MIAMI UNIVERSITY
The Graduate School

Certificate for Approving the Dissertation

We hereby approve the Dissertation

of

Racheal A. Devine

Candidate for the Degree

Doctor of Philosophy

_____
Mitch Balish, Ph.D., Director

_____
Eileen Bridge, Ph.D., Reader

_____
Rachael Morgan-Kiss, Ph.D., Reader

_____
Xiao-Wen Cheng, Ph.D., Reader

_____
Jack Vaughn, Ph.D., Graduate School Representative

ABSTRACT

REGULATORY FEATURES OF THE 5' UNTRANSLATED LEADER REGION OF *aroL* IN *ESCHERICHIA COLI* K12 AND THE sRNA, *ryhB,* IN *SHEWANELLA ONEIDENSIS* MR-1

by

Racheal A. Devine

RNA is an important regulator of gene expression within bacterial, eukaryotic, and archaeal cells. This work focuses on two aspects of RNA regulation: the first half investigates the role of regulatory features within the 5' untranslated leader region (UTR) of the *E. coli aroL* mRNA and the second half focuses on an sRNA in *S. oneidensis* MR-1. The 5'UTR of mRNAs contain information necessary for ribosome recognition and subsequent translation initiation. Translation initiation is a prominent part of gene expression, as it is the rate-limiting step of translation. The 70S ternary initiation complex contains initiator tRNA and the mRNA's start codon positioned in the P-site of the 70S ribosome. The Shine-Dalgarno (SD) sequence within the 5'UTR of the mRNA is an important feature that helps facilitate the initial interaction between the mRNA and the 30S subunit. Translation of mRNAs lacking an SD has been reported and suggests that alternative mechanisms of mRNA-30S interactions exist. The *aroL* mRNA contains a short open reading frame within its 5'UTR. Ribosome binding and expression assays showed that this open reading frame gets translated, and its translation affects *aroL* translation downstream. The upstream open reading frame binds 30S subunits in the absence of a canonical SD sequence. In this work, we have shown that multiple signals in the mRNA (upstream and downstream of the AUG) contribute to 30S binding to and translation from the AUG start codon. In this work we have also characterized an sRNA, a *ryhB* homologue, in *S. oneidensis* MR-1. sRNAs contain regulatory features in their sequence and structure that help regulate translation of bacterial mRNAs in response to environmental cues. Similar to what has been reported in *E. coli*, the sRNA is regulated in response to iron limitation in *S. oneidensis* MR-1 and may have a regulatory role in iron metabolism.

REGULATORY FEATURES OF THE 5' UNTRANSLATED LEADER REGION OF *aroL* IN *ESCHERICHIA COLI* K12 AND THE sRNA, *ryhB,* IN *SHEWANELLA ONEIDENSIS* MR-1

A DISSERTATION

Presented to the Faculty of

Miami University in partial

fulfillment of the requirements

for the degree of

Doctor of Philosophy

Department of Microbiology

by

Racheal A. Devine

The Graduate School
Miami University
Oxford, Ohio

2018

Dissertation Director: Mitch Balish, Ph.D.

Table of Contents

General Introduction……1

Chapter One……10

ABSTRACT……...11

IMPORTANCE……..…11

Introduction……...…12

Materials and Methods……..…14

Results……...….17

Ribosomes bind and initiate translation within the *aroL* untranslated leader……...17

aroL is translationally coupled to the uAUG-2/3 open reading frame……..……18

Identification of sequences needed for ribosome binding and expression from uAUG-2…….29

Proximity of the upstream signals to uAUG-2 is important for ribosome binding and expression……..……....41

Ribosome binding and expression from uAUG-2 involves additional sequence downstream of uAUG-2……42

The sequence surrounding uAUG-1 and uAUG-2 contributes to start site selection……...…..46

The upstream sequence elements required for uAUG-2 expression also promote expression of other coding regions……....51

Ribosomal protein S1 is required for 30S subunit binding to the *aroL* uAUG-1 and uAUG-2……..…52

Discussion……..59

Contribution of the upstream sequence (US) to expression……..59

Contribution of the downstream sequence (DR) to expression……..…..…..61

Start site selection (SSS)…….……..64

Model……..…65

Implications for regulation and expression……..….66

Chapter Two…….…..…..67

Introduction…….…..…68

Materials and Methods……71

Results……..….77

Description of *ryhB* locus in *S. oneidensis* MR-1……...…..77

S. oneidensis MR-1 growth and *ryhB* RNA levels in response to iron supplementation and chelation…….....……77

ryhB RNA levels in *S. oneidensis* MR-1 cells remained unchanged at 0, 7, 14, and 28 minutes after iron supplementation and chelation……...……..93

sodB RNA levels in mid-log phase *S. oneidensis* MR-1 cells did not change upon iron supplementation and chelation……...……..100

Construction of *ryhB* expression vectors…….105

Discussion……116

Appendix A……...……..120

Appendix B……..……..148

General Discussion…….……...178

References……...182

List of Tables

Table 1-1. Description of translational fusion constructs used in ribosome binding and expression studies……..22

Table 1-2. β-galactosidase activity of the *aroL* leader with segments of the leader sequence mutated (p.IF2,3mut1-14) or with nucleotide insertions (p.IF2,3.add1, p.IF2,3.add2, p.IF1.add1) or deletions (p.IF2,3.del1-7, p.IF1.del1) between the U rich region and uAUG-1…….....30

Table 1-3. β-galactosidase activity of *aroL* leader with regions downstream of uAUGs mutated (p.DR.mut1-4), or with nucleotide insertion (p.DR.add1, p.DR.add2) or deletions (p.DR.del1, p.DR.del2) between the uAUGs and downstream region 2 (DR2)……...…44

Table 1-4. β-galactosidase activity of the WT (with uAUG-1 [*aroL*Ld(u1)] or -2 [*aroL*Ld(u2)] in frame with *lacZ*) or mutated *aroL* leader (*aroL*LdMut(u2)) fused to internal fragments of the *tnaA*, and *aroL* genes (positions 382 to 531 of the *tnaA* coding region and 127 to 276 of the *aroL* coding region)…….…..…53

Table 2-1. Liquid MICs of *S. oneidensis* MR-1 cells grown in HBa media supplemented with or lacking 25 μM FeNTA and various concentrations of 2,2'-dipyridyl (DIP)…….....89

Table A-1. Descriptions and sequences of DNA and RNA oligonucleotides used in sequestration assays to identify sequences in the *aroL* 5' untranslated leader region responsible for ribosome recognition and binding to uAUG-2……124

Table B-1. Description of pBAD-*aroL* (leader) expression vectors…….……..154

List of Figures

Figure 1-1. Upstream and partial gene sequence of *E. coli* K12 *aroL*……..20

Figure 1-2. Primer extension inhibition (toeprint) analysis of 30S ribosomal subunit binding to the uAUG triplets of the upstream leader region and to the *aroL* AUG start codon.……25

Figure 1-3. β-galactosidase activity of the *aroL* leader + 16 codons (p.*aroL*.CDS), with various AUG knockouts or reading frames of the uAUGs fused to a *lacZ* reporter gene.....27

Figure 1-4. Primer extension inhibition (toeprint) analysis comparing 30S ribosomal subunit binding to uAUG-1 and uAUG-2 of the *aroL* upstream leader region in the presence of the two G to C substitutions (underlined) within the sequence GAUGGUAUG……...…....32

Figure 1-5. β-galactosidase activity (A) and primer extension inhibition (toeprint) assays (B) of the *aroL* leader (uAUGs 2,3 in frame with *lacZ*) with leader deletions…….....34

Figure 1-6. Primer extension inhibition (toeprint) analysis comparing 30S ribosomal subunit binding to uAUG-1 and uAUG-2 of the *aroL* upstream leader region in the presence of substitution mutations (p.IF2,3.mut13, p.IF2,3.mut14, p.DR.mut2), nucleotide insertions (p.IF2,3.add1), and nucleotide deletions (p.IF2,3.del3, p.DR.del2)…….....37

Figure 1-7. Primer extension inhibition (toeprint) analysis comparing 30S ribosomal subunit binding to uAUG-1 and uAUG-2 of the *aroL* upstream leader region in the presence of substitution mutations (p.IF2,3.mut2, p.IF2,3.KO2), nucleotide insertions (p.DR.add1), and nucleotide deletions (p.IF2,3.del2, p.IF2,3.del4, p.ΔGTATG.2, p.ΔGTATG.3).…….39

Figure 1-8. Primer extension inhibition (toeprint) analysis comparing 30S ribosomal subunit binding to uAUG-1 and uAUG-2 of the *aroL* upstream leader region in the presence of substitution mutations (p.IF2,3.KO1), nucleotide insertions (p.DR.add2), and nucleotide deletions (p.IF2,3.del7, p.ΔGTATG.3)…….47

Figure 1-9. β-galactosidase activity of the *aroL* leader with GUAUG-2 deleted (p.ΔGTATG.1) and 5 (p.ΔGTATG.2) or 10 (p.ΔGTATG.3) nucleotide insertions between the U-rich region and the uAUG.……..49

Figure 1-10. 30S subunits that lack ribosomal protein S1 do not bind to uAUG-1 or uAUG-2 of the *aroL* upstream leader on the WT (p.IF2,3) and mutant (p.IF2,3.mut13) *aroL* leader in primer extension inhibition (toeprint) assays……....55

Supplemental Figure 1-S1. Translational fusion constructs used in assays to measure ribosome binding and expression……..……57

Figure 2-1. Description of the *S. oneidensis* MR-1 *ryhB* sequence, the predicted RNA secondary structure, and multiple sequence alignment comparing *ryhB* in *S. oneidensis* MR-1 to *ryhB* in other Gammaproteobacteria……..…78

Figure 2-2. Primer extension analysis of *S. oneidensis* MR-1 *ryhB*……...……..83

Figure 2-3. *S. oneidensis* MR-1 growth curves in HBa minimal medium lacking (-FeNTA) or containing (+FeNTA) 25 μM FeNTA as an iron source in aerobic (A) or anaerobic (B) conditions……...…85

Figure 2-4. *S. oneidensis* MR-1 growth curves in HBa minimal medium lacking (-FeNTA) or containing (+FeNTA) 50 μM FeNTA as an iron source in aerobic (A) or anaerobic (B) conditions…….…….....87

Figure 2-5. qPCR analysis on *ryhB* levels in *S. oneidensis* MR-1 cells grown to early log phase in the presence of iron supplementation……..…..…….91

Figure 2-6. qPCR analysis on *ryhB* levels in *S. oneidensis* MR-1 cells grown to early log phase in the presence of iron chelation……..……..94

Figure 2-7. qPCR analysis on *ryhB* levels in *S. oneidensis* MR-1 cells at 0, 7, 14, and 28 minutes after iron supplementation……..…96

Figure 2-8. qPCR analysis on *ryhB* levels in *S. oneidensis* MR-1 cells at 0, 7, 14, and 28 minutes after addition of an iron chelator…….…..98

Figure 2-9. qPCR analysis on *ryhB* and *sodB* levels in *S. oneidensis* MR-1 cells grown to early log phase in the presence of iron supplementation……..101

Figure 2-10. qPCR analysis on *ryhB* and *sodB* levels in *S. oneidensis* MR-1 cells grown to early log phase in the presence of an iron chelator……...…103

Figure 2-11. Descriptions of the pBAD33-*ryhB* and pACYC-*ryhB* expression…….……107

Figure 2-12. Growth curves of *S. oneidensis* MR-1 cells in HBa medium with the empty pACYC vector and pACYC-*ryhB*……...……110

Figure 2-13. qPCR analysis of *ryhB* levels in *S. oneidensis* MR-1 cells containing the pACYC-*ryhB* expression vector……..…….….112

Figure 2-14. β-galactosidase assays were performed on *S. oneidensis* MR-1 cells containing the pBAD33-SDL*c*I-*lacZ* or pBAD33-LL*c*I-*lacZ* expression vectors…….…..…114

Figure A-1. Sequence of the *aroL* 5' untranslated leader region extending from the start site of transcription (+1) to the stop codon of the upstream open reading frame, overlapping the *aroL* start codon…….....122

Figure A-2. Primer extension inhibition (toeprint) analysis of 30S subunit binding to *aroL* uAUG-2 and the *aroL* AUG start codon in the presence of control oligonucleotides designed to hybridize with parts of the *aroL* leader……127

Figure A-3. Primer extension inhibition (toeprint) analysis of 30S subunit binding to *aroL* uAUG-2 and the AUG start codon in the presence of oligonucleotides designed to sequester parts of the *aroL* leader upstream of uAUG-2……..129

Figure A-4. Primer extension inhibition (toeprint) analysis of 30S subunit binding to *aroL* uAUG-2 and the AUG start codon in the presence of oligonucleotides at 200X or 400X concentrations over the *aroL* mRNA……...…131

Figure A-5. Primer extension inhibition (toeprint) analysis of 30S subunit binding to *aroL* uAUG-2 and the AUG start codon in the presence of DNA oligonucleotides designed to sequester the U-rich region of the *aroL* leader mRNA……..…….135

Figure A-6. Binding of DNA oligonucleotides to the 3' end of the 16S rRNA……..…140

Figure A-7. Primer extension inhibition (toeprint) analysis of 30S subunit binding to *aroL* uAUG-2 and the AUG start codon in the presence of DNA oligonucleotides designed to sequester the anti-Shine-Dalgarno sequence at the 3'end of the 16S rRNA of the 30S subunit.142

Figure A-8. Primer extension inhibition (toeprint) analysis of 30S subunit binding to *aroL* uAUG-2 and the AUG start codon in the presence of RNA oligonucleotides designed to prevent interactions between the *aroL* mRNA and the 3'end of the 16S rRNA of the 30S subunit……146

Figure B-1. Induction of the pBAD-uATG2-*lacZ* expression vector with various amounts of L-arabinose……...…...155

Figure B-2. Induction of the pBAD-LL*c*I-*lacZ* and pBAD-SDL*c*I-*lacZ* expression vectors with 0.05% of L-arabinose……...157

Figure B-3. Growth assays of *E. coli* RFS859 cells upon induction of pBAD expression vectors……..……....160

Figure B-4. Hydrophobicity analysis of the *aroL* upstream open reading frame……...163

Figure B-5. Vector map (A) and partial sequence (B) of pBAD-uATG2-*lacZ*……..……165

Figure B-6. Vector map (A) and partial sequence (B) of pBAD-uATG2……...167

Figure B-7. Vector map (A) and partial sequence (B) of pBAD-uATG2-*aroL*……..……..169

Figure B-8. Partial sequence of (A) pBAD-His-uATG2 and (B) pBAD-His-uATG2-*aroL*…..172

Figure B-9. Vector map and partial sequence of pBAD-SDL*c*I-*lacZ* and pBAD-LL*c*I-*lacZ*….174

General Introduction

For many years RNA molecules have been categorized into three classes based off of their roles in cellular biology: mRNA serves as a passive messenger between the genome and proteome, tRNA links the genetic code protein synthesis, and rRNA is a structural component of ribosomes. However, we are beginning to appreciate RNA also as a prominent player in gene regulation in bacterial, eukaryotic, and archaeal systems. Many of the regulatory features of RNAs are cis-acting and contained in the 5' or 3' untranslated regions (UTRs) of mRNA (Wagner and Simmons, 1994; Zeiler and Simon, 1996, Gerdes et al., 1997). Others, including micro (miRNA) or small (sRNA) RNAs, are transcribed independently from the molecule(s) that they regulate (Altuvia and Wagner, 2000, Lease and Belfort, 2000).

Many of the regulatory elements (e.g. sequences, secondary structure) found in the UTRs of mRNA function as "gates" that induce conformational changes to control transcription and translation. Riboswitches are an example of bacterial regulatory RNA sequences localized to the mRNA's 5'UTR and act as metabolic sensors (Mandal and Breaker, 2004). Binding of a ligand induces changes in secondary structure that lead to regulation of transcription and/or translation through premature termination of transcription or interference with ribosome-binding sites (Mandal and Breaker, 2004). RNA thermosensors are another example of elements in the 5'UTR that regulate ribosome binding and translation of the protein-encoding region of the mRNA (Hoe and Goguen, 1993; Johansson et al., 2002). These RNA elements respond to temperature increases by changes in secondary structure that make the ribosome binding site more accessible to the translation machinery (Hoe and Goguen, 1993; Johansson et al., 2002). In eukaryotes, specific RNA sequences within the 5'UTR of some viral RNAs and cellular mRNAs drive a cap- and scanning-independent mechanism of translation (Jang et al., 1988; Pelletier and Sonenberg, 1988; Glass and Summers, 1992; Sasaki and Nakashima; 2000, Jan, 2006; Kieft, 2008). These internal ribosome entry sites (IRES) have been identified in 39 viral RNAs and 85 cellular mRNAs (Baird et al., 2006) and therefore represent a prominent regulatory feature of translation initiation in eukaryotes.

This dissertation focuses on two examples of regulatory features of RNA. The first set of experiments described here (Chapter 1) explore the regulatory features located within the 5'UTR of the *aroL* mRNA of the bacterium *Escherichia coli.* The second set of experiments (Chapter 2) describes a small noncoding RNA (*ryhB* homologue) found in *Shewanella oneidensis* MR-1.

Canonical Translation

Canonical translation initiation in bacteria is a coordinated multistep process involving the 30S subunit, charged initiation tRNA (fMet-tRNA$^{(fMET)}$), initiation factors (IF-1, IF-2, and IF-3), and the mRNA. Complementary base pairing between the SD sequence located in the 5'UTR of the bacterial mRNA and the anti-Shine-Dalgarno (ASD) sequence located within the 3' terminus of the 16S rRNA helps to facilitate the initial binding of the 30S subunit to the mRNA (Shine and Dalgarno, 1974). Following the SD-ASD interaction, the initiation codon is moved into the P site where it interacts with the anticodon of the fMet-tRNA$^{(fMET)}$. The location of the SD relative to the start codon and the strength of the complementary pairing can influence this placement (Jacob et al., 1987; Dalboge et al*.,* 1988). The proper positioning of the mRNA's start codon at the P-site might not depend exclusively on the interaction between the SD and ASD, as crosslinking studies have identified interactions between mRNAs and the ribosome in addition to the SD and ASD (Rinke-Appel et al*.,* 1991; McCarthy and Brimacombe, 1994; Rinke-Appel et al., 1994).

IF-2 interacts with fMet-tRNA$^{(fMET)}$ and, in cooperation with IF-1, promotes proper positioning of the fMet-tRNA$^{(fMET)}$ to the P site in the 30S subunit (Laursen et al., 2005; Simonetti et al., 2008). Adjustment of codon-anticodon pairing is facilitated mainly by IF-3 (Simonetti et al., 2008), which also prevents binding of the 50S subunit. Upon the proper formation of the codon-anticodon interaction in the 30S subunit P site, IF-3 leaves the initiation complex, and the 50S subunit joins the 30S initiation complex to form the 70S initiation complex. The joining of the two subunits stimulates the hydrolysis of the GTP bound to IF-2, and IF-1 and -2 are released (Laursen et al., 2005; Simonetti et al., 2008). The resulting 70S initiation complex is now prepared for elongation on the mRNA.

Translation of leaderless mRNAs and non-canonical mRNAs

Although the SD-ASD complementary pairing plays an important role for initiation complex formation, translation of mRNAs lacking an SD occurs in bacteria and archaea (Chang et al., 2006). Leaderless mRNAs not only lack an SD sequence but they lack a 5'UTR. Translation of leaderless mRNA initiates at a start codon positioned immediately at the mRNA's 5' end; this AUG at the 5' end is somehow sufficient for ribosome recognition and binding (Moll et al*.,* 2004; Brock et al., 2008). Naturally occurring leaderless mRNAs were reported as early as 1976 with the lambda phage *c*I mRNA (Ptashne et al., 1976) and large numbers have been identified in bacteria and archaea (Bibb et al., 1994; Wu and Janssen, 1996; Benelli et al*.,* 2003).

In addition to leaderless mRNAs, there is another group of mRNAs that do not require an SD-ASD interaction for expression. These mRNAs have a 5'UTR but lack a canonical SD sequence. Comparison of candidate non-SD leadered mRNAs has not revealed a consensus sequence; rather, it is more likely that several different ribosome binding and translation initiation mechanisms are used (Chang et al., 2006; Hering et al*.,* 2009; Accetto and Avgustin, 2011). It has been suggested that mRNAs with SD-independent mechanisms may be as common as mRNAs with SD-dependent ribosome binding (Chang et al., 2006). Part of my work focuses on identifying the SD-independent ribosome binding signals that attract ribosomes to AUG triplets in the *aroL* 5'UTR, which may help us better understand the events of ribosome binding and translation initiation of mRNAs lacking an SD sequence.

Additional features within the mRNA that influence translation

A/U rich sequences within 5'UTRs enhance translation of mRNAs with SD sequences (McCarthy et al., 1986; Sleat et al., 1987; Gallie et al., 1987; Olins et al., 1988; Olins and Rangwala, 1989; Loechel et al., 1991; Zhang et al., 1992; Hirose and Sugiura, 2004; Hook-Barnard et al., 2007; Nafissi et al., 2012) and stimulate expression in the absence of SD sequences (Ivanov et al., 1995; Golshani et al., 2000). In most cases, the A/U element is not sufficient for translation initiation, but rather an essential element contributing to initiation (McCarthy et al., 1986; Loechel et al., 1991; Zhang et

al., 1992; Hirose and Sugiura, 2004; Hook-Barnard et al., 2007; Nafissi et al., 2012). Evidence supports that poly (A/U) tracts serve as recognition signals for binding of the 30S subunit via an interaction with ribosomal protein (r-protein) S1 (Boni et al., 1991; Ringquist et al., 1995). *E. coli* S1 is necessary for translation of most mRNAs in *E. coli* (Sorensen, et al., 1998), especially those that lack an SD sequence or contain a weak one (Boni et al., 1991; Farwell et al., 1992; Tedin et al., 1997).

In addition to A/U enhancing elements, other sequences influence translation of bacterial mRNAs. Adenine-rich and CA repeat sequences downstream of the start codon enhance translation of mRNAs with or without a 5'UTR (Martin-Farmer and Janssen, 1999; Brock et al., 2007). The downstream box (DB) is an extensively studied enhancing element located downstream of the initiation codon in a number of mRNAs (Sprengart et al., 1990; Nagai et al., 1991; Shean and Gottesman, 1992; Ito et al., 1993; Mitta et al., 1997). Although DBs have complementarity with the 16S rRNA, the complementary base pairing model has been invalidated (O'Connor et al., 1999) and the precise mechanism for ribosome-mRNA recognition has yet to be identified. In some cases, DBs can act independently of an SD or upstream enhancing element (Sprengart et al., 1996; Wu and Janssen, 1996; Winzeler and Shapiro, 1997; Zhang et al*.,* 1997) and the requirements for their position relative to the start codon are flexible (Sprengart et al., 1996).

Secondary structure can also be an influencing factor. A report by Gu and coworkers (2010) showed a prevalence of reduced mRNA stability near the start codons of genomes from bacteria, archaea, fungi, plants, and vertebrates. Their results strengthen the argument by Kudla et al., (2009) that secondary structure around the start codon is more influential than codon bias regarding gene expression.

Ribsome-mRNA Interactions

Crosslinking analyses have mapped non-SD-ASD interactions between the bacterial ribosome and SD-containing mRNAs. Specifically, nucleotides located between the SD and the start codon of an mRNA can be crosslinked to positions 665, 1360 and 1530 of the 16S rRNA (McCarthy and Brimacombe, 1994; Rinke-Appel et al*.,*

1994). Several positions on the 16S rRNA (e.g., 532, 1952, 1395, and 1402) can be crosslinked to nucleotides downstream of the start codon (Rinke-Appel et al., 1991; McCarthy and Brimacombe, 1994). Crosslinking analyses also show that the nucleotide positions -3, +2, and +11 (with the A of the AUG as position +1) contact r-proteins S1, S2, S7, S9, S11, S18, and S21 (La Teana et al., 1995). These interactions between the mRNA, rRNA, and r-proteins vary depending on the presence of IFs and fMet-tRNA$^{(fMET)}$ (Rinke-Appel et al., 1991; McCarthy and Brimacombe, 1994; Rinke-Appel et al., 1994). These data suggest that these r-proteins and conserved 16S rRNA nucleotides contribute to start codon selection in translation initiation.

The *aroL* mRNA of *E. coli* is an example of non-canonical translation initiation in bacteria

The *aroL* mRNA, encoding shikimate kinase II, contains a 125-nucleotide "untranslated" leader (5'UTR) that includes a canonical Shine-Dalgarno (SD) sequence upstream of the *aroL* start codon, and contains three additional AUG triplets (uAUG-1, uAUG-2 and uAUG-3). Primer extension inhibition (toeprint) assays indicate that 30S subunits bind to this upstream region (Brock et al., 2007). These results prompted my investigation (Chapter 1) into possible regulatory features residing within this 5'UTR.

General features of sRNAs

Bacterial sRNAs are regulatory entities, varying in length of 50-500 nts (Gottesman and Storz, 2011), which contribute to translation regulation by modifying mRNA stability and translation (Massé and Gottesman 2002; Gottesman, 2004; Gottesman and Storz, 2011). These RNA regulators are expressed in response to various environmental stimuli and are the bacterial equivalent of siRNAs or miRNAs of eukaryotes (Gottesman, 2004; Gottesman and Storz, 2011). Unlike siRNAs or miRNAs, bacterial sRNAs are transcribed as individual molecules that do not undergoe processing (Gottesman, 2004; Gottesman and Storz, 2011).

The majority of sRNAs act through RNA-RNA interactions by directly base-pairing with complementary regions of target mRNA, usually near the 5' end of the mRNA (Massé and Gottesman, 2002; Gottesman, 2004; Gottesman and Storz, 2011).

Both the sequences and structures of these sRNAs are believed to play a role in the interaction with their targets (Massé and Gottesman, 2002; Gottesman, 2004; Gottesman and Storz, 2011). Regions for base pairing within the sRNA are generally single stranded and often found in secondary loop structures (Massé and Gottesman, 2002; Gottesman, 2004; Gottesman and Storz, 2011). The majority of the sRNAs characterized to date inhibit translation, as the base pairing can prevent ribosome association and usually leads to degradation of the target mRNA through the action of RNaseE and sometimes RNase III (Massé and Gottesman, 2002; Gottesman, 2004; Gottesman and Storz, 2011). Consequently, pairing with the target mRNA leads to degradation of most of the sRNAs themselves.

Most sRNAs bind to and require the action of the chaperone protein Hfq (Gottesman, 2004; Gottesman and Storz, 2011), a hexameric protein whose ring-like structure is homologous to eukaryotic RNA splicing proteins (Schumacher et al., 2002; Sun et al., 2002; Sauter et al., 2003). This chaperone binds to A/U rich regions of single stranded RNA (sRNA and mRNA), typically next to stem loop structures (Moller et al*.,* 2002; Brescia, 2003). Hfq aids the stabilization of the sRNAs, as they fail to accumulate in cells lacking a functional Hfq (Zhang et al., 2003). Overexpression of the sRNA DsrA compensates for the absence of Hfq in an *hfq* mutant (Sledjeski et al., 2001). *In vivo* precipitation assays (Zhang et al., 2003) along with *in vitro* electrophoretic mobility shift assays (Gerdes et al., 1997; Moller et al., 2002; Vecerek et al*.,* 2003) support that Hfq can also bind the target mRNAs. The ability of Hfq to bind both the sRNA and its target mRNA suggests that in addition to stabilizing the RNAs, Hfq aids in bringing the two together and allows one sRNA to interact with multiple targets. The majority of sRNAs that do not require Hfq have been isolated from Gram-positive bacteria, although not all sRNAs from Gram-positive bacteria are Hfq-independent (Boisset et al., 2007; Heidrich et al., 2007; Christiansen et al., 2004, 2006). It is possible that the higher GC content found in some Gram-positive bacteria increases the stability of their sRNAs and negates the need for Hfq or, alternatively, that another chaperone fulfills the role.

sRNAs mediate transcript decay primarily through RNaseE. Similar to the Hfq binding sites, RNaseE recognizes and cleaves A/U rich regions of single stranded

RNAs (Massé et al., 2003). Two RNaseE mediated pathways have been proposed for mRNA instability. The first suggests that binding of the sRNA to the target mRNA prevents ribosomes from binding and leaves the exposed mRNA free to attack by RNasE (Wagner, 2009). The second suggests that the sRNA-mRNA duplex plays a more active role in activating RNasE (Bandyra et al., 2012). After the initial cleavage by RNaseE the degradation is completed by exonucleases (Massé et al., 2003).

Interactions between sRNAs and their targets

Cis-encoded sRNAs are transcribed from the DNA strand opposite their target (Wagner and Simmons, 1994; Zeiler and Simon, 1996), similar to the antisense regulation originally observed in plasmids, transposable elements, and bacteriophages. *Trans*-encoded sRNAs and their target RNAs, however, are transcribed from unlinked genes (Altuvia and Wagner, 2000) and are not necessarily located near the gene sequences of their targets on the genome (Altuvia and Wagner, 2000). Complementarity between *trans*-encoded sRNAs and their targets is not complete and each sRNA is capable of binding more than one target (Altuvia and Wagner, 2000). Binding of the sRNA can have different regulatory effects for different target mRNAs. For example, the sRNA *dsrA* binds and activates translation of *rpoS* mRNA but it also binds and inhibits translation of *hns* mRNA (Lease and Belfort, 2000). The extent of pairing between *trans*-acting sRNAs and their targets has not yet been fully studied but the mechanism by which *cis*-acting RNAs function may provide insight. Small regions of complementarity of *cis*-acting RNAs, usually in the loops of stem loop structures, form initial pairings with the target, which extend to larger interactions (Wagner and Simons, 1994; Gerdes et al., 1997).

Three sRNAs in *E. coli* are known to bind protein targets rather than mRNAs. RNA polymerase is partly regulated by binding to the 6S RNA (Wassarman and Storz, 2000) and CsrA is regulated by binding to CsrB and CsrC sRNAs (Liu et al., 1997; Majdalani et al*.,* 2001). 6S RNA regulation occurs at the level of transcription, as the structure of the 6S RNA resembles a sigma 70 promoter and sequesters the promoter binding regions of RNA polymerase (Wassarman and Storz, 2000). CsrB and CsrC sRNAs contain a sequence motif that is recognized by CsrA and act by sequestering

this translational regulator (Liu et al., 1997). It is not yet known whether Hfq plays a role in this mechanism of regulation, and if so, to what extent. The sRNAs GlmY and GlmZ are homologues, and whereas GlmZ base pairs with the *glmS* mRNA, GlmY interacts with the RNase adaptor protein (RapZ) to prevent its interaction with RNaseE and subsequent degradation of GlmY (Göpel et al., 2013).

sRNAs have intrinsic properties that allow them to elicit a strong and quick physiological response. As RNA molecules, they do not need to be translated into protein. Furthermore, most sRNAs can bind multiple target mRNAs (Massé and Gottesman 2002; Gottesman, 2004; Gottesman and Storz, 2011). As mentioned above, the majority of sRNAs are degraded along with their bound targets. This intrinsic shutdown mechanism suggests that sRNAs act stoichiometrically rather than catalytically and are active only while the signals for their expression are present (Gottesman, 2004).

The *ryhB* sRNAs of *E. coli* and *S. oneidensis* MR-1 are examples of RNA regulators

ryhB is a conserved non-coding sRNA originally identified in *E. coli* where its role in iron metabolism and regulation has been well characterized (Massé and Gottesman, 2002). Expression of *ryhB* is increased under conditions of cellular iron limitation (Massé et al., 2003). This sRNA targets mRNAs that encode nonessential Fe-containing proteins, allowing the cell to utilize the iron conservatively (Massé et al., 2003, 2005). We have identified a homologue to *ryhB* in *Shewanella oneidensis* MR-1. Although there are similarities at the sequence and structural levels to *E. coli ryhB,* our preliminary results suggest that *S. oneidensis* MR-1's *ryhB* may target mRNAs that are unique to this organism. Studying *ryhB* in *S. oneidensis* MR-1 may not only help us understand how iron is regulated in this organism but also provide further insight into how sRNAs in general interact with their targets.

The implications of studying *aroL* of *E. coli* and *ryhB* of *S. oneidensis* MR-1

The *aroL* and *ryhB* RNAs discussed above are very different from each other yet both have regulatory properties. Whereas the uORF expressed from the *aroL* 5'UTR

appears to play a role in expression of the downstream gene, *ryhB* is an RNA whose expression changes in response to intracellular iron. In this dissertation I describe our efforts to further characterize and understand the regulatory roles of these RNA molecules.

My first objective in Chapter 1 was to assess the implications of the upstream AUG triplets on expression of *aroL* downstream. My results support that translation of *aroL* is coupled to that of the uAUG-2 ORF. uAUG-2 lacks a canonical SD sequence positioned upstream, therefore, we reasoned that translation from this AUG occurred in a non-traditional fashion. My second objective was to investigate the features that contributed to ribosome recognition and binding to the upstream AUG. My results support a novel mechanism for translation initiation of the *aroL* uAUG-2 that requires multiple features located upstream of the AUG, nucleotides surrounding the AUG, and regions downstream. These results provide another piece of evidence that non-traditional or non-canonical translation initiation mechanisms are more common than originally thought and will help us broaden our understanding of this important step in protein synthesis.

My objectives In Chapter 2 were to characterize *ryhB* of *S. onedensis* MR-1 in terms of its size, sequence, and predicted secondary structure and to determine if expression of the sRNA in *S. oneidensis* MR-1 was regulated in response to iron, as observed in *E. coli.* My results from Chapter 2 show sequence similarity between *ryhB* in *S. oneidensis* MR-1 and *E. coli*. Furthermore, they support a role in the regulation of iron metabolism for *ryhB* of *S. oneidensis* MR-1 and they suggest that *ryhB* might target different mRNAs in *S. oneidensis* MR-1 compared to documented targets in *E. coli*. The work presented here implicates a role for sRNAs in the global physiology of this organism.

Chapter One

An upstream AUG triplet within the *aroL* untranslated leader region of *Escherichia coli* K12 utilizes multiple elements for translation initiation in the absence of a Shine-Dalgarno sequence

Running Title: Translation initiation of *aroL* uAUG-2 in *E. coli*

Authors: Racheal A. Devine, Leonard T. Buller and Gary R. Janssen

Affiliation: Department of Microbiology, Miami University, Oxford, Ohio, 45056

Address of Corresponding Author: Racheal Desmone-Devine, Nationwide Children's Hospital, 700 Children's Drive, Columbus, Ohio, 43205, Phone: 614-722-8456, Email: Racheal.Devine@nationwidechildrens.org

(Manuscript submitted to Journal of Bacteriology)

ABSTRACT

Shikimate kinase II, the enzyme encoded by *aroL* in *Escherichia coli*, plays a role in the synthesis of aromatic amino acids. We identified three AUG triplets (uAUG-1,-2,-3) within the 5' untranslated leader region (UTR) of *aroL* in *E. coli* that bound 30S ribosomal subunits. Translation of *aroL* was coupled to the uAUG-2 reading frame, whose stop codon overlaps the *aroL* start codon. Both the mutation of uAUG-2 and premature termination of translation from uAUG-2 decreased expression from the *aroL* start codon. Mutations including site-directed substitutions, deletions, and insertions, were introduced within the 5'UTR to identify the sequences responsible for recruitment of 30S subunits and subsequent translation from uAUG-2. Primer extension inhibition assays as well as fusions of the uAUG-2 reading frame to the *lacZ* reporter gene revealed that uAUG-2 bound 30S subunits and supported translation in a complex, non-canonical fashion involving sequences upstream and downstream of uAUG-2. We have identified a U-rich sequence (UUUUUUCUUUAC) upstream of uAUG-2 that is required for ribosome recruitment, most likely through an interaction with ribosomal protein S1.

IMPORTANCE

We demonstrated that translation initiation from uAUG-2 within the *Escherichia coli aroL* 5'UTR occurs in an intricate and nontraditional fashion. Translation initiation is the rate-limiting step in protein synthesis and dependent on signals typically found in the 5'UTR of prokaryotic mRNAs. Canonical binding of the mRNA to the 30S ribosomal subunit is facilitated by an interaction of the mRNA's Shine-Dalgarno (SD) sequence located upstream of the AUG start codon with the complementary anti-SD sequence on the 16s rRNA. While many bacterial mRNAs contain an SD sequence, the number of non-SD led mRNAs (those that lack a canonical SD sequence) and mRNAs that lack a 5'UTR altogether (leaderless mRNAs) is growing. Identification of more non-traditional translation initiation signals will provide a more comprehensive understanding of the mechanisms contributing to this fundamental process.

Introduction

Initiation, the rate-limiting step in translation, is dependent on signals present in the messenger RNA for start codon recognition by the ribosome. The features contributing to ribosome recognition and translation initiation of prokaryotic mRNAs are generally found in the 5' untranslated leader region (5'UTR). The most notable feature of the 5'UTR is the purine-rich Shine-Dalgarno (SD) sequence properly positioned upstream of the start codon, which shares complementarity and base pairs with the anti-Shine-Dalgarno (ASD) sequence near the 3' end of the 16S rRNA (Shine and Dalgarno, 1974; Hui and de Boer, 1987; Jacob et al., 1987; Yamagishi et al., 1987; Chen et al., 1994). Although the SD-ASD interaction is an important component of translation initiation, additional features of mRNAs have been shown to influence this process, including the start codon itself (Schneider et al., 1986; Vellanoweth and Rabinowitcz, 1992; O'Donnell and Janssen, 2001), secondary structures within the mRNA's translation initiation region (TIR) (de Smit and van Duin, 1994, 2003; Unoson, 2007; Nafissi et al., 2012), and sequences upstream (Nafissi et al., 2012; Olins et al., 1988; Gallie and Kado, 1989; Ivanov et al., 1995; Golshani et al., 1997) as well as downstream (Chen et al., 1994; Sprengart et al., 1990; Ito et al., 1993; Mitta et al., 1997; Brock et al., 2007; Loh et al., 2012) of the start codon.

Signals within the 5'UTR are not the sole determinant for mRNA-ribosome association and translation initiation as initiation and expression from leaderless mRNAs occur in the absence of an SD and 5'UTR (Ptashne et al., 1976; Wu and Janssen, 1996; Martin-Farmer and Janssen, 1999; Moll et al., 2004). In addition, large-scale genome analyses have revealed a large number of prokaryotic mRNAs whose 5'UTRs lack SD sequences (non-SD-led mRNAs), suggesting that SD-independent mechanisms of translation initiation are more prevalent than originally thought (Chang et al., 2006). Comparisons of non-SD-led mRNAs in Bacteria and Archaea have not identified a consensus sequence within the 5'UTR, suggesting that multiple mechanisms might exist for ribosome recognition and expression from these start codons (Chang et al., 2006; Hering et al., 2009; Accetto and Avgustin, 2011).

The *Escherichia coli aroL* gene encodes shikimate kinase II, an enzyme involved in the biosynthesis of aromatic amino acids (Millar et al., 1986). The *aroL* mRNA contains an untranslated leader region of 125 nucleotides (DeFeyter and Pittard, 1986; Millar et al., 1986) which includes a canonical Shine-Dalgarno (SD) sequence upstream of the *aroL* start codon. 30S ribosomal subunits bind to the *aroL* start codon, as well as to a site within the *aroL* 5'UTR, in a tRNA-dependent manner (Brock et al., 2007). The 30S subunit binding is localized to a region containing three additional upstream AUG triplets (uAUG-1, -2, -3) within *aroL*'s 5'UTR. We examined the uAUGs for their ability to bind ribosomes and initiate translation. We found that *aroL* expression was translationally coupled to uAUG-2 expression, and the coupling required *aroL*'s SD sequence. Ribosome binding and expression from uAUG-2 occurred in the absence of a canonical SD sequence. Deletion and substitution analyses were used to identify an upstream U-rich sequence and a downstream region as necessary for ribosome binding and expression from uAUG-2. In addition, the sequence directly surrounding uAUG-1 and uAUG-2 also influenced ribosome binding and expression. Ribosomal protein S1 was directly involved in 30S subunit recognition of these upstream uAUG triplets, as 30S subunits devoid of S1 did not produce a stable ternary toeprint complex. Addition of S1 to toeprint reactions containing the S1 deficient 30S subunits restored toeprints to uAUG-1 and uAUG-2. Taken together, we conclude that expression from uAUG-2 within the *aroL* 5'UTR occurs through ribosome recognition of non-traditional signals undetectable by conventional search approaches for translationa start sites, providing an example of features contributing to expression of a non-SD led coding sequence.

Materials and Methods

Bacterial strains. *E. coli* K12 total genomic DNA was used for the isolation of *aroL*, *lacZ*, and *tna* gene fragments. *E. coli* DH5α (New England Biolabs) was used as the host for all plasmid DNA manipulations. *E. coli* RFS859 (F^-, *thr*-1, *ara*C859, *leu*B6, Δ*lac*74, *tsx*-274, λ^-, *gyr*A111, *rec*A11, *rel*A1, *thi*-1) (Schleif, 1972) was used as the host strain for expression of *lac*Z fusion constructs. *E. coli* MRE600 (Wade and Robinson, 1966) was used as the host strain for ribosome isolation.

Reagents and recombinant DNA procedures. Oligonucleotides were purchased from Integrated DNA Technologies (IDT). Radiolabeled [γ-^{32}P]ATP (6,000 Ci/mmol, 150 mCi/mL) was purchased from Perkin Elmer. *Pfu* DNA polymerase (Stratagene), T4 DNA ligase and restriction endonucleases (New England Biolabs), AMV reverse transcriptase (Life Sciences), and RNase-free DNaseI (Ambion) were used according to the manufacturers' recommendations. T4 polynucleotide kinase (Wang and Shuman, 2001) and T7 RNA polymerase (Davanloo et al., 1984; Krishnan, 2010) were purified as described.

General cloning procedures, plasmid isolations, *E. coli* transformations, and other DNA manipulations were carried out in a standard manner (Sambrook et al., 1989). Site-directed mutagenesis using PCR and oligonucleotides containing specific nucleotide substitutions was utilized to generate (DNA) mutations. To measure expression from *aroL, lacZ* fusions were constructed with contained the 5'UTR plus 16 codons of *aroL* fused to *lacZ*. To measure expression from uAUG-2 and -3, a leader fragment containing the first 99 nucleotides of the 5'UTR were fused to *lacZ*. To measure expression from uAUG-1, a leader fragment containing the first 97 nucleotides of the 5'UTR were fused to *lacZ*. To measure expression from uAUG-1, site-directed mutagenesis was used to change the in-frame stop codons at positions three and eight from UGA to UGC.

Ribosome Isolation. Isolation of *E. coli* ribosomes and subunits were purified as previously described (Martin-Farmer and Janssen, 1996). Isolation of *E. coli* ribosomes and subunits deficient for protein S1 were purified as described by Krishnan (2010).

Purification of His-tagged S1. The overexpression and purification of ribosomal protein S1 from cells containing plasmid pHis$_6$-S1 (obtained from Robert Sauer) was performed as previously described (McGinness and Sauer, 2004).

β-galactosidase activity measurements. β-galactosidase assays were performed as described (Miller, 1992). Cultures of *E. coli* RFS 859 transformed with plasmids described here were grown in 2XYT (Kunkel et al., 1987) medium supplemented with ampicillin (200 μg/mL) at 37^{0}C. Assays were performed with cells grown to an OD_{600} of ~0.4 and represent the average value of triplicate assays from at least three independent cultures.

Radio-labeling of oligonucleotides. Oligonucleotides were end-labeled as described previously (Van Etten and Janssen, 1998).

Preparation of in vitro synthesized transcripts. RNAs used in toeprint assays were synthesized and purified as described (Fredrick and Noller, 2002). RNAs were transcribed from a T7 promoter contained on DNA templates. Transcription reactions were performed with T7 RNAP in 1X buffer (40 mM Tris pH 7.8, 25 mM $MgCl_2$, 1 mM spermidine, 0.01% Triton X-100, 5 mM each NTP, and 30 mM dithiothreitol). Reactions were incubated at 37^{0}C for 4 hours, after which EDTA was added to 40 mM. Samples were treated with DNaseI for 15 minutes at 37^{0}C. RNA was subsequently precipitated with ethanol and suspended in RNA loading dye (50% formamide, 0.05% bromophenol blue, 0.05% xylene cyanol). The RNA was subjected to polyacrylamide gel electrophoresis (PAGE; 6% acrylamide, 7 M urea) and the products were excised using ultraviolet shadowing. Minced gel slices were rocked overnight at room temperature in elution buffer (300 mM sodium acetate pH 5.2, 0.1% SDS, 1 mM EDTA) and the eluted RNA was phenol-extracted and precipitated with sodium acetate and ethanol.

Primer extension inhibition (toeprint) assays. Toeprint assays were performed essentially as described (Martin-Farmer and Janssen, 1999). DNA oligonucleotides were phosphorylated at the 5'-terminus using [γ-^{32}P]ATP (6,000 Ci/mmol, 150 mCi/mL; Perkin Elmer) and T4 PNK in 1X NEB kinase buffer (70 mM Tris-HCl pH 7.6, 10 mM

$MgCl_2$, 5 mM dithiothreitol) for 30 minutes at 37°C and annealed to 3'-termini of RNA in 1X SB-Mg2+ (60 mM NH_4Cl, 10 mM Tris acetate pH 7.4).

Annealed RNA was mixed with 30S subunits in the presence or absence of $tRNA^{fmet}$ for 15 minutes at 37°C in 1X SB (60 mM NH_4Cl, 10 mM Tris acetate pH 7.4, 10 mM magnesium acetate, 6 mM β-mercaptoethanol). Reverse transcriptase was added and samples were incubated for an additional 15 minutes at 37^0C. Reverse transcription was stopped by the addition of sodium acetate (70 mM) and ethanol (2.5X the volume). Complexes were precipitated at -80^0C and dissolved in loading dye (80% deionized formamide, 10 mM NaOH, 1 mM EDTA, and 0.5% bromophenol blue and xylene cyanol) and subjected to PAGE (6% acrylamide, 7 M urea) in 1X Tris/Borate/EDTA (TBE) buffer.

Results

Ribosomes bind and initiate translation within the *aroL* untranslated leader

Previous work reported tRNA-dependent ribosome binding to a site within *E. coli* K12's *aroL* 5'UTR (Brock et al., 2007); Figure 1-1 shows *aroL*'s 125-nucleotide 5'UTR and partial gene sequence. In addition to a canonical SD sequence (Fig. 1-1, underlined) positioned upstream of the *aroL* start codon, the *aroL* mRNA contains three upstream AUG triplets (uAUG-1, uAUG-2 and uAUG-3) within the 5'UTR located 59, 64, and 79 nucleotides, respectively, downstream from the transcriptional start site; none are preceded by a canonical SD sequence. The third codon of the uAUG-1 reading frame is a stop codon that overlaps uAUG-2. The uAUG-2 reading frame encodes a putative 21-amino acid peptide, with an UGA stop codon partially overlapping the downstream *aro*L AUG start codon, suggesting possible translational coupling of *aroL* to the uAUG-2 reading frame. uAUG-3 corresponds to codon 6 within the uAUG-2 reading frame and we refer to these overlapping reading frames as uAUG-2/3. Primer extension inhibition assays (toeprints) were used to assess 30S ribosomal subunit binding *in vitro* to the uAUG triplets. 30S subunits bind the *aroL* start codon and the uAUG triplets in a tRNA-dependent manner (Fig. 1-2). Toeprint signal intensity suggests that 30S subunit binding to uAUG-1 was minimal compared to binding to uAUG-2 and -3.

To determine if the 30S subunit binding (Fig. 1-2) was accompanied by translational activity, the *aroL* uAUG-1 (p.IF1) or uAUG-2/3 (p.IF2,3) reading frames (from transcriptional start site to 33/34 nucleotides downstream of uAUG-2) were fused in-frame (IF) to a *lacZ* reporter gene. Expression from the uAUG-*lacZ* fusions was compared to expression of the *aroL* coding sequence (16 codons) fused to *lacZ* (referred to as p.*aroL.*CDS), arbitrarily chosen as a reference point for comparison of expression levels. To distinguish uAUG-2 expression from uAUG-3, the uAUG-2/3 reading frame containing an AUG → AUC knockout (KO) mutation in uAUG-2 or uAUG-3 (p.IF2,3[KO2], p.IF2,3[KO3], respectively) was fused to *lacZ*. β-galactosidase assays were performed to measure expression from uAUG-1, -2, and -3 (Fig. 1-3). Plasmid constructs are listed and described in Table 1-1 and depicted in Supplemental Figure 1-

S1. uAUG-2 supported the highest *lacZ* expression of the uAUGs, representing 30% of the activity expressed from the downstream *aroL* CDS (set at 100%) (p.IF2,3[KO3]and p.*aroL*.CDS, respectively; Fig. 1-3). uAUG-1 supported a minimal amount of expression (8%; p.IF1), and uAUG-3 expression was negligible (p.IF2,3 [KO2]) (Fig. 1-3). Although the stop codon of the uAUG-1 open reading frame overlaps uAUG-2, the two open reading frames do not appear to be translationally coupled because mutation of uAUG-1 to AUC (p.IF2,3[KO1]) did not affect expression from uAUG-2 (compare p.IF2,3[KO1] to p.IF2,3; Fig. 1-3).

aroL is translationally coupled to the uAUG-2/3 open reading frame

To investigate the possible influence of the uAUG triplets on *aroL* expression, a series of *aroL-lacZ* translational fusions was made in which the uAUG triplets were changed to AUC (Fig. 1-S1A). Mutations are listed in Table 1-1. Mutation of uAUG-1 (p.KO1) or uAUG-3 (p.KO3) had a negligible effect on *aroL-lacZ* activity (when compared to p.*aroL*.CDS), suggesting that uAUG-1 and -3 do not impact *aroL* expression significantly (Fig. 1-3). However, mutation of uAUG-2 (p.KO2), or the double knockout of uAUG-1 and -2 (p.DKO), resulted in an approximate 30% reduction in *aroL-lacZ* expression. Interestingly, a triple knockout of all three uAUG triplets (p.TKO) reduced *aroL-lacZ* expression by approximately 75%. These data indicate that the uAUG triplets do contribute to downstream *aroL* expression. Toeprint assays verified that the mutated uAUG triplets do not bind 30S subunits (data not shown).

To determine whether the uAUGs themselves or translation of the uAUG-2/3 ORF were important for *aroL* expression, the eighth codon downstream of uAUG-2 (CCG) was mutated to a UGA stop codon (p.Stop in Fig. 1-3). Truncation of the uAUG-2/3 ORF, with termination of translating ribosomes at codon 8, resulted in a 65% decrease of *aroL-lacZ* expression, comparable to that seen with the triple uAUG knockout (p.TKO in Fig. 1-3). Preventing expression from uAUG-2 (by mutation to AUC) or terminating translation prematurely (by introducing a UGA stop codon) decreased expression from the *aroL* start codon, consistent with a coupling of *aroL* expression to translation of the uAUG-2/3 ORF. Mutation of the *aroL* SD sequence (p.SD.mut1 or p.SD.mut2; Table 1-1) to eliminate SD-dependent internal ribosome binding and

expression from the *aroL* start codon, and thereby coupling all *aroL* expression to uAUG-2 initiation, resulted in a 90% reduction in *aroL* expression, suggesting that translational coupling requires the SD sequence for positioning of the uAUG-2/3 terminating ribosome at the *aroL* start codon (Fig. 1-3).

Figure 1-1. **Upstream and partial gene sequence of *E. coli* K12 *aroL*.** The *aroL* 125-nucleotide untranslated leader contains a Shine-Dalgarno (SD) sequence (underlined) upstream of the *aroL* ATG start codon (red). The untranslated leader sequence contains three ATG triplets, labeled uATG-1, -2 and -3 (blue). The leader-encoded uATGs lack an upstream canonical SD sequence. The second uATG (i.e., uATG-2) encodes a putative polypeptide whose stop codon (TGA; boxed) overlaps the *aroL* start codon. Nucleotide positions from the transcriptional start site are indicated above the sequence.

