Texas Medical Center Library
DigitalCommons@The Texas Medical Center

UT GSBS Dissertations and Theses (Open Access) Graduate School of Biomedical Sciences

5-2011

MULTIPLE POSTTRANSCRIPTIONAL REGULATORY FEATURES CONTROL EXPRESSION OF ETHANOLAMINE UTILIZATION GENES IN ENTEROCOCCUS FAECALIS

Kristina A. Fox

Follow this and additional works at: http://digitalcommons.library.tmc.edu/utgsbs_dissertations

Part of the Other Microbiology Commons

Recommended Citation

Fox, Kristina A., "MULTIPLE POSTTRANSCRIPTIONAL REGULATORY FEATURES CONTROL EXPRESSION OF ETHANOLAMINE UTILIZATION GENES IN ENTEROCOCCUS FAECALIS" (2011). *UT GSBS Dissertations and Theses (Open Access)*. Paper 130.

This Dissertation (PhD) is brought to you for free and open access by the Graduate School of Biomedical Sciences at DigitalCommons@The Texas Medical Center. It has been accepted for inclusion in UT GSBS Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@The Texas Medical Center. For more information, please contact laurel.sanders@library.tmc.edu.

The TMC LIBRARY Health Sciences Resource Center

MULTIPLE POSTTRANSCRIPTIONAL REGULATORY FEATURES CONTROL
EXPRESSION OF ETHANOLAMINE UTILIZATION GENES
IN *ENTEROCOCCUS FAECALIS*

by

Kristina Ann Fox, B.S.

APPROVED:

____

Supervisory Professor: Danielle A. Garsin, Ph.D.

____

Theresa M. Koehler, Ph.D.

____

Barbara E. Murray, M.D.

____

Ann-Bin Shyu, Ph.D.

____

Ambro van Hoof, Ph.D.

APPROVED:

____

Dean, The University of Texas
Graduate School of Biomedical Sciences at Houston

MULTIPLE POSTTRANSCRIPTIONAL REGULATORY FEATURES CONTROL EXPRESSION OF ETHANOLAMINE UTILIZATION GENES IN *ENTEROCOCCUS FAECALIS*

A

DISSERTATION

Presented to the Faculty of

The University of Texas

Health Science Center at Houston and

The University of Texas M. D. Anderson Cancer Center

Graduate School of Biomedical Sciences

in Partial Fulfillment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

By

Kristina Ann Fox, B.S.

Houston, Texas

May, 2011

Acknowledgements

I would like to first thank my advisor Dr. Danielle Garsin. She gave me the independence to develop this project from the beginning and her insight turned this project into an interesting collaboration. Most importantly, she taught me the importance of patience and perseverance in achieving success as a scientist.

I am extremely grateful to my past committee members whose expertise in so many different areas was invaluable to my research: Dr. Adam Antebi, Dr. Stephen Norris, and Dr. Vasanthi Jarayaman. I am also very grateful to my current committee members Dr. Theresa Koehler, Dr. Barbara Murray, Dr. Ann-Bin Shyu, and Dr. Ambro van Hoof for their valuable assistance and guidance on my project. I would like to especially thank Dr. Ambro van Hoof for insightful discussion on structure and function of RNA.

I would like to thank Dr. Wade Winkler, Arati Ramesh, and Jennifer Stearns for their collaboration and insight.

I would like to thank all those who were kind enough to provide reagents especially Dr. Agathe Bourgogne in the Murray lab for technical assistance.

I would like to thank all the past and present members of the Garsin lab especially Arash Maadani, Dr. Violeta Chavez, Melissa Fellers, and Dr. Ransome van der Hoeven for all

their support. I would like to especially thank Dr. Sruti DebRoy for help in development of the model and providing data for this dissertation.

I would like to thank my family who provided continued support through the process. Finally, I would like to thank my fiancé Dr. John Latham. Without his love and support through all of the ups and downs of graduate school, I would not have been able to achieve this goal.

MULTIPLE POSTTRANSCRIPTIONAL REGULATORY FEATURES CONTROL EXPRESSION OF ETHANOLAMINE UTILIZATION GENES IN *ENTEROCOCCUS FAECALIS*

Publication No.__________

Kristina Ann Fox, B.S.

Supervisory Professor: Danielle A. Garsin, Ph.D.

Enterococcus faecalis is a Gram-positive bacterium that lives as a commensal organism in the mammalian gastrointestinal tract, but can behave as an opportunistic pathogen. Our lab discovered that mutation of the *eutK* gene attenuates virulence of *E. faecalis* in the *C. elegans* model host. *eutK* is part of the ethanolamine metabolic pathway which was previously unknown in *E. faecalis*. I discovered the presence of two unique posttranscriptional regulatory features that control expression of *eut* locus genes. The first feature I found is an AdoCBL riboswitch, a *cis*-acting RNA regulatory element that acts as a positive regulator of gene expression. The second feature I discovered is a unique two-component system, EutVW. The EutV response regulator contains an ANTAR family domain, which binds RNA to trigger transcriptional antitermination. I determined that induction of expression of several genes in the *eut* locus is dependent on ethanolamine, AdoCBL and the two-component system. AdoCBL and ethanolamine are both required for induction of *eut* locus gene expression. Additionally, I discovered *eutG* is regulated by a unique mechanism of antitermination. Both the AdoCBL riboswitch

and EutV response regulator control the expression of the downstream gene *eutG*. EutV potentially acts through a novel antitermination mechanism in which a dimer of EutV binds to a pair of mRNA stem loops forming an antitermination complex. My data show a unique mechanism by which two environmental signals are integrated by two different posttranscriptional regulators to regulate a single locus.

Table of Contents

Chapter 1: Introduction 1

Introduction to *Enterococcus faecalis* 2

Metabolism of *Enterococcus faecalis* 3

Initial identification of an *E. faecalis eut* locus gene in a screen for virulence factors in *C. elegans* 5

The ethanolamine metabolic pathway 5

Ethanolamine metabolism is potentially important for pathogenesis 11

Two-component systems in bacteria 13

ANTAR family of proteins 15

RNA riboswitches 16

Hypothesis, original goals and research summary 18

Chapter 2: Materials and Methods 20

E. faecalis strains and plasmids 21

E. faecalis growth conditions 21

Bioinformatic analysis 21

β-galactosidase assays 23

5'RACE..23

Construction of markerless mutations in *E. faecalis* 23

Quantitative reverse transcriptase PCR 27

Primer efficiency testing for qRT-PCR 27

Chapter 3: Bioinformatic analysis of the *eut* locus genes in *E. faecalis* 31

eut locus genes in *E. faecalis* 32

Ethanolamine metabolism..33

Analysis of *eut* locus genes present in *E. faecalis* 34

eut locus genes in two other Gram-positive pathogens 37

Identification of the AdoCBL riboswitch in *E. faecalis* 41

Domain analysis of the two-component system EutV and EutW in *E. faecalis*................45

Discussion...46

Chapter 4: Regulation of *eut* locus gene expression in response to AdoCBL and ethanolamine in OG1RF and *eutVW* strains 51

Development of a semi-defined medium for growth of *E. faecalis* showed that *eutP* is regulated by AdoCBL, ethanolamine and EutVW 53

Examining the expression of *eut* locus genes in response to AdoCBL and ethanolamine 59

Transcript analysis of *eut* locus genes 63

Discussion 67

Chapter 5: Regulation of *eutG* expression by ethanolamine, the ANTAR response regulator, and the AdoCBL riboswitch 71

EutV is an ANTAR 72

AmiR: a well studied ANTAR regulatory protein 75

NasR is an ANTAR regulatory protein 78

The universal ANTAR model 79

Preliminary testing of the ANTAR model using *eutP* 5'UTR 80

Expanding on the universal ANTAR model at *eutG* 83

Results 85

Mutational analysis of the transcriptional terminator 88

Mutational analysis of the P5 and P6 stem loops 96

Mutation of the P3 and P4 region of the *eutG* 5'UTR 102

Deletion of the *eut* riboswitch 108

Discussion 113

Chapter 6: Discussion and future directions 117

Identification of two posttranscriptional regulatory features by bioinformatics 119

Differential expression of *eut* locus genes 120

Potential positive feedback loop governs expression of *eut* regulators 121

Model of *eutG* regulation: the role of the P loops and the riboswitch 122

mRNA stability may play a role in *eut* locus expression 124

Alternative explanations for P3 and P4 mutatgeneis experiments 125

Control of *eut* expression by novel posttranscriptional regulators 126

Bibliography 130

Vita 143

List of Figures

Figure 1.1: Model of ethanolamine metabolism 8

Figure 2.1: Amplification plot of qRT-PCR reactions to test the efficiency of *eut* gene primer sets 30

Figure 3.1: *eut* locus organization of *E. faecalis* compared to *L. monocytogenes* and *C. perfringens*. 40

Figure 3.2: Graphical representation of domain analysis of the two-component system proteins EutV and EutW in *E. faecalis* 44

Figure 4.1: Media type and presence of AdoCBL and ethanolamine influence expression of *eutP::lacZ* in OG1RF and *eutVW* 55

Figure 4.2: qRT-PCR of *eutP, eutG, eutS*, and *eutA* showed *eut* locus expression is dependent on AdoCBL, ethanolamine, and *eutVW* 61

Figure 4.3: 5'RACE analysis of *eutP*, *eutG*, and *eutS* 65

Figure 5.1: Sequence of *eutG* 5'UTR with relevant features..74

Figure 5.2: Regulation of *eut* locus in *E. faecalis* 77

Figure 5.3: Universal model of ANTAR regulation 82

Figure 5.4: Modified model of ANTAR regulation for *eutG* 87

Figure 5.5: Terminator mutations 91

Figure 5.6: qRT-PCR of *eutG* in terminator mutants and wild-type 94

Figure 5.7: Schematic representation of mutations to P5 and P6 stem loops 99

Figure 5.8: qRT-PCR of *eutG* in P5 and P6 mutants 101

Figure 5.9: Schematic representation of P3 and P4 mutants 104

Figure 5.10: qRT-PCR of *eutG* in P3 and P4 mutants 106

Figure 5.11: Schematic of riboswitch deletion 110

Figure 5.12: qRT-PCR of *eutG* in riboswitch deletion 112

Figure 6: Model of *eutG* regulation 127

List of Tables

Table 2.1: *E. faecalis* strains and plasmids used..22

Table 2.2: List of primers used 24

Table 3.1: Genes present in the *eut* locus in *E. faecalis* 36

Chapter 1

Introduction

Introduction to *Enterococcus faecalis*

Enterococcus faecalis is a Gram-positive, non-motile, facultative anaerobe. *E. faecalis* grows in pairs or short chains. The natural reservoir for *E. faecalis* is in the gastrointestinal tract of most birds and mammals including humans (1, 78). In humans, *Enterococcus* species account for less than one percent of the total microbe population of the gastrointestinal tract (70). *E. faecalis* is a major cause of hospital-acquired infections and many traits of this bacterium lead to this phenomenon. *E. faecalis* is exceptionally resistant to harsh environments, contributing to its persistence in hospitals. For example, *E. faecalis* is able to survive in sodium chloride (NaCl) concentrations of up to 6.5 %, endure a broad range of pH conditions ranging from 4.0 to 9.6, and survive in temperatures as low as 10°C and above 45°C (72). Additionally, this organism is tolerant to desiccation and can survive on dry surfaces for long periods of time increasing the odds of infecting a new host (39). The success of *E. faecalis* as a pathogen is due in part to its resistance to many types of commonly used antibiotics (56). For example, *E. faecalis* are intrinsically resistant to certain β-lactam antibiotics. This may be due to the low affinity of the penicillin-binding proteins in the cell wall of *E. faecalis* for members of the β-lactam class of antibiotics (84). *E. faecalis* also can acquire antibiotic resistance by gaining genes required for antibiotic resistance through mobile DNA elements such as conjugative transposons and plasmids (56). For example, genes required for vancomycin resistance, which is considered to be an antibiotic of last resort, can be acquired through both conjugative transposons and plasmids (89). The combination of resilience and intrinsic resistance to many common antibiotics make *E. faecalis* a significant clinical concern as these organisms cause a wide variety of infections. The prevalence of

Enterococcus species as a common agent of hospital-acquired infection make the study of this organism of great importance as it is the second leading bacterial cause of hospital-acquired infections, behind *Staphylococcus aureus* (35). Enterococci commonly cause urinary tract infections (23). They are responsible for 10% of total urinary tract infections and up to 16% of urinary tract infections acquired in hospitals (23). Infections of the abdominal cavity are also caused by *E. faecalis* (40). *E. faecalis* can commonly cause bacteremia, blood stream infections, and is the third leading cause of nosocomial bacteremia (30). *E. faecalis* also can cause endocarditis and is responsible for up to 20% of all bacterial endocarditis (51).

Metabolism of *Enterococcus faecalis*

The survival of *E. faecalis* in harsh environments such as the intestinal tract of humans and animals may be facilitated by a flexible metabolism. *E. faecalis* is a member of the lactic acid bacteria. It generally uses homofermentive metabolism, meaning it can obtain energy when glucose is abundant and oxygen is limited. First, glucose is catabolized via the glycolysis pathway to pyruvate. Pyruvate is then reduced to lactic acid, which generates energy for the cell. Under certain circumstances *E. faecalis* can become heterofermentive, meaning pyruvate can be broken down into alcohol and lactic acid (52).

In addition to glucose, *E. faecalis* can utilize a wide variety of carbon sources for energy allowing *E. faecalis* propagation even when glucose is not available. Its genome contains more phosphotransfer systems (PTS) than any other sequenced bacterium (61). PTS

systems simultaneously import and phosphorylate different types of sugar molecules. Upon importation, the sugar molecules are fed into glycolysis usually after conversion to glucose-6-phosphate or fructose-6-phosphate (20). At least thirteen types of sugars including glucose, mannose, and maltose can be metabolized by *E. faecalis*.

Other types of carbohydrates can also be metabolized by *E. faecalis*. Mucins are a type of complex carbohydrate produced by intestinal epithelial cells. Mixtures of sugars, carbohydrates, and enzymatic digestion products from the host support the growth of *E. faecalis* (63).

E. faecalis can also utilize glycerol and has two pathways to metabolize it into a useful energy source. In the first pathway, ATP-dependent phosphorylation of glycerol by the glycerol kinase (GlpK) yields glycerol-3-phosphate (glycerol-3-P). In the second pathway, glycerol is first oxidized to dihydroxyacetone by a soluble NAD^+-dependent glycerol dehydrogenase. The dihydroxyacetone is then phosphorylated to dihydroxyacetone phosphate (DHAP) by dihydroxyacetone kinase and is fed into glycolysis (5, 38). Glycerol metabolism is hypothesized to be important for survival of *E. faecalis* within the eukaryotic host, as glycerol metabolism is important for *Listeria monocytogenes* survival inside eukaryotic cells (38).

The final carbon substrates that can be metabolized by *E. faecalis* are alpha-keto acids. Branched chain alpha-keto acids are carbon compounds that contain a keto group adjacent to the carboxylic acid group. In *E. faecalis*, these compounds are derived from

degradation of membrane lipids. *E. faecalis* can catabolize these alpha-keto acids into isovalerate using proteins expressed from the *bkd* gene cluster. The process of converting alpha-keto acids into isovalerate is analogous to pyruvate catabolism to acetate, which generates ATP by substrate level phosphorylation (82).

Initial identification of an *E. faecalis eut* locus gene in a screen for virulence factors in *C. elegans*

Our lab has a long-standing interest in the mechanisms underlying virulence in *E. faecalis*. To identify novel virulence determinants, an ordered library of insertion mutants in *E. faecalis* strain OG1RF was created using the Tn*917* transposon (29). This library consists of 540 insertion mutants, which disrupt approximately 23% of the nonessential genes in OG1RF (29). These mutants were tested in the *Caenorhabditis elegans* model to identify those attenuated for virulence (49). *C. elegans* has proven to be an effective model for identification of virulence factors in *E. faecalis* that may be relevant to a mammalian infection (28). The screen identified 23 transposon mutants with attenuated virulence phenotypes (49). The *eutK* mutant (annotated *pduJ* (29)) was of particular interest because of its involvement in the metabolism of ethanolamine, a previously unknown metabolic pathway in *E. faecalis*.

The ethanolamine metabolic pathway

Ethanolamine is a two-carbon alcohol amine. Ethanolamine and glycerol are catabolic products of the breakdown of phosphotidylethanolamine by phosphodiesterases (46). Phosphotidylethanolamine is an abundant phospholipid in both mammalian and bacterial

membranes. In the intestinal environment, ethanolamine is hypothesized to come from three sources: the diet of the host, membranes of host epithelial cells, and bacterial cell membranes (3, 64, 68). Ethanolamine utilization has been studied extensively in the Gram-negative bacterium *Salmonella typhimurium.* The following section describes the ethanolamine metabolic pathway in *S. typhimurium*.

Ethanolamine is converted to acetaldehyde and ammonia by an adenosylcobalamin (AdoCBL)-dependent ethanolamine lyase encoded by *eutB* and *eutC* (Figure 1.1) (6). *eutB* encodes the large subunit and *eutC* encodes the small subunit of this two-subunit protein (6). Acetaldehyde is a gas product of this reaction. Acetaldehyde is converted to acetyl coenzyme A (acetyl-CoA) by a *eutE*-encoded acetaldehyde oxidoreductase (Figure 1.1) (68, 69). Acetyl-CoA is used in a variety of different metabolic processes such as the tricarboxylic acid cycle (TCA), the glyoxylate cycle, or biosythesis of lipids. In addition, acetyl-CoA produced during metabolism of ethanolamine can also be converted to acetylphosphate by EutD, a phosphotransacetylase. Acetylphosphate can then be converted to acetate by AckA using substrate level phosphorylation generating ATP (10, 74). In an alternate branch of the ethanolamine metabolic pathway, acetaldehyde is converted into ethanol by an alcohol dehydrogenase encoded by *eutG* (43).

Several other proteins encoded in the *eut* locus play more indirect roles in the utilization of ethanolamine in *S. typhimurium*. AdoCBL is a necessary cofactor for the EutBC lyase (Figure 1.1). *eutT* encodes a corrinoid cobalamin adenosyltransferase, which adds the adenosyl moiety to the cobalamin molecule in an ATP-dependent manner.

Figure 1.1 Model of ethanolamine metabolism. In *S. typhimurium*, ethanolamine enters the cell through the permease EutH. Ethanolamine is catabolized into acetaldehyde by EutBC with AdoCBL as a cofactor. Acetaldehyde is further catabolized to acetyl-CoA by EutE or ethanol by EutG. Metabolism takes place in the microcompartment made of proteins EutSMNLK.

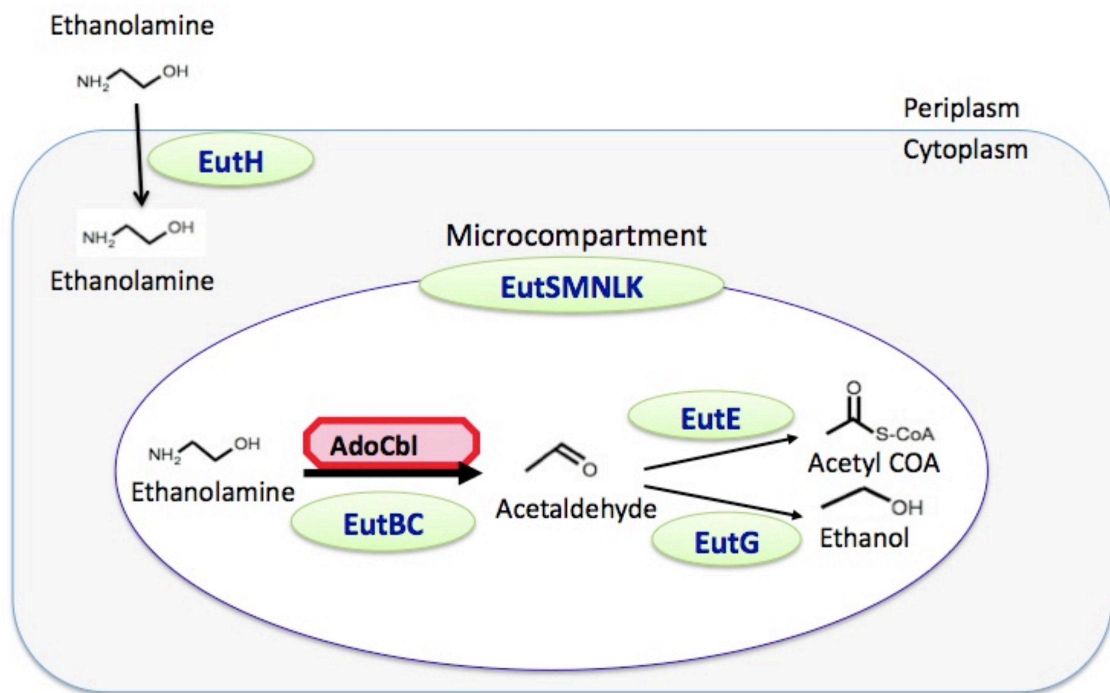


Figure 1.1

Adenosylation activates AdoCBL, which is used as a cofactor in the EutBC reaction (13). EutA encodes a reactivating factor, which reverses a suicidal inactivation reaction between AdoCBL and EutBC. A cleavage reaction of the Co-C bond present in the AdoCBL molecule occurs when it is bound to the enzyme. The result of this reaction is tight binding of AdoCBL to the lyase leading to the inactivation of the enzyme unless a reactivating factor, such as EutA, reverses the reaction (54). An inner membrane permease, which allows for transport of uncharged ethanolamine into the cells in low pH or low ethanolamine conditions, is encoded by the *eutH* gene (62). A positive transcriptional regulatory protein, which activates transcription of the *eut* operon in the presence of AdoCBL and ethanolamine, is encoded by *eutR* (67). The protein encoded by *eutP* has predicted GTPase activity and may be involved in cofactor reactivations (80). There are two open reading frames of unknown function in the *S. typhumurium eut* locus designated *eutQ* and *eutJ*.

In addition to the proteins described above, five additional proteins are encoded by the *eut* locus. These proteins make up a protein structure housing the metabolic enzymes that carry out ethanolamine catabolism. The five different structural Eut proteins are encoded by *eutSLKM* and *N*, and the structure is referred to as a bacterial microcompartment (Figure1.1) (43). This structure is also referred to as the carboxysome or the metabolosome in other studies and is akin to an organelle. Several possibilities for the function of this structure have been hypothesized. First, it may act as a barrier against diffusion of the acetaldehyde gas intermediate thus keeping it in a concentrated amount for more efficient utilization (42). The microcompartment may also sequester

acetaldehyde from the rest of the cell protecting against its toxicity (43). Lastly, the inner environment of the microcompartment may be more optimal for enzyme function due to a lower pH (42).

Bacterial microcompartments were first identified in chemoautotrophic bacteria (21). Microcompartments in these bacteria are filled with the enzyme ribulose bis-phosphate carboxylase oxygenase (RuBisCO) and were named carboxysomes. The proposed function of these structures was to optimize carbon fixation by creating a microenvironment rich in CO_2 and low in the competitive inhibitor of the reaction, molecular oxygen (14). The metabolic pathway for propanediol uses a similar microcompartment. The microcompartments of both the propanediol pathway and the ethanolamine pathway have been hypothesized to have similar functions (42). An open question regarding microcompartment function is how the substrates enter the inside of the compartment. Interestingly, a crystal structure study of the microcompartment protein EutL showed that it may be a gated molecule (77). Therefore, gated pores in the microcompartment structure may allow entry of the necessary molecules while preventing release of others.

Initial work on *S. typhimiurium* identified EutR as a regulatory protein of *eut* locus genes. EutR is a DNA-binding transcriptional activator. EutR acts as a positive regulator of transcription of the *eut* locus genes, activating transcription of the operon in the presence of ethanolamine and AdoCBL (67). In the absence of AdoCBL and ethanolamine, EutR is expressed at a low level from its own promoter. In the presence of AdoCBL and

ethanolamine, EutR is activated and induces expression of the operon containing the rest of the genes in the *eut* locus from a second promoter located upstream. In the presence of high levels of EutR, AdoCBL or ethanolamine individually lead to small changes in expression (67). The mechanism by which EutR can sense AdoCBL and ethanolamine is unknown. In addition, the mechanism by which EutR binds DNA has yet to be solved. However, recent bioinformatic analysis has identified a potential interaction site in the DNA (80).

