The exploration of small RNA regulation in *Pasteurella multocida*

Emily Laura Gulliver

Department of Microbiology

Infection and Immunity Program, Monash Biomedicine Discovery Institute

Monash University

Thesis submitted for the degree of Doctor of Philosophy at Monash University,

Victoria, Melbourne, Australia

January 2019

Table of Contents

Summary

Pasteurella multocida is a Gram-negative bacterium responsible for many important animal diseases. While a number of *P. multocida* virulence factors have been identified, very little is known about how gene expression and protein production is regulated in this organism. One mechanism by which bacteria regulate transcript abundance and protein production is riboregulation, which involves a specific interaction between a small RNA (sRNA) and an mRNA that acts to alter transcript stability and/or translational efficiency. In this study, transcriptomic analysis of the *P. multocida* strain VP161 revealed a putative sRNA with high identity to GcvB from *Escherichia coli* and *Salmonella enterica* serovar Typhimurium. High-throughput quantitative liquid proteomics was used to compare the proteomes of the *P. multocida* VP161 wild-type strain, a *gcvB* mutant and a GcvB overexpression strain. These analyses identified 47 proteins that displayed significant differential production after inactivation of *gcvB*, 37 of which showed increased production. Thus, GcvB predominantly acts to negatively regulate protein production in *P. multocida*. Of the 37 proteins that were repressed by GcvB, 27 were predicted to be involved in amino acid biosynthesis or transport. Bioinformatic analyses of putative *P. multocida* GcvB target mRNAs identified a strongly conserved 10 nucleotide consensus sequence, 5'-AACACAACAT-3', with the central eight nucleotides identical to the seed binding region present within GcvB mRNA targets in *E. coli* and *S.* Typhimurium. Using a defined set of seed region mutants, together with a two-plasmid reporter system, this sequence was confirmed to be critical for the binding of the *P. multocida* GcvB to the target mRNA, *gltA*, and the reduction in GltA production.

Hfq is a well-characterized RNA chaperone protein that is involved in bacterial riboregulation. Recently, a second RNA chaperone called ProQ was shown to play a critical role in stabilizing some sRNA/mRNA interactions. To assess the role of *P. multocida* ProQ in riboregulation, we used several high-throughput analyses, including proteomics, transcriptomics and UV-crosslinking, ligation, and sequencing of hybrids (UV-CLASH) to identify transcripts that may be bound and regulated by ProQ. These analyses identified that ProQ binds to, and stabilises, sRNA molecules but also shows strong binding to tRNAs. Two putative sRNA transcripts were identified that bound to ProQ, namely, Prrc13 and PMVP_0063. Both transcripts were stabilized by ProQ and bound to other RNA targets whilst bound to ProQ, including targets essential for the growth of *P. multocida* strain VP161; Prrc13 bound to adk, encoding adenylate kinase, and PMVP_0063 bound to *ftsH*, encoding an ATP-dependent zinc metallopeptidase. This indicates that these putative sRNAs are ideal targets for the development of therapeutic agents against *P. multocida* as targeting these sRNAs could lead to the dysregulation of essential genes.

General declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes one original paper published in a peer reviewed journal. The core theme of the thesis is small RNA regulation in *Pasteurella multocida*. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the Microbiology department under the supervision of Assoc. Prof. John Boyce and Dr Marina Harper.

The inclusion of the following co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research. In Chapter 2, Northern blotting was performed by Amy Wright, initial transcriptomic analyses were performed by Dr Deanna Deveson Lucan and Hfq co-immunoprecipitation was performed by Marianne Mégroz. For proteomics, sample preparation was performed by this candidate under the guidance of Amy Wright, mass spectrometry and initial data analysis was performed by Dr Oded Kleifeld and Dr Ralf B. Schittenhelm (Monash University Proteomics Platform). Final analyses of the proteomics data were performed by this candidate and David Powell (Monash University Bioinformatics Platform). In Chapter 3, UV-CLASh experiments were performed with the assistance of Dr Julia Wong and Brandon Sy, and initial bioinformatic analyses of these experiments was performed by Dr Jai Tree. All other laboratory work, including all other bioinformatic analyses was performed by this candidate. The PhD thesis and publication were prepared by this candidate, with major editing and concept input performed by the candidate's PhD supervisors, Dr Marina Harper and Assoc. Prof. John Boyce. Minor edits of the GcvB manuscript (Chapter 2) were performed by Dr Torsten Seemann and Dr Jürgen B. Bulitta.

Thesis Chapter	Publication Title	Status	Nature and % of student contribution	Co-author name(s) Nature and % of Co-author's contribution	Co-author(s), Monash student?
2	*Determination of the small RNA GcvB regulon in the Gram-negative bacterial pathogen Pasteurella multocida and identification of the GcvB seed binding region*	*Published*	*59%. Laboratory work (58%), concept, and writing first draft (60%)*	***Laboratory work*** *Amy Wright, laboratory work and assistance 15%* *Deanna Deveson Lucas, Laboratory work 5%* *Marianne Mégroz, Laboratory work 5%* *Oded Kleifeld, Experimental platform assistance and data analysis 6%* *Ralf B. Schittenhelm, experimental platform assistance and data analysis 6%* *David R. Powell, data analysis 5%* ***Manuscript preparation*** *Torsten Seemann paper editing 5%* *Jürgen B. Bulitta, paper editing 5%* *Marina Harper, concept and paper editing, 15%* *John D. Boyce, concept and paper editing 15%*	 *No* *No* *Yes* *No* *No* *No* *No* *No* *No* *No*

I have renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Student signature: *Emily Gulliver* **Date:** 04/01/2019

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor signature: [redacted] **Date:** 29/12/2018

Acknowledgements

People always warn you that doing a PhD is tough and that you will feel so low and disheartened, and that most people hate being in science by the time they're done. What people forget to tell you is that doing a PhD can be fun, and when things work it makes you feel like it's all worth it. I have had such an amazingly fun and exciting few years and it wouldn't be possible without all the people around me.

Firstly, I would like to thank John and Marina, because without them I wouldn't have a project at all. You have both encouraged me and guided me and have allowed me to grow into the scientist that I am now. John your compliment sandwiches and overly fancy rewording of all of my work have provided me with a lot of laughs, and Marina your determination to have my work perfect may have driven me crazy but you have taught me so much and your zany ways never cease to make me laugh.

Next, I would like to thank my Boycian family. This includes (in no particular order), Dee, Amy, Marianne, Jess, Michael, Tom, Tim, Caitlin, and Vicki. Each of you have made me laugh, helped me out, and just been there as friends as well as a helpful scientific team. When I first started in the lab I would never have imagined that the Boycians would be such a loving and tight-knit crew. Also, I better still be invited to game boards and morning teas once I'm gone!

I would also like to thank my other lab family, the LaboraTree/ Lynch mob. When I came to Sydney I was so worried, but you helped me to put my fears aside and welcomed me with open arms. Your help and guidance were invaluable as were your sight-seeing tours. I'm looking forward to the Lynch mob reunion in Canada!

My family and friends also deserve a huge thank-you. Although they live hours away from me, my parents and sister have always been there for a chat. I would like to thank you for supporting me, even when what I was doing made no sense to you and getting a real job seemed like a better option. My friends have been such a huge help during my PhD, whether they were giving advice or just being around for a fun time, I really appreciate it.

There are many others that deserve thanks, including the whole Microbiology department and the level 2 crew, particularly the PhD student cohort and Sherrie, it's really nice to have a group of people around who just want to share and help each other.

I'm sure there are other people who should be thanked, so if you were around and helped me learn something new or we just had a chat, thank-you.

Publications and conference proceedings

Publications

Gulliver EL, Wright A, Deveson Lucas D, Mégroz M, Kleifeld O, Schittenhelm R, Powell D, Seemann T, Bulitta J, Harper M et al. 2018. Determination of the small RNA GcvB regulon in the Gram-negative bacterial pathogen *Pasteurella multocida* and identification of the GcvB seed binding region. RNA **24**: 704-720. [Published]

Conference proceedings and presentations

(POSTER) "The role of the GcvB sRNA in the regulation of protein production in *Pasteurella multocida.",* presented at BacPath13: Molecular analysis of bacterial pathogens, (2015), Phillip Island, Australia

(POSTER) "The role of the GcvB sRNA in the regulation of protein production in *Pasteurella multocida.",* presented at Victorian Infection and Immunity Young Investigator Symposium (2015), Melbourne, Australia

(ORAL) "The role of the GcvB sRNA in the regulation of protein production in *Pasteurella multocida.",* presented at Monash University Microbiology Postgraduate Student Conference (2016), Melbourne, Australia

(ORAL) "The role of the GcvB sRNA in the regulation of protein production in *Pasteurella multocida."* presented at meRNA Monash University (2016), Melbourne, Australia

(POSTER) "The role of the GcvB sRNA in the regulation of protein production in *Pasteurella multocida.",* presented at Victorian Infection and Immunity Young Investigator Symposium (2016), Melbourne, Australia

(ORAL) "The role of the GcvB sRNA in the regulation of protein production in *Pasteurella multocida.",* presented at ASM-BD student awards night (2016), Melbourne, Australia

(ORAL) "The role of the RNA chaperone ProQ in *Pasteurella multocida.",* presented at BacPath14: Molecular analysis of bacterial pathogens (2017), Hahndorf, Australia

(ORAL) "The role of the RNA chaperone ProQ in *Pasteurella multocida.",* presented at Northwest Microbiology conference (2018), Manchester, United Kingdom

(POSTER) "The role of the RNA chaperone ProQ in *Pasteurella multocida.",* presented at RNA regulation in bacteria and archaea (2018), Seville, Spain

(ORAL) "The role of the RNA chaperone ProQ in *Pasteurella multocida.",* presented at BDI graduate student symposium (2018), Melbourne, Australia- Winner of the Merck showcase presentation

(ORAL) "The role of the RNA chaperone ProQ in *Pasteurella multocida",* presented at ASM student travel awards night (2018), Melbourne, Australia

(ORAL) "The role of the RNA chaperone ProQ in *Pasteurella multocida."* presented at meRNA Monash University (2018), Melbourne, Australia

(POSTER) "The role of the RNA chaperone ProQ in *Pasteurella multocida."* presented at ASM (2018), Brisbane, Australia

Chapter 1

Introduction

Chapter 1: Introduction

1.1 Introduction

Pasteurella multocida is a Gram-negative, encapsulated, coccobacillus that is commonly found as part of the natural flora in the respiratory tract of many animals, but it can also cause serious disease in a wide variety of animals and birds, including poultry, ungulates, swine, cats, dogs and humans (Wilkie et al. 2012). *P. multocida* strains can be differentiated according to the type of capsule and lipopolysaccharide (LPS) they produce using serological methods and more recently, PCR typing. Strains are classically differentiated into serogroups according to capsule type, A, B, C, E or F using a passive hemagglutination test (Carter 1952) and into LPS serotypes using a gel diffusion immunoprecipitation test (Heddleston and Rebers 1972). More recently, the classification of *P. multocida* strains into capsule type is undertaken using multiplex polymerase chain reaction (PCR) using primers specific for the five different capsule biosynthetic loci (Townsend et al. 2001). LPS typing is now conducted using a multiplex LPS genotyping system using primers that can identify all eight LPS assembly loci, L1 to L8 (Harper et al. 2015). *P. multocida* is the primary cause of many diseases including fowl cholera and haemorrhagic septicaemia but can also act as an opportunistic pathogen in a susceptible host following infection with another pathogen (Wilkie et al. 2012). These opportunistic infections are often involved in certain forms of soft tissue necrosis, gangrene, and upper respiratory tract diseases (Wilkie et al. 2012).

Fowl cholera, or avian cholera, is a systemic infection of poultry and other birds. Signs of disease include rhinitis, cough, and severe diarrhoea (Saif 2008). The severity of fowl cholera can range from acute to chronic, with more virulent strains of *P. multocida* able to cause rapid death (Saif 2008). Most cases of fowl cholera are caused by *P. multocida* strains belonging to capsular serogroup A, but strains belonging to capsular serogroup F and D have also been isolated from infected poultry (Wilkie et al. 2012). The disease may be transmitted via shedding from infected animals, or through contact with contaminated water, diseased carcasses or rodents (Mbuthia et al. 2008). *P. multocida* enters the host through the nasal and/or oral passages and colonises the upper respiratory tract. Following invasion into the bloodstream, *P. multocida* can spread systemically; the exact mechanisms this bacterium uses for invasion and spread is poorly understood. Systemic infection can lead to significant mortality, with lesions often found in the lungs and liver. Joints may also be affected, leading to arthritis (Rhoades and Rimler 1990).

Haemorrhagic septicaemia is a disease of ungulates, particularly cattle and buffalo (De Alwis 1999). It involves the systemic spread of *P. multocida* leading to inflammation and tissue necrosis (Carter and De Alwis 1989; Wilkie et al. 2012). Haemorrhagic septicaemia is typically transmitted by shedding from a

carrier animal; the bacterium then enters the upper respiratory tract of the new host to begin infection (De Alwis 1999). Cattle left untreated show fever, nasal discharge, swelling of the lower jaw, laboured breathing and excess salivation. In severe cases animals develop pneumonic pasteurellosis which has a 100% mortality rate (Wilkie et al. 2012). Haemorrhagic septicaemia is almost always caused by *P. multocida* strains with a type B or a type E capsule (Catry et al. 2005). In 2015, over 200,000 saiga antelopes died of haemorrhagic septicaemia following infection with *P. multocida* type B strain (Kock et al. 2018). *P. multocida* is found as a part of the normal oropharyngeal flora of these antelope, but it was hypothesized that, due to rising temperatures and humidity in the region, the bacterium flourished leading to a mass mortality event. These antelope are a critically endangered species and events such as this increase the likelihood of extinction (Kock et al. 2018).

Lower respiratory tract infections, or pneumonia, in ungulates can be caused by a variety of microorganisms, including *P. multocida.* This group of respiratory diseases is often called 'shipping fever' as the infection is often contracted and/or disease signs become apparent during transportation when animals are in very close proximity to each other (Storz et al. 2000). *P. multocida* strains with a type A capsule are the most common cause, but strains belonging to capsular group D and F may also cause this disease (Bethe et al. 2009). Shipping fever caused by *P. multocida* often follows initial colonisation of the lower respiratory tract by *Mycoplasma spp.* (Ciprian et al. 1988). In less severe cases, pulmonary macrophages may be able to clear the infection but in more severe cases, inflammation and tissue necrosis occurs and exudate is aspirated deep into the bronchi leading to respiratory failure and death (Wilkie et al. 2012).

Atrophic rhinitis is a self-limiting, upper respiratory tract disease caused by *P. multocida* capsular type D strains that produce Pasteurella Multocida Toxin (PMT) (Wilkie et al. 2012). The disease is most commonly seen in pigs, but has also been observed in rabbits, wild boar, and goats (Baalsrud 1987; DiGiacomo et al. 1991). Although *P. multocida* colonises the swine nasal mucosa, the levels are quite low. However, in the presence of the commensal/opportunistic pathogenic species *Bordetella bronchiseptica*, *P. multocida* has been shown to colonise at much higher levels (Rutter and Rojas 1982). The pathogenesis of this disease is driven by the causative agent, PMT, which enters host cells in the nasal cavity via endocytosis then activates downstream signalling cascades, causing a disruption of osteogenesis and increasing the amount of osteoclasts (Mullan and Lax 1998). Clinical signs of PMT intoxication in young pigs include facial distortion, turbinate bone destruction and retarded growth due to decreased feeding (Dominick and Rimler 1988; Martineau-Doize et al. 1991). The snout deformities caused by the action of PMT lead to an

abnormal opening up of the airways. It is thought that this allows more foreign bodies into the respiratory tract as animals suffering from atrophic rhinitis have an increased incidence of lower respiratory tract pneumonias (Martineau-Doize et al. 1990; Wilkie et al. 2012). Atrophic rhinitis has a high mortality, but also a high morbidity, as the snouts of surviving pigs are usually permanently deformed (Horiguchi 2012; Wilkie et al. 2012).

Snuffles is an upper respiratory tract infection caused by *P. multocida* that is problematic for farmed and domesticated rabbits (DiGiacomo et al. 1991). Signs of disease include chronic exudative rhinitis, coughing, fever and respiratory difficulty (Deeb et al. 1990; Guo et al. 2012). Without treatment the bacteria can be aspirated into the lower respiratory tract and cause serious pneumonia and sometimes death (DiGiacomo et al. 1991; Wilkie et al. 2012). Snuffles can be caused by *P. multocida* strains belonging to capsular type A and D and it is also common for opportunistic pathogens such as *Bordetella bronchiseptica* to be involved (Deeb et al. 1990; Wilkie et al. 2012). *S*nuffles is transferred by close animal contact and, as rabbits are usually tightly housed, treatment and prevention strategies should include the isolation of infected animals (Deeb et al. 1990; Wilkie et al. 2012).

In humans, *P. multocida* can cause localised wound infections following bites or scratches from cats or dogs that carry the bacteria as part of their normal nasopharyngeal flora (Dewhirst et al. 2012; Dewhirst et al. 2015). Sometimes, these infections progress to septicaemia, with very serious consequences. The bacterium is usually acquired by people who live closely with dogs or cats as pets (Talley et al. 2016). In some cases, *P. multocida* has caused systemic infection in immunocompromised patients who have been bitten, scratched or have had a surgical/open wound licked by a pet. In one study, more than 30 patients who had received a prosthetic joint replacement became infected with *P. multocida* following animals licking or contacting the surgical wound (Honnorat et al. 2016). Another study concluded that the majority of non-bite-wound associated *P. multocida* infections occurred in patients who had pets, indicating that animal contact was the likely source of the infection. Non-bite-wound associated *P. multocida* infections were more likely to progress to septicaemia, with mortality rates as high as 21% (Giordano et al. 2015). Interestingly, severe cases of septicaemia caused by *P. multocida* have been documented in patients who have not had contact with such animals. In Minnesota in 2014, over a short period of time, five people presented to hospital with *P. multocida* infections and three died (Talley et al. 2016). There was no evidence of transmission of disease between the patients while in hospital and it was hypothesized that each patient had acquired the infection at home, yet only one of the three patients who had died from the infection had any known animal contact (Talley et al. 2016).

1.2 Virulence factors

In order to survive and cause disease in a variety of host animals, including humans, *P. multocida* must control the temporal expression of a number of virulence factors. Some of the most well characterized of these include capsular polysaccharide that shrouds the bacterial cell, the highly variable LPS, various cell surface components such as adhesins and fimbriae, and the PMT toxin produced by *P. multocida* serogroup D strains that cause atrophic rhinitis (Harper et al. 2006).

1.2.1 Capsule

There are five different capsular serogroups, designated A, B, D, E, and F, produced by different *P. multocida* strains. Almost all virulent strains produce a capsule that protects the bacterium from the host innate immune system (Boyce and Adler 2000). Each capsule type consists of a structurally distinct polysaccharide (Carter 1967; Rimler and Rhoades 1987; Boyce et al. 2000). Type A capsules are comprised of hyaluronic acid (Rosner et al. 1992), which is also present in abundance in the host. Although there is a clear role for the type A capsule in preventing the action of the innate immune system, the role of this capsule in adhesion to host cells is unclear. Hyaluronic acid capsule was shown to bind to CD44 receptors in the extracellular matrix of host tissues, and this adhesion was predicted to be critical for subsequent systemic dissemination (Pruimboom et al. 1999). However, other studies have shown that the removal of the type A capsule from the bacterial cell surface increases bacterial cell adhesion to the pharyngeal mucosa of rabbits (Glorioso et al. 1982). Type D and F capsules contain the glycosaminoglycans heparin and chondroitin, respectively; both are chemically similar to the major component of type A capsules, hyaluronic acid (Rimler and Rhoades 1987; DeAngelis et al. 2002). Although capsule types B and E have not been structurally characterized, one study reported that galactose, mannose, and arabinose were identified as probable constituents of the serotype B capsule (Muniandy et al. 1992).

The genetic loci responsible for the biosynthesis of all *P. multocida* capsule types have been sequenced and each locus contains three distinct regions (Chung et al. 1998; Townsend et al. 2001). Region 1 encodes components of the ABC transporter system required for transport of the capsular polysaccharides to the outer membrane. Region 2 is the most variable region and encodes the proteins required for the production of the type-specific polysaccharides (Boyce and Adler 2000; Chung et al. 2001). In capsular type A strains, this region contains the *hyaEDCB* genes, which encode the proteins required for hyaluronic acid synthesis. Region 3 contains genes encoding proteins predicted to be required for anchoring the capsule to the cell surface (Boyce and Adler 2000). Genetic studies using capsular type A and B mutant strains of *P. multocida* showed that genes within Region 1 of the capsule locus are essential for capsule production. The type A and B capsule mutants were constructed by disrupting the gene encoding the ATP

transporter component; *hexA* in type A (Chung et al. 2001) and *cexA* in type B (Boyce and Adler 2000). Acapsular mutants of both type A and B strains were highly attenuated for virulence (Boyce and Adler 2000; Chung et al. 2001). The *P. multocida hexA* mutant showed a significant decrease in virulence in mice compared to the wild-type parent strain X73, with a 10^6-fold increase in 50% infective dose (ID_{50}), as well as a 400-fold increased sensitivity to active chicken serum compared to the parent strain (Chung et al. 2001). The type B strain M1404 *cexA* mutant was also highly attenuated for growth (approximately 10^5-fold increase in ID_{50}) and was four to six-fold more sensitive to phagocytosis by mouse peritoneal macrophages (Boyce and Adler 2000). These studies demonstrate unequivocally that the *P. multocida* capsule types A and B are essential for virulence and that they mediate resistance to innate host immunity.

Recent evidence indicates that *P. multocida* capsule has an antagonistic relationship with biofilm formation (Petruzzi et al. 2018). While definitive evidence linking *P. multocida* virulence and biofilms is currently lacking, biofilm formation is predicted to play a role, as it does in many other bacterial species (Krzysciak et al. 2014; Thummeepak et al. 2016). Biofilms can mediate bacterial adherence to biotic and abiotic surfaces and provide bacteria within the biofilm increased resistance to antibiotics and desiccation. Acapsular *hyaE* mutants generated in several different *P. multocida* type A wild-type strains were shown to form measurable and robust biofilms, in contrast to the wild-type parent strains that produced only weak biofilms (Petruzzi et al. 2018). These findings suggest that temporal/spatial regulation of capsule production may be important for *P. multocida* survival *in vivo*. Increased capsule production may be beneficial in some niches for increased resistance to elements of the host innate immune system but detrimental in other niches where biofilm formation and adherence are required for persistence.

1.2.2 LPS

LPS forms the majority of the outer leaflet of the outer membrane of Gram-negative bacterial cells. LPS is comprised of lipid A, which connects the LPS molecule to the outer membrane, an inner core polysaccharide and an outer core polysaccharide. In many Gram-negative bacteria, particularly those belonging to the Enterobacteriaceae family, this is followed by a repeating O-antigen. Like many mucosal Gram-negative pathogens, the *P. multocida* LPS has no O-antigen repeats; this type of LPS is often called lipooligosaccharide or LOS. LPS/LOS is also commonly known as endotoxin as the lipid A component elicits a strong inflammatory response and is highly toxic to humans and animals.

Since the 1970s, *P. multocida* strains have been commonly differentiated by “Heddleston” serology using chicken antibodies raised against LPS purified from the 16 Heddleston type strains. This serological differentiation was presumed to be driven by variations in the LPS structure (Heddleston and Rebers

1972). Recently, the LPS structures of all the *P. multocida* Heddleston type strains have been determined and each type strain does indeed produce one or more distinct LPS molecules (Figure 1.1.) (Harper et al. 2011). Most of the structural variability that is responsible for the antigenic variation observed is located within the outer core region of the LPS. However, almost all *P. multocida* strains examined simultaneously assemble a single outer core structure onto two distinct inner core glycoforms, A and B, that differ in the number of Kdo residues (glycoform A has one, glycoform B has two), the number of glucose residues (glycoform A has two, glycoform B has one) and in the number and position of phosphoethanolamine (PEtn) residues (Harper et al. 2007a). The genes responsible for synthesis of all *P. multocida* LPS structures identified to date have been identified. Those responsible for the outer core assembly are located within a single locus. In total there are eight distinct LPS genotypes (Figure 1.1.) (Harper et al. 2011; Harper et al. 2012, Harper et al. 2014). This is eight fewer than the number of Heddleston serotypes eight of the Heddleston LPS structures produced truncated variants, the result of point mutations in key LPS assembly genes (Harper et al. 2011; Harper et al. 2012, Harper et al. 2014).

The importance of LPS in virulence was first clearly demonstrated through analyses of two separate *P. multocida* signature-tagged mutagenesis (STM) libraries (Fuller et al. 2000; Harper et al. 2003b). Screening of an STM library in a septicaemic mouse model identified many genes essential for virulence including one that encoded a glycosyl transferase with high levels of identity to the *Haemophilus influenzae* transferase LgtC (Fuller et al. 2000). The LgtC transferase in *H. influenzae* adds galactose to a specific position in the LPS. It is also involved in phenotypic switching of the LPS epitope that leads to attenuation of virulence in *H. influenzae* (Hood et al. 1996). Screening of another STM library, this time in both a mouse and a chicken disease model, identified the gene *hptD* in strain VP161 as essential for virulence (Harper et al. 2003b). The HptD transferase was shown to add the third heptose to the LPS inner core (Harper et al. 2003b). Interestingly this mutant, expressing a highly truncated LPS structure, was avirulent in chickens but showed only a slightly decreased virulence in mice, indicating that a full length LPS structure is important for virulence in chickens but not for virulence in mice (Harper et al. 2003b). Further studies on the LPS produced by the *P. multocida* strain VP161 showed that directed inactivation of the transferase gene *pcgC*, essential for the addition of phosphocholine (PCho) to the distal end of the LPS, also resulted in strains with significantly reduced virulence. Chickens infected with 60 colony forming units (cfu) of the *pcgC* mutant took 26 hours longer to show clinical signs of fowl cholera, compared to chickens infected with an equivalent inoculum of the wild-type *P. multocida* parent strain VP161, proving that a complete LPS structure is essential for full virulence of strain VP161 (Harper et al. 2007b).

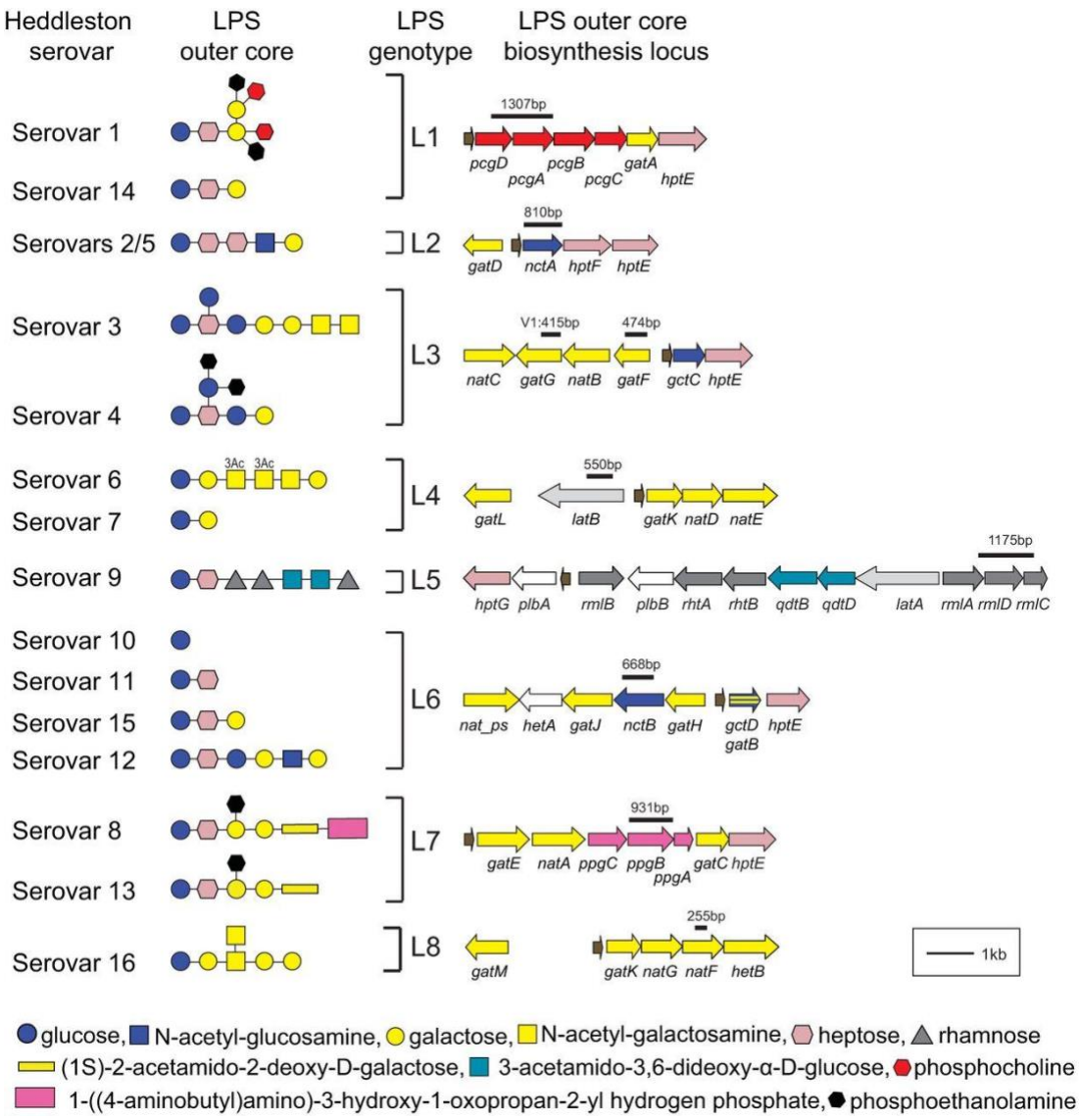

Figure 1.1. LPS outer core structure produced by each of the Heddleston serovar type strains and the genes responsible for LPS outer core biosynthesis in each strain. (Left) Schematic representation of the outer core LPS structures produced by each of the Heddleston serovar type strains. The last residue (glucose) of the conserved LPS inner core is shown on the far left as a reference point. Specific linkages between each of the residues are not shown. (Right) LPS genotype and genetic organization of each LPS outer core biosynthesis locus. The relative position and size of each genotype-specific PCR amplicon are shown above each LPS outer core biosynthesis locus. Each gene is colour coded according to its known/predicted role in LPS biosynthesis; *gctD* and *gatB* (yellow and blue striped) in locus L6 differ by only a single nucleotide and are involved in the addition of glucose or galactose, respectively, to the outer core heptose. The rpL31_2 gene, encoding ribosomal protein L31, is not involved in LPS biosynthesis and is coloured brown. Taken from Harper *et al.* 2015.

1.2.3 Adhesins and fimbriae

Adhesins and fimbriae are important structures found on the surface of many bacterial species. *P. multocida* produces a variety of surface proteins that may play a role in adherence to host cells (Wilkie et al. 2012). These include OmpA, which has been shown to bind to fibronectin and heparin in bovine kidney cells (Dabo et al. 2003), and ComE, which binds to fibronectin (Mullen et al. 2008). The genomes of some *P. multocida* strains contain up to two genes, *pfhB1* and *pfhB2,* which are predicted to encode filamentous haemagglutinin proteins (May et al. 2001). The *P. multocida* filamentous haemagglutinin protein is a known virulence factor as *pfhB* mutants are attenuated for virulence in poultry (Fuller et al. 2000; Tatum et al. 2005). In other species, such as *Bordetella pertussis*, filamentous haemagglutinin is an important outer membrane protein that allows for bacterial adhesion to host cells (Locht et al. 2001). However, the exact mechanism of action of the *P. multocida* filamentous haemagglutinin has not been determined but given the high level of amino acid identity to characterized filamentous haemagglutinins in other species, it is predicted that it also functions as an adhesin.

1.2.4 Pasteurella Multocida Toxin

PMT is a large toxin (146 kDa) that is produced by many serotype D strains and some type A strains of *P. multocida* (Frandsen et al. 1991). PMT is the causative agent of swine atrophic rhinitis; inoculation of pigs with purified PMT can recapitulate disease signs, including decreased weight gain, rough coat and atrophy of the snout (Ackermann et al. 1996). When given at a dose of 0.1 µg of PMT/ kg, the toxin was lethal to all inoculated pigs, and at a dose of 0.05 µg of PMT/ kg, 80% of pigs died (Ackermann et al. 1996). PMT enters host cells via endocytosis and then acts by activating heterotrimeric G proteins via deamination of the α-subunit (Orth et al. 2013). This leads to a dysregulation of several downstream signalling cascades, which results in inflammation, bone destruction (due to an increased number of osteoclasts), rearrangement of the host cell cytoskeleton and increased expression of calcium signalling pathways (Wilson et al. 1997). Consequently, swine affected by PMT develop tissue necrosis and bone deformation, which can eventually lead to death as they no longer feed effectively due to the bending of the snout (Orth and Aktories 2012).

1.2.5 Nutrient acquisition

To ensure survival within a host, invading microorganisms, including Gram-negative bacteria like *P. multocida,* must acquire a variety of nutrients. There are several mechanisms used by *P. multocida* to acquire and breakdown specific nutrients for use in cellular pathways, including enzymes required for the catabolism of hyaluronic acid and sialic acid and importantly, iron acquisition proteins (Rosner et al. 1992; Bosch et al. 2002a; Steenbergen et al. 2005). The acquisition and incorporation of amino acids is also an

important factor that determines the ability of *P. multocida* to grow and colonise the host, and as such it is predicted that the bacterium uses precisely-regulated uptake and biosynthesis pathways for the energy-efficient acquisition of the necessary amino acids (Boyce et al. 2002; Paustian et al. 2002).

P. multocida contains many genes involved in the acquisition of iron (May et al. 2001), which is vital for many bacterial enzymes to function. However, acquiring iron from the host is difficult as the concentration of free iron in the internal tissues of mammalian or avian hosts is very low (Paustian et al. 2001). To overcome this, the *P. multocida* PM70 genome encodes at least nine predicted haemoglobin binding proteins, including PfhR, DppA, PM0741, PM1081, PM1428, HbpA, HgbA, HemR and PM1282 (Bosch et al. 2004). Important iron acquisition proteins include haemoglobin binding protein A (HgbA) (May et al. 2001; Bosch et al. 2002b) and transferrin binding protein (Tbp), which binds transferrin and is produced by many strains of *P. multocida* to scavenge iron from host tissues (Shivachandra et al. 2005). The Tbp protein is commonly found in serogroup A and B strains of *P. multocida,* which infect a range of animals including birds, cattle, buffalo, sheep and goats. One study found that the *tbpA* gene was present in the genomes of more than 50% of the *P. multocida* strains examined (66/131 of animal isolates) (Shivachandra et al. 2005; Shirzad Aski and Tabatabaei 2016).

Hyaluronic acid and sialic acid are abundant on the surface of most host cells (Fraser et al. 1997; Wang and Brand-Miller 2003) and *P. multocida* serogroup A strains produce a hyaluronic acid capsule that is predicted to allow the bacterium to mimic host cells and thus evade detection by the innate immune system (Rosner et al. 1992). Studies on one *P. multocida* strain have also identified sialic acid as an outer surface component (Rosner et al. 1992; Steenbergen et al. 2005). Most strains of *P. multocida* also encode hydrolytic enzymes such as hyaluronidase and sialidases to break down hyaluronic acid and sialic acid, respectively (Carter and Chengappa 1980). Bacterial sialidases cleave sialic acid from a range of glycoproteins and carbohydrates, which can be self-derived or host-derived. The sialic acid may then be added to the bacterial cell surface, where it functions in host immune evasion, or it can be catabolised by the bacterium and used as an energy source (Steenbergen et al. 2005). Hyaluronidase is an enzyme used by many bacterial species to catabolise hyaluronic acid (usually derived from host cells) into hyaluronan for use as a carbon source (Marion et al. 2012). In other bacterial species, hyaluronidase is a key virulence factor as the breakdown of host hyaluronic acid damages host cells (Starr and Engleberg 2006). Although there is limited data on the exact role of these enzymes in *P. multocida*, it is predicted that by producing sialidases and/or hyaluronidases the bacterium can utilize these abundant host carbohydrates for use as

a source of nutrition or for surface decoration/host mimicry, allowing growth and persistence within the host.

As with most Gram-negative bacteria, the acquisition and production of amino acids is important for bacterial survival and growth, particularly during infection. *P. multocida* has several pathways to synthesise amino acids or transport them into the cell and many of the synthesis pathways are complex multi-step reactions (Paustian et al. 2002). This is exemplified by the histidine biosynthetic pathway, where it has been shown in other Gram-negative bacteria that there are at least nine histidine-specific reactions required to produce the amino acid from the starting molecule, fructose-6P. This histidine biosynthetic pathway also has multiple layers of regulation, including transcriptional repressors and specific attenuator sequences (Rossi et al. 2016). The importance of amino acid biosynthesis and transport proteins for *P. multocida* infection of a chicken host has been revealed by comparing the transcriptome of *in vivo* and *in vitro* grown bacteria using microarray (Boyce et al. 2002). These experiments found that ten of the seventeen genes with increased expression during growth in the chicken were involved in amino acid biosynthesis and transport, including *gltA*, *gdhA* and *dppA* (Boyce et al. 2002). Therefore, the acquisition and utilization of amino acids is likely tightly regulated and important for virulence in *P. multocida*.

1.3 Genomics

In 2001, the first complete *P. multocida* genome was published and was predicted to encode a total of 2014 proteins (May et al. 2001). The genome represented the avian isolate Pm70 that is classified as a capsule serogroup F and LPS serotype/genotype 3 strain. Since then, many other full or partial genomes of *P. multocida* strains have been sequenced (in both private and public colections), and many have been used to compare with the fully annotated genomes (Boyce et al. 2012; Moustafa et al. 2015). As of December 2018, *P. multocida* genome assemblies representing 176 strains are publicly available through the National Center for Biotechnology Information, 47 of which are complete. One multi-genome comparison identified a set of 1100 genes that were shared between selected strains and represented the core *P. multocida* genome (Boyce et al. 2012). The number of unique genes in each genome ranged from 73 genes in Pm70 to 474 genes in strain Anand1G (Boyce et al. 2012). Another study compared the genomes of haemorrhagic septicaemia isolates with four publicly available genomes (PM70, 3480, 36950 and HN06) representing isolates from a range of diseases. It found a core genome of 1824 genes and a further 96 genes that were present only in the haemorrhagic septicaemia isolates, indicating these genes may be required for haemorrhagic septicaemia disease (Moustafa et al. 2015). Finally, another study

comparing the genomes of virulent and avirulent strains of *P. multocida* identified 657 genes that were unique to the virulent strains examined and proposed that this set may provide a starting point for determining the subset of genes essential for virulence within *P. multocida* (Peng et al. 2017).

1.4 Gene regulation

Like all organisms, it is expected that the expression of many *P. multocida* genes is tightly regulated. The first annotation of the strain Pm70 genome identified approximately 80 genes that encoded predicted regulatory proteins, of which ten were similar to known two-component signal transduction system proteins in other bacteria and approximately 60 shared identity to known one-component regulatory systems (May et al. 2001). The number of predicted regulatory genes in *P. multocida* is quite small in comparison to the number of regulatory proteins encoded by *E. coli* and *Salmonella* sp*.,* which on average encode more than 300 and 600 regulatory proteins, respectively (May et al. 2001). However, detailed analysis of *P. multocida* regulatory proteins is still in progress and to date, only two regulatory proteins have been characterised, the global regulator Fis and the RNA chaperone protein Hfq (Steen et al. 2010; Mégroz et al. 2016).

1.4.1 Regulation of capsule production by Fis

Fis is a global transcriptional regulatory protein produced by many species of bacteria (Pan et al. 1996). In *E. coli* the amount of Fis produced within the cell is affected by nutrient abundance; during early exponential growth phase in fresh nutrient-rich media the levels of Fis are high (approximately 50, 000 molecules per cell), but during late exponential growth and when nutrients are depleted, the levels of Fis are low (Bradley et al. 2007). In *Pseudomonas sp.,* Fis regulates the expression of virulence factors, including those involved in motility and biofilm formation (Moor et al. 2014). Fis has also been shown to regulate quorum sensing in *Vibrio cholerae* (Lenz and Bassler 2007) and capsule production in *P. multocida* (Steen et al. 2010). The role of Fis in the regulation of *P. multocida* capsular gene expression was initially identified by bioinformatic comparison of the acapsular mutant genome with the parent (strain VP161) genome. A single point mutation in *fis* was identified in the genome of an acapsular mutant and importantly when the mutant was provided with an intact copy of *fis in trans,* capsule production was restored (Steen et al. 2010). This study also showed via microarray analysis (comparing the *fis* mutant with VP161) that Fis abundance in *P. multocida* affected the transcription of at least another 42 genes not involved in capsule synthesis, including the gene that encodes for the virulence-associated protein, filamentous haemagglutinin (Steen et al. 2010). Thus, *P. multocida* Fis is predicted to play a significant role in the regulation of multiple genes, including those encoding virulence factors.

1.4.2 RNA regulation of expression

In addition to transcriptional regulatory proteins, RNA molecules are also important regulators of protein production (Desnoyers et al. 2013). In bacteria, these regulatory RNA molecules include riboswitches, protein-binding RNAs, cis-encoded antisense RNAs, and trans-encoded small regulatory RNAs (hereafter called sRNAs) that are usually 40-400 bp in length (Desnoyers et al. 2013). Cis-encoded RNAs only bind to one mRNA target and generally do so with perfect base pairing. In contrast, trans-encoded sRNAs may be encoded some distance from the target gene and regulate the translation of multiple mRNA targets using only limited base-pair complementarity (Desnoyers et al. 2013). The sRNAs can act in one of several ways (Figure 1.2). Once bound to target mRNAs, a specific sRNA can either block translation, enhance translation, stabilize the mRNA, or induce mRNA degradation (Gottesman and Storz 2011; Desnoyers et al. 2013). As sRNAs generally have a number of mRNA targets, the interaction with each is via a short stretch of imperfect base pairing that usually requires a protein chaperone to facilitate the interaction (Chao and Vogel 2010; Gottesman and Storz 2011). Three chaperone proteins have been identified to date, Hfq, ProQ and CsrA. Each chaperone protein facilitates a specific set of sRNA-mRNA interactions (regulon) and often stabilize the sRNA molecules themselves (Smirnov et al. 2017). Although each chaperone protein is predicted to have its own regulon, some RNAs have been shown to interact with both the Hfq and ProQ chaperones (Smirnov et al. 2016).

1.4.2.1 Hfq: an sRNA chaperone

The importance of Hfq for sRNA action was first recognised in *E. coli* where it was shown that the sRNA OxyS could only function in the presence of a functional Hfq protein (Zhang et al. 1998). Hfq is a small protein that forms a ring-like, homo-hexamer, with three faces (proximal, distal, and rim) that allow for RNA binding (Faner and Feig 2013) (Schu et al. 2015) (Figure 1.3). The proximal face binds uridine rich sequences in sRNA transcriptional terminator sequences; studies using *E. coli* Hfq proximal face missense mutants showed that each mutant had a five- to ten-fold decrease in binding affinity to uridine-rich RNA targets (Sakai et al. 2013). The distal face of Hfq has been shown to bind to adenine rich sequences in RNA molecules, and more recently has been shown to bind to RNA molecules containing an ARN motif, where A is adenine, R represents a purine base and N represents any base (Schu et al. 2015). The third surface of Hfq is the recently characterised rim, which binds to UA-rich regions of RNA (Schu et al. 2015). Hfq-binding sRNAs can be differentiated into two classes, based on the faces of Hfq to which they bind (Figure 1.3). Class 1 sRNAs are those that bind to the proximal and rim faces of Hfq, allowing the mRNA target molecule to bind to the distal face. Class 2 sRNAs bind to the distal and proximal faces of the Hfq protein and the mRNA target binds to the rim face (Schu et al. 2015).

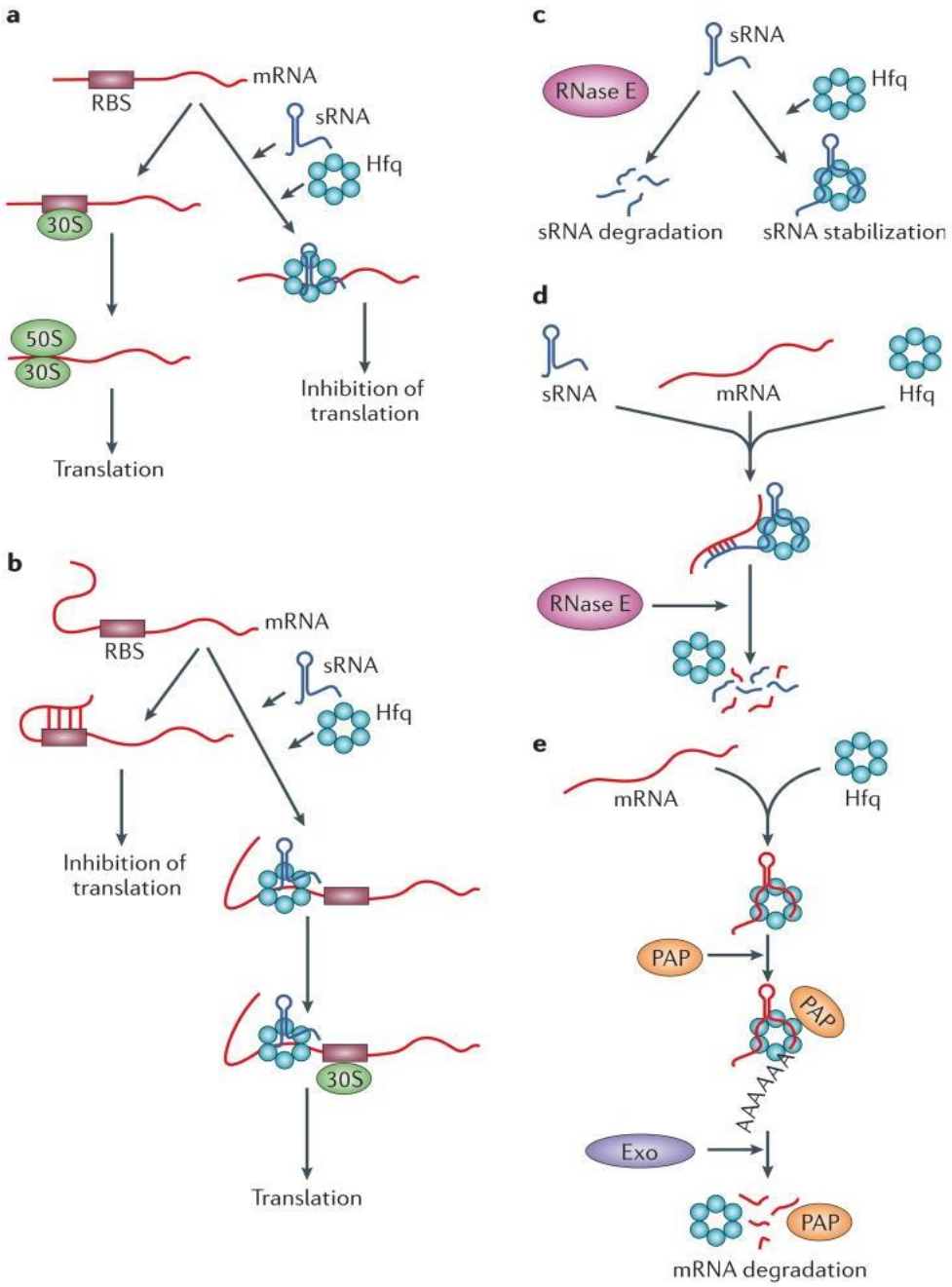

Figure 1.2. Mechanisms of Hfq action. a. Hfq in association with a small RNA (sRNA) may sequester the ribosome-binding site (RBS) of a target mRNA, thus blocking binding of the 30S and 50S ribosomal subunits and repressing translation. **b**. In some mRNAs, a secondary structure in the 5' untranslated region (UTR) can mask the RBS and inhibit translation. A complex formed by Hfq and a specific sRNA may activate the translation of one of these mRNAs by exposing the translation initiation region for 30S binding. **c.** Hfq may protect some sRNAs from ribonuclease cleavage, which is carried out by ribonuclease E (RNase E) in many cases. **d.** Hfq may induce the cleavage (often by RNase E) of some sRNAs and their target mRNAs. **e.** Hfq may stimulate the polyadenylation of an mRNA by poly (A) polymerase (PAP), which in turn triggers 3'-to-5' degradation by an exoribonuclease (Exo). In *Escherichia coli*, the exoribonuclease can be polynucleotide phosphorylase, RNase R or RNase II. Taken from Vogel & Luisi, 2011.

As mentioned above in section 1.4.2 (Figure 1.2), the binding of an Hfq-sRNA complex to its mRNA target can have a number of different outcomes, depending on the specific sRNA to which the Hfq protein binds and the position on the mRNA target where the sRNA binds. Firstly, the Hfq-sRNA complex can bind to a specific mRNA at the Shine Dalgarno sequence, commonly known as the ribosome binding site (RBS), located just upstream of the translation initiation/start codon. This Hfq-sRNA interaction blocks translation by preventing the ribosomal complex from binding to the mRNA (Figure 1.2a) (Vogel and Luisi 2011). Secondly, the Hfq-sRNA complex can bind within the secondary structure of the 5' untranslated region (5' UTR) present in some mRNA molecules that normally acts to occlude the RBS. By binding to the 5' UTR region the secondary structure is relaxed, exposing the RBS and allowing the ribosomal complex to bind and translation to occur (Figure 1.2b) (Vogel and Luisi 2011). Thirdly, Hfq-sRNA complexes can protect the mRNA target from the action of a ribonuclease such as RNase E or alternatively, induce nuclease-mediated degradation/breakdown of the mRNA (Figure 1.2c and 1d respectively). Lastly, in some interactions Hfq can facilitate mRNA degradation, without any prior sRNA interaction, by recruiting poly A polymerase to polyadenylate the mRNA target (Figure 1.2e) (Vogel and Luisi 2011; Faner and Feig 2013).

The regulatory role of Hfq has been examined in a range of bacterial species. In most bacteria studied to date, Hfq helps to regulate the expression of many genes involved in amino acid biosynthesis, amino acid transport, capsule biosynthesis, virulence, and in the response to stress (Chao and Vogel 2010; Sharma et al. 2011). In *E. coli* and *Salmonella* spp*.,* Hfq modulates the binding of more than 150 sRNAs to their mRNA targets, thereby regulating the expression of many proteins (Sittka et al. 2008b; Holmqvist et al. 2016). In *P. multocida*, a *hfq* mutant of the fowl cholera isolate VP161 was recently shown to have decreased capsule production and reduced fitness in mice (Mégroz et al. 2016). Transcriptional analysis revealed the VP161 *hfq* mutant had reduced expression of the genes essential for capsule production (most of the *hya* and *hex* genes) and altered expression of stress tolerance genes (*rpoE* and *rpoH*) and nitrate reductase genes *(nap* locus) (Mégroz et al. 2016). This key study indicated that the Hfq chaperone and associated sRNAs play important roles in regulating capsule production and the stress response in *P. multocida*.

1.4.2.2 ProQ: another important RNA chaperone

An RNA chaperone called ProQ was recently identified in *E. coli*. ProQ acts as a monomer, unlike Hfq which requires a hexamer to be formed, and contains two domains linked by a peptidase-sensitive linker (Smith et al. 2007). The N-terminal domain of ProQ shares identity to FinO-like proteins and has been shown to bind to RNA and to duplex two strands of RNA (Smith et al. 2007). The C-terminal section of ProQ contains a Hfq-like domain. This region also shares structural similarity to Tudor domains; it has a beta sheet

structure and is involved in RNA duplexing and strand exchange. This occurs when the protein catalyses the binding of two RNA molecules to bind to each other through complementary base pairing, followed by an ejection from the protein to allow the next RNA pair to be chaperoned (Smith et al. 2007; Chaulk et al. 2011).

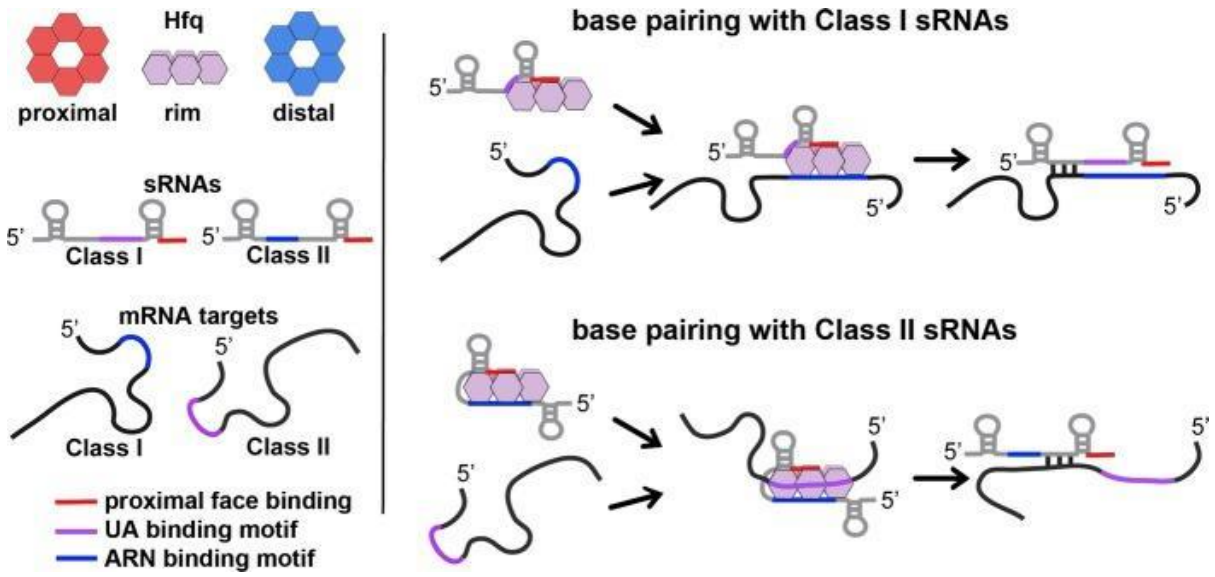

Figure 1.3. A model of alternative modes of RNA binding to Hfq. The cartoon model of the Hfq hexamer depicts the three RNA-binding surfaces of Hfq: proximal face (red), rim (purple), and distal face (blue). For sRNAs and mRNAs, elements in red represent sequences (rho-independent terminator) that bind to the proximal face, elements in purple represent UA sequences that bind the rim, and elements in blue represent ARN-binding motifs that bind to the distal face. The model depicts two alternative pathways for binding and regulation by sRNA-mRNA pairs. Class I sRNAs utilize a U-rich rho-independent terminator for binding the proximal face and a UA-binding motif for interaction with the rim. mRNA targets regulated by this class of sRNAs utilize ARN-binding motifs for interacting with the distal face of Hfq. Binding of the Class I sRNA and its corresponding mRNA to Hfq lead to base pairing and regulation and degradation of the sRNA. A second class of sRNAs (Class II) utilize the U-rich rho-independent terminator for binding the proximal face of Hfq and an ARN motif for binding to the distal face. mRNA targets regulated by this class of sRNAs contain UA-binding motifs that allow for binding to the rim of Hfq. Binding of the Class II sRNA and its corresponding mRNA to Hfq lead to base pairing and regulation, but not necessarily degradation of the sRNA. Taken from Schu et al. (2015)

The *Salmonella* ProQ was identified as an RNA chaperone protein using the Gradient profiling by sequencing (Grad-seq) method, where bacterial cell lysates are first separated on a glycerol gradient then RNA is extracted from different sections of the gradient and sequenced. The separation of some RNA species in the gradient is dependent on their binding to RNA chaperone proteins. RNA species bound to ProQ will cluster together, and separately from RNA species that are unbound, or are bound to other chaperones. Using this method, ProQ was predicted to interact with approximately 400 RNA transcripts, of which 18% were predicted sRNA molecules (Smirnov et al. 2016). Only two of the sRNAs identified also interacted with Hfq, indicating that ProQ has its own subset of sRNA molecules that it interacts with to regulate mRNA target expression. Interestingly, although sRNAs bind to Hfq via specific RNA sequences (see section 1.4.2.1), no specific binding sequences have been identified for ProQ-RNA interactions. Instead, it is thought that ProQ binds to its RNA targets via interactions with highly ordered RNA secondary structures (Smirnov et al. 2016). Other studies have identified that ProQ interacts with the terminator stem-loop of target transcripts which blocks the degradation of the transcript by RNaseIII (Holmqvist et al. 2018). ProQ also differs from Hfq as it binds sRNAs at a 1:1 ratio. The initial interaction is predicted to stretch the linker region of ProQ, allowing for more binding sites to be uncovered and extending the interaction with the RNA (Gonzalez et al. 2017). Although current evidence suggests that some RNA species can be bound to more than one RNA chaperone, evidence in *E. coli* suggests that there is a preference; the mRNA *malM* was able to bind to both Hfq and ProQ but electrophoretic mobility shift assays (EMSA) revealed that when both proteins were present *malM* preferentially bound to ProQ (Gonzalez et al. 2017).

Currently, only one ProQ associated sRNA-mRNA pair, RaiZ-*hupA*, has been characterised in detail. The RaiZ sRNA is generated via a cleavage event within the three prime untranslated region (3' UTR) of the *raiA* mRNA. Interaction with ProQ stabilizes RaiZ enhancing the ability of RaiZ to bind to its mRNA target *hupA,* which encodes the histone-like protein HU that acts in transcriptional regulation (Smirnov et al. 2017). The binding of the RaiZ-ProQ complex to the *hupA* mRNA blocks ribosome binding and leads to decreased HupA production (Smirnov et al. 2017).

Before being recognised as an important RNA chaperone, ProQ was thought to solely regulate the production of the ProP osmotic regulator protein. *E. coli proQ* mutants grown under high salinity conditions show decreased production of ProP, along with decreased biofilm production, decreased growth rate, and cell elongation (Smith et al. 2007; Sheidy and Zielke 2013; Kerr et al. 2014). These phenotypes can be rescued by complementation with *proQ* or via the addition of glycine betaine, which

rehydrates the cells (Kerr et al. 2014). The effects of *proQ* mutation on ProP production have been shown to be independent of the *proP* promoter site, as it was shown that ProQ did not bind to this region of the mRNA. ProQ was shown to bind to ribosomes during translation and it is hypothesised that this binding increases the rate of translation when ProP is being produced (Sheidy and Zielke 2013). However, ProQ can bind to ribosomes when other transcripts are being translated so it is predicted that ProQ also increases the translation of these transcripts (Sheidy and Zielke 2013). Whether an sRNA is involved in ProQ-mediated regulation of ProP is still unclear. One sRNA, Spot42, was identified as a putative activator of ProP production and another sRNA, IsrA, was shown to bind to ProQ. However, the 5' UTR of *proP* (where ProQ is likely to interact) is poorly conserved between species and a conserved sRNA seed binding region could not be identified. This lack of sequence conservation indicates that sRNA-mRNA base pairing is highly unlikely to occur (van Nues et al. 2015), supporting other studies (see above) that indicate that ProQ does not interact with sRNAs in the same manner as the Hfq chaperone.

1.4.2.3 CrsA: a third RNA chaperone

In contrast to Hfq and ProQ, the RNA chaperone protein CsrA does not facilitate the binding of an sRNA to an mRNA target. Instead CsrA binds to mRNA molecules using a well-defined GGA-binding sequence within a stem-loop region of the RNA. This binding allows for changes in translational efficiency, Rho-dependent termination or RNA stability (Dubey et al. 2005). The sRNA molecules involved in the CsrA-regulated regions usually contain several binding motifs that allow them to outcompete mRNA molecules by binding to CsrA. This CsrA-sRNA interaction stops CsrA interacting with the mRNA molecule (Dubey et al. 2005). CsrA has been implicated in the regulation of many bacterial virulence factors including those involved in biofilm formation and motility (Jackson et al. 2002; Pannuri et al. 2012; Yakhnin et al. 2013).

The formation of biofilms in *E. coli* is dependent on the production and export of poly-β-1,6-*N*-acetyl-D-glucosamine (PGA). The assembly and export of PGA requires a number of proteins including PgaA, a porin responsible for the transport of PGA to the cell surface, and NhaR, a transcriptional regulator that activates transcription of the *pga* locus (Jackson et al. 2002). CsrA downregulates the production of PGA by binding to *pgaA* and *nhaR* transcripts to block translation, thereby decreasing *E. coli* biofilm formation (Jackson et al. 2002; Pannuri et al. 2012). The CsrA RNA chaperone regulates motility within *E. coli* by binding to the *flhDC* mRNA, which encodes the motility master regulator $FlhD_4C_2$, that activates flagella production. The binding of CsrA to the *flhDC* mRNA protects the transcript from RNaseE degradation, allowing for increased flagella production and therefore increased motility (Yakhnin et al. 2013). Many

P. multocida genomes, including strain VP161 (PMVP_1313), encode a CsrA homolog but there have been no studies specifically focussed on this protein in this bacterium.

1.4.2.4 The importance of bacterial sRNA molecules

Bacterial sRNA molecules function by binding to a diverse range of mRNA targets in order to change transcript abundance or translational efficiency. Importantly, many sRNAs target mRNA molecules that encode virulence-associated genes, including those involved in biofilm formation, quorum sensing, capsule production and toxin production (Chambers and Sauer 2013; Papenfort and Vanderpool 2015; Pitman and Cho 2015; Perez-Reytor et al. 2016). The regulation of biofilm formation in at least seven different bacterial species is known to involve sRNAs. These species include *S.* Typhimurium, *Yersinia pseudotuberculosis, Yersinia pestis, Pseudomonas aeruginosa, V. cholera*, *Vibrio harveyi* and *E. coli* where 13 sRNA species have been shown to bind biofilm-associated mRNAs (Chambers and Sauer 2013). Each of these species used different types of sRNA molecules, however some mRNA targets were conserved across species, including *csgD* and *rpoS*, which encode for a transcriptional regulator and stress response sigma factor, respectively (Chambers and Sauer 2013).

Quorum sensing is the mechanism used by bacteria for inter-cell communication and is known to control biofilm formation in several bacteria (Perez-Reytor et al. 2016). Quorum sensing is also often under sRNA regulation (Perez-Reytor et al. 2016). In *Vibrio* spp., including *V. cholerae* and *V. harveyi*, the functionally redundant Qrr1-4 sRNA molecules bind to several mRNAs involved in quorum sensing, including *luxR* and *luxO* which encode the quorum sensing receptor and response regulator respectively. Once bound by a sRNA these transcripts are sequestered and degraded, leading to decreased protein production and reduced quorum sensing (Feng et al. 2015).

The production of capsule is also known to be regulated by sRNA molecules in many bacterial species. The hyaluronic acid capsule produced by *Streptococcus pyogenes* and *Streptococcus suis* and the colonic acid capsule produced by *E. coli,* are all under sRNA regulation (Sledjeski and Gottesman 1995; Pappesch et al. 2017; Xiao et al. 2017), but in each bacterial strain a different sRNA acts on different mRNA targets. For example, in *E. coli* the DsrA sRNA acts to relieve suppression of capsule production by the global transcriptional regulator H-NS whilst in *S. pyogenes* and *S. suis,* the MarS and rss04 sRNAs bind to their mRNA targets resulting in an increase in capsule production (Sledjeski and Gottesman 1995; Pappesch et al. 2017; Xiao et al. 2017). The production of a variety of toxins in Gram-positive bacteria are regulated by sRNAs (Pitman and Cho 2015). These include the sRNAs VirX and VirT, that regulate the production of the pore-forming toxin (perfringolysin A) and collagenase in *Clostridium perfingens*, RNAIII that regulates

heamolysin production and peptidoglycan hydrolase in *Staphylococcus aureus*, and FasX that regulates Streptolysin S production in *S. pyogenes* (Pitman and Cho 2015).

To date, there have been only three published studies that have analysed sRNAs in the *Pasteurellaceae* family (Amarasinghe et al. 2012; Santana et al. 2014; Rossi et al. 2016). In the periodontopathogen *Aggregatibacter actinomycetemcomitans*, four sRNA molecules (JA01 through to JA04) have been identified that are regulated by the transcriptional regulator Fur, and in *H. influenzae* the sRNA HrrF is also regulated by Fur (Amarasinghe et al. 2012; Santana et al. 2014). The JA03 sRNA was shown to regulate biofilm formation by *A. actinomycetemcomitans* during growth in low iron conditions, and through RNA sequencing (RNA-seq) analyses, HrrF was shown to regulate the expression of *H. influenzae* genes whose products are involved in amino acid biosynthesis, molybdate uptake, and deoxyribonucleotide synthesis (Amarasinghe et al. 2012; Santana et al. 2014). Bioinformatic analysis has also been used to identify regulatory RNAs, including sRNAs, encoded by *Actinobacillus pleuropneumoniae* (Rossi et al. 2016). This study used a variety of bioinformatic tools to identify 23 regulatory RNAs, of which 17 were validated experimentally. These RNAs were examined for conservation across *Pasteurellaceae* species and only three regulatory RNAs were found in all strains, namely RNaseP, RtT and tmRNA (Rossi et al. 2016). The well-known sRNA GcvB was identified in *A. pleuropneumoniae* strain L20 and was found to be conserved across 80% of *Pasteurellaceae* species using BLASTn and default paremeters (Rossi et al. 2016). However, the *P. multocida gcvB* was not identified in that study although it is located in the same position in the genome, relative to *gcvA,* and shares 78% nucleotide identity (64% coverage) with the *A. pleuropneumoniae* strain L20 *gcvB* molecule.

1.4.2.5 GcvB

The RNA-seq method was recently used in our laboratory to analyse the RNA transcripts produced by *P. multocida* strain VP161 grown to early-exponential, mid-exponential and late-exponential growth phases (Deveson Lucas, Harper and Boyce, unpublished). This analysis allowed for the identification of predicted intergenic regions with high levels of transcripts and subsequently >50 sRNA genes were putatively identified (Table 1.1; Mégroz, Boyce Laboratory, unpublished). One putative *P. multocida* sRNA showed sequence identity to the fully characterized GcvB sRNA from *E. coli* and *Salmonella enterica* serovar Typhi (Sharma et al. 2011). GcvB is a class I, Hfq-dependent, RNA that is conserved across many Gram-negative bacterial species. In *E. coli* and *Salmonella* sp., GcvB regulates the expression of a number of genes involved in amino acid biosynthesis and transport, including *oppA* and *dppA* (Pulvermacher et al. 2009; Sharma et al. 2011). The action of GcvB to actively suppress the production of amino acid biosynthesis and transport proteins when nutrients are abundant in the environment, allows bacteria like *E. coli* and *Salmonella* to conserve energy as other

transporters can transport the necessary amino acids when they highly abundant in the environment. The expression of GcvB is regulated by the GcvA and GcvR proteins in *E. coli* (Urbanowski et al. 2000). When the bacterium is growing in high nutrient conditions, particularly in high levels of glycine, GcvA acts to increase GcvB expression so that amino acid transport and biosynthesis proteins are repressed, and energy is conserved. However, when nutrient/glycine levels are low, GcvR is activated to form a complex with GcvA that represses GcvB expression, allowing the production of the necessary amino acid biosynthesis and transport proteins required for the acquisition and biosynthesis of nutrients (Urbanowski et al. 2000; Sharma et al. 2011). In *E. coli,* GcvB activity is also regulated by a prophage-encoded anti-sense sRNA, AgvB (Tree et al. 2014). AgvB, first identified through UV-crosslinking experiments involoving the sRNA chaperone Hfq, binds and represses GcvB activity in a Hfq-dependent manner (Tree et al. 2014). A similar antisense mechanism has also been identified in *Salmonella.* In this bacterium, an antisense molecule (or "sRNA sponge") called SroC, generated from the degradation of *gltI* transcripts, has been shown to bind and inhibit GcvB (Miyakoshi et al. 2015). Transcription of *gltI* itself is regulated by GcvB, and the binding of SroC to GcvB forms a negative feed-back loop (Miyakoshi et al. 2015).

GcvB controls a range of different bacterial phenotypes. In *Y. pestis, gcvB* mutants show decreased colony size and have a decreased growth rate *in vivo*. (McArthur et al. 2006). In *E. coli,* GcvB confers tolerance to acid stress via the activation of the stress response sigma factor RpoS, increases the levels of biofilm formation via the repression of *csgD,* and plays a role in LPS production through the repression of the PhoP/Q two-component signal transduction system (Jin et al. 2009; Boehm and Vogel 2012; Klein and Raina 2015). Additionally, *E. coli* GcvB and the transcriptional regulator protein Lrp act antagonistically to regulate biofilm production (Mika and Hengge 2014) and they control each other; GcvB can bind directly to *lrp* mRNA and decrease Lrp production but how Lrp negatively regulates GcvB production is not known. However, there are predicted Lrp recognition sequences within GcvB which indicate direct binding of Lrp to GcvB (Modi et al. 2011).

In both *E. coli* and *Salmonella,* GcvB is known to bind to the conserved sequence 5'-CACAACAT-3' within its mRNA targets (Sharma et al. 2011). This binding sequence has been verified using two-plasmid GFP or *lacZ* reporter systems in *E. coli* and *Salmonella* (Pulvermacher et al. 2009; Sharma et al. 2011). The reverse complement of the seed sequence is found within GcvB, including GcvB homologs encoded by many bacterial species within the Pasteurellaceae genus (Rossi et al. 2016). Though not identified in the Rossi study (due to the BLASTn default parameters used), it is likely that the GcvB homolog found in *P. multocida* will act in a similar manner and bind to mRNA targets through complementary base pairing.

Table 1.1. *P. multocida* strain VP161 putative sRNAs and other regions with high levels of transcript.
(Adapted from Mégroz, Boyce laboratory, unpublished)

Official name	sRNA homolog	VP161 Base Range	Strand	Size (bp)	PM70 base range	PM70 Strand	PM70 Identity with VP161	Confirmatory tests performed		
								Northern Blot	Primer Extension	5' RACE
Region of high expression in an intergenic region										
Prrc01	GcvB	569763:569943	-	181	651997:652177	-	99%	+ [a.]	+[a]	+ [a.]
Prrc02	HrrF	75918:76046	+	129	147414:147560	+	100%		+ [b.]	
Prrc04	-	1767438:1767636	-	199	1840114:1840312	-	98%			
Prrc05	-	65715:65877	+	163	137229:137391	+	100%			
Prrc06	-	550100:550259	+	160	632336:632495	+	99%	+ [c.]	+ [c.]	
Prrc07	-	552850:553040	+	191	635086:635276	+	100%			
Prrc10	-	608452:608650	+	199	700619:700815	+	98%		+ [c.]	
Prrc11	-	1376798:1377067	-	270	1455915:1456184	-	97%			
Prrc12	-	410793:410950	+	158	491412:491569	+	99%		+ [c.]	
Prrc13	-	410973:411122	+	150	491592:491741	+	99%	+ [a.]		
Prrc15	-	185627:185794	-	168	257147:257314	-	99%			
Prrc16	-	318927:319074	+	148	399001:399148	+	100%			
Prrc18	-	1357348:1357484	+	137	1436528:1436664	+	100%			
Prrc20	-	1789291:1789447	-	157	1866348:1866504	-	100%			
Prrc22	-	2061460:2061701	-	242	2156401:2156642	-	99%			
Prrc24	-	398700:398877	-	178	479325:479502	-	100%			
Prrc25	-	609792:609948	+	157	701998:702154	+	94%		+ [c.]	
Prrc26	-	1084517:1084649	-	133	1162214:1162346	-	100%			
Prrc27	-	1091959:1092105	-	147	1169748:1169894	-	99%			

Table 1.1 continued.

Official name	sRNA homolog	VP161 Base Range	Strand	Size (bp)	PM70 base range	PM70 Strand	PM70 Identity with VP161	Confirmatory tests performed		
								Northern Blot	Primer Extension	5' RACE
Region of high expression in an intergenic region										
Prrc30	-	1885283:1885435	-	153	1966084:1966237	-	96%			
Prrc32	-	2118915:2119058	+	144	2228108:2228251	+	100%			
Prrc34	-	115917:116083	+	167	187431:187596	+	98%			
Prrc40	-	927873:927996	-	124	1006260:1006387	-	92%			
Prrc41	-	356845:356988	+	144	436919:437062	+	100%			
Prrc42	-	1069162:1069406	+	245	1146878:1147122	+	100%			
Prrc45	-	2174849:2175193	+	345	29032:29375	+	97%			
Prrc47	-	1629186:1629264	+	79	1681788:1681866	+	100%			
Prrc48	-	1013049:1013165	-	117	1090566:1090682	-	100%			
Prrc49	-	1030980:1031115	-	136	1108489:1108620	-	98%			
Prrc50	-	1447720:1447906	+	187	1520193:1520379	+	100%			
Prrc51	-	490102:490402	+	301	571423:571571	+	92%			
Prrc55	-	195999:196033	-	35	267520:267553	+	100%			
Prrc56	-	709488:709529	-	42	804184:804224	+	100%			

Table 1.1 continued.

Region of high expression-antisense to an open reading frame							Confirmatory tests performed			Region of overlap (strain observed in)
Official name	**VP161 Base Range**	**Strand**	**Size (bp)**	**PM70 base range**	**PM70 Strand**	**PM70 Identity with VP161**	**Northern Blot**	**Primer Extension**	**5' RACE**	
Prrc03	893321:893467	-	147	971700:971885	-	100%	+ [b.]	+ [b.]		+10 to +189 of PMVP_0818 (VP161) [Opposite strand]
Prrc14	1640951:1641105	+	155	1076906:1077060	-	98%				unannotated ORF between PMVP_1552 and PMVP_1553 (VP161) [Opposite strand]
Prrc17	682405:682585	+	181	776980:777160	+	99%				Possible ORF with multispecies matches "hypothetical" (VP161)
Prrc19	1545667:1545827	-	161	1598604:1598764	-	100%				PMXST_1448 (X73) [Opposite strand]
Prrc21	1893872:1894096	-	225	1976805:1977029	-	99%				+1 to +73 of PM1754 (PM70) [Opposite strand]
Prrc33	646981:647160	+	180	738782:738961	+	100%				-16 to + 164 of PMVP_0608 (VP161) [Opposite strand]
Prrc44	1832310:1832480	+	171	1909329:1909498	+	98%				-30 to +141 of PMVP_1757 (VP161) [Opposite strand]
Prrc54	62136:62210	-	75	NA	NA	NA				+64 to +129 in PMVP_0059 (VP161) [Opposite strand]
Prrc57	722215:722252	-	38	816965:816992	+	96%				+129 to +98 of PMXST_00672 (X73) [Opposite strand]
Prrc61	1263950:1264121	+	172	1349633:1349675	-	100%				-122 to +51 of PM1151 (PM70) [Opposite strand]

Table 1.1 continued.