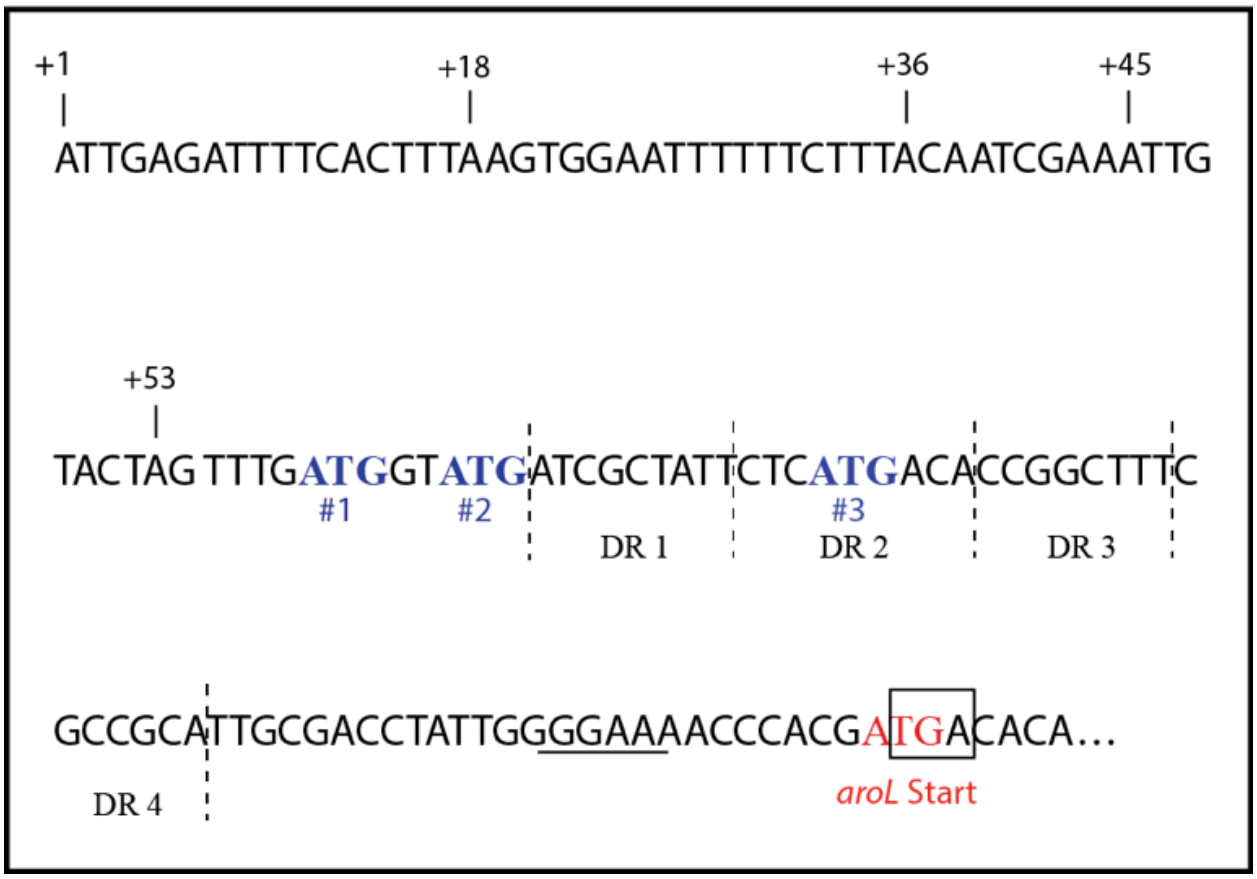

+1
+18
+36
+45
ATTGAGATTTTCACTTTAAGTGGAATTTTTTCTTTACAATCGAAATTG
+53
TACTAG TTTGATGGTATGATCGCTATTCTCATGACACCGGCTTTC
#1
#2
#3
DR 1
DR 2
DR 3
GCCGCATTGCGACCTATTGGGGAAAACCCACGATGACACA…
DR 4
aroL Start

Table 1-1. Description of translational fusion constructs used in ribosome binding and expression studies.

Plasmid Construct	Description
p.*aroL*.CDS	*aroL* leader + 16 codons of *aroL* fused to *lacZ*
p.KO1	*aroL* leader + 16 codons of *aroL* fused to *lacZ*; uATG-1 --> ATC
p.KO2	*aroL* leader + 16 codons of *aroL* fused to *lacZ*; uATG-2 --> ATC
p.KO3	*aroL* leader + 16 codons of *aroL* fused to *lacZ*; uATG-3 --> ATC
p.DKO	*aroL* leader + 16 codons of *aroL* fused to *lacZ*; uATG-1,2 --> ATC
p.TKO	*aroL* leader + 16 codons of *aroL* fused to *lacZ*; uATG-1,2,3 --> ATC
p.Stop	*aroL* leader + 16 codons of *aroL* fused to *lacZ;* TGA stop codon inserted 2 codons downstream from uATG-3
p.SD.mut1	*aroL* leader + 16 codons of *aroL* fused to *lacZ*; *aroL* SD sequence mutated to its complement, CCCCTTTT
p.SD.mut2	*aroL* leader + 16 codons of *aroL* fused to *lacZ*; *aroL* SD sequence mutated to GGCGAGAAT, which maintains the natural coding capacity
p.IF1	*aroL* leader fragment fused to *lacZ* 34nt downstream to uATG-2; uATG-1 in frame with *lacZ, T*GA stop codon -->TGC
p.IF2	*aroL* leader fragment fused to *lacZ* 12nt downstream to uATG-2 just before uATG-3; uATG-2 in frame with *lacZ*
p.IF2,3	*aroL* leader fragment fused to *lacZ* 33nt downstream to uATG-2; uATG-2,3 in frame with *lacZ*
p.IF2,3(KO1)	*aroL* leader fragment fused to *lacZ*; uATG-2,3 in frame with *lacZ; u*ATG-1 mutated to ATC
p.IF2,3(KO2)	*aroL* leader fragment fused to *lacZ*; uATG-2,3 in frame with *lacZ; u*ATG-2 mutated to ATC
p.IF2,3(KO3)	*aroL* leader fragment fused to lacZ; uATG-2,3 in frame with lacZ; uATG-3 mutated to ATC
p.Δ18	*aroL* leader fragment deleted to the +18 position from the transcriptional start fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.Δ36	*aroL* leader fragment deleted to the +36 position from the transcriptional start fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.Δ36.2	*aroL* leader fragment deleted to the +36 position from the transcriptional start fused to *lacZ*; CAATC at positions +2 - +6 substituted; uATG-2,3 in frame with *lacZ*
p.Δ36.3	*aroL* leader fragment deleted to the +36 position from the transcriptional start fused to *lacZ*; GAA at positions +7 - +9 substituted; uATG-2,3 in frame with *lacZ*
p.Δ36.4	*aroL* leader fragment deleted to the +36 position from the transcriptional start fused to *lacZ*; CAATCGAA at positions +2 - +9 substituted; uATG-2,3 in frame with *lacZ*
p.Δ45	*aroL* leader fragment deleted to the +45 position from the transcriptional start fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.Δ53	*aroL* leader fragment deleted to the +53 position from the transcriptional start fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.LLuATG-2	*aroL* leader fragment deleted to uATG-2 fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.IF2,3.mut1	*aroL* leader fragment with the sequence 'GATGG' mutated to 'CATGC' fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.IF2,3.mut2	*aroL* leader fragment with nucleotide sequence 'AAGT' substituted fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.IF2,3.mut3	*aroL* leader fragment with nucleotide sequence 'GGAA' substituted fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.IF2,3.mut4	*aroL* leader fragment with nucleotide sequence 'TTTTT' substituted fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.IF2,3.mut5	*aroL* leader fragment with nucleotide sequence 'TCT' substituted fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.IF2,3.mut6	*aroL* leader fragment with nucleotide sequence 'TTAC' substituted fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.IF2,3.mut7	*aroL* leader fragment with nucleotide sequence 'AATC' substituted fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.IF2,3.mut8	*aroL* leader fragment with nucleotide sequence 'GAA' substituted fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.IF2,3.mut9	*aroL* leader fragment with nucleotide sequence 'TTG' substituted fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.IF2,3.mut10	*aroL* leader fragment with nucleotide sequence 'TAC' substituted fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.IF2,3.mut11	*aroL* leader fragment with nucleotide sequence 'TAG' substituted fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.IF2,3.mut12	*aroL* leader fragment with nucleotide sequence 'TTTG' substituted fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.IF2,3.mut13	*aroL* leader fragment with nucleotide sequence 'TTTTTTCTTTAC' substituted fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.IF2,3.mut14	*aroL* leader fragment with nucleotide sequences 'AAGT' and 'TTTTTTCTTTAC' substituted fused to *lacZ*; uATG-2,3 in frame with *lacZ*

p.IF2,3.del1	*aroL* leader fragment with nucleotide sequence 'TTTTTTCTTTAC' deleted fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.IF2,3.del2	*aroL* leader fragment with nucleotide sequences 'AAGT' and 'TTTTTTCTTTAC' deleted fused to *lacZ*; uATG-2,3 in frame with *lacZ*
p.IF2,3.add1	*aroL* leader fragment with 5 nucleotides inserted between the T-rich region and uATG-1; uATG-2/3 in frame with *lacZ*
p.IF2,3.add2	*aroL* leader fragment with 15 nucleotides inserted between the T-rich region and uATG-1; uATG-2/3 in frame with *lacZ*
p.IF2,3.del3	*aroL* leader fragment with 5 nucleotides deleted between the T-rich region and uATG-1; uATG-2/3 in frame with *lacZ*
p.IF2,3.del4	*aroL* leader fragment with 15 nucleotides deleted between the T-rich region and uATG-2; uATG-2/3 in frame with *lacZ*
p.IF2,3.del5	*aroL* leader fragment with 15 nucleotides deleted between the T-rich region and uATG-1; uATG-3 mutated to ATC; uATG-2/3 in frame with *lacZ*
p.IF2,3.del6	*aroL* leader fragment with 15 nucleotides deleted between the T-rich region and uATG-1; uATG-2 mutated to ATC; uATG-2/3 in frame with *lacZ*
p.IF2,3.del7	*aroL* leader fragment with nucleotide sequence 'AAGT' deleted fused to *lacZ;* uATG-2,3 in frame with *lacZ*
p.IF1.add1	*aroL* leader fragment with 5 nucleotides inserted between the T-rich region and uATG-1; uATG-1 in frame with *lacZ*
p.IF1.del1	*aroL* leader fragment with 5 nucleotides deleted between the T-rich region and uATG-1; uATG-1 in frame with *lacZ*
p.DR.mut1(uATG-2)	*aroL* leader fragment with nucleotides 4-12 (region 1) of uATG-2 coding sequence (A of ATG is 1) substituted and fused to *lacZ*; uATG-2/3 in frame with *lacZ*
p.DR.mut2(uATG-2)	*aroL* leader fragment with nucleotides 13-21 (region 2) of uATG-2 coding sequence (A of ATG is 1) substituted and fused to *lacZ*; uATG-2/3 in frame with *lacZ*
p.DR.mut3(uATG-2)	*aroL* leader fragment with nucleotides 22-28 (region 3) of uATG-2 coding sequence (A of ATG is 1) substituted and fused to *lacZ*; uATG-2/3 in frame with *lacZ*
p.DR.mut4(uATG-2)	*aroL* leader fragment with nucleotides 29-36 (region 4) of uATG-2 coding sequence (A of ATG is 1) substituted and fused to *lacZ*; uATG-2/3 in frame with *lacZ*
p.DR.del1	*aroL* leader fragment with 2 codons of uATG-1 reading frame deleted; uATG-1 in frame with *lacZ*
p.DR.del2	*aroL* leader fragment with 2 codons of uATG-2/3 reading frame deleted; uATG-2/3 in frame with *lacZ*
p.DR.add1	*aroL* leader fragment with 2 codons of uATG-1 reading frame deleted and 5 nucleotides inserted between the T-rich region and uATG-1; uATG-1 in frame with *lacZ*
p.DR.add2	*aroL* leader fragment with 4 codons inserted between uATG-2 and region 2 (nucleotides 13-21) downstream; uATG-2/3 in frame with *lacZ*
p.ΔGTATG.1	*aroL* leader fragment with GTATG-2 deleted and fused to *lacZ*
p.ΔGTATG.2	*aroL* leader fragment with GTATG-2 deleted and 5 nucleotides inserted between the T-rich region and uATG; fused to *lacZ*
p.ΔGTATG.3	*aroL* leader fragment with GTATG-2 deleted and 10 nucleotides inserted between the T-rich region and uATG; fused to *lacZ*
p.*aroL*Ld(u1)-*aroL*IN	*aroL* leader fragment fused to an internal fragment of *aroL* containing an ATG plus 16 codons fused to *lacZ;* uATG-1 in frame with *lacZ*
p.*aroL*Ld(u2)-*aroL*IN	*aroL* leader fragment fused to an internal fragment of *aroL* containing an ATG plus 16 codons fused to *lacZ;* uATG-2 in frame with *lacZ*
p.*aroL*LdMut(u2)-*aroL*IN	*aroL* leader fragment with nucleotide sequences 'AAGT' and 'TTTTTTCTTTAC' substituted fused to internal fragment of *aroL* fused to *lacZ;* uATG-2 in frame with *lacZ*
p.*lac*Ld-*aroL*IN	*lac* leader fragment fused to an internal fragment of *aroL* containing an ATG plus 16 codons fused to *lacZ*
p.*aroL*Ld(u1)-*tna*IN	*aroL* leader fragment fused to an internal fragment of *tnaA* containing an ATG plus 16 codons fused to *lacZ;* uATG-1 in frame with *lacZ*
p.*aroL*Ld(u2)-*tna*IN	*aroL* leader fragment fused to an internal fragment of *tnaA* containing an ATG plus 16 codons fused to *lacZ;* uATG-2 in frame with *lacZ*
p.*aroL*LdMut(u2)-*tna*IN	*aroL* leader fragment with nucleotide sequences 'AAGT' and 'TTTTTTCTTTAC' substituted fused to internal fragment of *tna* fused to *lacZ;* uATG-2 in frame with *lacZ*
p.*lac*Ld-*tna*IN	*lac* leader fragment fused to an internal fragment of *tna* containing an ATG plus 16 codons fused to *lacZ*

Figure 1-2. Primer extension inhibition (toeprint) analysis of 30S ribosomal subunit binding to the uAUG triplets of the upstream leader region and to the *aroL* AUG start codon. Assays were performed with 30S subunits and p.*aroL*.CDS mRNA, in the presence (+) or absence (-) of tRNA. Toeprint signals for 30S subunits bound to uAUG-1, uAUG-2, uAUG-3 and the *aroL* AUG are indicated by an arrow and occur at the expected +16 position (with the A of AUG as +1). An *aroL* DNA sequencing ladder is on the left.

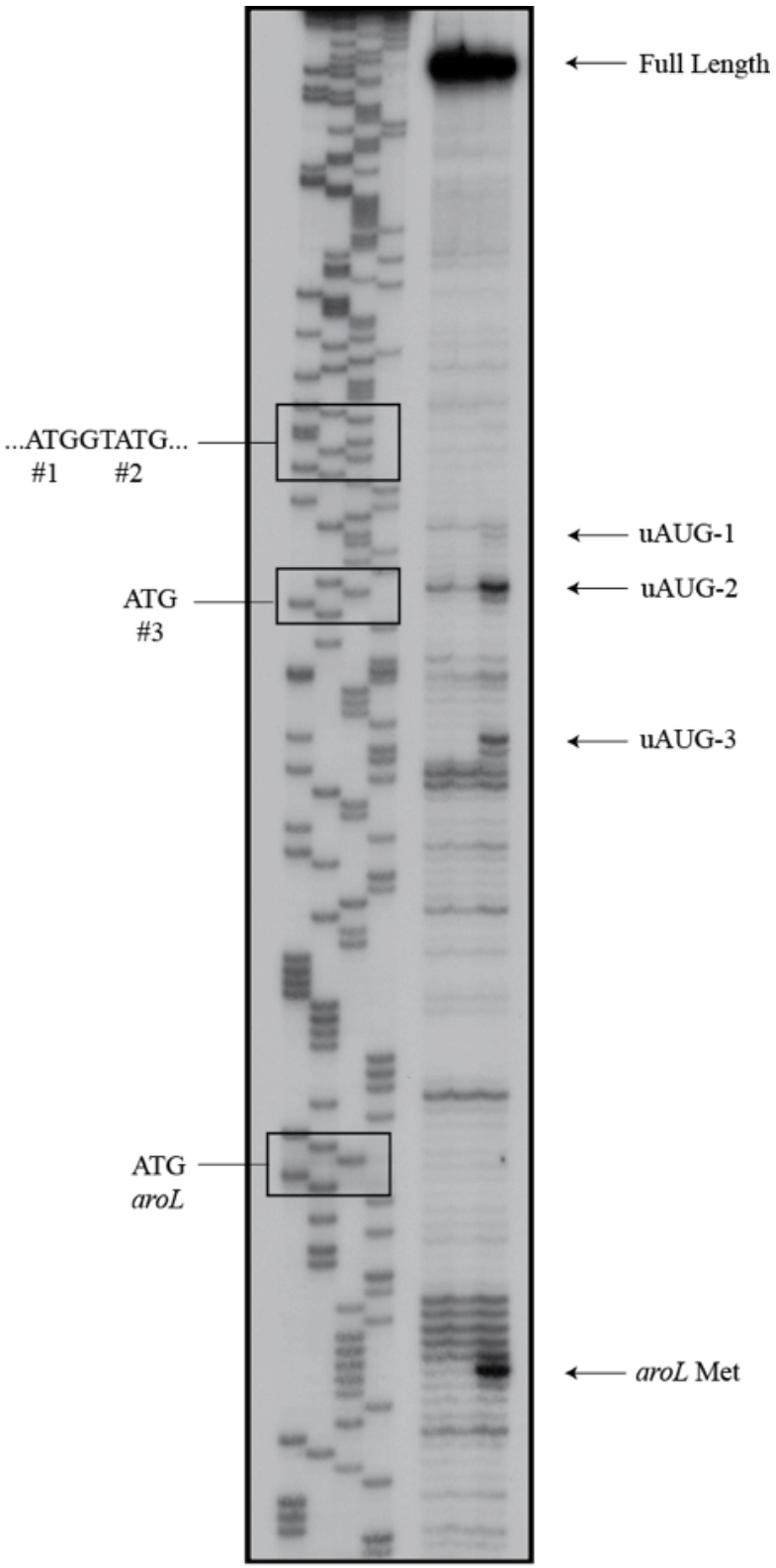
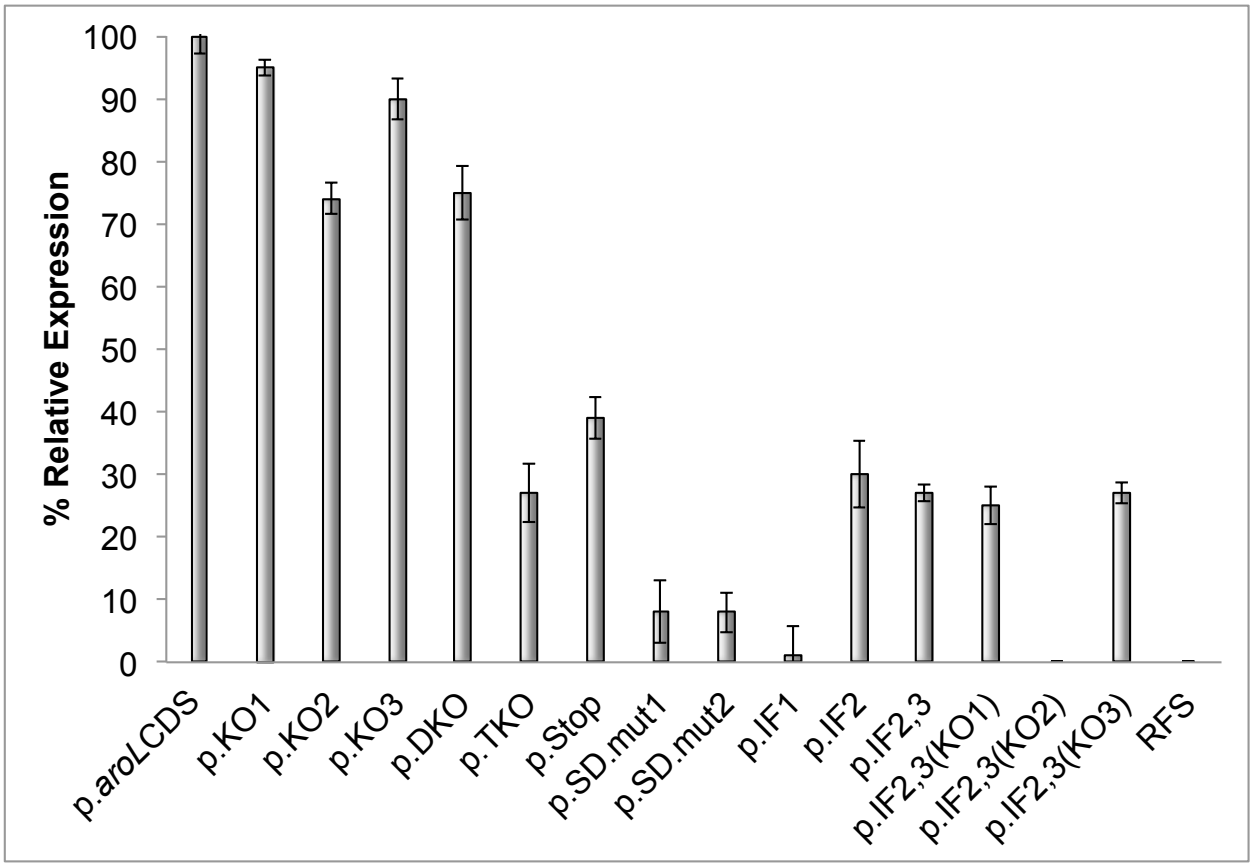

Full Length
...ATGGTATG...
#1 #2
uAUG-1
ATG
#3
uAUG-2
uAUG-3
ATG
aroL
aroL Met
G A T C
1
2
3
30S
-
+
+
tRNA fMET
+
-
+

Figure 1-3. β-galactosidase activity of the *aroL* leader + 16 codons (p.*aroL*.CDS), with various AUG knockouts or reading frames of the uAUGs fused to a *lacZ* reporter gene. Fusion constructs are depicted in Fig. 1-S1 and described in Table 1-1. *E. coli* RFS859 was the host strain for expression studies.


% Relative Expression
100
90
80
70
60
50
40
30
20
10
0
p.aroLCDS
p.KO1
p.KO2
p.KO3
p.DKO
p.TKO
p.Stop
p.SD.mut1
p.SD.mut2
p.IF1
p.IF2
p.IF2,3
p.IF2,3(KO1)
p.IF2,3(KO2)
p.IF2,3(KO3)
RFS

Identification of sequences needed for ribosome binding and expression from uAUG-2

Examination of the sequence upstream to uAUG-2 did not identify a canonical SD sequence, suggesting that ribosome binding occurs via a SD-independent mechanism. We initially ruled out the mRNA's GAUGG sequence, immediately upstream to uAUG-2 and containing uAUG-1, as acting as a SD sequence because substituting two G nucleotides to C (underlined in GAUGGUAUG, referred to as p.IF2,3.mut1) resulted in a 4-fold increase in expression from uAUG-2 (Table 1-2) and a stronger uAUG-2 toeprint signal (Fig. 1-4), indicating that this G-rich sequence does not function as a canonical SD sequence for uAUG-2 expression.

In an effort to localize the sequence(s) contributing to ribosome binding and expression from uAUG-2, translational fusions of the uAUG-2/3 reading frame to *lacZ* were constructed in which the 5'UTR was deleted progressively inward from the 5' end. The 5' end points of the deletion constructs, in relation to the transcriptional and translational starts, are shown in Fig. 1-1. Expression from the uAUG-1 reading frame was not measured in these constructs because of the lack of significant translational activity (Fig. 1-3) and weak 30S subunit binding (Fig. 1-2). Deleting the 5'UTR to +18 (p.Δ18) resulted in minimal loss of expression (~10%) compared to activity with the full 5'UTR (Fig. 1-5A). Deletion to +36 (p.Δ36) resulted in 70% reduction, and deletion to +45 (p.Δ45) and +53 (p.Δ53) led to a complete loss of expression from uAUG-2 (Fig. 1-5A). Interestingly, complete deletion of the leader sequence upstream of uAUG-2 (p.LLuATG-2) retained 40% of the activity measured with the full leader sequence present (Fig. 1-5A), suggesting that the truncated *aroL* mRNA, with a 5'-terminal uAUG-2, was recognized by ribosomes and translated as a leaderless message.

Toeprint assays revealed that 30S subunit binding to uAUG-2 decreased as the leader was deleted inward from the 5' end, with complete loss of binding at the +45 deletion end point (p.Δ45; Fig. 1-5B), further supporting the expression data that nucleotides important for ribosome binding and translation reside between +18 and +45 (corresponding to 19-46 nucleotides upstream of uAUG-2). Unexpectedly, no binding to uAUG-3 was observed in these toeprint assays.

Table 1-2. β-galactosidase activity of the *aroL* leader with segments of the leader sequence mutated (p.IF2,3mut1-14) or with nucleotide insertions (p.IF2,3.add1, p.IF2,3.add2, p.IF1.add1) or deletions (p.IF2,3.del1-7, p.IF1.del1) between the U rich region and uAUG-1. Expression of the unmutated leader was set to 100% with uAUG-2 (p.IF2,3) or uAUG-1 (p.IF1) in frame with *lacZ.* Fusion constructs are described in Table 1-1. *E. coli* RFS859 was the host strain for expression studies.

Plasmid Construct	% Expression
p.IF2,3	100 ± 8
p.IF2,3.mut1	400 ± 3
p.IF2,3.mut2	72 ± 4
p.IF2,3.mut3	110 ± 4
p.IF2,3.mut4	67 ± 4
p.IF2,3.mut5	70 ± 7
p.IF2,3.mut6	62 ± 6
p.IF2,3.mut7	98 ± 7
p.IF2,3.mut8	120 ± 7
p.IF2,3.mut9	150 ± 2
p.IF2,3.mut10	96 ± 2
p.IF2,3.mut11	98 ± 6
p.IF2,3.mut12	150 ± 20
p.IF2,3.mut13	30 ± 2
p.IF2,3.mut14	7 ± 5
p.IF2,3.del1	4 ± 4
p.IF2,3.del2	9 ± 4
p.IF2,3.add1	6 ± 5
p.IF2,3.add2	1 ± 8
p.IF2,3.del3	52 ± 2
p.IF2,3.del4	58 ± 10
p.IF2,3.del5	65 ± 5
p.IF2,3.del6	2 ± 1
p.IF2,3.del7	72 ± 5
p.IF1	100 ± 3
p.IF1.add1	90 ± 7
p.IF1.del1	100 ± 4
RFS	0 ± 4

Figure 1-4. Primer extension inhibition (toeprint) analysis comparing 30S ribosomal subunit binding to uAUG-1 and uAUG-2 of the *aroL* upstream leader region in the presence of the two G to C substitutions (underlined) within the sequence GAUGGUAUG. Assays were performed in the presence (+) or absence (-) of tRNA and 30S subunits, as indicated. Toeprint signals for 30S subunits bound to the uAUGs are indicated by arrows and occur at the expected +16 position (with the A of AUG as +1). Assays in lanes 1-3 were performed with the *aroL* WT leader (p.IF2,3) construct; lanes 4-5 were performed with p.IF2,3.mut1 construct.

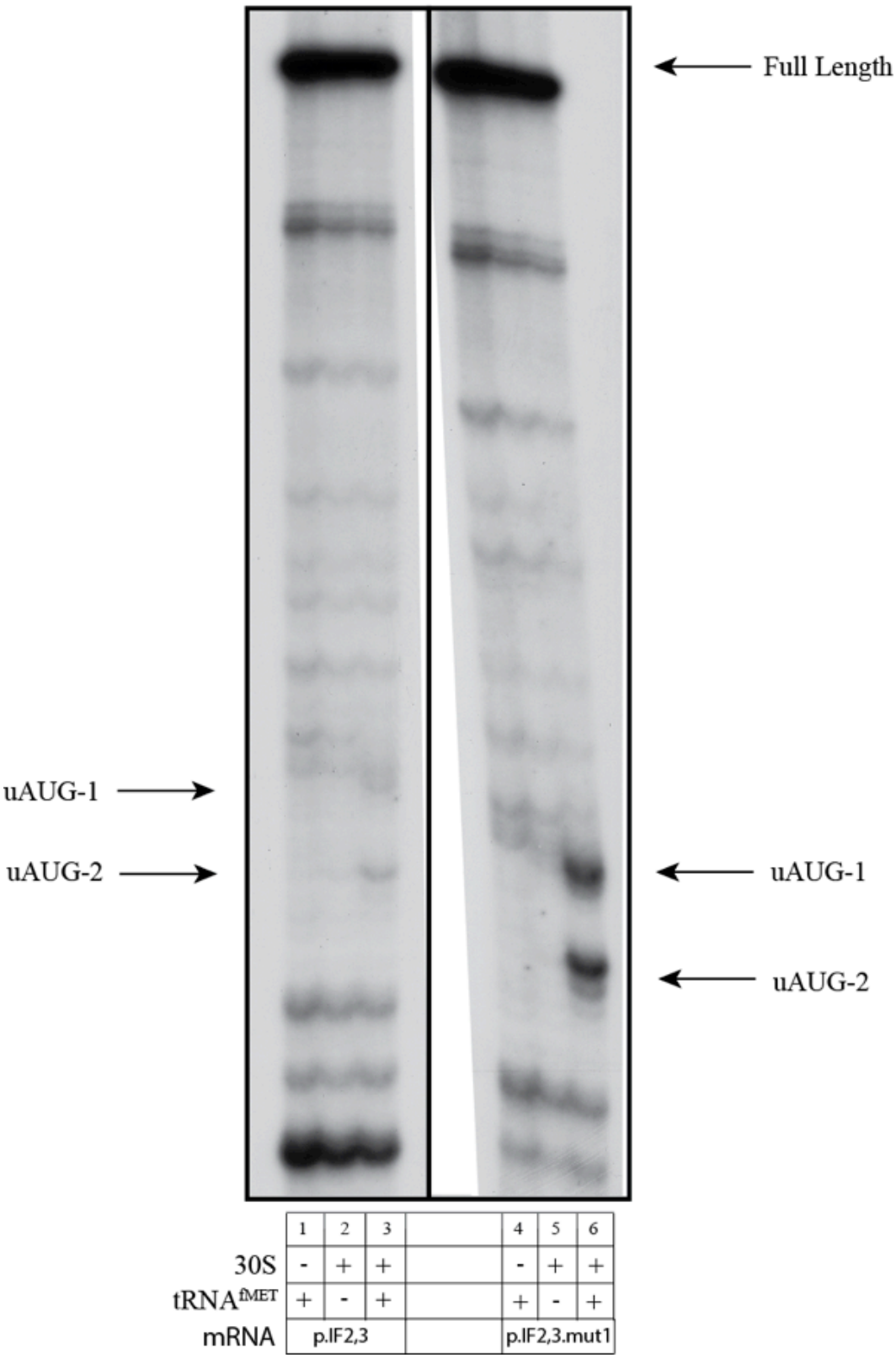
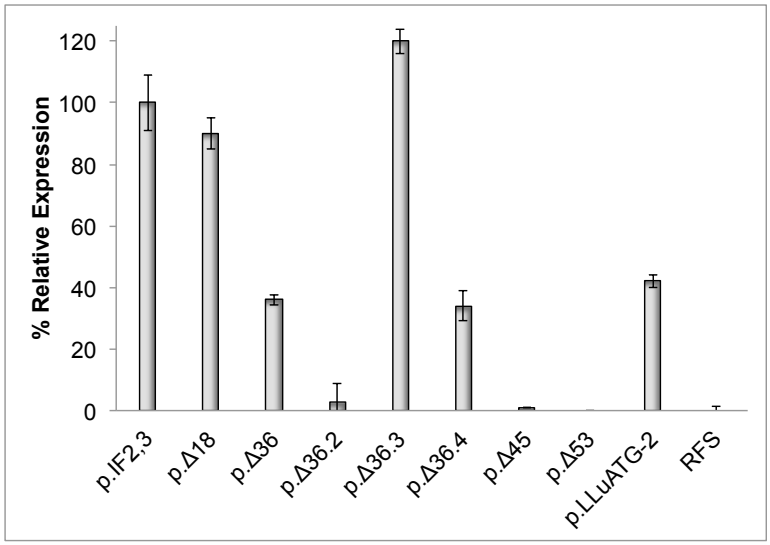
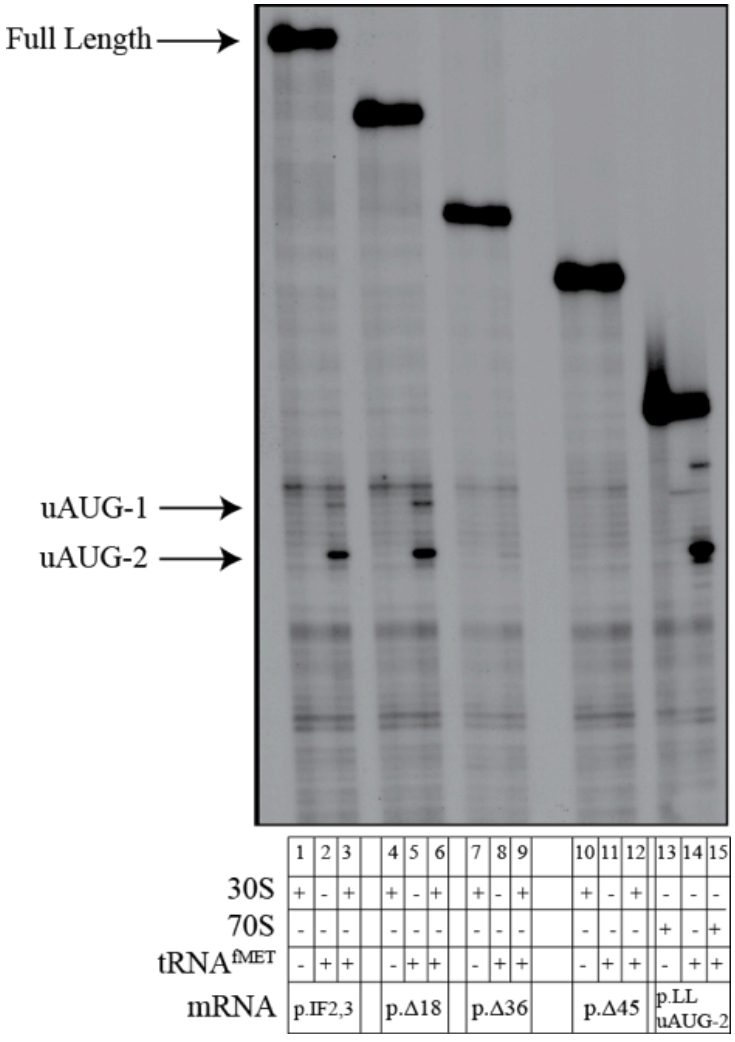

Full Length
uAUG-1
uAUG-2
uAUG-1
uAUG-2
1 2 3 4 5 6
30S - + + - + +
tRNA^fMET + - + + - +
mRNA p.IF2,3 p.IF2,3.mut1

Figure 1-5. β-galactosidase activity (A) and primer extension inhibition (toeprint) assays (B) of the *aroL* leader (uAUGs 2,3 in frame with *lacZ*) with leader deletions. Transcriptional start sites of deletions are shown in Fig. 1-1. Deletions constructs are described in Table 1-1. A. *E. coli* RFS859 was the host strain for expression studies. B. Toeprint assays were performed in the presence (+) or absence (-) of tRNA, 30S subunits or 70S ribosomes and *aroL* WT (full length leader) or partially deleted leader mRNAs (described in Table 1-1), as indicated. Toeprint signals for 30S subunits bound to the uAUGs are indicated by arrows and occurred at the expected +16 position (with the A of AUG as +1). Lanes 1-3 were performed with the WT full length leader (p.IF2,3); lanes 4-6 with the p.Δ18 shortened construct; 7-9 with the p.Δ36 shortened construct; lanes 10-12 with the p.Δ45 shortened construct; and lanes 13-15 with the p.LLuATG-2 construct.

A.

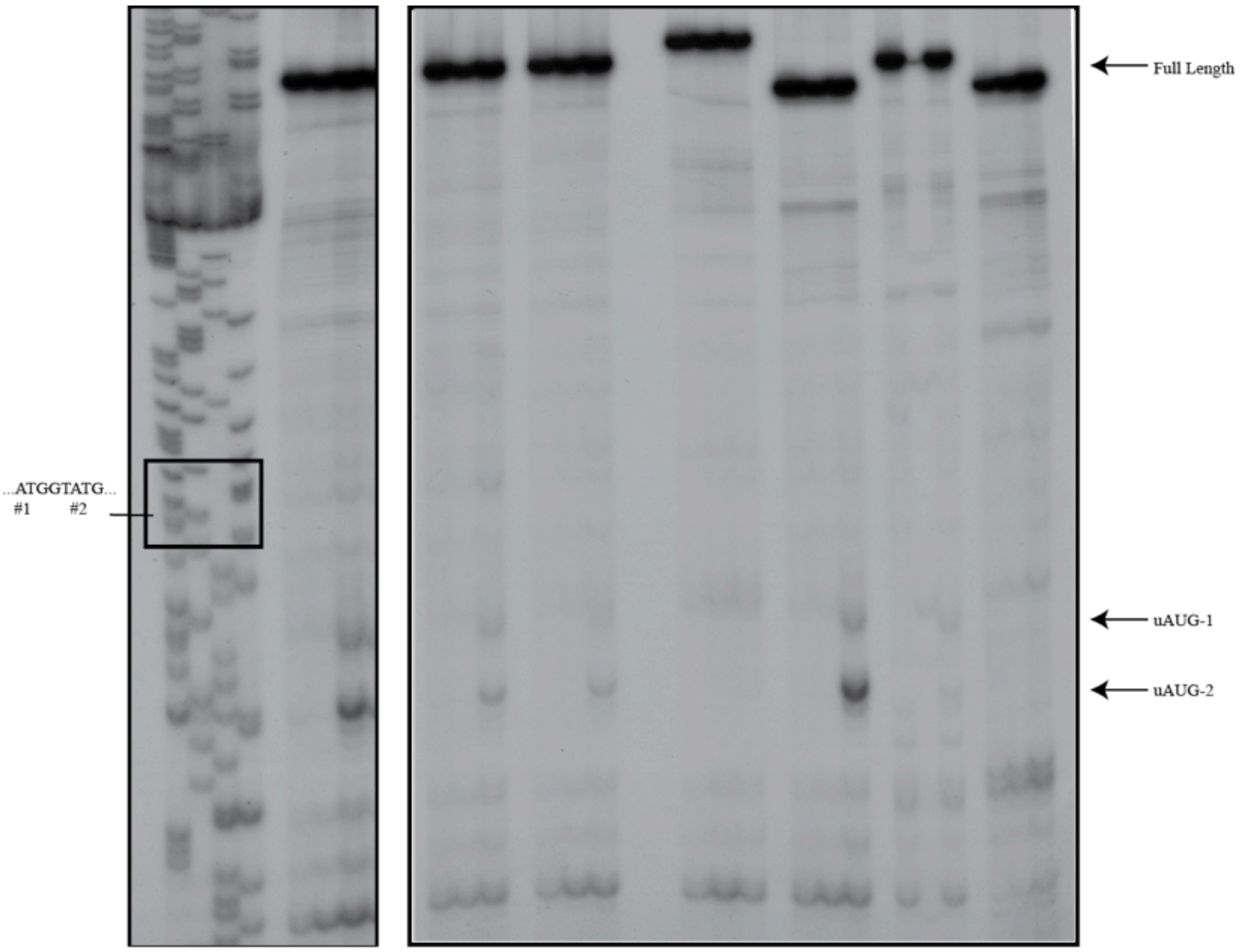

% Relative Expression
120
100
80
60
40
20
0
p.IF2,3
p.Δ18
p.Δ36
p.Δ36.2
p.Δ36.3
p.Δ36.4
p.Δ45
p.Δ53
p.LLuATG-2
RFS

B.


Full Length
uAUG-1
uAUG-2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
30S + - + + - + + - + + - + - - -
70S - - - - - - - - - - - - + - +
tRNA^fMET - + + - + + - + + - + + - + +
mRNA p.IF2,3 p.Δ18 p.Δ36 p.Δ45 p.LL uAUG-2

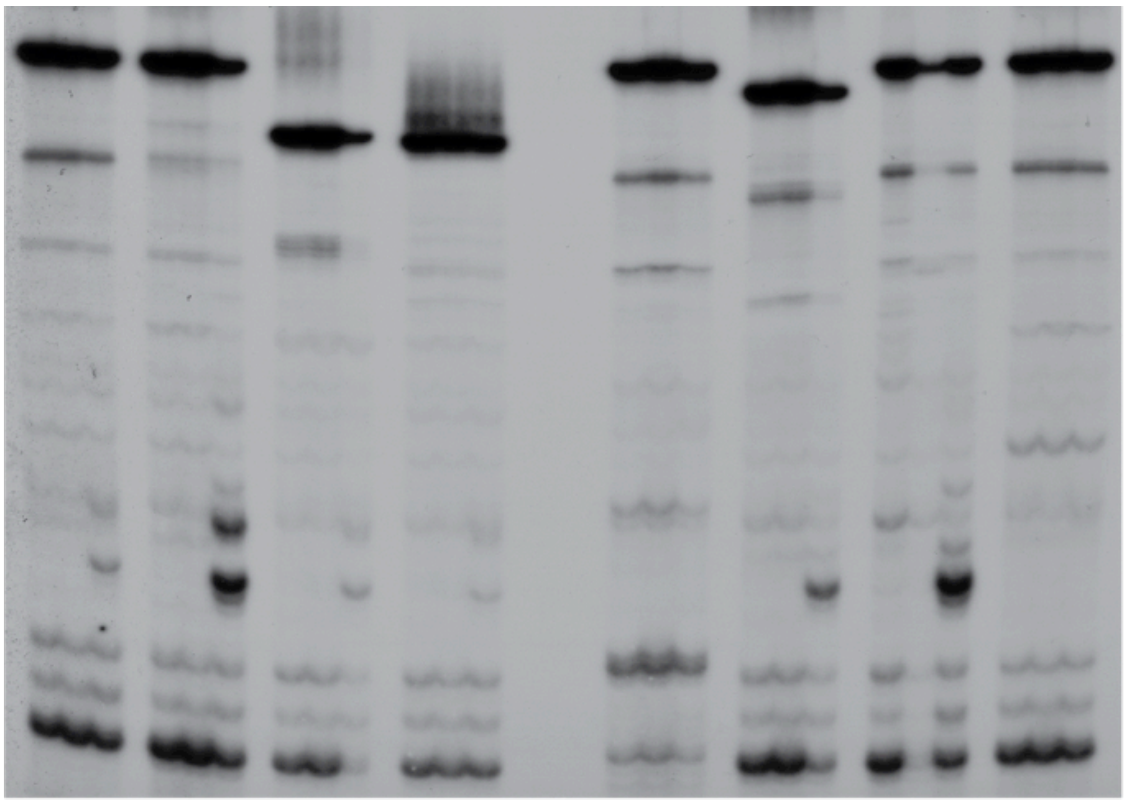
To further identify the sequences responsible for translation from uAUG-2, nucleotide substitutions were directed to the 5'UTR of uAUG-2-*lacZ* fusions where 3-5 nucleotide segments were mutated to their complement, with care not to introduce new AUG triplets. Mutations are listed in Table 1-1. β-galactosidase assays performed on cells expressing these fusions (Table 1-2) suggest that the sequence from position +26 to +37 (relative to the transcriptional start; Fig. 1-1) is responsible for translation from uAUG-2; within this region, mutation of nucleotide segments AAGU, UUUUU, UCU, and UUAC to their complements individually reduced expression from uAUG-2 by 30%, 33%, 20%, and 35%, respectively (Table 1-2, p.IF2,3.mut2, p.IF2,3.mut4, p.IF2,3.mut5, p.IF2,3.mut6, respectively). Mutation of the combined segments AAGU and UUUUUUCUUUAC as a whole reduced uAUG-2 expression by 93% (Table 1-2, p.IF2,3.mut14). In addition to mutating the combined leader sequences via substitutions, we also introduced site-directed deletions and measured expression from uAUG-2 (listed in Table 1-1). Deletion of AAGU (p.IF2,3.del7) or the combined segments AAGU and UUUUUUCUUUAC (p.IF2,3.del2) (Table 1-2) resulted in a 30% and 92% loss of expression, respectively, from uAUG-2, comparable to the loss measured with the substitution mutations. Toeprint assays showed that upon substituting (Fig. 1-6, compare lanes 1-3 to lanes 4-6 and 7-9) or deleting (Fig. 1-7, compare lanes 1-3 to lanes 10-12) the U-rich region from the *aroL* mRNA, 30S subunit binding decreased, also supporting the expression data. Unexpectedly, 30S subunit binding to uAUG-2 did not change upon deletion of the four-nucleotide sequence AAGU (Figure 1-8, compare lanes 1-3 to lanes 4-6), and the AAGU substitution resulted in an increase in 30S subunit binding to the mRNA (Fig. 1-7, compare lanes 1-3 to lanes 4-6). Regardless of this ribosome binding anomaly (see Discussion), these results clearly implicate the AAGU and the U-rich region of the *aroL* mRNA in ribosome binding and/or translation from uAUG-2.

Figure 1-6. Primer extension inhibition (toeprint) analysis comparing 30S ribosomal subunit binding to uAUG-1 and uAUG-2 of the *aroL* upstream leader region in the presence of substitution mutations (p.IF2,3.mut13, p.IF2,3.mut14, p.DR.mut2), nucleotide insertions (p.IF2,3.add1), and nucleotide deletions (p.IF2,3.del3, p.DR.del2). Descriptions of the mRNA constructs and the corresponding mutations are listed in Table 1-1. Assays were preformed in the presence (+) or absence (-) of tRNA and 30S subunits, as indicated. Toeprint signals for 30S subunits bound to the uAUG's are indicated by arrows and occurred at the expected +16 position (with the A of AUG as +1). An *aroL* leader DNA sequencing ladder is on the left.



	T A C G	1	2	3`		4	5	6	7	8	9		10	11	12	13	14	15	16	17	18	19	20	21
30S		-	+	+		-	+	+	-	+	+		-	+	+	-	+	+	-	+	+	-	+	+
$tRNA^{fMET}$		+	-	+		+	-	+	+	-	+		+	-	+	+	-	+	+	-	+	+	-	+
mRNA		p.IF2,3				p.IF2,3.mut14			p.IF2,3.mut13				p.IF2,3.add1			p.IF2,3.del3			p.DR.mut2			p.DR.del2		

Figure 1-7. Primer extension inhibition (toeprint) analysis comparing 30S ribosomal subunit binding to uAUG-1 and uAUG-2 of the *aroL* upstream leader region in the presence of substitution mutations (p.IF2,3.mut2, p.IF2,3.KO2), nucleotide insertions (p.DR.add1), and nucleotide deletions (p.IF2,3.del2, p.IF2,3.del4, p.∆GTATG.2, p.∆GTATG.3). Descriptions of the mRNA constructs and the corresponding mutations are listed in Table 1-1. Assays were performed in the presence (+) or absence (-) of tRNA and 30S subunits, as indicated. Toeprint signals for 30S subunits bound to the uAUGs are indicated by arrows and occurred at the expected +16 position (with the A of AUG as +1).



	1	2	3		4	5	6		7	8	9		10	11	12		13	14	15		16	17	18		19	20	21		22	23	24
30S	-	+	+		-	+	+		-	+	+		-	+	+		-	+	+		-	+	+		-	+	+		-	+	+
$tRNA^{fMET}$	+	-	+		+	-	+		+	-	+		+	-	+		+	-	+		+	-	+		+	-	+		+	-	+
mRNA	p.IF2,3				p.IF2,3. mut2				p.IF2,3.del4				p.IF2,3.del2				pDR.add1				p.ΔGTATG.3				p.ΔGTATG.2				p.IF2,3.KO2		

Interestingly, most of the sequences shown above to be important for expression are not present in the Δ36 construct (Fig. 1-1), yet we observed 30% expression from uAUG-2 in the absence of these sequences (Fig. 1-5A). We reasoned that, in the absence of the U-rich sequence, ribosomes recognized another sequence as a consequence of the shortened 5'UTR. To identify the sequence responsible, we introduced substitutions to the shortened leader and measured expression from uAUG-2 (Table 1-1, Fig. 1-5A). Substituting CAAUC (p.Δ36.2), representing positions +2 to +6 of the truncated sequence, to its complement abolished expression from uAUG-2. Mutating AAUC (+38- +41) in the full-length leader, however, did not affect expression from uAUG-2 (Table 1-2, p.IF2,3.mut7), suggesting that the ability of CAAUC to recruit ribosomes resulted from its position at the 5' end of the truncated mRNA.

Proximity of the upstream signals to uAUG-2 is important for ribosome binding and expression

In an effort to address whether there is a positional or spacing requirement of the upstream U-rich sequence for expression from uAUG-2, we inserted or deleted 5 or 15 nucleotides between the U-rich sequence and uAUG-2. As shown in Table 1-2, the position of the upstream U-rich sequence is important for efficient expression from uAUG-2. Deleting 5 (p.IF2,3.del3) or 15 (p.IF2,3.del4) nucleotides between the U-rich sequence and uAUG-2 reduced expression by 50%; increasing the distance by 5 (p.IF2,3.add1) or 15 (p.IF2,3.add2) nucleotides reduced expression by 95%. The intensity of toeprint signals mapping to uAUG-2 were consistent with the expression data. Although the 5- and 15-base deletions produced toeprint signals to uAUG-2 (Fig. 1-6, compare lanes 1-3 to lanes 13-15; Fig. 1-7, compare lanes 1-3 to lanes 7-9), the 5-base insertion resulted in a loss of toeprint signal (Fig. 1-6, compare lanes 1-3 to lanes 10-12).

Interestingly, these insertion and deletion mutations did not shift ribosome binding or expression from uAUG-2 to uAUG-1 or uAUG-3. For example, deleting 15 nucleotides between the upstream U-rich sequence and uAUG-2 of the mRNA moves uAUG-3 into the position normally occupied by uAUG-2, relative to the U-rich sequence, yet we did not detect any expression from uAUG-3. Individually mutating uAUG-2

(p.IF2,3.del6) or uAUG-3 (p.IF2,3.del5) allowed us to determine the contribution of each start codon to the measured level of expression. Even though the 15-nucleotide deletion reduced expression from uAUG-2 two-fold, translation still initiated exclusively from uAUG-2 over uAUG-3 (~50% and 0%, respectively; Table 1-2) and ribosomes continued to show a binding preference for uAUG-2 over uAUG-3 (data not shown). Furthermore, insertion of 5 nucleotides between the upstream U-rich sequence and uAUG-1 places uAUG-1 in uAUG-2's normal position relative to the upstream sequence. If proximity of the upstream sequence to uAUG-2 were the determining factor for selection of the translational start site, then shifting uAUG-1 into the position of uAUG-2 would shift translation to uAUG-1. We assessed expression from the uAUG-1 (p.IF1.add1) or uAUG-2 (p.IF2,3.add1) reading frames by *lacZ* fusions. Although the 5-nucleotide insertion decreased expression from uAUG-2 by 95%, there was not a concomitant increase in expression from uAUG-1 (Table 1-2 compare p.IF2,3 to p.IF2,3.add1 and p.IF1 to p.IF1.add1). These results indicate that we were not able to manipulate ribosomal preference from uAUG-2 to uAUG-1 or uAUG-3 by altering their distance to the upstream U-rich sequence and suggest that additional determinants for start site selection might reside downstream from uAUG-2 or in sequences flanking the uAUGs.

Ribosome binding and expression from uAUG-2 involves additional sequence downstream of uAUG-2

To determine whether sequences downstream of uAUG-2 might play a role in uAUG-2 expression, the 11 codons between uAUG-2 and the fusion site to *lacZ* in p.IF2,3 were divided into four downstream regions (DR1 → DR4) and mutated individually to their complementary sequences (Fig. 1-1) so as not to introduce additional start or stop codons, and assayed for LacZ activity. Downstream sequences important for expression from uAUG-2 were localized within or overlapping DR2 and DR4 (corresponding to positions +13-21 and +29-36, respectively, with the A of uAUG-2 as +1) (Fig. 1-1). Mutations to DR1 (p.DR.mut1) or DR3 (p.DR.mut3) had minimal effect on expression (Table 1-3). Mutations to DR2 (p.DR.mut2) or DR4 (p.DR.mut4) reduced expression from uAUG-2 by ~85% or 65%, respectively (Table 1-3). Consistent with the

expression data, mutation of DR2 resulted in a loss of toeprint signal for 30S subunit binding to uAUG-2 (Fig. 1-6, compare lanes 1-3 to lanes 16-18).

We altered the distance between uAUG-2 and DR2 to determine if position of the downstream sequence was important for proper uAUG-2 utilization. Deletion of codons 1 and 3 from DR1 nearly abolished expression from, and 30S subunit binding to, uAUG-2 (p.DR.del2, Table 1-3; Fig. 1-6, compare lanes 1-3 to lanes 19-21). Because the DR1 substitution mutations did not impact expression (Table 1-3, p.DR.mut1), deletion of the two DR1 codons suggested a positional requirement of the downstream sequences for uAUG-2 utilization. A similar reduction in uAUG-2 expression (p.DR.add2, Table 1-3) and ribosome binding (Fig. 1-8, compare lanes 1-3 to lanes 10-12) was observed when we inserted 4 additional codons between uAUG-2 and DR2, also suggesting a positional requirement for DR2. We also measured expression from uAUG-1 after deletion of codons 1 and 3 from DR1, reasoning that deletion of the 6 nucleotides might shift the downstream element to a better position for expression from uAUG-1 (p.DR.del1). Unexpectedly, the reduction in expression from uAUG-2 with this mutation did not correspond to an increase in expression from, or 30S subunit binding to, uAUG-1, even though uAUG-1 was essentially in the natural position of uAUG-2 (Table 1-3; Fig. 1-6, compare lanes 1-3 to lanes 19-21). Unexpectedly, combination of the 5-nucleotide insertion between the upstream U-rich sequences and uAUG-1 (described earlier) with deletion of codons 1 and 3 downstream of uAUG-2 did not increase expression from uAUG-1 (p.DR.add1, Table 1-3; Fig. 1-7, compare lanes 1-3 to lanes 13-15) even though this places uAUG-1 in the position of uAUG-2 relative to both the upstream sequences and DR2. These data suggest that proximity of the upstream and downstream sequences to uAUG-2 are not the only determinants for start site selection.

