured over a large LBA bioassay plate and allowed to solidify. *E. coli* strain ESS and *Serratia* strain ISTSO4 were used for the Car and *N*-AHL bioassays, respectively. A cork borer was used to cut wells in the bioassay plate and 200 μl of filter-sterilized culture supernatant was added to the wells. Plates were incubated for 24 hr at 30°C. The area of antibiosis of *E. coli* ESS surrounding the wells was measured for the Car bioassay plate and Car production expressed as the area of antibiosis (cm^2/OD_{600}). The area of pigmentation of *Serratia* ISTSO4 surrounding the wells was measured for the *N*-AHL bioassay plate and *N*-AHL production expressed as the area of pigmentation (cm^2/OD_{600}).

Car production was also assayed from colonies cultured on plate media with a top lawn of ESS. Overnight bacterial cultures were adjusted to an OD_{600} of 1.0 and 3 μl was spotted onto the plates. Halo size was examined after growth for 24 hr at 30°C. Assays for Car production from strains with a plasmid vector were modified by adding a final concentration of 50 μg/ml of potassium clavulanate (Sigma) to the LBA overlay of ESS (Bowden and Salmond, 2006). The clavulanate inhibited the β-lactamase activity which conferred ampicillin resistance to the plasmid vector pQE80oriT. The activity of this enzyme normally degrades Car.

2.12.3 Pectate lyase plate assay

Overnight cultures of bacterial strains were adjusted to an OD_{600} of 0.2, and 5 μl was spotted onto the Pel agar plate. Plates were incubated for 48-72 hr at 30°C and

covered with 7.5% w/v copper acetate solution for 1-2 hr. Pectate lyase activity was indicated by two halos surrounding the test colony, the innermost being darker.

2.12.4 Pectate lyase liquid assay

Liquid exoenzyme assays to measure the activity of secreted pectate lyase were performed as described previously (Coulthurst et al., 2005a). Cell cultures (1 ml) were sampled after 17 h of growth in PMB media at 25°C. Culture samples were centrifuged to pellet the cells and supernatant samples were carefully removed and kept on ice. 22.5 µL of supernatant sample was mixed with 876 µL Pel reaction mix [100 mM Tris–HCl (pH 8.5), 0.35 mM $CaCl_2$ and 0.24% sodium polygalacturonate] at 37°C, and the rate of increase of A_{235} measured immediately. The ΔA_{235}/min at 37°C was measured in a Unicam Heλios spectrophotometer, using the rate programme with five cycles of 5 s integration time.

2.12.5 Cellulase plate assay

Overnight cultures of the bacterial strains to be tested for cellulase activity were adjusted to an OD_{600} of 0.2, and 5 µl was onto the Cel agar plate. Plates were incubated for 48-72 hr at 30°C and covered with 0.2% w/v congo red solution for 20 min, bleached with 1 M NaCl for 20 min and then stained with 1 M HCl for 5 min. Cellulase activity was visualised as a red/purple halo surrounding the test colony on a blue background.

2.12.6 Cellulase liquid assay

Liquid exoenzyme assays to measure the activity of secreted cellulase were performed as described previously (Coulthurst et al., 2005a). Cell cultures (1 ml) were sampled after 17 h of growth in PMB media at 25°C. Culture samples were centrifuged to pellet the cells and supernatant samples were carefully removed and kept on ice.
Cellulase activity was measured using Ostazin Brilliant Red cellulose as described (Coulthurst et al., 2005a). 50 µL of each supernatant sample was mixed with 200 µL 2.5 mg/ml OBR-cellulose in phosphate buffer (40 mM K_2HPO_4, 15 mM KH_2PO_4) and incubated at 42°C for 1 hr. The reaction was stopped with 0.75 ml 2:1 ethanol–acetone. Following centrifugation (13,400 g, 5 min), the absorbance at 550 nm (A_{550}) of the

supernatant was measured. Activity was calculated as ΔA_{550}/min/ml and with medium-only as a blank.

2.12.7 Swimming and swarming motility plate assays

Swimming motility was assessed on tryptone swarm agar (TSA) plates (10 g l^{-1} Bacto tryptone, 5 g l^{-1} NaCl and 3 g l^{-1} Bacto agar). Swarming motility was measured on Eiken agar plates (10 g l^{-1} Bacto tryptone, 5 g l^{-1} NaCl, 5 g l^{-1} Bacto yeast extract and 7.5 g l^{-1} Eiken agar). Overnight bacterial cultures were adjusted to an OD_{600} of 0.2 for swimming motility plates and an OD_{600} of 1.0 was for swarming plates, and 3 µl was spotted onto the plates. Halo size was examined after growth for 16 hr at 30°C for swimming motility and 36 hr at 30°C for swarming motility.

2.12.8 Stress tolerance assays

Susceptibility of S39006 strains to salt, ethanol, and sodium dodecyl sulfate (SDS) was assayed similar to a previously described method (Sousa et al., 2010). Overnight cultures of S39006 strains were adjusted to an OD_{600} = 1.0 (corresponding to approximately 10^9 colony forming units (CFU) per ml) and serially diluted to 10^{-6}. 10 µl of each dilution was spotted onto the surface of LB agar supplemented with either 2.5% NaCl, 2.5% EtOH, 0.1% SDS or unsupplemented (as a control). The number of surviving CFU ml^{-1} was determined after incubation for 48 hours (for control and SDS plates) and 56 hours (for NaCl and EtOH plates).

Susceptibility of S39006 strains to hydrogen peroxide was determined by a disk diffusion assay similar to a previously described method (Lundberg et al., 1999). 15 µl of 3% H_2O_2 was spotted onto a 6 mm diameter paper disk and placed over a lawn of bacteria on LB agar. After incubation for 24 h at 30°C, the diameter of the zone of inhibition of bacterial growth was determined as a measure of susceptibility.

2.13 Gene expression assays

2.13.1 Promoter::*lacZ* fusion assay conditions

Promoter fragments of *prf* were cloned into the promoterless *lacZ* plasmid pRW50 and promoter activity assayed in *E. coli* DH5α. *E. coli* pRW50 contained no detectable β-galactosidase activity. Overnight cultures of strains tested were grown

with Tc selection and subcultured into 25 ml LB containing Tc. The bacterial cultures were incubated at 300 rpm at 37°C for 10 hr and cells from 1 ml of culture were pelleted by centrifugation at 13,000 rpm in a microcentrifuge and a β-galactosidase colorimetric assay performed.

2.13.2 β-galactosidase colorimetric assay

The colorimetric assay of β-galactosidase activity is based on the hydrolysis of *o*-nitrophenyl-β-D-galactopyranoside (ONPG) by β-galactosidase resulting in a colour change from colourless to yellow (Miller, 1972). This colour change is measured as the change in A_{420} with the rate of reaction a reflection of enzyme activity. β-galactosidase activity expressed as ΔA_{420} min^{-1} ml^{-1} OD_{600}^{-1}.

2.13.3 β-galactosidase fluorometric assay

lacZ reporter gene expression was also detected by the fluorescent signal resulting from cleavage by β-galactosidase of the fluorogenic substrate 4-Methylumbelliferyl beta-D-galactoside (MUG). 10 µl of cell culture was transferred to a 96-well plate and frozen at -80°C. Samples were thawed at room temperature and 100 µl of reaction mix (400 µg/ml lysozyme and 1× MUG in PBS) was added to each well and samples mixed by pipetting. The final concentration of MUG was 0.125 mg/ml. Fluorescent signal (relative light units per second, RLU/s) was detected in a SpectraMax Gemini XPS fluorescence microplate reader (Molecular Devices) using the following settings: excitation 360 nm, emission 450 nm, read intervals 30 sec/min for 20 min, temperature 37°C. β-galactosidase activity was expressed as RLU s^{-1} ml^{-1} OD_{600}^{-1}.

2.14 Potato virulence assay

Potato-rotting assays were performed as described previously (Fineran et al., 2007). An inoculum of 10 µl (OD_{600} = 0.1) was used to inoculate the potato tubers corresponding to approximately 10^6 colony forming units. For each experiment, the relative virulence of three replicates per strain was assessed using 4 potatoes per replicate for a total of 12 potatoes. Each tuber had two inoculation sites, one for each strain. Tubers were incubated at 25°C and after 96 h the amount of soft-rotted tissue was weighed.

2.15 *C. elegans* virulence assay and microscopy

Assays of S39006 killing of *C. elegans* were based on those used by Kurz *et al.* (2003). For strains expressing pGFP, the NGM agar was supplemented with ampicillin and 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) to select for plasmid maintenance and induce expression of GFP. For each test, at least 50 L4 stage N2 worms per bacterial strain were used. Plates were incubated at 25°C and scored for live worms every 24 h. Worms were considered dead when no longer responsive to touch and were transferred to new plates daily. During the first four days of infection, 3 worms per GFP-expressing strain were transferred to a glass slide spotted with 5 µl M9 buffer supplemented with 1 mM sodium azide to immobilize worms and inhibit expulsion of bacteria from the intestine. Worms were subsequently analyzed by light and fluorescence microscopy using an Olympus BX51 microscope.

For infection assays of the S39006 *rpoS* mutant the mutant worm strain DH26 (*fer-15*), which is conditionally sterile at 25°C, was used to avoid the potential confusion between generations. As a result, it was not necessary to transfer *fer-15* worms to a fresh plate each day. The time course of survival of *fer-15* worms on *Serratia marcescens* Db11 was previously shown to be essentially identical with that of wild type worms (Kurz et al., 2003). The DH26 nematode strain used in this work was provided by the Caenorhabditis Genetics Center, which is funded by the NIH National Center for Research Resources.

Survival curves were evaluated using Prism (version 5.0) and are considered statistically significant from each other when P-values are <0.05 using the log-rank test (also called the Mantel-Cox test).

2.16 Electron microscopy

Flagella of S39006 strains were visualized by transmission electron microscopy which was performed by David Goulding (WTSI). Strains were grown to early stationary phase and an aliquot of culture equivalent to an OD 0.5 in 1 ml was taken and cells pelleted gently to prevent shearing of flagella by centrifugation at 3,000 rpm in a tabletop microcentrifuge. Cells were resuspended in 1 ml PBS and delivered to WTSI for transmission electron microscopy. EM grids were prepared with Formvar/carbon

support films. 5µl of slightly turbid suspension (in double distilled water) was pipetted directly onto the film side of the grid and allowed to settle for 30 seconds. An equal volume of the heavy metal stain 1% ammonium molybdate (+ 0.5% trehalose) was added to the grid and immediately drained with cut Whatman no. 1 filter paper. After air drying the grid was observed in the FEI 120kV Spirit Biotwin TEM and images captured on a Tietz F415 CCD camera.

2.17 Bioinformatic analysis programs

Nucleotide and amino acid sequences were analyzed using a suite of bioinformatic programs. ClustalW (EBI) was used for sequence alignments. NCBI's Basic Local Alignment Search Tool (BLAST) was used to find homologous sequence in other organisms. InterProScan (EBI) scans the InterPro database of protein families to find matching functional domains to a query sequence. Secondary structure and disorder prediction was performed by PSIPRED and DISOPRED. DomPred predicts putative protein domains and their boundaries for a given protein sequence. PSIPRED, DISOPRED, and DomPred are tools provided by the Bioinformatics Group[2] at UCL (Marsden et al., 2002). FUGUE[3] and NCBI's conserved domain search[4] was used to identify homologous proteins based on secondary structure similarity. The computation of various physical and chemical parameters, such as the isoelectric point (pI) and molecular weight (MW), for a given protein was calculated using ProtParam, which is hosted on the ExPASy (Expert Protein Analysis System) Proteomics Server[5] of the Swiss Institute of Bioinformatics (SIB). Transmembrane helices were predicted by TMHMM 2.0 and signal peptide predicted by the SignalP 3.0 Server.

RNA sequences were analyzed using various tools. Rfam[6] (Sanger) searches a query sequence against a large collection of multiple sequence alignments and covariance models covering many common non-coding RNA families. RNAAnalyzer[7]

[2] Bioinformatic protein tools- http://bioinf.cs.ucl.ac.uk/web_servers/
[3] FUGUE- http://tardis.nibio.go.jp/fugue/prfsearch.html
[4] NCBI conserved domain search- http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
[5] ExPASy Proteomics Server- http://expasy.org/
[6] Rfam- http://rfam.sanger.ac.uk/
[7] RNAAnalyzer- http://rnaanalyzer.bioapps.biozentrum.uni-wuerzburg.de/server.html

searches for several known RNA structures. Riboswitch Finder[8] searches RNA/DNA for several known riboswitches. Tandem Repeats Finder[9] searches for tandem repeats. RNAfold[10] (Vienna RNA Web Server) and Mfold[11] (Rensselaer bioinformatics web server) were used to predict secondary structures of single stranded RNA sequences. RNAhybrid[12] was used to predict RNA duplex formation between small and target mRNAs. Noncoding RNAs, riboswitches, and rho-independent transcription terminators in S39006 were predicted by Paul Gardner (WTSI), curator of the Rfam database, using rfam_scan, tRNAscan-SE, rnammer, and RNIE.

Venn diagrams were made with VENNY[13].

[8] Riboswitch Finder- http://riboswitch.bioapps.biozentrum.uni-wuerzburg.de/
[9] Tandem Repeats Finder- http://tandem.bu.edu/trf/trf.html
[10] RNAfold- http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi
[11] Mfold server- http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form
[12] RNAhybrid- http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/
[13] VENNY- http://bioinfogp.cnb.csic.es/tools/venny/index.html

Chapter 3. *prf* regulates prodigiosin production

3.1 Introduction

Previous research investigating the regulatory network of prodigiosin identified an intergenic region referred to as **p**igment **r**egulatory **f**actor (*prf*) which when disrupted by a transposon insertion resulted in hyperpigmention. *prf* mutants were identified from screens of random transposon mutants of the S39006 WT strain and also *pigP* and *rsmB* mutants, bypassing the hypopigmented phenotype and restoring prodigiosin production. The aim was to investigate further the intergenic region of *prf* and characterize its function.

3.2 Results

3.2.1 Generation of *prf* mutants

Previous *prf* transposon mutants contained a chloramphenicol (Cm) antibiotic resistance marker, which limited the ability to create double mutants of *prf* with some other Pig regulator mutants marked with the same drug resistance. Therefore, new *prf* mutants were isolated from a random transposon mutagenesis screen using the transposon Tnkrcpn1 which contains the kanamycin resistance marker and a promoterless *lacZ* gene for transcriptional fusion studies (if inserted in the correct orientation).

Random transposon mutants were generated from a hypopigmented *pigQ* mutant (strain Pig17S). PigQ upregulates RsmB small RNA which sequesters RsmA allowing for the expression of prodigiosin (Figure 1.2). Since *prf* mutants were picked out from a transposon mutagenesis screen of *rsmB* bypass mutants, restoring pigment, then a *prf* mutation should also bypass a *pigQ* mutation. Around 20,000 transposon insertion transconjugants were generated and 126 mutants showing various degrees of increased pigment production were picked out for further analysis.

Random primed PCR was used to amplify and sequence across the transposon insertion site in the genome to identify the genes or region of DNA that have been disrupted by a transposon insertion. One of these nine mutants was identified with a

transposon insertion in the *prf* intergenic region (Table 3.1). By comparing the phenotype of this *prf* mutant with the other bank of hyperpigmented mutants, five more *prf* mutants were uncovered bringing the total to six *prf* mutants that were created by the random transposon bypass mutagenesis of the *pigQ* mutant (Figure 3.1). The other genes identified by the random PCR that were able to bypass the *pigQ* mutant included genes in five mutants involved in the phosphate transport system previously associated with a hyperpigmented phenotype (Slater et al., 2003), a gene in two mutants encoding the transcriptional regulator, *pigZ*, of an efflux pump (Gristwood et al., 2008), and a gene in one mutant involved in potassium efflux.

Table 3.1. Table of transposon mutants bypassing *pigQ*.

Mutant	Gene containing transposon insertion	Function
1	*pstC*	Phosphate transport system
2	*pigZ*	Transcriptional Regulator of an efflux pump
3	*pstA*	Phosphate transport system
4	*pstC*	Phosphate transport system
5	*pigZ*	Transcriptional Regulator of an efflux pump
6	*pstS*	Phosphate transport system
7	*prf*	Unknown, *argG-vfmE* intergenic region
8	*mscS*	Potassium efflux system
9	*pstS*	Phosphate transport system

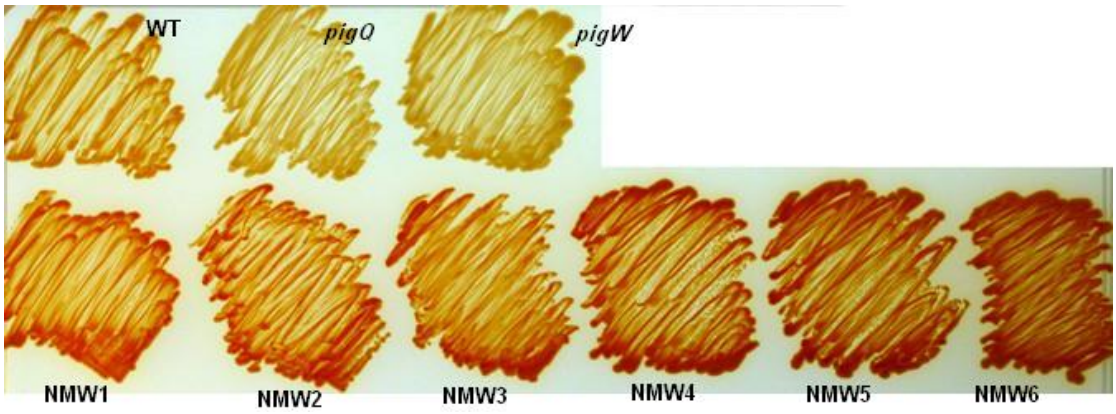


Figure 3.1. *prf* mutants isolated from a random transposon mutagenesis screen.

Hyperpigmented *prf* mutant strains, NMW1 to 6, were isolated from a random transposon mutagenesis of a *pigQ* mutant (strain Pig17S). *pigQ* and *pigW* mutants which are the histidine kinase and response regulator of a two-component system are decreased for pigment and therefore white. The WT is included for comparison.

The mutations from six *prf* transposon mutants (NMW1 to 6) were transduced into a WT (*lacA*) background to confirm that the hyperpigmented phenotype was due to the transposon insertion in *prf* and not a secondary mutation. The location of the transposon insertion in each mutant was determined by PCR amplification and sequencing across the insertion site. The transposons of the *prf* mutants (NMW1 to 6) had inserted into a putative intergenic region consisting of 477 bp between genes *vfmE* and *argG*. Initally, it was speculated that this region might contain a regulatory element such as a binding site, or might encode a small gene. Bioinformatic analysis of the region showed a potential open reading frame (ORF) predicted to encode a small protein of 86 aa. Furthermore, transposons in four of the six mutants formed active *lacZ* transcriptional fusions when inserted in the same orientation as the direction of the predicted ORF indicating that it may code for a small protein (Figure 3.2). Additionally, it was not possible to amplify by RT-PCR across the region from the end of *argG* into *prf* indicating there was no transcriptional read-through. However, transcripts for each of the individual genes could be amplified (N. Williamson, personal communication). 5'RACE (rapid amplification of cDNA ends) was performed to try and identify the transcription start site (TSS) of *prf* and the promoter region. However, the two attempts at RACE were unsuccessful in amplifying the *prf* transcript and instead ribosomal RNA was amplified.

Further assays were performed to investigate the phenotype of the *prf* mutants, in addition to hyperpigmentation. The *prf* mutants showed no changes in production of carbapenem, pectate lyase, or cellulase (data not shown). However, they exhibited production of an unknown yellow pigment and displayed increased swarming motility (Figure 3.3). Curiously, the hyperpigmented phenotype of the *prf* mutants could not be complemented by expression of the *prf* ORF or the entire intergenic region *in trans* (data not shown).

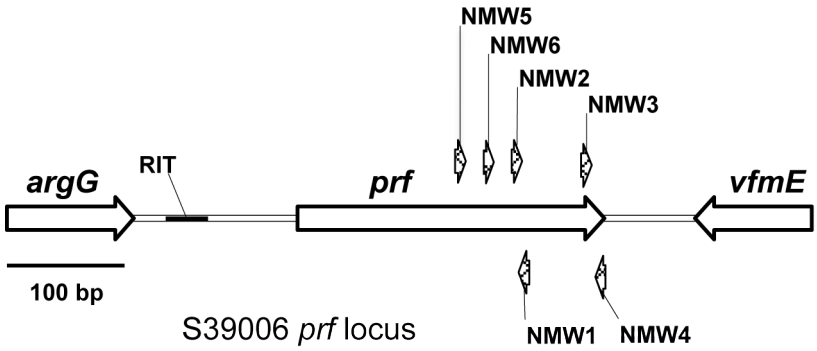


Figure 3.2. Genetic organization of the *prf* locus.

The predicted ORF, *prf*, is 261 nt potentially encoding an 86 aa protein. The region from the end of *argG* to the start of *prf* is 139 nt. The region between the ends of *prf* and *vfmE* is 77 nt. RIT is a rho-independent terminator predicted using RNIE. The hatched arrowheads above and below the *prf* ORF indicate the position and relative orientation of the transposons containing *lacZ* transcriptional fusions inserted in strains NMW1 to 6. Transposons in strains NMW2, 3, 5, and 6 formed active transcriptional fusions. The transposon insertion site is bordered on either side by a 9 nt repeat of the chromosomal sequence. The length between the two furthest transposon insertion sites, NMW4 and 5, is 120 bp.

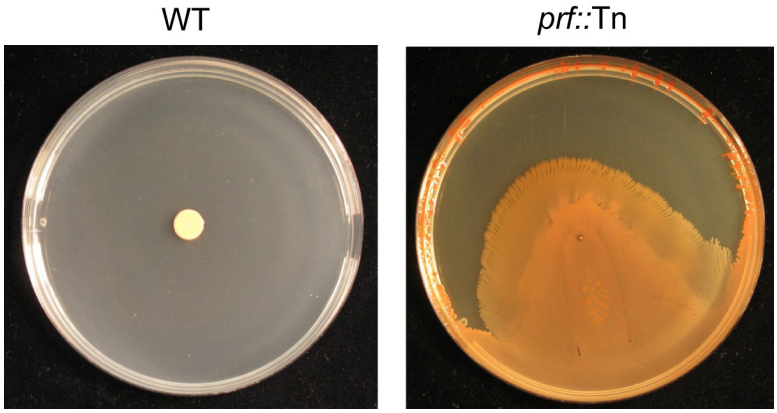


Figure 3.3. *prf* regulates swarming motility.

The *prf*::Tn mutant NMW5 displayed an increase in swarming motility compared to that of the WT on 0.75% LBA (Bacto agar).

3.2.2 Bioinformatic analysis of the *prf* region

A bioinformatic analysis of the *prf* ORF was conducted to assess if there was evidence that it was a valid protein coding gene, or if the region might code for a non-coding RNA. DNA and protein NCBI BLAST searches of the ORF did not reveal any homologues and resulted in only very weak matches. FUGUE, which searches protein sequences to identify matches based on structure, did not produce a match above the recommended cutoff. InterProScan, which searches the InterPro database of protein families for matches to a query sequence, did not yield any results. The amino acid sequence was then analysed in protein secondary structure prediction programs. The programs, PSIPRED and DISOPRED, predicted order across most of the protein with various structural motifs including coil, helix, and beta sheets (Figure 3.4). ProtParam calculated the molecular weight to be 9874.6 Daltons with a predicted isoelectric point of 10.58.

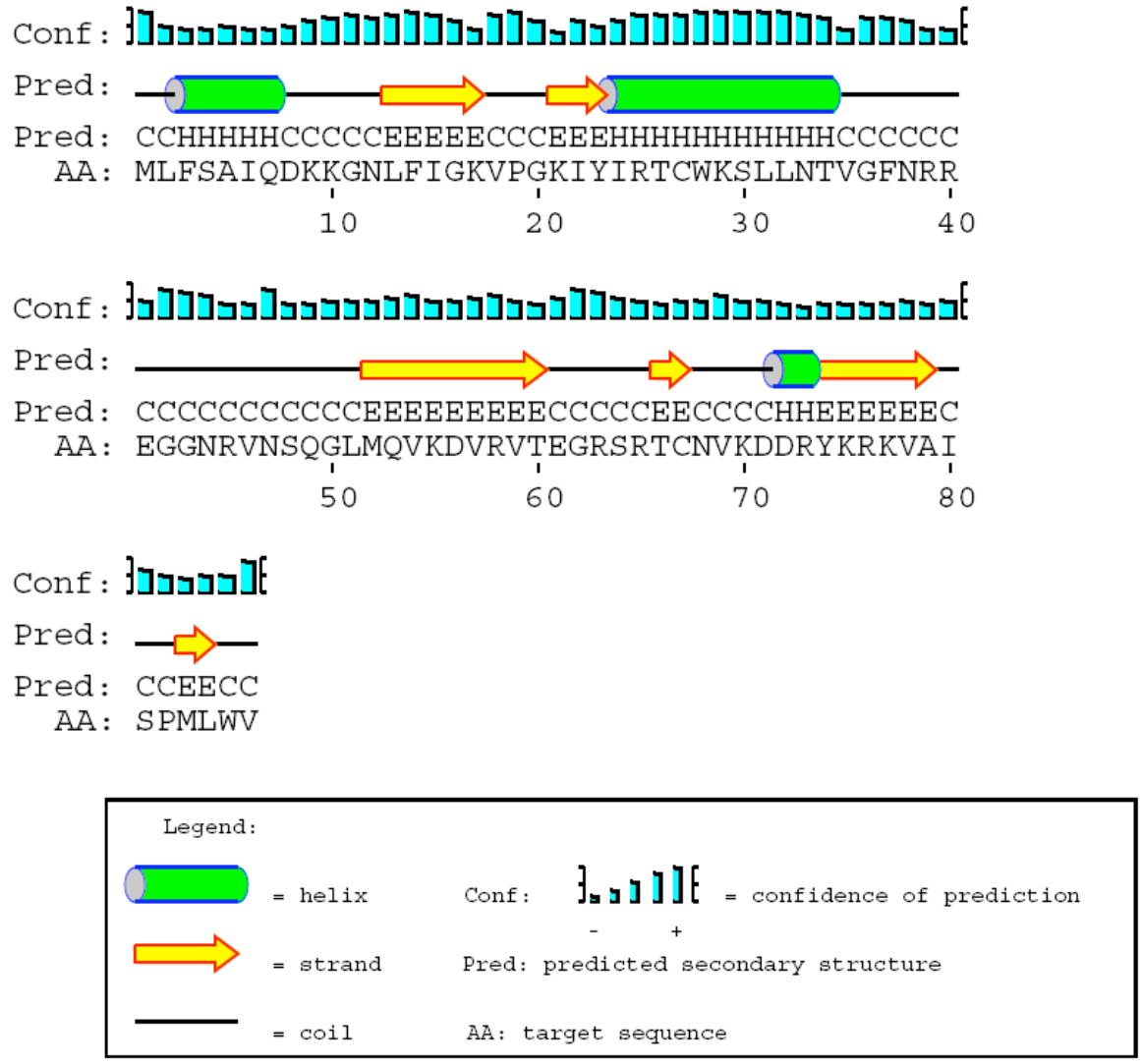


Figure 3.4. Protein secondary structure prediction of *prf*.

The secondary structure was predicted by PSIPRED and the graphical diagram created by DomPred. The blue bars above the protein depiction indicate the degree of confidence in the prediction.

Attempts to glean information using RNA bioinformatics programs did not reveal evidence that the *argG-vfmE* intergenic region codes for a non-coding RNA element. The sequence of the entire intergenic region was searched against the Rfam database of ncRNAs and put into RNAAnalyzer, which search for regulatory RNA elements, neither of which produced a result. Riboswitch Finder which searches queries for several known riboswitches likewise resulted in no hits. Tandem Repeats Finder Program also

produced no hits. Finally, the intergenic sequence was inputted into RNA folding programs to predict the RNA secondary structure. The programs, RNAfold and mfold, predicted structures but with no significantly known features. As a result, it is bioinformatically unclear whether *prf* codes for a sRNA or a small protein.

Additional programs were used to predict promoter sequences in the 139 nt region from the end of *argG* to the beginning of the putative start of *prf*. The entire intergenic region was put in two promoter prediction programs, BPROM and BDGP. BPROM predicted two promoter sequences, one candidate upstream of *prf* and a second downstream of *prf*. BDGP also predicted eight promoters, one of which matched the prediction from BPROM. This match is the most likely candidate promoter sequence which occurs 55 bp upstream of the putative *prf* start site (Figure 3.5).

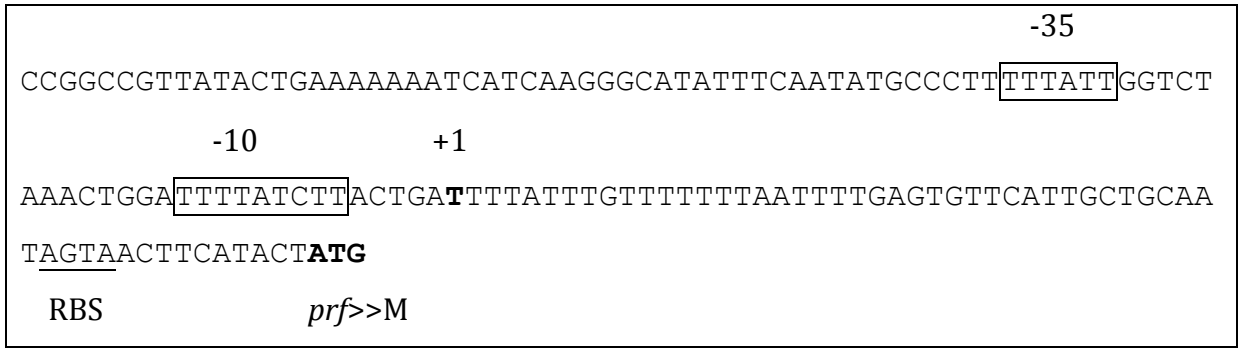


Figure 3.5. Candidate promoter sequence for *prf* ORF.

The candidate promoter sequence is the consensus between two promoter prediction programs, BPROM and BDGP. The predicted +1 transcription start site is highlighted in bold with the -35 and -10 elements highlighted in boxes. The likely RBS sequence is underlined with the *prf* start codon in bold.

3.2.3 Promoter analysis of *prf* ORF

lacZ transcriptional fusions in the *prf* putative ORF were functional when the transposon was inserted only in the same orientation as the ORF. Two attempts to identify the transcription start site of *prf* by 5'RACE were unsuccessful and resulted in amplification of ribosomal RNA. Therefore to determine whether the predicted candidate promoter (Figure 3.5) was driving transcription of *prf*, a 5' deletion analysis of the promoter region was conducted using an *E. coli* plasmid-based system. 5' truncations of the region spanning 239 nt from position -139, after the stop codon of

argG, to position +100 relative to the putative start of translation of *prf* were amplified and cloned upstream of a promoterless *lacZ* in the vector pRW50 forming constructs pNB8 to 12 (Figure 3.6A). *lacZ* expression measured as β-galactosidase activity was assessed from *E. coli* strains carrying the constructs. Surprisingly, β-galactosidase activity was maximal for construct pNB9 containing the promoter region spanning positions -46 to +100, while activity was decreased for constructs containing the predicted -10 or -35 elements or the full length region (Figure 3.6B). No activity was recorded for the construct, pNB8, containing only the putative *prf* coding sequence and the vector-only control. This could suggest that the *prf* promoter sites are in the 46 nt region upstream of the *prf* start codon. Sequence further upstream of this region in pNB10 to 12 could contain a regulatory region that suppresses the promoter activity of *prf*.

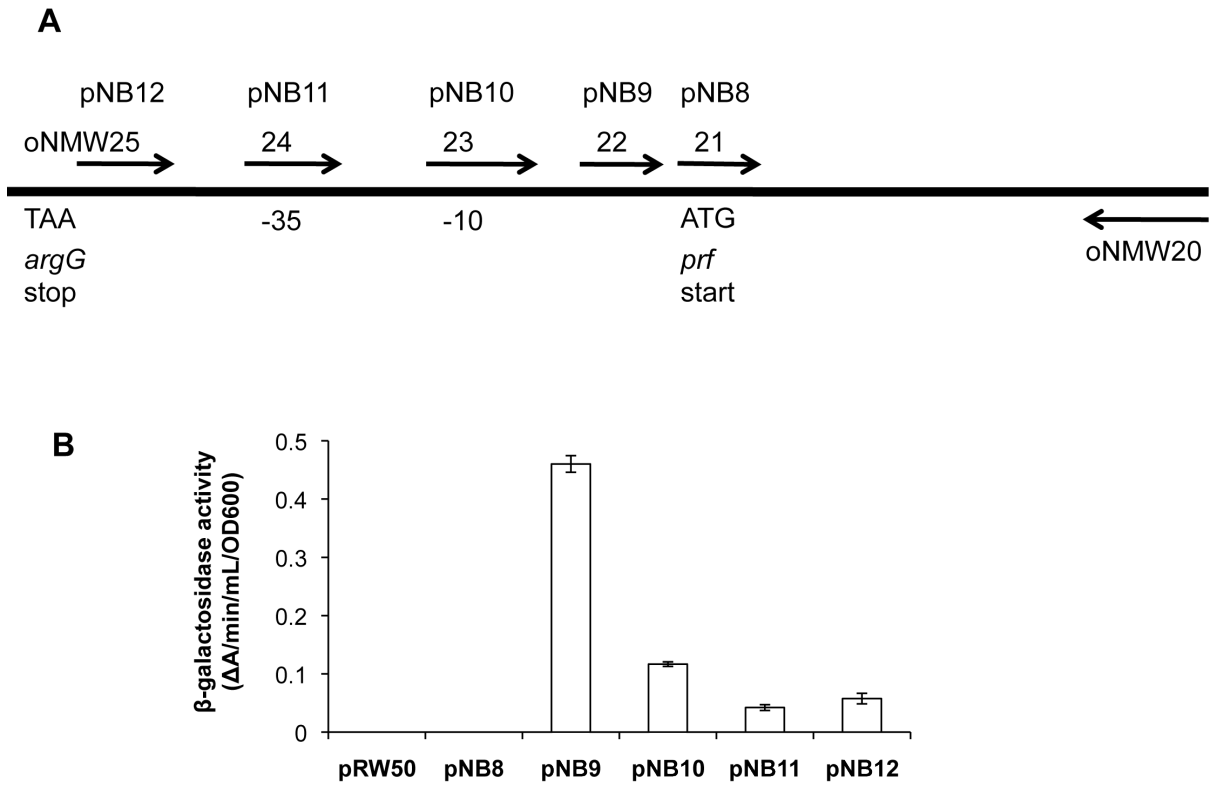


Figure 3.6. Promoter deletion analysis of *prf*.

A. Progressive 5' deletions of the *prf* promoter region were cloned into the vector pRW50 upstream of a promoterless *lacZ*. The region spans 239 nt from position -139, after the stop codon of *argG*, to position +100 relative to the putative start of translation of *prf*. Arrows

indicate the primers, oNMW20 to 25, used to amplify the progressive 5' deleted regions which correspond to the constructs, pNB8 to 12, listed above the relative left-hand primer. The predicted -10 and -35 elements are depicted.
B. *lacZ* expression from the promoter regions were measured as β-galactosidase activity in the heterologous host, *E. coli*, after 10 hours of growth.

3.2.4 The genomic context of *prf*

The genomic context of *prf* was investigated in order to try and glean information about its possible function from surrounding genes. During the course of this study the S39006 genome was sequenced and the context of surrounding genes determined. The *prf* locus was compared with the corresponding loci of the closely related species *Pectobacterium atroseptica* SCRI 1043 (*Pat*) and *Dickeya dadantii* 3937 (*Dda*; formerly *Erwinia chrysanthemi*) (Figure 3.7). The region upstream of *prf* was found to be similar to that in *Pat*. S39006 *argG*, which codes for argininosuccinate synthase, is 91% identical to its homologue in *Pat* and *ORF1* is 87% identical to *ECA0102*, which is similar to *araH*, a membrane protein of the L-arabinose operon in *S. typhimurium*. Downstream of *argG* in *Pat* are the quorum sensing genes, *expIR* which are homologues of *luxIR*.

Conversely, the region downstream of *prf* is more closely related to genes in *Dda*. S39006 VfmE is 85% identical to its homologue in *Dda*, which was the highest ranking result in an NCBI tblastx search. No match was revealed for VfmE in *Pat*. *vfmE* codes for a transcriptional regulator of the AraC family that regulates expression of plant cell wall degrading enzymes in *Dda* (Venkatesh et al., 2006). S39006 *untA*, which is an ABC transporter, is 87% identical to its homologue in *Dda*. Thus, the S39006 *prf* locus appears partly homologous to loci from each of two closely related species.

Additionally, it is surprising that the *prf* locus in S39006 is more closely related to loci in *Pectobacterium* and *Dickeya* spp. rather than to similar species within the same genus of *Serratia*. However, analysis of the genome sequence of S39006 indicates that it shares greater homology with species from these two genera than *Serratia* (J. Ramsay, personal communication). Searches using tblastx performed on the genomes of *Serratia proteamaculans* (*Spr*) and *Serratia marcescens* (*Sma*) revealed very weak matches to S39006 *argG*, which is 26% and 34% identical to *argG* in *Sma* (SMA4003) and *Spr* (Spro_4779), respectively. However, *argG* in *Sma* and *Spr* are 93% identical and appear

to be part of an operon of four genes (*argCBGH*). Searches for *vfmE* in *Sma* and *Spr* using tblastx revealed no matches in *Spr* and a weak match of 32% in *Sma* to SMA1625 which is also a predicted transcriptional regulator of the AraC family.

