

Insights into the Co-Evolution of Ribosomal Protein S15 with its Regulatory RNAs

Author: Betty L. Slinger

Persistent link: <http://hdl.handle.net/2345/bc-ir:106793>

This work is posted on [eScholarship@BC](#),
Boston College University Libraries.

Boston College Electronic Thesis or Dissertation, 2016

Copyright is held by the author, with all rights reserved, unless otherwise noted.

Boston College

The Graduate School of Arts and Sciences

Department of Biology

INSIGHTS INTO THE CO-EVOLUTION OF RIBOSOMAL PROTEIN S15
WITH ITS REGULATORY RNAS

a dissertation

by

BETTY L. SLINGER

submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

May 2016

© copyright by BETTY LOU SLINGER
2016

BOSTON COLLEGE
Graduate School of Arts & Sciences

The Dissertation of: Betty L. Slinger
(Student's Name)

Title: INSIGHTS INTO THE CO-EVOLUTION OF RIBOSOMAL PROTEIN
S15 WITH ITS REGULATORY RNAS

Submitted to the Department of: BIOLOGY

In partial fulfillment of the requirements for the degree of:

Doctor of Philosophy

In the Graduate School of Arts & Sciences has been read and approved by the
Committee:

Welkin Johnson
Chair

Thomas Chiles
Member

Eric Folker
Member

Charles Hoffman
Member

Michelle. M. Meyer
Advisor, Member

May 23, 2016
Date

ABSTRACT

INSIGHTS INTO THE CO-EVOLUTION OF RIBOSOMAL PROTEIN S15 WITH ITS REGULATORY RNAS

Betty L. Slinger

Thesis Advisor: Michelle M. Meyer

Ribosomes play a vital role in all cellular life translating the genetic code into functional proteins. This pivotal function is derived from its structure. The large and small subunits of the ribosome consist of 3 ribosomal RNA strands and over 50 individual ribosomal proteins that come together in a highly coordinated manner. There are striking differences between eukaryotic and prokaryotic ribosomes and many of the most potent antibacterial drugs target bacterial ribosomes (e.g. tetracycline and kanamycin). Bacteria spend a large amount of energy and nutrients on the production and maintenance of these molecular machines: during exponential growth as much as 40% of dry bacterial mass is ribosomes (Harvey 1970). Because of this, bacteria have evolved an elegant negative feedback mechanism for the regulation of their ribosomal proteins, known as autoregulation. When excess ribosomal protein is produced, unneeded for ribosome assembly, the protein binds a structured portion of its own mRNA transcript to prevent further expression of that operon. Autoregulation facilitates a quick response to changing environmental conditions and ensures economical use of nutrients.

My thesis has investigated the autoregulatory function of ribosomal protein S15 in diverse bacterial phyla. In many bacterial species, when there is excess S15 the protein interacts with an RNA structure formed in the 5'-UTR of its own mRNA transcript that

enables autoregulation of the S15-encoding operon, *rpsO*. For many ribosomal proteins (ex. L1, L20, S2) there is striking homology and often mimicry between the recognition motifs within the rRNA and the regulatory mRNA structure. However, this is not the case for S15-three different regulatory RNA structures have been previously described in *E. coli*, *G. stearothermophilus*, and *T. thermophilus* (Portier 1990, Scott 2001, Serganov 2003). These RNAs share little to no structural homology to one another, nor the rRNA, and they are narrowly distributed to their respective bacterial phyla, Gammaproteobacteria, Firmicutes, and Thermales. It is unknown which regulatory RNA structures control the expression of S15 outside of these phyla. Additionally, previous work has shown the S15 homolog from *G. stearothermophilus* is unable to regulate expression using the mRNA from *E. coli*. These observations formulate the crux of the question this thesis work endeavors to answer: What drove the evolution of such diverse regulatory RNA structures in these different bacteria?

In Chapter II, “Discover and Validate Novel Regulatory Structures for Ribosomal Protein S15 in Diverse Bacterial Phyla”, I present evidence for the *in silico* identification of three novel regulatory RNA structures for S15 and present experimental evidence that one of these novel structures is distinct from those previously described. In Chapter III, “Co-evolution of Ribosomal Protein S15 with Diverse Regulatory RNA Structures”, I present evidence that the amino acid differences in S15 homologs contribute to differences in mRNA binding profiles, and likely lead to the development of the structurally diverse array of the regulatory RNAs we observe in diverse bacterial phyla. In Chapter IV, “Synthetic *cis*-regulatory RNAs for Ribosomal Protein S15”, I investigate the derivation of novel *cis*-regulatory RNAs for S15 and find novel structures are readily-

derived, yet interact with the rRNA-binding face of S15. Together the work presented in this thesis advances our understanding of the co-evolution between ribosomal protein S15 and its regulatory RNAs in diverse bacterial phyla.

DEDICATION

To my Husband, Spencer,
For Everything, Forever!

ACKNOWLEDGEMENTS

There are a number of people without whom this dissertation would not be possible and to whom I am greatly indebted.

I would like to express deepest gratitude to Michelle M. Meyer, my mentor and PhD thesis advisor, who expertly guided me through my six years of graduate education at Boston College and who shared the excitement of five years of discovery. It was through her constant support, generous advice, and incredible patience that performing all of my experimental work, as well as publishing it and presenting it at scientific conferences was possible.