Region of high expression partially overlapping an open reading frame							Confirmatory tests performed			Region of overlap (strain observed in)
Official name	**VP161 Base Range**	**Strand**	**Size (bp)**	**PM70 base range**	**PM70 Strand**	**PM70 Identity with VP161**	**Northern Blot**	**Primer Extension**	**5' RACE**	
Prrc08	610053:610235	+	183	702252:702434	+	98%		+ [c.]		+56 to + 116 of PMVP_0574 (VP161)
Prrc09	2094010:2094182	-	173	2203205:2203377	-	100%				+1 to +159 of PM1963 (PM70)
Prrc23	376689:376928	+	240	457348:457587	+	99%				-124 to +116 of PMVP_0358 (VP161)
Prrc28	1333017:1333230	-	214	1416287:1416500	-	100%				-181 to +33 of PMXST_01228 (X73)
Prrc29	1880903:1881281	-	379	1961705:1962083	-	99%				-277 to + 102 of PMVP_1797 (VP161)
Prrc31	2088401:2088583	+	183	2197645:2197828	+	96%				-132 to + 111 of PMVP_1997 (VP161)
Prrc35	142159:142345	-	187	213676:213862	-	99%				-54 to + 73 of PMVP_0144 (VP161)
Prrc36	871928:872241	+	314	951900:952213	+	99%				-199 to + 115 of PM0805 (PM70)
Prrc37	1287819:1287949	-	131	1374408:1374538	-	100%				-79 to +52 of PM1180 (PM70)
Prrc38	1410949:1411159	+	211	1490063:1490273	+	100%				-91 to +120 of PMVP_0144 (VP161)
Prrc39	407402:407520	+	119	488024:488138	+	99%				+97 to + 121 of PMVP_0387 (VP161)
Prrc43	1250138:1250271	-	134	1334941:1335074	-	100%				-1 to -135 of PMVP_1152 (VP161)
Prrc46	1725190:1725432	-	243	1795017:1795259	-	100%				-111 to +132 of PMVP_1638 (VP161)

[a.] Performed by Emily Gulliver, [b.] Performed by Amy Wright, [c.] Performed by Marianne Mégroz.

1.4.2.6 Methods for identifying RNA-RNA and RNA-protein interactions

To determine direct RNA-RNA interacting partners several methods have been employed, including electrophoretic mobility shift assays (EMSA) and surface plasmon resonance (SPR). Two-plasmid fluorescent reporter systems, usually using the green fluorescent protein (GFP), have also been used. In *E. coli* this system was used to successfully characterise specific sRNA-mRNA binding interactions and to identify the short nucleotide sequence required for this interaction (Urban and Vogel 2007; Faner and Feig 2013). Within one cell, one plasmid expresses a recombinant copy of the sRNA being investigated and the second plasmid (within the same cell) expresses the *gfp* gene fused with sequence representing the 5' UTR region plus the first few codons of the predicted target gene (the sRNA binding site is usually within the 5' UTR region). If they are a true sRNA-mRNA binding pair, the sRNA will bind to a region within the target mRNA sequence (transcriptionally fused to *gfp*) resulting in altered levels of GFP-mediated fluorescence. This two-plasmid reporter system has been used to successfully show that *E. coli* GcvB binds to the target mRNA *dppA* (Urban and Vogel 2007). Subsequent uses of the reporter system have substituted GFP for a superior version called super-folder GFP (Corcoran et al. 2012). Superfolder GFP is more stable than GFP and folds more efficiently, even under unfavourable conditions, thus allowing for higher levels of fluorescence.

To identify large sets of sRNAs that interact with a particular RNA chaperone protein, such as Hfq or ProQ, several whole-genome methods can be used including co-immunoprecipitation, transcriptomics and proteomics. Transcriptomic and proteomic methods can be used to compare total RNA and protein production, respectively, between wild-type and specific mutant strains, thus identifying differentially expressed transcripts or proteins as possible targets of sRNA-chaperone regulation. However, not all changes detected by proteomics and transcriptomics may be the direct effect of a specific mutation. It is expected that in most bacteria, a network of regulation is in play so some differences in expression may be due to secondary effects. Recently, advances in UV-crosslinking methodology have allowed for the identification of RNAs that bind directly to protein chaperones such as Hfq (Holmqvist et al. 2016) and have allowed for the capture and sequencing of sRNA-mRNA pairs bound to the chaperone protein (Tree et al. 2014; Waters et al. 2017). The UV crosslinking and analysis of cDNA method (UV-CRAC) uses UV light to crosslink any RNA-protein interactions. The chaperone protein of interest is affinity tagged, usually with a 3 x FLAG-tag (DYKDDDDK) and a 6 x His-tag, so that any complexes can be precipitated using anti-FLAG beads- and Ni-NTA resin (Sy et al. 2018). Unbound and therefore exposed regions of the RNA molecules interacting/crosslinked to the chaperone protein are trimmed using RNase (RNA bound to the chaperone is protected) and then adapters are ligated to the trimmed RNA molecules. The chaperone protein is then

degraded by protease digestion, and cDNA generated from the released RNA is sequenced using a high-throughput technique (such as Illumina-seq). Mapping of the cDNA sequences to the genome allows for the identification of all chaperone-bound RNAs and their binding positions (Sy et al. 2018). A modification of this method, called UV-crosslinking, ligation and sequencing of hybrids (UV-CLASH), can also identify the RNA-RNA interactions that are facilitated by the chaperone protein (Figure 1.4). UV-CLASH includes an additional step that allows for the ligation of RNA pairs that are bound to the same chaperone protein molecule, thereby facilitating identification of precise RNA-RNA interacting regions. Data from UV-CLASH allows for the direct identification of sRNA-mRNA pairs and allows for interrogation of the entire RNA chaperone-dependent interaction network within a cell. This method was recently used to characterise the regulon of RNase E in *E. coli* and to identify the seed binding sequences within the sRNA species ChiX and RyhB (Waters et al. 2017).

1.5 Project aims

This study aims to expand the knowledge of sRNA regulation in the Gram-negative bacterium *P. multocida*, and to develop techniques that can be used to examine sRNA molecules and their interaction with protein chaperones. Specifically, we focused on the sRNA GcvB (Chapter 2) and the RNA chaperone protein ProQ (Chapter 3). We characterised the GcvB transcript and determined the GcvB regulon through bioinformatic analyses of the proteomic data obtained for the *gcvB* mutant. The putative seed binding region in GcvB and its predicted mRNA targets was identified and validated experimentally for the GcvB and one mRNA target, *gltA*, using the two-plasmid reporter GFP system and electrophoretic mobility shift assays (EMSA). The work presented in Chapter 3 focused on the role of the RNA chaperone protein ProQ in *P. multocida*. Here the proteome and transcriptome of a *proQ* mutant strain of *P. multocida* was examined to identify potential ProQ target RNA molecules. RNA molecules that bound ProQ, as well as ProQ-associated RNA-RNA pairs, were identified using UV-CLASH. Together, these analyses allowed for the identification of a predicted ProQ regulon in *P. multocida* strain VP161.

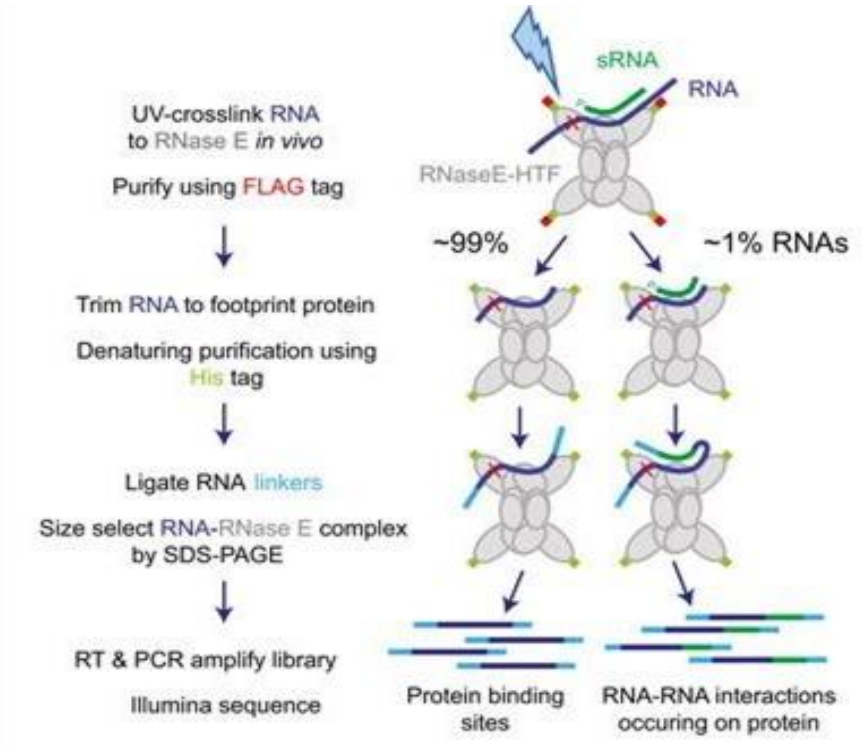

Figure 1.4. Schematic representation of the CLASH protocol for identification of RNA–RNA interactions. RNAs were UV-crosslinked to tagged RNA chaperone protein (RNAse E) *in vivo* and purified using M2 anti-FLAG resin. RNAs were trimmed using RNase A/T1 and further purified under denaturing conditions. RNA linkers were ligated to the immobilized RNA–protein complexes. Duplexed RNAs may be ligated into a single contiguous molecule (left, CLASH) that gives information on RNA–RNA interaction occurring on the chaperone protein. The remaining single RNAs reveal the site of chaperone protein binding within the transcriptome. Linker-ligated RNA–protein complexes were size-selected by SDS–PAGE and RNAs recovered for library preparation and sequencing. The schematic on the right represents the key steps in preparing UV-crosslinked RNA–protein complexes to map RNA–protein interactions sites (~99% of reads recovered), and RNA–RNA interaction sites (~1% of reads recovered). Colours correspond to key words in the flow diagram. Taken from Waters et al. (2017).

1.6 References

Ackermann MR, Register KB, Stabel JR, Gwaltney SM, Howe TS, Rimler RB. 1996. Effect of *Pasteurella multocida* toxin on physeal growth in young pigs. *Am J Vet Res* **57**: 848-852.

Amarasinghe JJ, Connell TD, Scannapieco FA, Haase EM. 2012. Novel iron-regulated and Fur-regulated small regulatory RNAs in *Aggregatibacter actinomycetemcomitans*. *Mol Oral Microbiol* **27**: 327-349.

Baalsrud KJ. 1987. Atrophic rhinitis in goats in Norway. *Vet Rec* **121**: 350-353.

Bethe A, Wieler LH, Selbitz HJ, Ewers C. 2009. Genetic diversity of porcine *Pasteurella multocida* strains from the respiratory tract of healthy and diseased swine. *Vet Microbiol* **139**: 97-105.

Boehm A, Vogel J. 2012. The *csgD* mRNA as a hub for signal integration via multiple small RNAs. *Mol Microbiol* **84**: 1-5.

Bosch M, Garrido E, Llagostera M, Perez de Rozas AM, Badiola I, Barbe J. 2002a. *Pasteurella multocida exbB, exbD* and *tonB* genes are physically linked but independently transcribed. *FEMS Microbiol Lett* **210**: 201-208.

Bosch M, Garrido ME, Llagostera M, Perez De Rozas AM, Badiola I, Barbe J. 2002b. Characterization of the *Pasteurella multocida hgbA* gene encoding a hemoglobin-binding protein. *Infect Immun* **70**: 5955-5964.

Bosch M, Garrido ME, Pérez De Rozas AM, Badiola I, Barbé J, Llagostera M. 2004. *Pasteurella multocida* contains multiple immunogenic haemin- and haemoglobin-binding proteins. *Vet Microbiol* **99**: 103-111.

Boyce JD, Adler B. 2000. The capsule is a virulence determinant in the pathogenesis of *Pasteurella multocida* M1404 (B:2). *Infect Immun* **68**: 3463-3468.

Boyce JD, Chung JY, Adler B. 2000. *Pasteurella multocida* capsule: composition, function and genetics. *J Biotechnol* **83**: 153-160.

Boyce JD, Seemann T, Adler B, Harper M. 2012. Pathogenomics of *Pasteurella multocida*. *Curr Top Microbiol Immunol* **361**: 23-38.

Boyce JD, Wilkie I, Harper M, Paustian ML, Kapur V, Adler B. 2002. Genomic scale analysis of *Pasteurella multocida* gene expression during growth within the natural chicken host. *Infect Immun* **70**: 6871-6879.

Bradley MD, Beach MB, de Koning AP, Pratt TS, Osuna R. 2007. Effects of Fis on *Escherichia coli* gene expression during different growth stages. *Microbiology* **153**: 2922-2940.

Carter GR. 1952. The type specific capsular antigen of *Pasteurella multocida*. *Can J Med Sci* **30**: 48-53.

Carter GR. 1967. Pasteurellosis: *Pasteurella multocida* and *Pasteurella hemolytica*. *Adv Vet Sci* **11**: 321-379.

Carter GR, Chengappa MM. 1980. Hyaluronidase production by type B *Pasteurella multocida* from cases of hemorrhagic septicemia. *J Clin Microbiol* **11**: 94-96.

Carter GR, De Alwis MCL. 1989. Haemorrhagic Septicaemia. In *Pasteurella and Pasteurellosis*, (ed. C Adlam, Rutter, J.M.), pp. 131-160. Academic Press Limited, London.

Catry B, Chiers K, Schwarz S, Kehrenberg C, Decostere A, de Kruif A. 2005. Fatal peritonitis caused by *Pasteurella multocida* capsular type F in calves. *J Clin Microbiol* **43**: 1480-1483.

Chambers JR, Sauer K. 2013. Small RNAs and their role in biofilm formation. *Trends Microbiol* **21**: 39-49.

Chao Y, Vogel J. 2010. The role of Hfq in bacterial pathogens. *Curr Opin Microbiol* **13**: 24-33.

Chaulk SG, Smith Frieday MN, Arthur DC, Culham DE, Edwards RA, Soo P, Frost LS, Keates RA, Glover JN, Wood JM. 2011. ProQ is an RNA chaperone that controls ProP levels in *Escherichia coli*. *Biochemistry* **50**: 3095-3106.

Chung JY, Wilkie I, Boyce JD, Townsend KM, Frost AJ, Ghoddusi M, Adler B. 2001. Role of capsule in the pathogenesis of fowl cholera caused by *Pasteurella multocida* serogroup A. *Infect Immun* **69**: 2487-2492.

Chung JY, Zhang Y, Adler B. 1998. The capsule biosynthetic locus of *Pasteurella multocida* A:1. *FEMS Microbiol Lett* **166**: 289-296.

Ciprian A, Pijoan C, Cruz T, Camacho J, Tortora J, Colmenares G, Lopez-Revilla R, de la Garza M. 1988. *Mycoplasma hyopneumoniae* increases the susceptibility of pigs to experimental *Pasteurella multocida* pneumonia. *Can J Vet Res* **52**: 434-438.

Corcoran CP, Podkaminski D, Papenfort K, Urban JH, Hinton JC, Vogel J. 2012. Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol Microbiol* **84**: 428-445.

Dabo SM, Confer AW, Quijano-Blas RA. 2003. Molecular and immunological characterization of *Pasteurella multocida* serotype A:3 OmpA: evidence of its role in *P. multocida* interaction with extracellular matrix molecules. *Microb Pathog* **35**: 147-157.

De Alwis MCL. 1999. *Haemorrhagic septicaemia / M.C.L. De Alwis*. Australian Centre for International Agricultural Research, Canberra, A.C.T.

DeAngelis PL, Gunay NS, Toida T, Mao WJ, Linhardt RJ. 2002. Identification of the capsular polysaccharides of Type D and F *Pasteurella multocida* as unmodified heparin and chondroitin, respectively. *Carbohydr Res* **337**: 1547-1552.

Deeb BJ, DiGiacomo RF, Bernard BL, Silbernagel SM. 1990. *Pasteurella multocida* and *Bordetella bronchiseptica* infections in rabbits. *J Clin Microbiol* **28**: 70-75.

Desnoyers G, Bouchard MP, Masse E. 2013. New insights into small RNA-dependent translational regulation in prokaryotes. *Trends Genet* **29**: 92-98.

Dewhirst FE, Klein EA, Bennett ML, Croft JM, Harris SJ, Marshall-Jones ZV. 2015. The feline oral microbiome: a provisional 16S rRNA gene based taxonomy with full-length reference sequences. *Vet Microbiol* **175**: 294-303.

Dewhirst FE, Klein EA, Thompson EC, Blanton JM, Chen T, Milella L, Buckley CM, Davis IJ, Bennett ML, Marshall-Jones ZV. 2012. The canine oral microbiome. *PLoS One* **7**: e36067.

DiGiacomo RF, Xu YM, Allen V, Hinton MH, Pearson GR. 1991. Naturally acquired *Pasteurella multocida* infection in rabbits: clinicopathological aspects. *Can J Vet Res* **55**: 234-238.

Dominick MA, Rimler RB. 1988. Turbinate osteoporosis in pigs following intranasal inoculation of purified *Pasteurella* toxin: histomorphometric and ultrastructural studies. *Vet Pathol* **25**: 17-27.

Dubey AK, Baker CS, Romeo T, Babitzke P. 2005. RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *RNA* **11**: 1579-1587.

Faner MA, Feig AL. 2013. Identifying and characterizing Hfq-RNA interactions. *Methods* **63**: 144-159.

Feng L, Rutherford ST, Papenfort K, Bagert JD, van Kessel JC, Tirrell DA, Wingreen NS, Bassler BL. 2015. A *qrr* noncoding RNA deploys four different regulatory mechanisms to optimize quorum-sensing dynamics. *Cell* **160**: 228-240.

Frandsen PL, Foged NT, Petersen SK, Bording A. 1991. Characterization of toxin from different strains of *Pasteurella multocida* serotype A and D. *Zentralbl Veterinarmed B* **38**: 345-352.

Fraser JR, Laurent TC, Laurent UB. 1997. Hyaluronan: its nature, distribution, functions and turnover. *J Intern Med* **242**: 27-33.

Fuller TE, Kennedy MJ, Lowery DE. 2000. Identification of *Pasteurella multocida* virulence genes in a septicemic mouse model using signature-tagged mutagenesis. *Microb Pathog* **29**: 25-38.

Giordano A, Dincman T, Clyburn BE, Steed LL, Rockey DC. 2015. Clinical features and outcomes of *Pasteurella multocida* infection. *Medicine* **94**: e1285.

Glorioso JC, Jones GW, Rush HG, Pentler LJ, Darif CA, Coward JE. 1982. Adhesion of type A *Pasteurella mulocida* to rabbit pharyngeal cells and its possible role in rabbit respiratory tract infections. *Infect Immun* **35**: 1103-1109.

Gonzalez G, Hardwick S, Maslen SL, Skehel JM, Holmqvist E, Vogel J, Bateman A, Luisi B, Broadhurst RW. 2017. Structure of the *Escherichia coli* ProQ RNA chaperone protein. *RNA* **23**: 696-711.

Gottesman S, Storz G. 2011. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol* **3**.

Guo D, Lu Y, Zhang A, Liu J, Yuan D, Jiang Q, Lin H, Si C, Qu L. 2012. Identification of genes transcribed by *Pasteurella multocida* in rabbit livers through the selective capture of transcribed sequences. *FEMS Microbiol Lett* **331**: 105-112.

Harper M, Boyce JD, Adler B. 2006. *Pasteurella multocida* pathogenesis: 125 years after Pasteur. *FEMS Microbiol Lett* **265**: 1-10.

Harper M, Boyce JD, Cox AD, St Michael F, Wilkie IW, Blackall PJ, Adler B. 2007a. *Pasteurella multocida* expresses two lipopolysaccharide glycoforms simultaneously, but only a single form is required for virulence: identification of two acceptor-specific heptosyl I transferases. *Infect Immun* **75**: 3885-3893.

Harper M, Boyce JD, Wilkie IW, Adler B. 2003. Signature-tagged mutagenesis of *Pasteurella multocida* identifies mutants displaying differential virulence characteristics in mice and chickens. *Infect Immun* **71**: 5440-5446.

Harper M, Cox A, St Michael F, Parnas H, Wilkie I, Blackall PJ, Adler B, Boyce JD. 2007b. Decoration of *Pasteurella multocida* lipopolysaccharide with phosphocholine is important for virulence. *J Bacteriol* **189**: 7384-7391.

Harper M, Cox AD, Adler B, Boyce JD. 2011. *Pasteurella multocida* lipopolysaccharide: The long and the short of it. *Vet Microbiol* **153**: 109-115.

Harper M, John M, Turni C, Edmunds M, St Michael F, Adler B, Blackall PJ, Cox AD, Boyce JD. 2015. Development of a rapid multiplex PCR assay to genotype *Pasteurella multocida* strains by use of the lipopolysaccharide outer core biosynthesis locus. *J Clin Microbiol* **53**: 477-485.

Harper M, St Michael F, John M, Steen J, van Dorsten L, Parnas H, Vinogradov E, Adler B, Cox AD, Boyce JD. 2014. Structural analysis of lipopolysaccharide produced by Heddleston serovars 10, 11, 12 and 15 and the identification of a new *Pasteurella multocida* lipopolysaccharide outer core biosynthesis locus, L6. *Glycobiology* **24**: 649-659.

Harper M, St Michael F, Vinogradov E, John M, Boyce JD, Adler B, Cox AD. 2012. Characterization of the lipopolysaccharide from *Pasteurella multocida* Heddleston serovar 9: identification of a proposed bi-functional dTDP-3-acetamido-3,6-dideoxy-alpha-D-glucose biosynthesis enzyme. *Glycobiology* **22**: 332-344.

Heddleston KL, Rebers PA. 1972. Fowl cholera: cross-immunity induced in turkeys with formalin-killed *in-vivo*-propagated *Pasteurella*. *Avian Dis* **16**: 578-586.

Holmqvist E, Li L, Bischler T, Barquist L, Vogel J. 2018. Global maps of ProQ binding *in vivo* reveal target recognition via RNA structure and stability control at mRNA 3' ends. *Mol Cell* **70**: 971-982.e976.

Holmqvist E, Wright PR, Li L, Bischler T, Barquist L, Reinhardt R, Backofen R, Vogel J. 2016. Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking *in vivo*. *Embo J* **35**: 991-1011.

Honnorat E, Seng P, Savini H, Pinelli PO, Simon F, Stein A. 2016. Prosthetic joint infection caused by *Pasteurella multocida*: a case series and review of literature. *BMC Infect Dis* **16**: 435.

Hood DW, Deadman ME, Jennings MP, Bisercic M, Fleischmann RD, Venter JC, Moxon ER. 1996. DNA repeats identify novel virulence genes in *Haemophilus influenzae*. *PNAS* **93**: 11121-11125.

Horiguchi Y. 2012. Swine atrophic rhinitis caused by *Pasteurella multocida* toxin and *Bordetella dermonecrotic* toxin. *Curr Top Microbiol Immunol* **361**: 113-129.

Jackson DW, Suzuki K, Oakford L, Simecka JW, Hart ME, Romeo T. 2002. Biofilm formation and dispersal under the influence of the global regulator CsrA of *Escherichia coli*. *J Bacteriol* **184**: 290-301.

Jin Y, Watt RM, Danchin A, Huang JD. 2009. Small noncoding RNA GcvB is a novel regulator of acid resistance in *Escherichia coli*. *BMC Genomics* **10**: 1471-2164.

Kerr CH, Culham DE, Marom D, Wood JM. 2014. Salinity-dependent impacts of ProQ, Prc, and Spr deficiencies on *Escherichia coli* cell structure. *J Bacteriol* **196**: 1286-1296.

Klein G, Raina S. 2015. Regulated control of the assembly and diversity of LPS by noncoding sRNAs. *Biomed Res Int* **2015**: 153561.

Kock RA, Orynbayev M, Robinson S, Zuther S, Singh NJ, Beauvais W, Morgan ER, Kerimbayev A, Khomenko S, Martineau HM et al. 2018. Saigas on the brink: Multidisciplinary analysis of the factors influencing mass mortality events. *Sci Adv* **4**: eaao2314.

Krzysciak W, Jurczak A, Koscielniak D, Bystrowska B, Skalniak A. 2014. The virulence of *Streptococcus* mutans and the ability to form biofilms. *Eur J Clin Microbiol Infect Dis* **33**: 499-515.

Lenz DH, Bassler BL. 2007. The small nucleoid protein Fis is involved in *Vibrio cholerae* quorum sensing. *Mol Microbiol* **63**: 859-871.

Locht C, Antoine R, Jacob-Dubuisson F. 2001. *Bordetella pertussis*, molecular pathogenesis under multiple aspects. *Curr Opin Microbiol* **4**: 82-89.

Marion C, Stewart JM, Tazi MF, Burnaugh AM, Linke CM, Woodiga SA, King SJ. 2012. *Streptococcus pneumoniae* can utilize multiple sources of hyaluronic acid for growth. *Infect Immun* **80**: 1390-1398.

Martineau-Doize B, Frantz JC, Martineau GP. 1990. Effects of purified *Pasteurella multocida* dermonecrotoxin on cartilage and bone of the nasal ventral conchae of the piglet. *Anat Rec* **228**: 237-246.

Martineau-Doize B, Menard J, Girard C, Frantz JC, Martineau GP. 1991. Effects of purified *Pasteurella multocida* dermonecrotoxin on the nasal ventral turbinates of fattening pigs: histological observations. *Can J Vet Res* **55**: 377-379.

May BJ, Zhang Q, Li LL, Paustian ML, Whittam TS, Kapur V. 2001. Complete genomic sequence of *Pasteurella multocida*, Pm70. *PNAS* **98**: 3460-3465.

Mbuthia PG, Njagi LW, Nyaga PN, Bebora LC, Minga U, Kamundia J, Olsen JE. 2008. *Pasteurella multocida* in scavenging family chickens and ducks: carrier status, age susceptibility and transmission between species. *Avian Pathol* **37**: 51-57.

McArthur SD, Pulvermacher SC, Stauffer GV. 2006. The *Yersinia pestis gcvB* gene encodes two small regulatory RNA molecules. *BMC Microbiol* **6**: 52.

Mégroz M, Kleifeld O, Wright A, Powell D, Harrison P, Adler B, Harper M, Boyce JD. 2016. The RNA-binding chaperone Hfq is an important global regulator of gene expression in *Pasteurella multocida* and plays a crucial role in production of a number of virulence factors, including hyaluronic acid capsule. *Infect Immun* **84**: 1361-1370.

Mika F, Hengge R. 2014. Small RNAs in the control of RpoS, CsgD, and biofilm architecture of *Escherichia coli*. *RNA Biol* **11**: 494-507.

Miyakoshi M, Chao Y, Vogel J. 2015. Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. *Embo J* **34**: 1478-1492.

Modi SR, Camacho DM, Kohanski MA, Walker GC, Collins JJ. 2011. Functional characterization of bacterial sRNAs using a network biology approach. *PNAS* **108**: 15522-15527.

Moor H, Teppo A, Lahesaare A, Kivisaar M, Teras R. 2014. Fis overexpression enhances *Pseudomonas putida* biofilm formation by regulating the ratio of LapA and LapF. *Microbiology* **160**: 2681-2693.

Moustafa AM, Seemann T, Gladman S, Adler B, Harper M, Boyce JD, Bennett MD. 2015. Comparative genomic analysis of asian haemorrhagic septicaemia-associated strains of *Pasteurella multocida* identifies more than 90 haemorrhagic septicaemia-specific genes. *PLoS One* **10**: e0130296.

Mullan PB, Lax AJ. 1998. *Pasteurella multocida* toxin stimulates bone resorption by osteoclasts via interaction with osteoblasts. *Calcif Tissue Int* **63**: 340-345.

Mullen LM, Bosse JT, Nair SP, Ward JM, Rycroft AN, Robertson G, Langford PR, Henderson B. 2008. *Pasteurellaceae* ComE1 proteins combine the properties of fibronectin adhesins and DNA binding competence proteins. *PLoS One* **3**: 22.

Muniandy N, Edgar J, Woolcock J, Mukkur T. 1992. Virulence, purification, structure, and protective potential of the putative capsular polysaccharide of *Pasteurella multocida* type 6: B. *Pasteurellosis in production animals*: 47-53.

Orth JH, Aktories K. 2012. Molecular biology of *Pasteurella multocida* toxin. *Curr Top Microbiol Immunol* **361**: 73-92.

Orth JH, Fester I, Siegert P, Weise M, Lanner U, Kamitani S, Tachibana T, Wilson BA, Schlosser A, Horiguchi Y et al. 2013. Substrate specificity of *Pasteurella multocida* toxin for alpha subunits of heterotrimeric G proteins. *Faseb J* **27**: 832-842.

Pan CQ, Finkel SE, Cramton SE, Feng JA, Sigman DS, Johnson RC. 1996. Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts. *J Mol Biol* **264**: 675-695.

Pannuri A, Yakhnin H, Vakulskas CA, Edwards AN, Babitzke P, Romeo T. 2012. Translational repression of NhaR, a novel pathway for multi-tier regulation of biofilm circuitry by CsrA. *J Bacteriol* **194**: 79-89.

Papenfort K, Vanderpool CK. 2015. Target activation by regulatory RNAs in bacteria. *FEMS Microbiol Rev* **39**: 362-378.

Pappesch R, Warnke P, Mikkat S, Normann J, Wisniewska-Kucper A, Huschka F, Wittmann M, Khani A, Schwengers O, Oehmcke-Hecht S et al. 2017. The regulatory small RNA MarS supports virulence of *Streptococcus pyogenes*. *Sci Rep* **7**: 12241.

Paustian ML, May BJ, Kapur V. 2001. *Pasteurella multocida* gene expression in response to iron limitation. *Infect Immun* **69**: 4109-4115.

Paustian ML, May BJ, Kapur V. 2002. Transcriptional response of *Pasteurella multocida* to nutrient limitation. *J Bacteriol* **184**: 3734-3739.

Peng Z, Liang W, Liu W, Chen H, Wu B. 2017. Genome characterization of *Pasteurella multocida* subspecies *septica* and comparison with *Pasteurella multocida* subspecies *multocida* and *gallicida*. *Arch Microbiol* **199**: 635-640.

Perez-Reytor D, Plaza N, Espejo RT, Navarrete P, Bastias R, Garcia K. 2016. Role of non-coding regulatory RNA in the virulence of human pathogenic Vibrios. *Front Microbiol* **7**: 2160.

Petruzzi B, Briggs RE, Tatum FM, Swords WE, De Castro C, Molinaro A, Inzana TJ. 2017. Capsular polysaccharide interferes with biofilm formation by *Pasteurella multocida* serogroup A". *MBio*. **8**: e01843-17. Erratum 2018. *MBio* **9**: e00176-18.

Pitman S, Cho KH. 2015. The mechanisms of virulence regulation by small noncoding RNAs in low GC Gram-Positive pathogens. *Int J Mol Sci* **16**: 29797-29814.

Pruimboom IM, Rimler RB, Ackermann MR. 1999. Enhanced adhesion of *Pasteurella multocida* to cultured Turkey peripheral blood monocytes. *Infect Immun* **67**: 1292-1296.

Pulvermacher SC, Stauffer LT, Stauffer GV. 2009. Role of the *Escherichia coli* Hfq protein in GcvB regulation of *oppA* and *dppA* mRNAs. *Microbiology* **155**: 115-123.

Rhoades KR, Rimler RB. 1990. *Pasteurella multocida* colonization and invasion in experimentally exposed turkey poults. *Avian Dis* **34**: 381-383.

Rimler RB, Rhoades KR. 1987. Serogroup F, a new capsule serogroup of *Pasteurella multocida*. *J Clin Microbiol* **25**: 615-618.

Rosner H, Grimmecke HD, Knirel YA, Shashkov AS. 1992. Hyaluronic acid and a (1→4]-β-D-xylan, extracellular polysaccharides of *Pasteurella multocida* (Carter type A) strain 880. *Carbohydr Res* **223**: 329-333.

Rossi CC, Bosse JT, Li Y, Witney AA, Gould KA, Langford PR, Bazzolli DM. 2016. A computational strategy for the search of regulatory small RNAs in *Actinobacillus pleuropneumoniae*. *RNA* **22**: 1373-1385.

Rutter J, Rojas X. 1982. Atrophic rhinitis in gnotobiotic piglets-differences in the pathogenicity of *Pasteurella multocida* in combined infections with *Bordetella bronchiseptica*. *Vet Rec* **110**: 531-535.

Saif Y. 2008. *Diseases of poultry*. Wiley-Blackwell, Ames.

Sakai Y, Abe K, Nakashima S, Yoshida W, Ferri S, Sode K, Ikebukuro K. 2013. Improving the gene-regulation ability of small RNAs by scaffold engineering in *Escherichia coli*. *ACS Synth Biol* **18**: 18.

Santana EA, Harrison A, Zhang X, Baker BD, Kelly BJ, White P, Liu Y, Munson RS, Jr. 2014. HrrF is the Fur-regulated small RNA in nontypeable *Haemophilus influenzae*. *PLoS One* **9**: e105644.

Schu DJ, Zhang A, Gottesman S, Storz G. 2015. Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. *EMBO J* **34**: 2557-2573.

Sharma CM, Papenfort K, Pernitzsch SR, Mollenkopf HJ, Hinton JC, Vogel J. 2011. Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol Microbiol* **81**: 1144-1165.

Sheidy DT, Zielke RA. 2013. Analysis and expansion of the role of the *Escherichia coli* protein ProQ. *PLoS One* **8**: e79656.

Shirzad Aski H, Tabatabaei M. 2016. Occurrence of virulence-associated genes in *Pasteurella multocida* isolates obtained from different hosts. *Microb Pathog* **96**: 52-57.

Shivachandra SB, Kumar AA, Amaranath J, Joseph S, Srivastava SK, Chaudhuri P. 2005. Cloning and characterization of *tbpA* gene encoding transferrin-binding protein (TbpA) from *Pasteurella multocida* serogroup B:2 (strain P52). *Vet Res Commun* **29**: 537-542.

Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TT, Hinton JCD, Vogel J. 2008. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* **4**: e1000163.

Sledjeski D, Gottesman S. 1995. A small RNA acts as an antisilencer of the H-NS-silenced *rcsA* gene of *Escherichia coli*. *PNAS* **92**: 2003-2007.

Smirnov A, Forstner KU, Holmqvist E, Otto A, Gunster R, Becher D, Reinhardt R, Vogel J. 2016. Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *PNASA* **113**: 11591-11596.

Smirnov A, Wang C, Drewry LL, Vogel J. 2017. Molecular mechanism of mRNA repression *in trans* by a ProQ-dependent small RNA. *EMBO J* **36**: 1029-1045.

Smith MN, Kwok SC, Hodges RS, Wood JM. 2007. Structural and functional analysis of ProQ: an osmoregulatory protein of *Escherichia coli*. *Biochemistry* **46**: 3084-3095.

Starr CR, Engleberg NC. 2006. Role of hyaluronidase in subcutaneous spread and growth of group A *Streptococcus*. *Infect Immun* **74**: 40-48.

Steen JA, Steen JA, Harrison P, Seemann T, Wilkie I, Harper M, Adler B, Boyce JD. 2010. Fis is essential for capsule production in *Pasteurella multocida* and regulates expression of other important virulence factors. *PLoS Pathog* **6**: 1000750.

Steenbergen SM, Lichtensteiger CA, Caughlan R, Garfinkle J, Fuller TE, Vimr ER. 2005. Sialic Acid metabolism and systemic pasteurellosis. *Infect Immun* **73**: 1284-1294.

Storz J, Lin X, Purdy CW, Chouljenko VN, Kousoulas KG, Enright FM, Gilmore WC, Briggs RE, Loan RW. 2000. *Coronavirus* and *Pasteurella* infections in bovine shipping fever pneumonia and Evans' criteria for causation. *J Clin Microbiol* **38**: 3291-3298.

Sy B, Wong J, Granneman S, Tollervey D, Gally D, Tree JJ. 2018. High-resolution, high-throughput analysis of Hfq-binding sites using UV crosslinking and analysis of cDNA (CRAC). *Methods Mol Biol* **1737**: 251-272.

Talley P, Snippes-Vagnone P, Smith K. 2016. Invasive *Pasteurella multocida* Infections - Report of Five Cases at a Minnesota Hospital, 2014. *Zoonoses Public Health* **63**: 431-435.

Tatum FM, Yersin AG, Briggs RE. 2005. Construction and virulence of a *Pasteurella multocida fhaB2* mutant in turkeys. *Microb Pathog* **39**: 9-17.

Thummeepak R, Kongthai P, Leungtongkam U, Sitthisak S. 2016. Distribution of virulence genes involved in biofilm formation in multi-drug resistant *Acinetobacter baumannii* clinical isolates. *Int Microbiol* **19**: 121-129.

Townsend KM, Boyce JD, Chung JY, Frost AJ, Adler B. 2001. Genetic organization of *Pasteurella multocida* cap Loci and development of a multiplex capsular PCR typing system. *J Clin Microbiol* **39**: 924-929.

Tree JJ, Granneman S, McAteer SP, Tollervey D, Gally DL. 2014. Identification of bacteriophage-encoded anti-sRNAs in pathogenic *Escherichia coli*. *Mol Cell* **55**: 199-213.

Urban JH, Vogel J. 2007. Translational control and target recognition by *Escherichia coli* small RNAs *in vivo*. *Nucleic Acids Res* **35**: 1018-1037.

Urbanowski ML, Stauffer LT, Stauffer GV. 2000. The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*. *Mol Microbiol* **37**: 856-868.

van Nues RW, Castro-Roa D, Yuzenkova Y, Zenkin N. 2015. Ribonucleoprotein particles of bacterial small non-coding RNA IsrA (IS61 or McaS) and its interaction with RNA polymerase core may link transcription to mRNA fate. *Nucleic Acids Res*. **44**: 2577-2592.

Vogel J, Luisi BF. 2011. Hfq and its constellation of RNA. *Nat Rev Microbiol* **9**: 578-589.

Wang B, Brand-Miller J. 2003. The role and potential of sialic acid in human nutrition. *Eur J Clin Nutr* **57**: 1351-1369.

Waters SA, McAteer SP, Kudla G, Pang I, Deshpande NP, Amos TG, Leong KW, Wilkins MR, Strugnell R, Gally DL et al. 2017. Small RNA interactome of pathogenic *E. coli* revealed through crosslinking of RNase E. *Embo J* **36**: 374-387.

Wilkie IW, Harper M, Boyce JD, Adler B. 2012. *Pasteurella multocida*: diseases and pathogenesis. *Curr Top Microbiol Immunol* **361**: 1-22.

Wilson BA, Zhu X, Ho M, Lu L. 1997. *Pasteurella multocida* toxin activates the inositol triphosphate signaling pathway in *Xenopus* oocytes via G(q)alpha-coupled phospholipase C-beta1. *J Biol Chem* **272**: 1268-1275.

Xiao G, Tang H, Zhang S, Ren H, Dai J, Lai L, Lu C, Yao H, Fan H, Wu Z. 2017. *Streptococcus suis* small RNA rss04 contributes to the induction of meningitis by regulating capsule synthesis and by inducing biofilm formation in a mouse infection model. *Vet Microbiol* **199**: 111-119.

Yakhnin AV, Baker CS, Vakulskas CA, Yakhnin H, Berezin I, Romeo T, Babitzke P. 2013. CsrA activates *flhDC* expression by protecting *flhDC* mRNA from RNase E-mediated cleavage. *Mol Microbiol* **87**: 851-866.

Zhang A, Altuvia S, Tiwari A, Argaman L, Hengge-Aronis R, Storz G. 1998. The OxyS regulatory RNA represses *rpoS* translation and binds the Hfq (HF-I) protein. *EMBO J* **17**: 6061-6068.

Chapter 2

Determination of the small RNA GcvB regulon in the Gram-negative bacterial pathogen *Pasteurella multocida* and identification of the GcvB seed binding region.

This chapter is based on Gulliver EL, Wright A, Deveson Lucas D, Mégroz M, Kleifeld O, Schittenhelm R, Powell D, Seemann T, Bulitta J, Harper M, Boyce JD. 2018. Determination of the small RNA GcvB regulon in the Gram-negative bacterial pathogen *Pasteurella multocida* and identification of the GcvB seed binding region. RNA 24:704–720. doi:10.1261/rna.063248.117. (Appendix 1)

Chapter 2: Determination of the small RNA GcvB regulon in the Gram-negative bacterial pathogen *Pasteurella multocida* and identification of the GcvB seed binding region.

2.1 Introduction

Pasteurella multocida is a Gram-negative, coccobacillus that is the causative agent of many economically important diseases, including fowl cholera, swine atrophic rhinitis, haemorrhagic septicaemia and various respiratory diseases of ungulates (Wilkie et al. 2012). *P. multocida* produces several virulence factors that are critical for the bacterium to cause disease. These include primary virulence factors, such as the polysaccharide capsule, lipopolysaccharide (LPS) and filamentous haemagglutinin as well as virulence-associated factors, such as proteins involved in iron and nutrient acquisition (Fuller et al. 2000; Bosch et al. 2002a; Harper et al. 2004; Boyce and Adler 2006). Appropriate regulation of these factors is likely critical for *P. multocida* survival. For example, during *P. multocida in vivo* growth, the bacteria must acquire and/or synthesise all necessary amino acids, many of which are not freely available in sufficient quantities (Boyce and Adler 2006). This requires the production of amino acid biosynthesis and transport proteins, the expression of which must be tightly regulated to ensure that there is a balance between energy input and expenditure.

Recently, we showed that the Hfq protein was essential for the appropriate expression of a range of proteins in the *P. multocida* serogroup A strain VP161, including those required for the biosynthesis of hyaluronic acid capsule which is a primary virulence factor (Mégroz et al. 2016). The Hfq protein is an RNA chaperone that directly interacts with particular small regulatory RNA (sRNA) molecules to facilitate their binding to specific mRNA targets. Non-coding sRNA molecules are generally 40-400 nucleotides long and regulate transcript/protein expression within bacteria by binding to target mRNA via complementary base pairing (Desnoyers et al. 2013). There is redundancy within the sRNA regulatory network, as one sRNA species may bind to many different mRNA targets and each mRNA target may be regulated by several sRNA species (Desnoyers et al. 2013). Depending on the type of interaction, the binding of a sRNA to a target mRNA may result in either inhibition or induction of protein production. The binding of the sRNA to the ribosome-binding site (RBS) of an mRNA target can block translation and therefore reduce protein production. Alternatively, sRNA binding can result in rapid mRNA degradation via induction of Ribonuclease E activity against double stranded RNA (Gottesman and Storz 2011). Less commonly, protein production can be enhanced via the binding of the sRNA to a natural secondary structure region in the

mRNA that normally acts to occlude the RBS. This sRNA-mRNA interaction leads to the unfolding of the secondary structure, allowing the ribosome greater access to the RBS in order to initiate translation (Gottesman and Storz 2011).

Comparative global transcriptomic and proteomic analyses of the *P. multocida* strain VP161 and an isogenic *hfq* mutant revealed that many genes displayed altered transcript expression, and/or altered protein production, when *hfq* was inactivated (Mégroz et al. 2016). Analysis of the transcriptional data also allowed for the identification of a number of intergenic regions encoding putative sRNAs (Mégroz, Boyce Laboratory, unpublished data). One putative sRNA identified in strain VP161, which is also encoded on the Pm70 genome (GenBank AE004439.1, position 652175 to 651999), exhibited high sequence identity to the Hfq-dependent sRNA GcvB. In *E. coli* and *Salmonella enterica* serovar Typhimurium (*S.* Typhimurium) GcvB has been shown to negatively regulate the production of proteins involved in amino acid transport and biosynthesis, such as the amino acid transporters ArgT, BrnQ, DppA, OppA, SstT, TppB and YaeC and the amino acid biosynthesis proteins GdhA, IlvC, IlvE, SerA and ThrL (Pulvermacher et al. 2008; Sharma et al. 2011). In *E. coli*, the expression of GcvB is intimately associated with the availability of glycine and GcvB expression is induced when nutrients, especially glycine, are abundant in the environment. The *gcvB* gene is adjacent to and transcribed divergently from *gcvA*, which encodes the GcvA protein that positively regulates both *gcvB* and the glycine cleavage operon *gcvTHP*. The activation of both the *gcvTHP* operon and *gcvB,* is repressed during growth in the absence of glycine due to the association between GcvA and the regulatory protein GcvR (Urbanowski et al. 2000). This interaction does not occur in the presence of glycine, leaving GcvA to act as an activator of *gcvB* and *gcvTHP* expression. Therefore, in *E. coli* and *S.* Typhimurium, during periods of low glycine abundance the decreased production of GcvB results in activation of the amino acid biosynthesis and transport proteins that are normally repressed by the GcvB sRNA (Urbanowski et al. 2000).

GcvB function has primarily been assessed in *E. coli* (Urbanowski et al. 2000; Pulvermacher et al. 2008; Coornaert et al. 2013), and *S.* Typhimurium (Sharma et al. 2011) with functional studies in other organisms limited to *Yersina pestis* (McArthur et al. 2006). *S.* Typhimurium *gcvB* mutants grow more slowly than the wild-type parent strain and *E. coli gcvB* mutants have a decreased ability to form biofilms (Sharma et al. 2007; Mika and Hengge 2014). Analysis of the *E. coli* and *S.* Typhimurium GcvB mRNA targets has facilitated the identification of a GcvB binding sequence (seed region), 5'-CACAACAT-3', that allows for base pairing between GcvB and its mRNA targets. The mRNA seed region is strongly conserved in the GcvB targets produced by both species (Sharma et al. 2007). The seed region sequence, 5'-AUGUUGUG-3', is

present in the GcvB expressed by both *S.* Typhimurium and *E. coli* (Sharma et al. 2011) and is the reverse complement of the seed region sequence present in the mRNA target molecules.

There is currently no information on the functional role of GcvB, its mRNA targets or mRNA binding interactions in any organisms from the *Pasteurellaceae* family. A bioinformatics screen of multiple genomes identified a putative GcvB homolog in *P. multocida* and a recent bioinformatics analysis of the related organism, *Actinobacillus pleuropneumoniae*, also identified a GcvB homolog and its expression was confirmed by Northern blotting (Sharma et al. 2007; Rossi et al. 2016). Another study in *Haemophilus influenzae* showed expression of GcvB was high when grown in the presence of primary normal human bronchial epithelial cells using RNA-seq (Baddal et al. 2015). However, neither study looked further into the function of GcvB. In this study, we report the characterization of GcvB in a highly pathogenic *P. multocida* strain and the identification of more than 30 targets. Furthermore, we identify the *P. multocida* GcvB seed region and use a two-plasmid green fluorescent protein (GFP) reporter system to confirm the binding interaction between *P. multocida* GcvB and one of its mRNA targets, *gltA*.

2.2 Materials and methods

2.2.1 Bacterial strains, media, plasmids and growth conditions

All bacterial strains and plasmids used in this study are listed in Table 2.1. *P. multocida* strains were routinely cultured in Heart Infusion (HI) broth (Oxoid). *E. coli* strains were routinely grown in Luria-Bertani (LB) broth (Oxoid). For solid media, 1.0-1.5 % (w/v) agar was added to the media. When required, media was supplemented with the appropriate antibiotics; Kanamycin (50 µg/mL), Spectinomycin (50 µg/mL) or Ampicillin (100 µg/mL).

2.2.2 DNA manipulations

Genomic DNA was purified using the Genomic DNA Extraction kit (RBC Bioscience Corp.) and plasmid DNA was extracted using NucleoSpin Plasmid Kit (Macherey-Nagel GmbH & Co KG), according to the manufacturer's instructions. For cloning experiments, restriction endonucleases (New England Biolabs) and ligase (Roche) were used according to the manufacturer's instructions. PCR amplifications were performed using Taq DNA polymerase (Roche) or Phusion high fidelity DNA polymerase (Roche) with oligonucleotides (primers) manufactured by Sigma-Aldrich. The primers used in this study are listed in Table 2.2, PCR products were purified using the NucleoSpin Gel and PCR Clean up Kit (Macherey-Nagel). Sequencing reactions where performed with whole genomic DNA, plasmid DNA or PCR products as template as previously described (Harper et al. 2013). DNA sequences were analysed using VectorNTI (Invitrogen), Clustal Omega (EBI) and BLAST (NCBI).

Table 2.1 Strains and plasmids used in this study

Strain or plasmid	Description	Source or Reference
Strains		
P. multocida strains		
AL2521	VP161 *hfq* TargeTron® mutant; Kan^R	(Mégroz et al. 2016)
AL2526	AL2521 containing pAL1108; Kan^R, Tet^R	(Mégroz et al. 2016)
AL2527	AL2521 containing pAL99T; Kan^R, Tet^R	(Mégroz et al. 2016)
AL2677	VP161 *gcvB* TargeTron® mutant; Kan^R	This study
AL2838	AL2521 containing pAL1266; Kan^R, Tet^R	This study
AL2862	AL2677 containing pPBA1100S	This study
AL2864	AL2677 containing pAL1190	This study
VP161	Virulent avian isolate ; Serotype A :1	(Wilkie et al. 2000)
E. coli strains		
AL1296	DH5α containing pAL99S	(Harper et al. 2013)
AL1995	DH5α containing pAL953	(Harper et al. 2013)
AL2224	DH5α containing pAL99T	(Harper et al. 2013)
AL2523	DH5α containing pAL1108	(Mégroz *et al.* 2016)
AL2659	DH5α containing pAL1170	This study
AL2678	DH5α containing pPBA1100S	This Study
AL2680	DH5α containing pAL1190	This study
AL2708	DH5α containing pREXY	This study
AL2713	DH5α containing pAL1197	This study
AL2772	DH5α containing pAL1240	This study
AL2781	DH5α containing pTEXY	This study
AL2789	DH5α containing pAL1257	This study
AL2799	DH5α containing pAL1257 & pREXY	This study
AL2805	DH5α containing pAL1257 & pAL1197	This study
AL2832	DH5α containing pAL1266	This study
AL2900	DH5α containing pAL1277	This study
AL2922	DH5α containing pAL1257 & pAL1277	This study
AL2944	DH5α containing pAL1290	This study

AL2953	DH5α containing pAL1290 & pAL1197	This study
AL2955	DH5α containing pAL1290 & pAL1277	This study
DH5α	*deoR endA1 gryA96 hsdR17*($r_k^- m_k^+$) *recA1 relA1 supE44 thi-1* (*lacZYA-argFV169*) φ80*lacZ* ΔM15, F^-	Bethesda Research Laboratory
Plasmids		
pAL99S	*P. multocida/E. coli* expression plasmid, $Spec^R$	(Harper et al. 2013)
pAL99T	*P. multocida/E. coli* expression plasmid, Tet^R	(Harper et al. 2013)
pAL953	*P. multocida* plasmid ($Spec^R$) containing TargeTron® group II intron with *aph3* (Kan^r). Targeted to *gatA*	(Harper et al. 2013)
pAL1108	pAL99T containing *hfq* from *P. multocida* VP161; Tet^R	(Mégroz et al. 2016)
pAL1170	*P. multocida* TargeTron® plasmid targeted to *gcvB;* Kan^R, $Spec^R$	This study
pAL1190	pPBA1100s containing *gcvB* from *P. multocida* VP161; $Spec^R$	This study
pAL1197	pREXY containing *gcvB* from *P. multocida* VP161 (pREXY::*gcvB*); $Spec^R$	This study
pAL1240	Modified pBR322 (Sigma-Aldrich) vector. BsgI site replaced with BssHII site, Amp^R, Tet^R	This study
pAL1257	pTEXY containing 5' end (38 bp upstream and 60 bp downstream of start codon) of VP161 *gltA*, fused to *sfGFP* (pTEXY::*gltA-sfGFP*); Amp^R	This study
pAL1266	*hfq* from *E. coli* DH5α cloned into pAL99T; Tet^R	This study
pAL1277	pREXY containing modified *gcvB* from *P. multocida* VP161. Has a 7 bp substitution within the seed binding region (pREXY::$gcvB_{MSR2}$); $Spec^R$	This study
pAL1290	pTEXY containing modified 5' end (38 bp upstream and 60 bp downstream of start codon) of *gltA* fromVP161. Has a 7 bp substitution within the seed binding region. Gene fusion with *sfGFP* (pTEXY::$gltA_{MSR1}$-*sfGFP*); Amp^R	This study
pBR322	*E. coli* cloning vector; Amp^R, Tet^R	Sigma-Aldrich
pPBA1100S	Modified pAL99S. The EcoRI fragment (240 bp) containing the *P. multocida* P_{tpi} promoter removed; $Spec^R$	This study
pCR2.1	*E. coli* cloning vector with 3'-T overhangs for TOPO cloning, Amp^R, Kan^R	ThermoFischer Scientific
pREXY	*P. multocida/E. coli* sRNA expression plasmid. pPBA1100S with a 96 bp fragment containing the *P. multocida* VP161 constitutive P_{tpi} promoter cloned into BamHI/HindIII, $Spec^R$	This study
pTEXY	GFP reporter/mRNA expression plasmid for sRNA/mRNA interaction studies. HindIII/EcoRI fragment containing commercially synthesized *sfGFP* and promoter, $P_{LtetO-1}$, cloned into HindIII/EcoRI sites of modified pBR322 (pAL1240); Amp^R, Tet^R	This study

Table 2.2 Oligonucleotides used in this study

Name	Sequence (5'-3')	Description
BAP612	GTAAAACGACGGCCAGT	pCR2.1 vector-specific primer used for sequencing across cloning site
BAP7565	TGAACGCAAGTTTCTAATTTCGATTTTGGTTCGATAGAGGAAA GTGTCT	EBS2 TargeTron® primer specific for *gcvB* mutagenesis
BAP7566	AAAAAAGCTTATAATTATCCTTAACCAACCAGAGTGTGCGCCC AGATAGGGTG	IBS TargeTron® primer specific for *gcvB* mutagenesis
BAP7567	CAGATTGTACAAATGTGGTGATAACAGATAAGTCCAGAGTAAT AACTTACCTTTCTTTGT	EBS1 TargeTron® primer specific for *gcvB* mutagenesis
BAP7585	CTCAATGGATCCTTTTTGATCTATAATATAGCG	Forward primer located upstream of VP161 *gcvB*; contains a BamHI site
BAP7586	AATCGGGTCGACAGCAATGTGAGCAGGTCTATG	Reverse primer located downstream of VP161 *gcvB*; contains a SalI site
BAP7632	CTTACCCGGGTCTAAAATGCGCGCATACTTAATG	Forward primer located upstream of VP161 *gcvB*; contains an XmaI site
BAP7633	AATCCCCGGGATAAAAAAACACCGCTCAATAGAGC	Reverse primer located downstream of VP161 *gcvB*; contains an XmaI site
BAP7638	ATAATCAAGCTTTCACAAAAAATTTTTTTAAATTTGCC	Forward primer for amplification of P_{tpi} promoter in VP161; contains a HindIII site
BAP7639	TAATTTGGATCCTATTAAAGTAATAAAAAATAACCGC	Reverse primer for amplification of P_{tpi} promoter in VP161; contains a BamHI site
BAP7719	AAGTCAGCGCGCACCATTATGTTCCGG	Forward primer for amplification of nucleotides 1653 to 4358 in pBR322. Anneals to region surrounding BsgI. Contains a BssHII site to replace BsgI.
BAP7720	AACATAATGGTGCGCGCTGACTTCCG	Reverse primer for amplification of nucleotides 1 to 1656 in pBR322. Anneals to region surrounding BsgI. Contains a BssHII site to replace BsgI.
BAP7721	CTTCAAGAATTCTCATGTTTGACAGC	Forward primer for amplification of nucleotides 1 to 1656 in pBR322. Anneals to region surrounding EcoRI. Contains an EcoRI site.
BAP7722	AACATGAGAATTCTTGAAGACGAAAGG	Reverse primer for amplification of nucleotides 1653 to 4358 in pBR322. Anneals to region surrounding EcoRI. Contains EcoRI site.
BAP7747	GTTAATTCTAGAGCAATAAAACACAACTTACTAAAAAC	Forward primer for amplification of VP161 DNA representing the region -38 to + 60 (relative to start codon) of the *P. multocida gltA*. Contains an XbaI site
BAP7748	AGAGCCAGATCTTACAGGTAGGTCATATTCACGTC	Reverse primer for amplification of VP161 DNA representing the region -38 to + 60 (relative to start codon) of the *P. multocida gltA*. Contains a BglII site
BAP7754	GATAAGTTACTCTGTTTGGTTTCCCAAA	Reverse primer for amplification of *gcvB* in 5' RACE inner PCR. Anneals 55 bp from *gcvB* start site
BAP7850	TGTTGGGGATCCCGCAGGCTGAATGTGTACAATTGAGACGTAT CGTGCG	Forward primer anneals 107 bp upstream of *hfq in E. coli*. Contains a BamHI site
BAP7851	GGGAACGTCGACTCGCTGGCTCCCCGTG	Reverse primer anneals 48 bp downstream of *hfq in E. coli*. Contains a SalI site
BAP7888	TGTTTGCATATTGTTTGGGAA	Forward internal *gcvB* primer, anneals 55 bp from start codon. Used to generate Northern blot probe.
BAP7889	GAGCGGTGTTTAACCAAAAGG	Reverse primer for *gcvB* amplification in 5' RACE outer PCR. Anneals 162 bp from *gcvB* start site

BAP7950	CGGACTTAAGTATGATCAACACGTTGCATATTGTTTGGG	Forward primer, anneals to *gcvB* sequence at the seed region. For SOE PCR. Contains a 7 nucleotide substitution of the seed region.
BAP7951	CAATATGCAACGTGTTGATCATACTTAAGTCCGAAACTCTTAACC	Reverse primer, anneals to *gcvB* sequence at the seed region. For SOE PCR. Contains a 7 nucleotide substitution of the seed region.
BAP7962	6-FAM-CTCTGTTTGGTTTCCCAAACAATATGC	Reverse primer located within *gcvB*, approximately 90 bp from the transcript start. Used for primer extension and contains a 5'- fluorescein amidite (FAM) label
BAP7964	GTTAATTCTAGAGCAATAAATGTGTTGTTACTAAAAAC	Reverse primer for amplification of VP161 DNA representing the region -38 to + 60 (relative to start codon) of the *P. multocida gltA*. Incorporates a 7 nucleotide substitution at the putative seed binding region and a BglII site.
BAP7957	TAATACGACTCACTATAGGGGAGCGGTGTTTAACCAAAAGG	Reverse primer for amplification of a 103 bp *gcvB* fragment. Used to generate Northern blotting probe. Contains a T7 RNA polymerase promoter sequence (20 bp) at the 5' end.
BAP8153	TAATACGACTCACTATAGGGTTAATGATTGGTAATTCCTTACTGGTTAAGA	Forward primer for amplification of *gcvB,* containing a T7 RNA polymerase promoter sequence (20 bp) and the 5' end
BAP8155	AAAAAAACCCGCTCAATAGGC	Reverse primer for amplification of *gcvB*
BAP8157	TAATACGACTCACTATAGGGGACACTGACTCTTTTAAGCTTTATAGTTAATTAAAATGCAATAAAACACAACTT	Forward primer for amplification of *gltA,* containing a T7 RNA polymerase promoter sequence (20 bp) and the 5' end
BAP8158	AAAGGCATTGTATTAAGAATGCACCTGCC	Reverse primer for amplification of *gltA*
BAP8166	AAGTATCCCAAATTTCCCTCTATTTAAAGAAAACGG	Forward primer for amplification of *gatA,* containing a T7 RNA polymerase promoter sequence (20 bp) and the 5' end
BAP8167	TAATACGACTCACTATAGGGAAAAAAGTTTTCTCAAACAGACCGCACTTTG	Reverse primer for amplification of *gatA*
BAP8190	6-FAM-GGGTCTAAAATAAATATACAGGAAGTGAAAA	Reverse primer located within *gcvB*, approximately 140 bp from the transcript start. Used for primer extension and contains a 5'- fluorescein amidite (FAM) label
EBS universal	CGAAATTAGAAACTTGCGTTCAGTAAAC	TargeTron® universal primer for re-targeting of intron.

2.2.3 Construction of a *P. multocida gcvB* mutant

To inactivate *gcvB* in the *P. multocida* strain VP161, TargeTron® mutagenesis (Sigma-Aldrich) was used as previously described (Steen et al. 2010) but with the following modifications. The group II intron within the *E. coli-P. multocida* TargeTron® shuttle vector, pAL953 (Harper et al. 2013), was retargeted to *gcvB* using the PCR amplification method described in the TargeTron® manual. The primers BAP7565, BAP7566 and BAP7567 (Table 2.2) were designed using the TargeTron® design site (Sigma-Aldrich). The resulting plasmid, pAL1170 (Table 2.1), was used to transform *P. multocida* strain VP161 by electroporation and mutants containing a TargeTron® group II intron insertion in *gcvB* were identified as previously described (Harper et al. 2013).

2.2.4 Construction of a *P. multocida* GcvB overexpression strain

The *gcvB* gene from *P. multocida* VP161 was amplified using BAP7585 and BAP7586 (Table 2.2), digested with BamHI and SalI, and cloned into similarly digested pPBA1100s. The resulting plasmid, pAL1190, and the empty vector pPBA1100s were separately used to transform the *P. multocida gcvB* mutant AL2677 via electroporation, producing strains AL2864 and AL2862 respectively (Table 2.1).

2.2.5 Heterologous expression of the *E. coli hfq* gene in *P. multocida*

The *hfq* gene from *E. coli* DH5α was amplified using BAP7850 and BAP7851 (Table 2.2), digested with BamHI and SalI then cloned into the *P. multocida* expression plasmid, pAL99T. The resulting plasmid, pAL1266, was used to transform the *P. multocida hfq* mutant AL2521 (Mégroz et al. 2016), producing the strain AL2838.

2.2.6 Hyaluronic acid capsule assay

P. multocida strains were grown in HI broth (in biological triplicate) supplemented with the appropriate antibiotics (where required) to mid-exponential growth phase (OD_{600} = 0.6). Capsule was extracted from washed cells and the amount of capsular material measured using a hyaluronic acid assay as described previously (Chung et al. 2001).

2.2.7 Response to acid stress

Acidic HI broth was prepared by addition of 37% (v/v) hydrochloric acid (HCl) to HI until pH 4.6 was reached. Triplicate overnight cultures were prepared for each *P. multocida* strain and supplemented with Kanamycin where required to maintain the plasmid. Each culture was diluted 1:100 in fresh HI broth and grown until early exponential phase (OD_{600} = 0.2). A 1 mL aliquot of this early exponential phase (OD_{600} = 0.2) culture was then added to 3 mL of acidified HI broth, without antibiotics, and incubated at 37^{o}C for

15 min with shaking. Following incubation, 12 mL of basic HI broth was added to neutralise the culture. Appropriate dilutions of each culture were plated onto HI agar in duplicate and after 16 h incubation colonies were enumerated.

2.2.8 Biofilm formation assay

Cultures representing each bacterial strain were grown to mid-exponential growth phase (OD_{600} = 0.6) then 100 µL of the diluted culture (1:100) was added to four wells of a sterile 96 well plate which was then incubated overnight at 37^{o}C without shaking to allow for biofilm formation. Following incubation, the plate was washed three times with dH_2O to remove planktonic bacteria. Remaining bacteria were stained with 125 µL of 0.1% (w/v) crystal violet and incubated for 10 min at room temperature. Excess stain was removed by washing with dH_2O three times. To resolubilize the crystal violet, 200 µL of 95% (v/v) EtOH was added to each well, incubated for 15 min and then mixed well. A 125 µL aliquot of each well was transferred to a well of an optically clear flat bottomed 96 well plate and the optical density determined using a Tecan Infinite M200 plate reader.

2.2.9 RNA extraction, qRT-PCR and whole-genome transcriptomic analyses by RNA-seq

P. multocida RNA extractions were performed as described previously (Boyce et al. 2002) but with the following modifications. Duplicate bacterial cultures were grown in HI broth to OD_{600} = 0.2 (early-exponential growth phase), OD_{600} = 0.6 (mid-exponential growth phase) or OD_{600} = 1.0 (late-exponential growth phase). Killing buffer was omitted from the RNA extraction method. Following DNase treatment of the samples, RNA was further purified by phenol: chloroform extraction using 5Prime phase lock gel tubes as per the manufacturer's instructions (Quanta Biosciences). RNA-seq library preparation, sequencing on an Illumina HiSeq, and mapping and differential expression analysis was carried out as previously described (Mégroz et al. 2016). For the RNA-seq analyses, the average number of reads mapped across samples was 5,647,690.83, and of these an average of 99.7% mapped to the *P. multocida* VP161 genome, giving an average read depth of 2701.76 reads per gene. qRT-PCR was performed using the Affinity script cDNA synthesis kit (Agilent) and Brilliant II SYBR green qPCR kit (Agilent) as per the manufacturer's instructions using the Eppendorf Realplex mastercycler. Reverse transcription reactions, both with and without reverse transcriptase (+RT and -RT respectively), were performed in biological triplicate, with each +RT reaction being measured in technical triplicate and each -RT reaction being measured in technical duplicate. Melt curve data was analysed to confirm that only a single product was formed in each reaction and –RT controls did not amplify any products within 10 cycles of the experimental reactions.

2.2.10 Northern blotting

Northern blotting analysis was performed using the DIG Northern starter kit version 10 (Roche) as per the manufacturer's instructions, with the following modifications. A total of 8 µg of RNA was separated by agarose/formaldehyde gel electrophoresis and the separated products transferred to a nylon membrane by capillary electrophoresis. A GcvB-specific probe was amplified from *P. multocida* VP161 genomic DNA using BAP7888 and BAP7957; BAP7957 contains a T7 RNA polymerase promoter sequence at the 5' end. The PCR product was then used in an *in vitro* transcription reaction using T7 RNA polymerase and 10X DIG labelled RNA mix (Promega).

2.2.11 Proteomics analysis

Total proteomes of the wild-type *P. multocida* VP161 and the *gcvB* mutant (in triplicate), were determined using nano-liquid chromatography coupled with tandem MS, following isotopic labelling with heavy and light formaldehyde as described previously (Mégroz et al. 2016).

Total proteomes of the wild-type *P. multocida* VP161, the *gcvB* mutant with empty vector and the GcvB overexpression strain were determined using label-free quantitative proteomics. Cells were grown in biological triplicate in HI broth to early-exponential growth phase (OD_{600} = 0.2) and pelleted by centrifugation. Cell pellets were lysed in 1% w/v SDC (sodium deoxycholate; Sigma), 100 mM Tris (pH = 8.1) and further homogenised on a Soniprep 150 Plus sonicator (MSE). The protein concentration was determined using a BCA assay kit (Pierce). A 200 µg aliquot of each total protein sample was denatured using 10 mM TCEP (Thermo Scientific) and free cysteine residues were alkylated with 40 mM chloroacetamide (Sigma). Trypsin Gold (Promega) was used to digest the proteins and SDC removed by extraction with water-saturated ethyl acetate. All samples were desalted using P-10 ZipTip columns (Agilent, OMIX-Mini Bed 96 C18), vacuum-dried and reconstituted in buffer A (0.1% formic acid, 2% acetonitrile) prior to mass spectrometry.

Using a Dionex UltiMate 3000 RSLCnano system equipped with a Dionex UltiMate 3000 RS autosampler, the samples were loaded via an Acclaim PepMap 100 trap column (100 µm x 2 cm, nanoViper, C18, 5 µm, 100Å; Thermo Scientific) onto an Acclaim PepMap RSLC analytical column (75 µm x 50 cm, nanoViper, C18, 2 µm, 100Å; Thermo Scientific). The peptides were separated using increasing concentrations of buffer B (80% acetonitrile / 0.1% formic acid) for 158 min and analyzed with a QExactive Plus mass spectrometer (Thermo Scientific) operated in data-dependent acquisition mode using in-house, LFQ-optimized parameters.

Acquired .raw files were analysed with MaxQuant (Cox et al., 2008) to globally identify and quantify proteins across the various conditions. Statistical analyses for identification of differentially produced proteins were performed using the Limma package within R studio, where FDR is derived from the Benjamini-Hochberg

procedure. Differentially produced proteins were identified as proteins with a ≥ 0.59 $\log_2$ fold-change (1.5-fold) and an FDR ≤ 0.05. The proteomics data have been deposited in ProteomeXchange via the PRIDE database with identifier PXD007719.

2.2.12 Fluorescent primer extension

Fluorescent primer extension was performed as described previously (Lloyd et al. 2005; Steen et al. 2010) with the following modifications. RNA was isolated from *P. multocida* VP161 at OD_{600} = 0.2. For cDNA synthesis, 10 µg of total RNA was used as template with the 6-carboxy fluorescein amidite (6-FAM) labelled primer, BAP7962 or BAP8190 (Table 2.2). Dried samples were analysed using an ABI 3730xl DNA Analyzer (Thermo Fisher Scientific) located at Australian Genome Research Facility (AGRF, Melbourne).

2.2.13 5' RACE

5' RACE was performed with 10 µg of RNA isolated from *P. multocida* VP161 using the Firstchoice® RLM-RACE Kit (Applied Biosystems) according to the manufacturer's instructions with the following modifications. The reverse transcription step was replaced with the cDNA synthesis protocol used for fluorescent primer extension (above) using the non-fluorescent GcvB-specific primer BAP7889. The cDNA generated was resuspended in 30 µl of nuclease-free water; 1 µl was used in the first, nested PCR using the primer BAP7889 together with the commercially supplied 5' RACE outer primer (Applied Biosystems). PCR reaction conditions were as follows; 94°C for 3 min, followed by 35 cycles consisting of 94°C 30 sec, 62°C 30 sec, 72°C 1 min, followed by a final extension step of 72°C for 7 min. The PCR product was then purified and 1 µl was used in the second, nested PCR using the primer BAP7754 and the commercially supplied 5' RACE inner primer (Applied Biosystems) with the same PCR reaction conditions as described above. The nested PCR products generated were cloned into the vector pCR2.1 using the TOPO TA cloning kit (Thermo Fischer Scientific) according to the manufacturer's instructions. The nucleotide sequences of the cloned inserts were then determined using the vector-specific primer BAP612 in Sanger sequencing reactions.

2.2.14 Co-immunoprecipitation of GcvB by Hfq

To test whether *P. multocida* GcvB bound Hfq, we used co-immunoprecipitation of total bacterial RNA by a FLAG-tagged Hfq, followed by high-throughput sequencing of the precipitated RNAs. Total RNA was prepared from *P. multocida* expressing a chromosomally-encoded, C-terminal 3xFLAG-tagged Hfq and as a control also from the wild-type *P. multocida* expressing native Hfq. FLAG-tagged Hfq, and any bound RNAs, were precipitated (three independent co-immunoprecipitation reactions) using anti-FLAG conjugated magnetic beads as previously described (Bilusic et al. 2014). RNA-seq library preparation,

sequencing on a NextSeq (Illumina), and mapping and differential expression analysis was carried out as previously described (Mégroz et al. 2016).

2.2.15 Construction of plasmids for the two-plasmid GFP reporter assays

To analyse *P. multocida* GcvB/*gltA* mRNA target interactions, a two-plasmid GFP reporter system was developed based on the previously described system of Urban and Vogel (Urban and Vogel 2007). This system required the construction of two expression vectors, pTEXY, required for the expression of the 5' end of the mRNA target, containing the GcvB seed/binding region fused to a gene encoding sfGFP (Corcoran et al. 2012), and pREXY, required for the expression of the sRNA molecule, GcvB. To generate pTEXY the unique BsgI site present in the *E. coli* plasmid pBR322 was first changed to a BssHII site using site-directed PCR mutagenesis to allow for future experiments that required this restriction site to be uniquely located in the mRNA-encoding DNA fragments. Two PCR products representing the pBR322 nucleotides 1 to 1656 (position of BsgI site) and nucleotides 1653 to 4358 were amplified by PCR. The first PCR reaction amplified the pBR322 nucleotides 1 to 1656 using BAP7721, which anneals to the EcoRI region and BAP7720, which anneals to the BsgI region but contains an altered sequence to incorporate a BssHII site instead of BsgI. The second PCR reaction amplified the pBR322 nucleotides 1653 to 4358 using BAP7722, which anneals to the EcoRI region and BAP7719, which anneals to the BsgI region but contains a BssHII site instead of BsgI. The PCR products were digested with EcoRI and BssHII, ligated, and the mixture used to transform competent *E. coli* DH5α, to generate the plasmid, pAL1240 (Table 2.1). To construct pTEXY, a pMAT plasmid containing a commercially synthesized DNA fragment (Life Technologies) encoding the superfolder GFP (*sfGFP*) gene (flanked by a HindIII and EcoRI restriction sites and under the control of the anhydrotetracycline (Atc)-inducible promoter, $P_{LtetO-1}$), was digested with HindIII and EcoRI. The DNA fragment containing *sfGFP* was then gel-purified and ligated to HindIII and EcoRI-digested pAL1240 to generate the plasmid pTEXY (Table 2.1).

The final expression plasmid containing the *gltA-sfGFP* fusion (pAL1257) was constructed by generating an XbaI/BglII-digested PCR fragment, using the primers BAP7747 and BAP7748, that represented the region -38 to + 60 (relative to start codon) of the *P. multocida gltA*. The PCR fragment was digested, purified then ligated into the XbaI and BglII-digested pTEXY (Table 2.1), such that expression would be under the control of the $P_{LtetO-1}$ promoter. The pTEXY plasmid, pAL1290, containing *gltA* with a mutated seed region ($gltA_{MSR1}$-*sfGFP*) was constructed in a similar manner with the exception that the forward primer BAP7964, containing the altered seed region sequence, was used for the amplification of the *gltA*-specific DNA (Table 2.2).

To generate the base plasmid, pPBA1100S, used for the construction of sRNA expression plasmid pREXY, the DNA region (240 bp) encoding the P_{tpi} promoter was removed from the pAL99S vector (Harper et al. 2013) using EcoRI digestion followed by re-ligation of the vector. This region was then replaced with a shorter DNA fragment (96 bp) containing the P_{tpi} promoter, amplified from *P. multocida* VP161 genomic DNA using BAP7638 and BAP7639 (containing HindIII and BamHI restriction sites, respectively), to ensure that transcription could begin as close as possible to the native sRNA (GcvB) start site. The BamHI/HindIII-digested PCR product was ligated to similarly-digested pPBA1100S and the ligation mix used to transform *E. coli* DH5α. Recombinant colonies were selected on HI agar containing 50 µg/µl spectinomycin. Correct recombinant plasmids were confirmed by restriction analysis and DNA sequencing and one plasmid with the correct sequence designated pREXY (Table 2.1).