Ethanolamine metabolism is potentially important for pathogenesis

Several studies of bacterial pathogens demonstrated that ethanolamine metabolism may be involved in virulence or survival within the host. In the Gram-positive pathogen *Listeria monocytogenes,* two studies suggest that ethanolamine may have a role in pathogenesis or host survival. Transcriptome analysis was performed to identify differentially regulated genes in *L. monocytogenes* cells when grown in a mouse intestinal model of infection. The authors found that all of the genes of the *eut* locus were strongly upregulated in the gut (79). A separate study examined the colonization of a series of *L. monocytogenes* mutants in Caco-2 cells, which are derived from human colon epithethelium. A mutant in *eutB,* encoding a component of the ethanolamine lyase, was severely deficient in colonization of these cells compared to wild type (38). Together, these experiments demonstrated a role for utilization of ethanolamine in *L. monocytogenes* host survival.

In the Gram-negative enteric pathogen *S. typhimurium*, global regulators of virulence influence expression of *eut* locus genes. The protein CsrA, which regulates expression of flagellar and virulence genes in *S. typhiumurium*, also upregulates expression of all of the *eut* genes (48). A second global regulator Fis, which regulates genes on several pathogenicity islands as well as genes important for flagellar synthesis, was also shown to positively influence expression of *eut* locus genes in a transcriptional profiling study (41). In another study, *S. typhiumurium eutE* and *eutG* mutants were less virulent than wild type in a mouse model of infection.(75).

In the Gram-positive pathogen *E. faecalis*, Fsr is a global virulence regulator. Fsr was found to be a negative regulator of the *eut* locus in a study to identify genes regulated by Fsr (9). This is in contrast to results in *S. typhimurium* in which global regulators of virulence positively regulated *eut* locus gene expression. *S. typhimurium* is a pathogen in the gut while *E. faecalis* is a commensal organism. This lifestyle difference within the intestine may explain the variation in regulatory patterns (27). Additionally, as described above, a transposon insertion mutant in *eutK* was attenuated in virulence the *C. elegans* model (49).

The instances described above suggest potential roles for the *E. faecalis eut* locus in mammalian pathogenesis. Interestingly, several instances of ethanolamine utilization being important in non-mammalian infection have been reported. The Gram-negative bacterium *Photorhabdus lunminescens*, lives as a symbiote in the nematode *Heterorhabdus bacteriophora.* This nematode is an insect pathogen. A screen using a

promoter trap library showed upregulation of *eutA, eutB*, and *eutC* in nematode infected insect hemolymph. These results suggest a potential role of ethanolamine genes for bacterial survival in this three-way infection (55).

Ethanolamine genes are also involved in pathogenesis in the plant pathogen *Erwinia chysanthemi*. A *eutR* mutant was unable to cause a systemic infection. However, this mutant was still able to engineer a localized infection. This suggests that ethanolamine metabolism could be important for infection dissemination in this organism (90).

Together, these studies show that the *eut* locus in many different pathogens has potential connections to virulence. However, many of these studies only imply an indirect role in infection for metabolism of ethanolamine. It is interesting that the ability to metabolize this carbon source provides a competitive advantage within the host environment.

Two-component systems in bacteria

Two-component systems allow bacteria to sense and respond to stimuli in the environment. Prototypically, an external stimulus is sensed by a histidine protein kinase (HK). Sensing of a specific signal causes an autophosphorylation of a conserved histidine residue by the conserved catalytic core of the enzyme (26). The activated HK then phosphorylates the cognate response regulator protein on a conserved aspartic acid residue in the receiver domain of the protein. Usually phosphorylation of the activated response regulator leads to modulation of a downstream event through its output domain (26). A wide variety of response regulator output domains have been described. The

most commonly occurring receiver domain binds to DNA where it modulates transcription. Less common response regulator output domains include enzymatic domains and RNA binding domains, which most commonly function as antiterminators (26).

E. faecalis strain V583 contains eighteen two-component systems (32), and the genome of strain OG1RF contains two additional two-component systems (8). In *E. faecalis*, two-component systems have been shown to control a wide variety of processes such as biofilm formation, antibiotic resistance, and stress tolerance (32). The two-component system controlling ethanolamine utilization is quite different from the other two-component systems found in *E. faecalis*.

The histidine kinase in the *eut* locus is EutW. The sensing domain is unique with no domain conservation. The autocatalytic domain and main catalytic residues are conserved, indicating that the histidine kinase likely functions in a manner similar to most HK proteins in phosphotransfer systems (32). In addition, it is the only HK within *E. faecalis* that has no predicted transmembrane domain indicating that it may respond to an intracellular signal rather than an outside signal (32).

The cognate response regulator is EutV. This protein contains a conserved receiver domain with a conserved phospho-accepting aspartic acid residue. The output domain is similar to the ANTAR family of proteins. ANTAR (AmiR and NasR transcription antitermination regulators) family proteins are predicted to prevent the formation of a

transcriptional termintor stem-loop through interaction with mRNA. This is the only predicted RNA-binding response regulator in the *E. faecalis* genome (32).

ANTAR family of proteins

ANTAR family response regulators are proteins that bind RNA and act as antitermination factors. They make up less than 1% of known response regulator proteins (73). The AmiR system in *Pseudomonas aeruginosa* and the NasR in *Klebsiella oxytoca* are two known examples (73).

The AmiC/AmiR regulatory pair controls expression of the amidase operon in *Pseudomonas aeruginosa*. The operon is expressed from a constitutive promoter that resembles an *Escherichia coli* sigma 70 promoter. A short non-coding leader RNA sequence exists containing the transcriptional terminator. In the presence of an inducing substrate such as acetamide, AmiC and AmiR form a complex. This complex interacts with a specific sequence upstream of the terminator sequence preventing formation of the terminator stem-loop (87).

NasR in *Klebsiella oxytoca* is the other ANTAR family protein studied to date. NasR differs from AmiR because it does not partner with a cognate sensor protein. NasR is activated directly by nitrate molecules, then positively regulates expression by inhibiting the formation of a transcriptional terminator in the *nasF* 5' UTR (16).

RNA riboswitches

A scan of the *E. faecalis eut* locus using the Ribex program uncovered a possible adenosylcobalamin riboswitch (2). This was interesting because ethanolamine degradation requires adenosylcobalamin as a cofactor. Riboswitches are highly structured non-coding RNAs located in the 5' leader regions of mRNAs. Riboswitches act *in cis* to control expression of genes located downstream. In order to regulate gene expression, a specific intracellular metabolite or ion binds to the riboswitch. The metabolite or ion binds in the aptamer domain, which is highly conserved in sequence and secondary structure. Binding elicits conformational changes resulting in the sequestration of sequences and preventing the formation of additional downstream structures. The downstream structures affected by the riboswitch are usually involved in transcription termination or access to a ribosome binding site (17). The affected gene(s) downstream are commonly involved in the synthesis, transport, or metabolism of the metabolite or ion.

Several classes of riboswitches have been identified to date. The first class of riboswitches is involved in responding to enzymatic cofactors, including AdoCBL, thiamine pyrophosphate, and flavin mononucleotide. A second class responds to amino acids and amino-sugars (17). This group includes lysine, glycine, and glucosamine-6-phosphate. Three distinct classes of riboswitches have been found that respond to S-adenosylmethionine (17). Additionally, a class of riboswitches responding to binding of either adenine or guanine has been identified (17, 22, 53, 88). Studies have assigned functional roles to orphan riboswitches, those that were identified bioinformatically but

had no initial predicted function. Several new classes of riboswitches were discovered that respond to cyclic diGMP, preQ1, *S*-adenosylhomocysteine, molybdenum cofactor, and arginine (17).

There are two main mechanisms of gene regulation used by riboswitches to control gene expression in bacteria. The first mechanism used by riboswitches is to control transcription termination. Intrinsic transcriptional terminators occur downstream of many riboswitches. In the absence of a riboswitch metabolite, an antiterminator structure is formed. Upon binding of a metabolite to the aptamer domain of the riboswitch, conformational change prevents antiterminator formation and the intrinsic terminator is formed. The second mechanism by which riboswitches regulate gene expression is through control of translation initiation. In the absence of a bound metabolite, the RNA adopts a structure in which the ribosome binding site (RBS) is accessible to the ribosome and translation can occur. When ligand is bound to the riboswitch an alternate structure is formed, sequestering the RBS and not allowing ribosomes to bind (17, 83).

Recent bioinformatic analysis has identified over 200 AdoCBL riboswitches in 67 bacterial genomes (4). The AdoCBL riboswitches described prior to my work serve as negative regulators of either transcription or translation (66). The function of two AdoCBL riboswitches in other organisms has been examined in detail in previous studies. These are the *btuB* riboswitch and the *cob* riboswich.

btuB encodes a protein that makes up part of an ABC transporter for cobalamin molecules. A riboswitch is found in the 5'UTR of this gene. This riboswitch was shown to have specific binding affinity for AdoCBL. Binding of AdoCBL to the riboswitch results in downregulation of *btuB*. Therefore, the *btuB* riboswitch acts as a negative regulator of expression (59). The second AdoCBL riboswitch described to date is the *cob* riboswitch. The *cob* operon encodes approximately 25 proteins required for synthesis of AdoCBL *de novo* (65). A riboswitch RNA found in the 5'UTR of this operon binds to AdoCBL and causes decreased gene expression of *cob* genes in high concentrations of AdoCBL (58). These results suggest that the *cob* riboswitch negatively regulates *cob* operon expression.

The two previously described riboswitches, *cob* and *btuB*, control expression of genes associated with synthesis and import of AdoCBL, respectively. The AdoCBL class of riboswitches acts as negative regulators of gene expression, likely because the synthesis and/or import of cobalamin molecules are unnecessary in high concentrations of AdoCBL.

Hypothesis, original goals and research summary

Our previous results showed that disruption of the *eut* locus in *E. faecalis* attenuated virulence in *C. elegans* (49). As this locus was uncharacterized in *E. faecalis*, I decided to elucidate how the *eut* locus is regulated. First, I characterized bioinformatically the *eut* locus and compared its structure to other organisms in which the *eut* locus was characterized previously. I identified a riboswitch and a two-component system as likely

regulatory features of the *eut* locus. These features are novel compared to the other studied *eut* loci in Proteobacteria such as *E. coli* and *Salmonella* species, which are regulated by a DNA-binding transcription factor, EutR. Based on these bioinformatic findings, my hypothesis was that two posttranscriptional regulators, the AdoCBL riboswitch and EutVW two-component system, work in concert to control expression of genes found in the operon. I investigated the function of these two regulators. First, I determined that ethanolamine and AdoCBL control expression in a positive manner. AdoCBL potentially acts through binding to the *eut* riboswitch. This is the first example of an AdoCBL riboswitch positively regulating gene expression. Second, I demonstrated that a two-component system, encoded by *eutV* and *eutW*, is also a positive regulator of *eut* locus gene expression. Lastly, I demonstrated that a unique mechanism of antitermination might be in place controlling gene expression. Overall, my results demonstrate that a novel regulatory mechanism is in place in the *eut* locus of *E. faecalis* that integrates two different molecular signals. This is the first described instance of a riboswitch and a two-component system together controlling expression of the same set of genes.

Chapter 2

Materials and Methods

E. faecalis strains

All strains and plasmids used in this study are listed in Table 2.1.

E. faecalis growth conditions

E. faecalis strains were grown shaking overnight in BHI medium in aerobic conditions at 37°C. For most experiments, cells were subcultured in a 1:25 dilution in M9HY medium. M9HY was made as described previously with the addition of 0.5% ribose (Sigma) and 33 mM ethanolamine (Sigma) and AdoCBL (Sigma). For experiments requiring anaerobic growth, culture tubes were placed in an anaerobic chamber (BD biosciences). The chamber was placed in a 37°C incubator and cells were grown without shaking. Anaerobic conditions were created in the chamber using gas packs (BD biosciences) and anaerobic conditions were verified with anaerobic indicator strips (BD biosciences) inside the chamber. For most experiments, anaerobically grown cells were removed from the chamber after 3.5 hours of incubation. Cells were collected by centrifugation and immediately frozen in a dry ice and ethanol bath.

Bioinformatic analysis

Gene sequences used in this analysis were obtained from the comprehensive microbial resource (CMR) from JCVI (18). Sequence similarity testing was done using BLASTP (36). Analysis of RNA features was performed using the Ribex program (2).

Table 2.1 *E. faecalis* strains and plasmids used

Strain	Resistance	Description
OG1RF	Rif^r	wild-type strain in these studies
eutVW	Rif^r	markerless deletion of *eutVW* two-component system, OG1RF background
KFF7	Rif^r Erm^r	*eutP::lacZ,* pKAF7 in OG1RF background
KFF9	Rif^r Erm^r	*lacZ* empty vector, pKAF11 in OG1RF
KFF66	Rif^r	*eutG* terminator deletion, OG1RF background
KFF69	Rif^r	*eutG* terminator stem mutation 1, OG1RF background
KFF72	Rif^r	*eutG* terminator stem mutation 2, OG1RF background
KFF75	Rif^r	*eutG* P5 AXXG mutation, OG1RF background
KFF77	Rif^r	*eutG* P6 AXXG mutation, OG1RF background
KFF79	Rif^r	*eutG* P5 stem mutation, OG1RF background
KFF 81	Rif^r	*eutG* P3 and P4 deletion, OG1RF background
KFF84	Rif^r	*eutG* P3 and P4 stem mutation, OG1RF background
KFF86	Rif^r	*eutG* P3 and P4 AXXG mutation, OG1RF background
KFF 88	Rif^r	*eut* riboswitch deletion, OG1RF background
Plasmid	**Resistance**	**Description**
pKAF6	Erm^r	LacZ empty vector
pKAF11	Erm^r	*eutP*::LacZ
pCJK47	Erm^r	shuttle vector for making markerless mutants
pKAF41	Erm^r	*eutG* terminator deletion in pCJK47
pKAF42	Erm^r	*eutG* terminator stem mutation 1 in pCJK47
pKAF43	Erm^r	*eutG* terminator stem mutation 2 in pCJK47
pKAF44	Erm^r	*eutG* P5 AXXG mutation in pCJK47
pKAF45	Erm^r	*eutG* P6 AXXG mutation in pCJK47
pKAF46	Erm^r	*eutG* P5 stem mutation in pCJK47
pKAF47	Erm^r	*eutG* P3 and P4 deletion in pCJK47
pKAF48	Erm^r	*eutG* P3 and P4 stem mutation in pCJK47
pKAF49	Erm^r	*eutG* P3 and P4 AXXG mutation in pCJK47
pKAF50	Erm^r	riboswitch deletion in pCJK47

β-galactosidase assays

E. faecalis cells were lysed with zirconia beads using a Mini Bead Beater for 2 minutes. Total protein content of cell lysates was determined using a BCA assay kit (Pierce) as indicated by the manufacturer's instructions. Cell lysates were incubated with *ortho*-nitrophenyl-β-galactoside (ONPG) and then ONP levels were measured using OD440 on a plate reader. Arbitrary units were calculated by dividing the ONP production by total protein.

5' Rapid amplification of cDNA ends (5'RACE)

OG1RF strain of *E. faecalis* was grown in anaerobic conditions. Cells were harvested after 3.5 hours of growth. RNA extraction was carried out using the Ribopure bacteria kit (Ambion) per the included instructions and 5'ends of transcripts were found using the 5'RACE kit from Invitrogen. RACE products were cloned into the pGEM-T vector and sequenced using vector specific primers (Table 2.2).

Construction of markerless mutations in *E. faecalis*

Mutant strains were made using a modification of the markerless deletion protocol using the *pheS* counterselection system (44). Mutation containing fragments used for cloning were made using two-step overlap extension PCR. Primers used for making each mutant are in Table 2.2. The mutagenic fragments were then cloned into pCJK47, the vector used to make the deletion strains. From this point the procedure described by Kristich et al, 2007, was used to complete the deletion (44). Final mutant strains were verified by sequencing.

Table 2.2: List of primers used

Name	Sequence	Gene and purpose
KF85	ATGGCGTTTAATGTTACTTGCTTAGG	qRT-PCR of *eutG* fwd
KF86	TAAGAGCATTGCATTCATTCGTC	qRT-PCR of *eutG* rev
KF88	ATTTTCTTGTGGAAGTTGGCTGA	*eutT* Reverse transcription
KF100	TTCCTCCTTAATAAAATACAAAAAGACACT T	*eutG* Gene specific 1 for 5'RACE
KF101	CCATGAATGGATATGTGACAATG	*eutG* Reverse transcription
KF102	AAAGTCGACGATATCCGTTTAGCGTTAACG CTAAG	*eutP* lacZ construction
KF116	GGAGCAATATCGCCAAGCAG	qRT-PCR of *eutA* fwd
KF117	GCACATTGTTGGCATTGCTTT	qRT-PCR of *eutA* rev
KF121	CAGACATACTTTAAGGTGGCGGACACA	*eutP* Gene specific 2 for 5'RACE
KF123	GGGAAAATACTACCAAATCCTGGTG	*eutP* Reverse transcription
KF132	ATGATAAAACACAGGCGGTGG	qRT-PCR of *eutP* fwd
KF133	CGCTGTGACGTTTAACGCAT	qRT-PCR of *eutP* rev
KF134	ACTTGTTGTGTTTGTAC	*eutP* Gene specific 1 for 5'RACE
KF135	TGTTGCACCCGCACTTTTTA	*eutT* Gene specific 1 for 5'RACE
KF138	AGATGTGAGTGCCTCTGTTGTGTTA	qRT-PCR of *gyrA* Fwd
KF139	TCGGTACGCCTTTTTCGATG	qRT-PCR of *gyrA* rev
KF147	AAAGGATCCCACCGATAGCTCCCATTAAAA TGATTCGTTT	*eutP* lacZ construction
KF160	CGAAGCGGCAATTATCGC	qRT-PCR of *eutS* fwd
KF161	CGTCACCAGAAATGACTACCGA	qRT-PCR of *eutS* reverse
KF196	CGTCATTGTACCTGAAATCTCCT	*eutG* Gene specific 2 for 5'RACE
KF 211	TCTAGACTCAACCGTCGCTAAGGATTC	Primer A for *eutG* deletion 1-6

KF212	CCCGGGCTTCTTCTTTTGGTAAACCATAGTC ACT	Primer D for *eutG* deletion 1-6
KF213	TCTAGAGGTTCAAAGTGCTGTAGAAACACA	Primer A for deletion 7-10
KF214	CCCGGGCAATTTGCTAACTGCGCATAACG	Primer D for deletion 7-10
KF215	TCAATTCCTCCTTAATAAAATACAAAAAGA CTTAAGAAACAAACGAAAAATCATCTG	Primer B for mutation 1
KF216	CGCTTCAGACAGATGATTTTTCGTTTGTTTC TTAAGTCTTTTTGTATTTTATTAAGG	Primer C for mutation 1
KF217	CCTTAATAAAATACAAAAAGACACTTCAAA AGTCGACAAAAATCATCTGTCTGAAGCGTC	Primer B for mutation 2
KF218	AAGCAAAGACGCTTCAGACAGATGATTTTT GTCGACTTTGAAGTGTCTTTTTGTATTTTA	Primer C for mutation 2
KF219	TTTCGTGTACAATGGCGTATACATAAGGAA GTTAACACGCTTCAGACAGATGATTTTTCG	Primer B for mutation 3
KF220	AACAAACGAAAAATCATCTGTCTGAAGCGT GTTAACTTCCTTATGTATACGCCATTGTAC	Primer C for mutation 3
KF221	CCTCCTTAATAAAATACAAAAAGACACTTC GAAGCGTCTTTGCTTCCTTATGTATACGCC	Primer B for mutation 4
KF222	GGCGTATACATAAGGAAGCAAAGACGCTTC GAAGTGTCTTTTTGTATTTTATTAAGGAGG	Primer C for mutation 4
KF223	CATCTGTCTGAAGCGTCTTTGCTTCCTTATC ATATGGCCATTGTACAC	Primer B for mutation 5
KF224	CTCTATGGATAAGGTTTCGTGTACAATGGC CATATGATAAGGAAGCAAAGACGCTTCAG A	Primer C for mutation 5
KF225	CTGAAGCGTCTTTGCTTCCTTATGTATACGC TTAAGTACACGAAACCTTATCCATAGAGT	Primer B for mutation 6
KF226	CGTTGTACTCTATGGATAAGGTTTCGTGTAC TTAAGCGTATACATAAGGAAGCAAAGACG	Primer C for mutation 6
KF227	GCCTTTAATTAATTGTCAATAAAAAAAGGG CCCAAAATTCCCAAAAGCACTTCCTCGTGC	Primer B for mutation 7
KF228	GCACGAGGAAGTGCTTTTGGGAATTTTGGG CCCTTTTTTTATTGACAATTAATTAAAGGC	Primer C for mutation 7
KF229	GGGAATATAGCAGTACGTCATTAGTACTGA AATTTCGAACAACATTGAAAGACCAGCCCA	Primer B for mutation 8
KF230	TGGGCTGGTCTTTCAATGTTGTTCGAAATTT CAGTACTAATGACGTACTGCTATATTCCC	Primer C for mutation 8
KF231	AGCAGTACGTCATAGTACCTGAAATCTCCT TCACTTAAGAAAGACCAGCCCAAAATTCCC	Primer B for mutation 9
KF232	TGGGCTGGTCTTTCTTAAGTGAAGGAGATT TCAGGTACTATGACGTACTGCTATATTCCC	Primer C for mutation 9
KF272	CTTGGGCTGGATTTTCAACACAAGCGTTGT TAACCTATCTGAAATTTACCATAA	Primer B for RS deletion

KF273	CTTATGGTAAATTTCAGATAGGTTAACCAA CGCTTGTGTTGAAAATCCCAGCCCAAG	Primer C for RS deletion

Quantitative reverse transcriptase PCR (qRT-PCR)

Cells were disrupted for two minutes using a bead beater and RNA extraction was performed as listed in the Ribopure kit (Ambion). Purified RNA was DNase treated with Turbo DNAfree (Ambion) using the manufacturers protocol for rigorous DNase treatment. cDNA synthesis was performed using Superscript II (Invitrogen) using the manufacturers protocol.

Gene expression was assessed using quantitative RT-PCR using primers listed in Table 2.2. Each reaction was set up as follows: 9μl of an appropriate dilution of cDNA was mixed with 1μl primer mix and 10μl SYBR green mastermix (Applied biosystems). Reactions were set in duplicate and primers had been previously verified for efficiency. Data analysis was done using the standard curve method and standardized to a housekeeping gene *gyrA* (34).

Primer efficiency testing for qRT-PCR

Primer sets for *eutP*, *eutG*, *eutS, eutA,* and *gyrA* were designed using the Primer Express software package by Applied Biosystems. Each amplicon was designed to be 100 base pairs long. To determine which primer sets had similar amplification efficiency, each primer set was used in a qRT-PCR reaction as described above. A known concentration of gDNA along with 1:10, 1:100, and 1:1000 dilutions of the gDNA were used as templates in the amplification reactions. The half-maximum amplification was determined using Applied Biosystems SDS software. If primer sets reached half-maximum amplification within one amplification cycle of each other and the maximum

delta reaction value was within 0.5 units, then the primer sets were determined to have similar amplification efficiencies. Only primers with similar amplification efficiencies were used in subsequent experiments. The amplification plot to determine the primer amplifiction efficiency for *eutP*, *eutG*, *eutS,* and *eutA* is shown in Figure 2.1.

Figure 2.1 Amplification plot of qRT-PCR reactions to test the efficiency of *eut* gene primer sets. Pink lines are *eutP* reactions, blue lines are *eutG* reactions, orange lines are *eutS* reactions, and green lines are *eutA* reactions. Each bracket is separate concentration of gDNA template. Bracket A is undiluted gDNA, bracket B is 1:10 gDNA, bracket C is 1:100 gDNA, bracket D is 1:1000 gDNA. The blue arrow indicates where maximum amplification was determined for each reaction.