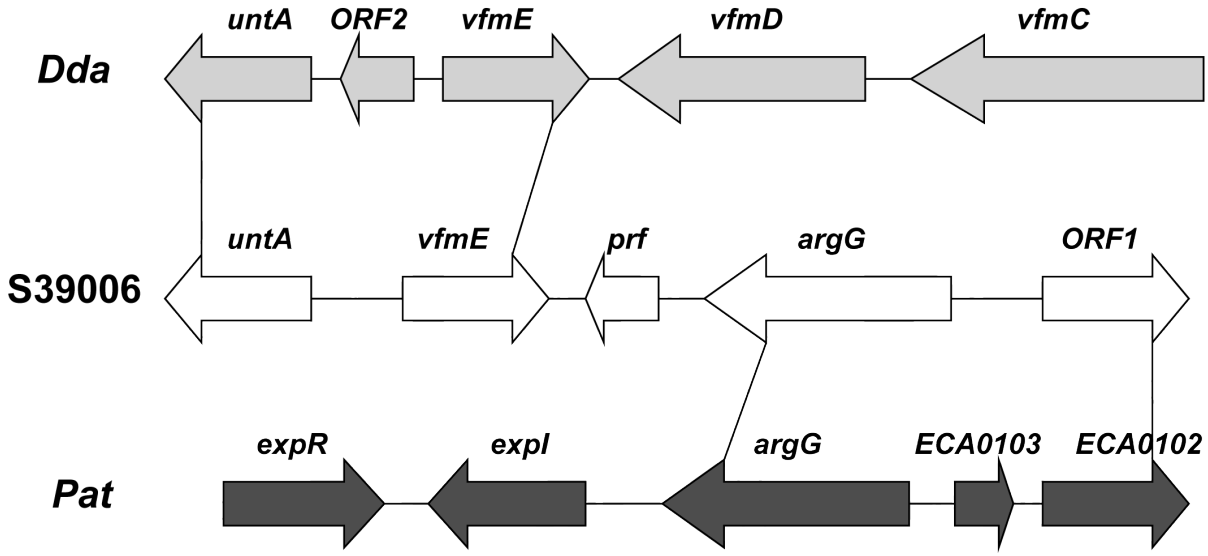


Figure 3.7. The genomic context of *prf*.

The *prf* locus is compared to the equivalent loci of two related species, *Pectobacterium atrosepticum* SCRI 1043 [Pat] and *Dickeya dadantii* 3937 [Dda]. Lines connecting genes of different loci indicate homology.

3.2.5 *prf* regulates the expression of *vfmE*

It was hypothesized that *prf* may be a regulatory factor of the neighbouring genes *vfmE* and/or *argG*. Therefore, the relative transcript levels of *vfmE* and *argG* in a *prf* mutant compared to WT were assessed by qRT-PCR. *vfmE* expression was shown to be increased 5-fold whereas no significant difference was observed for *argG* or *gyrB*, a housekeeping gene measured as a control (Figure 3.8). Additionally, the hyperswarming phenotype was reflected in a 3.5-fold increase in expression of *rhlA*. In a previous study from this lab it was shown that *rhlA* is required for surfactant production, which is essential for swarming motility (Williamson et al., 2008). *rhlA* encodes a protein homologous to *Pseudomonas aeruginosa* RhlA which is involved in the synthesis of surface-active 3-(3-hydroxyalkanoyloxy) alkanoic acids (HAAs), the precursors of rhamnolipid surfactants.

Furthermore, the interaction of *prf* with other factors in the wider regulatory network for secondary metabolism was explored. Double mutants of *prf*::Tn (strain NW108) and various Pig regulator transposon mutants were created by Neil Williamson. Pig production was measured from the double mutants and compared with that of the isogenic single mutants to determine if the *prf* phenotype was epistatic. *prf* was found to bypass mutations in *pigP*, *pigQ*, *pigU*, and *rsmB* and restore pigment production to various degrees (Figure 3.9). However, it was not able to bypass mutations in *rap* and *hfq* (data not shown).

Additionally, previous work investigating a *prf* mutant showed that increased production of prodigiosin resulted from a 2-fold increase in *pigA* transcription as determined using a *lacZ* transcriptional fusion. However, carbapenem production was unaffected (N. Williamson, personal communication). Furthermore, transcription of *pigP*, *pigQ*, *pigW*, *rap*, and *smaI* was shown not to be affected in a *prf* mutant. It was also shown in this study that a *pigQ* mutant does not affect *prf* transcription (data not shown).

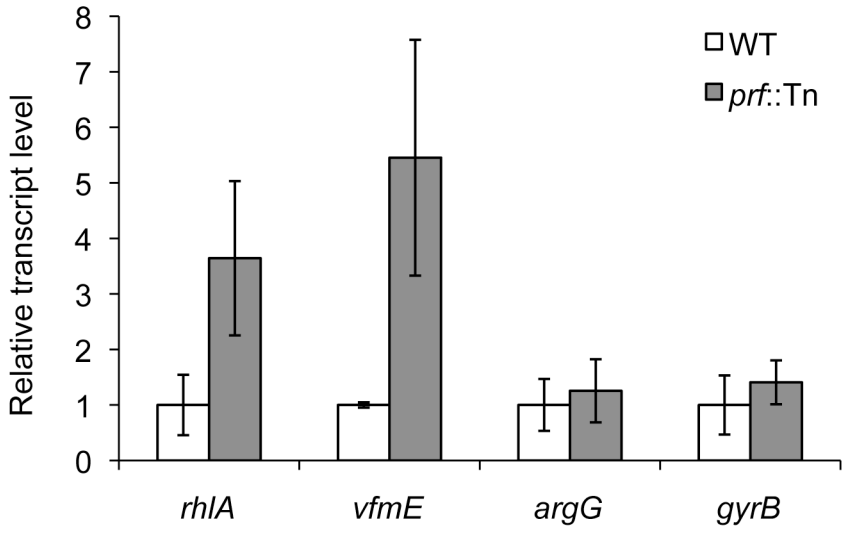


Figure 3.8. The impact of a *prf*::Tn mutation on transcript levels measured by qRT-PCR.

Gene expression for *prf*::Tn (NMW6) is measured as transcript levels relative to WT at early stationary phase growth. Values represent average gene expression ± SD from three independent experiments.

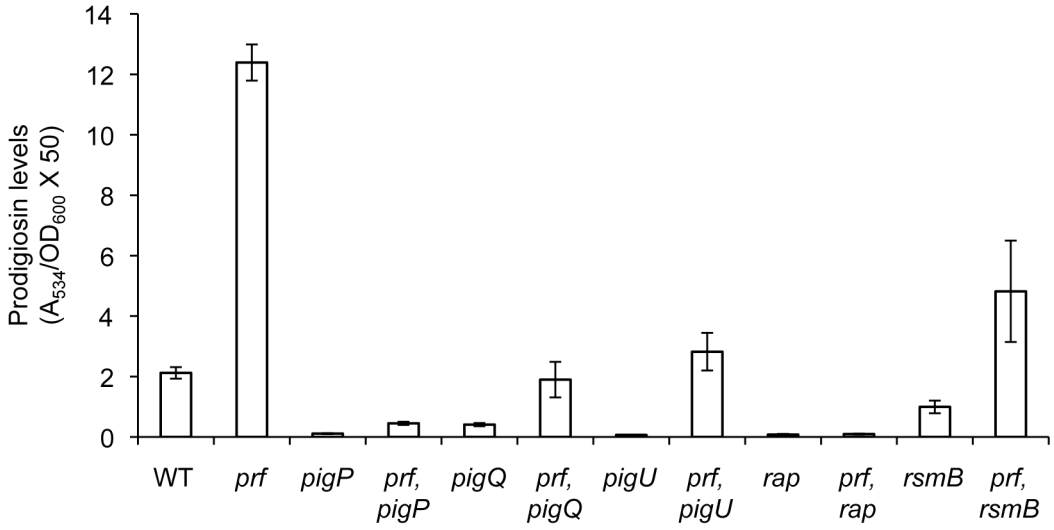


Figure 3.9. The *prf* phenotype is epistatic to *pigP*, *pigQ*, *pigU*, and *rsmB* mutations.

Prodigiosin production was measured in stationary phase from double mutants of *prf* carrying various Pig regulator mutants. Statistically significant increases in Pig production were observed for double mutants of *prf* and *pigP*, *pigQ*, *pigU*, and *rsmB*, compared to the isogenic single mutant. The strains assayed were NW108 (*prf*), HSPIG67 (*pigP*), NW109 (*prf*, *pigP*), HSPIG17 (*pigQ*), NW110 (*prf*, *pigQ*), HSPIG43 (*pigU*), NW111 (*prf*, *pigU*), RapL (*rap*), NW112 (*prf*, *rap*), HSPIG71 (*rsmB*), and NW113 (*prf*, *rsmB*). Data shown represent the mean ± SD of triplicate samples.

3.2.6 A *vfmE* mutant is decreased for prodigiosin production and epistatic over *prf*

Prior to the start of this study, the speculation that *prf* may regulate *vfmE* led Neil Williamson to create a *vfmE*::*cat* mutant (strain VFME), however it remained uncharacterized. Prodigiosin production was measured from the *vfmE*::*cat* mutant and shown to be decreased by half compared to the WT. Pig production was restored in the *vfmE* mutant by expression of *vfmE in trans* (Figure 3.10A). Moreover, expression of *vfmE in trans* in the WT strain doubled Pig production indicating that the increase in *vfmE* expression in the *prf* mutant contributes to its hyperpigmented phenotype. The *vfmE* mutation was also shown to have no effect on pectate lyase activity but was slightly decreased for cellulase activity (Figure 3.10BC).

To explore this result, a *prf*, *vfmE* double mutant was generated by transduction of the *prf::*Tn (Km^R) mutation (NMW6) into the *vfmE*::*cat* (Cm^R) strain. Because the two markers are located very close together in the chromosome (Figure 3.7), transductants were selected using both the antibiotics Km and Cm which resulted in colonies with red

or pink phenotypes identical to the red and pink phenotypes of the *prf* and *vfmE* single mutants (Figure 3.11). The same result was observed when the transduction was repeated in the opposite direction with the *vfmE::cat* mutation transduced into the *prf* mutant, except that the ratio of pink to red colonies was reversed. The ratio of pink to red colonies was greater than one when *prf*::Tn was the donor marker, whereas the ratio was less than one when *vfmE::cat* was the donor. Restreaking the red and pink colonies onto LBA plates with both Km^R and Cm^R or only Km^R selection maintained the phenotypes of the two variants. However, when the colonies were streaked onto LBA plates with only Cm^R selection or no antibiotic (Ab) selection the red variants gave rise to both pink and red colonies while the pink variants stayed pink. The pink and red variants still maintained Km^R and Cm^R resistance after growth on plates without Ab selection. Amplification by colony PCR across the *prf*::Tn insertion site in the double mutant was successful for the pink variant but not the red. Cumulatively, the evidence suggests that a not infrequent anomaly in the recombination of two close genetic markers is occurring in the selection of *prf*, *vfmE* double mutants. However, the pink variant is the true phenotype indicating that the hyperpigmented red phenotype of the *prf* mutant is solely due to the increase in *vfmE* expression which is reduced to the pink phenotype of the *vfmE* mutant when the double mutant is generated.

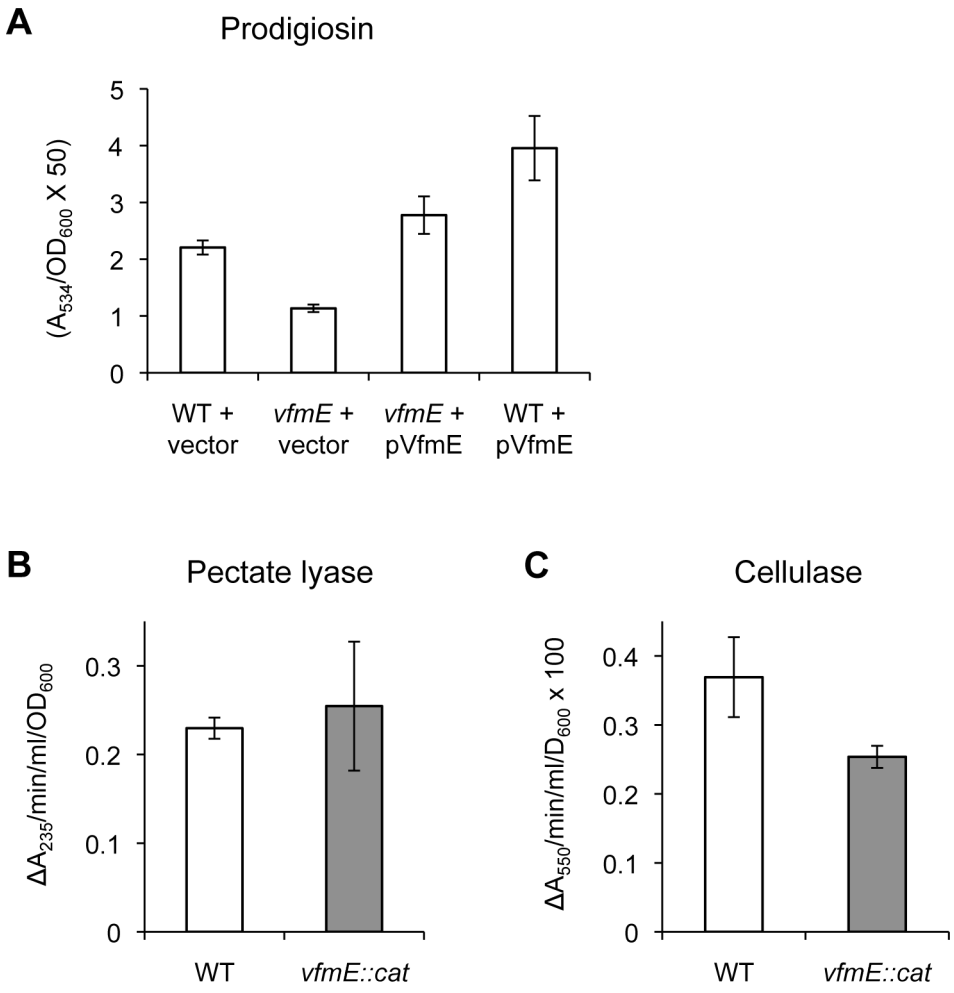


Figure 3.10. *vfmE* regulates prodigiosin and cellulase production.

A. Prodigiosin production was measured for the WT and *vfmE::cat* strains. The vector used was pQE80oriT and pVfmE refers to pNB13. B-C. The *vfmE* mutation has no effect on pectate lyase activity but is slightly decreased for cellulase.

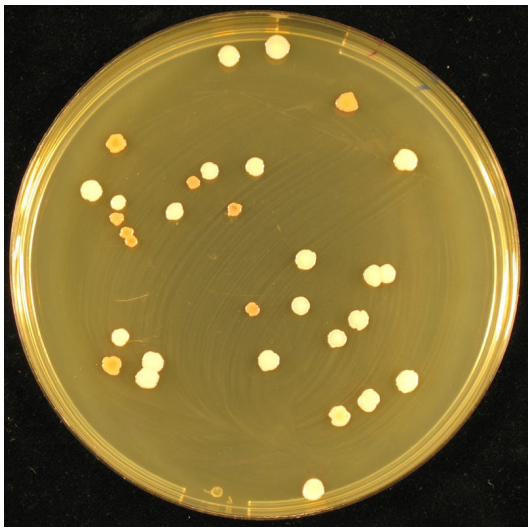


Figure 3.11. Selection of *prf*, *vfmE* transductants.

The *prf*::Tn (Km^R) mutation from strain NMW6 was transduced using the phage ϕOT8 into the recipient *vfmE::cat* (Cm^R) and transductants were selected on LBA using Km and Cm. The resulting colonies were a mixture of pink and red phenotypes.

3.2.7 *prf* is potentially a *cis*-antisense RNA

During the course of this study the transcriptome of S39006 was sequenced by RNA-seq (see Chapter 6), and the transcriptional profile of the *vfme-prf-argG* region was examined (Figure 3.12). Reads begin to map on the coding strand of the *prf* ORF 108 bp after the putative start of translation and span a region that includes the intergenic *prf-vfmE* region, the strand antisense to the *vfmE* ORF (558 bp), and cease mapping in a region upstream of *vfmE* (814 bp after the putative *prf* stop codon). Maximum coverage of reads was observed for the region just after the putative *prf* stop codon for a length of 300 bp spanning the intergenic region between *prf* and *vfmE* and the strand antisense to *vfmE*. Therefore, the RNA-seq data indicates that potentially a single 922 bp *cis*-antisense RNA (asRNA) primarily overlapping the *vfmE* ORF is transcribed beginning midway in the *prf* ORF. Interestingly, the region of *prf* in which transposon insertions have been identified from the *prf* mutants NMW1 to 6 are within the 5' region of the potential asRNA (Figure 3.12). Taken together, these results suggest that the hyperpigmented phenotype of a *prf* transposon mutant results from insertion of the transposon in the 5' region of the asRNA disrupting transcription of the asRNA and resulting in increased *vfmE* expression. The selection of *prf* transposon mutants in the

intergenic region between the *prf* and *vfmE* ORFs by former members of the lab lends weight to this hypothesis.

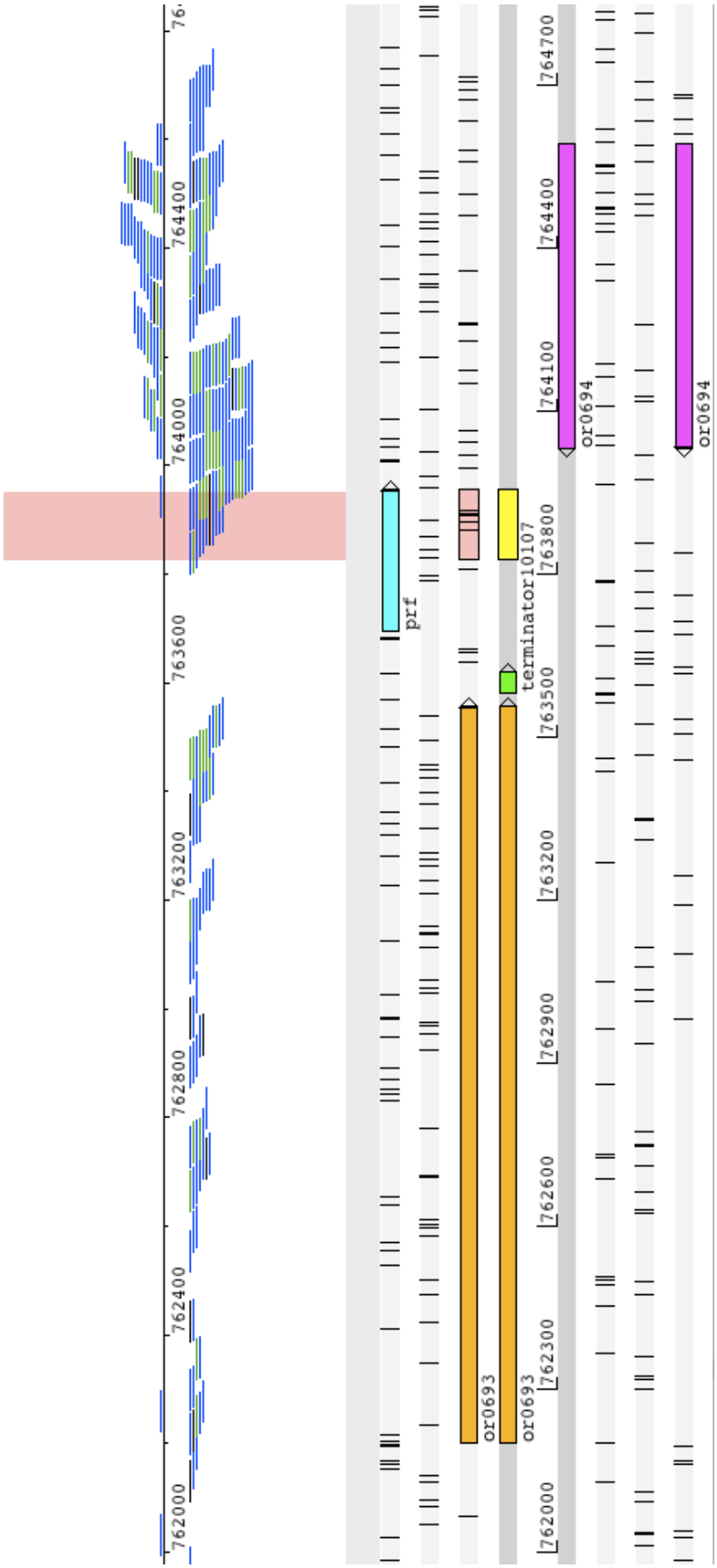


Figure 3.12. Transcriptional profile of the *prf* locus.

RNA-seq data from the S39006 transcriptome of the *argG-prf-vfmE* locus is displayed using BamView (top panel) in Artemis (bottom panel). The total coverage is shown as non-proper paired reads aligned against the two strands of the reference genome sequence with the annotation displayed underneath. Reads mapped to the top strand in BamView correspond to the bottom strand of the annotation in Artemis and vice versa. Non-proper paired reads map in a strand specific manner (see Chapter 6.2.3). The ORFs or0693 and or0694 refer to *argG* and *vfmE*, respectively. The region highlighted in yellow/pink spans the distance in which transposon insertions have been isolated from *prf* mutants NMW1 to 6.

3.3 Discussion

The initially mysterious function of the *argG-vfmE* region was revealed through the course of this study. Six *prf* transposon mutants (NMW1 to 6) were generated from a screen of mutants bypassing the white hypopigmented phenotype of a *pigQ* mutant (PIG17S). The *prf* mutants had a hyperpigmented and hyperswarming phenotype with no affect on carbapenem or exoenzyme production. The *prf* mutation was shown to be epistatic at various degrees to mutations in the Pig regulators *pigP*, *pigQ*, *pigU*, and *rsmB* but not *rap* and *hfq*. However, it was not possible to complement the *prf* mutation with plasmid vectors containing the entire *prf* intergenic region or just the predicted *prf* ORF. Although attempts to identify the transcription start site (TSS) by 5'RACE were unsuccessful, a 5' promoter deletion analysis identified a region spanning position -46 to +1 relative to the *prf* ORF as having promoter activity.

A *prf* mutation was shown to result in increased expression of the neighbouring *vfmE* gene. A *vfmE* mutant showed reduced prodigiosin production (by 50%) and this phenotype was complemented. Expression of *vfmE in trans* increased Pig production. Generation of a *prf*, *vfmE* mutant resulted in a phenotype identical to the *vfmE* mutant, indicating that the hyperpigmented phenotype of *prf* is solely due to increased *vfmE* expression.

RNA-seq data of the S39006 transcriptome suggested an interesting and unexpected answer to the question of how *prf* regulates *vfmE*. The mapping of reads indicated that a candidate *cis*-antisense RNA overlapping the *vfmE* ORF initiates transcription midway in the *prf* putative "ORF" and extends for 922 nt. It is presumed that the asRNA would regulate *vfmE* expression by forming complementary base pairs to the *vfmE* transcript preventing translation by the ribosome and inducing cleavage of the transcript by the endoribonuclease RNase III, which specifically digests double-stranded RNA. The selection of multiple *prf* mutants with transposon insertions in the 3' half of the putative ORF, which is the 5' region of the candidate asRNA, lends credence to the notion that the insertion disrupts the function of the asRNA and not the "ORF". Additionally, this explains why *prf* mutants with transposon insertions in the 5' half of the ORF have not been selected while mutants have been selected with insertions at the very 3' end of the ORF (such as NMW4) and in the intergenic region between the end of

the ORF and *vfmE*. If the ORF coded for a functional protein then there should be an equal, or similar, chance that a transposon disrupts the 5' half of the ORF in comparison to those identified in the 3' half, particularly since an insertion in the 5' half would certainly disrupt the translation of the ORF. However, the selection of multiple insertions in the 3' versus 5' region is statistically significant and suggests that the *prf* ORF is not a functional protein-coding gene. If the ORF were divided into two equal halves, the probability that all six insertions were identified in the 3' half is 1.56% according to the binomial distribution. However, there is a possibility that the translation start codon of *prf* is further downstream and that a shorter ORF produces a small protein that impacts upon *vfmE* expression. But the expression of an asRNA in the *prf* region also explains how the selection of transposon insertions at the very 3' end of the ORF and in the intergenic region before the end of *vfmE* would still function to disrupt the candidate asRNA and induce the hyperpigmented phenotype resulting from increased *vfmE* expression.

Further work is needed to verify the transcription of the candidate asRNA. Northern blotting will allow for the detection of both sense and antisense transcripts of *vfmE* in the WT versus *prf* mutants. Another attempt at 5'RACE, but redesigning the experiment based on recent knowledge from the transcriptional profile, should allow for the identification of the TSS of the antisense transcript.

Nevertheless, the presence of the candidate asRNA would also explain the mechanism by which *prf* regulates *vfmE* expression and, in consequence, pigment production. The asRNA would form complementary base pairs to the *vfmE* transcript resulting in degradation by the RNA double-strand specific endoribonuclease RNase III. An insertion in the *prf* region disrupts transcription of the asRNA but not the *vfmE* transcript resulting in increased *vfmE* expression which leads to the hyperpigmented phenotype. This is reflected in the increased Pig production of the WT strain expressing extra copies of *vfmE* on a plasmid vector. Disruption of the *vfmE* gene in a *prf* mutant dramatically reduces pigment production and results in a pink hypopigmented phenotype identical with the phenotype of a *vfmE* mutant proving that disruption of the asRNA is bypassed by a deletion mutation of *vfmE*. This leads to the question of how *vfmE*, an AraC-type transcriptional regulator, regulates prodigiosin production.

The role of *vfmE* has recently been explored in the related bacteria *Dickeya dadantii* 3937 and is mentioned in only three reports. The first report mentioning *vfmE* identified it as being increased 1.8-fold in a microarray experiment analyzing differential gene expression in a *Dda phoQ* mutant (Venkatesh et al., 2006). PhoQ is the sensor kinase of the PhoPQ two-component system that governs the expression of virulence factors in *Dda*. The report also stated that a *Dda vfmE* mutant was reduced 10-fold in production of pectinases, cellulase, and protease, although no data were provided. The second report investigated the role of PecS in plant pathogenicity. Using a microarray to define its regulons, *vfmE* was identified as being upregulated 2-fold in a *pecS* mutant (Hommais et al., 2008). PecS is a transcriptional regulator of the MarR family and controls production of various virulence factors and the *out* genes of the Type II secretion system. *S. ionantha* plant leaves infected with a *vfmE* mutant exhibited delayed symptoms, with more plants remaining symptomless compared to the results with the parent WT strain. The third report mentioning *vfmE* explored the regulation of the *cyt* genes and identified *vfmE* as a repressor (Costechareyre et al., 2010). The genome of *Dda* contains four *cyt* genes in an operon displaying homology to the *Bacillus thuringiensis* cytolytic (Cyt) toxins, which are involved in causing pathogenicity to insects. It was shown, using a *cytA*::*lacZ* transcriptional fusion, that expression of *cytA* is increased 13-fold in a *vfmE* mutant. In an infection assay, the *vfmE* mutant was lethal to the pea aphid insect, *Acyrthosiphon pisum*, similar to the WT strain.

Although the *Dda* VfmE protein is 85% identical to S39006 VfmE, a BLASTP search for a PecS homologue identified two proteins, or2179 and or4126, with 42% and 21% identity. Additionally, no homologues were identified for any of the four *Dda cyt* genes. Moreover, unlike the *Dda vfmE* mutant the S39006 mutant did not exhibit a dramatic effect on exoenzyme production, with no effect on pectate lyase activity and only a modest decrease in cellulase activity. Therefore it appears that *vfme* in S39006 plays a unique role, compared to its homologue in *Dda*, in regulating prodigiosin production. It will be interesting to explore in future studies the factors regulated by *vfmE* involved in regulating prodigiosin production.

Chapter 4. The RNA chaperone, Hfq, plays a key role in pathogenesis and secondary metabolite production

4.1 Introduction

Previous studies investigating the regulation of prodigiosin production revealed a complex regulatory network involving at least 20 genes, some of which also regulate carbapenem production (Fineran et al., 2005b; Williamson et al., 2005). Multiple environmental inputs, including quorum sensing (QS) (Slater et al., 2003) and phosphate and gluconate levels (Fineran et al., 2005a), are integrated through different genetic circuits to regulate the production of prodigiosin.

An additional regulatory factor that was more recently identified as a repressor of antibiotic production is the RNA-binding protein RsmA, the homologue of *E. coli* CsrA (Williamson et al., 2008). In *Pseudomonas aeruginosa* and *P. fluorescens*, RsmA plays an important role in regulation of virulence and biocontrol factor production (Lapouge et al., 2008; Lucchetti-Miganeh et al., 2008). RsmA represses expression of specific genes by binding to the 5' UTR of target mRNAs and preventing translation. This repression is relieved by the small, non-coding regulatory RNA (sRNA) RsmB, which binds to RsmA, titrating it away from target mRNAs (Majdalani et al., 2005). The involvement of this system of post-transcriptional regulation by small RNAs in the control of prodigiosin production led to the hypothesis that members from the larger family of Hfq-dependent *trans*-acting sRNAs may likewise play a role in the regulation of prodigiosin.

Hfq, a hexameric protein that forms a doughnut-like structure, is an RNA chaperone that facilitates the complementary base-pairing between sRNAs and target mRNAs. Increasing evidence shows that Hfq and dependent sRNAs play a fundamental role in regulation of stress response and pathogenesis (Chao and Vogel, 2010). To address the potential role of Hfq in secondary metabolism and pathogenesis of S39006, *hfq* deletion mutants were constructed and characterized.

4.2 Results

4.2.1 Construction of the *hfq* mutant

The genomic context of *hfq* is highly conserved across Gram-negative bacterial species. Degenerate primers were used to amplify the S39006 *hfq* gene by PCR, and the product was then sequenced. S39006 Hfq (105 aa) is 81% identical to the *E. coli* Hfq protein with complete identity in the N-terminal region from amino acids 1 to 73, which forms the core of the hexameric protein and its RNA-bindings sites (Figure 4.1A) (Link et al., 2009). The small divergence between the two proteins is exclusive to the C-terminal region that forms the tail of the protein. Sequence analysis showed that the *hfq* gene was in a cluster of ORFs orthologous to the *amiB-mutL-miaA-hfq-hflX-hflK-hflC* cluster of genes as found in *E. coli* (Figure 4.1B) (Tsui et al., 1994). In *E. coli*, transcription of *hfq* initiates from multiple promoter regions and *hfq* is co-transcribed with the downstream gene *hflX* (Tsui et al., 1996). Previous *hfq* deletion studies in other bacteria have shown that the phenotype of an *hfq* mutant is not due to polar affects on *hflX* transcription or other downstream genes. However, for further analysis two S39006 *hfq* deletion strains were constructed; Δ*hfq* (NMW8), which contains the entire coding region of *hfq* deleted, and *hfq::cat* (NMW9), in which a chloramphenicol resistance marker replaced the *hfq* coding region.

A

```
E_coli          MAKGQSLQDPFLNALRRERVPVSIYLVNGIKLQGQIESFDQFVILLKNTVSQMVYKHAIS 60
Pat             MAKGQSLQDPFLNALRRERVPVSIYLVNGIKLQGQIESFDQFVILLKNTVSQMVYKHAIS 60
Dda             MAKGQSLQDPFLNALRRKRVPVSIYLVNGIKLQGQIESFDQFVILLKNTVSQMVYKHAIS 60
Cro             MAKGQSLQDPFLNALRRERVPVSIYLVNGIKLQGQIESFDQFVILLKNTVSQMVYKHAIS 60
Spro            MAKGQSLQDPFLNALRRERVPVSIYLVNGIKLQGQIESFDQFVILLKNTVSQMVYKHAIS 60
Sma             MAKGQSLQDPFLNALRRERVPVSIYLVNGIKLQGQIESFDQFVILLKNTVSQMVYKHAIS 60
S39006          MAKGQSLQDPFLNALRRERVPVSIYLVNGIKLQGQIESFDQFVILLKNTVSQMVYKHAIS 60
                ****************:****************************************

E_coli          TVVPSRPVSHHSNNAGGG---TSSNYHHGSSAQNTSAQQ-DSEETE 102
Pat             TVVPSRPVSHHSNNPG-----GSNNYH-GSNTTAQQQSQ-DADDAE 99
Dda             TVVPSRPVSHHSNNP------GSNNYH-ANNQSAQQQPQQESDDAE 99
Cro             TVVPSRPVSHHSNNTGGG---TSSNYHHSNNAQGSSAPQ-DSEETE 102
Spro            TVVPSRPVSHHSNTPSG----GTSNYHHGNNPSAPQQPQQESDDAE 102
Sma             TVVPSRPVSHHSNNPSG----GSSNYHHGNNPSAQQQPQQESDDAE 102
S39006          TVVPSRPVSHHSNTPGTSSNPGTNNYH-GSNPAGQQQPQQDSDDAE 105
                *************..       :.*** ...    .  * :::::*
```

B

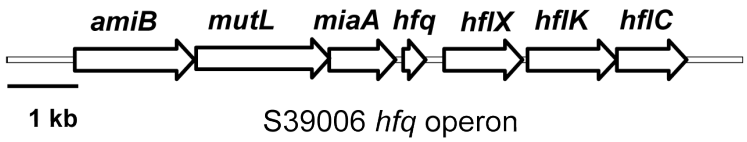


Figure 4.1. S39006 Hfq protein and *hfq* locus.

A. The ClustalW program was used to align Hfq proteins from *E. coli* K12, *Pectobacterium atrosepticum* SCRI 1043 [Pat], *Dickeya dadantii* 3937 [Dda], *Citrobacter rodentium* ICC 168 [Cro], *Serratia proteamaculans* 568 [Spr], *Serratia marcescens* strain Db11 [Sma], and *Serratia* 39006. An asterisk below the sequence alignment indicates a residue identical in all strains, a colon indicates a conserved substitution, and a period indicates a semiconserved substitution.
B. Genetic organization of the *hfq* locus in S39006.

4.2.2 Carbapenem and prodigiosin production is abolished in the *hfq* mutant

Colonies of S39006 wild type (WT) grown on LB agar are pink due to prodigiosin. During the construction of S39006 Δ*hfq* and *hfq::cat* strains, by marker exchange mutagenesis, the selection for the mutant strains resulted in colonies with a strikingly clear phenotypic difference compared to WT (Figure 4.2). In both cases, the Δ*hfq* and *hfq::cat* strains were completely white, indicating that production of prodigiosin was abolished. This was confirmed by assaying for prodigiosin levels throughout growth in LB media (Figure 4.3A). Additionally, cell filtrates produced in the same experiment were assayed for the production of Car and BHL/HHL. Although Car production was likewise abolished throughout growth (Figure 4.3B), synthesis of

BHL/HHL, the QS signaling molecule(s), was unaffected by the *hfq* deletion (data not shown). An *hfq* deletion caused a slightly extended lag phase after inoculation into fresh medium and the mutant reached stationary phase at a lower optical density compared with that of WT. Viable counts throughout growth showed no difference in cell viability for the *hfq* mutants and the WT. Finally, there was no difference in the colony morphology of the *hfq* mutants when grown on LB agar and, in all further assays, the Δ*hfq* and *hfq::cat* strains were indistinguishable.

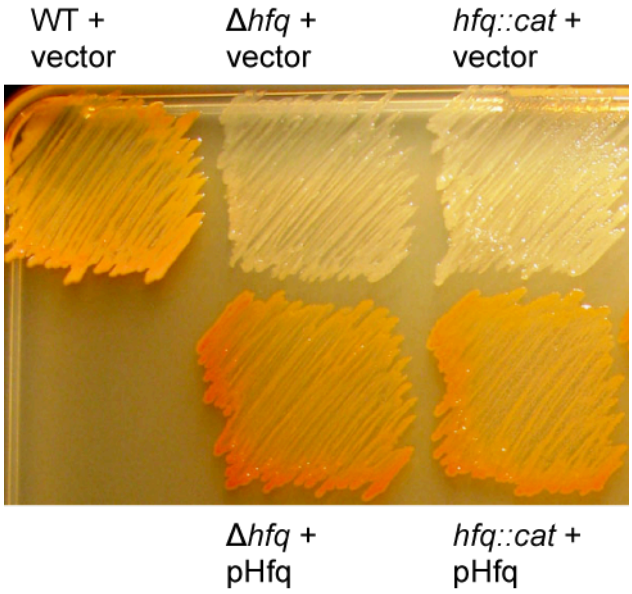


Figure 4.2. The white phenotype of the *hfq* mutants.

The pink color of the WT strain is indicative of prodigiosin production. The *hfq* mutants are white reflecting absence of prodigiosin production and this can be restored by expression of *hfq in trans*. The vector used was pFAJamp and pHfq refers to pNB18.