The GcvB expression plasmid, pAL1197 (Table 2.1), was constructed as follows. The region of the *P. multocida* VP161 genome encoding the putative *gcvB* was amplified from VP161 genomic DNA using the primers BAP7632 and BAP7633 (both containing an XmaI site). The purified, XmaI-digested PCR product was then ligated to XmaI-digested pREXY. The authenticity of the pAL1197 plasmid containing *gcvB*, was confirmed by PCR and DNA sequencing. The pREXY plasmid containing the mutated $gcvB_{MSR2}$ (pAL1277, Table 2.1) was constructed using splice overlap extension (SOE) PCR. Two PCR reactions were performed as follows. The reverse primer, BAP7951 and the forward primer BAP7950 (Table 2.2), that overlap and anneal to the *gcvB* gene region containing the predicted seed region, were paired with BAP7632 (forward primer located upstream of *gcvB*) and BAP7633 (reverse primer located downstream of *gcvB*), respectively. Primers BAP7951 and BAP7950 contained sequence that changed the *gcvB* seed region from 5'-GTTGTGT-3' to 5'-CAACACA-3'. The two PCR fragments, representing the 5' and 3' ends of *gcvB*, were combined using a second PCR amplification with primers BAP7632 and BAP7633 (Table 2.2) to produce the $gcvB_{MSR2}$ fragment. The PCR product was then purified, digested with XmaI, and ligated to XmaI-digested pREXY.

2.2.16 Whole-cell fluorescent measurements

E. coli DH5α strains containing both a pTEXY-based plasmid (5'mRNA-*sf*GFP fusion expression) and a pREXY-based plasmid (sRNA expression) were grown on LB agar supplemented with 50 µg/µl ampicillin and 50 µg/µl spectinomycin. Cells representing each strain were harvested from each plate (biological triplicate), resuspended in 1 mL of 1 x PBS buffer, and the volume adjusted to give a final OD_{600} of 2.0. A 200 µL aliquot of each cell suspension was added to a black flat bottomed 96-well microtiter plate (in triplicate). Fluorescence was measured using the Tecan Infinite M200 plate reader with an excitation/emission wavelength of 475/540 nm.

2.2.17 *In vitro* transcription

DNA from wild-type *P. multocida* VP161 was used as a template to amplify DNA fragments of *gcvB* (using BAP8153 and BAP8155), *gltA* (using BA8157 and BAP8158) and *gatA* (using BAP8166 and BAP8167) to be transcribed. The *gltA* and *gatA* fragments were used in an *in vitro* transcription reaction using the HiScribe™ T7 High Yield RNA Synthesis Kit (NEB) following the manufacturer's instructions. The *gcvB* fragment was transcribed using the HiScribe™ T7 High Yield RNA Synthesis Kit (NEB) but following the instructions for labelled nucleotides where 5' biotin labelled GMP was added in a 5:1 of 5'GMP to GTP.

2.2.18 Electrophoretic mobility shift assays (EMSA)

EMSA were performed as described previously (Morita et al. 2012) with the following modifications. Samples were mixed with 0.1 volume loading buffer consisting of 25% glycerol, 0.1% bromophenol blue in 0.5M Tris-Cl (pH6.8). The electrophoresed samples were transferred to a nylon membrane in 0.5 X Tris-borate-EDTA (TBE) buffer at 380 mA for 30 min. the RNA was then UV-crosslinked to the membrane via a 1 min exposure at 0.12 joules in a UV-transilluminator (UVITEC). The membrane was then washed for 2 min in wash buffer consisting of 0.003% Tween 20 in 1 x maleic acid buffer, (100 mM maleic acid, 375 mM NaCl and 457.5 mM NaOH, pH 7.5), incubated for 30 min in 25 mL blocking buffer consisting of 5 mL DIG-block (Roche) and 45 mL 1 x maleic acid buffer, then incubated in antibody solution (25 ml blocking buffer and 5 µl Precision Protein™ StrepTactin-HRP Conjugate (Bio-Rad)) for 1 h. The membrane was then washed 2 x 15 min in wash buffer (1 x maleic acid buffer and 0.3% Tween 20) and bands detected using the Amersham ECL Western Blotting Detection Kit (GE Healthcare Life).

2.2.19 Bioinformatic analysis

Comparison of nucleotide and protein sequences was performed using BLAST (Camacho et al. 2009). The Rfam database version 12.2 (Burge et al. 2013) was used to compare the *P. multocida gcvB* sequence to known RNAs. The MEME motif identification website with MEME motif finder version 4.11.2 (Bailey et al. 2009) was used to identify potential GcvB binding sites in putative mRNA targets and these were then further analysed using Clustal Omega (Sievers et al. 2011). Sequence data was analysed, and recombinant DNA molecules were designed using VectorNTI version 11 (Invitrogen). GcvB-regulated genes were mapped to the appropriate metabolic pathways using SmartTables (Travers et al. 2013) and pathway overview (Paley and Karp 2006; Karp et al. 2010) within the Biocyc database collection website (Caspi et al. 2016). Metabolic pathways were visualised using flow charts obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al. 2016). The Mfold webserver was used with default parameters to determine RNA secondary structures (Zuker 2003).

2.3 Results

2.3.1 Confirmation of GcvB expression in *P. multocida* using high-throughput transcriptomic analysis, Northern blotting and GcvB transcript analyses.

Previous bioinformatics analyses (Sharma et al. 2007) have identified a putative GcvB in *P. multocida* that contains the conserved R1 and R2 sequences common to all GcvB sRNA molecules (Figure 2.1). The *P. multocida gcvB* was located between *ivlE* (encoding a branched-chain amino acid aminotransferase) and *gcvA* (Figure 2.2A). To determine if GcvB is expressed in *P. multocida,* we analysed whole transcriptome RNA-sequencing (RNA-Seq) data generated from RNA isolated from *P. multocida* VP161 grown until the cultures reached an optical density at 600 nm (OD_{600}) of 0.2, 0.7 and 1.0, representing early-exponential, mid-exponential and late-exponential growth phases in biological duplicate. The putative *P. multocida* GcvB sRNA was expressed strongly during early-exponential and mid-exponential growth, with an average of 1490 and 2514 GcvB transcripts per million (TPM) total transcripts, respectively (Figure 2.2A). However, GcvB expression was reduced significantly by late-exponential growth (10-fold reduced expression compared to early exponential phase and 13-fold reduced expression compared to mid-exponential phase; false discovery rate [FDR] < 0.01) when only limited amounts of GcvB transcripts were produced (average of 209 GcvB TPM). The growth phase-dependent expression of the GcvB sRNA in *P. multocida* strain VP161 was confirmed by Northern blotting using a GcvB complementary strand-specific, RNA probe. The probe hybridized strongly with a fragment of the predicted size (~180 bp) of the GcvB sRNA transcript in the RNA isolated from VP161 cells in early-exponential growth phase, but only very weakly to RNA isolated from cells grown to late-exponential growth phase (Figure 2.2B).

Analysis of the *P. multocida* GcvB sequence revealed a putative rho-independent transcriptional terminator that corresponded to stem-loop 5 (SL5) present in the GcvB of *E. coli* and S. Typhimurium (Sharma et al. 2007)(Figure 2.1). The position of this putative stem-loop corresponded closely with the end of the RNA-seq transcript peak (Figure 2.2A) and we predict that this stem-loop defines the 3' end of the *P. multocida* GcvB. To determine the 5' end of the *gcvB* transcript, two independent methods were employed, primer extension and 5' RNA ligase mediated rapid amplification of cDNA ends (5' RLM-RACE). For primer extension experiments, RNA isolated from *P. multocida* VP161 was used as the template for cDNA synthesis with the fluorescently-labelled primer BAP7962 or primer BAP8190 (Table 2.2) that anneal ~90 bp and ~140 bp, respectively, from the predicted start of the *gcvB* transcript (as determined by the RNA-seq transcriptomic analyses). Fragment size analysis of the generated cDNA molecules identified a fragment of 87 nucleotides in length for the primer extension using BAP7962 and 140 nucleotides in length for primer extension using BAP8190. This data indicated that the *P. multocida* VP161 GcvB transcript

started with the sequence 5'-CUUAAUG-3', plus or minus the 5' C, which corresponds to the second nucleotide in the GcvB sequence from *E. coli* and *S.* Typhimurium (Figure 2.1). To determine if this was a *bone fide* transcript initiation site, we used 5' RLM-RACE. *P. multocida* VP161 RNA was first treated with calf intestine alkaline phosphatase and tobacco acid pyrophosphatase and then used as the template in nested PCRs to generate 5' adapter-ligated GcvB DNA fragments, which were then cloned into the plasmid pCR2.1. DNA sequencing of these cloned fragments using a vector-specific primer (BAP612) revealed that the *P. multocida* GcvB transcriptional start site was located two base pairs upstream of the *E. coli* and *S.* Typhimurium GcvB start sites (Figure 2.1A and 1B). Therefore, these data indicate that the *P. multocida* GcvB transcript begins with 5'-AUACUUAAUG-3'.

The secondary structure of the *P. multocida* GcvB was modelled (Figure 2.3A) using the Mfold webserver (Zuker 2003). While the predicted structure is very similar to the experimentally determined structure of the *S.* Typhimurium GcvB (Sharma et al. 2007), there are some notable differences. These include the observation that the stem-loop 1 (SL1) in the *P. multocida* GcvB is predicted to be significantly shorter than the SL1 in the *S.* Typhimurium GcvB; sequence alignment of the GcvB sRNAs from *P. multocida*, *E. coli*, *S.* Typhimurium and *Y. pestis* confirmed that the 5' region of the *P. multocida* GcvB is indeed shorter (Figure 2.1). The predicted *P. multocida* GcvB structure contains the SL2 and SL3 stem-loops between the conserved R1 and R2 G/U-rich linker regions, as is observed in the *S.* Typhimurium GcvB. However, the *P. multocida* GcvB has no predicted SL4 stem-loop, but rather the region between SL1 and R1 shows high complementarity to the region between R2 and SL5 and may form a long double-stranded section, although this remains to be experimentally verified (Figure 2.3A).

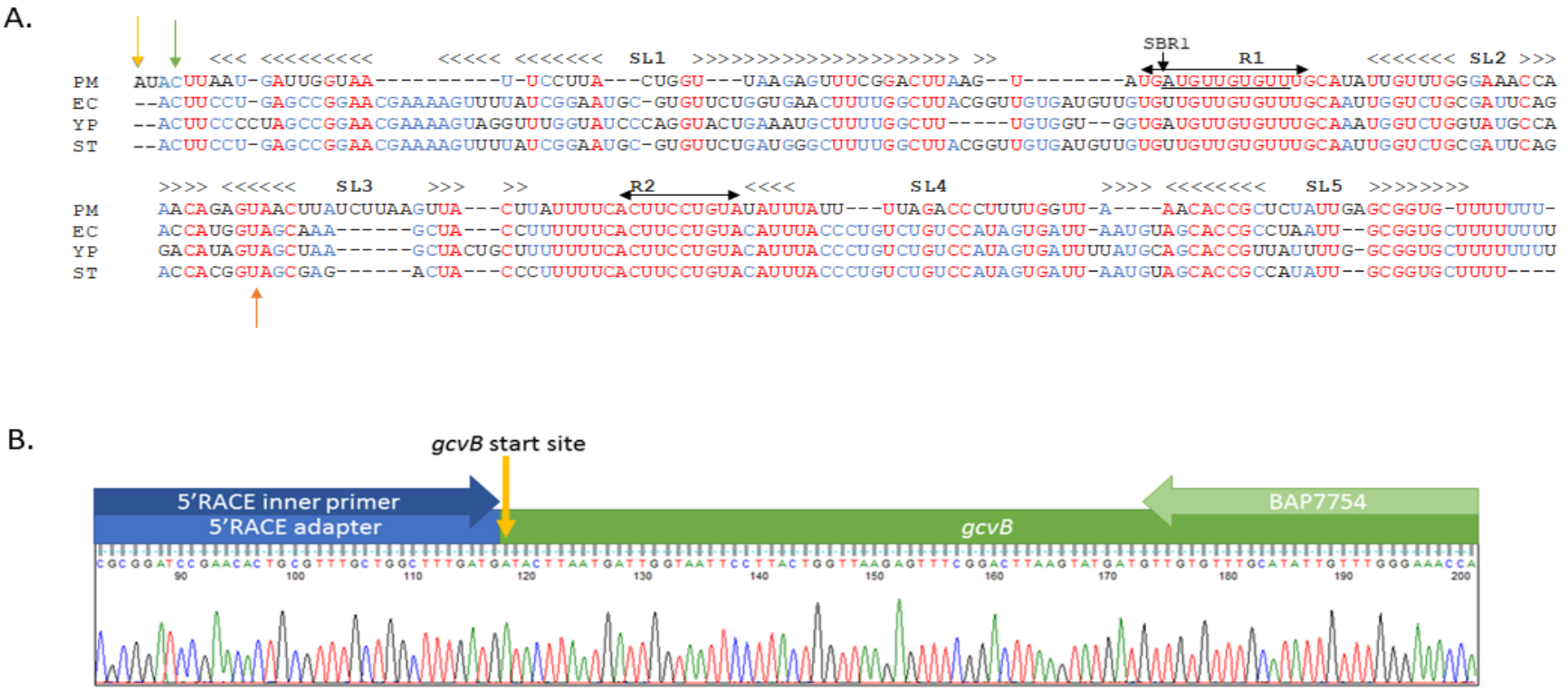

Figure 2.1. **A.** Nucleotide sequence alignment of GcvB from *P. multocida* (PM), *E. coli* (EC), *Y. pestis* (YP), and *S.* Typhimurium (ST). The previously identified *S.* Typhimurium GcvB stem-loop (SL) sequences (SL1-SL5), including their extent (< and >) and conserved R1 and R2 sequences (horizontal arrows) (Sharma et al. 2007) are shown above the alignment. Nucleotides in red are identical across all four GcvB sequences, nucleotides in blue are identical in three of the four GcvB sequences. The proposed *P. multocida* seed region is labelled SBR1 and underlined. The predicted *P. multocida* GcvB rho-independent terminator sequence is labelled SL5. The orange arrow designates the position of the TargeTron® intron insertion site in the *P. multocida* VP161 *gcvB* mutant. The green arrow indicates the predicted transcript start site for *P. multocida* GcvB as determined by primer extension. The yellow arrow indicates the predicted start site for *P. multocida* GcvB as determined by 5' RACE. **B.** DNA sequence generated from the GcvB 5' RACE nested PCR products. The position of the 5' RACE adapter (light blue) and the annealing position of the 5' RACE inner primer (dark blue) are adjacent to the 5' start site of *gcvB* (yellow arrow) and followed by the *gcvB* sequence (dark green) and the annealing site for the *gcvB* specific inner primer BAP7754 (light green).

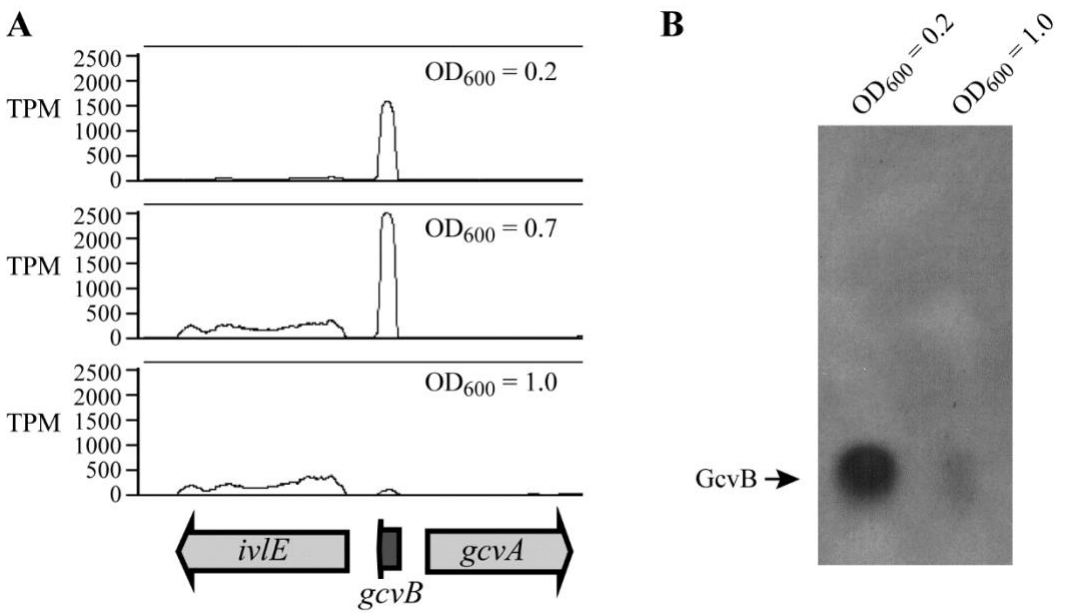

Figure 2.2. GcvB is expressed strongly at early- and mid-exponential growth phase but only weakly at late-exponential growth phase. **A.** Number of mapped transcripts per million (TPM) total transcripts that map to the genomic region surround *gcvB* as determined by whole genome RNA-seq and as visualised in Artemis (Sanger) genome viewer. The top panel shows the number of mapped reads using RNA derived from early-exponential growth phase (OD_{600} = 0.2) cells, the middle panel shows the number of mapped reads using RNA derived from mid-exponential growth phase (OD_{600} = 0.7) cells, and the bottom panel shows the number of mapped reads using RNA derived from late-exponential growth phase (OD_{600} = 0.2) cells. The extent and orientation of the genes *ivlE*, *gcvB* and *gcvA* are shown below the mapping graphs. B. *P. multocida* RNA (8 µg/ lane) isolated from early-exponential growth phase (OD_{600} = 0.2) cells and late-exponential growth phase (OD_{600} = 1.0) cells was used for Northern blotting together with a DIG-labelled, single-stranded RNA probe representing the sequence complementary to GcvB. The position of GcvB is shown at the left. Image has been modified to increase contrast.

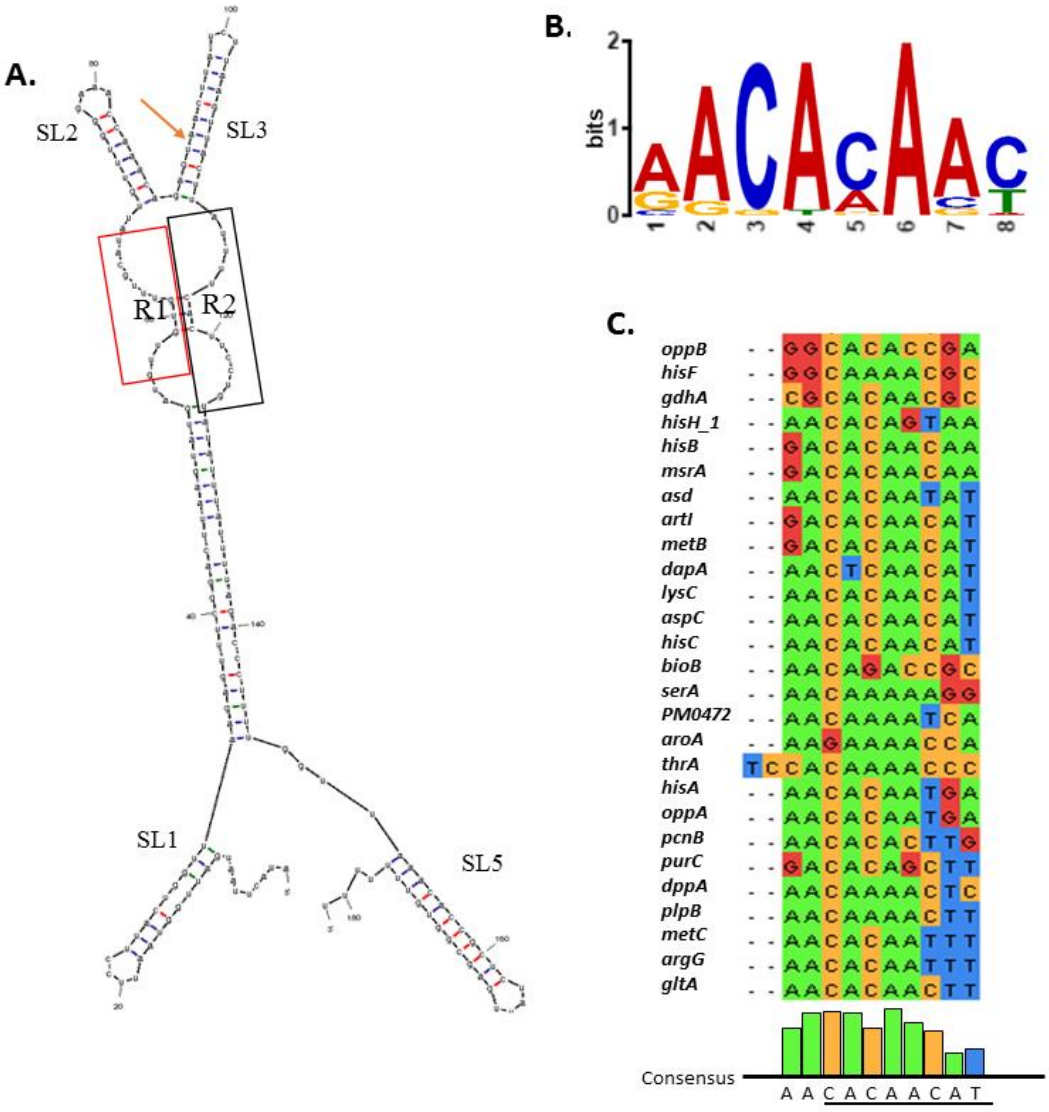

Figure 2.3. Predicted secondary structure of the *P. multocida* GcvB and the seed binding consensus motifs present in the 27 putative mRNA targets. **A.** Putative secondary structure of the *P. multocida* GcvB sRNA molecule as predicted by Mfold. The conserved R1 (red) and R2 (black) sequences are boxed and the proposed, SL1, SL2, SL3 and SL5 stem-loops are labelled. The position of the TargeTron® intron in the *P. multocida gcvB* mutant is also shown (orange arrow). **B.** Diagram of the GcvB mRNA target seed binding motif identified by MEME in 27 genes encoding putative GcvB mRNA targets. The letter height indicates the frequency of each base at each position. **C.** Sequence alignment of the 27 putative seed binding regions found by MEME motif finder in genes encoding the predicted GcvB mRNA targets. Nucleotides are highlighted with colour as follows to show the level of conservation; A, green; C, yellow; T, blue; G, red. The *P. multocida* GcvB consensus sequence is shown beneath the alignment with the *E. coli* and *S. Typhimurium* core GcvB-mRNA seed binding sequence underlined (Sharma et al. 2007).

2.3.2 GcvB predominately regulates amino acid biosynthesis and transport proteins in *P. multocida*

In order to determine the GcvB regulon in *P. multocida*, a VP161 *gcvB* mutant (AL2677; Table 2.1) was constructed using TargeTron® technology (Sigma-Aldrich). The intron insertion was located between nucleotides 92 and 93 of *gcvB* and within the predicted SL3 loop (Figure 2.1 and Figure 2.3A). To complement the mutation, the wild-type VP161 *gcvB* gene, together with its putative native promoter, was cloned into the *P. multocida* plasmid pPBA1100s to generate the plasmid pAL1190 (Table 2.1). This plasmid was used to transform the *P. multocida gcvB* mutant AL2677, producing the strain AL2864 (Table 2.1). As a control, pPBA1100s empty vector (Table 2.1) was also used to transform the *gcvB* mutant, generating the strain AL2862.

The level of *gcvB* expression in these strains was determined using qRT-PCR with all expression levels normalised to the expression of the housekeeping gene *gyrB* (Figure 2.4). The levels of *gcvB* expression in the wild-type strain (normalised to *gyrB*) were 0.24 ± 0.04 (n = 3, SEM) and 0.13 ± 0.03 at early- and mid-exponential growth phase, respectively. As expected, no expression of *gcvB* was measured in the *gcvB* mutant at either growth phase, as the primers used for the qRT-PCR spanned the point of the relatively large intron insertion. Surprisingly, the levels of *gcvB* expression in AL2864 (*gcvB* mutant provided with an intact copy of *gcvB* on the plasmid pAL1190) were 16.4 ± 2.5 and 37.6 ± 16.4, indicating a 69-fold increase in *gcvB* expression at early-exponential growth phase and a 289-fold increase at mid-exponential growth phase compared to expression in the wild-type strain. Thus, providing the *gcvB* mutant with functional *gcvB* in *trans* resulted in the significant overexpression of GcvB at both growth phases tested. Accordingly, the strain AL2864 was designated as a GcvB overexpression strain. It was predicted that that there would be an inverse relationship between the levels of expression of any GcvB-regulated genes in the *gcvB* overexpression strain and the levels of expression of the same genes in the *gcvB* deficient strains (*gcvB* mutant alone, or *gcvB* mutant containing empty vector). To test this, the overexpression strain was included in the proteomic analyses described below.

The survival and growth of the wild-type VP161, the GcvB-deficient strains (AL2677 and AL2862) and the GcvB overexpression strain (AL2864) was examined under several conditions. It was found that the *gcvB* mutant strains grew indistinguishably from the wild-type VP161 during growth in heart infusion (HI) broth (Figure 2.5A and 5B). The *gcvB* overexpression strain had a similar exponential growth rate (doubling time of 41.1 ± 0.5 min) to the wild-type strain (doubling time of 36.2 ± 2.1 min), however the lag-phase was increased by approximately 1.5 h (Figure 2.5B). There was no difference in the ability of the *gcvB* mutant, AL2677, and the wild-type VP161 to form biofilms during static growth (Figure 2.5C) and no difference in survival at low pH (HI broth, pH = 4.6 for 15 min) (Figure 2.5D).

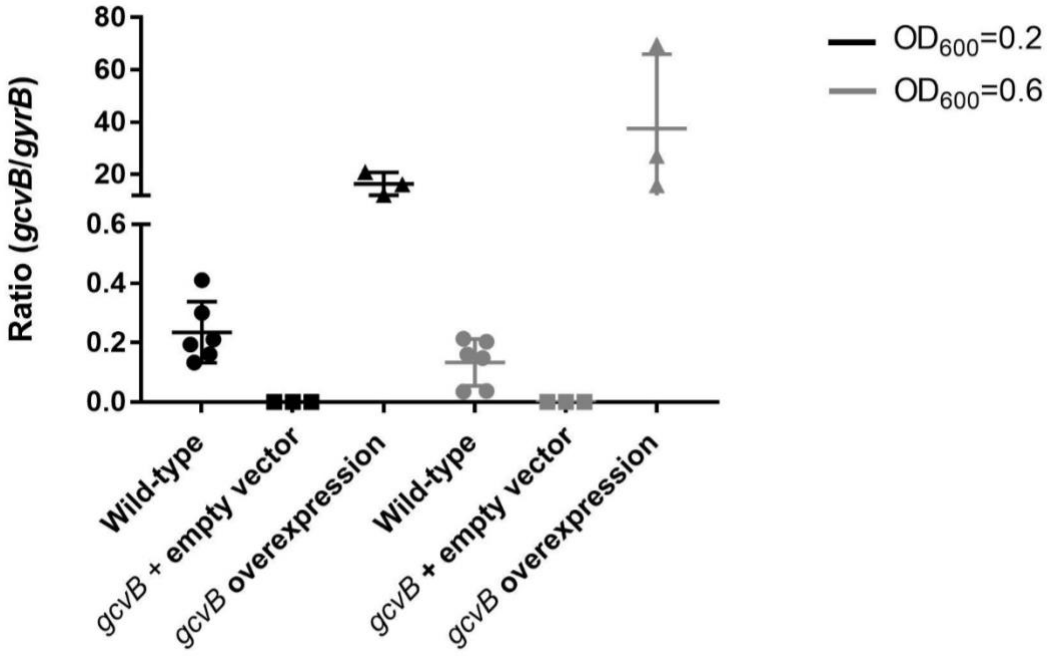

Figure 2.4. qRT-PCR analysing the expression of GcvB in wild-type *P. multocida* expression (represented by circles), the *gcvB* mutant with empty vector (square symbols) and the GcvB overexpression strains (triangle symbols). RNA was isolated from early exponential growth phase cells (OD_{600} = 0.2) (black symbols) and mid-exponential growth phase cells (OD_{600} = 0.6) (grey symbols). Expression was standardized to the expression of the house-keeping gene *gyrB*. Thick horizontal bars represent the mean ±SD.

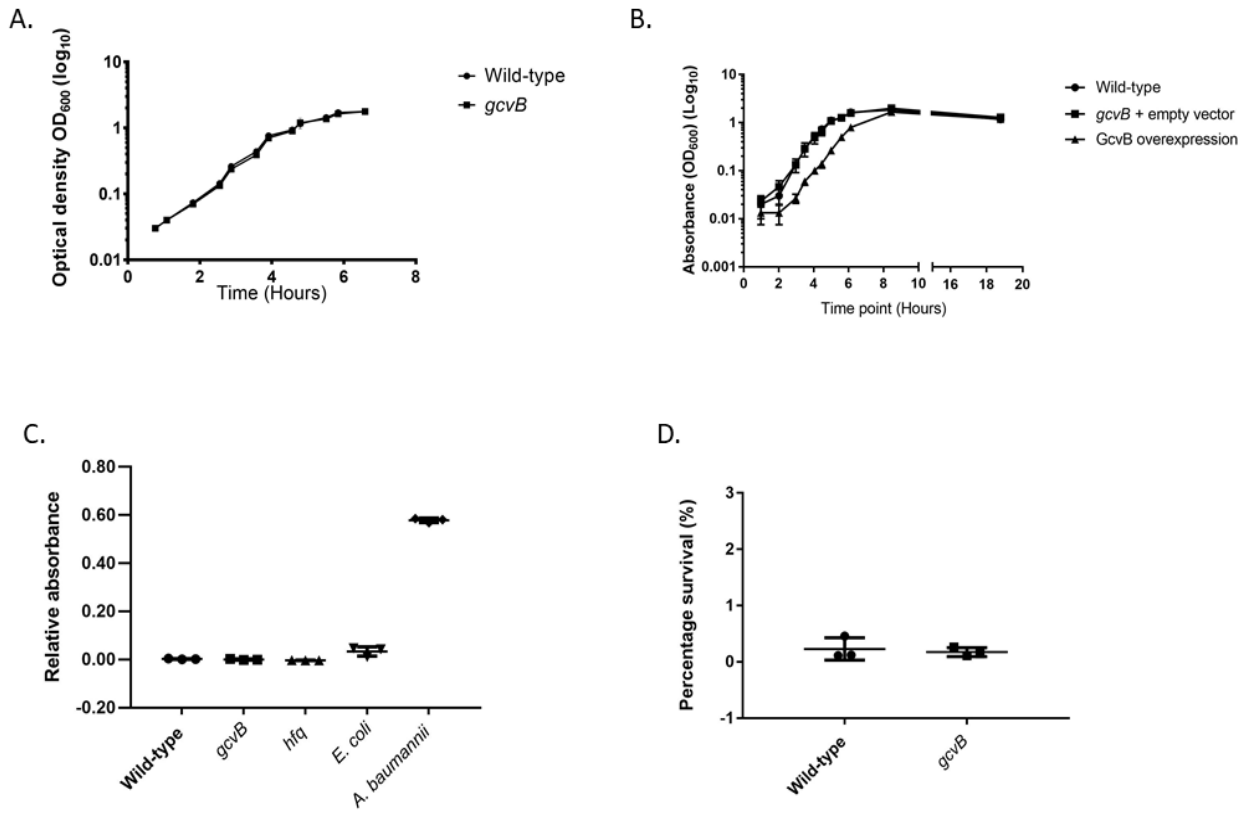

Figure 2.5. A. Growth curve of wild-type *P. multocida* VP161 (circles) and the *gcvB* mutant strain (squares) in Heart infusion broth, incubated with shaking at 37°C for 7 h. Data shown is mean ± SD (n = 3). B. Growth curve of wild-type *P. multocida* VP161 (circles), *gcvB* mutant containing empty vector (squares) and the *gcvB* overexpression strain (triangles) grown for 24 h under the same conditions as above. C. Relative absorbance compared to a no bacteria control, observed following a static crystal violet biofilm assay. The *P. multocida* wild-type VP161 (circles) was compared to the *gcvB* mutant strain (squares) and the *hfq* mutant strain (upright triangles). Controls included *E. coli* (upside-down triangles), an intermediate biofilm-forming species, and *A. baumannii* (diamonds), a strong biofilm-forming species. Horizontal lines represent mean ± SD (n = 3). D. Survival of *P. multocida* wild-type VP161 (circles) and the *gcvB* mutant strains (squares) following 15 minutes of acid stress at pH 4.6. Horizontal lines represent mean ±SD (n = 3).

We then analysed the protein expression profiles of the wild-type VP161, *gcvB* mutant (AL2677), *gcvB* mutant plus empty vector (AL2862) and GcvB overexpression strain (AL2864) using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) in biological triplicate. The initial experiment compared wild-type VP161 and the *gcvB* mutant using isotopically labelled samples. The second experiment compared wild-type, *gcvB* mutant plus empty vector and GcvB overexpression strain using label-free proteomics. For all experiments, cells were harvested at early-exponential growth phase, when *gcvB* is strongly expressed in the wild-type strain. In total, 1191 proteins were identified in the first experiment (using isotopic labelling) and 1540 proteins in the second (label-free); representing 57% and 74% respectively of the 2085 total proteins predicted to be encoded on the *P. multocida* VP161 genome (Boyce et al. 2012). Identified proteins were considered differentially expressed if they showed a ≥ 1.5-fold (≥ 0.59 $\log_2$) difference in production at an FDR of < 0.05 compared to wild-type VP161. Overall 36 proteins were measured as showing increased production in either of the two *gcvB* mutant strains analysed; 25 proteins in experiment 1, 28 in experiment 2 with 17 identified in both experiments (Appendix 2). Only 10 proteins showed decreased production in either of the two *gcvB* mutant strains analysed; two in experiment 1, eight in experiment 2 with none identified in both experiments (Appendix 3). In contrast, 218 proteins with altered production levels were identified in the GcvB overexpression strain, 75 with increased production and 143 with decreased production (Appendix 4 and 5).

We then compared the lists of differentially produced proteins identified in each of the GcvB-deficient strains (*gcvB* mutant strain AL2677, and *gcvB* mutant plus empty vector, AL2862) and the GcvB overexpression strain. A total of 27 proteins showed significantly increased production in one of the GcvB-deficient strains as well as inverse (decreased) production in the *gcvB* overexpression strain (Table 3). Of these 27 proteins, 17 proteins (71%) displayed increased production in both GcvB-deficient strains analysed (Table 3). A total of 10 proteins showed decreased production in one of the GcvB-deficient strains (AL2677 or AL2862) but none of these showed decreased production in both GcvB-deficient strains and only one (Tpl) showed inverse (increased) production in the *gcvB* overexpression strain.

Table 3. Proteins with increased production in either the *P. multocida gcvB* mutant strain (AL2677) or mutant with empty vector (AL2682) and decreased production in the *gcvB* overexpression strain (AL2684) as compared to the VP161 wild-type parent. Protein production ratio in each strain (relative to protein production in VP161) is shown as a $\log_2$ value with the corresponding false discovery rate (FDR) shown in brackets.

Protein name [a]	**VP161 locus tag, (Pm70 locus tag)**	**AL2677 ($\log_2$), (FDR)**	**AL2682 ($\log_2$), (FDR)**	**AL2684 ($\log_2$), (FDR)**	**Predicted protein function**	**Primary biochemical pathway/s**
ThrA	PMVP_0066, (PM0113)	0.59, (0.007)	0.47, (0.042)	-1.18, (0.0001)	Bifunctional aspartokinase I/homeserine dehydrogenase I	Isoleucine
ArtI	PMVP_0077, (PM124)	0.99, (0.002)	0.57, (0.059)	-1.63, (0.0001)	Arginine ABC transporter	Transporter – arginine
DppA	PMVP_0194, (PM0236)	1.89, (0.001)	1.87, (0.005)	-1.78, (0.0002)	Periplasmic dipeptide transport protein	Transporter- dipeptides
GltA	PMVP_0236, (PM0276)	1.25, (0.002)	1.02, (0.005)	-1.58, (0.0001)	Citrate synthase	TCA cycle, Glutamate
BioB	PMVP_0348, (PM0379)	0.85, (0.005)	0.64, (0.030)	-1.66, (0.0001)	Biotin synthase	Valine
PM0472	PMVP_0448, (PM0472)	1.08, (0.002)	0.74, (0.007)	-1.56, (0.00003)	PBP2_TAXI_TRAP_like_3 domain-containing protein	
MsrA	PMVP_0575, (PM0605)	0.65, (0.020)	0.73, (0.012)	-1.30, (0.0001)	Peptide methionine sulfoxide reductase	Methionine
AspC	PMVP_0593, (PM0621)	0.78, (0.010)	0.88, (0.005)	-1.38, (0.0001)	Aromatic amino acid aminotransferase	Tyrosine
MetC_2	PMVP_0791, (PM0794)	ND[b], (ND)	1.04, (0.013)	-2.06, (0.0001)	Cystathionine beta-lyase	Methionine
ArgG	PMVP_0809, (PM0813)	0.70, (0.006)	0.51, (0.082)	-1.63, (0.0001)	Arginosuccinate synthase	Arginine
PurC (HemH)	PMVP_0811, (PM0815)	0.82, (0.004)	1.03, (0.006)	-1.52, (0.0001)	Phosphoribosylaminoimidazole-succinocarboxamide synthase	De novo purine nucleotide synthesis
HisH_1	PMVP_0837, (PM0838)	1.05, (0.002)	1.17, (0.007)	-1.63, (0.0002)	Histidinol-phosphate aminotransferase	Histidine
AroA	PMVP_0838, (PM0839)	0.61, (0.008)	0.37, (0.200)	-1.03, (0.0007)	3-phosphoshikimate 1-carboxyvinyltransferase	Tyrosine
PcnB	PMVP_0865, (PM0864)	0.27, (0.059)	0.68, (0.035)	-0.78, (0.0033)	Poly (A) polymerase	
LysC	PMVP_0948, (PM0937)	1.07, (0.004)	1.27, (0.005)	-1.66, (0.0001)	Aspartate kinase	Lysine, threonine, methionine, homoserine, isoleucine

Table 3 continued.

Protein name [a]	VP161 locus tag, (Pm70 locus tag)	AL2677 (log_2), (FDR)	AL2682 (log_2), (FDR)	AL2684 (log_2), (FDR)	Predicted protein function	Primary biochemical pathway/s
MetB	PMVP_1008, (PM0995)	0.79, (0.047)	0.99, (0.035)	-2.11, (0.0001)	Cystathionine gamma-synthase	Methionine, lysine, threonine, homoserine
DapA	PMVP_1069, (PM1051)	0.83, (0.003)	0.86, (0.058)	-1.67, (0.0005)	Dihydrodipicolinate synthase	Lysine, threonine, methionine
HisC	PMVP_1220, (PM1199)	0.64, (0.030)	0.77, (0.035)	-1.11, (0.001)	Histidinol-phosphate aminotransferase	Histidine
HisB	PMVP_1221, (PM1200)	1.22, (0.002)	1.21, (0.005)	-1.33, (0.0002)	Histidinol-phosphatase	Histidine
HisA	PMVP_1224, (PM1203)	0.73, (0.008)	0.48, (0.275)	-2.02, (0.0002)	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase	Histidine
HisF	PMVP_1225, (PM1204)	0.71, (0.003)	0.72, (0.035)	-1.51, (0.0001)	Imidazoleglycerol phosphate synthase, cyclase subunit	Histidine
Asd	PMVP_1687, (PM1623)	0.89, (0.008)	0.63, (0.035)	-1.08, (0.0004)	Aspartate-semialdehyde dehydrogenase	Lysine, threonine, methionine, homoserine, isoleucine
SerA	PMVP_1723, (PM1671)	0.73, (0.004)	0.82, (0.035)	-1.20, (0.0009)	D-3-phosphoglycerate dehydrogenase	Serine, cyctine and Glycine
PlpB	PMVP_1787, (PM1730)	1.39, (0.002)	1.22, (0.009)	-1.28, (0.001)	Outer membrane lipoprotein	Transporter -Methionine
OppB	PMVP_1961, (PM1909)	0.63, (0.004)	0.20, (0.591)	-1.67, (0.0002)	Oligopeptide transport system permease protein OppB	Transporter-oligopeptides
OppA	PMVP_1962, (PM1910)	1.31, (0.002)	1.09, (0.005)	-1.55, (0.0001)	Periplasmic oligopeptides binding protein	Transporter-oligopeptides
GdhA	PMVP_2095, (PM0043)	2.27, (0.001)	2.48, (0.104)	-3.46, (0.0079)	Glutamate dehydrogenase	Glutamate synthesis, TCA cycle, Nitrate reduction

[a] Differentially expressed proteins were defined as those showing at least 1.5-fold change in production ($log_2 \geq 0.59$) with a FDR of less than 0.05.

[b] ND, no data available

The binding of GcvB with many mRNA targets in *E. coli* and *S.* Typhimurium is Hfq-dependent. To confirm *P. multocida* GcvB bound to Hfq we expressed a FLAG-tagged Hfq in *P. multocida* and used co-immunoprecipitation followed by high-throughput sequencing to identify precipitated RNAs (in triplicate samples). Sequences matching GcvB were recovered from the FLAG-tagged Hfq samples at high numbers, on average 774.7 reads per sample, but at significantly reduced numbers in the untagged control sample, an average of 39.3 reads per sample (FDR < 0.05). Therefore, we conclude that *P. multocida* GcvB can bind *P. multocida* Hfq. Given this information, the list of proteins identified as differentially produced in the *P. multocida gcvB* mutant analyses was compared to the list of proteins identified as differentially produced in the previously analysed *P. multocida hfq* mutant (Mégroz et al. 2016). Ten of the proteins that showed increased production in the *P. multocida hfq* mutant also showed increased production in both of the GcvB-deficient strains and a further five showed increased production in one of the GcvB-deficient strains (AL2677 or AL2682). These proteins were Asd, DapA, DppA, GdhA, GltA, HisC, HisH_1, IlvG, LysC, MetB, OppA, PlpB, PurC, RcpA and RsgA_2 (Appendix 2). One protein, SpeF, showed decreased production in the *hfq* mutant at mid-log growth phase (Mégroz et al. 2016) but increased production in the *gcvB* mutant, AL2677 (Appendix 2). The list of proteins with altered production in the *P. multocida* GcvB-deficient strains was also compared with the 54 GcvB-regulated targets identified in *E. coli* and *S.* Typhimurium (Sharma et al. 2011). Of these 54 known targets, 42 had homologs in the *P. multocida* genome and all but five of these were measured in our proteomics experiments However, only eight were identified as differentially produced in our proteomics experiments; namely, DppA, GdhA, LysC, OppA, OppB, PlpB, SerA and ThrA (Appendix 2).

The proteins identified as differentially produced (increased or decreased production) in *P. multocida* following inactivation of *gcvB* in either experiment (Appendix 2 and 3), were mapped to their metabolic pathways; amino acid biosynthesis proteins were observed to be highly over-represented (Figure 2.6; Fishers exact test; $p < 10^{-11}$). These amino acid biosynthesis proteins included 21 with increased production (ArgG, AroA, Asd, AspC, BioB, DapA, GdhA, GltA, HisH_1, HisA, HisB, HisC, HisF, HisG, IlvG, LysC, MetB, MetC_2, SerA, SpeF and ThrA) and three with decreased production (ArgC, LeuC and Tpl). Furthermore, another five proteins with increased production were predicted to be involved in the transport of amino acids or oligopeptides (ArtI, DppA, OppA, OppB and PlpB), as well as one protein with decreased production (ArtP). Thus, 27 of the 36 proteins negatively regulated by GcvB are involved in biosynthesis or transport of at least 13 different amino acids (Figure 2.6). Moreover, of the 17 proteins that displayed increased production in both GcvB-deficient strains and an equivalent decrease in the *gcvB* overexpression strain, only PurC (predicted to be involved in *de novo* purine biosynthesis), and PM0472 (an uncharacterized periplasmic binding protein containing a PBP2_TAXI_TRAP_like_3 domain), were not predicted to be involved in amino acid transport and metabolism.

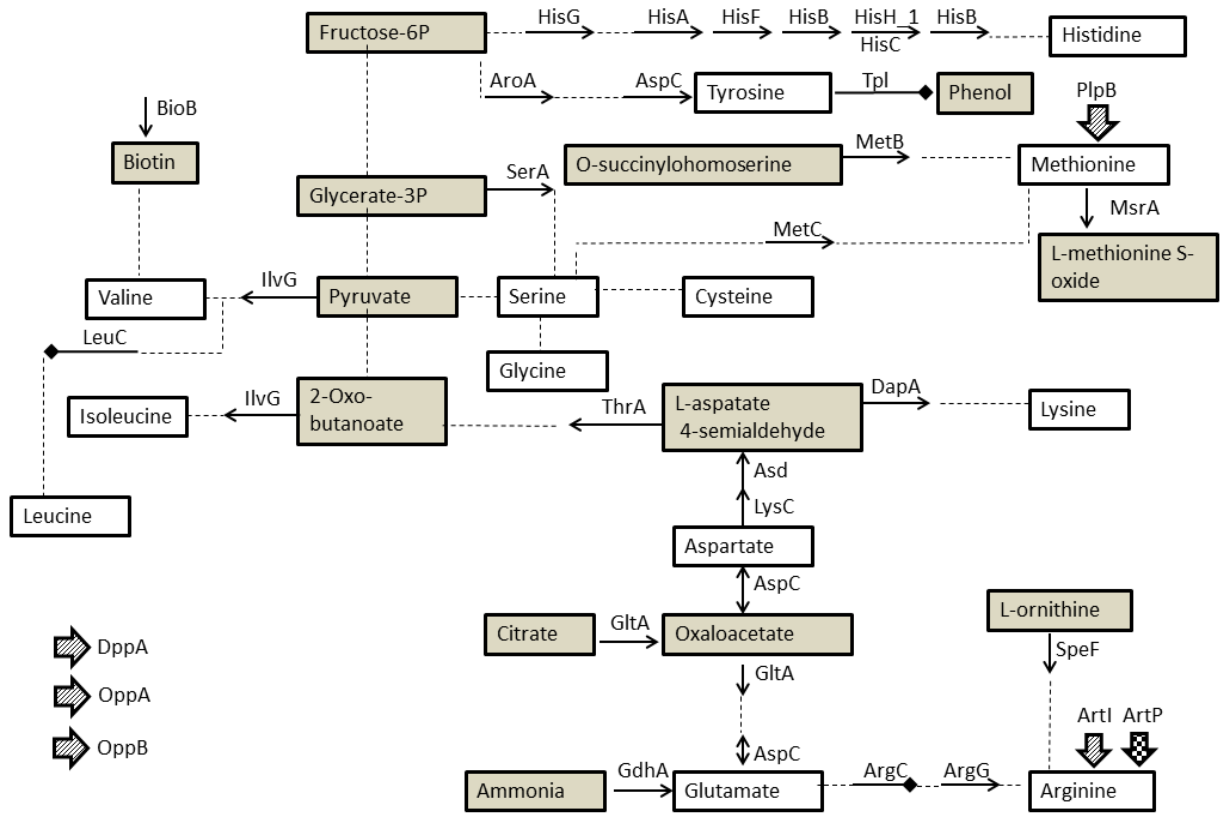

Figure 2.6. Amino acid biosynthesis pathways predicted to be affected by *gcvB* inactivation in *P. multocida* strain VP161. The amino acids whose biosynthesis is predicted to be regulated by GcvB are within white boxes. Amino acid biosynthesis proteins whose production is negatively regulated by GcvB are shown at the relevant pathway step with open-headed arrows. Proteins whose production is positively regulated by *gcvB* are indicated at the relevant pathway step with closed diamond-headed arrows. Large, diagonally striped, and checked arrows indicate predicted amino acid or oligo peptide transport proteins that are negatively and positively regulated by *gcvB*, respectively

2.3.3 Bioinformatic analyses identifies an extended GcvB seed region binding motif

In order to determine if each of the experimentally identified putative *P. multocida* GcvB targets contained a conserved region that may serve as a GcvB binding site, the DNA sequence starting 120 nt upstream of the start codon and continuing to 60 nt downstream of the start codon of each gene was examined for conserved sequence motifs using the Multiple Em for Motif Elicitation (MEME) tool (Bailey et al. 2009). Initially, all of the genes encoding the proteins identified as differentially produced following inactivation of *gcvB* (i.e. all proteins in Appendix 2 and Appendix 3) were examined. However, this analysis failed to identify a conserved motif across all proteins. We then constrained the target list to include only those proteins that showed increased differential production in either of the *gcvB* mutant strains and had a corresponding inverse production in the *gcvB* overexpression strain (Table 3). Using the DNA sequences (-120 to +60 nt) of these genes, a consensus sequence consisting of 5'-AACACAAC-3' (*E*-value: 3.2e-7) was identified in all targets (Figure 2.3B). The sequences around this identified motif were also aligned using Clustal Omega (Sievers et al. 2011), which revealed a highly conserved slightly extended consensus sequence (5'-AACACAACAT-3') (Figure 2.3C). Therefore, we predict that the *P. multocida* GcvB seed binding sequence is slightly longer than the *E. coli* and *S.* Typhimurium GcvB seed region, but that the eight central nucleotides (5'-CACAACAT-3') are identical. Importantly, the reverse complement of the extended *P. multocida* GcvB seed sequence, 5'-AUGUUGUGUU-3', is present within the sequence of GcvB; this sequence was identical to the same region in the *Y. pestis* GcvB and differed by just a single nucleotide compared to the same region in the *E. coli* and *S.* Typhimurium GcvB (Figure 2.1). The position of this consensus GcvB binding sequence was then mapped on each of the 27 mRNA targets (Figure 2.7). The binding sequence was located upstream of the predicted ribosome binding footprint [-39 to + 19 bp; (Huttenhofer and Noller 1994; Sharma et al. 2007)] in seven mRNA targets and was overlapping, or within the ribosome binding footprint, in 19 targets. In one target, *gdhA*, the binding sequence was downstream of the ribosome binding footprint (Figure 2.7). As four of the identified GcvB targets were known GcvB targets in *Salmonella* and *E. coli,* the putative seed binding regions of the *P. multocida* targets were compared to the known GcvB seed binding regions in *oppA, dppA, serA* and *gdhA* encoded by *Salmonella* (Sharma et al. 2007; Sharma et al. 2011). It was found that the seed region for the *P. multocida oppA* was located at the same position relative to the seed region of *oppA* in *Salmonella* and contained a similar sequence (Sharma et al. 2007). The seed region for the *P. multocida serA* was located close to the seed region position reported for *serA* in *Salmonella* but the sequence was dissimilar (Sharma et al. 2011). In contrast, the predicted seed regions for *P. multocida dppA* and *gdhA* were found in different locations to

those reported for the equivalent genes in *Salmonella* and had only limited sequence similarity (Sharma et al. 2007; Sharma et al. 2011).

In order to determine if the identified GcvB seed region was present in all mRNAs encoding proteins predicted to be regulated by GcvB, the corresponding DNA sequences (-120 to +60 nt) were visually inspected. Two more seed sequences were identified that exactly matched the consensus sequence generated with MEME and these were located in *glpQ* and *leuC*, positioned at 34 nt and 0 nt upstream of the start codon, respectively. The remaining putative targets had no sites with less than one or two mismatched nucleotides at critical positions (3 and 6).

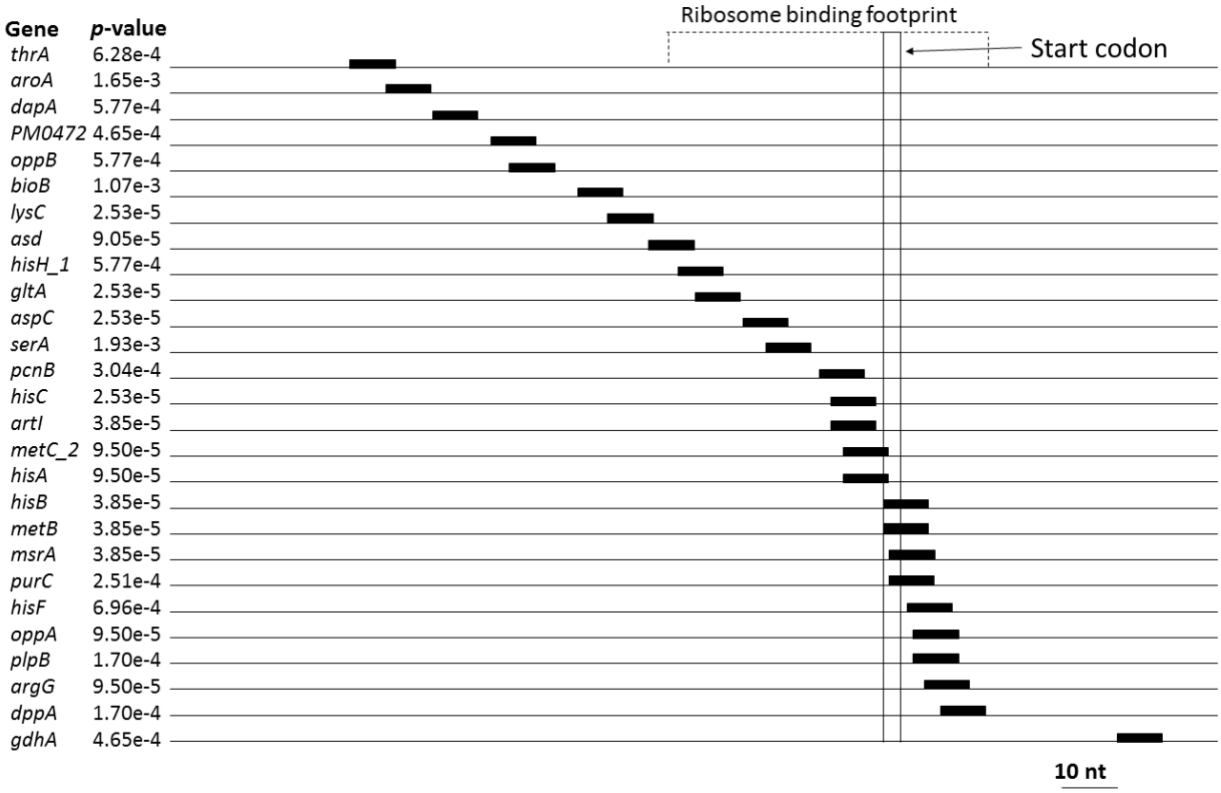

Figure 2.7. Schematic representation of the position of the predicted GcvB seed binding regions (black boxes) in 27 GcvB mRNA targets (listed), relative to the start codon. The *p*-values generated by MEME motif finder are shown at the left and give the likelihood of each identified motif occurring in the analysed sequence fragment by chance. The predicted ribosome footprint is indicated by the dashed line (top). A scale bar is shown at the bottom.

2.3.4 Modification of a two-plasmid GFP reporter system to detect *P. multocida* sRNA-mRNA interaction in *E. coli*

In order to experimentally confirm that the conserved sequence 5'-AACACAACAT-3' contained the *P. multocida* GcvB seed region, sRNA/mRNA interaction experiments using two recombinant plasmids were conducted in *E. coli* strain DH5α, based on a previously described two-plasmid GFP reporter system (Urban and Vogel 2007). *P. multocida* Hfq shares 92.7% identity with two thirds of the *E. coli* Hfq protein (amino acids 1-73) but shares only 13.7% identity with the C-terminal region of *E. coli* Hfq (amino acids 74-102). Therefore, before using this system, we first assessed whether *E. coli* Hfq could act as a chaperone for *P. multocida* sRNA molecules. A *P. multocida* expression plasmid containing a functional copy of the *E. coli* DH5α *hfq* (pAL1266, Table 2.1) was used to transform the *P. multocida* VP161 *hfq* mutant, which produces only low levels of hyaluronic acid capsule compared to the parent strain VP161 (Mégroz et al. 2016). When the *P. multocida hfq* mutant was complemented with pAL1266 (expressing *E. coli hfq*), capsule production was restored to the same level as that observed when the *hfq* mutant was complemented with the native *P. multocida hfq* gene (Figure 2.8). Thus, these data show that the native *E. coli* Hfq molecule can appropriately chaperone *P. multocida* sRNAs, allowing *E. coli* to be used as the host cell for the *P. multocida* sRNA-mRNA interaction studies described below.

To produce a two-plasmid GFP reporter system for our experiments, two expression vectors were constructed, designated pREXY and pTEXY (Table 2.1). The pREXY plasmid is a shuttle vector used for the expression of *P. multocida* sRNAs (GcvB in this case) in either *E. coli* or *P. multocida* and contains a *P. multocida tpi* promoter upstream of the multiple cloning site (MCS). The second plasmid, pTEXY, is used for the transcriptional and translational coupling of the mRNA target with superfolder GFP (sfGFP) under the control of the tetracycline promoter ($P_{Ltet0-1}$).

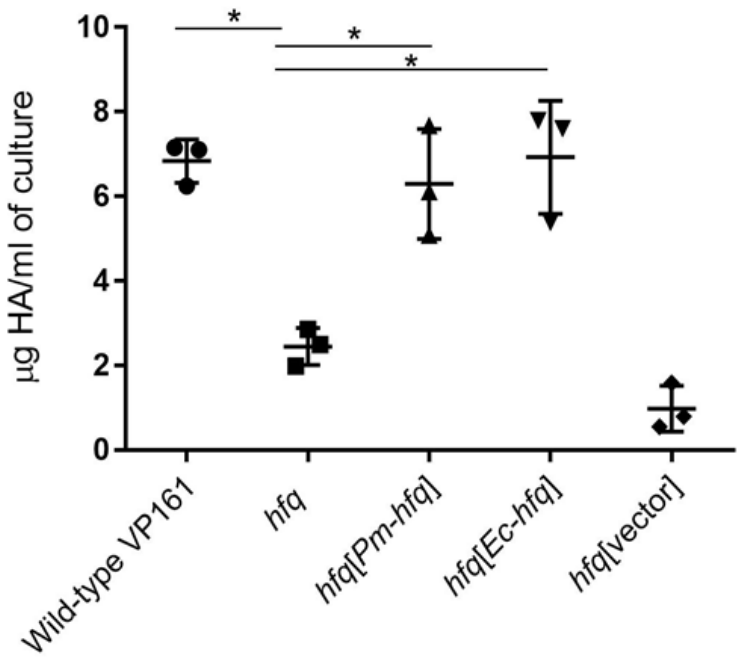

Figure 2.8. Hyaluronic acid capsule production in the *P. multocida hfq* mutant containing a functional copy of *hfq* from *E. coli* or *P. multocida*. The amount of hyaluronic acid capsular material produced during mid-exponential growth by *P. multocida* wild-type strain VP161, *P. multocida hfq* mutant (*hfq*), *P. multocida hfq* mutant complemented with a functional copy of the native *hfq* from *P. multocida* (*hfq*[*Pm-hfq*]), *P. multocida hfq* mutant containing a functional copy of the *hfq* from *E. coli* (*hfq*[*Ec-hfq*]) or the *P. multocida hfq* mutant containing empty vector (*hfq*[vector]). Each data point shows a single hyaluronic acid measurement. Thick horizontal bars represent the mean and error bars show ± 1 SD (n = 3). * = p <0.05 using Student's T-test.

2.3.5 GcvB inhibits GltA production via complementary binding between the predicted seed regions in GcvB and *gltA.*

For recombinant expression of *P. multocida gcvB* sRNA in *E. coli*, the entire *gcvB* gene from *P. multocida* strain VP161 was PCR-amplified and cloned into the MCS of pREXY, generating the GcvB expression plasmid pAL1197. For recombinant expression of a predicted *gcvB* target region, a *P. multocida* fragment containing 38 bp upstream and the first 60 bp of *gltA* was cloned into the XbaI and BglII sites of pTEXY, located between $P_{Ltet0-1}$ and sf*GFP* to produce a *gltA-sfGFP* translational fusion. This plasmid was named pAL1257 (Table 2.1). The recombinant plasmids, or vector only, were used in various combinations to transform competent *E. coli* DH5α. Restriction digest analysis and DNA sequencing confirmed all transformants contained the correct plasmids.

The *E. coli* strain containing both the pTEXY::*gltA-sfGFP* expression plasmid and the empty pREXY vector (no GcvB) was highly fluorescent, but the strain containing both the pTEXY::*gltA-sfGFP* plasmid and the pREXY::*gcvB* expression plasmid showed significantly reduced fluorescence ($p < 0.0005$; Figure 2.9). Thus, expression of GcvB represses production of the GltA-sfGFP fusion protein, as would be expected for a *bone fide* GcvB target mRNA.

In order to confirm that the GcvB-mediated repression of GltA expression was specifically due to complementary base pairing between the predicted seed regions, two modified plasmids were constructed and tested for fluorescence in the two-plasmid GFP reporter system. Firstly, the putative central seed sequence in the *gltA* upstream region was replaced with a nucleotide sequence identical to the central seed region of the GcvB sRNA (UGUGUUG) to generate the plasmid pTEXY::$gltA_{MSR1}$-*sfGFP* (pAL1290; Table 2.1). The *E. coli* strain containing this plasmid, with the *gltA* seed region mutation, and the pREXY::*gcvB* plasmid showed levels of fluorescence indistinguishable from the fluorescence of the strains containing pTEXY::*gltA-sfGFP* and empty pREXY (no GcvB). This indicates that GcvB was unable to repress the production of GltA following the mutation of the *gltA* seed region. Secondly, the plasmid pREXY::$gcvB_{MSR2}$ (pAL1277; Table 2.1) was generated, encoding a modified *gcvB* that contained a nucleotide sequence identical to the seed region of *gltA* mRNA target (ACACAAC), instead of the GcvB seed region (UGUGUUG). The *E. coli* strain containing both this plasmid and the pTEXY::*gltA-sfGFP* showed levels of fluorescence indistinguishable from the fluorescence of the strains containing pTEXY::*gltA-sfGFP* and empty pREXY (no GcvB). Thus, GcvB-mediated repression of *gltA* expression was also abrogated by mutation of the *gcvB* sRNA seed region. Finally, we tested the fluorescence of the *E. coli* strain containing both of the mutated plasmids, pTEXY::$gltA_{MSR1}$-*sfGFP* and pREXY::$gcvB_{MSR2}$, containing swapped seed regions but which are still complementary to each other. The strain containing these plasmids showed

significantly reduced fluorescence compared to each of the strains containing the following plasmid pairs; pTEXY::*gltA-sfGFP* and empty pREXY, pTEXY::$gltA_{MSR1}$-*sfGFP* and pREXY::*gcvB* and pTEXY::*gltA-sfGFP* and pREXY::$gcvB_{MSR2}$. Therefore, when the seed regions of both the GcvB sRNA and the mRNA target are mutated but in a complementary fashion, GcvB-mediated repression is restored, confirming that there is a direct interaction between the two predicted seed binding regions and that this level of binding is sufficient for the repression of GcvB expression.

In order to confirm this direct interaction between GcvB and *gltA,* EMSA experiments were performed. DNA fragments representing GcvB, *gltA,* and *gatA* (negative control) were amplified from *P. multocida* VP161 genomic DNA using the appropriate primer pairs (see Table 2.1), then transcribed using *in vitro* transcription reactions. To detect interaction between molecules, biotin was incorporated into the 5' end of the GcvB transcript via the addition of 5'biotin labelled GMP nucleotide to each *in vitro* transcription reaction. The GcvB transcript was separately mixed with the *gltA* transcript or the *gatA* transcript (control) at three different ratios (1:1, 1:2 and 1:4) and the mixtures subjected to gel electrophoresis. Samples were then transferred to a nylon membrane to detect biotin labelled bands (Figure 2.10). The results showed that, in addition to a band representing unbound GcvB, there was a much larger band present in samples containing the *gltA* transcript and the intensity of this band increased with increasing amounts of *gltA*. In contrast, only the band representing unbound GcvB was observed in lanes containing GcvB mixed with the *gatA* transcript. Thus, these experiments show that GcvB specifically binds to the *gltA* transcript.

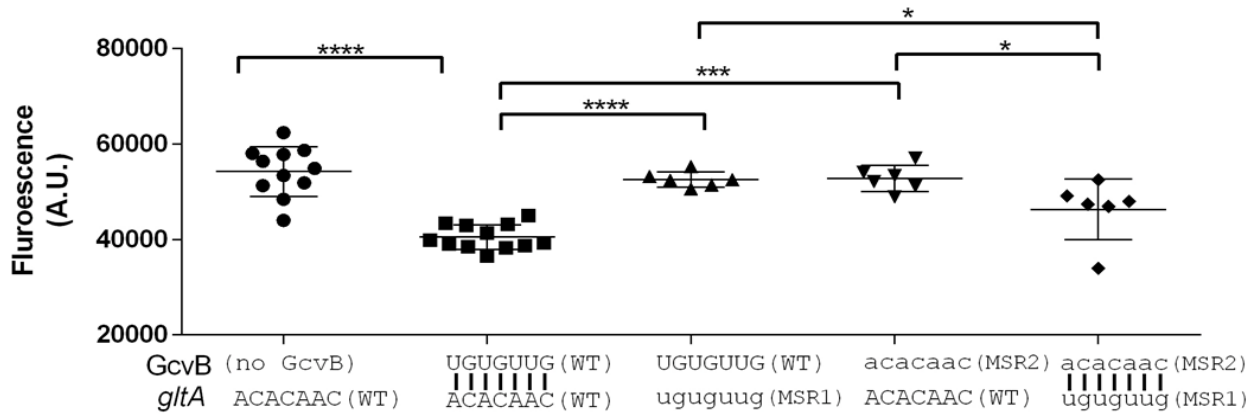

Figure 2.9. Super-folder green fluorescent protein (sfGFP) production in *E. coli* strains containing different plasmid pairs. Each *E. coli* strain harboured one pREXY sRNA expression plasmid derivative and one pTEXY mRNA::*sf*GFP reporter derivative. The top line of the x-axis label shows the sequence of the native (WT) or mutated (MSR2) seed region within the recombinant *gcvB* in the plasmid pREXY::*gcvB* or pREXY::$gcvB_{MSR2}$, respectively. A pREXY vector only control (no GcvB) was also included in the study. The bottom line of the x-axis shows the sequence of the native (WT) or mutated (MSR1) seed region within the recombinant *gltA* fused to the sfGFP gene in the pTEXY::*gltA-sfGFP* or pTEXY::$gltA_{MSR1}$-*sfGFP*, respectively. Wild-type seed sequence is shown in all capitals, mutated seed sequence is shown in lower case. Vertical lines between the text show if complementary base pairing is predicted between the sRNA and mRNA seed sequence. The amount of sfGFP-mediated fluorescence for each recombinant *E. coli* strain was measured (475/540nm ex/em) and each data point shows the amount of fluorescence emitted by a single strain. The long horizontal bars show the mean of the replicate data and error bars show ±1 SD (n = 6 or 12). * = p-value <0.05, *** = p-value <0.005, **** = p-value <0.0005.

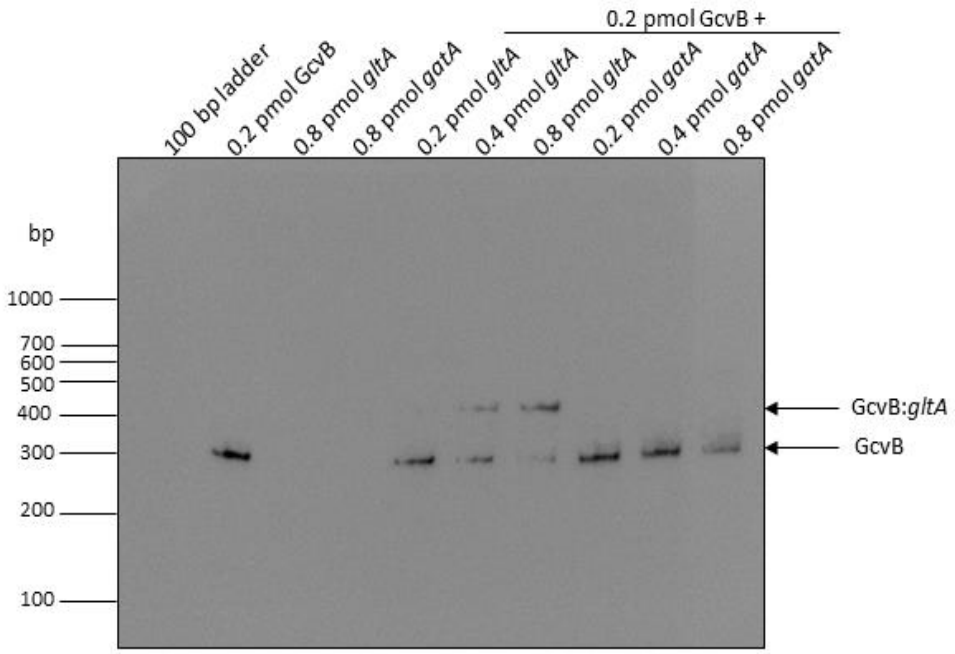

Figure 2.10. Duplex formation of biotin labelled GcvB when incubated at 70°C for 5 min then 30°C for 1 h with *gltA* or *gatA* at a 1:1, 1:2 or 1:4 ratio. Samples were electrophoresed on a 4% acrylamide gel and transferred to a nylon membrane where biotin was detected using a HRP-conjugated streptavidin antibody.

2.4 Discussion

In this study, we have shown using deep sequencing transcriptomic analyses and Northern blotting that the *P. multocida* GcvB is strongly expressed at early- and mid-exponential growth phase but displays highly reduced expression during late-exponential growth. Although qRT-PCR showed higher expression of GcvB at early-exponential growth phase whereas RNA-seq showed higher levels during mid-exponential growth phase, at both phases expression is still high when compared to the levels at late-exponential growth phase. These data correlate well with the known expression profile of GcvB in *E. coli* (Argaman et al. 2001) and *S.* Typhimurium (Sharma et al., 2007) and support the predicted function of GcvB as a repressor that acts primarily during growth under nutrient-rich conditions. We also examined the 5' start of the GcvB sRNA expressed by the *P. multocida* strain VP161 using both primer extension and 5' RACE. Primer extension identified the starting base as being positioned 1-2 bp downstream from the known start for the GcvB transcript in *E. coli* and *S.* Typhimurium. In contrast, experiments using 5' RACE identified the transcript start was located 2 bp upstream of the start in *E. coli* and *S.* Typhimurium. The 5' RACE method is considered the superior method for determining transcript starts as the 5' end of the RNA is protected from degradation by the addition of an adapter. Therefore, we conclude the *P. multocida* GcvB begins with the sequence 5'-AUACUUAAU-3'.

In order to identify the *P. multocida* GcvB regulon, we analysed the proteome of the wild-type strain, a *gcvB* mutant, a *gcvB* mutant containing empty vector and a GcvB overexpression strain. Nearly four times as many *P. multocida* proteins were identified as differentially produced in the *gcvB* overexpression strain than in the GcvB-deficient strains. Quantitative qRT-PCR showed that the level of GcvB in the overexpression strain was increased by approximately 70-fold compared to the wild-type strain at early-exponential growth phase when the proteomics was performed. Therefore, we propose that the overexpression of GcvB to this level may lead to some off-target effects via non-specific binding, as has been observed for other sRNAs (Storz et al. 2011).

The *P. multocida gcvB* mutant displayed normal growth in rich medium, was unaffected by acid stress and showed no change in phenotype (compared to the parent strain) with respect to biofilm formation. This is in contrast to what has been reported for other species; *gcvB* mutants constructed in *E. coli, S.* Typhimurium and *Y. pestis* all show a decreased growth rate in rich media, and inactivation of *gcvB* in *E. coli* results in cells with decreased biofilm formation and decreased tolerance to acid stress (McArthur et al. 2006; Sharma et al. 2007; Jin et al. 2009; Mika and Hengge 2014). It is perhaps unsurprising that the *P. multocida gcvB* mutant did not show a change in the ability to form a biofilm. Indeed, our data suggest

that wild-type VP161 forms very poor single species biofilms. Moreover, with respect to acid tolerance, *P. multocida* is considered a bite wound and respiratory/mucosal pathogen and, unlike enteric organisms, is unlikely to encounter strongly acidic conditions. However, the *P. multocida gcvB* overexpression strain did show an increased lag-phase during growth when compared to wild-type VP161. It has been previously shown that during lag phase the glycolysis pathway is predominantly used to produce energy (Rolfe et al. 2012). An important enzyme in the Krebbs cycle is citrate synthase, which in *P. multocida* is encoded by the GcvB target, *gltA.* Therefore, the increase in lag phase displayed by the *P. multocida gcvB* overexpression strain may in part be the result of decreased production of GltA due to increased GcvB binding to *gltA* transcripts. However, as this strain significantly overexpresses GcvB, it is acknowledged that the likely off-target effects on multiple other proteins may also play a role.

Of the proteins that showed either increased (36) or decreased (10) production in the GcvB-deficient *P. multocida* strains analysed, 31 (27 increased, 4 decreased) were predicted to be involved in amino acid biosynthesis and transport, and pathway analyses indicated that GcvB specifically affects the biosynthesis of at least 13 different amino acids. Therefore, our data suggest that the *P. multocida* GcvB acts primarily to repress the production and transport of amino acids during the early growth stages, likely as a means to conserve energy when nutrients are abundant. In *E. coli* and *S.* Typhimurium the role of GcvB is also to repress amino acid biosynthesis and transport when nutrients are in plentiful supply. However, in these species GcvB shows a preponderance for regulation of amino acid transporters [>60% of GcvB targets; (Sharma et al. 2011)]. In *P. multocida* this situation appears to be reversed, with the majority of the regulated proteins (~75%) being directly involved in the biosynthesis of amino acids. A comparison of the targets regulated by GcvB in *P. multocida*, *E. coli* and *S.* Typhimurium identified four that were GcvB-regulated in all three species (GdhA, OppA, SerA and DppA), two targets that were GcvB-regulated in both *P. multocida* and *S.* Typhimurium (PlpB and ThrA) and two targets that were GcvB-regulated in both *P. multocida* and *E. coli* (OppB and LysC). Thus, while the general function of GcvB as a controller of amino acid biosynthesis and transport has been conserved across the species, the precise GcvB targets show significant diversity.