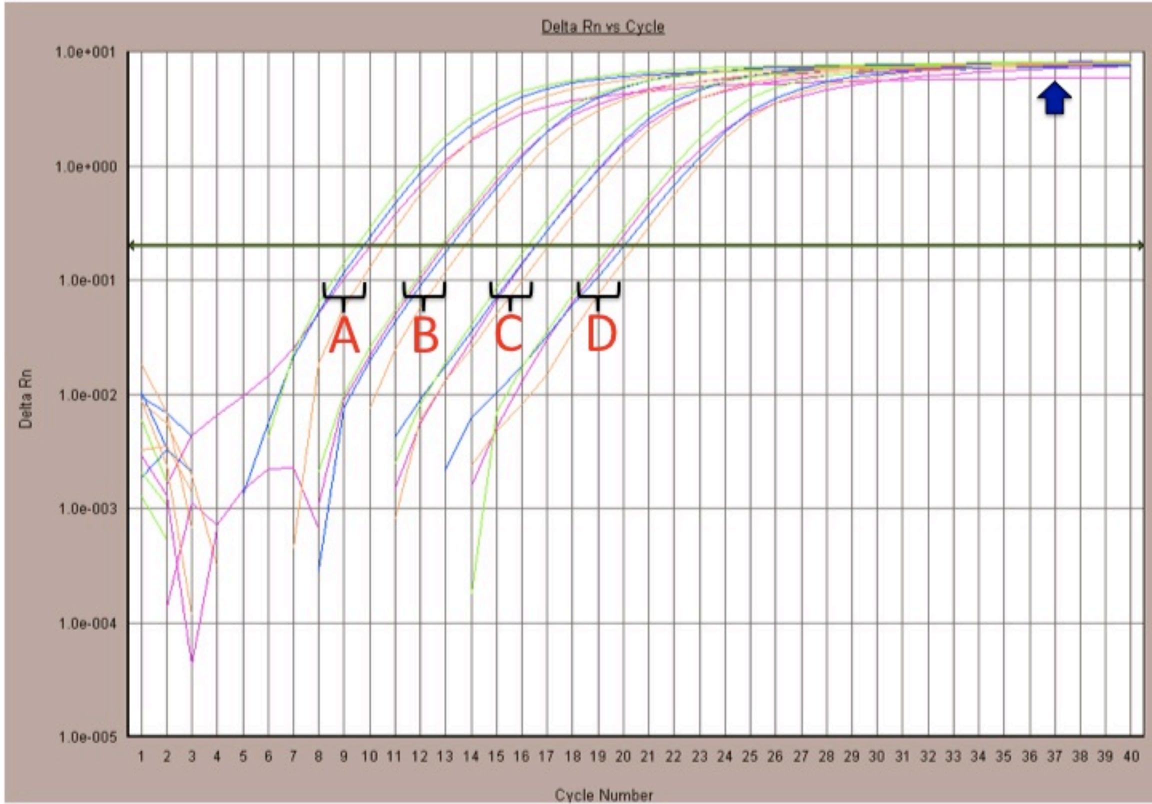


Figure 2.1

Chapter 3

Bioinformatic analysis of *eut* locus genes in *E. faecalis*

eut* locus genes in *E. faecalis

Our lab previously identified a number of *E. faecalis* transposon insertion mutants with attenuated virulence in a *C. elegans* model of infection (49). One mutant uncovered was in the gene originally annotated *pduJ*, which is involved in the metabolism of propanediol (49). My analysis below indicated that *pduJ* had been incorrectly annotated and was in fact *eutK,* a gene involved in ethanolamine metabolism. My discovery prompted me to perform a thorough bioinformatic analysis to correctly identify the genes present in the *eut* locus.

Ethanolamine metabolism was identified as potentially involved in bacterial virulence (49), making it an interesting target to study. While ethanolamine utilization genes had been studied in *S. typhiumurium*, little was known about the *eut* locus genes outside of this organism. Therefore, a bioinformatic analysis of the *eut* locus is important to understand the components of ethanolamine metabolism present in *E. faecalis*. The following section describes the bioinformatic analysis I performed annotating the genes found in the *eut* locus in *E. faecalis*. In addition, this section includes my comparisons of the *E. faecalis eut* locus to the *eut* locus in several other Gram-positive pathogens. Finally, I describe the identification of two interesting regulatory mechanisms in this locus; an AdoCBL responsive riboswitch and a two-component system that has RNA binding and antiterminator function.

Ethanolamine metabolism

Ethanolamine utilization has been studied extensively in the Gram-negative, enteric bacterium *Salmonella typhimurium*. In this organism, the 17 genes required for this process are found together on the chromosome within the *eut* locus (11). The essential enzymes required for metabolism reside in a multi-protein complex called a bacterial microcompartment. This complex is composed of five different Eut proteins encoded by the *eutS, L, K, M,* and *N* genes (43).

Ethanolamine metabolism in *S. typhimurium* is a multistep process that occurs within the microcompartment. Briefly, ethanolamine is converted to acetaldehyde by a two-subunit ethanolamine lyase encoded by the genes *eutB* and *eutC.* This reaction requires AdoCBL as a cofactor (7). The EutBC lyase can undergo a suicidal inactivation reaction in which a crucial bond of AdoCBL is cleaved, but the coenzyme remains tightly bound to the lyase rendering it inactive. A reactivating factor, encoded by *eutA,* reverses coenzyme B12 inactivation and restores normal enzyme function (54). Next, the acetaldehyde dehydrogenase EutE, converts acetaldehyde to acetyl coenzyme A (acetyl-CoA) (37). Alternatively, acetaldehyde is converted to ethanol by the alcohol dehydrogenase EutG depending on the cellular requirements (43) . Acetyl-CoA is used subsequently in a variety of different metabolic processes such as the TCA cycle, the glyoxylate cycle, or lipid biosynthesis (11).

Several additional proteins encoded in the *eut* locus play a more indirect role in the utilization of ethanolamine in *S. typhimurium*. A corrinoid cobalamin

adenosyltransferase, encoded by *eutT*, makes the active form of coenzyme B12, which is used as a cofactor in the EutBC lyase reaction (13). An inner membrane permease, EutH, allows transport of ethanolamine into the cells under low pH or low ethanolamine conditions (62). The positive transcriptional regulatory protein EutR activates transcription of the *eut* operon in the presence of coenzyme B12 and ethanolamine (67).

Analysis of *eut* locus genes present in *E. faecalis*

I first performed an analysis of the locus where we found the *eutK* insertion mutant to better understand its molecular genetic context. I discovered that many of the genes in the locus containing *eutK* were misannotated when the genome of *E. faecalis* strain V583 sequence was published. Several genes were annotated as part of the propanediol operon. Propanediol is a three-carbon compound, which has a similar metabolic pathway. Like ethanolamine, it is degraded by a multi-subunit lyase and this enzymatic process also takes place inside a multi-protein microcompartment. Therefore, an undiscerning analysis may not distinguish the two operons. However, propanediol utilization is not found in *E. faecalis*. The evidence that the genes present in the *eut* locus were in fact for ethanolamine metabolism and not propanediol metabolism is the presence of the *eutBC* lyase genes (Table 3.1) and the absence of propanediol lyase genes.

I properly annotated the genes and examined the locus structure and organization in *E. faecalis*. I used BLAST analysis (36) and conserved domain analysis (CDD) (50) to properly identify the genes located within the *eut* locus. I next compared the *E. faecalis* *eut* locus composition to the *eut* locus in *S. typhimurium* (summarized in Table 3.1).

Table 3.1 Genes present in the *eut* locus in *E. faecalis*. The first column is gene names in the order they occur in the *E. faecalis* chromosome from 5' to 3'. The second column is gene functions. The third column lists whether homologs are present in *S. typhiumurium*. The fourth column lists BLAST E. value between *S. typhiumurium* and *E. faecalis,* if applicable. Genes marked with an asterisk are present in *S. typhimurium* but are not present in *E. faecalis*.

Gene Name	Function	Present in *S. typhimurium*	Blast e. value
eutP	unknown	Y	4e-17
eutT	corrinoid cobalamin adenosyltransferase	Y	7e-13
eutG	alcohol dehydrogenase	Y	2e-66
eutS	microcompartment structural protein	Y	2e-21
eutV	response regulator	N	n/a
eutW	histidine kinase	N	n/a
eutA	lyase reactivating factor	Y	2e-75
eutB	lyase large subunit	Y	5e-145
eutC	lyase small subunit	Y	7e-53
eutL	microcompartment structural protein	Y	3e-62
eutM	microcompartment structural protein	Y	2e-20
eutE	acetaldehydride dehydrogenase	Y	6e-47
eutK	microcompartment structural protein	Y	0.002
eutX1	unknown	N	n/a
eutX2	potential structural protein	N	n/a
eutX3	unknown	N	n/a
eutN	microcompartment structural protein	Y	1e-06
eutH	ethanolamine permease	Y	5e-98
eutQ	unknown	Y	7e -18
*eutR**	transcriptional regulator	Y	n/a
*eutD**	phosphotransacetylase	Y	n/a

Table 3.1

I discovered that the main metabolic components of the *S. typhimurium eut* locus, the EutBC lyase and its reactivating factor EutA, are also present in *E. faecalis*. In addition, I found that the enzymatic components EutE and EutG are present in *E. faecalis*. I determined that each of the microcompartment proteins EutKLMNS are also present. Finally, I discovered that the outer membrane permease, EutH, is also in *E. faecalis*. However, I found that there are several elements of the *E. faecalis eut* locus that differ from the *eut* locus of *S. typhimurium*. Missing from *E. faecalis,* but present in *S. typhimurium,* are the genes for phosphotransacetylase, *eutD,* and for the transcriptional regulator, *eutR*. I also discovered that several genes are present in *E. faecalis* that were absent in *S. typhimurium*. There is a two-component system, that I named EutV and EutW, as well as three open reading frames of unknown function downstream of *eutK* (Table 3.1). Although many genes are conserved between the two organisms, I found that the overall organization of the genes in the loci is different. My discovery of the absence of a gene encoding the regulatory protein EutR and the presence of genes encoding a two-component system suggested that a unique, undescribed regulatory mechanism is in place in *E. faecalis*.

eut locus genes in two other Gram-positive pathogens

At first, I limited my comparison of the *E. faecalis eut* locus to *S. typhimurium*. To expand my analysis and determine if the *E. faecalis eut* locus is conserved in other Gram-positive species, I used BLAST analysis. I discovered that the Gram-positive pathogens *L. monocytogenes* and *C. perfringens* each contain *eut* locus genes. Next, I used BLAST analysis and CDD analysis to examine the entire *eut* locus in two species, *Listeria*

monocytogenes strain EGD-e and *Clostridium perfringens* strain 13. I found that the genomic context of the *eut* locus of each species is different. Upstream of the *eut* locus in *E. faecalis* is an ABC transporter for the import of cobalamin molecules that are precursors to adenosylcobalamin, the cofactor in the EutBC lyase reaction. I did not find this transporter in the genomes of either *L. monocytogenes* or *C. perfringens*. In *L. monocytogenes*, I found the *eut* locus in tandem with the locus for propanediol utilization. The first gene for ethanolamine utilization is *eutG,* which is proceeded by a large intergenic space. My analysis showed that the gene *eutP* is not present in this species. Different from the other species I studied, the *C. perfringens* genes surrounding the *eut* locus seem completely unrelated to the *eut* locus or metabolism. I determined that the central genes in the *eut* locus from *eutV* to *eutK* are conserved between the three species. In *L. monocytogenes* and *C. perfringens*, *eutT* is located in the later portion of the locus following *eutK*. I found that all three loci encode the two-component system EutVW and lack *eutD* and *eutR* (Figure 3.1).

The differences that I discovered in the *eut* locus between *E. faecalis* and the other two species examined in this analysis may be due to differences in the acquisition of cobalamin. Unlike *E. faecalis*, both *L. monocytogenes* and *C. perfringens* contain the genes necessary to synthesize cobalamin molecules; therefore, a transporter may not be necessary in these species. All three Gram-positive species I examined encode the EutVW two-component regulatory system and lack the EutR transcriptional regulator. This suggested that there are divergent mechanisms of regulation controlling *eut* gene expression. The transcription-factor-like, DNA-binding protein EutR is likely the

Figure 3.1 ***eut*** **locus organization of** ***E. faecalis*** **compared to** ***L. monocytogenes*** **and** ***C. perfringens*****.**

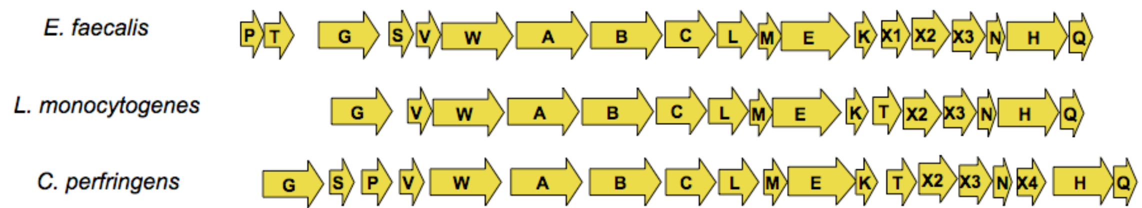


Figure 3.1

regulator that Gram-negative bacteria control *eut* locus gene expression. In Gram-positive bacteria, a two-component system is likely used to regulate *eut* locus gene expression.

Identification of the AdoCBL Riboswitch in *E. faecalis*

The 5'UTR of *eutG* in *E. faecalis* is unusually long, at 491 base pairs. Therefore, I hypothesized that the 5'UTR may contain important regulatory features such as transcriptional terminators. To investigate this possibility, I used the Ribex software to examine the 5'UTR sequence for potential regulatory features (2). Ribex is a web-based software program that identifies RNA regulatory elements, transcriptional attenuators, and open reading frames. I used this program to analyze the sequence of the entire *eut* locus in *E. faecalis*. My analysis uncovered a predicted structured RNA element similar to an AdoCBL riboswitch within the 5' UTR of *eutG*. Using this program, I also predicted that there are several transcriptional terminators in the *eut* locus. These terminators are found upstream of *eutP*, *eutG*, *eutS*, and *eutA*.

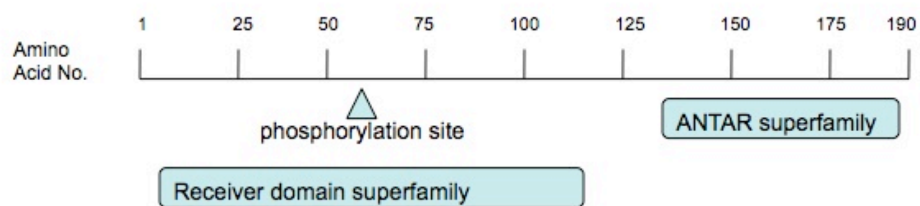
Riboswitches are structured RNAs found in the 5'UTR of genes that act *in cis* to control downstream transcription or translation. A metabolite binds to the first part of the RNA, the aptamer domain. This leads to a change in the structure of the downstream portion of the RNA known as the expression platform. The expression platform usually contains a transcriptional terminator or ribosome-binding site that is controlled by altering the RNA structure. This change in the RNA structure leads to changes in downstream gene expression. Previously studied AdoCBL riboswitches control the expression of genes for

the synthesis and import of AdoCBL and precursor molecules. In the presence of AdoCBL, these riboswitches negatively regulate transcription of the downstream genes. Previous global analyses to bioinformatically identify AdoCBL riboswitches in bacteria had failed to identify a riboswitch within the *eut* locus. I discovered that the context of the *eut* riboswitch I identified is quite different in that it is likely to control enzymes that require AdoCBL to function. I therefore postulated that it may have a unique positive regulatory role in *eut* transcription and represent a new class of AdoCBL riboswitches. In addition to a riboswitch, I also uncovered a predicted transcriptional terminator in the 5'UTR of *eutG*. Therefore, I predicted that this riboswitch functions to control the stability of the transcriptional terminator rather than the alternative function of altering accessibility to a ribosome-binding site.

In order to determine whether this RNA element is conserved, I used BLAST analysis and sequence alignment to show that this structure is also present in *Listeria* species. This demonstrated that the riboswitch is a conserved element used to respond to the presence of AdoCBL. Surprisingly, I found that the Gram-positive *C. perfringens*, despite having a similar locus organization, did not contain this RNA element. This suggested that other organisms may have a different method for integrating the AdoCBL signal. All together, my analysis indicated that there is a previously unknown and conserved riboswitch in the *eut* locus suggesting that this is important for metabolism across some prokaryotes.

Figure 3.2 Graphical representation of domain analysis of the two-component system proteins EutV and EutW in *E. faecalis*.

A. EutV domain analysis



B. EutW domain analysis

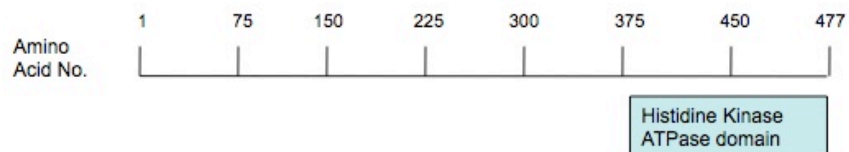


Figure 3.2

Domain analysis of the two-component system EutV and EutW in *E. faecalis*

The two-component system present in the *eut* locus in *E. faecalis* was previously identified (32, 33). However, the authors of the publications referencing this two-component system did not speculate on its function. The authors did show the two-component system to be different from the seventeen other two-component systems in *E. faecalis* because it is the only one predicted to have a soluble histidine kinase. The other two-component systems in *E. faecalis* have predicted transmembrane domains (32, 33).

To characterize the specific domains present in the two-component system, I used CDD to identify conserved domains present within each protein (Summarized in Figure 3.2). I found that EutV contains a C-terminal ANTAR family domain and the N-terminus of the protein contains a conserved receiver domain common to response regulators. Members of the ANTAR protein family have RNA-binding activity. ANTAR family members are thought to activate gene expression by preventing the formation of an intrinsic terminator stem loop present in the 5' untranslated leader sequences of the genes they regulate (73). In the putative receiver domain, I discovered that EutV contains a conserved predicted site of phosphorylation on aspartic acid 54.

I performed domain analysis of EutV's cognate sensor kinase, EutW, revealing that the C-terminus contains a conserved histidine kinase ATPase domain while the N-terminus lacks similarity to any protein family in the database. I predicted that this unique N-terminal domain might be the sensor for ethanolamine. Overall, my analysis showed that

the EutV and EutW two-component system contains the conserved domains present in typical two-component systems. Therefore, the two-component system likely functions via a phosphotransfer mechanism. From my domain analysis I predicted that a signal, likely ethanolamine, activates EutW inducing autophosphorylation on a conserved histidine residue. Next, EutW activates EutV by phosphotransfer to a conserved aspartic acid residue. Finally, I predicted that EutW functions as an antiterminator, interacting with nascent mRNA to prevent the formation of transcriptional terminators.

Discussion

Our lab has a strong interest in understanding the mechanisms underlying virulence in *E. faecalis*. A screen for *E. faecalis* mutants attenuated for virulence in *C. elegans* revealed that a *pduJ* is potentially important for virulence (29) Upon further analysis, I found that *pduJ* was actually *eutK,* which is part of the *eut* locus that encodes proteins used in the breakdown of ethanolamine. The *eut* locus had only been studied previously in *S. typhimurium* and *E. coli*. Like in *E. faecalis*, the *eut* locus also plays a potential role in *S. typhimurium* virulence making it an attractive target to study. Because much of the *E. faecalis eut* locus was annotated incorrectly, I performed an analysis using bioinformatic methods to identify the genes present in the *eut* locus as well as potential regulatory features. I found that many genes present in the *eut* locus in *E. faecalis* are also present in *S. typhimurium* indicating conservation of this pathway across prokaryotes. The main enzymatic genes are present in both species, suggesting that enzymatic functions of the metabolic pathway are the same. However, my analysis uncovered an AdoCBL riboswitch and an unusual two-component system in the *eut* locus. This led me to

hypothesize that a novel mechanism of regulation governs the *eut* locus in *E. faecalis* compared to *eut* loci in other organisms. Since my initial analysis, we have found that the two-component system is responsive to ethanolamine and responsible for expression of the *eut* locus (25). In addition, we demonstrated that AdoCBL binds to the region of the 5'UTR that contains a predicted riboswitch structure providing strong evidence for the existence of a true AdoCBL riboswitch. By comparing the *eut* locus in *E. faecalis* to several other Gram-positive organisms, I showed that a conserved mechanism of regulation of *eut* locus genes may exist across Gram-positive organisms that contain this locus.

After I performed my analysis of the *eut* locus across several different Gram-positive bacteria, a study was published performing a global analysis of *eut* loci across all bacteria with currently sequenced genomes (80). As the goal of the study by Tsoy et al. was to globally characterize the *eut* locus, they discovered that the loci containing genes for metabolism of ethanolamine were more diverse than I demonstrated in my analysis of *eut* loci in a limited number of Gram-positive bacteria (80). The data presented in the Tsoy et al. study support my hypothesis that differences in regulatory elements exist between Gram-positive and Gram-negative bacteria. The investigators found that in *Enterobacteriaceae* and in some *Betaproteobacteria*, the EutR regulatory protein is present (80). The *Enterobacteriaceae* family of bacteria contains *S. typhimurium*. The *Firmicutes*, which includes Enterococci, Listeria and Clostridium, contain the EutVW regulatory system. In addition, Tsoy et al. found the two-component system has a high rate of co-inheritance with cobalamin biosynthesis. However, *E. faecalis* does not

synthesize cobalamin, but rather likely imports this molecule. Import of cobalmin has not been shown experimentally however it is inferred from the absence of cobalamin synthesis genes and presences of genes used for import of cobalamin in other species. Also, this global analysis of *eut* loci identified certain *Actinobacteria* and *Proteobacteria* that contain a shortened version of the *eut* locus, which contain the EutBC lyase component (80). This shortened *eut* locus sometimes occurs with either the transporter EutH or *eat* and sometimes with the regulatory protein EutR (80). This suggested that in these bacteria the metabolic microcompartment may not be necessary for metabolism of ethanolamine. This idea was also supported by a study showing that the microcompartment genes were not necessary for ethanolamine metabolism in *S. typhimurium* when other enzymatic genes were overexpressed (11).

At the time I identified the riboswitch within the *eut* locus, the bioinformatic methods employed during global searches for riboswitches were unable to identify this structure (58). However, new methods for bioinformatic searches did identify this element confirming my analysis (81). The initial bioinformatic method to identify riboswitches used an algorithm that searched degenerate matches to a representative motif that was made using the *cob* and *btu* riboswitches from *E. coli* and *S. typhimurium*. This bioinformatic analysis obtained ninety-two matches for additional AdoCBL riboswitches (58). These analysis methods may have been too stringent to find the *eut* locus because of its divergence from the previously identified AdoCBL riboswitches.

Later work by Barrick and Breaker with different phylogenetic algorithms, using covariance models that were trained on sequence alignments followed by analysis of appropriate genomic context, were used to identify additional riboswitches (4). This analysis also uncovered the AdoCBL riboswitch in the *E. faecalis eut* locus. In addition, Barrick and Breaker found that AdoCBL was the most widely distributed type of riboswitch among different classes of bacteria (4). This study identified an AdoCBL aptamer core, consisting of several conserved base pairing interactions (4). Structural predictions performed by Dr. Wade Winkler's laboratory showed that several of base pairing interactions were maintained suggesting there is a functional relationship to canonical AdoCBL riboswitches (25). However two paired region interactions, referred to as P1 and P2, in the canonical AdoCBL aptamer core were missing. In addition, an extra paired region in the 3'end of the *eut* riboswitch was present (25). These differences from the canonical AdoCBL riboswitch are suggestive of a new subclass of riboswitch.

The information provided by the phylogenetic and bioinformatic analysis of the *eut* locus and riboswitch features was crucial information needed to develop a hypothesis for testing regulation of the *eut* locus. This work uncovered a novel combination of regulatory features that may function together on the posttranscriptional level to control gene expression. From this analysis I hypothesized that the two-component system is activated by ethanolamine. Activation results in the response regulator interacting with the nascent RNA to affect antitermination and allow *eut* gene expression. Furthermore, my bioinformatic analysis led me to postulate that the *eut* locus contains an AdoCBL riboswitch, which affects gene expression via an antitermination mechanism. Subsequent

experiments described in Chapter 4 tests my hypothesis, and work done in Chapter 5 further tests the mechanism of antitermination. Specifically, as will be shown in the next chapter, I demonstrated that AdoCBL, ethanolamine and the two-component system were necessary for expression of the *eut* genes in *E. faecalis*.

Chapter 4

Regulation of *eut* locus gene expression by AdoCBL and ethanolamine in OG1RF and *eutVW* strains

E. faecalis contains a large chromosomal locus for the metabolism of ethanolamine. As described in Chapter 3, I discovered what appeared to be several unique regulatory features present in the *E. faecalis eut* locus that are conserved in other Gram-positive bacteria. First, an AdoCBL riboswitch is present in the locus upstream of the gene *eutG*. This riboswitch is hypothesized to integrate a signal from AdoCBL to control gene expression of the genes downstream of the riboswitch. I hypothesized that the AdoCBL riboswitch likely acts as a positive regulatory element, influencing the stability of a transcriptional terminator upstream of *eutG*. The second unique regulatory feature of the *eut* locus is a two-component system composed of the histidine kinase EutW and the response regulator EutV. EutW is the only soluble histidine kinase in *E. faecalis*. EutV contains an output domain in the ANTAR family, which bind RNA and has antiterminator activity. I hypothesized a mechanism by which the two-component system regulates *eut* transcription. First, ethanolamine induces autophosphorylation of EutW, causing EutW to transfer the phosphate group to a conserved aspartic acid residue in the receiver domain of EutV. The EutV response regulator then activates expression of genes in the *eut* locus by causing antitermination at a series of four terminators present in the operon. These regulatory elements likely work together to integrate multiple signals and control the expression of genes in the *eut* locus on the posttranscriptional level. The following chapter describes experiments to test the hypothesis I formed about the mechanism of *eut* locus regulation in *E. faecalis*.