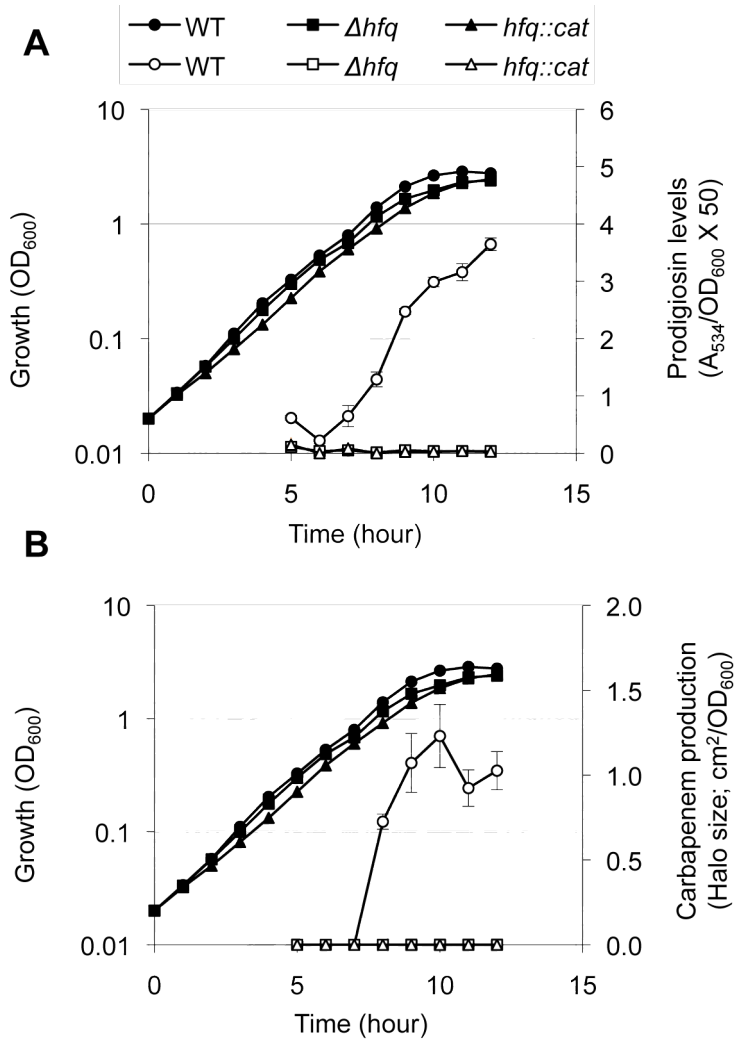


Figure 4.3. Pig and Car production measured throughout growth for the S39006 *hfq* mutants.

Data in both graphs are from the same experiment and represent the mean ± SD of triplicate samples. Growth in LB of the WT and *hfq* mutant strains, *Δhfq* and *hfq::cat*, was measured by taking the OD_{600} values every hour and is shown in both graphs depicted by filled symbols. Pig and Car production are shown by open symbols. (A) Prodigiosin production was assayed by measuring the absorbance of cell lysate, (B) whereas cell filtrate was assayed for Car production which is represented by halo area of the zone of antibiosis of an *E. coli* sensor strain.

4.2.3 Swimming and swarming motility are impaired in the *hfq* mutant

The effect of loss of Hfq on swimming and swarming motility was examined. S39006 WT swarms on 0.75% LBA containing Eiken agar, which is more wettable than Bacto agar (Harshey, 2003). Compared with the parent strain, the *hfq* mutants showed reduced swimming and did not swarm (Figure 4.4). Expression of *hfq in trans* restored

swimming motility whereas swarming was partially complemented, probably reflecting the sensitivity of the assay conditions. In another experiment liquid growth media were shown to induce a tendril-like zone of abiosis which was concomitant with surfactant production required for swarming (data not shown). Because of the role of Hfq in motility, the relative expression of the *rhlA* and *flhC* genes in the Δ*hfq* strain was measured by qRT-PCR. In a previous study, it was shown that homologues of *rhlA* and *flhC* were essential for swarming and swimming motility, respectively (Williamson et al., 2008). *rhlA* encodes a protein homologous to *Pseudomonas aeruginosa* RhlA which is involved in the synthesis of surface-active 3-(3-hydroxyalkanoyloxy) alkanoic acids (HAAs), the precursors of rhamnolipid surfactants, and *flhC* is the master regulator of the flagella operon in *E. coli*. The qRT-PCR results indicated that *rhlA* expression was decreased in the Δ*hfq* strain while, surprisingly, expression of *flhC* was increased (Figure 4.9).

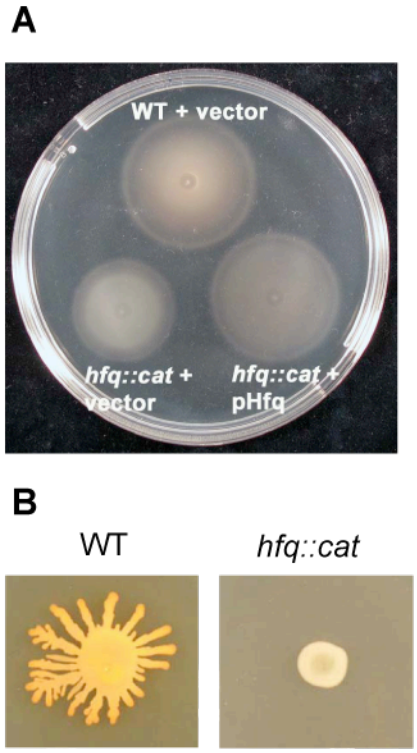


Figure 4.4. The impact of an *hfq* deletion on swimming and swarming.

The *hfq* mutant is decreased for swimming (A) and swarming motility (B). The vector used was pQE80oriT and pHfq refers to pNB36.

4.2.4 The role of Hfq in stress tolerance

Previous studies have shown that Hfq contributes to tolerance of various stresses, including nutrient deprivation, membrane perturbation, high osmolarity, and oxidation (Chao and Vogel, 2010). Therefore, the ability of the WT and *hfq* mutants to grow and survive under different stress inducing conditions was compared. In minimal medium with glucose as the sole carbon source, Hfq did not play a noticeable role in the ability of S39006 to grow under nutrient limiting conditions (Figure 4.5A). More interestingly, the slight lag in growth of Δ*hfq* compared with that of WT in LB medium was reversed in minimal medium, with WT cells showing a delayed entry into exponential phase. Viable cell counts from stationary phase after 30 hr of growth showed no differences (data not shown). Additional assays of stress inducing factors showed no difference between the WT and *hfq*::*cat* strain for growth and survival on media supplemented with NaCl (2.5%) - to test for the effect of high osmolarity, and SDS (0.1%) and EtOH (2.5%) to test for the effect of membrane perturbing agents. The only

observation of note was that *hfq*::*cat* cells stressed with SDS formed smaller colonies compared with those of WT and this phenotype was complemented by the expression of *hfq in trans* (data not shown). Finally, the ability of the *hfq::cat* strain to withstand oxidative stress was tested using an H_2O_2 disk diffusion assay on a lawn of bacteria. Compared with the WT, the *hfq::cat* strain showed a larger halo of inhibition of growth from the H_2O_2 which could be complemented by the expression of *hfq in trans* (Figure 4.5B). The results of the different assays indicate that Hfq plays a role in withstanding oxidative and membrane perturbation stresses but not osmotic stress or nutrient deprivation.

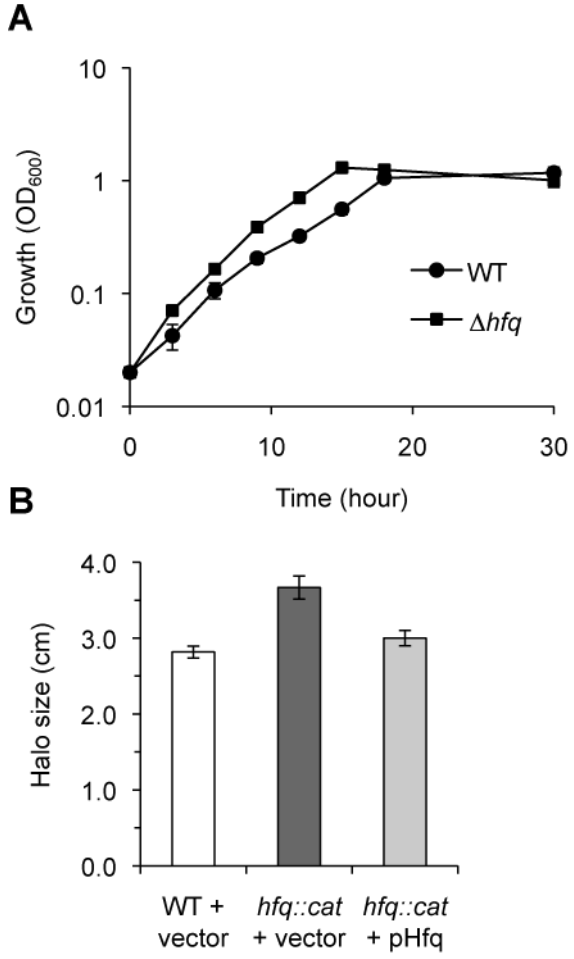


Figure 4.5. The *hfq* mutant resists nutritional stress but is susceptible to oxidative stress.

(A) The affect of nutritional stress was observed by growing the WT and *Δhfq* strains in minimal media supplemented with glucose. (B) Resistance to oxidative stress was observed using an

H_2O_2 disk diffusion assay on agar plates spread with overnight cultures. The diameter of the zone of inhibition of growth by H_2O_2 was measured as halo size, and the difference between the WT and the *hfq* mutant is significant with a *P* value of <0.001 according to a t-test. Data shown in (A) and (B) represent the mean ± SD of triplicate samples.

4.2.5 The *hfq* mutant can be complemented heterologously

The ability of exogenously expressed Hfq orthologues from closely related species to restore prodigiosin production was investigated. *E. coli hfq* is reported to have multiple promoters contained within the upstream intergenic region and the *miaA* gene (Tsui et al., 1996). Therefore, the *hfq* gene with the upstream region of *miaA*, which would include putative native promoters, was cloned into the low-copy number vector pFAJamp in order to complement the *hfq* mutants. This strategy to complement the *hfq* mutants with the native S39006 *hfq* gene was used as well as with orthologous *hfq* genes from two closely related species , *E. coli* and *Citrobacter rodentium*, (Figure 4.1) which also have genetically conserved operons. The level of prodigiosin production restored in the complemented strains was dependent on the different *hfq* genes (Figure 4.6A). S39006 *hfq* restored prodigiosin levels to approximately 35% of WT levels whereas *E. coli* and *C. rodentium hfq* restored prodigiosin levels to 80% and 60% of WT levels. This could have been due to differences in the C-terminal region of the respective Hfq proteins. Alternatively the results could have been attributable to differential expression of *hfq* by the different native promoters for each construct. The *hfq* genes were therefore re-cloned under the T5 promoter of the medium-copy vector, pQE80oriT, in order to control for the level of *hfq* expression. The results showed that, when expressed from the same promoter the S39006 and *E. coli hfq* orthologues restored prodigiosin production to approximately the same level at 45% of the WT level whereas the level of prodigiosin for *C. rodentium hfq* remained at 60% of the WT level (Figure 4.6A). Overexpression of S39006 *hfq* in WT and *hfq::cat* resulted in decreased cell viability and prodigiosin production (data not shown). Taken together, these results show that there is a positive correlation between the level of *hfq* expression and prodigiosin production, but only up to a certain point since an over abundance of Hfq impacts on cell viability and decreases prodigiosin production. This suggests that the stoichiometry of Hfq to target RNA is important for activation of prodigiosin production.

Finally, expression of *hfq in trans* also restored production of the carbapenem production (Figure 4.6).

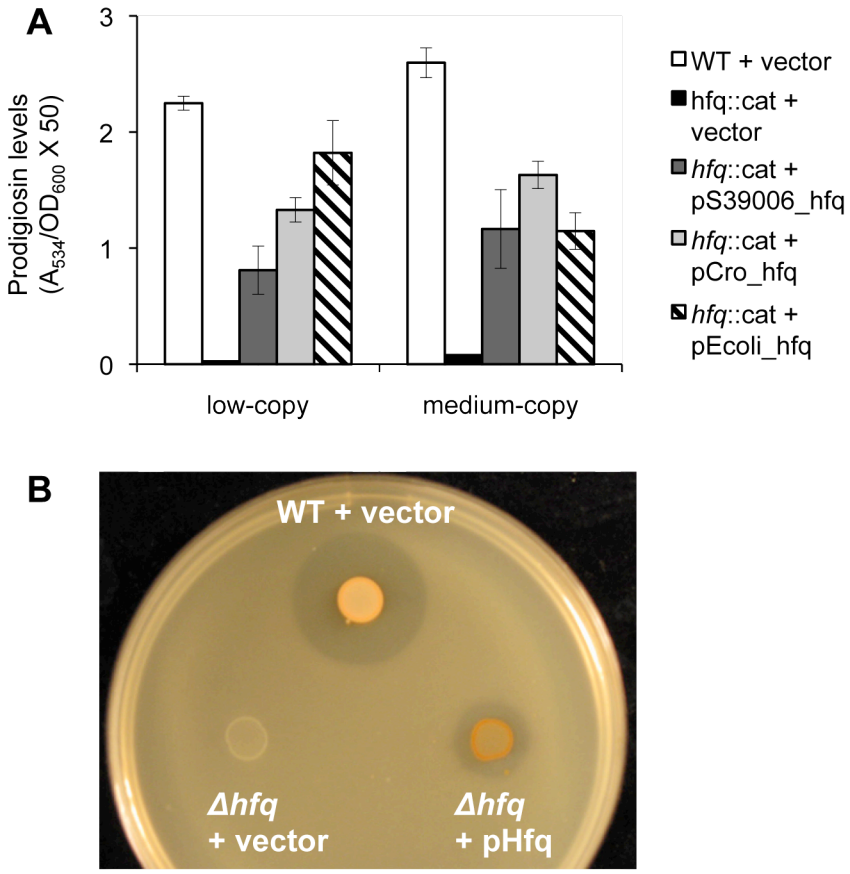


Figure 4.6. Heterologous complementation of the *hfq* mutant.

A. Orthologues of *hfq* with its native promoters from S39006, *C. rodentium*, and *E. coli* were cloned into the vector pFAJamp (low-copy) to give, respectively, plasmids pNB18, pNB22, and pNB35. In contrast, the CDS of the *hfq* orthologs were cloned under the same T5 promoter in the expression vector pQE80oriT (medium-copy) to give, respectively, plasmids pNB36, pNB37, and pNB38. In the figure legend, vector, p39006_hfq, pCro_hfq, and pEcoli_hfq refer respectively to pFAJamp, pNB18, pNB22, and pNB35 for columns labeled low-copy. Similarly, the labels refer to pQE80oriT, pNB36, pNB37, and pNB38 for columns labeled medium-copy. Prodigiosin production was measured from strains grown to stationary phase.

B. Expression of *hfq in trans* in the *Δhfq* strain restored carbapenem production, which is represented by the halo area of the zone of antibiosis of an *E. coli* sensor strain. The vector used was pQE80oriT and pHfq refers to pNB36.

4.2.6 The *hfq* mutant phenotype is epistatic

In two rounds of transposon mutagenesis it was attempted, unsuccessfully, to identify bypass mutants of Δ*hfq* that were upregulated for prodigiosin production (data not shown). This suggested that Hfq might play a dominant role in the regulatory hierarchy of prodigiosin production. Moreover, transduction of the *hfq*::*cat* marker into the S39006 *pstS*::Tn and *rsmA*::Tn strains which are hyperpigmented, resulted in hypostasis of the phenotype of the two mutants and epistasis of the white phenotype of the *hfq* mutant (Figure 4.7). *pstS* and *rsmA* are involved in phosphate regulation and post-transcriptional control, respectively (Slater et al., 2003; Williamson et al., 2008).

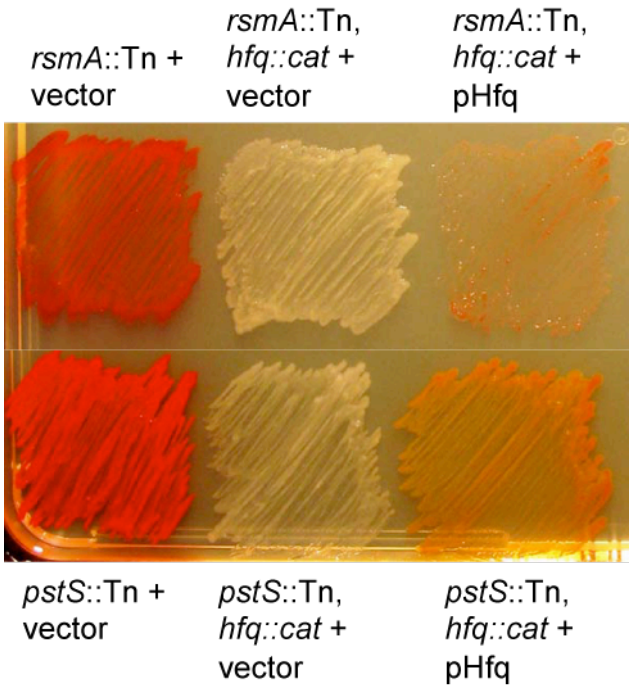


Figure 4.7. An *hfq* mutation is epistatic for prodigiosin production.

The *hfq* mutation (NMW9) resulting in a white phenotype is epistatic over the *rsmA* (NMW7) and *pstS* (ROP2) hyperpigmented transposon insertion mutants. Expression of *hfq in trans* partially restored pigment production. The vector used was pFAJamp and pHfq refers to pNB18.

4.2.7 Hfq regulates the LuxR-type QS regulators, *smaR* and *carR*

Hfq and dependent small regulatory RNAs have been shown to play a role in the version of QS in *Vibrio harveyi* and *V. cholerae*. This system is based on the production of multiple autoinducers and detection by membrane-bound sensor kinases which relay sensory information via a phosphoregulatory cascade (Lenz et al., 2004; Bejerano-Sagie

and Xavier, 2007). It was postulated that Hfq may play a role in the unrelated LuxIR-type QS system of S39006. To investigate the role of Hfq in regulation of the QS response regulators in S39006, the *hfq::cat* marker was transduced into the reporter strains, *smaR::lacZ* and *carR::lacZ*, and reporter strains for three other Pig regulators and the *pigA* and *carA* genes of the prodigiosin and carbapenem biosynthetic clusters. β-galactosidase activity was assayed as a reflection of corresponding gene expression. In agreement with the striking phenotype of the S39006 *hfq* mutants, *pigA* and *carA* expression was down-regulated (Figure 4.8). Interestingly, *smaR* and *carR* expression was also significantly reduced whereas expression of the four other Pig regulators tested, *pigP*, *rap*, *rsmA*, and *prf* was not affected (data not shown).

The results of the transcriptional fusion experiments were also validated by qRT-PCR (Figure 4.9). Relative expression of *gyrB* was measured as a control between the WT and Δ*hfq* strains. The expression of *pigB* was also decreased, reflecting *pigA-O* operon expression (Slater et al., 2003). Expression of *smaI* was not altered, consistent with our earlier data showing that Hfq does not regulate BHL/HHL production. Expression of the *luxS*, *rsmA*, and *rpoS* genes was also examined. *luxS* encodes autoinducer 2 (AI-2) synthase which generates the QS signaling molecule AI-2. As with *smaI*, the *hfq* mutation had no impact on the levels of the *luxS* transcript. Likewise, the expression of *rsmA* was unaffected by *hfq*. *rpoS* encodes the stationary phase sigma factor, σ^S, and is positively regulated by Hfq-dependent small RNAs (Soper et al., 2010). Loss of *hfq* results in decreased *rpoS* transcript levels in *E. coli* (Muffler et al., 1996) and *Salmonella typhimurium* (Brown and Elliott, 1996) due to the protection from degradation by small RNA binding to *rpoS* mRNA (McCullen et al., 2010). This effect was confirmed in S39006, implying that post-transcriptional regulation of *rpoS* by Hfq is conserved.

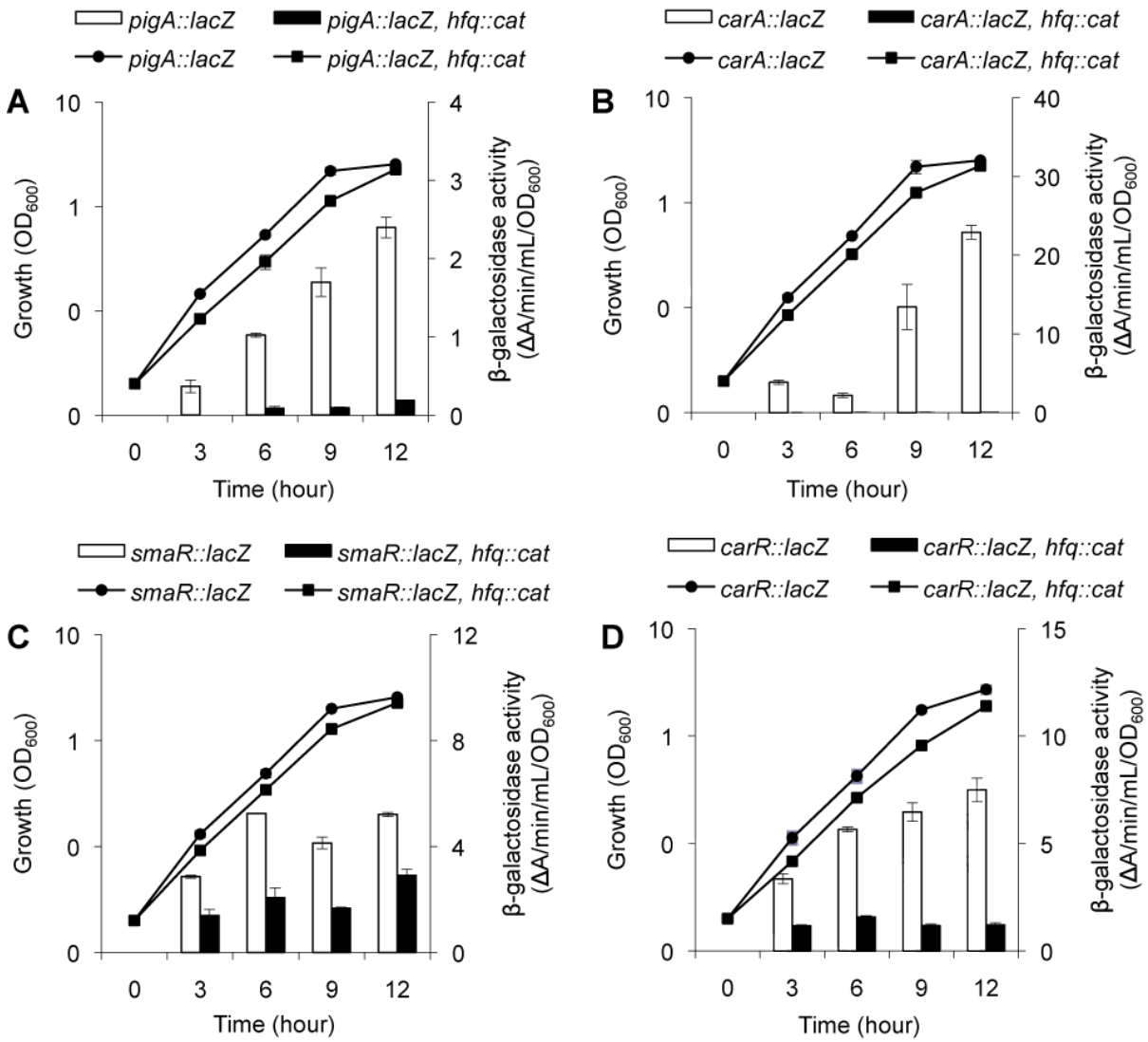


Figure 4.8. The impact of an *hfq* deletion on transcript levels using transcriptional fusion assays.

hfq regulates Car and Pig production by controlling the expression of the biosynthetic genes, *carA-H* and *pigA-O*, and the activators *smaR* and *carR*. β-galactosidase activity was measured from chromosomal *pigA::lacZ* (strain MCP2L) (A), *carA::lacZ* (MCA54) (B), *smaR::lacZ* (SP21) (C) and *carR::lacZ* (MCR14) (D) transcriptional fusions in a WT background (white columns) or in strains containing the *hfq::cat* marker (black columns). The lines in the graphs represent the growth curves of the corresponding strains. Data shown are the means ± SD of at least three independent experiments.

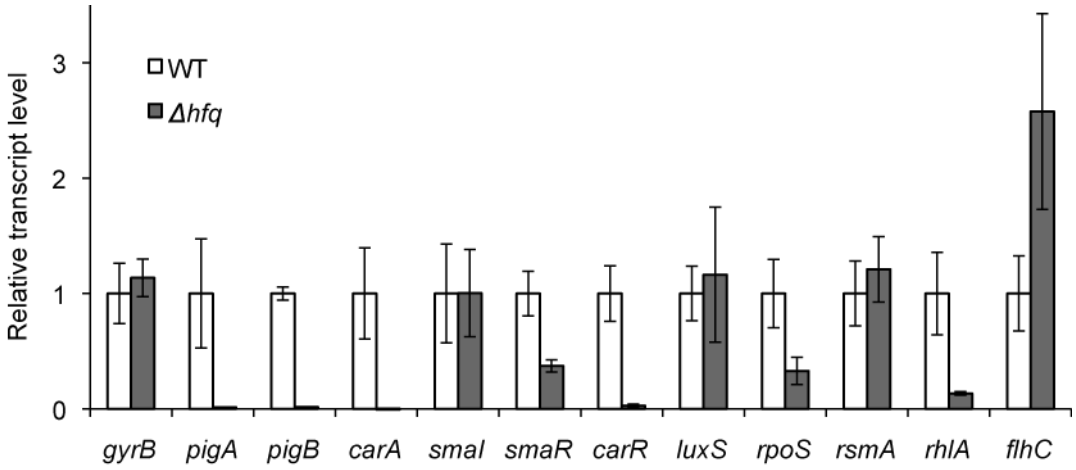


Figure 4.9. The impact of an *hfq* deletion on transcript levels measured by qRT-PCR.

Gene expression for the *Δhfq* strain was measured as transcript levels relative to WT at early stationary phase growth. Values represent average gene expression ± SD from three independent experiments except values for genes *smaI* and *flhC* which were determined from two independent experiments.

4.2.8 Hfq binding *in vitro* to 5' UTR of mRNAs

Hfq is known to regulate mRNAs by binding directly to the 5' UTR. In order to investigate the potential interaction of Hfq with mRNAs of genes of interest, electrophoretic gel-mobility shift assays (EMSA) were performed to assess the binding affinity of purified *E. coli* Hfq protein for target mRNAs. The N-terminal region (1-73 aa), which forms the proximal and distal faces of the doughnut-like structure, is involved in binding to mRNA and sRNA, whereas the C-terminal extension is thought to play only a small, or no, role in riboregulation (Olsen et al., 2010). Because of the total conservation of the N-terminal region of the Hfq protein between *E. coli* and S39006, and the availability of purified *E. coli* Hfq from the lab of Ben Luisi, the protein was used in EMSAs to determine affinity for mRNAs of genes from S39006. As a control, the affinity was initially measured of *E. coli* Hfq to a short truncated region of the S39006 *rpoS* RNA leader region and coding sequence similar to the same region of *E. coli rpoS* RNA, as in a previous study (Figure 4.10) (Updegrove et al., 2008). The low equilibrium dissociation constant (K_d) of *E. coli* Hfq for S39006 *rpoS* (~20 nM), as well as the separation of six Hfq/*rpoS* complexes with increasing Hfq concentration, was similar to the previously

observed result for *E. coli rpoS*. The affinity of Hfq for other mRNA species from S39006 was measured. The 5' UTR and an initial portion of the coding sequence was *in vitro* transcribed from the known transcriptional start sites for *smaI*, *smaR*, *carR*, *pigA*, and *carA* with *smaI* used as a non-target control. *E. coli* Hfq showed a strong affinity for *carA* RNA with an estimated K_d of ~20 nM and formed six Hfq/*carA* complexes. In contrast there was relatively weak affinity for *smaI* RNA with an estimated K_d of ~70 nM and two types of Hfq/*smaI* complexes, suggesting that Hfq may directly regulate *carA* mRNA *in vivo* (Figure 4.11). Because Hfq has a strong affinity for RNA, at high concentrations Hfq can bind to non-target specific mRNA *in vitro* such as the *smaI* 5' UTR (Link et al., 2009). The affinity for *smaR*, *carR*, and *pigA* was intermediate with an estimated K_d between 50-55 nM (Figure 4.10). The secondary structure for the 5' region of *carA* mRNA, predicted using mfold, showed single-stranded AU-rich tracts for which Hfq is known to have high affinity (Lorenz et al., 2010). This technique to examine the comparative affinities of Hfq for different RNA species is not quantitative which would have required the use of a competitive control.

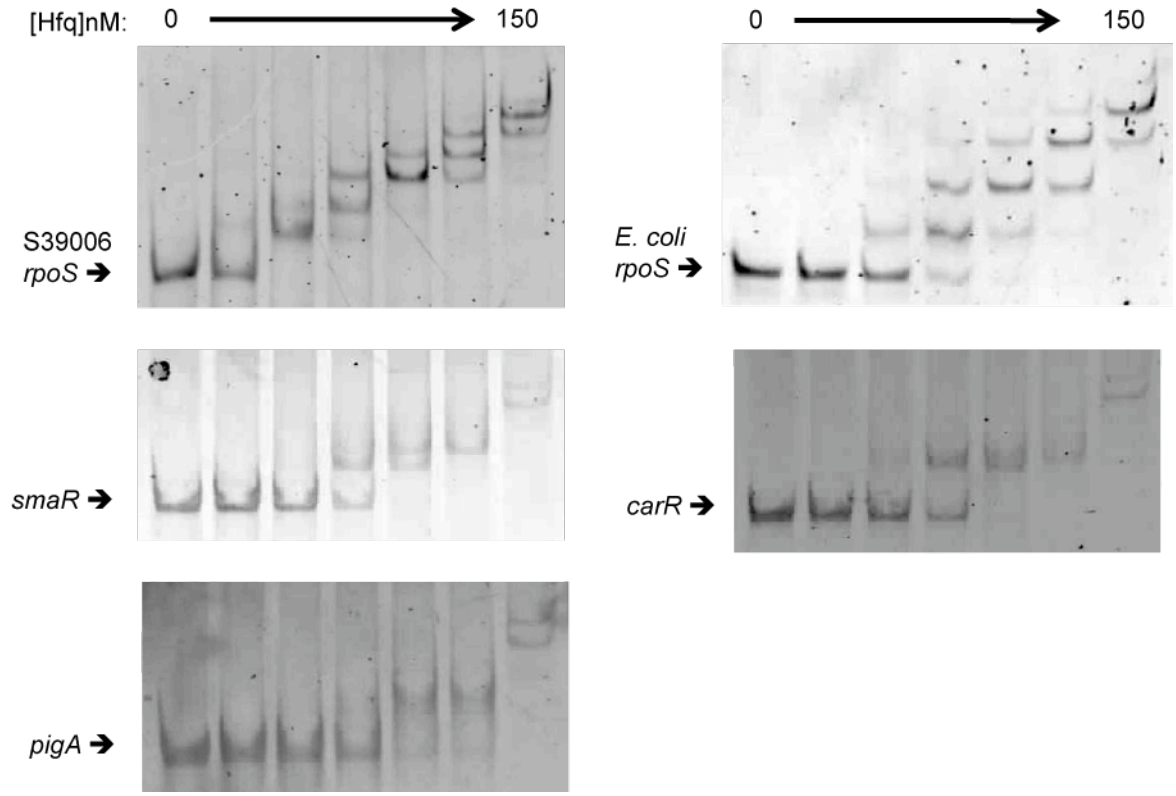


Figure 4.10. Hfq-RNA binding *in vitro*.

E. coli Hfq binds the 5' UTR *in vitro* of S39006 and *E. coli rpoS*, *smaR*, *carR*, and *pigA*. 10 nM of in vitro transcribed RNA was incubated with increasing nM concentrations of Hfq hexamer corresponding to 0, 20, 40, 60, 80, 100, and 150 nM in lanes 1-7. Following 10 min incubation at room temperature, samples were resolved on a native 5% gel.

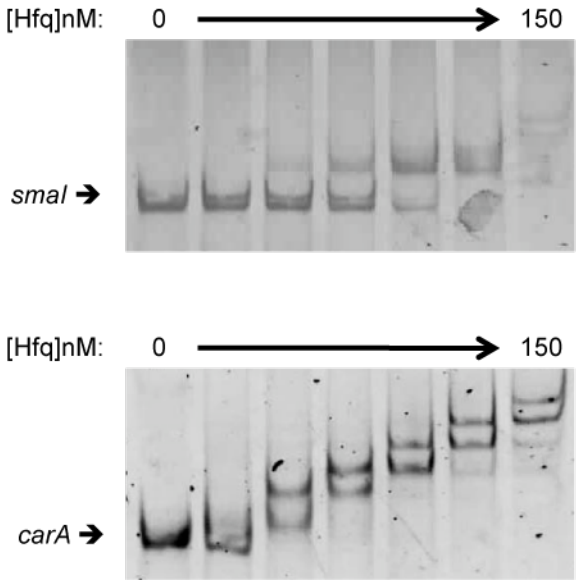


Figure 4.11. *E. coli* Hfq binds *carA* 5' UTR *in vitro*.

10 nM of *in vitro* transcribed *smaI* and *carA* RNA was incubated with increasing nM concentrations of Hfq hexamer corresponding to 0, 20, 40, 60, 80, 100, and 150 nM in lanes 1-7. Following 10 min incubation at room temperature, samples were resolved on a native 5% gel.

4.2.9 The *hfq* mutant is attenuated for virulence in *C. elegans* and potato

Hfq has been shown to regulate virulence factors and secretion systems in *E. coli*, *S. typhimurium*, and several other species (Chao and Vogel, 2010), and it was sought to determine if the S39006 *hfq* mutants were attenuated for pathogenesis using the model hosts, potato and *C. elegans*. The *hfq::cat* mutant strain was significantly reduced for the plant cell wall degrading exoenzymes, pectate lyase and cellulase (Figure 4.12A-B). The decrease in exoenzyme production in the *hfq::cat* strain was reflected in a significant reduction in the levels of rot in potato (Figure 4.12C-D). In order to determine whether the reduction in rot was due to reduced production of exoenzymes or an inability of the *hfq* mutant to grow in the potato, the CFU/g rot was measured and it was shown that there was no significant impact on growth of the *hfq* mutants in the potato environment (data not shown).

S39006 is also virulent in the worm, *C. elegans*. The rate of killing by the *hfq::cat* mutant was significantly reduced compared with that of WT (Figure 4.13A), and these results were also confirmed using the Δ*hfq* strain (data not shown). The survival median (representing the time required for 50% of nematodes to die) was 3 days for

WT, 8 days for the *hfq::cat* strain, 6 days for the complemented *hfq::cat* strain, and 10 days for the conventional control strain, *E. coli* OP50 (OP50) (Kurz et al., 2003). Additionally, worms fed on WT were impaired for growth compared with worms fed on the *hfq* mutants which reached the full adult size at the same rate as worms fed on OP50 (data not shown). Moreover, it was previously shown that *C. elegans* avoided lawns of the pathogenic bacterium, *Serratia marcescens* (Pujol et al., 2001). A similar behavior toward lawns of S39006 WT, but not the *hfq* mutants, was observed while tracking worm survival. It was also shown that the absence of prodigiosin production by the *hfq* mutants was not a factor in causing attenuation of virulence since no difference was observed in the rate of survival of worms fed on the WT (Pig+) or strain MCP2L (*pigA::lacZ*; Pig–) (Figure 4.14). Moreover, another mutant increased for prodigiosin production was attenuated for virulence in *C. elegans*, thereby providing additional evidence that prodigiosin does not appear to play a major role in pathogenesis in this particular strain (see Chapter 5).

The colonization of the worm lumen was visualized by fluorescence microscopy using worms fed on GFP-expressing strains of WT, *hfq::cat*, or OP50 from day 1 through 5 of the survival experiment. Normally, bacterial cells pass from the mouth to the base of the pharynx where the grinder disrupts the cells before entering the intestinal lumen. Nevertheless, the lumen of the worm fed on WT was distended and full of GFP expressing bacteria (Figure 4.13E-F). In contrast, the lumen of the worm fed on OP50 showed little detectable GFP, while the lumen of worms fed on the *hfq::cat* strain was clear of bacteria on day 2 but showed an accumulation of cells by day 4 of the experiment, indicating delayed colonization of the intestine when compared with the WT (Figure 4.13B-D).