The production of the histidine biosynthesis proteins HisA, HisB, HisC, HisF, HisG and HisH_1 was strongly increased (fold-change ranging from 1.6 to 2.3-fold) in *P. multocida* lacking a functional *gcvB*; five of these proteins are predicted to be encoded within a single operon. In other bacteria, histidine production is regulated by multiple mechanisms including repression of transcription initiation and attenuation (Kulis-Horn et al. 2014), but to our knowledge GcvB has not been previously linked with control of histidine

biosynthesis. HisD encodes a histidinol dehydrogenase that has also been bioinformatically predicted to be a target of GcvB in the related *Pasteurellaceae* species *A. pleuropneumoniae* (Rossi et al. 2016). Moreover, specific attenuator sequences that target histidine production have been identified in *A. pleuropneumoniae* (Rossi et al. 2016).

Of the 27 GcvB targets shown in Table 3, 71% also showed increased production in a *P. multocida* strain VP161 *hfq* mutant (Mégroz et al. 2016). This indicates that the action of the *P. multocida* GcvB on many of the putative mRNA targets is dependent on the chaperone activity of Hfq, which mediates the docking of an sRNA onto its mRNA target. The reliance of GcvB on Hfq for binding to certain mRNA targets has also been demonstrated in *E. coli* (Pulvermacher et al. 2008). Compared to protein levels in the wild-type VP161, the predicted GcvB target SpeF showed increased production in the *gcvB* mutant during early-exponential growth, when *P. multocida* GcvB has been shown to be most active. In contrast, SpeF showed decreased production in the *P. multocida hfq* mutant during mid-exponential growth, indicating that other sRNAs may act upon SpeF at later growth phases.

Previously, it was proposed that during the late stages of *P. multocida* infection the *in vivo* environment is nutrient poor (Boyce et al. 2002). Under these conditions, we would predict that the levels of *gcvB* gene expression would be low, thus allowing the expression of *gcvB* mRNA targets involved in amino acid biosynthesis and transport. Supporting this prediction, four of the genes encoding putative GcvB mRNA targets, *aspC*, *dppA, gdhA* and *gltA* had increased expression (fold changes ranging from 1.8 to 11.3) during *in vivo* growth in chickens (Boyce et al. 2002). It is possible that GcvB plays an important role in the regulation of these targets *in vivo*. However, as yet we have no direct evidence of reduced GcvB expression during growth *in vivo* as the previous microarray experiments (Boyce et al. 2002) did not include DNA spots representing any sRNAs. In our current study, the glutamate dehydrogenase, GdhA, was identified as the most highly differentially produced protein (5-fold increase) following GcvB inactivation. GdhA catalyses the conversion of L-glutamate to 2-oxoglutarate, releasing NH_3 and NADPH which then allows for the production of all amino acids within the cell (Reitzer 1996). Interestingly, a *P. multocida gdhA* mutant belonging to the capsular type B and LPS serotype/genotype 2, was attenuated for virulence and was used as an effective vaccine against haemorrhagic septicaemia in buffalo (Rafidah et al. 2012).

Comparative bioinformatic analysis using the gene sequences for 27 of the *P. multocida* GcvB mRNA targets, allowed for the identification of the predicted GcvB sRNA seed region (initial sRNA-mRNA binding site) consisting of 5'-AACACAACAT-3'. This sequence was highly conserved in a large number of the putative mRNA targets (Figure 2.3C) and the complementary sequence of this seed region was present in

the *P. multocida* GcvB sRNA. This seed binding sequence is two nucleotides longer than the characterised seed binding regions of the GcvB sRNA molecules encoded by *E. coli* and *S.* Typhimurium but importantly contains the same core region sequence, 5'-CACAACAT-3' (Urbanowski et al. 2000; McArthur et al. 2006; Pulvermacher et al. 2008; Sharma et al. 2011). The internal section of this binding region was confirmed as essential for GcvB interaction with the target *gltA* mRNA using the GFP translational reporter assay in *E. coli*, where substitution of these bases in either the mRNA target, *gltA*, or the sRNA, GcvB, decreased the interaction between the RNA molecules. Complementary substitution of bases in the sRNA and mRNA target allowed for restoration of binding efficiency and a concomitant decrease in GFP production.

The predicted seed region within each of the negatively regulated *P. multocida* GcvB target mRNAs was mapped relative to the start codon of the gene. Similar to what has been observed in other bacteria (Bobrovskyy et al. 2015), most of the GcvB-specific seed regions mapped within the ribosome binding footprint, which we predict would allow GcvB to occlude the RBS and block translation of the target mRNA. However, some were located upstream of the ribosome binding footprint, this included the seed region sequence in *thrA* which was located approximately 43 nucleotides upstream of the ribosome binding footprint. The distal position of the seed region relative to the RBS has also been noted in some GcvB mRNA targets in *E. coli* and *S.* Typhimurium. In these instances, it is thought that the CA-rich sequence within the seed region acts as a translational enhancer element and the binding of GcvB to this region blocks this enhanced translation (Sharma et al. 2007; Yang et al. 2014). Interestingly, twelve of the *P. multocida* mRNA targets (*hisB*, *metB*, *purC*, *msrA*, *metC_2, hisA, plpB, dppA, argG*, *oppA, hisF* and *gdhA*) had the seed region sequence located on or after the start codon. Binding of an sRNA molecule soon after the translational start codon on the mRNA target is predicted to affect ribosome binding and translation of a gene because the ribosome footprint can extend from the -39 to the +19 nucleotide (Huttenhofer and Noller 1994; Sharma et al. 2007). Indeed, inhibitory interactions between the sRNA RybB and the mRNA target *ompN* in *S.* Typhimurium occur at +5 to +20 nucleotides from the start codon (Bouvier et al. 2008).

The predicted seed region within the *gdhA* mRNA (position + 40) is significantly downstream of the ribosome footprint region. Previous studies in *S.* Typhimurium have suggested that the sRNA-mRNA interactions between *gdhA* and GcvB may include a second highly conserved GcvB binding site called R2 (Figure 2.1). There is very limited evidence that R2 is definitively involved in any specific sRNA-mRNA binding interactions, as the study showed that deleting the R2 region of GcvB did not abrogate binding to the *gdhA* transcript (Sharma et al. 2011; Melamed et al. 2016). The R2 region is located downstream of

the primary seed region and is present in GcvB from all bacterial species analysed (Figure 2.1), (Sharma et al. 2011; Melamed et al. 2016). Future work will assess the importance of this R2 region in *P. multocida*.

This study has characterised the GcvB regulon in *P. multocida* strain VP161 and identified the seed binding regions required for interaction between GcvB and its targets. Many of the mRNA targets identified are required for the biosynthesis and transport of amino acids. Thus, the correct temporal expression of GcvB is likely to be important for growth of this pathogen in a nutrient depleted environment, such as *in vivo* during late-stage infection. While the GcvB target-binding site is well conserved between *P. multocida* and *E. coli*, and the GcvB-regulated genes in both species are primarily involved in amino acid biosynthesis and transport, the precise genes controlled by GcvB in the two species are quite different. These data are the first functional characterisation of sRNA regulation in the *Pasteurellaceae* family; future studies will focus on identifying the role of GcvB and other sRNAs *in vivo* during *P. multocida* infections.

2.5 References

Baddal B, Muzzi A, Censini S, Calogero RA, Torricelli G, Guidotti S, Taddei AR, Covacci A, Pizza M, Rappuoli R et al. 2015. Dual RNA-seq of nontypeable *Haemophilus influenzae* and host cell transcriptomes reveals novel insights into host-pathogen cross talk. *MBio* **6**: e01765-01715.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202-208.

Bilusic I, Popitsch N, Rescheneder P, Schroeder R, Lybecker M. 2014. Revisiting the coding potential of the *E. coli* genome through Hfq co-immunoprecipitation. *RNA Biol* **11**: 641-654.

Bobrovskyy M, Vanderpool CK, Richards GR. 2015. Small RNAs regulate primary and secondary metabolism in Gram-negative bacteria. *Microbiol Spectr* **3**.

Bosch M, Garrido E, Llagostera M, Perez de Rozas AM, Badiola I, Barbe J. 2002. *Pasteurella multocida exbB, exbD* and *tonB* genes are physically linked but independently transcribed. *FEMS Microbiol Lett* **210**: 201-208.

Bouvier M, Sharma CM, Mika F, Nierhaus KH, Vogel J. 2008. Small RNA binding to 5' mRNA coding region inhibits translational initiation. *Mol Cell* **32**: 827-837.

Boyce JD, Adler B. 2006. How does *Pasteurella multocida* respond to the host environment? *Curr Opin Microbiol* **9**: 117-122.

Boyce JD, Seemann T, Adler B, Harper M. 2012. Pathogenomics of *Pasteurella multocida*. *Curr Top Microbiol Immunol* **361**: 23-38.

Boyce JD, Wilkie I, Harper M, Paustian ML, Kapur V, Adler B. 2002. Genomic scale analysis of *Pasteurella multocida* gene expression during growth within the natural chicken host. *Infect Immun* **70**: 6871-6879.

Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, Bateman A. 2013. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res* **41**: D226-232.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA et al. 2016. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* **44**: D471-480.

Chung JY, Wilkie I, Boyce JD, Townsend KM, Frost AJ, Ghoddusi M, Adler B. 2001. Role of capsule in the pathogenesis of fowl cholera caused by *Pasteurella multocida* serogroup A. *Infect Immun* **69**: 2487-2492.

Coornaert A, Chiaruttini C, Springer M, Guillier M. 2013. Post-transcriptional control of the *Escherichia coli* PhoQ-PhoP two-component system by multiple sRNAs involves a novel pairing region of GcvB. *PLoS Genet* **9**: 3.

Corcoran CP, Podkaminski D, Papenfort K, Urban JH, Hinton JC, Vogel J. 2012. Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol Microbiol* **84**: 428-445.

Desnoyers G, Bouchard MP, Masse E. 2013. New insights into small RNA-dependent translational regulation in prokaryotes. *Trends Genet* **29**: 92-98.

Fuller TE, Kennedy MJ, Lowery DE. 2000. Identification of *Pasteurella multocida* virulence genes in a septicemic mouse model using signature-tagged mutagenesis. *Microb Pathog* **29**: 25-38.

Gottesman S, Storz G. 2011. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol* **3**.

Harper M, Cox AD, St Michael F, Wilkie IW, Boyce JD, Adler B. 2004. A heptosyltransferase mutant of *Pasteurella multocida* produces a truncated lipopolysaccharide structure and is attenuated in virulence. *Infect Immun* **72**: 3436-3443.

Harper M, St Michael F, John M, Vinogradov E, Steen JA, van Dorsten L, Steen JA, Turni C, Blackall PJ, Adler B et al. 2013. *Pasteurella multocida* Heddleston serovar 3 and 4 strains share a common lipopolysaccharide biosynthesis locus but display both inter- and intrastrain lipopolysaccharide heterogeneity. *J Bacteriol* **195**: 4854-4864.

Huttenhofer A, Noller HF. 1994. Footprinting mRNA-ribosome complexes with chemical probes. *Embo J* **13**: 3892-3901.

Jin Y, Watt RM, Danchin A, Huang JD. 2009. Small noncoding RNA GcvB is a novel regulator of acid resistance in *Escherichia coli*. *BMC Genomics* **10**: 1471-2164.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**: D457-462.

Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, Lee TJ, Kaipa P, Gilham F, Spaulding A, Popescu L et al. 2010. Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief Bioinform* **11**: 40-79.

Kulis-Horn RK, Persicke M, Kalinowski J. 2014. Histidine biosynthesis, its regulation and biotechnological application in *Corynebacterium glutamicum*. *Microb Biotechnol* **7**: 5-25.

Lloyd AL, Marshall BJ, Mee BJ. 2005. Identifying cloned *Helicobacter pylori* promoters by primer extension using a FAM-labelled primer and GeneScan analysis. *J Microbiol Methods* **60**: 291-298.

McArthur SD, Pulvermacher SC, Stauffer GV. 2006. The *Yersinia pestis gcvB* gene encodes two small regulatory RNA molecules. *BMC Microbiol* **6**: 52.

Megroz M, Kleifeld O, Wright A, Powell D, Harrison P, Adler B, Harper M, Boyce JD. 2016. The RNA-binding chaperone Hfq is an important global regulator of gene expression in *Pasteurella multocida* and plays a crucial role in production of a number of virulence factors, including hyaluronic acid capsule. *Infect Immun* **84**: 1361-1370.

Melamed S, Peer A, Faigenbaum-Romm R, Gatt YE, Reiss N, Bar A, Altuvia Y, Argaman L, Margalit H. 2016. Global mapping of small RNA-target interactions in bacteria. *Mol Cell* **63**: 884-897.

Mika F, Hengge R. 2014. Small RNAs in the control of RpoS, CsgD, and biofilm architecture of *Escherichia coli*. *RNA Biol* **11**: 494-507.

Morita T, Maki K, Aiba H. 2012. Detection of sRNA-mRNA interactions by electrophoretic mobility shift assay. *Methods Mol Biol* **905**: 235-244.

Paley SM, Karp PD. 2006. The Pathway Tools cellular overview diagram and Omics Viewer. *Nucleic Acids Res* **34**: 3771-3778.

Pulvermacher SC, Stauffer LT, Stauffer GV. 2008. The role of the small regulatory RNA GcvB in GcvB/mRNA posttranscriptional regulation of *oppA* and *dppA* in *Escherichia coli*. *FEMS Microbiol Lett* **281**: 42-50.

Rafidah O, Zamri-Saad M, Shahirudin S, Nasip E. 2012. Efficacy of intranasal vaccination of field buffaloes against haemorrhagic septicaemia with a live *gdhA* derivative *Pasteurella multocida* B:2. *Vet Rec* **171**: 175.

Reitzer LJ. 1996. Ammonia assimilation and the biosynthesis of glutamine, glutamate, aspartate, asparagine, l-alanine, and d-alanine. In *Escherichia coli and Salmonella: cellular and molecular biology (F C Neidhardt, R Curtiss III, J L Ingraham, E C C Lin, K B Low, B Magasanik, W S Reznikoff, M Riley, M Schaechter, and H E Umbarger (ed))*, Vol 1, pp. 391-407. ASM Press, Washington, D. C.

Rolfe MD, Rice CJ, Lucchini S, Pin C, Thompson A, Cameron AD, Alston M, Stringer MF, Betts RP, Baranyi J et al. 2012. Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *J Bacteriol* **194**: 686-701.

Rossi CC, Bosse JT, Li Y, Witney AA, Gould KA, Langford PR, Bazzolli DM. 2016. A computational strategy for the search of regulatory small RNAs in *Actinobacillus pleuropneumoniae*. *RNA* **22**: 1373-1385.

Sharma CM, Darfeuille F, Plantinga TH, Vogel J. 2007. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev* **21**: 2804-2817.

Sharma CM, Papenfort K, Pernitzsch SR, Mollenkopf HJ, Hinton JC, Vogel J. 2011. Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol Microbiol* **81**: 1144-1165.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539.

Steen JA, Steen JA, Harrison P, Seemann T, Wilkie I, Harper M, Adler B, Boyce JD. 2010. Fis is essential for capsule production in *Pasteurella multocida* and regulates expression of other important virulence factors. *PLoS Pathog* **6**: 1000750.

Storz G, Vogel J, Wassarman KM. 2011. Regulation by small RNAs in bacteria: expanding frontiers. *Mol Cell* **43**: 880-891.

Travers M, Paley SM, Shrager J, Holland TA, Karp PD. 2013. Groups: knowledge spreadsheets for symbolic biocomputing. *Database* **2013**: bat061.

Urban JH, Vogel J. 2007. Translational control and target recognition by *Escherichia coli* small RNAs *in vivo*. *Nucleic Acids Res* **35**: 1018-1037.

Urbanowski ML, Stauffer LT, Stauffer GV. 2000. The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*. *Mol Microbiol* **37**: 856-868.

Wilkie IW, Harper M, Boyce JD, Adler B. 2012. *Pasteurella multocida*: diseases and pathogenesis. *Curr Top Microbiol Immunol* **361**: 1-22.

Yang Q, Figueroa-Bossi N, Bossi L. 2014. Translation enhancing ACA motifs and their silencing by a bacterial small regulatory RNA. *PLoS Genet* **10**: 2.

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

Chapter 3

The role of the RNA-chaperone protein ProQ in the Gram-negative bacterium *Pasteurella multocida*.

Chapter 3: The role of the RNA-chaperone protein ProQ in the Gram-negative bacterium *Pasteurella multocida*.

3.1 Introduction

Pasteurella multocida is a bacterial pathogen that causes a wide range of diseases in a number of important production animal species. These diseases include fowl cholera in birds, haemorrhagic septicaemia and shipping fever in ungulates, as well as atrophic rhinitis and pneumonia in swine (Wilkie et al. 2012). *P. multocida* produces many virulence factors in order to cause disease. These include capsule (either A, B D, E or F serogroup) and lipopolysaccharide (LPS) that both afford antigenic variability and help protect against the innate immune system, as well as numerous iron acquisition systems, filamentous haemagglutinin and other adhesins (Fuller et al. 2000; Bosch et al. 2002a; Harper et al. 2004; Boyce and Adler 2006). Many of these virulence factors have been well characterised but exactly how their production is regulated is poorly understood. Understanding how bacteria regulate the production of crucial virulence factors is an important step in the development of novel methods to combat infection.

It is becoming increasingly clear that bacteria utilize small RNA (sRNA) molecules as regulators of protein production. This type of regulation involves the sRNA molecule binding to an mRNA target to modulate either translational efficiency or transcript abundance. The efficient binding of these two RNA species (sRNA and mRNA target) usually requires the presence of an RNA chaperone protein, as base pairing between the RNA species normally involves only 7-10 nt and is often an imperfect match. The best characterised of these RNA chaperone proteins is Hfq. The Hfq 3D structure, its RNA binding characteristics, and the sRNAs that interact with this molecule have been determined in a range of bacteria, including *E. coli*, *Salmonella enterica* serovar Typhimurium and *P. multocida* (Sauter et al. 2003; Sittka et al. 2007; Sittka et al. 2008a; Mégroz et al. 2016). In *P. multocida,* a *hfq* mutant showed reduced *in vivo* fitness, altered expression of capsule, filamentous haemagglutinin and a number of proteins used to modify the LPS structure (Mégroz et al. 2016). These data indicate that sRNA regulation plays a critical role in *P. multocida* pathogenesis. However, the identification and characterisation of the full cohort of *P. multocida* sRNAs has not yet been completed.

While Hfq is clearly an important RNA chaperone in bacteria, several other bacterial RNA chaperone proteins have been identified that contribute significantly to the RNA regulatory network. These other RNA binding proteins, including ProQ, assist in the interaction between a unique subset of sRNAs and

mRNA targets (Attaiech et al. 2017). ProQ was originally identified in *E. coli* as an osmoregulatory protein that controlled the production of a proline pump called ProP (Chaulk et al. 2011). However, recent research in *E. coli* and *S.* Typhimurium has demonstrated that ProQ acts as an important RNA chaperone (Smirnov et al. 2016; Gonzalez et al. 2017). Structural analysis of the *E. coli* ProQ has identified three distinct domains, a large N-terminal FinO-like domain, a C-terminal Tudor domain (Hfq-like), and a linker domain that joins the two aforementioned domains (Gonzalez et al. 2017). The FinO-like domain is highly conserved amongst ProQ proteins and shares structural and functional characteristics with the FinO RNA chaperone that is encoded on the *E. coli* IncF plasmid (Glover et al. 2015). The C-terminal domain of ProQ shares similarities with Hfq-like domains but recent structural data indicate that it shares more identity to eukaryotic Tudor domains (Gonzalez et al. 2017). In other proteins, Tudor domains function to bind ribonuclear proteins and they contain Zn^{2+} finger domains that allow them to bind RNA (Ponting 1997).

Recent work involving techniques that identify RNA species bound specifically to RNA binding proteins, such as gradient profiling by sequencing (Grad-seq), and cross-linking immunoprecipitation sequencing (CLIP-seq), have shown that ProQ functions similarly to Hfq (Smirnov et al. 2016; Holmqvist et al. 2018). The *S.* Typhimurium ProQ was confirmed to be an RNA chaperone that interacts with approximately 400 RNA transcripts, of which 18% were sRNA molecules. Of these sRNAs, only two were experimentally confirmed to also interact with Hfq. It was concluded that *S.* Typhimurium ProQ stabilised the interaction of a specialised subset of sRNA molecules and their mRNA targets (Smirnov et al. 2016). Further studies showed that ProQ primarily binds at the terminator stem loop structure in the 3' untranslated region (3' UTR) of mRNAs, where it stabilizes the transcript by blocking the action of RNaseIII (Holmqvist et al. 2018). However, while Hfq is known to bind to a linear/base pair motif, no such binding motif could be identified in ProQ targets and it was hypothesised that ProQ binds to a structural motif (Holmqvist et al. 2018). This was supported by the observation that each of the *S.* Typhimurium ProQ-binding sRNAs were highly structured, unlike the sRNAs that interact with Hfq (Smirnov et al. 2016). ProQ binds to sRNA molecules at a 1:1 ratio and it is predicted that this interaction causes the linker region in ProQ to stretch, straightening the RNA and allowing the binding sites to be uncovered (Gonzalez et al. 2017). Only a few RNA molecules known to bind ProQ have been fully investigated, including the MalM sRNA in *E. coli* that can bind to both Hfq and ProQ. However, electrophoretic mobility shift assays (EMSA) showed that when both Hfq and ProQ were present, the MalM sRNA preferentially bound to ProQ (Gonzalez et al. 2017).

To date, only the interaction of one ProQ-associated sRNA-mRNA pair, Raiz-*hupA*, has been characterised. In *S.* Typhimurium, RaiZ acts as an sRNA molecule once it is cleaved from the 3' UTR of the *raiA* mRNA.

The RaiZ sRNA is then stabilized by ProQ, allowing RaiZ to bind to its mRNA target *hupA,* which encodes the histone-like protein HU that acts in transcriptional regulation (Smirnov et al. 2017). The interaction of RaiZ with *hupA* blocks ribosome binding, leading to decreased HupA production (Smirnov et al. 2016).

There have been no investigations involving ProQ in any species within the Pasteurellaceae family. However, bioinformatic analyses of the available *P. multocida* genomes has identified that a ProQ homolog is encoded in all genomes. Understanding the function of ProQ, and the sRNAs under its control, could uncover crucial layers of regulation within *P. multocida* and other bacterial pathogens within the same family.

3.2 Materials and methods

3.2.1 Bacterial strains, media, plasmids and growth conditions

All bacterial strains and plasmids used in this study are listed in Table 3.1. Strains were grown as previously described (Gulliver et al. 2018). For low-iron growth, cells were grown in peptone water containing 1% glucose that had been pre-treated with dipyridyl at a concentration of either 100 µM or 150 µM in order to deplete the media of available iron.

3.2.2 DNA manipulations

All DNA manipulations were performed as per Gulliver et al. (2018). Oligonucleotides used in this study are listed in Table 3.2.

3.2.3 Construction of a *P. multocida proQ* mutant, overexpression and complementation strains

To inactivate *proQ* in the *P. multocida* strain VP161, TargeTron® mutagenesis (Sigma-Aldrich) was used as previously described (Steen et al. 2010), with the following modifications. The specificity of the group II intron within the *E. coli-P. multocida* TargeTron® shuttle vector, pAL953 (Table 3.1) (Harper et al. 2013), was retargeted to *proQ* using the splice overlap extension (SOE) PCR method described previously in the TargeTron® manual. Primers BAP7971, BAP7972 and BAP7973 (Table 3.2) were designed using the TargeTron® design site (Sigma-Aldrich). The plasmid containing the *proQ*-targeted intron, pAL1291 (Table 3.1), was used to transform *P. multocida* strain VP161 by electroporation and mutants containing a TargeTron® group II intron insertion in *proQ* were identified as previously described (Harper et al. 2013). The resultant strain was designated AL2973 (Table 3.1)

To produce a *proQ* complementation strain, the region of the *P. multocida* wild-type strain VP161 genome containing the *proQ* gene and the predicted native promoter was PCR-amplified using Taq polymerase with primers BAP8573 and BAP8574 (Table 3.2). The 1.2 kb PCR product generated was digested with BamHI and SalI and ligated to the BamHI and SalI digested vector pPBA1100S (Table 3.1). The ligation mix was used to transform *E. coli* DH5α. Transformants were screened for the correct plasmid using colony PCR with primers BAP2679 and BAP612 (Table 3.2). Plasmids were extracted from three transformants that were positive by colony PCR and separate sequencing reactions were performed using primers that flanked the inserted DNA (BAP2679 and BAP612, Table 3.2). One plasmid with the correct sequence was designated pAL1449 (Table 3.1) and was used to transform the *P. multocida proQ* mutant AL2973, resulting in the complementation strain, AL3357 (Table 3.1). As controls, the empty vector was also used to transform the *proQ* mutant strain and wild-type VP161, resulting in the strains AL3358 and AL3356 (Table 3.1), respectively.

For overexpression of ProQ, the intact *P. multocida* VP161 *proQ* gene was amplified using BAP7977 and BAP7978 (Table 3.2). The resultant fragment was cloned into the BamHI and SalI sites of the expression plasmid pAL99S (Table 3.1), such that *proQ* expression would be under the control of the constitutive, *P. multocida tpi* promoter. Transformants were screened by colony PCR, using oligonucleotides that flanked the multiple cloning site, and three plasmids positive by colony PCR were isolated from the transformants and sequenced using primers that flanked the insert region (BAP612 and BAP2679) (Table 3.2). One correct plasmid was designated pAL1294 (Table 3.1) and was used to transform the *P. multocida proQ* mutant strain AL2973 to generate the *proQ* overexpression strain, AL2978 (Table 3.1). In order to assess the possibility of polar effects of *proQ* inactivation, a *proQ* mutant derivative was also constructed that overexpressed the two genes directly downstream of *proQ*, namely, *prc* and *ycbB*. To do this, the two genes were PCR-amplified from *P. multocida* wild-type VP161 as a single fragment using BAP8384 and BAP8385 (Table 3.2) to produce a 3827 bp product. This fragment was then digested with EcoRI and XmaI and ligated to the EcoRI and XmaI sites of pREXY (Table 3.1). The ligation was used to transform *E. coli* DH5α, and colonies were screened by colony PCR using primers that flanked the insert region (BAP612 and BAP2679, Table 3.2). Plasmid was isolated from three colonies that were PCR positive; sequencing was then performed on these plasmids with flanking primers (BAP612 or BAP2679) and/or internal primers (BAP8422-BAP8425) (Table 3.2). One correct plasmid was designated pAL1387 (Table 3.1) and was used to transform the *P. multocida proQ* mutant strain AL2973. The resultant *prc/ycbB* overexpression strain was designated AL3214 (Table 3.1).

3.2.4 Proteomics analysis

Proteomic analysis was performed using label-free quantitative proteomics as per Gulliver et al. (2018), using biological triplicates of the wild-type parent *P. multocida* strain VP161, and *proQ* mutant strain (AL2973), grown to an OD_{600} of 0.6.

3.2.5 RNA extraction, quantitative reverse transcription-PCR (qRT-PCR), and whole genome transcriptomic analyses using RNA sequencing (RNA-seq)

RNA extractions were performed as per Gulliver *et al.* (2018), except strains were grown in biological triplicate to an OD_{600} of 0.6 before RNA extraction. Quantitative RT-PCR reactions were performed as per Gulliver *et al.* (2018), using the following strains; *P. multocida* wild-type strain VP161, VP161 *proQ* mutant (AL2973), VP161 *proQ* overexpression strain (AL2978) and the VP161 *prc/ycbB* overexpression strain (AL3214) (Table 3.1). Non-strand-specific RNA sequencing was performed as per Mégroz et al. (2016), using the wild-type parent strain *P. multocida* VP161 and the *proQ* mutant (strain AL2973, Table 3.1). Strand-specific RNA sequencing was performed using the SureSelect strand-specific RNA library

preparation kit (Agilent) as per the manufacturer's instructions with RNA from the following strains; *P. multocida* wild-type VP161 containing empty vector (AL3356), VP161 *proQ* mutant containing empty vector (AL3358) and the *proQ* mutant with pAL1449 containing an intact copy of *proQ* (AL3357) (Table 3.1). All libraries were sequenced on an Illumina NextSeq by Micromon Services (Monash University).

3.2.6 Northern blotting

Northern blotting was performed as per Rio (2014), with the following modifications. DIG-labelled RNA probes were synthesized by *in vitro* transcription from PCR products; the reverse primer used in each initial PCR contained a T7 promoter sequence. PCR products were generated using genomic VP161 DNA as template with primers BAP8571 and BAP8572, specific for PMVP_0063, or primers BAP8610 and BAP8611, specific for Prrc13 (Table 3.2). *In vitro* transcription was performed as per the manufacturer's instructions, using the DIG Northern starter kit version 10 (Roche). Detection of hybridizing fragments was performed as per Gulliver *et al*. (2018).

3.2.7 5' Rapid amplification of cDNA ends (RACE)

The 5' transcriptional start site of *prc* was determined using 5' RACE as per Gulliver *et al.* (2018), using the primer BAP8516 (reverse, located 138 bp downstream of ATG start) with the commercially supplied outer primer in a first round PCR and then the primer BAP8517 (reverse primer, anneals 90 bp downstream of the predicted ATG start) with the commercially supplied inner primer in a second round PCR.

Table 3.1. Strains and plasmids used in this study.

Strain or plasmid	Description	Source or Reference
Strains		
P. multocida		
VP161	Serotype A:1 virulent strain, avian isolate	(Wilkie et al. 2000)
AL571	VP161 *pcgC* mutant, Kan^R	(Harper et al. 2007b)
AL829	Complemented *pcgC* mutant; Kan^R, $Spec^R$	(Harper et al. 2007b)
AL1354	VP161 *petL* TargeTron® mutant; Kan^R	(Harper et al. 2017)
AL2234	VP161 *hyaD* TargeTron® mutant; Kan^R	(Mégroz et al. 2016)
AL2973	VP161 *proQ* TargeTron® mutant; Kan^R	This study
AL2978	AL2973 containing pAL1294; Kan^R, $Spec^R$	This study
AL2994	AL2973 containing pAL99S; Kan^R, $Spec^R$	This study
AL3067	VP161 *proQ/hyaD* TargeTron® double mutant; Kan^R	This study
AL3068	AL3067 containing pAL1339; $Spec^R$	This study
AL3069	AL3067 containing pAL1332; $Spec^R$	This study
AL3214	AL2973 containing pAL1387; $Spec^R$	This study
AL3356	VP161 containing pPBA1100S; $Spec^R$	This study
AL3357	AL2973 containing pAL1449; $Spec^R$	This study
AL3358	AL2973 containing pPBA1100S; $Spec^R$	This study
E. coli		
AL1995	DH5α containing pAL953	(Harper et al. 2013)
AL1296	DH5α containing pAL99S	(Harper et al. 2013)
AL2227	DH5α containing pAL1069; Kan^R, $Spec^R$	(Mégroz et al. 2016)
AL2708	DH5α containing pREXY	This study
AL2970	DH5α containing pAL1291; Kan^R, $Spec^R$	This study
AL3057	DH5α containing pAL1332; $Spec^R$	This study
AL3058	DH5α containing pAL1333; $Spec^R$	This study
AL3064	DH5α containing pAL1337; $Spec^R$	This study
AL3065	DH5α containing pAL1338; $Spec^R$	This study
AL3066	DH5α containing pAL1339; $Spec^R$	This study
AL3214	DH5α containing pAL1387; $Spec^R$	This study
AL3354	DH5α containing pAL1449; $Spec^R$	This study
DH5α	*deoR endA1 gryA96 hsdR17*($r_k^- m_k^+$) *recA1 relA1 supE44 thi-1* (*lacZYA-argFV169*) φ80*lacZ* ΔM15, F^-	Bethesda Research Laboratory

Plasmids		
pAL99S	*P. multocida/E. coli* expression and shuttle plasmid, Spec^R	(Harper et al. 2013)
pAL953	*P. multocida* plasmid (Spec^R) containing TargeTron® group II intron with *aph3* (Kan^R). Intron targeted to *gatA*	(Harper et al. 2013)
pAL1069	pAL953 with the TargeTron® group II intron re-targeted to *hyaD*	(Mégroz et al. 2016)
pAL1291	pAL953 with the TargeTron® group II intron re-targeted to *proQ*	This study
pAL1294	pAL99S containing the intact *P. multocida* VP161 *proQ* gene cloned into the SalI and BamHI sites using primers BAP7977 and BAP7978	This study
pAL1332	pREXY containing the intact *P. multocida* VP161 *proQ* gene cloned into the XmaI and EcoRI sites using primers BAP8088 and BAP8089	This study
pAL1333	pREXY containing the *P. multocida* VP161 *proQ* gene without the stop codon, cloned into the XmaI and EcoRI sites using primers BAP8088 and BAP8090	This study
pAL1337	pAL1069 with *aph3* (Kan^R) removed from TargeTron® group II intron. Used to generate to markerless *hyaD* TargeTron® mutant	This study
pAL1338	pAL1333 containing a DNA fragment encoding a His-TEV-tag cloned in-frame with and downstream of *proQ.* His-TEV-encoding fragment generated from oligonucleotides BAP8098 and BAP8099 and cloned into the EcoRI site of pAL1333	This study
pAL1339	pAL1338 containing a DNA fragment encoding a 3xFLAG-tag cloned in-frame with and downstream of the His-TEV-tagged ProQ. 3xFLAG-tag encoding fragment generated from oligonucleotides BAP8100 and BAP8101 and cloned into the EcoRI site of pAL1338	This study
pAL1387	pREXY containing intact *prc* and *ycbB* genes from wild-type *P. multocida* VP161, cloned into the EcoRI and XmaI sites using primers BAP8384 and BAP8385	This study
pAL1449	pPBA1100S containing 1.2 kb *proQ* fragment from wild-type *P. multocida* VP161 (including predicted promoter region), cloned into the BamHI and SalI sites using primers BAP8573 and BAP8574	This study
pPBA1100S	*P. multocida/E. coli* shuttle plasmid used to express recombinant proteins or sRNAs under the control of their predicted native promoters, Spec^R	(Gulliver et al. 2018)
pREXY	*P. multocida/E. coli* expression/shuttle plasmid used to express recombinant proteins or sRNAs under the control of a *P. multocida* constitutive promoter. Constructed by cloning a 96-bp region containing P_{tpi} promoter from strain VP161 into BamHI/HindIII sites of pPBA1100S using primers BAP7638 and BAP7639, Spec^R	(Gulliver et al. 2018)

Table 3.2. Oligonucleotides used in this study

Name	Sequence (5'-3') [a, b]	Description
BAP612 (universal)	GTAAAACGACGGCCAGT	Universal primer located upstream of the multiple cloning site (pPBA1100S) and the the *tpi* promoter (pAL99, pREXY)
BAP2067	GGAAGGAACAGTTTCTCTGGATTG	Forward primer specific for *P. multocida* VP161 *hyaD*. Used to PCR-amplify region containing TargeTron® insertion site in *hyaD*
BAP2679	TTGTGTGGAATTGTGAGCGGA	Reverse primer located downstream of multiple cloning site in pAL99, pPBA1100S and pREXY
BAP7971	AAAAAAGCTTATAATTATCCTTAGTAAGCAAAACAGTGC GCCCAGATAGGGTG	IBS TargeTron® primer specific for *proQ* mutagenesis
BAP7972	CAGATTGTACAAATGTGGTGATAACAGATAAGTCAAAAC ACATAACTTACCTTTCTTTGT	EBS1d TargeTron® primer specific for *proQ* mutagenesis
BAP7973	TGAACGCAAGTTTCTAATTTCGATTCTTACTCGATAGAGG AAAGTGTCT	EBS2 TargeTron® primer specific for *proQ* mutagenesis
BAP7977	GCCACG**GGATCC**AAACAGAGCCACATTCTATCCTGTCT	Forward primer for PCR amplification of *proQ*; beginning 107 bp upstream of the *proQ* start codon. Contains a BamHI site
BAP7978	TTGTCG**GTCGAC**GATTTTCTTCCGTAGGTTGTGGCGTA	Reverse primer for PCR amplification of *proQ*; ending 230 bp downstream of the *proQ* stop codon. Contains a SalI site
BAP8088	GCTTTT**CCCGGG**GTAAATTAATGTAT	Forward primer for PCR-amplification of *proQ*. Primer includes the *proQ* start codon and an XmaI site
BAP8089	ACTTAT**GAATTC**TTATGCAAACAGATGTTCA	Reverse primer for amplification of *proQ*. Primer includes the native *proQ* stop codon and an EcoRI site
BAP8090	TATTTC**GAATTC**TGCAAACAGATGTTC	Reverse primer for amplification of *proQ*. In same position as BAP8089 but native stop codon excluded. Contains an EcoRI site
BAP8098	AATTAATGGAGCACCATCACCATCACCATGATTATGATAT TCCAACTACTGCTAGCGAGAATTTGTATTTTCAGG	Sense strand encoding His TEV tag; used with BAP8099 to produce a double stranded fragment with ends compatible with an EcoRI site. Used to construct pAL1333
BAP8099	AATTCACCCTGAAAATACAAATTCTCGCTAGCAGTAGTTG GAATATCATAATCATGGTGATGGTGATGGTGCTCC	Antisense strand encoding His TEV tag; used with BAP8098 as above
BAP8100	AATTGGACTACAAAGATGACGACGATAAAGACTACAAA GATGACGACGATAAAGACTACAAAGATGACGACGATA	Sense strand encoding 3xFLAG tag; used to anneal with BAP8101 to produces double stranded fragment with ends compatible with an EcoRI site. Use to construct pAL1338

BAP8101	AATTCTCATTTATCGTCGTCATCTTTGTAGTCTTTATCGTC GTCATCTTTGTAGTCTTTATCGTCGTCATCTTTG	Antisense strand encoding 3xFLAG; used with BAP8100 as above
BAP8384	GATGAA**CCCGGG**ACAAAATGCCATCAAATAA	Forward primer flanking *prc* and *ycbB*. Contains an XmaI site for cloning into pREXY
BAP8385	CTCTCT**GAATTC**TTTCATGCTTAGTTTGACC	Reverse primer flanking *prc* and *ycbB*. Contains an EcoRI site for cloning into pREXY
BAP8422	GAGAAGTGGATCCGCACACAA	Internal forward primer 1 for sequencing *prc*
BAP8423	GGCGCATTAACCGAAGCGG	Internal forward primer 2 for sequencing *prc*
BAP8424	CCCGAATTTGTCGCATTAAATGAAGAGC	Internal forward primer 3 for sequencing *prc*
BAP8425	GAAGAGGAACGTCTTGCAGCAG	Internal forward primer 4 for sequencing *ycbB*.
BAP8516	ATTTTCTTCCGTAGGTTGTGGCGTAA	5' RACE outer primer specific to the *prc* transcript of *P. multocida* VP161
BAP8517	GGCTGTACTGCCTCGACAAGATTAAAGCTCAA	5' RACE inner primer specific to the *prc* transcript of *P. multocida* VP161
BAP8571	CGCACTGATTGAAAACAAGG	Forward primer for PCR amplification of PMVP_0063 region. Anneals 26-bp upstream of predicted start codon. For *in vitro* transcription production of a riboprobe
BAP8572	<u>TAATACGACTCACTATAG</u>GGCATTAAGGGCTTTCCCCAGT	Reverse PCR primer, anneals 179-bp downstream of predicted start codon for PMVP_0063. For *in vitro* transcription production of a riboprobe. Contains T7 RNA polymerase promoter sequence at 5' end
BAP8573	ATTCAG**GGATCC**GCCACGTGTAATAAACAGAGCCA	Forward primer for PCR amplification of *proQ*, including the native promoter region. Contains a BamHI site for cloning into pPBA1100S
BAP8574	TAATTG**GTCGAC**TTCCGTAGGTTGTGGCGTAAC	Reverse primer for PCR amplification of *proQ* with the native promoter region. Contains a SalI site for cloning into pPBA1100S
BAP8610	TGGACGAAAGTAAGTGCAAGAA	Forward primer for PCR amplification of an 82-bp fragment of the sRNA Prrc13. For *in vitro* transcription production of a riboprobe
BAP8611	<u>TAATACGACTCACTATAG</u>CTTCCGTGCCTGTAACGAAT	Reverse primer for amplification of an 82-bp fragment representing the putative sRNA Prrc13. Contains T7 RNA polymerase promoter sequence at 5' end. For *in vitro* transcription production of a riboprobe
EBS universal	CGAAATTAGAAACTTGCGTTCAGTAAAC	TargeTron® universal primer used for Sanger sequencing and PCR-based re-targeting of TargeTron® group II intron.

[a.] All restriction sites are shown in bold

[b.] T7 RNA polymerase promotor sequences are underlined

3.2.8 Hyaluronic acid capsule assay

The amount of hyaluronic acid capsule produced by the *P. multocida* serogroup A strain VP161 and its derivatives (AL2973, AL2978, AL2994 and AL2234, Table 3.1) was determined as per Chung et al. (2001).

3.2.9 Carbohydrate silver staining and phosphocholine detection

The overall profile of the LPS produced by selected *P. multocida* strains was visualized using carbohydrate silver staining following the separation of whole cell lysates via PAGE using a 15% polyacrylamide gel as per Harper et al. (2007b). Immunoblotting was used to detect the presence of phosphocholine (PCho) on the LPS as described previously (Harper et al. 2007b). Briefly, a 1/800 dilution of TEPC-15 primary antibody in TBS-tween was added to the membrane and incubated for 1 h at 37°C. Following three 10 min washes in TBS-tween at room temperature, a 1/1000 dilution of goat anti-mouse immunoglobulin A horseradish peroxidase conjugate in TBS-tween was added to the membrane and incubated for 1 h at room temperature. The membrane was washed a further three times for 10 min in TBS-tween before detection using Amersham ECL Western Blotting Detection Reagent (GE Healthcare). Immunoblots were visualized using the Fujifilm LAS-3000 image reader or by exposure to X-ray film (Kodak).

3.2.10 PlpE immunoblotting

To detect the presence of the *P. multocida* outer membrane protein PlpE, immunoblotting of whole cell lysate samples was performed as per Hatfaludi et al. (2012), with the following modifications. The wild-type strain *P. multocida* VP161, *proQ* mutant (AL2973), *proQ* overexpression strain (AL2978), and *proQ* mutant containing empty vector strain (AL2994) (Table 3.1) were grown in biological triplicate in heart infusion (HI) media to mid-exponential growth phase (OD_{600} = 0.6). The primary antibody, chicken anti-recombinant PlpE (Hatfaludi et al. 2012), was used at a 1/250 dilution, whilst the secondary antibody, donkey anti-chicken IgY, was used at a 1/1000 dilution. Antibodies bound to the membrane were detected as described above.

3.2.11 Fowlicidin-1 sensitivity assay

Sensitivity to the chicken antimicrobial peptide Fowlicidin-1 (RVKRVWPLVIRTVIAGYNLYAIKKK) (Xiao et al. 2006) was determined as per Harper et al. (2007a) with the following modifications. The wild-type strain *P. multocida* VP161, the *proQ* mutant (AL2973) plus a positive control (Fowlicidin-1 sensitive *petL* mutant AL1354, Table 3.1) were grown to OD_{600} = 0.4 in HI broth before 500 µL of each strain was pelleted by centrifugation at 13000 x *g* for 2 min, followed by resuspension of the cell pellet in buffer A (10 mM phosphate, 30% HI broth). Cells were then added to a standard 96-well tray in equal volume (25 µl) to fowlicidin-1 at a range of concentrations (0-5 µM). Cells were then incubated for 3 h at 37°C before serial dilution and plating onto HI agar to determine the number of surviving bacteria.

3.2.12 Construction of strains for UV-CLASH

To produce a His-TEV-FLAG (HTF)-tagged ProQ in a *P. multocida* strain that was non-virulent and therefore safe for use in the available UV-CLASH apparatus (University of New South Wales), the following manipulations were performed. A *P. multocida* VP161 *hyaD* mutant had been shown previously to be acapsular and highly attenuated (Mégroz, Boyce Laboratory, unpublished). Therefore, to attenuate the *proQ* mutant AL2793, a previously constructed TargeTron® plasmid (pAL1069, Table 3.1) targeted to the VP161 *hyaD* gene (Mégroz et al. 2016), was modified to remove the Kanamycin gene present on the intron and thus make the introduced mutation markerless. The kanamycin gene was removed by restriction enzyme digestion of pAL1069 with MscI followed by re-ligation of the plasmid with T4 DNA ligase. The resulting plasmid, pAL1337, was used for TargeTron® mutagenesis of the *proQ* mutant (AL2793, Table 3.1) as described previously (Gulliver et al. 2018). Transformants cured of the TargeTron® plasmid were screened for the correct intron insertion into the genome by colony PCR with BAP2067 (located upstream of the target region in *hyaD*) paired with EBS universal primer (reverse primer located within the TargeTron® intron, Table 3.2). Mutants that were positive by colony PCR were then sequenced using genomic DNA as template with EBS universal primer (Table 3.2) to confirm the site of intron insertion. Southern blotting was also employed using DIG-labelled probes specific for the kanamycin gene or the group II intron to confirm that there was only a single intron insertion, and this had occurred at the correct site. One of several confirmed *hyaD*/*proQ* double mutants was assigned the strain name AL3067 (Table 3.1) and used for further study.

In order to express a HTF-tagged ProQ protein for co-immunoprecipitation, the *proQ* gene was PCR-amplified (without its native stop codon) using Phusion high-fidelity DNA polymerase (NEB) and the primers BAP8088 and BAP8090 (Table 3.2). The PCR product generated was digested with XmaI/EcoRI and ligated to into XmaI/EcoRI digested pREXY to generate the interim plasmid, pAL1333 (Table 3.1). To insert the His-TEV-tag at the 3' end of *proQ* into pAL1333, 50 µM of primers BAP8098 and BAP8099 (Table 3.2) were combined in annealing buffer (1 M Tris- HCl pH 8.0, 10mM EDTA and 5 M NaCl) and annealed using the following thermal cycling conditions; 95°C for 5 min, followed by 70 cycles of 1 min with the temperature reduced by 1°C per cycle (e.g. 94°C for 1 min, 93°C for 1 min, etc), followed by a 12°C hold. The resulting double stranded DNA fragment was ligated to EcoRI digested pAL1333. The primers were designed such that the ligation of the product to pAL1333 regenerated an EcoRI site at the 3' end but not at the 5' end. Colonies were screened using colony PCR with vector primers BAP2679 and BAP612 that flanked the insert (Table 3.2). The correct plasmid was designated pAL1338 (Table 3.1). Next, a DNA fragment encoding a 3x FLAG-tag was constructed by annealing the primers BAP8100 and BAP8101 as

described above. The resultant double stranded fragment was inserted into the EcoRI site of pAL1338. Colonies were screened as above and the resultant plasmid containing a His-TEV-3xFLAG(HTF)-tagged *proQ* gene was designated pAL1339. A second plasmid was also constructed for use as the negative, untagged ProQ, control. For the control plasmid, the primer BAP8088 was used with primer BAP8089 to amplify *proQ,* which included the sequence of the native stop codon. The PCR product generated was cloned into pREXY, in the same manner as described above for pAL1333, to generate the plasmid, pAL1332 (Table 3.1). For use in the UV-CLASH experiments, the plasmids pAL1339 (encoding HTF-tagged ProQ) and pAL1332 (control encoding untagged ProQ) were then separately used to transform, via electroporation, the *P. multocida hyaD*/*proQ* double mutant (AL3067) to generate the strains AL3068 and AL3069, respectively (Table 3.1). The expression of HTF-tagged ProQ in AL3068 was confirmed using a Western immunoblot with an anti-FLAG antibody.

3.2.13 Preparation of UV-CLASH libraries

UV-CLASH was performed as per Waters et al. (2017), with the following modifications. Cells from AL3068 and AL3069 (containing the plasmid pAL1339 encoding HTF-tagged ProQ and the plasmid pAL1332 encoding untagged ProQ, respectively) were grown to OD_{600} = 0.8 in HI media in biological triplicate prior to UV-crosslinking. As per Waters et al. (2017) cells were lysed, and HTF-tagged ProQ:RNA complexes were extracted on anti-FLAG resin. The FLAG-tag on the HTF-tagged ProQ was then cleaved off using the TEV enzyme, and the protein:RNA complexes were further purified using a Ni-NTA slurry. The attached RNAs were then trimmed, radiolabelled and adapter ligated, prior to the complexes being run on an acrylamide gel. The products of the correct size were purified and the tagged-ProQ proteins were degraded using proteinase K. The RNA was then reverse transcribed into cDNA before being PCR amplified, and sent for single-end 100-bp HiSeq2500 sequencing (Ramiciotti centre, UNSW).

3.2.14 Analysis of binding and CLASH hybrids

CLASH hybrids were analysed as per Waters et al. (2017) using the *hyb* package (Travis et al. 2014). Analysis of RNA molecules binding to ProQ was determined as per Holmqvist et al. (2016). The list of ProQ binding RNA species and ProQ bound RNA:RNA hybrids was further analysed using Cytoscape (Shannon et al. 2003) and Intergrated Genome Browser (Freese et al. 2016) to visualise bound RNA species and their interactions.

3.3 Results

3.3.1 The *P. multocida proQ* mutant shows normal growth and osmotic tolerance

In order to determine the role of ProQ within *P. multocida,* an insertional *proQ* mutant was constructed in the *P. multocida* strain VP161 (AL2973). The growth of the *proQ* mutant was compared to the growth of the wild-type parent strain VP161 by culturing both (in biological triplicate) in HI broth. The growth curves generated for the two strains were indistinguishable, indicating that the *proQ* mutant displayed normal growth *in vitro* (Figure 3.1A). In *E. coli*, ProQ has previously been shown to be involved in the production of ProP, an osmoregulatory protein pump, leading to changes in osmoregulation (Smith et al. 2007). To determine if *proQ* inactivation affected osmoregulation in *P. multocida*, the wild-type *P. multocida* strain VP161 and the *proQ* mutant (AL2973) were grown in HI broth containing 300 mM NaCl to induce osmotic stress. Optical density readings (OD_{600}) were taken every 30-60 min and viable counts were taken at 3 h, 6 h and 24 h (Figure 3.1B & 3.1C). There was no difference observed in growth rate or viability of the wild-type and *proQ* mutant under these conditions, indicating that *proQ* inactivation does not affect osmoregulation in *P. multocida*.

3.3.2 The effect of *P. multocida* ProQ inactivation on global protein production

To determine the role that ProQ plays in the regulation of protein production within *P. multocida*, the proteomes of the *P. multocida* wild-type VP161 and *proQ* mutant (AL2793) were compared at mid-exponential growth phase (OD_{600} = 0.6), using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). Samples were prepared from biological triplicate cultures representing each strain. Proteins were identified as being differentially produced if they showed a ≥ 1.5-fold change ($\log_2$ = 0.59) in production with a false discovery rate (FDR) of less than 0.05. The analysis identified 20 proteins with differential production. The ProQ protein was identified as having strongly decreased production (>230-fold) in the *proQ* mutant, confirming that the TargeTron® insertion had successfully inactivated the *proQ* gene. Thirteen other proteins showed decreased production in the *proQ* mutant, namely; Tpl, PM1457, TnaA, RecN, GalT, PM0165, PM1325, MenE, GlpQ, MglB, Prc, RecA and PM0452 (Table 3.3). Six proteins showed increased production in the *proQ* mutant, namely; PM0834, PlpE, PM1854, PM0337, PM0336, and FxsA (Table 3.4).

The proteins that showed differential production in the *proQ* mutant (AL2973) were grouped according to their predicted biological pathways/general function within the cell (Table 3.3 and Table 3.4). Four proteins were predicted to be involved in iron transport or metabolism (inorganic ion transport or metabolism), two of which (PM1457 and PM0452) showed decreased production, whilst the remaining two (PM0337 and PM0336) showed increased production. Seven proteins were predicted to be involved

in the transport and metabolism of other substrates; two in the transport and metabolism of amino acids (Tpl and TnaA), three in carbohydrate metabolism and transport, (PM1325, MglB and PM0834), and two in lipid metabolism and transport (MenE and PM0165). Eight proteins were predicted to be involved in other cellular processes, including cell wall/membrane biogenesis (Prc, PlpE and FxsA), energy production and conservation (GalT, GlpQ and PM1854) and replication and recombination (RecA and RecN).

As the production of four proteins predicted to be involved in iron transport and metabolism was altered in the *proQ* mutant, this strain was examined for changes in its ability to grow under low iron conditions. Biological triplicate cultures of wild-type VP161 and *proQ* mutant were grown in peptone water containing 1% glucose that had been pre-treated with dipyridyl at a concentration of either 100 µM or 150 µM in order to deplete the media of available iron. The optical density (OD_{600}) of the cultures was measured at regular intervals for 24 h. Growth curve analysis revealed that there was no difference in growth between the strains at either concentration of dipyridyl tested (Figure 3.1D).

Protomic data indicated that the outer membrane protein, PlpE, showed increased production (1.67-fold) in the *proQ* mutant. An immunoblot blot was performed in an effort to detect this altered production. Whole cell lysates of the *proQ* mutant (AL2973), the *proQ* overexpression strain (AL2978), the *proQ* mutant containing the empty vector, pAL99S (AL2994) and the wild-type strain VP161 (Figure 3.2A) were included in the study. Densitometry analysis of the amount of bound PlpE antibodies was then conducted (Figure 3.2B). A Coomassie stained SDS-PAGE gel (Figure 3.2A), containing an identical amount of each whole cell lysate, was used as a comparison. The densitometry analysis indicated that PlpE production in the *proQ* mutant strain was approximately 2-fold higher than the wild-type strain and the *proQ: proQ* complemented strain but this difference fell short of statistical significance ($p = 0.07$).

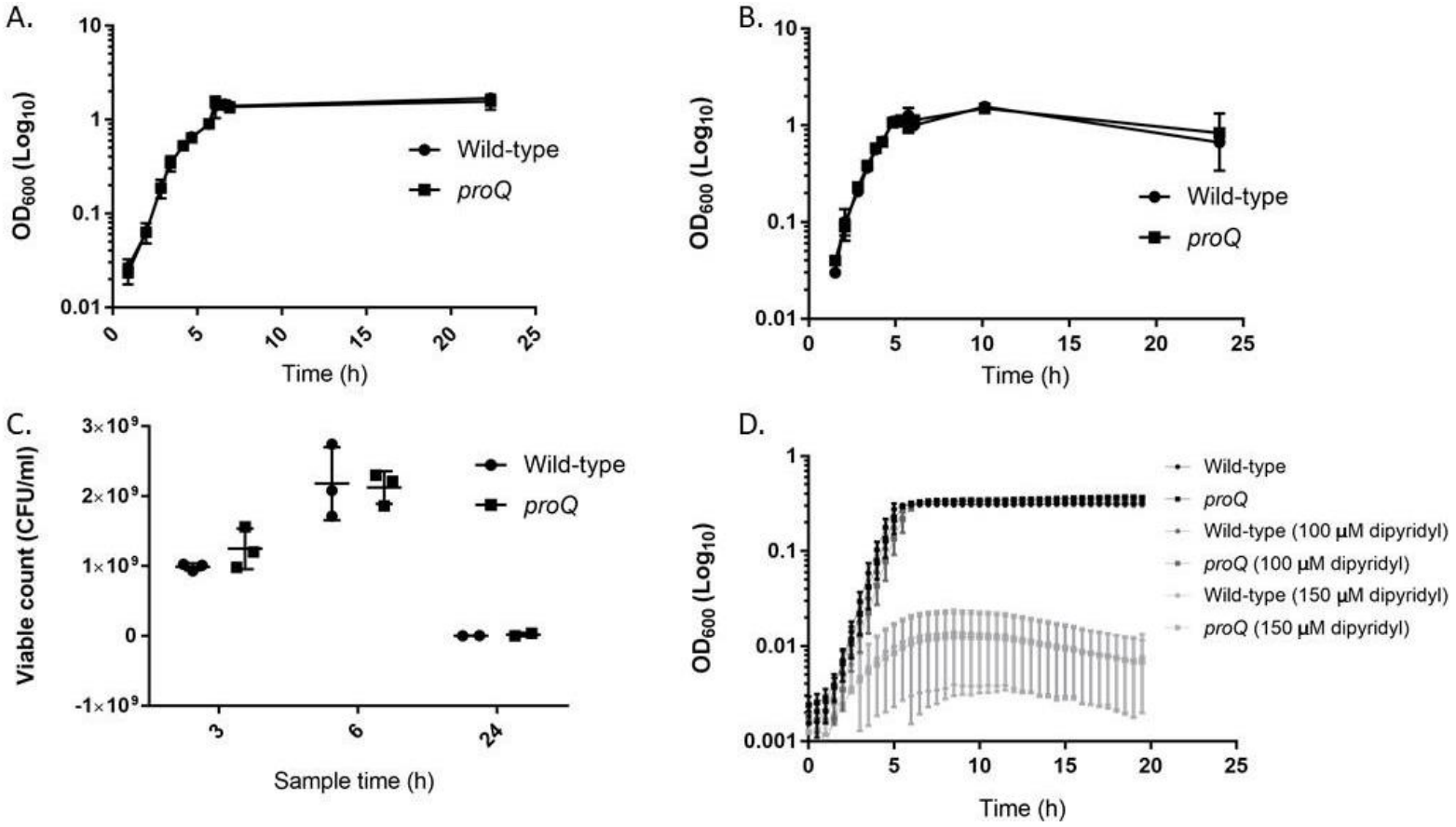

Figure 3.1. Growth kinetics and viability of the wild-type *P. multocida* VP161 and *proQ* mutant in different growth media A. Growth curve of wild-type *P. multocida* VP161 (circles) and the *proQ* mutant (AL2973) (squares) over 24 h when incubated in Heart infusion (HI) broth with shaking at 37°C. Data shown are mean ± SD (n = 3). **B.** Growth curve of wild-type *P. multocida* VP161 (circles) and the *proQ* mutant (AL2973) (squares) over 24 h when incubated in HI broth containing 300 mM NaCl with shaking at 37°C for 24 h. Data shown are mean ± SD (n =3). **C.** Viable counts of the wild-type *P. multocida* VP161 strain and *proQ* mutant (AL2973) taken at 3, 6 and 24 h during growth in HI broth containing 300 mM NaCl. Data shown are mean ± SD (n = 3). **D.** Growth curve of wild-type *P. multocida* VP161 (circles) and the *proQ* mutant (AL2973) (squares) over 24 h, with shaking at 37°C, when grown in peptone water containing 1% glucose, either untreated (0 µM [black]), or pre-treated with dipyridyl (100 µM [dark grey] or 150 µM [light grey]). Data shown are mean ± SD (n = 3).

Table 3.3. Proteins with decreased production in the *P. multocida proQ* mutant (AL2973) as compared to production of the same protein in the *P. multocida* VP161 wild-type parent.

Protein name[a]	VP161 locus tag (PM70 locus tag)[b]	Log_2 fold-change, (FDR)	Predicted protein product	General function prediction
PM0165	PMVP_0119, (PM0165)	-1.47, (0.0183)	Conserved hypothetical protein	Lipid transport and metabolism
ProQ	PMVP_0227, (PM0268)	-7.85, (0.001)	Putative solute/DNA competence effector	Signal transduction mechanisms genes
Prc	PMVP_0228, (PM0269)	-0.96, (0.0143)	Carboxy-terminal protease	Cell wall/membrane biogenesis
RecN	PMVP_0298, (PM0332)	-1.57, (0.0042)	DNA repair protein RecN	Replication, recombination and repair
MenE	PMVP_0326, (PM0357)	-1.10, (0.0315)	O-succinylbenzoic acid--coA ligase	Lipid transport and metabolism
PM0452	PMVP_0425, (PM0452)	-0.73, (0.0425)	Periplasmic iron binding protein	Inorganic ion transport and metabolism
Tpl	PMVP_0807, (PM0811)	-3.50, (0.0143)	Tyrosine phenol-lyase, TPL	Amino acid transport and metabolism
GalT	PMVP_1051, (PM1036)	-1.51, (0.0189)	Galactose-1-phosphate uridylyltransferase	Energy production and conversion
MglB	PMVP_1053, (PM1038)	-0.97, (0.0313)	Galactose ABC transporter, periplasmic-binding protein	Carbohydrate transport and metabolism
PM1325	PMVP_1353, (PM1325)	-1.15, (0.0212)	Binding protein component precursor of ABC ribose transporter	Carbohydrate transport and metabolism
TnaA	PMVP_1470, (PM1420)	-1.61, (0.0212)	Tryptophanase	Amino acid transport and metabolism
GlpQ	PMVP_1495, (PM1444)	-1.01, (0.0221)	Glycerophosphodiester phosphodiesterase	Energy production and conversion
PM1457	PMVP_1508, (PM1457)	-2.57, (0.0078)	Iron (III) transport system substrate-binding protein	Inorganic ion transport and metabolism
RecA	PMVP_1870, (PM1817)	-0.88, (0.0189)	RecA bacterial DNA recombination proteins	Replication, recombination and repair genes

[a] Differentially expressed proteins were defined as those showing at least 1.5-fold decreased production ($log_2 \leq -0.59$) with a false discovery rate (FDR) of less than 0.05.

[b] The PMVP locus tag (for VP161) is shown first followed by locus tag of the closest ortholog in *P. multocida* strain Pm70, the genome of which is publicly available.

Table 3.4. Proteins with increased production in the *P. multocida proQ* mutant (AL2973) as compared to the VP161 wild-type parent.

Protein name[a]	**VP161 locus tag (PM70 locus tag)** [b]	**Log_2 fold-change, (FDR)**	**VP161 product**	**General function prediction**
PM0336	PMVP_0302, (PM0336)	1.3, (0.0167)	Hemoglobin binding protein B	Inorganic ion transport and metabolism
PM0337	PMVP_0304, (PM0337)	1.01, (0.0402)	Hemoglobin/transferrin/lactoferrin receptor protein	Inorganic ion transport and metabolism
PM0834	PMVP_0833, (PM0834)	0.71, (0.0402)	Pts system, mannose-specific iiab component	Carbohydrate transport and metabolism
FxsA	PMVP_1127, (PM1105)	1.95, (0.0359)	FxsA protein, putative	Cell wall/membrane biogenesis
PlpE	PMVP_1573, (PM1517)	0.74, (0.0212)	Outer membrane lipoprotein PlpE, putative	Cell wall/membrane biogenesis
PM1854	PMVP_1907, (PM1854)	0.75, (0.0172)	Iron-sulfur cluster binding protein	Energy production and conversion

[a] Differentially expressed proteins were defined as those showing at least 1.5-fold increased production ($log_2 \geq 0.59$) with a false discovery rate (FDR)of less than 0.05.

[b] The PMVP locus tag (for VP161) is shown first followed by the locus tag of the closest ortholog in *P. multocida* strain Pm70, the genome of which is publicly available.

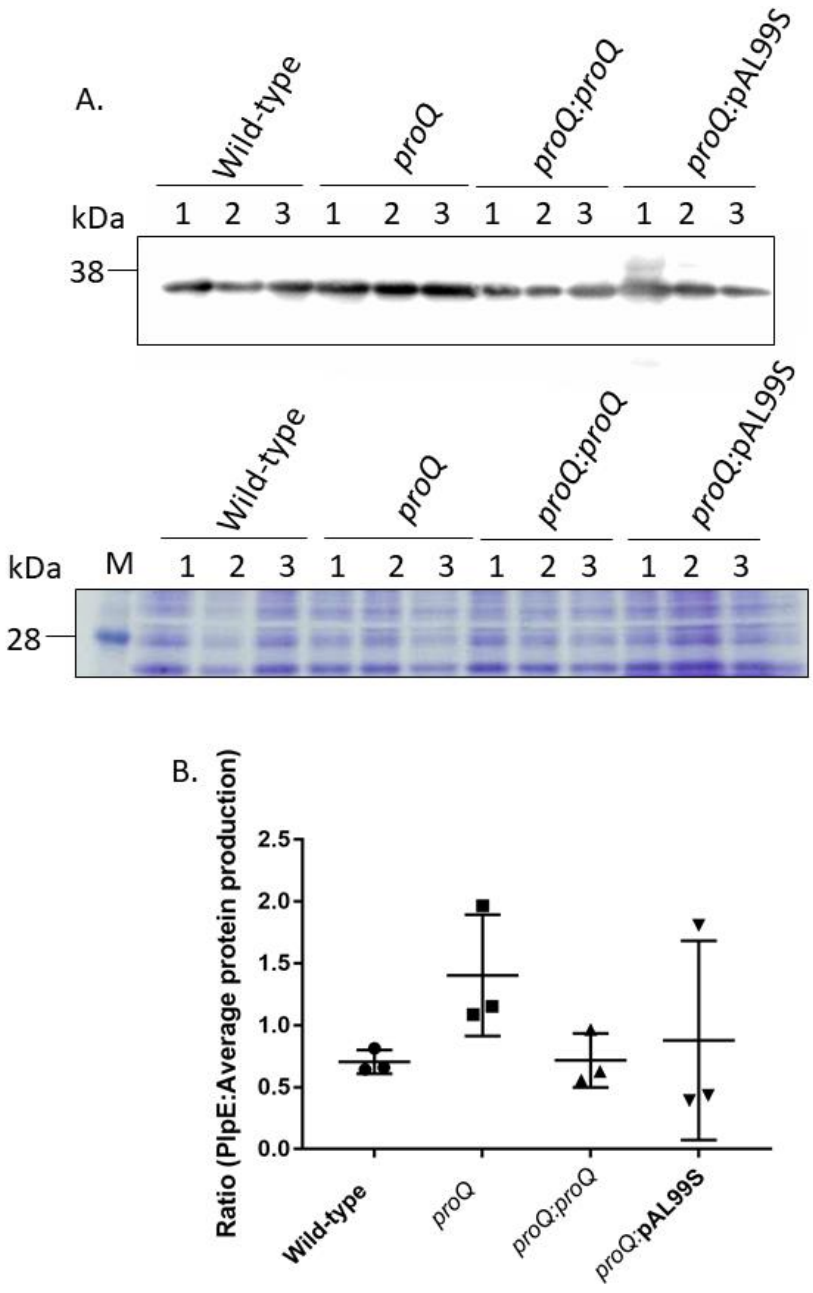

Figure 3.2. A. Western immunoblot using PlpE antibodies (top image) and a Coomassie stained SDS-PAGE gel of the same whole cell lysate samples (in biological triplicate). Whole cell lysates were derived from *P. multocida* wild-type VP161 (wild-type), the *proQ* mutant (*proQ*), the *proQ* overexpression strain (*proQ: proQ*) and the *proQ* mutant containing empty vector (*proQ*: pAL99S). For densitometry, proteins between 20 kDa and 30 kDa on the Coomassie stained gel were used for standardisation **B.** Graphical representation of the densitometry data generated from the PlpE Western blot containing *P. multocida* wild-type VP161 (circles), *proQ* mutant (squares), *proQ* overexpression strain (upright triangles) and *proQ* mutant containing empty vector (upside-down triangles). Data shown are mean ± SD, with n = 3.

3.3.3 The effect of *proQ* inactivation on *P. multocida* gene expression

To determine the role of ProQ in RNA transcript stability and/or degradation, the transcriptome of the *P. multocida proQ* mutant (AL2973) was compared to the wild-type *P. multocida* strain VP161. Biological triplicate samples of each strain were grown to mid-exponential phase (OD_{600} = 0.6), then total RNA was extracted. Samples were depleted of ribosomal RNA and the remaining RNA was used as template for cDNA synthesis which was then ligated to the appropriate adapters and sequenced. Transcripts were determined to be differentially expressed if there was a ≥ 2-fold change ($\log_2 \geq 1$) in expression (FDR < 0.05). Using this criteria, 131 transcripts were identified as differentially expressed between the two strains. Of these, 96 showed decreased expression (Table 3.5) and 35 showed increased expression (Table 3.6) in the *proQ* mutant. The differentially expressed transcripts were grouped according to the predicted general function within the cell (Table 3.5 and 3.6) and notably this grouping showed that 27 tRNAs with increased expression in the *proQ* mutant (Table 3.6). Seventeen putative sRNAs also showed altered expression; five had increased expression in the *proQ* mutant (PMVP_2523, Prrc11, Prrc25, Prrc40, and Prrc44, Table 3.6) and 12 showed decreased expression in the *proQ* mutant (Prrc05, Prrc06, Prrc10, Prrc12, Prrc15, Prrc19, Prrc20, Prrc24, Prrc30, Prrc32, Prrc42 and Prrc50, Table 3.5). Interestingly, the expression of the gene encoding the known RNA chaperone protein, Hfq, was also significantly decreased ($\log_2$ fold-change = -1.94) in the *proQ* mutant (Table 3.5).

The transcriptomic analysis identified an increase in the expression of *hexB* in the *proQ* mutant ($\log_2$ fold-change = 1.00, FDR = 0.014) (Table 3.5), which encodes a component of the hyaluronic acid capsule export system (Chung et al. 1998). A decrease in expression of *hssB* in the *proQ* mutant strain was also observed ($\log_2$ fold-change = -1.17, FDR = 0.0044) (Table 3.6). The gene *hssB* encodes a secondary hyaluronan synthase which may play a role in capsule biosynthesis under some as yet specified conditions (Deangelis and White 2004). To determine if these changes impacted *P. multocida* capsule synthesis and/or transport, extracellular hyaluronic acid capsule production in the *P. multocida proQ* mutant (AL2973) was compared to production in the wild-type *P. multocida* strain VP161. As controls, the *proQ* overexpression strain (AL2978), the *proQ* mutant containing empty vector (AL2994) and the acapsular *hyaD* mutant (AL2234) were included in the assay. The *hyaD* mutant (Mégroz et al. 2016) was used as a negative control as it shows highly reduced capsule production (Figure 3.3A). The results showed that was no difference in capsule production observed between the wild-type VP161, the *proQ* mutant and the *proQ* overexpression strain.

Table 3.5. Transcripts with decreased expression in the *P. multocida proQ* mutant (AL2973) compared to expression in the VP161 wild-type parent strain.