My appreciation also extends to my laboratory colleagues, especially Shermin Pei and Arianne Babina, for your constant encouragement, support and making our lab a positive, good-humored, and intellectually-stimulating environment to perform research.

Thank you to my thesis committee members, Tom Chiles, Welkin Johnson, Eric Folker, and Charlie Hoffman for your invaluable advice and guidance.

Thank you to my fellow graduate students and friends who make this journey through grad school about the journey, and not the destination. Thanks for all the Thirsty Thursdays, running 5K's with me for the beer tent, and all the game nights!

I am indebted to my parents, Ron and Elsie Slinger, and sisters, Mary and Laura, for their unequivocal support and love.

And finally, I acknowledge my husband, Spencer, who is my champion and who has blessed me with a life of joy and love in the hours when lab lights were off.

TABLE OF CONTENTS

ABSTRACT	i
DEDICATION	iv
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	xi
DATASETS	xiv

CHAPTER I

The Role of Ribosomal Protein S15 in the Prokaryotic Ribosome	2
<i>Cis</i> -acting RNA Elements Regulate Gene Expression in Bacteria	6
Evolution of RNA-Protein Regulatory Interactions	10
The S15-mRNA Interaction in Bacteria	14
FIGURES & LEGENDS	19
Figure 1.1 Complexity of The Prokaryotic Ribosome	19
Figure 1.2 S15-rRNA Binding Interaction	20
Figure 1.3 Conservation of Residues and Nucleotides Involved in S15-rRNA recognition	21
Figure 1.4 Autoregulation of R-Protein S15	22
Figure 1.5 Phylogenetic Analysis of Regulatory RNAs	23
Figure 1.6 <i>rpsO</i> Regulatory RNAs	24

CHAPTER II

INTRODUCTION	26
RESULTS & DISCUSSION	27
Non-coding Regulatory RNA Discovery Using Comparative Genomics	27
RNAs Identified Are Diverse in Sequence and Secondary Structure	29
RNA from Alphaproteobacterium <i>Rhizobium radiobacter</i> Specifically Interacts with S15 Protein	32
Truncation Analysis Suggests that Rra-S15 Minimal Binding Site Includes Both H1 and H2	33
Mutation Analysis Suggests a Potential Binding-site	35
The <i>R. radiobacter</i> RNA Allows Regulation in Response to S15 <i>in vivo</i>	35
Structural Probing Confirms Predicted Secondary Structure of Rra-RNA	37
CONCLUSION	40

MATERIALS & METHODS	42
FIGURES, & LEGENDS	48
Figure 2.1 Overview of Comparative Genomic Pipeline: Genomic Analysis for Illuminating Structured RNA.....	48
Figure 2.2 Consensus Diagrams of Novel Putative RNA Structures and Individual Examples Used for Transcriptomic Analysis	49
Figure 2.3 Transcriptomic Analysis of Alphaproteobacterial and Actinobacterial RNAs.....	50
Figure 2.4 Nitrocellulose Binding Assays Identify Regions Important for Rra-S15 Binding.....	51
Figure 2.5 5'-RACE for <i>rpsO</i> Transcription Start Site in <i>R. radiobacter</i>	52
Figure 2.6 <i>In vivo</i> Regulation Assays Validate Rra-RNA Regulates Gene Expression	53
Figure 2.7 Footprinting Assays on Rra-RNA	54
Figure 2.8 Updated Phylogenetic Tree for Regulatory RNAs of Ribosomal Protein S15	55
Additional File 1: Alignment of Alphaproteobacterial RNA	56
Additional File 2: Alignment of Actinobacterial RNA.....	56
Additional File 3: Alignment of Chlamydia RNA.....	56
<u>CHAPTER III</u>	
INTRODUCTION	58
RESULTS AND DISCUSSION.....	59
<i>LacZ in vivo</i> Reporter Assay Validates mRNA-S15 Regulatory Interactions	59
Regulation Assays Confirm Native mRNA-S15 Interactions	63
Regulation Assays Reveal Specific RNA Recognition Patterns.....	65
<i>In vitro</i> Binding Assays Show Distinct Recognition Profiles for S15 Homologs	67
S15 Homologs Recognize mRNAs Via Distinct Motifs.....	70
S15 Homologs Interacting with Non-homologous mRNA Regulators Have Different Conservation Patterns	75
Mutated <i>G. kaustophilus</i> S15 Shows Altered Specificity.....	77
CONCLUSION.....	78
MATERIALS & METHODS	80
FIGURES, & LEGENDS	86
Figure 3.1 <i>E. coli</i> $\Delta rpsO$ Confirmation	86
Figure 3.2 pBS3-RNA Plasmid Diagram.....	87