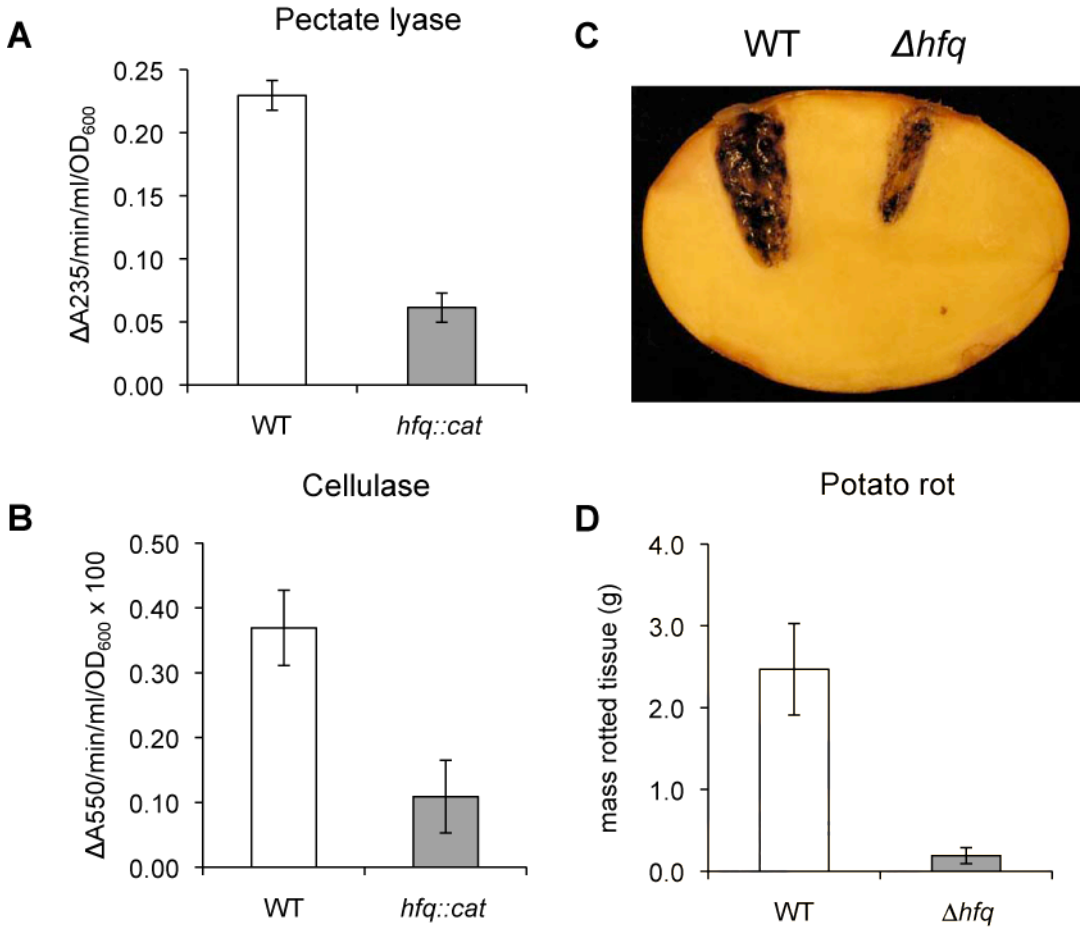


Figure 4.12. An *hfq* mutant is decreased for exoenzyme production and attenuated for virulence in potatoes.

A-B. The amount of pectate lyase and cellulase activity in the culture supernatant of WT and the *hfq::cat* mutant was measured after 17 h growth at 25°C in PMB Medium. Measurement of enzyme activity is described in the experimental procedures. Bars represent the mean ± SE from three independent experiments.

C-D. Comparison of the virulence of the WT and Δ*hfq* strains using the potato tuber rotting assay. Potatoes were inoculated with 10^6 cells of each strain and incubated at 25 °C for 4 days. The area of rot from a sample potato (C) is stained with iodine and the amount of rotten tissue generated is shown below (D). Bars indicate mean ± SE (n = 12). The results were repeatable using the *hfq::cat* strain (data not shown).

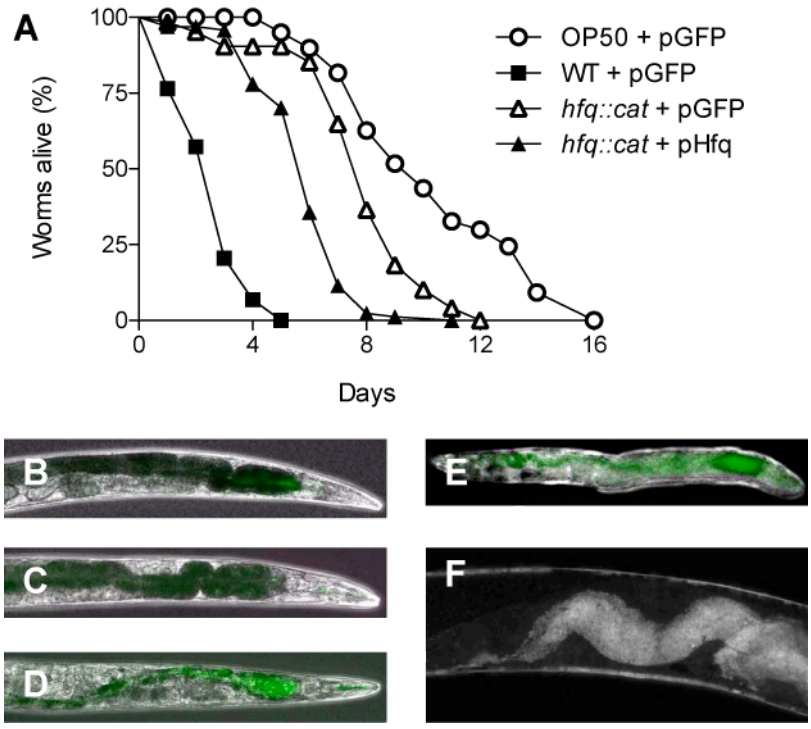


Figure 4.13. An *hfq* mutant is attenuated for virulence in *C. elegans*.

A. Kinetics of killing of *C. elegans* infected by S39006 WT, *hfq::cat*, and complemented *hfq::cat* are shown. pHfq refers to pNB36. Included for comparison is the survival of worms fed on the normal feeder strain *E. coli* OP50. Fifty N2 hermaphrodites were used for each strain and were grown at 25°C on NGM plates.

B-F. Images of worms fed on strains (B) *E. coli* OP50 for 4 days, (C) *hfq::cat* for 2 days, (D) *hfq::cat* for 4 days, and (E-F) WT for 1 day. The lumen of the worm fed on WT is distended and full of GFP expressing bacteria. In contrast, the lumen of the worm fed on OP50 shows little detectable GFP while the lumen of worms fed on the *hfq::cat* strain accumulated in the intestine from day 2 to 4. Images are of worms at 40x magnification for (B-E) and 100x for (F). For (B-E) the phase contrast images were merged with the green fluorescent images, whereas (F) is the fluorescent image with the white color representing green fluorescence.

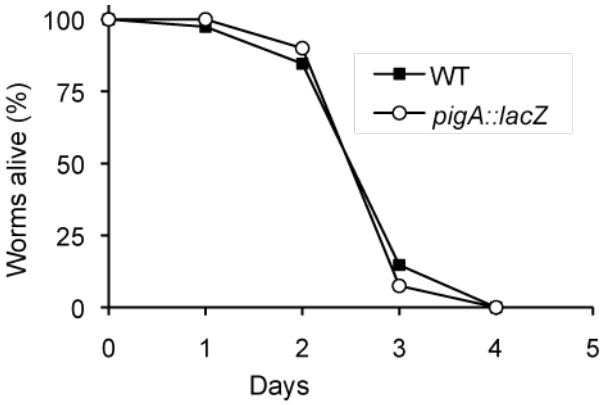


Figure 4.14. Absence of prodigiosin does not affect survival of *C. elegans* on S39006.

Kinetics of killing of *C. elegans* infected by S39006 WT (Pig^+) and strain MCP2L (*pigA::lacZ*; Pig^-). Fifty worms of the temperature-sensitive strain, DH26, were used for each strain and were grown at 25°C using NGM plates.

4.3 Discussion

A growing body of evidence has shown that Hfq is a post-transcriptional regulator of gene expression that plays an important role in regulating stress response and pathogenesis in a variety of bacterial species (Chao and Vogel, 2010). In this study, the role of Hfq in the Gram-negative bacterium, *Serratia* 39006, was investigated. In addition to affecting stress tolerance and virulence in animal and plant hosts, Hfq was found to play a major role in regulating the production of the bioactive secondary metabolites, prodigiosin and carbapenem, partly through regulating the transcript levels of the LuxR-type QS response regulators, SmaR and CarR.

Initial assays showed that the *hfq* mutant has a mild effect on growth in LB media resulting in a slight lag in entry into exponential and stationary phases. Surprisingly, the mild effect on growth was reversed when the S39006 *hfq* mutant was grown in nutrient-limiting conditions, entering exponential phase and reaching stationary phase before WT. This is contrary to the observation that *hfq* deletion mutants in *E. coli* and *V. cholerae* have a growth defect in both minimal medium and LB medium (Ding et al., 2004; Shakhnovich et al., 2009). This indicates that the role of Hfq in carbon metabolism is species-dependent and indicates that the S39006 *hfq* is neither an auxotroph nor lacking in growth due to a nutrient deficiency.

In addition to withstanding the stress of nutrient limiting growth conditions, the S39006 *hfq* mutant was not impacted for viability by the stresses of high salt, ethanol, or SDS which is in contrast to previous studies showing that *hfq* deletion mutants of *Burkholderia cepacia* and *E. coli* are sensitive to these stresses (Sousa et al., 2010). However, the S39006 *hfq* mutant was slightly sensitive to oxidative stress induced by H_2O_2. Thus, the role Hfq plays in resistance to various stresses appears to also differ among bacterial species. The sensitivity of *hfq* mutants to stress is thought to contribute to attenuation of virulence. Therefore the delayed colonization of the worm intestine by the S39006 *hfq* mutant could be a result of sensitivity to the defense response of the worm or other stress conditions not assayed in this study but encountered within the worm.

Studies of the related pathogen, *Serratia marcescens*, showed that it kills *C. elegans* by an infection-like process and that *C. elegans* exhibits the behavior of avoiding lawns of *Serratia marcescens*. Similarly, S39006 WT appears to also kill worms by an infection-like process of the intestine and induces lawn avoidance behavior in the worm. It was also observed that worms did not avoid lawns of the *hfq* mutant indicating that the factor inducing lawn avoidance is down-regulated in the *hfq* mutant. This factor was identified in the related strain, *S. marcescens* Db10, as serrawettin W2, a secreted surfactant encoded by the *swrA* gene - forming a large locus (> 17kb) predicted to encode a nonribosomal peptide synthase (NRPS) (Pradel et al., 2007). Interestingly, a BLAST search for homologues of *swrA* in S39006 identified an operon of genes which code for NRPSs and collectively is ~17.3 kb long. A cyclic peptide transporter precedes this cluster of genes suggesting that the product synthesized by these genes is exported from the cell. Also the *rhlA* gene involved in production of another surfactant, which is required for swarming, is downregulated in the absence of Hfq and may be involved in inducing lawn avoidance behavior in *C. elegans*. Thus, this study demonstrates that Hfq is also a regulator of surfactant production involved in host-pathogen interactions.

Moreover, the *hfq* mutant was also attenuated for virulence in potato and this was reflected in a decrease in activity for the plant cell wall degrading exoenzymes, pectate lyase and cellulase, which are exported by the Type II secretion system (T2SS) (Barnard et al., 2007). While all previous studies of *hfq* deletion mutants have looked at

the role of Hfq in animal pathogenesis, to the knowledge of the author, this is the first report of a role for Hfq in plant pathogenesis. Therefore, Hfq and possible dependent sRNAs play a role in regulating secretion and/or production of exoenzymes involved in plant pathogenesis.

Also related to the role of Hfq in virulence, the *hfq* mutants were impaired for swimming, and swarming motility was abolished . The decrease in swarming motility is probably due to the significant decrease in *rhlA* expression that was previously shown to be a swarming-dependent gene involved in surfactant production (Williamson et al., 2008). Surprisingly, *flhC*, the master regulator of the flagellar operon, was upregulated in the Δ*hfq* mutant despite the *hfq* mutant being impaired for swimming motility. This is in contrast to the role Hfq plays in flagellum production in *S. typhimurium*. In the latter host, 87% of the flagellar genes were downregulated in an *hfq* mutant, which was impaired for swimming motility, and Hfq was shown to bind to *flhDC* mRNA (Sittka et al., 2008). Therefore, Hfq may play a unique role in regulating flagellum production in S39006.

Key phenotypes of the S39006 *hfq* mutants are the loss of prodigiosin and carbapenem production and the corresponding regulation by Hfq of the LuxR-type regulators, SmaR and CarR. Studies from this lab have shown that multiple environmental signals are involved in regulating secondary metabolism in this strain, including quorum sensing, gluconate, and phosphate levels (Thomson et al., 2000; Fineran et al., 2005a; Gristwood et al., 2009). Despite extensive transposon mutagenesis in previous studies to identify genes involved in the regulation of prodigiosin, surprisingly, a mutant with a transposon insertion in the *hfq* gene was never found. However, selection for transposon mutants affected in *rsmA* (also a small RNA-binding protein involved in post-transcriptional regulation) and its target small RNA *rsmB* indicated that Hfq and its dependent small RNAs also may also be involved in prodigiosin production (Williamson et al., 2008). Interestingly, the phenotype of an *rsmA* mutant is hyperpigmentation and increased swarming, while the *hfq* mutant is hypopigmented and decreased for swarming. Therefore, it appears these two global regulators of gene expression play opposing roles in the regulation of several key phenotypes within S39006.

Investigation of the transcriptional control of secondary metabolism by Hfq revealed a decrease in transcript levels of the *car* and *pig* clusters and the QS response regulators, *smaR* and *carR*, as measured by transcriptional fusions and qRT-PCR. Additionally, it was shown that there were no changes in the transcript levels of three other regulators of prodigiosin production, including *rsmA*. Moreover, BHL/HHL levels were unaffected in the *hfq* mutant as were transcript levels of *smaI* and *luxS*, the latter controlling production of autoinducer-2 (AI-2) involved in interspecies communication. Therefore, this study confirms the link between Hfq and the QS system in S39006 through modulation of the LuxR-type regulators, SmaR and CarR, independently of the impact of the BHL/HHL cognate molecule. It suggests that Hfq positively regulates the transcripts of *smaR* and *carR* mRNA which may occur via Hfq-dependent sRNAs that activate translation, resulting in an increased stability of the transcripts and protection from degradation. This type of positive post-transcriptional regulation by Hfq is similar to what has been observed for *rpoS* mRNA by the sRNAs, DsrA, RprA, and ArcZ in *E. coli* and *S. typhimurium* (Majdalani et al., 2002; Majdalani et al., 2005; Papenfort et al., 2009). Additional evidence for this suggestion is provided by the decrease in *smaR* and *carR* transcript levels in the S39006 *hfq* mutant - as measured by qRT-PCR data similar to the decreased levels observed for *rpoS* transcript. The decrease in *rpoS* transcript levels can be explained by the absence of Hfq and sRNA binding to the transcript. Binding of Hfq and sRNAs to rpoS mRNA protects it from endonucleases and activates translation (Folichon et al., 2003).

Electrophoretic gel shift assays of *E. coli* Hfq with the 5' UTR regions of *carR* and *smaR* mRNA species did not result in a strong binding affinity compared to that of the control *rpoS* mRNA. Therefore further work is needed to show whether Hfq directly regulates these mRNA species. Surprisingly, Hfq bound to *carA* 5' UTR RNA with a low estimated K_D similar to that for the *rpoS* transcript. A secondary structure prediction for the 5' UTR of *carA* showed that it contained a single-stranded region with three ARN repeats and two AU loop regions which has been shown to be bound by Hfq in previous studies (Soper and Woodson, 2008; Link et al., 2009).

During the course of this study, a short 124 bp non-coding sequence was reported to activate the heterologous expression in *E. coli* of the *hap* gene cluster, which

is responsible for prodigiosin production in *Hahalla cheujensis*, and was reported to be Hfq-dependent (Kwon et al., 2010). A BLAST search for the non-coding region did not identify any homologous sequence in the S39006 genome. Moreover, a 1.1 kb region from the 3' end of the *pig* cluster expressed *in trans* was unable to restore pigment production in the *hfq* mutant (data not shown). A bioinformatic search based on the genetic context of sRNAs in closely related organisms has revealed in the S39006 genome several known conserved core sRNAs sharing sequence similarity with RprA, CyaR, RybB, RyhB, MicA, SgrS, and SpoT 42. Future work will focus on identifying Hfq-dependent sRNAs and their target mRNAs involved in the regulation of prodigiosin production in S39006.

S39006 Hfq was shown to regulate secondary metabolism through positive regulation of the QS response regulators, CarR and SmaR. While a decrease in *carR* expression explains the decrease in levels of expression of the *car* operon, the decrease in *smaR* expression alone does not explain the decrease in levels of expression of the *pig* cluster. SmaR is an atypical LuxR-type regulator that functions by a derepression mechanism; the *pig* operon being derepressed at high cell density in response to an increased concentration of BHL/HHL (Figure 4.15) (Thomson et al., 2000; Tsai and Winans, 2010). Therefore, Hfq must regulate other factors, directly or indirectly, that positively regulate expression of the *pig* cluster. It was investigated whether the decrease in *rpoS* transcript levels in the *hfq* mutant may result in the decreased prodigiosin production. However, expression of *rpoS in trans* in the *hfq* mutant did not restore prodigiosin production (data not shown).

In summary, this study identifies Hfq as an essential regulator of secondary metabolism and virulence in S39006. It identifies two LuxR-type regulators and a surfactant that are positively regulated by Hfq and suggests the involvement of Hfq-dependent sRNAs in the regulation of secondary metabolism. It also demonstrates that the two major post-transcriptional regulators, Hfq and RsmA, play opposing roles in the regulation of key phenotypes.

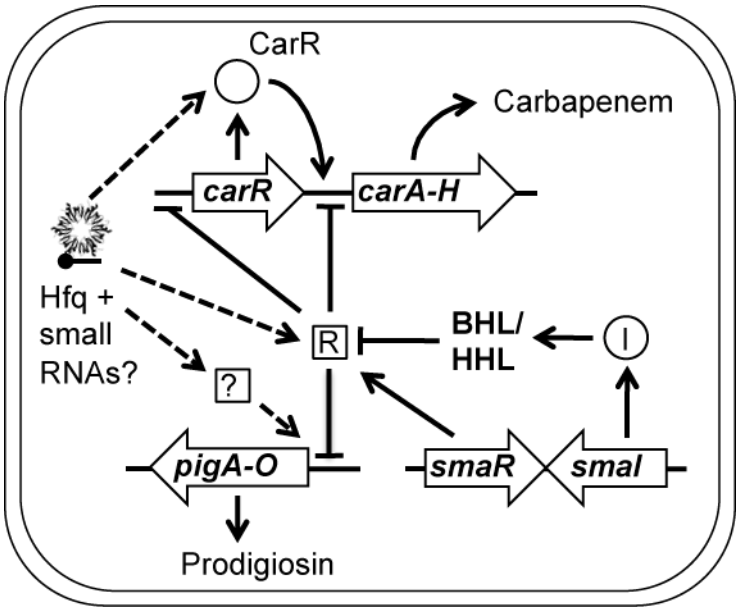


Figure 4.15. Model of the S39006 quorum sensing circuit.

The S39006 QS system is encoded by the *smaIR* cluster. SmaI synthesizes the QS signalling molecules, BHL/HHL. At low cell density, SmaR represses the expression of the *pig* and *car* clusters and *carR*. At high cell density, the increasing concentration of BHL/HHL induces SmaR to derepress from its target genes. Hfq and potential small RNAs stabilize the mRNA transcripts of *carR* and *smaR*. Moreover, Hfq is needed for the expression of an unknown positive regulator of the *pig* cluster or activates the cluster directly. Activation is depicted by pointed arrowheads and repression by flat arrowheads. Dashed lines indicate putative regulation by Hfq.

Chapter 5. The stationary phase sigma factor RpoS regulates secondary metabolism and pathogenesis

5.1 Introduction

In S39006, Hfq was shown to be a major regulator of carbapenem and prodigiosin production, stress tolerance, and animal and plant pathogenesis (see Chapter 4). Loss of *hfq* was shown to result in decreased *rpoS* transcript levels, initially in *E. coli* (Muffler et al., 1996) and *Salmonella typhimurium* (Brown and Elliott, 1996) due to the protection from degradation by small RNA binding to *rpoS* mRNA (McCullen et al., 2010). Likewise, an S39006 *hfq* deletion mutant resulted in decreased transcript levels of *rpoS*, which encodes σ^S, the stationary phase alternative sigma factor subunit of RNA polymerase. Loss of Hfq abolishes prodigiosin and carbapenem production and attenuates virulence in the animal and plant hosts, *C. elegans* and potato. Therefore, it was decided to investigate whether the phenotypes regulated by Hfq were mediated through its control of *rpoS*.

5.2 Results

5.2.1 Prodigiosin and carbapenem production is increased in the *rpoS* mutant

A mutant with a transposon insertion in *rpoS* was identified from a random transposon mutagenesis screen. The *rpoS*::Tn mutation was transduced into a clean WT genetic background generating mutant strain NMW25. An assessment across growth showed that the production of prodigiosin and carbapenem was elevated in the *rpoS* mutant with Pig levels increased by over 2-fold (Figure 5.1A-B). This was reflected in an increased transcription of the *pigA* and *carA* genes of the prodigiosin and carbapenem biosynthetic clusters measured using transcriptional fusions, although transcription of *carA* was only significantly elevated in stationary phase (Figure 5.1C-D). However, no difference was observed for lactone production or the rate of growth, except that the *rpoS* mutant entered stationary phase an hour earlier and reached a lower optical

density. Viable counts throughout growth showed no differences (data not shown). Finally, there was no difference in the colony morphology of the *rpoS* mutant.

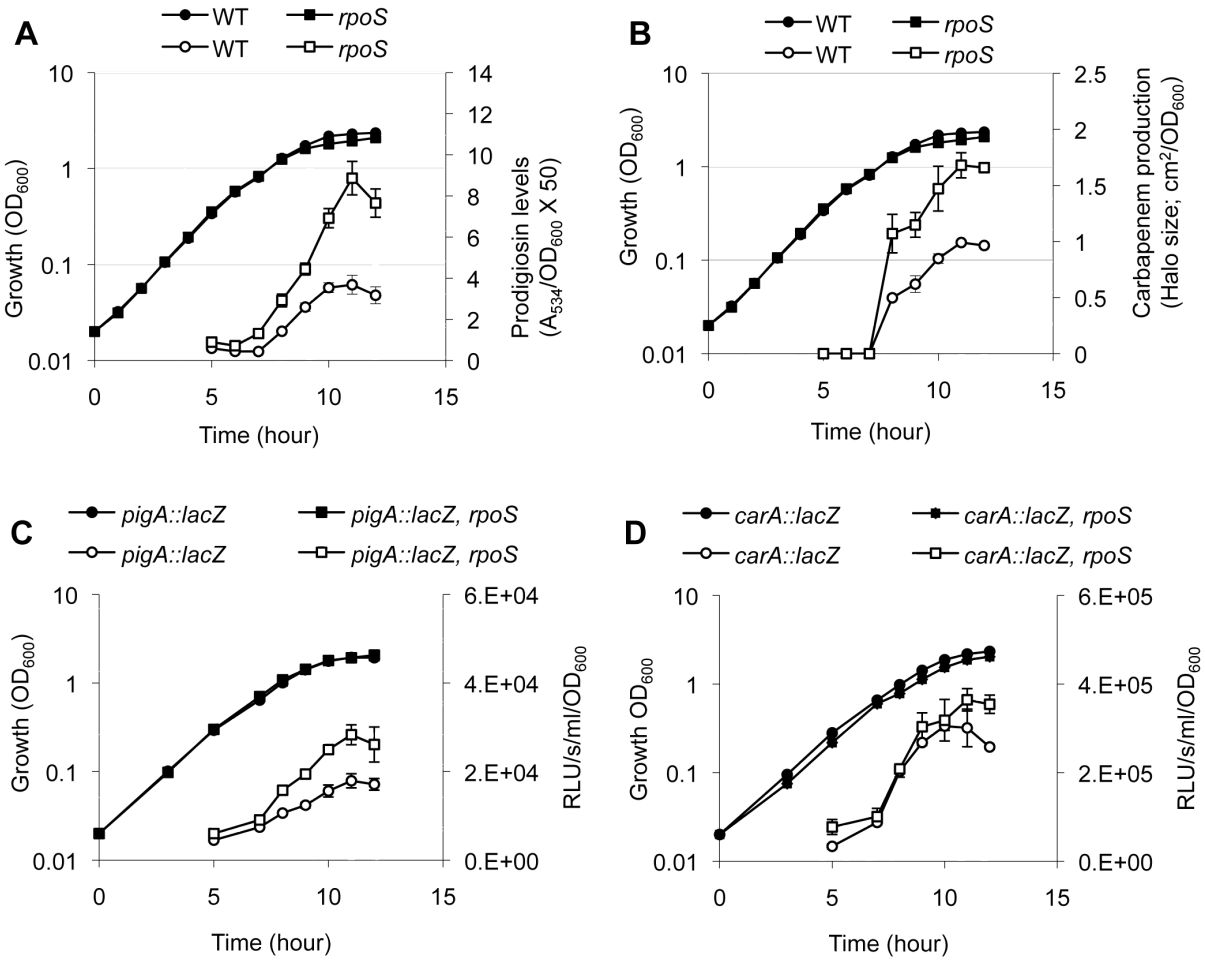


Figure 5.1. An *rpoS* mutant shows increased production of Pig and Car.

A-B. Growth in LB of the WT and *rpoS::*Tn mutant strain NMW25, was measured by taking the OD_{600} values every hour and is shown in both graphs depicted by filled symbols. Pig and Car production are shown by open symbols. (A) Prodigiosin production was assayed by measuring the absorbance of cell lysate, (B) whereas cell filtrate was assayed for Car production which is represented by halo area of the zone of antibiosis of an *E. coli* sensor strain.

C-D. *rpoS* regulates Car and Pig production by controlling the expression of the biosynthetic genes, *carA-H* and *pigA-O*. β-galactosidase activity shown by open symbols was measured from chromosomal *pigA::lacZ* (strain MCP2L) (A) and *carA::lacZ* (MCA54) (B) transcriptional fusions in a WT background or in strains containing the *rpoS*::Tn mutation. Lines with filled symbols represent the growth curves of the corresponding strains. Data shown in all graphs are the mean ± SD of at least three independent experiments.

5.2.2 RpoS is a repressor of swimming and swarming motility

The effect of loss of *rpoS* on swimming and swarming motility was examined. Compared with the parent strain, the *rpoS* mutant showed no difference in swimming motility but reduced motility was observed for the complemented strain (Figure 5.2A). S39006 WT does not swarm on 0.75% LBA containing Bacto agar but the *rpoS* mutant displayed swarming motility (Figure 5.2B). Attempts to complement the strain gave inconsistent results which is probably due to the sensitivity of the assay conditions.

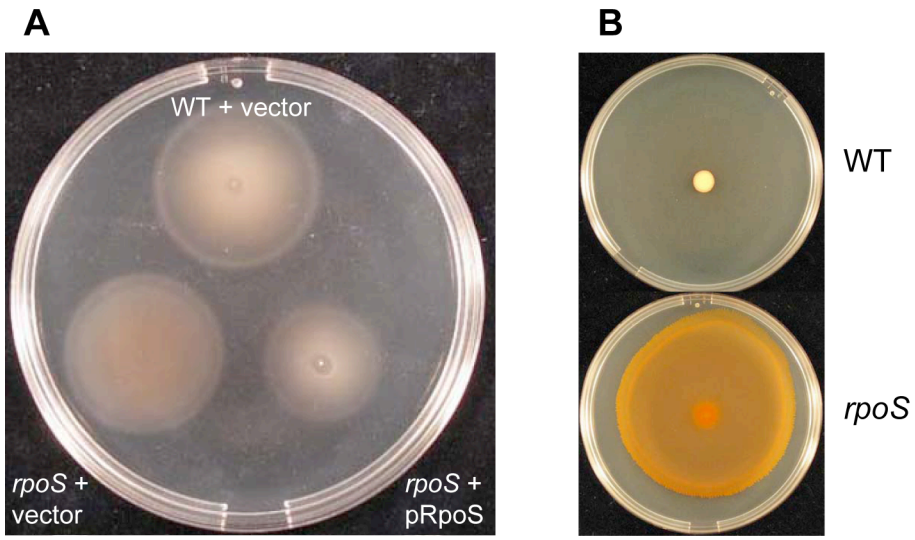


Figure 5.2. *rpoS* suppresses swimming and swarming motility.

Expression of RpoS *in trans* decreased swimming motility (A) and the *rpoS* mutant (NMW25) is increased for swarming motility (B). The vector used was pQE80oriT and pRpoS refers to pNB43.

5.2.3 The role of *rpoS* in stress tolerance

E. coli RpoS has been shown to play a critical role in coordinating the expression of genes involved in withstanding various stresses, including oxidative stress and starvation (Dong and Schellhorn, 2010). Therefore, the *rpoS* mutant was investigated for its ability to withstand different stresses. Unlike the *hfq* mutant, the *rpoS* mutant was unable to grow in minimal medium (MM) with glucose as the sole carbon source. However, colonies grew when cells were diluted in LB media and plated onto the MM plates suggesting that the mutant was auxotrophic for essential amino acids. This was

confirmed by supplementing the media with casamino acids which allowed for the normal growth of the *rpoS* mutant. It was also observed that addition of casamino acids depleted for vitamins resulted in increased pigment production in both the WT and *rpoS* mutants indicating that perhaps particular vitamins may suppress pigment production (Figure 5.3). Additionally, the *rpoS* mutant cultured on MM plates resulted in spontaneous revertants that could grow on MM without the addition of amino acids.

Additional assays of stress inducing factors showed no difference between the WT and the *rpoS* mutant for growth and survival on media supplemented with SDS (0.1%) and EtOH (2.5%) to test for the effect of membrane perturbing agents. However, the *rpoS* mutant was reduced 10-fold for growth compared to WT when grown on media supplemented with NaCl (2.5%) - to test for the effect of high osmolarity, a phenotype which was complemented (data not shown).

Finally, the ability of the *rpoS* mutant to withstand oxidative stress was tested using an H_2O_2 disk diffusion assay on a lawn of bacteria. Compared with the WT, the *rpoS* mutant showed a marginally larger halo of inhibition of growth from the H_2O_2 which could be complemented by the expression of *rpoS in trans* (Figure 5.4). This was in contrast to the much larger halo observed for the *hfq* mutant and indicates that *rpoS* plays a small role in resistance to oxidative stress, when compared to the role of *hfq*.

The results of the different assays indicate that RpoS plays a role in withstanding nutritional, osmotic, and oxidative stresses but not membrane perturbation stresses.

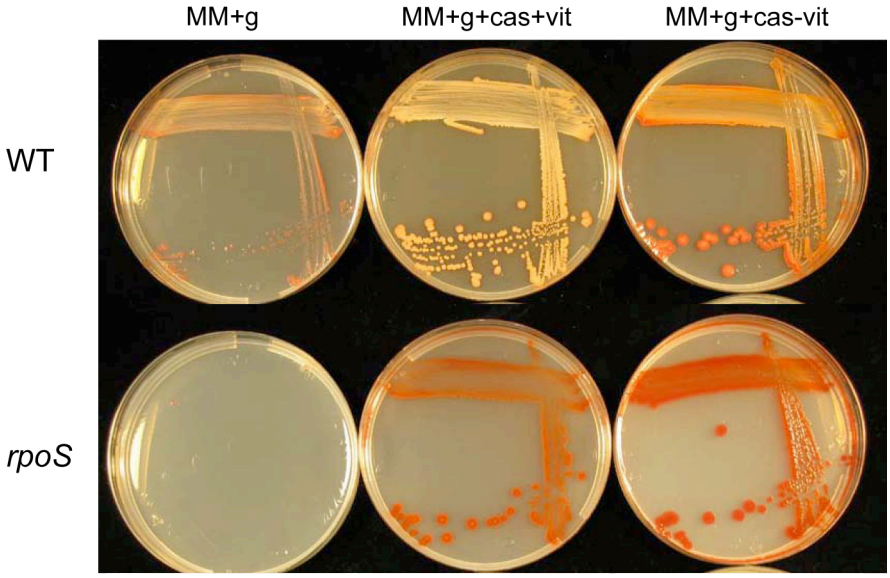


Figure 5.3. Growth of the *rpoS* mutant on minimal media.

The WT and *rpoS* mutant (NMW25) was cultured on MM with glucose as the sole carbon source (MM+g) and supplemented with casamino acids (+cas). Two different types of casamino acids were tested, one from a standard preparation (+vit) and the second depleted for vitamins (-vit). Plates shown are after 5 days of incubation.

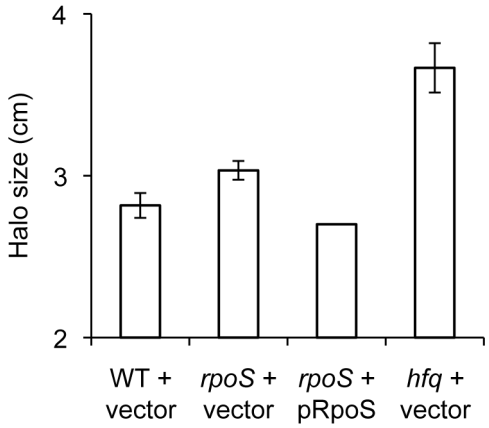


Figure 5.4. The *rpoS* mutant is susceptible to oxidative stress.

Resistance to oxidative stress was observed using an H_2O_2 disk diffusion assay on agar plates spread with overnight cultures. The diameter of the zone of inhibition of growth by H_2O_2 was measured as halo size, and the difference between the WT and the *rpoS* mutant (NMW25) is significant with a *P* value of <0.05 according to a t-test. The vector used was pQE80oriT and pRpoS refers to pNB43. The phenotype of the *rpoS* mutant was compared to the phenotype of the *hfq::cat* mutant (NMW9). Data shown represent the mean ± SD of triplicate samples.

5.2.4 *rpoS* is epistatic over other regulators of prodigiosin production

The interaction of *rpoS* with other factors in the wider regulatory network for secondary metabolism was explored. Double mutants of *rpoS*::Tn and various Pig regulator transposon mutants were generated by transduction. Pig production was measured in the double mutants and compared to the isogenic single mutants to determine if the *rpoS* phenotype was epistatic. *rpoS*::Tn was found to bypass mutations in *pigQ* and *pigW*, which are part of a two-component system, and restore pigment production (Figure 5.5). However, it was not able to bypass mutations in *pigP*, *rap*, *rsmA* and *hfq* (data not shown).

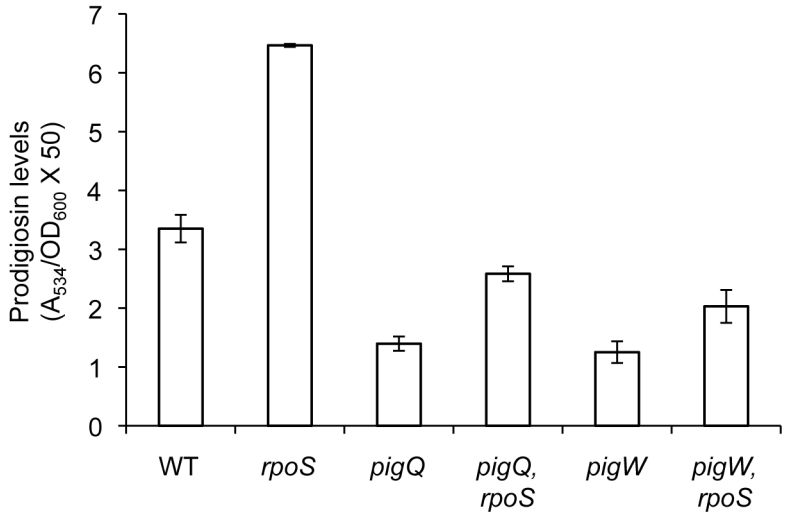


Figure 5.5. Deletion of *rpoS* function restores pigment production in the *pigQW* two-component system mutants.

Prodigiosin production measured at stationary phase from double mutants of *rpoS* with the Pig regulators, *pigQ* and *pigW*, were increased for Pig production compared to the isogenic single mutants. The strains assayed were the WT, NMW25 (*rpoS*), PIG17S (*pigQ*), NMW27 (*rpoS*, *pigQ*), PIG62S (*pigW*), and NMW28 (*rpoS*, *pigW*). Data shown represent the mean ± SD of triplicate samples.

5.2.5 The small RNAs DsrA and RprA regulate prodigiosin production

Previous research has shown that the *E. coli rpoS* mRNA contains an inhibitory stem loop structure in the 5' UTR occluding the ribosome binding site from translation initiation (Brown and Elliott, 1997). Hfq facilitates the complementary base pairing of two small regulatory RNAs, DsrA and RprA, to the 5' UTR weakening the structure of the

inhibitory stem loop and resulting in increased *rpoS* translation (Majdalani et al., 1998; Majdalani et al., 2002). The *E. coli rpoS* mRNA has been shown to have an unusually long 567 nt 5' UTR with transcription initiating midway within the upstream *nlpD* gene (Takayanagi et al., 1994). A comparison of the hypothetical S39006 *rpoS* 5' UTR with the homologous region in *E. coli* showed a high degree of conservation with complete identity for the overlapping DsrA/RprA interaction site indicating that small RNA homologues in S39006 may positively regulate the translation of *rpoS*. Small RNAs are located in intergenic regions and a comparison of the genomic context of *E. coli rprA* (105 bp) to the homologous region in S39006 identified a conserved region (111 bp) with a similar predicted secondary structure (Figure 5.6A-B). Interestingly, S39006 *rprA* showed a high degree of variation in the 5' half of the sequence whereas the 3' half showed a near complete identity to its *E. coli* homologue. The RNAhybrid algorithm was used to predict potential RNA duplexes formed between S39006 RprA and *rpoS* mRNA and identified two short pairing regions in RprA which overlapped but differed with the pairing of *E. coli* RprA (Figure 5.6C) (Rehmsmeier et al., 2004). However, no homologue for *dsrA* was found after a genome comparison and after scanning the S39006 genome for conserved small RNAs using Rfam (Gardner et al., 2009).