VP161 locus tag, (PM70 locus tag)[a.b.c.]	**Gene name**	**Log_2 fold-change, (FDR)[d.]**	**Predicted product**	**General function prediction**
PMVP_0041, (NA)		-1.43, (0.0081)	Hypothetical	No function prediction
PMVP_0061, (NP)		-1.73, (0.0031)	RNA-directed DNA polymerase	No function prediction
PMVP_0062, (NP)		-2.19, (0.0011)	XRE-family transcriptional regulator	No function prediction
PMVP_0063, (NP)		-1.20, (0.0108)	Hypothetical	No function prediction
PMVP_0153, (PM0199)	*frdC*	-1.34, (0.0025)	Fumarate reductase subunit FrdC	Energy production and conversion genes
PMVP_0164, (PM0209)	PM0209	-1.14, (0.0007)	Sigma 70 family RNA polymerase sigma factor	Transcription genes
PMVP_0228, (PM0269)	*prc*	-1.83, (0.00003)	Carboxy-terminal protease	Cell wall/membrane biogenesis genes
PMVP_0229, (PM0270)	*ycbB*	-1.20, (0.0008)	Murein L, D-transpeptidase	Cell wall/membrane biogenesis genes
PMVP_0253, (NA)	*mliC*	-1.06, (0.0112)	Lysozyme inhibitor	No function prediction
PMVP_0255, (PM0294)	PM0294	-1.26, (0.0034)	Hypothetical protein PM0294	No function prediction
PMVP_0269, (NA)		-1.68, (0.0176)	Hypothetical	No function prediction
PMVP_0271, (NA)		-2.62, (0.0229)	Hypothetical	No function prediction
PMVP_0321, (PM0352)	*fur*	-1.53, (0.0008)	Ferric uptake regulation protein	Inorganic ion transport and metabolism genes
PMVP_0323, (PM0354)	PM0354	-1.16, (0.0095)	LexA regulated protein	No function prediction
PMVP_0341, (PM0372)	*sdh*	-1.07, (0.0077)	Saccharopine dehydrogenase family protein	Amino acid transport and metabolism genes
PMVP_0385, (PM0414)	*hssB*	-1.17, (0.0044)	Heparosan synthase	Cell wall/membrane biogenesis genes
PMVP_0386, (PM0415)	PM0415	-1.28, (0.0131)	Conserved hypothetical protein	No function prediction
PMVP_0396, (PM0424)	PM0424	-2.75, (0.0011)	Conserved hypothetical protein	No function prediction
PMVP_0463, (NA)		-1.53, (0.0031)	SMI1/KNR4 family protein	No function prediction
PMVP_0468, (PM0496)	PM0496	-1.37, (0.0015)	DUF1436 domain containing protein	No function prediction

VP161 locus tag, (PM70 locus tag)[a.b.c.]	Gene name	Log_2 fold-change, (FDR)[d.]	Predicted product	General function prediction
PMVP_0469, (PM0497)	PM0497	-1.79, (0.001)	DUF1436 domain containing protein	No function prediction
PMVP_0470, (PM0498)	PM0498	-2.82, (0.0025)	Hypothetical protein PM0498	No function prediction
PMVP_0471, (PM0500)	PM0500	-1.45, (0.0024)	SMI1/KNR4 family protein	No function prediction
PMVP_0473, (NP)		-1.17, (0.0031)	Hypothetical	No function prediction
PMVP_0551, (PM0576)	*hemR*	-1.14, (0.0011)	TonB dependent haemoglobin family receptor	Inorganic ion transport and metabolism genes
PMVP_0567, (NA)		-2.24, (0.0144)	Hypothetical	No function prediction
PMVP_0585, (PM0613)	PM0613	-1.37, (0.0046)	Phage holin family protein	No function prediction
PMVP_0586, (PM0614)	PM0614	-1.39, (0.0082)	Hypothetical protein PM0614	No function prediction
PMVP_0642, (PM0670)	PM0670	-1.09, (0.0007)	Soluble cytochrome b56, putative	Energy production and conversion genes
PMVP_0672, (PM0697)	PM0697	-1.02, (0.0192)	DUF4391 domain containing protein	No function prediction
PMVP_0693, (PM0714)	*hsf_1*	-1.11, (0.036)	Adhesin	Intracellular trafficking and secretion genes
PMVP_0756, (NP)		-1.48, (0.0007)	MoxR-like ATPases	No function prediction
PMVP_0757, (NP)		-2.76, (0.0006)	VWA domain containing protein	No function prediction
PMVP_0785, (NA)	*gntR*	-2.06, (0.0019)	LacI family transcriptional regulator	Transcription genes
PMVP_0789, (NP)		-1.84, (0.0015)	TRAP transporter small permease	No function prediction
PMVP_0813, (NP)	*petG* (pseudogene)	-1.91, (0.0008)	Phosphoethanolamine transferase	No function prediction
PMVP_0829, (PM0830)	PM0830	-1.41, (0.0023)	DUF896 domain containing protein	No function prediction
PMVP_0841, (PM0842)	PM0842	-1.27, (0.0089)	Conserved hypothetical protein	No function prediction
PMVP_0846, (PM0847)	*tadC*	-1.58, (0.0016)	Type II secretion system F family protein	Cell motility genes
PMVP_0849, (PM0850)	*tadZ*	-1.64, (0.0032)	Pilus assembly protein	Intracellular trafficking and secretion genes
PMVP_0854, (PM0855)	*flp1*	-1.93, (0.0019)	Flp family type Ivb pilin	Intracellular trafficking and secretion genes

VP161 locus tag, (PM70 locus tag)[a.b.c.]	Gene name	Log_2 fold-change, (FDR)[d.]	Predicted product	General function prediction
PMVP_0875, (PM0873)	*purU*	-1.05, (0.0056)	Formyltetrahydrofolate deformylase	Nucleotide transport and metabolism genes
PMVP_0882, (PM0880)	PM0880	-1.72, (0.0092)	Membrane protein	No function prediction
PMVP_0909, (PM0906)	*hfq*	-1.94, (0.0019)	RNA-binding protein Hfq	No function prediction
PMVP_0921, (NP)	*alsK*	-1.02, (0.0008)	Allulose kinase	Transcription genes
PMVP_0922, (NP)	*alsE*	-1.62, (0.002)	D-allulose ribulose-phosphate 3-epimerase	Carbohydrate transport and metabolism genes
PMVP_0935, (NA)	scRNA	-2.35, (0.0007)	Enterocidin A/B family lipoprotein	No function prediction
PMVP_0997, (PM0984)	PM0984	-1.12, (0.0052)	DUF1295 domain containing protein	No function prediction
PMVP_1018, (PM1004)	PM1004	-1.23, (0.0418)	Hypothetical protein PM1004	No function prediction
PMVP_1079, (PM1060)	PM1060	-1.18, (0.0018)	Ycek/YidQ family lipoprotein	No function prediction
PMVP_1115, (PM1095)	PM1095	-1.12, (0.0021)	Hypothetical protein PM1095	No function prediction
PMVP_1122, (PM1101)	PM1101	-1.09, (0.0065)	Hypothetical protein PM1101	No function prediction
PMVP_1125, (NA)		-1.41, (0.001)	Hypothetical	No function prediction
PMVP_1159, (NP)	*pcgC*	-1.13, (0.0263)	CTP: phosphocholine cytidylyltransferase	Cell wall/membrane biogenesis genes
PMVP_1192, (NA)		-1.73, (0.0179)	Hypothetical	No function prediction
PMVP_1205, (PM1187)	*exbD*	-1.78, (0.0032)	TonB system transport protein	Intracellular trafficking and secretion genes
PMVP_1218, (NP)		-1.05, (0.0147)	Histidine phosphate family protein	No function prediction
PMVP_1222, (NP)		-1.55, (0.0014)	HNH endonuclease	No function prediction
PMVP_1226, (NP)		-1.38, (0.0082)	Hypothetical	No function prediction
PMVP_1258, (NP)		-1.17, (0.0031)	Hypothetical	No function prediction
PMVP_1372, (NA)		-2.11, (0.0088)	Hypothetical	No function prediction
PMVP_1424, (NA)		-1.07, (0.0132)	Hypothetical	No function prediction
PMVP_1447, (NP)	*araC*	-1.31, (0.002)	Arabinose operon transcriptional regulator AraC	Amino acid transport and metabolism genes

VP161 locus tag, (PM70 locus tag)[a.b.c.]	Gene name	Log_2 fold-change, (FDR)[d.]	Predicted product	General function prediction
PMVP_1449, (NP)		-1.08, (0.0007)	Sugar ABC transporter ATP-binding protein	No function prediction
PMVP_1451, (NP)		-1.36, (0.004)	ABC transporter permease	No function prediction
PMVP_1475, (NA)		-1.00, (0.0241)	Hypothetical	No function prediction
PMVP_1530, (PM1478)	PM1478	-1.02, (0.0099)	Hypothetical protein PM1478	No function prediction
PMVP_1533, (PM1481)	PM1481	-1.61, (0.0007)	Bacterial regulatory protein, arsR family, putative	Transcription genes
PMVP_1539, (PM1487)	PM1487	-1.60, (0.0058)	Hypothetical	No function prediction
PMVP_1553, (NA)		-1.32, (0.0055)	Unique hypothetical protein	No function prediction
PMVP_1554, (NA)		-1.84, (0.0105)	Hypothetical	No function prediction
PMVP_1570, (NP)		-1.72, (0.0012)	Addiction molecule antidote protein, HigA family	No function prediction
PMVP_1595, (PM1543)	PM1543	-1.02, (0.0015)	Hypothetical protein PM1543	No function prediction
PMVP_1631, (NA)		-1.22, (0.0162)	Hypothetical	No function prediction
PMVP_1658, (PM1603)	PM1603	-1.28, (0.001)	Conserved hypothetical protein	No function prediction
PMVP_1697, (PM1647)	*iolE*	-1.63, (0.0032)	Sugar phosphate isomerase/epimerase	No function prediction
PMVP_1704, (PM1654)	PM1654	-1.33, (0.0159)	Hypothetical protein PM1654	No function prediction
PMVP_1709, (PM1658)	*ahpA*	-1.36, (0.001)	Hemolysin regulation protein AhpA	No function prediction
PMVP_1739, (NA)		-1.12, (0.0208)	Hypothetical	No function prediction
PMVP_1832, (NP)		-1.15, (0.0019)	Hypothetical	No function prediction
PMVP_1987, (PM1934)	PM1934	-1.16, (0.0195)	Hypothetical	No function prediction
PMVP_2045, (PM2008)	*pilW*	-1.07, (0.0168)	Type IV fimbrial biogenesis protein	Cell motility genes
Prrc05, (NA)	Prrc05	-1.26, (0.0096)	Prrc05 putative sRNA	sRNA

VP161 locus tag, (PM70 locus tag)[a.b.c.]	Gene name	Log_2 fold-change, (FDR)[d.]	Predicted product	General function prediction
Prrc06, (NA)	Prrc06	-2.25, (0.0019)	Prrc06 putative sRNA	sRNA
Prrc09, (NA)	Prrc09	-1.09, (0.0173)	hypothetical (see Table 1.1)	No function prediction
Prrc10, (NA)	Prrc10	-2.45, (0.0141)	Prrc10 putative sRNA	sRNA
Prrc12, (NA)	Prrc12	-1.43, (0.0008)	Prrc12 putative sRNA	sRNA
Prrc15, (NA)	Prrc15	-1.61, (0.0033)	Prrc15 putative sRNA	sRNA
Prrc19, (NA)	Prrc19	-2.24, (0.001)	Prrc19 putative sRNA	sRNA
Prrc20, (NA)	Prrc20	-1.40, (0.001)	Prrc20 putative sRNA	sRNA
Prrc24, (NA)	Prrc24	-1.19, (0.002)	Prrc24 putative sRNA	sRNA
Prrc30, (NA)	Prrc30	-2.60, (0.0022)	Prrc30 putative sRNA	sRNA
Prrc32, (NA)	Prrc32	-1.43, (0.0469)	Prrc32 putative sRNA	sRNA
Prrc37, (NA)	Prrc37	-1.01, (0.001)	Hypothetical (see Table 1.1)	No function prediction
Prrc42, (NA)	Prrc42	-3.15, (0.0001)	Prrc42 putative sRNA	sRNA
Prrc50, (NA)	Prrc50	-1.58, (0.0012)	Prrc50 putative sRNA	sRNA

[a] Differentially expressed transcripts were defined as those showing at least 2-fold change in production ($log_2 \leq -1$) with an FDR of less than 0.05.

[b.] NA= Not annotated but sequence present, NP= No sequence present

[c.] All predicted sRNAs can be found in Table 1.1

[d.] Transcript expression ratio is shown as a log_2 value with the corresponding false discovery rate (FDR) shown in brackets.

Table 3.6. Transcripts with increased expression in the *P. multocida proQ* mutant (AL2973) compared to the VP161 wild-type parent strain.

VP161 locus tag, (PM70 locus tag)[a.b.c.]	Gene name	Log_2 fold-change, (FDR)[d.]	Predicted product	General function prediction
PMVP_0260, (PM_t45)	tRNA Ala	2.22, (0.002)	tRNA Ala	Translation genes
PMVP_0439, (PM_t10)	tRNA His	1.12, (0.007)	tRNA His	Translation genes
PMVP_0626, (PM0654)	tRNA Val	1.25, (0.006)	tRNA Val	Translation genes
PMVP_0628, (PM0655)	*msmB*	1.42, (0.002)	Cold shock-like protein CspC-related protein	Transcription genes
PMVP_0657, (PM_t18)	tRNA Met	1.04, (0.007)	tRNA Met	Translation genes
PMVP_0773, (PM0780)	*hexB*	1.00, (0.014)	HexB; hyaluronic acid capsule transport	Cell wall/membrane biogenesis genes
PMVP_0961, (PM_t26)	tRNA Ser	5.72, (0.025)	tRNA Ser	Translation genes
PMVP_1009, (PM_t33)	tRNA Leu	1.53, (0.001)	tRNA Leu	Translation genes
PMVP_1065, (PM_t32)	tRNA Cys	1.06, (0.001)	tRNA Cys	Translation genes
PMVP_1077, (PM_t36)	tRNA Ser	1.73, (0.002)	tRNA Ser	Translation genes
PMVP_1137, (NA)	tRNA	1.57, (0.014)	tRNA	Translation genes
PMVP_1286, (PM1262)	*thiM*	1.14, (0.006)	Hydroxyethylthiazole kinase	Coenzyme transport and metabolism genes
PMVP_1287, (PM1263)	PM1263	1.01, (0.004)	Thiamine biosynthesis protein, putative	Inorganic ion transport and metabolism genes
PMVP_1322, (PM1298)	rpL19	1.06, (0.001)	50S ribosomal protein L19	Translation genes
PMVP_1349, (PM_t42)	tRNA Phe	2.69, (0.001)	tRNA Phe	Translation genes
PMVP_1608, (PM1556)	*comF*	1.41, (0.001)	Competence protein F	No function prediction
PMVP_1647, (PM1592)	*napF*	1.21, (0.001)	Ferredoxin-type protein NapF	Energy production and conversion genes
PMVP_1650, (PM1595)	*napG*	1.02, (0.001)	Quinol dehydrogenase periplasmic component	Energy production and conversion genes
PMVP_1782, (PM_t48)	tRNA Trp	2.11, (0.004)	tRNA Trp	Translation genes
PMVP_1805, (PM_t52)	tRNA Gly	2.74, (0.003)	tRNA Gly	Translation genes

PMVP_1807, (PM_t54)	tRNA Thr	1.48, (0.007)	tRNA Thr	Translation genes
PMVP_1831, (PM_t55)	tRNA SeC	3.12, (0.01)	tRNA SeC	Translation genes
PMVP_1964, (PM1912)	rpL32	1.59, (0.001)	50S ribosomal protein L32	Translation genes
PMVP_1984, (PM1932)	tRNA Arg	1.40, (0.002)	tRNA Arg	Translation genes
PMVP_2500, (NA)		1.03, (0.008)	Putative glycine riboswitch	No function prediction
PMVP_2515, (NA)		1.23, (0.009)	Putative lysine riboswitch	No function prediction
PMVP_2522, (NA)		2.76, (0.0001)	Putative 6S RNA represses sigma 70 transcripts in stationary phase	No function prediction
PMVP_2523, (NA)		6.02, (0.001)	Putative sRNA involved in protein secretion	No function prediction
Prrc11, (NA)	Prrc11	1.35, (0.004)	Prrc11	sRNA
Prrc25, (NA)	Prrc25	1.54, (0.008)	Prrc25	sRNA
Prrc40, (NA)	Prrc40	1.36, (0.046)	Prrc40	sRNA
Prrc44, (NA)	Prrc44	1.68, (0.025)	Prrc44	sRNA
tRNA Val, (PM_t41)	tRNA Val	1.41, (0.001)	tRNA Val	Translation genes
tRNA Ser, (PM_t30)	tRAN Ser	2.05, (0.023)	tRNA Ser	Translation genes
tRNA Gly, (PM_t31)	tRNA Gly	2.35, (0.001)	tRNA Gly	Translation genes

[a] Differentially expressed transcripts were defined as those showing at least 2-fold change in production ($\log_2 \geq 1$) with an FDR of less than 0.05.

[b.] NA= Not annotated but sequence present, NP= No sequence present

[c.] All predicted sRNAs can be found in Table 1.1

[d.] Transcript expression ratio is shown as a $\log_2$ value with the corresponding false discovery rate (FDR) shown in brackets.

Several transcripts identified as having decreased expression were involved in cell wall biogenesis, including, *prc*, *ycbB*, *pglA*, PM1289, as well as two involved in the post-assembly modification of LPS; the Phosphocholine (PCho) biosynthesis gene *pcgC* and the lipid A-specific phosphoethanolamine (PEtn) transferase gene *petL* (Table 3.6). To determine if decreased expression of *pcgC* and *petL* altered the LPS produced by the *P. multocida proQ* mutant (AL2973), the *proQ* mutant was compared with the wild-type parent strain VP161 for amount and length of LPS produced, for the addition of PCho to the LPS, and for susceptibility to the host antimicrobial peptide Fowlicidin-1, whose activity is known to be affected by the presence of PEtn and PCho (Harper et al. 2007b). Firstly, overall LPS production was compared in the wild-type VP161, *proQ* mutant (AL2973), overexpression strain (AL2978) and *proQ* mutant containing empty vector (AL2994) using a carbohydrate silver stain (Figure 3.3B). No differences were observed between the wild-type and *proQ* mutants. PCho production was then analysed by Western blotting using the TEPC-15 antibody which binds directly to PCho residues (Figure 3.3C). As expected there was a complete loss of PCho in the *pcgC* mutant, AL571, used as a control (Harper et al. 2007b), and production was restored in the *pcgC* complemented strain (AL829). However, there was no difference observed between the wild-type and *proQ* mutants. To assess susceptibility to Fowlicidin-1, the wild-type VP161, the *proQ* mutant and, as a control, the fowlicidin sensitive VP161 *petL* mutant (AL1354), were grown in biological triplicate to OD_{600} = 0.4 and then incubated for 3 h at 37°C in media containing Fowlicidin-1 at varying concentrations (0, 0.625, 1.25 and 2.5 μg/ml). Pre- and post-treatment viable counts were compared to determine the percentage of survivors (Figure 3.3D). The *proQ* mutant was recovered at an equal level to the wild-type VP161 parent strain at all Fowlicidin-1 concentrations tested, whilst the control strain (VP161 *petL* mutant) showed decreased survival at 2.5 μg/mL of Fowlicidin-1 as has been observed previously (Harper et al. 2017).

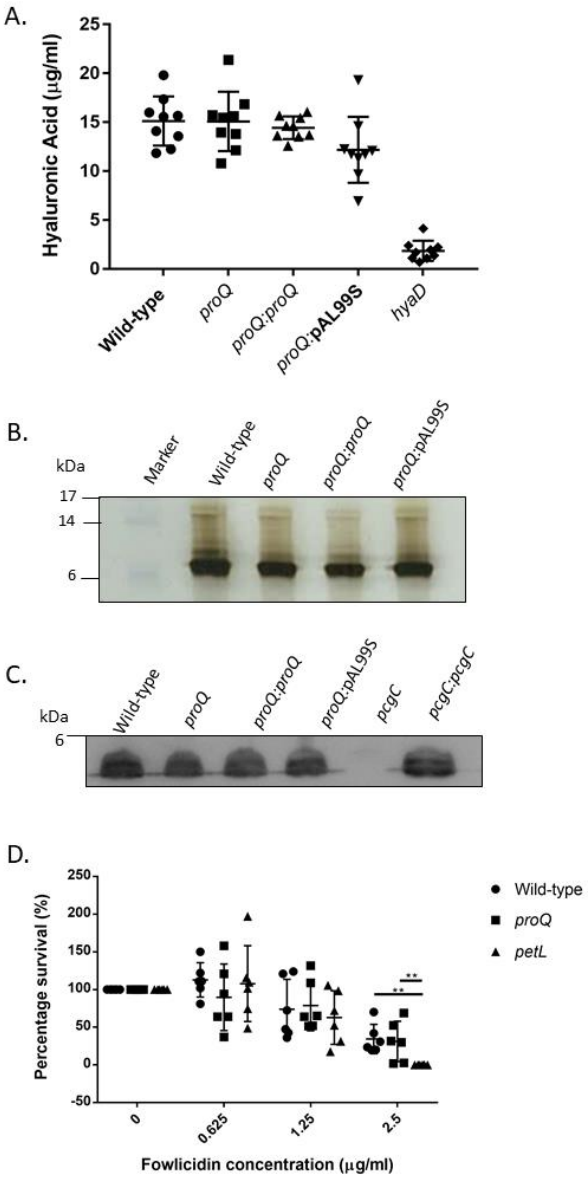

Figure 3.3. Capsule and LPS production in the *P. multocida* wild-type strain and *proQ* mutant and complemented strains. A. Amount of hyaluronic acid capsule (µg/ml) produced in the *P. multocida* wild-type VP161 (Wild-type) (circles), *proQ* mutant (AL2973) (*proQ*) (squares), *proQ* overexpression strain (AL2978) (*proQ: proQ*) (upright triangles), *proQ* mutant containing empty vector (AL2994) (*proQ:* pAL99S) (upside-down triangles), and acapsular *hyaD* mutant (AL2234) (*hyaD*)(diamonds). Data shown are mean ± SD, with n=9. **B.** Carbohydrate silver-stained PAGE of whole cell lysates produced by the *P. multocida* wild-type VP161 (wild-type), *proQ* mutant (*proQ*)(AL2973), *proQ* overexpression strain (*proQ: proQ*) (AL2978) and *proQ* mutant containing empty vector (*proQ*: pAL99S) (AL2994). **C.** Immunoblot detection of phosphocholine (using TEPC-15 antibody) in the *P. multocida* wild-type VP161 (wild-type), *proQ* mutant (*proQ*) (AL2973), *proQ* overexpression strain (*proQ: proQ*) (AL2978), *proQ* mutant containing empty vector (*proQ*: pAL99S) (AL2994), PCho deficient *pcgC* mutant (*pcgC*) (AL571) and the complemented *pcgC* strain (*pcgC: pcgC*) (AL829). **D.** Percentage survival of *P. multocida* wild-type VP161 (circles), *proQ* mutant (squares) (AL2973), and fowlicidin sensitive *petL* mutant (triangles) (AL1354) after treatment with 0, 0.625, 1.25, or 2.5 µg/ml of fowlicidin for 3 h at 37°C. Data shown are mean ± SD, with n=6, **= $p \leq 0.005$ using a Student's t-test.

3.3.4 Transcriptomic analysis of a *P. multocida* complemented *proQ* strain

In order to determine if the transcriptional changes observed in the *proQ* mutant (AL2973) were specifically due to the loss of ProQ alone, a second analysis was performed that included the *proQ* mutant provided with the plasmid pAL1449, encoding an intact copy of *proQ* on the plasmid pPBA1100s under the control of its native promoter, to generate the strain AL3357. As controls, the *proQ* mutant, AL2973, and the wild-type *P. multocida* parent strain VP161 were provided with empty vector to produce strains AL3358 and AL3356, respectively. The three strains were grown in biological triplicate until mid-exponential growth phase (OD_{600} = 0.6). Cells were harvested, RNA was extracted and processed into a strand-specific cDNA library and sequenced. Genes were determined to be differentially expressed if they showed a ≥ 2-fold change in expression (≥ 1 $\log_2$) compared to the control, with an FDR < 0.05 (Table 3.7).

A comparison of the transcriptome of the *proQ* mutant and wild-type strain, both with empty vector (AL3358 and AL3356, respectively), revealed there were 31 differentially expressed transcripts. Four transcripts showed increased expression and 27 showed decreased expression in the *proQ* mutant with empty vector compared to the parent strain with empty vector. This differential expression pattern was reversed (FDR < 0.05 for AL3358 vs AL3357, Table 3.7) for 24 of the 31 genes when the *proQ* mutant was provided with a functional copy of *proQ* (strain AL3357) indicating the change in the expression for these genes was specifically associated with ProQ abundance. Comparison of the *proQ* mutant complemented with an intact copy of *proQ* (AL3357) with the wild-type parent strain containing empty vector (AL3356) revealed 91 differentially expressed genes; six had decreased expression in the complemented *proQ* mutant and 85 had increased expression, indicating that providing *proQ* on a plasmid may have resulted in the overabundance of ProQ in strain AL3356 (Table 3.7). This was supported by the finding that 64 (70%) of the 91 differentially expressed genes identified in the above comparison showed a reversed pattern of expression when the transcript data for this strain was directly compared to the transcript data derived from the *proQ* mutant containing empty vector (FDR < 0.05 for AL3357 vs AL3358, Table 3.7).

In total, there were 126 transcripts that showed differential expression between the *proQ* mutant containing empty vector (AL3358) and *proQ* mutant provided with a functional copy of *proQ* (AL3357); 113 had increased expression in the complemented *proQ* strain and 13 had decreased expression. Of the differentially expressed genes/transcripts identified in the above comparisons, only the gene *prc* (PMVP_0228; located immediately downstream of *proQ*) was differentially expressed in all strains with altered levels of ProQ. The *prc* transcript showed a 2.44-fold reduced expression ($\log_2$ fold-change of -1.29, Table 3.7) in the *proQ* mutant with empty vector (AL3358) when compared to the expression in wild-type containing empty vector (AL3356). In contrast, the *prc* transcript had 21.6-fold increased expression

(log_2 fold-change of 4.43, Table 3.7) when the *proQ* mutant was provided with an intact copy of *proQ* (AL3357) compared to the level of transcript in the wild-type. There were 79 transcripts found to be differentially expressed in any two of the comparison groups, however three transcripts showed an expression change in the same direction for both the *proQ* mutant and complemented strains when compared to wild-type, indicating that the change in expression was not due to the amount of ProQ. The list of differentially expressed genes was then compared to those identified in the initial round of RNA-seq (section 3.3.3 above) and it was found that 15 transcripts were shared between the two data sets. These included; PMVP_0063, PMVP_0164, *prc* (PMVP_0228), *ycbB* (PMVP_0229), *mliC* (PMVP_0253), tRNA His (PMVP_0439), *hsf_1* (PMVP_0693), *hfq* (PMVP_0909), PMVP_0997, PMVP_1115, PMVP_1704, PMVP_1987, Prrc09, Prrc12 and Prrc32, (Table 3.5-7).

The complete list of 166 differentially expressed genes was generated by collating data from the following comparisons; *proQ* mutant vs wild-type parent (AL3358 vs AL3356), ProQ complementation strain vs wild-type (AL3357 vs AL3356) and *proQ* mutant vs ProQ complementation strain (AL 3358 vs AL3357). Using this data, overrepresented cellular pathways were identified. These included amino acid metabolism (16 differentially expressed genes), carbohydrate metabolism and transport (18 genes), and inorganic ion transport or metabolism (11 genes) (Table 3.7). Also identified as overrepresented were eight putative sRNAs (Prrc12, Prrc13, Prrc32, Prrc25, Prrc08, Prrc17, Prrc46 and Prrc49) (Table 3.7, Table 1.1), and 19 tRNAs (tRNA His (PMVP_0439), tRNA Val (PMVP_0626), tRNA Leu (PMVP_1009), tRNA Cys (PMVP_1065), tRNA Ser (PMVP_1077), tRNA Phe (PMVP_1349), tRNA Trp (PMVP_1782), tRNA Gly (PMVP_1805), tRNA Thr (PMVP_1807), tRNA SeC (PMVP_1831), tRNA Arg (PMVP_1984), tRNA Val (PM_t41), tRNA Ser (PM_t30), tRNA Gly (PM_t31), tRNA Leu (PMVP_0262), tRNA Arg (PMVP_0438), tRNA Met (PMVP_0657), tRNA Tyr (PMVP_1806), and tRNA Met) (Table 3.7). Furthermore, ABC transporters were significantly over-represented (FDR = 0.0016) with 20 identified, namely, *artP, artI, artQ, artM, potD_1, metE, hbpA, rbsD, mglB, mglA, mglC*, *lrsA*, *rbsB_2*, *rbsA_2*, *fecC, fecB, pstA, pstS*, PM1266, and *rbsC* (Table 3.7). In summary, these data show that changes in ProQ abundance in *P. multocida* results in the differential expression of a variety of transcripts, particularly those representing sRNAs, tRNAs and ABC transporter proteins.

Table 3.7. Transcripts with differential expression in the *P. multocida proQ* mutant containing empty vector (AL3358) or complemented *proQ* strain (AL3357) compared to the VP161 wild-type parent strain with empty vector (AL3356).

VP161 locus tag, (PM70 locus tag) [a.b.c.]	**Gene name**	**AL3358 log_2 fold-change, (FDR)** [d.e.]	**AL3357 log_2 fold-change, (FDR)** [d. e]	**AL3358 v AL3357 log_2 fold-change, (FDR)** [d. e]	**Predicted product**	**General function prediction**
PMVP_0011, (PM0064)	*grcA*	1.33, (0.172)	-0.23, (0.785)	1.56, (0.03291)	Autonomous glycyl radical cofactor GrcA	No function prediction
PMVP_0060, (PM0108)	PM0108	0.25, (0.497)	1.03, (0.001)	-0.78, (0.00391)	DUF72 domain containing protein	No function prediction
PMVP_0063, (NP)		0.11, (0.771)	1.99, (0.00004)	-1.87, (0.00004)	Hypothetical	No function prediction
PMVP_0076, (PM0123)	*artP*	-1.42, (0.038)	0.09, (0.883)	-1.52, (0.00607)	Arginine ABC transporter, ATP-binding protein	Amino acid transport and metabolism genes
PMVP_0077, (PM0124)	*artI*	-1.4, (0.046)	0.02, (0.979)	-1.42, (0.00991)	Arginine ABC transporter, periplasmic-binding protein	Amino acid transport and metabolism genes
PMVP_0078, (PM0125)	*artQ*	-1.39, (0.063)	-0.14, (0.835)	-1.25, (0.02793)	Arginine ABC transporter, permease protein	Amino acid transport and metabolism genes
PMVP_0079, (PM0126)	*artM*	-1.40, (0.095)	-0.14, (0.852)	-1.26, (0.04239)	Arginine ABC transporter, permease protein	Amino acid transport and metabolism genes
PMVP_0083, (PM0130)	*fecC*	1.17, (0.022)	0.64, (0.105)	0.54, (0.10802)	Iron(iii) dicitrate transport system permease protein FecC	Inorganic ion transport and metabolism genes
PMVP_0084, (PM0131)	*fecB*	1.07, (0.015)	0.50, (0.112)	0.57, (0.04858)	Iron-dicitrate transporter substrate-binding subunit	Inorganic ion transport and metabolism genes
PMVP_0110, (PM0156)	*rbsD*	-0.70, (0.304)	-1.51, (0.016)	0.81, (0.14439)	D-ribose pyranase	Carbohydrate transport and metabolism genes
PMVP_0164, (PM0209)	PM0209	0.05, (0.921)	1.43, (0.0001)	-1.38, (0.00006)	Sigma 70 family RNA polymerase sigma factor	Transcription genes
PMVP_0180, (PM0223)	*dcaA*	-0.27, (0.621)	1.22, (0.004)	-1.49, (0.00066)	Phosphoethanolamine transferase	Cell wall/membrane biogenesis genes
PMVP_0201, (PM0243)	*mepM*	-0.52, (0.485)	1.70, (0.011)	-2.22, (0.0016)	Peptidase family M23/M37 domain protein	Cell wall/membrane biogenesis genes
PMVP_0218, (PM0260)	*potD_1*	-1.95, (0.002)	0.03, (0.945)	-1.98, (0.0001)	Spermidine/putrescine ABC transporter, periplasmic-binding protein	Amino acid transport and metabolism genes
PMVP_0227, (PM0268)	*proQ*	0.84, (0.009)	6.62, (0.000000004)	-5.78, (0.00000001)	Putative solute/DNA competence effector	Signal transduction mechanisms genes

VP161 locus tag, (PM70 locus tag) [a.b.c.]	Gene name	AL3358 log_2 fold-change, (FDR) [d.e.]	AL3357 log_2 fold-change, (FDR) [d. e]	AL3358 v AL3357 log_2 fold-change, (FDR) [d. e]	Predicted product	General function prediction
PMVP_0228, (PM0269)	*prc*	-1.29, (0.002)	4.43, (0.0000002)	-5.72, (0.00000001)	Carboxy-terminal protease	Cell wall/membrane biogenesis genes
PMVP_0229, (PM0270)	*ycbB*	-1.34, (0.014)	-1.18, (0.006)	-0.16, (0.66122)	Murein L, D-transpeptidase	Cell wall/membrane biogenesis genes
PMVP_0232, (PM0272)	PM0272	-0.24, (0.699)	1.16, (0.008)	-1.40, (0.00202)	Metallo-beta-lactamase superfamily metallohydrolase	No function prediction
PMVP_0233, (PM0273)	PM0273	-1.97, (0.03)	-0.66, (0.308)	-1.31, (0.04744)	TRAP dicarboxylate transporter- DctM subunit	Carbohydrate transport and metabolism genes
PMVP_0253, (NA)	*mliC*	-0.70, (0.009)	0.99, (0.0003)	-1.68, (0.000002)	Lysozyme inhibitor	No function prediction
PMVP_0262, (PM_t06)	tRNA Leu	-0.62, (0.651)	1.65, (0.049)	-2.27, (0.01042)	tRNA Leu	Translation genes
PMVP_0274, (PM0307)	*csy3*	0.92, (0.136)	-0.38, (0.489)	1.31, (0.01038)	Type I-F CRISPR-associated protein Csy3	No function prediction
PMVP_0292, (PM0326)	*mltB*	-0.25, (0.542)	0.75, (0.016)	-1.00, (0.0023)	Transglycosylase SLT domain protein	Cell wall/membrane biogenesis genes
PMVP_0303, (PM0336)	PM0336	0.79, (0.118)	1.03, (0.014)	-0.24, (0.491)	TonB dependent receptor	Inorganic ion transport and metabolism genes
PMVP_0313, (PM0344)	*argD*	-1.29, (0.049)	0.14, (0.796)	-1.43, (0.00626)	Bifunctional N-succinyldiaminopimelate-aminotransferase/acetylornithine transaminase protein	Amino acid transport and metabolism genes
PMVP_0374, (PM0404)	PM0404	0.06, (0.92)	1.34, (0.0004)	-1.28, (0.00028)	Membrane protein	No function prediction
PMVP_0389, (PM0417)	PM0417	1.00, (0.136)	1.26, (0.022)	-0.26, (0.56398)	Helix-turn-helix transcriptional regulatory protein	Transcription genes
PMVP_0391, (PM0419)	PM0419	0.05, (0.926)	1.65, (0.0001)	-1.59, (0.00006)	Metal dependent hydrolase	No function prediction
PMVP_0392, (PM0420)	*metE*	-0.15, (0.656)	1.04, (0.001)	-1.19, (0.00012)	5-methyltetrahydropteroyltriglutamate --homocysteine methyltransferase	Amino acid transport and metabolism genes
PMVP_0401, (PM0428)	PM0428	-0.22, (0.463)	1.23, (0.0001)	-1.45, (0.00002)	K+-dependent Na+/Ca+ exchanger related-protein	Inorganic ion transport and metabolism genes
PMVP_0407, (PM0434)	*pstA*	0.10, (0.909)	1.01, (0.015)	-0.91, (0.01658)	Phosphate ABC transporter, permease protein	Inorganic ion transport and metabolism genes

VP161 locus tag, (PM70 locus tag) [a.b.c.]	Gene name	AL3358 log_2 fold-change, (FDR) [d.e.]	AL3357 log_2 fold-change, (FDR) [d. e]	AL3358 v AL3357 log_2 fold-change, (FDR) [d. e]	Predicted product	General function prediction
PMVP_0409, (PM0436)	*pstS*	0.56, (0.231)	1.18, (0.004)	-0.62, (0.04228)	Phosphate-binding periplasmic protein precursor	Inorganic ion transport and metabolism genes
PMVP_0422, (PM0449)	PM0449	0.43, (0.228)	1.00, (0.002)	-0.57, (0.02598)	Transporter permease	Defense mechanisms genes
PMVP_0438, (PM_t09)	tRNA Arg	-0.20, (0.867)	1.25, (0.048)	-1.45, (0.0191)	tRNA Arg	Translation genes
PMVP_0439, (PM_t10)	tRNA His	0.12, (0.891)	1.10, (0.024)	-0.98, (0.03)	tRNA His	Translation genes
PMVP_0461, (PM0485)	PM0485	0.31, (0.301)	1.09, (0.0004)	-0.77, (0.00191)	Cys-tRNA (Pro)deacylase	Translation genes
PMVP_0507, (NP)	*torC*	1.67, (0.231)	-0.46, (0.728)	2.13, (0.04228)	Cytochrome c-type protein TorC	Energy production and conversion genes
PMVP_0519, (NP)		-0.71, (0.056)	2.13, (0.00003)	-2.84, (0.000001)	ABC transporter substrate binding protein	Inorganic ion transport and metabolism genes
PMVP_0566, (PM0592)	*hbpA*	0.72, (0.239)	-0.99, (0.021)	1.71, (0.00113)	Heme-binding lipoprotein	Amino acid transport and metabolism genes
PMVP_0574, (NA)		0.29, (0.637)	1.68, (0.001)	-1.39, (0.00228)	Autotransporter adhesin	No function prediction
PMVP_0581, (PM0610)	PM0610	0.21, (0.651)	1.00, (0.006)	-0.79, (0.01424)	AmiS/UreI family transporter family	No function prediction
PMVP_0594, (PM0622)	*moaE*	0.64, (0.199)	1.51, (0.001)	-0.87, (0.01429)	Molybdopterin converting factor, subunit 2	Coenzyme transport and metabolism genes
PMVP_0598, (PM0626)	*yvcK*	0.07, (0.877)	1.03, (0.001)	-0.96, (0.00157)	Uridine diphosphate N-acetylglucosamine-binding protein	No function prediction
PMVP_0626, (PM0654)	tRNA Val	-0.29, (0.834)	1.86, (0.016)	-2.15, (0.00554)	tRNA Val	Translation genes
PMVP_0627, (NA)		0.05, (0.931)	3.09, (0.0000001)	-3.04, (0.00000004)	Unique hypothetical protein	No function prediction
PMVP_0645, (NA)		-0.004, (0.992)	1.02, (0.006)	-1.03, (0.00321)	Unique hypothetical protein	No function prediction
PMVP_0647, (PM0674)	PM0674	0.10, (0.702)	1.12, (0.00005)	-1.02, (0.00006)	Hypothetical protein PM0674	No function prediction
PMVP_0654, (PM0681)	PM0681	-0.33, (0.394)	1.10, (0.001)	-1.43, (0.00013)	Divalent metal cation transporter	Inorganic ion transport and metabolism genes

VP161 locus tag, (PM70 locus tag) [a.b.c.]	Gene name	AL3358 log_2 fold-change, (FDR) [d.e.]	AL3357 log_2 fold-change, (FDR) [d. e]	AL3358 v AL3357 log_2 fold-change, (FDR) [d. e]	Predicted product	General function prediction
PMVP_0657, (PM_t18)	tRNA Met	0.67, (0.206)	1.08, (0.013)	-0.41, (0.22749)	tRNA Met	Translation genes
PMVP_0664, (PM0689)	PM0689	-0.03, (0.951)	1.07, (0.001)	-1.10, (0.00073)	CPBP family intermembrane metalprotease	No function prediction
PMVP_0693, (PM0714)	*hsf_1*	0.46, (0.248)	1.06, (0.003)	-0.60, (0.02824)	Adhesin	Intracellular trafficking and secretion genes
PMVP_0694, (PM0715)	*greA*	-0.08, (0.864)	1.01, (0.001)	-1.09, (0.00043)	Transcription elongation factor GreA	Transcription genes
PMVP_0700, (PM0721)	*ttrA*	0.61, (0.284)	1.07, (0.021)	-0.45, (0.24161)	Tetrathionate reductase subunit A	Energy production and conversion genes
PMVP_0713, (PM0734)	*htrA*	-0.58, (0.122)	1.36, (0.001)	-1.94, (0.00006)	Do family serine endopeptidase	Posttranslational modification, protein turnover, chaperones genes
PMVP_0725, (PM0745)	PM0745	0.29, (0.233)	1.08, (0.0001)	-0.78, (0.00043)	TonB-dependent receptor	Inorganic ion transport and metabolism genes
PMVP_0744, (PM0763)	PM0763	-0.27, (0.517)	1.53, (0.0001)	-1.80, (0.00002)	Putative L-ascorbate 6-phosphate lactonase	No function prediction
PMVP_0786, (NP)	*idnD*	0.26, (0.571)	1.56, (0.0002)	-1.30, (0.00038)	L-idonate-5-dehydrogensase	No function prediction
PMVP_0787, (NP)	*ydfG*	-0.15, (0.864)	0.91, (0.048)	-1.06, (0.01817)	NADP-dependent 3-hydroxy acid dehydrogenase YdfG	Energy production and conversion
PMVP_0800, (PM0803)	PM0803	1.21, (0.034)	0.45, (0.308)	0.76, (0.06022)	TonB dependent receptor C-terminal region subfamily	No function prediction
PMVP_0805, (PM0809)	PM0809	0.42, (0.258)	1.09, (0.002)	-0.67, (0.01562)	DUF535 domain containing protein	No function prediction
PMVP_0878, (PM0876)	*ptsG*	-0.59, (0.132)	0.57, (0.058)	-1.16, (0.00117)	PTS permease for N-acetylglucosamine and glucose	Carbohydrate transport and metabolism genes
PMVP_0909, (PM0906)	*hfq*	-0.29, (0.2)	0.73, (0.001)	-1.02, (0.00009)	RNA-binding protein Hfq	No function prediction
PMVP_0938, (PM0927)	*tcdA*	-0.02, (0.97)	0.99, (0.011)	-1.01, (0.00625)	tRNA cyclic N6 threonyl-carbamoyl adenosine (37) synthase	Coenzyme transport and metabolism genes

VP161 locus tag, (PM70 locus tag) [a.b.c.]	Gene name	AL3358 log_2 fold-change, (FDR) [d.e.]	AL3357 log_2 fold-change, (FDR) [d. e]	AL3358 v AL3357 log_2 fold-change, (FDR) [d. e]	Predicted product	General function prediction
PMVP_0943, (PM0932)	PM0932	-0.09, (0.854)	0.94, (0.002)	-1.03, (0.00081)	M20 family peptidase	Amino acid transport and metabolism genes
PMVP_0958, (PM0946)	PM0946	-0.23, (0.411)	1.35, (0.00005)	-1.58, (0.00001)	Tellurite resistance TerB family protein	No function prediction
PMVP_0977, (PM0964)	PM0964	0.19, (0.846)	1.62, (0.006)	-1.43, (0.00589)	Conserved hypothetical protein	Intracellular trafficking and secretion genes
PMVP_0978, (PM0965)	*pulG*	0.26, (0.624)	1.76, (0.0003)	-1.50, (0.00039)	Type II secretion system pseudopilin PulG	Cell motility genes
PMVP_0995, (PM0982)	PM0982	0.52, (0.014)	-0.49, (0.006)	1.02, (0.00002)	DUF3298 domain containing protein	No function prediction
PMVP_0997, (PM0984)	PM0984	0.01, (0.989)	1.13, (0.007)	-1.12, (0.00508)	DUF1295 domain containing protein	No function prediction
PMVP_0998, (PM0985)	*cmoA*	0.01, (0.99)	1.75, (0.0003)	-1.74, (0.00018)	Carboxy-S-adenosyl-L-methionine synthase	Secondary metabolites biosynthesis, transport and catabolism genes
PMVP_1009, (PM_t33)	tRNA Leu	-0.45, (0.067)	1.08, (0.0001)	-1.53, (0.00001)	tRNA Leu	Translation genes
PMVP_1019, (PM1005)	PM1005	0.10, (0.926)	1.17, (0.021)	-1.08, (0.02107)	Hypothetical protein PM1005	No function prediction
PMVP_1023, (PM1009)	*wbjD*	0.56, (0.261)	1.10, (0.008)	-0.54, (0.0986)	UDP-N-acetylglucosamine 2-epimerase	Cell wall/membrane biogenesis genes
PMVP_1027, (PM1013)	PM1013	-0.004, (0.994)	1.01, (0.022)	-1.01, (0.01446)	N-acetyltransferase	Translation genes
PMVP_1031, (PM1017)	PM1017	0.24, (0.755)	1.11, (0.028)	-0.87, (0.04822)	Low molecular weight phosphotyrosine protein phosphatase	Signal transduction mechanisms genes
PMVP_1043, (PM1028)	*ndk*	-0.24, (0.199)	0.79, (0.0001)	-1.03, (0.00001)	Nucleoside diphosphate kinase	Nucleotide transport and metabolism genes
PMVP_1053, (PM1038)	*mglB*	-1.09, (0.141)	0.21, (0.735)	-1.30, (0.02263)	Galactose ABC transporter, periplasmic-binding protein	Carbohydrate transport and metabolism genes
PMVP_1054, (PM1039)	*mglA*	-1.26, (0.011)	-0.51, (0.122)	-0.74, (0.02968)	Galactose/methyl galaxtoside transporter ATP-binding protein	Carbohydrate transport and metabolism genes

VP161 locus tag, (PM70 locus tag) [a.b.c.]	Gene name	AL3358 log_2 fold-change, (FDR) [d.e.]	AL3357 log_2 fold-change, (FDR) [d. e]	AL3358 v AL3357 log_2 fold-change, (FDR) [d. e]	Predicted product	General function prediction
PMVP_1055, (PM1040)	*mglC*	-1.11, (0.015)	-0.31, (0.335)	-0.80, (0.01707)	Beta-methylgalactoside transporter inner membrane component	Carbohydrate transport and metabolism genes
PMVP_1057, (PM1042)	*petL*	-0.69, (0.228)	0.76, (0.135)	-1.45, (0.00508)	Phosphoethanolamine transferase specific for Lipid A	Cell wall/membrane biogenesis genes
PMVP_1064, (PM1048)	*gshAB*	-0.25, (0.408)	1.00, (0.001)	-1.25, (0.00008)	Bifunctional glutamate--cysteine ligase/glutathione synthetase	Cell wall/membrane biogenesis genes
PMVP_1065, (PM_t32)	tRNA Cys	-0.29, (0.505)	2.61, (0.0001)	-2.91, (0.00002)	tRNA Cys	Translation genes
PMVP_1076, (PM1058)	PM1058	0.05, (0.93)	1.76, (0.00003)	-1.71, (0.00002)	PRO domain containing protein	No function prediction
PMVP_1077, (PM_t36)	tRNA Ser	-0.15, (0.842)	1.97, (0.0005)	-2.12, (0.00018)	tRNA Ser	Translation genes
PMVP_1082, (PM1063)	*mltR*	0.28, (0.255)	1.65, (0.000001)	-1.37, (0.00001)	Mannitol repressor protein	Transcription genes
PMVP_1083, (PM1064)	PM1064	-1.14, (0.025)	-0.61, (0.105)	-0.52, (0.11544)	5'-nucleotidase lipoprotein e(P4) family protein	No function prediction
PMVP_1090, (PM1070)	PM1070	0.75, (0.179)	-0.53, (0.238)	1.28, (0.00612)	DUF406 family protein	No function prediction
PMVP_1115, (PM1095)	PM1095	-0.09, (0.755)	1.05, (0)	-1.14, (0.00006)	Hypothetical protein PM1095	No function prediction
PMVP_1126, (PM1104)	*mtrF*	-1.33, (0.086)	-0.13, (0.852)	-1.21, (0.03799)	AbgT putative transporter family subfamily	Coenzyme transport and metabolism genes
PMVP_1137, (NA)	tRNA	-0.60, (0.546)	1.08, (0.093)	-1.68, (0.01299)	tRNA	No function prediction
PMVP_1142, (PM1118)	*argC*	-1.44, (0.046)	-0.07, (0.92)	-1.37, (0.01424)	N-acetyl-gamma-glutamyl-phosphate reductase	Amino acid transport and metabolism genes
PMVP_1143, (PM1119)	*argB*	-1.19, (0.03)	-0.22, (0.601)	-0.96, (0.01826)	Acetylglutamate kinase	Amino acid transport and metabolism genes
PMVP_1212, (PM1194)	PM1194	-1.24, (0.015)	-0.76, (0.044)	-0.49, (0.16874)	Class I SAM-dependent methyltransferase	Secondary metabolites biosynthesis, transport and catabolism genes
PMVP_1233, (PM1211)	PM1211	-1.27, (0.032)	-1.01, (0.029)	-0.26, (0.53368)	Transglutaminase-like superfamily domain protein	Amino acid transport and metabolism genes

VP161 locus tag, (PM70 locus tag) [a.b.c.]	Gene name	AL3358 log_2 fold-change, (FDR)[d.e.]	AL3357 log_2 fold-change, (FDR) [d. e]	AL3358 v AL3357 log_2 fold-change, (FDR) [d. e]	Predicted product	General function prediction
PMVP_1271, (PM1248)	PM1248	1.42, (0.033)	1.41, (0.011)	0.01, (0.98251)	Gluconolactonase precursor	Carbohydrate transport and metabolism genes
PMVP_1281, (PM1258)	PM1258	-0.64, (0.14)	0.50, (0.124)	-1.14, (0.00226)	Conserved hypothetical protein TIGR00645	No function prediction
PMVP_1282, (PM1259)	PM1259	-1.41, (0.019)	-0.21, (0.661)	-1.21, (0.00882)	MFS transporter	Carbohydrate transport and metabolism genes
PMVP_1283, (PM1260)	*thiE*	-1.38, (0.015)	-0.53, (0.177)	-0.84, (0.03586)	Thiamine-phosphate pyrophosphorylase	Coenzyme transport and metabolism genes
PMVP_1284, (PM1261)	*thiD*	-1.52, (0.03)	-0.63, (0.212)	-0.89, (0.0697)	Phosphomethylpyrimidine kinase	Coenzyme transport and metabolism genes
PMVP_1290, (PM1266)	PM1266	-1.11, (0.214)	0.66, (0.301)	-1.77, (0.00994)	ABC transporter, ATP-binding protein	Inorganic ion transport and metabolism genes
PMVP_1298, (PM1274)	*lsrA*	-0.60, (0.283)	0.50, (0.211)	-1.10, (0.01027)	Autoinducer 2 ABC transporter ATP binding protein LsrA	Carbohydrate transport and metabolism genes
PMVP_1342, (PM1315)	*corA*	0.15, (0.837)	1.18, (0.007)	-1.03, (0.00939)	Magnesium and cobalt transport protein CorA	Inorganic ion transport and metabolism genes
PMVP_1349, (PM_t42)	tRNA Phe	-0.67, (0.631)	3.06, (0.002)	-3.73, (0.00048)	tRNA Phe	Translation genes
PMVP_1350, (PM1322)	*tcmP*	0.42, (0.122)	1.11, (0.0002)	-0.69, (0.00214)	Class I SAM-dependent methyltransferase	Secondary metabolites biosynthesis, transport and catabolism genes
PMVP_1370, (PM1341)	*uhpT*	-0.23, (0.729)	1.30, (0.006)	-1.54, (0.00156)	Hexose phosphate transport protein	Carbohydrate transport and metabolism genes
PMVP_1390, (PM1359)	*gntP_2*	-0.06, (0.935)	1.05, (0.004)	-1.10, (0.00177)	Gluconate permease	Carbohydrate transport and metabolism genes
PMVP_1397, (PM1366)	PM1366	-0.31, (0.612)	0.72, (0.08)	-1.03, (0.01382)	3-hydroxyisobutyrate dehydrogenase, putative	Lipid transport and metabolism genes
PMVP_1411, (PM1371)	PM1371	-1.41, (0.009)	-0.12, (0.77)	-1.30, (0.00118)	Xylulokinase	Carbohydrate transport and metabolism genes
PMVP_1412, (PM1372)	PM1372	-1.38, (0.009)	-0.18, (0.629)	-1.20, (0.00178)	PfkB family carbohydrate kinase family	Carbohydrate transport and metabolism genes

VP161 locus tag, (PM70 locus tag) [a.b.c.]	Gene name	AL3358 log_2 fold-change, (FDR) [d.e.]	AL3357 log_2 fold-change, (FDR) [d. e]	AL3358 v AL3357 log_2 fold-change, (FDR) [d. e]	Predicted product	General function prediction
PMVP_1413, (PM1373)	*kbaY*	-1.35, (0.011)	-0.17, (0.674)	-1.18, (0.00321)	Ketose 1,6-bisphosphate aldolase	Carbohydrate transport and metabolism genes
PMVP_1414, (PM1374)	PM1374	-1.46, (0.01)	-0.07, (0.868)	-1.39, (0.00144)	D-lyxose /D-mannose family sugar isomerase	Carbohydrate transport and metabolism genes
PMVP_1415, (PM1375)	*frk*	-1.44, (0.011)	-0.13, (0.77)	-1.31, (0.00285)	Fructokinase	Transcription genes
PMVP_1417, (PM1377)	*rbsB_2*	-1.39, (0.039)	-0.23, (0.722)	-1.16, (0.01948)	Periplasmic binding proteins and sugar binding domain of the LacI family., putative	Carbohydrate transport and metabolism genes
PMVP_1418, (PM1378)	*rbsC*	-1.69, (0.011)	-0.91, (0.048)	-0.78, (0.04475)	Ribose ABC transporter permease	Carbohydrate transport and metabolism genes
PMVP_1419, (PM1379)	*rbsA_2*	-1.91, (0.033)	-0.98, (0.15)	-0.92, (0.10613)	Ribose transport ATP-binding protein rbsa	Carbohydrate transport and metabolism genes
PMVP_1479, (PM1428)	PM1428	-0.28, (0.364)	1.12, (0)	-1.40, (0.00003)	TonB-dependent receptor	Inorganic ion transport and metabolism genes
PMVP_1494, (PM1443)	*glpT*	-1.23, (0.015)	-0.33, (0.378)	-0.89, (0.01486)	Glycerol-3-phosphate transporter	Carbohydrate transport and metabolism genes
PMVP_1504, (PM1453)	*adh2*	1.36, (0.353)	-0.61, (0.513)	1.97, (0.0475)	Bifunctional acetaldehyde-CoA/alcohol dehydrogenase	Energy production and conversion genes
PMVP_1514, (PM1463)	*trpG_2*	0.03, (0.958)	1.10, (0.001)	-1.07, (0.00062)	Para-aminobenzoate synthase component II	Amino acid transport and metabolism genes
PMVP_1529, (PM1477)	PM1477	-1.52, (0.179)	0.24, (0.799)	-1.76, (0.03586)	NAD(P)/FAD dependent oxidoreductase	No function prediction
PMVP_1574, (NA)		0.58, (0.342)	-1.12, (0.043)	1.70, (0.00284)	Hypothetical	No function prediction
PMVP_1575, (PM1518)	*plpP*	0.68, (0.056)	-1.21, (0.001)	1.89, (0.00003)	Outer membrane lipoprotein PlpP, putative	No function prediction
PMVP_1608, (PM1556)	*comF*	-0.12, (0.681)	1.24, (0)	-1.36, (0.00002)	Competence protein F	No function prediction
PMVP_1620, (PM1568)	PM1568	-0.80, (0.258)	0.22, (0.701)	-1.01, (0.04668)	DUF465 domain containing protein	No function prediction
PMVP_1623, (PM1571)	*rimO*	0.75, (0.011)	-0.82, (0.001)	1.57, (0.00001)	30S ribosomal protein S12 methylthiotransferase RimO	Translation genes

VP161 locus tag, (PM70 locus tag) [a.b.c.]	Gene name	AL3358 log_2 fold-change, (FDR) [d.e.]	AL3357 log_2 fold-change, (FDR) [d. e]	AL3358 v AL3357 log_2 fold-change, (FDR) [d. e]	Predicted product	General function prediction
PMVP_1638, (PM1584)	*rpoH*	-0.47, (0.206)	1.16, (0.003)	-1.64, (0.00019)	RNA polymerase factor sigma-32	Transcription genes
PMVP_1688, (NA)		0.05, (0.899)	1.16, (0.00004)	-1.12, (0.00003)	Hypothetical	No function prediction
PMVP_1704, (PM1654)	PM1654	-0.29, (0.732)	1.37, (0.012)	-1.66, (0.00308)	Hypothetical protein PM1654	No function prediction
PMVP_1712, (NA)		-0.36, (0.696)	1.49, (0.014)	-1.86, (0.00332)	Hypothetical	No function prediction
PMVP_1762, (PM1707)	*nanM*	-0.66, (0.228)	0.62, (0.168)	-1.28, (0.0062)	YjhT family mutarotase	No function prediction
PMVP_1764, (PM1709)	PM1709	-0.61, (0.23)	0.50, (0.226)	-1.11, (0.00885)	TRAP dicarboxylate transporter- DctP subunit subfamily, putative	Carbohydrate transport and metabolism genes
PMVP_1782, (PM_t48)	tRNA Trp	-0.58, (0.656)	1.36, (0.101)	-1.94, (0.02106)	tRNA Trp	Translation genes
PMVP_1787, (PM1730)	*plpB*	0.16, (0.757)	-0.91, (0.009)	1.07, (0.00285)	Outer membrane lipoprotein 2 precurso	Inorganic ion transport and metabolism genes
PMVP_1805, (PM_t52)	tRNA Gly	-0.04, (0.964)	1.50, (0.007)	-1.54, (0.00354)	tRNA Gly	Translation genes
PMVP_1806, (PM_t53)	tRNA Tyr	0.05, (0.951)	1.36, (0.008)	-1.31, (0.00728)	tRNA Tyr	Translation genes
PMVP_1807, (PM_t54)	tRNA Thr	-0.09, (0.948)	1.41, (0.052)	-1.49, (0.03003)	tRNA Thr	Translation genes
PMVP_1831, (PM_t55)	tRNA SeC	0.44, (0.576)	1.21, (0.029)	-0.77, (0.09584)	tRNA SeC	Translation genes
PMVP_1833, (NP)		-0.08, (0.832)	1.26, (0.0001)	-1.33, (0.00002)	Hypothetical	No function prediction
PMVP_1837, (PM1787)	*rseB*	-0.24, (0.663)	0.78, (0.046)	-1.02, (0.00983)	Periplasmic negative regulator of sigmaE	Signal transduction mechanisms genes
PMVP_1838, (PM1788)	*mclA*	-0.41, (0.368)	1.03, (0.01)	-1.43, (0.00097)	Sigma-E factor negative regulatory protein	Signal transduction mechanisms genes
PMVP_1839, (PM1789)	*rpoE*	-0.45, (0.092)	0.60, (0.012)	-1.05, (0.00024)	RNA polymerase sigma factor RpoE	Transcription genes
PMVP_1871, (PM1818)	PM1818	0.54, (0.206)	-0.92, (0.014)	1.46, (0.00054)	Hypothetical virulence factor protein PM1818	No function prediction

VP161 locus tag, (PM70 locus tag) [a.b.c.]	Gene name	AL3358 log_2 fold-change, (FDR) [d.e.]	AL3357 log_2 fold-change, (FDR) [d. e]	AL3358 v AL3357 log_2 fold-change, (FDR) [d. e]	Predicted product	General function prediction
PMVP_1872, (PM1819)	*srfB*	0.33, (0.233)	-0.81, (0.002)	1.14, (0.00011)	Virulence factor SrfB	No function prediction
PMVP_1964, (PM1912)	rpL32	-0.35, (0.639)	0.77, (0.123)	-1.13, (0.02434)	50S ribosomal protein L32	Translation genes
PMVP_1984, (PM1932)	tRNA Arg	-0.54, (0.369)	0.68, (0.104)	-1.22, (0.00667)	tRNA Arg	Translation genes
PMVP_1987, (PM1934)	PM1934	0.00, (0.992)	1.04, (0.001)	-1.05, (0.00054)	Hypothetical	No function prediction
PMVP_1998, (PM1959)	*leuD*	0.26, (0.342)	1.14, (0.0001)	-0.87, (0.00043)	3-isopropylmalate dehydratase, small subunit	Amino acid transport and metabolism genes
PMVP_1999, (PM1960)	*leuC*	0.21, (0.61)	1.28, (0.001)	-1.08, (0.00106)	3-isopropylmalate dehydratase, large subunit	Amino acid transport and metabolism genes
PMVP_2023, (PM1985)	*tsf*	-0.19, (0.777)	-1.07, (0.013)	0.88, (0.02417)	Translation elongation factor Ts	Translation genes
PMVP_2091, (PM0039)	*ars*	0.52, (0.246)	1.39, (0.001)	-0.87, (0.00667)	Metallo-beta-lactamase superfamily domain protein	No function prediction
PMVP_2095, (PM0043)	*gdhA*	-0.24, (0.463)	0.80, (0.003)	-1.04, (0.00038)	Glutamate dehydrogenase	Amino acid transport and metabolism genes
PMVP_2500, (NA)		-0.17, (0.752)	1.06, (0.011)	-1.22, (0.00342)	Putative glycine riboswitch	No function prediction
PMVP_2517, (NA)		-1.29, (0.098)	0.16, (0.804)	-1.45, (0.01654)	Putative his leader sequence	No function prediction
Prrc08, (NA)	Prrc08	0.34, (0.656)	2.34, (0.0004)	-2.00, (0.00048)	Prrc08 putative sRNA	sRNA
Prrc09, (NA)	Prrc09	-0.65, (0.389)	0.38, (0.497)	-1.03, (0.04506)	hypothetical (see Table 1.1)	No function prediction
Prrc12, (NA)	Prrc12	-0.48, (0.031)	0.53, (0.006)	-1.01, (0.00005)	Prrc12 putative sRNA	sRNA
Prrc13, (NA)	Prrc13	-2.07, (0.021)	0.90, (0.168)	-2.98, (0.00046)	Prrc13 putative sRNA	sRNA
Prrc17, (NA)	Prrc17	-0.96, (0.19)	0.32, (0.56)	-1.28, (0.01877)	Prrc17 putative sRNA	sRNA
Prrc25, (NA)	Prrc25	-0.76, (0.579)	1.31, (0.112)	-2.07, (0.01834)	Prrc25 putative sRNA	sRNA
Prrc32, (NA)	Prrc32	0.43, (0.594)	1.63, (0.007)	-1.21, (0.01639)	Prrc32 putative sRNA	sRNA
Prrc46, (NA)	Prrc46	-1.06, (0.438)	1.25, (0.133)	-2.31, (0.01378)	Prrc46 putative sRNA	sRNA
Prrc49, (NA)	Prrc49	-0.13, (0.926)	1.30, (0.054)	-1.43, (0.02613)	Prrc49 putative sRNA	sRNA

VP161 locus tag, (PM70 locus tag) [a.b.c.]	Gene name	AL3358 log_2 fold-change, (FDR) [d.e.]	AL3357 log_2 fold-change, (FDR) [d. e]	AL3358 v AL3357 log_2 fold-change, (FDR) [d. e]	Predicted product	General function prediction
tRNA Met, (PM_t25)	tRNA Met	0.17, (0.911)	2.69, (0.001)	-2.52, (0.00098)	tRNA Met	Translation genes
tRNA Val, (PM_t41)	tRNA Val	0.17, (0.766)	3.50, (0.000001)	-3.33, (0.000001)	tRNA Val	Translation genes
tRNA Ser, (PM_t30)	tRNA Ser	-0.14, (0.936)	3.33, (0.001)	-3.47, (0.00026)	tRNA Ser	Translation genes
tRNA Gly, (PM_t31)	tRNA Gly	-0.42, (0.423)	1.19, (0.003)	-1.61, (0.00035)	tRNA Gly	Translation genes

[a] Differentially expressed transcripts were defined as those showing at least 2-fold change in production ($log_2 \leq -1$ or $log_2 \geq 1$) with an FDR of less than 0.05.

[b.] NA= Not annotated but sequence present, NP= No sequence present

[c.] All predicted sRNAs can be found in Table 1.1

[d.] Transcript expression ratio is shown as a log_2 value with the corresponding false discovery rate (FDR) shown in brackets.

[e.] Shaded boxes indicate a statistically significant difference in expression.

3.3.4 Determination of ProQ-bound RNA species

The transcriptomics experiments described above identified many transcripts that showed altered expression in response to *proQ* inactivation or overexpression. However, these analyses cannot identify RNA species that interact directly with ProQ. Therefore, a co-immunoprecipitation strategy (UV-CLASH; Waters et al. 2017) was used to identify those RNA species that directly interact with ProQ (Figure 3.4). In order to perform the UV-CLASH experiments, a *P. multocida proQ* mutant was provided with a plasmid encoding a recombinant version of ProQ that was His-TEV-FLAG (HTF)-tagged at the 3' end to allow for co-immunoprecipitation. As a requirement for use in the UV crosslinking apparatus in the Tree laboratory, University of New South Wales where these experiments were performed, the *proQ* mutant was made avirulent by inactivation of *hyaD* (using a markerless TargeTron® intron), an essential capsule biosynthesis gene (Chung et al. 1998). The double *proQ/hyaD* mutant (AL3067) was assessed by Southern blotting, using probes to detect the intron (with or without the kanamycin gene) or to detect only the kanamycin gene (Figure 3.5A-D). Hybridisation with a kanamycin-specific probe identified a 2.4 kb DraI fragment, and an 8.5 kb EcoRV fragment, a profile that corresponded to a successful insertional inactivation of *proQ* (Figure 3.5A & C). Hybridisation with the intron-specific probe (Figure 3.5D) identified a 1.8 kb EcoRV fragment and a 1.4 kb DraI fragment correlating to an intron insertion in *hyaD* (Figure 3.5B), and a 2.4 kb DraI fragment and an 8.5 kb EcoRV fragment correlating to an intron insertion in *proQ* (Figure 3.5A).

The confirmed *proQ/hyaD* double mutant was then transformed with a plasmid expressing a HTF-tagged ProQ protein (pAL1339) to generate the strain AL3068. A plasmid encoding an untagged recombinant ProQ was also constructed (pAL1338) and used to transform the double mutant, to generate the control strain AL3069. Both strains (AL3068 and AL3069) were then used in UV-crosslinking experiments. RNA species that crosslinked with ProQ were recovered by multistep co-immunoprecipitation of the tagged ProQ (anti-His followed by anti-FLAG antibodies). The isolated RNA was reverse transcribed into cDNA, sequenced and the resultant data mapped to the VP161 genome to identify the regions representing transcripts that had increased association with the tagged ProQ protein. In addition to the above procedure, and in order to identify interacting RNAs (Figure 3.4), an intermolecular RNA ligation step was included (after UV cross-linking and prior to co-immunoprecipitation and ligation of adapters for sequencing), which ligated interacting RNA molecules so that they formed hybrid RNA molecules. Detailed bioinformatic analyses of the cDNA sequences generated from these ligated products allowed for the identification of any hybrid sequences (i.e. continuous sequence that matched two different genome positions), which would indicate two RNA species directly interacting with each other and/or with the same ProQ molecule.

Analysis of the sequencing data generated from the co-immunoprecipitation data identified 73 RNA species that showed a significant association with the tagged-ProQ protein, as compared to the untagged control (Appendix 6). Of these, 28 encoded known ribosomal proteins, including 18 of the 31 RNAs that encode components of the 50s subunit and 10 of the 21 RNAs that encode components of the 30s ribosomal subunit. Thus, RNA encoding ribosomal components were highly over-represented (p = 2.2e-16) in those molecules that interact with ProQ. In addition, seven tRNA species showed significant binding to ProQ, namely tRNA Met (PMVP_0657), tRNA Met (PM_t25), tRNA Cys (PMVP_1065), tRNA Ser (PMVP_1077), PMVP_1137, tRNA Val (PM_t41) and tRNA Trp (PMVP_1782). The ProQ protein was also shown to bind to six predicted sRNA molecules, two of which had been previously identified within *P. multocida* (Prrc02 and Prrc13; Table 1.1) and four of which were identified for the first time (Prrc54-57; Table 1.1).

An analysis of ProQ-bound RNA species identified that ProQ predominately binds to the beginning of the terminator stem loop at the 3' end of RNA transcripts; however, there were many other ProQ binding sites in different locations on other targets (Figure 3.5E). The sequences of molecules that bound to ProQ were all examined for the presence of a conserved sequence motif and although no clear sequence motif could be identified the analysis revealed that a large proportion of the ProQ-bound RNA could form stem loops that potentially acted as terminators. Further examination showed that these regions of the stem-loop structures were A-rich, and the bases could pair to form the stem of each potential terminator. (Figure 3.5F).

As noted above, to identify RNA species that directly interact with each other at a single ProQ molecule, an RNA ligation step was included in the sample processing (Figure 3.4). The analysis of the sequences generated revealed 79 unique hybrid molecules or RNA pairs. These included sRNAs, tRNAs, mRNAs, RNAs from the 5' or 3'-UTR of mRNA transcripts, and tmRNA. Some of the interacting RNA species identified by the CLASH analysis were not annotated in the VP161 genome, and were mostly found to be probable antisense RNAs, therefore they were termed misc_x (miscellaneous_a-h, Table 3.8). All of the RNA binding partners were then mapped and grouped according to RNA type using Cytoscape (Figure 3.6). Several RNA species were identified to interact with multiple transcripts. For example, hybrids were identified between the 3' UTR of *ompA* (PM0786) and RNaseP (PMVP_2505), tRNA Tyr (PMVP_1806), tRNA Cys (PMVP_1065), Prrc10 (sRNA), and *hupA.* The binding of the 3' UTR of *ompA* to multiple transcripts suggests that this region may be clipped to form a regulatory sRNA molecule that interacts with ProQ and multiple mRNA targets, as has been observed in *Salmonella* where the ProQ-dependent sRNA RaiZ is derived from the 3' end of the *raiA* mRNA (Smirnov et al. 2017).

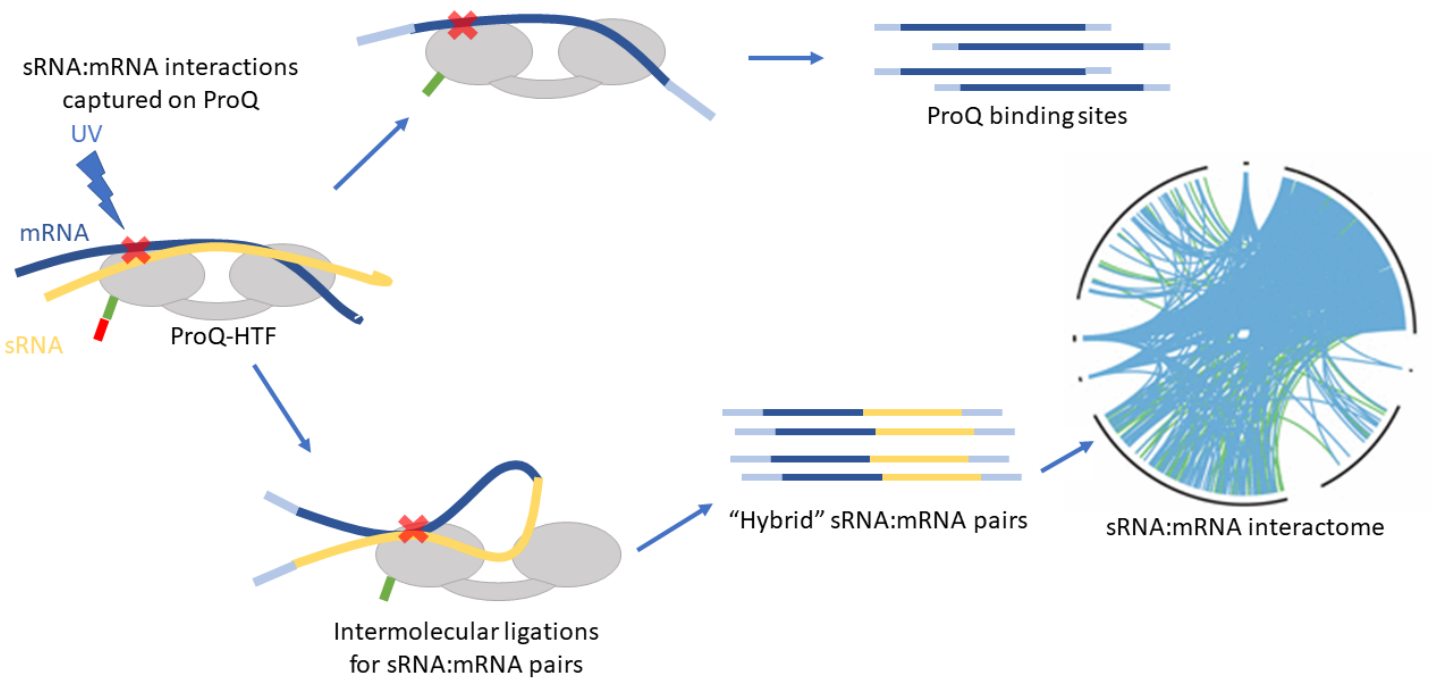

Figure 3.4. Schematic representation of the ProQ UV-CLASH protocol. The His-TEV-3xFLAG-tagged (green and red boxes) ProQ protein (grey) binds and facilitates the interaction between one or more mRNAs (dark blue) and sRNAs (yellow). The CLASH protocol UV-crosslinks (blue thunderbolt) the interacting RNA species to the ProQ protein. The ProQ-RNA complexes are co-immunoprecipitated (top panel), trimmed and adapters (light blue) are ligated. Adapter-ligated RNA can then be sequenced to identify ProQ binding sites. To identify directly interacting RNA molecules, before co-immunoprecipitation an intermolecular ligation step is performed (bottom panel). Interacting mRNA-sRNA pairs can be ligated together to form hybrid RNA pairs (dark blue and yellow lines). These hybrids can then be sequenced and a map of the ProQ mediated sRNA-mRNA interactome can be constructed. Adapted from Waters et al. (2017)

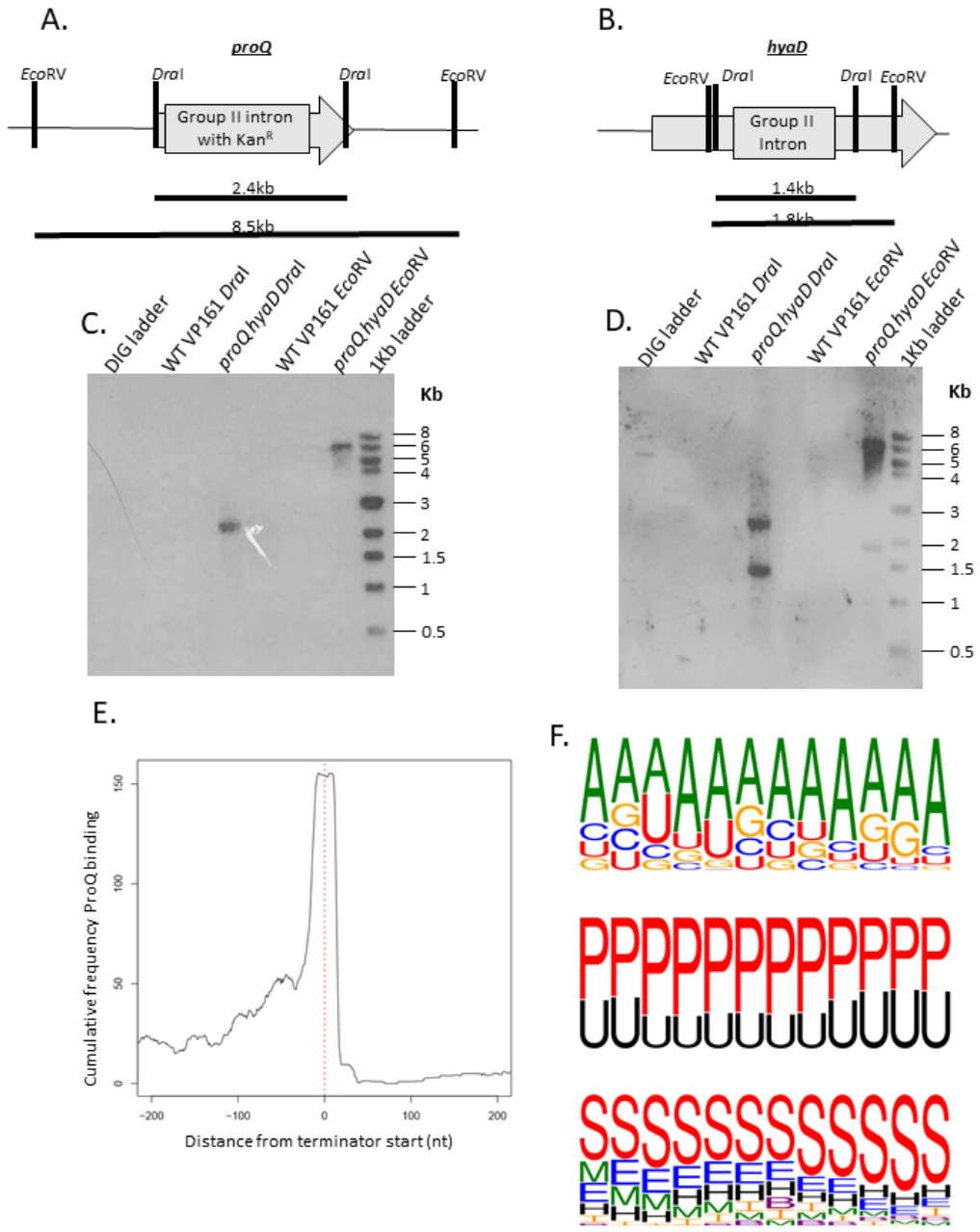

Figure 3.5. A and **B.** Schematic representation of the *proQ* and *hyaD* genes depicting the expected fragments following digestion with EcoRV and DraI. **C** Southern blot analysis of the EcoRV and DraI cut *P. multocida* wild-type and double *proQ* and *hyaD* mutant genomes, probed with a kanamycin cassette probe. **D** Southern blot analysis of the EcoRV and DraI cut *P. multocida* wild-type and double *proQ* and *hyaD* mutant genomes, probed with a group-II intron probe. The intron probe bound to two bands; the larger band represents the Kan^R intron in *proQ* and the smaller band represents the markerless intron in *hyaD*. **E.** ProQ binding site frequency as a function of position on the transcripts relative to the beginning of the predicted terminator stem-loop sequence. **F.** ProQ binding sequence (top), pairing (middle) and structure (bottom) motifs (where S= stem, E= external bulge, B= bluge loop, H= hairpin loop, and I-internal loop) as determined from transcripts bound at the beginning of the terminator stem-loop as identified using GraphProt.