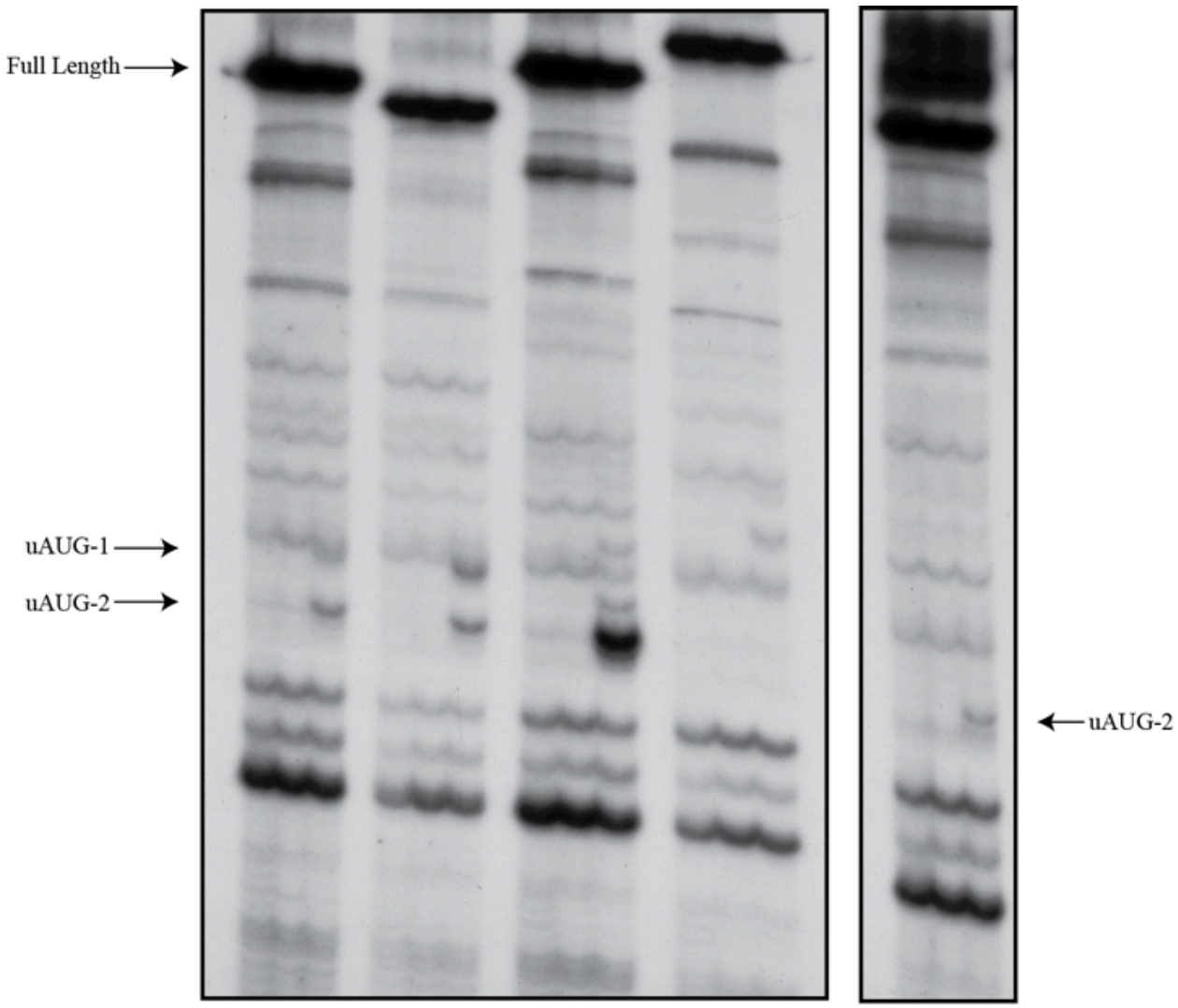
Table 1-3. β-galactosidase activity of *aroL* leader with regions downstream of uAUGs mutated (p.DR.mut1-4), or with nucleotide insertion (p.DR.add1, p.DR.add2) or deletions (p.DR.del1, p.DR.del2) between the uAUGs and downstream region 2 (DR2). Expression of the unmutated leader was set to 100% with uAUG-2 (p.IF2,3) or uAUG-1 (p.IF1) in frame with *lacZ.* Fusion constructs are described in Table 1-1. Mutated regions downstream of uAUGs are depicted in Fig. 1-1. *E. coli* RFS859 was the host strain for expression studies.

Plasmid Construct	% Expression
p.IF2,3	100 ± 3
p.DR.mut1	97 ± 10
p.DR.mut2	17 ± 2
p.DR.mut3	72 ± 5
p.DR.mut4	35 ± 10
p.DR.del2	3 ± 9
p.DR.add2	5 ± 8
p.IF1	100 ± 9
p.DR.del1	66 ± 7
p.DR.add1	60 ± 4
RFS	0 ± 1

The sequence surrounding uAUG-1 and uAUG-2 contributes to start site selection

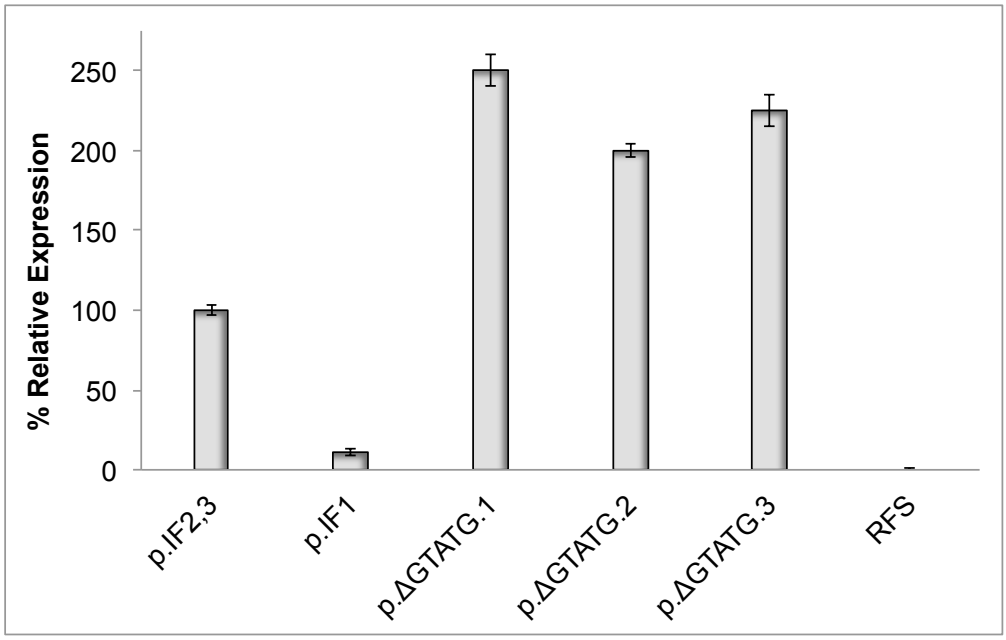
We next directed our attention to the nucleotide context surrounding uAUG-1 and -2 for a possible role in expression and/or start site selection. In an attempt to direct translation initiation to uAUG-1, we deleted the sequence GUAUG (containing uAUG-2; Fig. 1-1), thereby placing uAUG-1 as the start codon of the uAUG-2/3 ORF (p.ΔGTATG.1); this placement maintains the natural distance between uAUG-1 and the upstream U-rich sequence and brings the downstream region into its normal position relative to uAUG-2. This change resulted in increased expression measured from uAUG-1 compared to when in its native context (compare p.IF1 to p.ΔGTATG.1in Fig. 1-9). Next, using the GUAUG deletion construct (p.ΔGTATG.1), we inserted 5 (p.ΔGTATG.2) or 10 (p.ΔGTATG.3) nucleotides between the upstream U-rich sequence and uAUG-1. A 5-nucleotide insertion would position uAUG-1 in the normal position of uAUG-2 relative to the U-rich and DR sequences and might be expected to increase expression. A 10-nucleotide insertion would shift uAUG-1 five nucleotides downstream of where uAUG-2 is normally positioned and might be expected to decrease expression (recall that when the spacing between the upstream U-rich sequence and uAUG-2 was increased by five nucleotides expression from uAUG-2 dropped to 5%). Surprisingly, increasing the spacing between the upstream U-rich sequence and uAUG-1 by 5 or 10 nucleotides had minimal effects on expression compared to the GUAUG deletion alone (Fig. 1-9). Toeprint assays of 30S subunit binding to uAUG-1 are consistent with the expression data (Fig. 1-7, lanes 16-21; Fig. 1-8, lanes 7-9). The GUAUG deletion appeared to relieve the spacing requirement of the upstream U-rich sequence and uAUG-1. The GUAUG sequence in the mRNA was inhibitory for expression from uAUG-1.

Figure 1-8. Primer extension inhibition (toeprint) analysis comparing 30S ribosomal subunit binding to uAUG-1 and uAUG-2 of the *aroL* upstream leader region in the presence of substitution mutations (p.IF2,3.KO1), nucleotide insertions (p.DR.add2), and nucleotide deletions (p.IF2,3.del7, p.ΔGTATG.3). Descriptions of the mRNA constructs and the corresponding mutations are listed in Table 1-1. Assays were performed in the presence (+) or absence (-) of tRNA and 30S subunits, as indicated. Toeprint signals for 30S subunits bound to the uAUG's are indicated by arrows and occurred at the expected +16 position (with the A of AUG as +1).



	1	2	3		4	5	6		7	8	9		10	11	12		13	14	15
30S	-	+	+		-	+	+		-	+	+		-	+	+		-	+	+
$tRNA^{fMET}$	+	-	+		+	-	+		+	-	+		+	+	+		+	-	+
mRNA	p.IF2,3				p.IF2,3.del7				p.ΔGTATG.3.				p.DR.add2				p.IF2,3.KO1		

Figure 1-9. β-galactosidase activity of the *aroL* leader with GUAUG-2 deleted (p.ΔGTATG.1) and 5 (p.ΔGTATG.2) or 10 (p.ΔGTATG.3) nucleotide insertions between the U-rich region and the uAUG. Expression of the *aroL* leader with uAUG-2 in frame with *lacZ* (p.IF2,3) was set to 100%. Descriptions of the constructs are presented in Table 1-1. *E. coli* RFS859 was the host strain for expression studies.


% Relative Expression
250
200
150
100
50
0
p.IF2,3
p.IF1
p.ΔGTATG.1
p.ΔGTATG.2
p.ΔGTATG.3
RFS

To address whether the uAUG-1 and uAUG-2 start codons themselves were influencing ribosome recognition and expression from each other, we mutated uAUG-2 reasoning that the mere presence of the adjacent, competing uAUG-2 start codon was inhibiting expression from uAUG-1. However, mutation of uAUG-2 to AUC did not result in increased ribosome recognition of uAUG-1 (Fig. 1-7, compare lanes 1-3 to lanes 22-24). Likewise, mutation of uAUG-1 to AUC had no effect on translation from uAUG-2 (p.IF2,3(KO1), Fig. 1-3; Fig. 1-8, compare lanes 1-3 to lanes 13-15). These results suggest that, if acting as start codons, uAUG-1 does not detract from expression of uAUG-2, and uAUG-2 does not diminish expression from uAUG-1, and suggests that the sequence containing or surrounding uAUG-1 and uAUG-2 plays a role in ribosome recognition of the uAUGs and/or selection of the translational start site. This notion is further supported by the increase in expression from uAUG-2 upon substitution of the SD-like sequence described earlier (i.e., two G→C substitutions, underlined in the sequence GAUGGUAUG and containing uAUG-1 and -2, resulted in a four-fold increase in expression from uAUG-2; p.IF2,3.mut1, Table 1-2).

The upstream sequence elements required for uAUG-2 expression also promotes expression of other coding regions

To determine if the translation signals of the *aroL* upstream leader region support ribosome binding and expression outside of its native context, the *aroL* leader sequence from +1 to +66 was added upstream to gene-internal fragments of *E. coli*'s *aroL* (*aroL*IN, positions 127 to 276; DeFeyter and Pittard, 1986) and *tna* (*tna*IN, positions 382 to 531; Deeley and Yanofsky, 1982) fused to *lacZ*. Fusions were done such that either uAUG-1 (u1) or uAUG-2 (u2) was in frame with *lacZ*. These gene-internal fragments are not expected to contain sequences, signals or features associated with a translational start site; ribosome recognition and expression would relate directly to the added *aroL* leader sequence. For comparison, we also fused the *E. coli* SD-containing *lac* leader to these internal fragments. These constructs are described in Table 1-1.

The *aroL* 5'UTR fragment supported expression of both *aroL*IN and *tna*IN from uAUG-2 (p.*aroL*Ld(u2)-*aroL*IN and p.*aroL*Ld(u2)-*tna*IN, respectively, Table 1-4). Mutation of AAGU and UUUUUUCUUUAC to its complement within the *aroL* leader

(LdMut) reduced expression from uAUG-2 with both coding sequences to 5% (p.*aroL*LdMut(u2)-*aroL*IN and p.*aroL*LdMut(u2)-*tna*IN, Table 1-4). These results further support a role for the *aroL* leader sequence as a stand-alone signal for ribosome recruitment and translation. In its native context, expression from uAUG-1 was less than 10% of that measured from uAUG-2 (Fig. 1-3); however, uAUG-1 expressed *aroL*IN and *tna*IN at 38% and 67%, respectively, of that measured from uAUG-2, indicating that in the absence of the *aroL* downstream region (DR) ribosomes are less discriminatory in their choice of a start codon. Surprisingly, the *aroL* leader provided for significantly higher expression than observed with the SD-containing *lacZ* leader upstream to *aroL*IN and *tna*IN (Table 1-4).

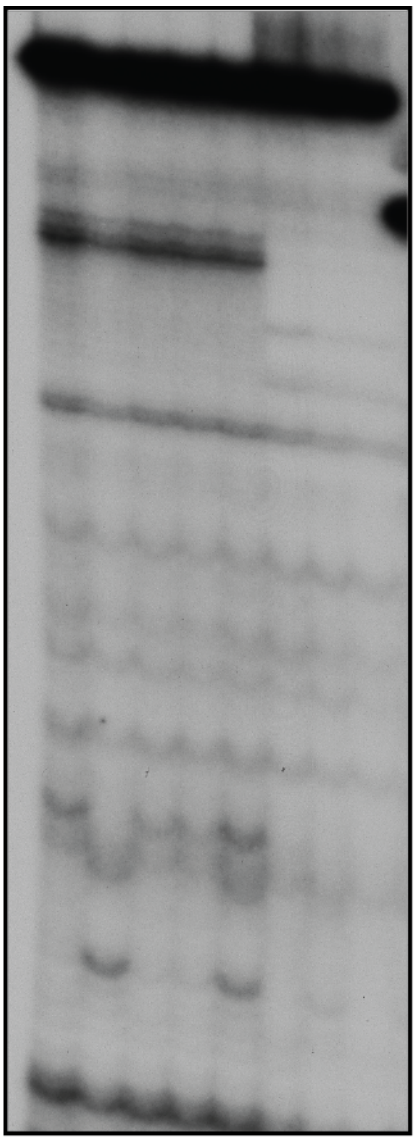
Ribosomal protein S1 is required for 30S subunit binding to the *aroL* uAUG-1 and uAUG-2

Based on the U-richness of *aroL*'s 5'UTR, we hypothesized that ribosomal protein S1 is required for 30S subunit binding and expression from uAUG-2. We repeated our toeprint assays using 30S subunits that were stripped of S1. As shown in Fig. 1-10, 30S subunits devoid of S1 were unable to produce toeprint signals at *aroL's* uAUGs with the wild type 5'UTR or the 5'UTR containing mutations to the U-rich sequence (Fig. 1-10, p.IF2,3, compare lanes 2 and 3; p.IF2,3.mut13, compare lanes 7 and 8). Addition of free S1 (5X over 30S subunit concentration) to reactions containing the S1-deficient 30S subunits restored toeprint signals to uAUG-1 and uAUG-2, indicating that the lack of 30S subunit binding was due to the absence of S1 (Fig. 1-10, p.IF2,3, compare lanes 2-5). Although these results suggest that S1 is required for ribosomal recognition of uAUG-1 and uAUG-2, we have yet to identify specific interactions between the mRNA and S1.

Table 1-4. β-galactosidase activity of the WT (with uAUG-1 [*aroL*Ld(u1)] or -2 [*aroL*Ld(u2)] in frame with *lacZ*) or mutated *aroL* leader (*aroL*LdMut(u2)) fused to internal fragments of the *tnaA*, and *aroL* genes (positions 382 to 531 of the *tnaA* coding region and 127 to 276 of the *aroL* coding region). The internal fragments were also fused to the *E. coli* SD-containing *lac* leader (*lac*Ld) for comparison. *E. coli* RFS859 was the host strain for expression studies.

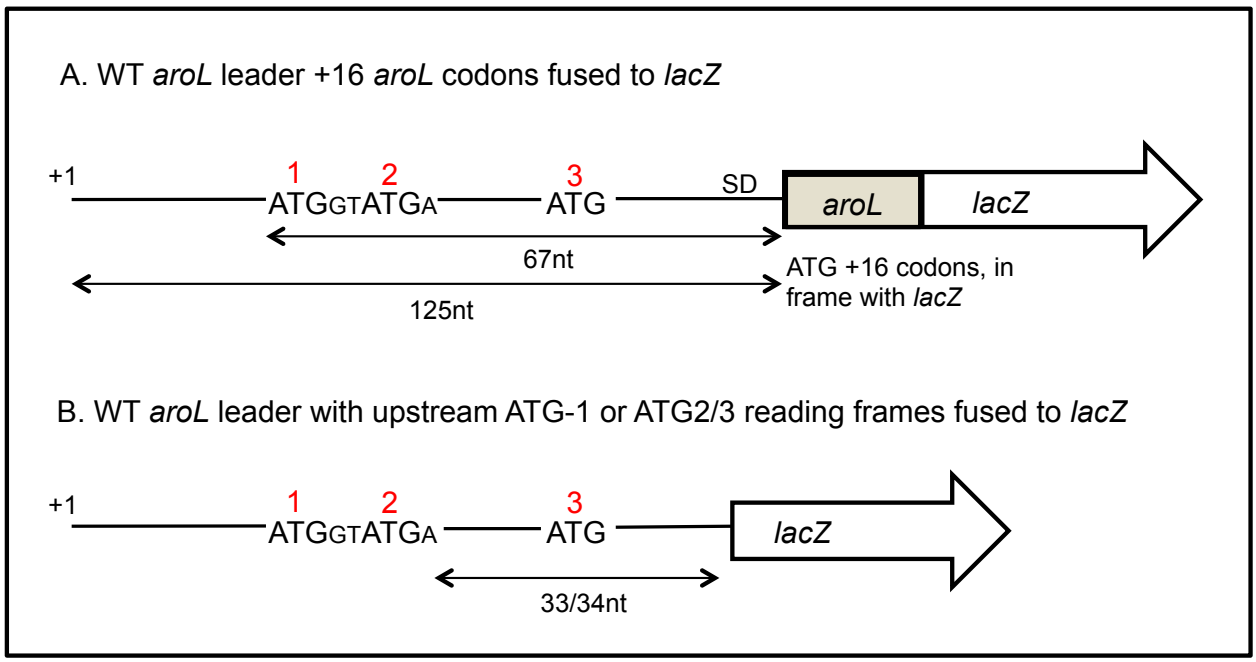
Plasmid Construct	Miller Units
p.IF2,3	4242 ± 200
p.IF2,3.mut14	515 ± 23
p.*aroL*Ld(u1)-*tna*IN	750 ± 20
p.*aroL*Ld(u2)-*tna*IN	1990 ± 50
*p.aroL*LdMut(u2)-*tna*IN	101 ± 10
p.*lac*Ld-*tna*IN	150 ± 25
p.*aroL*Ld(u1)-*aro*LIN	300 ± 80
p.*aroL*Ld(u2)-*aroL*IN	445 ± 70
*p.aroL*LdMut(u2)-*aroL*IN	21 ± 4
p.*lac*Ld-*aroL*IN	35 ± 3
RFS	5 ± 1

Figure 1-10. 30S subunits that lack ribosomal protein S1 do not bind to uAUG-1 or uAUG-2 of the *aroL* upstream leader on the WT (p.IF2,3) and mutant (p.IF2,3.mut13) *aroL* leader in primer extension inhibition (toeprint) assays.
Purified S1 was added to reactions containing the deficient 30S subunits at concentrations of 2X and 5X over 30S subunit concentration (lanes 4 and 5). Toeprint signals for 30S subunits bound to the uAUG's are indicated by arrows and occurred at the expected +16 position (with the A of AUG as +1).



	1	2	3	4	5	6	7	8
30S	-	+	-	-	-	-	+	-
30S (-S1)	-	-	+	+	+	-	-	+
$tRNA^{fMET}$	+	+	+	+	+	+	+	+
S1	-	-	-	2x	5x	-	-	-
mRNA	p.IF2,3					p.IF2,3.mut13		

Supplemental Figure 1-S1. **Translational fusion constructs used in assays to measure ribosome binding and expression.** Transcription of *aroL* was provided by a *lac* promoter. A. Translational fusion of the WT *aroL* leader and 16 codons of *aroL* to *lacZ* with the *aroL* ATG start codon in frame with *lacZ*. Site-directed mutagenesis was used to mutate uATG-1 (p.KO1), uATG-2 (p.KO2), uATG-3 (p.KO3), uATG-1 and uATG-2 (p.DKO), or all three uATGs (p.TKO) to ATC. An TGA stop codon was inserted 2 codons downstream from uATG-3 to make the (p.Stop) construct. B. Leader fragments in which the uATG-1 (p.IF1), uATG-2,3 (p.IF2,3), or uATG-2,3 (with uATG-2 [p.IF2,3(KO2)] or uATG-3 [p.IF2,3(KO3)] mutated to ATC) reading frames fused to the *lacZ* reporter gene. The TGA stop codon of the uATG-1 reading frame was mutated to TGC to assess contribution of the sequence within the uATG-2/3 reading frame on ribosome binding and expression.


A. WT *aroL* leader +16 *aroL* codons fused to *lacZ*
+1
1
2
3
ATGGTATGA
ATG
SD
aroL
lacZ
67nt
125nt
ATG +16 codons, in frame with *lacZ*
B. WT *aroL* leader with upstream ATG-1 or ATG2/3 reading frames fused to *lacZ*
+1
1
2
3
ATGGTATGA
ATG
lacZ
33/34nt

Discussion

Our results demonstrate ribosome binding to and translation from a previously unknown upstream start codon (uAUG-2) within the *aroL* "untranslated" leader (5'UTR) when canonical ribosome binding signals are removed. Expression of the uAUG-2-initiated ORF was dependent on sequences located several nucleotides upstream and downstream of the start codon, as well as nucleotides immediately adjacent to uAUG-2. The *aroL* coding sequence is translationally coupled to expression of a putative 21-amino acid peptide of unknown function that is encoded by the uAUG-2 ORF, with the *aroL* SD sequence needed for coupling. Non-traditional translation signals, such as the one described in this paper, are likely to be involved in ribosome recruitment to translational start sites for a number of genes that are reported to lack canonical, SD-like translation signals (Chang et al., 2006).

Contribution of the upstream sequence (US) to expression

Deletion or substitution of *aroL*'s upstream U-rich sequence (UUUUUUCUUUAC), or portions thereof, led to a significant decrease in ribosome binding and expression from uAUG-2. This U-rich sequence also showed position specificity; altering its spacing to uAUG-2 disrupted ribosome binding and expression from uAUG-2. Our 5- and 15-nucleotide additions and deletions were introduced at different positions on the mRNA, but yielded the same trend in binding and expression from uAUG-2. Deletion of the U-rich sequence brings a secondary U-rich region, located further upstream (Fig. 1-1), into position for possible expression; however, this region did not restore ribosome binding or expression, suggesting that differences in the sequence or its position were not suitable for expression from uAUG-2. Also, deletion of this secondary U-rich region (p.Δ18 construct; Table 1-1; Fig. 1-1) resulted in minimal reduction in ribosome binding and expression from uAUG-2, suggesting it does not normally play a role in uAUG-2 expression.

Nucleotide substitution or deletion of the AAGT sequence (positions +18 → +21 of the *aroL* untranslated leader) resulted in a 30% reduction in uAUG-2 expression. Toeprint assays, however, showed increased ribosome binding to uAUG-2 in the

presence of the AAGT mutation (i.e., strongly enhanced in the substitution construct, mildly enhanced with the deletion). Although the AAGU sequence impacted expression, it is not clear how it affected ribosome binding and/or translation initiation or how our *in vitro* observations relate to *in vivo* expression levels.

Addition of a portion of *aroL*'s 5'UTR, corresponding to +1 through uAUG-2, upstream of gene fragments internal to the *aroL* or *tna* coding regions resulted in translation from uAUG-2. Expression from these gene-internal fragments, presumably devoid of any translation initiation signals, suggests strongly that this region of *aroL*'s 5'UTR contains stand-alone translation signals that recruit ribosomes to the uAUG-2 start codon and the U-rich region is required for this activity. Although the majority of translation initiated from uAUG-2, substantial expression occurred from uAUG-1, suggesting that ribosomes were less discriminatory in their start site selection when the sequence normally downstream of uAUG-2 was missing. Surprisingly, the *aroL* 5'UTR fragment provided for substantially higher expression of the *aroL* and *tnaA* fragments than did the SD-containing *lac* leader. Although it is unclear why the *lac* leader was unable to stimulate higher expression, one obvious possibility is that the *aroL* and *tnaA* internal fragments lack features found naturally at a translation initiation site; the *lac* leader might be compromised in its ability to recruit ribosomes without features that occur at natural initiation sites.

Although the *aroL* U-rich (UUUUUUCUUUAC) mRNA sequence is deficient in A residues, the U-richness has some resemblance to a class of A/U-rich translation enhancer elements reported previously (McCarthy et al., 1986; Gallie et al., 1987; Sleat et al., 1987; Olins et al., 1988; Olins, 1989; Loechel et al., 1991; Zhang and Deutscher, 1992; Hirose and Sugiura, 2004; Hook-Barnard et al., 2007; Nafissi et al., 2012). Devoid of G nucleotides, these A/U elements share no homology with canonical SD sequences and they vary in sequence composition, size, distance to the start codon, and ability to influence translation initiation. In most cases, the A/U element contributes to initiation in conjunction with additional elements such as SD sequences or secondary structure (Loechel et al., 1991; Zhang and Deutscher, 1992; Hirose and Sugiura, 2004; Hook-Barnard et al., 2007; Nafissi et al., 2012); however, some stimulate translation in the

presence or absence of an SD sequence (Ivanov et al., 1995; Golshani et al., 1997). A search of the *E. coli* chromosome for other examples of the *aroL* U-rich sequence indicates that this specific sequence is unique to the *aroL* leader, although variations in sequence and size of U-rich regions can be found (David Ream, personal communication).

Our results support a role for S1 in ribosome recognition of uAUG-1 and -2, as 30S subunits devoid of S1 did not produce toeprint signals on the uAUGs. Whereas deleting this U-rich region led to an almost complete loss of expression from uAUG-2, substitutions that mutated the sequence to its complement reduced expression to 30%. Because S1 interacts with poly(A) tracts (Kalapos et al., 1997), it is possible that substituting the U residues for A only weakened the interaction with S1, and this weak interaction was able to support 30% expression. S1 is necessary for translation of most mRNAs in *E. coli* (Sørensen et al., 1998), especially those that lack or contain a weak SD sequence (Boni et al., 1991; Farwell et al., 1992; Tedin et al., 1997), and there is evidence supporting A/U-rich upstream elements serving as recognition signals for 30S binding via an S1 interaction (Boni et al., 1991; Ringquist et al., 1995).

Contribution of the downstream sequence (DR) to expression

The U-rich region of *aroL*'s 5'UTR's is not the only element controlling ribosome binding and expression from uAUG-2, as mutation or displacement of the downstream region (DR) can result in near total loss of expression. However, the DR alone was not sufficient to stimulate ribosome binding and expression, as mutations and deletions of the U-rich region reduced expression up to 93%. The sequence of downstream sub-regions 2 (DR-2) and 4 (DR-4), as well as the DR position relative to uAUG-2, were critical components for expression from uAUG-2. Interestingly, deletion of two codons in DR-1 moved the DR six nucleotides closer and essentially eliminated expression from uAUG-2 and also reduced expression from uAUG-1, even though uAUG-1 was now only one nucleotide away from the natural positioning of uAUG-2 from the DR.

As discussed above, expression from the *aroL* and *tna* gene-internal fragments demonstrate that *aroL*'s 5'UTR recruited ribosomes to sequences that are not part of

natural initiation sites and that the U-rich region was needed for ribosome binding and expression. However, these internal fragments were expressed at different levels from each other and at substantially lower levels than measured from uAUG-2 in its natural context, possibly due to the lack of DR stimulatory signals. Visual inspection of the downstream region revealed no obvious similarity to the sequence residing downstream to uAUG-2 in *aroL*'s 5'UTR. Expression in the absence of a DR raises the question of whether the DR functions with the U-rich sequence for ribosome recruitment or whether it functions to help direct the ribosome to the translational start site. Although it is not known how the DR affects expression from uAUG-2, *in vitro* ribosome binding patterns correlated well with the *in vivo* expression data, suggesting that the DR influences ribosome binding and/or translation initiation rather than elongation. Additional work is needed to determine how the *aroL* DR contributes to ribosome binding and/or expression from uAUG-2.

Several examples of sequences downstream of start codons that influence gene expression have been reported. For example, a downstream box (DB) region has been reported to enhance translation for a number of mRNAs (Sprengart et al., 1990; Nagai et al., 1991; Shean and Gottesman, 1992; Ito et al., 1993; Mitta et al., 1997). In contrast to the position-specific *aroL* DR sequence, the requirement for the position of the DB sequence relative to the start codon appears to be somewhat relaxed (O'Connor et al., 1999). Although DB sequences are reported to share various degrees of complementarity to the 16S rRNA, the originally proposed base-pairing model has been invalidated (O'Connor et al., 1999) and the mechanism for DB stimulation of mRNA translation has yet to be identified. Other mechanisms proposed for the influence of downstream sequences on expression include codon usage (Robinson et al., 1984; Bulmer, 1988; Gao et al., 1997; Sørensen et al., 1998), peptidyl-tRNA drop-off (Dong et al., 1996; Kuroda and Maliga, 2001), and the presence (Vellanoweth and Rabinowitz, 1992; de Smit and van Duin, 1994; Eyre-Walker, 1996) or absence (Kudla et al., 2009; Gu et al., 2010) of secondary structure.

Work by Loh and coworkers (2012) supports a role for the first 20 nucleotides of the *prfA* coding region in translation initiation. They suggest that this region serves as a

ribosome standby site, allowing the ribosome to bind, wait for the SD to become available (which is sequestered in secondary structure), and diffuse in the 5' direction to contact the SD and initiate translation (Loh et al., 2012). Standby binding, in general, occurs independent of the SD on open and single-stranded regions of the mRNA flanking the hairpin structure that contains the ribosome binding site (de Smit and van Duin, 2003; Unoson, 2007). However, experimental evidence supporting the standby model is lacking.

Computational analysis (Mfold www.mfold.rna.albany.edu/?=mfold) of the *aroL* 5'UTR's secondary structure suggested that uAUG-1 and uAUG-2 are contained within a hairpin and the sequence downstream from uAUG-2 is an open, unstructured region of mRNA (data not shown). If the upstream U-rich region in the *aroL* leader is acting as the primary contact site between the mRNA and ribosome, the downstream region might act as a standby site, allowing the ribosome to associate and wait for the structure containing uAUG-2 to open. This downstream region might also contact the ribosome in such a way as to stabilize an mRNA-S1 interaction, anchoring the mRNA so that uAUG-2 is in the proximity of the P-site. Experimental evidence is needed to test the structural elements of this model.

In an effort to determine the pathway of mRNA through the ribosome, crosslinking studies revealed non SD-ASD interactions between the ribosome (16S rRNA and ribosomal proteins) and mRNA at residues upstream as well as downstream of the start codon (Rinke-Appel et al., 1991; McCarthy, 1994; Rinke-Appel et al., 1994; La Teana et al., 1995). In the absence of an SD sequence, perhaps these interactions play a larger role in placement of the uAUG in the ribosomal P site. The downstream regions DR.2 and DR.4 might strengthen or facilitate these interactions. In addition, the downstream regions might make alternative contacts with the rRNA or ribosomal proteins based off of the mRNA sequence or structure. Mutations that change the sequence or deletions that shift the downstream sequence out of position reduced expression from uAUG-2, most likely by preventing or weakening these interactions.

Start site selection (SSS)

Our 30S ribosome binding assays and expression assays showed a preference for uAUG-2 over uAUG-1. This preference remained constant in the presence of the substitution and deletion analyses of the upstream U-rich and downstream sequence, and when we manipulated the spacing between the U-rich and downstream sequence to the uAUGs. An extended interaction between uAUG-2 and $tRNA^{fMet}$ might explain start site selection of uAUG-2 over uAUG-1. uAUG-1 is directly preceded by a G residue while uAUG-2 is proceeded by a U residue. An interaction between the A at position 37 of the initiator tRNA ($tRNA^{fMet}$) and the U at the -1 position of the mRNA has been proposed to explain such observations (Ganoza et al., 1978; Ganoza et al., 1985; Esposito et al., 2003).

In addition to the sequences upstream and downstream of the uAUGs, the sequence directly surrounding uAUG-1 and -2 appears to be influencing 30S subunit binding and expression. Mutation of two G residues surrounding uAUG-1 (underlined in GAUGGUAUG) resulted in increased binding to both uAUG-1 and -2, although we still saw a binding preference for, and increased expression from, uAUG-2, suggesting that this region exerts an inhibitory influence on the upstream uAUG utilization. G residues around the start codon generally have a negative effect on translation (De Boer et al., 1983; Hui et al., 1984). Deletion of the sequence GUAUG (underlined in GAUGGUAUG) also resulted in increased 30S subunit binding and expression from the remaining uAUG. Interestingly, deletion of this sequence relieved the distance constraint on the upstream U-rich region as well as the downstream sequence. We have not tested whether the distance constraint would be relieved by mutation of the two G residues surrounding uAUG-1. In general, the sequence flanking the AUG start codon influences translation in mammalian, yeast, and prokaryotic cells (Hui et al., 1984; Kozak, 1986; Kozak, 1987; Looman and Kuivenhoven, 1993; Stenström et al., 2001; Kozak, 2005; Pisarev et al., 2006; Rangan et al., 2008; Dvir et al., 2013). Although various mutations to nucleotides surrounding start codons have been documented, the particular effects tend to be influenced by neighboring context and may be specific to particular mRNAs.

It is possible that this G-rich region surrounding uAUG-1 and -2 acts like an SD sequence and base pairs to the 3' end of the 16S RNA (5'- …**UUUG**AU**GGU**AUGA…-3'). The inhibitory effect on 30S binding to and expression we observed might be a result of the position of the sequence (overlapping uAUG-1 and directly adjacent to uAUG-2). The position relative to the start codon as well as the sequence are important for efficient translation initiation (Sedlácek et al., 1979; Weiss et al., 1988; Ringquist et al., 1992; Chen et al., 1994; Jin et al., 2006). If spacing is a determinant, we would expect to see more binding to uAUG-3 as this codon is in a better position from this pseudo SD-like sequence. We believe that 30S binding to uAUG-3 involves a slightly different mechanism than those described above for uAUG-1 and -2, including a region of the mRNA further downstream (Devine and Janssen, unpublished data).

It is also possible that the increased binding to and expression from uAUG-2 was due to disruption of secondary structure around the translation initiation region. Secondary structure within the translation initiation region inhibits expression by acting as a physical barrier between the ribosome and start codon or SD sequence (de Smit and van Duin, 1990, 1994; Unoson, 2007; Gu et al., 2010; Kozak, 2005). Computational analysis of secondary structure (Mfold www.mfold.rna.albany.edu/?=mfold) suggests that the G substitutions and GUAUG deletion relieve some of the secondary structure surrounding the uAUG, making it more accessible to ribosomes (data not shown). This computer analysis will be followed up experimentally in the near future.

Model

Translation initiation on uAUG-2 of the *aroL* leader occurs via a non-traditional mechanism that involves signals upstream and downstream of the start codon. We believe that the U- rich region upstream of uAUG-2 is recognized by ribosomal protein S1 that "catches" the mRNA and brings it closer to the ribosomal decoding site. The DR interacts with rRNA and/or ribosomal proteins so as to stabilize the ribosome: mRNA interaction and enhance uAUG placement in the P site. Our data do not distinguish whether these events occur sequentially or cooperatively.

Implications for regulation and expression

The number of non-SD led mRNAs in prokaryotic genomes has been underestimated (Chang et al., 2006) and the mechanism(s) by which translation initiation occurs on these mRNAs has been underappreciated. Translation initiation from the *aroL* uAUG-2 appears to involve a unique and complex mechanism that involves nontraditional sequences upstream, downstream, and directly flanking the uAUG-2 start codon. We have yet to show direct interactions between these regions of the mRNA with the ribosome and identify the specific sites of contact with the ribosome, nor do we understand the selective pressure that may have given rise to the mechanism of translation initiation at uAUG-2. Perhaps this mechanism is a way to maintain low expression from the uAUG-2 reading frame, reflecting a need to keep expression of the putative peptide low. The U-rich region partially overlaps a TYR DNA repressor binding site (Lawley and Pittard, 1994). The cell might capitalize on this T-rich sequence in the DNA as part of the repressor binding site and the U-rich sequence in the mRNA as an S1 binding site for two independent functions, both of which serve to control expression from uAUG-2 and *aroL*. As the literature continues to present additional examples of mRNAs lacking SD sequences, we can better understand the selective pressures and regulatory features provided by non-SD control of translation initiation such as the uAUG-2 of the *aroL* untranslated leader.

Chapter Two

ryhB in *Shewanella oneidensis* MR-1 shares sequence similarity to *ryhB* in *Escherichia coli* and is regulated by iron

Introduction

Small, noncoding RNAs (sRNAs) are a growing class of bacterial regulatory molecules (reviewed in Gottesman and Storz, 2011). sRNAs play a regulatory role in many aspects of cellular physiology and it has been suggested that in the future roles of sRNAs will be discovered in almost every global bacterial response (Gottesman and Storz, 2011). These regulatory molecules bind to target mRNA via complementary base pairing, usually on the 5' end of the mRNA. Most sRNAs characterized to date require the protein, Hfq, as a cofactor to help stabilize the sRNA and facilitate the proper interactions between the sRNA and its targets (Gottesman, 2004; Gottesman and Storz, 2011). This pairing usually results in translation inhibition and leads to RNaseE-mediated degradation of the target message (Massé and Gottesman, 2002; Gottesman, 2004). The sRNA itself is usually degraded along with the target message.

Originally identified and studied thoroughly in *Escherichia coli, ryhB* is an sRNA that has a key role in iron metabolism. This sRNA is essential for strict regulation of iron uptake and storage by *E. coli* and itself is regulated by the master regulator of iron metabolism, the ferric uptake regulator (Fur protein) (Massé et al., 2005). Under iron-replete conditions inside the cell the Fur protein represses the expression of iron uptake genes and *ryhB* by binding the promoter regions with the ferrous iron cofactor (Fe^{2+}) (Hantke, 1981; Hantke and Braun, 1997). The repression of these genes is relieved as Fur becomes inactivated under iron-deplete conditions in the cell. Expression of *ryhB* results in decreased expression of a handful of genes including Fe superoxide dismutase, iron-binding enzymes in the TCA cycle, and succinate dehydrogenase (Massé and Gottesman, 2002; Massé et al., 2005). Complementary pairing of the *trans*-encoded sRNA leads to degradation of these target mRNAs (Massé et al., 2003). *ryhB* requires Hfq for stability and proper function and RNaseE for degradation of the target mRNAs (Massé et al*.,* 2003, 2005). Expression of *ryhB* provides for a quick cellular response to iron-deplete conditions by down-regulating the expression of nonessential Fe-containing proteins and ferritins. Thus, *ryhB* enables the cell to use its limited iron supply sparingly.

ryhB homologs have been found in other enterobacterial species. These sRNAs contain a Fur binding site in the promoter region and also share a conserved region involved in pairing to target mRNAs (Gottesman et al., 2007). Although they share many target mRNAs, they do not bind all of the same target messages (Gottesman et al., 2007). *ryhB*-like sRNAs have been identified in *Pseudomonas aeruginosa, Neisseria meningitides, and Bacillus subtilis.* These Fur-regulated sRNAs are considered *ryhB*-like rather than *ryhB* homologs because although they share similar targets, they do not share sequence similarity (Wilderman et al., 2004; Mellin et al., 2007; Gaballa et al., 2008). It is unclear whether these sRNAs evolved independently or arose through rapid divergence. There is evidence of sRNA duplication in several bacteria species (Wilderman et al*.,* 2004; Gottesman et al., 2007), which most likely played a role in proliferation of these *ryhB* homologs and *ryhB*-like sRNAs.

One such *ryhB* homolog was identified in the Gram-negative, facultative anaerobe, *Shewanella oneidensis* MR-1. Since it was isolated from the sediments of Lake Oneida, New York (Myers and Nealson 1988a,b), this mineral-reducing bacterium has received a significant amount of attention for its eclectic respiratory capacity. *S. oneidensis* MR-1 has the ability to reduce a myriad of compounds including nitrate, fumarate, Mn(IV) oxides, iron (III) oxides, and trimethylamine oxide (TMAO) and plays critical roles not only in the environmental cycling of these metals (Nealson et al., 1991; Myers and Nealson, 1990) but also in heavy metal immobilization, mineral dissolution, and bioremediation (Myers and Nealson, 1997; Nealson et al., 2002; Ward et al., 2003; Thormann et al., 2004, 2005).

The *ryhB* homolog of *S. oneidensis* MR-1 is upregulated in a *fur* mutant, suggesting a role in iron metabolism (Wan et al., 2004). Further analysis shows similarity to the *E.coli ryhB* on both sequence and structural levels. Despite these similarities there are key differences between *ryhB* of *S. oneidensis* MR-1 and *E. coli*. The former is significantly longer than the latter. The additional nucleotides may provide additional binding sites for targets specific to *S. oneidensis* MR-1. In *E. coli*, the *sodB* mRNA is negatively regulated by *ryhB* (Massé et al., 2003), whereas this mRNA does not appear to be a target of *ryhB* in *S. oneidensis* MR-1. We hypothesize that *ryhB* in *S.*

oneidensis MR-1 has attributes and targets that are unique to this bacterium. Because this organism has impressive environmental implications for its use in immobilizing heavy metals and bioremediation (Myers and Nealson, 1998; Nealson et al., 2002; Ward et al., 2003; Thormann et al., 2004, 2005) it is important not only to define the regulatory features of *ryhB* but also to better understand how sRNAs in general regulate the global physiology of *S. oneidensis* MR-1.

Materials and Methods

Reagents and recombinant DNA procedures. Oligonucleotides were purchased from Integrated DNA Technologies. Radiolabeled [γ-^{32}P]ATP (6,000 Ci/mmol, 150 mCi/mL) was purchased from Perkin Elmer. *Pfu* DNA polymerase (Stratagene), T4 DNA ligase and restriction endonucleases (New England Biolabs), AMV reverse transcriptase (Life Sciences), and RNase-free DNase I (Ambion) were used according to the manufacturers' recommendations. T4 polynucleotide kinase (Wang and Shuman, 2001) was purified as described. The TRIzol® Reagent was purchased from Invitrogen. Antibiotics were purchased from Fisher Scientific. 2,2'-dipyridyl was purchased from Sigma-Aldrich and added to growth media at various concentrations.

E. coli DH5α (New England Biolabs) was used as the host for all plasmid DNA manipulations. *E. coli* ET124567 (*dam*-, *dcm*-) (MacNeil et al., 1992) cells were transformed with plasmids and propagated, and plasmids were isolated from these transformants. *S. oneidensis* MR-1 cells were electroporated in the presence of the plasmids. *E. coli* cells were cultured at 37^{0}C and *S. oneidensis* MR-1 cells at 30^{0}C. Media containing *S. oneidensis* MR-1 cells was supplemented with rifampin (10 μg/mL). Cells were grown in either Luria-Burtani (Bertani, 1951), 2XYT (Kunkel et al., 1987), or MR-1 HBa (High Biomass) minimal media. HBa medium contained the following ingredients: 0.3 mM piperazine-N,N′ -bis(2-ethanesulfonic acid) (PIPES), 7.5 mM NaOH, 28 mM NH_4Cl, 1.3 mM KCl, 4.4 mM NaH_2PO_4, 30 mM Na_2SO_4, 10 mL/L minerals [nitrilotriacetic acid (1.5 g/L), $MgSO_4$ (3.0 g/L), $MnSO_4 \cdot H_2O$ (0.5 g/L), NaCl (1.0 g/L), $FeSO_4 \cdot 7H_2O$ (0.1 g/L), $CaCl_2 \cdot 2H_2O$ (0.1 g/L), $CoCl_2 \cdot 6H_2O$ (0.1 g/L), $ZnCl_2$ (0.1 g/L), $CuSO_4 \cdot 5H_2O$ (0.01 g/L), $AlK(SO_4)_2 \cdot 12H_2O$ (0.01 g/L), H_3BO_3 (0.01 g/L), Na_2MoO_4 (0.03 g/L), $NiCl_2 \cdot 6H_2O$ (0.03 g/L), $Na_2WO_4 \cdot 2H_2O$ (0.03 g/L)], 10 mL/L vitamins [biotin (0.002 g/L), folic acid (0.002 g/L), pyridoxine HCl (0.01 g/L), riboflavin (0.005 g/L), thiamine (0.005 g/L), nicotinic acid (0.005 g/L), pantothenic acid (0.005 g/L), B-12 (0.0001 g/L), *p*-aminobenzoic acid (0.005 g/L), thioctic acid (0.005 g/L)], and 90 mM lactate (aerobic growth). Ten mL/L amino acids [glutamic acid, arginine, and serine (2.0 g/L each)], $CaCl_2$ (0.7 mM), and fumarate (120 mM, anaerobic growth) were added after

autoclaving. For anaerobic growth, the media was prepared by boiling under O_2-free N_2 gas.

General cloning procedures, plasmid isolations, *E. coli* transformations, and other DNA manipulations were carried out in a standard manner (Sambrook et al., 1989). To select for plasmid maintenance, antibiotics were used at the following concentrations unless otherwise noted: kanamycin (50 μg/mL), chloramphenicol (3 μg/mL). *ryhB* from *S. oneidensis* MR-1 was cloned directly at the transcriptional start site of the *lac* promoter in plasmid pACYC(Kan^R) and P_{BAD} in pBAD33(Cm^R) (Guzman et al., 1995). To obtain pACYC-*ryhB* the *ryhB* gene fragment was amplified by PCR from S*. oneidensis* MR-1 genomic DNA using the oligonucleotides ryhB 5' (5'-GCGTTCCAAAACTCATCTTTAACTC-3') and ryhB 3'ClaI which contained an added 5' *Cla*I site (5'-ATCGAT-3') (5'-CCATCGATCAGCAGGATATAGCGATTGGTGTC-3'). The amplified product was cut with *Cla*I and ligated into pM1108 (+*Eco*RV/*Cla*I). PCR was used to amplify a DNA fragment from the above ligation using primers lac.pcr4 (5'-CAGAATTCTGGCACGACAGGTTTCCC-3') and ryhB 3'SalI, which contained an added *Sal*I site (5'-GTCGAC-3') (5'-ACGCGTCGACGACATGGTTAGCCTAATGCGCC-3'). The product was digested with *Eco*RI and *Sal*I and ligated into pACYC (+*Eco*RI and *Sal*I). Transformants were selected on LB agar with kanamycin. To obtain pBAD33-*ryhB*, PCR was used to generate a DNA fragment from pACYC-*ryhB* using primers 5'BglII, which contained an added *Bgl*II site (5'-AGATCT-3') (5'-ACGCAGATCTCTGTAACAAAGCGGG-3) and ryhB 3'SalI. The amplified product was cut with *Bgl*II and *Sal*I and ligated into pBAD-uAUG2-*lacZ* (see Appendix B) (+*Bgl*II and *Sal*I). Transformants were selected on LB agar with chloramphenicol and screened on LB agar with chloramphenicol and 1% L-arabinose.

Radiolabeling of oligonucleotides. Oligonucleotides were end-labeled as described previously (Van Etten and Janssen, 1998). The sequence of the radiolabeled primer used in the primer extension assay and DNA sequencing ladder reaction was 5'-AGCCGGATGATGAATCCGGCC-3'.

DNA and RNA isolation. Genomic DNA and total RNA were isolated using the TRIzol® Reagent and protocol from Life Technologies. Centrifugations were performed at 12,000

x *g* at 4^0C unless otherwise noted. In general, a 5-mL suspension of cells was harvested by centrifugation and lysed in 1 mL TRIzol® Reagent by pipetting. Following a 5-minute incubation at room temperature, 0.2 mL chloroform was added and the sample was incubated at room temperature for 3 additional minutes. The sample was centrifuged for 15 minutes. The aqueous phase was used to isolate total RNA and the interphase/organic phenol-chloroform phase was used for the isolation of DNA. To isolate RNA, 0.5 mL of 100% isopropanol was added to the aqueous phase, the sample was incubated at room temperature for 10 minutes and centrifuged for 10 minutes. The RNA pellet was washed with 1 mL 75% ethanol and centrifuged at 7,500 x *g* for 5 minutes. RNA samples were treated with DNase I prior to use in qPCR. To isolate DNA, 0.3 mL of 100% ethanol was added and the sample incubated at room temperature for 3 minutes, then centrifuged at 2,000 x *g* for 5 minutes. The DNA pellet was washed with 1 mL sodium citrate/ethanol solution (0.1 M sodium citrate in 10% ethanol, pH 8.5), incubated for 30 minutes at room temperature, and centrifuged at 2,000 x *g* for 5 minutes. Next, 2 mL of 75% ethanol was added to the DNA pellet, the sample was incubated for 20 minutes at room temperature and centrifuged at 2,000 x *g* for 5 minutes.

Primer extension reaction. Initially, 40 μg total RNA and 2 pmol of kinase-treated oligonucleotide were co-precipitated using 2 M sodium acetate and 100% ethanol, and suspended in 30 μL hybridization solution (0.4 M NaCl, 40 mM PIPES pH 6.4, 1 mM EDTA, 80% [v/v] deionized formamide). Samples were heated to 80^0C for 10 minutes and vortexed. Hybridization of the oligonucleotide to the RNA was performed by heating to 80^0C for 10 minutes and then lowering the temperature to 55^0C (annealing temperature of oligonucleotide). Annealing proceeded overnight. The annealing reaction was precipitated by adding 45 μL 12 M sodium acetate, 145 μL DEPC-treated H_2O, and 800 μL100% ethanol. Samples were incubated at -80^0C for 20 minutes, centrifuged, and washed with cold 100% ethanol. During the remaining procedures the samples and all reagents added were kept cold. The pellets were carefully suspended in 10 μL PIPES/NaCl (10 mM PIPES pH 6.4, 400 mM NaCl). Extension of the annealed primer was initiated by adding 80 μL of S1 buffer (50 mM Tris pH 8.2, 10 mM dithiothreitol, 6

mM $MgCl_2$, 25 μg/mL actinomycin D, 500 μM each of dATP, dGTP, dCTP, dTTP) and 0.5 μL of AMV reverse transcriptase. Reactions were incubated at 42^0C for 60 minutes. Samples were then precipitated by adding 9 μL 3 M sodium acetate and 100 μL phenol/chloroform/isoamyl alcohol (25:24:1). Upon centrifugation for 10 minutes, the aqueous phase was transferred to a fresh tube and 250 μL 100% ethanol was added. Samples were incubated at -80^0C for 20 minutes and centrifuged for 15 minutes. Pellets were dried, dissolved in 10 μL loading dye (10 mM NaOH, 1 mM EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue, 80% deionized formamide) and subjected to PAGE (6% acrylamide, 7M urea) in 1X Tris/Borate/EDTA (TBE) buffer.

DNA sequencing ladder. The purified DNA template used in the sequencing reaction was generated using genomic DNA isolated from *S. oneidensis* MR-1 cultures and the forward 5'-GGGCTTCAAGCTGGGTTATGAAC-3' and reverse 5'-AGCCGGATGATGAATCCGGCC-3' primers. The sequencing ladder was prepared using the USB® Thermo Sequenase™ Cycle Sequencing Kit and protocol from Affymetrix. Briefly, 200 ng genomic DNA was added to a reaction mix containing 2 μL reaction buffer (260 mM Tris-HCl pH 9.5, 65 mM $MgCl_4$), 2 uL radiolabeled primer (5'-AGCCGGATGATGAATCCGGCC-3' at 4 μM), 2 μL of DNA polymerase (4 units/μL) and 10.5 uL H_2O. Four μL of this reaction was added to 4 μL each dideoxy termination mix (ddATP, ddTTP, ddGTP, ddCTP). After fifty cycles of PCR amplification (95^0C 30 seconds, 55^0C 30 seconds, 72^0C 60 seconds), 4 μL stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF) was added to each reaction. Samples were heated for 2 minutes at 75^0C before being subjected to PAGE (6% acrylamide, 7 M urea) in 1X TBE.