Figure 3.3 Background β -galactosidase Expression of the <i>E. coli</i> <i>ArpsO</i> Strain	88
Figure 3.4 Native mRNA-S15 Regulatory RNA-S15 Interactions.....	89
Figure 3.5 All + and – L-arabinose Miller Units	90
Figure 3.6 Non-native Regulatory RNA-S15 Interactions	91
Figure 3.7 Table of Inter-Species Nitrocellulose Binding Assays.....	92
Figure 3.8 Cartoon Representation of S15-mRNA Binding Sites	93
Figure 3.9 Mutation to Gk-mRNA-M2 Strengthens Model of Interaction.....	94
Figure 3.10 S15 Protein Conservation in Gammaproteobacteria and Firmicutes	95
Figure 3.11 Chimeric <i>G. kaustophilus</i> S15 with <i>E. coli</i> S15	96
Figure 3.12 Sequences and Primers	97
Figure 3.13 Statistics for Data: Native and Non-Native Interactions	102
Figure 3.14 Statistics for Gk-S15 Mutants Assays	104

CHAPTER IV

INTRODUCTION	107
RESULTS	108
<i>In vitro</i> Selection of RNA Aptamers for Gk-S15	108
Four Synthetic RNAs Regulate Gene Expression with Gk-S15.....	110
RNAs Regulate Gene Expression with S15 Homolog from <i>T. thermophilus</i> but not <i>E. coli</i>	111
Elucidation of the Gk-S15 Binding Face	112
Footprinting Experiments Elucidate Nucleotides Important in 11-1 Binding Gk-S15	113
Mutagenesis Experiments Confirm Gk-S15 Binding Regions in 11-1.....	115
Footprinting Experiments of Non-regulatory RNA 11-2	117
DISCUSSION	118
MATERIALS & METHODS	121
FIGURES & LEGENDS.....	125
Figure 4.1 SELEX Overview	125
Figure 4.2 Individual RNAs Isolated from Round 11 Have Diverse Sequence and Predicted Structure.....	126
Figure 4.3 RNAs from Round 11 Bind Gk-S15	127
Figure 4.4 <i>In vivo</i> Regulation Assays for Synthetic RNAs	128
Figure 4.5 All Miller Units, + and - L-arabinose.....	129
Figure 4.6 RNA 11-1 Interacts with Same Face of Gk-S15 as Gk-mRNA	130

Figure 4.7 RNAs 11-1 and 11-2 bind Tt-S15, not Ec-S15.....	131
Figure 4.8 Structure Probing Elucidates the Secondary Structure of RNA 11-1....	132
Figure 4.9 <i>In vitro</i> Binding Assays and Mutagenesis Confirms the Predicted RNA 11-1 Structure.....	133
Figure 4.10 All RNA 11-1 Mutant Binding Assays with Gk-S15.....	134
Figure 4.11 Binding Curves for All RNA 11-1 Mutants with Gk-S15.....	135
Figure 4.12 <i>In vitro</i> Mutagenesis Assays with Gk-S15 Plotted to an Alternative Potential Secondary Structure for RNA 11-1	136
Figure 4.13 Footprinting Data Mapped to the Alternative Structure for RNA 11-1	137
Figure 4.14 Structure Probing Assays Elucidate the Secondary Structure of RNA 11-2 in the Presence and Absence of Gk-S15	138
Figure 4.15 Binding curves for Gk-S15 binding assays with RNA 11-2-M1	139
Figure 4.15 Table of Primers and Reference Sequences	140
Figure 4.16 Statistics for Data on Figure 4.4	143

CHAPTER V

A Summary: Ribosomal Protein S15 Co-evolved with its Regulatory RNAs.....	145
Evolution of Distinct Regulatory RNAs for Ribosomal Protein S15	148
Ribosomal Protein Biosynthesis as a Novel Antibacterial Target	153
A Broader Context: Many Structural Solutions to this Biological Problem.....	156

APPENDIX

CHAPTER AI

INTRODUCTION	163
RESULTS	164
RNAiFold Determines All RNA Sequences for the Target Structure, PLMVd Hammerhead Ribozyme.....	164
Hammerhead Candidate Sequences Self-Cleave	165
Rate of Cleavage Differs for the 10 Computationally Designed Hammerheads	166
Designed Hammerhead Functions Within a Larger Rationally Designed RNA	167
DISCUSSION	169
MATERIALS & METHODS	172
Computational Methods.....	172
Design of Modular Hammerhead Within Another Structure.....	173

Experimental Validation	174
Kinetics	176
T1-RNase Cleavage	176
TABLES, FIGURES & LEGENDS	178
Figure A1.1 RNAiFold Sequence Target PLMVd Hammerhead Ribozyme	178
Figure A1.2 Table of HH Sequences	179
Figure A1.3 Cleavage Assays of 10 Hammerhead Ribozymes	180
Figure A1.4 Table of Kinetics	181
Figure A1.5 HH1 and HH7 Cleavage Time Series.....	182
Figure A1.6 Cleavage Time Series for HH2-6, and HH8-10.	183
Figure A1.7 Modular Placement of XPT-Riboswitch with Hammerhead Ribozyme	184
Figure A1.8 Cleavage Assay and Kinetics Data for Modular RNA containing XPT- Riboswitch and Hammerhead.....	185
Figure A1.9 Structure Probing Confirms Modular RNA Cleaves at Expected Site	186
ABBREVIATIONS & DEFINITIONS	187
REFERENCES	188

LIST OF FIGURES

Figure 1.1 Complexity of The Prokaryotic Ribosome

Figure 1.2 S15-rRNA Binding Interaction

Figure 1.3 Conservation of Residues and Nucleotides Involved in S15-rRNA recognition