Table 3.8. Miscellaneous RNAs identified through UV-CLASH

Name	Start base	End base	Size	Strand	Sequence	Extra information
misc_a	2200531	2200597	67	+	TTGTACCGGACCGGGATAATCTAATA AATCAGTGTATAGATAGCCGACTAAA TAAGCTGGGATCGCT	Antisense to *fpbP* (PMVP_2103)
misc_b	439595	439615	21	+	TTCTGGCTGACCGAGGAGGCTT	Antisense to PM0445 (PMVP_0418)
misc_c	1663991	1663966	26	-	ATCAGAAATCGCAAAAATCACACCGC	Antisense to *ftsY* (PMVP_1576)
misc_d	1860367	1860426	60	+	TAAATAAGTGGCGGAACGGACGGGA CTCGAACCCGCGACC	Between tRNA Trp (PMVP_1782) and 16s ribosomal RNA (PMVP_1783)
misc_e	637375	637338	38	-	ATATTTTCTAAGGCTTTGAGCATGAG ATTCCCTAAGTT	Antisense to *yvcK* (PMVP_0598)
misc_f	1764163	1764242	80	+	GAAATACACACCATGGCATCGGCAC AGTGTGCATTGACCATATATTCCACA CTATCGGCAATTAAATCACGACTTGG TAA	Antisense to *ilvD* (PMVP_1680)
misc_g	2210455	2210423	33	-	TACATTGCACCGGTACTCGCACCGCT AATCGCA	Antisense to *pfhB1* (PMVP_2110)
misc_h	1269211	1269249	39	+	CCGAAAAGTGCGGTGTTGCTGTGGTT GCATTTGCTAAAA	Antisense to *xylB* (PMVP_1296)

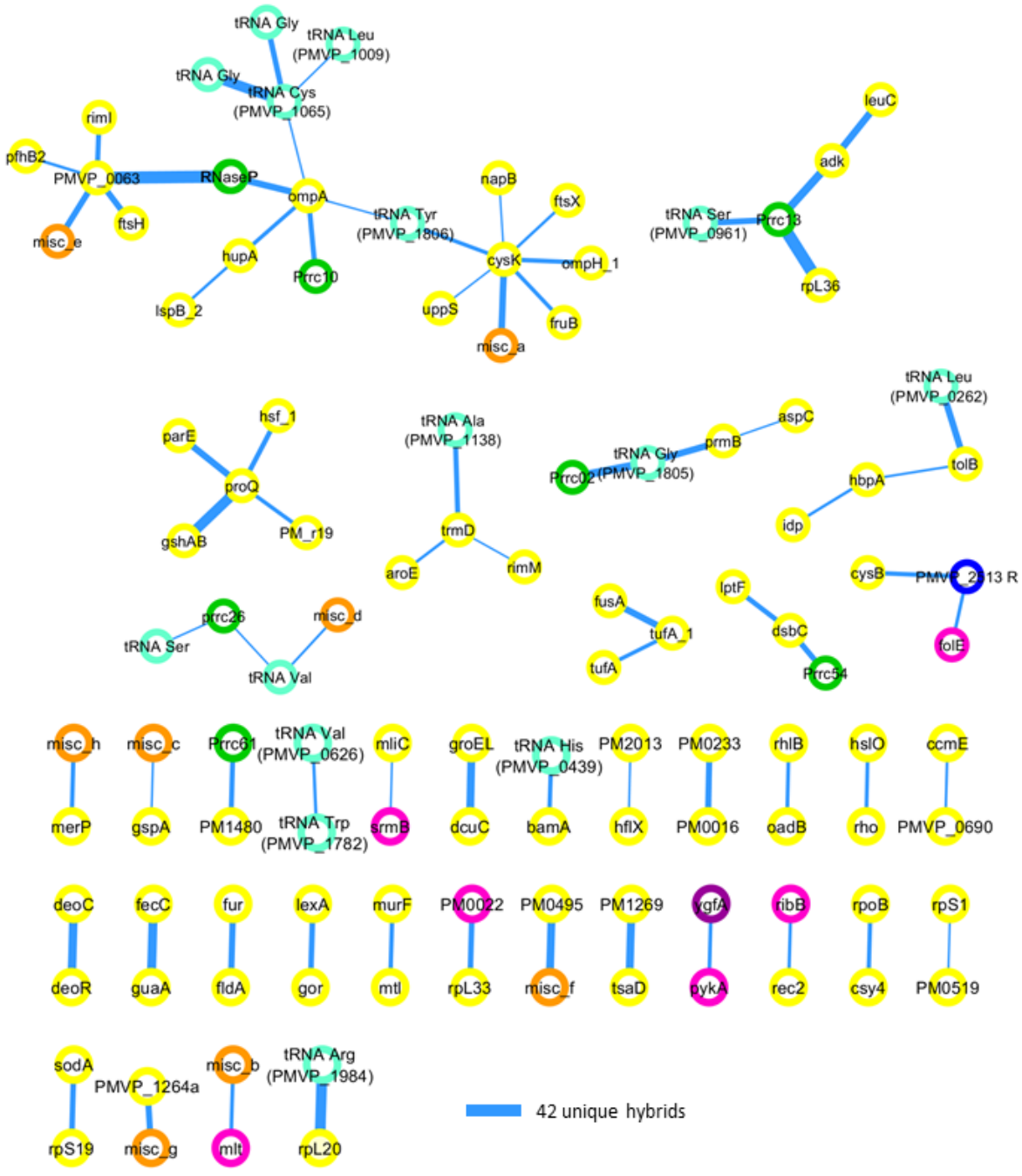

Figure 3.6. The ProQ regulon in *P. multocida* VP161. Connection network of ProQ bound RNAs (nodes), showing interactions between mRNAs (yellow), putative sRNAs (green), tRNAs (blue), 3' UTRs (pink), 5' UTRs (purple), tmRNAs (navy blue), miscellaneous RNAs (orange) (Table 3.8) and their binding partners. RNA species found in hybrids are joined (blue lines) and the thickness of each line indicates the number of hybrids identified.

3.3.5 Effect of *proQ* mutation on the downstream gene *prc*

In *E. coli*, the *prc* transcriptional start site is located within the coding sequence of *proQ* and mutagenesis of *proQ* has been shown to affect transcription of *prc* (Kerr et al. 2014). Therefore, it was possible that TargeTron® insertional mutagenesis of *proQ* affected the transcription of *prc* in *P. multocida*. In order to determine if the *prc* transcriptional start site in *P. multocida* was also within the *proQ* coding sequence, the 5' end of the *prc* transcript was determined using 5' RACE on RNA isolated from wild-type *P. multocida* VP161 cells grown to an OD_{600} = 0.6 in rich medium. The data showed that the 5' start of the *prc* transcript and the *proQ* open reading frame overlapped by 291 nucleotides (Figure 3.7A). However, the TargeTron® intron in the *proQ* mutant is located 176 nucleotides upstream of the start of the *prc* transcriptional start, indicating that *prc* transcription initiation should be unaffected by the intron insertion. To confirm that *prc* transcription was unaffected by *proQ* mutation, a *proQ* mutant/*prc* overexpression strain (AL3214) was constructed by adding a plasmid that constitutively expresses *prc* (pAL1387) and *ycbB* (PMVP_0229) to the *proQ* mutant (AL2973). Both *prc* and *ycbB* were cloned as they are predicted to be co-transcribed. Following this, qRT-PCR was performed, comparing the expression of *prc* and the 5' region of *proQ* (before the TargeTron® insertion) in the wild-type *P. multocida* VP161, *proQ* mutant (AL2973), *proQ* complementation/overexpression strain (AL2978) and *proQ* mutant/*prc* overexpression strain (AL3214) (Figure 3.7B). The data indicated that mutation of *proQ* alone (AL2973) did not change the expression of *prc*. However, transcript levels representing the 5' end of *proQ* increased by over 2-fold, indicating that ProQ may act to negatively regulate its own production. Both the *proQ* and *prc* overexpression strains (AL2978 and AL3214, respectively) showed the expected increase of *proQ* or *prc* expression respectively, compared to wild-type (~300 fold each). Interestingly, the overexpression of *proQ* in strain AL2978 led to a 3-fold increase in *prc* expression. Moreover, the overexpression of *prc* (in strain AL3214) led to a 2-fold increase in abundance of the 5' region of the *proQ* transcript (Figure 3.7B). Together these data indicate that the two transcripts may positively regulate each other. The UV-CLASH data was examined to determine if the ProQ protein co-immunoprecipitated with the *prc* transcript to enable this regulation, but no significant binding between the two molecules was detected.

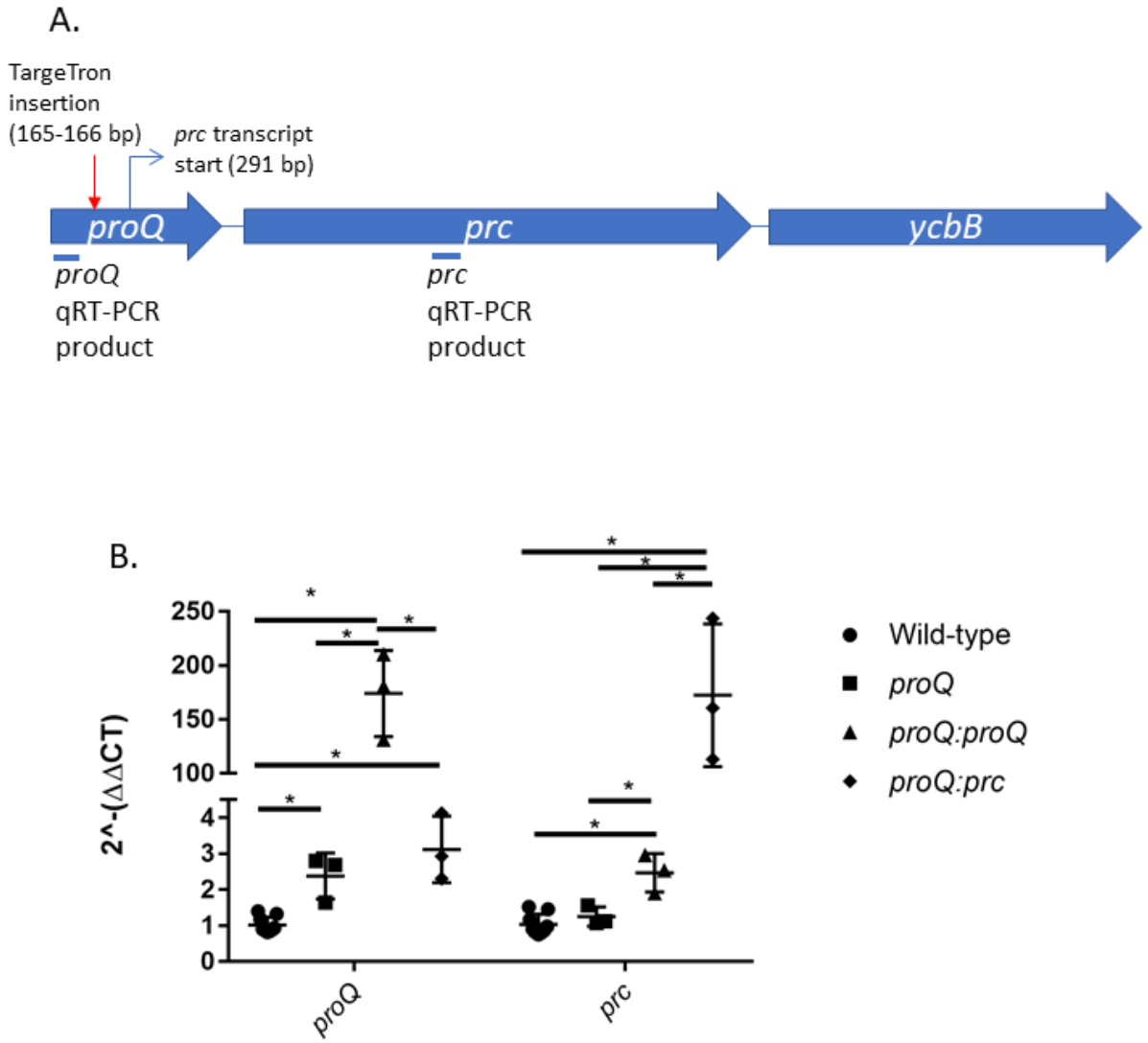

Figure 3.7. A. Schematic representation of the *proQ, prc* and *ycbB* genes (large blue arrows) including the TargeTron® insertion site within AL2973 (red arrow), the *prc* transcriptional start site identified through 5' RACE (small blue right-angled arrow), and the location of the *proQ* and *prc* specific PCR products produced by qRT-PCR. **B.** Levels of *proQ* and *prc* transcript identified through qRT-PCR, standardised to the house-keeping gene *gyrB*. The wild-type *P. multocida* VP161 strain (circles), was compared to the *proQ* mutant (AL2973), *proQ* overexpression strain (*proQ: proQ*) (AL2978), and *proQ* mutant complemented with *prc* (*proQ: prc*) (AL3214). Data shown is mean ±SD, n=3, *=p<0.05 using Student's T-test.

3.3.7 ProQ stabilizes the PMVP_0063 transcript

The co-immunoprecipitation analyses identified one transcript that was strongly associated with ProQ (Figure 3.8A) and corresponded to a region in the *P. multocida* VP161 genome encoding a hypothetical gene PMVP_0063. This hypothetical gene is found in the genomes of only four other *P. multocida* strains, namely RCAD0259, NCTC10382, PMTB2.1, and Razi 0002. Similarly, the predicted protein encoded by this gene is not commonly found within other bacterial species, as it has only been identified in *Haemophilus* and *Neisseria* species as a hypothetical protein with no conserved domains. Analysis of the PMVP_0063 secondary structure (SWISS-model analysis) revealed that a small region of the protein shared limited identity (25/79 residues, 31.7% identity) with proteins belonging to the Snf7 protein family. The Snf7 family of proteins are involved in protein transport into vacuoles in eukaryotic cells (Wemmer et al. 2011) but no homologs of the Sn7 proteins have yet been identified/reported in prokaryotes. Comparative transcriptomic analyses showed that the levels of PMVP_0063 transcript were decreased by 2.3-fold (log_2 fold-change of -1.2, Table 3.5) in the *proQ* mutant (AL2973) compared to expression levels in the wild-type parent strain VP161. In addition, the abundance of the PMVP_0063 transcript increased by 4-fold (log_2 fold-change = 1.98; Table 3.7) in the complemented *proQ* mutant/overexpression strain (AL3357) compared to the expression levels in the wild-type with empty vector (strain AL3356). Taken together, these data suggest that in the wild-type VP161 ProQ binds and stabilizes the PMVP_0063 transcript. To demonstrate this experimentally, Northern blotting was used to assess the levels of PMVP_0063 transcript in the wild-type *P. multocida* strain containing empty vector (AL3356), the *proQ* mutant containing empty vector (AL3358) and the *proQ* complementation/overexpression strain (AL3357) (Figure 3.8C). A PMVP_0063-specific DIG-labelled RNA probe was used, and two hybridizing transcripts were detected by Northern blotting, with the band of approximately 230 nt correlating with the predicted size of the PMVP_0063 transcript, and a second transcript of approximately 260 nt that may correspond to the primary transcript, indicating that the functional PMVP_0063 transcript is clipped from this longer transcript. Densitometry was used to assess the relative levels of the predicted 230 nt PMVP_0063 transcript. Significantly more PMVP_0063 transcript was identified in the *proQ* complementation/overexpression strain (AL3357) than in the wild-type strain or the *proQ* mutant (both with empty vector, Figure 3.8D). There was slightly less PMVP_0063 transcript detected in the *proQ* mutant compared to the wild-type, but this was not statistically significant ($p = 0.13$). The predicted ProQ binding position on the PMVP_0063 transcript was mapped relative to the PMVP_0063 secondary structure, as predicted by RNAfold (Lorenz et al. 2011) and visualised by VARNA (Darty et al. 2009; Figure 3.8B). ProQ is predicted to bind to the central region of the PMVP_0063 transcript, in a region with a

predicted strong secondary structure with several stem-loops. These data suggest that ProQ may bind to a central stem-loop-associated region on the PMVP_0063 transcript and block its degradation.

RNA-RNA hybrid analysis identified interactions between the PMVP_0063 transcript and five other transcripts, namely PMVP_0006 (PfhB2), which encodes filamentous haemagglutinin, *rimI,* which encodes the ribosomal protein alanine N-acetyltransferase, *ftsH,* which is an essential gene in *P. multocida* strain VP161 (Smallman, Boyce Laboratory, unpublished) that encodes an ATP-dependent zinc metalloprotease, RNaseP, which is a ribonuclear RNA processing enzyme, and misc_e, an antisense RNA to *yvcK* (PMVP_0598) (Figure 3.6; Table 3.8). Each hybrid identified was analysed for regions of base pair complementarity using RNAHybrid. In each case, a region of significant complementary was identified between the two molecules (Figure 3.9A-E), supporting the proposition that these molecules do directly interact with each other.

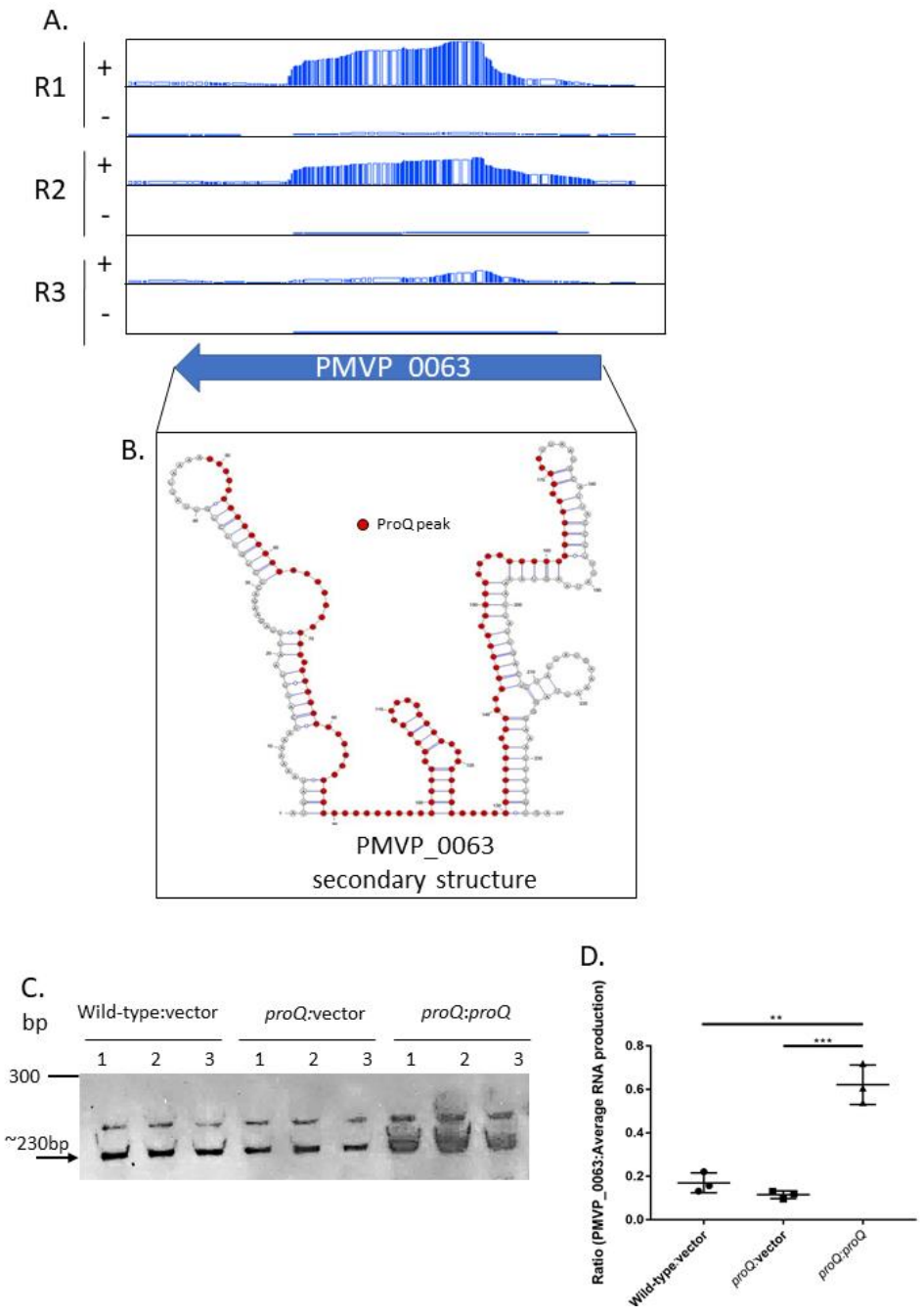

Figure 3.8. A. RNA-seq read coverage over the PMVP_0063 transcript following co-immunoprecipitation of RNA samples from strains expressing either tagged-ProQ (+) or untagged-ProQ (-); co-immunoprecipitation was performed in biological triplicate (R1-R3). **B.** The secondary structure of the PMVP_0063 transcript as predicted using RNAfold and indicating the ProQ binding region as determined by the co-immunoprecipitation data (red circles). **C.** Northern blot detection (in biological triplicate) of the PMVP_0063 transcript (~230 bp) in RNA from the wild-type *P. multocida* VP161 containing empty vector (AL3356), *proQ* mutant containing empty vector (AL3358) and *proQ* complementation/overexpression strain (AL3357). **D.** Densitometry from PMVP_0063 Northern blot, using total RNA production determined by SYBR stained rRNA, for standardisation, data shown are mean ± SD, n=3, **= $p < 0.005$ and *** = $p < 0.005$ as per a Student's T-test.

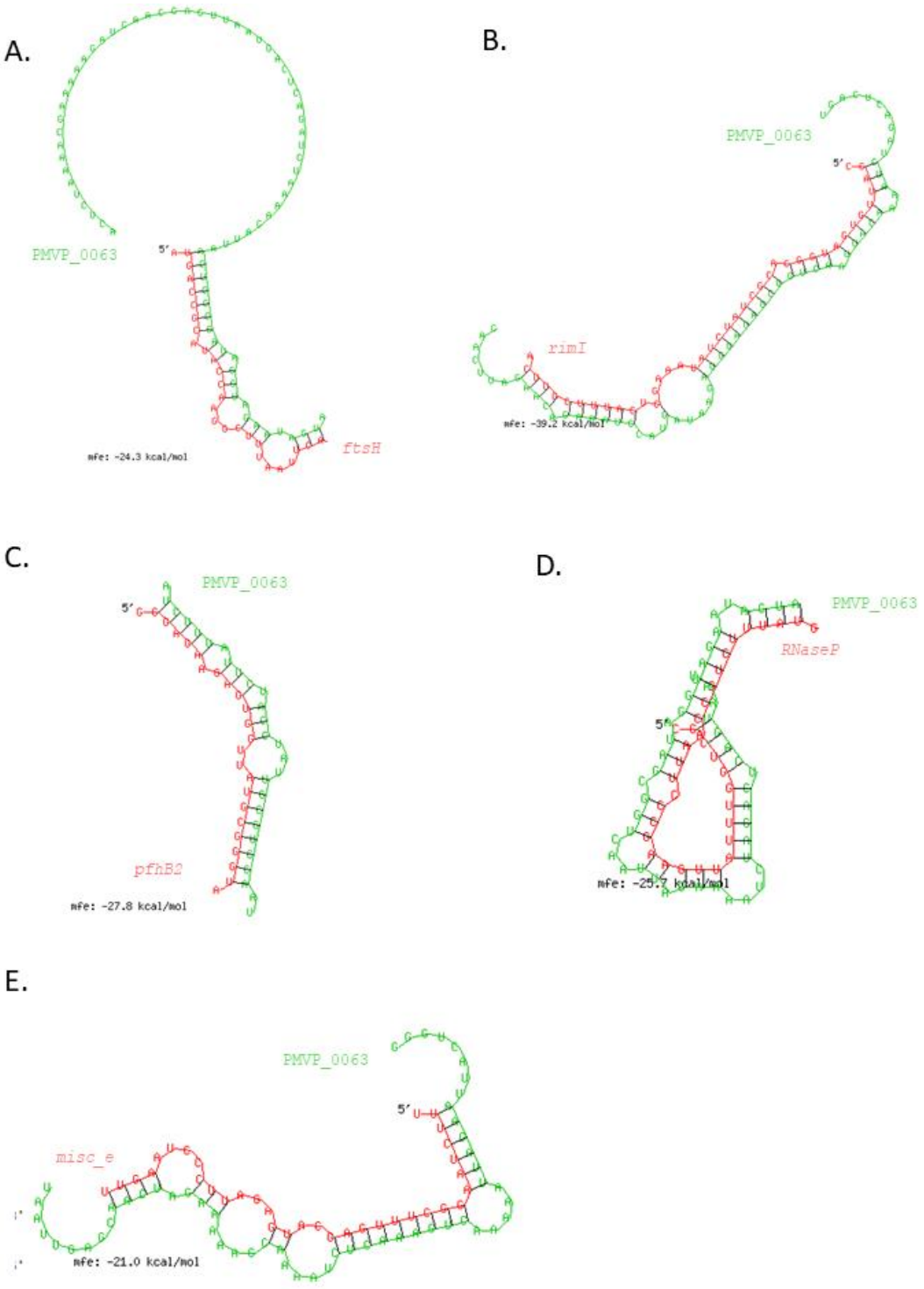

Figure 3.9. Putative binding interactions for RNA-RNA hybrids containing PMVP_0063 (green) and either **A.** *ftsH* (red) **B.** *rimI* (red) **C.** *pfhB2* (red) **D.** RNaseP (red) or **E.** misc_e (red) formed during UV-CLASH. Putative interactions were visualised using BiBiServ2: RNAhybrid.

3.3.8 The sRNA Prrc13 is stabilized by ProQ

The putative Prrc13 sRNA is located immediately upstream of PMVP_0391 encoding a metal-dependent hydrolase belonging to the YdjM superfamily of proteins. It was first identified using RNA-seq analysis of RNA isolated from *P. multocida* strain VP161, that was further enriched for small RNAs (< 200 bp) prior to library preparation (Mégroz, Table 1.1). Prrc13 was found in higher quantities in the sRNA-enriched data set, compared to abundance in the non-size selected samples. This putative sRNA was also differentially expressed in the *proQ* mutant containing empty vector (AL3358), where it showed a 4.2-fold reduction in expression (log_2 fold-change of -2.07), compared to the wild-type strain containing empty vector (AL3356). There was also a corresponding 1.9-fold increase in Prrc13 expression in the *proQ* mutant containing the complementation/overexpression plasmid (AL3357) compared to the wild-type strain containing empty vector strain. Analysis of the UV-CLASH experiment showed that ProQ bound strongly to the Prrc13 sRNA (Figure 3.10A). Taken together these data indicate that ProQ may act to stabilize the Prrc13 transcript. To confirm this, a Northern blot was performed to compare the levels of Prrc13 across these strains. Each strain was grown in biological triplicate to mid-exponential growth phase (OD_{600} = 0.6), RNA was extracted, and Northern blotting performed using a Prrc13-specific DIG-labelled riboprobe. The expression of Prrc13 was significantly reduced in the *P. multocida proQ* mutant compared to both the wild-type and *proQ* complementation/overexpression strains (Figure 3.10C and D). Furthermore, the *proQ* overexpression strain expressed more Prrc13 than the wild-type strain, but this was not statistically significant (p = 0.08; Figure 3.10C & D).

The UV-CLASH binding data allowed for identification of the likely ProQ binding sequence on the Prrc13 sRNA; ProQ bound at the 3' end of the transcript at a strong stem-loop structure (Figure 3.10A & B). The UV-CLASH analysis also identified hybrids that formed between Prrc13 and three other transcripts, namely *adk*, rpL36 (encoding a ribosomal protein), and the serine tRNA encoded by PMVP_0961 (Figure 3.11A-C). The gene *adk* encodes the adenylate kinase protein that catalyses the reaction of ATP to AMP and is essential in many bacteria (Glaser et al. 1975), including *P. multocida* strain VP161 (Thomas Smallman, Boyce laboratory, unpublished). The position that Prrc13 binds *adk* overlaps with the ProQ binding site and shows strong base pair complementarity (Figure 3.11A.). The *adk* transcript is highly expressed in *P. multocida* (~600 tpm in wild-type cells grown in HI media to OD_{600} = 0.6). However, adk was not identified as co-immunoprecipitated by ProQ and was not measured as differentially expressed in the transcriptomic or proteomic analyses of the *P. multocida proQ* mutant. This suggests that *adk* is regulated through other levels of regulation, which is supported by the fact that UV-CLASH also showed that it interacted with the *leuC* RNA. This *adk-leuC* interaction occurred at a different location of *adk* from the Prrc13 interaction.

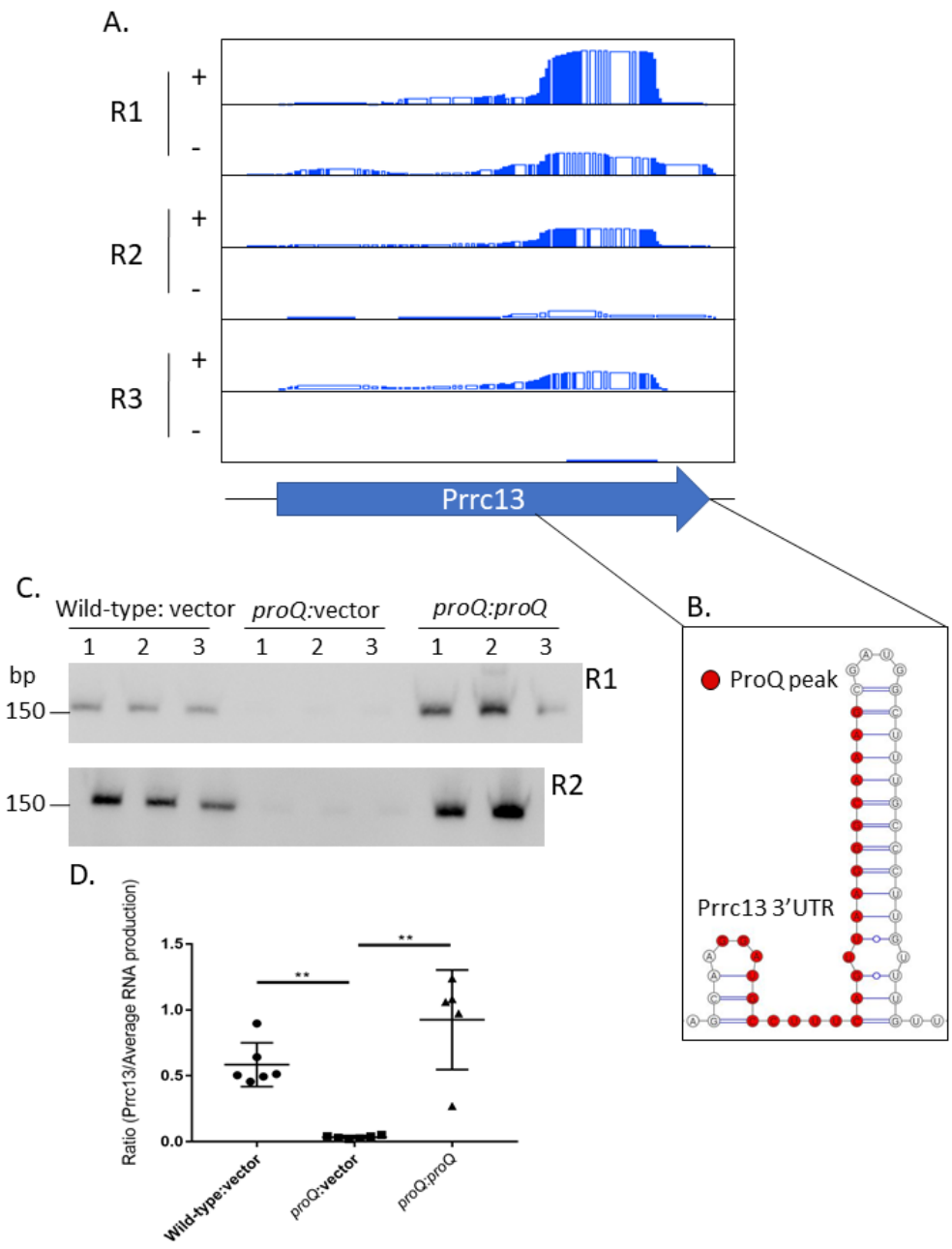

Figure 3.10. A. RNA-seq read coverage over the Prrc13 transcript following co-immunoprecipitation (biological triplicate) of RNA samples from strains expressing either tagged-ProQ (+) or untagged-ProQ (-) (R1-R3). **B.** The secondary structure of the 3' UTR of Prrc13 as predicted using RNAfold. The location of ProQ binding is indicated (red circles). **C.** Northern blot detection of Prrc13 (150 bp). RNA was isolated from the wild-type P. multocida VP161 containing empty vector (AL3356), the proQ mutant containing empty vector (AL3358) and the proQ complementation/overexpression strain (AL3357). **D.** Densitometry performed on Northern blots. Prrc13 expression is normalised to the average RNA production. Data shown are mean ± SD, n= 5-6, ** = $p < 0.005$ as per a Mann-Whitney test.

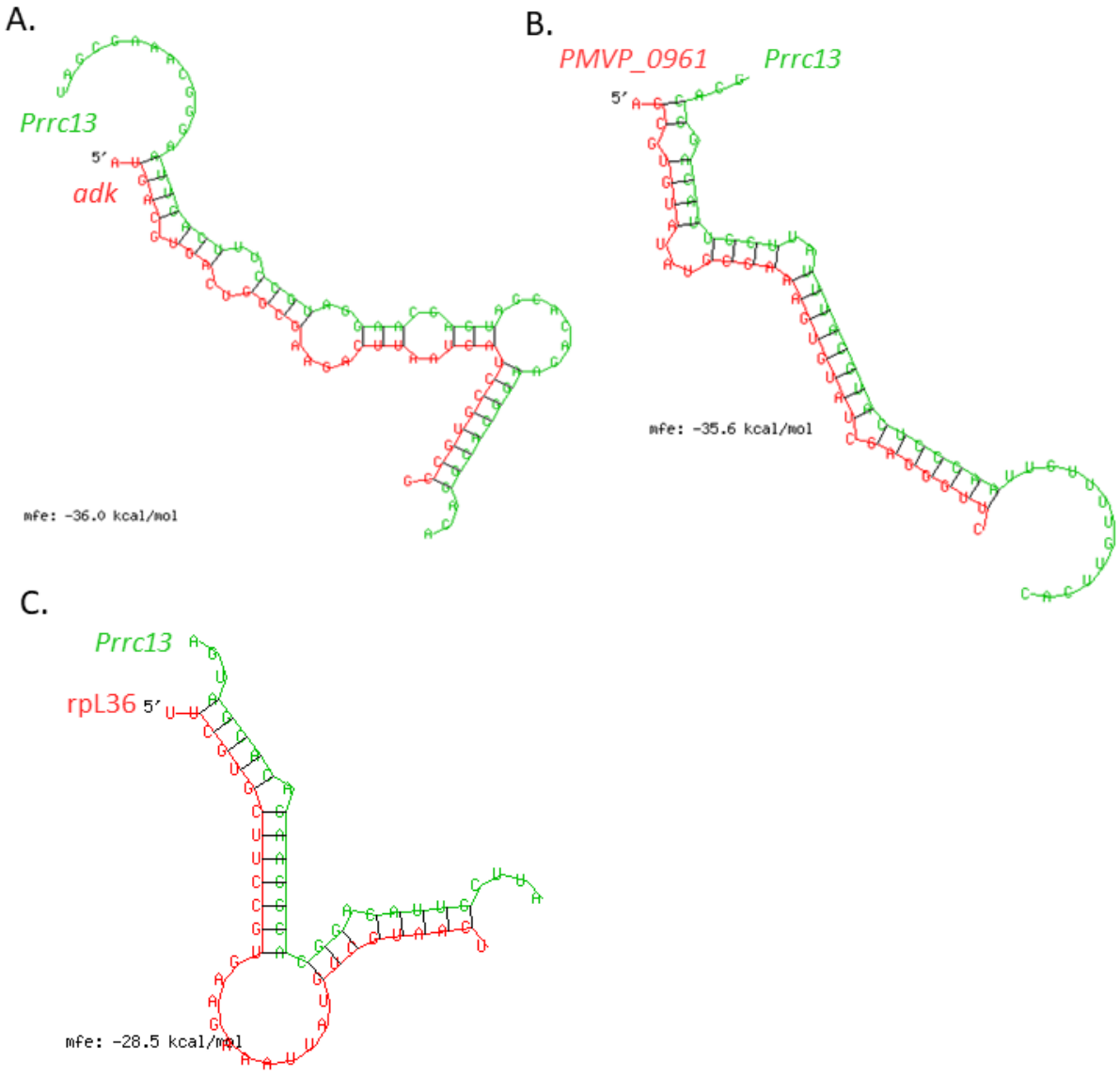

Figure 3.11. Putative binding interactions for RNA-RNA hybrids containing Prrc13 (green) and either **A.** *adk* (red) **B.** PMVP_ 0961(red) or **C.** rpL36 (red) formed during UV-CLASH. Putative interactions were visualised using BiBiServ2: RNAhybrid.

3.4 Discussion

This study used multiple high-throughput methods, including whole-cell proteomics, transcriptomics and UV-CLASH to identify that ProQ in *P. multocida* is an RNA chaperone protein that acts to regulate the expression of many transcripts, in some cases via direct binding to, and stabilization of, target RNA species. To determine the function of ProQ in *P. multocida,* a *proQ* mutant and a complemented/overexpression *proQ*:*proQ* strain were compared to the wild-type parent strain, VP161, using a wide range of assays. Unlike the ProQ from *E. coli, P. multocida* ProQ does not appear to be essential for normal osmoregulation as there was no difference in growth between the wild-type and *proQ* mutant strains of *P. multocida* in media containing 300 mM NaCl (Figure 3.1B and 1C). In *E. coli*, the role of ProQ in osmoregulation is to bind to the *proP* mRNA and the ribosome during translation and increase the production of a major facilitator superfamily (MFS) transporter ProP (Kunte et al. 1999). ProP is an osmosensor and osmoregulator that acts to transport osmolytes into the cell. When ProP is non-functional, or available only in low amounts, the cell loses the ability to osmoregulate appropriately and displays defects in growth (Kunte et al. 1999). It is important to note that within the available genomes for *P. multocida*, including the genome of strain VP161, there is no annotated gene that shares significant identity with known *proP* genes. The *P. multocida* VP161 gene encoding a product with the closest identity to the *E. coli* ProP is PMVP_1282 (30% amino acid identity). Bioinformatic analysis indicates that PMVP_1282 encodes a putative major facilitator superfamily (MFS) transporter but shares more identity (51% amino acid identity) with another *E. coli* MFS protein, YhjE. PMVP_1282 had decreased expression in the *proQ* mutant strain with empty vector compared to the wild-type (Log_2 fold-change = -1.41, FDR = 0.019); however, the data presented here suggest that it is highly unlikely that *P. multocida* ProQ controls bacterial survival under osmotic stress.

Studies focusing on the effects of *proQ* mutation on the levels of ProP within *E. coli* noted that mutation of the *proQ* gene had an off-target effect on the downstream gene *prc,* as the *prc* transcript start site was located within the *proQ* gene (Kerr et al. 2014). In *P. multocida*, the *prc* transcript also initiates within the *proQ* coding sequence but the the intron insertion site in the *proQ* mutant constructed in this study was a significant distance upstream of the prc transcript start site and therefore the position of the intron should not have directly affected *prc* transcript initiation. Despite this, proteomic and transcriptomic analyses indicated that inactivation of *proQ* resulted in both decreased *prc* transcript expression and Prc protein production, and complementation with intact *proQ* increased *prc* transcript levels. However, subsequent qRT-PCR analysis failed to detect any reduction in *prc* transcript levels in the *proQ* mutant; but an increase in *prc* expression was observed following qRT-PCR analysis of the *proQ* mutant

complemented with intact *proQ*. Together, these data suggest that ProQ may be having a direct and positive effect on *prc* expression. Furthermore, overexpression of *prc* in a *proQ* mutant background led to increased transcription of *proQ*. These data indicate that ProQ and Prc act to regulate each other; however, the exact mechanism of regulation was not determined. Examination of the UV-CLASH data identified no interactions between ProQ and the *prc* transcript, but it is possible that regulation may be occurring at the level of protein-DNA binding or may be indirect via a second regulatory molecule. To further characterize this interaction, more binding analyses could be performed to establish if the Prc protein is directly interacting with the ProQ protein or *proQ* transcript. The binding analyses could include EMSA, biolayer interferometry (BLitz) or surface plasmon resonance (SPR) analyses, where the ProQ and/or Prc proteins could be mixed with the *prc* or *proQ* transcripts and differences in gel shift or refracted light from the bound molecules could be observed.

Prc is a C-terminal peptidase that hydrolyses the peptide bond at the C-termini of proteins to allow for further processing. Prc-specific targets appear to differ between bacterial species (Bandara et al. 2005). An example is the cleavage of the 11 terminal residues from the precursor penicillin-binding protein 3 (PBP3) in *E. coli* (Keiler et al. 1995). In *P. multocida* strain VP161, Prc is predicted to be involved in cell wall/membrane biogenesis (based on COGs grouping), and the gene encoding Prc was amongst the nine (*hssB, hexB*, PM1289, *ycbB*, *prc*, *mepM*, *mltB*, *wbjD*, and *petL*) predicted to encode proteins involved in cell wall/membrane biogenesis that were differentially expressed in the *P. multocida proQ* mutant. Although HexB is involved in capsule transport (Chung et al. 1998) and PetL and PcgC are involved in LPS modification (May et al. 2001; Harper et al. 2017), there was no observable difference in capsule or LPS production between the wild-type and the *proQ* mutant strains. This is perhaps not unexpected as there are a large number of other enzymatic reactions and steps involved in the production of each of these structures and there are likely other methods of regulation involved. Indeed, the global regulator Fis has already been identified as important for capsule production, as has the RNA chaperone, Hfq (Steen et al. 2010).

Several high-throughput methods were used to identify transcripts that may bind to, and be regulated by, ProQ. Two different transcriptomics analyses were performed. The first analysis directly compared the *proQ* mutant with the wild-type strain. The second analysis compared expression of three strains; the wild-type strain containing the empty vector, the *proQ* mutant containing either empty vector, and the *proQ* mutant containing a plasmid constitutively expressing a recombinant version of *proQ* (*proQ* overexpression strain). In addition, comparative proteomics analysis of the *proQ* mutant and the wild-

type strain was undertaken, and UV-CLASH was used to identify ProQ-RNA binding interactions. Across both transcriptomic analyses there were a total of 43 transcripts with increased expression in the *proQ* mutant strains compared to the wild-type strains, whilst there were 104 transcripts with decreased expression. This indicates that in general ProQ may be acting more as a stabilizer of transcripts in *P. multocida* than a facilitator of degradation, however this stabilisation may be due to secondary effects of ProQ loss. These combined analyses identified a total of 17 predicted sRNAs as differentially expressed, of which five were identified in more than one analysis, namely Prrc10, Prrc12, Prrc13, Prrc25 and Prrc32. Of these Prrc10 and Prrc13 were identified in at least one of the transcriptomic analyses and both bound to ProQ and/or formed hybrids with other RNA species as determined using UV-CLASH. The remaining sRNAs were identified in both sets of transcriptomic analyses but were not preferentially co-immunoprecipitated with ProQ in the UV-CLASH experiment. Interestingly, of the sRNAs identified, almost all appeared to be stabilized by ProQ as they showed decreased transcript abundance in the *proQ* mutant strains. Although, when looking at all of the differentially expressed transcripts, the data were not fully consistent across the datasets as only 13 transcripts showed the same increased or decreased expression across both transcriptomic experiments and only one of the 13 also showed a corresponding change in protein production as detected using proteomics. However, it should be recognized that each data set has its own limitations. For example, while only a small number of the sRNAs were shown to be co-immunoprecipitated by ProQ, it is important to note that many RNA-protein interactions are transient and may not have been captured during the UV-CLASH (Soper et al. 2011). Nevertheless, the discrepancies between data sets may indicate that ProQ is not directly regulating all of these transcripts, including the above sRNAs.

The high-throughput data analyses identified that the expression of 26 tRNA species was affected by the abundance of ProQ. Of these, two were found across all four data sets, namely PMVP_1065 and PMVP_1782, which encode cystine and tryptophan tRNAs, respectively. The remaining 24 tRNAs were found in at least two datasets with the majority (14) identified as differentially expressed in the RNA-seq/ transcriptomic analyses; one (tRNA His-PMVP_0439) with increased expression in the *proQ* mutant across both experiments, and 13 with increased expression in the first analysis (*proQ* mutant vs wild-type), but decreased expression in the second (*proQ* mutant with empty vector vs wild-type with empty vector). This may be due to the difference in RNA library preparation as the initial analysis was performed on a library prepared using a kit that generated non-stranded data, whereas the second analysis was performed using a kit that generated strand-specific data. It would be interesting to compare the output of both kits using the same RNA samples to determine if this discrepancy is an artefact of the library preparation. Of the 13

tRNAs that showed different expression across the two transcriptomics experiments, seven were also identified by UV-CLASH as interacting with ProQ, namely, tRNA Val (PMVP_0626), tRNA Leu (PMVP_1009), tRNA Cys (PMVP_1065), tRNA Ser (PMVP_1077), tRNA Trp (PMVP_1782), tRNA Gly (PMVP_1805) and tRNA Arg (PMVP_1984). It has been suggested that interactions between an RNA binding protein such as ProQ and tRNA may be an artefact of tRNAs being highly abundant in the cell (Waters et al. 2017). However, our data show that some of the tRNAs species listed above were expressed at relatively low levels, such as tRNA Cys (PMVP_0626) with only 4-26 transcripts per million (tpm), and tRNA Trp (PMVP_1782) with 4-20 tpm. This indicates that ProQ may bind to these tRNAs in a *bona fide* manner to alter their abundance, but conflicting RNA-seq measurements make it difficult to determine if ProQ is stabilizing or degrading these tRNA species. To determine the true interaction between *P. multocida* ProQ and these tRNAs, a rifampicin RNA degradation assay could be performed. Rifampicin blocks transcription, allowing for the direct determination of transcript stability. Such an experiment would compare the levels of selected tRNAs in the wild-type, *proQ* mutant and *proQ* overexpression strain at given timepoints following rifampicin treatment. Common interaction events between ProQ and tRNAs were not evident in *Salmonella*; only three tRNAs were identified by other co-immunoprecipitation methods (CLIP-seq) as binding to ProQ and the predominant RNA species that bound to ProQ was 3' UTR transcripts (Holmqvist et al. 2018). However, in *E. coli*, fragments of RNA cleaved from pre-tRNA during their maturation into functional tRNA have been shown to form new sRNA molecules (Lalaouna et al. 2015). The formation of sRNAs from tRNA fragments may be occurring in *P. multocida* and ProQ may be the chaperone involved in this form of sRNA regulation. Together, our data suggest that there could be some real ProQ-tRNA interactions but with the limited current data we can not rule out the possibility that the identified interactions are artefactual.

A group of 25 transcripts that encode ribosomal proteins were identified as being either differentially expressed or bound by ProQ. However, unlike the data for sRNAs and tRNAs, there was minimal overlap between the data generated by the two different types of experiments. Two of these ribosomal-encoding transcripts, rpL32 (PMVP_1964) and rpL19 (PMVP_1322), were identified as differentially expressed in the RNA-seq data and rpL32 (PMVP_1964) was also identified as interacting with ProQ in the UV-CLASH. Unlike the majority of transcripts that encode ribosomal proteins, the rpL32 (PMVP_1964) transcript was not highly expressed (~10-20 tpm) which gives some confidence as to the likelihood that the ProQ-rpL32 interaction may be genuine. It is known that transcripts that are highly expressed may be non-specifically co-immunoprecipitated/bound to RNA chaperone proteins in UV-CLASH experiments and because of this they have often been excluded from further analysis (Waters et al. 2017). It is therefore likely that many

of the predicted *P. multocida* ProQ interactions with ribosomal-encoding RNAs are artefactual but further work is needed to confirm this. Such analyses could include EMSA or BLitz to determine if a direct interaction between ProQ and the ribosomal RNA species occurs.

While sRNAs and tRNAs were highly enriched in multiple datasets, there was low correlation seen between datasets for a number of the other transcripts. Indeed, only a few transcripts, including *prc*, Prrc13 and PMVP_0063, were observed to be differentially expressed in more than two datasets. A number of different factors might contribute to the lack of correlation between the different experiments. Firstly, each experiment measures slightly different components, proteomics measures final protein amounts and transcriptomics measures only RNA abundance but there will only be close correlation between these two experiments when there is little post-transcriptional regulation. RNA binding proteins often act post-transcriptionally, giving a plausible explanation as to why transcript and corresponding protein abundance were in general poorly correlated. Furthermore, the UV-CLASH experiments measure direct RNA-protein binding but evidence of some interaction, or loss of this interaction, does not necessarily equate to a change in the final amount of transcript or protein and some interactions may not even be captured as many will be transient.

It is also highly feasible that ProQ regulates the expression/production of other regulatory proteins, which in turn may affect the transcript or protein levels from genes under their direct control. Indeed, the gene encoding a major sRNA chaperone Hfq showed decreased expression in the *proQ* mutant strain (AL2973) compared to the wild-type strain VP161 (log_2-fold change = -1.9) and the *hfq* transcript was observed to bind ProQ in the UV-CLASH experiments; however, there were more reads bound to the untagged-ProQ that the tagged-ProQ, so this was not recognised as significant. Similarly, our data indicate that ProQ is involved in the expression of six sigma factor-associated genes, including *rpoE*, *mclA*, *rpoH*, *rsbE*, PM0209, and PMVP_2522. Two of the genes are located next to each other on the VP161 genome and are co-transcribed; *rpoE* which encodes the RpoE sigma factor, followed by *mclA* which encodes the cognate negative regulator of RpoE. Transcripts from both genes bound to ProQ and showed increased expression in the *proQ* overexpression strain compared to the wild-type. Thus, ProQ is likely to control the expression/production of other *P. multocida* regulatory proteins and we predict that the regulatory networks are complex and sensitive to precise growth conditions.

Our data clearly show that ProQ plays a crucial role in the regulation of PMVP_0063 and Prrc13. PMVP_0063 is a hypothetical protein that, based on bioinformatics analysis of available *P. multocida* genomes, is only present on the genome of strain VP161 and a small number of other *P. multocida*

isolates. Whole proteomics analyses detected the PMVP_0063 protein was produced during growth *in vitro* in rich medium (2 peptides) and modelling of the predicted protein (SWISS-model) indicates that part of the protein shares similarity with a Snf7 domain, which has previously only been identified in eukaryotic proteins. These Snf7 family proteins play a role in protein sorting and protein transport between lysosomes and the cytosol (Howard et al. 2001). The ProQ protein was shown to bind the PMVP_0063 transcript at a predicted central stem-loop region, and RNA-seq analysis and preliminary Northern blotting experiments indicated that ProQ binding stabilized the transcript. Two hybridizing transcripts were detected by Northern blotting, one correlating with the predicted size of the PMVP_0063 transcript, and a second slightly longer transcript that may correspond to a primary transcript. It is possible that the longer/initial transcript had an extended upstream region that is cleaved (not necessarily by ProQ) though it is not known what role it may play. To determine if the larger product does correspond to an initial transcript as described, further Northern blots could be performed with a probe that anneals to a sequence within the putative extended 5' region of the transcript. If the extended transcript is detected but not the shorter one, it could be concluded that larger transcript contains PMVP_0063 with 5' untranslated region, and that this extension is removed to form the active PMVP_0063 transcript. It would also be of interest to use 5' RACE to confirm the native 5' transcriptional site and to identify any internally processed 5' ends.

Analysis of the UV-CLASH data indicated that the PMVP_0063 transcript bound to other transcripts encoding an essential ATP-dependent zinc-metalloprotease (*ftsH*), a ribosomal protein alanine N-acetyl transferase (*rimI*), PfhB2 filamentous haemagglutinin (PMVP_0006), the ribonuclear protein RNaseP (PMVP_2505) and a previously unannotated RNA product, misc_e (Table 3.8). The binding of PMVP_0063 to multiple transcripts in the UV-CLASH experiments indicate that PMVP_0063 may encode a sRNA that regulates their expression; however, the corresponding proteins/transcripts were not identified as differentially expressed in the RNA-seq or proteomics experiments.

Prrc13 is a novel sRNA that was first identified as a highly expressed intergenic transcript located between *ner* (PMVP_0390) and PMVP_0391 (Marianne Mégroz, Boyce Laboratory, unpublished). Recently, it has been shown that Prrc13 does not bind to the most well-characterised RNA chaperone protein Hfq (Marianne Mégroz, Boyce Laboratory, unpublished) and the data presented here show ProQ interacts with a region in Prrc13 that corresponds to the beginning of a predicted 3' terminator stem-loop. The binding of ProQ to the terminator region of transcripts has been observed for ProQ produced by *Salmonella* and *E. coli* and in these species generally leads to stabilization of the transcript (Holmqvist et

al. 2018). Similarly, in our study we showed that the *P. multocida* ProQ stabilized the Prrc13 sRNA, as in the absence of ProQ the Prrc13 transcript had significantly reduced abundance and this was restored by providing the *proQ* mutant with an intact copy of *proQ* in trans. In *Salmonella,* ProQ was shown to protect RNA transcripts by blocking the action of RNaseIII (Holmqvist et al. 2018). The data presented here suggest that the *P. multocida* ProQ is acting in a similar manner to protect Prrc13 from degradation, although it is possible that ProQ acts to increase transcription of Prrc13. Rifampicin RNA degradation assays would help clarify the situation by confirming that the Prrc13 transcript is being specifically degraded in the absence of ProQ. Prrc13 is highly likely to function as an sRNA as, in the presence of ProQ, it bound to three other transcripts encoding a serine tRNA (PMVP_0961), a ribosomal protein (rpL36) and an adenylate kinase (*adk*) that converts ATP to AMP within cells and is essential in *P. multocida* strain VP161 (Thomas Smallman, Boyce Laboratory, unpublished). It is worth nothing that rpL36 is highly expressed and the co-immunoprecipitation of this RNA could be an artefact of high expression rather than representing a true ProQ-Prrc13 interaction. As observed with PMVP_0063 binding to its RNA targets, there was no corresponding change in expression of the putative Prrc13 targets when the levels of Prrc13 were altered. This could indicate that there are multiple layers of regulation that control the expression of these transcripts.

To confirm direct interactions between each of Prrc13 and PMVP_0063 and their respective predicted target mRNA species, EMSA and/or the two-plasmid GFP reporter assay could be used (in a similar manner to the analysis of GcvB binding interactions detailed in Chapter 2). Both methods would identify if binding is occurring between the RNA species; EMSA analyses would detect any RNA-RNA binding between Prrc13 or PMVP_0063 and the target transcripts and the GFP reporter assay would be able to quantitatively assess if the binding of the sRNA to its mRNA target leads to reduced or increased protein production. Additionally, a *PMVP_0063* or *prrc13* mutant strain and overexpression strain could be constructed then subjected to transcriptomic and proteomic analyses to identify putative targets. Together these experiments would show if PMVP_0063 and Prrc13 act as sRNAs. Confirming the RNA sizes, secondary structure and other attributes would allow for a clearer understanding of each sRNA. Importantly, both Prrc13 and PMVP_0063 bound to transcripts that encoded essential proteins, *adk* and *ftsH* respectively. If the sRNA regulation of these transcripts was manipulated through the addition of DNA mimics, such as peptide-conjugated phosphorodiamidate morpholino oligomers (PPMOs), then it is possible that *P. multocida* viability could be affected. PPMOs are nuclease resistant, synthetic nucleotide mimics that bind irreversibly to targets. PPMO therapeutics have been used successfully in the past to target essential genes such as *acpP* in *Acinetobacter baumannii*, and also to increase the sensitivity of *A. baumannii* to

meropenem via targeting NDM-1 (Geller et al. 2013b; Sully et al. 2017). For our purposes PPMOs would act through mimicking sRNA binding, and block the production of these essential proteins, leading to cell death. Therefore, such inhibitors could form the basis of a new therapeutic for the treatment of *P. multocida* infection.

3.5 References

Attaiech L, Glover JN, Charpentier X. 2017. RNA chaperones step out of Hfq's shadow. *Trends Microbiol* **25**: 247-249.

Bandara AB, Sriranganathan N, Schurig GG, Boyle SM. 2005. Carboxyl-terminal protease regulates *Brucella suis* morphology in culture and persistence in macrophages and mice. *J Bacteriol* **187**: 5767-5775.

Bosch M, Garrido E, Llagostera M, Perez de Rozas AM, Badiola I, Barbe J. 2002. *Pasteurella multocida exbB, exbD* and *tonB* genes are physically linked but independently transcribed. *FEMS Microbiol Lett* **210**: 201-208.

Boyce JD, Adler B. 2006. How does *Pasteurella multocida* respond to the host environment? *Curr Opin Microbiol* **9**: 117-122.

Chaulk SG, Smith Frieday MN, Arthur DC, Culham DE, Edwards RA, Soo P, Frost LS, Keates RA, Glover JN, Wood JM. 2011. ProQ is an RNA chaperone that controls ProP levels in *Escherichia coli*. *Biochemistry* **50**: 3095-3106.

Chung JY, Wilkie I, Boyce JD, Townsend KM, Frost AJ, Ghoddusi M, Adler B. 2001. Role of capsule in the pathogenesis of fowl cholera caused by *Pasteurella multocida* serogroup A. *Infect Immun* **69**: 2487-2492.

Chung JY, Zhang Y, Adler B. 1998. The capsule biosynthetic locus of *Pasteurella multocida* A:1. *FEMS Microbiol Lett* **166**: 289-296.

Darty K, Denise A, Ponty Y. 2009. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**: 1974-1975.

Nowlan H. Freese, David C. Norris, and Ann E. Loraine. 2016. Integrated Genome Browser: Visual analytics platform for genomics. *Bioinformatics* **14**:2089-95.

Fuller TE, Kennedy MJ, Lowery DE. 2000. Identification of *Pasteurella multocida* virulence genes in a septicemic mouse model using signature-tagged mutagenesis. *Microb Pathog* **29**: 25-38.

Geller BL, Marshall-Batty K, Schnell FJ, McKnight MM, Iversen PL, Greenberg DE. 2013. Gene-silencing antisense oligomers inhibit *Acinetobacter* growth *in vitro* and *in vivo*. *J Infect Dis* **208**: 1553-1560.

Glaser M, Nulty W, Vagelos PR. 1975. Role of adenylate kinase in the regulation of macromolecular biosynthesis in a putative mutant of *Escherichia coli* defective in membrane phospholipid biosynthesis. *J. Bacteriol* **123**: 128-136.

Glover M, J. N., Chaulk SG, Edwards RA, Arthur D, Lu J, Frost LS. 2015. The FinO family of bacterial RNA chaperones. *Plasmid* **78**: 79-87.

Gonzalez G, Hardwick S, Maslen SL, Skehel JM, Holmqvist E, Vogel J, Bateman A, Luisi B, Broadhurst RW. 2017. Structure of the *Escherichia coli* ProQ RNA chaperone protein. *RNA* **23**: 696-711.

Gulliver EL, Wright A, Deveson Lucas D, Megroz M, Kleifeld O, Schittenhelm R, Powell D, Seemann T, Bulitta J, Harper M et al. 2018. Determination of the small RNA GcvB regulon in the Gram-negative bacterial pathogen *Pasteurella multocida* and identification of the GcvB seed binding region. *RNA*. **24**: 704-720.

Harper M, Boyce JD, Cox AD, St Michael F, Wilkie IW, Blackall PJ, Adler B. 2007a. *Pasteurella multocida* expresses two lipopolysaccharide glycoforms simultaneously, but only a single form is required for virulence: identification of two acceptor-specific heptosyl I transferases. *Infect Immun* **75**: 3885-3893.

Harper M, Cox A, St Michael F, Parnas H, Wilkie I, Blackall PJ, Adler B, Boyce JD. 2007b. Decoration of *Pasteurella multocida* lipopolysaccharide with phosphocholine is important for virulence. *J Bacteriol* **189**: 7384-7391.

Harper M, Cox AD, St Michael F, Wilkie IW, Boyce JD, Adler B. 2004. A heptosyltransferase mutant of *Pasteurella multocida* produces a truncated lipopolysaccharide structure and is attenuated in virulence. *Infect Immun* **72**: 3436-3443.

Harper M, St Michael F, John M, Vinogradov E, Steen JA, van Dorsten L, Steen JA, Turni C, Blackall PJ, Adler B et al. 2013. *Pasteurella multocida* Heddleston serovar 3 and 4 strains share a common lipopolysaccharide biosynthesis locus but display both inter- and intrastrain lipopolysaccharide heterogeneity. *J Bacteriol* **195**: 4854-4864.

Harper M, Wright A, St Michael F, Li J, Deveson Lucas D, Ford M, Adler B, Cox AD, Boyce JD. 2017. Characterization of two novel lipopolysaccharide phosphoethanolamine transferases in *Pasteurella multocida* and their role in resistance to cathelicidin-2. *Infect Immun* **85:** e00557-17.

Hatfaludi T, Al-Hasani K, Gong L, Boyce JD, Ford M, Wilkie IW, Quinsey N, Dunstone MA, Hoke DE, Adler B. 2012. Screening of 71 *P. multocida* proteins for protective efficacy in a fowl cholera infection model and characterization of the protective antigen PlpE. *PLoS One* **7**: e39973.

Holmqvist E, Li L, Bischler T, Barquist L, Vogel J. 2018. Global maps of ProQ binding *in vivo* reveal target recognition via RNA structure and stability control at mRNA 3' ends. *Mol Cell* **70**: 971-982.e976.

Holmqvist E, Wright PR, Li L, Bischler T, Barquist L, Reinhardt R, Backofen R, Vogel J. 2016. Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking *in vivo*. *Embo J* **35**: 991-1011.

Howard TL, Stauffer DR, Degnin CR, Hollenberg SM. 2001. CHMP1 functions as a member of a newly defined family of vesicle trafficking proteins. *J Cell Sci* **114**: 2395-2404.

Keiler KC, Silber KR, Downard KM, Papayannopoulos IA, Biemann K, Sauer RT. 1995. C-terminal specific protein degradation: activity and substrate specificity of the Tsp protease. *Protein Sci* **4**: 1507-1515.

Kerr CH, Culham DE, Marom D, Wood JM. 2014. Salinity-dependent impacts of ProQ, Prc, and Spr deficiencies on *Escherichia coli* cell structure. *J Bacteriol* **196**: 1286-1296.

Kunte HJ, Crane RA, Culham DE, Richmond D, Wood JM. 1999. Protein ProQ influences osmotic activation of compatible solute transporter ProP in *Escherichia coli* K-12. *J. Bacteriol* **181**: 1537-1543.

Lalaouna D, Carrier MC, Semsey S, Brouard JS, Wang J, Wade JT, Masse E. 2015. A 3' external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise. *Mol Cell* **58**: 393-405.

Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**: 26.

May BJ, Zhang Q, Li LL, Paustian ML, Whittam TS, Kapur V. 2001. Complete genomic sequence of *Pasteurella multocida*, Pm70. *PNAS* **98**: 3460-3465.

Mégroz M, Kleifeld O, Wright A, Powell D, Harrison P, Adler B, Harper M, Boyce JD. 2016. The RNA-binding chaperone Hfq is an important global regulator of gene expression in *Pasteurella multocida* and plays a crucial role in production of a number of virulence factors, including hyaluronic acid capsule. *Infect Immun* **84**: 1361-1370.

Ponting CP. 1997. Tudor domains in proteins that interact with RNA. *Trends Biochem Sci* **22**: 51-52.

Rio DC. 2014. Northern blots for small RNAs and microRNAs. *Cold Spring Harb Protoc* **2014**: 793-797.

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

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**(11):2498-504

Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TT, Hinton JC, Vogel J. 2008. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* **4**: e1000163.

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

Smirnov A, Forstner KU, Holmqvist E, Otto A, Gunster R, Becher D, Reinhardt R, Vogel J. 2016. Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *PNAS* **113**: 11591-11596.

Smirnov A, Wang C, Drewry LL, Vogel J. 2017. Molecular mechanism of mRNA repression *in trans* by a ProQ-dependent small RNA. *EMBO J* **36**: 1029-1045.

Smith MN, Kwok SC, Hodges RS, Wood JM. 2007. Structural and functional analysis of ProQ: an osmoregulatory protein of Escherichia coli. *Biochemistry* **46**: 3084-3095.

Soper TJ, Doxzen K, Woodson SA. 2011. Major role for mRNA binding and restructuring in sRNA recruitment by Hfq. *RNA* **17**: 1544-1550.

Steen JA, Steen JA, Harrison P, Seemann T, Wilkie I, Harper M, Adler B, Boyce JD. 2010. Fis is essential for capsule production in *Pasteurella multocida* and regulates expression of other important virulence factors. *PLoS Pathog* **6**: 1000750.

Sully EK, Geller BL, Li L, Moody CM, Bailey SM, Moore AL, Wong M, Nordmann P, Daly SM, Sturge CR et al. 2017. Peptide-conjugated phosphorodiamidate morpholino oligomer (PPMO) restores carbapenem susceptibility to NDM-1-positive pathogens *in vitro* and *in vivo*. *J. Antimicrob. Chemother.* **72**: 782-790.

Travis AJ, Moody J, Helwak A, Tollervey D, Kudla G. 2014 Hyb: abioinformatics pipeline for the analysis of CLASH (crosslinking, ligationand sequencing of hybrids) data. *Methods* **65**:263–273

Waters SA, McAteer SP, Kudla G, Pang I, Deshpande NP, Amos TG, Leong KW, Wilkins MR, Strugnell R, Gally DL et al. 2017. Small RNA interactome of pathogenic *E. coli* revealed through crosslinking of RNase E. *EMBO J* **36**: 374-387.

Wemmer M, Azmi I, West M, Davies B, Katzmann D, Odorizzi G. 2011. Bro1 binding to Snf7 regulates ESCRT-III membrane scission activity in yeast. *J Cell Biol* **192**: 295-306.

Wilkie IW, Grimes SE, O'Boyle D, Frost AJ. 2000. The virulence and protective efficacy for chickens of *Pasteurella multocida* administered by different routes. *Vet Microbiol* **72**: 57-68.

Wilkie IW, Harper M, Boyce JD, Adler B. 2012. *Pasteurella multocida*: diseases and pathogenesis. *Curr Top Microbiol Immunol* **361**: 1-22.

Xiao Y, Cai Y, Bommineni YR, Fernando SC, Prakash O, Gilliland SE, Zhang G. 2006. Identification and functional characterization of three chicken cathelicidins with potent antimicrobial activity. *J Biol Chem* **281**: 2858-2867.

Chapter 4

General discussion and future directions

Chapter 4: General discussion and future directions

The importance of riboregulation has now been demonstrated in a number of bacterial species, as has the role of specific sRNAs and RNA chaperone proteins (Gottesman and Storz 2011). Riboregulation has been shown to play crucial roles in the regulation of many different cellular processes that can impact cell survival and virulence (Jin et al. 2009; Pitman and Cho 2015). While riboregulation has been the subject of detailed characterisation in enteric organisms, such as *E. coli* and *Salmonella* ssp, there has been only limited characterisation of regulatory RNA molecules and RNA chaperone proteins in bacteria within the family Pasteurellaceae. Very recent data from within our laboratory have shown that the *P. multocida* sRNA chaperone protein Hfq plays an important role in the production of capsule and other virulence factors such as filamentous haemagglutinin (Chung et al. 2001; Mégroz et al. 2016). These initial data suggest that regulatory sRNAs are likely to be critical for controlling expression of *P. multocida* virulence factors. Thus, components of the riboregulatory system are potential targets for the design of novel anti-virulence therapeutics. This project aimed to explore riboregulation in *P. multocida*, in particular the sRNA species GcvB and the sRNA chaperone protein ProQ. Throughout this work, new techniques were developed, and significant insight gained to begin to unravel the *P. multocida* sRNA regulatory network and build a base of data on critical riboregulatory molecules which could be targets for subsequent drug development.

In Chapter 2, the role of the *P. multocida* sRNA species GcvB was examined. GcvB is well conserved across many Gram-negative bacterial species, including *E. coli* and *Salmonella* (Sharma et al. 2011). We showed that the *P. multocida* GcvB acts to primarily regulate the production of proteins involved in amino acid biosynthesis. In contrast, GcvB in other species regulates the production of proteins involved in amino acid transport (Sharma et al. 2011). Indeed, of the *P. multocida* GcvB targets that showed reduced production by global proteomics analyses, only eight of 46 were the same as known GcvB targets in *E. coli* and *Salmonella*. It is of interest that across many bacterial species GcvB has retained its primary function of controlling genes involved in amino acid biosynthesis and transport, but the specific target genes show significant cross-species diversity. Previous work on GcvB in other species has identified a conserved seed binding region of 5'-ACACAAC-3' located within target mRNAs to which GcvB binds via complementary base pairing (Sharma et al. 2011). The *P. multocida* GcvB seed region on the mRNA targets, identified through bioinformatic analyses, electrophoretic mobility shift assays (EMSA) and two-plasmid fluorescent GFP reporter systems, showed an extended consensus consisting of 5'-AACACAACAT-3'. Taken together, these analyses clearly showed that although the role of GcvB in *P. multocida* is similar to that in other species, there are many subtle differences that make the *P. multocida* GcvB and its regulon unique.

The *P. multocida gcvB* mutant strain did not show a growth defect compared to the wild-type strain when grown *in vitro* but it was not tested *in vivo* and future studies should include competition assays and/or direct virulence assays in laboratory mice or in chickens, a natural *P. multocida* host. For the competition experiments, similar numbers of wild-type and *gcvB* mutant bacteria would be inoculated into the intraperitoneal region of the mouse or chicken and blood samples would be taken at the terminal stage of disease and plated onto non-selective media. Colonies would then be patched onto selective media to determine the proportion of mutant to wild-type bacteria (Harper et al. 2003a). For determining direct virulence of the *gcvB* mutant compared to the wild-type, similar numbers of each strain would be used to separately inoculate groups of chickens or mice. The animals would then be closely observed for signs of disease and euthanised when they were incapable of survival. Virulence would be determined by the number of animals that survived the infection (Harper et al. 2007b). Assessing the role of GcvB in virulence or growth *in vivo* would allow us to determine if GcvB could be used to develop a therapeutic. Additionally, the bacterial RNA and protein expression levels during infection could be examined to determine the effects of GcvB during growth *in vivo*. Regardless, the techniques used to study GcvB can be used to explore other sRNAs and enable the identification of sRNAs, and their precise interacting sequences, that regulate the production of proteins involved in virulence or other processes that may be appropriate targets for therapeutics. Molecules could then be designed to appropriately target these virulence-associated sRNA-mRNA interactions. Such molecules could be either sRNA mimics (bind to the mRNA target at the normal sRNA binding site) or sRNA inhibitors (bind to the sRNA and inhibit sRNA binding at the normal site) depending on the type of action desired. These molecules would need to bind RNA with strong affinity, be able to traverse the bacterial membranes to gain access to the cytoplasm and be resistant to degradation. One such family of molecules that fulfil these criteria are peptide-conjugated phosphorodiamidate morpholino oligomers (PPMOs) (Summerton 2017) that are synthetic nucleic acid mimics, usually 8-10 nt in length, that have nuclease resistant backbones (Summerton 2017). Studies in *E. coli, Salmonella* and *A. baumannii* have used PPMOs successfully to treat bacterial infections in mouse models (Mitev et al. 2009; Geller et al. 2013a; Sully et al. 2017). In each case, the target for the PPMO was either an essential gene in the species, or a gene that enables antibiotic resistance, and binding of the PPMO inhibited translation of the mRNA. As yet, no studies have used PPMOs to target the sRNA molecules themselves, although these molecules would seem to be ideal targets as the sequences used to bind each target gene are critical for each sRNA to exert its regulatory affect. Thus, PPMOs can be designed to have the same seed/binding sequence as the sRNA allowing it to act like the sRNA to regulate mRNA targets, or the PPMO can have a seed/binding sequence similar to the mRNA targets of the chosen

sRNA so that it acts as an antisense molecule to the sRNA, preventing the sRNA binding to its mRNA targets. Once synthesised, each PPMO would be tested for its ability to bind to the targets using EMSA, and their ability to alter protein production could be examined using the fluorescent GFP reporter system developed in this work. If the PPMO was able to successfully alter protein production *in vivo*, it could then be tested in an animal infection model to determine if it was able to inhibit *P. multocida* growth *in vivo* or attenuate its virulence.

In Chapter 3 the role of the predicted *P. multocida* RNA chaperone protein, ProQ was examined. At the beginning of this study there was almost no data on the role of ProQ as an RNA chaperone. Indeed, there are currently only three published reports on the characterisation of ProQ-interacting sRNAs in *Salmonella* and *E. coli*. In order to comprehensively assess the role of ProQ in *P. multocida* and identify its global interacting targets, several high-throughput methods were employed; transcriptomics, proteomics, and UV-crosslinking and sequencing of hybrids (UV-CLASH). These experiments clearly showed that ProQ interacts with many RNA molecules and its confirmed stabilization of some targets confirms that *P. multocida* ProQ does indeed act as an RNA chaperone. ProQ was shown to have its own regulon of targets that were quite different from those that are known to bind to the well characterised sRNA chaperone protein, Hfq (Smirnov et al. 2016). The cross-linking analyses showed that the *P. multocida* ProQ bound primarily at the beginning of terminator stem-loop sequences on transcripts and the transcriptomic and proteomic analyses suggested than this binding mostly protected the transcripts from degradation. This ProQ-dependent target protection has also been observed in *Salmonella* and *E. coli* where the binding of ProQ to transcripts blocked degradation of the 3'-end by RNaseIII (Holmqvist et al. 2018). *P. multocida* ProQ was also shown to bind to many putative sRNAs and tRNAs and to regulate their abundance within the cell. Again, in most cases ProQ acted to stabilize sRNA transcripts but there was some conflicting data generated from the different high-throughput techniques used in this study. Confirmation of direct ProQ binding to tRNAs awaits further experimentation. From the initial, non-stranded RNA-seq analysis, comparing the wild-type *P. multocida* strain VP161 to the *proQ* mutant strain, the amount of many tRNAs was increased in the absence of ProQ. In contrast, the stranded RNA-seq analysis, which compared the wild-type and *proQ* mutant strains (both containing empty vector), indicated that tRNA levels were increased in the presence of ProQ. Further experiments, such as rifampicin degradation assays (Desnoyers and Masse 2012), should be performed to determine if the binding of ProQ to tRNA transcripts is enabling or blocking degradation or tRNA processing. These assays could be used to compare the relative amount of tRNA degradation in the wild-type *P. multocida* strain and the *proQ* mutant. Each strain would be grown and treated with rifampicin to stop protein production and RNA

samples taken at chosen time points. Northern blotting would then be used to determine the amount of each chosen tRNA in the samples at each point. The results of these experiments would show if the presence of ProQ is affecting the half-life of specific tRNA species and whether tRNA-ProQ interactions result in stabilization or destabilization.