Development of a semi-defined medium for growth of *E. faecalis*

My first experimental goal to test the hypotheses developed in the previous chapter was to examine the response of *eut* locus genes to AdoCBL and ethanolamine. Therefore, I needed to grow *E. faecalis* cells in a defined growth medium in which ethanolamine and AdoCBL could be added exogenously. It was shown that *E. faecalis* growth could be supported in a completely defined medium (57). However in my hands, the *E. faecalis* cells exhibited very little growth despite using the exact same media conditions.

While I was searching for appropriate growth conditions, I received a personal communication from Dr. Agathe Bourgogne that growth in brain heart infusion broth (BHI) supplemented with horse serum may lead to transcriptional changes in genes in the *eut* locus. To determine the role of serum on *eut* locus expression, I measured changes in the gene *eutP* upon addition of serum. I constructed a translational fusion of *eutP* to *lacZ* (referred to as *eutP::lacZ* throughout this chapter) and transformed this construct into OG1RF and Δ*eutVW* backgrounds. I grew the cells in either BHI or BHI supplemented with 40% horse serum, and performed β-galactosidase assays to measure *eutP* gene expression. I discovered that *eutP::lacZ* expression increased 3-fold in BHI with 40% horse serum compared to growth in BHI alone, and this transcriptional response to serum is abrogated upon deletion of the two-component system (Figure 4.1). The transcriptional response of *eutP::lacZ* to serum was relatively small, but it showed differential regulation of *eut* locus genes under certain conditions and the dependence on

Figure 4.1 Media type and presence of AdoCBL and ethanolamine influence expression of *eutP::lacZ* in OG1RF and *eutVW* Expression of *eutP::lacZ* in wild-type (blue) and Δ*eutVW* (red) in various media. Expression of *eutP::lacZ* in wild-type cells grown in M9HY, in *eutVW* cells grown in BHI serum, in *eutVW* cells grown in M9HY, and in *eutVW* cells grown in M9HY with ethanolamine and AdoCBL was below the limit of detection for this experiment.

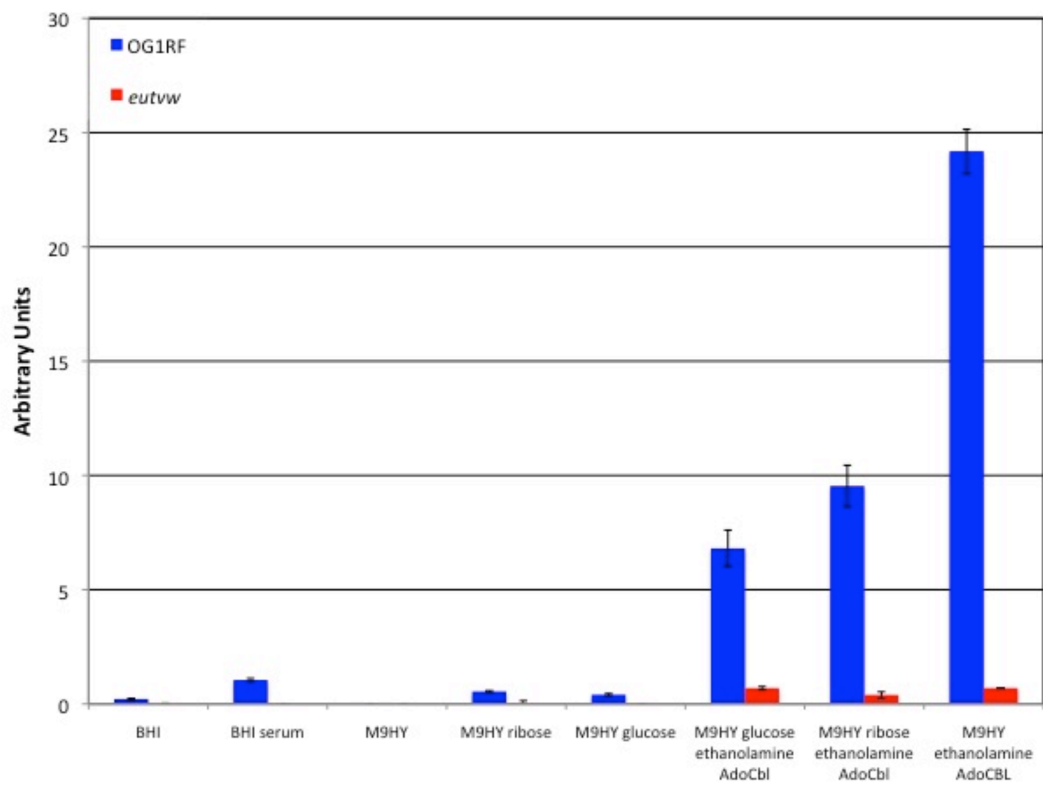


Figure 4.1

the EutVW two-component system for differential regulation. In addition, the response of *eut* locus genes to serum is suggestive of a role for the *eut* locus in virulence.

These data showed, for the first time, that a *eut* locus gene is regulated by the two-component system EutVW. However, several limitations to these growth conditions made the conditions suboptimal for continued study of *eut* locus regulation. First, BHI and BHI with serum are rich media. Therefore, the amount of ethanolamine or AdoCBL present in the media is unknown and cannot be accurately controlled. Secondly, the overall expression level of *eutP* in these conditions was low (Figure 4.1). Defined media that support the growth of *E. faecalis*, allow for more robust *eut* locus gene expression, and permit the testing of the effects of exogenous addition of ethanolamine and AdoCBL would be preferred.

A study was published showing that an *E. faecalis eut* locus gene responded to exogenous addition of ethanolamine and AdoCBL in a semi-defined medium containing yeast extract (19). In cells grown in strict anaerobic conditions, Del Papa and Perego showed that transcription of a reporter fusion of the promoter found upstream of *eutS* increased upon addition of AdoCBL and ethanolamine (19). Although the addition of yeast extract is not optimal because it adds unknown components to the growth medium, the authors were successful in examining transcriptional changes to *eut* locus genes with the exogenous addition of ethanolamine and AdoCBL.

A limitation of the Del Papa and Perego study was that a mutant in the two-component system was unable to grow in the semi-defined medium supplemented with only ethanolamine as a source of carbon (19). Therefore, the conditions that they developed are not useful to examine transcriptional changes that occur between the *eutVW* deletion and wild-type cells. In addition, it was important that I develop growth conditions in which transcriptional changes in response to exogenous addition of ethanolamine and AdoCBL could be monitored as well as support the growth of the *eutVW* mutant.

I created a modified version of the semi-defined conditions described above in order to support growth of Δ*eutVW*. I supplemented the medium with ribose as a secondary carbon source. This medium is referred to as M9HY with ribose. Ribose was chosen as a secondary carbon source because ribose, unlike glucose, was shown to avoid the effects of carbon catabolite repression in *E. faecalis* (45). As well as the addition of ribose, I changed the concentration of ethanolamine used in the assay. The previous study used 100 mM ethanolamine in their medium, which I found inhibited the growth of the OG1RF strain. Therefore, I lowered the concentration to 33 mM. A strain called V583 was used in the previous study, and this strain difference may explain the difference in tolerance to ethanolamine.

To determine if the new growth medium would permit me to examine the role of AdoCBL and ethanolamine in *eut* gene regulation in OG1RF and the *eutVW,* I performed β-galactosidase assays to examine expression of *eutP::lacZ* in wild-type and in the *eutVW* deletion mutant in the following media conditions. I used M9HY with ribose, and

M9HY with ribose supplemented with exogenous AdoCBL and ethanolamine. All cells were grown anaerobically. In wild-type cells, the β-galactosidase activity of *eutP::lacZ* in M9HY with ribose, ethanolamine, and AdoCBL was approximately 18-fold higher than the activity in M9HY with ribose only (Figure 4.1). In the *eutVW* deletion strain, there was no significant change in expression between the two media conditions (Figure 4.1). These data demonstrate that my newly developed growth medium, M9HY with ribose, is sufficient to examine the effect of exogenous addition of ethanolamine and AdoCBL on *eut* gene expression.

eutP::lacZ activity in M9HY with only ethanolamine and AdoCBL was substantially higher than expression of M9HYwith ribose, AdoCBL, and ethanolamine with 24.1 arbitrary units and 9.1 arbitrary units, respectively (Figure 4.1). The decreased expression seen upon the addition of ribose may have some repressive effects to *eut* gene expression. However, similar to the results in the Del Papa and Perego study (19), *eutVW* deletion did not grow in M9HY with only ethanolamine and AdoCBL.

I did an additional experiment in which I compared *eutP::lacZ* expression in cells grown in M9HY supplemented with ribose to those grown in M9HY supplemented with glucose. Expression of *eutP::lacZ* was higher in ribose than in glucose, 9.1 arbitrary units vs. 6.8 arbitrary units, respectively. This difference in expression may be indicative of catabolite repression. Because of this potential complication, experiments were continued with ribose supplementation.

Using this growth medium, I was able to determine whether the presence of EutV and EutW was required for expression of *eutP::lacZ* induced by ethanolamine and AdoCBL by comparing the amount of expression achieved in wild-type cells compared to the *eutVW* deletion mutant. In the *eutVW* strain grown in M9HY + ribose with ethanolamine and AdoCBL, the expression of *eutP::lacZ* did not increase as it did in wild-type cells (Figure 4.1). These results suggest that the change in expression of *eutP::lacZ* in the presence of AdoCBL and ethanolamine are dependent on EutVW.

Examining the expression of *eut* locus genes in response to AdoCBL and ethanolamine

In the section above, I described the growth medium I developed to examine expression of the *eut* locus genes that are both conducive to growth of a *eutVW* deletion mutant and sensitive to ethanolamine and AdoCBL. Using this medium, I showed that the transcription of *eutP::lacZ* is dependent on AdoCBL, EutVW, and ethanolamine. My next goal was to expand my analysis and examine the expression of more *eut* locus genes. My data in Chapter 3 describe four transcriptional terminators located in the *eut* locus. They are located upstream of *eutP*, *eutG*, *eutS*, and *eutA*. My hypothesis was that the two-component system controls expression by antitermination at each one of these terminators. To test this hypothesis, I examined gene expression using qRT-PCR of *eutP*, *eutG*, *eutS*, and *eutA* in M9HY + ribose with either AdoCBL or ethanolamine, and with the addition of both AdoCBL and ethanolamine. The primer sets used for amplification were also optimized so they had similar amplification efficiencies (Figure 2.1). Therefore, total expression could be compared between primer sets. Similar to the results

Figure 4.2. qRT-PCR of *eutP, eutG, eutS*, and *eutA* showed *eut* locus expression is dependent on AdoCBL, ethanolamine, and *eutVW*. Expression levels of *eutP* (blue), *eutG* (red), *eutS* (green) and *eutA* (purple) in qRT-PCR experiments of wild-type and *eutVW* cells grown in no addition, ethanolamine, AdoCBL and AdoCBL plus ethanolamine.

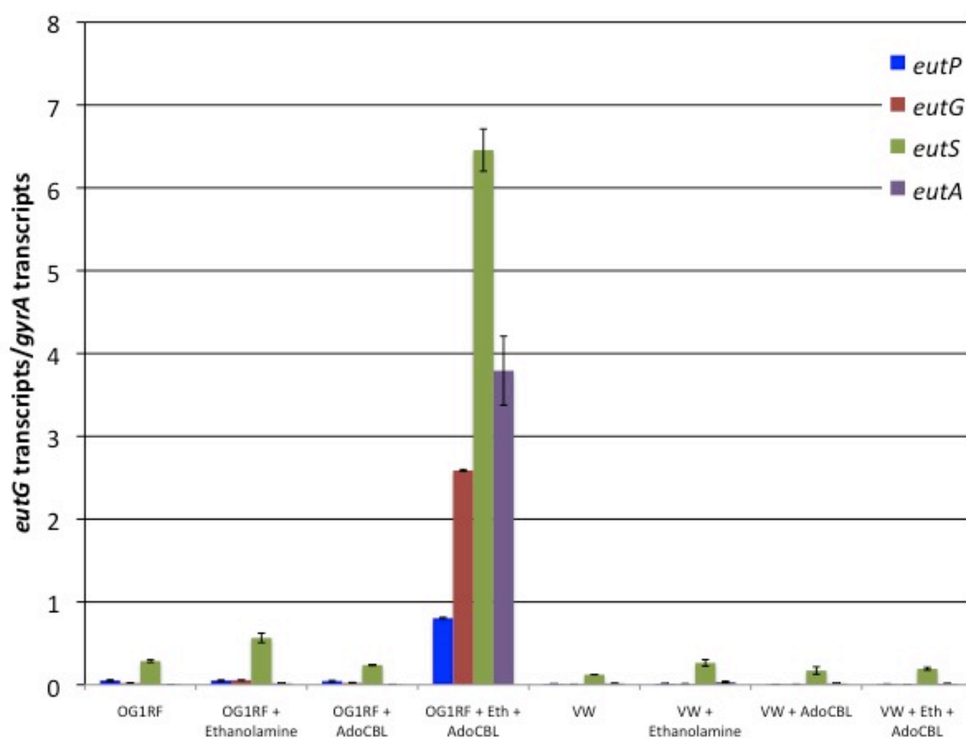


Figure 4.2

of the β-galactosidase assays of *eutP::lacZ* above, *eutP* expression was increased approximately 16-fold in the presence of ethanolamine and AdoCBL compared to media lacking these supplements (Figure 4.2). This increase in expression of *eutP* in the presence of ethanolamine and AdoCBL was not seen upon addition of AdoCBL only or ethanolamine only (Figure 4.2). This pattern of expression is similar for *eutG*, *eutS*, and *eutA* (Figure 4.2). A large increase in expression of these genes occurred when ethanolamine and AdoCBL were added together indicating that their addition induced gene expression (Figure 4.2). These data show that both ethanolamine and AdoCBL together are responsible for the increase in *eutP, eutG, eutS*, *and eutA* expression. Although the expression pattern the four genes follow is the same, the total transcript levels of each gene differ. *eutG* expression in inducing conditions was approximately four-fold greater than *eutP* expression (Figure 4.2). *eutS* expression was eight-fold greater than *eutP* expression, and *eutA* expression was 4.75-fold greater than *eutP* expression (Figure 4.2). These data suggest that rather than just an on/off mechanism controlling transcription of the *eut* locus genes, a more intricate pattern of regulation exists. Interestingly, the basal expression in non-inducing conditions also varies, as *eutS* basal transcription in non-inducing conditions was about 10-fold higher than that of the other genes (Figure 4.2). The *eutS* transcript likely contains the two-component system and may need to be present at a higher level within the cell to respond to the signal.

In the β-galactosidase assays of *eutP* described above (Figure 4.1), expression of *eutP::lacZ* was substantially lower in the *eutVW* two-component system deletion mutant. These results indicate that the attenuation of termination observed upon addition of

ethanolamine and AdoCBL is dependent on the two-component system. To examine the effects of the two-component system on other *eut* locus genes, I performed qRT-PCR on *eutP*, *eutG, eutS*, and *eutA* in the *eutVW* deletion strain using the same media conditions described above. In every gene tested, the increase in expression seen upon addition of AdoCBL and ethanolamine in wild-type cells was abrogated in the Δ*eutVW* strain (Figure 4.2).

These data demonstrated that the increase of *eutP, eutG, eutS*, and *eutA* transcript levels upon the addition of AdoCBL and ethanolamine is dependent on the presence of EutV and EutW. Differences in total transcript levels suggest that differential regulation exists at each gene and there is not just a single on/off mechanism controlling expression of the entire locus. *eutP*, *eutG*, *eutS*, and *eutA* each have a predicted promoter and terminator directly preceding the open reading frame (Figure 5.2). It is a reasonable postulate that the difference in transcript levels of these genes is likely due to differences in rates of transcriptional initiation, differences in rates of termination, or changes to mRNA stability.

Transcript analysis of *eut* locus genes

There is a predicted promoter region in the 5'UTR of *eutP*, *eutG* and *eutS* (Figure 4.3). There is also a predicted promoter upstream of *eutA,* but it may overlap with the coding region of the upstream gene *eutW*. Each promoter is upstream of a predicted terminator indicating that each gene is likely controlled by a transcription antitermination mechanism. My goal was to determine if each predicted promoter corresponded to a

Figure 4.3 5'RACE analysis of *eutP*, *eutG*, and *eutS*. Promoter sequences are underlined and marked with -35 and -10. Predicted transcriptional start sites are bold and marked with +1.

-35 -10 +1

eutP TAATTTTGACAAGTTTTTTGGAAACAGGTAAGATAAAGGCA**A**AGAA

-35 -10 +1

eutG ATTGACAATTAATTAAAGGCGTTGTACTCTATGGTGGATA**A**GGTT

-35 -10 +1

eutS ATTGACAGGATGAAATCGATAACATATAATGTGGGT**A**GCACA

Figure 4.3

transcriptional start site. I performed 5' Rapid Amplification of cDNA ends (5'RACE) to find the transcript ends of *eutP*, *eutG*, and *eutS* that are possibly indicative of a transcriptional start site. For *eutP*, I found that a transcript end was present upstream of *eutP* open reading frame and eight base pairs downstream of a predicted -10 sequence (Figure 4.3). The end of the *eutG* transcript was five base pairs downstream of a predicted -10 sequence (Figure 4.3). Finally, I discovered that a transcript end was present in front of *eutS*, eight base pairs downstream of a predicted -10 sequence of a putative promoter (Figure 4.3). All of these transcript ends represent possible transcriptional start sites. The fact that they lie downstream of good promoter sequences sequence makes it likely, but one would need to perform additional experiments to confirm these are actual transcriptional start sites and not the resulting products of transcript processing. Assuming these are transcriptional start sites, differences in the rates of initiation at these promoters may contribute to the difference in total transcript levels seen for these three genes by qRT-PCR (Figure 4.2).

One possible experiment to determine whether the 5' ends found by 5' RACE are from primary transcripts or processed transcripts is to use 5'tagRACE (24). This is similar to traditional 5' RACE except the RNA is first treated with a tag that ligates only to 5' monophosphate RNA ends, which are indiciative of processed transcripts (24). This tag prevents the addition of an additional tag, thus ensuring that the second tag labels only primary, unprocessed transcripts. Reverse transcription is then performed to make cDNA. Therefore, any amplification products produced by a PCR reaction of the cDNA

using primers that anneal to the second tag and a gene specific primer are primary unprocessed transcripts (24).

My attempts to uncover a 5'end for *eutA* proved to be unsuccessful, suggesting there may not be a promoter upstream of this gene. A *lacZ* fusion to this intergenic region generated by Dr. DebRoy was also inactive, which supports this hypothesis (personal communication, Dr. Sruti DebRoy). It is possible that a large transcript exists starting from the *eutP* promoter and continuing down to the end of the *eut* locus. In addition, several other small transcripts may exist starting from the other promoters identified within the *eut* locus.

These data demonstrate that three transcript ends that likely represent transcriptional start sites exist in the *eut* locus upstream of *eutP*, *eutG*, and *eutS* and indicate that several transcripts are potentially produced. A more thorough analysis to confirm that these are indeed start sites, and identify other promoters and the transcripts present in the *eut* locus in the future will help us better understand *eut* locus transcription.

Discussion

The data presented in this chapter describe the development of an appropriate growth medium for examining the transcriptional response to exogenous ethanolamine and AdoCBL in *E. faecalis*. Ribose supplementation of a previously described semi-defined medium allowed for grown of the *eutVW* deletion. This medium allowed me to compare expression of *eut* locus genes between *eutVW* and wild-type cells. I showed that

expression of *eutP::lacZ* was increased greatly in this medium upon supplementation with AdoCBL and ethanolamine. However, expression in the ribose supplemented medium was not as high as in media lacking ribose or glucose supplement, suggesting that genes may still be under control of carbon catabolite repression or another unknown regulatory mechanism. I examined transcription of four *eut* locus genes, *eutP*, *eutG*, *eutS*, and *eutA*. The presence of both AdoCBL and ethanolamine induced high expression of all of these genes. Importantly, this increase in expression was dependent on EutV and EutW as the increase in expression was abrogated in a mutant containing a deletion that disrupts both genes.

Interestingly, the transcript levels of each of the genes varied by several fold compared to one another. In a model in which transcription starts upstream of *eutP* and continues down to the end of the locus, it would be expected that each of the genes would have the same or less total transcript levels than *eutP*. However, this was not the case. Expression was higher in *eutG* and highest in *eutS* suggesting the presence of additional transcriptional start sites and additive total transcripts. 5' RACE analysis showed that a potential transcript start site exists upstream of both *eutG* and *eutS*, likely accounting for increased expression of these two genes over *eutP*. *eutA* does not likely contain a promoter as a 5' cDNA end could not be found using 5' RACE. In addition, reporter fusions to the 5'UTR of *eutA* were non-functional (personal communication from Dr. DebRoy). Therefore *eutA* is likely transcribed from the promoter upstream of *eutS*. This may explain why expression of *eutA* is less than *eutS* in inducing conditions (Figure 4.2).

The total number and length of the transcripts from the *eut* locus is not yet known. It is unknown whether a large transcript containing all *eut* locus genes is made as well as several smaller transcripts. Northern analysis will be useful in the future to examine total transcript lengths of *eut* locus transcripts.

My initial bioinformatic analysis described in Chapter 3 led to a series of hypotheses about the function of regulatory elements of the *eut* locus. First, I predicted that the two-component system was activated by ethanolamine and controlled expression of genes in the *eut* locus by a mechanism of antitermination at a series of terminators located at *eutP*, *eutG*, *eutS* and *eutA*. Our collaborators and another independent group showed that *in vitro* ethanolamine induced activation of EutW and phosphotransfer to EutV (19, 25). I showed that ethanolamine was required for expression of *eut* locus genes (along with AdoCBL). Additionally, I showed that this increase in expression was dependent on EutV and EutW. These data support my hypothesis that EutV and EutW control expression of *eut* locus genes in the presence of ethanolamine. Further work needs to be done to examine the mechanism of antitermination by EutV. Another prediction made by my hypothesis is that AdoCBL interacts with the riboswitch leading to antitermination of the *eutG* terminator and resulting in increased expression of *eut* locus genes. *In vitro* experiments done by our collaborators showed that AdoCBL bound to the *eut* riboswitch RNA and binding of AdoCBL to riboswitch RNA lead to changes in the riboswitch structure (25). They also showed using *in vitro* transcription assays that some relief of termination occurred in the presence of AdoCBL, but the relief of termination was incomplete. These results suggest that other factors are also involved in transcription

termination. EutV is potentially needed for complete antitermination of the *eutG* terminator. My experiments demonstrated that AdoCBL was needed for induction of expression of *eutG* as well as the other *eut* locus genes *eutP*, *eutS*, and *eutA*. This suggests that the role of AdoCBL may be more complex than just leading to instability at the *eutG* terminator. In addition, the AdoCBL riboswitch may be involved in a positive feedback loop regulating expression of *eutV* and *eutW*.

Together, these data have led me to form some interesting new ideas about the mechanism of *eut* locus regulation. First, the *eut* locus may be transcribed on multiple transcripts. There may be a long transcript from *eutP* down through the end of the *eut* locus that is present in non-inducing conditions. This long transcript may result in the basal level of transcription seen without the addition of AdoCBL and ethanolamine. This idea is supported by the fact that the basal transcript levels of *eutP*, *eutG*, and *eutA* were similar. However, *eutS* transcription existed at a higher basal level suggesting that in non-inducing conditions, an additional transcript starting from the *eutS* promoter exists that likely contains *eutV* and *eutW*. Having more of the two-component system protein present in non-inducing conditions may help to respond to changes in intracellular ethanolamine. Interestingly, in inducing conditions total transcript levels of *eutG* were much higher than *eutP* and *eutS* transcript levels were higher than *eutG*. These results suggest that active transcription from each of the promoters is increased individually in induction conditions. Northern analysis of genes in the *eut* locus will help to understand how *eut* locus genes are transcribed.

These data also suggest that an interesting mechanism of antitermination is employed by EutV to control expression at the terminators. It is likely that transcription begins from the three putative promoters in the *eut* locus but is terminated by the transcriptional terminators in conditions in which the two-component system is not activated. Under inducing conditions, EutV likely acts as an antiterminator leading to the increased expression of *eut* locus genes. I hypothesize that differences in total transcript levels of *eut* locus genes may be due to the presence of multiple transcripts. However, these changes could additionally be due to the efficiency of antitermination by EutV at each terminator. Future experiments to identify how the EutV protein interacts with mRNA and if the affinity for interaction at each terminator differs will provide further insight.