Nevertheless, the complete conservation of the DsrA/RprA interaction site in the S39006 *rpoS* 5' UTR indicated that *E. coli* DsrA could pair with S39006 *rpoS* mRNA and regulate its translation. To test this, the S39006 *rpr*A and the E. coli *dsrA* sequences were cloned into plasmid vectors and expressed *in trans* in S39006 WT. It was shown that RpoS down-regulates pigment production; therefore positive regulation of *rpoS* mRNA translation by DsrA/RprA should result in a suppression of pigment production. This was confirmed by the expression of DsrA/RprA *in trans* in the WT which resulted in decreased pigment levels (Figure 5.7). Moreover, pigment production was almost abolished by inducing an overexpression of the small RNAs from the plasmid vectors. However growth of the WT was only marginally decreased by overexpression of the small RNAs whereas overexpression of the *rpoS* CDS from the same plasmid vector caused a 10-fold reduction in growth (data not shown). This is explained by the current model that the increased translation of *rpoS* by the small RNAs is limited by the fixed amount of the *rpoS* message naturally expressed in the cell, whereas expression of *rpoS*

directly from the plasmid vector results in a much greater translation of the protein which affects the growth of the cell.

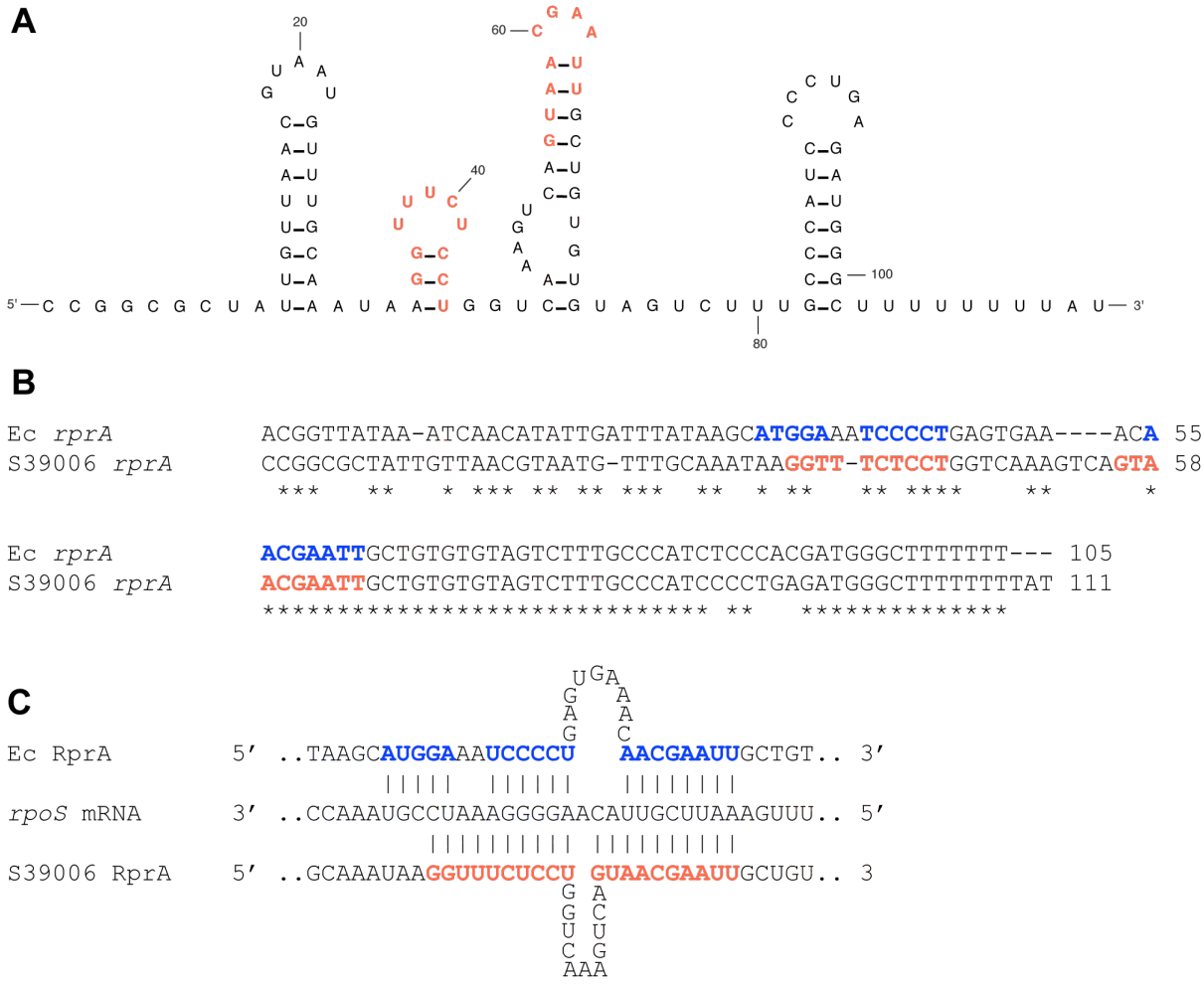


Figure 5.6. Structure, alignment, and pairing of RprA.

A. The most energetically favoured structure of S39006 RprA predicted by MFOLD is shown (Zuker, 2003).

B. An alignment of the *E. coli* and S39006 RprA DNA sequences was produced using ClustalW.

C. Shown is a comparison of the pairing of the *E. coli* and S39006 RprA sRNAs to the section of the *rpoS* mRNA 5' UTR conserved in both E. coli and S39006. The region of the *rpoS* mRNA involved in sRNA pairing is from position -111 to -85 relative to the *rpoS* translation start. *E. coli* RprA-*rpoS* mRNA pairing is shown as described previously (Majdalani et al., 2002), and the S39006 RprA-*rpoS* mRNA pairing is based on the prediction by the RNAhybrid algorithm. *E. coli* RprA forms a duplex of 19 bp while S39006 RprA is predicted to form a duplex of 20 bp differing from the *E. coli* pairing by 7 bp. S39006 and *E. coli* RprA nucleotides involved in pairing with *rpoS* are shown in red and blue, respectively.

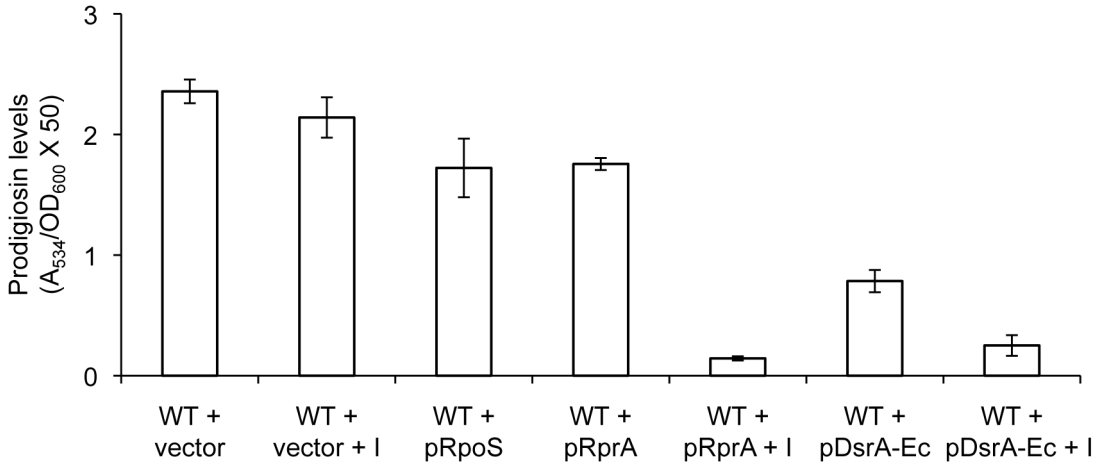


Figure 5.7. RprA and DsrA sRNAs repress pigment production.

Prodigiosin production was measured after 12 hours of growth of the WT carrying different plasmid constructs expressing RpoS (pNB43), RprA (pNB49), or *E. coli* DsrA (pNB50). IPTG (+ I) was added at the start of growth to induce an overexpression of the DsrA and RprA.

5.2.6 The *rpoS* mutant is attenuated for virulence in animal, but not plant, hosts

RpoS has been shown to regulate virulence factors in *E. coli*, *S. typhimurium*, and several other species involved in both animal and plant pathogenesis (Dong and Schellhorn, 2010), and it was sought to determine if the S39006 *rpoS* mutant was attenuated for pathogenesis using the model hosts, potato and *C. elegans*. The *rpoS* mutant was significantly altered for production of the plant cell wall degrading exoenzymes, decreased for cellulase but increased for pectate lyase activity (Figure 5.8). Nevertheless, the altered production of exoenzymes for the *rpoS* mutant did not lead to an affect on plant pathogenesis since no difference was observed in the level of rot production in potato compared to WT (data not shown).

However, the *rpoS* mutant was attenuated for virulence in the worm, *C. elegans*. The rate of killing by the *rpoS* mutant was significantly reduced compared with that of WT (Figure 5.9). The survival median (representing the time required for 50% of nematodes to die) was 4 days for WT, 8 days for the *rpoS*::Tn strain, 7 days for the partially complemented *rpoS*::Tn strain, 9 days for the *hfq::cat* strain and 9 days for the conventional control strain, *E. coli* OP50 (OP50) (Kurz et al., 2003). Additionally, worms fed on WT were impaired for growth compared with worms fed on the *rpoS* mutant which reached the full adult size at the same rate as worms fed on OP50 (data not

shown). Moreover, it was previously shown that *C. elegans* avoided lawns of the pathogenic bacterium, *Serratia marcescens* (Pujol et al., 2001). A similar behavior toward lawns of S39006 WT and the *rpoS* mutant, but not the *hfq* mutant, was observed while tracking worm survival. Moreover, that the *rpoS* mutant is increased for prodigiosin production but attenuated for virulence in *C. elegans* provides additional evidence that prodigiosin does not appear to play a role in pathogenesis in this particular strain.

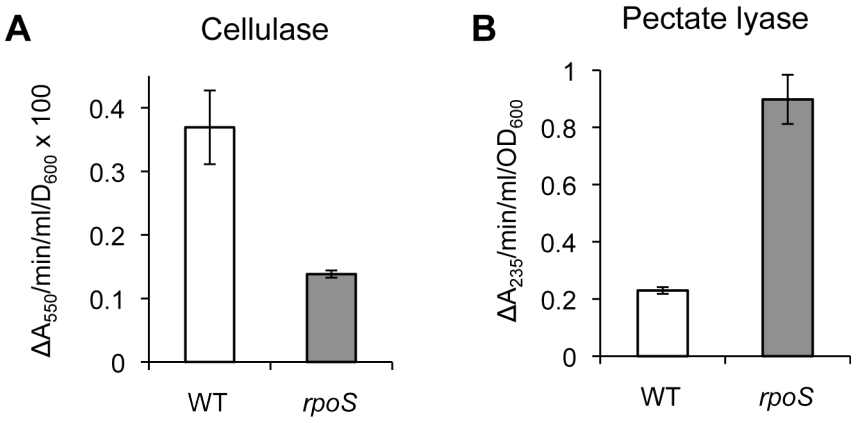


Figure 5.8. The *rpoS* mutant is altered for exoenzyme production.

The amount of cellulase and pectate lyase activity in the culture supernatant of WT and the *rpoS* mutant (NMW25) was measured after 17 h growth at 25°C in PMB Medium. Bars represent the mean ± SE from three independent experiments.

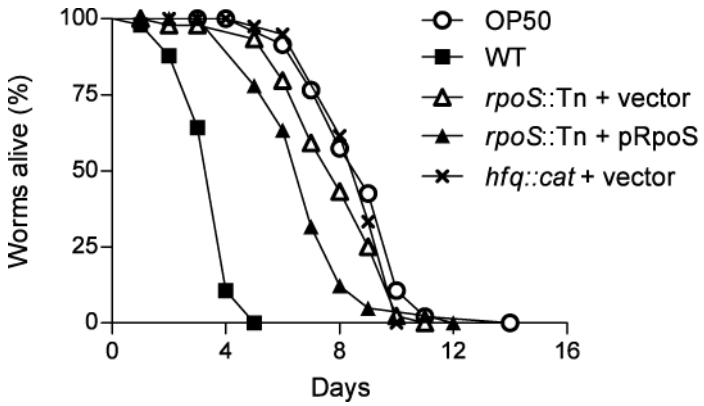


Figure 5.9. The *rpoS* mutant is attenuated for virulence in C. elegans.

Kinetics of killing of *C. elegans* infected by S39006 WT, *rpoS*::Tn, complemented *rpoS*::Tn, and *hfq::cat* are shown. pRpoS refers to pNB43. Included for comparison is the survival of worms fed

on the normal feeder strain *E. coli* OP50. Fifty N2 hermaphrodites were used for each strain and were grown at 25°C using NGM plates.

5.3 Discussion

A growing body of evidence has shown that the RpoS sigma factor plays an important role in regulating stress response and pathogenesis in a variety of bacterial species (Dong and Schellhorn, 2010). In this study, the role of RpoS in the Gram-negative bacterium, *Serratia* 39006, was investigated. In addition to affecting stress tolerance and virulence in animal but not plant hosts, RpoS was found to be a repressor of both swimming and swarming motility and to play a major role in regulating the production of the bioactive secondary metabolites, prodigiosin and carbapenem.

Initial assays showed that the *rpoS* mutant was not affected for growth rate but entered stationary slightly earlier. Unlike the *hfq* mutant, which could grow in nutrient-limiting conditions, the *rpoS* mutant could not be cultured in the same minimal medium which is consistent with the observation that *rpoS* mutants in related bacteria die off rapidly (Lange and Hengge-Aronis, 1994; Mukherjee et al., 1998). It appears that the *rpoS* mutant is auxotrophic for the production of essential amino acids since the addition of casamino acids to the medium allowed growth of the strain. Interestingly, casamino acids depleted for certain vitamins resulted in increased pigmentation indicating that pathways involved in vitamin transport or metabolism influence prodigiosin production. In addition to starvation, the *rpoS* mutant was also affected by osmotic stress and, to a lesser degree, by oxidative stress.

The key phenotype of the S39006 *rpoS* mutant is the increase of prodigiosin and carbapenem production independent of an affect on BHL/HHL levels. Additionally transcription of the *pigA* and *carA* biosynthetic genes was increased, reflecting *pigA-O* and *carA-G* operon expression. Studies from this lab have shown that multiple environmental signals sensed by various genetic factors are involved in regulating secondary metabolism, including quorum sensing, gluconate, and phosphate levels (Thomson et al., 2000; Fineran et al., 2005a; Gristwood et al., 2009). One of these genetic systems involved in regulating prodigiosin production includes the PigQW two-component signal transduction system homologous to the BarA/UvrY system in *E. coli*. PigQW activate the expression of the sRNA, RsmB, which interacts with the RNA binding

protein RsmA and prevents it from binding to and repressing target mRNAs (Williamson et al., 2008). Mutations in *pigQ* or *pigW* results in a significant decrease in pigment levels due to the inactivation of RsmB expression and continued repression by RsmA of its mRNA targets including those involved in prodigiosin production. In *E. coli*, mutants deficient in *barA* or *uvrY* displayed reduced *rpoS* transcription (Hengge, 2008). If this function were conserved for *pigQW* in S39006 then it would be expected that decreased levels of *rpoS* would result in increased pigment levels, contrary to the observed result. Analysis of double mutants of *rpoS* with *pigQ* and *pigW* showed that an *rpoS* mutation could bypass mutations in *pigQ* and *pigW* and restore pigment production, indicating that *rpoS* may be regulated by *pigQ* and *pigW*.

However, the *rpoS* mutation was not able to bypass mutations in the other Pig regulators *pigP*, *rap*, *rsmA* and *hfq*. It was surprising that the *rpoS* mutation did not bypass an *hfq* deletion. Hfq positively regulates the translation of *rpoS* mRNA by the dependent sRNAs, DsrA, RprA, and ArcZ in *E. coli* and *S. typhimurium* (Majdalani et al., 2002; Majdalani et al., 2005; Papenfort et al., 2009). The Hfq-mediated regulation of *rpoS* in S39006 was previously shown by the decreased *rpoS* transcript levels in an *hfq* mutant, and *E. coli* Hfq bound to a section of the S39006 *rpoS* 5' UTR containing the conserved sRNA pairing sites (see Chapter 4). Moreover, a homologue of RprA in S39006 was identified. The small RNA mediated regulation of prodigiosin production via the positive regulation of *rpoS* mRNA translation by S39006 RprA and *E. coli* DsrA was confirmed by expression of these sRNAs *in trans* resulting in decreased pigment levels. Thus, it is surprising that, despite a decrease in *rpoS* transcript levels, the *hfq* mutant showed no prodigioisin and carbapenem production and showed decreased swarming, whereas the *rpoS* mutant was upregulated for both secondary metabolism and swarming. Additionally, expression of *rpoS in trans* in the *hfq* mutant did not restore prodigiosin production (data not shown). This indicates that Hfq-mediated regulation of secondary metabolism is independent of its affect on *rpoS* expression.

It has been shown previously that S39006 kills *C. elegans* by an infection-like process of the intestine and that *C. elegans* exhibits the behavior of avoiding lawns of S39006 (see Chapter 4). The *rpoS* mutant was shown to be attenuated for virulence in *C. elegans* although to a lesser degree than the *hfq* mutant. Moreover, it was observed that

worms continued to avoid lawns of the *rpoS* mutant whereas worms did not avoid lawns of the *hfq* mutant indicating that the factor inducing lawn avoidance is down-regulated in the *hfq* but not *rpoS* mutant. This factor was identified in the related strain, *S. marcescens* Db10, as serrawettin W2, a secreted surfactant encoded by the *swrA* gene predicted to encode a nonribosomal peptide synthase (NRPS) (Pradel et al., 2007). A BLAST search identified a homologue of *swrA* in S39006 preceded by a cyclic peptide transporter suggesting that the product is synthesized and exported from the cell. Also expression of the swarming-dependent *rhlA* gene involved in production of another surfactant has been observed to be elevated in mutants increased for swarming motility such as the *rpoS* mutant. It may additionally be involved in inducing lawn avoidance behaviour in *C. elegans*. Thus, this study demonstrates that the *rpoS* mutant is attenuated for virulence in an animal host despite being increased for prodigiosin production and swarming motility.

However, the *rpoS* mutant was not attenuated for virulence in plants unlike the *hfq* mutant indicating that Hfq regulates plant pathogenesis independent of its affect on *rpoS* expression. Activity of the plant cell wall degrading exoenzymes were differentially regulated in the *rpoS* mutant with activity of cellulase decreased and pectate lyase increased. However, an *rpoS* mutation in other plant pathogens showed different results. An *rpoS* mutant of *Pectobacterium carotovorum* (formerly *Erwinia*) was more virulent for infection in celery and tobacco, but not potato, whereas mutants of the rice and tomota pathogens *Burkholderia plantarii* and *Ralstonia solanacearum* were attenuated for pathogenesis (Dong and Schellhorn, 2010).

In summary, this study demonstrates that RpoS regulates secondary metabolism and animal pathogenesis in S39006. It identifies the Hfq-dependent RprA sRNA and demonstrates its ability to influence prodigiosin production through positive regulation of *rpoS* translation, as suggested by the current model. It also demonstrates that the Hfq-mediated regulation of secondary metabolism and plant pathogenesis is independent of Hfq regulation of *rpoS* translation.

Chapter 6. Systems-wide investigation of the global RNA binding proteins, Hfq and RsmA.

6.1 Introduction

Previously it was shown that the RNA-binding proteins and post-transcriptional regulators, Hfq and RsmA, play a global role in regulating antibiotic production, motility, and pathogenesis (see Chapter 4)(Williamson et al., 2008). An *hfq* deletion resulted in cessation of antibiotic production and a major attenuation of plant and animal pathogenesis. This suggested that Hfq-dependent processes, such as regulation by *trans*-acting small regulatory RNAs, are involved in the regulation of these phenotypes. The identification of sRNAs has rapidly increased in recent years by the use of new high-throughput sequencing technologies. These allow sequencing of a library of cDNAs reverse transcribed from extracted RNA (referred to as RNA-seq) and generate a profile of the whole transcriptome (Sorek and Cossart, 2009). Additionally, RNA-seq can be adapted to analyze a subset of the transcriptome such as Hfq-associated messenger and small RNAs (Sittka et al., 2008).

In this study, sequencing of single-stranded (ss) cDNA and quantitative proteomics was used to generate a profile of the whole transcriptome and proteome of S39006 WT and *hfq* and *rsmA* mutants. A conserved hypothetical protein was identified as being regulated by a candidate riboswitch with an impact on prodigiosin production. Differential expression analysis of the transcriptome and proteome of *hfq* and *rsmA* mutants was used to identify genes and pathways regulated by Hfq and RsmA with some correlation between RNA and protein expression. The results indicate that Hfq and RsmA directly or indirectly regulate 4% and 19% of the genome with some correlation between transcription and protein expression. Pathways affected include those involved in antibiotic regulation, virulence, flagella synthesis, and surfactant production. The combination of transcriptomics and proteomics in this study provides a systems-level understanding of Hfq and RsmA regulation and identifies similarities and differences in the regulons of two major global regulators.

6.2 Results

6.2.1 Sequencing, assembly, and annotation of the S39006 genome

During the course of this study sequencing and assembly of the S39006 genome (~5 Mb) was led by Peter Fineran at the University of Otago, NZ, in a joint Otago/Cambridge project with Josh Ramsay leading efforts in Cambridge. A manuscript describing the work is in preparation for publication. A brief description of its assembly will be mentioned here. The S39006 genome was sequenced using 454 pyrosequencing with an average read length of 200 bp in NZ and with an average read length of 400 bp at the DNA Sequencing Facility in the Department of Biochemistry, Cambridge, UK. A *de novo* assembly of the S39006 genome into about 400 contigs was generated using the program Newbler by P. Fineran. The orientation of contigs was determined by trying to align contigs to the related genome of *Pectobacterium atrosepticum* SCRI1043, and gaps between contigs were closed by amplification by PCR across contig ends and Sanger sequencing.

The method of 454 sequencing is problematic for resolving homopolymeric stretches of DNA and artifactual insertions and deletions are common (MacLean et al., 2009). To correct for indels and single-nucleotide polymorphisms (SNPs) in the assembly of the S39006 genome, extracted genomic DNA was subjected to single-end sequencing (36 bp read length) on the Illumina GAIIX platform and mapped to the reference genome generated from the 454 sequencing by the Eastern Sequence and Informatics Hub (EASIH), Cambridge, UK. Bioinformatic analysis by EASIH resulted in correction of 202 indels and 258 SNPs with four regions containing indels that resemble organism variation or assembly error. Mapping of the Illumina reads resulted in over 100× coverage across most of the genome. However 13 very short contigs (<500 bp) from the 454 assembly were identified with no coverage, which is most likely the result of a misassembly, and therefore these contigs were removed from the assembly. The improved genome assembly consisted of 4,945,974 bp spread across 20 contigs with the six largest contigs accounting for 4.943 Mb or 99.9% of the sequence. The remaining 14 contigs were less than 1 kb in size. The contigs were ordered from the longest to the

shortest and concatenated into a single sequence with 50×N added between each contig to delineate contig boundaries.

The incomplete genome was annotated by the computerized Oak Ridge Genome Annotation and Analysis pipeline hosted by Oak Ridge National Laboratory, USA. The S39006 genome has a GC content of 49.2% and 4418 candidate protein-encoding genes. This version of the genome was used for the mapping of RNA-seq reads and for the generation of the protein database used to identify peptides from the proteomic experiments. However, gaps between contigs are still being closed and the genome assembly is almost completed at the time of writing.

6.2.2 Sequencing the transcriptome of the S39006 WT and the *hfq* and *rsmA* mutants

To characterize the transcriptome of the S39006 WT and the *Δhfq* (NMW8) and *rsmA*::Tn (NMW7) mutants, total RNA was isolated from cultures of three replicates per strain grown to early stationary phase (T = 10 hr; or OD = 2.0 for the WT), when antibiotic production is maximal. As a quality control measure, to ensure sufficient removal of genomic DNA and the integrity of the extracted RNA samples, qRT-PCR was performed to measure the relative expression of genes known to be differentially regulated in the *hfq* and *rsmA* mutants (Figure 6.6). It was shown in a previous study that the homologues of *rhlA* and *flhC*, essential for swarming and swimming motility, respectively, were increased in a different S39006 *rsmA* mutant which was also observed for the *rsmA* mutant used in this study (Williamson et al., 2008). Likewise, transcript levels were decreased for genes *pigA* and *pigB* of the *pigA-O* prodigiosin biosynthetic cluster, *carA* of the carbapenem biosynthetic cluster, and the QS *luxR*-type regulators, *smaR* and *carR*, reflecting the decrease in expression observed using transcriptional fusions of these genes in an *hfq* mutant background (see Chapter 4.2.7).

Strand-specific cDNA libraries for sequencing were prepared from the RNA samples using the method of single-stranded cDNA synthesis (Croucher et al., 2009). Ss-cDNAs were synthesized and sent to the Wellcome Trust Sanger Institute (WTSI) for adaptor ligation and paired-end sequencing on the Illumina GAIIX sequencer with a read length of 76 bp. The paired-end reads were mapped to the reference genome using

the MAQ alignment program by Nicholas Croucher (WTSI). Paired reads normally map to opposite strands of the genome. Therefore the reverse complement of the second read in proper pairs, corresponding to the 3' end of the cDNA molecule, was mapped in order that both the first and second reads map to the same strand of the genome in a strand-specific manner.

Table 6.1. Analysis of RNA-seq data mapped to the S39006 genome.

Flowcell ID_lane	4577_1	4577_2	4583_1	4577_3	4577_5	4583_2	4577_6	4577_7	4583_3
Library	WT-1	WT-2	WT-3	hfq-1	hfq-2	hfq-3	rsmA-1	rsmA-2	rsmA-3
Total reads ($\times 10^6$)	61.8	48.1	41.2	53.6	34.1	44.3	47.5	43.2	17.5
Reads mapped ($\times 10^6$)	27.9	26.5	26.9	31.7	20.1	28.1	4.2	7.1	8.5
Total bases ($\times 10^9$)	4.70	3.66	3.13	4.08	2.59	3.36	3.61	3.28	1.33
Bases mapped ($\times 10^9$)	2.12	2.01	2.04	2.41	1.53	2.14	0.32	0.54	0.65
% Read/bases mapped	45%	55%	65%	59%	59%	64%	9%	16%	49%
% Repetitive reads	41%	53%	63%	58%	59%	62%	5%	15%	44%
% Reads in pairs	39%	51%	60%	55%	56%	60%	8%	15%	42%
% Reads in proper pairs	29%	49%	57%	54%	55%	58%	8%	14%	40%

6.2.3 Non-strand specific mapping anomaly of RNA-seq reads

RNA-seq reads mapped in an almost non-strand-specific manner throughout the genome. This includes indiscriminate mapping in non-coding regions suggesting that there was low-level genomic DNA present in the RNA sample that was undetectable by conventional PCR despite the quality control steps (described in methods section 2.7) and qRT-PCR to ensure removal of genomic DNA. However, the low level gDNA contamination was not uniform across the genome. It is possible that the method

employed is highly selective for even the smallest amount of undetectable genomic DNA since adaptor ligation favours double-stranded versus ss-cDNA. Nevertheless, a majority of reads mapped in a strand-specific manner where coding sequences (CDSs) were predicted. Surprisingly, when a filter was applied to hide proper paired reads mapping to the genome in BamView, the remaining non-proper paired reads all mapped in a strand-specific manner to the CDSs and indiscriminate mapping to non-coding regions was nearly eliminated (Figure 6.1A-D). The reason is not clear for proper paired reads mapping in a non-strand specific manner compared to non-proper paired and single reads. Proper paired reads are defined as reads where the first and second reads map to opposite strands of the genome (i.e. opposite orientations before read processing) and map at a distance apart on the genome relative to the cDNA library insert size of 200–250 bp. Non-proper paired reads may consist of reads with an inversion (i.e. reads with the same orientation that map to the same strand on the genome), or are paired reads that map far apart on the genome reflecting an impossibly large insert size. An additional anomaly from the sequencing data was the presence of single reads, which do not have a paired read.

It is known that the method employed of ss-cDNA sequencing contains technical limitations and a comparison of various methods for strand-specific RNA-seq purposefully excluded the method developed by Croucher *et al.* (2009) (Levin et al., 2010). These technical limitations, which were studied by Croucher *et al*, include the generation of chimeric cDNAs due to intramolecular or intermolecular annealing of cDNAs during the steps of adaptor ligation to prepare the cDNA library for sequencing.

Library preparation is a two-step process consisting first of end-repair of fragmented cDNA followed by adaptor ligation (Figure 6.2A-B). During the end repair reaction, duplexed regions, such as those formed by the annealing of cDNA molecules, are subject to 3’-end processing. The 3’-> 5’ exonuclease activity of the Klenow fragment degrades 3' overhangs, and T4 polymerase primes polymerization from the 3' end to fill in 5’ overhangs. It is during this step that chimeric cDNAs are generated by complementary strand synthesis primed by annealed cDNAs to produce a reverse complement of the annealed cDNA’s 5’-end. In accordance with the Illumina paired-end sequencing protocol, forked adaptors were ligated to DNA so that the adaptor sequence

at the 5'-end is distinct from the adaptor sequence at the 3'-end (Figure 6.2C). Therefore, the 5'-end of all paired sequence reads should maintain the same orientation relative to the original RNA transcript while the sequence read of the 3'-end is potentially an artefact of the library preparation. Additionally, artifactual paired reads can also be generated whereby highly abundant transcripts, such as rRNA, can anneal as a function of concentration rather than sequence complementarity to mRNA and prime synthesis of a chimeric mRNA-rRNA cDNA molecule. Thus, while the first pair maps to the coding sequence in the genome the second pair will map to an rRNA operon at a distance considerably greater than the insert size (Croucher et al., 2009).

Application of RNA-seq to ss-cDNA samples from the bacteria *Salmonella bongori* and *Streptococcus pneumoniae* and the yeast *Schizosaccharomyces pombe* showed that 3'-end processing was sufficiently small not to interfere with aligning paired reads to the reference genome sequence (Croucher et al., 2009). For *S. pneumoniae*, 60% of paired-end reads could be mapped in a conventional manner. For about 20% of the paired-reads the first and second read had the same orientation, while the remaining 20% had opposite orientations, as expected, but mapped at a distance considerably larger than the insert size with most of the reverse reads mapping to the highly expressed rRNA operons.

However, it was observed for the S39006 RNA-seq data that the first or forward reads in paired reads did not map in a strand-specific manner which may be due to the low-level genomic DNA contamination in the cDNA library, as observed by the indiscriminate mapping in non-coding regions. Low-level gDNA contamination explains this in some instances but not in others, and although there appears to be gDNA contamination in the RNA-seq reads, it is not uniform across the genome. A good example of this is indicated by the transcriptional profile of the region encoding tmRNA. A high peak of reads was observed on both strands of tmRNA with little to no reads mapping to the left of the peak but with a medium amount of reads to the right of the peak mapping in a relatively strand-specific manner but with a clear signal on the antisense strand of the CDSs (Figure 6.1E). Various filters were applied to try and understand how the reads were mapped. The transcriptional profile of the region was observed after application of the filters to show reads mapped corresponding to only

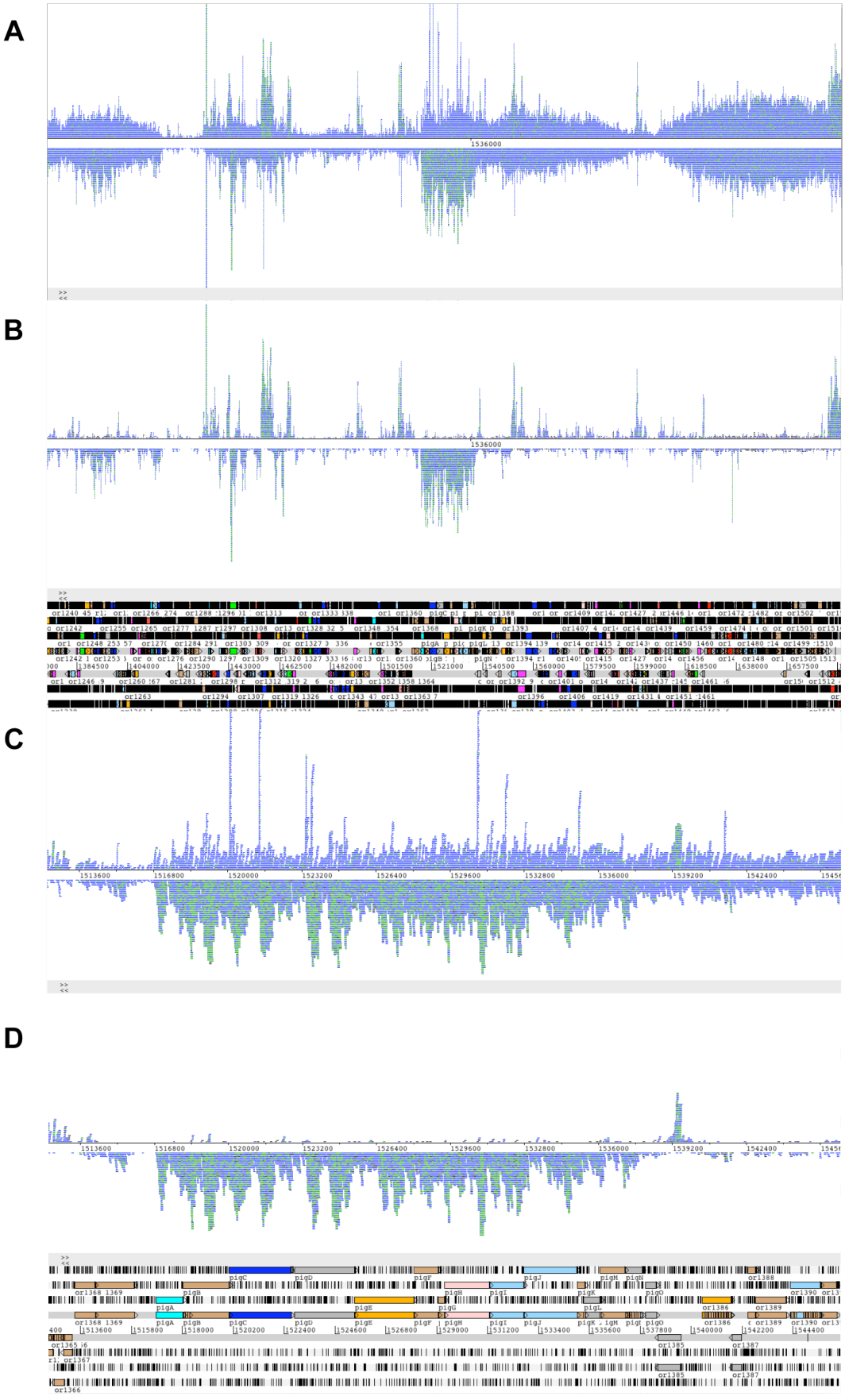
the second reads of paired reads, the first reads of paired reads, or all non-proper paired reads such as those corresponding to inversions or single reads (Figure 6.1F-H).

Since the first reads of paired reads should maintain the same orientation relative to the original RNA molecule, then application of the filter to show only the first reads of pairs should alter the transcriptional profile so that it appears more strand-specific. However, showing only the first or second reads of mapped pairs did not significantly alter the transcriptional profile of tmRNA which still maintained a double peak of reads on opposite strands of the gene. Although, viewing only the first reads of mapped pairs did reduce antisense mapping of reads for the CDSs to the right of the tmRNA peak more than viewing only the second reads of pairs. Nevertheless, only application of the filter to hide all proper paired reads to show only the remaining non-proper paired reads entirely eliminates almost all non-strand-specific mapping including the non-strand-specific peak mapping to tmRNA. This was observed throughout the length of the genome. One speculative explanation for the anomaly of non-strand specific mapping of proper paired reads is that, after synthesis of cDNA using random hexamers, the reactions was purified using Sephadex gel filtration spin columns. The columns exclude most but not all of the hexamers from the purified product (personal communication, GE Healthcare technical support). Therefore, hexamers present during the steps of library preparation will be able to randomly prime synthesis of the complementary strand of cDNA molecules preventing the ability to discriminate the original orientation of cDNA molecules.

Nevertheless, by applying the filter to hide proper paired reads it was still possible to identify some highly abundant known non-coding RNAs, candidate *cis*-regulatory RNA elements, and a *cis*-antisense RNA (See Chapter 3.2.7). However, the observed technical limitations of ss-cDNA sequencing limited the ability to identify novel non-coding RNAs and to precisely map transcription start sites at a single-nucleotide resolution due to reduced sequencing depth after application of the proper pair filter. The identification of ncRNAs and the determination of TSSs are possible using the approach of differential RNA-seq (dRNA-seq) developed by the Vogel group which enriches for the 5'-ends of primary transcripts by depleting processed transcripts such as mature rRNA and tRNA using a terminator exonuclease that degrades 5'

monosphate but not 5' triphosphate RNA (Sharma et al., 2010). High cDNA coverage clusters at the 5'-end of RNA transcripts resulting in a transcriptional profile with an elevated sharp 5' flank at a single-nucleotide level to reveal the TSS, whereas other RNA-seq studies have not been able to unequivocally assign TSSs. Similar to the latter studies, the S39006 RNA-seq data revealed the relative site of transcription initiation and operon structure but reads did not align to the genome with a clear 5' flank at the TSSs of *pigA*, *carA*, *smaI*, *smaR*, and *carR* which have been determined previously by primer extension (Slater et al., 2003; Fineran et al., 2005b).