Preparation of electrocompetent cells. All centrifugations were performed at 5,000 x *g* at 4^0C. A 50-mL overnight culture of *S. oneidensis* MR-1 cells grown in LB (Rif) was diluted into 1 L of HBa minimal medium with rifampin and incubated at 30^0C until an OD_{600} of ~0.6 was reached. Cells were then incubated on ice for 30 minutes, split into four 250-mL centrifuge bottles, and centrifuged for 20 minutes. After removal of the supernatants, cells were suspended in 1 L cold H_2O and centrifuged for 20 minutes. The supernatant was removed and cells were suspended in 500 mL cold H_2O and

centrifuged for 20 minutes. Cell pellets were suspended in 30 mL of 10% glycerol and centrifuged for 20 minutes before being suspended in a final volume of 6 mL 10% glycerol.

Electroporation. DNA (100 ng) was electroporated into competent *S. oneidensis* MR-1 cells using the following settings: 0.75 kV, 400Ω, and 25 μF. Cells were recovered in LB medium overnight at room temperature and then plated onto LB plates supplemented with rifampin and chloramphenicol and incubated at 30^0C.

β-galactosidase assays. β-galactosidase assays were performed as described (Miller, 1992). 2XYT medium (+/- chloramphenicol) was inoculate from overnight cultures (OD_{600} = 0.05) and grown to an OD_{600} of ~0.4. Cultures were then split into three tubes, containing 0%, 0.05%, or 1% L-arabinose. Samples were pulled at 0, 30, 60, 120, and 240 minutes after L-arabinose induction and assayed for β-galactosidase activity.

Quantitative PCR. Quantitative PCR (qPCR) was performed using the QuantiTect SYBR Green RT-PCR Kit (QIAGEN). Samples and reagents were kept on ice during preparation. A reaction mixture was prepared containing 25 μL 2x QuantiTect SYBR Green RT-PCR Master Mix (HotStartTaq DNA Polymerase, QuantiTect SYBR Green RT-PCR Buffer, dNTP mix, fluorescent dyes), primer sets to a final concentration of 0.5 μM each, 0.5 μL QuantiTect RT Mix, 100 ng RNA, and H_2O to a final volume of 50 μL. Negative controls without RNA or QuantiTect RT Mix were included with each run. qPCR was performed on the Rotorgene 3000 which was programed for the following conditions: 50^0C for 30 minutes to support reverse transcription, 95^0C for 15 minutes to allow for PCR initial activation, and 40 cycles of a 3-step cycling which included a 15-second denaturation at 94^0C, a 30-second annealing at 55.8^0C, and a 30-second extension at 72^0C. A melting curve was performed to verify the specificity and identity of RT-PCR products between 65^0C to 95^0C. The *ryhB* primer set included: 5'-CTGATGACTGGTAATCTGAC-3' and 5'-GCCGGATGATGAATCCGG-3', the *fur* primer set included: 5'-GCCACGAGTCAAGATCCTAG-3' and 5'-CACACGGTAGACTGTTGC-3, the *tufB* primer set included: 5'-GGAGCGTGAGCGCGGTATTACC-3' and 5'-

GTCCATCTGTGCAGCACCAGTG-3', and the *sodB* primer set included: 5'-CGGCAAGCATCACAACACC-3' and 5'-CCTGCGAAATCGGTTCCTTCG-3'.

Results

Description of *ryhB* locus in *S. oneidensis* MR-1

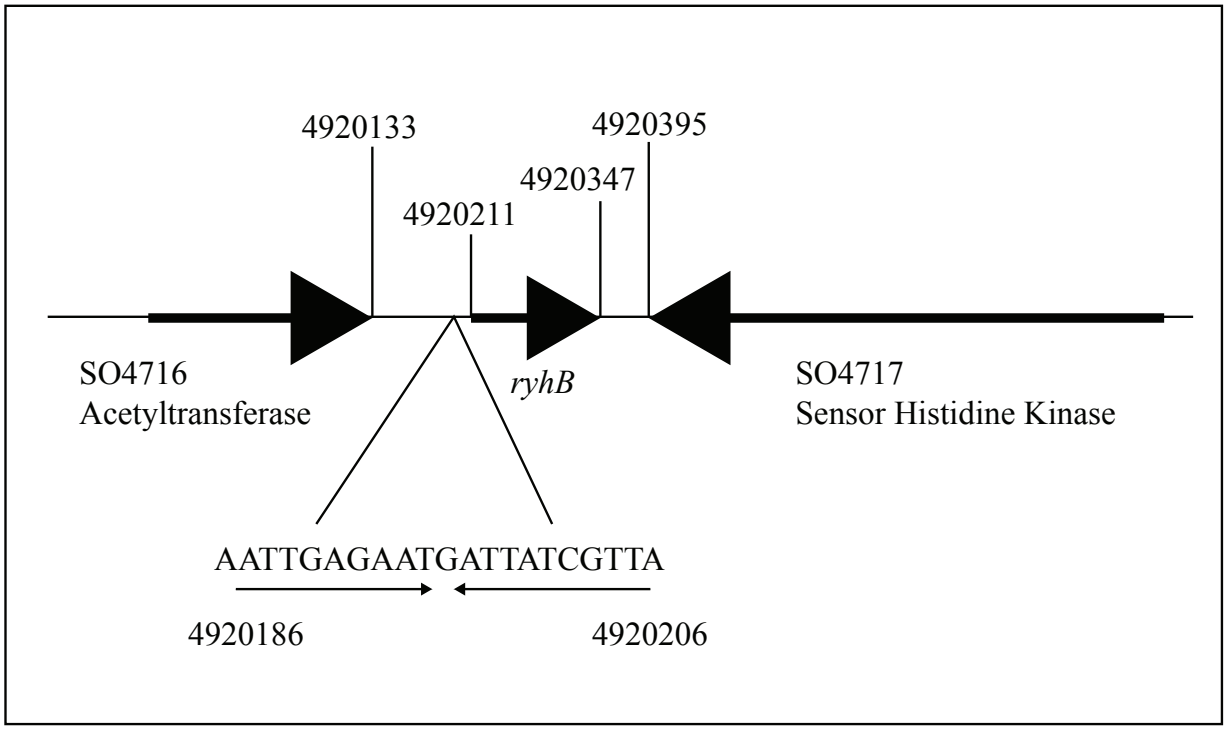
S. oneidensis MR-1 *ryhB* is predicted to encode an sRNA of 137 or 136 nucleotides in length, depending upon the start. Figure 2-1 shows the *ryhB* locus on the *S. oneidensis* MR-1 genome with the predicted Fur box sequence located upstream. The *ryhB* transcriptional start site was mapped by primer extension to a G or an adjacent C (Fig. 2-2). A factor-independent stem-loop transcriptional terminator was predicted by computational analyses at positions 101-119 of the RNA (Fig. 2-1B), which contains a G-C rich stem-loop directly followed by a series of U residues. Although a comparison of *ryhB* in *S. oneidensis* to *ryhB* in other Gammaproteobacteria showed a core region of high sequence identity (starred nucleotides in Fig. 2-1C), the majority of the *S. oneidensis* MR-1 *ryhB* sequence did not share high sequence identity (Fig. 2-1C). The sequences contained in the loops at positions 1-27 and 62-69 of the *S. oneidensis* MR-1 RNA (Fig. 2-1B and C) did not share high sequence similarity to the other organisms. These regions may target mRNAs unique to *S. oneidensis* MR-1.

S. oneidensis MR-1 growth and *ryhB* RNA levels in response to iron supplementation and chelation

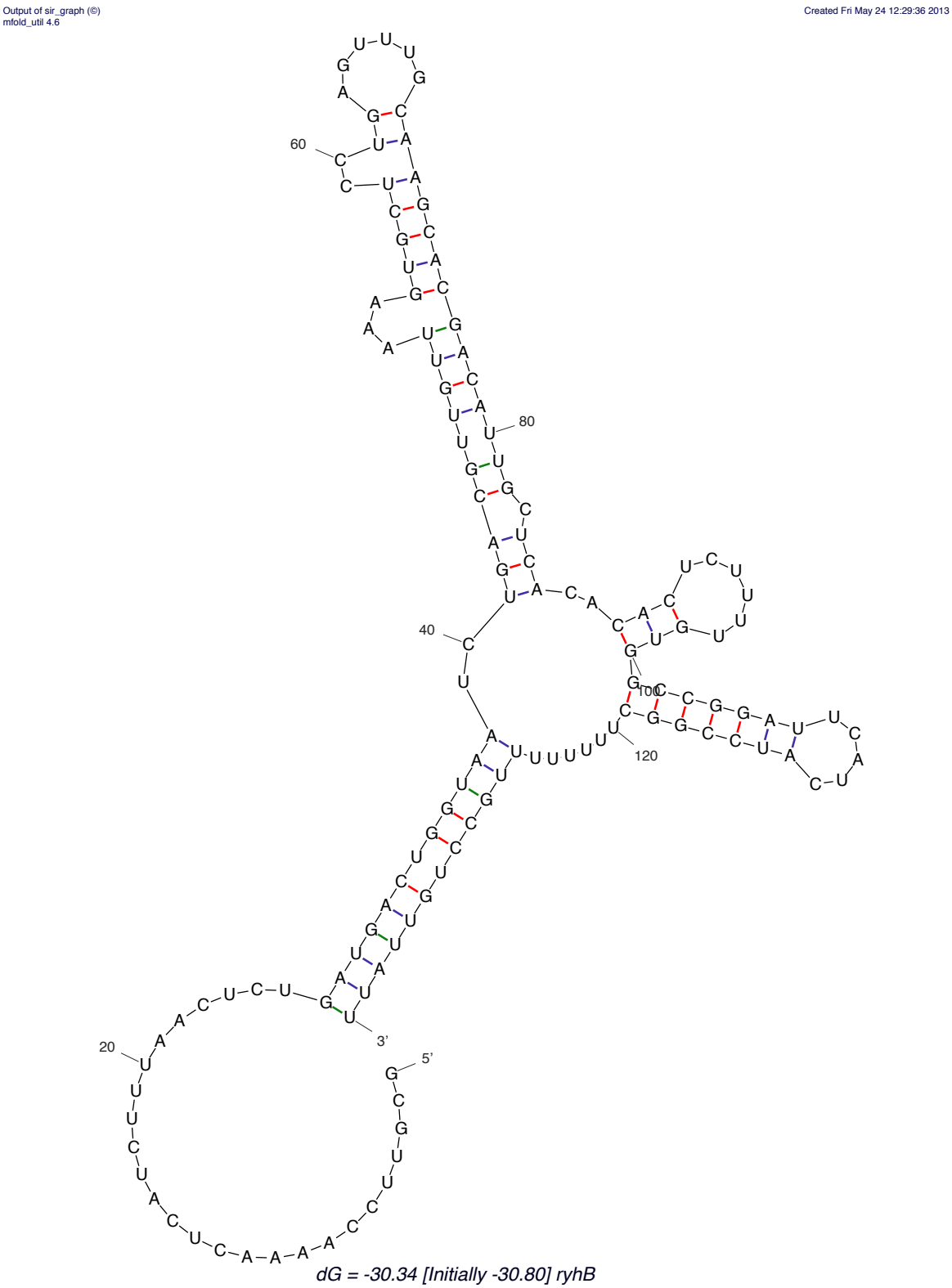
The putative Fur box upstream of *ryhB* suggests that the RNA is regulated in response to iron (Bagg et al., 1987; de Lorenzo et al., 1988; Escolar et al., 1997). If so, then *ryhB* levels would differ under iron-replete and -deplete conditions. A series of growth curves was performed to assess the growth phenotype of *S. oneidensis* MR-1 cells in the presence or absence of an iron source, FeNTA, both aerobically and anaerobically (Fig. 2-3 and Fig. 2-4). The presence of the iron source at 25 or 50 μM did not affect the patterns of aerobic (Fig. 2-3A and Fig. 2-4A) or anaerobic growth (Fig. 2-3B and Fig. 2-4B). Next, we examined *ryhB* RNA levels under iron-replete or -deplete conditions. Quantitative PCR (qPCR) was used to compare *ryhB* RNA levels from cells grown to mid-log phase aerobically in media containing 0, 25, or 50 μM FeNTA. *ryhB* levels were decreased 17-fold in the presence of 25 μM FeNTA and 9-fold in the presence of 50 μM FeNTA (Fig. 2-5).

Figure 2-1. Description of the *S. oneidensis* MR-1 *ryhB* sequence, the predicted RNA secondary structure, and multiple sequence alignment comparing *ryhB* in *S. oneidensis* MR-1 to *ryhB* in other Gammaproteobacteria. A. *ryhB* locus and surrounding genes in the chromosome of *S. oneidensis* MR-1. The positions in the DNA are indicated. A predicted Fur box has been identified upstream of *ryhB;* its sequence, predicted hairpin, and DNA positions are shown. B. Predicted secondary structure of *S. oneidensis* MR-1 *ryhB* generated from Mfold, (http://mfold.rna.albany.edu/?q=mfold). C. Multiple sequence alignment comparing *ryhB* in *S. oneidensis MR-1* to *ryhB* in other Gammaproteobacteria generated from CLUSTALW (http://www.genome.jp/tools/clustalw/). D. Phylogenetic tree containing *S. oneidensis* MR-1 and other Gammaproteobacteria that encode *ryhB* in their genome generated from CLUSTALW (http://www.genome.jp/tools/clustalw/). *S. oneidensis* MR-1 is outlined.

A.

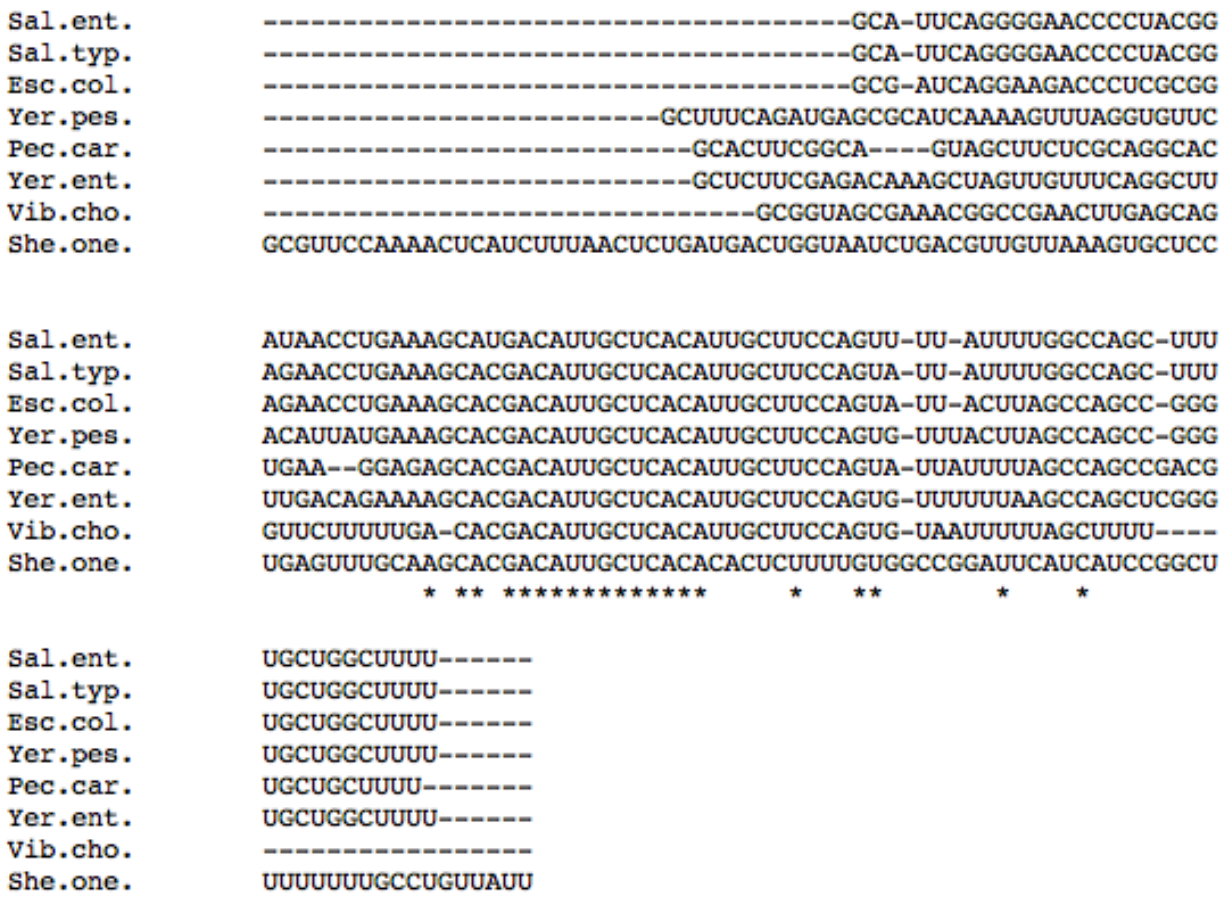

4920133
4920395
4920347
4920211
SO4716
Acetyltransferase
ryhB
SO4717
Sensor Histidine Kinase
AATTGAGAATGATTATCGTTA
4920186
4920206

B.



dG = -30.34 [Initially -30.80] ryhB

C.


Sal.ent. ----------GCA-UUCAGGGGAACCCCUACGG
Sal.typ. ----------GCA-UUCAGGGGAACCCCUACGG
Esc.col. ----------GCG-AUCAGGAAGACCCUCGCGG
Yer.pes. ----------GCUUUCAGAUGAGCGCAUCAAAAGUUUAGGUGUUC
Pec.car. ----------GCACUUCGGCA----GUAGCUUCUCGCAGGCAC
Yer.ent. ----------GCUCUUCGAGACAAAGCUAGUUGUUUCAGGCUU
Vib.cho. ----------GCGGUAGCGAAACGGCCGAACUUGAGCAG
She.one. GCGUUCCAAAACUCAUCUUUAACUCUGAUGACUGGUAAUCUGACGUUGUUAAAGUGCUCC
Sal.ent. AUAACCUGAAAGCAUGACAUUGCUCACAUUGCUUCCAGUU-UU-AUUUUGGCCAGC-UUU
Sal.typ. AGAACCUGAAAGCACGACAUUGCUCACAUUGCUUCCAGUA-UU-AUUUUGGCCAGC-UUU
Esc.col. AGAACCUGAAAGCACGACAUUGCUCACAUUGCUUCCAGUA-UU-ACUUAGCCAGCC-GGG
Yer.pes. ACAUUAUGAAAGCACGACAUUGCUCACAUUGCUUCCAGUG-UUUACUUAGCCAGCC-GGG
Pec.car. UGAA--GGAGAGCACGACAUUGCUCACAUUGCUUCCAGUA-UUAUUUUAGCCAGCCGACG
Yer.ent. UUGACAGAAAAGCACGACAUUGCUCACAUUGCUUCCAGUG-UUUUUUAAGCCAGCUCGGG
Vib.cho. GUUCUUUUUGA-CACGACAUUGCUCACAUUGCUUCCAGUG-UAAUUUUUAGCUUUU----
She.one. UGAGUUUGCAAGCACGACAUUGCUCACACACUCUUUUGUGGCCGGAUUCAUCAUCCGGCU
* ** ********** * ** * *
Sal.ent. UGCUGGCUUUU------
Sal.typ. UGCUGGCUUUU------
Esc.col. UGCUGGCUUUU------
Yer.pes. UGCUGGCUUUU------
Pec.car. UGCUGCUUUU-------
Yer.ent. UGCUGGCUUUU------
Vib.cho. ----------
She.one. UUUUUUUGCCUGUUAUU

D.

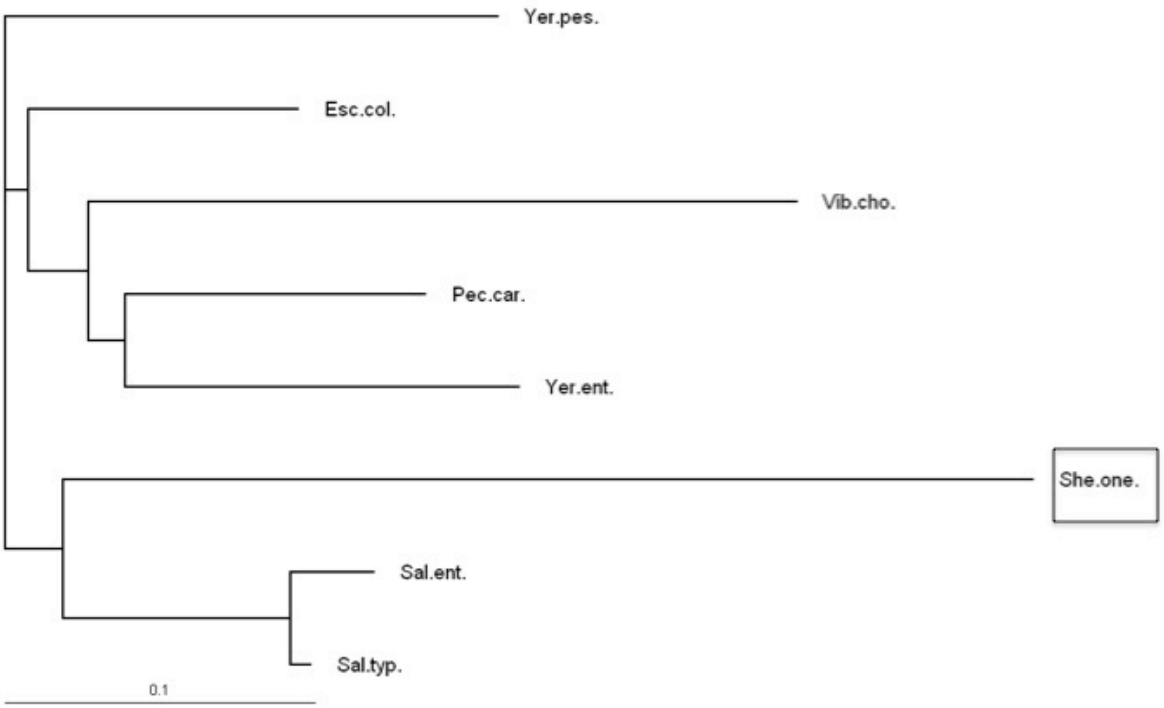

Yer.pes.
Esc.col.
Vib.cho.
Pec.car.
Yer.ent.
She.one.
Sal.ent.
Sal.typ.
0.1

Figure 2-2. Primer extension analysis of *S. oneidensis* MR-1 *ryhB*. The 5' end of the RNA was compared to a DNA sequencing ladder and mapped to nucleotides G and C. Lane 1, *S. oneidensis* MR-1 RNA; lane 2, *E. coli* RNA (negative control).

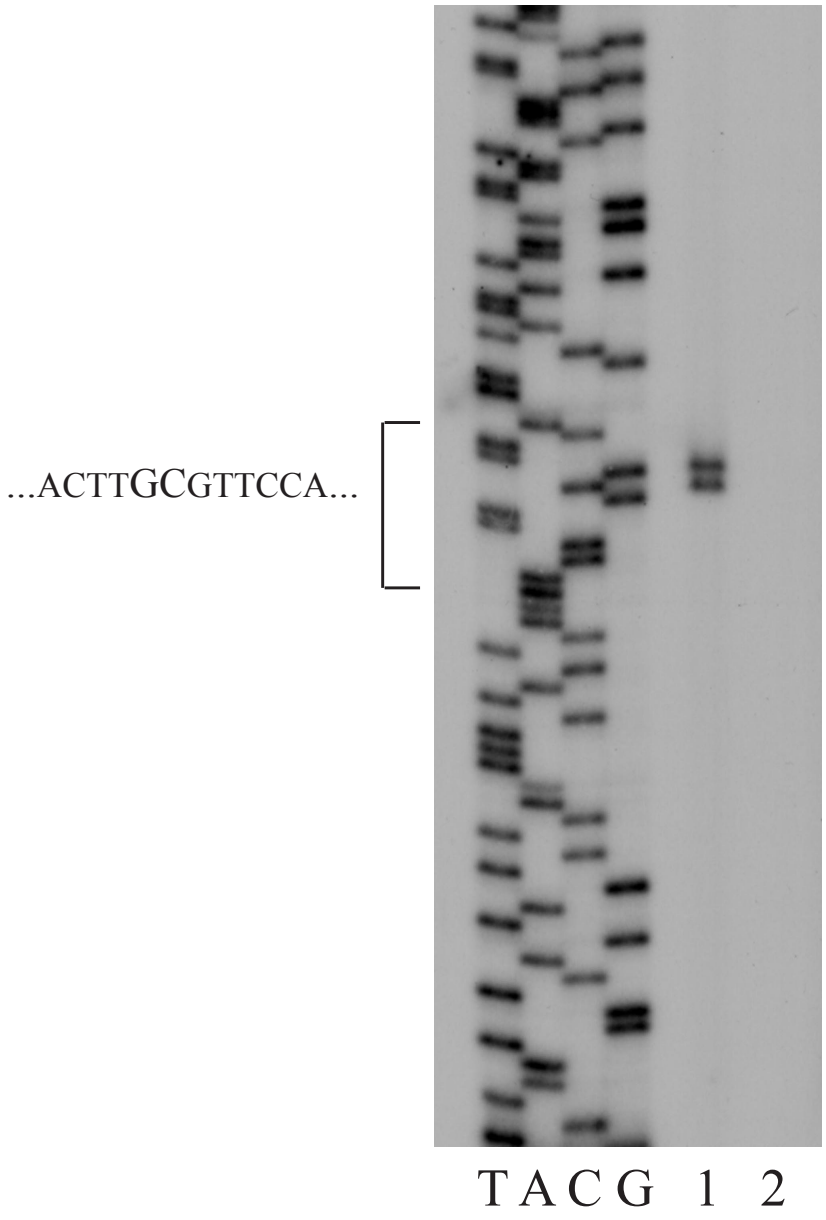
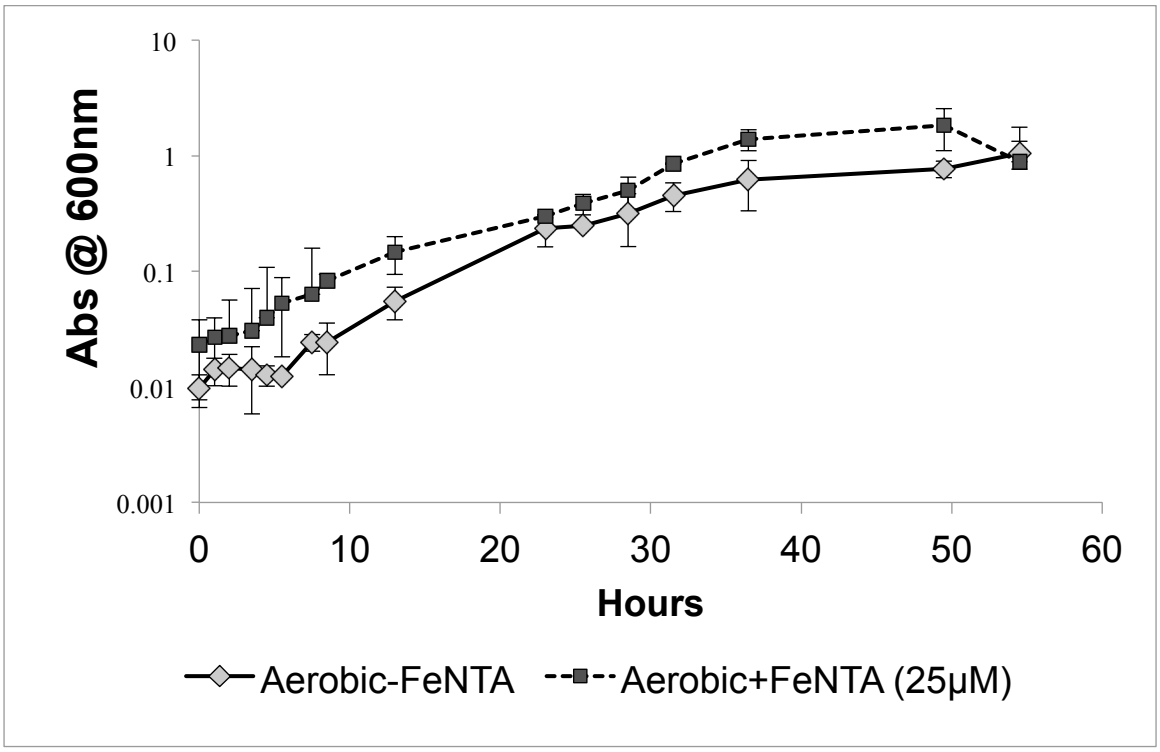

...ACTTGCGTTCCA...
T A C G 1 2

Figure 2-3. *S. oneidensis* MR-1 growth curves in HBa minimal medium lacking (-FeNTA) or containing (+FeNTA) 25 μM FeNTA as an iron source in aerobic (A) or anaerobic (B) conditions.

A.


10
1
0.1
0.01
0.001
Abs @ 600nm
0
10
20
30
40
50
60
Hours
Aerobic-FeNTA
Aerobic+FeNTA (25μM)

B.

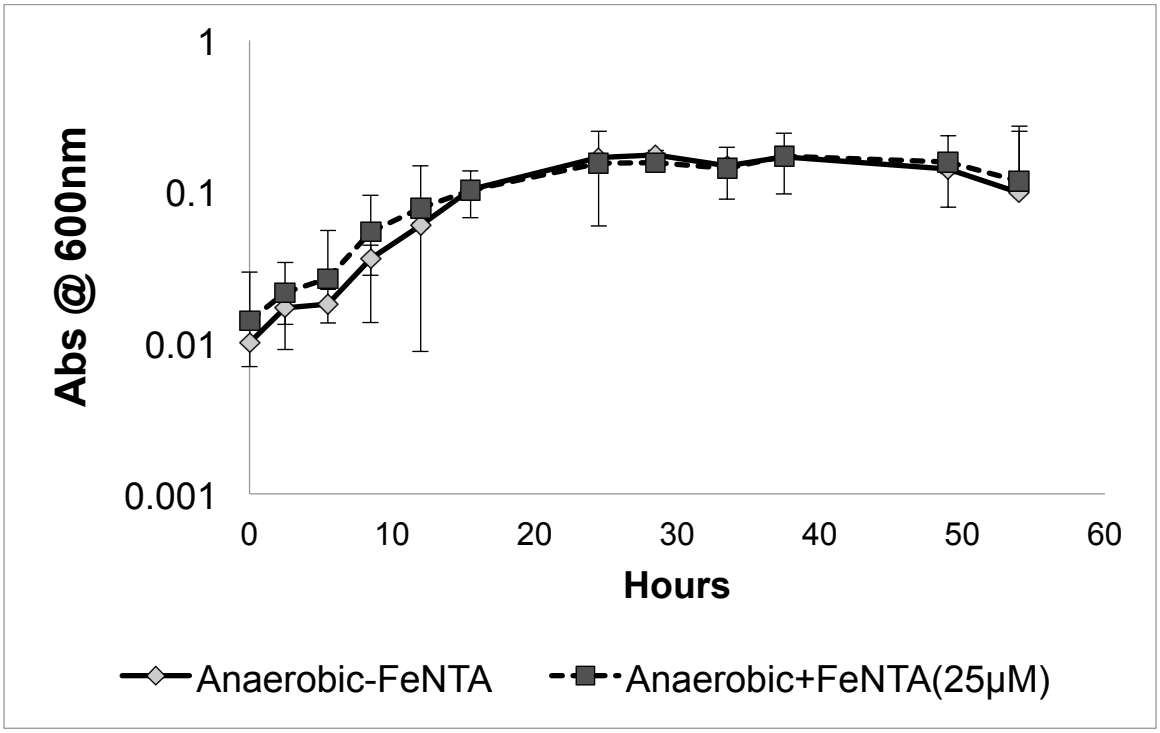
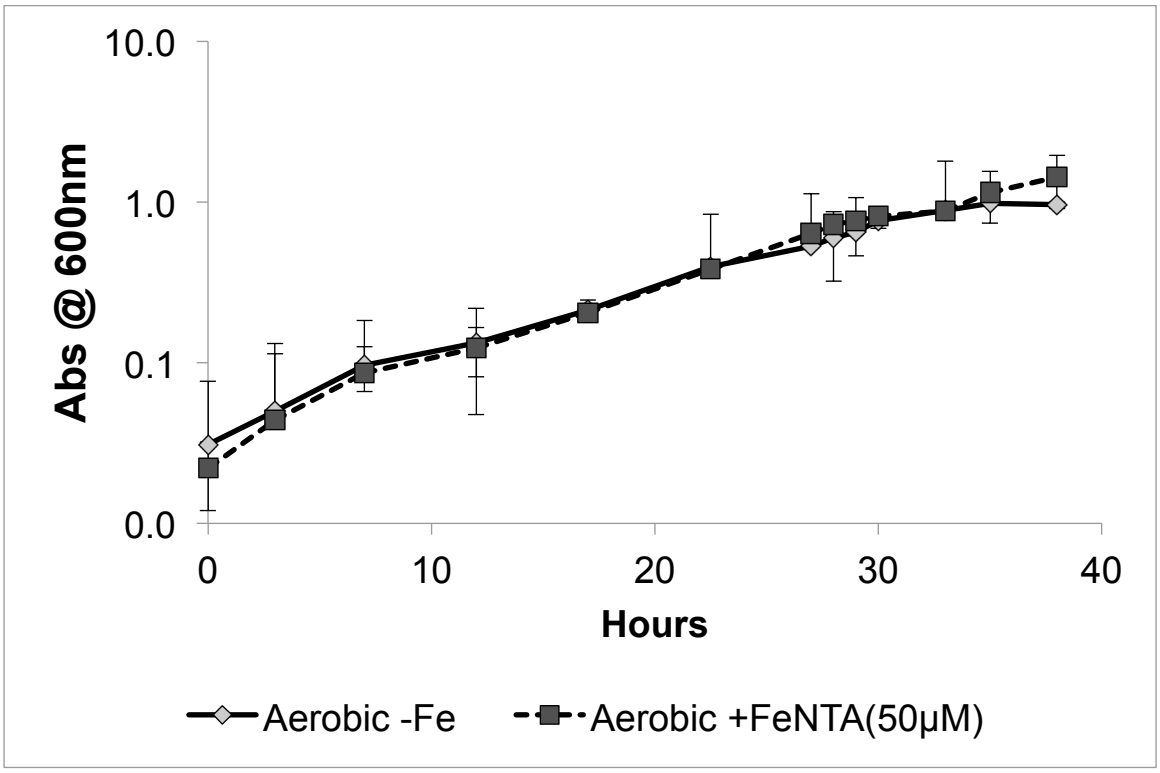

1
0.1
0.01
0.001
Abs @ 600nm
0
10
20
30
40
50
60
Hours
Anaerobic-FeNTA
Anaerobic+FeNTA(25μM)

Figure 2-4. *S. oneidensis* MR-1 growth curves in HBa minimal medium lacking (-FeNTA) or containing (+FeNTA) 50 μM FeNTA as an iron source in aerobic (A) or anaerobic (B) conditions.

A.


10.0
1.0
0.1
0.0
Abs @ 600nm
0
10
20
30
40
Hours
Aerobic -Fe
Aerobic +FeNTA(50μM)

B.

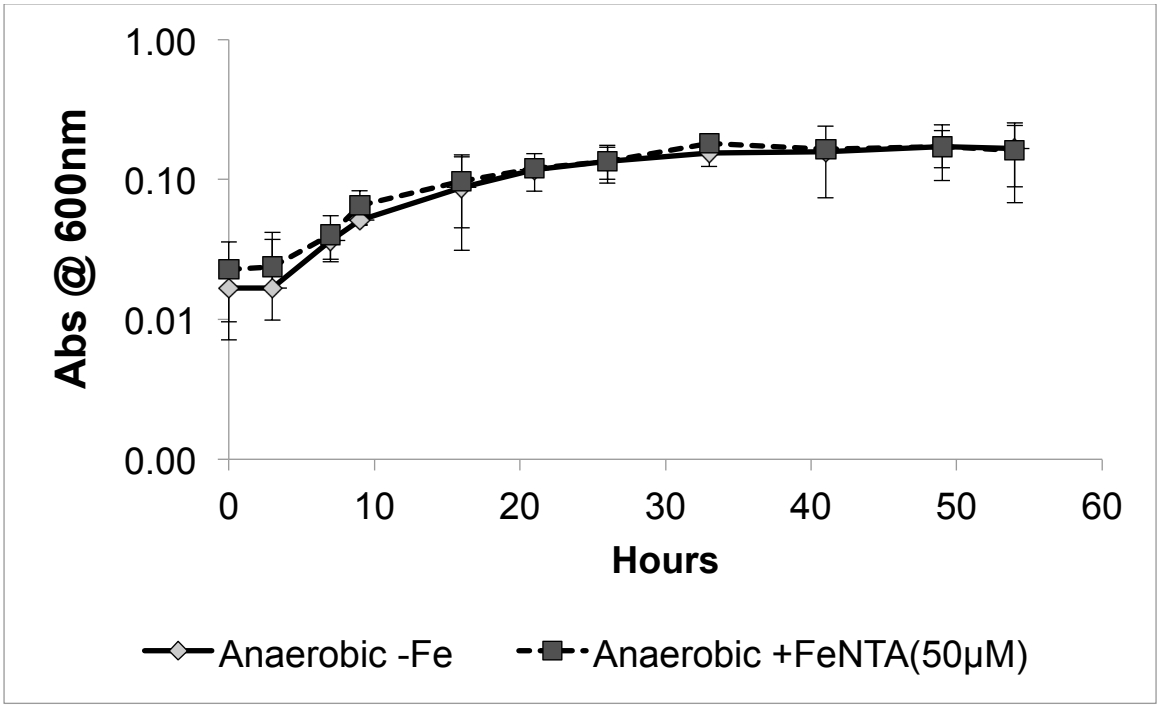
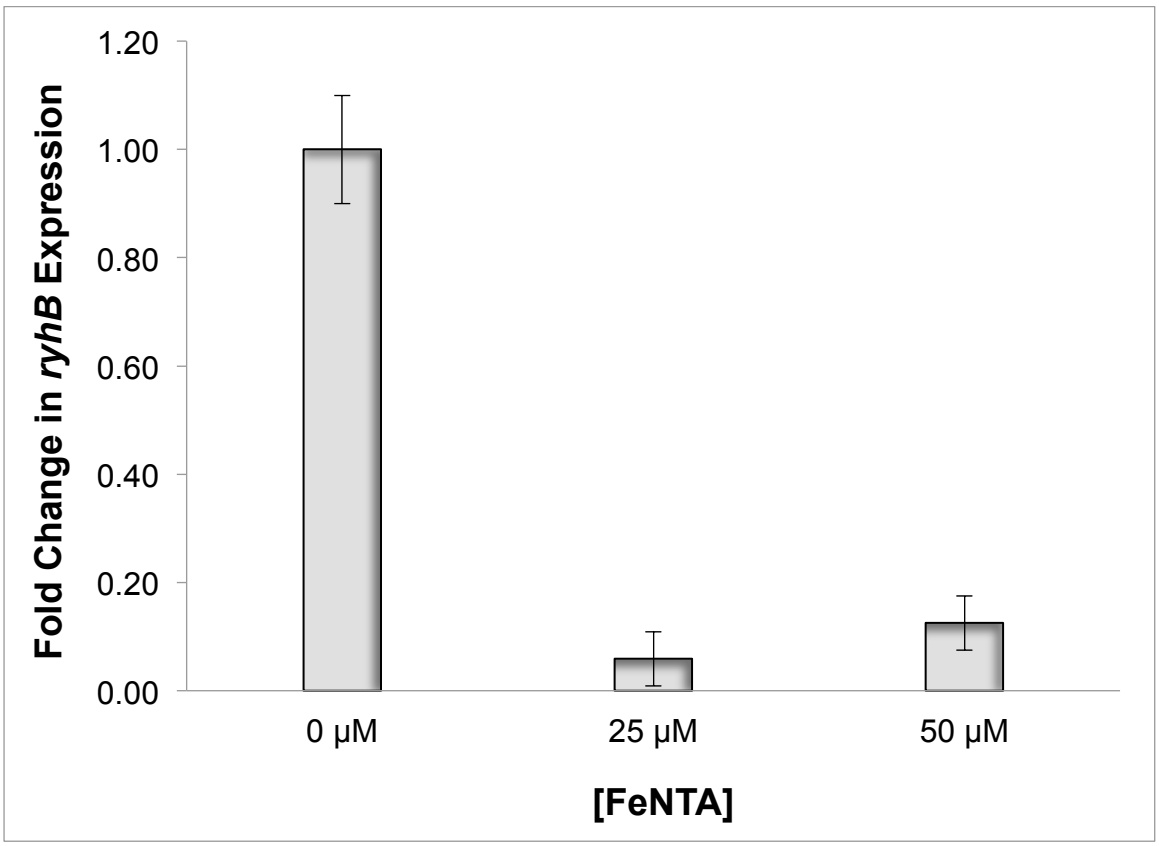

1.00
0.10
0.01
0.00
Abs @ 600nm
0
10
20
30
40
50
60
Hours
Anaerobic -Fe
Anaerobic +FeNTA(50μM)

Table 2-1. Liquid MICs of *S. oneidensis* MR-1 cells grown in HBa media supplemented with or lacking 25 μM FeNTA and various concentrations of 2,2'-dipyridyl (DIP).

HBa + 50 µM FeNTA + DIP		HBa - FeNTA + DIP	
[DIP] mM	**Growth**	**[DIP] mM**	**Growth**
0	+++	0	+++
0.050	+++	0.050	+++
0.075	+++	0.075	+++
0.100	+++	0.100	+
0.125	+	0.125	-
0.150	+	0.150	-
0.175	-	0.175	-

Figure 2-5. qPCR analysis on *ryhB* levels in *S. oneidensis* MR-1 cells grown to early log phase in the presence of iron supplementation. RNA was extracted from *S. oneidensis* MR-1 cells grown to early log phase in the presence of 0, 25, or 50 μM FeNTA. *ryhB* levels were normalized to the *tufB* housekeeping mRNA. Bars represent the fold change in *ryhB* expression from cells grown in the presence of FeNTA (25 or 50 μM) compared to cells grown in the absence of FeNTA, which was set to 1.


1.20
1.00
0.80
0.60
0.40
0.20
0.00
Fold Change in ryhB Expression
0 μM
25 μM
50 μM
[FeNTA]

Parallel experiments were performed using the iron chelator 2,2'-dipyridyl (DIP) to mimic an iron-deplete condition. Liquid MICs were performed to monitor *S. oneidensis* MR-1 growth in the presence of DIP concentrations ranging from 0.05-0.175 mM (Table 2-1). *ryhB* transcript levels from cells grown to mid-log phase in media containing 0, 0.05, 0.1, or 0.125 mM of DIP were determined through qPCR analysis. *ryhB* levels were increased 2.5-fold in the presence of 0.05 mM DIP, 7-fold in the presence of 0.1 mM DIP, and 2-fold in the presence of 0.125 mM DIP (Fig. 2-6). This increase in *ryhB* accompanied by the decreased *ryhB* in cells grown in the presence of FeNTA (Fig. 2-5) suggests that *ryhB* in *S. oneidensis* MR-1 is regulated in response to iron.

ryhB RNA levels in *S. oneidensis* MR-1 cells remained unchanged at 0, 7, 14, and 28 minutes after iron supplementation and chelation

To determine how quickly the cells can adjust *ryhB* expression in response to iron-replete or -deplete conditions, we assayed *ryhB* levels immediately following the addition of 25 μM FeNTA or 0.01 mM DIP to the media. Cells were grown in medium lacking an iron source to mid-log phase, at which time a supplement of FeNTA was added to 25 μM. Cell samples were harvested at 0, 7, 14, and 28 minutes after FeNTA supplementation and subjected to qPCR. Cells harvested at 0 minutes after FeNTA supplementation were compared to those not given any FeNTA as a control. When compared to levels at 0 minutes, *ryhB* levels were not increased at 7, 14, or 28 minutes after FeNTA supplementation (Fig. 2-7). We repeated this experiment but initially grew cells to mid-log phase in medium containing 25 μM FeNTA, added 0.1 mM DIP, and harvested cell samples at 0, 7, 14, and 28 minutes after DIP supplementation. Cells harvested at 0 minutes after DIP supplementation were compared to those not given any DIP as a control. qPCR analyses show that *ryhB* levels did not change at 7, 14, or 28 minutes after DIP supplementation when compared to levels at 0 minutes (Fig. 2-8). Our results indicate that it takes longer than 28 minutes for *S. oneidensis* MR-1 cells to change *ryhB* regulation in response to the iron-replete or -deplete conditions we tested.

Figure 2-6. qPCR analysis on *ryhB* levels in *S. oneidensis* MR-1 cells grown to early log phase in the presence of iron chelation. RNA was extracted from *S. oneidensis* MR-1 cells grown to early log phase in the presence of 0, 0.05, 0.1, 0.125 mM of 2,2'-dipyridyl (DIP). *ryhB* levels were normalized to the *tufB* housekeeping mRNA. Bars represent the fold change in *ryhB* expression from cells grown in the presence of DIP (0.05, 0.1, 0.125 mM) compared to cells grown in the absence of DIP, which was set to 1.

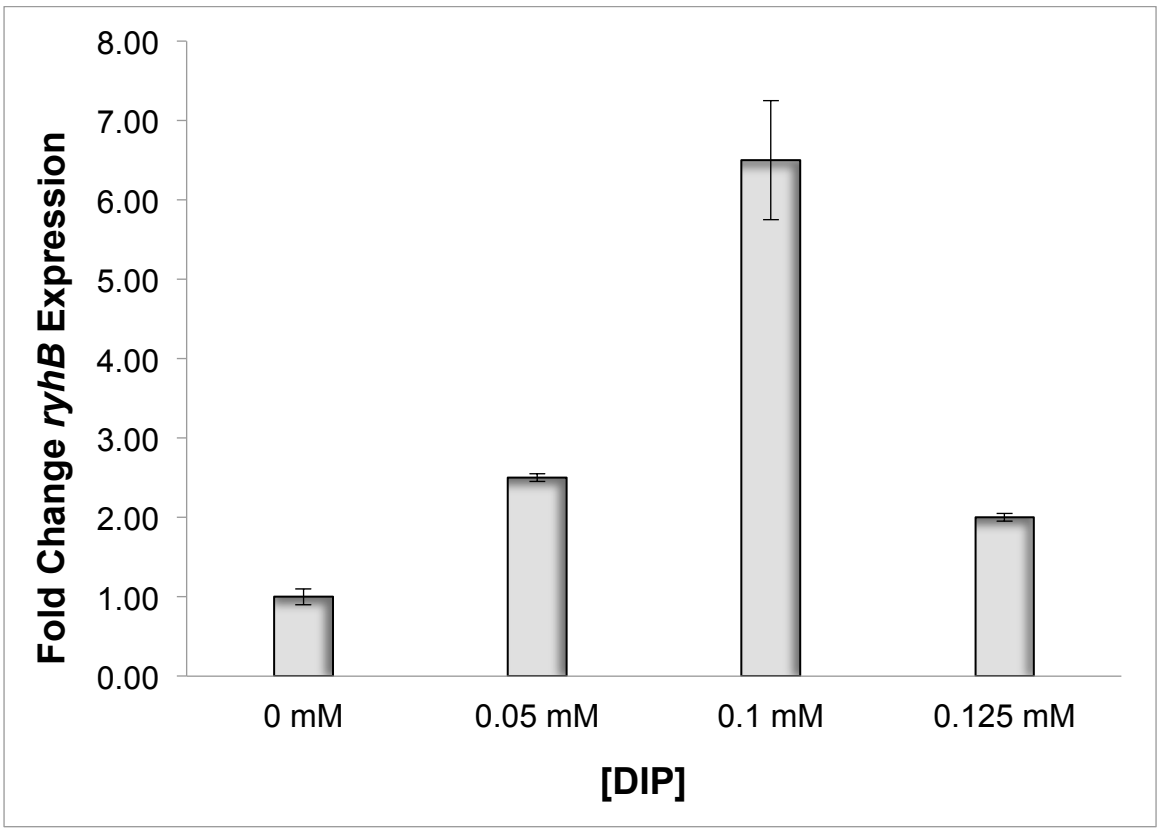

8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
Fold Change ryhB Expression
0 mM
0.05 mM
0.1 mM
0.125 mM
[DIP]

Figure 2-7. qPCR analysis on *ryhB* levels in *S. oneidensis* MR-1 cells at 0, 7, 14, and 28 minutes after iron supplementation. RNA was extracted from *S. oneidensis* MR-1 cells, which were grown to early log phase, and then given a supplement of 25 μM FeNTA and harvested at 0, 7, 14, and 28 minutes after FeNTA supplementation. Bars represent the fold change in *ryhB* expression from cells harvested at 7, 14, and 28 minutes after 25 μM FeNTA supplementation compared to cells harvested at 0 minutes, which was set to 1.

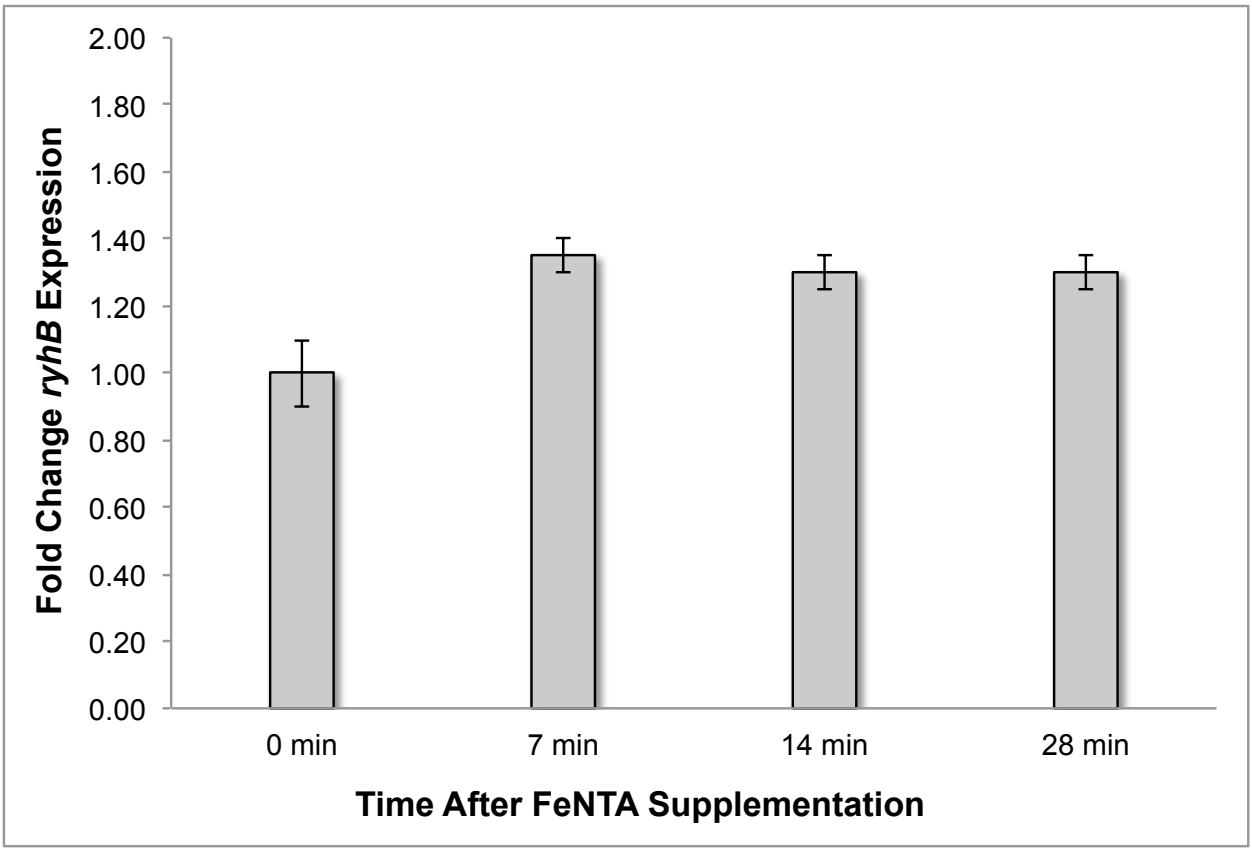
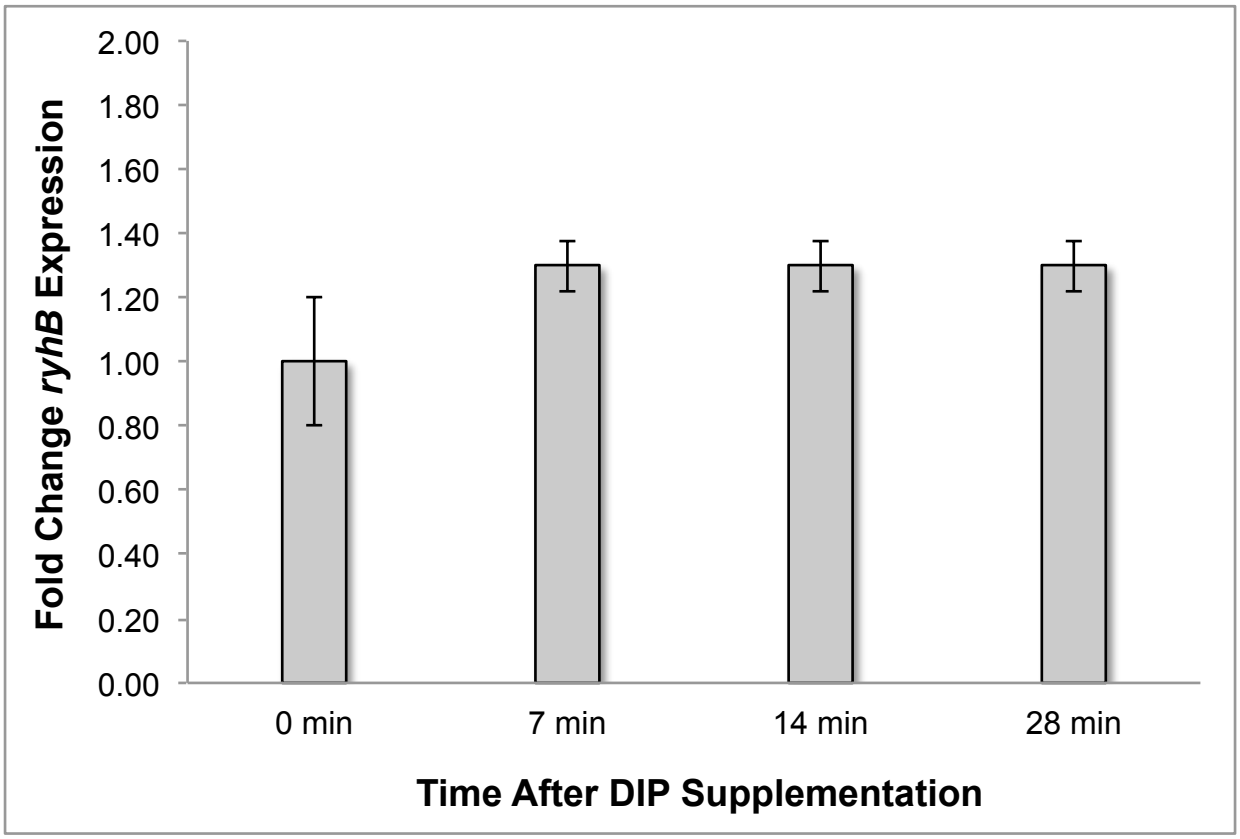

2.00
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Fold Change ryhB Expression
0 min
7 min
14 min
28 min
Time After FeNTA Supplementation

Figure 2-8. qPCR analysis on *ryhB* levels in *S. oneidensis* MR-1 cells at 0, 7, 14, and 28 minutes after addition of an iron chelator. RNA was extracted from *S. oneidensis* MR-1 cells which were grown to early log phase and then given a supplement of 0.1 mM 2,2'-dipyridyl (DIP) and harvested at 0, 7, 14, and 28 minutes after DIP supplementation. Bars represent the fold change in *ryhB* expression from cells harvested at 7, 14, and 28 minutes after DIP supplementation compared to cells harvested at 0 minutes, which was set to 1.