Figure 1.4 Autoregulation of R-Protein S15

Figure 1.5 Phylogenetic Analysis of Regulatory RNAs

Figure 1.6 *rpsO* Regulatory RNAs

Figure 2.1 Overview of Comparative Genomic Pipeline: Genomic Analysis for Illuminating Structured RNA

Figure 2.2 Consensus Diagrams of Novel Putative RNA Structures and Individual Examples Used for Transcriptomic Analysis

Figure 2.3 Transcriptomic analysis of Alphaproteobacterial and Actinobacterial RNAs

Figure 2.4 Nitrocellulose Binding Assays Identify Regions Important for Rra-S15 Binding

Figure 2.5 5'-RACE for *rpsO* Transcription Start Site in *R. radiobacter*

Figure 2.6 *In vivo* Regulation Assays Validate Rra-RNA Regulates Gene Expression

Figure 2.7 Footprinting Assays on Rra-RNA

Figure 2.8 Updated Phylogenetic Tree for Regulatory RNAs of Ribosomal Protein S15

Figure 3.1 *E. coli* $\Delta rpsO$ Confirmation

Figure 3.2 pBS3-RNA Plasmid Diagram

Figure 3.3 Background β -galactosidase Expression of the *E. coli* $\Delta rpsO$ Strain

Figure 3.4 Native mRNA-S15 Regulatory RNA-S15 Interactions

Figure 3.5 All + and – L-arabinose Miller Units

Figure 3.6 Non-native Regulatory RNA-S15 Interactions

Figure 3.7 Table of Inter-Species Nitrocellulose Binding Assays

Figure 3.8 Cartoon Representation of S15-mRNA Binding Sites

Figure 3.9 Mutation to Gk-mRNA-M2 Strengthens Model of Interaction

Figure 3.10 S15 Protein Conservation in Gammaproteobacteria and Firmicutes

Figure 3.11 Chimeric *G. kaustophilus* S15 with *E. coli* S15

Figure 3.12 Sequences and Primers

Figure 3.13 Statistics for Data: Native and Non-Native Interactions

Figure 3.14 Statistics for Gk-S15 Mutants Assays

Figure 4.1 SELEX Overview

Figure 4.2 Individual RNAs Isolated from Round 11 Have Diverse Sequence and Predicted Structure

Figure 4.3 RNAs from Round 11 Bind Gk-S15

Figure 4.4 *In vivo* Regulation Assays for Synthetic RNAs

Figure 4.5 All Miller Units, + and - L-arabinose

Figure 4.6 RNA 11-1 Interacts with Same Face of Gk-S15 as Gk-mRNA

Figure 4.7 RNAs 11-1 and 11-2 bind Tt-S15, not Ec-S15

Figure 4.8 Structure Probing Elucidates the Secondary Structure of RNA 11-1

Figure 4.9 *In vitro* Binding Assays and Mutagenesis Confirms the Predicted RNA 11-1 Structure

Figure 4.10 All RNA 11-1 Mutant Binding Assays with Gk-S15

Figure 4.11 Binding Curves for All RNA 11-1 Mutants with Gk-S15

Figure 4.12 *In vitro* Mutagenesis Assays with Gk-S15 Plotted to an Alternative Potential Secondary Structure for RNA 11-1

Figure 4.13 Footprinting Data Mapped to the Alternative Structure for RNA 11-1

Figure 4.14 Structure Probing Assays Elucidate the Secondary Structure of RNA 11-2 in the Presence and Absence of Gk-S15

Figure 4.15 Binding curves for Gk-S15 binding assays with RNA 11-2-M1

Figure 4.15 Table of Primers and Reference Sequences

Figure 4.16 Statistics for Data on Figure 4.4

Figure A1.1 RNAiFold Sequence Target PLMVd Hammerhead Ribozyme

Figure A1.2 Table of HH Sequences

Figure A1.3 Cleavage Assays of 10 Hammerhead Ribozymes

Figure A1.4 Table of Kinetics

Figure A1.5 HH1 and HH7 Cleavage Time Series

Figure A1.6 Cleavage Time Series for HH2-6, and HH8-10

Figure A1.7 Modular Placement of XPT-Riboswitch with Hammerhead Ribozyme

Figure A1.8 Cleavage Assay and Kinetics Data for Modular RNA containing XPT-Riboswitch and Hammerhead

Figure A1.9 Structure Probing Confirms Modular RNA Cleaves at Expected Site

DATASETS

Additional File 1: Alignment of Alphaproteobacterial RNA

Additional File 2: Alignment of Actinobacterial RNA

Additional File 3: Alignment of Chlamydia RNA

CHAPTER I

Introduction

The Role of Ribosomal Protein S15 in the Prokaryotic Ribosome

Ribosomes play a vital role in all living organisms translating the genetic code into functional proteins. All domains of life contain this ribonucleoprotein complex, so the last universal common ancestor almost certainly contained an ancient form of the ribosome (Root-Bernstein 2015, Fox 2010). It is astounding what this one complex, and more specifically the RNA within this complex, has accomplished in shaping life as we know it for over a billion years. Ribosomes have catalyzed every peptidyl transferase reaction in every naturally-derived protein (Nissen 2000). Put simply, life would not be possible in any domain on Earth without ribosomes.