During this study, two transcripts were identified, namely PMVP_0063 and Prrc13, that showed differential expression in the *P. multocida proQ* mutant and may act as ProQ-dependent sRNAs. PMVP_0063 is only found in the genomes of four species of *P. multocida*, including *P. multocida* strain VP161, and the encoded protein shares a small region of identity with proteins belonging to the Snf7 protein family. The Snf7 family of proteins are involved in sorting of other proteins into vacuoles in eukaryotic cells (Wemmer et al. 2011), but no homologs of these proteins have yet been characterised within prokaryotes. The PMVP_0063 transcript was clearly stabilized by the presence of ProQ, as confirmed by Northern blot analysis, and was shown by CLASH analysis to bind to several other transcripts within the cell, indicating that the PMVP_0063 transcript may be acting as a ProQ-dependent sRNA. There is now clear precedent for a transcript to both encode a functional protein and also act as a sRNA (Vanderpool et al. 2011). For example, in *E. coli* the SgrS transcript acts both as a sRNA and as the template for a functional protein (Wadler and Vanderpool 2007), in *Pseudomonas aeruginosa* the sRNA PhrS encodes a hypothetical protein (Sonnleitner et al. 2008), RNAIII in *Staphylococcus aureus* encodes δ-hemolysin (Williams and Harper 1947), and the SR1 sRNA in *Bacillus subtilis* encodes the SR1P protein (Gimpel et al. 2010). Analysis of available proteomics data clearly show that PMVP_0063 is translated, as peptides corresponding to this protein were identified in VP161. To determine if the PMVP_0063 transcript can also act as a sRNA it would firstly be important to determine the exact transcript length(s) associated with PMVP_0063, as this current study showed via Northern blotting that a PMVP_0063-specific probe detected two fragments (Chapter 3). Further Northern blotting targeting the potential 5' extension of the PMVP_0063 transcript would allow confirmation of whether there are indeed two different transcripts that span the PMVP_0063 gene, and if the longer transcript is extended at the 5' or 3' end. To differentiate whether the longer transcript is extended at the 5' or 3' end, different probes could be used in Northern blotting to define the regions of each transcript that hybridize uniquely. To more accurately identify the start and end points of each transcript 5' and 3' RACE could be used. For these experiments, RNA should be extracted from wild-type *P. multocida*, treated with calf intestinal phosphatase (CIP) and tobacco acid pyrophosphatase (TAP), which will ensure that only the unprocessed ends of RNA can be adapter ligated. These adapter ligated RNA molecules are used for reverse transcription so that only native unprocessed transcripts are synthesised into cDNA before the transcript

of interest is amplified and sequenced. If the longer transcript contains a 5' extension, and the shorter transcript is a processed product of this longer transcript, then only the native 5' end will be identified using this procedure. If the experiment is repeated without the TAP treatment, then the 5' ends of both transcripts should be identified; this would allow both native and processed ends to be confirmed.

Once the different transcripts corresponding to PMVP_0063 are accurately defined it would then be necessary to confirm any regulatory effect of these transcripts on putative interacting targets. The PMVP_0063 transcript was shown via CLASH to bind to several other transcripts within the cell that may represent direct mRNA targets of the putative PMVP_0063 sRNA. However, none of the potential targets showed transcript and/or associated protein abundance changes in the *proQ* mutant strain compared to wild-type. This lack of mRNA differential expression following loss of ProQ may be due to multiple layers of regulation occurring. This includes the possibility that ProQ regulates other, as yet unidentified, sRNAs that also bind to these transcripts with different effects. To determine if there is a direct relationship between PMVP_0063 and the putative mRNA targets, confirmation that there is direct binding between the PMVP_0063 transcript and these putative target mRNAs is required. This could be done using EMSA, BLitz and/or two-plasmid fluorescent GFP reporter assays in the presence and absence of ProQ; EMSA and the two plasmid GFP reporter assays were successfully used to confirm the GcvB-gltA interaction (Chapter 2). Furthermore, if direct interactions were confirmed, then specific interacting nucleotides could be identified using wild-type and mutated sequences using the two-plasmid reporter assay. It would be of particular interest to investigate the interaction between PMVP_0063 and *ftsH* as this gene encodes an ATP-dependant metalloprotease that in other bacteria is known to be required for correct cell division (Tomoyasu et al. 1995). Importantly, *ftsH* has recently been identified as an essential gene in *P. multocida* strain VP161 (Smallman-Boyce laboratory, unpublished data). If PMVP_0063 does genuinely interact with *ftsH*, and this interaction is essential for normal production of FtsH, then there is the possibility to exploit this by designing a specific PPMO that would interfere with this interaction. Finally, if the PMVP_0063 transcript was confirmed as a likely sRNA affecting expression of other mRNA targets, then a *P. multocida* PMVP_0063 mutant and/or overexpression strain could be analysed genetically (transcriptomics) and phenotypically (proteomics and specific phenotypic assays) to determine if these regulatory networks play a role in *P. multocida* growth or the production of virulence factors.

The role of the predicted sRNA Prrc13 should also be further examined. To identify the exact start of the Prrc13 transcript 5' RACE should be employed. The role of Prrc13 in regulation can be examined by constructing both a mutant and an overexpression strain of Prrc13, followed by phenotypic, proteomic

and transcriptomic analyses to identify the Prrc13 regulon. As observed for PMVP_0063, Prrc13 also showed interactions with other RNA species when bound to ProQ, giving further evidence that Prrc13 is indeed acting as a sRNA. Of the observed interactions, the binding between Prrc13 and *adk* is of interest as *adk* has also been identified as essential for growth in *P. multocida* strain VP161 (Smallman-Boyce laboratory, unpublished data). This interaction should be validated, using the previously described methods and as with PMVP_0063, there is the possibility of designing PPMOs to interfere with this interaction.

UV-CLASH identified a number of RNA molecules that interact with PMVP_0063 and Prrc13 but it is possible that other mRNA targets remain unidentified. Future transcriptomic and proteomic analyses of *P. multocida* strains with each putative sRNA inactivated may allow more targets to be identified. However, to get a comprehensive picture of which transcripts each sRNA is binding to, an alternative method called global small non-coding RNA target identification by ligation and sequencing (GRIL-seq) could be employed (Han et al. 2016). GRIL-seq involves co-expression of the sRNA of interest together with expression of a T4 RNA-ligase protein within the same strain. Importantly the sRNA being examined is tagged with a poly A tag to allow purification of the sRNA and any interacting transcripts. When the T4 RNA-ligase is expressed, it will ligate any two RNA species that are directly binding/interacting with each other (e.g. sRNA/mRNA pair), thus forming a stable RNA hybrid molecule (Han et al. 2016). Total RNA can then be extracted from the cells and the sRNA of interest, potentially bound to its mRNA targets, can be immunoprecipitated using poly-T tailed beads. The isolated RNA-hybrids can then be sequenced, allowing the mRNA targets to be identified (Han et al. 2016).

As GRIL-seq relies on already knowing the target sRNA, this method could be used to define global interacting targets for PMVP_0063 and Prrc13 if they were shown to function as sRNAs (as outlined above). However, it could not be used more generally to identify all interacting sRNA-mRNA pairs. To interrogate the entire sRNA-mRNA regulatory network in an unbiased manner another method called high-throughput GRIL-seq (Hi-GRIL-seq) could be used as it does not rely on previous knowledge of RNA chaperones (like the CLASH method used in this study) or sRNAs present within the species (Zhang et al. 2017). Hi-GRIL-seq involves expressing the T4 RNA-ligase within a wild-type cell followed by sequencing of total RNA isolated from the cell, with a focus on identifying RNA-RNA hybrids (Zhang et al. 2017). Hi-GRIL-seq can give a fully comprehensive snapshot of the whole RNA interactome within a given cell and can allow for the identification of several sRNAs and their target mRNA species (Zhang et al. 2017).

In summary, the aim of this project was to gain knowledge about the riboregulatory network within *P. multocida*, with the ultimate goal to identify specific sRNA-mRNA interactions, which would allow for a clearer understanding of how the bacteria regulates protein production, especially those proteins involved in *in vivo* growth and virulence. To this end, the work presented in this thesis involved detailed characterisation of the role of one sRNA and one RNA chaperone. Characterization of the GcvB sRNA identified the regulon of the sRNA and confirmed direct nucleotide interactions critical for GcvB action. Characterization of the RNA chaperone ProQ identified a large number of RNA species that interact with ProQ and identified putative novel sRNAs and their targets. Importantly, the role of ProQ in stabilizing two RNA molecules was confirmed by Northern blotting. The information and techniques that have been developed in this project have increased our understanding of gene regulation in *P. multocida* and may in the future allow for the development of therapeutics to combat *P. multocida* infections. In turn, the knowledge gained from studying this species may allow for the production of similar therapeutics targeting other serious Gram-negative bacterial infection.

References

Chung JY, Wilkie I, Boyce JD, Townsend KM, Frost AJ, Ghoddusi M, Adler B. 2001. Role of capsule in the pathogenesis of fowl cholera caused by *Pasteurella multocida* serogroup A. *Infect Immun* **69**: 2487-2492.

Desnoyers G, Masse E. 2012. Activity of small RNAs on the stability of targeted mRNAs *in vivo*. *Methods Mol Biol* **905**: 245-255.

Geller BL, Marshall-Batty K, Schnell FJ, McKnight MM, Iversen PL, Greenberg DE. 2013. Gene-silencing antisense oligomers inhibit *Acinetobacter* growth *in vitro* and *in vivo*. *J Infect Dis* **208**: 1553-1560.

Gimpel M, Heidrich N, Mader U, Krugel H, Brantl S. 2010. A dual-function sRNA from *B. subtilis*: SR1 acts as a peptide encoding mRNA on the *gapA* operon. *Mol Microbiol* **76**: 990-1009.

Gottesman S, Storz G. 2011. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol* **3**.

Han K, Tjaden B, Lory S. 2016. GRIL-seq provides a method for identifying direct targets of bacterial small regulatory RNA by *in vivo* proximity ligation. *Nat Microbiol* **2**: 16239.

Harper M, Boyce JD, Wilkie IW, Adler B. 2003. Signature-tagged mutagenesis of *Pasteurella multocida* identifies mutants displaying differential virulence characteristics in mice and chickens. *Infect Immun* **71**: 5440-5446.

Harper M, Cox A, St Michael F, Parnas H, Wilkie I, Blackall PJ, Adler B, Boyce JD. 2007. Decoration of *Pasteurella multocida* lipopolysaccharide with phosphocholine is important for virulence. *J Bacteriol* **189**: 7384-7391.

Holmqvist E, Li L, Bischler T, Barquist L, Vogel J. 2018. Global maps of ProQ binding *in vivo* reveal target recognition via RNA structure and stability control at mRNA 3' ends. *Mol Cell* **70**: 971-982.e976.

Jin Y, Watt RM, Danchin A, Huang JD. 2009. Small noncoding RNA GcvB is a novel regulator of acid resistance in *Escherichia coli*. *BMC Genomics* **10**: 1471-2164.

Mégroz M, Kleifeld O, Wright A, Powell D, Harrison P, Adler B, Harper M, Boyce JD. 2016. The RNA-binding chaperone Hfq is an important global regulator of gene expression in *Pasteurella multocida* and plays a crucial role in production of a number of virulence factors, including hyaluronic acid capsule. *Infect Immun* **84**: 1361-1370.

Mitev GM, Mellbye BL, Iversen PL, Geller BL. 2009. Inhibition of intracellular growth of *Salmonella enterica* serovar Typhimurium in tissue culture by antisense peptide-phosphorodiamidate morpholino oligomer. *Antimicrob Agents Chemother* **53**: 3700-3704.

Pitman S, Cho KH. 2015. The mechanisms of virulence regulation by small noncoding RNAs in low GC Gram-Positive pathogens. *Int J Mol Sci* **16**: 29797-29814.

Sharma CM, Papenfort K, Pernitzsch SR, Mollenkopf HJ, Hinton JC, Vogel J. 2011. Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol Microbiol* **81**: 1144-1165.

Smirnov A, Forstner KU, Holmqvist E, Otto A, Gunster R, Becher D, Reinhardt R, Vogel J. 2016. Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *PNAS* **113**: 11591-11596.

Sonnleitner E, Sorger-Domenigg T, Madej MJ, Findeiss S, Hackermuller J, Huttenhofer A, Stadler PF, Blasi U, Moll I. 2008. Detection of small RNAs in *Pseudomonas aeruginosa* by RNomics and structure-based bioinformatic tools. *Microbiology* **154**: 3175-3187.

Sully EK, Geller BL, Li L, Moody CM, Bailey SM, Moore AL, Wong M, Nordmann P, Daly SM, Sturge CR et al. 2017. Peptide-conjugated phosphorodiamidate morpholino oligomer (PPMO) restores carbapenem susceptibility to NDM-1-positive pathogens *in vitro* and *in vivo*. *J Antimicrob Chemother* **72**: 782-790.

Summerton JE. 2017. Invention and early history of morpholinos: from pipe dream to practical products. *Methods Mol Biol* **1565**: 1-15.

Tomoyasu T, Gamer J, Bukau B, Kanemori M, Mori H, Rutman AJ, Oppenheim AB, Yura T, Yamanaka K, Niki H et al. 1995. *Escherichia coli* FtsH is a membrane-bound, ATP-dependent protease which degrades the heat-shock transcription factor sigma 32. *EMBO J* **14**: 2551-2560.

Vanderpool CK, Balasubramanian D, Lloyd CR. 2011. Dual-function RNA regulators in bacteria. *Biochimie* **93**: 1943-1949.

Wadler CS, Vanderpool CK. 2007. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *PNAS* **104**: 20454-20459.

Wemmer M, Azmi I, West M, Davies B, Katzmann D, Odorizzi G. 2011. Bro1 binding to Snf7 regulates ESCRT-III membrane scission activity in yeast. *J Cell Biol* **192**: 295-306.

Williams RE, Harper GJ. 1947. Staphylococcal haemolysins on sheep-blood agar with evidence for a fourth haemolysin. *J Pathol Bacteriol* **59**: 69-78.

Zhang YF, Han K, Chandler CE, Tjaden B, Ernst RK, Lory S. 2017. Probing the sRNA regulatory landscape of *P. aeruginosa*: post-transcriptional control of determinants of pathogenicity and antibiotic susceptibility. *Mol Microbiol* **106**: 919-937.

Appendices

Appendix 1.

Determination of the small RNA GcvB regulon in the Gram-negative bacterial pathogen *Pasteurella multocida* and identification of the GcvB seed binding region

EMILY L. GULLIVER,[1] AMY WRIGHT,[1] DEANNA DEVESON LUCAS,[1] MARIANNE MÉGROZ,[1] ODED KLEIFELD,[2] RALF B. SCHITTENHELM,[2] DAVID R. POWELL,[3] TORSTEN SEEMANN,[1,4] JÜRGEN B. BULITTA,[5] MARINA HARPER,[1,6] and JOHN D. BOYCE[1,6]

[1]Infection and Immunity Program, Monash Biomedicine Discovery Institute and Department of Microbiology, Monash University, Clayton, Victoria 3800, Australia
[2]Monash Biomedical Proteomics Facility, Monash Biomedicine Discovery Institute and Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia
[3]Monash Bioinformatics Platform, Monash University, Clayton, Victoria 3800, Australia
[4]Victorian Life Sciences Computation Initiative, The University of Melbourne, Parkville, Victoria 3052, Australia
[5]Department of Pharmaceutics, College of Pharmacy, University of Florida, Orlando, Florida 32827, USA

ABSTRACT

Pasteurella multocida **is a Gram-negative bacterium responsible for many important animal diseases. While a number of *P. multocida* virulence factors have been identified, very little is known about how gene expression and protein production is regulated in this organism. Small RNA (sRNA) molecules are critical regulators that act by binding to specific mRNA targets, often in association with the RNA chaperone protein Hfq. In this study, transcriptomic analysis of the *P. multocida* strain VP161 revealed a putative sRNA with high identity to GcvB from *Escherichia coli* and *Salmonella enterica* serovar Typhimurium. High-throughput quantitative liquid proteomics was used to compare the proteomes of the *P. multocida* VP161 wild-type strain, a *gcvB* mutant, and a GcvB overexpression strain. These analyses identified 46 proteins that displayed significant differential production after inactivation of *gcvB*, 36 of which showed increased production. Of the 36 proteins that were repressed by GcvB, 27 were predicted to be involved in amino acid biosynthesis or transport. Bioinformatic analyses of putative *P. multocida* GcvB target mRNAs identified a strongly conserved 10 nucleotide consensus sequence, 5′-AACACAACAT-3′, with the central eight nucleotides identical to the seed binding region present within GcvB mRNA targets in *E. coli* and *S.* Typhimurium. Using a defined set of seed region mutants, together with a two-plasmid reporter system that allowed for quantification of sRNA–mRNA interactions, this sequence was confirmed to be critical for the binding of the *P. multocida* GcvB to the target mRNA, *gltA*.**

Keywords: GcvB; sRNA; *Pasteurella multocida*; Hfq; GltA

INTRODUCTION

Pasteurella multocida is a Gram-negative, coccobacillus that is the causative agent of many economically important diseases, including fowl cholera, swine atrophic rhinitis, hemorrhagic septicemia, and various respiratory diseases of ungulates (Wilkie et al. 2012). *P. multocida* produces several virulence factors that are critical for the bacterium to cause disease. These include primary virulence factors, such as the polysaccharide capsule, lipopolysaccharide (LPS), and filamentous hemagglutinin as well as virulence-associated factors, such as proteins involved in iron and nutrient acquisition (Fuller et al. 2000; Bosch et al. 2002; Harper et al. 2004; Boyce and Adler 2006). Appropriate regulation of these factors is likely critical for *P. multocida* survival. For example, during *P. multocida* in vivo growth, the bacteria must acquire and/or synthesize all necessary amino acids, many of which are not freely available in sufficient quantities (Boyce and Adler 2006). This requires the production of amino acid biosynthesis and transport proteins, the expression of which must be tightly regulated to ensure that there is a balance between energy input and expenditure.

[6]**These authors contributed equally to this work.**
Corresponding author: john.boyce@monash.edu
Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.063248.117.

Recently, we showed that the Hfq protein was essential for the appropriate expression of a range of proteins in the *P. multocida* serogroup A strain VP161, including those required for the biosynthesis of hyaluronic acid capsule which is a primary virulence factor (Mégroz et al. 2016). The Hfq protein is an RNA chaperone that directly interacts with particular small regulatory RNA (sRNA) molecules to facilitate their binding to specific mRNA targets. Noncoding sRNA molecules are generally 40–400 nucleotides (nt) long and regulate transcript/protein expression within bacteria by binding to target mRNA via complementary base pairing (Desnoyers et al. 2013). There is redundancy within the sRNA regulatory network, as one sRNA species may bind to many different mRNA targets and each mRNA target may be regulated by several sRNA species (Desnoyers et al. 2013). Depending on the type of interaction, the binding of a sRNA to a target mRNA may result in either inhibition or induction of protein production. The binding of the sRNA to the ribosome-binding site (RBS) of an mRNA target can block translation and therefore reduce protein production. Alternatively, sRNA binding can result in rapid mRNA degradation via induction of Ribonuclease E activity against double-stranded RNA (Gottesman and Storz 2011). Less commonly, protein production can be enhanced via the binding of the sRNA to a natural secondary structure region in the mRNA that normally acts to occlude the RBS. This sRNA–mRNA interaction leads to the unfolding of the secondary structure, allowing the ribosome greater access to the RBS in order to initiate translation (Gottesman and Storz 2011).

Comparative global transcriptomic and proteomic analyses of the *P. multocida* strain VP161 and an isogenic *hfq* mutant revealed that many genes displayed altered transcript expression, and/or altered protein production, when *hfq* was inactivated (Mégroz et al. 2016). Analysis of the transcriptional data also allowed for the identification of a number of intergenic regions encoding putative sRNAs (M Mégroz, unpubl.). One putative sRNA identified in strain VP161, which is also encoded on the Pm70 genome (GenBank AE004439.1, position 652175 to 651999), exhibited high sequence identity to the Hfq-dependent sRNA GcvB. In *E. coli* and *Salmonella enterica* serovar Typhimurium (*S.* Typhimurium) GcvB has been shown to negatively regulate the production of proteins involved in amino acid transport and biosynthesis, such as the amino acid transporters ArgT, BrnQ, DppA, OppA, SstT, TppB, and YaeC and the amino acid biosynthesis proteins GdhA, IlvC, IlvE, SerA, and ThrL (Pulvermacher et al. 2008; Sharma et al. 2011). In *E. coli*, the expression of GcvB is intimately associated with the availability of glycine, and GcvB expression is induced when nutrients, especially glycine, are abundant in the environment. The *gcvB* gene is adjacent to and transcribed divergently from *gcvA*, which encodes the GcvA protein that positively regulates both *gcvB* and the glycine cleavage operon *gcvTHP*. The activation of both the *gcvTHP* operon and *gcvB*, is repressed during growth in the absence of glycine due to the association between GcvA and the regulatory protein GcvR (Urbanowski et al. 2000). This interaction does not occur in the presence of glycine, leaving GcvA to act as an activator of *gcvB* and *gcvTHP* expression. Therefore, in *E. coli* and *S.* Typhimurium, during periods of low glycine abundance the decreased production of GcvB results in activation of the amino acid biosynthesis and transport proteins that are normally repressed by the GcvB sRNA (Urbanowski et al. 2000).

GcvB function has primarily been assessed in *E. coli* (Urbanowski et al. 2000; Pulvermacher et al. 2008; Coornaert et al. 2013), and *S.* Typhimurium (Sharma et al. 2011) with functional studies in other organisms limited to *Yersinia pestis* (McArthur et al. 2006). *S.* Typhimurium *gcvB* mutants grow more slowly than the wild-type parent strain and *E. coli gcvB* mutants have a decreased ability to form biofilms (Sharma et al. 2007; Mika and Hengge 2014). Analysis of the *E. coli* and *S.* Typhimurium GcvB mRNA targets has facilitated the identification of a GcvB binding sequence (seed region), 5′-CACAACAT-3′, that allows for base pairing between GcvB and its mRNA targets. The mRNA seed region is strongly conserved in the GcvB targets produced by both species (Sharma et al. 2007). The seed region sequence, 5′-AUGUUGUG-3′, is present in the GcvB expressed by both *S.* Typhimurium and *E. coli* (Sharma et al. 2011) and is the reverse complement of the seed region sequence present in the mRNA target molecules.

There is currently no information on the functional role of GcvB, its mRNA targets, or mRNA binding interactions in any organisms from the *Pasteurellaceae* family. A bioinformatics screen of multiple genomes identified a putative GcvB homolog in *P. multocida*, and a recent bioinformatics analysis of the related organism, *Actinobacillus pleuropneumoniae*, also identified a GcvB homolog and its expression was confirmed by northern blotting (Sharma et al. 2007; Rossi et al. 2016). Another study in *Haemophilus influenzae* showed expression of GcvB was high when grown in the presence of primary normal human bronchial epithelial cells using RNA-seq (Baddal et al. 2015). However, neither study looked further into the function of GcvB. In this study, we report the characterization of GcvB in a highly pathogenic *P. multocida* strain and the identification of more than 30 targets. Furthermore, we identify the *P. multocida* GcvB seed region and use a two-plasmid green fluorescent protein (GFP) reporter system to confirm the binding interaction between *P. multocida* GcvB and one of its mRNA targets, *gltA*.

RESULTS

Confirmation of GcvB expression in *P. multocida* using high-throughput transcriptomic analysis, northern blotting, and GcvB transcript analyses

Previous bioinformatics analyses (Sharma et al. 2007) have identified a putative GcvB in *P. multocida* that contains the

]

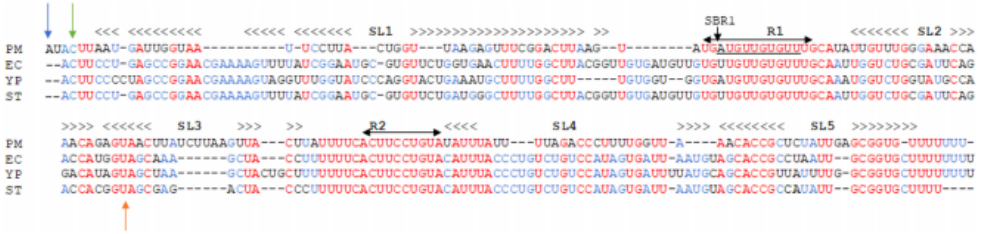

FIGURE 1. Nucleotide sequence alignment of GcvB from *P. multocida* (PM), *E. coli* (EC), *Y. pestis* (YP), and *S.* Typhimurium (ST). The previously identified *S.* Typhimurium GcvB stem–loop (SL) sequences (SL1-SL5), including their extent (< and >) and conserved R1 and R2 sequences (*horizontal* arrows) (Sharma et al. 2007), are shown *above* the alignment. Nucleotides in red are identical across all four GcvB sequences; nucleotides in blue are identical in three of the four GcvB sequences. The proposed *P. multocida* seed region is labeled SBR1 and underlined. The predicted *P. multocida* GcvB rho-independent terminator sequence is labeled SL5. The orange arrow designates the position of the TargeTron intron insertion site in the *P. multocida* VP161 *gcvB* mutant. The green arrow indicates the predicted transcript start site for *P. multocida* GcvB as determined by primer extension. The blue arrow indicates the predicted start site for *P. multocida* GcvB as determined by 5′ RACE.

conserved R1 and R2 sequences common to all GcvB sRNA molecules (Fig. 1). The *P. multocida gcvB* was located between *ivlE* (encoding a branched-chain amino acid aminotransferase) and *gcvA* (Fig. 2A). To determine if GcvB is expressed in *P. multocida,* we analyzed whole-transcriptome RNA-sequencing (RNA-seq) data generated from RNA isolated from *P. multocida* VP161 grown until the cultures reached an optical density at 600 nm (OD_{600}) of 0.2, 0.7, and 1.0, representing early-exponential, mid-exponential and late-exponential growth phases in biological duplicate. The putative *P. multocida* GcvB sRNA was expressed strongly during early-exponential and mid-exponential growth, with an average of 1490 and 2514 GcvB transcripts per million (TPM) total transcripts, respectively (Fig. 2A). However, GcvB expression was reduced significantly by late-exponential growth (10-fold reduced expression compared to early exponential phase and 13-fold reduced expression compared to mid-exponential phase; false discovery rate [FDR] < 0.01) when only limited amounts of GcvB transcripts were produced (average of 209 GcvB TPM). The growth phase-dependent expression of the GcvB sRNA in *P. multocida* strain VP161 was confirmed by northern blotting using a GcvB complementary strand-specific, RNA probe. The probe hybridized strongly with a fragment of the predicted size (~180 bp) of the GcvB sRNA transcript in the RNA isolated from VP161 cells in early-exponential growth phase, but only very weakly to RNA isolated from cells grown to late-exponential growth phase (Fig. 2B).

Analysis of the *P. multocida* GcvB sequence revealed a putative rho-independent transcriptional terminator that corresponded to stem–loop 5 (SL5) present in the GcvB of *E. coli* and *S.* Typhimurium (Fig. 1; Sharma et al. 2007). The position of this putative stem–loop corresponded closely with the end of the RNA-seq transcript peak (Fig. 2A) and we predict that this stem–loop defines the 3′ end of the *P. multocida* GcvB. To determine the 5′ end of the *gcvB* transcript, two independent methods were used, primer

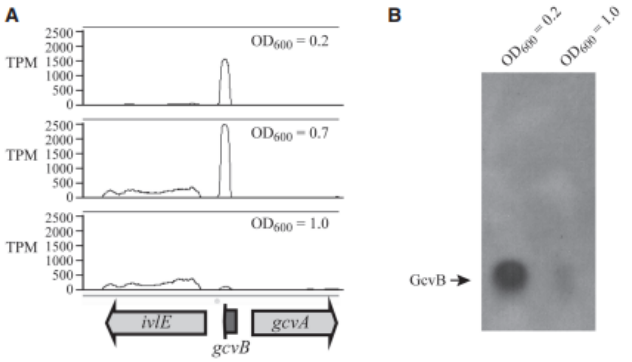

FIGURE 2. GcvB is expressed strongly at early- and mid-exponential growth phases but only weakly at late-exponential growth phase. (*A*) Number of mapped transcripts per million (TPM) total transcripts that map to the genomic regions surrounding *gcvB* as determined by whole-genome RNA-seq and as visualized in Artemis (Sanger) genome viewer. The *top* panel shows the number of mapped reads using RNA derived from early-exponential growth phase (OD_{600} = 0.2) cells, the *middle* panel shows the number of mapped reads using RNA derived from mid-exponential growth phase (OD_{600} = 0.7) cells, and the *bottom* panel shows the number of mapped reads using RNA derived from late-exponential growth phase (OD_{600} = 1.0) cells. The extent and orientation of the genes *ivlE*, *gcvB*, and *gcvA* are shown *below* the mapping graphs. (*B*) *P. multocida* RNA (8 μg/lane) isolated from early-exponential growth phase (OD_{600} = 0.2) cells and late-exponential growth phase (OD_{600} = 1.0) cells was used for northern blotting together with a DIG-labeled, single-stranded RNA probe representing the sequence complementary to GcvB. The position of GcvB is shown at the *left*. Image has been modified to increase contrast.

extension and 5′ RNA ligase-mediated rapid amplification of cDNA ends (5′ RLM-RACE). For primer extension experiments, RNA isolated from *P. multocida* VP161 was used as the template for cDNA synthesis with the fluorescently labeled primer BAP7962 or primer BAP8190 (Supplemental Table S1) that anneal ~90 bp and ~140 bp, respectively, from the predicted start of the *gcvB* transcript (as determined by the RNA-seq transcriptomic analyses). Fragment size analysis of the generated cDNA molecules identified a fragment of 87 nt in length for the primer extension using BAP7962 and 140 nt in length for primer extension using BAP8190. These data indicated that the *P. multocida* VP161 GcvB transcript started with the sequence 5′-CUUAAUG-3′, plus or minus the 5′ C, which corresponds to the second nucleotide in the GcvB sequence from *E. coli* and *S.* Typhimurium (Fig. 1). To determine if this was a bona fide transcript initiation site, we used 5′ RLM-RACE. *P. multocida* VP161 RNA was first treated with calf intestine alkaline phosphatase and tobacco acid pyrophosphatase and then used as the template in nested PCRs to generate 5′ adapter-ligated GcvB DNA fragments, which were then cloned into the plasmid pCR2.1. DNA sequencing of these cloned fragments using a vector-specific primer (BAP612) revealed that the *P. multocida* GcvB transcriptional start site was located 2 bp upstream of the *E. coli* and *S.* Typhimurium GcvB start sites (Fig. 1; Supplemental Fig. 1). Therefore, these data indicate that the *P. multocida* GcvB transcript begins with 5′-AUACUUAAUG-3′.

The secondary structure of the *P. multocida* GcvB was modeled (Fig. 3A) using the Mfold webserver (Zuker 2003). While the predicted structure is very similar to the experimentally determined structure of the *S.* Typhimurium GcvB (Sharma et al. 2007), there are some notable differences. These include the observation that the stem–loop 1 (SL1) in the *P. multocida* GcvB is predicted to be significantly shorter than the SL1 in the *S.* Typhimurium GcvB; sequence alignment of the GcvB sRNAs from *P. multocida*, *E. coli*, *S.* Typhimurium and *Y. pestis* confirmed that the 5′ region of the *P. multocida* GcvB is indeed shorter (Fig. 1). The predicted *P. multocida* GcvB structure contains the SL2 and SL3 stem–loops between the conserved R1 and R2 G/U-rich linker regions, as is observed in the *S.* Typhimurium GcvB. However, the *P. multocida* GcvB has no predicted SL4 stem–loop, but rather the region between SL1 and R1 shows high complementarity to the region between R2 and SL5 and may form a long double-stranded section, although this remains to be experimentally verified (Fig. 3A).

GcvB predominately regulates amino acid biosynthesis and transport proteins in *P. multocida*

In order to determine the GcvB regulon in *P. multocida*, a VP161 *gcvB* mutant (AL2677) (Supplemental Table S2) was constructed using TargeTron technology (Sigma-Aldrich). The intron insertion was located between nucleotides 92 and 93 of *gcvB* and within the predicted SL3 loop (Fig. 1,

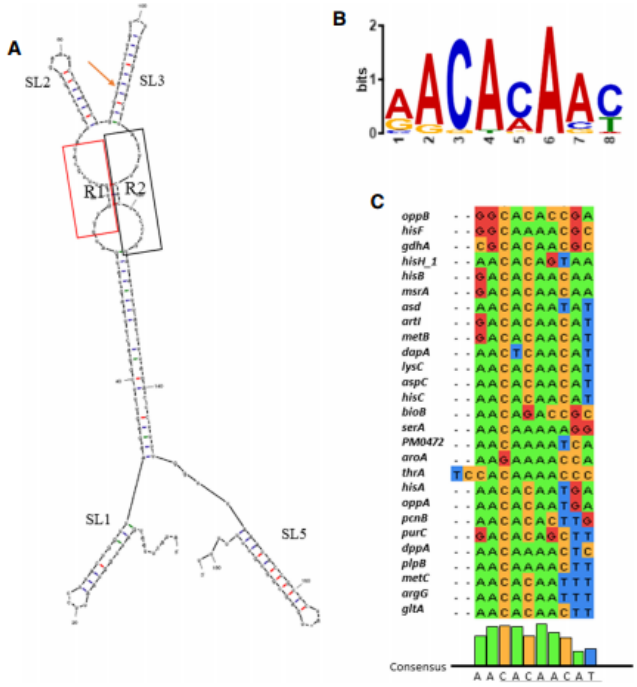

FIGURE 3. Predicted secondary structure of the *P. multocida* GcvB and the seed binding consensus motifs present in the 27 putative mRNA targets. (*A*) Putative secondary structure of the *P. multocida* GcvB sRNA molecule as predicted by Mfold. The conserved R1 (red) and R2 (black) sequences are boxed and the proposed SL1, SL2, SL3, and SL5 stem–loops are labeled. The position of the TargeTron intron in the *P. multocida gcvB* mutant is also shown (orange arrow). (*B*) Diagram of the GcvB mRNA target seed binding motif identified by MEME in 27 genes encoding putative GcvB mRNA targets. The letter height indicates the frequency of each base at each position. (*C*) Sequence alignment of the 27 putative seed binding regions found by MEME motif finder in genes encoding the predicted GcvB mRNA targets. Nucleotides are highlighted with color as follows to show the level of conservation: (A) green; (C) yellow; (T) blue; and (G) red. The *P. multocida* GcvB consensus sequence is shown *beneath* the alignment, with the *E. coli* and *S. Typhimurium* core GcvB-mRNA seed binding sequence underlined (Sharma et al. 2007).

Fig. 3A). To complement the mutation, the wild-type VP161 *gcvB* gene, together with its putative native promoter, was cloned into the *P. multocida* plasmid pPBA1100s to generate the plasmid pAL1190 (Supplemental Table S2). This plasmid was used to transform the *P. multocida gcvB* mutant AL2677, producing the strain AL2864 (Supplemental Table S2). As a control, pPBA1100s empty vector (Supplemental Table S2) was also used to transform the *gcvB* mutant, generating the strain AL2862.

The level of *gcvB* expression in these strains was determined using qRT-PCR with all expression levels normalized to the expression of the housekeeping gene *gyrB* (Fig. 4). The levels of *gcvB* expression in the wild-type strain (normalized to *gyrB*) were 0.24 ± 0.04 ($n = 3$, SEM) and 0.13 ± 0.03 at early- and mid-exponential growth phases, respectively. As expected no expression of *gcvB* was measured in the *gcvB* mutant at either growth phase, as the primers used for the qRT-PCR spanned the point of the relatively large intron insertion. Surprisingly, the levels of *gcvB* expression in AL2864 (*gcvB* mutant provided with an intact copy of *gcvB* on the plasmid pAL1190) were 16.4 ± 2.5 and 37.6 ± 16.4, indicating a 69-fold increase in *gcvB* expression at early-exponential growth phase and a 289-fold increase at mid-exponential growth phase compared to expression in the wild-type strain. Thus, providing the *gcvB* mutant with functional *gcvB* in *trans* resulted in the significant overexpression of GcvB at both growth phases tested. Accordingly, the strain AL2864 was designated as a GcvB overexpression strain. It was predicted that that there would be an inverse relationship between the levels of expression of any GcvB-regulated genes in the *gcvB* overexpression strain and the levels of expression of the same genes in the *gcvB* deficient strains (*gcvB* mutant alone, or *gcvB* mutant containing empty vector). To test this, the strain was included in the proteomic analyses described below.

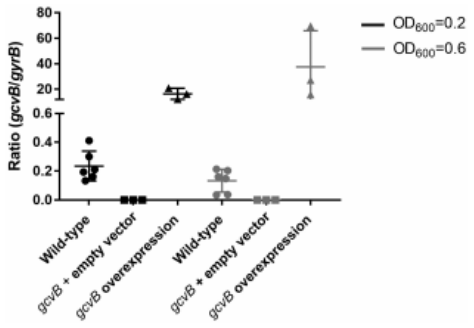

FIGURE 4. qRT-PCR analyzing the expression of GcvB in wild-type *P. multocida* expression (represented by circles), the *gcvB* mutant with empty vector (square symbols), and the GcvB overexpression strains (triangle symbols). RNA was isolated from early exponential growth phase cells ($OD_{600} = 0.2$) (black symbols) and mid-exponential growth phase cells ($OD_{600} = 0.6$) (gray symbols). Expression was standardized to the expression of the housekeeping gene *gyrB*. Thick *horizontal* bars represent the mean ± SD.

The survival and growth of the wild-type VP161, the GcvB-deficient strains (AL2677 and AL2862) and the GcvB overexpression strain (AL2864) was examined under several conditions. It was found that the *gcvB* mutant strains grew indistinguishably from the wild-type VP161 during growth in heart infusion (HI) broth (Fig. 5A,B). The *gcvB* overexpression strain had a similar exponential growth rate (doubling time of 41.1 ± 0.5 min) to the wild-type strain (doubling time of 36.2 ± 2.1 min); however, the lag-phase was increased by ~1.5 h (Fig. 5B). There was no difference in the ability of the *gcvB* mutant, AL2677, and the wild-type VP161 to form biofilms during static growth (Fig. 5C) and no difference in survival at low pH (HI broth, pH = 4.6 for 15 min) (Fig. 5D).

We then analyzed the protein expression profiles of the wild-type VP161, *gcvB* mutant (AL2677), *gcvB* mutant plus empty vector (AL2862), and GcvB overexpression strain (AL2864) using liquid chromatography coupled to tandem mass spectrometry (LC–MS/MS) in biological triplicate. The initial experiment compared wild-type VP161 and the *gcvB* mutant using isotopically labeled samples. The second experiment compared wild-type, *gcvB* mutant plus empty vector and GcvB overexpression strain using label-free proteomics. For all experiments, cells were harvested at early-exponential growth phase when *gcvB* is strongly expressed in the wild-type strain. In total, 1191 proteins were identified in the first experiment (using isotopic labeling) and 1540 proteins in the second (label-free); representing 57% and 74%, respectively, of the 2085 total proteins predicted to be encoded on the *P. multocida* VP161 genome (Boyce et al. 2012). Identified proteins were considered differentially expressed if they showed a ≥1.5-fold (≥0.59 $\log_2$) difference in production at an FDR of <0.05 compared to wild-type VP161. Overall 36 proteins were measured as showing increased production in either of the two *gcvB* mutant strains analyzed; 25 proteins in experiment 1, 28 in experiment 2 with 17 identified in both experiments (Supplemental Table S3). Only 10 proteins showed decreased production in either of the two *gcvB* mutant strains analyzed; two in experiment 1, eight in experiment 2 with none identified in both experiments (Supplemental Table S4). In contrast, 218 proteins with altered production levels were identified in the GcvB overexpression strain, 75 with increased production, and 143 with decreased production (Supplemental Tables S5, S6).

We then compared the lists of differentially produced proteins identified in each of the GcvB-deficient strains (*gcvB* mutant strain AL2677, and *gcvB* mutant plus empty vector, AL2862) and the GcvB overexpression strain. A total of 27 proteins showed significantly increased production in one of the GcvB-deficient strains as well as inverse (decreased) production in the *gcvB* overexpression strain (Table 1). Of these 27 proteins, 17 proteins (71%) displayed increased

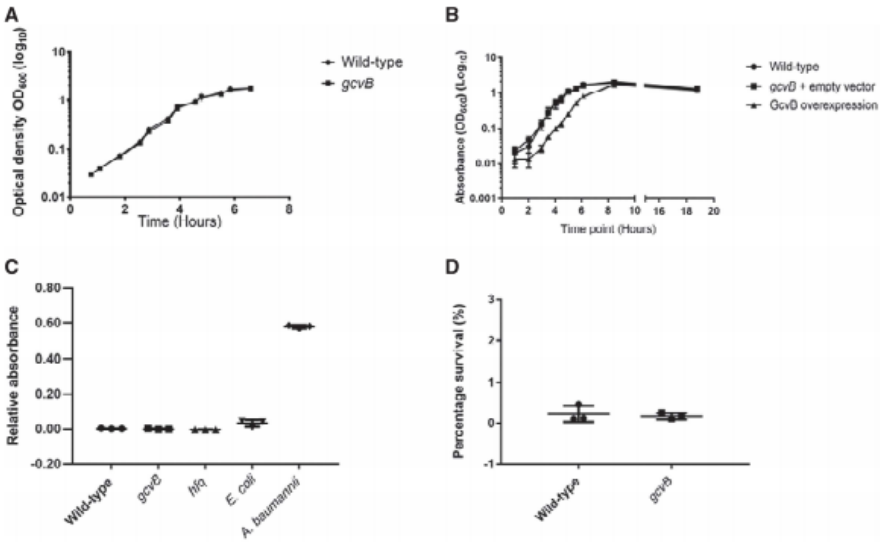

FIGURE 5. (*A*) Growth curve of wild-type *P. multocida* VP161 (circles) and the *gcvB* mutant strain (squares) in Heart infusion broth, incubated with shaking at 37°C for 7 h. Data shown are mean ± SD ($n = 3$). (*B*) Growth curve of wild-type *P. multocida* VP161 (circles), *gcvB* mutant containing empty vector (squares), and the *gcvB* overexpression strain (triangles) grown for 24 h under the same conditions as above. (*C*) Relative absorbance compared to a no bacteria control, observed following a static crystal violet biofilm assay. The *P. multocida* wild-type VP161 (circles) was compared to the *gcvB* mutant strain (squares) and the *hfq* mutant strain (upright triangles). Controls included *E. coli* (upside-down triangles), an intermediate biofilm-forming species, and *A. baumannii* (hexagons), a strong biofilm-forming species. *Horizontal* lines represent mean ± SD ($n = 3$). (*D*) Survival of *P. multocida* wild-type VP161 (circles) and the *gcvB* mutant strain (squares) following 15 min of acid stress at pH 4.6. *Horizontal* lines represent mean ± SD ($n = 3$).

production in both of the GcvB-deficient strains analyzed (Table 1). A total of 10 proteins showed decreased production in one of the GcvB-deficient strains (AL2677 or AL2862), but none of these showed decreased production in both GcvB-deficient strains and only one (Tpl) showed inverse (increased) production in the *gcvB* overexpression strain.

P. multocida GcvB binds Hfq and many of the GcvB mRNA targets are unique

The binding of GcvB with many mRNA targets in *E. coli* and *S.* Typhimurium is known to be Hfq-dependent. To confirm *P. multocida* GcvB bound to Hfq we expressed a FLAG-tagged Hfq in *P. multocida* and used coimmunoprecipitation followed by high-throughput sequencing to identify precipitated RNAs (in triplicate samples). Sequences matching GcvB were recovered from the FLAG-tagged Hfq samples at high numbers, on average 774.7 reads per sample, but at significantly reduced numbers in the untagged control sample, an average of 39.3 reads per sample (FDR < 0.05). Therefore, we conclude that *P. multocida* GcvB can bind *P. multocida* Hfq. Given this information, the list of proteins identified as differentially produced in the *P. multocida gcvB* mutant analyses was compared to the list of proteins identified as differentially produced in the previously analyzed *P. multocida hfq* mutant (Mégroz et al. 2016). Ten of the proteins that showed increased production in the *P. multocida hfq* mutant also showed increased production in both of the GcvB-deficient strains and a further five showed increased production in one of the GcvB-deficient strains (AL2677 or AL2682). These proteins were Asd, DapA, DppA, GdhA, GltA, HisC, HisH_1, IlvG, LysC, MetB, OppA, PlpB, PurC, RcpA, and RsgA_2 (Supplemental Table S3). One protein, SpeF, showed decreased production in the *hfq* mutant at mid-log growth phase (Mégroz et al. 2016) but increased production in the *gcvB* mutant, AL2677 (Supplemental Table S3). The list of proteins with altered production in the *P. multocida* GcvB-deficient strains was also compared with the 54 GcvB-regulated targets identified in *E. coli* and *S.* Typhimurium (Sharma et al. 2011). Of these 54 known targets, 42 had homologs in the *P. multocida* genome and all but five of these were measured in our proteomics experiments. However, only eight were identified as differentially produced in our proteomics experiments; namely, DppA, GdhA, LysC, OppA, OppB, PlpB, SerA, and ThrA (Supplemental Table S3).

TABLE 1. Proteins with increased production in either the *P. multocida gcvB* mutant strain (AL2677) or mutant with empty vector (AL2682) and decreased production in the *gcvB* overexpression strain (AL2684) when compared to the VP161 wild-type parent

Protein name[a]	VP161 locus tag, (Pm70 locus tag)	AL2677 (log_2), (FDR)	AL2682 (log_2), (FDR)	AL2684 (log_2), (FDR)	Predicted protein function	Primary biochemical pathway/s
ThrA	PMVP_0066, (PM0113)	0.59, (0.007)	0.47, (0.042)	−1.18, (0.0001)	Bifunctional aspartokinase I/homeserine dehydrogenase I	Isoleucine
ArtI	PMVP_0077, (PM124)	0.99, (0.002)	0.57, (0.059)	−1.63, (0.0001)	Arginine ABC transporter	Transporter–arginine
DppA	PMVP_0194, (PM0236)	1.89, (0.001)	1.87, (0.005)	−1.78, (0.0002)	Periplasmic dipeptide transport protein	Transporter–dipeptides
GltA	PMVP_0236, (PM0276)	1.25, (0.002)	1.02, (0.005)	−1.58, (0.0001)	Citrate synthase	TCA cycle, Glutamate
BioB	PMVP_0348, (PM0379)	0.85, (0.005)	0.64, (0.030)	−1.66, (0.0001)	Biotin synthase	Valine
PM0472	PMVP_0448, (PM0472)	1.08, (0.002)	0.74, (0.007)	−1.56, (0.00003)	PBP2_TAXI_TRAP_like_3 domain-containing protein	
MsrA	PMVP_0575, (PM0605)	0.65, (0.020)	0.73, (0.012)	−1.30, (0.0001)	Peptide methionine sulfoxide reductase	Methionine
AspC	PMVP_0593, (PM0621)	0.78, (0.010)	0.88, (0.005)	−1.38, (0.0001)	Aromatic amino acid aminotransferase	Tyrosine
MetC_2	PMVP_0791, (PM0794)	ND[b] (ND)	1.04, (0.013)	−2.06, (0.0001)	Cystathionine β-lyase	Methionine
ArgG	PMVP_0809, (PM0813)	0.70, (0.006)	0.51, (0.082)	−1.63, (0.0001)	Argininosuccinate synthase	Arginine
PurC (HemH)	PMVP_0811, (PM0815)	0.82, (0.004)	1.03, (0.006)	−1.52, (0.0001)	Phosphoribosylaminoimidazole-succinocarboxamide synthase	De novo purine nucleotide synthesis
HisH_1	PMVP_0837, (PM0838)	1.05, (0.002)	1.17, (0.007)	−1.63, (0.0002)	Histidinol-phosphate aminotransferase	Histidine
AroA	PMVP_0838, (PM0839)	0.61, (0.008)	0.37, (0.200)	−1.03, (0.0007)	3-phosphoshikimate 1-carboxyvinyltransferase	Tyrosine
PcnB	PMVP_0865, (PM0864)	0.27, (0.059)	0.68, (0.035)	−0.78, (0.0033)	Poly(A) polymerase	
LysC	PMVP_0948, (PM0937)	1.07, (0.004)	1.27, (0.005)	−1.66, (0.0001)	Aspartate kinase	Lysine, threonine, methionine, homoserine, isoleucine
MetB	PMVP_1008, (PM0995)	0.79, (0.047)	0.99, (0.035)	−2.11, (0.0001)	Cystathionine γ-synthase	Methionine, lysine, threonine, homoserine
DapA	PMVP_1069, (PM1051)	0.83, (0.003)	0.86, (0.058)	−1.67, (0.0005)	Dihydrodipicolinate synthase	Lysine, threonine, methionine
HisC	PMVP_1220, (PM1199)	0.64, (0.030)	0.77, (0.035)	−1.11, (0.001)	Histidinol-phosphate aminotransferase	Histidine
HisB	PMVP_1221, (PM1200)	1.22, (0.002)	1.21, (0.005)	−1.33, (0.0002)	Histidinol-phosphatase	Histidine
HisA	PMVP_1224, (PM1203)	0.73, (0.008)	0.48, (0.275)	−2.02, (0.0002)	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase	Histidine
HisF	PMVP_1225, (PM1204)	0.71, (0.003)	0.72, (0.035)	−1.51, (0.0001)	Imidazoleglycerol phosphate synthase, cyclase subunit	Histidine
Asd	PMVP_1687, (PM1623)	0.89, (0.008)	0.63, (0.035)	−1.08, (0.0004)	Aspartate-semialdehyde dehydrogenase	Lysine, threonine, methionine, homoserine, isoleucine
SerA	PMVP_1723, (PM1671)	0.73, (0.004)	0.82, (0.035)	−1.20, (0.0009)	D-3-phosphoglycerate dehydrogenase	Serine, cysteine and glycine
PlpB	PMVP_1787, (PM1730)	1.39, (0.002)	1.22, (0.009)	−1.28, (0.001)	Outer membrane lipoprotein	Transporter-Methionine
OppB	PMVP_1961, (PM1909)	0.63, (0.004)	0.20, (0.591)	−1.67, (0.0002)	Oligopeptide transport system permease protein OppB	Transporter—oligopeptides
OppA	PMVP_1962, (PM1910)	1.31, (0.002)	1.09, (0.005)	−1.55, (0.0001)	Periplasmic oligopeptides binding protein	Transporter—oligopeptides
GdhA	PMVP_2095, (PM0043)	2.27, (0.001)	2.48, (0.104)	−3.46, (0.0079)	Glutamate dehydrogenase	Glutamate synthesis, TCA cycle, nitrate reduction

Protein production ratio in each strain (relative to protein production in VP161) is shown as a log_2 value with the corresponding false discovery rate (FDR) shown in brackets.
[a]Differentially expressed proteins were defined as those showing at least 1.5-fold change in production ($log_2 \geq 0.59$) with a FDR of less than 0.05.
[b]ND, no data available.

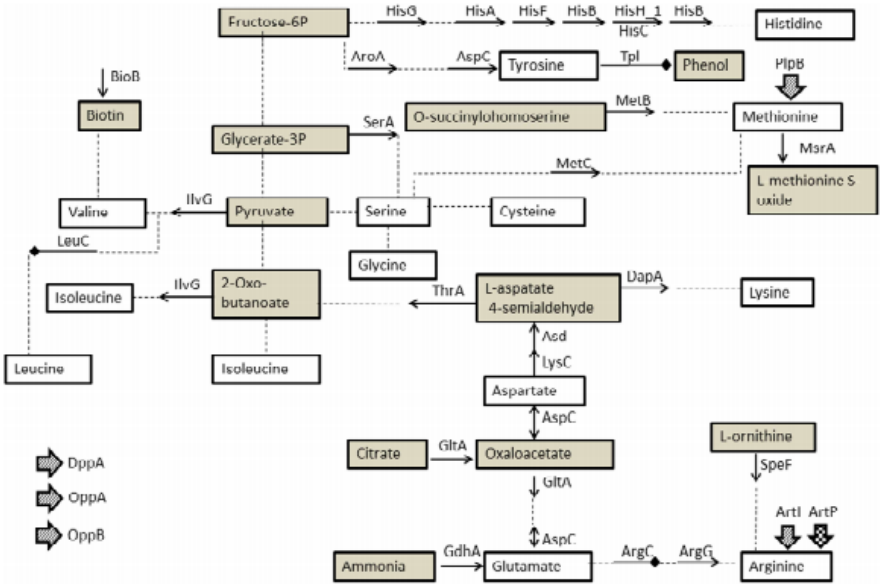

FIGURE 6. Amino acid biosynthesis pathways predicted to be affected by *gcvB* inactivation in *P. multocida* strain VP161. The amino acids whose biosynthesis is predicted to be regulated by GcvB are within white boxes. Amino acid biosynthesis proteins whose production is negatively regulated by GcvB are shown at the relevant pathway step with open-headed arrows. Proteins whose production is positively regulated by *gcvB* are indicated at the relevant pathway step with closed diamond-headed arrows. Large, diagonally striped, and checked arrows indicate predicted amino acid or oligo peptide transport proteins that are negatively and positively regulated by *gcvB*, respectively.

The proteins identified as differentially produced (increased or decreased production) in *P. multocida* following inactivation of *gcvB* in either experiment (Supplemental Tables S3, S4), were mapped to their metabolic pathways; amino acid biosynthesis proteins were observed to be highly overrepresented (Fig. 6; Fishers exact test; $P < 10^{-11}$). These amino acid biosynthesis proteins included 21 with increased production (ArgG, AroA, Asd, AspC, BioB, DapA, GdhA, GltA, HisH_1, HisA, HisB, HisC, HisF, HisG, IlvG, LysC, MetB, MetC_2, SerA, SpeF, and ThrA) and three with decreased production (ArgC, LeuC, and Tpl). Furthermore, another five proteins with increased production were predicted to be involved in the transport of amino acids or oligopeptides (ArtI, DppA, OppA, OppB, and PlpB), as well as one protein with decreased production (ArtP). Thus, 27 of the 36 proteins negatively regulated by GcvB are involved in biosynthesis or transport of at least 14 different amino acids (Fig. 6). Moreover, of the 17 proteins that displayed increased production in both GcvB-deficient strains and an equivalent decrease in the *gcvB* overexpression strain, only PurC (predicted to be involved in de novo purine biosynthesis), and PM0472 (an uncharacterized periplasmic binding protein containing a PBP2_TAXI_TRAP_like_3 domain), were not predicted to be involved in amino acid transport and metabolism.

Bioinformatic analyses identify an extended GcvB seed region binding motif

In order to determine if each of the experimentally identified putative *P. multocida* GcvB targets contained a conserved region that may serve as a GcvB binding site, the DNA sequence starting 120 nt upstream of the start codon and continuing to 60 nt downstream from the start codon of each gene was examined for conserved sequence motifs using the Multiple Em for Motif Elicitation (MEME) tool (Bailey et al. 2009). Initially, all of the genes encoding the proteins identified as differentially produced following inactivation of *gcvB* (i.e., all proteins in Supplemental Tables S3 and S4) were examined. However, this analysis failed to identify a conserved motif across all proteins. We then constrained the target list to include only those proteins that showed increased differential production in either of the *gcvB* mutant strains and had a corresponding inverse production in the *gcvB* overexpression strain (Table 1). Using the DNA sequences (−120 to +60 nt) of these genes, a consensus sequence consisting of 5′-AACACAAC-3′ (*E*-value: 3.2×10^{-7}) was identified in all targets (Fig. 3B). The sequences around this identified motif were also aligned using Clustal Omega (Sievers et al. 2011), which revealed a highly conserved slightly extended

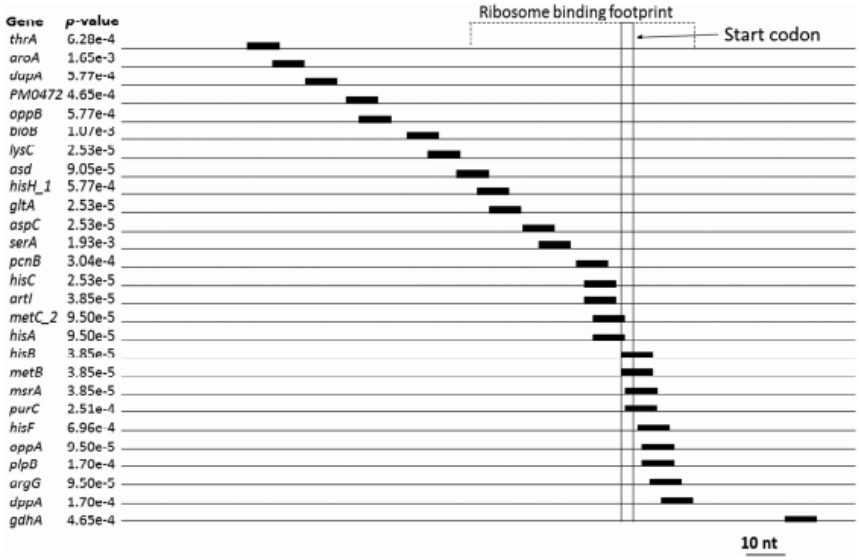

FIGURE 7. Schematic representation of the position of the predicted GcvB seed binding regions (black boxes) in 27 GcvB mRNA targets (listed), relative to the start codon. The *P*-values, generated by MEME motif finder, are shown at the *left* and give the likelihood of each identified motif occurring in the analyzed sequence fragment by chance. The predicted ribosome footprint is indicated by the dashed line (*top*). A scale bar is shown at the *bottom*.

consensus sequence (5′-AACACAACAT-3′) (Fig. 3C). Therefore, we predict that the *P. multocida* GcvB seed binding sequence is slightly longer than the *E. coli* and *S.* Typhimurium GcvB seed region, but that the eight central nucleotides (5′-CACAACAT-3′) are identical. Importantly, the reverse complement of the extended *P. multocida* GcvB seed sequence, 5′-AUGUUGUGUU-3′, is present within the sequence of GcvB; this sequence was identical to the same region in the *Y. pestis* GcvB and differed by just a single nucleotide compared to the same region in the *E. coli* and *S.* Typhimurium GcvB (Fig. 1). The position of this consensus GcvB binding sequence was then mapped on each of the 27 mRNA targets (Fig. 7). The binding sequence was located upstream of the predicted ribosome binding footprint [−39 to + 19 bp (Hüttenhofer and Noller 1994; Sharma et al. 2007)] in seven mRNA targets and was overlapping, or within the ribosome binding footprint, in 19 targets. In one target, *gdhA*, the binding sequence was downstream from the ribosome binding footprint (Fig. 7). As four of the identified GcvB targets were known GcvB targets in *Salmonella* and *E. coli,* the putative seed binding regions of the *P. multocida* targets were compared to the known GcvB seed binding regions in *oppA*, *dppA*, *serA*, and *gdhA* encoded by *Salmonella* (Sharma et al. 2007, 2011). It was found that the seed region for the *P. multocida oppA* was located at the same position relative to the seed region of *oppA* in *Salmonella* and contained a similar sequence (Sharma et al. 2007). The seed region for the *P. multocida serA* was located close to the seed region position reported for *serA* in *Salmonella* but the sequence was dissimilar (Sharma et al. 2011). In contrast, the predicted seed regions for *P. multocida dppA* and *gdhA* were found in different locations to those reported for the equivalent genes in *Salmonella* and had only limited sequence similarity (Sharma et al. 2007, 2011).

In order to determine if the identified GcvB seed region was present in all mRNAs encoding proteins predicted to be regulated by GcvB, the corresponding DNA sequences (−120 to +60 nt) were visually inspected. Two more seed sequences were identified that exactly matched the consensus sequence generated with MEME, and these were located in *glpQ* and *leuC*, positioned at 34 and 0 nt upstream of the start codon, respectively. The remaining putative targets had no sites with less than one or two mismatched nucleotides at critical positions (3 and 6).

Modification of a two-plasmid GFP reporter system to detect *P. multocida* sRNA–mRNA interaction in *E. coli*

In order to experimentally confirm that the conserved sequence 5′-AACACAACAT-3′ contained the *P. multocida* GcvB seed region, sRNA/mRNA interaction experiments using two recombinant plasmids were conducted in *E. coli*

strain DH5α, based on a previously described two-plasmid GFP reporter system (Urban and Vogel 2007). *P. multocida* Hfq shares 92.7% identity with two-thirds of the *E. coli* Hfq protein (amino acids 1–73) but shares only 13.7% identity with the C-terminal region of *E. coli* Hfq (amino acids 74–102). Therefore, before using this system, we first assessed whether *E. coli* Hfq could act as a chaperone for *P. multocida* sRNA molecules. A *P. multocida* expression plasmid containing a functional copy of the *E. coli* DH5α *hfq* (pAL1266, Supplemental Table S2) was used to transform the *P. multocida* VP161 *hfq* mutant, which produces only low levels of hyaluronic acid capsule compared to the parent strain VP161 (Mégroz et al. 2016). When the *P. multocida hfq* mutant was complemented with pAL1266 (expressing *E. coli hfq*), capsule production was restored to the same level as that observed when the *hfq* mutant was complemented with the native *P. multocida hfq* gene (Fig. 8). Thus, these data show that the native *E. coli* Hfq molecule can appropriately chaperone *P. multocida* sRNAs, allowing *E. coli* to be used as the host cell for the *P. multocida* sRNA–mRNA interaction studies described below.

To produce a two-plasmid GFP reporter system for our experiments, two expression vectors were constructed, designated pREXY and pTEXY (Supplemental Table S2). The pREXY plasmid is a shuttle vector used for the expression of *P. multocida* sRNAs (GcvB in this case) in either *E. coli*

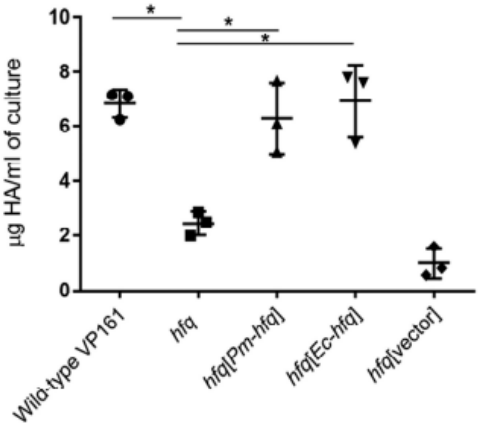

FIGURE 8. Hyaluronic acid capsule production in the *P. multocida hfq* mutant containing a functional copy of *hfq* from *E. coli* or *P. multocida*. The amount of hyaluronic acid capsular material produced during mid-exponential growth by *P. multocida* wild-type strain VP161, *P. multocida hfq* mutant (*hfq*), *P. multocida hfq* mutant complemented with a functional copy of the native *hfq* from *P. multocida* (*hfq*[*Pm-hfq*]), *P. multocida hfq* mutant containing a functional copy of the *hfq* from *E. coli* (*hfq*[*Ec-hfq*]), or the *P. multocida hfq* mutant containing empty vector (*hfq*[vector]). Each data point shows a single hyaluronic acid measurement. Thick *horizontal* bars represent the mean and error bars show ±1 SD ($n = 3$). (*) $P < 0.05$ using Student's *t*-test.

or *P. multocida* and contains a *P. multocida tpi* promoter upstream of the multiple cloning site (MCS). The second plasmid, pTEXY, is used for the transcriptional and translational coupling of the mRNA target with superfolder GFP (sfGFP) under the control of the tetracycline promoter (P_{tet0-1}).

GcvB inhibits GltA production via complementary binding between the predicted seed regions in GcvB and *gltA*

For recombinant expression of *P. multocida gcvB* sRNA in *E. coli*, the entire *gcvB* gene from *P. multocida* strain VP161 was PCR-amplified and cloned into the MCS of pREXY, generating the GcvB expression plasmid pAL1197. For recombinant expression of a predicted *gcvB* target region, a *P. multocida* fragment containing 38 bp upstream and the first 60 bp of *gltA* was cloned into the XbaI and BglII sites of pTEXY, located between P_{tet0-1} and sf*GFP* to produce a *gltA-sfGFP* translational fusion. This plasmid was named pAL1257 (Supplemental Table S2). The recombinant plasmids, or vector only, were used in various combinations to transform competent *E. coli* DH5α. Restriction digest analysis and DNA sequencing confirmed all transformants contained the correct plasmids.

The *E. coli* strain containing both the pTEXY::*gltA-sfGFP* expression plasmid and the empty pREXY vector (no GcvB) was highly fluorescent, but the strain containing both the pTEXY::*gltA-sfGFP* plasmid and the pREXY::*gcvB* expression plasmid showed significantly reduced fluorescence ($P < 0.0005$; Fig. 9). Thus, expression of GcvB represses production of the GltA-sfGFP fusion protein, as would be expected for a bona fide GcvB target mRNA.

In order to confirm that the GcvB-mediated repression of GltA expression was specifically due to complementary base pairing between the predicted seed regions, two modified plasmids were constructed and tested for fluorescence in the two-plasmid GFP reporter system. Firstly, the putative central seed sequence in the *gltA* upstream region was replaced with a nucleotide sequence identical to the central seed region of the GcvB sRNA (UGUGUUG) to generate the plasmid pTEXY::$gltA_{MSR1}$-*sfGFP* (pAL1290; Supplemental Table S2). The *E. coli* strain containing this plasmid, with the *gltA* seed region mutation, and the pREXY::*gcvB* plasmid showed levels of fluorescence indistinguishable from the fluorescence of the strains containing pTEXY::*gltA-sfGFP* and empty pREXY (no GcvB). This indicates that GcvB was unable to repress the production of GltA following the mutation of the *gltA* seed region. Secondly, the plasmid pREXY::$gcvB_{MSR2}$ (pAL1277; Supplemental Table S2) was generated, encoding a modified *gcvB* that contained a nucleotide sequence identical to the seed region of *gltA* mRNA target (ACACAAC), instead of the GcvB seed region (UGUGUUG). The *E. coli* strain containing both this plasmid and the pTEXY::*gltA-sfGFP* showed levels of fluorescence indistinguishable from the fluorescence of the strains

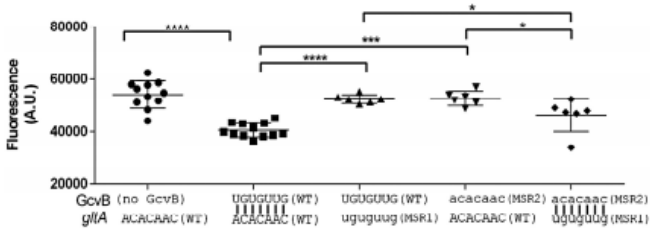

FIGURE 9. Superfolder green fluorescent protein (sfGFP) production in *E. coli* strains containing different plasmid pairs. Each *E. coli* strain harbored one pREXY sRNA expression plasmid derivative and one pTEXY mRNA::*sf*GFP reporter derivative. The *top* line of the *x*-axis label shows the sequence of the native (WT) or mutated (MSR2) seed region within the recombinant *gcvB* in the plasmid pREXY::*gcvB* or pREXY::$gcvB_{MSR2}$, respectively. A pREXY vector only control (no GcvB) was also included in the study. The *bottom* line of the *x*-axis shows the sequence of the native (WT) or mutated (MSR1) seed region within the recombinant *gltA* fused to the sfGFP gene in the pTEXY::*gltA-sfGFP* or pTEXY::$gltA_{MSR1}$-*sfGFP*, respectively. Wild-type seed sequence is shown in all capitals, mutated seed sequence is shown in lower case. *Vertical* lines between the text show if complementary base pairing is predicted between the sRNA and mRNA seed sequence. The amount of sfGFP-mediated fluorescence for each recombinant *E. coli* strain was measured (475/540 nm ex/em), and each data point shows the amount of fluorescence emitted by a single strain. The long *horizontal* bars show the mean of the replicate data, and error bars show ±1 SD ($n = 6$ or 12). (*) *P*-value < 0.05, (***) *P*-value < 0.005, (****) *P*-value < 0.0005.

containing pTEXY::*gltA-sfGFP* and empty pREXY (no GcvB). Thus, GcvB-mediated repression of *gltA* expression was also abrogated by mutation of the *gcvB* sRNA seed region. Finally, we tested the fluorescence of the *E. coli* strain containing both of the mutated plasmids, pTEXY::$gltA_{MSR1}$-*sfGFP* and pREXY::$gcvB_{MSR2}$, containing swapped seed regions but which are still complementary to each other. The strain containing these plasmids showed significantly reduced fluorescence compared to each of the strains containing the following plasmid pairs: pTEXY::*gltA-sfGFP* and empty pREXY, pTEXY::$gltA_{MSR1}$-*sfGFP* and pREXY::*gcvB*, and pTEXY::*gltA-sfGFP* and pREXY::$gcvB_{MSR2}$. Therefore, when the seed regions of both the GcvB sRNA and the mRNA target are mutated but in a complementary fashion, GcvB-mediated repression is restored, confirming that there is a direct interaction between the two predicted seed binding regions and that this level of binding is sufficient for the repression of GcvB expression.

DISCUSSION

In this study we have shown, using deep sequencing transcriptomic analyses and northern blotting, that the *P. multocida* GcvB is strongly expressed at early- and mid-exponential growth phases but displays highly reduced expression during late-exponential growth. These data correlate well with the known expression profile of GcvB in *E. coli* (Argaman et al. 2001) and *S.* Typhimurium (Sharma et al. 2007) and support the predicted function of GcvB as a repressor that acts primarily during growth under nutrient-rich conditions. We also examined the 5′ start of the GcvB sRNA expressed by the *P. multocida* strain VP161 using both primer extension and 5′ RACE. Primer extension identified the starting base as being positioned 1–2 bp downstream from the known start for the GcvB transcript in *E. coli* and *S.* Typhimurium. In contrast, experiments using 5′ RACE identified the transcript start was located 2 bp upstream of the start in *E. coli* and *S.* Typhimurium. The 5′ RACE method is considered the superior method for determining transcript starts as the 5′ end of the RNA is protected from degradation by the addition of an adapter. Therefore, we conclude the *P. multocida* GcvB begins with the sequence 5′-AUACUUAAU-3′.

In order to identify the *P. multocida* GcvB regulon, we analyzed the proteome of the wild-type strain, a *gcvB* mutant, a *gcvB* mutant containing empty vector and a GcvB overexpression strain. Nearly four times as many *P. multocida* proteins were identified as differentially produced in the *gcvB* overexpression strain than in the GcvB-deficient strains. Quantitative qRT-PCR showed that the level of GcvB in the overexpression strain was increased by approximately 70-fold compared to the wild-type strain at early-exponential growth phase when the proteomics was performed. Therefore, we propose that the overexpression of GcvB to this level may lead to some off-target effects via nonspecific binding, as has been observed for other sRNAs (Storz et al. 2011).

The *P. multocida gcvB* mutant displayed normal growth in rich medium, was unaffected by acid stress and showed no change in phenotype (compared to the parent strain) with respect to biofilm formation. This is in contrast to what has been reported for other species; *gcvB* mutants constructed in *E. coli, S.* Typhimurium and *Y. pestis* all show a decreased growth rate in rich media, and inactivation of *gcvB* in *E. coli* results in cells with decreased biofilm formation and decreased tolerance to acid stress (McArthur et al. 2006; Sharma et al. 2007; Jin et al. 2009; Mika and Hengge 2014). It is perhaps unsurprising that the *P. multocida gcvB* mutant did not show a change in the ability to form a biofilm. Indeed, our data suggest that wild-type VP161 forms very poor single species biofilms. Moreover, with respect to acid tolerance, *P. multocida* is considered a bite wound and respiratory/mucosal pathogen and, unlike enteric organisms, is unlikely to encounter strongly acidic conditions. However, the *P. multocida gcvB* overexpression strain did show an increased lag-phase during growth when compared to wild-type VP161. It has been previously shown that during lag phase the glycolysis pathway is predominantly used to produce energy (Rolfe et al. 2012). An important enzyme in the Krebbs cycle is citrate synthase, which in *P. multocida* is encoded by the GcvB target, *gltA*. Therefore, the increase in lag phase displayed by

the *P. multocida gcvB* overexpression strain may in part be the result of decreased production of GltA due to increased GcvB binding to *gltA* transcripts. However, as this strain significantly overexpresses GcvB, it is acknowledged that the likely off-target effects on multiple other proteins may also play a role.

Of the proteins that showed either increased (36) or decreased (10) production in the GcvB-deficient *P. multocida* strains analyzed, 31 (27 increased, four decreased) were predicted to be involved in amino acid biosynthesis and transport, and pathway analyses indicated that GcvB specifically affects the biosynthesis of at least 14 different amino acids. Therefore, our data suggest that the *P. multocida* GcvB acts primarily to repress the production and transport of amino acids during the early growth stages, likely as a means to conserve energy when nutrients are abundant. In *E. coli* and *S.* Typhimurium the role of GcvB is also to repress amino acid biosynthesis and transport when nutrients are in plentiful supply. However, in these species GcvB shows a preponderance for regulation of amino acid transporters [>60% of GcvB targets (Sharma et al. 2011)]. In *P. multocida* this situation appears to be reversed, with the majority of the regulated proteins (~75%) being directly involved in the biosynthesis of amino acids. A comparison of the targets regulated by GcvB in *P. multocida*, *E. coli*, and *S.* Typhimurium identified four that were GcvB-regulated in all three species (GdhA, OppA, SerA, and DppA), two targets that were GcvB-regulated in both *P. multocida* and *S.* Typhimurium (PlpB and ThrA) and two targets that were GcvB-regulated in both *P. multocida* and *E. coli* (OppB and LysC). Thus, while the general function of GcvB as a controller of amino acid biosynthesis and transport has been conserved across the species, the precise GcvB targets show significant diversity.