Some of the reads mapped to intergenic sequences which were predicted to code for known non-coding RNAs such as tRNAs, RNaseP, and tmRNA. Nevertheless, despite the limited information obtained about TSSs and new non-coding RNAs, accurate differential expression analysis between the WT and the *hfq* and *rsmA* mutant strains was still possible.


A
B
C
D

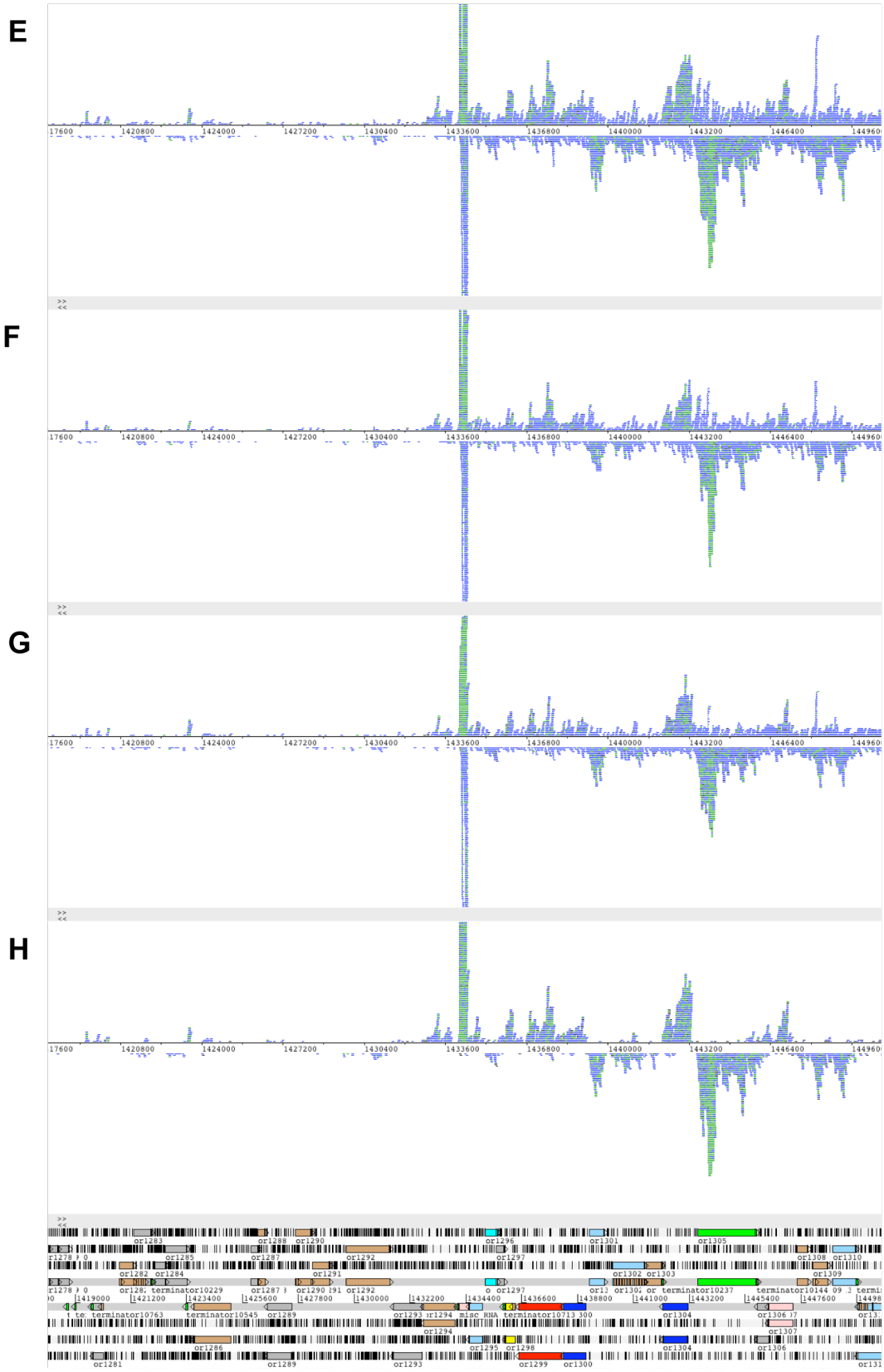

E
F
G
H

Figure 6.1. RNA-seq data displayed in BamView and Artemis.

RNA-seq reads from library WT-1 are shown mapped to the S39006 genome. The total coverage is shown as raw reads aligned against the two strands of the reference genome sequence in BAM strand stack view in the Artemis genome browser with the annotation displayed underneath. Reads mapped to the top strand in BamView correspond to the bottom strand of the annotation in Artemis and vice versa.

A-B. A zoomed out view of a 300 kb region of the genome centred around the *pigA-O* cluster comparing the transcriptional profile of total reads mapped in (A) with the reads mapped after application of the proper pair filter in (B).

C-D. A zoomed in view of the *pigA-O* cluster showing the same comparison of total reads mapped in (C) with reads mapped after application of the filter (D).

E-F. A high peak of reads is observed on both strands of the tmRNA gene (363 bp; centre of the screenshots) when total reads are mapped (E) compared to only one peak of reads on the coding strand of tmRNA after application of the proper pair filter. The filter hides proper paired reads so that only non-proper paired reads and single reads are shown (H). For comparison, screenshots are shown of reads mapped after application of the filter to hide the first reads of pairs so that only the second reads are shown (F) and vice versa, filtering second reads so that only the first reads of pairs are shown mapped (G).

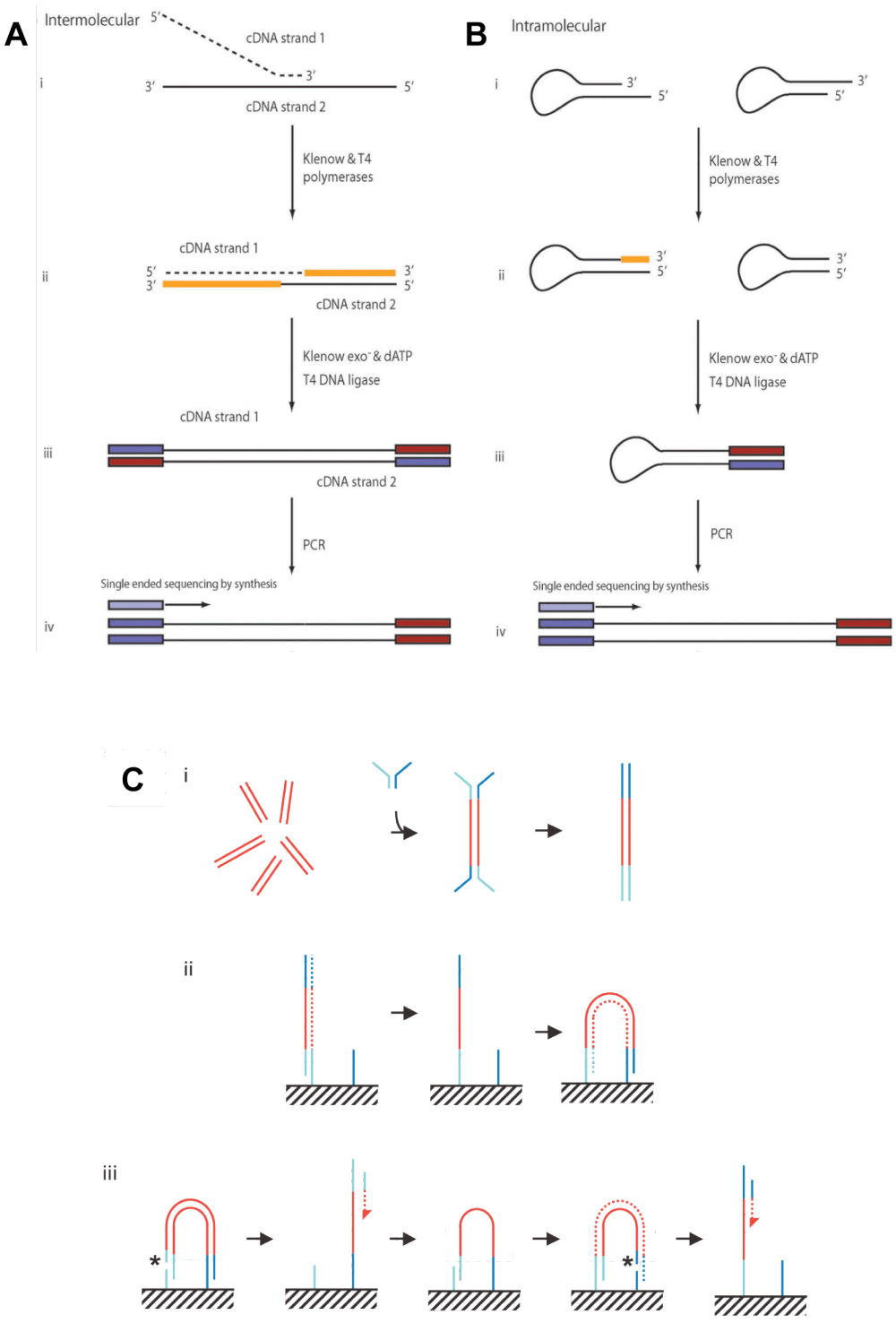


Figure 6.2. Illustration of ss-cDNA library preparation and paired end sequencing.

A-B. The schematic illustrates the ss-cDNA library preparation consisting of the end repair reaction using Klenow and T4 polymerases, followed by addition of a single 3' dATP using Klenow exo⁻ and ligation of sequencing adaptors by T4 DNA ligase. Intermolecular (A) or intramolecular (B) annealing of cDNAs formed duplexes that can prime second strand synthesis.

Forked adaptors are ligated to the double-stranded DNA and the 5' and 3'-adaptor sequences are indicated by blue and red, respectively. The resulting library can be sequenced by single-end or paired-end sequencing. This figure is adapted from Croucher *et al.* (2009).
C. An illustration of sequencing by synthesis for paired-end DNA on the Illumina GA sequencer. (i) Forked adaptors are ligated to blunt-end DNA to mark the 5' and 3'-ends. (ii) The DNA fragments hybridize to oligos on the flowcell surface. The surface bound oligos prime synthesis of a new strand; the original strand is then removed by denaturation. The adaptor sequence on the 3' end of the single-stranded DNA fragment is annealed to new surface-bound oligos forming a bridge which serves as a new site for synthesis of a second strand. Multiple cycles of annealing, extension and denaturation in a process referred to as bridge amplification results in double-stranded DNA and generates a polony, i.e. a clonal amplification of a single DNA molecule. (iii) The DNA in each polony is linearized by cleavage within one adaptor sequence and denatured to generate single-stranded template. The sequencing primer is added and the first read is sequenced by synthesis. For paired read sequencing, the products of the first read are removed by denaturation and the remaining template strand is amplified to re-synthesize its complementary strand. The DNA is linearized in this second round by cleavage of the second adaptor sequence. The complementary strand now provides the template for sequencing of the second read. This figure is adapted from Bentley *et al.* (2008).

6.2.4 Identification of a *cis*-acting regulatory RNA element

While observing the transcriptional profile of the *pigA-O* cluster it was noted that a peak of reads mapped at the end of the cluster on the opposite strand in the region around the translation start of a conserved hypothetical protein (Figure 6.1D). A closer examination of this CDS (or1385; 951 bp; 316 aa) showed that the reads mapped predominantly to a region from -60 to +250 bp relative to the translation start in the replicates of WT and both the *hfq* and *rsmA* mutants (Figure 6.3A-B). The transcriptional profile suggests that the peak of reads may indicate the presence of a candidate *cis*-regulatory RNA element encoded within the mRNA of the CDS that regulates transcription elongation. The high abundance of reads in the UTR and 5' region of or1385 followed by a significant and distinct drop in reads indicates that there is a high level of transcription initiation followed by a high rate of premature transcription termination about a quarter of the way into the CDS.

In order to investigate the function and regulation of the CDS, the region where reads mapped in high abundance (-154 bp to +260 bp), the ORF (+1 to +951) and the ORF preceded by the 5' UTR (-154 bp to +951 bp) were cloned into plasmid vectors and expressed *in trans* in the WT to determine if there was an impact on prodigiosin

production. Interestingly, pigment levels were increased when the CDS was expressed on its own whereas there was no effect when the CDS was expressed with the 5' UTR or when the region spanning the peak of reads was expressed on its own (Figure 6.3C). Moreover, when overexpression of the vectors was induced by the addition of IPTG, only the vector expressing the CDS on its own caused a six-fold decrease in growth whereas there was no effect for the vector with the ORF preceded by the 5'UTR. Therefore, the result may indicate the presence of a candidate novel *cis*-acting regulatory RNA element encoded within the 5' UTR that prevents transcription elongation of the RNA transcript and expression of the protein.

The protein of or1385 is predicted to contain two domains with the second C-terminal domain identified as conserved for the domain of unknown function PF04016 (DUF364) Pfam family (Figure 6.4). Moreover, expression of the protein was confirmed by peptides mapping to or1385 detected by mass spectrometry in the global quantitative proteomic experiment carried out as part of this study (see section 6.2.5). The Dhaf4260 protein from *Desulfitobacterium hafniense* DCB-2 is the only solved protein structure of the DUF364 family, homologues of which are present in proteobacteria, firmicutes, actinobacteria, cyanobacteria, thermotogae and archaea (Miller et al., 2010). The crystal structure of Dhaf4260 revealed that DUF364 is itself a novel combination of two well-known domains, an enolase N-terminal-like fold followed by a Rossmann-like domain, which form a unique catalytic site at the domain interface. An analysis of the genetic context and interdomain cleft suggests that DUF364 enzymes are involved in heavy metal transport and facilitate the synthesis of flavins, pterins, or similar compounds that chelate heavy metals (Miller et al., 2010). Therefore it may be possible that the substrate of or1385 or another relevant metabolite is the cue detected by the candidate *cis*-regulatory element to activate or repress transcription elongation.

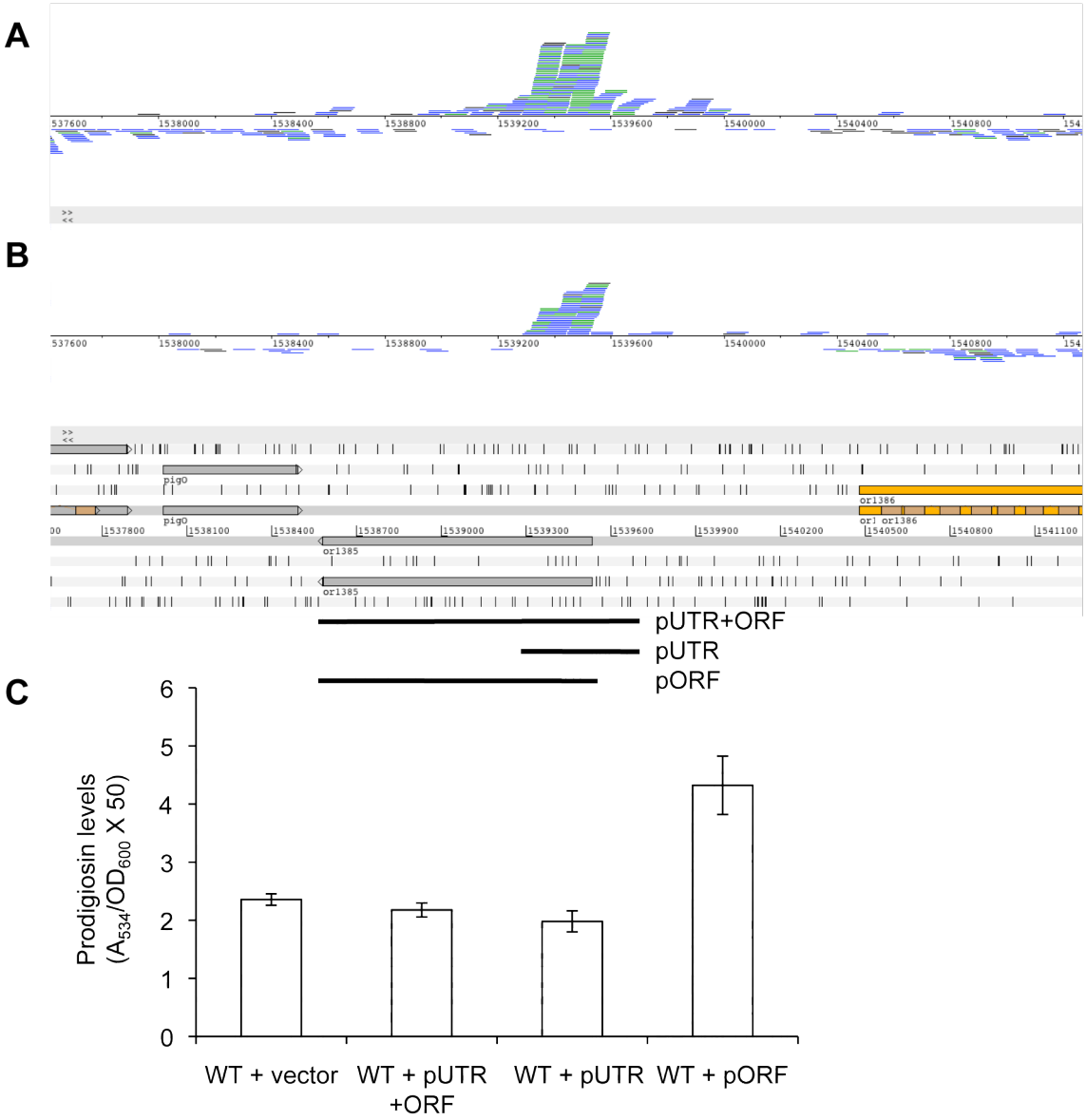


Figure 6.3. RNA-seq mapping suggests a candidate *cis*-acting regulatory RNA element.

A-B. RNA-seq reads after application of the proper pair filter are shown mapped to the conserved hypothetical CDS (or1385) on the reverse strand at the end of the *pigA-O* cluster. The cumulative coverage is shown for the three replicates of the WT (A) and the *hfq* mutant (B). A peak was also observed in the replicates of *rsmA* (data not shown). The coverage is shown as raw reads aligned against the two strands of the reference genome sequence in BAM strand stack view in the Artemis genome browser with the annotation displayed underneath. Reads mapped to the top strand in BamView correspond to the bottom strand of the annotation in Artemis and vice versa.

C. Various regions of or1385 were cloned in the same manner into the vector pQE80oriT under the T5 promoter and expressed *in trans* in the WT to determine an effect on pigment levels.

pNB46 contains the ORF preceded by the 5' UTR (-154 bp to +951 bp relative to the translation start codon); pNB47 contains the region where the peak of reads mapped in high abundance (-154 bp to +260 bp); and pNB48 contains only the ORF (+1 to +951).

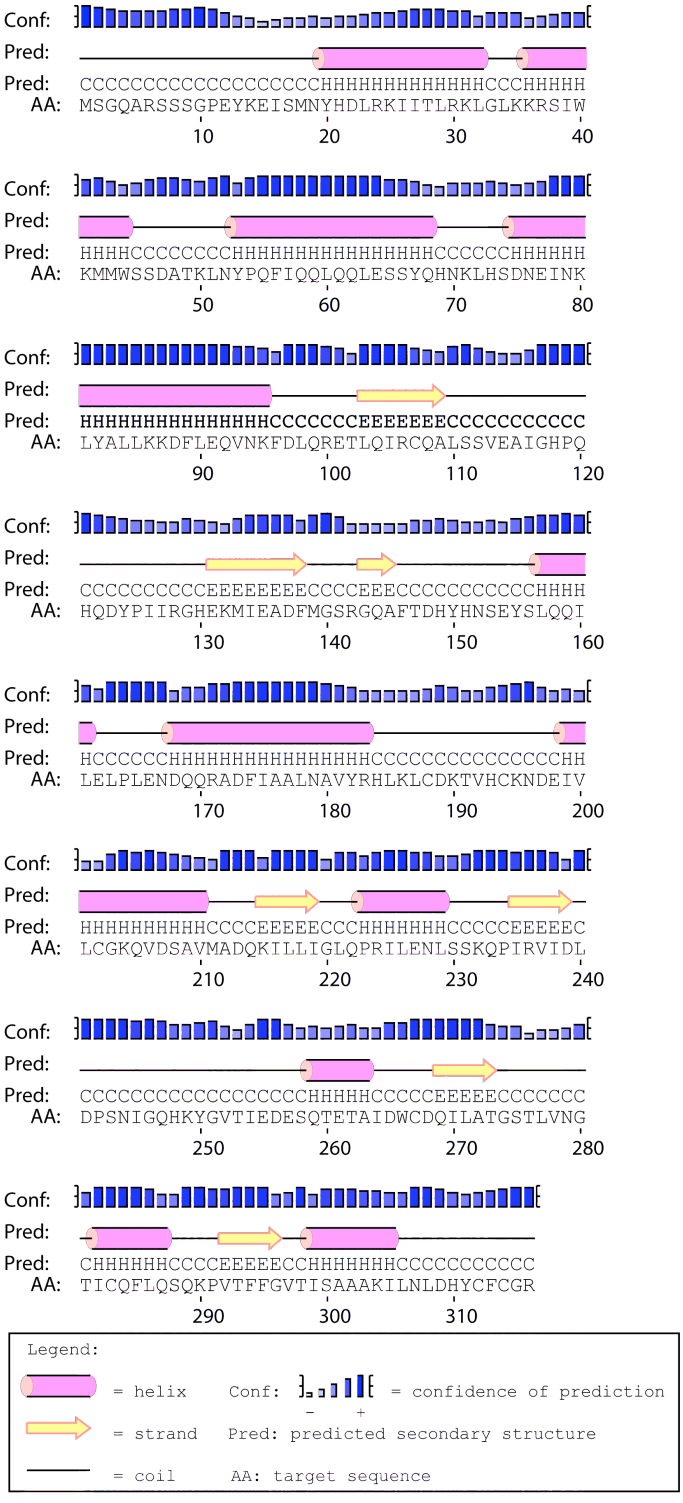


Figure 6.4. Secondary structure prediction of or1385.

The secondary structure of or1385 (316 aa) was predicted by PSIPRED and the graphical diagram created by DomPred. Two domains were predicted with a domain boundary at aa 88, and aa 140-309 was conserved for DUF364.

6.2.5 Differential expression analysis of the *hfq* and *rsmA* mutants

6.2.5.1 Transcript levels are comparable between RNA-seq and qRT-PCR

Digital gene expression was quantified from the RNA-seq data and analysis of differential expression between the WT and the mutants of *hfq* and *rsmA* was performed using the DEseq package in R which models count data using a negative binomial distribution (Anders and Huber, 2010). Analysis using DEseq was performed by Adam Reid (WTSI). *P* values were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) correction. At an FDR of 1%, 86 genes (14 decreased; 72 increased) were identified as significantly different in the *hfq* mutant while 504 genes (339 decreased; 65 increased) were identified in the *rsmA* mutant (Figure 6.5A-B). The number of genes identified in the *hfq* mutant was less than expected compared to the 785 genes or 18% of the *Salmonella typhimurium* altered in an *hfq* mutant (Sittka et al., 2008). This is attributable to the increased variance of the replicates used in the DEseq analysis as illustrated by the atypical scatter plot of gene expression for the *hfq* mutant compared to the *rsmA* mutant (Figure 6.5A-B). The results for specific genes correlated with relative transcript levels determined by qRT-PCR confirming that the analysis by DEseq of the RNA-seq data provides accurate quantitative estimates of fold change in the mutant versus the WT (Figure 6.6; Table 6.2). For the *hfq* and *rsmA* mutants 14 and 7 genes, respectively, were examined by qRT-PCR and the comparison with the RNA-seq data had a Pearson correlation coefficient > 0.90 and > 0.85. Of the eight genes determined as differentially expressed in the *hfq* mutant by qRT-PCR, six were corroborated by the RNA-seq analysis, five with an FDR < 5% and one with an FDR < 10%. Of the four genes determined as differentially expressed in the *rsmA* mutant by qRT-PCR, three were corroborated by the RNA-seq analysis with an FDR < 0.001%. The *pigA* gene was differentially expressed in the RNA-seq analysis but not by qRT-PCR in the *rsmA* mutant due to increased variation in expression in the WT. However the *pigB* gene was identified as differentially expressed according to both methods and since it is operonic with *pigA* then it confirms that *pigA* transcript levels are also reduced. As a result of the comparison of the differential analysis by qRT-PCR and RNA-seq and because of the increased variation in the *hfq*

RNA-seq data, an FDR threshold of 10% and 5% was used for the *hfq* and *rsmA* mutants, respectively. At an FDR threshold of 10%, 178 genes (30 decreased, 148 increased) are differentially expressed in the *hfq* mutant, which represents 4.0% of the predicted 4418 ORFs in the genome. And at an FDR threshold of 5%, 846 genes (523 decreased, 323 increased) are differentially expressed in the *rsmA* mutant, which represents 19.1% of the genome.

Table 6.2. Comparison of the fold change (log_2 ratio) between RNA-seq and qRT-PCR.

	hfq/WT				*rsmA*/WT			
Gene	qPCR	*P* val	RNA-seq	adj *P* val[1]	qPCR	*P* val	RNA-seq	adj *P* val[1]
gyrB	0.18	<50%	0.18	<100%	-0.32	<35%	0.12	<80%
pigA	-6.17	<5%	-3.13	<0.001	-0.56	<35%	-0.98	<0.1%
pigB	-5.91	<0.1%	-2.22	<1%	-0.99	<5%	-1.21	<0.001
carA	-8.76	<0.1%	-5.02	<0.001	—	—	—	—
smaI	0.01	<100%	0.18	<100%	-0.38	<50%	0.19	<90%
smaR	-1.42	<5%	0.22	<100%	-1.36	<5%	-0.59	<60%
carR	-5.14	<0.01%	-4.10	<5%	—	—	—	—
luxS	0.22	<70%	-0.06	<100%	—	—	—	—
rpoS	-1.60	<5%	-1.50	<10%	—	—	—	—
ompA	-0.26	<50%	0.74	<85%	—	—	—	—
ompX	0.06	<90%	0.55	<70%	—	—	—	—
rsmA	0.27	<45%	0.29	<100%	—	—	—	—
rhlA	-2.90	<5%	-0.82	<90%	5.23	<0.01%	2.23	<0.001
flhC	1.37	<10%	1.14	<35%	2.29	<0.01%	2.24	<0.001

[1] *P* values were adjusted for multiple testing according to the method of Benjamini-Hochberg.

6.2.5.2 Proteome-transcriptome correlations

The results of the RNA-seq differential analysis were further validated using quantitative proteomics. Protein was extracted from the WT and *hfq* strains grown under the same conditions as those used in the RNA-seq analysis (i.e. early stationary phase), and global changes in protein expression were investigated using an LC-MS/MS approach with isotope tags for relative and absolute quantification (iTRAQ). Peptides mapped to 1,369 proteins in total and 486 proteins met a cut-off of two or more mapped peptides of which 36 (23 decreased; 13 increased) were differentially expressed (*P* value < 0.1). The higher cut-off of *P* value < 0.1 was used due to the increased variance from one of the replicates being an outlier for 75% of the proteins identified. 14 of the 36 proteins were also differentially expressed in the *hfq* RNA-seq dataset (FDR < 0.1) (Figure 6.5C). A comparison of the fold changes (log_2 ratio) showed the direction of change was the same for 13 proteins (8 decreased; 5 increased) which gave a Pearson correlation coefficient > 0.90. Expression of one gene was decreased at the level of protein production but increased for the mRNA transcript, which may indicate a post-transcriptional effect on regulation.

The *rsmA* mutant proteome was also investigated using the same iTRAQ-based approach, but samples used in the analysis were grown to late stationary phase (T = 12 hr) compared to early stationary phase for samples investigated by RNA-seq (T = 10 hr) (N. Williamson and J. Ramsay, unpublished). A comparison of the differentially expressed genes in the RNA-seq (FDR < 0.5) and iTRAQ datasets (FDR < 0.1) identified 83 genes that were differentially expressed at an RNA and protein level (Figure 6.5D). The fold changes of 73 genes were in the same direction at an RNA and protein level (59 decreased; 14 increased), and had a Pearson correlation coefficient > 0.85. For the 10 genes which were regulated in different directions at the level of RNA and protein, 4 were increased for transcript levels but decreased for protein expression, and 6 were decreased for transcript levels but increased for protein expression. The *pigA* and *pigB* genes were among the latter 6 genes and was decreased 0.51 and 0.43-fold for RNA and increased 1.71 and 1.96-fold for protein. *pigB* was also confirmed by qPCR to be decreased for mRNA. Therefore this indicates that altered post-transcriptional regulation of the *pig* operon may explain increased translation of protein despite

decreased RNA levels resulting in greater prodigiosin production and the observed hyperpigmented phenotype of the *rsmA* mutant.

Differentially expressed genes in the RNA-seq and protein iTRAQ datasets for the *hfq* and *rsmA* mutants were compared to identify co-regulated genes (Figure 6.5E). The biosynthetic genes *pigA* and *carABC* were the four genes altered at both an RNA and protein level in both mutants. The combined RNA and protein datasets for each mutant was compared and showed that 134 genes were co-regulated by Hfq and RsmA (Figure 6.5F).

The *pig* and *car* biosynthetic clusters were examined further in the *hfq* and *rsmA* mutants (Table 6.3). For the *rsmA* mutant, additional genes in the *pig* cluster (*pigCDH*) were decreased 0.57 to 0.68-fold at the level of RNA while *pigF* was increased at the level of protein providing further evidence that the increased prodigiosin production of an *rsmA* mutant may be due to altered post-transcriptional regulation. Additionally, another possibility is that substrates of the *pig* enzymes may also be more readily available. However, genes in the *car* cluster were consistently decreased at both an RNA (*carABCDF*; fold change 0.21-0.40) and protein level (*carABC*; fold change 0.46-0.51) in the *rsmA* mutant. However, due to interference from surfactant production, it has not been possible to accurately measure carbapenem production for the *rsmA* mutant by measuring the zone of inhibition in a plate assay, and therefore it is not possible to correlate changes in *car* expression to carbapenem production (N. Williamson; personal communication). For the *hfq* mutant, which is abolished for antibiotic production, the results were straightforward and showed that the *pig* and *car* genes were significantly decreased at both an RNA and protein level. In summary, there is significant consistency for differential expression at the transcriptomic and proteomic levels in the *hfq* and *rsmA* mutants with a greater degree of overlap observed for the *rsmA* mutant.

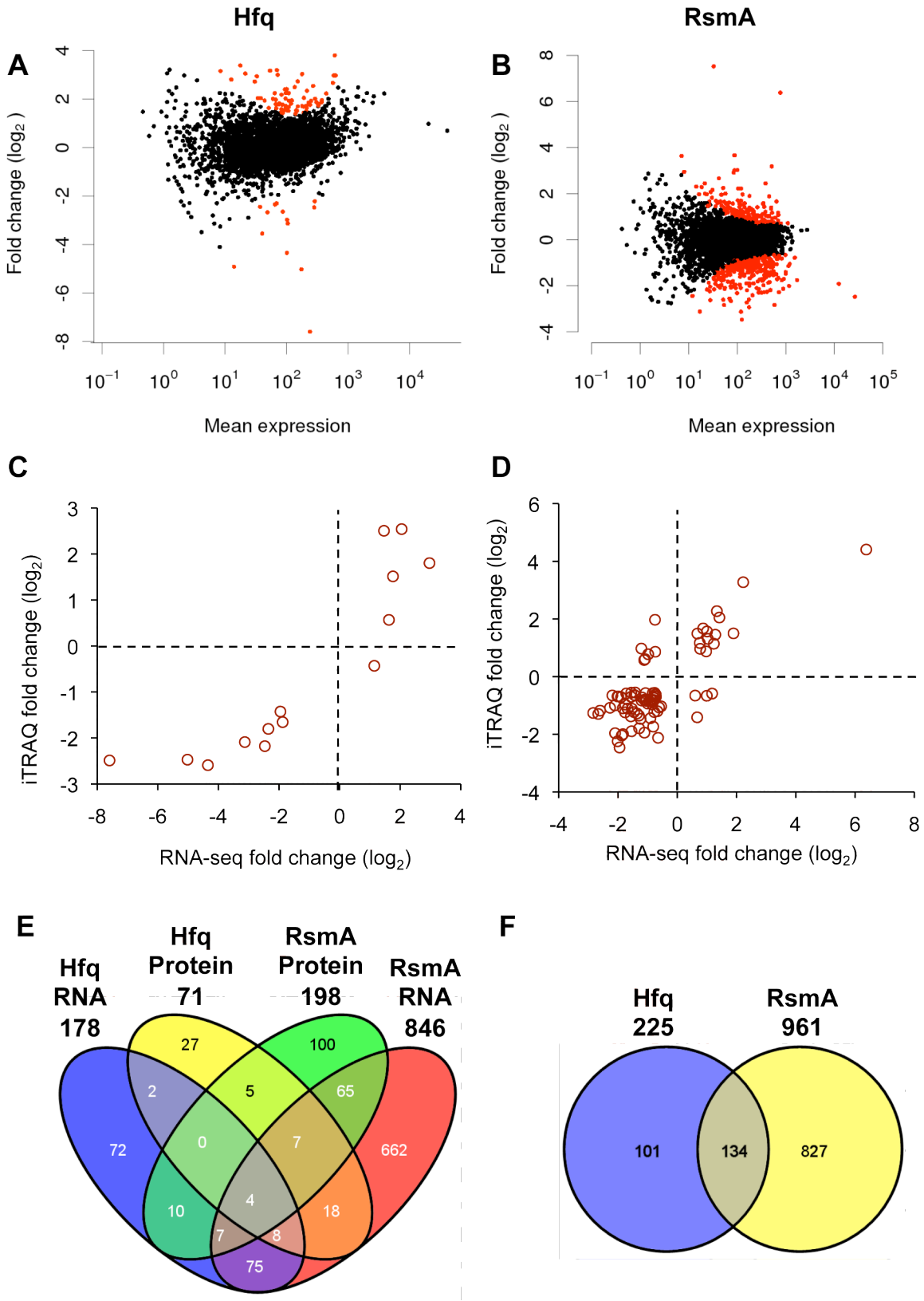


Figure 6.5. Analysis of differential gene expression in the *hfq* and *rsmA* mutants.

A-B. The R package, DEseq, was used to analyze differential gene expression of the RNA-seq data. A scatter plot of the fold change (log_2 ratio) versus mean expression is shown for the *hfq* (A) and *rsmA* (B) mutants. The red circles indicate genes identified as differentially expressed with a 1% false discovery rate according to the method of the Benjamini-Hochberg multiple testing adjustment. Replicates 1 and 2 of WT, 1 and 3 of *hfq*, and 2 and 3 of *rsmA* were used in the analysis. These graphs were prepared by Adam Reid (WTSI).
C-D. The RNA-seq and iTRAQ-proteomic datasets for the *hfq* (C) and *rsmA* (B) mutants were compared by plotting the fold change (log_2 ratio) of differentially expressed genes. 13 and 73 genes had fold changes in the same direction in the *hfq* and *rsmA* datasets, respectively, which gave a Pearson correlation coefficient > 0.90 and >0.85.
E-F. The Venn diagrams illustrate the overlap between the RNA-seq and protein iTRAQ datasets for the *hfq* and *rsmA* mutants to identify co-regulated genes (E). The uniquely regulated genes in the RNA and protein datasets for each mutant was combined to produce a list of all the genes that were significantly differentially expressed. The combined datasets for each mutant was also compared to identify the overlap between them (F).