There are differences between the overall make-up of eukaryotic and prokaryotic ribosomes (Klinge 2012). This makes prokaryotic ribosomes an excellent target for antibacterial pharmaceuticals. Many of our most potent classes of antibacterials target bacterial ribosomes, with little effect to those of an infected eukaryotic host (ex. tetracycline and kanamycin). Because the ribosome is such an important biological complex, and because prokaryotic ribosomes are of the best targets for novel antibacterial pharmaceuticals, it is of vital importance to understand which structural components comprise prokaryotic ribosomes, and to understand how bacteria regulate the production of these components.

The pivotal function of the prokaryotic ribosome is derived from its structure, which was first described at atomic resolution in 2000 (for review see Schmeing 2009). Three ribosomal RNA strands and over 50 individual ribosomal proteins (r-proteins) come together in a highly coordinated manner, forming the small (30S) ribosomal subunit, and the large (50S) ribosomal subunit, (Figure 1.1). The structure of the 30S

subunit, which reads the mRNA message, was first determined from the bacterium *Thermus thermophilus* (Figure 1.1A (top), Wimberly 2000). The active site where peptide bond formation occurs, the structure of the 50S subunit was first described in atomic detail in the bacterium *Deinococcus radiodurans* (Harms 2001) (Figure 1.1A (bottom)). The importance of these discoveries was recognized in 2009 by the awarding of the Nobel Prize in Chemistry. Additional crystal structures of a ribosome complexed with tRNAs and mRNAs provided additional details of the translation process (Schmeing 2009). Of note, the ribosome does not contain any r-proteins within close proximity to the active site where peptidyl transfer occurs (Korostelev 2006). These observations and subsequent studies indicate the ribosome is a ribozyme (i.e. RNA enzyme, Nissen 2000), the rRNA itself is responsible for polypeptide chain creation and the r-proteins solely scaffold the rRNA (Figure 1.1A, B). These proteins are generally small (~ 3 – 50 kDa, Zengel & Lindahl 1994), relatively unstructured or alpha helical, and positively charged, allowing them to associate with the negatively charged rRNA sugar-phosphate backbone (Figure 1.1C). The result is a ribonucleoprotein complex made up of hundreds of highly specific RNA-protein interactions that collectively assemble into ribosomes who translate mRNA messages into nascent polypeptide chains in all bacteria.

Through extensive *in vitro* characterization studies r-protein S15's function within the bacterial ribosome has been elucidated. The primary role for S15 is to stabilize the 16S rRNA during assembly of the small ribosomal subunit (Sykes 2010, Mulder 2010). It is one of the first proteins recruited and its binding of the 16S rRNA triggers a cascade of structural rearrangements of the rRNA, enabling the recruitment of additional r-proteins (Held 1974, Jagannathan 2003). Isothermal titration calorimetry studies indicate that in

the absence of S15 the S6:S18 dimer do not bind rRNA. This suggests that formation S15-16S rRNA complex directly affects the ability of S6, S18, S11 and S21 to bind 16S rRNA (Recht 2001). Additionally, the crystal structure of the bacterial ribosome shows S15 helps bridge the 30S ribosomal subunit at its interface with the 50S (Yusupov 2001).

In vivo ribosome characterization studies have further determined the function of S15 in a cellular context. An in-frame chromosomal deletion of *rpsO* yields a viable, yet slow-growing *E. coli* strain even at the permissive temperature (37°C) (Mathy 2004). This knockout organism is cold-sensitive, a typical characteristic of bacterial strains with ribosomal assembly defects (Guthrie 1969, Dammel 1993). However, r-proteins S6 and S18 were found in ribosomes isolated from this strain (Bubunenko 2006). Therefore, functional ribosomes must be assembling, albeit at a slower rate. A ribosome profiling experiment of the $\Delta rpsO$ strain showed few 70S ribosomes or polysomes were present, even under permissive growth conditions (Bubunenko 2006). Because the strain is viable, and because S15 is known to be a subunit association protein, it suggests that there is decreased stability of the interaction between the small and large ribosomal subunits in the absence of S15. All of these findings have elucidated that S15's primary function is to bind 16S rRNA that not only enables the binding of additional r-proteins, but also enables the bridging of the two ribosomal subunits themselves.

The crystal structure, as well as numerous mutagenesis studies, have elucidated the S15-recognition site in 16S rRNA (Figure 1.2A & B, Batey 1996a, Batey 1996b, Mougél 1988, Powers 1995, Nikulin 2000). The rRNA contains a bipartite S15 recognition site that is formed where helices H20, H21, and H22 come together forming a three helix junction (3HJ, Serganov 1996, Batey 1996a, Batey 1996b, Nikulin 2000).

These studies indicate that the first and primary S15-recognition site is within the 3HJ at the base of H22 where a base triple GGC resides. Distal to this base triple, the base of H22 provides a shallow groove that also provides contacts for S15-binding, though these are not nucleotide-specific (Serganov 1996)). The secondary S15-recognition site is a G•U/G-C motif located ~1 helical turn, distal to the 3HJ in H22 (Benard 1998). Mutations to the G•U/G-C motif were more tolerated, suggesting this is a secondary stabilizing motif for S15 (Serganov 2001). Thus, S15 primarily recognizes the 3HJ of the rRNA, and then S15 binds the G•U/G-C motif, which stabilizes the correct conformation of the rRNA for subsequent binding by S6, S18, and, ultimately, to the formation of the small ribosomal subunit.