2.00
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Fold Change ryhB Expression
0 min
7 min
14 min
28 min
Time After DIP Supplementation

sodB RNA levels in mid-log phase *S. oneidensis* MR-1 cells did not change upon iron supplementation and chelation

Superoxide dismutase (SodB) is an iron-containing protein whose mRNA is a known target of *ryhB* in *E. coli* (Massé et al., 2005). Previous research showed a decreased level of *S. oneidensis* MR-1 *sodB* mRNA in a *fur* mutant (Wan et al., 2004). Therefore, we hypothesized that the decreased *sodB* levels may be a result of increased expression of *ryhB* in the absence of the Fur repressor. We examined *sodB* levels in cells grown to mid-log phase in the presence of 25 μM FeNTA to those grown in the absence of the iron source. qPCR was used to analyze *ryhB* and *sodB* levels from cells grown to mid-log phase in the presence of 25 μM FeNTA compared to those grown without the iron source. As expected, *ryhB* levels decrease in the presence of FeNTA (Fig. 2-9). No change was observed in *sodB* levels (Fig. 2-9). Similar analyses were performed on cells grown to mid-log phase in the presence of 0.1 mM DIP (Fig. 2-10). Whereas levels of *ryhB* were increased in the presence of DIP, *sodB* levels remained unchanged (Fig. 2-10). These results suggest that either the *sodB* mRNA is not a target of *ryhB* in *S. oneidensis* MR-1 or that the levels of *ryhB* were not increased high enough under the conditions tested to have an effect on *sodB* mRNA levels. It is also possible that our experimental design does not capture the complex regulation of *sodB* under the conditions tested.

Figure 2-9. qPCR analysis on *ryhB* and *sodB* levels in *S. oneidensis* MR-1 cells grown to early log phase in the presence of iron supplementation. RNA was extracted from *S. oneidensis* MR-1 cells grown to early log phase in the presence of 25 μM FeNTA. *ryhB* and *sodB* levels were normalized to the *tufB* housekeeping mRNA. Bars represent the fold change in *ryhB* and *sodB* expression from cells grown in the presence of 25 μM FeNTA compared to cells grown in the absence of FeNTA, which was set to 1.

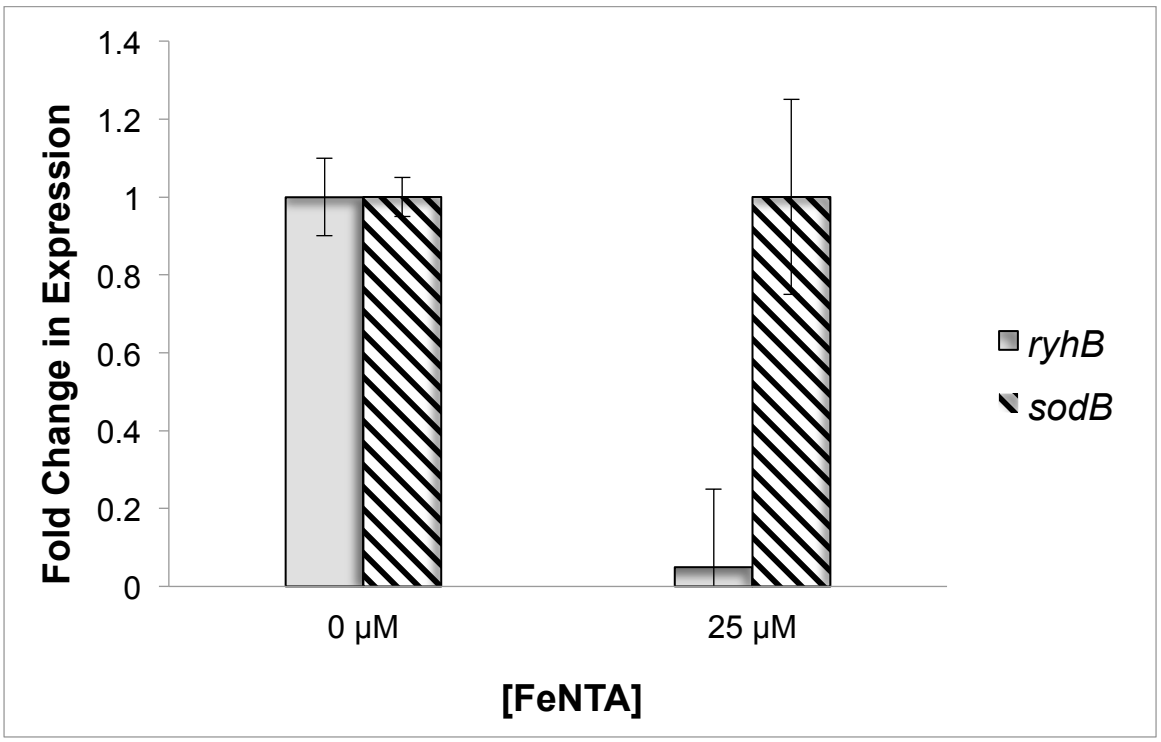
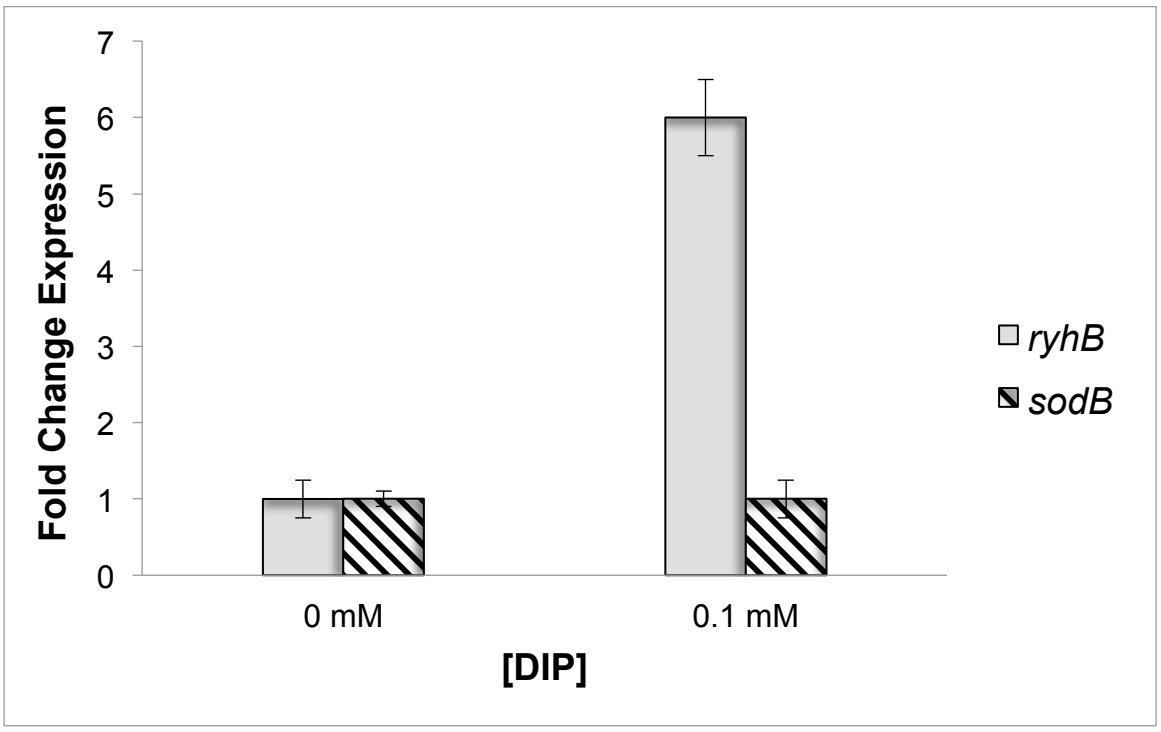

1.4
1.2
1
0.8
0.6
0.4
0.2
0
Fold Change in Expression
0 μM
25 μM
[FeNTA]
ryhB
sodB

Figure 2-10. qPCR analysis on *ryhB* and *sodB* levels in *S. oneidensis* MR-1 cells grown to early log phase in the presence of an iron chelator. RNA was extracted from *S. oneidensis* MR-1 cells grown to early log phase in the presence of 2,2'-dipyridyl (DIP). *ryhB* and *sodB* levels were normalized to the *tufB* housekeeping mRNA. Bars represent the fold change in *ryhB* and *sodB* expression from cells grown in the presence of 0.1 mM DIP compared to cells grown in the absence of DIP, which was set to 1.


Fold Change Expression
7
6
5
4
3
2
1
0
0 mM
0.1 mM
[DIP]
ryhB
sodB

Construction of *ryhB* expression vectors

To achieve higher levels of *ryhB* in *S. oneidensis* MR-1, we designed expression vectors that placed transcription of the RNA under the control of the P_{lac} and P_{BAD} promoters. Our intention was to use these vectors to induce high levels of *ryhB* and perform large scale transcriptome analyses to identify those mRNAs whose levels were significantly reduced as a result of increased *ryhB*. First, we designed the vector, pACYC-*ryhB* (Kan^R), which placed *ryhB* under the transcriptional control of the *E. coli lac* promoter (Fig. 2-11B). Growth curves were obtained for cells containing pACYC or pACYC-*ryhB* in the presence or absence of 25 μM FeNTA. Cells that contained either the pACYC or the pACYC-*ryhB* expression vector experienced a lag in growth compared to cells that did not contain any vector regardless of FeNTA supplementation (Fig. 2-12). Cells containing the empty pACYC vector grew slightly slower than those containing pACYC-*ryhB* in the absence of FeNTA, but eventually reached a similar OD_{600} (Fig. 2-12A).

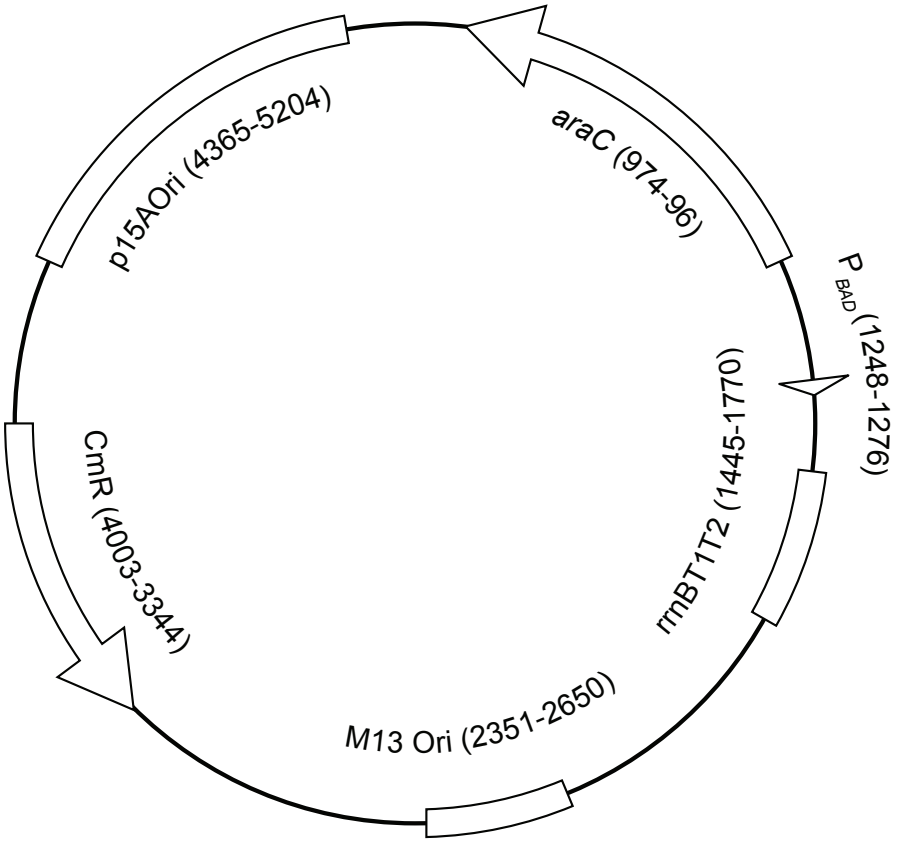
qPCR was utilized to compare levels of *ryhB* in cells grown in medium lacking FeNTA that contained the empty vector or the *ryhB* expression vector. Cells containing pACYC-*ryhB* only had a 2-fold increase in *ryhB* levels compared to cells that did not contain either vector (Fig. 2-13). We observed a greater increase simply by growing wild-type cells in the presence of DIP (Fig. 2-6). The presence of pACYC alone did not have any effect on *ryhB* levels (Fig. 2-13). The pACYC-*ryhB* vector provided *ryhB* transcription under control of the P_{lac} promoter at constant, high activity. It is possible that this constant and high expression of *ryhB* (from the time of media inoculation through mid-log phase when the cells were harvested) negatively impacted fitness, causing cells to reduce expression by rearranging the plasmid vector, resulting in only a 2-fold increase in expression (Fig. 2-13).

pBAD33-*ryhB* (Cm^R) was constructed to place *ryhB* transcription under the control of the arabinose inducible promoter, P_{BAD} (Fig. 2-11A, C, D). Modified versions of this vector were generated which contained the *lacZ* fusions of the lambda repressor gene, *c*I, under transcriptional control of P_{BAD} (see Appendix B, Materials and Methods, Table B-1, Fig. B-9). Expression of these vectors in *E. coli* was induced with L-

arabinose and confirmed by measuring β-galactosidase activity (see Appendix B, Table B-1, Fig. B-9). β-galactosidase assays were performed on *S. oneidensis* MR-1 cells expressing the pBAD33-LL-*c*I-*lacZ* (Cm^R) expression vector to estimate expression levels from the induced promoter. The *c*I sequence is cloned at the transcriptional start position in this vector and is a good representation of *ryhB* expression at this position as both are transcribed from the +1 position in the vector. Early log phase cells were induced with 0, 0.2, and 1% of L-arabinose. Samples were harvested at 0, 30, 60, and 120 minutes after L-arabinose induction and assayed for β-galactosidase activity. We did not observe any β-galactosidase activity (Fig. 2-14), indicating that this expression vector was inactive in *S. oneidensis* MR-1 cells. To rule out the possibility that the lack of activity resulted from the inability of the cells to express leaderless messages efficiently, we assayed *S. oneidensis* MR-1 cells that contained pBAD33-SDL*c*I-*lacZ* (Cm^R), in which we added a 5' untranslated leader sequence to the *c*I gene (see Appendix B, Fig. B-2). The assay was repeated and again, we did not observe any β-galactosidase activity (Fig. 2-14), suggesting that the lack of activity of pBAD33-LL*c*I-*lacZ* did not result from an inability to express the leaderless message in *S. oneidensis* MR-1 cells. It is more likely that the there was a problem inducing the promoter, P_{BAD}. It is also possible that the expression vectors were subjected to rearrangement by the cells.

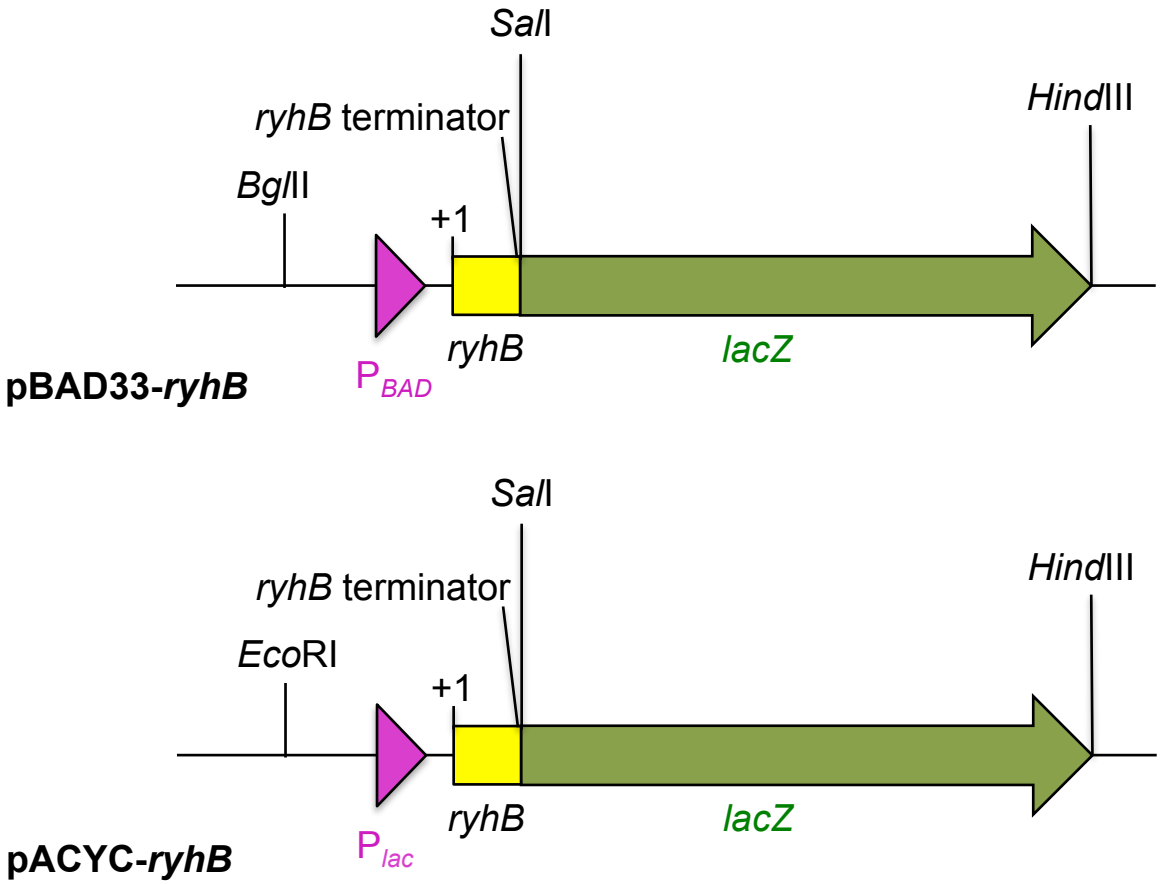
Figure 2-11. Descriptions of the pBAD33-*ryhB* and pACYC-*ryhB* expression vectors. A. Map of the plasmid, pBAD33. B. Partial map of the pBAD33-*ryhB* and pACYC-*ryhB* expression constructs showing the *Bgl*II, *Eco*RI, *Sal*I, and *Hind*III restriction sites, P_{BAD} and P_{lac} promoters, and *ryhB* sequence including the predicted intrinsic terminator. C. Partial sequence of the pBAD33-*ryhB* construct showing the *Sal*I restriction site and *ryhB* sequence beginning at the transcriptional start site (+1).

A.


p15AOri (4365-5204)
araC (974-96)
P BAD (1248-1276)
rrnBT1T2 (1445-1770)
CmR (4003-3344)
M13 Ori (2351-2650)

pBAD33

B.



C.

+1

...GTTTCTCCATGCGTTCCAAAACTCATCTTTAACTCTGATGACTGGTAATCTGACGTTGTTAAAGTGCTC

ryhB

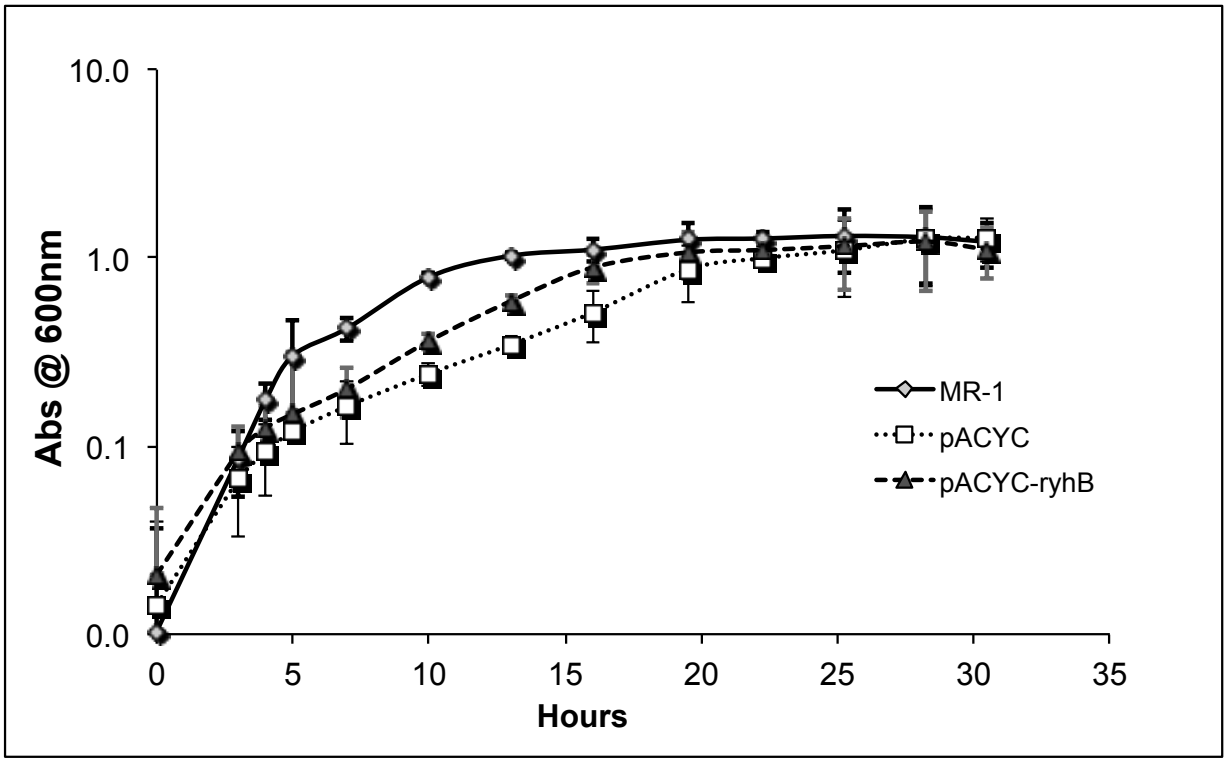
CTGAGTTTGCAAGCACGACATTGCTCACACACTCTTTTGTGGCCGGATTCATCATCCGGCTTTTTTTTG

CCTGTTATTCAGCGAATAGGCGCATTAGGCTAACCATGTCGTCGACACGGATTCACTGG...

SalI

Figure 2-12. Growth curves of *S. oneidensis* MR-1 cells in HBa medium with the empty pACYC vector and pACYC-*ryhB*. Cells grown in HBa without FeNTA (A) or with 25 μM FeNTA (B). Growth was monitored on a spectrophotometer at 600 nm.

A.


10.0
1.0
0.1
0.0
Abs @ 600nm
0
5
10
15
20
25
30
35
Hours
MR-1
pACYC
pACYC-ryhB

B.

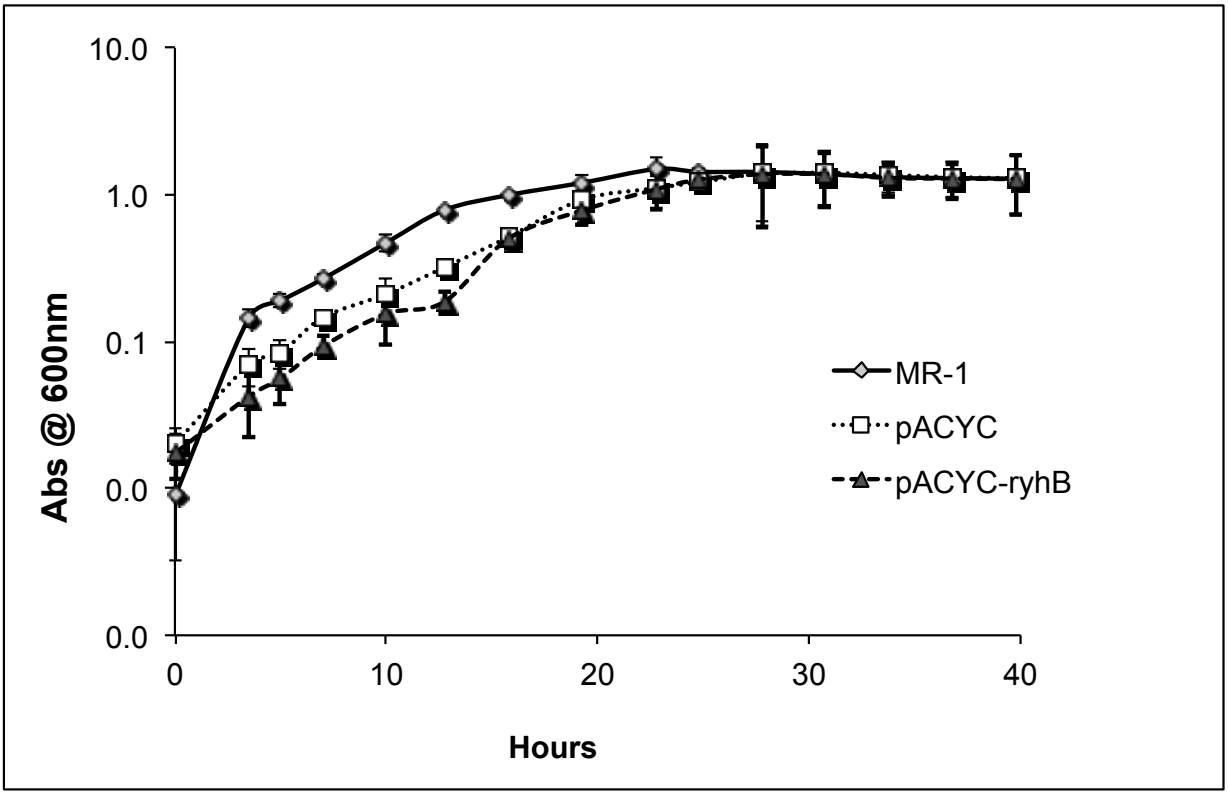

10.0
1.0
0.1
0.0
Abs @ 600nm
0
10
20
30
40
Hours
MR-1
pACYC
pACYC-ryhB

Figure 2-13. qPCR analysis of *ryhB* levels in *S. oneidensis* MR-1 cells containing the pACYC-*ryhB* expression vector. RNA was extracted from *S. oneidensis* MR-1 cells with pACYC-*ryhB* in HBa medium (-FeNTA). *ryhB* levels were normalized to the *tufB* housekeeping mRNA. Bars represent the fold increase in *ryhB* expression from cells containing the pACYC empty vector and the pACYC-*ryhB* expression vector compared to expression from cells without vector, which was set to 1.

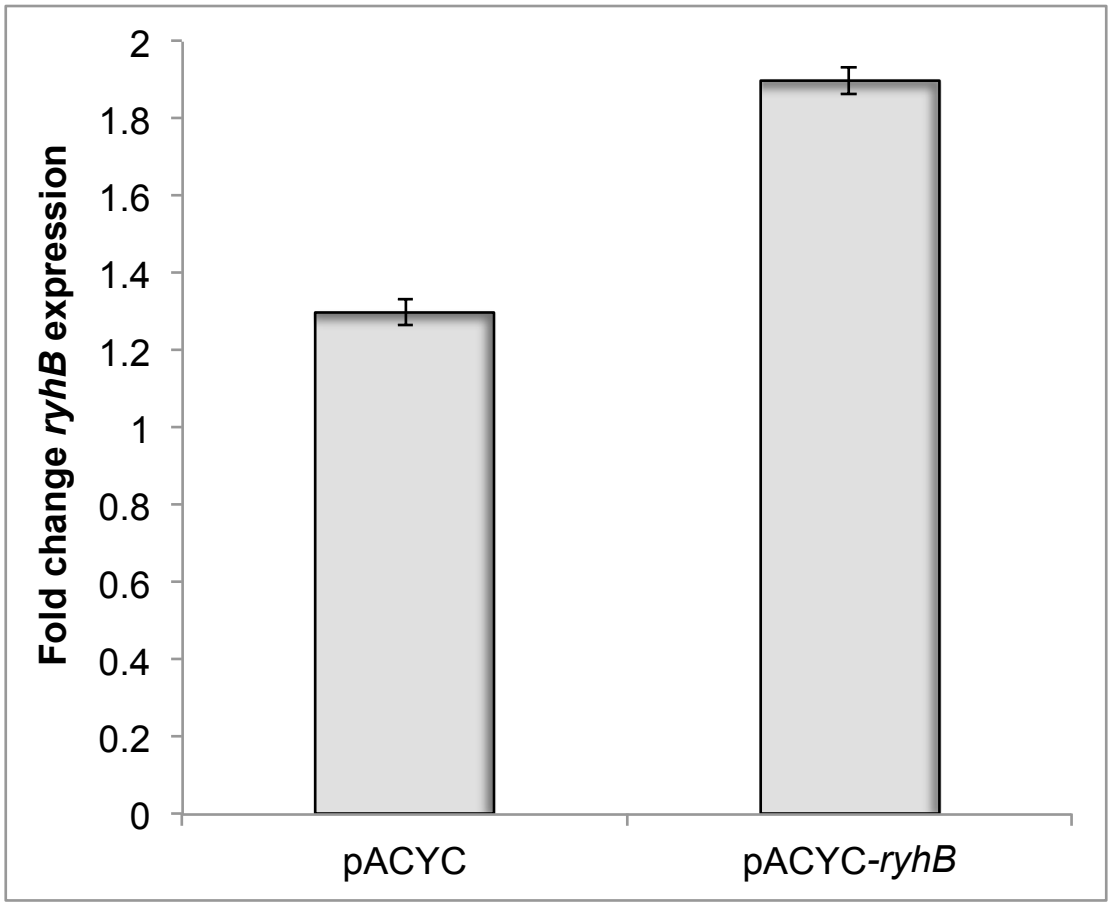
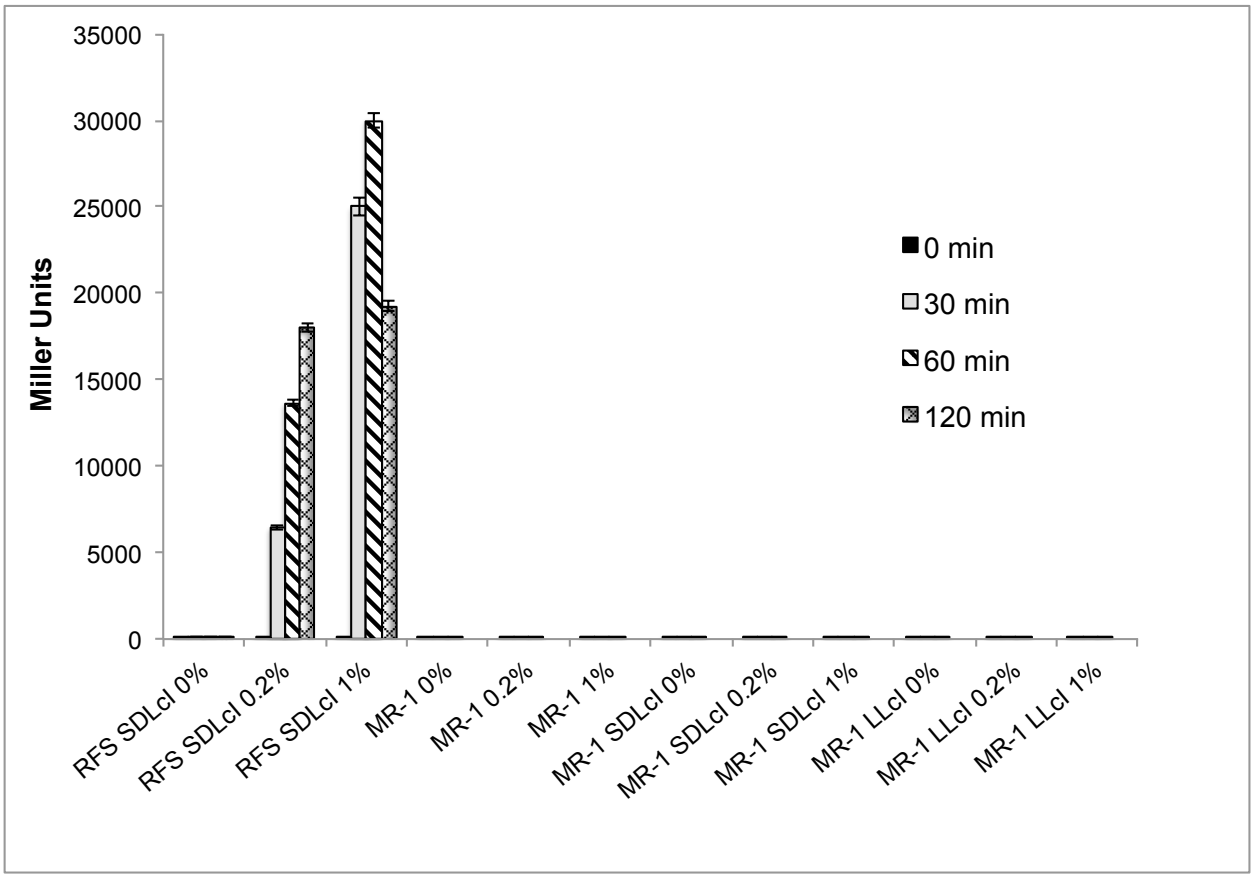

Fold change *ryhB* expression
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
pACYC
pACYC-*ryhB*

Figure 2-14. β-galactosidase assays were performed on *S. oneidensis* MR-1 cells containing the pBAD33-SDL*c*I-*lacZ* or pBAD33-LL*c*I-*lacZ* expression vectors. Cells were grown to early log phase in 2XYT media and then L-arabinose was added at concentrations of 0, 0.2, and 1%. Samples were pulled at 0, 30, 60, and 120 minutes after L-arabinose induction and assayed for β-galactosidase activity. *E. coli* RFS859 containing both expression vectors was used as a positive control strain.


35000
30000
25000
20000
15000
10000
5000
0
Miller Units
0 min
30 min
60 min
120 min
RFS SDLcI 0%
RFS SDLcI 0.2%
RFS SDLcI 1%
MR-1 0%
MR-1 0.2%
MR-1 1%
MR-1 SDLcI 0%
MR-1 SDLcI 0.2%
MR-1 SDLcI 1%
MR-1 LLcI 0%
MR-1 LLcI 0.2%
MR-1 LLcI 1%

Discussion

ryhB is an sRNA that was originally identified in *E. coli* and characterized for its role in iron regulation. A number of homologues and *ryhB*-like sRNAs have since been discovered in a variety of organisms ranging from other closely related enterobacterial species to distantly related Gram-positive organisms. A *ryhB* homolog was identified in *S. oneidensis* MR-1 (Wan et al., 2004) that shares some sequence similarity to *ryhB* in *E. coli* and appears to be regulated in response to iron, similar to the regulation in *E. coli* (Massé and Gottesman, 2002, Massé et.al., 2005). A palindromic sequence typical of a Fur box has been identified in the promoter region of this *ryhB* homologue (Wan et al., 2004), further suggesting a role in iron metabolism. We have yet to show experimentally that Fur recognizes the sequence and binds to this region of the DNA, but we predict that Fur binds to the promoter of *ryhB* in *S. oneidensis* MR-1 when iron is bountiful and represses expression of this sRNA.

Since our results indicate that *ryhB* expression changes in response to iron availability, we tried to capture a reference of time required for cells to induce these changes. We did not observe any changes in expression within the first 28 minutes in response to supplementation of the iron source (FeNTA) or iron chelator (DIP). It is possible, given that *S. oneidensis* MR-1 has a generation time of 3 hours in the growth medium, that our time frame was not long enough for cells to adjust to the new conditions. These experiments could be repeated, with samples taken intermittently over longer durations to determine when the cells begin to regulate *ryhB*.

Given the regulatory nature of sRNAs to act on multiple targets and employ diverse mechanisms (Altuvia, 2004, Gottesman, 2004) of action we predict that *ryhB* in *S. oneidensis* MR-1 acts quickly once it is initially expressed. sRNAs do not need to be translated in order to be active, therefore the ability of this sRNA to induce a rapid global response once activated might obviate the need for the organism to induce expression immediately after exposure to the iron chelator.

Although the *ryhB* we identified in *S. oneidensis* MR-1 shares characteristics with *ryhB* of *E. coli*, there are key differences between the two, which may imply differences

in regulation within the cell. Regarding sequence similarities, the former is predicted to be 47 nucleotides longer than the latter. The 5'-end of the sRNA has been mapped using primer extension analysis but the 3'-end has been predicted through computational analysis of factor-independent stem loop terminators. Mapping of the 3'-end of this sRNA needs to be performed to verify its precise size. The sequences at positions 1-27 and 62-69 of the *S. oneidensis* MR-1 sRNA do not share high similarity with those of the sRNA in *E. coli*. Computational analysis (mfold) predicts that these regions are not sequestered in secondary structure stem loops, which would make them available to bind mRNA targets. The individuality of these sequences suggests that their target mRNAs might be different than those targets of *E. coli ryhB*.

One of the *ryhB*-regulated genes documented in *E. coli* is *sodB*, encoding an Fe-containing superoxide dismutase (Massé and Gottesman, 2002). Our results suggest that *sodB* is not a target of *ryhB* in *S. onediensis* MR-1. One explanation is that the change in *ryhB* expression we observed was not sufficient to alter the *sodB* target. *ryhB* might have a preference for other target mRNAs, binding to them more quickly; a higher concentration of the sRNA might be needed to bind a wider variety of target mRNAs. Another explanation is that *sodB* is not a target of *ryhB* in *S. oneidensis* MR-1. Even under conditions of iron limitation, *S. oneidensis* MR-1 cells might not be able to forgo the protective role of the Fe-containing SodB. *E. coli* has additional Fe-independent superoxide dismutases that can substitute for SodB in its protective role against oxidative damage (Imlay and Linn, 1988, Compan and Touati, 1993).

In this study not only were we interested in characterizing this sRNA in *S. oneidensis* MR-1 but we also wanted to identify its target mRNAs. In order to ensure that we could achieve concentrations of *ryhB* high enough to interact with all of the targets we placed transcription of the sRNA under control of the high expression vectors, the P_{lac} and P_{BAD} promoters, which had been used previously in *S. oneidensis* MR-1 studies (Myers and Myers, 1997, Bordi, et al., 2003, Thormann, et al., 2006, Gunton et al., 2007).

We were not able to efficiently produce high levels of *ryhB* from either of the expression vectors in *S. oneidensis* MR-1. The range of conditions used to perform the

assays (% inducer, induction time, medium) should have met the requirements for vector expression. The inability to induce expression from the plasmids most likely resulted from a problem that arose once the plasmids were electroporated into the *S. oneidensis* MR-1 cells because we were able to produce efficient expression from the inducible promoter in *E. coli*, suggesting that the vectors were designed properly. One plausible explanation is that the expression vectors were subjected to rearrangement. When we extracted and analyzed them via gel electrophoresis, we did not observe the expected banding pattern; rather, we saw many extraneous bands and an unusual amount of DNA smearing (data not shown), suggesting that the plasmid DNA that originally went into the cells was not the same that was extracted. Mutation of the *ryhB* locus on the chromosome would be a more direct approach to study the role of the sRNA in cellular physiology and identify mRNA targets whose expression is increased in the absence of the regulatory sRNA.

Microarray analysis would allow for a global approach to identifying targets of *ryhB* in *S. oneidensis* MR-1. Genes that are downregulated in cells expressing *ryhB* from the expression vector would be candidate *ryhB* target mRNAs. The resulting data could be cross-referenced with the genes that are downregulated in the *fur* mutant (Wan et al., 2004) as inhibition by *ryhB*, which itself is upregulated in the *fur* mutant, would offer a likely explanation. Some of the downregulated genes identified as *ryhB* targets might actually be indirect targets of the sRNA as overproduction of *ryhB* may increase the iron concentration inside the cell by down-regulating many Fe-binding proteins, leading to Fur repression of these genes. One way to distinguish between the direct and indirect *ryhB* targets is to overexpress *ryhB* in a *fur* mutant. Those genes that are downregulated upon *ryhB* overexpression in the wild-type cells but whose expression did not change upon *ryhB* overexpression in the *fur* mutant would be likely indirect targets of the sRNA.

Once candidate targets of *ryhB* in *S. oneidensis* MR-1 are identified, they could be used to further characterize how the sRNA works. We mentioned earlier that the unpaired regions of this sRNA might represent the area of the RNA that binds to the

targets. This hypothesis could be tested by mutating nucleotides within the region and testing to see if a change in the pattern of regulation of the candidate mRNAs results.

ryhB and the *ryhB* homologues require the chaperone Hfq for proper function (Massé et al., 2003, 2005; Gottesman and Storz, 2011). Although the *S. oneidensis* MR-1 genome encodes Hfq, it is not yet clear whether *ryhB* in this organism also requires the chaperone. Furthermore, we have yet to determine whether RNase E is solely responsible for the degradation of the target mRNAs, like *ryhB* in *E. coli*, or if RNase III is also used.

Our characterization of *S. oneidensis* MR-1 *ryhB* supports a resemblance to *ryhB* of *E. coli* on the sequence level and function in response to iron availability. Although these two sRNAs share some sequence similarity, they are not identical and the regions unique to *ryhB* in *S. oneidensis* MR-1 probably recognize target mRNAs that the *E. coli ryhB* does not. These two sRNAs have likely evolved to suit each organism's physiology optimally.

Appendix A

Identification of sequences required for ribosome binding to uAUG-2 using oligonucleotides to sequester regions of the *aroL* 5'UTR or the 3' end of the 16S rRNA

For all toeprint assays described in this section, we used T7 RNA polymerase to synthesize *aroL* mRNA *in vitro* that included the full *aroL* leader plus 16 codons of the *aroL* coding sequence. Refer to Chapter 1 (Materials and Methods) for details regarding transcription with T7 RNA polymerase.

Identification of sequences required for ribosome binding to uAUG-2 using oligonucleotides complementary to regions of the *aroL* mRNA in toeprint assays (*aroL* leader sequestration toeprints)

In addition to the deletion and substitution analysis of the *aroL* untranslated leader described in Chapter 1, we took another approach to identify the sequences responsible for ribosome recognition and binding to uAUG-2. Using multiple strategies at the same time increased the odds of successfully identifying recognition signals within the sequence. We reasoned that if we identified the same recognition signals via two independent methods, our argument for their role in ribosome recognition and binding could be strengthened.

We performed toeprint assays using mRNA pre-bound with DNA oligonucleotides that were complementary to short overlapping regions of the *aroL* mRNA, upstream of uAUG-2. We predicted that the oligonucleotide would anneal to the mRNA, thereby sequestering the sequence needed for ribosome binding, and would result in a loss of toeprint signal for uAUG-2. The inhibitory effects of these sequestering oligonucleotides were analyzed individually and in combination. DNA oligonucleotides were designed that were complementary to various regions of the mRNA, labeled A-H (Fig. A-1 and Table A-1). Oligonucleotides A, G, and H were controls. Oligonucleotide A was designed as a negative control, complementary to a region spanning +1 - +12 of the *aroL* mRNA (+1 taken as the transcriptional start site); deletion analysis (see Chapter 1,

Fig. 1-5) suggested this region was not needed for 30S subunit binding to, or expression from, uAUG-2. Oligonucleotide G was designed as a positive control, complementary to the region spanning +58 - +69 of the mRNA and containing the uAUG-2 start site; binding of this oligonucleotide to the *aroL* mRNA would sequester uAUG-2 and thus prevent 30S subunit binding. Oligonucleotide H was designed as a second positive control, complementary to positions +110 - +121 and spanning the *aroL* SD sequence. The expression data described in Chapter 1 (Fig. 1-3) show the importance of the SD sequence for a*roL* expression; therefore, sequestering it with an oligonucleotide is predicted to result in decreased binding to the *aroL* AUG start codon. Oligonucleotides B-F are test oligonucleotides that are complementary to various regions within +13 - +47 of the *aroL* mRNA and partially overlap each other. This region was implicated by the deletion analysis (see Chapter 1, Fig. 1-5) as containing the sequence responsible for 30S subunit binding and expression from uAUG-2. Specifically, oligonucleotide B was complementary to positions +13 - +24, C to positions +22 - +32, D to positions + 25 - +36, E to positions +30 - +41, and F to positions +36 - +47 of the *aroL* mRNA (Fig. A-1 and Table A-1).

Figure A-1. Sequence of the *aroL* mRNA 5' untranslated leader region extending from the start site of transcription (+1) to the stop codon of the upstream open reading frame, overlapping the *aroL* start codon. Horizontal lines (purple, A-H) indicate annealing sites for complementary DNA oligonucleotides to block ribosome binding in the toeprint sequestration assays. Ribosome binding and expression studies suggest the ribosome binding signals lie in between or partially overlapping the +26 and +45 positions of the leader region upstream of the uAUG triplets. The uAUGs in the leader are numbered (blue) and the *aroL* AUG start codon is highlighted in red.

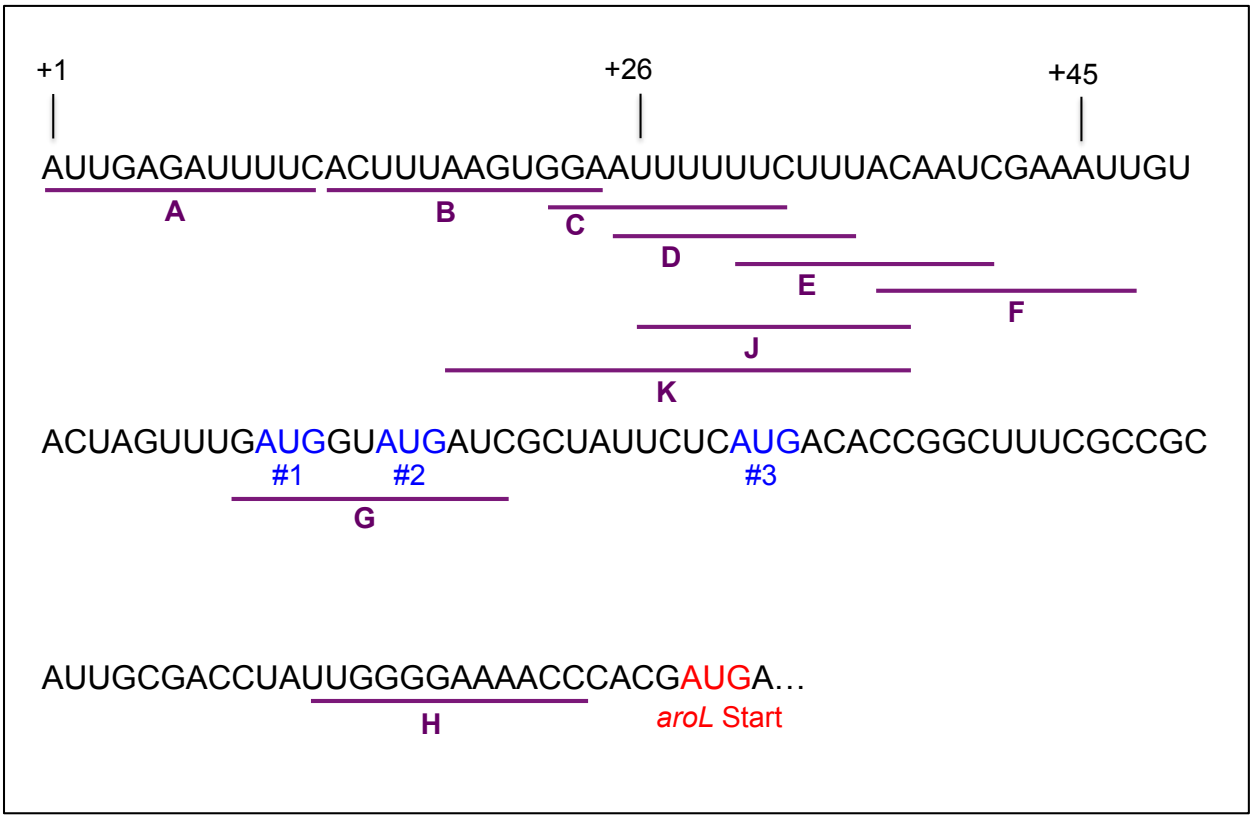

+1
+26
+45
AUUGAGAUUUUCACUUUAAGUGGAAUUUUUUCUUUACAAUCGAAAUUGU
A
B
C
D
E
F
J
K
ACUAGUUUGAUGGUAUGAUCGCUAUUCUCAUGACACCGGCUUUCGCCGC
#1
#2
#3
G
AUUGCGACCUAUUGGGGAAAACCCACGAUGA…
H
aroL Start

Table A-1. Descriptions and sequences of DNA and RNA oligonucleotides used in sequestration assays to identify sequences in the *aroL* 5' untranslated leader region responsible for ribosome recognition and binding to uAUG-2. Sites of complementarity to the *aroL* mRNA are depicted in Figure A-1.

Oligonucleotide	Oligonucleotide Sequence	Description (+1 is the *aroL* transcriptional start site)
A	5'-GAAAATCTCAAT-3'	Negative control DNA oligo; complementary to positions +1 - +12 of *aroL* mRNA
B	5'-TCCACTTAAAGT-3'	Test DNA oligo; complementary to positions +13 - +24 of *aroL* mRNA
C	5'-GAAAAAATTCC-3'	Test DNA oligo; compelementary to positions +22 - +32 of *aroL* mRNA
D	5'-TAAAGAAAAAAT-3'	Test DNA oligo; complementary to positions +25 - +35 of *aroL* mRNA
E	5'-GATTGTAAAGAA-3'	Test DNA oligo; complementary to positions +30 - +41 of *aroL* mRNA
F	5'-AATTTCGATTGT-3'	Test DNA oligo; complementary to positinos +36 - +47 of *aroL* mRNA
G	5'-GATCATACCATC-3'	Positive control DNA oligo; complementary to positions +58 - +69 of *aroL* mRNA
H	5'-GGTTTTCCCCAA-3'	Positive control DNA oligo; complementary to positions +110 - +121 of *aroL* mRNA, spanning *aroL* SD
J	5'-GTAAAGAAAAAA-3'	Test DNA oligo; complementary to positions +26 - +37 of *aroL* mRNA
K	5'-GTAAAGAAAAAATTCCACTT-3'	Test DNA oligo; complementary to positions +18 - +37 of *aroL* mRNA
16S.9mer ASD	5'-TAAGGAGGT-3'	Positive control DNA oligo; Complementary to 3' end of *E.coli* 16S rRNA
16S.11mer	5'-GTGATCCAACC-3'	Positive control DNA oligo; Complementary to 3' end of *E.coli* 16S rRNA
16S.13mer	5'-TAAGGAGGTGATC-3'	Positive control DNA oligo; Complementary to 3' end of *E.coli* 16S rRNA
16S.13mer	5'-UAAGGAGGUGAUC-3'	Positive control RNA oligo; Complementary to 3' end of *E.coli* 16S rRNA
RNA.Neg	5'-AAAAAAGAAAUG-3'	Negative control RNA oligo; represents complement of positions +26 - +37 (U-rich region) of *aroL* mRNA
UpRBS2	5'-UUUUUUCUUUAC-3'	Test RNA oligo; represents positions +26 - +37 (U-rich region) of *aroL* mRNA
lac.comp2	5'-ATTAAGTTGGGTAACGCCAG-3'	Radiolabeled and annealed to the *aroL* mRNA to prime reverse transcription

To perform the toeprint assays, we followed the procedure essentially as described in Chapter 1 with the following modification. After annealing radiolabeled lac.comp2 primer to *aroL* mRNA, the potentially sequestering test oligonucleotides were added at 20X and 100X the mRNA concentration and incubated with the mRNA an additional 15 minutes at 37^0C. After this additional incubation for oligonucleotide annealing, $tRNA^{fMet}$, dNTPs, and 30S subunits were added and the toeprint reactions were carried out according to protocol (see Chapter 1, Materials and Methods).