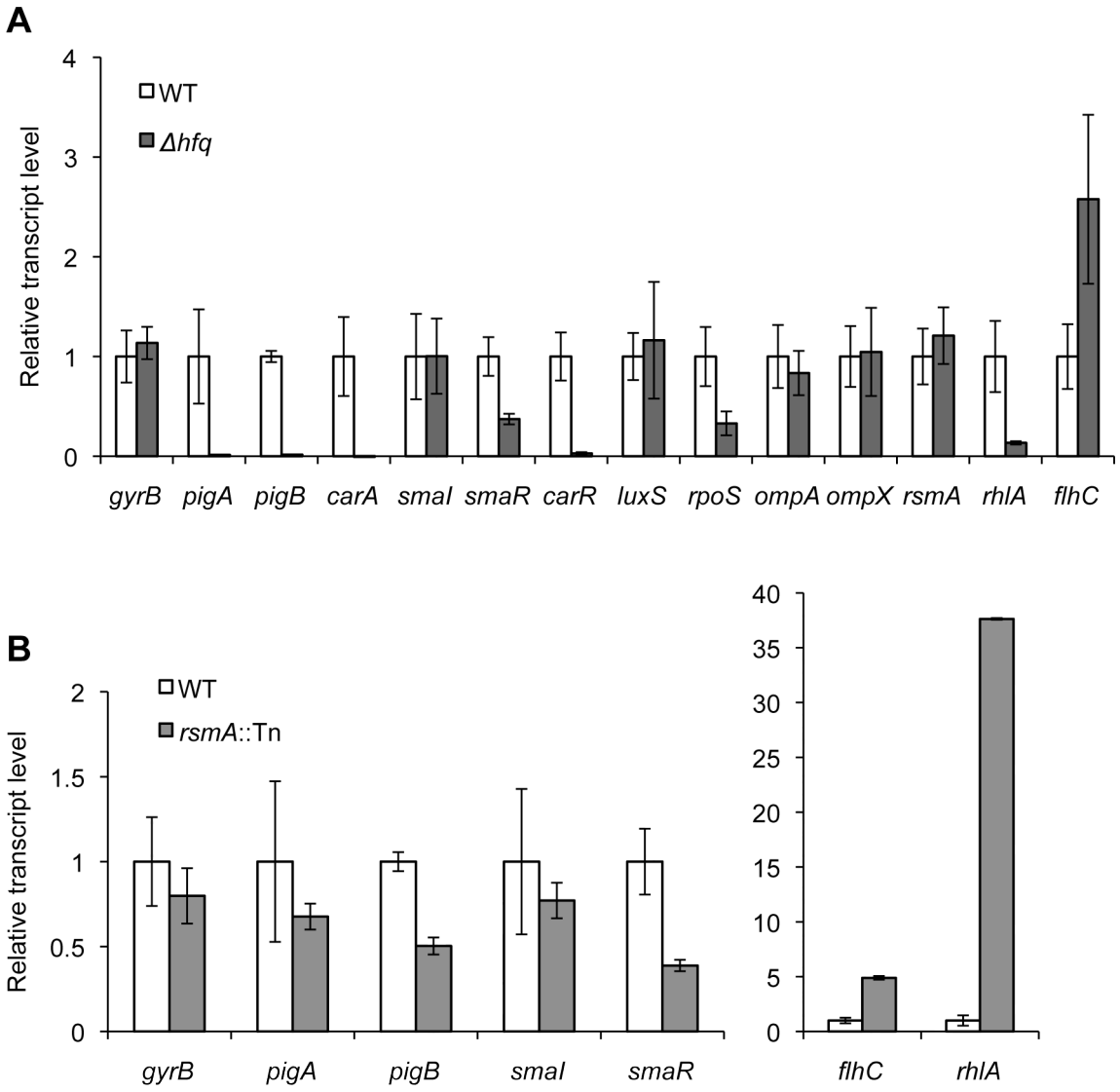


Figure 6.6. qRT-PCR showing differential gene expression.

Gene expression for *Δhfq* (A) and *rsmA*::Tn (B) is measured as transcript levels relative to WT at early stationary phase growth. Values represent average gene expression ± SD from three independent experiments except values for genes *smaI* and *flhC* which were determined from two independent experiments.

Table 6.3. A selection of differentially expressed genes in the *hfq* and *rsmA* mutants.

Gene	*hfq* RNA[1]	*hfq* Protein[1]	*rsmA* RNA[1]	*rsmA* Protein[1]	Description
Pig cluster					
or1370	0.11	0.24	0.51	1.71	PigA;
or1371	0.21	—	0.43	1.96	PigB;
or1372	0.26	0.37	0.57	—	PigC;
or1373	0.27	0.32	0.68	—	PigD;
or1374	—	0.32	—	—	PigE;
or1375	—	0.41	—	1.55	PigF;
or1377	—	—	0.67	—	PigH;
or1385	—	—	0.56	—	Uncharacterized protein next to the end of the *pig* cluster;
Car cluster					
or2790	0.06	—	—	—	CarR;
or2791	0.03	0.18	0.30	0.51	CarA;
or2792	0.05	0.17	0.31	0.46	CarB;
or2793	0.01	0.18	0.21	0.47	CarC;
or2794	0.16	—	0.40	—	CarD;
or2796	0.13	—	0.22	—	CarF;
Additional secondary metabolism genes					
or1962	—	—	2.31	—	Non-ribosomal peptide synthase;
or1963	—	—	2.63	—	Polyketide synthase;
or1964	—	—	3.21	—	Non-ribosomal peptide synthase;
or1938	—	—	0.36	—	Putative non-ribosomal peptide synthesis thioesterase;
or3501	—	—	2.30	—	Similar to proteins involved in antibiotic biosynthesis;
Regulators of prodigiosin production					
or0460	—	—	0.53	—	PhoU protein;
or0464	—	—	0.65	0.48	PstS protein;
or2894	3.39	2.85	—	—	PigU, Transcriptional regulator, LysR family;
or3257	—	—	0.34	0.68	Rap, Transcriptional regulator slyA;
or0694	—	—	0.49	—	VfmE, Transcriptional regulator, AraC family;
Protein export					
or0052	—	—	0.59	—	Preprotein translocase subunit secY;
or0773	2.49	—	—	—	Preprotein translocase, SecE subunit;

Gene	*hfq* RNA[1]	*hfq* Protein[1]	*rsmA* RNA[1]	*rsmA* Protein[1]	Description
Type II secretion system					
or2809	—	—	1.92	—	OutD;
or2810	—	—	2.08	—	OutE;
or2813	—	—	2.43	—	OutH;
or2814	—	—	2.99	—	OutI;
or3742	—	—	2.51	—	Putative major pilin subunit;
or3743	—	—	3.15	—	Type II secretion system protein E;
or3744	—	—	1.81	—	Type II secretion system protein;
Type IV pilus system					
or0866	—	—	1.57	—	Type IV pilus secretin PilQ;
or1482	—	—	1.63	—	PilM, type IV pilus gene cluster
or1484	—	—	1.65	—	PilO, type IV pilus gene cluster
or1485	—	—	1.47	—	PilP, type IV pilus gene cluster
or1486	—	—	1.44	—	PilQ, type IV pilus gene cluster
or1491	—	—	1.52	—	PilV, bacterial shufflon protein, type IV pilus gene cluster
Virulence factors					
or0146	—	—	0.49	—	Pectate lyase/Amb allergen;
or0147	0.37	—	0.43	—	Pectate lyase/Amb allergen;
or2372	—	—	0.71	—	Pectate lyase;
or3495	—	—	4.67	—	Pectate lyase L;
or3494	—	—	0.36	—	Cellulase Z;
or3490	—	2.25	—	—	Putative virulence effector protein SrfB;
or3491	3.46	—	—	—	Putative virulence effector protein SrfA;
or1355	—	—	183.69	—	Putative thermostable hemolysin;
Sigma factors					
or0342	—	—	0.49	—	RpoH, heat shock sigma factor;
or0881	—	—	0.66	—	RpoN, nitrogen-limitation sigma factor;
or1081	—	—	0.44	—	RpoD, housekeeping sigma factor;
or3124	—	—	3.13	—	RpoF, flagellar sigma factor;
or4142	0.35	—	0.34	0.27	RpoS, starvation/stationary phase sigma factor;
Stress response					
or0679	—	—	0.42	—	Superoxide dismutase;
or3439	2.41	—	0.65	—	Tellurite resistance protein;
or3440	—	—	0.61	—	Tellurium resistance protein TerA;
or3442	—	—	0.49	—	Integral membrane protein TerC;
or3443	—	—	0.54	—	Tellurium resistance protein terD;

Gene	*hfq* RNA[1]	*hfq* Protein[1]	*rsmA* RNA[1]	*rsmA* Protein[1]	Description
Outer membrane proteins					
or1614	—	—	0.49	0.57	Outer membrane chaperone Skp (OmpH);
or2178	—	—	0.44	—	Virulence-related outer membrane protein;
or2708	—	—	0.38	—	OmpA domain protein transmembrane region-containing protein;
or2771	—	—	12.38	—	Polysaccharide biosynthesis protein;
or3235	—	—	4.94	—	Porin LamB type;
or3258	—	—	0.38	—	Outer membrane lipoprotein Pcp;
or3378	—	—	0.66	—	Filamentous hemagglutinin family outer membrane protein;
or3572	—	—	0.51	—	Outer membrane protein W;
or4361	2.64	—	—	—	Outer membrane assembly lipoprotein YfiO;
Efflux pumps					
or1356	—	—	7.52	—	Drug resistance transporter, Bcr/CflA subfamily;
or1931	—	—	1.60	—	Probable multidrug efflux system transmembrane protein;
or1939	0.16	—	0.11	—	Putative Drug/metabolite exporter family;
or1961	—	—	2.24	—	Inner membrane component of tripartite multidrug resistance system;
or2954	—	—	2.05	—	Drug resistance transporter, Bcr/CflA subfamily;
or2956	—	—	1.69	—	Efflux transporter, RND family, MFP subunit;
or2957	—	—	1.49	—	Transporter, hydrophobe/amphiphile efflux-1 (HAE1) family;
or2958	—	—	1.69	—	RND efflux system, outer membrane lipoprotein, NodT family;
or4387	—	—	0.53	—	Outer membrane efflux protein;
Electron transport					
or2895	3.53	—	—	—	NADH-quinone oxidoreductase subunit A;
or2896	3.23	—	—	—	NADH-quinone oxidoreductase, B subunit;

Gene	*hfq* RNA[1]	*hfq* Protein[1]	*rsmA* RNA[1]	*rsmA* Protein[1]	Description
or2897	—	—	0.56	—	NADH-quinone oxidoreductase subunit C/D;
or2898	2.20	—	—	—	NADH-quinone oxidoreductase, E subunit;
or2899	2.38	—	—	—	NADH-quinone oxidoreductase, F subunit;
or2901	2.90	—	0.62	—	NADH dehydrogenase I subunit H;
or2903	—	—	0.44	—	NADH dehydrogenase (Quinone);
or2904	2.98	—	—	—	NADH-quinone oxidoreductase subunit K;
or2906	3.05	—	—	—	Proton-translocating NADH-quinone oxidoreductase, chain M;
or2907	—	—	0.30	—	Proton-translocating NADH-quinone oxidoreductase, chain N;
Flagellar and chemotaxis motility locus					
or3069	—	1.63	0.36	—	Methyl-accepting chemotaxis sensory transducer;
or3070	4.63	—	5.43	—	Flagellar transcriptional activator;
or3071	—	—	4.73	—	Flagellar transcriptional activator FlhC;
or3074	3.40	—	—	1.90	CheA signal transduction histidine kinase;
or3075	4.10	—	—	2.05	CheW protein;
or3076	3.83	—	2.35	2.20	Methyl-accepting chemotaxis sensory transducer;
or3077	4.52	—	—	1.78	MCP methyltransferase, CheR-type;
or3078	2.47	—	—	—	Response regulator receiver modulated CheB methylesterase;
or3079	3.93	—	—	—	Chemotaxis protein;
or3080	2.78	—	1.89	—	Chemotaxis phosphatase, CheZ;
or3082	3.11	—	1.86	—	FlhA protein;
or3083	4.27	—	—	—	FlhE protein;
or3084	2.87	—	—	—	FlgN family protein;
or3087	7.98	—	4.00	—	Flagellar basal-body rod protein;
or3088	9.14	—	2.99	—	Flagellar basal-body rod protein FlgC;
or3089	8.15	—	2.94	—	Flagellar hook capping protein;
or3090	6.34	—	1.70	—	Fagellar hook-basal body protein;
or3091	6.64	—	2.38	—	Flagellar basal-body rod protein FlgF;
or3092	5.32	—	2.35	—	Flagellar basal-body rod protein FlgG;
or3093	5.27	—	1.87	—	Flagellar L-ring protein;
or3094	4.77	—	2.33	—	Flagellar P-ring protein;

Gene	*hfq* RNA[1]	*hfq* Protein[1]	*rsmA* RNA[1]	*rsmA* Protein[1]	Description
or3095	4.73	—	1.88	—	Flagellar rod assembly protein/muramidase FlgJ;
or3096	2.55	—	1.62	—	Flagellar hook-associated protein FlgK;
or3097	3.85	—	1.72	—	Flagellar hook-associated protein 3;
or3103	3.40	—	—	—	Flagellar motor switch protein FliM;
or3104	3.10	—	—	—	Flagellar basal body-associated protein FliL;
or3105	4.69	—	—	—	Flagellar hook-length control protein;
or3107	2.67	—	2.03	—	Flagellar protein export ATPase FliI;
or3108	4.73	—	2.44	—	Flagellar assembly protein FliH/Type III secretion system HrpE;
or3109	—	—	2.37	—	Flagellar motor switch protein FliG;
or3110	5.64	—	2.21	—	Flagellar M-ring protein FliF;
or3111	8.26	—	3.93	—	Flagellar hook-basal body complex protein fliE;
or3113	—	—	2.73	—	Flagellar protein;
or3114	3.65	—	1.62	—	Flagellar hook-associated protein 2;
or3115	—	—	—	2.80	Flagellin;
or3125	3.38	—	4.46	—	Protein FliZ;
Putative prophage 1					
or2489	2.45	—	—	—	Tail E family protein;
or2491	—	—	—	1.88	Tail sheath protein;
or2503	3.43	—	—	—	Tail assembly chaperone gp38;
or2504	3.27	—	—	—	Tail Collar domain protein;
or2521	—	—	2.21	—	CI repressor;
or2553	—	—	0.46	0.57	Integration host factor subunit alpha;
Putative prophage 2					
or3624	—	—	1.87	—	Putative uncharacterized protein;
or3626	—	—	1.65	—	Putative uncharacterized protein;
or3638	—	—	1.58	—	Putative uncharacterized protein;
or3639	—	—	1.68	—	Putative uncharacterized protein;
or3641	—	—	1.71	—	Putative uncharacterized protein;
or3646	—	—	1.56	—	Putative uncharacterized protein;
or3648	—	—	1.47	—	Phage-associated protein, HI1409 family;
or3651	—	—	1.79	—	Putative uncharacterized protein;
or3659	—	—	1.88	—	Lysozyme;
or3686	—	—	1.97	1.83	Single-stranded DNA-binding protein;
or3687	—	—	7.68	—	RNA-directed DNA polymerase;

Gene	*hfq* RNA[1]	*hfq* Protein[1]	*rsmA* RNA[1]	*rsmA* Protein[1]	Description
Putative prophage 3					
or4027	—	—	3.10	—	Tail assembly chaperone gp38;
or4028	—	—	3.90	—	Putative tail fiber protein;
or4029	—	—	2.86	—	Putative bacteriophage protein;
or4030	—	—	2.65	—	Putative bacteriophage protein;
or4031	—	—	2.13	—	Putative bacteriophage protein;
or4032	—	—	2.45	—	Putative bacteriophage protein;
or4036	—	—	2.06	—	Putative uncharacterized protein;
or4037	—	—	1.94	—	Putative bacteriophage protein;
or4038	—	—	2.20	—	Putative uncharacterized protein;
or4039	—	—	2.25	—	Lytic transglycosylase, catalytic;
or4040	—	—	2.44	—	Putative uncharacterized protein;
or4041	—	—	2.61	—	Putative bacteriophage protein;
or4042	—	—	2.01	—	Putative uncharacterized protein;
or4043	—	—	2.07	—	Putative bacteriophage protein;
or4044	—	—	2.10	—	Putative uncharacterized protein;
or4045	—	—	1.98	—	Putative bacteriophage protein;
or4046	—	—	1.93	—	Hypothetical phage protein;
or4047	—	—	1.72	—	Putative bacteriophage protein;
or4048	—	—	2.05	—	Putative bacteriophage protein;
or4049	—	—	2.17	—	Putative bacteriophage protein;
or4050	—	—	2.24	—	Putative uncharacterized protein STY2037;
or4051	—	—	2.01	—	Putative bacteriophage protein;
or4052	—	—	2.11	—	Phage-associated protein, family;
or4053	—	—	1.69	—	Putative bacteriophage protein;
or4054	—	—	1.72	—	Putative bacteriophage protein;
or4056	—	—	2.30	—	Predicted CDS Pa_2_6160;
or4057	—	—	1.83	—	Putative uncharacterized protein;
or4058	—	—	1.69	—	Lysozyme;
or4060	—	—	1.72	—	Putative DNA adenine methylase;
or4061	—	—	1.70	—	Gifsy-1 prophage RegQ;
or4064	—	—	1.67	—	NinB protein;
or4066	—	—	1.57	—	Putative cytosine-specific modification methylase;
or4070	—	—	1.63	—	DNA replication protein DnaC;
or4071	—	—	1.65	—	Conserved domain protein;
or4072	—	—	1.71	—	Predicted protein;
or4073	—	—	1.69	—	Regulatory protein CII;
or4079	—	—	1.70	—	DNA polymerase III, epsilon subunit;
or4081	—	—	1.52	—	Putative uncharacterized protein;
or4083	—	—	1.56	—	Putative phage integrase;

Gene	*hfq* RNA[1]	*hfq* Protein[1]	*rsmA* RNA[1]	*rsmA* Protein[1]	Description
Additional genes of interest					
or0363	—	—	1.49	—	Cellulose synthase operon C domain protein;
or0698	—	—	0.11	—	Gas vesicle protein;
or1349	—	0.65	0.44	—	QueF, 7-cyano-7-deazaguanine reductase, queuosine biosynthesis;
or1353	—	—	8.04	—	Taurine catabolism dioxygenase TauD/TfdA;
or1354	—	—	12.62	—	AMP-dependent synthetase and ligase;
or1714	—	—	83.17	21.22	Putative exported protein;
or2280	—	—	4.68	9.67	RhlA, surfactant biosynthesis;
or3590	—	—	0.37	0.68	H-NS protein;

[1]Fold changes were calculated as the ratio of mutant to WT. Differentially expressed genes were identified with an FDR threshold of 10% for RNA levels of *hfq* and protein levels of *rsmA*. A 5% FDR threshold was used for RNA levels of *rsmA*. *hfq* protein levels were identified with a *P* value < 10%.

6.2.5.3 Further observations from global changes in gene expression

Analysis of the RNA-seq data for the *hfq* mutant showed that expression of 22 of 35 genes involved in flagella production, which are clustered in the genome and are contained in various operons on either strand, was increased from 2.5 to 9.1-fold, despite a decrease in swimming and swarming motility. In order to explore this result, an extract of flagellar proteins was examined by SDS-PAGE analysis and major protein bands were identified by mass spectrometry (MALDI-TOF) (Figure 6.7A). The results suggest flagellin levels were increased significantly in the Δ*hfq* mutant as well as in the *rsmA* and *rpoS* mutants which displayed increased motility. The RNA-seq of the *rsmA* mutant also revealed that expression of 19 of the 35 flagellar genes were increased from 1.6 to 4.0-fold as well as the flagellar sigma factor, *rpoF*, and iTRAQ analysis showed flagellin protein was increased (Table 6.3). However, a control strain with a mutation in *flhC* gave a significantly reduced band corresponding to flagellin protein. *flhC* is located in the flagellar master regulator operon, and in a previous study the control strain was shown by electron microscopy to possess no flagella (Williamson et al., 2008). Additionally, the second gel band was identified as a porin protein, which was increased in the *hfq* mutant. Although, complementation of the *hfq* mutant was able to

restore swimming motility, expression of Hfq *in trans* did not reduce flagellin and porin to WT levels. Additionally, electron microscopy of the WT, *hfq* mutant, and complemented strain showed no significant differences in the presence of flagella on the cell surface (Figure 6.7B). This is in contrast to the role Hfq plays in flagellum production in *S. typhimurium* where 87% of the flagellar genes were downregulated in an *hfq* mutant which was impaired for swimming motility, and Hfq was shown to bind to *flhDC* mRNA (Sittka et al., 2008). Also, no flagella were present in an electron micrograph of an *E. coli hfq* mutant (Simonsen et al., 2011). However, the species-specific role of Hfq in the regulation of motility is illustrated by the increased swarming motility of a *Y. pseudotuberculosis hfq* mutant as well as increased biosurfactant production (Schiano et al., 2010). Therefore, this raises further questions regarding the role of Hfq in the regulation of flagellum production in S39006, and the mechanism explaining the decrease in motility despite an increase in flagellar gene expression.

Another set of genes with increased transcript levels in the *hfq* mutant include seven of 13 genes (2.2 to 3.5-fold) in an operon encoding an NADH dehydrogenase, whereas five of the genes were decreased 0.31 to 0.62-fold in the *rsmA* mutant. This operon is homologous to the *nuo* operon in *E. coli* encoding the membrane protein NADH:ubiquinone oxidoreductase I (NDH-1) which is one of two NADH deyhydrogenases the second being (NDH-2). NDH-1 and NDH-2 catalyzes the transfer of an electron from NADH to ubiquinone as part of the electron transport chain in aerobic respiration. However, only NDH-1 catalyzed electron flow generates an electrochemical gradient, while NDH-2 was shown to be the predominant respiratory source of H_2O_2 and O^{-}_{2} although it was also shown that H_2O_2 is primarily formed from a source outside the respiratory chain in *E. coli* (Messner and Imlay, 1999; Seaver and Imlay, 2004). Nevertheless, the increased levels of the *nuo* operon may contribute to the increased sensitivity to H_2O_2 in the S39006 *hfq* mutant. In contrast, a similar protein, the NADPH-dependent quinone oxidoreductase, *qor*, was found to be decreased at the level of RNA and protein production in the *S. typhimurium hfq* mutant (Sittka et al., 2008; Ansong et al., 2009), and *E. coli* Hfq was shown to be associated with *qor* DNA (Updegrove et al., 2011). These results suggest that Hfq could mediate the regulation of aerobic respiration and the transfer of electrons to the quinone pool within the cell.

Immediately upstream of the *nuo* operon is a LysR-family transcriptional regulator which was increased at the level of RNA (3.4-fold) and protein (2.85-fold) in the *hfq* mutant. This regulator is a homologue of HexS and HexA (74% and 82% identical) (hyperproduction of exoenzymes) from *S. marcescens* 274 and *P. atrosepticum* SCRI1043 (*Pat*), respectively. A *hexS* mutant was increased for transcription of prodigiosin and HexS was shown to bind to the *pigA* promoter and reduce prodigiosin production (Tanikawa et al., 2006). Additionally, expression of HexA from *Pat* and the related *P. carotovorum* subsp. *carotovorum* (*Pcc*) in S39006 repressed prodigiosin and carbapenem production (Harris et al., 1998). Surprisingly, a *hexA* (*pigU*) transposon mutant (strain HSPIG43) in S39006 with an insertion in the promoter region upstream of the CDS was decreased for prodigiosin production; although it is not known how the intergenic transposon insertion affected the levels of *pigU* (Fineran et al., 2005b). Nevertheless, the increased levels of PigU in the *hfq* mutant most likely contribute to the repression of prodigiosin production. *Pcc* HexA also represses exoenzyme production and binds to the promoter of *pelC* encoding a pectate lyase (Harris et al., 1998), and it represses the production of *rpoS* and RsmB, the small RNA antagonist of RsmA (Mukherjee et al., 2000). The HexA homologue in *E. coli*, LrhA, was shown to repress *rpoS* expression at the level of translation and represses the transcription of the sRNA RprA (Peterson et al., 2006). Therefore, increased PigU in the *hfq* mutant may contribute to the decrease in *rpoS* transcript levels and exoenzyme production.

Several additional regulators of pigment production that have been identified in previous studies were also differentially expressed and decreased in the *rsmA* mutant. These include *rap*, *pstS*, *phoU*, *rpoS*, and *vfmE*. *rap* is a transcriptional regulator of the SlyA family which is an activator of secondary metabolism (Thomson et al., 1997). *pstS* and *phoU* form part of the *pstSCAB-phoU* operon which, if deleted, causes increased hyperpigmentation (Slater et al., 2003). *rpoS* is a sigma factor which is a repressor of secondary metabolism (see Chapter 5). And *vfmE* is an AraC-type regulator which is an activator of secondary metabolism (see Chapter 3).

Secretion and pili systems involved in exporting virulence factors and attachment, respectively, were also identified as differentially expressed in the *rsmA* mutant. The S39006 encodes putative Type I to III secretion systems (T1SS to T3SS).

The *rsmA* mutant was increased for expression of genes involved in T2SS (1.8 to 3.2-fold) and Type IV pilus (T4P) genes (1.4 to 1.5-fold), whereas a cluster of genes involved in T3SS was unaffected. S39006 contains a gene cluster most closely related to the Type II Out secretion system in species of *Dickeya dadantii* and *D. zeae* (Sandkvist, 2001). Four genes in this cluster were increased for expression in the *rsmA* mutant. However, SecY is decreased which is part of the SecYEG machinery that translocates proteins across the inner membrance into the periplasm for secretion by the T2SS.

Further investigation of the T4P genes revealed that S39006 possesses a Type IV pilus 11 gene cluster (*pilLMNOPQRSUVW*; 10,806 bp) of which 5 are increased in expression in the *rsmA* mutant. The *pil* cluster resides within an 82.1 kb putative pathogenicity island (PAI) and is most closely related to the *pil* cluster within the *Yersinia* adhesion pathogenicity island (YAPI; 98 kb) found in *Y. pseudotuberculosis* and *Y. enterocolitica*, but not *Y. pestis* (Collyn et al., 2004). The *pil* gene cluster is also found within the *Salmonella enterica* pathogenicity island SPI-7 (134 kb), the *Pat* horizontally acquired island HAI-2 (97.6 kb) (Bell et al., 2004; Seth-Smith, 2008), and the recently sequenced *Erwinia* strain Ejp617 which causes bacterial shoot blight of pear in Japan (Park et al., 2011). However, the S39006 *pil* cluster does not appear to be present in related *Serratia* species. The S39006 PAI contains the characteristics features of a PAI which include: i) insertion between a (phenylalanine) tRNA and a 50 bp partially duplicated copy (similar to YAPI and SPI-7), ii) a G+C content of 52.9% which differs from the genome (49.2%), and iii) a number of mobile genetic elements such as integrases and transposases. The S39006 PAI carries 85 ORFs which include a type-I restriction modification system and CRISPR-associated proteins.

The *S. enterica pil* cluster was shown to encode type IVB pili and a *pil*$^{-}$ mutant was significantly reduced for invasion into human intestinal cells (Zhang et al., 2000). However, a *Y. pseudotuberculosis pil*$^{-}$ mutant was delayed in the killing of mice when administered orally, but was as virulent as the parental strain when administered intravenously. While a survey for the presence of the *pil* locus showed that it was present in only 38 of 92 (41%) tested strains (Collyn et al., 2002). Therefore, Collyn *et al.* (2002) suggested that the *pil* cluster may form an “adaptation pathogenicity island” allowing the bacteria to colonize a wider variety of hosts.

Secreted virulence factors were identified as differentially expressed in the *hfq* and *rsmA* mutants. In the *rsmA* mutant the expression of three genes involved in pectate lyase production was decreased whereas the mRNA of Pectate lyase L was increased 4.67-fold which is consistent with the 8.44-fold increase in secreted Pel activity relative to the WT (Williamson *et al.*, in preparation). However, cellulase expression was decreased although secreted Cel activity was marginally increased 1.31-fold. Cellulase and pectate lyase are exported by the T2SS and the increased expression of the Out Type 2 system may explain the increase in Cel activity despite a decrease in expression. Interestingly, the gene with the highest fold change at 184-fold is an uncharacterized protein (or1355) that shares 44% aa similarity to the δ-thermostable haemolysin (δ-VPH) of *V. cholerae* and *parahaemolyticus* species. However, a second gene (or0539) in S39006 with 50% similarity to δ-VPH was not differentially expressed. The *Vibrio* species contain four families of haemolysins, including TDH of *V. parahaemolyticus* and HlyA of *V. cholerae* which are closely associated with virulence through lysis of cells, notably erythrocytes (Zhang and Austin, 2005). However, the role of some haemolysins, such as δ-VPH, is not clear. A study examining the role of δ-VPH in *V. cholerae* O1 showed that it is a 22.8 kD protein with haemolytic activity on sheep erythrocytes (Fallarino et al., 2002). Haemolytic activity was observed in both whole-cell lysates and supernatant fractions despite the absence of an identifiable signal peptide. Similarly, or1355 is predicted to encode a 22.6 kD protein but no signal peptide was identified. The increase in expression of δ-VPH, the Out T2SS and the Pil T4SS could contribute to the increased virulence of an *rsmA* mutant in the host *C. elegans* (N. Williamson, in preparation).

Additionally, or1355 is part of a putative operon of four genes (or1353-6) all of which were among the top ten genes with highest fold-change in the RNA-seq dataset of the *rsmA* mutant. Proteins encoded by the four genes are most similar to proteins from the plant endosymbiont, *Pseudomonas putida* W619, and various soil species isolates of *Burkholderia*. The genes forming the operon (in addition to or1355) include: or1353 encoding taurine catabolism dioxygenase (TauD/TfdA) which in *E. coli* allows the utilization of taurine as a sulphur source under conditions of sulphate starvation;

or1354 encoding an AMP-dependent synthetase and ligase; and or1356 encoding a drug resistance transporter of the Bcr/CflA subfamily.

Additional categories of genes that are altered for expression in the *rsmA* mutant include efflux transporters important in drug resistance, outer membrane proteins such as increased expression of a polysaccharide biosynthesis protein, sigma factors, and stress response genes. But perhaps one of the most interesting observations for the *rsmA* mutant is the increased expression in three putative prophage clusters encoding bacteriophage or related genes, including the CI repressor, CII regulator, reverse transcriptase, and lysozyme. These results suggest that RsmA regulates lysogeny and that the absence of *rsmA* results in increased expression of prophage genes and possibly a switch into the lytic cycle.

Further genes of interest that were of particular interest in the *rsmA* dataset include or1714 which had the second highest fold-change at 83-fold in the RNA-seq dataset and was matched by a 21-fold increase in protein. It encodes a small putative exported protein (125 aa) and a conserved domain search indicates that it has a helix-hairpin-helix repeat region. Members of this domain subfamily include competence protein ComEA for transformation by exogenous DNA. or0698 encoding a gas vesicle protein was decreased by a factor of 10 in the *rsmA* mutant and or1349 encoding QueF involved in queuosine biosynthesis was reduced by about half in both the *rsmA* and *hfq* mutants. These results suggest RsmA regulates diverse phenotypes including genes connected with the production of gas vesicle and queuosine products.

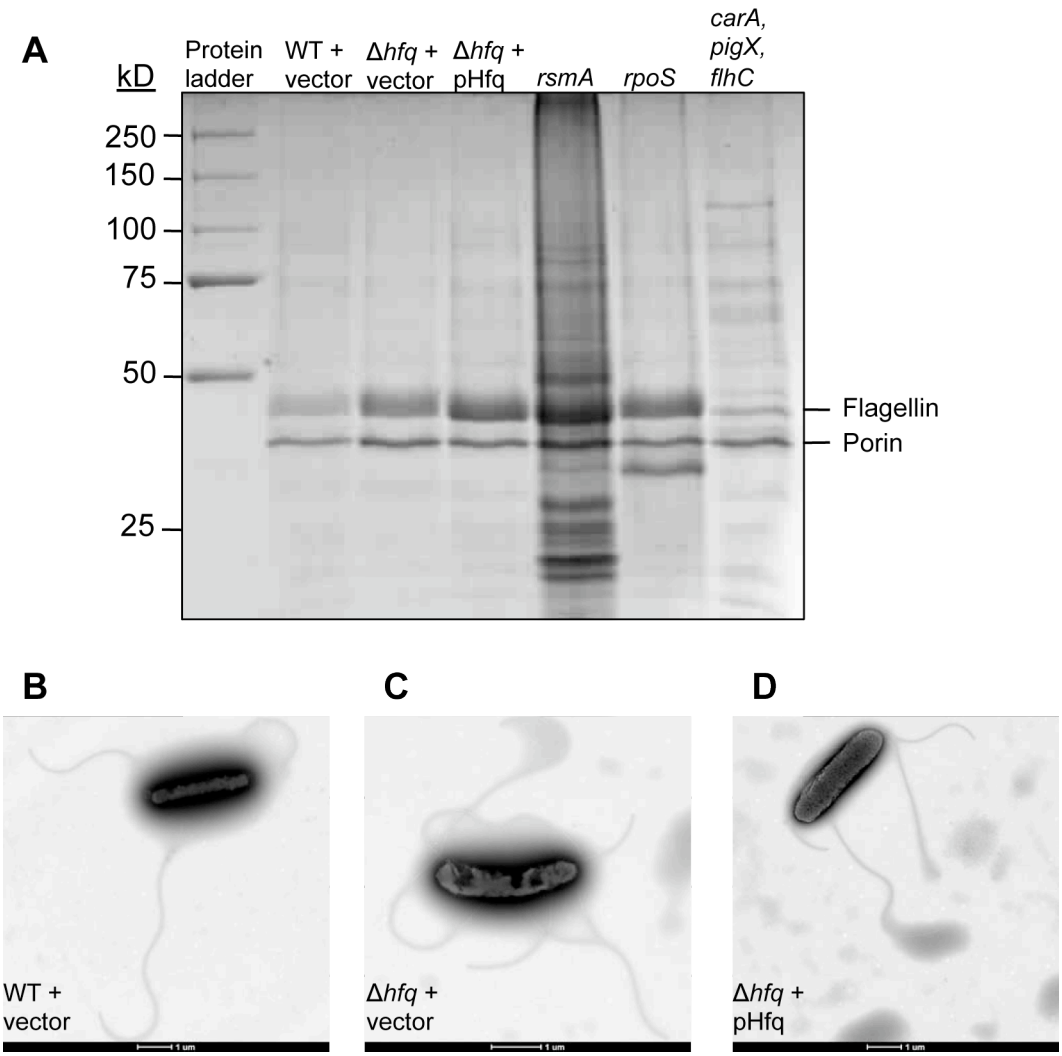


Figure 6.7. SDS-PAGE analysis of flagellin.

A. Extracts of flagella were analyzed by SDS-PAGE. Two abundant protein bands were identified by mass spectrometry (MALDI-TOF) as Flagellin (or3115, 41.4 kD) and a Porin (or3201, 38.8 kD). B-D. Transmission electron micrographs of WT + vector (B), Δ*hfq* + vector (C), and Δ*hfq* + pHfq (D) are shown with the scale bar at the bottom of each image indicating a length of 1 μm. Samples were grown to early stationary phase the same as those used in the RNA-seq. The vector used was pQE80oriT and pHfq refers to pNB36. The strains analyzed were NMW9 (Δ*hfq*), NMW7 (*rsmA*), NMW25 (*rpoS*), and NWX21 (*carA*, *pigX*, *flhC*).

6.3 Conclusion

In this study the transcriptomes of S39006 WT and the *hfq* and *rsmA* mutants were characterized and differential expression determined. Transcriptomic changes were correlated with altered protein expression which allowed for the characterization of the regulons of two major RNA-binding post-transcriptional regulators. Genes in the categories of secondary metabolism and motility were co-regulated and differentially expressed in both the *hfq* and *rsmA* mutants. Surprisingly, expression of the *pig* cluster was reduced at the level of RNA in both mutants but increased at the level of protein in the *rsmA* mutant consistent with the hyperpigmented phenotype. This suggests that RsmA regulates translation of the Pig cluster. The second co-regulated category of genes include the flagellar and chemotaxis genes which are increased in expression in both mutants. Increased flagellar expression is consistent with increased swimming and swarming motility in the *rsmA* mutant, whereas the *hfq* mutant was decreased for motility. This indicates that *hfq* mutant cells are not able to coordinate movement using their flagella. The differential expression analysis also identified Type II secretion and Type IV pilus systems and virulence factors increased for expression in the *rsmA* mutant explaining the increased virulence for this strain in both potato tuber rot and *C. elegans* infection assays.

The RNA-seq allowed identification of non-coding RNAs and potential *cis*-acting regulatory RNA elements, although this was limited by the technical difficulties experienced with the strand-specific protocol used in this study. Experimental observations suggested the presence of a candidate regulatory RNA and *cis*-regulatory element. A candidate *cis*-acting antisense RNA for the transcriptional regulator, *vfmE*, was identified (Figure 3.12; pg. 85). Transposon mutants were selected with insertions in the intergenic region downstream of this gene. These insertions disrupted transcription of the antisense RNA without altering *vfmE* transcription leading to increased transcript levels of *vfmE*. The increased expression of *vfmE* caused increased production of prodigiosin.

The candidate *cis*-acting regulatory RNA element explored in this study was also shown to influence prodigiosin production. The peak of reads overlapping the UTR and start of the ORF or1385 which encodes a conserved hypothetical protein show that the

cis-acting regulatory RNA element is encoded within the mRNA of the ORF and regulates transcriptional elongation. Bioinformatic analysis of or1385 indicates that it most likely encodes an enzyme and therefore it is speculated whether the regulatory RNA element is a riboswitch that responds to a metabolic cue related to the function of the ORF.

In summary, this study characterizes the regulons of Hfq and RsmA, two global regulatory RNA-binding proteins, at both an RNA and protein level using transcriptomic and proteomic methods. This study has revealed new insights into their regulatory targets. Furthermore, the characterization of regulatory RNA elements raises new questions about their role in antibiotic regulation that can be explored in future work.

Chapter 7. Summary

Bacteria are exposed to a wide variety of environmental conditions and must be able to sense and respond to changes in such factors as nutrient levels and temperature. Cells possess the machinery to sense diverse stimuli and process the information to precisely alter gene expression in reaction to extracellular conditions. Regulation of transcription by protein transcriptional regulators is the most well-studied mechanism of regulation. However, recent studies are increasingly demonstrating the importance of post-transcriptional regulation which include mechanisms that regulate transcription elongation and termination, translation initiation and termination, and mRNA stability (Winkler and Breaker, 2005). These mechanisms involved the use of *trans*-acting small regulatory RNAs, *cis*-acting antisense RNAs, and *cis*-acting regulatory RNA elements encoded within mRNAs, such as riboswitches, in the 5' UTR.

Serratia sp. 39006 is a Gram-negative bacterium that produces the red-pigmented antibiotic, prodigiosin, which provides the advantage to explore the wide variety of genetic regulatory mechanisms that influence pigment levels. This study investigated the role of post-transcriptional regulation by uncovering the role of the global regulatory RNA-binding protein Hfq in the control of antibiotic and exoenzyme production and pathogenesis. It was established that Hfq control of antibiotic production and plant pathogenesis was independent of its regulation of the sigma factor RpoS. Additionally, the result that the QS-response regulators, SmaR and CarR, are regulated by Hfq suggests that Hfq-dependent small RNAs may play a role in QS regulation by increasing transcript stability of SmaR and CarR.