Chapter 5

Regulation of *eutG* expression by ethanolamine, the ANTAR response regulator, and the AdoCBL riboswitch

My data in the previous chapter showed that expression of *eutG* is dependent of AdoCBL, ethanolamine, and the two-component system (Figure 4.2). The mechanism by which these multiple inputs control expression may lie in a series of regulatory features found in the unusually long 5'UTR of *eutG*. These features include an AdoCBL riboswitch, two ANTAR substrate domains (regions in the RNA predicted to bind to ANTAR family proteins), and an intrinsic transcriptional terminator (The *eutG* 5'UTR sequence and features are depicted in Figure 5.1). This chapter will focus on experimentally dissecting how these features work together to control *eutG* gene expression. We have a working model for ANTAR function without the presence of a riboswitch. This work both examines and expands on the current model.

EutV is an ANTAR family protein

Two-component systems (TCS) are a very common tool bacteria use for sensing and responding to environmental and intracellular signals. Canonically, a sensor histidine kinase autophosphorylates in the presence of the inducer, then transfers its phosphoryl group to a response regulator protein. Phosphorylation activates the response regulator and the activated response regulator affects downstream processes through a variety of output domains.

Response regulator proteins can contain different output domains, most commonly the regulator will bind to DNA and change expression on the transcriptional level (47). EutV belongs to the rare ANTAR family of response regulators, making up less than 1% of known response regulators. This type of regulator binds to RNA and has

Figure 5.1 Sequence of *eutG* 5'UTR with relevant features. The riboswitch, P3, P4, P5 and P6 are highlighted, as well as the promoter sequence.

```
eutT Stop codon                                                  {-----riboswitch---------------------------------------
TAAACAGTTGCTTTTCTAAAAAACTTATGGTAAATTTCAGATAGATTAAACGAATATTTTATGAAAAATATGAATGGAAGCCAGTGAGAATCTGGCACGG  < 100
        10        20        30        40        50        60        70        80        90

--------------------------------------------------------------------------------------------------}
TCCCGCCACTGTGAAGAAGCAAGGTTGCTTTTAAGTCAGGTCTTTTCATTTTTCATTATTGGGCATGCTGTTTCGAGGCAAAACAGGATGTTTCTTAACA  < 200
        110       120       130       140       150       160       170       180       190

                                                                                        P3
ACGCTTGTGTTGAAAATCCAGCCCAAGATATTTGTATTAATCCAATTAATGGCACGAGGAAGTGCTTTTGGGAATTTTGGGCTGGTCTTTCAATGTTGAA  < 300
        210       220       230       240       250       260       270       280       290

          P4                            -35                  -10 promoter               P5
GGAGATTTCAGGTACAATGACGTACTGCTATATTCCCTTTTTTTATTGACAATTAATTAAAGGCGTTGTACTCTATGGATAAGGTTTCGTGTACAATGGC  < 400
        310       320       330       340       350       360       370       380       390

          P6
GTATACATAAGGAAGCAAAGACGCTTCAGACAGATGATTTTTCGTTTGTTTGAAGTGTCTTTTTGTATTTTATTAAGGAGGAATTGAGATG  < 491
        410       420       430       440       450       460       470       480       490

Features
1-3 =   eutT stop codon
57-196 = AdoCBL eut riboswitch
284-303 = P3 (paired region in green highlight)
311-324 = P4 (paired region in blue highlight)
345-352 = Promoter -10 sequence
365-375 = Promoter -35 sequence
389-406 = P5 (paired region in green highlight)
412-425 = P6 (paired region in blue highlight)
419-458 = Terminator stem/loop (paired region underlined)
478-482 = Ribosome binding site
489-491 = eutG start codon
```

Figure 5.1

antitermination activity (73). ANTAR family proteins act on intrinsic transcriptional terminators; their association with nascent transcripts prevents terminator formation, thus having a positive overall effect on transcription of downstream genes (73).

There are four transcriptional terminators present in the *eut* locus, in the 5'UTR of *eutP, eutG, eutS* and *eutA*. Therefore, we hypothesized that EutV regulates the genes of the *eut* locus by an antitermination mechanism. EutW is activated by ethanolamine and in turn activates EutV by phosphorylation. Our model predicts that EutV interacts with the 5'UTRs in the nascent mRNAs that contain these putative terminators. The interaction is predicted to occur with sequences adjacent and part of the terminator thereby preventing terminator formation and allowing read-through (Figure 5.2) (27).

AmiR: a well studied of ANTAR regulatory protein

A well-studied ANTAR family protein is AmiR in *Pseudomonas aeruginosa.* The amidase operon in *P. aeruginosa* encodes genes that allow for the metabolism of small amide compounds. Amidase operon expression is regulated by the atypical regulatory system AmiC and AmiR (86). In the absence of inducing ligands (acetamide or propiamide) or presence of a co-repressor (butyramide), two molecules of AmiC form a complex with a dimer of AmiR, preventing AmiR antiterminator activity (60). When inducing ligands are present, AmiC changes conformation, and no longer hinders the antiterminator function of AmiR. AmiR then interacts with the UTR of *amiE* causing antitermination, and allowing for transcription of the downstream amidase operon genes. Mutational studies of the RNA that interacts with the regulatory protein have elucidated

Figure 5.2 Regulation of *eut* locus in *E. faecalis.* Ethanolamine causes autophosphorylation of EutW followed by phosphotransfer to EutV. The activation of two-component system leads to disruption of four transcriptional terminators. AdoCBL also promotes antitermination through interaction with the riboswitch.

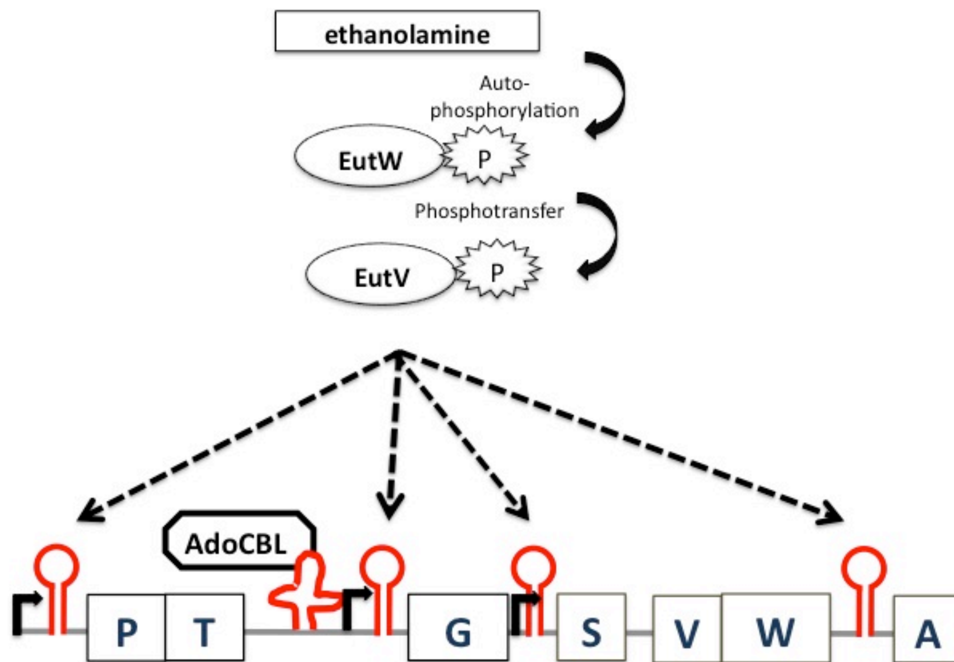


Figure 5.2

some of the sequence requirements for the interaction, but did not define a specific antiterminator structure that explains the mechanism (60).

NasR is an ANTAR regulatory protein

NasR is a regulatory protein in *Klebsiella oxytoca* that governs expression of the nitrate assimilation operon composed of the genes *nasFEDCBA* (15). NasR is an ANTAR family protein; it differs from AmiR because it does not partner with a cognate sensor protein. NasR is activated directly by nitrate molecules, then positively regulates expression by inhibiting the formation of a transcriptional terminator in the *nasF* 5' UTR (16). Mutations in the *nasF* leader again provide some information on what sequences are required, but do not define a specific antiterminator structure that defines the mechanism (16). The *nasF* leader mRNA consists of a transcriptional terminator and a second stem loop upstream. The authors postulate that binding of NasR to the first stem loop disrupts formation of the terminator structure, by a mechanism that does not involve the formation of an alternate antiterminator. A typical antitermination mechanism would form an alternate structure, which contains parts of the sequence in the terminator structure, thus preventing formation of the terminator. The authors argue against this idea because mutations they made in what they believe to be the alternate structure confer no change to antitermination (15, 16).

Studies of both NasR and AmiR examined how the ANTAR family protein interacted with the mRNA and both showed that there were some sequences and possible secondary structural features required for protein RNA interaction, but no true mechanism of

antitermination was found by either team of investigators. In the case of NasR, the authors even reject the idea that an antiterminator structure was formed (16).

The universal ANTAR model

In Chapter 3, I showed that a conserved sequence is found in the UTR. This sequence overlaps the terminator of several *eut* locus genes in *E. faecalis* as well as the *eut* locus genes in other species such as *Listeria monocytogenes* and *Clostridium difficile.* Interestingly, the same sequence is also in the UTR of an AmiR controlled gene in *P. aeruginosa*. This sequence is AGCAANGRRGCUY (N= A, C, G or T, R= A or G, Y= C or T) (25). Folding predictions by the Mfold program using an extended sequence containing this motif found two hairpin loops fold in the RNA (Figure 5.3). These putative hairpins are downstream of predicted promoters, and a section of the second loop overlaps the 5' end of the intrinsic terminator sequences. This indicates that the hairpins could be part of an antitermination structure. A second key feature of the UTR is a conserved motif present on the "loop" of each stem loop (Figure 5.3). An A residue in the first position on the loop and a G residue in the fourth position are highly conserved through many ANTAR substrate UTRs (referred to as AXXG motif throughout). These sequences are found also in NasR and AmiR substrate UTRs. In *amiE*, a small stem loop is found upstream of the terminator and an additional stem loop is found overlapping the terminator which forms the second hairpin loop. The sequence of the *nasF* UTR contains a large predicted stem loop upstream of the transcriptional terminator (Figure 5.3). This first stem loop structure was larger than other ANTAR substrates (about 13 bases instead of 6), and the primary sequence differs quite a bit from the other UTRs. However,

careful examination of the *nasF* UTR has uncovered two small hairpin stem loops containing AXXG motifs that can form as an alternate secondary structure. Because these sequences are highly conserved, this sequence is a proposed recognition sequence for all ANTAR proteins. The following describes the universal model of ANTAR regulation.

In a non-induced state, nascent mRNA forms an intrinsic terminator causing transcription termination (Figure 5.3). In the induced state, a dimer ANTAR protein interacts with the two stem loops via the AXXG motif of the nascent mRNA. This locks in a secondary structure that prevents the formation of the terminator and RNA polymerase can proceed with transcription. Dimerization of the ANTAR domain-containing protein is also likely, with each subunit of the dimer interacting with one loop at the conserved AXXG motif (Figure 5.3). In our system, ethanolamine activates EutW by inducing autophosphorylation. EutW activates EutV by phosphotransfer, then we hypothesize that the ANTAR domain-containing EutV protein dimerizes and interacts with the two small hairpin loops at the conserved AXXG motif within the nascent mRNA UTR of *eutP, eutG, eutS*, and *eutA*. This interaction between EutV and the AXXG motif prevents the formation of an intrinsic terminator and expression of downstream genes can continue.

Preliminary testing of the ANTAR model using *eutP* 5' UTR

This section describes unpublished work performed by Dr. Sruti DebRoy, a postdoctoral fellow in the Garsin laboratory. The ANTAR hypothesis rests on the control of a transcriptional terminator that forms in the nascent RNA. Dr. DebRoy created a *eutP*

Figure 5.3 Universal model of ANTAR regulation. A. The EutVW TCS is not activated and terminator is formed. B. TCS is activated by ethanolamine then EutV dimerizes and interacts with stem loops at conserved AXXG motifs.

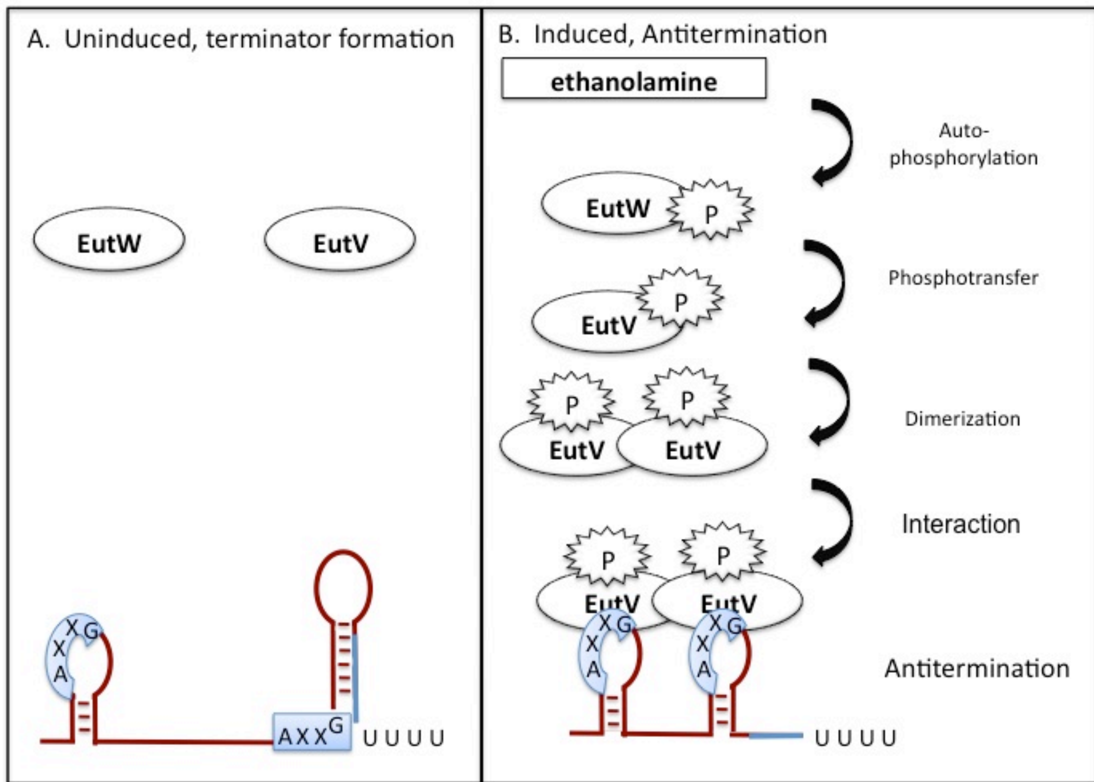


Figure 5.3

5’UTR lacking a transcriptional terminator, and fused this sequence to a promoterless *lacZ*. In a wild-type strain, *lacZ* was expressed at a low level in media lacking both AdoCBL and ethanolamine. When both AdoCBL and ethanolamine are present, *lacZ* expression was increased. Using β-galactosidase assays, she found constitutive expression of the *eutP* terminator deletion reporter construct. Surprisingly, *lacZ* expression occurred in her assay even in conditions lacking the inducers AdoCBL and ethanolamine while there was little expression from the native *eutP* 5’UTR in these conditions. To examine the function of the P1 stem loop, she engineered a *eutP* 5’UTR in which the P1 stem loop had been deleted and fused this sequence to a promoterless *lacZ* for β-galactosidase assays. Deletion of the P1 stem loop reduced expression in inducing conditions with addition of AdoCBL and ethanolamine. In addition, she engineered two additional P1 stem loop with mutation of either the A or G conserved bases of the AXXG motif. In these two P1 stem loop point mutants, *lacZ* was also uninducible indicative of the importance of these conserved residues for *eutP* expression. Experiments to determine the function of the P2 loop by mutational analysis are currently being performed in the lab by Dr. DebRoy, and we do not yet have clear conclusions about the role of this potential structure.

In summary, either deletion of the P1 or individual point mutants in the AXXG motif of this loop renders *eutP* uninducible. The data provide preliminary support for our model that the sequence containing P1 and P2 stem loop form functional structures and provide part of the rationale for my experiments on the *eutG* leader described below.

Expanding on the universal ANTAR model at *eutG*

In common with *eutP* and other ANTAR substrate UTRs, *eutG* UTR contains an intrinsic terminator, along with the hairpin loops analogous to P1 and P2 that overlap with the terminator stem. The stem loops at *eutG* are referred to as P5 and P6, because they are the fifth and sixth pair regions found in the *eut* locus. However, *eutG* UTR contains many other features that make this more complex and our model needs to be expanded to account for this complexity.

First, the AdoCBL riboswitch is found in the *eutG* 5'UTR, approximately 220 bases upstream of the terminator. A second region of putative hairpin loops (referred to as P3 and P4) is found downstream of the riboswitch and upstream of the first loops. Approximately sixty base pairs separate the two 'pairs'.

In vitro experiments by the Winkler lab using an in-line probing assay showed that the riboswitch bound AdoCBL. *In vitro* transcription run-off assays demonstrated that increased read-through of the terminator occurred with increasing concentrations of AdoCbl (25). In addition, AdoCBL was necessary to induce *eutG* expression in cells as I show in my experiments described below. These data support a model in which efficient antitermination in the *eutG* leader requires AdoCBL binding to the riboswitch.

The additional putative stem loops, P3 and P4, do not have an obvious role; they are not near an intrinsic terminator like the other stem loops in other ANTAR substrate UTRs.

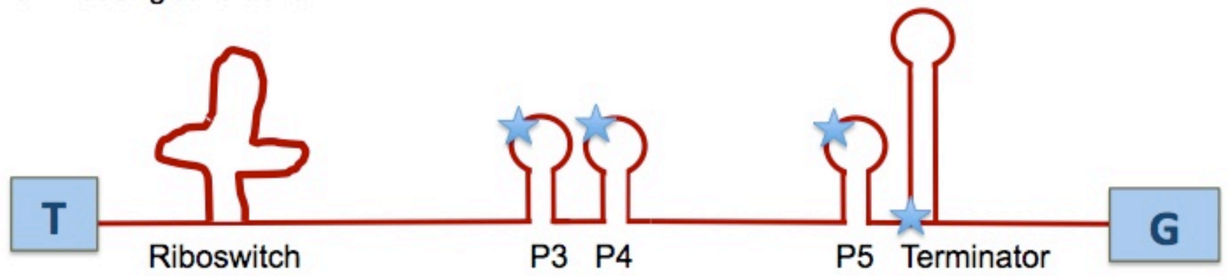
These stem loops may serve as a second site for interaction with a second dimer of EutV. The interaction of a dimer of EutV with P3 and P4 may recruit or promote binding of a second dimer to P5 and P6. P5 and P6 binding by EutV leads to the formation of the antitermination complex. Therefore, the combination of riboswitch and ANTAR protein binding results in complete antitermination at *eutG*. It is unclear whether the interaction of EutV or AdoCBL with the 5' UTR needs to happen sequentially or if they occur independently. But together the ethanolamine signal, mediated by EutV, and the AdoCBL signal synergistically lead to expression of *eutG* (Figure 5.4).

I developed a series of mutations in *eutG* 5' UTR to examine the model of *eutG* regulation described above. There are several important components to this model. First is the intrinsic terminator, which we believe ultimately is inhibited from forming by EutV and the riboswitch. Next, are the downstream hairpins (P5 and P6), which are thought to be a part of the antitermination complex. Third are the upstream hairpins P3 and P4, which we believe lead to cooperativety between two EutV dimers. Lastly is the riboswitch, which has a synergistic regulatory effect with the two-component system.

Results

I previously examined the expression of *eutG* in wild-type cells (Figure 4.2). I performed qRT-PCR to measure expression levels of *eutG* in cells grown in media containing AdoCBL, ethanolamine, both AdoCBL and ethanolamine, or neither additive. I found

Figure 5.4 Modified Model of ANTAR regulation for *eutG*. Non-inducing conditions- No interaction of AdoCBL or EutV with the nascent transcript and terminator formation occurs. Inducing conditions - The interaction of AdoCBL with the riboswitch and EutV-P dimers with two stem loop pairs (P3,P4 and P5,P6) cause destabilizing of the terminator structure. Antitermination occurs and transcription proceeds through *eutG*.



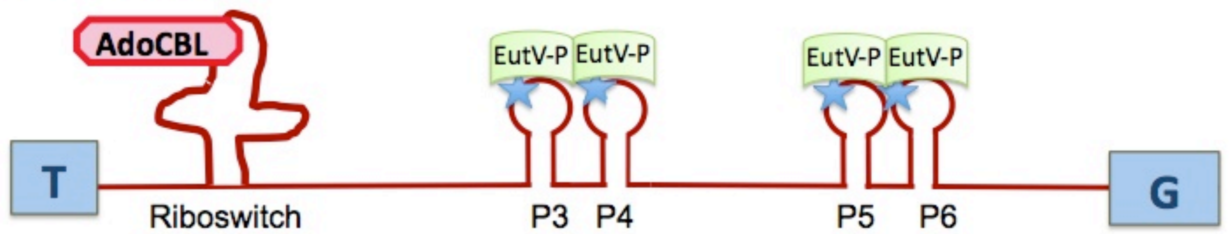


Figure 5.4

that the presence of AdoCBL and ethanolamine together caused a 20-30 fold increase in *eutG* expression. The other three conditions, AdoCBL only, ethanolamine only, and lacking AdoCBL and ethanolamine, still had a basal level of transcription, indicating that even in these non-inducing conditions, a small amount of transcription can proceed.

Mutational analysis of the transcriptional terminator

Ultimately the features described: P3, P4, P5, P6, and the riboswitch are predicted to control the stability of transcriptional terminator. Without the terminator present or with mutations that disrupt pairing, transcription from the upstream promoter should be constitutive. The terminator is predicted to be 46 bases long with 20 base pairing interactions making the stem. The terminator stem is also followed by a run of Us which is a conserved feature of intrinsic terminators (31, 85).

The first mutation I designed and constructed deleted most of the bases in the terminator while maintaining the bases predicted to be in P6. Second, I made a mutation in which the base pairing had been disrupted in five base pair interactions in the lower part of the stem. I used mFold to examine the change in stability by calculating the change in ΔG. The ΔG of the *eutG* terminator stem was -26.50 while in the mutant it changed to -15.40. It is possible that the top of the stem may still form in this mutant and be able to cause termination. However, this is unlikely because if a smaller stem still forms the run of Us would no longer be following the stem. Lastly, I made a mutant in which base pairing was interrupted for five bases at the top half of the terminator stem. The ΔG was changed from -26.50 to -12.60. Using mFold predictions, it is likely the bottom pairing

interactions may still occur. A schematic of these terminator mutants is diagrammed in Figure 5.5.

Preliminary experiments by Dr. DebRoy on the *eutP* intrinsic terminator showed that upon deletion of the terminator constitutive expression occurred through the *eutP* coding region. I therefore predicted a similar result for the terminator deletion mutants I constructed within *eutG*. Transcription of *eutG* starts from a predicted upstream promoter, proceeding either *eutP* or *eutG,* as supported by 5'RACE analysis (Figure 4.3). With no terminator to stop the RNA polymerase, transcription should occur through the downstream gene *eutG*. This should occur regardless of the media conditions tested. For the stem base and the stem top mutants I also predicted that the expression of *eutG* would be constitutive. However, there is a caveat to this prediction. It is possible that a partial stem may form in either stem mutant, as predicted using mFold. In stem mutant 1 this may not influence transcription termination because the run of Us no longer follows the stem. However in stem mutant 2 and the terminator deletion, any residual stem formation will be followed by a run of Us and may influence transcription termination. Therefore, expression of *eutG* may not be as robust as in the complete terminator deletion mutant and stem mutant 2. I examined the expression of *eutG* in each of these mutations using qRT-PCR as described in Chapter 3. I performed these experiments in the four media conditions as described in Chapter 3. I compared *eutG* expression in each mutant to *eutG* expression in the wild-type strain under the same conditions.

Figure 5.5 Terminator mutations. Schematic representation of terminator mutants constructed. A. wild-type terminator sequence. B. Stem mutation 1. C. Stem mutation 2 . D. Terminator deletion.