Addition of the positive control oligonucleotide H abolished binding to the *aroL* AUG start codon, both at 20X and 100X over mRNA concentration (Fig. A-2, lanes 4 and 5). Likewise, addition of the positive control oligonucleotide G abolished binding to uAUG-2 both at 20X and 100X over mRNA concentration (Fig. A-2, lanes 6 and 7). Addition of the negative control oligonucleotide A at 20X or 100X over mRNA concentration did not change 30S subunit binding to uAUG-2 or the *aroL* AUG, as expected (Fig. A-2, lanes 8 and 9).

Addition of test oligonucleotides B, C, D, and E at concentrations of 20X and 100X did not result in a change in toeprint signal to uAUG-2 (Fig. A-3, lanes 18-28). Addition of oligonucleotide F at 100X over mRNA concentration, however, resulted a slight decrease in the uAUG-2 toeprint signal (Fig. A-3, lanes 31-32).

We repeated the assays using a 400X concentration of oligonucleotide to mRNA and saw an increased reduction in 30S subunit binding to uAUG-2. The control oligonucleotides H, G, and A (Fig. A-4, lanes 2-4) gave similar binding patterns to those observed at 20X or 100X concentrations (Fig. A-2). Although addition of oligonucleotides B, C, D or E (Fig. A-4, lanes 5-8) did not affect binding to uAUG-2, addition of oligonucleotide F (Fig. A-4, lane 9) substantially decreased the uAUG-2 toeprint signal. The reactions containing oligonucleotide E (Fig. A-4, lanes 17 and 19) showed a slightly reduced uAUG-2 toeprint signal and those containing oligonucleotide F (Fig. A-4, lanes 18, 20, 21) showed a pronounced reduction in signal. The data suggest that oligonucleotides E and F bound and sequestered the sequence needed for 30S subunit recognition and/or binding to uAUG-2. None of the test oligonucleotides B-F affected binding to the *aroL* AUG start (Fig. A-4, lanes 5-9).

Figure A-2. Primer extension inhibition (toeprint) analysis of 30S subunit binding to *aroL* uAUG-2 and the *aroL* AUG start codon in the presence of control oligonucleotides designed to hybridize with parts of the *aroL* leader. The mRNA used in these assays included the full *aroL* leader plus 16 codons of the *aroL* coding sequence. 30S subunits were added at a concentration of 10x mRNA. Oligonucleotides were added at 20X and 100X concentrations over mRNA. The positions of the full-length message, toeprint signals for uAUG-2, uAUG-3 and *aroL* AUG start codon, and predicted position of annealed oligonucleotides are indicated by arrows. DNA oligonucleotide sequences are listed in Table A-1. Sites of complementarity to the *aroL* mRNA are depicted in Figure A-1.

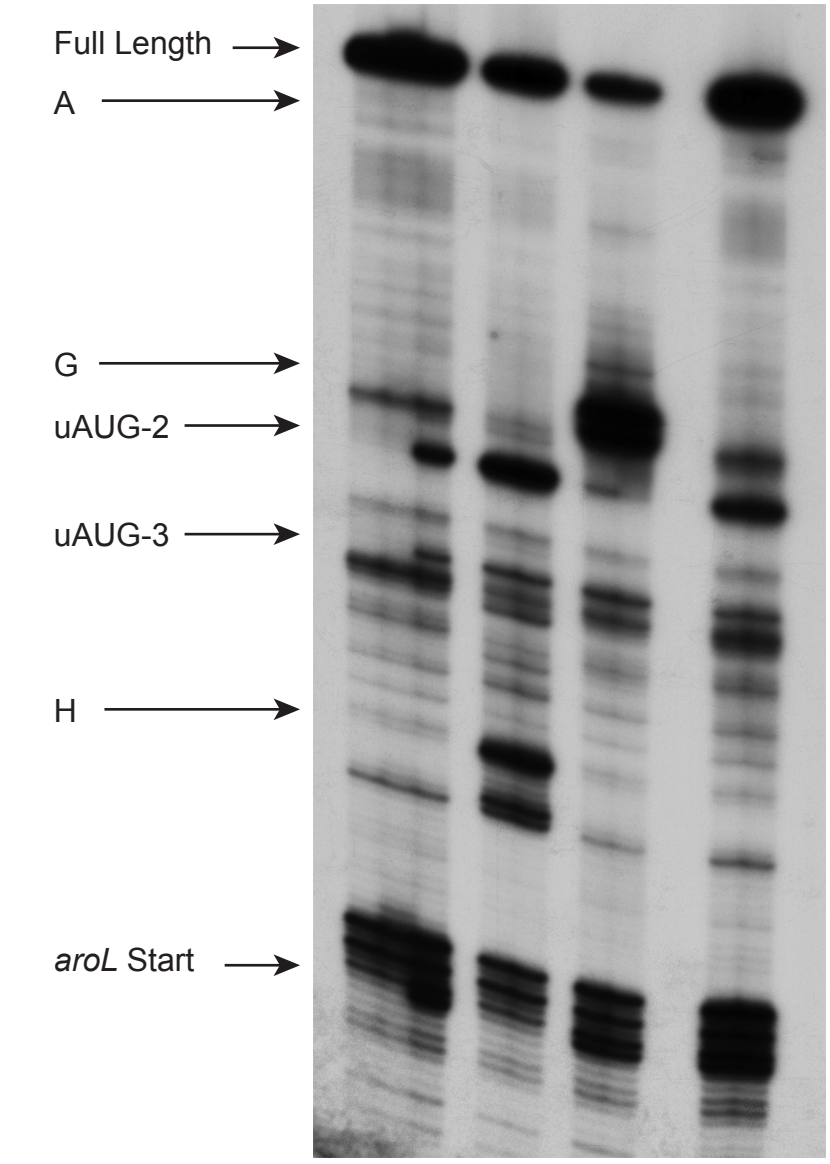
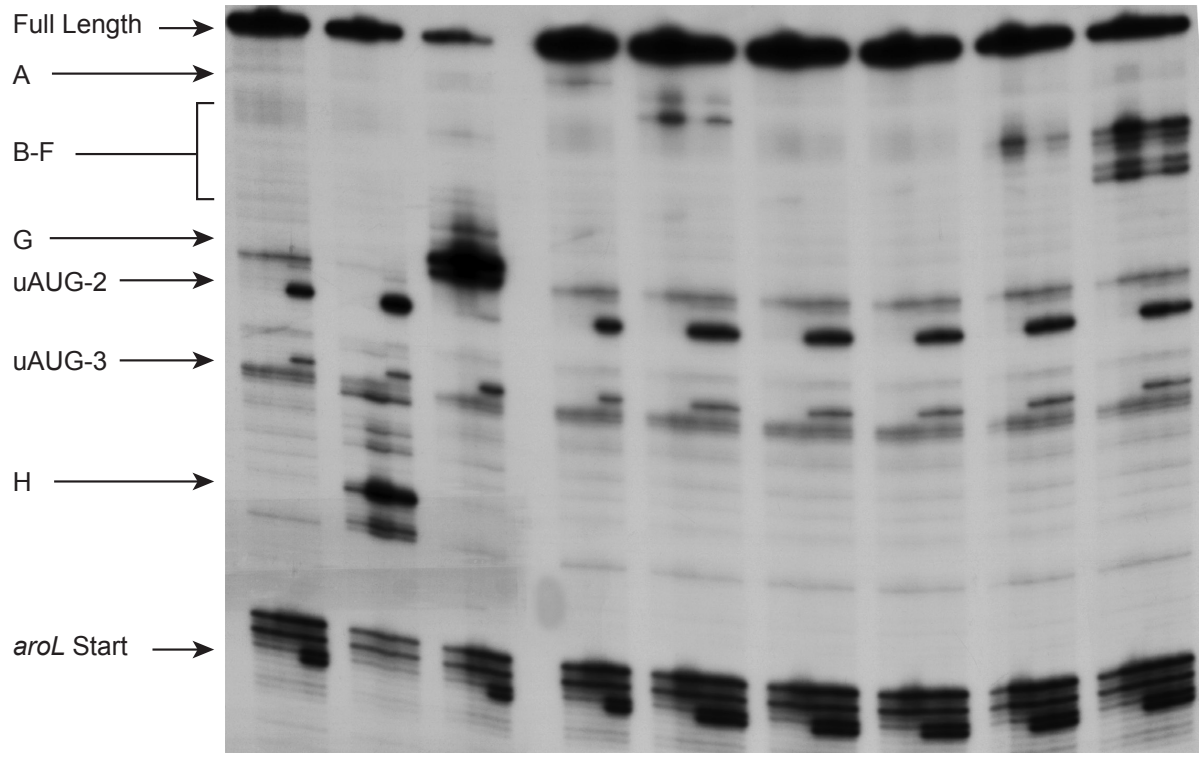
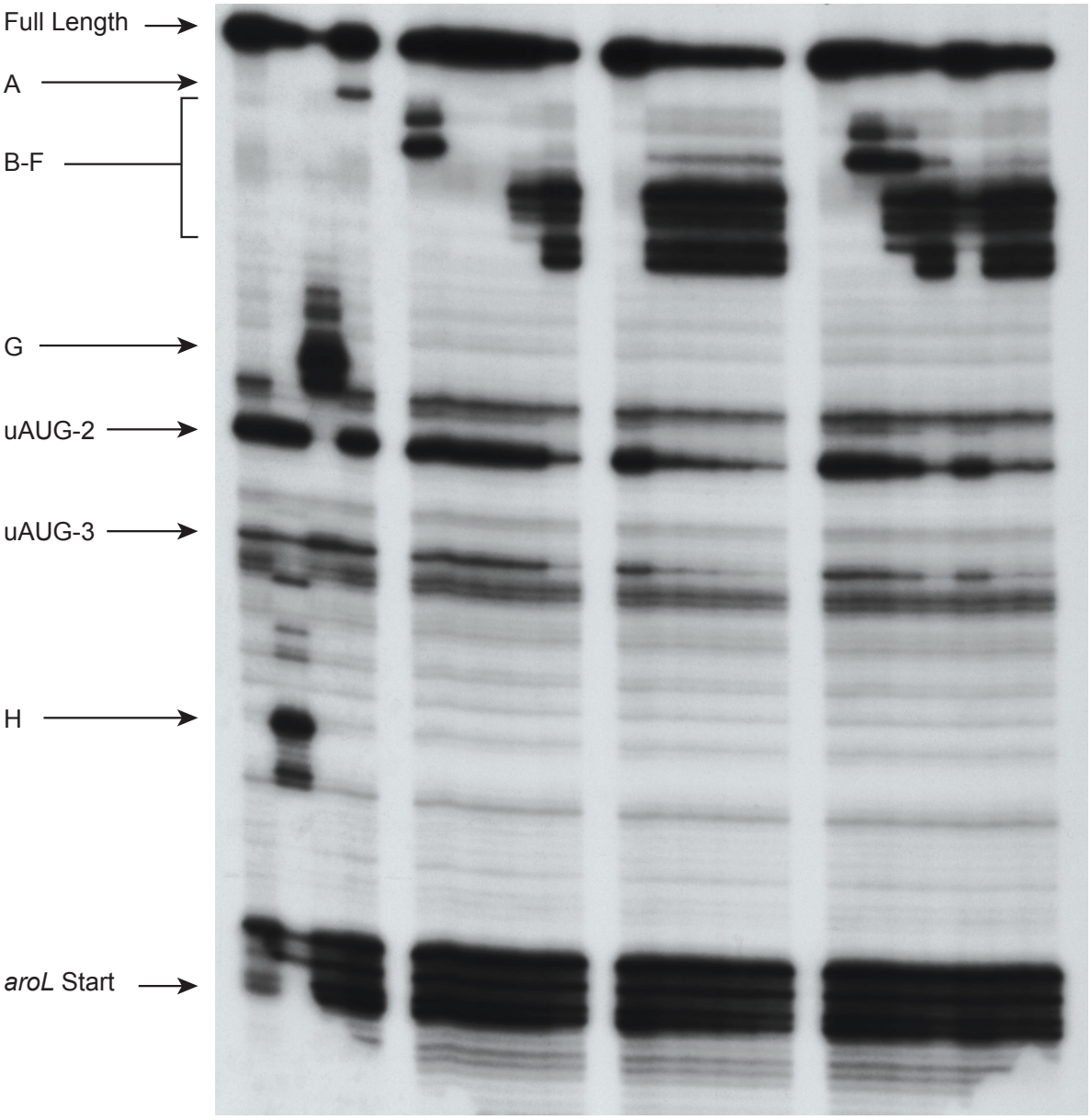

Full Length
A
G
uAUG-2
uAUG-3
H
aroL Start
1 2 3 4 5 6 7 8 9
30S – + + + + + + + +
tRNAfmet + – + + + + + + +
DNA oligonucleotides – – – H H G G A A
20 100 20 100 20 100

Figure A-3. Primer extension inhibition (toeprint) analysis of 30S subunit binding to *aroL* uAUG-2 and the AUG start codon in the presence of oligonucleotides designed to sequester parts of the *aroL* leader upstream of uAUG-2. The mRNA used in these assays included the full *aroL* leader plus 16 codons of the *aroL* coding sequence. 30S subunits were added at a concentration of 10x mRNA. Control oligonucleotides A,G, and H were added at 100X concentration over mRNA. Test oligonucleotides B-F were added at 20X and 100X concentrations over mRNA. The positions of the full-length message, toeprint signals for uAUG-2, uAUG-3 and *aroL* AUG start codon, and predicted position of annealed oligonucleotides are indicated by arrows. DNA oligonucleotide sequences are listed in Table A-1. Sites of complementarity to the *aroL* mRNA are depicted in Figure A-1.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
30S	–	+	+	–	–	+	–	–	+	–	–	+	–	–	+	+	–	–	+	+	–	–	+	+	–	–	+	+	–	–	+	+
$tRNA^{fmet}$	+	–	+	–	–	+	–	–	+	–	–	+	–	–	+	+	–	–	+	+	–	–	+	+	–	–	+	+	–	–	+	+
DNA oligonucleotides	–	–	–	H	H	H	G	G	G	A	A	A	B	B	B	B	C	C	C	C	D	D	D	D	E	E	E	E	F	F	F	F
20x	–	–	–	+	–	–	+	–	–	+	–	–	+	–	+	–	+	–	+	–	+	–	+	–	+	–	+	–	+	–	+	–
100x	–	–	–	–	+	+	–	+	+	–	+	+	–	+	–	+	–	+	–	+	–	+	–	+	–	+	–	+	–	+	–	+

Figure A-4. Primer extension inhibition (toeprint) analysis of 30S subunit binding to *aroL* uAUG-2 and the AUG start codon in the presence of oligonucleotides at 200X or 400X concentrations over the *aroL* mRNA. The mRNA used in these assays included the full *aroL* leader plus 16 codons of the *aroL* coding sequence. 30S subunits were added at a concentration of 10x mRNA. All oligonucleotides were added at a 200X or 400X concentrations over mRNA individually or in combinations. The positions of the full-length message, toeprint signals for uAUG-2, uAUG-3 and *aroL* AUG start codon, and predicted position of annealed oligonucleotides are indicated by arrows. DNA oligonucleotide sequences are listed in Table A-1. Sites of complementarity to the *aroL* mRNA are depicted in Figure A-1.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
30S	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
tRNA^{fmet}	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
DNA oligonucleotides	–	H 400	G 400	A 400	B 400	C 400	D 400	E 400	F 400	–	B-F 200	BDF 200	B-F 400	BDF 400	–	BD 400	BE 400	BF 400	CE 400	CF 400	DF 400

When we compared all of the reactions containing a potentially sequestering oligonucleotide, each reaction contained specific bands corresponding in position to where the RNA:DNA hybrid would cause premature termination of the reverse transcriptase during cDNA synthesis. We took these signals as evidence that the oligonucleotides were bound to the mRNA in the predicted position. Upon examination of the toeprint reactions that contained oligonucleotides C or D (Fig. A-4, lanes 6 and 7), we did not observe any specific banding pattern consistent with cDNA termination that mapped to the position of bound oligonucleotide. We therefore were not confident that these oligonucleotides bound the mRNA, which would have resulted in the unchanged toeprint signal observed for uAUG-2.

We repeated the assays with the modification that the oligonucleotides were added to the reaction during the annealing of radiolabeled primer to the mRNA. During the annealing reaction, the mRNA, radiolabeled primer, and sequestering oligonucleotides were heated to 65^0C for 3 minutes followed by a 20-minute incubation at 50^0C to promote annealing and then cooled to 4^0C. We reasoned that these conditions would be more conducive for the sequestering oligonucleotides to bind the mRNA. The elevated temperature could open up secondary structures that would otherwise prevent the sequestering oligonucleotide from binding, and as the temperature fell the oligonucleotides would be available to access the mRNA as it began to fold. The *aroL* mRNA bound with the radiolabeled primer and test oligonucleotide was then used in subsequent toeprint assays that were carried out as described above.

Adding the oligonucleotides during the annealing reaction did not change the toeprint patterns described above (data not shown). Oligonucleotides C and D still did not appear to bind the mRNA, as suggested by the absence of bands corresponding to the predicted position of the annealed oligonucleotide. Although we are not able to make any conclusions regarding oligonucleotides C and D, the decrease in toeprint signal to uAUG-2 upon addition of oligonucleotides E and F to the reaction supported the deletion analysis (see Chapter 1, Fig. 1-5) and allowed us to narrow down the

region of *aroL* sequence responsible for binding and subsequent expression from uAUG-2.

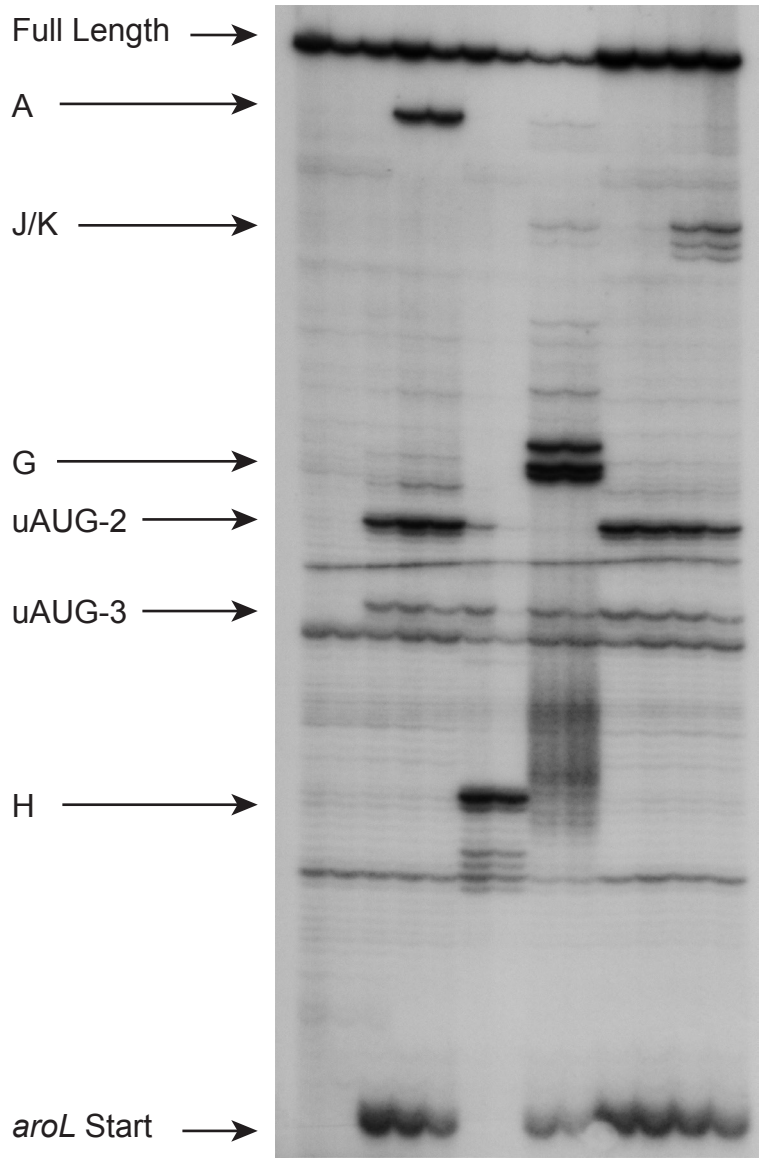
Use of oligonucleotides complementary to the U-rich region of the *aroL* leader in toeprint assays to measure 30S subunit binding to uAUG-2 (U-rich sequestration toeprints)

Once we determined that the U-rich region of the *aroL* leader contributed to ribosome binding and expression from uAUG-2 (data from concurrent substitution and deletion analyses), we repeated the sequestration toeprint assays using DNA oligonucleotides complementary to this region. We intended to use this assay in addition to the deletion and substitution analyses to further support a role for this region in ribosome binding and expression.

In addition to oligonucleotides A, G, and H, which served as controls and are described earlier in this section, we designed test oligonucleotides J and K (Fig. A-1 and Table A-1) to specifically sequester the U-rich region of the *aroL* leader mRNA. Toeprint assays were performed as described (see Chapter 1, Materials and Methods) with the modification that the mRNA was incubated with the sequestering oligonucleotides for 15 minutes at 37^0C after the radiolabeled lac.comp2 primer (Table A-1) was annealed to the *aroL* mRNA. Sequestering oligonucleotides were added at 30X and 600X over the *aroL* mRNA concentration. After this additional incubation, $tRNA^{fMet}$, dNTPs, and 30S subunits were added and the toeprint reactions were carried out as previously described (see Chapter 1, Materials and Methods).

As expected for the negative control, oligonucleotide A had no effect on 30S subunit binding to uAUG-2 (Fig. A-5, lanes 4 and 5), and both positive control oligonucleotides G (Fig. A-5, lanes 8 and 9) and H (Fig. A-5, lanes 6 and 7) reduced binding to uAUG-2 and the *aroL* AUG start codon, respectively. Surprisingly, oligonucleotide J had no effect on ribosome binding to uAUG-2 (Fig. A-5, lanes 10 and 11) at either concentration, and oligonucleotide K decreased binding only slightly at 600X over mRNA concentration (Fig. A-5, lanes 12 and 13).

Figure A-5. Primer extension inhibition (toeprint) analysis of 30S subunit binding to *aroL* uAUG-2 and the AUG start codon in the presence of DNA oligonucleotides designed to sequester the U-rich region of the *aroL* leader mRNA. The mRNA used in these assays included the full *aroL* leader plus 16 codons of the *aroL* coding sequence. 30S subunits were added at a concentration of 10x over mRNA and incubated with oligonucleotides for 15 minutes at 37^0C. The positions of the full-length message, toeprint signals for uAUG-2, uAUG-3 and *aroL* AUG start codon, and predicted position of annealed oligonucleotides are indicated by arrows. DNA oligonucleotide sequences are listed in Table A-1. Sites of complementarity to the *aroL* mRNA are depicted in Figure A-1.


Full Length
A
J/K
G
uAUG-2
uAUG-3
H
aroL Start
1 2 3 4 5 6 7 8 9 10 11 12 13
30S – + + + + + + + + + + + +
tRNAfmet + – + + + + + + + + + + +
DNA oligonucleotides – – – A 60 A 300 H 60 H 300 G 60 G 300 J 60 J 300 K 60 K 300

We suspect that the lack of effect on binding to uAUG-2 was based on an inability of these oligonucleotides to efficiently bind the *aroL* mRNA. Through the course of these sequestration toeprint assays we have been able to visualize banding patterns representing termination of reverse transcription by the bound oligonucleotide. We did not observe any banding pattern in reactions containing oligonucleotide J (Fig. A-5, lanes 10 and 11) and only faint banding in reactions containing K (Fig. A-5, lanes 12 and 13), indicating that these oligonucleotides did not bind the mRNA stably or were displaced by the elongating AMV reverse transcriptase. We had similar difficulties binding oligonucleotides complementary to parts of these regions in previous assays (oligonucleotides C and D used in *aroL* leader sequestration toeprints, Fig. A-4, lanes 6 and 7).

We repeated these assays with a couple of alterations to promote binding of the oligonucleotides to the mRNA. First, we tried adding the test oligonucleotides to the annealing reaction during which the radiolabeled primer is annealed to the mRNA. The mRNA bound with the radiolabeled primer and test-sequestering oligonucleotide was then used in toeprint assays that were carried out as described earlier. Unfortunately, addition of the sequestering oligonucleotide to the annealing reaction did not increase its binding to the mRNA, as evidenced by no band corresponding to an annealed oligonucleotide; also the toeprint signals to uAUG-2 did not change (data not shown). Reactions containing the control oligonucleotides A, G, or H showed the expected results.

We tried another reaction modification in an effort to increase the efficiency of oligonucleotides J and K binding to the *aroL* mRNA. This time, we performed the toeprint assays at 28^0C. We reasoned that 37^0C, the temperature at which we normally perform the assay, might be too high to allow stable binding of the small oligonucleotides. Therefore, the sequestering test oligonucleotides were incubated with the mRNA for 15 minutes at 28^0C after the radiolabeled lac.comp2 primer (Table A-1) was annealed to the mRNA. fMet-tRNA, dNTPs, and 30S subunits were then added and the toeprint reactions were carried out as described earlier except all incubations for 30S subunit binding and cDNA synthesis by reverse transcriptase were done at 28^0C.

The results from these toeprint reactions were inconclusive, as we did not get any toeprint signal or significant full-length cDNA formed (data not shown). We concluded that reverse transcription activity was likely to be hindered at this low temperature.

Our substitution and deletion mutation strategies described extensively in Chapter 1 both strongly support that the U-rich region upstream of uAUG-2 is necessary for 30S binding to uAUG-2. We therefore decided not to further pursue the use of these complementary oligonucleotides to assess 30S binding patterns.

Use of RNA oligonucleotides representing the U-rich region of the *aroL* leader in toeprint assays to measure 30S binding to uAUG-2

Because we were able to use DNA oligonucleotides to sequester regions of the *aroL* mRNA contributing to 30S subunit binding to uAUG-2 and the *aroL* AUG (Figs. A-3-A-5), we reasoned that we could use RNA oligonucleotides to sequester or saturate binding sites on the 30S subunit involved in recognition of the U-rich region of the *aroL* mRNA that contributes to uAUG-2 binding. In an effort to provide support for an interaction between the 30S subunit and the U-rich region of the *aroL* leader that contributes to binding and expression from uAUG-2, we used RNA oligonucleotides representing the U-rich sequence in toeprint assays with 30S subunits. Although the U-rich sequence is not complementary to the 3' end of the 16S rRNA, we believe it nonetheless makes contact with some region or component of the ribosome. By performing toeprints with 30S subunits prebound with excess RNA oligonucleotide, we predicted that the ribosome contact site would be saturated with the U-rich oligonucleotide and not available to interact with the *aroL* U-rich region on the mRNA, thereby reducing a toeprint signal to uAUG-2.

Before conducting this assay with our test RNA oligonucleotides representing the *aroL* U-rich sequence, we first performed the assays with DNA oligonucleotides that were complementary to the 3' end of the *E. coli* 16S rRNA. Sequestering the 16S rRNA 3' end, containing the anti-SD sequence (ASD), in an RNA:DNA hybrid would render the ASD sequence unavailable for pairing with an mRNA's SD sequence and reduce the intensity of a toeprint signal that reflects mRNA-ribosome binding. These served as our

controls and included: 16S.9mer ASD, 16S.11mer, and 16S.13mer (Table A-1 and Fig. A-6). These oligonucleotides were incubated with 30S subunits at 20X and 100X the 30S subunit concentration for 15 minutes at 37^0C. The 30S subunits were then used in toeprint assays as described earlier. Incubation with the 16S.9mer and 16S.13mer reduced the toeprint signal to the *aroL* AUG start codon at both 20X and 100X concentrations (Fig. A-7, lanes 4-5 and 8-9).

The 16S.11mer did not have an effect on 30S subunit binding to the *aroL* start codon at either 20X or 100X concentration (Fig. A-7, lanes 6 and 7). Interestingly, addition of the 16S.9mer and 16S.13mer oligonucleotides also reduced the toeprint signals to uAUG-2 (Fig. A-7, lanes 4-5 and 8-9), implying that 30S subunit recognition to uAUG-2 also involves an interaction with the ASD sequence at the 3' end of the 16S rRNA. The *aroL* leader upstream to uAUG-2 does not contain a strong SD sequence, and the region of the leader important for binding to uAUG-2 (implicated by the mutation analyses presented in Chapter 1) does not contain obvious complementarity to the 3' end of the 16S rRNA. Therefore, we interpret this loss of toeprint signal to uAUG-2 as a result of the oligonucleotides binding to the 16S rRNA, thus making the 30S subunit mRNA tunnel inaccessible for proper placement of uAUG-2 in the P-site.

These experiments supported our ability to use oligonucleotides to sequester the ASD sequence on the 30S subunit and reduce mRNA binding. Next, we used a similar approach in using RNA oligonucleotides in an effort to sequester the 30S subunit site responsible for contacting the U-rich region of the *aroL* mRNA. We predicted that if this contact site was unavailable to contact the a*roL* mRNA, the toeprint signal to uAUG-2 would decrease. The assays were repeated using the negative control RNA oligonucleotide RNANeg (Table A-1), the positive control RNA oligonucleotide 16S.13mer, and our test RNA oligonucleotide UpRBS2 (Table A-1), which represents the U-rich region of the *aroL* leader. RNA oligonucleotides were added at 30X and 600X concentrations over 30S subunits. 30S subunits were pre-incubated with the RNA oligonucleotides for 15 minutes at 37^0C prior to use in toeprint assays.

Figure A-6. Binding of DNA oligonucleotides to the 3' end of the 16S rRNA. The secondary structure of the 16S rRNA is shown on the left. The 3' domain has been enlarged on the right and shows the regions of complementarity of the DNA oligonucleotides (colored lines). DNA oligonucleotide sequences are listed in Table A-1. Image is adapted from XRNA GALLERY (rna.ucsc.edu/rnacenter/xrna/xrna_gallery.html).

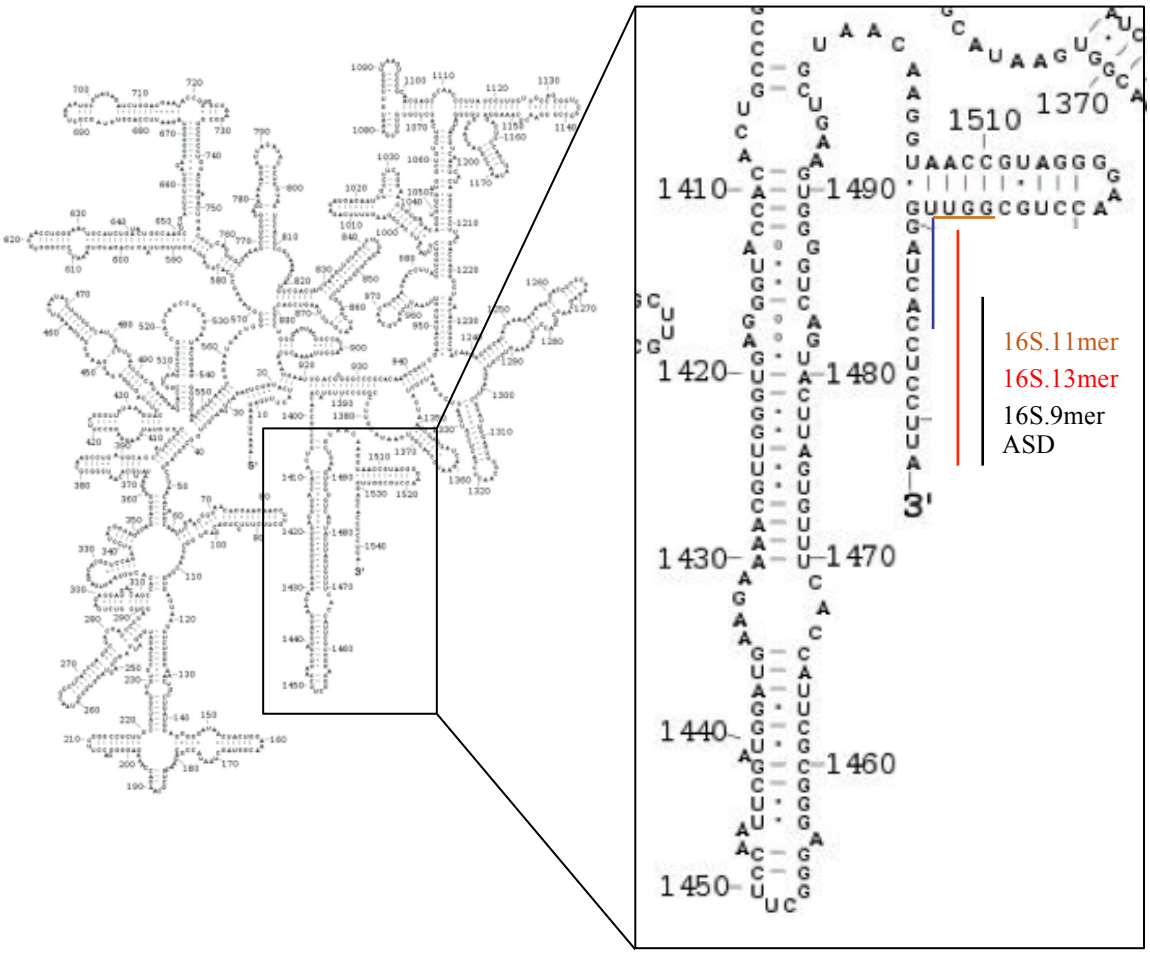
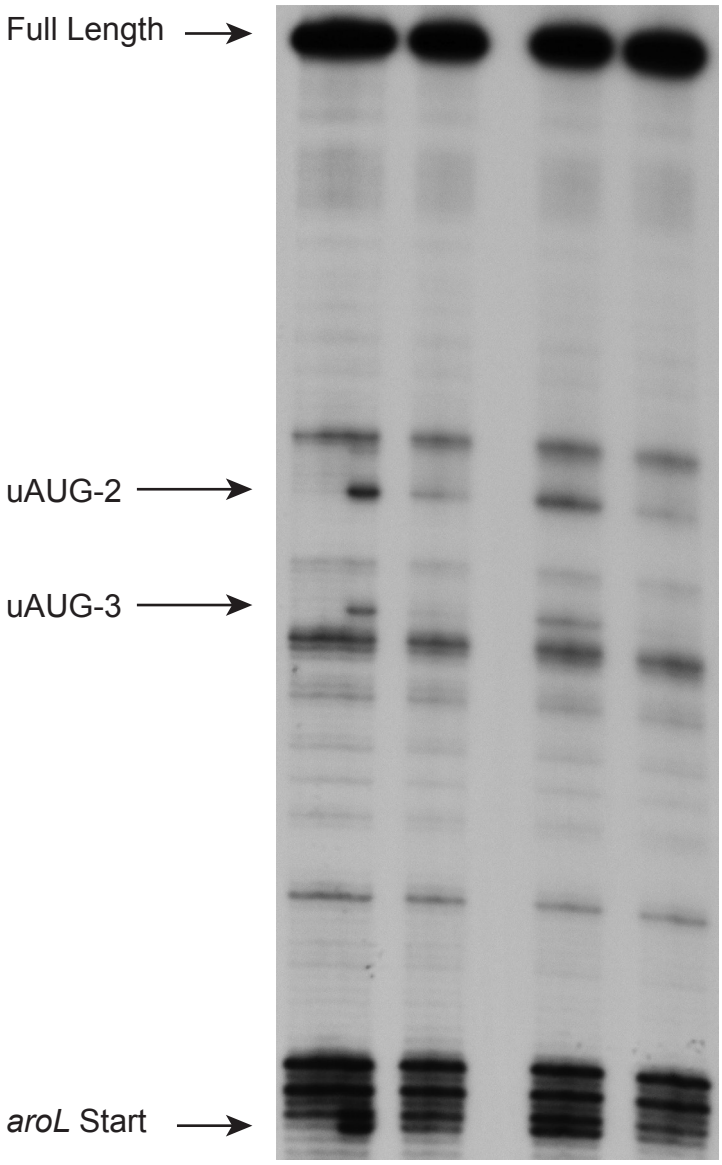

1410
1490
1420
1480
1430
1470
1440
1460
1450
1510
1370
3'
16S.11mer
16S.13mer
16S.9mer
ASD

Figure A-7. Primer extension inhibition (toeprint) analysis of 30S subunit binding to *aroL* uAUG-2 and the AUG start codon in the presence of DNA oligonucleotides designed to sequester the anti-Shine-Dalgarno sequence at the 3'end of the 16S rRNA of the 30S subunit. The mRNA used in these assays included the full *aroL* leader plus 16 codons of the *aroL* coding sequence. 30S subunits were added at a concentration of 10x over mRNA and were incubated with oligonucleotides (20X and 100X concentrations over 30S subunits) for 15 minutes at 37^0C before being used in toeprint assays. The positions of the full-length message and toeprint signals for uAUG-2, uAUG-3 and *aroL* AUG start codon are indicated by arrows. DNA oligonucleotide sequences are listed in Table A1. Complementarity to the 16S rRNA is shown in Figure A-6.



	1	2	3	4	5	6	7	8	9
30S	–	+	+	+	+	+	+	+	+
tRNA^{fmet}	+	–	+	+	+	+	+	+	+
16S.9mer	–	–	–	20	100	–	–	–	–
16S.11mer	–	–	–	–	–	20	100	–	–
16S.13mer	–	–	–	–	–	–	–	20	100

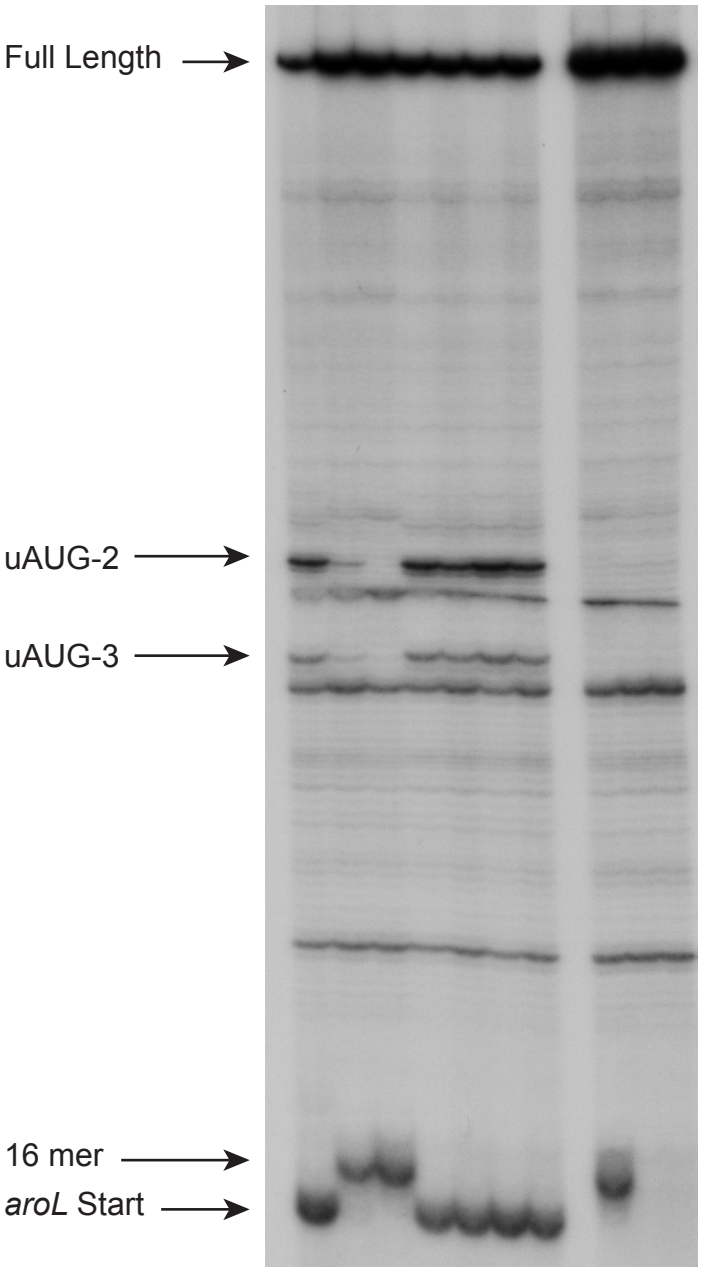
Addition of the control RNA oligonucleotides gave the predicted results: the positive control, 16S.13mer, decreased 30S subunit binding to the *aroL* AUG start (Fig. A-8, lanes 2 and 3) and the negative control, RNANeg, did not alter 30S subunit binding to uAUG-2 (Fig. A-8, lanes 4 and 5). Reactions that contained the positive control (Fig. A-8, lanes 2 and 3) had a strong extraneous band slightly above the position of the *aroL* AUG toeprint signal. This same signal was apparent in reactions that lacked 30S subunits (Fig. A-8, lane 8) and was most likely due to the oligonucleotide interacting with the mRNA in a way that resulted in premature termination of cDNA synthesis by reverse transcriptase. It was difficult to tease apart whether the lack of toeprint signal (Fig. A-8, lanes 2 and 3) resulted from the 16S.13mer oligonucleotide binding the mRNA, thereby preventing an interaction with the 30S subunit, or the oligonucleotide binding the 16S rRNA, preventing an interaction with the mRNA. We have used this oligonucleotide in previous reactions (data not shown) which gave reproducible results (showing that addition of the oligonucleotide reduced toeprint signal to the *aroL* AUG, implying an interaction between the oligonucleotide and the 16SrRNA), and therefore, we think that reduced toeprint signal is a result of both the oligonucleotide binding the mRNA and the oligonucleotide interacting with the 3' end of the 16S rRNA. However, addition of the test oligonucleotide, upRBS2, did not change the toeprint pattern to any AUG (Fig. A-8, lanes 6 and 7); therefore, we were unable to use this assay to support or confirm an interaction between the U-rich region of the *aroL* leader and the 30S subunit. Although it is possible that the concentrations used were not high enough to saturate the contact site on the 30S subunit, it is also possible that the assay conditions were not optimized to allow for the interaction between the oligonucleotide and 30S subunit.

Conclusion

Using DNA oligonucleotides complementary to the *aroL* mRNA we were able to sequester part of the region in the mRNA which contributes to 30S subunit binding to uAUG-2. Addition of oligonucleotide F (Fig. A-4, lanes 9, 11-14, 18, 20-21) resulted in the greatest reduction in uAUG-2 toeprint signal and addition of oligonucleotide E reduced the signal slightly (Fig. A-4, lanes 17 and 19). These oligonucleotides are each complementary to part of the U-rich region of the *aroL* mRNA identified by nucleotide

substitution and deletion analyses as being responsible for expression from uAUG-2 in Chapter 1. Oligonucleotides D (Fig. A-4, lane 7) and J, K (Fig. A-5, lanes 10-13) were complementary to the entire U-rich region yet did not significantly decrease the toeprint signal to uAUG-2, presumably due to an inability to efficiently bind the mRNA. Although it is possible that the interactions of these oligonucleotides with the mRNA were hindered by the conditions of the toeprint assays, it is also possible that this region of the mRNA is involved in secondary structure that prevented oligonucleotide binding. Such secondary structure might be part of the mechanism by which the 30S subunit recognizes and binds the mRNA. The toeprint signal of uAUG-3 was affected by oligonucleotides E,F, (Fig. A-4) and K (Fig. A-5), but unpublished data suggest that sequences further downstream within the *aroL* coding region also contribute to 30S subunit binding to uAUG-3.

Figure A-8. Primer extension inhibition (toeprint) analysis of 30S subunit binding to *aroL* uAUG-2 and the AUG start codon in the presence of RNA oligonucleotides designed to prevent interactions between the *aroL* mRNA and the 3'end of the 16S rRNA of the 30S subunit. The mRNA used in these assays included the full *aroL* leader plus 16 codons of the *aroL* coding sequence. 30S subunits were added at a concentration of 10x mRNA and were incubated with oligonucleotides (at 60X or 300X concentration over 30S subunits) for 15 minutes at 37^0C before being used in toeprint assays. The positions of the full-length message and toeprint signals for uAUG-2, uAUG-3 and *aroL* AUG start codon are indicated by arrows. RNA oligonucleotide sequences are listed in Table A-1.



	1	2	3	4	5	6	7	8	9	10
30S	–	+	+	+	+	+	+	–	–	–
tRNA^{fmet}	+	+	+	+	+	+	+	–	–	–
16S mer	–	60	300	–	–	–	–	300	–	–
RNA neg	–	–	–	60	300	–	–	–	300	–
upRBS2	–	–	–	–	–	60	300	–	–	300

Appendix B

Investigation of physiological roles for the *aroL* upstream ORF-encoded peptide during aromatic amino acid biosynthesis in *Escherichia coli* K12

Introduction

Despite their small size, researchers are beginning to assign functions to small proteins in both prokaryotes and eukaryotes (Alix and Blanc-Potard, 2009, Rosenberg and Desplan, 2010). The short open reading frames (ORFs) that encode small proteins are sometimes expressed as part of an operon (reviewed in Hobbs et al., 2011); the genomic context within which short ORFs are found may provide clues regarding their function. Proteins produced from short coding regions, identified in both Gram-positive and Gram-negative bacteria, are typically between 25-50 amino acids in length and function in a variety of physiological processes. These small proteins have been assigned roles as regulators (Alix and Blanc-Potard, 2008, Handler et al., 2008, Wadler and Vanderpool, 2007, Cunningham and Burkholder, 2009), stabilizing factors (Schneider et al., 2007, Gassel et al., 1999), membrane components (Ramamurthi et al., 2009, Gaballa et al., 2008), signal peptides (Lopez et al., 2009; Bauer and Dicks, 2005, Cogen et al., 2010), metal chaperones (Gaballa et al., 2008), and toxins (Fozo et al., 2008, Unoson and Wagner, 2008, Fozo et al., 2010). The occurrence and significance of small ORFs has been underestimated and genome-wide analyses will most likely identify more of them.

Because *aroL* expression is coupled to expression of the upstream ORF (uORF), (see Chapter 1, Fig. 1-3), it is likely that if the encoded peptide was produced and stable, it would have an *aroL*-related function. Shikimate kinase II, encoded by *aroL*, catalyzes the fifth step in the shikimate pathway to convert erythrose-4-phosphate and phosphoenolpyruvate to chorismate, the common precursor for synthesis of aromatic amino acids (Berlyn and Giles, 1969). The putative peptide might be involved in some aspect of chorismate or aromatic amino acid biosynthesis.

To address a possible role for the *aroL* putative peptide in aromatic amino acid synthesis, we assessed the growth phenotype of cells when the uORF was highly expressed from an expression vector when shifted from medium containing aromatic amino acids (i.e., tryptophan, tyrosine and phenylalanine) to medium lacking aromatic amino acids. We reasoned that if the peptide was produced and stable it would aid the cells in this physiological transition. We expressed the *aroL* uORF (and putative peptide product) from the arabinose inducible promoter, P_{BAD}, to increase expression in a controlled manner.

The results presented here show no difference in growth phenotype between cells expressing the *aroL* uORF from the P_{BAD} promoter and those not expressing the uORF from P_{BAD}. Although we did not observe an impact of expression of the *aroL* uORF during transition to growth in the absence of provided aromatic amino acids, it is possible that expression (and potential production of the peptide) influences some other aspect of aromatic amino acid synthesis. To identify specific cellular components that interact with the putative peptide product from the uORF, His-tagged vectors were constructed. Our intentions were to identify proteins that interacted with the peptide and then address how those interactions impact cellular physiology.

The assays described in this appendix were designed with the assumption that the putative peptide encoded within the *aroL* uORF is produced. We have yet to show that the putative peptide is present and stable in the cell once expressed from the vectors. Further, we have yet to show that the putative peptide is produced from the chromosome.

Materials and Methods

Genomic DNA from *E. coli* K12 was used for the isolation of gene fragments containing the *aroL* untranslated leader. *E. coli* DH5α [New England Biolabs (NEB)] was used as the host for all plasmid DNA manipulations. *E. coli* RFS859 (F^-, *thr*-1, *ara*C859, *leu*B6, Δ*lac*74, *tsx*-274, λ^-, *gyr*A111, *rec*A11, *rel*A1, *thi*-1) (Schleif, 1972) was used as the host strain for expression of *lacZ* fusion constructs. Cells were grown in M9 minimal medium (Gerhardt, et al., 1994) with glycerol 0.2% (vol/vol), threonine (80 µg/mL),

leucine (20 µg/mL), thiamine (40 µg/mL) and the addition or absence of aromatic amino acids tryptophan (20 µg/mL), tyrosine (20 µg/mL), and phenylalanine (20 µg/mL). Antibiotics were from Sigma-Aldrich and Fisher Scientific. Amino acids and L-arabinose were from Sigma-Aldrich.

The general cloning protocols were performed in a standard manner (Sambrook et al., 1989). Restriction endonucleases and DNA modifying enzymes were purchased from NEB. Chloramphenicol (Cm; 30 µg/mL) and ampicillin (Amp; 200 µg/mL) were used to select for plasmid maintenance. In order to achieve tight regulation, fast induction, and high expression of the uORF, we placed the *aroL* leader sequence under control of the arabinose-inducible P_{BAD} promoter. L-arabinose was used at the indicated concentrations to induce the promoter activity of P_{BAD}. The name and description of the expression vectors and plasmids are listed in Table B-1.

Fusions of the *aroL* uORF were first made to a *lacZ* reporter gene in order to verify expression and determine the induction kinetics of the P_{BAD} promoter. As described in Chapter 1 (Materials and Methods), the *aroL* leader and gene fragments were first subcloned into the pBR322-derived plasmid pA904 (Amp^R; transcription provided by the *E. coli lac* promoter;) and fused to the fifth codon of *lacZ*. Fragments of *aroL-lacZ* from pA904 were then amplified by PCR and cloned into pBAD33 (Cm^R) (Guzman, et al., 1995). Fusions were constructed with the *aroL* uORF sequence beginning with uATG-2 plus 11 codons or uATG-3 plus 6 codons of the open reading frame fused to *lacZ* (named pBAD-uATG2-*lacZ* and pBAD-uATG3-*lacZ*, respectively). Cultures were grown to an OD_{600} of ~0.4 and split into individual flasks with various amounts of L-arabinose. Arabinose induction of P_{BAD} expression from these fusions was monitored over time by β-galactosidase assays (Miller, 1992). Assays were performed in triplicate from three independent cultures grown at 37^0C.

After confirmation of expression from the *aroL-lacZ* fusions (above), the *aroL* uORF sequence (beginning with either uATG-2 or uATG-3 (see Chapter 1, Fig. 1-1) and extending through the stop codon) was subcloned into the multiple cloning site of the pBAD expression vector, pBAD33 (Cm^R) (Guzman, et al., 1995), and named pBAD-uATG2 and pBAD-uATG3, respectively. Another plasmid, pBAD-uATG2-*aroL*, was

constructed to place the *aroL* uORF sequence starting at uATG-2 and the entire *aroL* coding sequence under the control of P_{BAD}. Translation in these expression vectors was under control of the pBAD33 Shine-Dalgarno (SD) sequence (5'-…AGGAGGA…-3'); this ribosome binding site (RBS) along with the P_{BAD} promoter has been shown to provide efficient regulation and high levels of expression (Guzman, et al., 1995). Use of the pBAD33 RBS also eliminated regulation of uORF expression by the unusual ribosome binding signals described for the *aroL* 5'UTR (see Chapter 1, Results). We used these constructs to monitor the growth phenotype of *E. coli* in minimal medium upon expression of the uORF (and potential production of the encoded peptide).

To detect an effect of the expression of the *aroL* uORF on the growth phenotype as cells transition from growth with aromatic amino acids to growth without, the following procedure was used. Cells were grown in M9 minimal medium with the aromatic amino acid supplement (see above) until early log phase (OD_{600} of ~0.2; Fig. B-3A). The cultures were then split into two. Cells from the first half were collected by centrifugation (8000 x *g*, 5 min), washed and resuspended in fresh medium without the aromatic amino acid supplement and expression of the *aroL* uORF from the P_{BAD} promoter was induced with 0.05% L-arabinose. Cells from the second half of the culture were continued in M9 plus supplements and induced (0.05% L-arabinose) for 50 minutes, followed by centrifugation (8000 x *g*, 5 min), washing, and resuspension in fresh medium without the aromatic amino acid supplement, as described above. In this manner, an abundance of peptide is expected to be present in the second half culture prior to transition to growth medium lacking aromatic amino acids.

Derivatives of pBAD-uATG2 and pBAD-uATG2-*aroL* were constructed in which a 6XHis-tag (5'-CATCACCATCACCATCAC-3') was placed at the 5'-terminus of the uORF sequence (pBAD-His-uATG2 and pBAD-His-uATG2-*aroL*). These constructs were expected to produce high levels of the putative peptide with a His-tag at the N-terminus and were intended to be used as the bait in pulldown assays with an anti-6XHis antibody (Abcam) to identify putative binding partners (Arifuzzaman, et al., 2006).