During the course of this study new methods were developed for strand-specific RNA sequencing using advanced high-throughput sequencing technologies. This allowed for the first time to explore the whole transcriptomes of various bacterial species to identify new small and antisense RNAs, mRNAs with long 5' UTRs potentially encoding *cis*-regulatory elements (e.g. riboswitches), and transcription start sites at a single nucleotide level (Croucher and Thomson, 2010). In collaboration with the WTSI, RNA-seq was applied to the study of the whole transcriptomes of S39006 WT and the *hfq* and *rsmA* mutants. Although the method of directional RNA-seq used in this study

limited the ability to identify non-coding RNAs, it was possible to identify and provide support from experiments for two candidate regulatory RNA elements that influence prodigiosin production. The candidate antisense RNA identified by RNA-seq provides an explanation for the mechanism by which transposon mutants (identified at the start of the study) with multiple insertions in an intergenic region downstream of *vfmE* resulted in a hyperpigmented phenotype. Differential expression analysis was also performed on the *hfq* and *rsmA* RNA-seq datasets.

Furthermore, a comparison of changes in the transcriptome to changes in the proteome (using the iTRAQ-based quantitative proteomic approach) allowed the detailed characterization of the regulons of RsmA and Hfq which shared a large number of co-regulated genes. The differential expression analysis revealed that flagellar gene expression and production was increased in the *hfq* mutant despite a decrease in motility, and Type II secretion and Type IV pilus systems and virulence factor expression was increased in the *rsmA* mutant.

The two main questions raised by the results of this study for future exploration include, i) the identification and characterization of both *cis* and *trans*-acting regulatory RNAs in the regulation of prodigiosin production, and ii) the determination of direct RNA targets of Hfq and RsmA. The continued innovation and development in the area of high-throughput sequencing will allow for the rapid investigation of these questions. The re-sequencing of the whole transcriptome of S39006 WT using more robust strand-specific RNA-seq methods will lead to the identification of novel regulatory RNAs and precisely map TSSs to a nucleotide level. Sequencing of cDNA libraries of RNAs co-immunoprecipitated with epitope-tagged Hfq and RsmA will readily reveal their targets and possibly their RNA binding sites. Continued research in this area will increase the understanding of the mechanism of post-transcriptional regulation and its importance in the regulation of the key phenotypes of antibiotic production and pathogenesis.

References

Aballay, A., and Ausubel, F.M. (2002) *Caenorhabditis elegans* as a host for the study of host-pathogen interactions. *Curr Opin Microbiol* **5**: 97-101.

Aballay, A., Yorgey, P., and Ausubel, F.M. (2000) *Salmonella typhimurium* proliferates and establishes a persistent infection in the intestine of *Caenorhabditis elegans*. *Curr Biol* **10**: 1539-1542.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402.

Anders, S., and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.

Ansong, C., Yoon, H., Porwollik, S., Mottaz-Brewer, H., Petritis, B.O., Jaitly, N. et al. (2009) Global systems-level analysis of Hfq and SmpB deletion mutants in *Salmonella*: implications for virulence and global protein translation. *PLoS ONE* **4**: e4809.

Avery, L., and Shtonda, B.B. (2003) Food transport in the *C. elegans* pharynx. *J Exp Biol* **206**: 2441-2457.

Bainton, N.J., Stead, P., Chhabra, S.R., Bycroft, B.W., Salmond, G.P., Stewart, G.S., and Williams, P. (1992) *N*-(3-oxohexanoyl)-L-homoserine lactone regulates carbapenem antibiotic production in *Erwinia carotovora*. *Biochem J* **288 (Pt 3)**: 997-1004.

Barnard, A.M.L., Bowden, S.D., Burr, T., Coulthurst, S.J., Monson, R.E., and Salmond, G.P.C. (2007) Quorum sensing, virulence and secondary metabolite production in plant soft-rotting bacteria. *Philos Trans R Soc Lond, B, Biol Sci* **362**: 1165-1183.

Bejerano-Sagie, M., and Xavier, K.B. (2007) The role of small RNAs in quorum sensing. *Curr Opin Microbiol* **10**: 189-198.

Bell, K.S., Sebaihia, M., Pritchard, L., Holden, M.T.G., Hyman, L.J., Holeva, M.C. et al. (2004) Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. *Proc Natl Acad Sci USA* **101**: 11105-11110.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G. et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**: 53-59.

Bérdy, J. (2005) Bioactive microbial metabolites. *J Antibiot* **58**: 1-26.

Bowden, S.D., and Salmond, G.P.C. (2006) Exploitation of a beta-lactamase reporter gene fusion in the carbapenem antibiotic production operon to study adaptive evolution in *Erwinia carotovora*. *Microbiology* **152**: 1089-1097.

Brescia, C.C., Mikulecky, P.J., Feig, A.L., and Sledjeski, D.D. (2003) Identification of the Hfq-binding site on DsrA RNA: Hfq binds without altering DsrA secondary structure. *RNA* **9**: 33-43.

Brosch, M., Yu, L., Hubbard, T., and Choudhary, J. (2009) Accurate and sensitive peptide identification with Mascot Percolator. *J Proteome Res* **8**: 3176-3181.

Brown, L., and Elliott, T. (1996) Efficient translation of the RpoS sigma factor in *Salmonella typhimurium* requires host factor I, an RNA-binding protein encoded by the *hfq* gene. *J Bacteriol* **178**: 3763-3770.

Brown, L., and Elliott, T. (1997) Mutations that increase expression of the *rpoS* gene and decrease its dependence on hfq function in *Salmonella typhimurium*. *J Bacteriol* **179**: 656-662.

Bycroft, B., Maslen, C., Box, S., Brown, A., and Tyler, J. (1987) The isolation and characterisation of (3R,5R)- and (3S,5R)-carbapenam-3-carboxylic acid from *Serratia* and *Erwinia* species and their putative biosynthetic role. *J Chem Soc, Chem Commun VL* -: 1623-1625.

Carver, T., Böhme, U., Otto, T.D., Parkhill, J., and Berriman, M. (2010) BamView: viewing mapped read alignment data in the context of the reference sequence. *Bioinformatics* **26**: 676-677.

Castrillo, J.I., Zeef, L.A., Hoyle, D.C., Zhang, N., Hayes, A., Gardner, D.C.J. et al. (2007) Growth control of the eukaryote cell: a systems biology study in yeast. *J Biol* **6**: 4.

Chao, Y., and Vogel, J. (2010) The role of Hfq in bacterial pathogens. *Curr Opin Microbiol*.

Chawrai, S.R., Williamson, N.R., Salmond, G.P.C., and Leeper, F.J. (2008) Chemoenzymatic synthesis of prodigiosin analogues--exploring the substrate specificity of PigC. *Chem Commun (Camb)*: 1862-1864.

Christiansen, J.K., Larsen, M.H., Ingmer, H., Søgaard-Andersen, L., and Kallipolitis, B.H. (2004) The RNA-binding protein Hfq of *Listeria monocytogenes*: role in stress tolerance and virulence. *J Bacteriol* **186**: 3355-3362.

Collyn, F., Billault, A., Mullet, C., Simonet, M., and Marceau, M. (2004) YAPI, a new *Yersinia pseudotuberculosis* pathogenicity island. *Infect Immun* **72**: 4784-4790.

Collyn, F., Léty, M.-A., Nair, S., Escuyer, V., Ben Younes, A., Simonet, M., and Marceau, M. (2002) *Yersinia pseudotuberculosis* harbors a type IV pilus gene cluster that contributes to pathogenicity. *Infect Immun* **70**: 6196-6205.

Costechareyre, D., Dridi, B., Rahbé, Y., and Condemine, G. (2010) Cyt toxin expression reveals an inverse regulation of insect and plant virulence factors of *Dickeya dadantii*. *Environ Microbiol* **12**: 3290-3301.

Couillault, C., and Ewbank, J.J. (2002) Diverse bacteria are pathogens of *Caenorhabditis elegans*. *Infect Immun* **70**: 4705-4707.

Coulthurst, S., Lilley, K., and Salmond, G. (2005a) Genetic and proteomic analysis of the role of *luxS* in the enteric phytopathogen, *Erwinia carotovora*. *Mol Plant Pathol* **7**: 31-45.

Coulthurst, S.J., Barnard, A.M.L., and Salmond, G.P.C. (2005b) Regulation and biosynthesis of carbapenem antibiotics in bacteria. *Nat Rev Micro* **3**: 295-306.

Coulthurst, S.J., Lilley, K.S., Hedley, P.E., Liu, H., Toth, I.K., and Salmond, G.P.C. (2008) DsbA plays a critical and multifaceted role in the production of secreted virulence factors by the phytopathogen *Erwinia carotovora* subsp. *atroseptica*. *J Biol Chem* **283**: 23739-23753.

Croucher, N., Fookes, M., Perkins, T., Turner, D., Marguerat, S., Keane, T. et al. (2009) A simple method for directional transcriptome sequencing using Illumina technology. *Nucleic Acids Res*.

Croucher, N.J., and Thomson, N.R. (2010) Studying bacterial transcriptomes using RNA-seq. *Curr Opin Microbiol* **13**: 619-624.

Crow, M.A. (2001) The genetic regulation of pigment and antibiotic biosynthesis in *Serratia* sp. In *Department of Biochemistry*. Cambridge: The University of Cambridge.

Cunning, C., Brown, L., and Elliott, T. (1998) Promoter substitution and deletion analysis of upstream region required for rpoS translational regulation. *J Bacteriol* **180**: 4564-4570.

Darby, C., Cosma, C.L., Thomas, J.H., and Manoil, C. (1999) Lethal paralysis of *Caenorhabditis elegans* by *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* **96**: 15202-15207.

Demain, A.L. (2006) From natural products discovery to commercialization: a success story. *J Ind Microbiol Biotechnol* **33**: 486-495.

Demain, A.L., and Adrio, J.L. (2008) Contributions of microorganisms to industrial biology. *Mol Biotechnol* **38**: 41-55.

Demarre, G., Guérout, A.-M., Matsumoto-Mashimo, C., Rowe-Magnus, D.A., Marlière, P., and Mazel, D. (2005) A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPalpha) conjugative machineries and their cognate *Escherichia coli* host strains. *Res Microbiol* **156**: 245-255.

Ding, Y., Davis, B.M., and Waldor, M.K. (2004) Hfq is essential for *Vibrio cholerae* virulence and downregulates sigma expression. *Mol Microbiol* **53**: 345-354.

Dombrecht, B., Vanderleyden, J., and Michiels, J. (2001) Stable RK2-derived cloning vectors for the analysis of gene expression and gene function in gram-negative bacteria. *Mol Plant-Microbe Interact* **14**: 426-430.

Dong, T., and Schellhorn, H.E. (2010) Role of RpoS in virulence of pathogens. *Infect Immun* **78**: 887-897.

Evans, T.J., Crow, M.A., Williamson, N.R., Orme, W., Thomson, N.R., Komitopoulou, E., and Salmond, G.P.C. (2010) Characterization of a broad-host-range flagellum-dependent phage that mediates high-efficiency generalized transduction in, and between, *Serratia* and *Pantoea*. *Microbiology* **156**: 240-247.

Fallarino, A., Attridge, S.R., Manning, P.A., and Focareta, T. (2002) Cloning and characterization of a novel haemolysin in *Vibrio cholerae* O1 that does not directly contribute to the virulence of the organism. *Microbiology* **148**: 2181-2189.

Fantappiè, L., Metruccio, M.M.E., Seib, K.L., Oriente, F., Cartocci, E., Ferlicca, F. et al. (2009) The RNA chaperone Hfq is involved in stress response and virulence in *Neisseria meningitidis* and is a pleiotropic regulator of protein expression. *Infect Immun* **77**: 1842-1853.

Fineran, P.C. (2006) Investigation of a regulatory network controlling secondary metabolism in *Serratia*. In *Department of Biochemistry*. Cambridge: The University of Cambridge.

Fineran, P.C., Everson, L., Slater, H., and Salmond, G.P.C. (2005a) A GntR family transcriptional regulator (PigT) controls gluconate-mediated repression and defines a new, independent pathway for regulation of the tripyrrole antibiotic, prodigiosin, in *Serratia*. *Microbiology* **151**: 3833-3845.

Fineran, P.C., Williamson, N.R., Lilley, K.S., and Salmond, G.P.C. (2007) Virulence and prodigiosin antibiotic biosynthesis in *Serratia* are regulated pleiotropically by the GGDEF/EAL domain protein, PigX. *J Bacteriol* **189**: 7653-7662.

Fineran, P.C., Slater, H., Everson, L., Hughes, K., and Salmond, G.P.C. (2005b) Biosynthesis of tripyrrole and beta-lactam secondary metabolites in *Serratia*: integration of quorum sensing with multiple new regulatory components in the control of prodigiosin and carbapenem antibiotic production. *Mol Microbiol* **56**: 1495-1517.

Folichon, M., Arluison, V., Pellegrini, O., Huntzinger, E., Régnier, P., and Hajnsdorf, E. (2003) The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation. *Nucleic Acids Res* **31**: 7302-7310.

Fox, E.M., and Howlett, B.J. (2008) Secondary metabolism: regulation and role in fungal biology. *Curr Opin Microbiol* **11**: 481-487.

Franze de Fernandez, M.T., Eoyang, L., and August, J.T. (1968) Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. *Nature* **219**: 588-590.

Gardner, P.P., Daub, J., Tate, J.G., Nawrocki, E.P., Kolbe, D.L., Lindgreen, S. et al. (2009) Rfam: updates to the RNA families database. *Nucleic Acids Res* **37**: D136-140.

Geng, J., Song, Y., Yang, L., Feng, Y., Qiu, Y., Li, G. et al. (2009) Involvement of the post-transcriptional regulator Hfq in *Yersinia pestis* virulence. *PLoS ONE* **4**: e6213.

Gravato-Nobre, M.J., and Hodgkin, J. (2005) *Caenorhabditis elegans* as a model for innate immunity to pathogens. *Cell Microbiol* **7**: 741-751.

Gristwood, T., Fineran, P.C., Everson, L., and Salmond, G.P.C. (2008) PigZ, a TetR/AcrR family repressor, modulates secondary metabolism via the expression of a putative four-component resistance-nodulation-cell-division efflux pump, ZrpADBC, in *Serratia* sp. ATCC 39006. *Mol Microbiol* **69**: 418-435.

Gristwood, T., Fineran, P.C., Everson, L., Williamson, N.R., and Salmond, G.P. (2009) The PhoBR two-component system regulates antibiotic biosynthesis in *Serratia* in response to phosphate. *BMC Microbiol* **9**: 112.

Gristwood, T., McNeil, M.B., Clulow, J.S., Salmond, G.P.C., and Fineran, P.C. (2011) PigS and PigP Regulate Prodigiosin Biosynthesis in *Serratia* via Differential Control of Divergent Operons, Which Include Predicted Transporters of Sulfur-Containing Molecules. *J Bacteriol* **193**: 1076-1085.

Hansen, A.-M., and Kaper, J.B. (2009) Hfq affects the expression of the LEE pathogenicity island in enterohaemorrhagic *Escherichia coli*. *Mol Microbiol* **73**: 446-465.

Harris, S.J., Shih, Y.L., Bentley, S.D., and Salmond, G.P. (1998) The *hexA* gene of *Erwinia carotovora* encodes a LysR homologue and regulates motility and the expression of multiple virulence determinants. *Mol Microbiol* **28**: 705-717.

Harshey, R.M. (2003) Bacterial motility on a surface: many ways to a common goal. *Annu Rev Microbiol* **57**: 249-273.

Heeb, S., and Haas, D. (2001) Regulatory roles of the GacS/GacA two-component system in plant-associated and other gram-negative bacteria. *Mol Plant-Microbe Interact* **14**: 1351-1363.

Hengge, R. (2008) The two-component network and the general stress sigma factor RpoS (sigma S) in *Escherichia coli*. *Adv Exp Med Biol* **631**: 40-53.

Herrero, M., de Lorenzo, V., and Timmis, K.N. (1990) Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria. *J Bacteriol* **172**: 6557-6567.

Hommais, F., Oger-Desfeux, C., Van Gijsegem, F., Castang, S., Ligori, S., Expert, D. et al. (2008) PecS is a global regulator of the symptomatic phase in the phytopathogenic bacterium *Erwinia chrysanthemi* 3937. *J Bacteriol* **190**: 7508-7522.

Izu, H., Adachi, O., and Yamada, M. (1997) Gene organization and transcriptional regulation of the *gntRKU* operon involved in gluconate uptake and catabolism of *Escherichia coli*. *J Mol Biol* **267**: 778-793.

Kaniga, K., Delor, I., and Cornelis, G.R. (1991) A wide-host-range suicide vector for improving reverse genetics in gram-negative bacteria: inactivation of the *blaA* gene of *Yersinia enterocolitica*. *Gene* **109**: 137-141.

Karp, N.A., Huber, W., Sadowski, P.G., Charles, P.D., Hester, S.V., and Lilley, K.S. (2010) Addressing accuracy and precision issues in iTRAQ quantitation. *Mol Cell Proteomics*.

Kulesus, R.R., Diaz-Perez, K., Slechta, E.S., Eto, D.S., and Mulvey, M.A. (2008) Impact of the RNA chaperone Hfq on the fitness and virulence potential of uropathogenic *Escherichia coli*. *Infect Immun* **76**: 3019-3026.

Kurz, C.L., Chauvet, S., Andrès, E., Aurouze, M., Vallet, I., Michel, G.P.F. et al. (2003) Virulence factors of the human opportunistic pathogen *Serratia marcescens* identified by in vivo screening. *EMBO J* **22**: 1451-1460.

Kwon, S.-K., Park, Y.-K., and Kim, J.F. (2010) Genome-wide screening and identification of factors affecting the biosynthesis of prodigiosin by *Hahella chejuensis*, using *Escherichia coli* as a surrogate host. *Appl Environ Microbiol* **76**: 1661-1668.

L. Riddle, D. (1997) *C. elegans* II. 1222.

Labrousse, A., Chauvet, S., Couillault, C., Kurz, C.L., and Ewbank, J.J. (2000) *Caenorhabditis elegans* is a model host for *Salmonella typhimurium*. *Curr Biol* **10**: 1543-1545.

Lange, R., and Hengge-Aronis, R. (1994) The cellular concentration of the sigma S subunit of RNA polymerase in *Escherichia coli* is controlled at the levels of transcription, translation, and protein stability. *Genes Dev* **8**: 1600-1612.

Lapouge, K., Schubert, M., Allain, F.H.-T., and Haas, D. (2008) Gac/Rsm signal transduction pathway of gamma-proteobacteria: from RNA recognition to regulation of social behaviour. *Mol Microbiol* **67**: 241-253.

Lease, R.A., and Woodson, S.A. (2004) Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. *J Mol Biol* **344**: 1211-1223.

Lenz, D.H., Mok, K.C., Lilley, B.N., Kulkarni, R.V., Wingreen, N.S., and Bassler, B.L. (2004) The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell* **118**: 69-82.

Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N. et al. (2010) Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods*.

Link, T.M., Valentin-Hansen, P., and Brennan, R.G. (2009) Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA. *Proc Natl Acad Sci USA* **106**: 19292-19297.

Lorenz, C., Gesell, T., Zimmermann, B., Schoeberl, U., Bilusic, I., Rajkowitsch, L. et al. (2010) Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts. *Nucleic Acids Res*.

Lucchetti-Miganeh, C., Burrowes, E., Baysse, C., and Ermel, G. (2008) The post-transcriptional regulator CsrA plays a central role in the adaptation of bacterial pathogens to different stages of infection in animal hosts. *Microbiology* **154**: 16-29.

Lundberg, B.E., Wolf, R.E., Dinauer, M.C., Xu, Y., and Fang, F.C. (1999) Glucose 6-phosphate dehydrogenase is required for *Salmonella typhimurium* virulence and resistance to reactive oxygen and nitrogen intermediates. *Infect Immun* **67**: 436-438.

Lybecker, M.C., Abel, C.A., Feig, A.L., and Samuels, D.S. (2010) Identification and function of the RNA chaperone Hfq in the Lyme disease spirochete *Borrelia burgdorferi*. *Mol Microbiol*.

MacLean, D., Jones, J.D.G., and Studholme, D.J. (2009) Application of 'next-generation' sequencing technologies to microbial genetics. *Nat Rev Micro* **7**: 287-296.

Mahajan-Miklos, S., Tan, M.W., Rahme, L.G., and Ausubel, F.M. (1999) Molecular mechanisms of bacterial virulence elucidated using a *Pseudomonas aeruginosa-Caenorhabditis elegans* pathogenesis model. *Cell* **96**: 47-56.

Majdalani, N., Hernandez, D., and Gottesman, S. (2002) Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. *Mol Microbiol* **46**: 813-826.

Majdalani, N., Vanderpool, C.K., and Gottesman, S. (2005) Bacterial small RNA regulators. *Crit Rev Biochem Mol Biol* **40**: 93-113.

Majdalani, N., Cunning, C., Sledjeski, D., Elliott, T., and Gottesman, S. (1998) DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proc Natl Acad Sci USA* **95**: 12462-12467.

Mallo, G.V., Kurz, C.L., Couillault, C., Pujol, N., Granjeaud, S., Kohara, Y., and Ewbank, J.J. (2002) Inducible antibacterial defense system in *C. elegans*. *Curr Biol* **12**: 1209-1214.

Marsden, R.L., McGuffin, L.J., and Jones, D.T. (2002) Rapid protein domain assignment from amino acid sequence using predicted secondary structure. *Protein Sci* **11**: 2814-2824.

Mattatall, N.R., and Sanderson, K.E. (1996) *Salmonella typhimurium* LT2 possesses three distinct 23S rRNA intervening sequences. *J Bacteriol* **178**: 2272-2278.

McCullen, C.A., Benhammou, J.N., Majdalani, N., and Gottesman, S. (2010) Mechanism of positive regulation by DsrA and RprA small noncoding RNAs: pairing increases translation and protects *rpoS* mRNA from degradation. *J Bacteriol* **192**: 5559-5571.

McNealy, T.L., Forsbach-Birk, V., Shi, C., and Marre, R. (2005) The Hfq homolog in *Legionella pneumophila* demonstrates regulation by LetA and RpoS and interacts with the global regulator CsrA. *J Bacteriol* **187**: 1527-1532.

Meibom, K.L., Forslund, A.-L., Kuoppa, K., Alkhuder, K., Dubail, I., Dupuis, M. et al. (2009) Hfq, a novel pleiotropic regulator of virulence-associated genes in *Francisella tularensis*. *Infect Immun* **77**: 1866-1880.

Messner, K.R., and Imlay, J.A. (1999) The Identification of Primary Sites of Superoxide and Hydrogen Peroxide Formation in the Aerobic Respiratory Chain and Sulfite Reductase Complex of *Escherichia coli*. *J Biol Chem* **274**: 10119-10128.

Metcalf, W.W., Jiang, W., Daniels, L.L., Kim, S.K., Haldimann, A., and Wanner, B.L. (1996) Conditionally replicative and conjugative plasmids carrying *lacZ* alpha for cloning, mutagenesis, and allele replacement in bacteria. *Plasmid* **35**: 1-13.

Miller, J.H. (1972) *Experiments in molecular genetics*. New York, NY: Cold Spring Harbour Laboratory Press.

Miller, M.D., Aravind, L., Bakolitsa, C., Rife, C.L., Carlton, D., Abdubek, P. et al. (2010) Structure of the first representative of Pfam family PF04016 (DUF364) reveals enolase and Rossmann-like folds that combine to form a unique active site with a possible role in heavy-metal chelation. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **66**: 1167-1173.

Miller, W.G., and Lindow, S.E. (1997) An improved GFP cloning cassette designed for prokaryotic transcriptional fusions. *Gene* **191**: 149-153.

Młynarczyk, A., Młynarczyk, G., Pupek, J., Bilewska, A., Kawecki, D., Łuczak, M. et al. (2007) *Serratia marcescens* isolated in 2005 from clinical specimens from patients with diminished immunity. *Transplantation Proceedings* **39**: 2879-2882.

Muffler, A., Fischer, D., and Hengge-Aronis, R. (1996) The RNA-binding protein HF-I, known as a host factor for phage Qbeta RNA replication, is essential for *rpoS* translation in *Escherichia coli*. *Genes Dev* **10**: 1143-1151.

Mukherjee, A., Cui, Y., Ma, W., Liu, Y., and Chatterjee, A.K. (2000) *hexA* of *Erwinia carotovora* ssp. *carotovora* strain Ecc71 negatively regulates production of RpoS and *rsmB* RNA, a global regulator of extracellular proteins, plant virulence and the quorum-sensing signal, *N*-(3-oxohexanoyl)-L-homoserine lactone. *Environ Microbiol* **2**: 203-215.

Mukherjee, A., Cui, Y., Ma, W., Liu, Y., Ishihama, A., Eisenstark, A., and Chatterjee, A.K. (1998) RpoS (Sigma-S) controls expression of *rsmA*, a global regulator of secondary metabolites, harpin, and extracellular proteins in *Erwinia carotovora*. *J Bacteriol* **180**: 3629-3634.

Ng, W., and Bassler, B. (2009) Bacterial Quorum-Sensing Network Architectures. *Annu Rev Genet*.

Olsen, A.S., Møller-Jensen, J., Brennan, R.G., and Valentin-Hansen, P. (2010) C-Terminally Truncated Derivatives of *Escherichia coli* Hfq Are Proficient in Riboregulation. *J Mol Biol*.

Papenfort, K., and Vogel, J. (2010) Regulatory RNA in bacterial pathogens. *Cell Host Microbe* **8**: 116-127.

Papenfort, K., Said, N., Welsink, T., Lucchini, S., Hinton, J.C.D., and Vogel, J. (2009) Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. *Mol Microbiol* **74**: 139-158.

Park, D.H., Thapa, S.P., Choi, B.-S., Kim, W.-S., Hur, J.H., Cho, J.M. et al. (2011) Complete genome sequence of Japanese *erwinia* strain ejp617, a bacterial shoot blight pathogen of pear. *J Bacteriol* **193**: 586-587.

Peterson, C.N., Carabetta, V.J., Chowdhury, T., and Silhavy, T.J. (2006) LrhA regulates *rpoS* translation in response to the Rcs phosphorelay system in *Escherichia coli*. *J Bacteriol* **188**: 3175-3181.

Poulter, S., Carlton, T.M., Spring, D.R., and Salmond, G.P.C. (2011) The *Serratia* LuxR family regulator CarR(39006) activates transcription independently of cognate quorum sensing signals. *Mol Microbiol*.

Pradel, E., and Ewbank, J.J. (2004) Genetic models in pathogenesis. *Annu Rev Genet* **38**: 347-363.

Pradel, E., Zhang, Y., Pujol, N., Matsuyama, T., Bargmann, C.I., and Ewbank, J.J. (2007) Detection and avoidance of a natural product from the pathogenic bacterium *Serratia* marcescens by *Caenorhabditis elegans*. *Proc Natl Acad Sci USA* **104**: 2295-2300.

Pujol, N., Link, E.M., Liu, L.X., Kurz, C.L., Alloing, G., Tan, M.W. et al. (2001) A reverse genetic analysis of components of the Toll signaling pathway in *Caenorhabditis elegans*. *Curr Biol* **11**: 809-821.

Rehmsmeier, M., Steffen, P., Hochsmann, M., and Giegerich, R. (2004) Fast and effective prediction of microRNA/target duplexes. *RNA (New York, NY)* **10**: 1507-1517.

Robertson, G.T., and Roop, R.M. (1999) The *Brucella abortus* host factor I (HF-I) protein contributes to stress resistance during stationary phase and is a major determinant of virulence in mice. *Mol Microbiol* **34**: 690-700.

Sambrook, J., F. Fritsch, E., and Maniatis, T. (1989) Molecular cloning: a laboratory manual. 1659.

Sandkvist, M. (2001) Type II secretion and pathogenesis. *Infect Immun* **69**: 3523-3535.

Sauter, C., Basquin, J., and Suck, D. (2003) Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*. *Nucleic Acids Res* **31**: 4091-4098.

Schiano, C.A., Bellows, L.E., and Lathem, W.W. (2010) The small RNA chaperone Hfq is required for the virulence of *Yersinia pseudotuberculosis*. *Infect Immun* **78**: 2034-2044.

Seaver, L.C., and Imlay, J.A. (2004) Are Respiratory Enzymes the Primary Sources of Intracellular Hydrogen Peroxide? *J Biol Chem* **279**: 48742-48750.

Seth-Smith, H.M.B. (2008) SPI-7: *Salmonella*'s Vi-encoding Pathogenicity Island. *J Infect Dev Ctries* **2**: 267-271.

Shakhnovich, E.A., Davis, B.M., and Waldor, M.K. (2009) Hfq negatively regulates type III secretion in EHEC and several other pathogens. *Mol Microbiol* **74**: 347-363.

Shapira, M., and Tan, M.-W. (2008) Genetic analysis of *Caenorhabditis elegans* innate immunity. *Methods Mol Biol* **415**: 429-442.

Sharma, A.K., and Payne, S.M. (2006) Induction of expression of *hfq* by DksA is essential for *Shigella flexneri* virulence. *Mol Microbiol* **62**: 469-479.

Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A. et al. (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**: 250-255.

Simonsen, K.T., Nielsen, G., Bjerrum, J.V., Kruse, T., Kallipolitis, B.H., and Møller-Jensen, J. (2011) A Role for the RNA Chaperone Hfq in Controlling Adherent-Invasive *Escherichia coli* Colonization and Virulence. *PLoS ONE* **6**: e16387.

Sittka, A., Pfeiffer, V., Tedin, K., and Vogel, J. (2007) The RNA chaperone Hfq is essential for the virulence of *Salmonella typhimurium*. *Mol Microbiol* **63**: 193-217.

Sittka, A., Lucchini, S., Papenfort, K., Sharma, C.M., Rolle, K., Binnewies, T.T. et al. (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* **4**: e1000163.

Slater, H., Crow, M., Everson, L., and Salmond, G.P.C. (2003) Phosphate availability regulates biosynthesis of two antibiotics, prodigiosin and carbapenem, in *Serratia* via both quorum-sensing-dependent and -independent pathways. *Mol Microbiol* **47**: 303-320.

Soper, T., Mandin, P., Majdalani, N., Gottesman, S., and Woodson, S.A. (2010) Positive regulation by small RNAs and the role of Hfq. *Proc Natl Acad Sci USA* **107**: 9602-9607.

Soper, T.J., and Woodson, S.A. (2008) The *rpoS* mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. *RNA* **14**: 1907-1917.

Sorek, R., and Cossart, P. (2009) Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat Rev Genet*.

Sousa, S.A., Ramos, C.G., Moreira, L.M., and Leitão, J.H. (2010) The *hfq* gene is required for stress resistance and full virulence of *Burkholderia cepacia* to the nematode *Caenorhabditis elegans*. *Microbiology* **156**: 896-908.

Stock, A.M., Robinson, V.L., and Goudreau, P.N. (2000) Two-component signal transduction. *Annu Rev Biochem* **69**: 183-215.

Sun, X., Zhulin, I., and Wartell, R.M. (2002) Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res* **30**: 3662-3671.

Takayanagi, Y., Tanaka, K., and Takahashi, H. (1994) Structure of the 5' upstream region and the regulation of the *rpoS* gene of *Escherichia coli*. *Mol Gen Genet* **243**: 525-531.

Tan, M.W., Mahajan-Miklos, S., and Ausubel, F.M. (1999a) Killing of *Caenorhabditis elegans* by *Pseudomonas aeruginosa* used to model mammalian bacterial pathogenesis. *Proc Natl Acad Sci USA* **96**: 715-720.

Tan, M.W., Rahme, L.G., Sternberg, J.A., Tompkins, R.G., and Ausubel, F.M. (1999b) *Pseudomonas aeruginosa* killing of *Caenorhabditis elegans* used to identify *P. aeruginosa* virulence factors. *Proc Natl Acad Sci USA* **96**: 2408-2413.

Tanikawa, T., Nakagawa, Y., and Matsuyama, T. (2006) *Transcriptional downregulator hexS controlling prodigiosin and serrawettin W1 biosynthesis in Serratia marcescens*.

Tenor, J.L., and Aballay, A. (2008) A conserved Toll-like receptor is required for *Caenorhabditis elegans* innate immunity. *EMBO Rep* **9**: 103-109.

Thomson, N.R., Crow, M.A., McGowan, S.J., Cox, A., and Salmond, G.P. (2000) Biosynthesis of carbapenem antibiotic and prodigiosin pigment in *Serratia* is under quorum sensing control. *Mol Microbiol* **36**: 539-556.

Thomson, N.R., Cox, A., Bycroft, B.W., Stewart, G.S., Williams, P., and Salmond, G.P. (1997) The rap and hor proteins of *Erwinia*, *Serratia* and *Yersinia*: a novel subgroup in a growing superfamily of proteins regulating diverse physiological processes in bacterial pathogens. *Mol Microbiol* **26**: 531-544.

Tsai, C.-S., and Winans, S.C. (2010) LuxR-type quorum-sensing regulators that are detached from common scents. *Mol Microbiol* **77**: 1072-1082.

Tsui, H.C., Leung, H.C., and Winkler, M.E. (1994) Characterization of broadly pleiotropic phenotypes caused by an *hfq* insertion mutation in *Escherichia coli* K-12. *Mol Microbiol* **13**: 35-49.

Tsui, H.C., Feng, G., and Winkler, M.E. (1996) Transcription of the *mutL* repair, *miaA* tRNA modification, *hfq* pleiotropic regulator, and *hflA* region protease genes of *Escherichia coli* K-12 from clustered Esigma32-specific promoters during heat shock. *J Bacteriol* **178**: 5719-5731.

Updegrove, T., Wilf, N., Sun, X., and Wartell, R.M. (2008) Effect of Hfq on RprA-rpoS mRNA pairing: Hfq-RNA binding and the influence of the 5' *rpoS* mRNA leader region. *Biochemistry* **47**: 11184-11195.

Updegrove, T.B., Correia, J.J., Chen, Y., Terry, C., and Wartell, R.M. (2011) The stoichiometry of the *Escherichia coli* Hfq protein bound to RNA. *RNA (New York, NY)*.

Venkatesh, B., Babujee, L., Liu, H., Hedley, P., Fujikawa, T., Birch, P. et al. (2006) The *Erwinia chrysanthemi* 3937 PhoQ sensor kinase regulates several virulence determinants. *J Bacteriol* **188**: 3088-3098.

Vining, L.C. (1992) Secondary metabolism, inventive evolution and biochemical diversity--a review. *Gene* **115**: 135-140.

Vogel, J. (2009) A rough guide to the non-coding RNA world of *Salmonella*. *Mol Microbiol* **71**: 1-11.

Williamson, N.R., Fineran, P.C., Leeper, F.J., and Salmond, G.P.C. (2006) The biosynthesis and regulation of bacterial prodiginines. *Nat Rev Micro* **4**: 887-899.

Williamson, N.R., Fineran, P.C., Ogawa, W., Woodley, L.R., and Salmond, G.P.C. (2008) Integrated regulation involving quorum sensing, a two-component system, a GGDEF/EAL domain protein and a post-transcriptional regulator controls swarming and RhlA-dependent surfactant biosynthesis in *Serratia*. *Environ Microbiol* **10**: 1202-1217.

Williamson, N.R., Fineran, P.C., Gristwood, T., Chawrai, S.R., Leeper, F.J., and Salmond, G.P.C. (2007) Anticancer and immunosuppressive properties of bacterial prodiginines. *Future microbiology* **2**: 605-618.

Williamson, N.R., Simonsen, H.T., Ahmed, R.A.A., Goldet, G., Slater, H., Woodley, L. et al. (2005) Biosynthesis of the red antibiotic, prodigiosin, in *Serratia*: identification of a novel 2-methyl-3-n-amyl-pyrrole (MAP) assembly pathway, definition of the terminal condensing enzyme, and implications for undecylprodigiosin biosynthesis in *Streptomyces*. *Mol Microbiol* **56**: 971-989.

Winkler, W.C., and Breaker, R.R. (2005) Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol* **59**: 487-517.

Worrall, J.A.R., Górna, M., Crump, N.T., Phillips, L.G., Tuck, A.C., Price, A.J. et al. (2008) Reconstitution and analysis of the multienzyme *Escherichia coli* RNA degradosome. *J Mol Biol* **382**: 870-883.

Zaborin, A., Romanowski, K., Gerdes, S., Holbrook, C., Lepine, F., Long, J. et al. (2009) Red death in *Caenorhabditis elegans* caused by *Pseudomonas aeruginosa* PAO1. *Proc Natl Acad Sci USA* **106**: 6327-6332.

Zhang, X.-H., and Austin, B. (2005) Haemolysins in *Vibrio* species. *J Appl Microbiol* **98**: 1011-1019.

Zhang, X.L., Tsui, I.S., Yip, C.M., Fung, A.W., Wong, D.K., Dai, X. et al. (2000) *Salmonella enterica* serovar Typhi uses type IVB pili to enter human intestinal epithelial cells. *Infect Immun* **68**: 3067-3073.

Zhang, Y. (2008) Neuronal mechanisms of *Caenorhabditis elegans* and pathogenic bacteria interactions. *Curr Opin Microbiol* **11**: 257-261.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406-3415.