Using the crystal structure generated from the S15-rRNA interaction in *T. thermophilus*, the amino acids within S15 that interact with rRNA were elucidated (Figure 1.2C, Nikulin 2000). S15's overall structure is comprised of 4 alpha helical chains, with three loop regions, all of which fold together into a 12 kDa globular protein. Specific residues residing in two distinct regions on one face of the protein recognize the 3HJ and G•U/G-C motifs within the rRNA. The 3HJ of rRNA binds residues in both the loop 1 and C-terminal part of alpha helix 3. Residues T21, G22, Q27, Y68, and R71 were shown to make direct contacts with the 3HJ, while residues K4, K7, Q8, D20, T24, R34, R64, and E72 interact with the region surrounding the 3HJ of the rRNA. Residues that contact the G•U/G-C region of rRNA are located in the loop 2 region of S15. Residues H41 and D48 directly contact the G-C base pair, and S51 directly interacts with G-U non-canonical base pair via a water molecule. The binding of S15 to the 3HJ, followed by

binding to the G•U/G-C motif, ensures proper rRNA conformation for additional r-protein recruitment.

This S15-rRNA binding interaction is essential to bacterial viability, and not surprisingly this interaction is conserved across bacterial phyla to ensure proper ribosome assembly and function (Figure 1.3). A comparative analysis of this rRNA region indicates that the nucleotides comprising the 3HJ and the G•U/G-C are conserved, as is the number of base pairs separating the two recognition motifs (Figure 1.3A). This is further evidence that the motifs and the distance separating them are important for S15 recognition. Additionally, regions of the S15 amino acid sequence that interact with these motifs are conserved across bacteria (Figure 1.3B-D). The conservation of the players within S15 and the region of rRNA it binds underlines the importance of this interaction to prokaryotic ribosome function.

***Cis*-acting RNA Elements Regulate Gene Expression in Bacteria**

RNA-based regulation is a common way life forms monitor their environment and respond quickly to changes therein by altering their gene expression. Bacteria have invented a variety of ways utilizing RNA to modulate transcription, translation, mRNA stability, DNA maintenance and silencing (for review see Waters 2009). One of the most common and widespread methods is the use of *cis*-acting RNA elements, often called riboswitches, which modulate gene expression in bacteria without the need for protein co-factors. These RNAs are highly structured and usually localized to the untranslated regions of bacterial transcripts where they bind an effector ligand to regulate gene

expression of the operon in which they reside. The number of biological processes these RNAs control and variety of the effector ligands with which they interact continues to grow. However, the general mechanism of action is the same for all. RNA-ligand binding results in global structural rearrangements to the RNA transcript so as to mask or reveal important downstream transcription or translation elements of the mRNA.

The widespread nature of prokaryotic *cis*-regulatory RNAs became clear in the early 2000s through systematic computational searches using sequenced microbial genomes (Livny 2007). An ever increasing number of RNAs continue to be identified in bacteria (Charpentier 2015), as well as some eukaryotes (Yadav 2015). Many of these RNAs interact with metabolite ligands to control the expression of proteins involved in fundamental metabolic processes. Metabolite-responsive riboswitches for adenosylcobalamin (AdoCbl, Lundrigan 1991, Schaffer 2014), thiamin pyrophosphate (TPP, Croft 2007), lysine (Garst 2008), glycine (Ruff 2014), flavin mononucleotide (FMN, Wickiser 2005b), guanine (Batey 2004), adenine (Mandal 2003), glucosamine-6-phosphate (GlcN6P, Winkler 2004), 7-aminoethyl 7-deazaguanine (preQ₁, Roth 2007, Meyer 2008), and *S*-adenosylmethionine (SAM, Winkler 2003) have been described. The term riboswitch has grown to encompass those RNAs that change gene expression in response to metal ions (Cromie 2006, Dann 2007), and temperature (Klinkert 2009), tRNAs (T-Box, Gutierrez-Preciado 2009) The number of ligands that interact with riboswitches continues to grow and underscores the importance of RNA-based regulation in the control of many biological processes in bacteria.

Bacteria control gene expression using a variety of riboswitch-mediated mechanisms. One of the most common mechanisms is transcription termination (Barrick

2007, Breaker 2012), for example, the FMN riboswitch in the *fibD* operon of *B. subtilis* (Wickiser 2005). The formation of a strong stem followed by a run of uridine residues establishes an intrinsic transcription terminator that causes RNA polymerase to stall transcription and eventually to release the DNA template and nascent RNA product (Gusarov 1999, Yarnell 1999). There are examples of riboswitches that control translation initiation, for example, the preQ₁ riboswitch (Eichhorn 2014). Mutually-exclusive base-paired structures are exploited by riboswitches to control ribosome access to the ribosome binding site (RBS) or the Shine-Dalgarno (SD) sequence, thereby regulating translation initiation. Another interesting way bacteria use riboswitches to control gene expression couples ligand-binding to ribozyme cleavage activity. For example, the GlcN6P ribozyme (Brooks 2009) and the hammerhead ribozyme (Pan 2003).