For future work expressing sRNAs (e.g., the *ryhB* sRNA described in Chapter 2) or leaderless mRNAs from the P_{BAD} promoter, we constructed pBAD33 derivatives

containing fragments of the leaderless *c*I gene from bacteriophage lambda or a *lac*-leadered version of the *c*I gene. After induction with L-arabinose, expression from P_{BAD} was monitored by β-galactosidase assays, as described above. Plasmids LL*c*I-*lacZ* (leaderless *c*I; Amp^R) and pSDL*c*I-*lacZ* (*lac*-leadered *c*I; Amp^R), encoding *c*I codons 1-16 fused to the fifth codon of *lacZ* (O'Donnell and Janssen, 2001), were used as templates to PCR amplify and clone the leaderless and *lac*-leadered *c*I-*lacZ* fragments into pBAD33 (Cm^R) (Guzman et al., 1995), resulting in pBAD-LL*c*I-*lacZ* and pBAD-SDL*c*I-*lacZ*, respectively. For pBAD-LL*c*I-*lacZ*, the A of the ATG start codon was cloned directly into the P_{BAD} transcriptional start site (i.e., +1 position) such that induction of P_{BAD} would be expected to produce a leaderless *c*I message with its AUG start codon at the mRNA's 5'-terminus; expression of LacZ activity from this construct indicates that the *c*I-*lacZ* mRNA transcribed from P_{BAD} is translated as a leaderless mRNA. For pBAD-SDL*c*I-*lacZ*, the P_{BAD} promoter initiates transcription at the 5'-terminus of the *lac* untranslated leader and *c*I translation is under control of the *lac* RBS (5'-…AGGA…-3').

Vector maps are depicted in Figures B-5-B-9 in the Results section of this appendix.

Results

The function of the putative peptide encoded in the *aroL* uORF, if any, is unknown. Because the uORF is contained within the same transcription unit as the *aroL* CDS and our observation that *aroL* is translationally coupled to expression of the upstream *aroL* ORF (see Chapter 1, Fig. 1-3), we considered it possible that expression of the *aroL* uORF (and production of the putative peptide) is somehow linked to induction of shikimate kinase expression and biosynthesis of aromatic amino acids. In order to elucidate a possible function for the putative peptide encoded within the *aroL* leader, we overexpressed the uORF and assessed the ability of cells to transition to conditions that required aromatic amino acid synthesis. Cells were initially grown in the presence of aromatic amino acids, resulting in repression of *aroL* transcription (Lawley and Pittard, 1994), and therefore, also the uORF (and potential peptide product), and then transitioned to medium lacking aromatic amino acids, thereby inducing *aroL* transcription. We monitored this transition in cells containing, or lacking, the P_{BAD} – expressed *aroL* uORF by observing growth phenotypes, with particular attention to the

time needed for cells to adjust to the new medium and resume growing at an exponential rate.

To optimize conditions for expression of the uORF (and putative peptide production), we placed the *aroL* uATG-2 (pBAD-uATG2-*lacZ)* or uATG-3 (pBAD-uATG3-*lacZ)* open reading frames (see Fig. 1-1, 1-S1) under control of the arabinose-inducible P_{BAD} promoter and fused the *aroL* uORF to a *lacZ* reporter gene. Promoter activity and uORF expression were assessed by performing β-galactosidase assays on cells containing the fusion constructs in M9 minimal medium in the presence or absence of aromatic amino acids. Expression was induced by adding various amounts of L-arabinose for a variable amount of time. 0.05% L-arabinose was sufficient for supporting maximal expression from both uAUG-2 and uAUG-3; increasing the inducer concentration to 1% did not result in further increases in expression (Fig. B-1). Overall, expression appeared to increase as the length of induction time increased, peaking at 50 min after addition of inducer. These trends were similar regardless of which uAUG start codon was used (pBAD-ATG3-*lacZ* data not shown). Translation from uAUG-3 represents a shortened version of the putative peptide produced from uAUG-2 (both uAUGs are in the same reading frame; see Fig. 1-1). Comparison of results obtained with both uORFs (pBAD-ATG2*,* pBAD-ATG3) could help identify the functional regions of the peptide. Any differences observed in the growth phenotype, would suggest that the amino acids omitted from the shortened version (pBAD-ATG3-*lacZ*) are necessary.

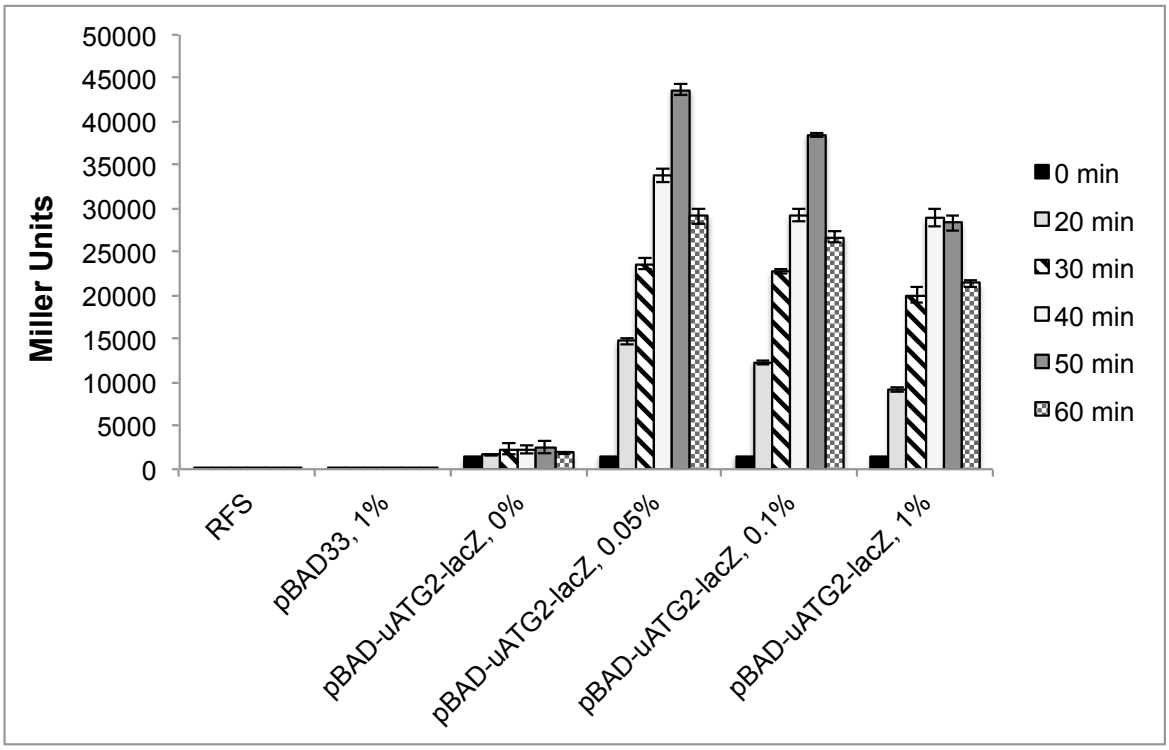
As a comparison, and to quantitate expression of leaderless messages from the inducible P_{BAD} promoter, the arabinose induction experiments were repeated with the pBAD-LL*c*I-*lacZ* (Fig. B-2A) and pBAD-SDL*c*I-*lacZ* fusions (Fig. B-2B). Expression was measured and continued to increase 20, 40, and 60 minutes after addition of 0.05% L-arabinose. These results are consistent with the ability of leaderless messages to be efficiently expressed from the P_{BAD} promoter upon induction, although expression was ~6-8x lower than the *lac*-leadered equivalent (Fig. B-2B). Further, the results suggest that sRNAs can be expressed efficiently from the promoter when the first nucleotide of the sRNA is cloned into the +1 position of the pBAD expression vector.

Table B-1. Description of pBAD-*aroL* (leader) expression vectors.

Plasmid Construct	Description
pBAD33	Expression vector; araBAD promoter; CmR
pBAD-uATG2-*lacZ*	*aroL* upstream ORF sequence beginning with uATG-2 plus 11 codons of the coding sequence fused to *lacZ*
pBAD-uATG3-*lacZ*	*aroL* upstream ORF sequence beginning with uATG-3 plus 6 codons of the coding sequence fused to *lacZ*
pBAD-uATG2	*aroL* upstream ORF sequence beginning with uATG-2 through stop codon
pBAD-uATG3	*aroL* upstream ORF sequence beginning with uATG-3 through stop codon
pBAD-uATG2-*aroL*	*aroL* upstream ORF sequence beginning with uATG-2 extending through full *aroL* gene through *aroL* stop codon
pBAD-His-uATG2	Derivative of pBAD-uATG2 with 'CATCACCATCACCATCAC' His-tag at the 5'-terminus of the upstream ORF
pBAD-His-uATG2-*aroL*	Derivative of pBAD-uATG2-*aroL* with 'CATCACCATCACCATCAC' His-tag at the 5'-terminus of the upstream ORF
pBAD-SDL*c*I-*lacZ*	*lac*-leadered *c*I (16 codons) fused to *lacZ*
pBAD-LL*c*I-*lacZ*	Leaderless *c*I (16 codons) fused to *lacZ*

Figure B-1. Induction of the pBAD-uATG2-*lacZ* expression vector with various amounts of L-arabinose. Cultures with pBAD-uATG2-*lacZ* were grown in M9 minimal medium with (A) or without (B) the addition of aromatic amino acids (tyrosine 20 µg/mL, tryptophan 20 µg/mL, phenylalanine 20 µg/mL) to an OD_{600} of ~0.4. Cultures were then split equally into separate flasks and L-arabinose was added to final concentrations of 0%, 0.05%, 0.1%, and 1%. Control cultures (RFS and pBAD33) were given 1% of the inducer. Samples were pulled at 0 min, 20 min, 30 min, 40 min, 50 min, and 60 min after induction and measured for β-galactosidase activity. *E. coli* RFS859 is the host strain.

A.


Miller Units
50000
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
0 min
20 min
30 min
40 min
50 min
60 min
RFS
pBAD33, 1%
pBAD-uATG2-lacZ, 0%
pBAD-uATG2-lacZ, 0.05%
pBAD-uATG2-lacZ, 0.1%
pBAD-uATG2-lacZ, 1%

B.

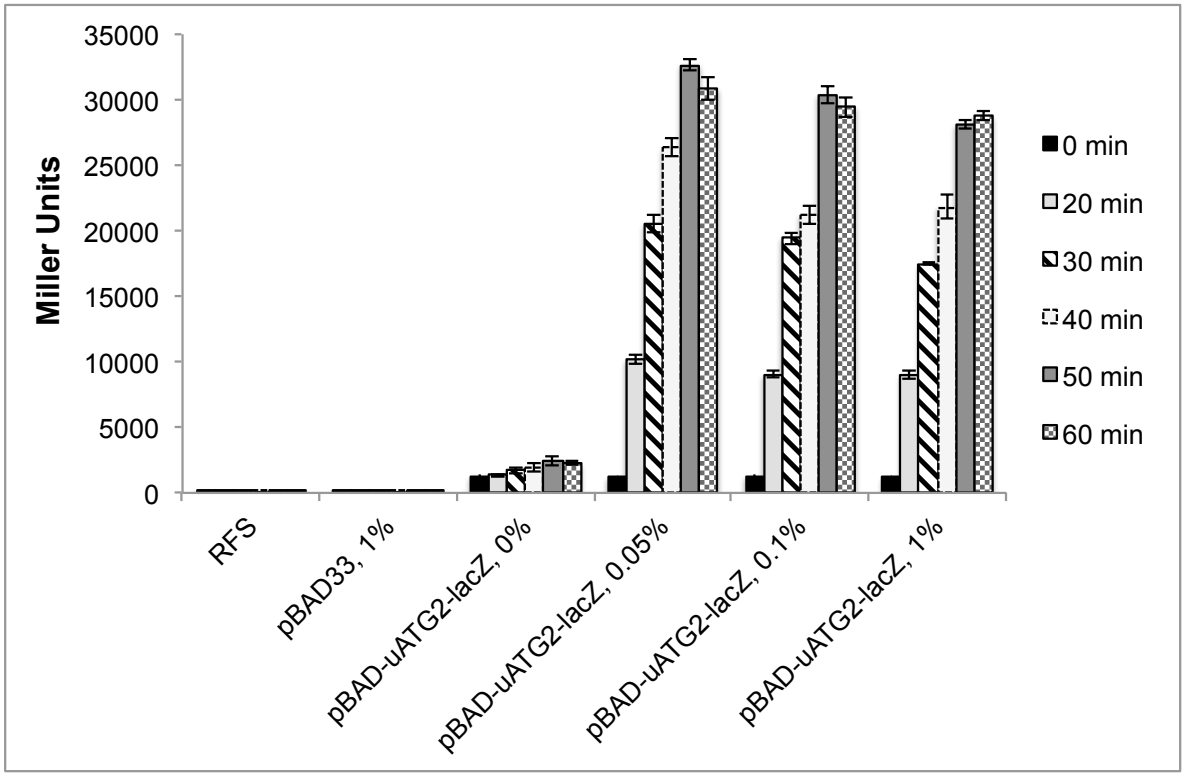
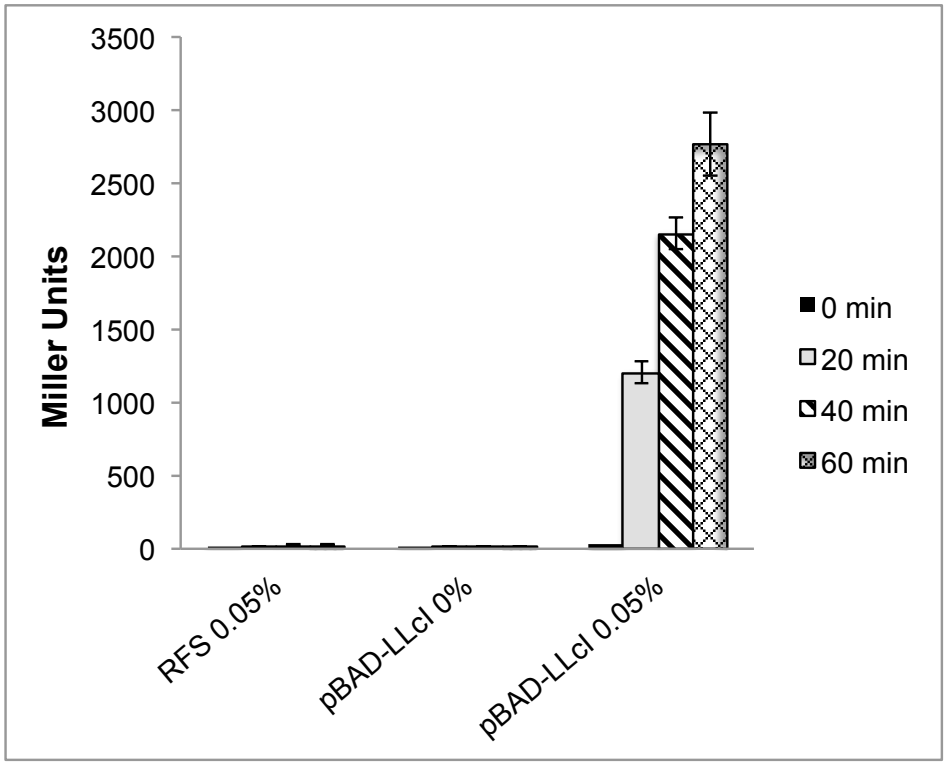
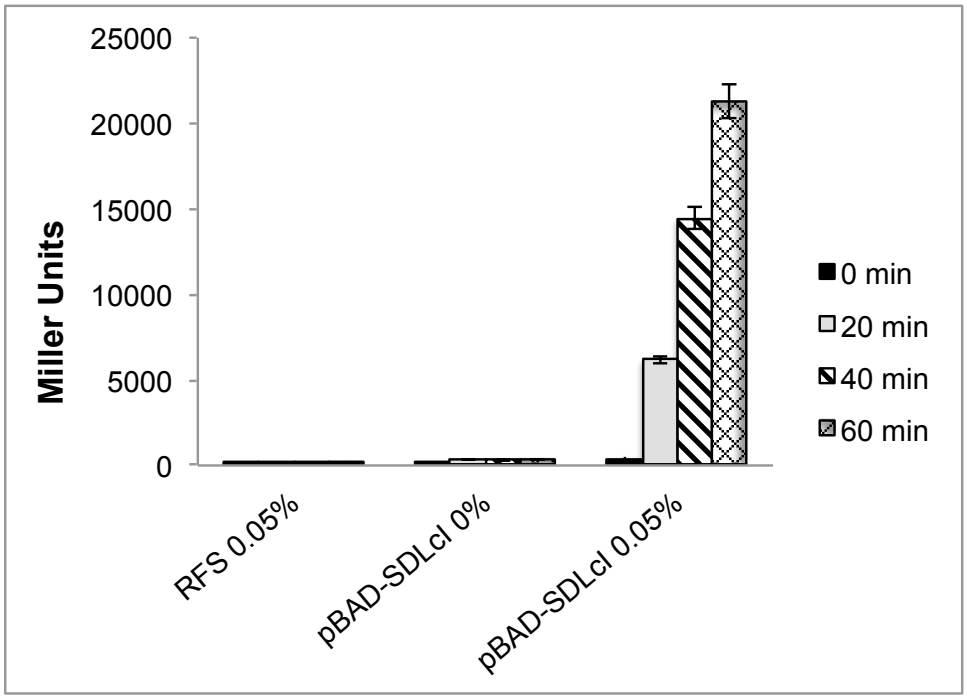

Miller Units
35000
30000
25000
20000
15000
10000
5000
0
0 min
20 min
30 min
40 min
50 min
60 min
RFS
pBAD33, 1%
pBAD-uATG2-lacZ, 0%
pBAD-uATG2-lacZ, 0.05%
pBAD-uATG2-lacZ, 0.1%
pBAD-uATG2-lacZ, 1%

Figure B-2. Induction of the pBAD-LL*c*I-*lacZ* and pBAD-SDL*c*I-*lacZ* expression vectors with 0.05% of L-arabinose. Cultures with pBAD-LL*c*I-*lacZ* (A) or pBAD-SDL*c*I-*lacZ* (B) were grown in M9 minimal medium with aromatic amino acids (tyrosine 20 µg/mL, tryptophan 20 µg/mL, phenylalanine 20 µg/mL) to an OD_{600} of ~0.4. Cultures were then split equally into separate flasks and L-arabinose was added to final concentrations of 0% and 0.05%. Control cultures (RFS) were given 0.05% of the inducer. Samples were pulled at 0 min, 20 min, 40 min, and 60 min after induction and measured for β-galactosidase activity. *E. coli* RFS859 is the host strain.

A.


3500
3000
2500
2000
1500
1000
500
0
Miller Units
0 min
20 min
40 min
60 min
RFS 0.05%
pBAD-LLcI 0%
pBAD-LLcI 0.05%

B.

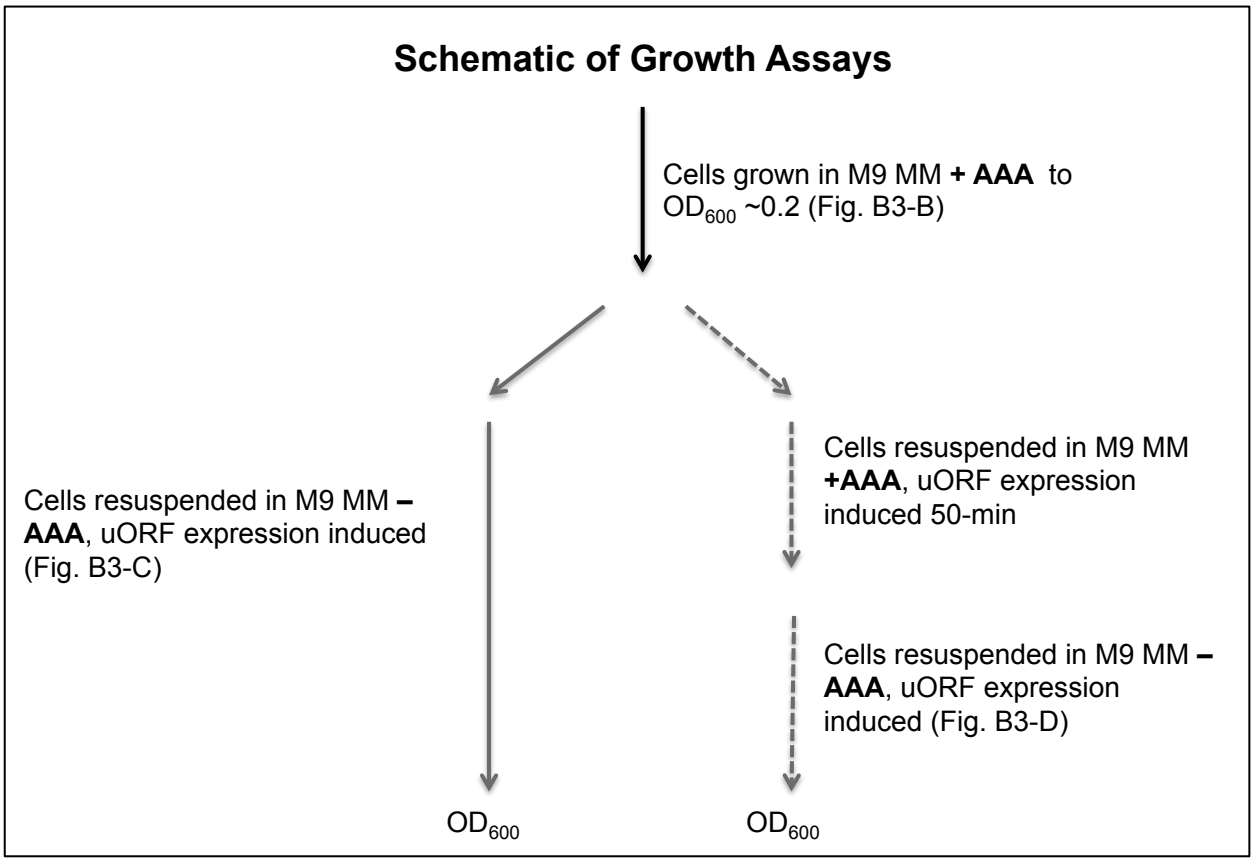

25000
20000
15000
10000
5000
0
Miller Units
0 min
20 min
40 min
60 min
RFS 0.05%
pBAD-SDLcI 0%
pBAD-SDLcI 0.05%

Next, we examined the growth phenotype of cells expressing the *aroL* uORF from the P_{BAD} promoter using the optimal inducer concentration and expression time identified from the β-galactosidase assays described above. Because *aroL* is involved in aromatic amino acid production, we predicted that increased levels of expression of the uORF would give cells an advantage when transitioning from medium containing aromatic amino acids to medium lacking them. To test this prediction, the growth phenotype of a culture in which expression of the *aroL* uORF was induced for 50 minutes before removing aromatic amino acids was compared to a culture in which the uORF was induced at the same time as amino acid limitation. Growth of these cultures was monitored into stationary phase (Fig. B-3). Expression from uAUG-2 (pBAD-uATG2), induced either before or directly after the shift to medium without aromatic amino acids, did not change the growth pattern. The growth rates and total growth were similar regardless of induction or pre-induction of expression from uAUG-2. Similar results were obtained when the experiment was repeated with cultures induced with 1% L-arabinose (data not shown).

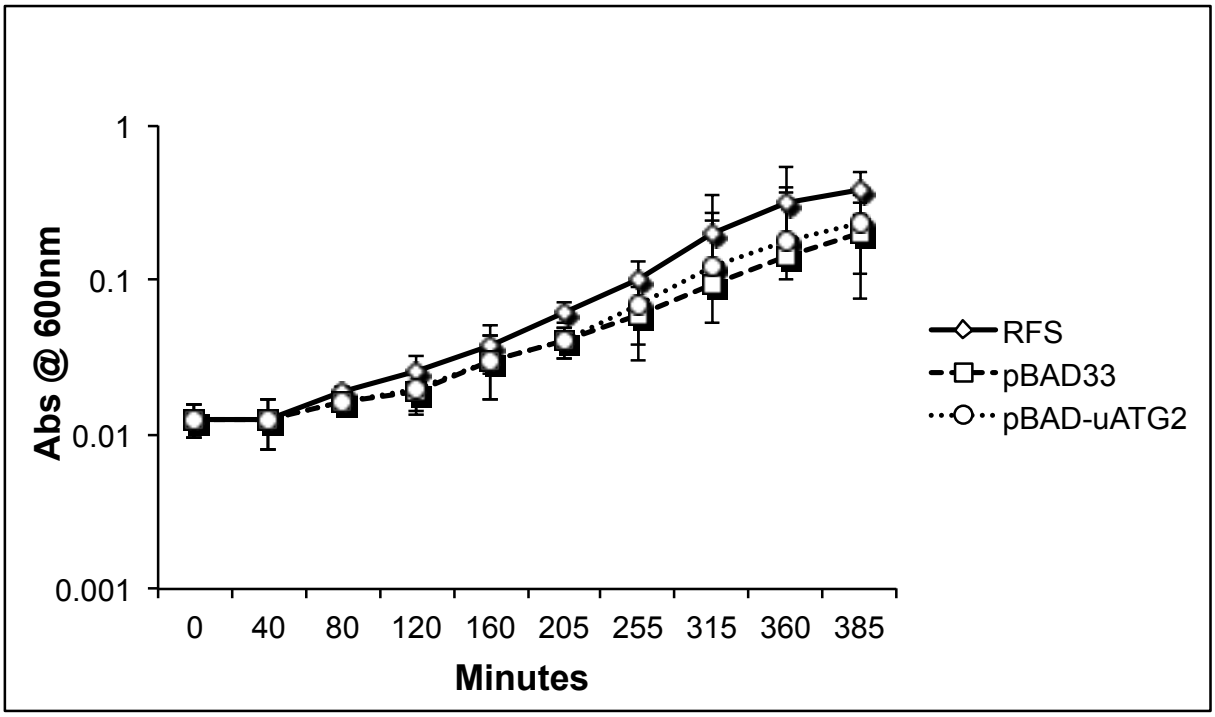
Because *aroL* translation is coupled to translation of the uORF (see Chapter 1, Fig. 1-3), it is possible that the two translation products (shikimate kinase and the putative peptide) are linked functionally; therefore, a growth effect from translation of uAUG-2 might be exerted only with concomitant *aroL* translation and production of shikimate kinase. The growth experiments described above were repeated with pBAD-uATG-2-*aroL*, in which expression of both the *aroL* uORF and *aroL* CDS were induced from the P_{BAD} promoter. No changes in either rates or total growth were observed upon induction of the *aroL* peptide and *aroL* CDS (data not shown).

Figure B-3. Growth assays of *E. coli* RFS859 cells upon induction of pBAD expression vectors. A. Schematic of growth assays. B. Uninduced *E. coli* RFS859 cultures with no vector, pBAD33, or pBAD-uATG2 were grown in M9 minimal medium with aromatic amino acids (tyrosine 20 μg/mL, tryptophan 20 μg/mL, phenylalanine 20 μg/mL) to early log phase OD_{600} of ~0.2. Cultures were then split and induced with 0.05% L-arabinose prior to or following transition into M9 minimal medium lacking aromatic amino acids. C. Cells were spun down, washed and resuspended in fresh medium lacking the aromatic amino acids prior to induction. D. Cells were induced for 50 minutes before being washed and resuspended in the fresh medium lacking the aromatic amino acids.

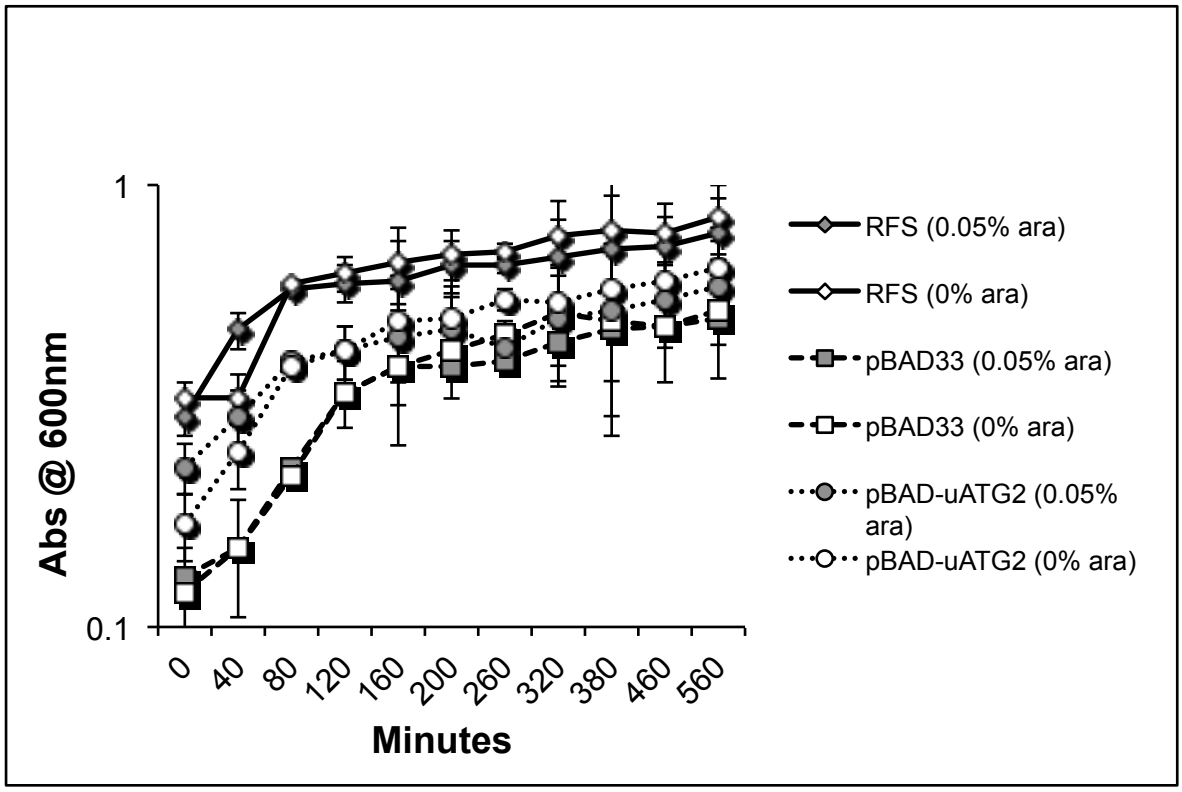
A.


Schematic of Growth Assays
Cells grown in M9 MM + AAA to OD600 ~0.2 (Fig. B3-B)
Cells resuspended in M9 MM – AAA, uORF expression induced (Fig. B3-C)
Cells resuspended in M9 MM +AAA, uORF expression induced 50-min
Cells resuspended in M9 MM – AAA, uORF expression induced (Fig. B3-D)
OD600
OD600

B.


Abs @ 600nm
1
0.1
0.01
0.001
0
40
80
120
160
205
255
315
360
385
Minutes
RFS
pBAD33
pBAD-uATG2

C.


1
0.1
Abs @ 600nm
0
40
80
120
160
200
260
320
380
460
560
Minutes
RFS (0.05% ara)
RFS (0% ara)
pBAD33 (0.05% ara)
pBAD33 (0% ara)
pBAD-uATG2 (0.05% ara)
pBAD-uATG2 (0% ara)

D.

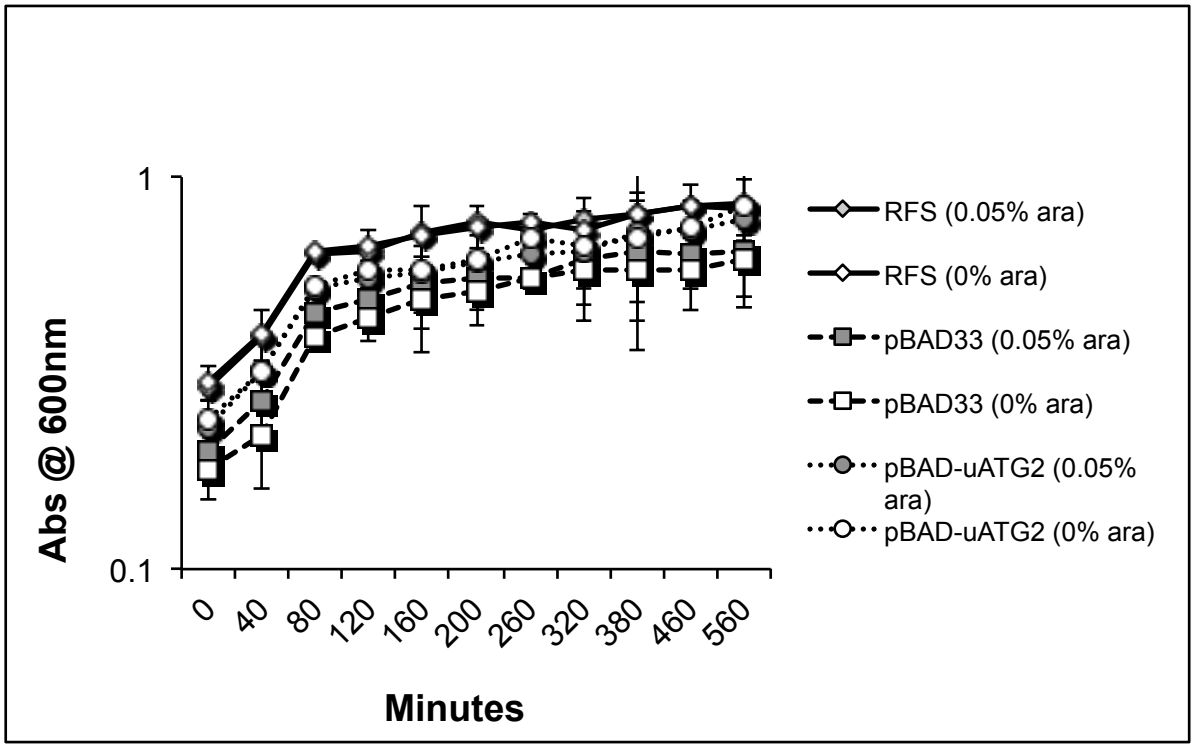

1
0.1
Abs @ 600nm
0
40
80
120
160
200
260
320
380
460
560
Minutes
RFS (0.05% ara)
RFS (0% ara)
pBAD33 (0.05% ara)
pBAD33 (0% ara)
pBAD-uATG2 (0.05% ara)
pBAD-uATG2 (0% ara)

Figure B-4. Hydrophobicity analysis of the *aroL* upstream open reading frame. GenScript: The Biology CRO Peptide Property Calculator (https://www.genscript.com/ssl-bin/site2/peptide_calculation.cgi) was utilized to determine the hydrophobicity of the putative peptide sequence encoded in the *aroL* upstream open reading frame. The putative peptide sequence contains an abundance of hydrophobic, uncharged residues (green).

Hydrophilicity Analysis:

Peptide	Charge	Attribute
MIAILMTPAFAALRPIGENPR	1	basic

Note:

- Red: acidic residues, like D E and C-terminal -COOH
- Blue: basic residues, like R K H and N-terminal -NH2
- Green: hydrophobic uncharged residues, like F I L M V W A and P
- Black: other residues, like G S T C N Q and P
- **Z:** Unrecognized codes are replaced of 'Z'.

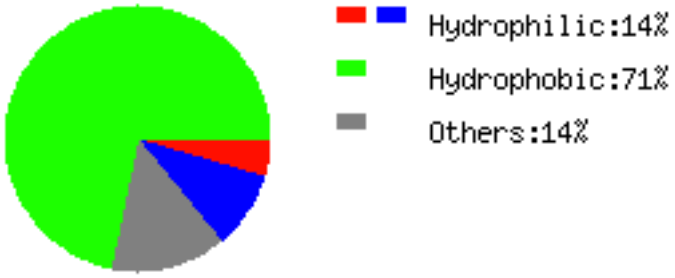
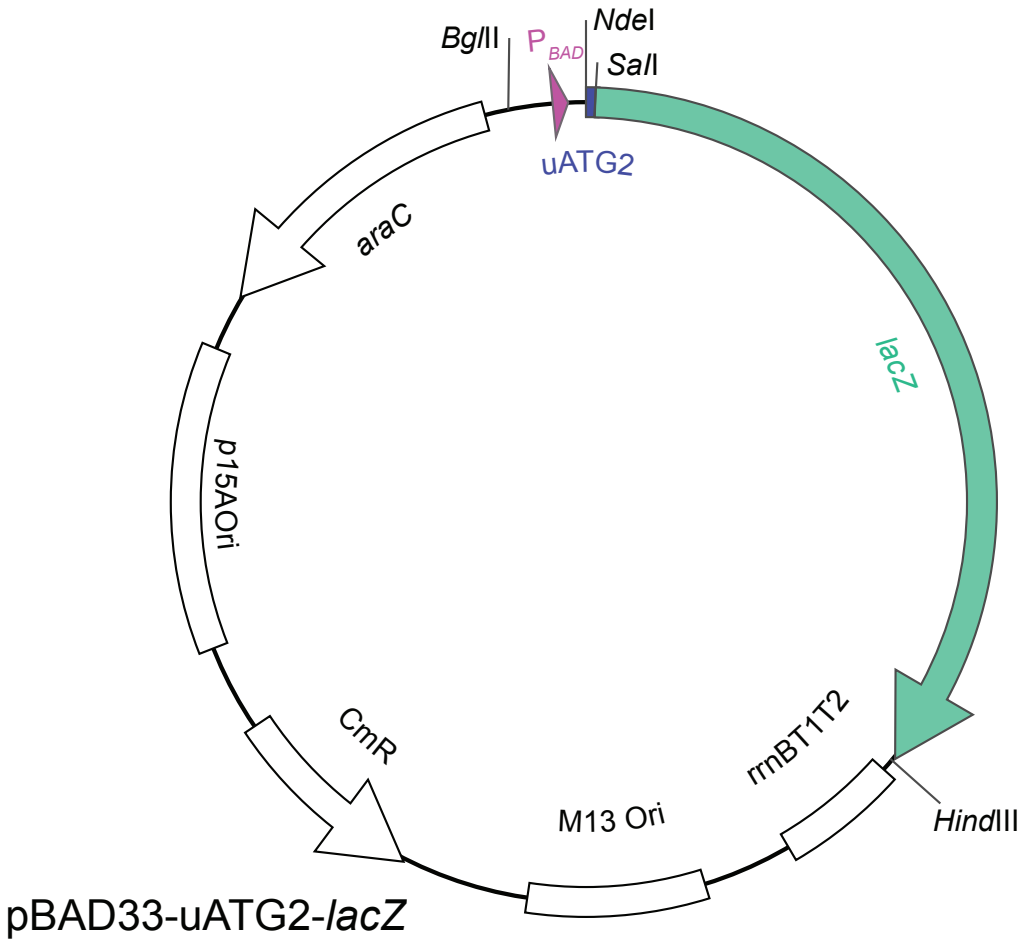


Figure B-5. Vector map (A) and partial sequence (B) of pBAD-uATG2-*lacZ*. The *aroL* leader was originally subcloned into pA904 (*lac* promoter, Amp^R) and fused to the fifth codon of *lacZ.* The fusion of the *aroL* upstream open reading frame, beginning with uATG-2 through *lacZ* was amplified using primers to generate a 5' *Nde*I site and 3' *Hind*III site. The DNA product was digested with *Nde*I and *Hind*III and ligated into pBAD33 to generate pBAD-uATG2-*lacZ*. A second construct was designed in which the upstream open reading frame beginning with uATG-3 was in frame with *lacZ*, named pBAD-uATG3-*lacZ* and not shown. The orientation of the P_{BAD} promoter (pink), uATG-2 open reading frame (blue) and *lacZ* (green) are shown in A. The transcriptional start site (+1, bold), ribosome binding site (underlined), *Nde*I and *Sal*I restriction sites (underlined), and parts of the upstream open reading frame and *lacZ* coding sequences are shown in B. The uATG-2 start codon is part of the *Nde*I site (boxed).

A.



pBAD33-uATG2-*lacZ*

B.

+1

....GTTTCTCCATACCCGTTTTTTTGGGCTAGCAGGAGGAATTCCATATGATCGCTATTCTCATGACACC

RBS NdeI

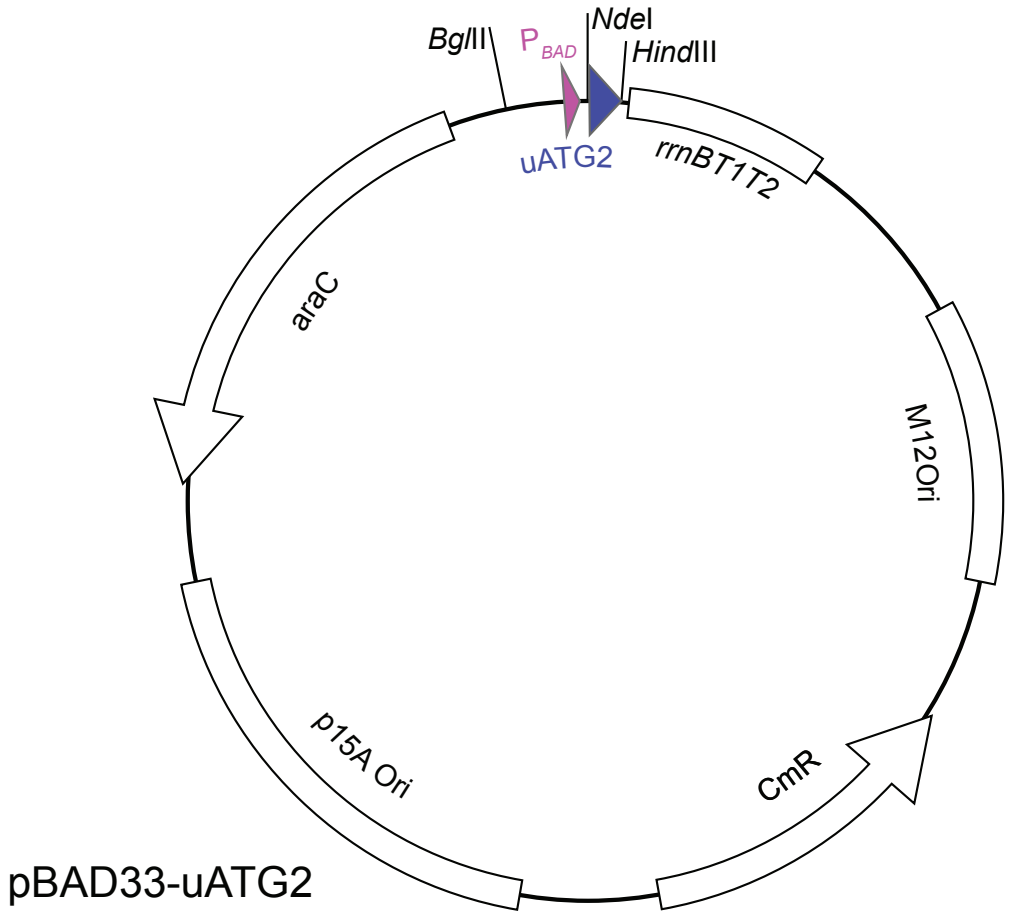
GGCTTTCGCCGCAGTCGACACGGATTCACTGGCCGTCG…

SalI

lacZ

Figure B-6. Vector map (A) and partial sequence (B) of pBAD-uATG2. The uATG-2 coding region through the stop codon (an additional stop codon was added directly adjacent to the first) was amplified from *E. coli* K12 genomic DNA using primers to generate a 5' *Nde*I site and a 3' *Hind*III site. The DNA product was digested with *Nde*I and *Hind*III and ligated into pBAD33 to generate pBAD-uATG2. A second construct was designed in which the upstream open reading frame beginning with uATG-3 through the stop codon was cloned into pBAD33, named pBAD-uATG3 and not shown. The orientation of the P_{BAD} promoter (pink) and uATG-2 open reading frame (blue) are shown in A. The transcriptional start site (+1, bold), ribosome binding site (underlined), *Nde*I and *Hind*III restriction sites (underlined), and the upstream open reading frame sequence is shown in B. The uATG-2 start codon is part of the *Nde*I site (boxed). The two adjacent stop codons of the uATG-2 coding region are boxed.

A.



B.

+1

....GTTTCTCCAT**A**CCCGTTTTTTTGGGCTAGCAGGAGGAATTCCATATGATCGCTATTCTCATGACACCGGC

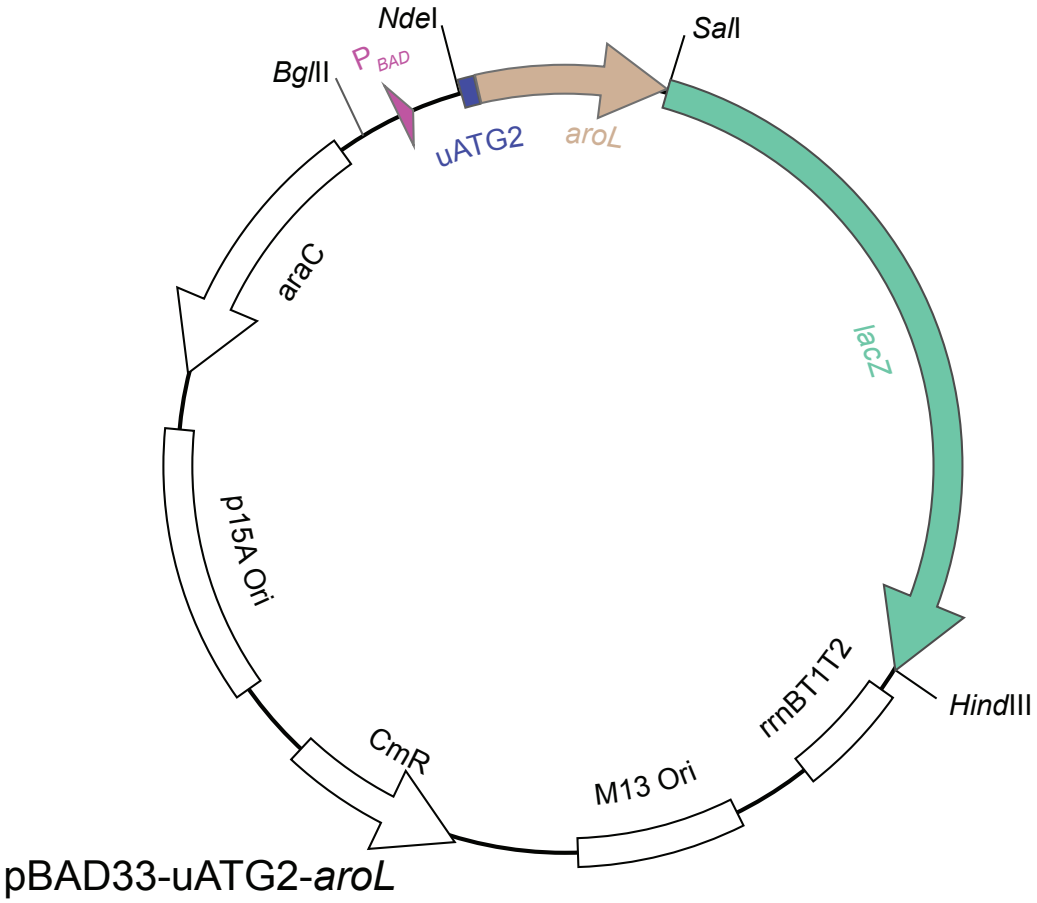
RBS ***Nde*I**

TTTCGCCGCATTGCGACCTATTGGGGAAAACCCACGATAATAAAAGCTTGGC…

***Hind*III**

Figure B-7. Vector map (A) and partial sequence (B) of pBAD-uATG2-*aroL*. The uATG-2 coding region through *aroL* was amplified from *E. coli* K12 genomic DNA using primers to generate a 5' *Nde*I site and a 3' *Sal*I site. The DNA product was digested with *Nde*I and *Sal*I and ligated into pBAD33-uATG2-*lacZ* to generate pBAD-uATG2-*aroL*. The *aroL* stop codon is directly adjacent to the *Sal*I site. The orientation of the P_{BAD} promoter (pink), uATG-2 open reading frame (blue), aroL (brown) and *lacZ* (green) are shown in A. The transcriptional start site (+1, bold), ribosome binding site (underlined), *Nde*I and *Sal*I restriction sites (underlined), and both the upstream open reading frame and *aroL* coding sequences are shown in B. The uATG-2 start codon is part of the *Nde*I site (boxed). The uATG-2 stop codon overlapping the *aroL* start codon and the *aroL* stop codon are boxed.

A.



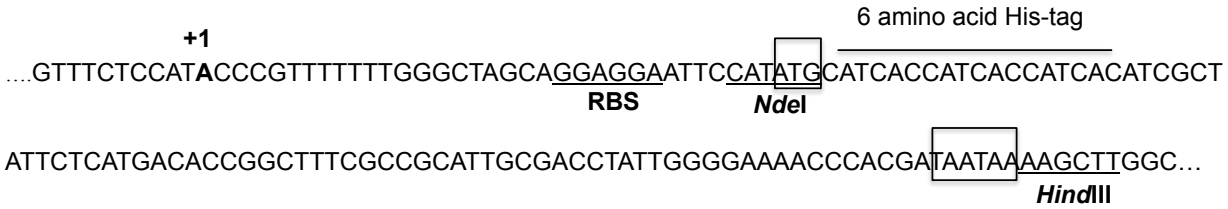
pBAD33-uATG2-*aroL*

B.


+1
....GTTTCTCCATACCCGTTTTTTTGGGCTAGCAGGAGGAATTCCATATGATCGCTATTCT
RBS
NdeI
CATGACACCGGCTTTCGCCGCATTGCGACCTATTGGGGAAAACCCACGATGACACAAC
CTCTTTTTCTGATCGGGCCTCGGGGCTGTGGTAAAACAACGGTCGGAATGGCCCTTGC
CGATTCGCTTAACCGTCGGTTTGTCGATACCGATCAGTGGTTGCAATCACAGCTCAATA
TGACGGTCGCGGAGATCGTCGAAAGGGAAGAGTGGGCGGGATTTCGCGCCAGAGAA
ACGGCGGCGCTGGAAGCGGTAACTGCGCCATCCACCGTTATCGCTACAGGCGGCGGC
ATTATTCTGACGGAATTTAATCGTCACTTCATGCAAAATAACGGGATCGTGGTTTATTTGT
GTGCGCCAGTATCAGTCCTGGTTAACCGACTGCAAGCTGCACCGGAAGAAGATTTACG
GCCAACCTTAACGGGAAAACCGCTGAGCGAAGAAGTTCAGGAAGTGCTGGAAGAACG
CGATGCGCTATATCGCGAAGTTGCGCATATTATCATCGACGCAACAAACGAACCCAGCC
AGGTGATTTCTGAAATTCGCAGCGCCCTGGCACAGACGATCAATTGTTGAGTCGACAC
SalI
GGATTCA…

Figure B-8. Partial sequence of (A) pBAD-His-uATG2 and (B) pBAD-His-uATG2-*aroL*. A. The uATG-2 coding region through the stop codon (an additional stop codon was added directly adjacent to the first) was amplified from *E. coli* K12 genomic DNA using primers to generate a 5' *Nde*I site, a 6-amino acid N-terminal His-tag, and a 3' *Hind*III site. The DNA product was digested with *Nde*I and *Hind*III and ligated into pBAD33 to generate pBAD-His-uATG2. B. The uATG-2 coding region through *aroL* was amplified from *E. coli* K12 genomic DNA using primers to generate a 5' *Nde*I site, 6 amino acid N-terminal His-tag, and a 3' *Sal*I site. The DNA product was digested with *Nde*I and *Sal*I and ligated into pBAD33-uATG2-*lacZ* to generate pBAD-His-uATG2-*aroL*. The transcriptional start site (+1, bold), ribosome binding site (underlined), *Nde*I, *Hind*III and *Sal*I restriction sites (underlined), start and stop codons (boxed) and the 6-amino acid N-terminal His-tag (labeled above) are shown in A and B. The uATG-2 start codon is part of the *Nde*I site (boxed).

A.


6 amino acid His-tag
+1
....GTTTCTCCATACCCGTTTTTTTGGGCTAGCAGGAGGAATTCCATATGCATCACCATCACCATCACATCGCT
RBS
NdeI
ATTCTCATGACACCGGCTTTCGCCGCATTGCGACCTATTGGGGAAAACCCACGATAATAAAAGCTTGGC…
HindIII

B.