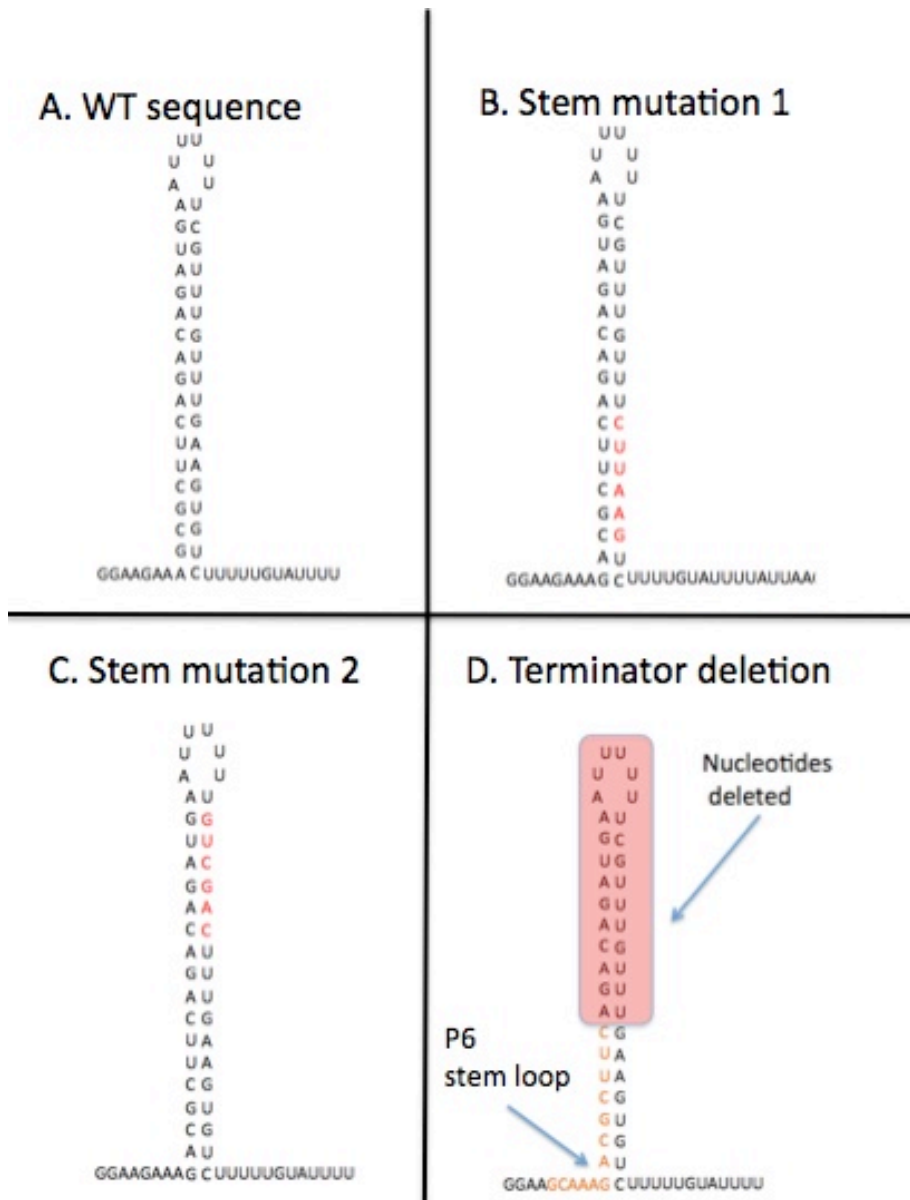


Figure 5.5

In wild-type cells, low level (basal) expression of *eutG* was seen in the non-inducing conditions containing either AdoCBL, ethanolamine, or lacking either additive. When both AdoCBL and ethanolamine were added to the growth medium, *eutG* expression increased 20-30 fold. Upon deletion of the terminator, the expression levels of *eutG* in normally non-inducing conditions were dramatically higher compared to wild-type in the same conditions. These data suggest that deletion of the terminator leads to constitutive expression. Interestingly, under inducing conditions (both AdoCBL and ethanolamine) there was evidence of a small but statistically significant (p value = 0.032) induction, as *eutG* expression increased by 25%. Potentially, an undeleted portion of the terminator may remain partially active. Another explanation is that expression driven from the upstream *eutP* promoter (Figure 4.6), which is still wild-type and therefore inducible, is responsible for this small increase under inducing conditions.

Stem mutant 1, that disrupts the bottom pairing bases of the terminator, displayed a similar pattern of *eutG* expression to the mutant containing deletion of the entire terminator. Constitutive expression of *eutG* was seen in normally non-inducing conditions in this mutant indicating that the bottom pairing bases are required for terminator function. In inducing conditions, there was an approximately 30% increase in *eutG* expression. This small increase in expression seen upon AdoCBL and ethanolamine was statistically significant (p-value = 0.0082). This result could indicate that either the terminator is not completely disrupted or there is additive expression from the *eutP* promoter (Figure 5.6).

Figure 5.6 qRT-PCR of *eutG* in terminator mutants and wild-type. Mutation of the terminator results in constitutive expression of *eutG* in non-inducing conditions. *eutG* expression in AdoCBL and ethanolamine is slightly higher than under non-inducing conditions in the terminator deletion (red) and stem mutation 1 (green). Error bars indicate standard deviation. Etn = ethanolamine.

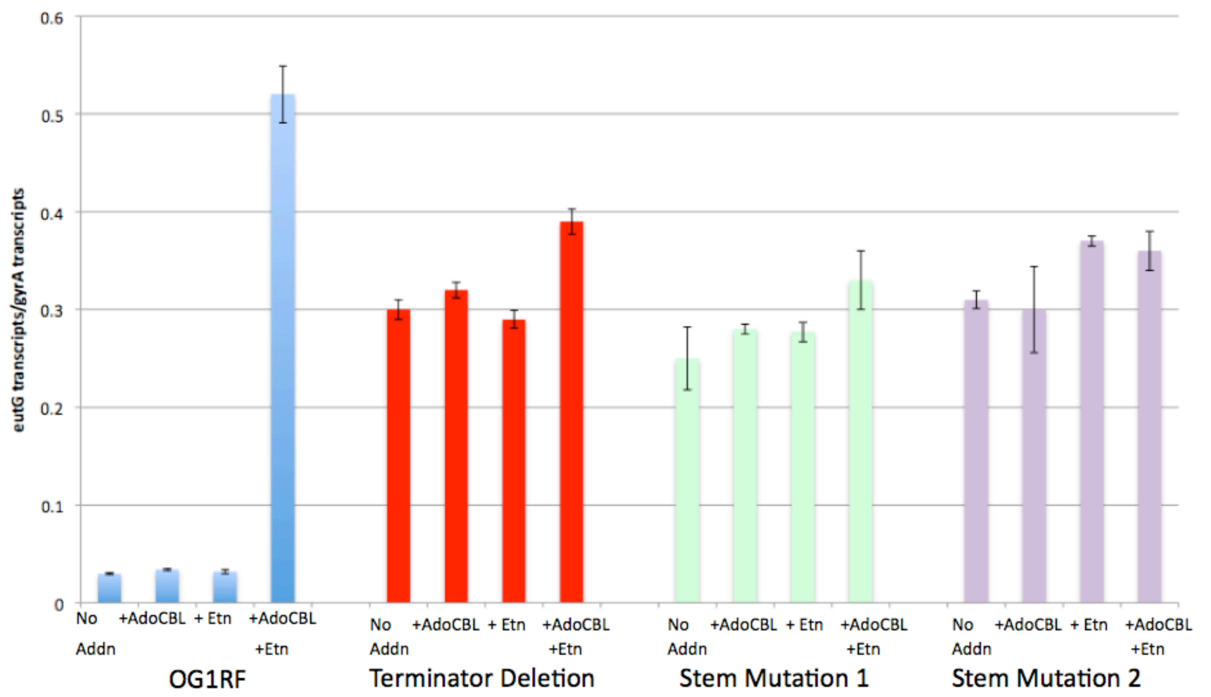


Figure 5.6

Stem mutant 2 contained a disruption in the pairing at the top of the terminator and displayed a different phenotype compared to the other previously described mutants. As seen in the other two terminator mutants, higher levels of *eutG* expression were seen under non-inducing conditions. Additionally, expression increased approximately 20% in the presence of both AdoCBL and ethanolamine (p value = 0.014). However, cells grown with only ethanolamine did not follow the same pattern observed with the other two terminator mutants. The addition of only ethanolamine to the media resulted in an approximately 20% increase in expression compared to cells grown in media lacking both AdoCBL and ethanolamine (p value = 0.0051) (Figure 5.6).

Overall, there was constitutive expression of *eutG* in mutants with a disruption or a deletion in the putative terminators, providing evidence for the formation of this structure and its importance in regulating *eutG* expression. One important difference between my studies of terminator deletions in the 5'UTR of *eutG* and Dr. DebRoy's similar studies of terminator mutations in the 5'UTR of *eutP* is that in all my terminator deletion mutants, expression of *eutG* underwent a slight increase in the presence of both AdoCBL and ethanolamine. As stated above, this may suggest that my mutations do not completely disrupt terminator function, or alternatively, that the upstream *eutP* promoter is contributing to the overall gene expression observed. Disrupting the *eutP* promoter would be one method by which we could distinguish between these two possibilities. Of particular interest is stem mutant 2, which completely loses dependence on the ethanolamine inducer but retains partial dependence on AdoCBL. This result could indicate that the part of the terminator that remains is still sensitive to the antitermination

effects of EutV, but not the riboswitch, potentially hinting at a dissectible difference between how these two antitermination mechanisms both act on this terminator. This makes sense considering that the part of the terminator that remains does contain the P6 loop, which I am hypothesizing is what EutV binds.

Mutational analysis of the P5 and P6 stem loops

The universal ANTAR model posits that a dimer of EutV interacts with two stem loop pairs, locking in a structure that prevents the hairpin terminator from forming. The presence of the conserved AXXG motif on the loop of each stem is also required for interaction. Therefore without the stem loops, or if the AXXG sequence is mutated, the antitermination complex will not form but the transcription terminator will, leading to no induction of expression of the downstream genes. Preliminary experiments by Dr. DebRoy on the P1 loop of *eutP* fit with this hypothesis as mutation or deletion of the P1 loop greatly inhibits antitermination.

To examine if the P5/P6 region in *eutG* functions in accordance with the model, I made three mutations in the *eutG* P5/P6 region. The first mutant changed the sequence containing the AXXG motif of P5 from CAAUGC to CUUAAG, which is a AflII restriction endonuclease cleavage site (Figure 5.7B). According to our model, this mutant should reduce transcription due to a lack of formation of an antitermination complex.

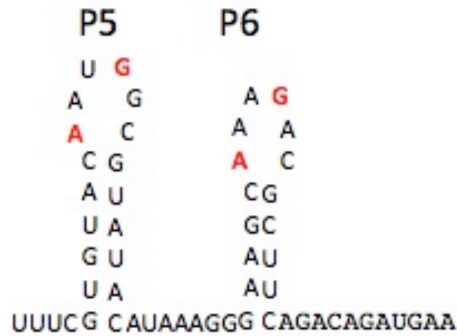
I also constructed a mutant in which the sequence containing the AXXG motif in P6 was mutated from GCAAAG to GUUAAC, again changing both conserved bases. This mutation also eliminated a predicted base pairing interaction in the stem (Figure 5.7C). According to the model, expression of *eutG* will not be induced in any conditions because the antitermination complex will not form.

The third mutant has changes of predicted pairing bases in the P5 stem to non-pairing bases (Figure 5.7D). Based on the model, if the P5 stem is not formed, little *eutG* expression will be seen.

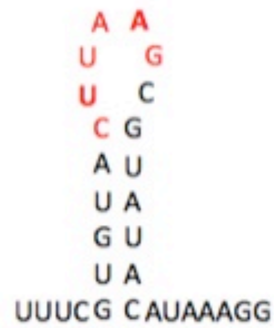
Mutants 1 and 2 individually changed the AXXG motif of P5 and P6 respectively. To investigate the role of the P5 and P6 AXXG motifs in *eutG* expression, I used qRT-PCR to determine the expression of *eutG* in the two mutants that had individual changes to the AXXG of P5 or P6. Both mutants had low transcript levels of *eutG* in all conditions tested (No addition, AdoCBL, ethanolamine, and AdoCBL and ethanolamine). When I compared the *eutG* expression levels in the mutants to that measured in wild-type, I found that the mutants displayed similar expression level to the wild-type locus in non-inducing conditions (Figure 5.8). Therefore, the conserved AXXG motif on both P5 and P6 are likely required for proper formation of the antiterminator complex at *eutG*. In P5 there was a slight increase in transcription in the inducing conditions. A similar increase was observed in P6, but this result was not consistent between experiments (Figure 5.8). This may indicate that a slight level of induction occurs in these mutants.

Figure 5.7 Schematic representation of mutations to P5 and P6 stem loops. A. Wild-type P5 and P6 sequence. B. AXXG mutation of P5. C. AXXG mutation of P6. D. Mutation to the stem of P5 from pairing to non-pairing bases. Red letters indicate bases that were mutated. Red Bold letters indicate the conserved A and G of the AXXG motif.

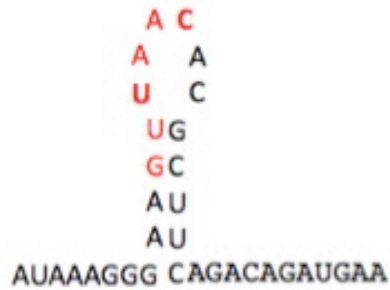
A. Wild type P5 and P5 stem loop sequence.



B. P5 AXXG motif mutation



C. P6 AXXG motif mutation



D. P5 Stem mutation

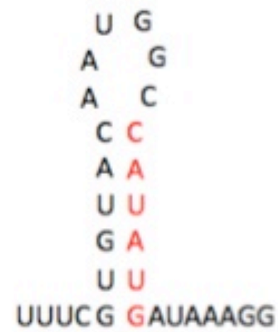


Figure 5.7

F**igure 5.8 qRT-PCR of *eutG* in P5 and P6 mutants.** Expression of *eutG* is not inducible when mutations are made to P5 and P6 AXXG motif. Mutation to the P5 stem is also renders *eutG* uninducible. Error bars indicate standard deviation. Etn = ethanolamine.

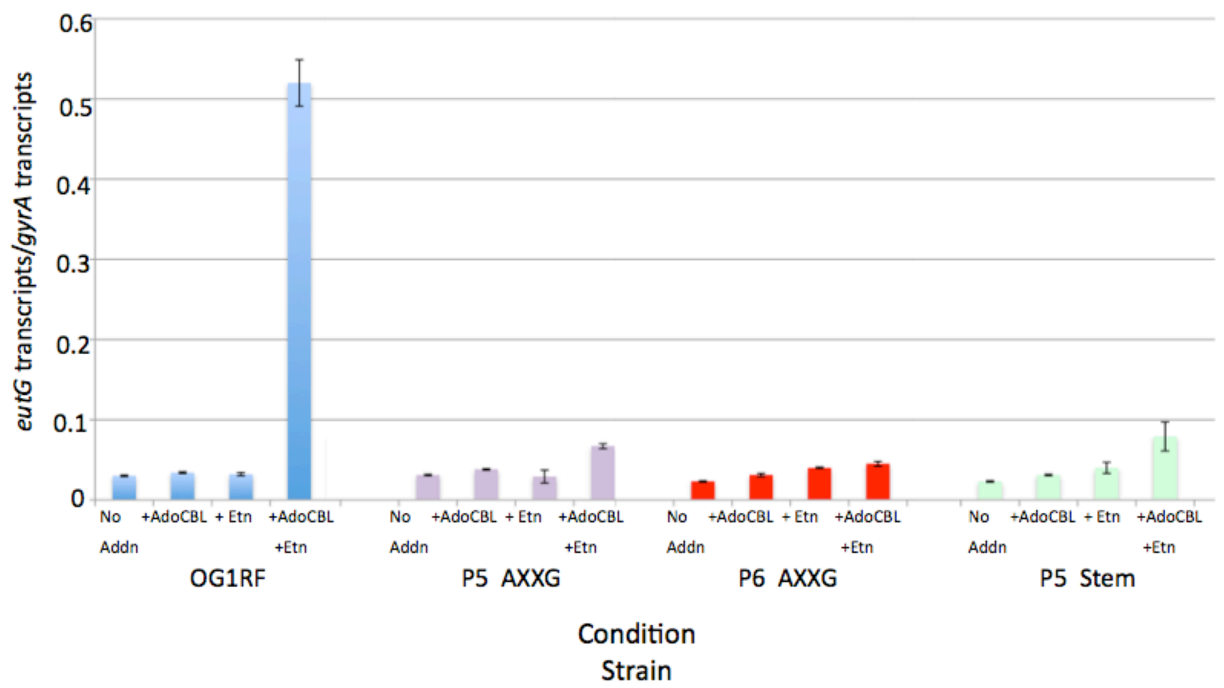


Figure 5.8

My third mutant disrupted pairing interactions within the P5 stem. Similar to my results with the AXXG motif mutants, there was also similar expression levels in non-inducing vs. inducing conditions in the mutant that disrupts pairing interactions within the P5 stem. Therefore the formation of a hairpin structure is likely necessary for proper antitermination complex formation (Figure 5.8).

Mutation of the P3 and P4 region of *eutG* 5'UTR

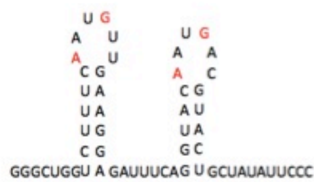
A unique feature of the *eutG* UTR is the presence of a putative second pair of stem loops, P3 and P4. One potential hypothesis is that EutV dimer binding at P3 and P4 promotes binding at P5 and P6 through a cooperative interaction between two EutV dimers. The following section describes mutations I constructed to assess the function of P3 and P4 in *eutG* expression (Figures 4.9 and 4.10).

As shown in Figure 5.9, the first mutation I constructed was a complete deletion of the P3 and P4 regions. The model predicts that in the absence of P3 and P4, EutV will not bind to the region and result in loss of the cooperative interaction that brings EutV to P5 and P6. This will lead to reduced antitermination complex formation at the *eutG* terminator and therefore less *eutG* expression. The second mutant changed the pairing bases to non-pairing bases in both P3 and P4 stems simultaneously. In this situation it is predicted that EutV cannot form a complex with the P3 and P4 stem loops and again would result in reduced expression of *eutG* due to the antiterminator complex not forming. The third mutation substituted the AXXG motif in both P3 and P4 loops simultaneously. P3 was changed from AAUG to UUAA and P4 was changed from AAUG to UUAU. It is

Figure 5.9 Schematic representation of P3 and P4 mutants. A. Wild-type P3 and P4 sequence. B. Deletion of P3 and P4. C. Mutation of P3 and P4 AXXG motifs. D. Mutation of pairing bases in the stem of P3 and P4.

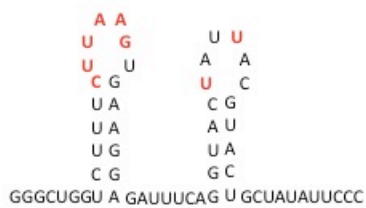
Fig 5.9

A. Wild-type P3 and P4 stem loop sequence.



B. Deletion of P3 and P4 stem loop sequence.

C. Mutation of P3 and P4 AXXG motifs



D. Mutation of P3 and P4 stem interactions

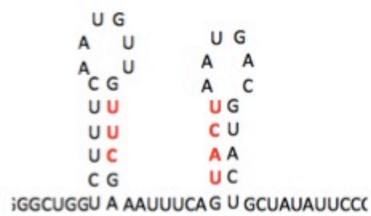


Figure 5.9

Figure 5.10 qRT-PCR of *eutG* in P3 and P4 mutants. In a deletion of P3 and P4 (red) *eutG* is uninducible. *eutG* expression resembled wild-type patterns in AXXG mutants and stem disruption mutants. Error bars indicate standard deviation.

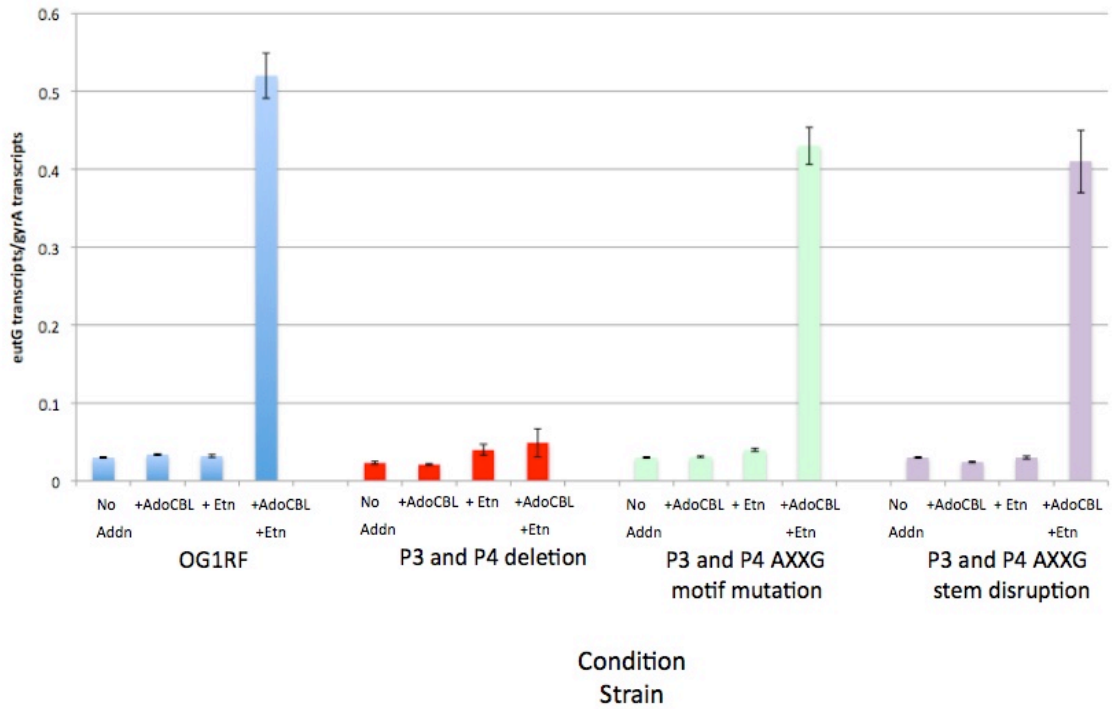


Figure 5.10

predicted that the changes will lead to reduced *eutG* expression because of loss of EutV binding that leads to loss of a cooperative interaction that brings EutV to P5 and P6 the antitermination complex. This is supported by my previous data showing that in the P5 and P6 mutant, changing the AXXG motif led to little expression of *eutG*. I tested these mutants as previously described for *eutG* expression (Figure 5.10). The second P3/P4 region mutation, which disrupts binding interactions within the stem, had an expression profile that resembled wild-type in that low expression was seen in non-inducing conditions and a 20-fold increase in expression occurred in inducing conditions. This result is in contrast to our model of expression, which predicts that *eutG* induced expression would be reduced because of a cooperative interaction.

The third mutation, which changes the ANTAR substrate binding sequences AXXG motifs, also failed to produce a phenotype different from wild-type. In this mutant, *eutG* was expressed at a low level in non-inducing conditions and was expressed about 20-fold higher in inducing conditions. These results suggest that P3 and P4 are not needed for *eutG* expression.

The first mutation deleted the entire region containing P3 and P4. My initial prediction was that *eutG* expression would be reduced due to a loss of a cooperative interaction between EutV at P3 and P4 and EutV at P5 and P6. As a result the antitermination complex would not form as efficiently. As I showed with the stem mutant and the AXXG mutant, mutation of the P3 and P4 did not lead to a loss of expression of *eutG* in inducing conditions. Initially, I predicted that a deletion of P3 and P4 would result in

reduced *eutG* expression due to the loss of a cooperative interaction promoting EutV binding to P5 and P6. In light of the data from the P3 and P4 AXXG mutant and stem mutant, I predict that the deletion will function like wild-type because the P3 and P4 sequence seems to not be required for *eutG* expression.

However, in the P3 and P4 deletion mutant, *eutG* expression in inducing conditions was the same as in non-inducing conditions. *eutG* expression was not induced upon the addition of AdoCBL and ethanolamine. In light of the results of mutant 2 and 3, this result suggests that this part of the RNA is important for function, even if the more detailed mutations do not fit with our initial hypothesis. Alternatively, the deletion could cause a deleterious change in the overall secondary structure of the 5'UTR or affect its stability. Further mutants will need to be generated and tested to understand the role (if any) of this part of the 5'UTR.