A simple riboswitch composed of a single regulatory structure will respond only to its target ligand (or a close chemical analog). Several things can affect a single riboswitch's ability to regulate including RNA folding kinetics (slower folder may regulate slower), and binding affinity (poor affinity leads to poor riboswitching). Modern organisms have found several ways to overcome performance limitations of metabolite-sensing RNAs. Tandem-arranged riboswitch configurations expand the functional capability of *cis*-regulatory RNAs. In *Vibrio cholera*, the glycine riboswitch has two aptamers with a single expression platform that function cooperatively. Cooperative binding narrows the ligand-sensing dynamic range (Mandal 2004). In *Bacillus anthracis*, two TPP riboswitches (Sudarsan 2006, Welz 2007) act independently yet respond to the same ligand (TPP), which enables tighter regulatory control. In *Candidatus pelagibacter*

ubique (Poiata 2009) a SAM-II riboswitch exists in tandem with a SAM-V riboswitch. They respond to the same ligand (*S*-adenosylmethionine), yet act independently, yielding a tighter control over gene expression. Finally, there are cases where tandem riboswitches contain different RNA architectures and respond to completely different ligands. In *Bacillus clausii*, the *metE* mRNA contains a SAM-I riboswitch followed by a riboswitch that responds to AdoCbl (Sudarsan 2006). Each act independently and is associated with its own intrinsic terminator stem; therefore, binding of either riboswitch to its ligand results in termination of transcription for the entire operon. These examples highlight the incredible complexity of function achieved by RNA-based *cis*-regulatory devices in bacteria.

Cis-acting regulatory RNAs can also regulate the expression of operons by interacting with protein products of that operon in a *cis*-regulatory mechanism that mirrors that of riboswitches (Gelfand 2005). Once enough of a specific protein is produced, the excess binds the 5' untranslated leader region of its mRNA. Binding induces structural changes to the mRNA transcript so as to compete with ribosome binding or stall translation initiation. Generally, one RNA-protein interaction results in gene regulation; however, these RNAs are generally not classified as riboswitches because of their interaction with a protein. Instead this type of RNA-based regulation is called autogenous regulation or autoregulation, and the RNAs referred to as *cis*-regulatory RNAs.

Bacteria often control the expression of their ribosomal proteins in this manner, which allows them to maintain the correct stoichiometric amounts of ribosomal components. The process is best described in the model organism, *E. coli* (Figure 1.4,

Zengel 1994). More than half of the genes encoding ribosomal proteins in *E. coli* (r-proteins) are localized to twelve operons and the expression of the gene products from these operons is controlled by specific autoregulatory RNAs. (Lindahl 1986, Fu 2013). This allows a microbe to respond quickly to changing environmental conditions and ensures economical use of nutrients. RNA's central role in the transcription and translation process, coupled with the near limitless structures into which an RNA can fold make it ideal for autoregulatory purposes. While most of the r-protein autoregulatory RNAs in *E. coli* have been described, it is now important to go beyond this model organism and characterize this interaction.

Evolution of RNA-Protein Regulatory Interactions

RNA is now appreciated to participate in almost all aspects of biology, beyond carrying the genetic message from DNA to be translated into proteins. Over the last two decades numerous regulatory RNA have been discovered that range in function from directing development in eukaryotes (Amaral 2008), to controlling bacterial virulence (Johansson 2003) and metabolism (Dambach 2009). The diversity of biological functions RNA performs as well as the fact that RNA catalyzes reactions as a ribozyme strongly suggests that RNA has played a central role in controlling cellular processes since the beginning of life.

The characteristics of modern riboswitches suggest *cis*-regulatory RNAs could be descendants of an ancient sensory system. RNA uses the same four types of monomers to form its selective ligand-binding pockets, those four monomers are found in all forms of

life. Members of all experimentally validated riboswitches bind their ligands without the need for protein co-factors, suggesting riboswitches evolved before rudimentary translational machinery evolved, and use similar RNA-based sensors to control expression. There are riboswitches that sense compounds thought to be relics from an RNA World, the SAM and preQ₁ riboswitches (White 1976, Benner 1989).

Because only four types of nucleic acids are used by RNA, the RNA sequences and structures can be strikingly well-conserved over great evolutionary distances (Grundy 1998, Sudarsan 2003, Nahvi 2004). The widespread phylogenetic distribution and structural conservation of some riboswitch classes, AdoCbl, FMN and TPP suggest ancient beginnings and indicate these structures may have a common ancestor. For example, the TPP riboswitch is found in eukaryotes, such as plants and fungi (Yadav 2015), in addition to bacteria (Sudarsan 2006). Many widespread riboswitch classes contain complex tertiary architectures that are unlikely to have emerged independently during evolution.

RNA structures that are found to be more narrowly distributed, require smaller sequence space, or have less complex secondary structures may represent more modern inventions. Tracking the evolutionary trajectory and ultimately classifying these RNAs remains challenging. For example, riboswitches interacting with purines (Kim 2007) or preQ₁ (Roth 2007) may have emerged as recent inventions that perform novel regulatory functions, or may be reinventions of a conserved binding pocket that are presented with novel tertiary architecture.