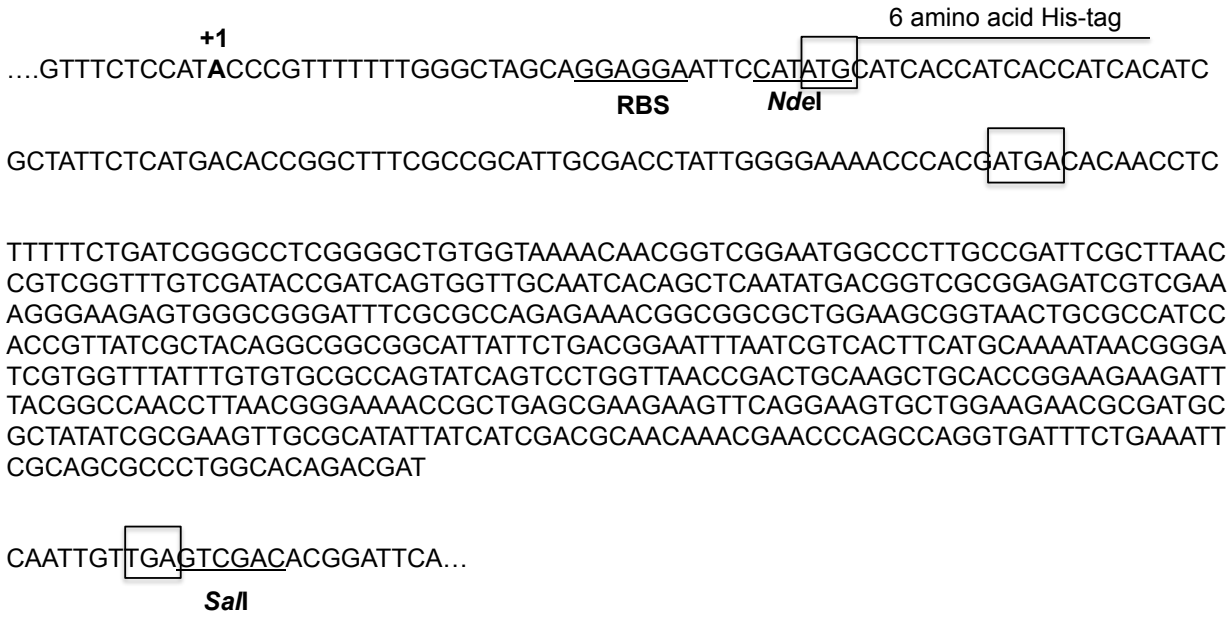
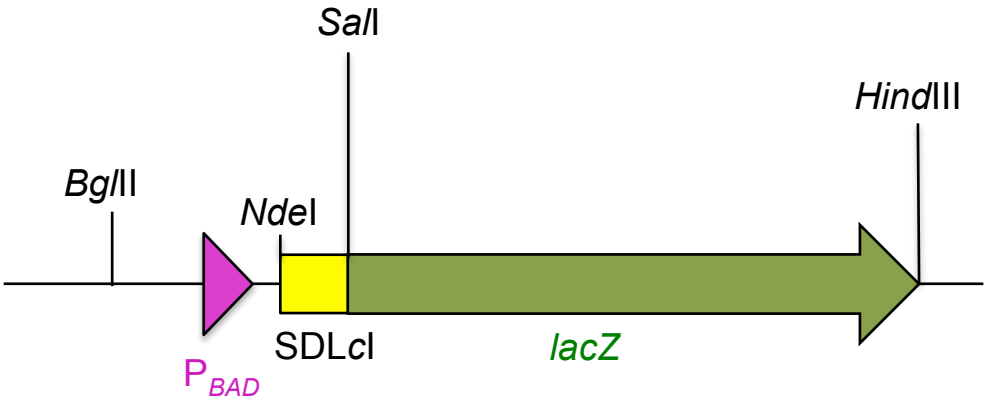

6 amino acid His-tag
+1
....GTTTCTCCATACCCGTTTTTTTGGGCTAGCAGGAGGAATTCCATATGCATCACCATCACCATCACATC
RBS
NdeI
GCTATTCTCATGACACCGGCTTTCGCCGCATTGCGACCTATTGGGGAAAACCCACGATGACACAACCTC
TTTTTCTGATCGGGCCTCGGGGCTGTGGTAAAACAACGGTCGGAATGGCCCTTGCCGATTCGCTTAAC
CGTCGGTTTGTCGATACCGATCAGTGGTTGCAATCACAGCTCAATATGACGGTCGCGGAGATCGTCGAA
AGGGAAGAGTGGGCGGGATTTCGCGCCAGAGAAACGGCGGCGCTGGAAGCGGTAACTGCGCCATCC
ACCGTTATCGCTACAGGCGGCGGCATTATTCTGACGGAATTTAATCGTCACTTCATGCAAAATAACGGGA
TCGTGGTTTATTTGTGTGCGCCAGTATCAGTCCTGGTTAACCGACTGCAAGCTGCACCGGAAGAAGATT
TACGGCCAACCTTAACGGGAAAACCGCTGAGCGAAGAAGTTCAGGAAGTGCTGGAAGAACGCGATGC
GCTATATCGCGAAGTTGCGCATATTATCATCGACGCAACAAACGAACCCAGCCAGGTGATTTCTGAAATT
CGCAGCGCCCTGGCACAGACGAT
CAATTGTTGAGTCGACACGGATTCA…
SalI

Figure B-9. Vector map and partial sequence of pBAD-SDL*c*I-*lacZ* and pBAD-LL*c*I-*lacZ*. Codons 1-16 of a lac-leadered *c*I (A and B) or leaderless *c*I (C and D) gene were amplified using primers to generate a 5' *Nde*I site and a 3' *Sal*I site from pSDL*c*I-*lacZ* (lac-leadered, Amp^R) and LL*c*I-*lacZ* (leaderless, Amp^R) encoding *c*I codons 1-16 fused to the fifth codon of *lacZ* (O'Donnell and Janssen, 2001). The DNA product was digested with *Nde*I and *Sal*I and ligated into pBAD-uATG2-*lacZ* to generate pBAD-SDL*c*I-*lacZ,* as shown in A and B. The DNA product was digested with *Nde*I and *Sal*I and ligated into a modified version of pBAD33, which placed the A of the ATG *c*I start codon at the transcriptional start of the vector to generate pBAD-LL*c*I-*lacZ,* as shown in C and D. The transcriptional start site (+1, bold), start codon (boxed), ribosome binding site (underlined), *Nde*I and *Sal*I restriction sites (underlined), *c*I open reading frame and part of the *lacZ* coding sequence (arrow) are shown. The *c*I start codon is part of the *Nde*I site (boxed).

A.



pBAD33-SDLcI-*lacZ*

B.

+1

….GTTTCTCCAT**A**TAACAATTTCACACAGGAAACAGCATATGAGCACAAAAAAGAAACCATTAACACAA

RBS **NdeI**

GAGCAGCTTGAGGACGCAGTCGACACGGATTCACTGGCCGTCG…

SalI

lacZ

Discussion

We began our investigation of a function for the *aroL* uORF-encoded peptide with the growth assays based on *aroL*'s involvement in aromatic amino acid biosynthesis (Davis and Mingioli, 1953; Berlyn and Giles, 1969; Ely and Pittard, 1979). Expression from uAUG-2 from the inducible P_{BAD} promoter did not appear to aid cells as they transitioned into a medium lacking the aromatic amino acids. These assays were performed with *E. coli* RFS cells that contain a chromosomal *aroL* gene and upstream ORF, in addition to *aroL* uORF expressed from the pBAD expression vector. The data suggest that if expression of the uORF (and putative peptide product) is needed for some aspect of aromatic amino acid biosynthesis other than its influence on *aroL* expression, as shown in Chapter 1, then expression of the chromosome-encoded uORF (and putative peptide) is sufficient for the transition and the presence of more uORF expression does not increase or decrease the cellular fitness.

It is important to note that although we were able to measure expression of the *aroL* uORF when induced from the pBAD vectors, we did not show that a peptide product was present and stable in the cell. We can not rule out the possibility that it was degraded before an effect was observed. It is also possible that the putative peptide product works as part of a complex and without subsequent increased expression of its counterparts, no change in phenotype was observed.

The size and hydrophobicity of the putative peptide (Fig. B-4) is consistent with a class of small hydrophobic regulatory molecules (encoded by small open reading frames) that interact with membrane proteins (Alix and Blanc-Potard, 2009). Examples include the adaptor MgtR that promotes the degradation of MgtC by AAA+ proteases (Alix and Blanc-Potard, 2008); KdpF, which stabilizes the KdpFABC transporter complex (Greie and Altendorf, 2007; Gassel et al., 1999); and YneN, which interacts with the PhoQ sensor kinase (Eguchi et al., 2007). The *aroL* upstream putative peptide may also localize to the membrane where it regulates the transport of the aromatic amino acids or pathway substrates or intermediates into the cell. Assuming that the peptide was produced and stable in the cell, a His-tagged version would provide a way to identify cellular location. Therefore, we placed a 6-amino acid His-tag at the 5'-

terminus of the uATG-2 ORF under control of the inducible P_{BAD} promoter. Our intention was to fractionate cells expressing the uORF and probe for the tagged peptide in the membrane and the soluble cytoplasmic fractions using an anti-His antibody. Fluorescent protein fusions can be used as an alternative approach to visualize localization *in vivo*. It is also possible that the putative peptide has a more subtle function that does not pertain to aromatic amino acid production, and requires a broader approach to identify its target(s). The His-tagged version of the peptide will be used as the bait in pull down assays to identify binding partners (Ahmed et al., 2008; Arifuzzaman et al., 2006).

General Discussion

The experiments performed in this work were designed to address the regulatory features of the *aroL* mRNA in *E. coli* and the *ryhB* sRNA in *S. oneidensis* MR-1. The 5'UTR of the *aroL* mRNA contains additional AUG triplets upstream of the *aroL* start codon (uAUGs) and the data presented in this dissertation show that these uAUGs influence *aroL* translation. The uAUGs themselves are regulated by a complex translation initiation mechanism that involves multiple sequences upstream, downstream, and surrounding them. *ryhB* identified in *S. oneidensis* MR-1 shares some sequence similarity to *ryhB* of *E. coli* but might regulate additional target mRNAs.

The uAUGs in the 5'UTR of the *aroL* mRNA bind 30S subunits and influence *aroL* expression in the absence of a canonical Shine-Dalgarno sequence

In this study we identified three AUG (uAUG-1, -2, -3) triplets that bind 30S subunits in the 5'UTR region of the *aroL* leader upstream of the *aroL* start codon and *aroL* SD sequence. Our results support translational coupling of *aroL* to the uAUG-2 reading frame. The uAUG-2 reading frame encodes a putative 21-amino acid peptide, although we have yet to determine if this peptide is stable and define its function in the cell. A series of deletion and substitution mutations was designed to identify the sequences required for ribosome recognition and expression from uAUG-2 in the absence of a canonical SD sequence. Our data support that multiple signals are required for efficient 30S binding and expression from uAUG-2, which include a U-rich sequence upstream and a sequence downstream of the uAUG. Furthermore, the sequence surrounding uAUG-1 and -2 also influences ribosome recognition.

The U-rich region upstream of uAUG-2 most likely contributes to 30S binding and expression through an interaction with ribosomal protein, S1. There is ample evidence to support an interaction between S1 and A/U-rich upstream elements to enhance weak SD sequences or serve as the sole recognition signal for 30S binding (Boni et al., 1991; Farwell et al., 1992; Ringquist et al., 1995; Tedin et al., 1997). We have shown that 30S

binding to the uAUGs is dependent on S1, but we have yet to show a specific interaction between the ribosomal protein and the U-rich region of the mRNA.

Two sub-regions (DR-2 and DR-4) downstream of uAUG-2 also appear to influence expression. There are numerous examples of sequences downstream of start codons that influence gene expression. Predicted secondary structure analysis suggests that the downstream region acts as a standby site where the ribosome associates and waits for the release of structure occluding uAUG-2. It is also possible that the downstream regions DR-2 and DR-4 are required to promote non SD-ASD interactions between the ribosome and mRNA in order to place uAUG-2 in the ribosomal P site. Crosslinking analysis has shown a number of interactions between the ribosome and nucleotides of the mRNA upstream and downstream of start codons (Rinke-Appel et al., 1991; McCarthy and Brimacombe, 1994; Rinke-Appel et al., 1994; La Teanna et al., 1995). DR-2 and DR-4 might enhance these interactions or they might make alternative contacts with the ribosomal proteins or ribosomal RNA. The interactions between DR-2 and DR-4 and the ribosome might also help stabilize the association of the ribosome with the U-rich region upstream of uAUG-2 to enhance placement in the P site.

The nucleotide sequence directly surrounding uAUG-1 and uAUG-2 also influences 30S binding and expression from uAUG-2. Mutating two G residues (bold, **G**AUG**G**UAUG) or deleting part of the surrounding sequence (underlined, AUGGUAUG) resulted in increased binding and expression from the uAUG. We favor two hypotheses to explain these observations. The first states that these mutations have disrupted secondary structure that normally obstructs the uAUG. The other is that the G rich region surrounding the uAUGs represents a pseudo-SD sequence and base pairs to the 16S RNA, placing uAUG-2 in a less favorable position to the P site. Improper positioning of SD sequences relative to the start codon can change ribosome recognition and expression patterns in a handful of genes tested (Sedlácek et al., 1979; Weiss et al., 1988; Chen et al., 1994; Jin et al., 2006).

Identification and characterization of a RyhB homolog in *S. oneidensis* MR-1

The experiments described in the second half of this thesis focused on characterizing an sRNA previously identified in *S. oneidensis* MR-1 (Wan et al., 2004). This sRNA shares similarity on a sequence and predicted secondary structure level to *ryhB* in *E. coli* and related organisms. *ryhB* is expressed in *E. coli* in response to iron limitation in the cell (Massé and Gottesman, 2002); therefore, we examined expression levels of the sRNA in *S. oneidensis* MR-1 cells under conditions of iron limitation and iron excess. As expected, our results showed an increase in expression under iron-deplete conditions and decrease in expression under iron-replete conditions. These results support a role for the sRNA in iron regulation in *S. oneidensis* MR-1. We also identified a predicted Fur box in the promoter of the sRNA, which further supports a role in iron regulation in the cell. We have yet to show an interaction between Fur and this sequence on the *S. oneidensis* MR-1 genomic DNA.

Although *ryhB* identified in *S. oneidensis* MR-1 shares sequence similarity with that of *E. coli*, the two sRNAs are not identical. Predicted secondary structure analysis suggests that the regions of sequence that are unique to the *S. oneidensis* MR-1 *ryhB* are in unpaired, open regions, making them available to pair with target mRNAs. We predict that the two sRNAs will have different mRNA targets in *E. coli* and *S. oneidensis* MR-1 based off of each organism's unique physiological needs. We have yet to determine if *ryhB* in *S. oneidensis* MR-1 requires Hfq for proper recognition of and RNaseE for degradation of its target mRNAs.

Concluding Remarks

Despite recent advances in RNA biology, we have yet to understand how extensively this molecule is used as a regulator in the cell and to define all of its roles in physiology and mechanisms of action. The discovery of small non-coding RNAs has helped explain regulatory phenomena that had previously perplexed researchers. The *aroL* mRNA contains unusual features in its 5'UTR that contribute to gene expression. Numerous mechanisms in addition to the traditional SD sequence are utilized by mRNAs for ribosome recognition and expression, suggesting that the mRNAs

themselves help regulate levels of the gene product. The results presented in this dissertation provide additional insight for implications of RNA as a regulator in the cell.

References

Accetto, T. and Avgustin, G. 2011. Inability of *Prevotella bryantii* to form a functional Shine-Dalgarno interaction reflects unique evolution of ribosome binding sites in Bacteroidetes. *PLoS ONE* **6**: 1-9.

Ahmed, N.T., Gao, C., Lucker, B.F., Cole, D.G. and Mitchell, D.R. 2008. ODA16 aids axonemal outer row dynein assembly through an interaction with the intraflagellar transport machinery. *J. Cell Biol.* **183**: 313-322.

Alix, E. and Blanc-Potard, A.B. 2007. MgtC a key player in intramacrophage survival. *Trends Microbiol*. **15**: 252-256

Alix, E. and Blanc-Potard, A.B. 2008. Peptide-assisted degradation of the Salmonella MgtC virulence factor. *EMBO J* **27**: 546-557.

Alix, E. and Blanc-Potard, A.B. 2009. Hydrophobic peptides: novel regulators within bacterial membrane. *Mol. Microbiol.* **72**(1): 5-11.

Altuvia, S. 2004. Regulatory small RNAs: the key to coordinating global regulatory circuits. *J. Bacteriol*. **186**(20): 6679-6680.

Altuvia, S. and Wagner, E.G. 2000. Switching on and off with RNA. *Proc. Natl. Acad. Sci.* **97**: 9824–9826.

Arifuzzaman, M., Maeda, M., Itoh, A., Nishikata, K., Takita, C., Saito, R., Ara, T., Nakahigashi, K., Huang, H-C., Hirai, A., Tsuzuki, K., Nakamura, S., Noka-Nakamichi, M., Kitagawa, M., Tomita, M., Kanaya, S., Wada, C. and Mori, H. 2006. Large-scale identification of protein-protein interaction of *Escherichia coli* K12. *Genome Research* **16**(5): 696-691.

Baird, S.D., Turcotte, M., Korneluk, R.G. and Holcik, M. 2006. Searching for IRES. *RNA* **12**(10): 1755-1785.

Bagg, A. and Beilands, J.B. 1987. Ferric uptake regulation protein acts as a repressor, employing iron (II) as a cofactor to bind the operator of an iron transport operon in *Escherichia coli*. *Biochemistry* **26**: 5471-5477.

Bandyra, K.J., Said, N., Pfeiffer, V., Gorna, M.W., Vogel, J., Luisi, B.F. 2012. The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNaseE. *Mol. Cell* **47**(6): 943-953.

Bauer, R. and Dicks, L.M. 2005. Mode of action of lipid II-targeting antibiotics. *Int. J. Food Microbiol*. **101**: 201-216.

Benelli, D., Maone, E. and Londei, P. 2003. Two different mechanisms for ribosome/mRNA interaction in archaeal translation initiation. *Mol. Microbiol*. **50**(2): 635-43.

Berlyn, M.D. and Giles, N.H. 1969. Organization of enzymes in the polyaromatic synthetic pathway: separability in bacteria. *J. Bacteriol.* **99**: 222-230.

Bertani, G. 1951. Studies on lysogenesis I.: The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol*. **62**: 293.

Bibb, M.J., White, J., Ward, J.M. and Janssen, G.R. 1994. The mRNA for the 23S rRNA methylase encoded by the *ermE* gene of *Saccharopolyspora erythraea* is translated in the absence of a conventional ribosome-binding site. *Mol. Microbiol.* **14**: 533-545.

Boisset, S., Geissmann, T., Huntzinger, E., Fechter, P., Bendridi, N,. Possedko, M., Chevalier, C., Helfer, A.C., Benito, Y. and Jacquier, A. 2007. *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Genes Dev*. **21**: 1353–1366.

Boni, I.V., Isaeva, D.M., Musychenko, M.L. and Tzareva, N.V. 1991. Ribosome-messenger recognition: mRNA target sites for ribosomal protein S1. *Nucleic Acids Res.* **19**: 155-162.

Bordi, C., Ibbo-Nivol, C., Mejean, V. and Patte, J-C. 2003. Effects of ISSo2 insertions in structural and regulatory genes of the trimethylamine oxide reductase of *Shewanella oneidensis*. *J. Bacteriol*. **185**(6): 2042-2045.

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

Brock, J.E., Paz, R.L., Cottle, P. and Janssen, G.R. 2007. Naturally occurring adenines within mRNA coding sequences affect ribosome binding and expression in *Escherichia coli. J. Bacteriol.* **189**: 501-510.

Brock, J.E., Pourshahian, S., Giliberti, J., Limbach, P.A. and Janssen, G.R. 2008. Ribosomes bind leaderless mRNA in *Escherichia coli* through recognition of their 5'-terminal AUG. *RNA* **14**(10): 2159-2169.

Bulmer, M. 1988. Codon usage and intragenic position. *Journal of Theoretical Biology*. **133**: 67-71.

Chang, B., Halgamuge, S. and Tang. S-L. 2006. Analysis of SD sequences in completed microbial genomes: non-SD-led genes are as common as SD-led genes. *Gene* **373**: 90-99.

Chen, H., Bjerknes, M., Kumar, R. and Jay, E. 1994. Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res*. **22:** 4953-4957.

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

Christiansen, J.K., Nielsen, J.S., Ebersbach, T., Valentin-Hansen, P., Sogaard-Andersen, L. and Kallipolitis, B.H. 2006. Identification of small Hfq-binding RNAs in *Listeria monocytogenes*. *RNA* **12**: 1383–1396.

Cogen, A.L., Yamasaki, K., Muto, J., Sanchezz, K.M., Crotty Alexander, L., Tanios, J., Lai, Y., Kim, J.E., Nizet, V. and Gallo, R.L. 2010. *Staphylococcus epidermidis* antimicrobial delta-toxin (phenol-soluble modulin-gamma) cooperates with host antimicrobial peptides to kill group A Streptococcus. *PLoS ONE* **5**: e8557.

Compan, I. and Touati, D. 1993. Interaction of six global transcription regulators in expression of manganese superoxide dismutase in *Escherichia coli* K-12. *J. Bacteriol*. **175**: 1687-96.

Cunningham, K.A. and Burkholder, W.F. 2009. The histidine kinase inhibitor Sda binds near the site of autophosphorylation and may sterically hinder autophosphorylation and phosphotransfer to spo0F. *Mol Microbiol.* **71**: 659-677.

Dalboge, H., Carlsen, S., Jensen, E.B., Christensen, T. and Dahl, H.H. 1988. Expression of recombinant growth hormone in *Escherichia coli*: effect of the region between the Shine-Dalgarno sequence and the ATG initiation codon. *DNA* **7**: 399-405.

Davanloo, P., Rosenber, A.H., Dunn, J.J. and Studier, F.W. 1984. Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **81**: 2035-2039.

Davis, B.D. and Mingioli, E.S. 1953. Aromatic biosynthesis. VII. Accumulation of two derivatives of shikimic acid by bacterial mutants. *J. Bacteriol.* **66**: 129-136.

De Boer, H.A., Comstock, L.J., Hui, A., Wong, E. and Vasser, M. 1983a. Portable Shine-Dalgarno regions; nucleotides between the Shine-Dalgarno sequence and the start codon affect the translation efficiency. *Gene Amplif. Anal*. **3**: 103-116.

de Lorenzo, V., Giovannini, F., Herrero, M. and Neilands, J.B. 1988. Operator sequences of the aerobactin operon of plasmid ColV-K30 binding the ferric uptake regulation (fur) repressor. *J. Bacteriol.* **169:** 2624-2630.

de Smit, M.H., and van Duin, J. 1990. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 7668-7672.

de Smit, M.H., and van Duin, J. 1994. Control of translation by mRNA secondary structure in *Escherichia coli.* A quantitative analysis of literature data. *J. Mol. Biol.* **244**: 144-150.

de Smit, M.H., and van Duin, J. 2003. Translational standby sites: how ribosomes may deal with the rapid folding kinetics of mRNA. *J. Mol. Biol.* **331**: 737-743.

DeFeyter, R.C and Pittard, J. 1986. Genetic and molecular analysis of *aroL*, the gene for shikimate kinase II in *Escherichai coli* K12. *J. Bacteriol.* **165**: 226-232.

Deeley, M.C. and Yanofsky, C. 1982. Transcription initiation at the tryptophanase promoter of *Escherichia coli* K-12. *J. Bacteriol.* **151**: 942-951.

Dong, H., Nilsson, L. and Kurland, C.G. 1996. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.* 260: 649-663.

Dvir, S., Velten, L., Sharon, E., Zeevi, D., Carey, L.B., Weinberger, A. and Segala, E. 2013. Deciphering the rules by which 5'-UTR sequences affect protein expression in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **110**(30): E2792-2801.

Eguchi, Y., Itou, J., Yamane, M., Demizu, R., Yamato, F. and Okada, A. 2007. B1500, a small membrane protein, connects the two-component systems EvgS/EvgA and PhoQ/PhoP in *Escherichia coli*. *Proc Natl Acad Sci U.S.A*. **104**: 18712-18717.

Eguchi, Y., Okada, T., Minagawa, S., Oshima, T., Mori, H. and Yamamoto, K. 2004. Signal transduction cascade between EvgA/EvgS and PhoP/PhoQ two-component systems of *Escherichia coli. J. Bacteriol.* **186**: 3006-3014.

Ely, B. and Pittard, J. 1979. Aromatic amino acid biosynthesis: regulation of shikimate kinase in *Escherichia coli* K-12. *J. Bacteriol.* **138**: 933-943.

Escolar, L., de Lorenzo, V. and Perez-Marin J. 1997. Binding of the fur (ferric uptake regulator) repressor of *Escherichia coli* to arrays of the GATAAT sequence. *J. Mol. Biol.* **283**: 537-547.

Eyre-Walker. A. 1996. Synonymous codon bias is related to gene length in *Escherichia coli*: selection for translational accuracy*? Mol. Biol. Evol.* **13**: 864-872.

Farwell, M.A., Roberts, M.W. and Rabinowitz, J.C. 1992. The effect of ribosomal protein S1 from *Escherichia coli* and *Micrococus luteus* on protein synthesis in vitro by *E.coli* and *Bacillus subtilis*. *Mol. Microbiol.* **6**: 3375-3383.

Fozo, E.M., Hemm, M.R. and Storz, G. 2008. Small toxic proteins and the antisense RNAs that repress them. *Microbiol. Mol. Biol. Rev.* **72**: 579-589.

Fozo, E.M., Makarova, K.S., Shabalina, S.A., Yutin, N., Koonin, E.V. and Storz, G. 2010. Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families. *Nucleic Acids Res*. **38**: 3743-3759.

Fredrick, K. and Noller, H.F. 2002. Accurate translocation of mRNA by the ribosome requires a peptidyl group or its analog on the tRNA moving into the 30S P site. *Mol. Cell.* **9**: 1125-31.

Gaballa, A., Antelmann, H., Aguilar, C., Khakh, S.K., Song, K.B., Smaldone, G.T. and Helmann, J.D. 2008. The *Bacillus subtilis* iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. *Proc Natl Acad Sci U.S.A.* **105**: 11927-11932.

Gallie, D.R. and Kado, C.I. 1989. A translational enhancer derived from tobacco mosaic virus is functionally equivalent to a Shine-Dalgarno sequence. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 129-132.

Gallie, D.R., Sleat, D.E., Watts, J.W., Turner, P.C. and Wilson, T.M.A. 1987. The 5'-leader sequence of tobacco mosaic virus RNA enhances the expression of foreign gene transcripts in vitro and in vivo. *Nucleic Acids Res.* **15:** 3257-3272.

Ganoza, M.C., Fraser, A.R. and Neilson, T. 1978. Nucleotides contiguous to AUG affect translation initiation. *Biochemistry* **17**: 2769-2775.

Ganoza, M.C., Marliere, P., Kofoid, E.C. and Louis, B.G. 1985. Initiator tRNA may recognize more than the initiation codon in mRNA: A model for translation initiation. *Proc. Natl. Acad. Sci. U.S.A*. **82**: 4587-4591.

Gao, W., Tyagi, S., Kramer, F.R. and Goldman, E. 1997. Messenger RNA release from ribosomes during 5'-translational blockage by consecutive low-usage arginine but not leucine codons in *Escherichia coli. Mol. Microbiol.* 25: 707-716.

Gassel, M., Mollenkamp, T., Puppe, W. and Altendorf, K. 1999. The KdpF subunit is part of the K(+)-translocating Kdp complex of *Escherichia coli* and is responsible for stabilization of the complex in vitro. *J. Biol. Chem.* **274**: 37901-37907.

Gerdes, K., Gultyaev, A.P., Franch, T., Pedersen, K. and Mikkelsen, N.D. 1997. Antisense RNA-regulated programmed cell death. *Annu. Rev. Genet*. **31**: 1–31.

Gerhardt, P., Murray, R.G.E., Wood, W.A. and Krieg, N.R. 1994. Methods for General and Molecular Bacteriology. American Society for Microbiology Press, Washington, DC.

Glass, M.J. and Summers, D.F. 1992. A cis-acting element within the hepatitis A virus 5' non-coding region required for in vitro translation. *Virus Res*. **26**: 15-31.

Golshani, A., Golomehova, V., Mironova, R., Ivanov, I.G., and AbouHaidar, M.G. 1997. Does the epsilon sequence of phage T7 function as an initiator for the translation of CAT mRNA in *Escherichia coli*? *Biochem. Biophys. Res. Commun*. **236:** 253-256.

Golshani, A., Kolev, V., AbouHaidar, M.G. and Ivanov, I.G. 2000. Epsilon as an initiator of translation of CAT mRNA in *Escherichia coli. Biochem. Biophys. Res. Commun*. **273**: 528-531.

Göpel, Y., Papenfort, K., Reichenbach, B., Vogel, J. and Görke, B. 2013. Targeted decay of a regulatory small RNA by an adaptor protein for RNaseE and counteraction by an anti-adaptor RNA. *Genes Dev.* **27**(5): 552-564.

Gottesman, S. 2004. The small RNA regulators of *Escherichia coli*: roles and mechanisms. *Annual Rev. Microbiology* **58:** 303-328.

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

Greie, J.C. and Altendorf, K. 2007. The K+-translocating KdpFABC complex from *Escherichia coli*: a P-type ATPase with unique features. .*J. Bioenerg. Biomembr*. **39**: 397-402.

Gu, W., Zhou, T. and Wilke, C.O. 2010. A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Computational Biology* **6**: 1-8.

Gunton, J.E., Gilmour, M.W., Baptista, K.P., Lawley, T.D. and Taylor, D. 2007. Interaction between the co-inherited TraG coupling protein and the TraJ membrane-associated protein of the H-plasmid conjugative DNA transfer system resembles chromosomal DNA translocases. *Microbiology* **153**: 428-441.

Guzman, L.M., Belin, D., Carson, M.J. and Beckwith, J. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**(14): 4121-30.

Handler, A.A., Lim, J.E. and Losick, R. 2008. Peptide inhibitor of cytokinesis during sporulation in *Bacillus subtilis*. *Mol. Microbiol*. **68**: 588-599.

Hantke, K. 1981. Regulation of ferric iron transport in *Escherichia coli* K12: isolation of a constitutive mutant. *Mol. Gen. Genet*. **182:** 288-292.

Hanktke, K., and Braun, V. 1997. Control of bacterial iron transport by regulatory proteins, p. 11-44. In S. Silver and W. Walden (ed.), Metal ions in gene regulation. Chapman and Hall, New York, N.Y.

Heidrich, N., Moll, I. and Brantl, S. 2007. *In vitro* analysis of the interaction between the small RNA SR1 and its primary target *ahrC* mRNA. *Nucleic Acids Res*. **35**: 4331–4346.

Hering, O., Brenneis, M., Beer, J., Suess, B. and Soppa, J. 2009. A novel mechanism for translation initiation operates in haloarchaea. *Mol. Microbiol.* **71**: 1451-1463.

Hirose, T. and Sugiura, M. 2004. Multiple elements required for translation of plastid *atpB* mRNA lacking the Shine-Dalgarno sequence. *Nucleic Acids Res.* **32**: 3503-3510.

Hobbs, E.C., Fontaine, F., Yin, X. and Storz, G. 2011. An expanding universe of small proteins. *Current Opinion in Microbiology* **14:** 167-173.

Hoe, N.P. and Goguen, J.D. 1993. Temperature sensing in *Yersinia pestis*: translation of the LcrF activator protein is thermally regulated. *J. Bacteriol*. **175**(24): 7901-7909.

Hook-Barnard, I.G., Brickman, T.J. and McIntosh, M.A. 2007. Identification of an AU-rich translational enhancer within the *Escherichia coli fepB* leader RNA. *J. Bacteriol.* **289**: 4028-37.

Hui, A. and de Boer, H.A. 1987. Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli*. *PNAS* **84**: 4762-6.

Hui, A., Hayflick, J., Dinkelspiel, K. and de Boer, H.A. 1984. Mutagenesis of the three bases preceding the start codon of the β-galactosidase mRNA and its effect on translation in *Escherichia coli*. *EMBO J*. **3**: 623-29.

Imlay, J.A. and Linn, S. 1988. DNA damage and oxygen radical toxicity. *Science* **240**(4857): 1302-1309.

Ito, K., Kawakami, K. and Nakamura, Y. 1993. Multiple control of *Escherichia coli* lysyl-tRNA synthetase expression involves a transcriptional repressor and translational enhancer element*. Proc. Natl. Acad. Sci. U.S.A*. **90**: 302-306.

Ivanov, I.G., Alexandrova, R.A., Dragulev, B.P. and AbouHaidar, M.G. 1995. A second putative mRNA binding site on the *Escherichia coli* ribosome. *Gene* **160**: 75-79.

Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**(2): 283-292.

Kozak, M. 1987. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. Mol. Biol.* **196**(4): 947–950.

Kozak, M. 2005. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361**: 13-37.

Jacob, W.F., Santer, M. and Dahlberg, A.E. 1987. A single base change in the Shine-Dalgarno region of 16S rRNA of *Escherichia coli* affects translation of many proteins. *Proc. Natl. Acad. Sci U. S. A.* **84**: 4757-4761.

Jan, E. 2006. Divergent IRES elements in invertebrates. *Virus Res.* **119**: 16-28.

Jang, S.K., Krausslich, H.G., Nicklin, M.J., Duke, G.M., Palmenberg, A.C. and Wimmer, E. 1988. A segment of the 5' nontranslated region of encephalomyocarditis viral RNA directs internal entry of ribosomes during in vitro translation. *J. Virol.* **62**(8): 2636-2643.

Jin, H., Zhao, Q., Gonzalez de Valdivia, E.I., Ardell, D.H., Stenstrm, M. and Isaksson, L.A. 2006. Influences on gene expression in vivo by a Shine-Dalgarno sequence. *Mol. Microbiol*. **60**(2): 480-92.

Johansson, J., Mandin, P., Renzoni, A., Chiaruttini, C. Springer, M. and Cossart, P. 2002. An RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes. Cell* **110**(5): 551-61.

Kalapos, M.P., Paulus, H. and Sarkar, N. 1997. Identification of ribosomal preotein S1 as a poly(A) binding protein in *Escherichia coli. Biochimie*. **79**: 493-502.

Kieft, J.S. 2008. Viral IRES RNA structures and ribosome interactions. *Trends Biochem Sci.* **33**(6): 274-283.

Krishnan, K.M. 2010. Ribosome-mRNA interactions that contribute to recognition and binding of a 5'-terminal AUG start codon. Miami University, Oxford, Ohio.

Kudla, G., Murray, A.W., Tollervey, D. and Plotkin, J.B. 2009. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**: 255-258.

Kunkel, T.A., Roberts, J.D. and Zakour, R.A. 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Meth. Enzymol.* **154**: 367-382.

Kuroda, H. and Maliga. P. 2001. Complementarity of the 16S rRNA penultimate stem with sequences downstream of the AUG destabilizes the plastid mRNAs. *Nucleic Acids Res.* **29**: 970-975.

La Teana, A., Gualerzi, C.O. and Brimacombe, R. 1995. From stand-by to decoding site. Adjustment of the mRNA on the 30S ribosomal subunit under the influence of the initiation factors. *RNA* **1**: 772-782.

Laursen, B.S., Sorensen, H.P., Mortensen, K.K. and Sperling-Petersen, H.U. 2005. Initiation of protein synthesis in bacteria. *Microbiol. Mol. Biol. Rev.* **69:** 101-123.

Lawley, B., and Pittard, A.J. 1994. Regulation of *aroL* expression by TyrR protein and Trp repressor in *Escherichia coli* K-12. *J. Bacteriol*. **176**: 6921-6930.

Lease, R.A. and Belfort, M. 2000. Riboregulation by DsrA RNA:trans-actions for global economy. *Mol. Microbiol.* **38**: 667–672.

Liu, M.Y., Gui, G., Wei, B., Preston, J.F. III and Oakford, L. 1997. The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli. J. Biol. Chem.* **272**: 17502–10.

Loechel, S., Inamine, J.M. and Ping-chuan, H. 1991. Novel translation initiation region from *Mycoplasma genitalium* that functions in *Escherichia coli*. *Nucleic Acids Res.* **19:** 6905-6911.

Loh, E., Memarpou, F., Vaitkevicius, K., Kallipolitis, B.H., Johansson, J. and Sonden, B. 2012. An unstructured 5'-coding region of the prfA mRNA is required for efficient translation. *Nucleic Acids Res*. **40**: 1818-1827.

Looman, A.C. and Kuivenhoven, J.A. 1993. Influence of the three nucleotides upstream of the initiation codon on expression of the *Escherichia coli lacZ* gene in *Saccharomyces cerevisiae*. *Nucleic Acids Res*. **21**(18): 4268–4271.

Lopez, D., Vlamakis, H., Losick, R. and Kolter, R. 2009. Paracrine signaling in a bacterium. *Genes Dev.* **23**: 1631-1638.

MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H. and MacNeil, T. 1992. Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111**: 61-8.

Mandal, M. and Breaker, R.R. 2004. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **5**: 451-463.

Majdalani, N., Chen, S., Murrow, J., St John, K. and Gottesman, S. 2001. Regulation of RpoS by a novel small RNA: the characterization of RprA. *Mol. Microbiol.* **39**: 1382–94.

Martin-Farmer, J. and Janssen, G.R. 1999. A downstream CA repeat sequence increases translation from leadered and unleadered mRNA in *Escherichia coli. Mol. Microbiol.* **31**: 1025-1038.

Massé, E. and Gottesman, S. 2002. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli. Proc. Natl. Acad. Sci.* U.S.A. **99**: 4620-4625.

Massé, E., Escorcia, F.E. and Gottesman, S. 2003. Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli. Genes Dev*. **17**: 2374-2383.

Massé, E., Vanderpool, C.K. and Gottesman, S. 2005. Effect of RhyB small RNA on global iron use in *Escherichia coli. J. Bacteriol.* **187**(20): 6962-6971.

McCarthy, J.E. and Brimacombe, R. 1994. Prokaryotic translation: the interactive pathway leading to initiation. *Trends Genet.* **10**: 402-407.

McCarthy, J.E., Sebald, W., Gross, G. and Lammers, R. 1986. Enhancement of translational efficiency by the *Escherichia coli atpE* translational initiation region: its fusion with two human genes. *Gene* **41(**2-3): 201-6.

McGinness, K.E. and Sauer, R.T. 2004 Ribosomal protein S1 binds mRNA and tmRNA similarly but plays distinct roles in translation of these molecules. *Proc. Natl. Acad. Sci. U.S.A.* **101**: 13454-9.

Mellin, J.R., Goswami, S., Grogan, S., Tjaden, B. and Genco, C.A. 2007. A novel Fur- and Iron-regulated small RNA, NrrF, is required for indirect Fur-medicated regulation of the *sdhA* and *sdhC* genes in *Neisseria meningitidis*. *J. Bacteriol*. **189**: 3686–3694.

Millar, G., Lewendon, A., Hunter, M.G. and Coggins, J.R. 1986. The cloning and expression of the *aroL* gene from *Escherichia coli* K12. Purification and complete amino acid sequence of shikimate kinase II, the *aroL*-gene product. *Biochem J.* **237**: 427-437.

Miller, J. 1992. In A Short Course in Bacterial Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Mitta, M., Fang, L. and Inouye, M. 1997. Deletion analysis of *cspA* of *Escherichia coli:* requirement of the AT-rich UP element for *cspA* transcription and the downstream box in the coding region for its cold shock induction. *Mol. Microbiol.* **26:** 321-335.

Moll, I., Hirokawa, G., Kiel, M.C., Kaji, A. and Blasi, U. 2004. Translation initiation with 70S ribosomes: an alternative pathway for leaderless mRNAs. *Nucleic Acids Res.* **32:** 3354-3363.

Moller, T., Franch, T., Hojrup, P., Keene, D.R. and Bachinger, H.P. 2002. Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol. Cell* **9**: 23-30.

Myers, C.R. and Nealson, K.H. 1988a. Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor. *Science* **240**: 1319-1321.

Myers, C.R. and Nealson, K.H. 1988b. Microbial reduction of manganese oxides: interactions with iron and sulfur. *Geochimica et Cosmochimica Acta* **52**: 2727-2732.

Myers, C.R. and Nealson, K.H. 1990. Respiration-linked proton translocation coupled to anaerobic reduction of manganes(IV) and iron(III) in *Shewanella putrefaciens* MR-1. *J. Bacteriol*. **172**: 6232-6238.

Myers, C.R., and Myers, J.M. 1997. Replication of plasmids with the p15A origin in *Shewanella putrefaciens* MR-1. *Letters in Applied Microbiology* **24**: 221-225.

Nafissi, M., Chau, J., Xu, J. and Johnson, R. 2012. Robust translation of the nucleoid protein Fis requires a remote upstream AU element and is enhanced by RNA secondary structure. *J. Bacteriol*. 2458-2469.

Nagai, H., Yuzawa, H. and Yura, T. 1991. Interplay of two cis acting mRNA regions in translational control of sigma 32 synthesis during the heat shock response of *E.coli*. *Proc. Natl. Acad. Sci. U.S.A*. **88**: 10515-10519.

Nealson, K.H., Myers, C.R. and Wimper, B.B. 1991. Isolation and identification of manganese-reducing bacteria and estimates of microbial Mn(IV)-reducing potential in the Black Sea. *Deep-Sea Research* **38**(Suppl 2): S907-S920.

Nealson, K.H., Belz, A. and McKee, B. 2002. Breathing metals as a way of life: geobiology in action. *Antonie Leeuwenhoek* **81**: 215-222.

O'Connor, M., Asai, T., Squires, C. and Dahlbert, A. 1999. Enhancement of translation by the downstream box does not involve base pairing of mRNA with the penultimate stem sequence of 16S rRNA. *Proc. Natl. Acad. Sci. U.S.A*. **96**: 8973-8978.

O'Donnell, S. M., and Janssen, G.R. 2001. The initiation codon effects ribosome binding and translational efficiency in *Escherichia coli* of *c*I mRNA with or without the 5' untranslated leader. *J. Bacteriol.* **183**: 1277-1283.

Olins, P.O., Devine, S.C., Rangwala, S.H. and Kavka, K.S. 1988. The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in *Escherichia coli. Gene* **73**: 227-235.

Olins, P.O. and Rangwala, S.H. 1989. A novel sequence element derived from bacteriophage T7 mRNA acts as an enhancer of translation of the *lacZ* gene in *Escherichia coli. J. Biol. Chem*. **264**: 16973-16976.

Pelletier, J., and Sonenberg, N. 1988. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334**: 320-325.

Pisarev, A.V., Kolupaeva, V.G., Pisareva, V.P., Merrick, W.C., Hellen, C.U.T. and Pestova, T.V. 2006. Specific functional interactions of nucleotides at key −3 and +4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. *Genes Dev.* **20**(5): 624–636.

Ptashne, M., Backman, M.K., Humayum, M.Z., Jeffery, A., Maurer, R., Meyer, B. and Sauer, R.T. 1976. Autoregulation and function of a repressor in bacteriophage lambda. *Science* **194**: 156-161.

Ramamurthi, K.S., Lecuyer, S., Stone, H.A. and Losick, R. 2009. Geometric cue for protein localization in a bacterium. *Science* **323**: 1354-1357.

Rangan, L., Vogel, C. and Srivastava, A. 2008. Analysis of contect surrounding translation initiation site from complete genome model of plants. *Mol. Biotech.* **39(**3): 201-13.

Ringquist, S., Jones, T., Snyder, E.E., Gibson, T., Boni, I. and Gold, L. 1995. High-affinity RNA ligands to *Escherichia coli* ribosomes and ribosomal protein S1: comparison of natural and unnatural binding sites. *Biochemistry*. **34**: 3640-3648.

Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G.D. and Gold, L. 1992. Translation initiation in *Escherichia coli*: sequences within the ribosome binding site. *Mol. Microbiol.* **6**(9): 1219-1229.

Rinke-Appel, J., Junke, N., Stade, K. and Brimacombe, R. 1991. The path of mRNA through the *Escherichia coli* ribosome; site-directed cross-linking of mRNA analogues carrying a photoreactive label at various points 3' to the decoding site. *EMBO J.* **10**: 2195-2202.

Rinke-Appel, J., Junke, N., Brimacombe, R., Lavrik, I., Dokudovskaya, S., Dontsova, O. and Bogdanov, A. 1994. Contacts between 16S ribosomal RNA and mRNA, within the spacer region separating the AUG initiator codon and the Shine-Dalgarno sequence; a site-directed cross-linking study. *Nucleic Acids Res*. **22**: 3018-3025.

Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarranton, G. and Stephens, P. 1984. Codon usage can affect efficiency of translation of genes in *Escherichia coli. Nucleic Acids Res.* **12**: 6663-6671.

Rosenberg, M.I. and Desplan, C. 2010. Hiding in plain sight. *Science* **329**: 284-285.

Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual Second Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sasaki, J. and Nakashima, N. 2000. Methionine-independent initiation of translation in the capsid protein of an insect RNA virus*. Proc. Natl. Sci. U.S.A.* **97**: 1512-1515.

Sauter, C., Basquin, J. and Suck, D. 2003. Sm-like proteins in eubacteria: the crystal structure of the Hfq protein from *Escherichia coli. Nucleic Acids Res*. **31**: 4091-98.

Schleif, R. 1972. Fine-structure deletion map of the *Escherichia coli* L-arabinose operon. *Proc. Natl. Acad. Sci. U. S. A.* **69**: 3479-3484.

Schneider, D., Volkmer, T. and Rogner, M. 2007. PetG and PetN, but not PetL, are essential subunits of the cytochrome b_6f complex from Synechocystis PCC 6803. *Res. Microbiol.* **158**: 45-50.

Schneider, T.D., Stormo, G.D. and Gold, I. 1986. Information content of binding sites on nucleotide sequences*. J. Mol. Biology*. **188**: 415-431.

Schumacher, M.A., Pearson, R.F., Moller, T., valentine-Hansen, P. and Brennan, R.G. 2002. Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J*. **21**: 3546-56.

Sedlácek, J., Fábry, M., Rychlík, I., Volny, D. and Vitek, A. 1979. The arrangement of nucleotides in the coding regions of natural templates. *Mol. Gen. Genet.* **172**(1): 31-6.

Shean, C. S. and Gottesman, M.E. 1992. Translation of the prophage lambda *c*I transcript. *Cell* **70**: 513-22.

Shine, J. and Dalgarno, L. 1974. The 3'-Terminal Sequence of *Escherichia coli* 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proc. Nat. Acad. Sci. U.S.A.* **71**: 1342-1346.

Simonetti, A., Marzi, S., Myasnikov, A.G., Fabbretti, A., Yusupov, M., Gualerzi, C.O. and Klaholz, B.P. 2008. Structure of the 30S translation initiation complex. *Nature* **455**: 416-420.

Sleat, D.E., Gallie, D.R., Jefferson, R.A., Bevan, M.W., Turner, P.C. and Wilson, T.M.A. 1987. Characterization of the 5'-leader sequence of tobacco mosaic virus RNA as a general enhancer of translation in vitro. *Gene* **60:** 217-225.

Sledjeski, D.D., Whitman, C. and Zhang, A. 2001. Hfq is necessary for regulation by the untranslated RNA DsrA. *J. Bacteriol*. **183**: 1997-2005.

Sørensen, M.A., Fricke, J. and Pedersen, S. 1998. Ribosomal protein S1 is required for translation of most, if not all, natural mRNAs in *Escherichia coli* in vivo. *J. Mol. Biol.* 561-569.

Sprengart, M.L., Fatscher, H.P. and Fuchs, E. 1990. The initiation of translation in *E.coli*: apparent base pairing between the 16S rRNA and downstream sequences of the mRNA. *Nucleic Acids Res*. **18**: 1719-23.

Sprengart, M.L., Fuchs, E. and Porter, A.G. 1996. Downstream box: an efficient and independent translation initiation signal in *Escherichia coli. EMBO J.* **15**: 665-674.

Stenström, C.M., Holmgren, E. and Isaksson, L.A. 2001. Cooperative effets by the initiation codon and its flanking regions on translation initiation. *Gene* **273**: 259-265.

Sun, X., Zhulin, I. and Wartell, R.M. 2002. Predicted structure and phyletic distribution of the RNA-binding protein Hfq. *Nucleic Acids Res*. **30**: 3662-71.

Tedin, K., Resch, A. and Blasi, U. 1997. Requirements for ribosomal protein S1 for translation initiation of mRNAs with and without a 5' leader sequence. *Mol. Microbiol.* **25**: 189-199.

Thormann, K.M., Saville, R., Shukla, S. and Spormann, A.M. 2004. Initial phases of biofilm formation in *Shewanella oneidensis* MR-1. *J. Bacteriol.* **186**: 8096-8104.

Thormann, K.M., Saville, R., Shukla, S. and Spormann, A.M. 2005. Induction of rapid detachment in *Shewanella oneidensis* MR-1 biofilms. *J. Bacteriol.* **187**: 1014-1021.

Thormann, K.M., Duttler, S., Saville, R.M., Hydo, M., Shukla, S., Hayakawa, Y. and Spromann, A.M. 2006. Control of formation and cellular detachment from *Shewanella oneidensis* MR-1 biofilms by cyclic di-GMP. *J. Bacteriol.* **188**(7): 2681-2691.

Unoson, C. and Wagner, E.G. 2007. Dealing with stable structures at ribosome binding sites. Bacterial translation and ribosome standy. *RNA Biology* **4**: 113-117.

Unoson, C. and Wagner, E.G. 2008. A small SOS induced toxin is targeted against the inner membrane in *Escherichia coli*. *Mol. Microbiol.* **70**: 258-270.

Van Etten, W.J. and Janssen, G.R. 1998. An AUG initiation codon, not codon-anticodon complementarity, is required for the translation of unleadered mRNA in *Escherichia coli. Mol. Microbiol.* **27**: 987-1001.

Vecerek, B., Moll, I., Afonyushkin, T., Kaberdin, V. and Blasi, U. 2003. Interaction of the RNA chaperone Hfq with mRNAs: direct and indirect roles of Hfq in iron metabolism of *Escherichia coli*. *Mol. Microbiol.* **50**: 897-909.

Vellanoweth, R.L. and Rabinowitz, J.C. 1992. The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli* in vivo*. Mol. Microbiol*. **6**: 1105-14.

Wadler, C.S. and Vanderpool, C.K. 2007. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *Proc. Natl. Acad. Sci. U.S.A.* **1-4**: 20454-20459.

Wagner, E.G. Kill the messenger: bacterial antisense RNA promotes mRNA dacay. 2009. *Nat. Struct. Mol. Biol.* **16**: 804-806.

Wagner, E.G. and Simons, R.W. 1994. Antisense RNA control in bacteria, phages, and plasmids. *Annu. Rev. Microbiol.* **48**: 713–742.

Wade, H.E. and Robinson, H.K. 1966. Magnesium ion-independent ribonucleic acid depolymerases in bacteria. *Biochem. J.* **101**: 467-479.

Wan, X-F, VerBerkmoes, N., McCue, L., Stanek, D., Connelly, H., Hauser, L., Wu, L., Liu, X., Yan, T., Leaphart, A., Hettich, R., Zhou, J. and Thompson, D. 2004. Transcriptomic and proteomic characterization of the Fur modulon in the metal-reducing bacterium *Shewanella oneidensis. J. Bacteriol.* **186**(24): 8385-8400.

Wang, L.K. and Shuman, S. 2001. Domain structure and mutational analysis of T4 polynucleotide kinase. *J. Biol. Chem.* **276**: 26868-26874.

Ward, M.J., Fu, Q.S., Rhoads, K.R., Yeung, C.H., Spormann, A.M and Criddle, C.S. 2003. A derivative of the menaquinone precursor 1,4-dihydroxy-2napthoate is involved in the reductive transformation of carbon tetrachloride by aerobically grown *Shewanella oneidensis* MR-1. *Appl. Microbiol Biotechnol.* **63**: 571-577.

Wassarman, K.M. and Storz, G. 2000. 6S RNA regulates *E. coli* RNA polymerase activity. *Cell* **101**: 613–23.

Weiss, R.B., Dunn, D.M., Dahlberg, A.E., Atkins, J.F. and Gesteland, R.F. 1988. Reading frame switch caused by base-pair formation between the 3' end of the 16S rRNA and the mRNA during elongation of protein synthesis in *Escherichia coli. EMBO J.* **7**(5): 1503-7.

Wilderman, P.J., Sowa, N.A., FitzGerald, D.J., FitzGerald, P.C., Gottesman, S., Ochsner, U.A. and Vasil, M.L. 2004. Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. *Proc Natl Acad Sci* **101**: 9792–9797.

Winzeler, E. and Shapiro, L. 1997. Translation of the leaderless Caulobacter *dnaX* mRNA. J. Bacteriol. 179(12): 3981-3988.

Wu, C.J. and Janssen, G.R. 1996. Translation of vph mRNA in *Streptomyces lividans* and *Escherichia coli* after removal of the 5' untranslated leader. *Mol Microbiol.* **22**: 339-355.

Yamagishi, M., de Boer, H.A. and Nomura, M. 1987. Feedback regulation of rRNA synthesis. A mutational alteration in the anti-Shine-Dalgarno region of the 16S rRNA gene abolishes regulation. *J Mol Biol*. **198**: 547-50.

Zeiler, B.N. and Simon, R.W. 1996. Control by antisense RNA. In: Lin ECC, Lynch AS, editors. Regulation of gene expression in *Escherichia coli*. Austin, TX: R.G. Landers Company; pp. 67–83.

Zhang, A., Wassarman, K.M., Rosenow, C., Tjaden, B.C., Storz, G. and Gottesman, S. 2003. Global analysis of small RNA and mRNA targets of Hfq. *Mol. Microbiol*. **50**: 1111-24.

Zhang, J. and Deutzcher, M.P. 1992. A uridine-rich sequence required for translation of prokaryotic mRNA. *Proc Nat Acad Sci U.S.A*. **89**: 2605-9.

Zhang, Y.B., Ayalew, S. and Lacks, S. 1997. The *rnhB* gene encoding RNase HII of *Streptococcus pneumonia* and evidence of conserved motifs in eukaryotic genes. *J. Bacteriol.* **179**: 3828-3836.