The production of the histidine biosynthesis proteins HisA, HisB, HisC, HisF, HisG, and HisH_1 was strongly increased (fold change ranging from 1.6- to 2.3-fold) in *P. multocida* lacking a functional *gcvB*; five of these proteins are predicted to be encoded within a single operon. In other bacteria, histidine production is regulated by multiple mechanisms including repression of transcription initiation and attenuation (Kulis-Horn et al. 2014), but to our knowledge GcvB has not been previously linked with control of histidine biosynthesis. HisD encodes a histidinol dehydrogenase that has also been bioinformatically predicted to be a target of GcvB in the related *Pasteurellaceae* species *A. pleuropneumoniae* (Rossi et al. 2016). Moreover, specific attenuator sequences that target histidine production have been identified in *A. pleuropneumoniae* (Rossi et al. 2016).

Of the 27 GcvB targets shown in Table 1, 71% also showed increased production in a *P. multocida* strain VP161 *hfq* mutant (Mégroz et al. 2016). This indicates that the action of the *P. multocida* GcvB on many of the putative mRNA targets is dependent on the chaperone activity of Hfq, which mediates the docking of an sRNA onto its mRNA target. The reliance of GcvB on Hfq for binding to certain mRNA targets has also been demonstrated in *E. coli* (Pulvermacher et al. 2008). Compared to protein levels in the wild-type VP161, the predicted GcvB target SpeF showed increased production in the *gcvB* mutant during early-exponential growth, when *P. multocida* GcvB has been shown to be most active. In contrast, SpeF showed decreased production in the *P. multocida hfq* mutant during mid-exponential growth, indicating that other sRNAs may act upon SpeF at later growth phases.

Previously, it was proposed that during the late stages of *P. multocida* infection the in vivo environment is nutrient poor (Boyce et al. 2002). Under these conditions, we would predict that the levels of *gcvB* gene expression would be low, thus allowing the expression of *gcvB* mRNA targets involved in amino acid biosynthesis and transport. Supporting this prediction, four of the genes encoding putative GcvB mRNA targets, *aspC*, *dppA*, *gdhA*, and *gltA* had increased expression (fold changes ranging from 1.8 to 11.3) during in vivo growth in chickens (Boyce et al. 2002). It is possible that GcvB plays an important role in the regulation of these targets in vivo. However, as yet we have no direct evidence of reduced GcvB expression during growth in vivo as the previous microarray experiments (Boyce et al. 2002) did not include DNA spots representing any sRNAs. In our current study, the glutamate dehydrogenase, GdhA, was identified as the most highly differentially produced protein (fivefold increase) following GcvB inactivation. GdhA catalyzes the conversion of L-glutamate to 2-oxoglutarate, releasing NH_3 and NADPH which then allows for the production of all amino acids within the cell (Reitzer 1996). Interestingly, a *P. multocida gdhA* mutant belonging to the capsular type B and LPS serotype/genotype 2, was attenuated for virulence and was used as an effective vaccine against hemorrhagic septicemia in buffalo (Rafidah et al. 2012).

Comparative bioinformatic analysis using the gene sequences for 27 of the *P. multocida* GcvB mRNA targets, allowed for the identification of the predicted GcvB sRNA seed region (initial sRNA-mRNA binding site) consisting of 5′-AACACAACAT-3′. This sequence was highly conserved in a large number of the putative mRNA targets (Fig. 3C) and the complementary sequence of this seed region was present in the *P. multocida* GcvB sRNA. This seed binding sequence is 2 nt longer than the characterized seed binding regions of the GcvB sRNA molecules encoded by *E. coli* and *S.* Typhimurium but importantly contains the same core region sequence, 5′-CACAACAT-3′ (Urbanowski et al. 2000; McArthur et al. 2006; Pulvermacher et al. 2008; Sharma et al. 2011). The internal section of this binding region was confirmed as essential for GcvB interaction with the target *gltA* mRNA using the GFP translational reporter assay in *E. coli*, where substitution of these bases in either the mRNA target, *gltA*, or the sRNA, GcvB, decreased the interaction between the RNA molecules. Complementary substitution of bases in the sRNA and mRNA target allowed for restoration of binding efficiency and a concomitant decrease in GFP production.

The predicted seed region within each of the negatively regulated *P. multocida* GcvB target mRNAs was mapped relative to the start codon of the gene. Similar to what has been observed in other bacteria (Bobrovskyy et al. 2015), most of the GcvB-specific seed regions mapped within the ribosome binding footprint, which we predict would allow GcvB to occlude the RBS and block translation of the target mRNA. However, some were located upstream of the ribosome binding footprint; this included the seed region sequence in *thrA*, which was located ~43 nt upstream of the ribosome binding footprint. The distal position of the seed region relative to the RBS has also been noted in some GcvB mRNA targets in *E. coli* and *S.* Typhimurium. In these instances it is thought that the CA-rich sequence within the seed region acts as a translational enhancer element and the binding of GcvB to this region blocks this enhanced translation (Sharma et al. 2007; Yang et al. 2014). Interestingly, 12 of the *P. multocida* mRNA targets (*hisB*, *metB*, *purC*, *msrA*, *metC_2*, *hisA*, *plpB*, *dppA*, *argG*, *oppA*, *hisF*, and *gdhA*) had the seed region sequence located on or after the start codon. Binding of an sRNA molecule soon after the translational start codon on the mRNA target is predicted to affect ribosome binding and translation of a gene because the ribosome footprint can extend from the −39 to the +19 nt (Hüttenhofer and Noller 1994; Sharma et al. 2007). Indeed, inhibitory interactions between the sRNA RybB and the mRNA target *ompN* in *S.* Typhimurium occur at +5 to +20 nt from the start codon (Bouvier et al. 2008).

The predicted seed region within the *gdhA* mRNA (position +40) is significantly downstream from the ribosome footprint region. Previous studies in *S.* Typhimurium have suggested that the sRNA–mRNA interactions between *gdhA* and GcvB may include a second highly conserved GcvB binding site called R2 (Fig. 1). There is very limited evidence that R2 is definitively involved in any specific sRNA-mRNA binding interactions, as the study showed that deleting the R2 region of GcvB did not abrogate binding to the *gdhA* transcript (Sharma et al. 2011; Melamed et al. 2016). The R2 region is located downstream from the primary seed region and is present in GcvB from all bacterial species analyzed (Fig. 1; Sharma et al. 2011; Melamed et al. 2016). Future work will assess the importance of this R2 region in *P. multocida*.

This study has characterized the GcvB regulon in *P. multocida* strain VP161 and identified the seed binding regions required for interaction between GcvB and its targets. Many of the mRNA targets identified are required for the biosynthesis and transport of amino acids. Thus, the correct temporal expression of GcvB is likely to be important for growth of this pathogen in a nutrient depleted environment, such as in vivo during late-stage infection. While the GcvB target-binding site is well conserved between *P. multocida* and *E. coli*, and the GcvB-regulated genes in both species are primarily involved in amino acid biosynthesis and transport, the precise genes controlled by GcvB in the two species are quite different. These data are the first functional characterization of sRNA regulation in the *Pasteurellaceae* family; future studies will focus on identifying the role of GcvB and other sRNAs in vivo during *P. multocida* infections.

MATERIALS AND METHODS

Bacterial strains, media, plasmids, and growth conditions

All bacterial strains and plasmids used in this study are listed in Supplemental Table S2. *P. multocida* strains were routinely cultured in Heart Infusion (HI) broth (Oxoid). *E. coli* strains were routinely grown in Luria-Bertani (LB) broth (Oxoid). For solid media, 1.0%–1.5% (w/v) agar was added to the media. When required, media were supplemented with the appropriate antibiotics; kanamycin (50 µg/mL), spectinomycin (50 µg/mL), or ampicillin (100 µg/mL).

DNA manipulations

Genomic DNA was purified using the Genomic DNA Extraction Kit (RBC Bioscience Corp.), and plasmid DNA was extracted using the NucleoSpin Plasmid Kit (Macherey-Nagel GmbH & Co. KG), according to the manufacturer's instructions. For cloning experiments, restriction endonucleases (New England Biolabs) and ligase (Roche) were used according to the manufacturer's instructions. PCR amplifications were performed using Taq DNA Polymerase (Roche) or Phusion High Fidelity DNA Polymerase (Roche) with oligonucleotides (primers) manufactured by Sigma-Aldrich. The primers used in this study are listed in Supplemental Table S1. PCR products were purified using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel). Sequencing reactions were performed with whole genomic DNA, plasmid DNA or PCR products as template as previously described (Harper et al. 2013). DNA sequences were analyzed using VectorNTI (Invitrogen), Clustal Omega (EBI), and BLAST (NCBI).

Construction of a *P. multocida gcvB* mutant

To inactivate *gcvB* in the *P. multocida* strain VP161, TargeTron mutagenesis (Sigma-Aldrich) was used as previously described (Steen et al. 2010) but with the following modifications. The group II intron within the *E. coli–P. multocida* TargeTron shuttle vector, pAL953 (Harper et al. 2013), was retargeted to *gcvB* using the PCR amplification method described in the TargeTron manual. The primers BAP7565, BAP7566, and BAP7567 (Supplemental Table S1) were designed using the TargeTron design site (Sigma-Aldrich). The resulting plasmid, pAL1170 (Supplemental Table S2), was used to transform *P. multocida* strain VP161 by electroporation, and mutants containing a TargeTron group II intron insertion in *gcvB* were identified as previously described (Harper et al. 2013).

Construction of a *P. multocida* GcvB overexpression strain

The *gcvB* gene from *P. multocida* VP161 was amplified using BAP7585 and BAP7586 (Supplemental Table S1), digested with

BamHI and SalI, and cloned into similarly digested pPBA1100s. The resulting plasmid, pAL1190, and the empty vector pPBA1100s were separately used to transform the *P. multocida gcvB* mutant AL2677 via electroporation, producing strains AL2864 and AL2862, respectively (Supplemental Table S2).

Heterologous expression of the *E. coli hfq* gene in *P. multocida*

The *hfq* gene from *E. coli* DH5α was amplified using BAP7850 and BAP7851 (Supplemental Table S1), digested with BamHI and SalI then cloned into the *P. multocida* expression plasmid, pAL99T. The resulting plasmid, pAL1266, was used to transform the *P. multocida hfq* mutant AL2521 (Mégroz et al. 2016), producing the strain AL2838.

Hyaluronic acid capsule assay

P. multocida strains were grown in HI broth (in biological triplicate) supplemented with the appropriate antibiotics (where required) to mid-exponential growth phase ($OD_{600} = 0.6$). Capsule was extracted from washed cells and the amount of capsular material measured using a hyaluronic acid assay as described previously (Chung et al. 2001).

Response to acid stress

Acidic HI broth was prepared by addition of 37% (v/v) hydrochloric acid (HCl) to HI until pH 4.6 was reached. Triplicate overnight cultures were prepared for each *P. multocida* strain and supplemented with Kanamycin where required to maintain the plasmid. Each culture was diluted 1:100 in fresh HI broth and grown until early exponential phase ($OD_{600} = 0.2$). A 1 mL aliquot of this early exponential phase ($OD_{600} = 0.2$) culture was then added to 3 mL of acidified HI broth, without antibiotics, and incubated at 37°C for 15 min with shaking. Following incubation, 12 mL of basic HI broth was added to neutralize the culture. Appropriate dilutions of each culture were plated onto HI agar in duplicate and after 16 h incubation colonies were enumerated.

Biofilm formation assay

Cultures representing each bacterial strain were grown to mid-exponential growth phase ($OD_{600} = 0.6$), then 100 µL of the diluted culture (1:100) was added to four wells of a sterile 96-well plate which was then incubated overnight at 37°C without shaking to allow for biofilm formation. Following incubation, the plate was washed three times with dH_2O to remove planktonic bacteria. Remaining bacteria were stained with 125 µL of 0.1% (w/v) crystal violet and incubated for 10 min at room temperature. Excess stain was removed by washing with dH_2O three times. To resolubilize the crystal violet, 200 µL of 95% (v/v) EtOH was added to each well, incubated for 15 min and then mixed well. A 125 µL aliquot of each well was transferred to a well of an optically clear flat bottomed 96-well plate and the optical density determined using a Tecan Infinite M200 plate reader.

RNA extraction, qRT-PCR, and whole-genome transcriptomic analyses by RNA-seq

P. multocida RNA extractions were performed as described previously (Boyce et al. 2002) but with the following modifications. Duplicate bacterial cultures were grown in HI broth to $OD_{600} = 0.2$ (early-exponential growth phase), $OD_{600} = 0.6$ (mid-exponential growth phase), or $OD_{600} = 1.0$ (late-exponential growth phase). Killing buffer was omitted from the RNA extraction method. Following DNase treatment of the samples, RNA was further purified by phenol:choloform extraction using 5Prime phase lock gel tubes as per the manufacturer's instructions (Quanta Biosciences). RNA-seq library preparation, sequencing on an Illumina HiSeq, and mapping and differential expression analysis were carried out as previously described (Mégroz et al. 2016). For the RNA-seq analyses, the average number of reads mapped across samples was 5647690.83, and of these an average of 99.7% mapped to the *P. multocida* VP161 genome, giving an average read depth of 2701.76 reads per gene. qRT-PCR was performed using the AffinityScript cDNA Synthesis Kit (Agilent) and Brilliant II SYBR Green qPCR Kit (Agilent) as per the manufacturer's instructions using the Eppendorf Realplex Mastercycler. Reverse transcription reactions, both plus and minus reverse transcriptase (+RT and −RT, respectively), were performed in biological triplicate, with each +RT reaction being measured in technical triplicate and each −RT reaction being measured in technical duplicate. Data were analyzed to ensure melt curves identified that only a single product was formed in each reaction and −RT controls did not amplify any products within 10 cycles of the experimental reactions.

Northern blotting

Northern blotting analysis was performed using the DIG Northern Starter Kit version 10 (Roche) as per the manufacturer's instructions, with the following modifications. A total of 8 µg of RNA was separated by agarose/formaldehyde gel electrophoresis and the separated products transferred to a nylon membrane by capillary electrophoresis. A GcvB-specific probe was amplified from *P. multocida* VP161 genomic DNA using BAP7888 and BAP7957; BAP7957 contains a T7 RNA polymerase promoter sequence at the 5′ end. The PCR product was then used in an in vitro transcription reaction using T7 RNA polymerase and 10X DIG labeled RNA mix (Promega).

Proteomics analysis

Total proteomes of the wild-type *P. multocida* VP161 and the *gcvB* mutant (in triplicate), were determined using nano-liquid chromatography coupled with tandem MS, following isotopic labeling with heavy and light formaldehyde as described previously (Mégroz et al. 2016).

Total proteomes of the wild-type *P. multocida* VP161, the *gcvB* mutant with empty vector and the GcvB overexpression strain were determined using label-free quantitative proteomics. Cells were grown in biological triplicate in HI broth to early-exponential growth phase ($OD_{600} = 0.2$) and pelleted by centrifugation. Cell pellets were lysed in 1% w/v SDC (sodium deoxycholate; Sigma-Aldrich), 100 mM Tris (pH = 8.1) and further homogenized on a Soniprep 150 Plus Sonicator (MSE). The protein concentration

was determined using a BCA Assay Kit (Pierce). A 200 µg aliquot of each total protein sample was denatured using 10 mM TCEP (Thermo Scientific), and free cysteine residues were alkylated with 40 mM chloroacetamide (Sigma-Aldrich). Trypsin Gold (Promega) was used to digest the proteins and SDC removed by extraction with water-saturated ethyl acetate. All samples were desalted using P-10 ZipTip columns (Agilent, OMIX-Mini Bed 96 C18), vacuum-dried and reconstituted in buffer A (0.1% formic acid, 2% acetonitrile) prior to mass spectrometry.

Using a Dionex UltiMate 3000 RSLCnano system equipped with a Dionex UltiMate 3000 RS autosampler, the samples were loaded via an Acclaim PepMap 100 trap column (100 µm × 2 cm, nanoViper, C18, 5 µm, 100 Å; Thermo Scientific) onto an Acclaim PepMap RSLC analytical column (75 µm × 50 cm, nanoViper, C18, 2 µm, 100 Å; Thermo Scientific). The peptides were separated using increasing concentrations of buffer B (80% acetonitrile/0.1% formic acid) for 158 min and analyzed with a QExactive Plus mass spectrometer (Thermo Scientific) operated in data-dependent acquisition mode using in-house, LFQ-optimized parameters.

Acquired .raw files were analyzed with MaxQuant (Cox and Mann 2008) to globally identify and quantify proteins across the various conditions. Statistical analyses for identification of differentially produced proteins were performed using the Limma package within R studio, where FDR is derived from the Benjamini–Hochberg procedure. Differentially produced proteins were identified as proteins with a ≥ 0.59 $\log_2$ fold change and an FDR ≤ 0.05. The proteomics data have been deposited in ProteomeXchange via the PRIDE database with identifier PXD007719.

Fluorescent primer extension

Fluorescent primer extension was performed as described previously (Lloyd et al. 2005; Steen et al. 2010) with the following modifications. RNA was isolated from *P. multocida* VP161 at $OD_{600} = 0.2$. For cDNA synthesis, 10 µg of total RNA was used as template with the 6-carboxy fluorescein amidite (6-FAM) labeled primer, BAP7962 or BAP8190 (Supplemental Table S1). Dried samples were analyzed using an ABI 3730xl DNA Analyzer (Thermo Fisher Scientific) located at the Australian Genome Research Facility (AGRF, Melbourne).

5′ RACE

5′ RACE was performed with 10 µg of RNA isolated from *P. multocida* VP161 using the Firstchoice RLM-RACE Kit (Applied Biosystems) according to the manufacturer's instructions with the following modifications. The reverse transcription step was replaced with the cDNA synthesis protocol used for fluorescent primer extension (above) using the nonfluorescent GcvB-specific primer BAP7889. The cDNA generated was resuspended in 30 µL of nuclease-free water; 1 µL was used in the first, nested PCR using the primer BAP7889 together with the commercially supplied 5′ RACE outer primer (Applied Biosystems). PCR reaction conditions were as follows; 94°C for 3 min, followed by 35 cycles consisting of 94°C 30 sec, 62°C 30 sec, 72°C 1 min, followed by a final extension step of 72°C for 7 min. The PCR product was then purified and 1 µL was used in the second, nested PCR using the primer BAP7754 and the commercially supplied 5′ RACE inner primer (Applied Biosystems) with the same PCR reaction conditions as described above. The nested PCR products generated were cloned into the vector pCR2.1 using the TOPO TA Cloning Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The nucleotide sequences of the cloned inserts were then determined using the vector-specific primer BAP612 in Sanger sequencing reactions.

Coimmunoprecipitation of GcvB by Hfq

To test whether *P. multocida* GcvB bound Hfq, we used coimmunoprecipitation of total bacterial RNA by a FLAG-tagged Hfq, followed by high-throughput sequencing of the precipitated RNAs. Total RNA was prepared from *P. multocida* expressing a chromosomally encoded, C-terminal 3xFLAG-tagged Hfq and as a control also from the wild-type *P. multocida* expressing native Hfq. FLAG-tagged Hfq, and any bound RNAs, were precipitated (three independent coimmunoprecipitation reactions) using anti-FLAG conjugated magnetic beads as previously described (Bilusic et al. 2014). RNA-seq library preparation, sequencing on a NextSeq (Illumina), and mapping and differential expression analysis was carried out as previously described (Mégroz et al. 2016).

Construction of plasmids for the two-plasmid GFP reporter assays

To analyze *P. multocida* GcvB/*gltA* mRNA target interactions, a two-plasmid GFP reporter system was developed based on the previously described system of Urban and Vogel (2007). This system required the construction of two expression vectors, pTEXY, required for the expression of the 5′ end of the mRNA target, containing the GcvB seed/binding region fused to a gene encoding sfGFP (Corcoran et al. 2012), and pREXY, required for the expression of the sRNA molecule, GcvB. To generate pTEXY the unique BsgI site present in the *E. coli* plasmid pBR322 was first changed to a BssHII site using site-directed PCR mutagenesis to allow for future experiments that required this restriction site to be uniquely located in the cloned mRNA DNA fragments. Two PCR products representing the pBR322 nucleotides 1 to 1656 (position of BsgI site) and nucleotides 1653–4358 were amplified by PCR. The first PCR reaction amplified the pBR322 nucleotides 1–1656 using BAP7721, which anneals to the EcoRI region and BAP7720, which anneals to the BsgI region but contains an altered sequence to incorporate a BssHII site instead of BsgI. The second PCR reaction amplified the pBR322 nucleotides 1653–4358 using BAP7722, which anneals to the EcoRI region and BAP7719, which anneals to the BsgI region but contains a BssHII site instead of BsgI. The PCR products were digested with EcoRI and BssHII, ligated, and the mixture used to transform competent *E. coli* DH5α, to generate the plasmid, pAL1240 (Supplemental Table S2). To construct pTEXY, a pMAT plasmid containing a commercially synthesized DNA fragment (Life Technologies) encoding the *sfGFP* gene (flanked by a HindIII and EcoRI restriction sites and under the control of the anhydrotetracycline (Atc)-inducible promoter, $P_{LtetO-1}$), was digested with HindIII and EcoRI. The DNA fragment containing *sfGFP* was then gel-purified and ligated to HindIII and EcoRI-digested pAL1240 to generate the plasmid pTEXY (Supplemental Table S2).

The final expression plasmid containing the *gltA-sfGFP* fusion (pAL1257) was constructed by generating an XbaI/BglII-digested PCR fragment, using the primers BAP7747 and BAP7748, that represented the region −38 to +60 (relative to start codon) of the

P. multocida gltA. The PCR fragment was digested, purified then ligated into the XbaI and BglII–digested pTEXY (Supplemental Table S2), such that expression would be under the control of the $P_{LtetO-1}$ promoter. The pTEXY plasmid, pAL1290, containing *gltA* with a mutated seed region (*gltA*$_{MSR1}$-*sfGFP*) was constructed in a similar manner with the exception that the forward primer BAP7964, containing the altered seed region sequence, was used for the amplification of the *gltA*-specific DNA (Supplemental Table S1).

To generate the base plasmid, pPBA1100S, used for the construction of sRNA expression plasmid pREXY, the DNA region (240 bp) encoding the P_{ipi} promoter was removed from the pAL99S vector (Harper et al. 2013) using EcoRI digestion followed by religation of the vector. This region was then replaced with a shorter DNA fragment (96 bp) containing the P_{ipi} promoter, amplified from *P. multocida* VP161 genomic DNA using BAP7638 and BAP7639 (containing HindIII and BamHI restriction sites, respectively), to ensure that transcription could begin as close as possible to the native sRNA (GcvB) start site. The BamHI/HindIII-digested PCR product was ligated to similarly digested pPBA1100S and the ligation mix used to transform *E. coli* DH5α. Recombinant colonies were selected on HI agar containing 50 µg/µL spectinomycin. Correct recombinant plasmids were confirmed by restriction analysis and DNA sequencing and one plasmid with the correct sequence designated pREXY (Supplemental Table S2).

The GcvB expression plasmid, pAL1197 (Supplemental Table S2), was constructed as follows. The region of the *P. multocida* VP161 genome encoding the putative *gcvB* was amplified from VP161 genomic DNA using the primers BAP7632 and BAP7633 (both containing an XmaI site). The purified, XmaI-digested, PCR product was then ligated to XmaI-digested pREXY. The authenticity of the pAL1197 plasmid containing *gcvB*, was confirmed by PCR and DNA sequencing. The pREXY plasmid containing the mutated *gcvB*$_{MSR2}$ (pAL1277, Supplemental Table S2) was constructed using splice overlap extension (SOE) PCR. Two PCR reactions were performed as follows. The reverse primer, BAP7951 and the forward primer BAP7950 (Supplemental Table S1), that overlap and anneal to the *gcvB* gene region containing the predicted seed region, were paired with BAP7632 (forward primer located upstream of *gcvB*) and BAP7633 (reverse primer located downstream from *gcvB*), respectively. Primers BAP7951 and BAP7950 contained sequence that changed the *gcvB* seed region from 5′-GTTGTGT-3′ to 5′-CAACACA-3′. The two PCR fragments, representing the 5′ and 3′ ends of *gcvB*, were combined using a second PCR amplification with primers BAP7632 and BAP7633 (Supplemental Table S1) to produce the *gcvB*$_{MSR2}$ fragment. The PCR product was then purified, digested with XmaI, and ligated to XmaI-digested pREXY.

Whole-cell fluorescent measurements

E. coli DH5α strains containing both a pTEXY-based plasmid (5′ mRNA-*sf*GFP fusion expression) and a pREXY-based plasmid (sRNA expression) were grown on LB agar supplemented with 50 µg/µL ampicillin and 50 µg/µL spectinomycin. Cells representing each strain were harvested from each plate (biological triplicate), resuspended in 1 mL of 1 × PBS buffer, and the volume adjusted to give a final OD_{600} of 2.0. A 200 µL aliquot of each cell suspension was added to a black flat bottomed 96-well microtiter plate (in triplicate). Fluorescence was measured using the Tecan Infinite M200 plate reader with an excitation/emission wavelength of 475/540 nm.

Bioinformatic analysis

Comparison of nucleotide and protein sequences was performed using BLAST (Camacho et al. 2009). The Rfam database version 12.2 (Burge et al. 2013) was used to compare the *P. multocida gcvB* sequence to known RNAs. The MEME motif identification website with MEME motif finder version 4.11.2 (Bailey et al. 2009) was used to identify potential GcvB binding sites in putative mRNA targets and these were then further analyzed using Clustal Omega (Sievers et al. 2011). Sequence data were analyzed and recombinant DNA molecules were designed using VectorNTI version 11 (Invitrogen). GcvB-regulated genes were mapped to the appropriate metabolic pathways using SmartTables (Travers et al. 2013) and pathway overview (Paley and Karp 2006; Karp et al. 2010) within the Biocyc database collection website (Caspi et al. 2016). Metabolic pathways were visualized using flow charts obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al. 2016). The Mfold webserver was used with default parameters to determine RNA secondary structures (Zuker 2003).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

This work was supported in part by the Australian Research Council discovery project grant number 150103715. E.G. is supported by a Monash University MBio Postgraduate Discovery Scholarship.

Received August 2, 2017; accepted February 1, 2018.

REFERENCES

Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EG, Margalit H, Altuvia S. 2001. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol* **11:** 941–950.

Baddal B, Muzzi A, Censini S, Calogero RA, Torricelli G, Guidotti S, Taddei AR, Covacci A, Pizza M, Rappuoli R, et al. 2015. Dual RNA-seq of nontypeable *haemophilus influenzae* and host cell transcriptomes reveals novel insights into host-pathogen cross talk. *MBio* **6:** e01765–15.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37:** W202–W208.

Bilusic I, Popitsch N, Rescheneder P, Schroeder R, Lybecker M. 2014. Revisiting the coding potential of the *E. coli* genome through Hfq co-immunoprecipitation. *RNA Biol* **11:** 641–654.

Bobrovskyy M, Vanderpool CK, Richards GR. 2015. Small RNAs regulate primary and secondary metabolism in Gram-negative bacteria. *Microbiol Spectr* **3**. doi: 10.1128/microbiolspec.MBP-0009-2014.

Bosch M, Garrido E, Llagostera M, Pérez de Rozas AM, Badiola I, Barbé J. 2002. *Pasteurella multocida exbB, exbD* and *tonB* genes are physically linked but independently transcribed. *FEMS Microbiol Lett* **210:** 201–208.

Bouvier M, Sharma CM, Mika F, Nierhaus KH, Vogel J. 2008. Small RNA binding to 5′ mRNA coding region inhibits translational initiation. *Mol Cell* **32:** 827–837.

Boyce JD, Adler B. 2006. How does *Pasteurella multocida* respond to the host environment? *Curr Opin Microbiol* **9:** 117–122.

Boyce JD, Wilkie I, Harper M, Paustian ML, Kapur V, Adler B. 2002. Genomic scale analysis of *Pasteurella multocida* gene expression

during growth within the natural chicken host. *Infect Immun* **70:** 6871–6879.

Boyce JD, Seemann T, Adler B, Harper M. 2012. Pathogenomics of *Pasteurella multocida*. *Curr Top Microbiol Immunol* **361:** 23–38.

Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, Bateman A. 2013. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res* **41:** D226–D232.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10:** 421.

Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, et al. 2016. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* **44:** D471–D480.

Chung JY, Wilkie I, Boyce JD, Townsend KM, Frost AJ, Ghoddusi M, Adler B. 2001. Role of capsule in the pathogenesis of fowl cholera caused by *Pasteurella multocida* serogroup A. *Infect Immun* **69:** 2487–2492.

Coornaert A, Chiaruttini C, Springer M, Guillier M. 2013. Post-transcriptional control of the *Escherichia coli* PhoQ-PhoP two-component system by multiple sRNAs involves a novel pairing region of GcvB. *PLoS Genet* **9:** e1003156.

Corcoran CP, Podkaminski D, Papenfort K, Urban JH, Hinton JC, Vogel J. 2012. Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol Microbiol* **84:** 428–445.

Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26:** 1367–1372.

Desnoyers G, Bouchard MP, Massé E. 2013. New insights into small RNA-dependent translational regulation in prokaryotes. *Trends Genet* **29:** 92–98.

Fuller TE, Kennedy MJ, Lowery DE. 2000. Identification of *Pasteurella multocida* virulence genes in a septicemic mouse model using signature-tagged mutagenesis. *Microb Pathog* **29:** 25–38.

Gottesman S, Storz G. 2011. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harb Perspect Biol* **3:** a003798.

Harper M, Cox AD, St Michael F, Wilkie IW, Boyce JD, Adler B. 2004. A heptosyltransferase mutant of *Pasteurella multocida* produces a truncated lipopolysaccharide structure and is attenuated in virulence. *Infect Immun* **72:** 3436–3443.

Harper M, St Michael F, John M, Vinogradov E, Steen JA, van Dorsten L, Steen JA, Turni C, Blackall PJ, Adler B, et al. 2013. *Pasteurella multocida* Heddleston serovar 3 and 4 strains share a common lipopolysaccharide biosynthesis locus but display both inter- and intrastrain lipopolysaccharide heterogeneity. *J Bacteriol* **195:** 4854–4864.

Hüttenhofer A, Noller HF. 1994. Footprinting mRNA-ribosome complexes with chemical probes. *EMBO J* **13:** 3892–3901.

Jin Y, Watt RM, Danchin A, Huang JD. 2009. Small noncoding RNA GcvB is a novel regulator of acid resistance in *Escherichia coli*. *BMC Genomics* **10:** 165.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44:** D457–D462.

Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, Lee TJ, Kaipa P, Gilham F, Spaulding A, Popescu L, et al. 2010. Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief Bioinform* **11:** 40–79.

Kulis-Horn RK, Persicke M, Kalinowski J. 2014. Histidine biosynthesis, its regulation and biotechnological application in *Corynebacterium glutamicum*. *Microb Biotechnol* **7:** 5–25.

Lloyd AL, Marshall BJ, Mee BJ. 2005. Identifying cloned *Helicobacter pylori* promoters by primer extension using a FAM-labelled primer and GeneScan analysis. *J Microbiol Methods* **60:** 291–298.

McArthur SD, Pulvermacher SC, Stauffer GV. 2006. The *Yersinia pestis gcvB* gene encodes two small regulatory RNA molecules. *BMC Microbiol* **6:** 52.

Mégroz M, Kleifeld O, Wright A, Powell D, Harrison P, Adler B, Harper M, Boyce JD. 2016. The RNA-binding chaperone Hfq is an important global regulator of gene expression in *Pasteurella multocida* and plays a crucial role in production of a number of virulence factors, including hyaluronic acid capsule. *Infect Immun* **84:** 1361–1370.

Melamed S, Peer A, Faigenbaum-Romm R, Gatt YE, Reiss N, Bar A, Altuvia Y, Argaman L, Margalit H. 2016. Global mapping of small RNA-target interactions in bacteria. *Mol Cell* **63:** 884–897.

Mika F, Hengge R. 2014. Small RNAs in the control of RpoS, CsgD, and biofilm architecture of *Escherichia coli*. *RNA Biol* **11:** 494–507.

Paley SM, Karp PD. 2006. The Pathway Tools cellular overview diagram and Omics Viewer. *Nucleic Acids Res* **34:** 3771–3778.

Pulvermacher SC, Stauffer LT, Stauffer GV. 2008. The role of the small regulatory RNA GcvB in GcvB/mRNA posttranscriptional regulation of *oppA* and *dppA* in *Escherichia coli*. *FEMS Microbiol Lett* **281:** 42–50.

Rafidah O, Zamri-Saad M, Shahirudin S, Nasip E. 2012. Efficacy of intranasal vaccination of field buffaloes against haemorrhagic septicaemia with a live *gdhA* derivative *Pasteurella multocida* B:2. *Vet Rec* **171:** 175.

Reitzer LJ. 1996. Ammonia assimilation and the biosynthesis of glutamine, glutamate, aspartate, asparagine, l-alanine, and d-alanine. In *Escherichia coli and Salmonella: cellular and molecular biology* (ed. Neidhardt FC, et al.), Vol. 1, pp. 391–407. ASM Press, Washington, DC.

Rolfe MD, Rice CJ, Lucchini S, Pin C, Thompson A, Cameron AD, Alston M, Stringer MF, Betts RP, Baranyi J, et al. 2012. Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *J Bacteriol* **194:** 686–701.

Rossi CC, Bossé JT, Li Y, Witney AA, Gould KA, Langford PR, Bazzolli DM. 2016. A computational strategy for the search of regulatory small RNAs in *Actinobacillus pleuropneumoniae*. *RNA* **22:** 1373–1385.

Sharma CM, Darfeuille F, Plantinga TH, Vogel J. 2007. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev* **21:** 2804–2817.

Sharma CM, Papenfort K, Pernitzsch SR, Mollenkopf HJ, Hinton JC, Vogel J. 2011. Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol Microbiol* **81:** 1144–1165.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7:** 539.

Steen JA, Steen JA, Harrison P, Seemann T, Wilkie I, Harper M, Adler B, Boyce JD. 2010. Fis is essential for capsule production in *Pasteurella multocida* and regulates expression of other important virulence factors. *PLoS Pathog* **6:** e1000750.

Storz G, Vogel J, Wassarman KM. 2011. Regulation by small RNAs in bacteria: expanding frontiers. *Mol Cell* **43:** 880–891.

Travers M, Paley SM, Shrager J, Holland TA, Karp PD. 2013. Groups: knowledge spreadsheets for symbolic biocomputing. *Database (Oxford)* **2013:** bat061.

Urban JH, Vogel J. 2007. Translational control and target recognition by *Escherichia coli* small RNAs *in vivo*. *Nucleic Acids Res* **35:** 1018–1037.

Urbanowski ML, Stauffer LT, Stauffer GV. 2000. The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*. *Mol Microbiol* **37:** 856–868.

Wilkie IW, Harper M, Boyce JD, Adler B. 2012. *Pasteurella multocida*: diseases and pathogenesis. *Curr Top Microbiol Immunol* **361:** 1–22.

Yang Q, Figueroa-Bossi N, Bossi L. 2014. Translation enhancing ACA motifs and their silencing by a bacterial small regulatory RNA. *PLoS Genet* **10:** e1004026.

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

Determination of the small RNA GcvB regulon in the Gram-negative bacterial pathogen *Pasteurella multocida* and identification of the GcvB seed binding region

Emily L. Gulliver, Amy Wright, Deanna Deveson Lucas, et al.

RNA 2018 24: 704-720 originally published online February 12, 2018
Access the most recent version at doi:10.1261/rna.063248.117

Appendix 2.

Proteins with increased production in either of the tested *P. multocida gcvB* mutant strains as compared to the VP161 wild-type parent. Protein production ratio in the GcvB overexpression strain is also shown for comparison.

Protein name [a, b]	VP161 locus tag (Pm70 locus tag)	Protein production ratio in the *gcvB* mutant (AL2677) (log_2)	FDR	Protein production ratio in the *gcvB* mutant with empty vector (AL2682) (log_2)	FDR	Protein production ratio in the GcvB overexpression strain (AL2684) (log_2)	FDR	Predicted protein function (Primary biochemical pathway/s)	Protein production ratio in the *hfq* mutant at early/mid-log growth phase (log_2) [c]	Ortholog controlled by GcvB in *E. coli* (E) and/or *S.* Typhimurium (S) [d]
ThrA	PMVP_0066 (PM0113)	0.59	0.007	0.47	0.042	-1.18	0.0001	Bifunctional aspartokinase I/homeserine dehydrogenase I (Isoleucine)	NC/NC	S
ArtI	PMVP_0077 (PM124)	0.99	0.002	0.57	0.059	-1.63	0.0001	Arginine ABC transporter (Transporter – arginine)	NC/NC	
DppA	**PMVP_0194 (PM0236)**	**1.89**	**0.001**	**1.87**	**0.005**	**-1.78**	**0.0002**	**Periplasmic dipeptide transport protein (Transporter- dipeptides)**	**2.59/2.44**	**E/S**
GltA	**PMVP_0236 (PM0276)**	**1.25**	**0.002**	**1.02**	**0.005**	**-1.58**	**0.0001**	**Citrate synthase (TCA cycle, Glutamate)**	**1.30/NC**	
BioB	**PMVP_0348 (PM0379)**	**0.85**	**0.005**	**0.64**	**0.030**	**-1.66**	**0.0001**	**Biotin synthase (Valine)**	**NC/NC**	
PM0472	**PMVP_0448 (PM0472)**	**1.08**	**0.002**	**0.74**	**0.007**	**-1.56**	**0.00003**	**PBP2_TAXI_TRAP_like_3 domain-containing protein**	**NC/NC**	
MsrA	**PMVP_0575 (PM0605)**	**0.65**	**0.02**	**0.73**	**0.012**	**-1.30**	**0.0001**	**Peptide methionine sulfoxide reductase (Methionine)**	**NC/NC**	
AspC	**PMVP_0593 (PM0621)**	**0.78**	**0.01**	**0.88**	**0.005**	**-1.38**	**0.0001**	**Aromatic amino acid aminotransferase (Tyrosine)**	**NC/NC**	
RsgA_2	PMVP_0639 (PM0667)	0.25	0.268	0.74	0.007	1.04	0.0002	Ferritin	NC/ 1.17	

MetC_2	PMVP_0791 (PM0794)	ND[e]	ND	1.04	0.013	-2.06	0.0001	Cystathionine beta-lyase (Methionine)	NC / NC	
PM0803	PMVP_0800 (PM0803)	0.06	0.484	0.67	0.035	1.37	0.0001	TonB dependent receptor C-terminal region	NC/ NC	
SpeF	PMVP_0802 (PM0806)	0.62	0.059	1.80	0.007	1.44	0.0031	Orthenine decarboxylase (Arginine)	NC / -3.09	
ArgG	PMVP_0809 (PM0813)	0.70	0.006	0.51	0.082	-1.63	0.0001	Arginosuccinate synthase (Arginine)	NC/NC	
PurC (HemH)	**PMVP_0811 (PM0815)**	**0.82**	**0.004**	**1.03**	**0.006**	**-1.52**	**0.0001**	**Phosphoribosylaminoimidazole-succinocarboxamide synthase (De novo purine nucleotide synthesis)**	**NC/1.99**	
HisH_1	**PMVP_0837 (PM0838)**	**1.05**	**0.002**	**1.17**	**0.007**	**-1.63**	**0.0002**	**Histidinol-phosphate aminotransferase (Histidine)**	**1.26/1.41**	
AroA	PMVP_0838 (PM0839)	0.61	0.008	0.37	0.200	-1.03	0.0007	3-phosphoshikimate 1-carboxyvinyltransferase (Tyrosine)	NC/NC	
RcpA	PMVP_0851 (PM0852)	ND	ND	0.88	0.008	1.16	0.0002	Type II/IV secretion system secretin	1.39/ NC	
PcnB	PMVP_0865 (PM0864)	0.27	0.059	0.68	0.035	-0.78	0.0033	Poly (A) polymerase	NC/ NC	
LysC	**PMVP_0948 (PM0937)**	**1.07**	**0.004**	**1.27**	**0.005**	**-1.66**	**0.0001**	**Aspartate kinase (Lysine, threonine, methionine, homoserine, isoleucine)**	**1.33/1.21**	**E**
MetB	**PMVP_1008 (PM0995)**	**0.79**	**0.047**	**0.99**	**0.035**	**-2.11**	**0.0001**	**Cystathionine gamma-synthase (Methionine, lysine, threonine, homoserine)**	**1.08/NC**	
DapA	PMVP_1069 (PM1051)	0.83	0.003	0.86	0.058	-1.67	0.0005	Dihydrodipicolinate synthase (Lysine, threonine, methionine)	1.11/1.01	
PM1128	PMVP_1145	ND	ND	1.27	0.005	1.19	0.0004	Dithiol-disulfide isomerase	NC/ NC	

	(PM1128)									
HisG	PMVP_1213 (PM1195)	0.62	0.167	0.75	0.044	-0.57	0.0313	ATP phosphoribosyltransferase (Histidine)	NC/ NC	
HisC	**PMVP_1220 (PM1199)**	**0.64**	**0.03**	**0.77**	**0.035**	**-1.11**	**0.0010**	**Histidinol-phosphate aminotransferase (Histidine)**	**1.03/NC**	
HisB	**PMVP_1221 (PM1200)**	**1.22**	**0.002**	**1.21**	**0.005**	**-1.33**	**0.0002**	**Histidinol-phosphatase (Histidine)**	**NC/NC**	
HisA	PMVP_1224 (PM1203)	0.73	0.008	0.48	0.275	-2.02	0.0002	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (Histidine)	NC/NC	
HisF	**PMVP_1225 (PM1204)**	**0.71**	**0.003**	**0.72**	**0.035**	**-1.51**	**0.0001**	**Imidazoleglycerol phosphate synthase, cyclase subunit (Histidine)**	**NC/NC**	
GlpQ	PMVP_1495 (PM1444)	0.31	0.352	0.77	0.007	0.97	0.0002	Glycerophosphodiester phosphodiesterase	NC/ NC	
IlvG	PMVP_1683 (PM1628)	0.35	0.217	0.71	0.005	0.07	0.6505	Acetolactate synthase (Valine, Leucine, Isoleucine)	1.16/ NC	
Asd	**PMVP_1687 (PM1623)**	**0.89**	**0.008**	**0.63**	**0.035**	**-1.08**	**0.0004**	**Aspartate-semialdehyde dehydrogenase (Lysine, threonine, methionine, homoserine, isoleucine)**	**1.03/NC**	
SerA	**PMVP_1723 (PM1671)**	**0.73**	**0.004**	**0.82**	**0.035**	**-1.20**	**0.0009**	**D-3-phosphoglycerate dehydrogenase (Serine, cyctine and Glycine)**	**NC/NC**	**E/S**
PlpB	**PMVP_1787 (PM1730)**	**1.39**	**0.002**	**1.22**	**0.009**	**-1.28**	**0.0010**	**Outer membrane lipoprotein (Transporter -Methionine)**	**1.32/1.09**	**S**
PM1791	PMVP_1841 (PM1791)	0.35	0.615	0.62	0.026	1.68	0.00003	High affinity choline transporter	NC/ NC	

OppB	PMVP_1961 (PM1909)	0.63	0.004	0.20	0.591	-1.67	0.0002	Oligopeptide transport system permease protein OppB (Transporter-oligopeptides)	NC/NC	E
OppA	**PMVP_1962 (PM1910)**	**1.31**	**0.002**	**1.09**	**0.005**	**-1.55**	**0.0001**	**Periplasmic oligopeptides binding protein (Transporter-oligopeptides)**	**1.26/1.17**	**E/S**
GdhA	PMVP_2095 (PM0043)	2.27	0.001	2.48	0.104	-3.46	0.0079	Glutamate dehydrogenase (Glutamate synthesis, TCA cycle, Nitrate reduction)	2.0/2.7	E/S

[a] Differentially expressed proteins were defined as those showing at least 1.5-fold increased production ($\log_2 \geq 0.59$) with a false discovery rate (FDR) of less than 0.05.

[b] Rows in bold indicate those proteins identified as statistically differentially expressed in both *gcvB* mutant strains.

[c] Relative protein production in the *P. multocida* VP161 *hfq* mutant compared to the wild-type strain. *P. multocida hfq* mutant protein production data is taken from Mégroz et al. (Mégroz et al. 2016). NC = no change in protein production.

[d] Proteins with orthologs known to be controlled by GcvB in *E. coli* and/or *S.* Typhimurium are indicated with an (E) and/or an (S), respectively. *E. coli* and *S.* Typhimurium *gcvB* mutant protein production data is taken from Sharma *et al.* (Sharma et al. 2011) and Pulvermacher *et al.* (Pulvermacher et al. 2009).

[e] ND, no data available

Appendix 3.

Proteins with decreased production in either of the tested *P. multocida gcvB* mutant strains as compared to the VP161 wild-type parent. Protein production ratio in the GcvB overexpression strain is also shown for comparison.

Protein name[a]	**VP161 locus tag (PM70 locus tag)**	**Protein production ratio in the *gcvB* mutant (AL2677) (log_2)**	**FDR**	**Protein production ratio in the *gcvB* mutant with empty vector (AL2682) (log_2)**	**FDR**	**Protein production ratio in the *gcvB* overexpression strain (AL2684) (log_2)**	**FDR**	**Predicted protein function (Primary biochemical pathway/s)**	**Protein production ratio in the *hfq* mutant at early/mid-log growth phase (log_2)** [b]	**Ortholog controlled by GcvB in *E. coli* (E) or *S.* Typhimurium (S)** [c]
ArtP	PMVP_0076 (PM0123)	-0.07	0.394	-0.61	0.034	0.33	0.077	arginine ABC transporter, ATP-binding protein (arginine)	NC/NC	
PqqL	PMVP_0801 (PM0804)	ND[d]	ND	-1.47	0.035	-0.69	0.134	Zinc protease	NC/NC	
Tpl	PMVP_0807 (PM0811)	-0.73	0.014	0.66	0.151	1.41	0.002	tyrosine phenol-lyase (tyrosine)	NC/NC	
CysS	PMVP_0957 (PM0945)	-0.04	0.451	-0.66	0.005	-0.52	0.002	cysteinyl-tRNA synthetase (fatty acid/ lipid biosynthesis)	NC/NC	
ArgC	PMVP_1142 (PM1118)	0.06	0.721	-0.65	0.035	-1.00	0.001	n-acetyl-gamma-glutamyl-phosphate reductase (arginine)	-1.44/NC	
PM1217	PMVP_1239 (PM1217)	-0.77	0.020	-0.10	0.814	0.05	0.863	Very similar to cellulose synthase catalytic	2.27/NC	

								subunit, putative	
PM1266	PMVP_1290 (PM1266)	ND	ND	-1.01	0.026	-0.56	0.053	ABC transporter, ATP-binding protein (inorganic ion transport)	NC/NC
PM1682	PMVP_1734 (PM1682)	ND	ND	-0.76	0.012	-0.80	0.001	hypothetical protein PM1682	NC/NC
PM1790	PMVP_1840 (PM1790)	-0.23	0.136	-0.62	0.035	-0.83	0.001	conserved hypothetical protein (electron transport, TCA cycle)	NC/NC
LeuC	PMVP_1999 (PM1960)	ND	ND	-1.11	0.035	-1.15	0.005	3-isopropylmalate dehydratase, large subunit (leucine)	1.15/NC

[a] Differentially expressed proteins were defined as those showing at least 1.5-fold decreased production ($\log_2 \leq -0.59$) with a false discovery rate (FDR) of less than 0.05.

[b] Relative protein production in the *P. multocida* VP161 *hfq* mutant compared to the wild-type strain. *P. multocida hfq* mutant protein production data is taken from Mégroz et al. (Mégroz et al. 2016). NC = no change in protein production.

[c] Proteins with orthologs known to be controlled by GcvB in *E. coli* and/or *S.* Typhimurium are indicated with an (E) and/or an (S), respectively. *E. coli* and *S.* Typhimurium *gcvB* mutant protein production data is taken from Sharma *et al.* (Sharma et al. 2011) and Pulvermacher *et al.* (Pulvermacher et al. 2009).

[d] ND, no data available

Appendix 4.

Proteins with increased production in the *P. multocida* GcvB overexpression strain (AL2684) as compared to the VP161 wild-type parent.

Protein name	**PMVP Locus Tag (PM70 Locus Tag)**	**Protein production ratio in the *gcvB* overexpression strain (AL2684) (log_2)** [a]	**FDR**	**Predicted protein function**
PflB	PMVP_0022 (PM0075)	0.63	0.0017	Formate acetyltransferase
PM0134	PMVP_0088 (PM0134)	0.75	0.0028	Conserved hypothetical protein
FtsI	PMVP_0090 (PM0136)	0.79	0.0003	Penicillin-binding protein 3
PM0211	PMVP_0166 (PM0211)	0.97	0.0009	Conserved hypothetical protein
DcuC	PMVP_0188 (PM0230)	1.49	0.00001	C4-dicarboxylate anaerobic carrier
PM0270	PMVP_0229 (PM0270)	0.62	0.0013	Hypothetical protein PM0270
LldD	PMVP_0248 (PM0288)	1.22	0.0002	Alpha-hydroxy-acid oxidizing enzyme
PMVP_0261	PMVP_0261 (NA[b])	2.08	0.0001	Hypothetical
PM0305	PMVP_0272 (PM0305)	0.82	0.0016	Hypothetical protein PM0305
PM0306	PMVP_0273 (PM0306)	1.27	0.0004	Hypothetical protein PM0306
PM0307	PMVP_0274 (PM0307)	1.25	0.0006	Hypothetical protein PM0307
PM0336	PMVP_0302 (PM0336)	0.73	0.0017	Hemoglobin binding protein B
PM0337 PS[c]	PMVP_0304 (PM0337)	1.41	0.0001	Hemoglobin binding protein B
PM0337 PS[c]	PMVP_0305 (PM0337)	1.44	0.0000	Hemoglobin binding protein B
Grx2	PMVP_0487 (PM0518)	0.91	0.0040	Glutaredoxin 2
TorA	PMVP_0506 (PM1793)	1.21	0.0008	Trimethylamine-N-oxide reductase
PM0537	PMVP_0510 (PM0537)	0.88	0.0045	Putative sulfite oxidase subunit
GcvA	PMVP_0542 (PM0567)	1.80	0.0001	Glycine cleavage system transcriptional activator
PM0568	PMVP_0543 (PM0568)	1.50	0.0001	Putative RNA 2'-O-ribose methyltransferase
HemR	PMVP_0551 (PM0576)	0.88	0.0010	TonB dependent receptor

HbpA	PMVP_0566 (PM0592)	0.79	0.0024	ABC-transporter
PM0612	PMVP_0584 (PM0612)	0.86	0.0043	Conserved hypothetical protein
PM0652	PMVP_0624 (PM0652)	0.77	0.0019	Oxidoreductase, Gfo/Idh/moca family
RsgA_2	PMVP_0639 (PM0667)	1.05	0.0002	Ferritin
DnaJ	PMVP_0719 (PM0740)	0.91	0.0002	Chaperone protein
PM0741	PMVP_0720 (PM0741)	1.22	0.0000	Hypothetical protein PM0741
PM0803	PMVP_0800 (PM0803)	1.37	0.0001	TonB dependent receptor C-terminal region subfamily
SpeF	PMVP_0802 (PM0806)	1.44	0.0031	Ornithine decarboxylase
Tpl	PMVP_0807 (PM0811)	1.41	0.0016	Tyrosine phenol-lyase
RcpA	PMVP_0851 (PM0852)	1.17	0.0002	Type II/IV secretion system secretin
PM0853	PMVP_0852 (PM0853)	0.71	0.0015	Conserved hypothetical protein
PM0903	PMVP_0906 (PM0903)	0.75	0.0008	N-acetylmuramoyl-L-alanine amidase
Hfq	PMVP_0909 (PM0906)	1.51	0.0045	RNA-binding protein
PM1325	PMVP_0925 (PM1325)	1.01	0.0007	Hypothetical protein PM1325
PM0928	PMVP_0939 (PM0928)	1.03	0.0006	Membrane-bound lytic transglycosylase A
PM0979	PMVP_0992 (PM0979)	1.04	0.0002	Hypothetical protein PM0979
MglB	PMVP_1053 (PM1038)	1.07	0.0010	Galactose ABC transporter, periplasmic-binding protein
PM1069	PMVP_1089 (PM1069)	0.99	0.0045	Hypothetical protein PM1069
DeaD	PMVP_1134 (PM1112)	1.08	0.0001	ATP-dependent RNA helicase
Pnp	PMVP_1136 (PM1114)	0.82	0.0003	Polynucleotide phosphorylase/polyadenylase
PM1128	PMVP_1145 (PM1128)	1.19	0.0004	Dithiol-disulfide isomerase
PM1211	PMVP_1233 (PM1211)	1.19	0.0001	Transglutaminase-like superfamily domain protein
AraD	PMVP_1266 (PM1244)	0.71	0.0045	L-ribulose-5-phosphate 4-epimerase
PM1246	PMVP_1268 (PM1246)	0.72	0.0008	Hexulose-6-phosphate synthase sgbh, putative

SpeE	PMVP_1422 (PM1381)	0.60	0.0017	SpeE
TnaA	PMVP_1470 (PM1420)	1.71	0.0001	Tryptophanase
GlpQ	PMVP_1495 (PM1444)	0.97	0.0002	Glycerophosphodiester phosphodiesterase
PM1457	PMVP_1508 (PM1457)	1.31	0.0028	Periplasmic iron-binding protein
CarA	PMVP_1557 (PM1502)	0.86	0.0041	Carbamoyl-phosphate synthase, small subunit
PM1677	PMVP_1729 (PM1677)	0.91	0.0033	Transmembrane flavin adenine dinucleotide binding protein
Trx	PMVP_1760 (PM1705)	1.61	0.0001	Thioredoxin
PM1791	PMVP_1841 (PM1791)	1.68	0.0000	High-affinity choline transport protein
NarP	PMVP_1863 (PM1810)	0.70	0.0023	Nitrate/nitrite response regulator protein
RpL32	PMVP_1964 (PM1912)	0.82	0.0003	50S ribosomal protein L32
PM1928	PMVP_1980 (PM1928)	0.66	0.0033	Conserved hypothetical protein
PM0016	PMVP_2069 (PM0016)	0.69	0.0028	Hypothetical protein PM0016
PM0042	PMVP_2094 (PM0042)	1.51	0.0006	Conserved hypothetical protein

[a] Differentially expressed proteins were defined as those showing at least 1.5-fold increased production ($\log_2 \geq 0.59$) with a false discovery rate (FDR) of less than 0.05.

[b] NA, not applicable as gene not present in PM70 genome

[c] PS, pseudogene in VP161

Appendix 5.

Proteins with decreased production in the *P. multocida* GcvB overexpression strain as compared to the VP161 wild-type parent.

Protein name	**PMVP Locus Tag (PM70 Locus Tag)**	**Protein production ratio in the *gcvB* overexpression strain (AL2684) (log_2) [a]**	**FDR**	**Predicted protein function**
ThrA	PMVP_0066 (PM0113)	-1.18	0.0001	Bifunctional aspartokinase I/homeserine dehydrogenase I
PmbA	PMVP_0073 (PM0120)	-1.02	0.0001	Metalloprotease
ArtI	PMVP_0077 (PM0124)	-1.63	0.0001	Arginine ABC transporter
RecR	PMVP_0160 (PM0206)	-0.80	0.0010	Recombination protein RecR
SecF	PMVP_0184 (PM0226)	-0.82	0.0002	Preprotein translocase subunit SecF
YajC	PMVP_0186 (PM0228)	-0.93	0.0004	Preprotein translocase, YajC subunit
DppA	PMVP_0194 (PM0236)	-1.78	0.0002	Periplasmic dipeptide transport protein
PotD_1	PMVP_0218 (PM0260)	-0.74	0.0028	Spermidine/putrescine ABC transporter, periplasmic-binding protein
PotD_2	PMVP_0219 (PM0261)	-0.69	0.0045	Spermidine/putrescine ABC transporter, periplasmic-binding protein
GltA	PMVP_0236 (PM0276)	-1.58	0.0001	Citrate synthase
AccA	PMVP_0252 (PM0292)	-1.11	0.0001	Acetyl-coa carboxylase, carboxyl transferase subunit alpha
MetG	PMVP_0268 (PM0303)	-0.82	0.0002	Methionyl-tRNA synthetase
RecN	PMVP_0298 (PM0332)	-1.97	0.0001	DNA repair protein
BioB	PMVP_0348 (PM0379)	-1.66	0.0001	Biotin synthase
FdhE	PMVP_0376 (PM0405)	-1.03	0.0032	Formate dehydrogenase accessory protein
Pgi	PMVP_0388 (PM0416)	-0.82	0.0004	Glucose-6-phosphate isomerase
PM0472	PMVP_0448 (PM0472)	-1.56	0.0000	PBP2_TAXI_TRAP_like_3 domain containing protein
PM0476	PMVP_0452 (PM0476)	-0.95	0.0009	7-cyano-7-deazaguanine reductase
PM0478	PMVP_0454 (PM0478)	-1.86	0.0001	Conserved hypothetical protein

IvlE	PMVP_0541 (PM0566)	-1.13	0.0005	Branched-chain amino acid aminotransferase
TrxB	PMVP_0548 (PM0573)	-0.89	0.0013	Thioredoxin reductase
Plp4	PMVP_0561 (PM0586)	-0.62	0.0045	Lipoprotein (mlp)
MsrA	PMVP_0575 (PM0605)	-1.30	0.0001	Peptide methionine sulfoxide reductase
PepN	PMVP_0590 (PM0618)	-1.01	0.0001	Aminopeptidase N
AspC	PMVP_0593 (PM0621)	-1.38	0.0001	Aromatic amino acid aminotransferase
OsmY	PMVP_0621 (PM0649)	-1.10	0.0004	Osmotically-inducible protein
MsmB	PMVP_0628 (PM0655)	-0.90	0.0036	Cold shock-like protein cspc-related protein
AroF	PMVP_0637 (PM0665)	-1.01	0.0029	Phospho-2-dehydro-3-deoxyheptonate aldolase
PM0675	PMVP_0648 (PM0675)	-0.68	0.0023	N-acetyl-D-glucosamine kinase
PM0696	PMVP_0671 (PM0696)	-1.12	0.0010	SNF2 and others N-terminal domain subfamily
NrdB	PMVP_0698 (PM0719)	-1.24	0.0001	Ribonucleotide-diphosphate reductase subunit beta
RbfA	PMVP_0738 (PM0757)	-0.73	0.0012	Ribosome-binding factor A
PM0785	PMVP_0778 (PM0785)	-0.86	0.0005	LemA_like domain protein
PM0787	PMVP_0780 (PM0787)	-1.41	0.0001	Hypothetical protein PM0787
MetC_2	PMVP_0791 (PM0794)	-2.06	0.0001	Cystathionine beta-lyase
ArgG	PMVP_0809 (PM0813)	-1.63	0.0001	Argininosuccinate synthase
PurC	PMVP_0811 (PM0815)	-1.52	0.0001	Phosphoribosylaminoimidazole-succinocarboxamide synthase
PrfC	PMVP_0814 (PM0816)	-0.72	0.0008	Peptide chain release factor 3
SerC	PMVP_0836 (PM0837)	-1.21	0.0001	Phosphoserine aminotransferase
HisH_1	PMVP_0837 (PM0838)	-1.63	0.0002	Histidinol-phosphate aminotransferase
AroA	PMVP_0838 (PM0839)	-1.03	0.0007	3-phosphoshikimate 1-carboxyvinyltransferase
PcnB	PMVP_0865 (PM0864)	-0.78	0.0033	Poly(A) polymerase
FolK	PMVP_0866 (PM0865)	-1.19	0.0021	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase

ProA	PMVP_0947 (PM0936)	-0.75	0.0006	Gamma-glutamyl phosphate reductase
LysC	PMVP_0948 (PM0937)	-1.66	0.0001	Aspartate kinase
AspS	PMVP_0996 (PM0983)	-0.69	0.0011	Aspartyl-tRNA synthetase
TrxM	PMVP_1007 (PM0994)	-1.40	0.0001	Thioredoxin
MetB	PMVP_1008 (PM0995)	-2.11	0.0001	Cystathionine gamma-synthase
PM0996	PMVP_1010 (PM0996)	-0.97	0.0022	ABC transporter, ATP-binding protein
DapA	PMVP_1069 (PM1051)	-1.67	0.0005	Dihydrodipicolinate synthase
Pur	PMVP_1071 (PM1053)	-1.29	0.0019	Phosphoribosylglycinamide formyltransferase 2
AspA	PMVP_1124 (PM1103)	-0.68	0.0009	Aspartate ammonia-lyase
ArgC	PMVP_1142 (PM1118)	-1.00	0.0007	N-acetyl-gamma-glutamyl-phosphate reductase
PcgA	PMVP_1157 (NA[b])	-0.59	0.0026	Choline kinase
Crp	PMVP_1174 (PM1157)	-0.72	0.0009	Camp-regulatory protein
RpL34	PMVP_1179 (PM1162)	-1.12	0.0010	50S ribosomal protein L34
GlnA	PMVP_1193 (PM1175)	-0.95	0.0007	Glutamine synthetase, type I
HisC	PMVP_1220 (PM1199)	-1.11	0.0010	Histidinol-phosphate aminotransferase
HisB	PMVP_1221 (PM1200)	-1.33	0.0002	Imidazoleglycerol-phosphate dehydratase / histidinol-phosphatase
HisA	PMVP_1224 (PM1203)	-2.02	0.0002	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase
HisF	PMVP_1225 (PM1204)	-1.51	0.0001	Imidazoleglycerol phosphate synthase, cyclase subunit
HisIE	PMVP_1228 (PM1206)	-0.81	0.0019	Histidine biosynthesis bifunctional protein
PM1243	PMVP_1265 (PM1243)	-1.21	0.0008	Conserved hypothetical protein subfamily
ThiE	PMVP_1283 (PM1260)	-1.58	0.0009	Thiamine-phosphate pyrophosphorylase
PM1264	PMVP_1288 (PM1264)	-1.77	0.0002	Hypothetical protein PM1264
IlvC	PMVP_1309 (PM1284)	-1.00	0.0029	Ketol-acid reductoisomerase
ApbE	PMVP_1363 (PM1334)	-1.38	0.0000	FAD: protein FMN transferase

RbsA_2	PMVP_1419 (PM1379)	-0.62	0.0023	Ribose transport ATP-binding protein
Nqr6	PMVP_1420 (NA)	-1.32	0.0005	Nqr6 subunit of Na-translocating NADH-quinone reductase complex beta-subunit
PM1469	PMVP_1520 (PM1469)	-1.12	0.0010	Hypothetical protein PM1469
PM1470	PMVP_1521 (PM1470)	-1.41	0.0045	Conserved hypothetical protein
PM1500	PMVP_1555 (PM1500)	-1.27	0.0013	Conserved hypothetical protein
Zwf	PMVP_1601 (PM1549)	-0.95	0.0002	Glucose-6-phosphate 1-dehydrogenase
DevB	PMVP_1602 (PM1550)	-0.61	0.0029	6-phosphogluconolactonase
PM1590	PMVP_1645 (PM1590)	-0.64	0.0007	Hypothetical protein PM1590
NapD	PMVP_1648 (PM1593)	-1.34	0.0007	Nitrate reductase
NapA	PMVP_1649 (PM1594)	-0.63	0.0025	Periplasmic nitrate reductase precursor
NapB	PMVP_1652 (PM1597)	-2.06	0.0004	Periplasmic nitrate reductase
Tal_1	PMVP_1657 (PM1602)	-1.64	0.0000	Transaldolase B
Idp	PMVP_1661 (PM1606)	-1.22	0.0001	Isocitrate dehydrogenase, NADP-dependent
PM1626	PMVP_1681 (PM1626)	-0.80	0.0021	Hypothetical protein PM1626
Asd	PMVP_1687 (PM1632)	-1.08	0.0004	Aspartate-semialdehyde dehydrogenase
SerB	PMVP_1708 (PM1657)	-1.21	0.0001	Phosphoserine phosphatase
SerA	PMVP_1723 (PM1671)	-1.20	0.0008	D-3-phosphoglycerate dehydrogenase
PM1682	PMVP_1734 (PM1682)	-0.80	0.0013	Hypothetical protein PM1682
PM1702	PMVP_1755 (PM1702)	-0.67	0.0049	Conserved hypothetical protein
ClpB	PMVP_1759 (PM1704)	-1.06	0.0025	ATP-dependent Clp protease, atpase subunit
PlpB	PMVP_1787 (PM1730)	-1.28	0.0010	Outer membrane lipoprotein
PM1790	PMVP_1840 (PM1790)	-0.83	0.0014	Conserved hypothetical protein
PM1805	PMVP_1857 (PM1805)	-0.91	0.0013	Hypothetical protein PM1805
Pgk	PMVP_1913 (PM1860)	-0.75	0.0005	Phosphoglycerate kinases

Fba	PMVP_1914 (PM1861)	-1.52	0.0001	Fructose-bisphosphate aldolase
PM1874	PMVP_1926 (PM1874)	-0.79	0.0002	Conserved hypothetical protein
PM1875	PMVP_1927 (PM1875)	-1.62	0.0001	Conserved hypothetical protein
ProB	PMVP_1948 (PM1896)	-0.77	0.0006	Gamma-glutamyl kinase
Psd	PMVP_1951 (PM1899)	-1.17	0.0004	Phosphatidylserine decarboxylase proenzyme
OppF	PMVP_1958 (PM1906)	-1.29	0.0008	Oligopeptide ABC transporter, ATP-binding protein
OppD	PMVP_1959 (PM1907)	-0.97	0.0040	Oligopeptide transport ATP-binding protein
OppB	PMVP_1961 (PM1909)	-1.67	0.0002	Oligopeptide transport system permease protein
OppA	PMVP_1962 (PM1910)	-1.55	0.0001	Periplasmic oligopeptide-binding protein
FbpA	PMVP_2104 (PM0051)	-1.01	0.0024	Iron binding protein

[a] Differentially expressed proteins were defined as those showing at least 1.5-fold decreased production ($\log_2 \leq -0.59$) with a false discovery rate (FDR) of less than 0.05.

[b] NA, not applicable as gene not present in PM70 genome

Appendix 6.

Transcripts identified to bind to ProQ through UV-CLASH.

VP161 locus tag (PM70 locus tag)	Gene name	Predicted product	RNA Type	General Function prediction
PMVP_0063, (NP)		Hypothetical	mRNA	No function prediction
PMVP_0087, (PM0133)	*mraZ*	Transcriptional regulator MraZ	mRNA	Cell cycle control, mitosis and meiosis genes
PMVP_0167, (PM0212)	PM0212	Conserved hypothetical protein	mRNA	No function prediction
PMVP_0172, (PM0217)	*smpB*	SsrA-binding protein	mRNA	Posttranslational modification, protein turnover, chaperones genes
PMVP_0252, (PM0292)	*accA*	Acetyl-CoA carboxylase, carboxyl transferase subunit alpha	mRNA	Lipid transport and metabolism genes
PMVP_0310, (PM0342)	*cafA*	Cytoplasmic axial filament protein	Antisense mRNA	Translation genes
PMVP_0316, (PM0347)	rpL21	50S ribosomal protein L21	mRNA	Translation genes
PMVP_0415, (PM0442)	PM0442	Hypothetical	mRNA	No function prediction
PMVP_0528, (PM0554)	*lpp*	Glycine zipper TM2 domain containing protein	mRNA	Cell wall/membrane biogenesis genes
PMVP_0573, (PM0604)	rpL20	50S ribosomal protein L20	mRNA	Translation genes
PMVP_0599, (PM0627)	*nlpC*	Lipoprotein	mRNA	Cell wall/membrane biogenesis genes
PMVP_0625, (PM0653)	*pykA*	Pyruvate kinase	3’ UTR	Carbohydrate transport and metabolism genes
PMVP_0647, (PM0674)	PM0674	Hypothetical protein PM0674	mRNA	No function prediction
PMVP_0657, (PM_t18)	tRNA Met	tRNA Met	tRNA	Translation genes
PMVP_0690, (NA)		Hypothetical	mRNA	No function prediction
PMVP_0696, (PM0717)	*nrdA*	Ribonucleotide-diphosphate reductase subunit alpha	mRNA	Nucleotide transport and metabolism genes
PMVP_0779, (PM0786)	*ompA*	Outer membrane protein OmpA	mRNA	Cell wall/membrane biogenesis genes
PMVP_0792, (PM0795)	*tsaA*	Antioxidant, AhpC/Tsa family	mRNA	Posttranslational modification, protein turnover, chaperones genes
PMVP_0797, (PM0800)	*himD*	Integration host factor, beta subunit	mRNA	Replication, recombination and repair genes
PMVP_0898, (PM0894)	*dlaT*	Dihydrolipoamide acetyltransferase	mRNA	Energy production and conversion genes
PMVP_0910, (PM0907)	*hflX*	GTPase	mRNA	No function prediction
PMVP_0949, (PM0938)	*purA*	Adenylosuccinate synthetase	3’ UTR	Nucleotide transport and metabolism genes

PMVP_0979, (PM0966)	*pal*	Peptidoglycan associated lipoprotein Pal	mRNA	Cell wall/membrane biogenesis genes
PMVP_0992, (PM0979)	PM0979	Membrane protein	mRNA	No function prediction
PMVP_0993, (PM0980)	PM0980	YebC/PmpR family DNA binding transcriptional regulator	mRNA	No function prediction
PMVP_1013, (PM0999)	PM0999	Putative nicotinate phosphoribosyltransferase	3' UTR	Coenzyme transport and metabolism genes
PMVP_1065, (PM_t32)	tRNA Cys	tRNA Cys	tRNA	Translation genes
PMVP_1077, (PM_t36)	tRNA Ser	tRNA Ser	tRNA	Translation genes
PMVP_1129, (PM1107)	*groEL*	60 kd chaperonin	mRNA	Posttranslational modification, protein turnover, chaperones genes
PMVP_1137, (NA)	tRNA	tRNA	tRNA	No function prediction
PMVP_1185 (NA)-1186 (PM1168)	Hypothetic al - menG		mRNA	
PMVP_1195, (PM1177)	rpL9	50S ribosomal protein L9	mRNA	Translation genes
PMVP_1196, (PM1178)	rpS18	30S ribosomal protein S18	mRNA	Translation genes
PMVP_1241, (PM1219)	*secA*	Preprotein translocase, SecA subunit	3' UTR	Intracellular trafficking and secretion genes
PMVP_1246, (PM1224)	*aroK*	Shikimate kinase	mRNA	Amino acid transport and metabolism genes
PMVP_1262, (PM1239)	rpS21	30S ribosomal protein S21	mRNA	Translation genes
PMVP_1266, (PM1244)	*araD*	L-ribulose-5-phosphate 4-epimerase	mRNA	Carbohydrate transport and metabolism genes
PMVP_1319 (PM1294a)-1320 (PM1296)	*rpsP-rimM*	30S ribosomal protein S16 - 16s rRNA processing protein rimm	mRNA	Translation genes
PMVP_1348, (PM1321)	*mltC*	Membrane-bound lytic murein transglycosylase C	mRNA	Cell wall/membrane biogenesis genes
PMVP_1362, (NA)		Hypothetical	mRNA	No function prediction
PMVP_1379, (PM1349)	*slyD*	FkbP-type peptidyl-prolyl cis-trans isomerase	mRNA	Posttranslational modification, protein turnover, chaperones genes
PMVP_1388, (PM1357)	*tufA*	Elongation factor Tu	mRNA	Translation genes
PMVP_1431, (PM1389)	rpL17	50S ribosomal protein L17	3' UTR	Translation genes
PMVP_1432, (PM1390)	*rpoA*	DNA-directed RNA polymerase subunit alpha	mRNA	Transcription genes
PMVP_1435, (PM1393)	rpS13	Ribosomal protein S13/S18	mRNA	Translation genes

PMVP_1439 (PM1398)-1440 (PM1399)	rpS5 -rpL18	30S ribosomal protein S5 - 50S ribosomal protein L18	mRNA	Translation genes
PMVP_1444, (PM1403)	rpL5	50S ribosomal protein L5	mRNA	Translation genes
PMVP_1445 (PM1405)-1446 (PM1406)	rpL24-rpL14	50S ribosomal protein L24-L14	mRNA	Translation genes
PMVP_1457 (PM1408)-1458 (PM1409)	rpL16-rpS3	50S ribosomal protein L16-30S ribosomal protein S3	mRNA	Translation genes
PMVP_1460 (PM1411)-1461 (PM1412)	rpS19-rpL2	30S ribosomal protein S19-50S ribosomal protein L2	mRNA	Translation genes
PMVP_1462, (PM1413)	rpL23	50S ribosomal protein L23	mRNA	Translation genes
PMVP_1464 (PM1415)-1465 (PM1416)	rpL3-rpS10	50S ribosomal protein L3-30S ribosomal protein S10	mRNA	Translation genes
PMVP_1532, (PM1480)	PM1480	Rhodanese-like domain protein	mRNA	Inorganic ion transport and metabolism genes
PMVP_1707, (PM1656)	PM1656	YajQ family cyclic di-GMP-binding protein	mRNA	No function prediction
PMVP_1746, (PM1693)	*cysK*	Cystine synthase A	mRNA	Amino acid transport and metabolism genes
PMVP_1782, (PM_t48)	tRNA Trp	tRNA Trp	tRNA	Translation genes
PMVP_1790, (PM1732)	*hupA*	DNA-binding protein HU-alpha	mRNA	Replication, recombination and repair genes
PMVP_1800, (PM1742)	rpL1	50S ribosomal protein L1	mRNA	Translation genes
PMVP_1801, (PM1743)	rpL11	50S ribosomal protein L11	mRNA	Translation genes
PMVP_1838 (PM1788)-1839 (PM1789)	*mclA-rpoE*	Sigma-E factor negative regulatory protein-RNA polymerase sigma factor RpoE	mRNA	
PMVP_1964, (PM1912)	rpL32	50S ribosomal protein L32	mRNA	Translation genes
PMVP_1986, (PM1933)	*folD*	Methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase	mRNA	Coenzyme transport and metabolism genes
PMVP_2022, (PM1984)	rpS2	30S ribosomal protein S2	mRNA	Translation genes
PMVP_2023, (PM1985)	*tsf*	Translation elongation factor Ts	mRNA	Translation genes

PMVP_2069, (PM0016)	PM0016	Hypothetical protein PM0016	mRNA	No function prediction
Prrc37, (NA)	Prrc37	Hypothetical (see Table 1.1)		No function prediction
Prrc54, (NA)	Prrc54	sRNA	sRNA	No function prediction
Prrc55, (NA)	Prrc55	sRNA	sRNA	No function prediction
Prrc56, (NA)	Prrc56	sRNA	sRNA	No function prediction
Prrc57, (NA)	Prrc57	sRNA	sRNA	No function prediction
Prrc58, (PM_t25)	tRNA Met	tRNA Met	tRNA	Translation genes
Prrc59, (PM_t41)	tRNA Val	tRNA Val	tRNA	Translation genes