Deletion of the *eut* riboswitch

The final mutant I made was to test the modified *eutG* ANTAR hypothesis was a deletion of the riboswitch. The riboswitch incorporates the signal from AdoCBL into the system. We hypothesize that both the EutV interactions and the AdoCBL riboswitch influence the stability of the terminator. In previous experiments, AdoCBL and ethanolamine synergistically activate *eutG* expression. Therefore, integration of both signals is needed for antitermination to occur. I predict that without the riboswitch present, antitermination will not occur properly and induction of *eutG* will not be seen.

Figure 5.11 Schematic of riboswitch deletion. The top panel shows the wild-type locus. The bottom panel shows the *eut* riboswitch deletion locus.

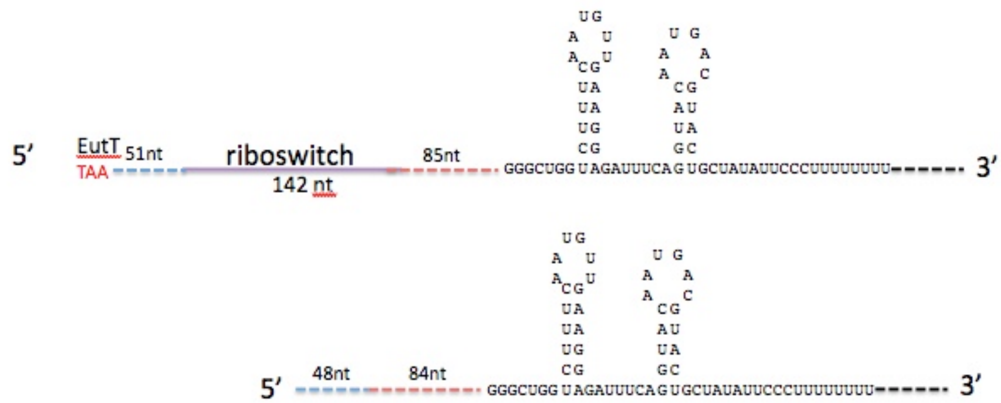


Figure 5.11

F**igure 5.12 qRT-PCR of *eutG* in riboswitch deletion.** The deletion of the riboswitch renders *eutG* uninducible in all conditions tested. Error bars indicate standard deviation.

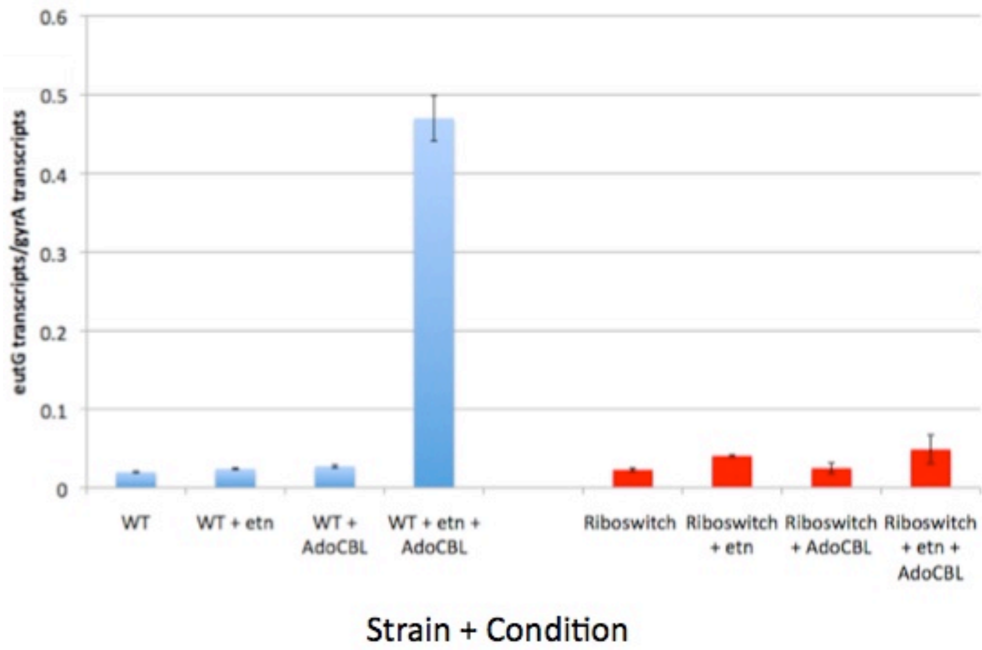


Figure 5.12

The riboswitch deletion construct eliminated 145 bases from the UTR of *eutG* (Figure 5.11). qRT-PCR of *eutG* in this mutant showed that expression of *eutG* in the riboswitch deletion was non-inducible (Figure 5.12). The lack of *eutG* expression in this mutant is most likely due to the fact that the functional riboswitch is needed for antitermination.

Discussion

My model of regulation at *eutG* predicts that a dimer of EutV interacts with P3 and P4, then recruits a second dimer of EutV to P5 and P6 through a cooperative interaction. P5 and P6 interaction with EutV creates the antitermination complex, and the transcriptional terminator is not formed. Also, this antitermination depends on AdoCBL interacting with the riboswitch resulting in antitermination.

My data confirm that the P5 and P6 stem loops and conserved AXXG motifs are necessary for expression of *eutG*. However, mutation of the putative P3 and P4 loops have no effect on *eutG* expression and suggest that these predicted structures are not valid. But importantly, complete elimination of this region of RNA prevents induction, suggesting that this part of the RNA is doing something that is important for the overall functionality of this 5' UTR.

The first experiments in this study examined the function of a transcriptional terminator in the 5'UTR of *eutG*. These experiments showed that mutation and deletion of this terminator led to expression of *eutG* in normally non-inducing conditions. However, a slight induction was still present. One potential explanation is that multiple promoters

are responsible for the total expression of *eutG* was measured. A promoter was mapped upstream of *eutP* that may transcribe all the way down to the *eutG* coding region and beyond. A second promoter was mapped to the *eutG* 5'UTR. In non-inducing conditions the upstream terminator will remain intact and function normally. Therefore transcription was terminated upstream of *eutP* and little to no transcription proceeded to *eutG*. However, transcription starting from a second promoter that is normally terminated at *eutG* was not terminated in the mutant, leading to the expression seen in these experiments. This led to the interesting hypothesis that expression seen from *eutG* in inducing conditions is the additive effect of transcription from two promoters.

My experimental results indicate that the P5 and P6 putative stem loop are required for expression of *eutG* while P4 and P3 are not required for expression and may not exist as predicted. In accordance with our initial model of ANTAR regulation, a dimer of EutV potentially binds to the small stem loops formed by P5 and P6 at the AXXG motif. This action would theoretically lock in an antiterminator structure preventing the formation of the transcriptional terminator in the nascent RNA. In opposition to the model, these data showed that this interaction is not dependent on, or augmented by, EutV binding to the possible P3 and P4 loops. The experiments on P5 and P6 are the second example of functional studies on the *eut* ANTAR substrates and support Dr. DebRoy's work on the P1 and P2 stem loops of *eutP*. Overall, my data provide support for our model on how ANTARs promote antitermination as described in Figure 5.3.

My experiments with directed mutation of P3 and P4 oppose the idea that the P3 and P4 stem loops lead to a cooperative interaction between a second dimer of EutV at P5 and P6. This indicates that P5 and P6 are independent of the putative P3 and P4 loops. The data from the mutant containing the deletion of P3 and P4 showed that the sequence was important for the expression of *eutG* but probably not in the manner we originally predicted. The overall RNA secondary structure of the UTR may be important for proper function. A large deletion, such as the one I generated to cover the putative P3 and P4 stem loops, could also prevent another important interaction from occurring because the RNA is not the proper length.

My final experiment showed that in a riboswitch deletion strain expression of *eutG* was not inducible. This is in agreement with the model that both the riboswitch and the two-component system control the stability of the terminator. The two structures are both required to cause antitermination and without either structure, the system in not inducible.

These data showed that *eutG* was unresponsive to AdoCBL in a mutant with deletion of the riboswitch and suggest that the riboswitch is likely the only input for AdoCBL sensing in the system. *C. difficile* contains a *eut* locus. However, an AdoCBL riboswitch is not found in this organism. Therefore an additional mechanism may be in place in these organisms to sense AdoCBL. If another input for AdoCBL existed, expression of *eut* locus genes may still be responsive to AdoCBL in the absence of a riboswitch. These data suggest that in *E. faecalis* this is not the case, and AdoCBL riboswitch is the only sensor of AdoCBL feeding into the *eut* locus.

Most riboswitches are found alone, controlling their downstream genes through the sensing of one signal (17). There have been identified cases of tandem interactions of two riboswitches in the same UTR, allowing multiple signal integration (76). The *eut* locus provides the first example of a metabolite binding riboswitch interacting with an RNA binding regulatory protein. This is a novel mechanism by which two different signals are integrated to control expression.

Chapter 6

Discussion and future directions

My studies showed a complex regulatory mechanism governs expression of the *eut* locus genes in *E. faecalis*. By bioinformatically analyzing the *eut* locus gene sequences and comparing them to other species, I discovered the presence of unique regulators in *E. faecalis*. These regulators are an AdoCBL riboswitch and a two-component system that control expression through an antitermination mechanism. My bioinformatic analysis led me to the hypothesis that the two-component system is activated by ethanolamine resulting in the response regulator interacting with the nascent RNA to cause antitermination and allowing *eut* gene expression. Testing this hypothesis, I discovered that increased expression of four *eut* locus genes, *eutP*, *eutG*, *eutS*, and *eutA* was dependent upon AdoCBL, ethanolamine, and EutVW. Surprisingly, the total transcript levels of each of the genes varied significantly compared to one another, suggesting a previously unknown regulatory mechanism to control *eut* gene expression. My 5' RACE analysis showed that putative transcriptional start site exists upstream of *eutG*, *eutS,* and *eutP* and could account for the differences in expression.

These results led to a model to explain the mechanism by which the AdoCBL riboswitch and EutV cause antitermination of *eutG*. In this model, a dimer of EutV interacts with P3 and P4. Upon this interaction, EutV recruits a second dimer of EutV to P5 and P6 through a cooperative interaction. Interaction of P5 and P6 with EutV creates the antitermination complex, and the transcriptional terminator is not formed. In addition, our model predicts that antitermination also depends upon AdoCBL interaction with the riboswitch. In testing this model, I determined that P5 and P6 stem loops played a necessary role in *eutG* antitermination. My data also showed that formation of P3 and P4

stem loops were not necessary for antitermination. However, deletion of the sequence rendered *eutG* non-inducible. Finally, I showed that the *eut* riboswitch was necessary for antitermination of *eutG*. These results revealed a novel antitermination mechanism in which *eutG* terminator stability is controlled by the riboswitch and the two-component system acting through the P5 and P6 stem loops. All together, my data uncovered a novel regulatory mechanism that regulates *eut* locus expression and ethanolamine metabolism.

Identification of two posttranscriptional regulatory features by bioinformatics

The results of my bioinformatic study uncovered two potential new regulators of *eut* locus gene expression, the *eut* AdoCBL riboswitch and the EutVW two-component system. In addition, my analysis identified a novel class of AdoCBL riboswitches. Secondly, I discovered that *eut* gene regulation in *E. faecalis* employs a two-component system with an RNA-binding response regulator. The RNA binding domain belongs to the ANTAR family of proteins, which function as antiterminators. Together, the potential concerted regulation by a two-component system and a riboswitch has never been described in the literature. My additional bioinformatic studies identified a conserved sequence in the 5'UTR of many genes that are substrates of ANTAR regulators. Further studies predicting the structure of this RNA found that two hairpins may form in this region and may potentially interact with dimerized EutV. This sequence is the predicted ANTAR substrate domain.

Future work expanding on this bioinformatic analysis to identify new ANTAR regulatory proteins and ANTAR RNA substrate domains in other microbes would be important to determine the degree that this new mechanism of regulation is conserved. A

bioinformatic search for ANTAR domain proteins was carried out previously (73). However, this study was performed in 2002 and many new bacterial genome sequences have been published since. Therefore, an updated search may reveal new examples of proteins with ANTAR domains.

Differential expression of *eut* locus genes

Using data from my bioinformatic analysis, I hypothesized that *eut* locus genes were positively regulated by ethanolamine and AdoCBL. Using qRT-PCR to test the expression of several *eut* locus genes, I confirmed this hypothesis. I showed that in the presence of both of these compounds expression of *eutP, eutG, eutS,* and *eutA* are increased. Interestingly, these data also showed differential total transcript levels for *eutP, eutG, eutS*, and *eutA*. The total expression levels of the genes tested varied by several fold. The 5'UTR of each of these genes contained a putative promoter followed by a transcriptional terminator. We now have evidence that supports the presence of active promoters in front of *eutP, eutG* and *eutS,* but not *eutA.* The presence of these three promoters combined with the variation in transcript levels in each gene suggests that a number of transcripts might be produced within the *eut* locus.

It is possible that differences in the initiation of transcription and the antitermination efficiency will vary at each promoter/terminator pair. There are sequence differences within the promoters. Interestingly, the *eutS* promoter contains a match to the consensus Gram-positive promoter. The closer a promoter sequence is to the consensus, the more efficiently it operates (12). This may explain why *eutS* transcripts levels are much higher

than the other *eut* locus genes tested. It will be interesting in the future to determine the number of transcripts present in the *eut* locus as well as the efficiency of transcription initiation from each *eut* locus promoter. To examine if the difference in trancript levels for each gene is due to promoter strength, a LacZ fusion could be made to each promoter individually. The LacZ fusions should be made to include just the promoter region, not any of the other potential regulatory features in the leader RNA. Then activity of each promoter could be examined using β-galactosidase assays. Any difference in expression in these assays could then be attributed to the strength of the promoter.

Efficiency of antitermination may also explain the differences in total transcript levels between *eut* locus genes. In future studies, it would be interesting to determine the changes in efficiency of antitermination in order to understand the transcription differences. Using an *in vitro* approach would allow the affinity of the interaction between EutV and the substrate RNA to be determined. In addition, such an *in vitro* approach would also lend insight into the rate of antitermination.

Potential positive feedback loop governs expression of *eut* regulators

An interesting finding in my studies of *eut* gene expression was that the basal level of *eutS* expression was substantially higher compared to the other genes examined. In addition, total *eutS* transcript levels were highest in induced conditions and this induction was dependent on EutVW. *eutS* is hypothesized to be transcribed on the same transcript as the two-component system genes *eutV* and *eutW*. Since EutVW was required to induce expression of the transcript containing *eutS, eutV, and eutW*, this suggests that

EutVW regulates its own expression through a positive feedback loop. In addition, AdoCBL was also required for the increased *eutS* expression suggesting it also feeds into the loop regulating expression of *eutVW*. A potential model for this feedback loop is as follows. First, ethanolamine causes activation of the EutV through EutW. Both activated EutV and the presence AdoCBL result in antitermination of *eutG* and *eutS*. This allows transcription to continue through the *eutG* coding region and into the *eutS, V, and W* coding regions, thus increasing the net concentration of EutVW. This leads to increased levels of two-component system proteins, which can then be activated, leading to more transcription of *eutV* and *eutW*. In this model, transcription starts upstream of *eutG* explaining why AdoCBL was necessary for positive feedback of EutVW to occur. Interestingly, EutR, the regulator of *eut* locus expression in *S. typhimurium*, is also hypothesized to be regulated by a positive feedback loop (67). The caveat to this model is that *eutV* and *eutW* co-transcription with *eutS* has not been shown experimentally. Examining expression of *eutV* and *eutW* individually will be important test this model. Teasing out the molecular mechanisms necessary for this positive feedback loop will be an interesting future direction.

Model of *eutG* regulation: the role of the P loops and the riboswitch

I initially proposed a model in which antitermination of *eutG* occurs through two stem loop pairs, P3/P4 and P5/P6, and the AdoCBL riboswitch. In this model, I hypothesized that a cooperative interaction occurs in which binding of a EutV dimer to P3 and P4 promotes EutV binding to P5 and P6. However, my data do not support this model. Mutations in the P5 and P6 conserved AXXG motifs abrogated induction of *eutG*

suggesting they are important for antitermination. In contrast, P3 and P4 mutations had wild-type levels of *eutG* expression suggesting that formation of these structures is not required for antitermination and cooperativity does not occur. Interestingly, the sequence that contains the predicted P3 and P4 stem loop structure is important for functional antitermination, as *eutG* was non-inducible in a mutant with a deletion of this entire region. This result suggests that this sequence is necessary for antitermination to occur, but not through the cooperative interaction of P3 and P4 with EutVW. Potentially, a secondary structure change necessary for antitermination is dependent on the P3 and P4 sequence. A large deletion of the sequence, such as the P3/P4 deletion, may prevent this interaction from occurring. In such a deletion, the RNA is not the proper length. This may result in changes to the RNA secondary structure and in a lack of antitermination. However, mutation of the P3 and P4 stem loop bases may not interfere in RNA secondary structure changes and antitermination persists. Therefore, further studies are still needed to understand how this region of the leader RNA contributes to antitermination.

Expression of *eutG* was partially dependent on the presence of the AdoCBL. The AdoCBL signal is integrated into the system via the riboswitch. I found that in a deletion of the riboswitch, there was no induction of expression of *eutG* indicating that the riboswitch is necessary for antitermination of *eutG*. In addition, *eutG* expression was also dependent on EutV antitermination. These results suggest that binding of AdoCBL to the riboswitch, together with EutV antitermination, is required for *eutG* expression. Binding of AdoCBL to the riboswitch may modulate the secondary structure allowing EutV to bind and cause antitermination.

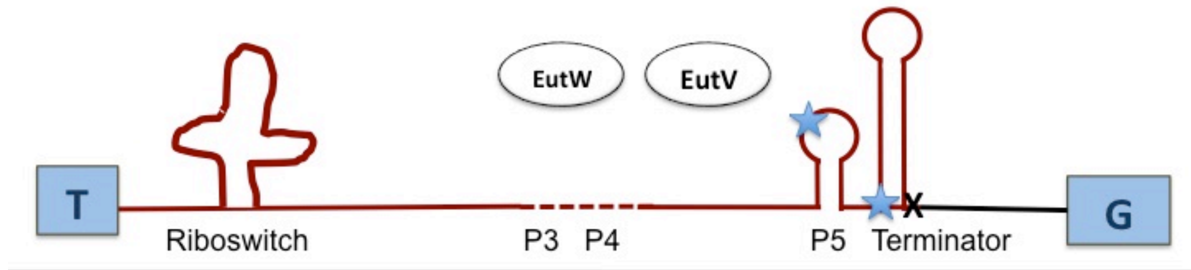
A potential overall model to describe the regulation of *eutG* is as follows. In non-inducing conditions, the transcriptional terminator is formed and *eutG* expression is attenuated (Figure 6, top panel). In the presence of AdoCBL and ethanolamine, AdoCBL binds to riboswitch RNA. This causes a conformational change in the riboswitch mRNA. This conformational change may include an interaction with the top bases in transcriptional terminator that promotes antitermination. In addition, activated EutV as a dimer interacts with P5 and P6 to form the antitermination structure. The combination of secondary structure changes in the mRNA caused by AdoCBL binding to the riboswitch and P5 and P6 stem loop interaction with activated EutV are needed for antitermination and increased expression of *eutG* (Figure 6, bottom panel). The hypothesized interaction between the riboswitch RNA and the transcriptional terminator is supported by my experiment showing that in stem mutant 2, disruption of the top of the stem loop led to increased *eutG* expression with AdoCBL only. Another piece of data supporting a potential long-range interaction is that deletion of P3 and P4 led to loss of expression. The P3 and P4 sequence may be necessary for proper folding of the mRNA to facilitate an interaction between the riboswitch and the terminator.

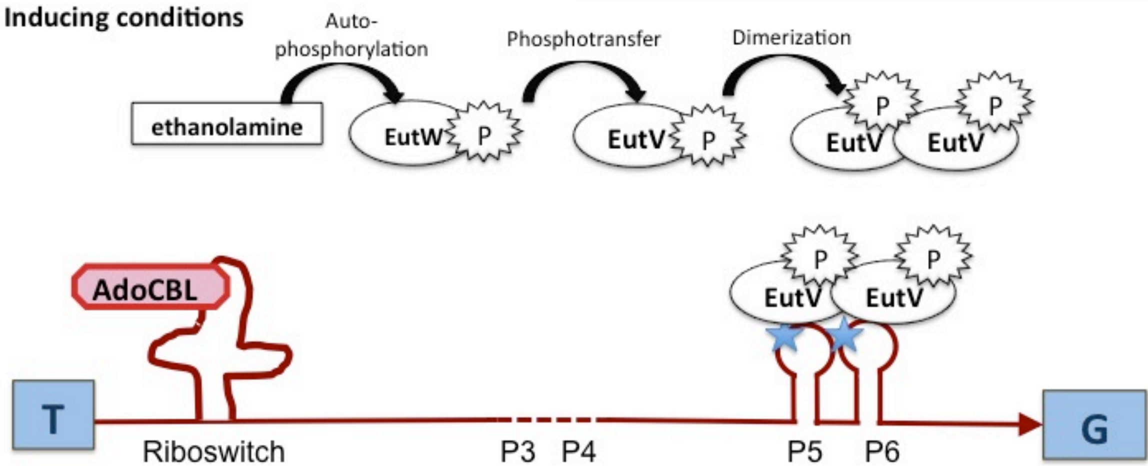
mRNA stability may play a role in *eut* locus expression

Control of mRNA stability by modulating susceptibility to RNA degrading enzymes may provide an alternate explanation for how the *eut* riboswitch controls the mRNA levels of downstream genes. The S-adenosyl-methionine (SAM) riboswitch in *Bacillus subtilis* was shown to control mRNA stability by altering binding affinity of an endoribonuclease (71). The SAM riboswitch controls expression using a termination/antitermination

Figure 6. Model of *eutG* regulation. Top Panel: Non-inducing conditions- No interaction of AdoCBL or EutV with the nascent transcript and terminator formation occurs. Bottom panel: Inducing conditions - The interaction of AdoCBL with the riboswitch and a EutV-P dimer with stem loop pair P5 and P6. Both AdoCBL binding and EutV-P binding are necessary for antitermination to occurs and transcription to proceed through *eutG*. Dashed line indicates P3 and P4 sequence which is important for function but it is unknown if stem loops are formed.

Non-inducing conditions





regulation strategy. In low concentrations of SAM, the predominant mRNA species is the antiterminated form. The authors showed that the antiterminated form of mRNA was

a poor substrate for the endonuclease, RNase Y. As a result, transcripts remained stable. In high concentrations of SAM, the predominant mRNA species was the terminated form. This RNA was subject to cleavage by RNase Y and degradation by the exonuclease RNaseJ 1(71). It will be important in the future to determine if the *eut* locus riboswitch functions via a mechanism for mRNA stability control. *E. faecalis* contains an orthologue of RNase Y (18). Examining *in vitro* susceptibility of both terminated and non-terminated transcripts to cleavage by the RNase Y would determine if the riboswitch regulates *eut* locus transcript stability in this manner.

Alternative explanation for P3 and P4 mutagenesis experiments

Some of my P stem loop mutagenesis results seem to argue against my predicted model. Deletion of P3 and P4 led to loss of induction phenotype; while the P3 and P4 stem loop mutants and P3 and P4 AXXG mutants had a wild-type phenotype. The P3 and P4 AXXG mutants contain mutations of most of the bases that are deleted in the P3 and P4 deletion mutants. Therefore, I expected both the deletion and point mutants to have a similar phenotype. One explanation for these data is that the length of the sequence is important for folding and interaction, which would be shortened in the deletion mutant but not in the point mutants. Alternatively, mutagenesis of the P3 and P4 stem loop may have led to additional mutations that were outside of the sequenced region. Each strain was sequenced only through the mutated sequence and the junctions with the native sequence. One way to address this possibility is through complementation analysis by re-introducing the wild-type sequence into the strain via the same counter selection process used to create the strain. Additionally, repeating the P loop experiments in several

independent isolates of each mutant would determine if the phenotype I observed is specific to a single isolate.

Control of *eut* expression by novel posttranscriptional regulators

It is well known that bacteria are able to integrate multiple signals to control gene expression. Transcriptional regulation of genes can be controlled by multiple inputs by several DNA-binding transcription factors. Riboswitches are also known to integrate multiple signals as some riboswitches have been found in tandem controlling expression of the same locus (4). My results demonstrate a novel mechanism for integrating multiple signals to control expression of a single locus. I discovered that regulation of expression of the *eut* locus required integration of signals by two distinct posttranscriptional regulators. This is the first example of a riboswitch and an antitermination factor protein together integrating multiple environmental signals at a single locus to control expression of genes.

Bibliography

1. **Aarestrup FM, B. P., Witte W** 2002. p. 55–99. *In* C. D. Gilmore MS , Courvalin P , Dunny GM , Murray BE , Rice LB (ed.), The Enterococci: Pathogenesis, Molecular Biology, and Antibiotic Resistance. Am Soc Microbiol Press, Washington, DC.
2. **Abreu-Goodger, C., and E. Merino.** 2005. RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. Nucleic Acids Res **33:**W690-2.
3. **Albright, F. R., D. A. White, and W. J. Lennarz.** 1973. Studies on enzymes involved in the catabolism of phospholipids in *Escherichia coli*. J Biol Chem **248:**3968-77.
4. **Barrick, J., and R. Breaker.** 2007. The distributions, mechanisms, and structures of metabolite-binding riboswitches. Genome Biology **8:**R239.
5. **Bizzini, A., C. Zhao, A. Budin-Verneuil, N. Sauvageot, J.-C. Giard, Y. Auffray, and A. Hartke.** Glycerol Is Metabolized in a Complex and Strain-Dependent Manner in *Enterococcus faecalis*. J. Bacteriol. **192:**779-785.
6. **Blackwell, C. M., F. A. Scarlett, and J. M. Turner.** 1976. Ethanolamine catabolism by bacteria, including *Escherichia coli*. Biochem Soc Trans **4:**495-7.
7. **Blackwell, C. M., and J. M. Turner.** 1978. Microbial metabolism of amino alcohols. Formation of coenzyme B12-dependent ethanolamine ammonia-lyase and its concerted induction in *Escherichia coli*. Biochem J **176:**751-7.

8. **Bourgogne, A., D. A. Garsin, X. Qin, K. V. Singh, J. Sillanpaa, S. Yerrapragada, Y. Ding, S. Dugan-Rocha, C. Buhay, H. Shen, G. Chen, G. Williams, D. Muzny, A. Maadani, K. A. Fox, J. Gioia, L. Chen, Y. Shang, C. A. Arias, S. R. Nallapareddy, M. Zhao, V. P. Prakash, S. Chowdhury, H. Jiang, R. A. Gibbs, B. E. Murray, S. K. Highlander, and G. M. Weinstock.** 2008. Large scale variation in *Enterococcus faecalis* illustrated by the genome analysis of strain OG1RF. Genome Biol **9:**R110.

9. **Bourgogne, A., S. G. Hilsenbeck, G. M. Dunny, and B. E. Murray.** 2006. Comparison of OG1RF and an isogenic fsrB deletion mutant by transcriptional analysis: the Fsr system of *Enterococcus faecalis* is more than the activator of gelatinase and serine protease. J Bacteriol **188:**2875-84.

10. **Brinsmade, S. R., and J. C. Escalante-Semerena.** 2004. The eutD gene of *Salmonella enterica* encodes a protein with phosphotransacetylase enzyme activity.Journal of Bacteriology **186:**1890-2.

11. **Brinsmade, S. R., T. Paldon, and J. C. Escalante-Semerena.** 2005. Minimal functions and physiological conditions required for growth of salmonella enterica on ethanolamine in the absence of the metabolosome. Journal of Bacteriology **187:**8039-46.

12. **Browning, D. F., and S. J. Busby.** 2004. The regulation of bacterial transcription initiation. Nat Rev Microbiol **2:**57-65.

13. **Buan, N. R., S.-J. Suh, and J. C. Escalante-Semerena.** 2004. The eutT gene of *Salmonella enterica* Encodes an oxygen-labile, metal-containing ATP:corrinoid adenosyltransferase enzyme. Journal of Bacteriology **186:**5708-14.

14. **Cannon, G. C., C. E. Bradburne, H. C. Aldrich, S. H. Baker, S. Heinhorst, and J. M. Shively.** 2001. Microcompartments in Prokaryotes: Carboxysomes and Related Polyhedra. Appl. Environ. Microbiol. **67:**5351-5361.

15. **Chai, W., and V. Stewart.** 1998. NasR, a novel RNA-binding protein, mediates nitrate-responsive transcription antitermination of the *Klebsiella oxytoca* M5al nasF operon leader in vitro. J Mol Biol **283:**339-51.

16. **Chai, W., and V. Stewart.** 1999. RNA sequence requirements for NasR-mediated, nitrate-responsive transcription antitermination of the *Klebsiella oxytoca* M5al nasF operon leader. J Mol Biol **292:**203-16.

17. **Dambach, M. D., and W. C. Winkler.** 2009. Expanding roles for metabolite-sensing regulatory RNAs. Curr Opin Microbiol **12:**161-9.

18. **Davidsen, T., E. Beck, A. Ganapathy, R. Montgomery, N. Zafar, Q. Yang, R. Madupu, P. Goetz, K. Galinsky, O. White, and G. Sutton.** The comprehensive microbial resource. Nucleic Acids Research **38:**D340-D345.

19. **Del Papa, M. F., and M. Perego.** 2008. Ethanolamine Activates a Sensor Histidine Kinase Regulating Its Utilization in *Enterococcus faecalis*. J. Bacteriol. **190:**7147-7156.

20. **Deutscher, J., C. Francke, and P. W. Postma.** 2006. How Phosphotransferase System-Related Protein Phosphorylation Regulates Carbohydrate Metabolism in Bacteria. Microbiol. Mol. Biol. Rev. **70:**939-1031.

21. **Drews, G., and W. Niklowitz.** 1956. Beiträge zur Cytologie der Blaualgen. Archives of Microbiology **24:**147-162.

22. **Edwards, T. E., D. J. Klein, and A. R. Ferre-D'Amare.** 2007. Riboswitches: small-molecule recognition by gene regulatory RNAs. Curr Opin Struct Biol **17:**273-9.

23. **Felmingham, D., A. P. R. Wilson, A. I. Quintana, and R. N. Grüneberg.** 1992. Enterococcus Species in Urinary Tract Infection. Clinical Infectious Diseases **15:**295-301.

24. **Fouquier d'Herouel, A., F. Wessner, D. Halpern, J. Ly-Vu, S. P. Kennedy, P. Serror, E. Aurell, and F. Repoila.** A simple and efficient method to search for selected primary transcripts: non-coding and antisense RNAs in the human pathogen *Enterococcus faecalis.* Nucleic Acids Res **39:**e46.

25. **Fox, K. A., A. Ramesh, J. E. Stearns, A. Bourgogne, A. Reyes-Jara, W. C. Winkler, and D. A. Garsin.** 2009. Multiple posttranscriptional regulatory mechanisms partner to control ethanolamine utilization in *Enterococcus faecalis*. Proc Natl Acad Sci USA **106:**4435-40.

26. **Gao, R., and A. M. Stock.** 2009. Biological insights from structures of two-component proteins. Annu Rev Microbiol **63:**133-54.

27. **Garsin, D. A.** Ethanolamine utilization in bacterial pathogens: roles and regulation. Nat Rev Micro **8:**290-295.

28. **Garsin, D. A., C. D. Sifri, E. Mylonakis, X. Qin, K. V. Singh, B. E. Murray, S. B. Calderwood, and F. M. Ausubel.** 2001. A simple model host for identifying Gram-positive virulence factors. Proc Natl Acad Sci USA **98:**10892-7.

29. **Garsin, D. A., J. Urbach, J. C. Huguet-Tapia, J. E. Peters, and F. M. Ausubel.** 2004. Construction of an *Enterococcus faecalis* Tn917-mediated-gene-

disruption library offers insight into Tn917 insertion patterns. Journal of Bacteriology **186:**7280-9.

30. **Graninger, W., and R. Ragette.** 1992. Nosocomial Bacteremia Due to *Enterococcus faecalis* without Endocarditis. Clinical Infectious Diseases **15:**49-57.

31. **Grundy, F. J., and T. M. Henkin.** 2006. From Ribosome to Riboswitch: Control of Gene Expression in Bacteria by RNA Structural Rearrangements. Critical Reviews in Biochemistry & Molecular Biology **41:**329-338.

32. **Hancock, L., and M. Perego.** 2002. Two-Component Signal Transduction in *Enterococcus faecalis.* J. Bacteriol. **184:**5819-5825.

33. **Hancock, L. E., and M. Perego.** 2004. Systematic inactivation and phenotypic characterization of two-component signal transduction systems of *Enterococcus faecalis* V583. Journal of Bacteriology **186:**7951-8.

34. **Heid, C. A., J. Stevens, K. J. Livak, and P. M. Williams.** 1996. Real time quantitative PCR. Genome Research **6:**986-994.

35. **Hidron, Alicia I., Jonathan R. Edwards, J. Patel, Teresa C. Horan, Dawn M. Sievert, Daniel A. Pollock, and Scott K. Fridkin.** 2008. NHSN Annual Update: Antimicrobial‐Resistant Pathogens Associated With Healthcare‐Associated Infections: Annual Summary of Data Reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2006–2007 • Infection Control and Hospital Epidemiology **29:**996-1011.

36. **Johnson, M., I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, and T. L. Madden.** 2008. NCBI BLAST: a better web interface. Nucleic Acids Research **36:**W5-W9.

37. **Jones, P. W., and J. M. Turner.** 1984. Interrelationships Between the Enzymes of Ethanolamine Metabolism in *Escherichia coli*. J Gen Microbiol **130:**299-308.

38. **Joseph, B., K. Przybilla, C. Stühler, K. Schauer, J. Slaghuis, T. M. Fuchs, and W. Goebel.** 2006. Identification of *Listeria monocytogenes* genes contributing to intracellular replication by expression profiling and mutant screening. Journal of Bacteriology **188:**556-68.

39. **Kampf, G., and A. Kramer.** 2004. Epidemiologic background of hand hygiene and evaluation of the most important agents for scrubs and rubs. Clinical Microbiology Reviews **17:**863 - 893.

40. **Kaye, D.** 1982. Enterococci: Biologic and Epidemiologic Characteristics and In Vitro Susceptibility. Arch Intern Med **142:**2006-2009.

41. **Kelly, A., M. D. Goldberg, R. K. Carroll, V. Danino, J. C. Hinton, and C. J. Dorman.** 2004. A global role for Fis in the transcriptional control of metabolism and type III secretion in *Salmonella enterica* serovar Typhimurium. Microbiology **150:**2037-53.

42. **Kerfeld, C. A., S. Heinhorst, and G. C. Cannon.** Bacterial microcompartments. Annu Rev Microbiol **64:**391-408.

43. **KOFOID, E., C. RAPPLEYE, I. STOJILJKOVIC, and J. ROTH.** 1999. The 17-Gene Ethanolamine (eut) Operon of *Salmonella typhimurium* Encodes Five

Homologues of Carboxysome Shell Proteins. Journal of Bacteriology **181:**5317-5329.

44. **Kristich, C. J., J. R. Chandler, and G. M. Dunny.** 2007. Development of a host-genotype-independent counterselectable marker and a high-frequency conjugative delivery system and their use in genetic analysis of *Enterococcus faecalis*. Plasmid **57:**131-44.

45. **Kristich, C. J., C. L. Wells, and G. M. Dunny.** 2007. A eukaryotic-type Ser/Thr kinase in *Enterococcus faecalis* mediates antimicrobial resistance and intestinal persistence. Proceedings of the National Academy of Sciences **104:**3508-3513.

46. **Larson, T. J., M. Ehrmann, and W. Boos.** 1983. Periplasmic glycerophosphodiester phosphodiesterase of *Escherichia coli,* a new enzyme of the glp regulon. Journal of Biological Chemistry **258:**5428-5432.

47. **Laub, M. T., and M. Goulian.** 2007. Specificity in Two-Component Signal Transduction Pathways. Annual Review of Genetics **41:**121-145.

48. **Lawhon, S. D., J. G. Frye, M. Suyemoto, S. Porwollik, M. McClelland, and C. Altier.** 2003. Global regulation by CsrA in *Salmonella typhimurium*. Mol Microbiol **48:**1633-45.

49. **Maadani, A., K. A. Fox, E. Mylonakis, and D. A. Garsin.** 2007. *Enterococcus faecalis* mutations affecting virulence in the Caenorhabditis elegans model host. Infect Immun **75:**2634-7.

50. **Marchler-Bauer, A., S. Lu, J. B. Anderson, F. Chitsaz, M. K. Derbyshire, C. DeWeese-Scott, J. H. Fong, L. Y. Geer, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, J. D. Jackson, Z. Ke, C. J. Lanczycki, F. Lu, G. H.

Marchler, M. Mullokandov, M. V. Omelchenko, C. L. Robertson, J. S. Song, N. Thanki, R. A. Yamashita, D. Zhang, N. Zhang, C. Zheng, and S. H. Bryant. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Research **39:**D225-D229.

51. **Megran, D. W.** 1992. Enterococcal Endocarditis. Clinical Infectious Diseases **15:**63-71.

52. **Mehmeti, I., M. Jonsson, E. M. Fergestad, G. Mathiesen, I. F. Nes, and H. Holo.** Transcriptome, proteome and metabolite analysis of a lactate dehydrogenase negative mutant of *Enterococcus faecalis* V583. Appl. Environ. Microbiol.**:**AEM.02485-10.

53. **Montange, R. K., and R. T. Batey.** 2008. Riboswitches: emerging themes in RNA structure and function. Annual review of biophysics **37:**117-33.

54. **Mori, K., R. Bando, N. Hieda, and T. Toraya.** 2004. Identification of a reactivating factor for adenosylcobalamin-dependent ethanolamine ammonia lyase. journal of bacteriology **186:**6845-54.

55. **Munch, A., L. Stingl, K. Jung, and R. Heermann.** 2008. Photorhabdus luminescens genes induced upon insect infection. BMC Genomics **9:**229.

56. **Murray, B. E.** 1990. The life and times of the Enterococcus. Clin. Microbiol. Rev. **3:**46-65.

57. **Murray, B. E., K. V. Singh, R. P. Ross, J. D. Heath, G. M. Dunny, and G. M. Weinstock.** 1993. Generation of restriction map of *Enterococcus faecalis* OG1 and investigation of growth requirements and regions encoding biosynthetic function. Journal of Bacteriology **175:**5216-23.

58. **Nahvi, A., J. E. Barrick, and R. R. Breaker.** 2004. Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. Nucleic Acids Res **32:**143-50.

59. **Nahvi, A., N. Sudarsan, M. S. Ebert, X. Zou, K. L. Brown, and R. R. Breaker.** 2002. Genetic control by a metabolite binding mRNA. Chem Biol **9:**1043.

60. **Norman, R. A., C. L. Poh, L. H. Pearl, B. P. O'Hara, and R. E. Drew.** 2000. Steric hindrance regulation of the *Pseudomonas aeruginosa* amidase operon. J Biol Chem **275:**30660-7.

61. **Paulsen, I. T., L. Banerjei, G. S. A. Myers, K. E. Nelson, R. Seshadri, T. D. Read, D. E. Fouts, J. A. Eisen, S. R. Gill, J. F. Heidelberg, H. Tettelin, R. J. Dodson, L. Umayam, L. Brinkac, M. Beanan, S. Daugherty, R. T. DeBoy, S. Durkin, J. Kolonay, R. Madupu, W. Nelson, J. Vamathevan, B. Tran, J. Upton, T. Hansen, J. Shetty, H. Khouri, T. Utterback, D. Radune, K. A. Ketchum, B. A. Dougherty, and C. M. Fraser.** 2003. Role of Mobile DNA in the Evolution of Vancomycin-Resistant *Enterococcus faecalis*. Science **299:**2071-2074.

62. **Penrod, J. T., C. C. Mace, and J. R. Roth.** 2004. A pH-sensitive function and phenotype: evidence that EutH facilitates diffusion of uncharged ethanolamine in *Salmonella enterica*. Journal of Bacteriology **186:**6885-90.

63. **Pultz, N. J., L. C. Hoskins, and C. J. Donskey.** 2006. Vancomycin-Resistant Enterococci May Obtain Nutritional Support by Scavenging Carbohydrate

Fragments Generated During Mucin Degradation by the Anaerobic Microbiota of the Colon. Microbial Drug Resistance **12:**63-67.

64. **Randle CL, A. P., Dittmer JC.** 1969. The phosphoglyceride composition of Gram-negative bacteria and the changes in composition during growth. Biochim Biophys Acta **187:**214-20.

65. **Ravnum, S., and D. I. Andersson.** 2001. An adenosyl–cobalamin (coenzyme-B12)-repressed translational enhancer in the cob mRNA of *Salmonella typhimurium*. Molecular Microbiology **39:**1585-1594.

66. **Rodionov, D. A., A. G. Vitreschak, A. A. Mironov, and M. S. Gelfand.** 2003. Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. J Biol Chem **278:**41148-59.

67. **Roof, D. M., and J. R. Roth.** 1992. Autogenous regulation of ethanolamine utilization by a transcriptional activator of the eut operon in *Salmonella typhimurium*. Journal of Bacteriology **174:**6634-43.

68. **Roof, D. M., and J. R. Roth.** 1988. Ethanolamine utilization in *Salmonella typhimurium*. Journal of Bacteriology **170:**3855-63.

69. **Roof, D. M., and J. R. Roth.** 1989. Functions required for vitamin B12-dependent ethanolamine utilization in *Salmonella typhimurium*. Journal of Bacteriology **171:**3316-23.

70. **Sghir, A., G. Gramet, A. Suau, V. Rochet, P. Pochart, and J. Dore.** 2000. Quantification of Bacterial Groups within Human Fecal Flora by Oligonucleotide Probe Hybridization. Appl. Environ. Microbiol. **66:**2263-2266.

71. **Shahbabian, K., A. Jamalli, L. Zig, and H. Putzer.** 2009. RNase Y, a novel endoribonuclease, initiates riboswitch turnover in Bacillus subtilis. EMBO J **28:**3523-3533.

72. **Sherman, J. M., J. C. Mauer, and P. Stark.** 1937. *Streptococcus fecalis*. J. Bacteriol. **33:**275-282.

73. **Shu, C. J., and I. B. Zhulin.** 2002. ANTAR: an RNA-binding domain in transcription antitermination regulatory proteins. Trends Biochem Sci **27:**3-5.

74. **Starai, V. J., J. Garrity, and J. C. Escalante-Semerena.** 2005. Acetate excretion during growth of *Salmonella enterica* on ethanolamine requires phosphotransacetylase (EutD) activity, and acetate recapture requires acetyl-CoA synthetase (Acs) and phosphotransacetylase (Pta) activities. Microbiology (Reading, Engl) **151:**3793-801.

75. **Stojiljkovic, I., A. J. Baumler, and F. Heffron.** 1995. Ethanolamine utilization in *Salmonella typhimurium*: nucleotide sequence, protein expression, and mutational analysis of the cchA cchB eutE eutJ eutG eutH gene cluster. J. Bacteriol. **177:**1357-1366.

76. **Sudarsan, N., M. C. Hammond, K. F. Block, R. Welz, J. E. Barrick, A. Roth, and R. R. Breaker.** 2006. Tandem Riboswitch Architectures Exhibit Complex Gene Control Functions. Science **314:**300-304.

77. **Tanaka, S., M. R. Sawaya, and T. O. Yeates.** Structure and mechanisms of a protein-based organelle in *Escherichia coli.* Science **327:**81-4.

78. **Tannock, G. W., Cook, G. .** 2002. *In* C. D. Gilmore MS , Courvalin P , Dunny GM , Murray BE , Rice LB (ed.), The Enterococci: Pathogenesis, Molecular

Biology, and Antibiotic Resistance, vol. . Am Soc Microbiol Press, Washington, DC.

79. **Toledo-Arana, A., O. Dussurget, G. Nikitas, N. Sesto, H. Guet-Revillet, D. Balestrino, E. Loh, J. Gripenland, T. Tiensuu, K. Vaitkevicius, M. Barthelemy, M. Vergassola, M.-A. Nahori, G. Soubigou, B. Regnault, J.-Y. Coppee, M. Lecuit, J. Johansson, and P. Cossart.** 2009. The Listeria transcriptional landscape from saprophytism to virulence. Nature **459:**950-956.

80. **Tsoy, O., D. Ravcheev, and A. Mushegian.** 2009. Comparative Genomics of Ethanolamine Utilization. J. Bacteriol. **191:**7157-7164.

81. **Wakeman, C. A., W. C. Winkler, and C. E. Dann.** 2007. Structural features of metabolite-sensing riboswitches. Trends Biochem Sci **32:**415-24.

82. **Ward, D. E., C. C. van der Weijden, M. J. van der Merwe, H. V. Westerhoff, A. Claiborne, and J. L. Snoep.** 2000. Branched-Chain alpha -Keto Acid Catabolism via the Gene Products of the bkd Operon in *Enterococcus faecalis*: a New, Secreted Metabolite Serving as a Temporary Redox Sink. J. Bacteriol. **182:**3239-3246.

83. **Wickiser, J. K., W. C. Winkler, R. R. Breaker, and D. M. Crothers.** 2005. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. Mol Cell **18:**49-60.

84. **Williamson, R., C. Le Bouguenec, L. Gutmann, and T. Horaud.** 1985. One or Two Low Affinity Penicillin-binding Proteins May Be Responsible for the Range of Susceptibility of *Enterococcus faecium* to Benzylpenicillin. J Gen Microbiol **131:**1933-1940.

85. **Wilson, K. S., and P. H. von Hippel.** 1995. Transcription termination at intrinsic terminators: the role of the RNA hairpin. Proceedings of the National Academy of Sciences **92:**8793-8797.

86. **Wilson, S., and R. Drew.** 1991. Cloning and DNA sequence of amiC, a new gene regulating expression of the *Pseudomonas aeruginosa* aliphatic amidase, and purification of the amiC product. J. Bacteriol. **173:**4914-4921.

87. **Wilson, S. A., S. J. Wachira, R. E. Drew, D. Jones, and L. H. Pearl.** 1993. Antitermination of amidase expression in *Pseudomonas aeruginosa* is controlled by a novel cytoplasmic amide-binding protein. EMBO J **12:**3637-42.

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

89. **Wood, A. J. J., and B. E. Murray.** 2000. Vancomycin-Resistant Enterococcal Infections. New England Journal of Medicine **342:**710-721.

90. **Yang, S., N. T. Perna, D. A. Cooksey, Y. Okinaka, S. E. Lindow, A. M. Ibekwe, N. T. Keen, and C. H. Yang.** 2004. Genome-wide identification of plant-upregulated genes of Erwinia chrysanthemi 3937 using a GFP-based IVET leaf array. Mol Plant Microbe Interact **17:**999-1008.

Vita

Kristina Ann Fox was born in Pontiac, Michigan on September 8, 1981. Kristina is the daughter of Sharon Upperman and James Fox and has a younger brother James Fox Jr. After matriculating from Milford High School in 1999, she attended Central Michigan University in Mt. Pleasant, Michigan majoring in Biology. During college, she performed research in Dr. Alm laboratory in the Department of Biology. Kristina graduated from Central Michigan University in December of 2003 with a degree of Bachelor of Science in Biology. In August of 2004, she entered the University of Texas Graduate School of Biomedical Sciences in Houston, Texas. There she joined the laboratory of Dr. Danielle Garsin in the Department of Microbiology and Molecular Genetics at the University of Texas Health Sciences Center at Houston for her dissertation research. She also joined the graduate program in Microbiology and Molecular Genetics. Kristina is engaged to John Andrew Latham.

Publications:

Fox KA, Ramesh A, Stearns JE, Bourgogne A, Reyes-Jara A, Winkler WC, Garsin DA. Multiple posttranscriptional regulatory mechanisms partner to control ethanolamine utilization in Enterococcus faecalis. *Proc Natl Acad Sci*. 2009 Mar 17;106(11):4435-40.

Bourgogne A, Garsin DA, Qin X, Singh KV, Sillanpaa J, Yerrapragada S, Ding Y, Dugan-Rocha S, Buhay C, Shen H, Chen G, Williams G, Muzny D, Maadani A, Fox KA,

Gioia J, Chen L, Shang Y, Arias CA, Nallapareddy SR, Zhao M, Prakash VP, Chowdhury S, Jiang H, Gibbs RA, Murray BE, Highlander SK, Weinstock GM. Large scale variation in Enterococcus Faecalis illustrated by the genome analysis of strain OG1RF. *Genome Biol*. 2008;9(7):R110.

Bourgogne A, Singh KV, Fox KA, Pflughoeft KJ, Murray BE, Garsin DA. EbpR is important for biofilm formation by activating expression of the endocarditis and biofilm-associated pilus operon (ebpABC) of Enterococcus Faecalis OG1RF. *J Bacteriol*. 2007 Sep;189(17):6490-3.

Maadani A, Fox KA, Mylonakis E, Garsin DA. Enterococcus faecalis mutations affecting virulence in the Caenorhabditis elegans model host. *Infect Immun*. 2007 May;75(5):2634-7.

Permanent address:

2250 Holly Hall #240

Houston, TX 77054