To add to the complexity of describing the evolutionary trajectory of these RNAs, for many RNAs it appears that there are different secondary or tertiary structures that

accomplish very similar or identical biological functions in different bacterial phyla. A classic example of the structural diversity that has arisen across different bacterial phyla are the many distinct riboswitch classes that bind the small molecule *S*-adenosylmethionine (SAM) (Valley 2012, Burge 2012, Lindgreen 2014, Dambach 2009, Smith 2014). From structural data it is clear that at least three of these RNAs interact with their ligand (SAM) in fundamentally different ways, suggesting completely independent derivation (Wang 2008, Corbino 2005, Fuchs 2006). Furthermore, this example is far from unique. Two distinct riboswitch classes interact with the second messenger *c*-di-GMP (Sudarsan 2008), and three such classes with the nucleoside prequeosine-1 (Roth 2007, Meyer 2008, McCown 2014).

The existence of multiple unique RNA architectures responsible for analogous biological functions is not limited to RNA-small molecule interactions. This phenomenon is also apparent for regulatory RNAs interacting with protein partners. Multiple mRNA regulatory structures have been identified that perform autoregulation in response to ribosomal proteins L20, S4, and S15 (Ban 2014, Fu 2013, Deiorio-Haggar 2013, Chapter II, Slinger 2014). From even this small set of RNA-protein interactions, we see that distinct RNA architectures in different bacterial phyla can successfully perform analogous biological functions by interacting with homologous protein binding partners. In some cases, there is obvious similarity between the mRNA and rRNA binding-sites, suggesting that the protein recognizes the same tertiary structural features (Choonee 2007, Guillier 2005). However, there are several examples where this similarity is not obvious (Scott 2001, Grundy 1992, Tang 1989, Phillippe 1993). In such cases, it remains unclear how much of the mRNA structural diversity observed is due to independent

derivation of the similar tertiary structure, or if differences between homologous protein partners lead to distinct RNA-binding profiles.

Due to the complexity of RNA-protein interactions, and the challenges associated with in-depth characterization of RNA binding sites, relatively few studies have assessed how the specificities of RNA-binding proteins may be conserved, or altered over evolutionary time. Many eukaryotic RNA-binding proteins appear to have conserved recognition motifs (Ray 2014). However, there may be multiple modes of binding for a single protein (e.g. PUF (Pumilio and FBF) RNA-binding proteins), and minor changes to a protein sequence can have specific effects on RNA recognition (Valley 2012). Due to the nature of the genetic code, the direct impacts of genomic change on the structure of proteins and RNA are very different. RNA secondary structure is more conserved than sequence within RNA families (Burge 2012). Amino acid sequences of proteins tend to be much more highly conserved than nucleotide sequences of structured RNAs, and it is often difficult or impossible to follow the vertical inheritance of any but the most conserved structured RNAs (e.g. the ribosome) across large evolutionary distances (Lindgreen 2014).

Little work has been done to explore and characterize the RNAs interacting with r-proteins beyond model organism *E. coli*. A comparative genomics analysis was conducted to study the evolution of riboregulatory structures, and to identify unknown structures from non-model species (Fu 2013). From this study it was clear that although the autogenous method of r-protein regulation is widely distributed to many bacterial phyla, the specific regulatory RNA structures themselves are not widely distributed. Of the ten riboregulators originating in *E. coli*, three were found to be widely distributed

over many eubacterial phyla: L1, L10, and S2. The remaining seven RNA structures (interacting with S1, S4, S7, S8, S15, L4 and L20 proteins) were found to be narrowly distributed to a few orders of Gammaproteobacteria. The scattered distribution of the RNA structures, and the low frequency of identification of some regulatory RNAs may point toward horizontal transfer, multiple inventions, or a lack of sensitivity in the homology search methodologies. While the phenomena of autoregulation is widely distributed, it appears bacteria from phyla beyond Gammaproteobacteria have invented different means with which to carry it out.

We are now beginning to understand how many different structures autoregulate expression of r-protein S15. Three different RNA structures have been previously described in *E. coli*, *T. thermophilus*, and *G. stearothermophilus* (Portier 1990, Scott 2001, Serganov 2003), which are found to be narrowly distributed to Gammaproteobacteria, Thermales, and Firmicutes, respectively (Figure 1.5). The wide distribution of autoregulation as a regulatory mechanism for ribosomal protein synthesis coupled with the narrow distribution of the S15-interacting regulatory RNAs, and the existence of alternative RNA structures, strongly suggests that many similar such mechanisms remain to be discovered in other bacterial phyla.

The S15-mRNA Interaction in Bacteria

Beyond its role in rRNA scaffolding, S15 autoregulates the *rpsO* operon by binding a structured *cis*-regulatory RNA located in the 5' untranslated region (Figure 1.4, Lindahl 1986). Though the S15 amino acid sequence is conserved (Figure 1.3), the RNAs

with which it interacts in a regulatory capacity are not. As stated previously, three different regulatory RNAs from the species *E. coli*, *G. stearothermophilus*, and *T. thermophilus* (Portier 1990, Scott 2001, Serganov 2003) have been experimentally described. In some cases, the mRNA-S15 interaction has been well-characterized, such as *E. coli* (Lindahl 1986, Portier 1990, Philippe 1990). In the other species, additional work is needed to elucidate the regulation-specific nucleotides and residues important to autoregulation in these different species.

The first RNA was described in 1990, in the species *E. coli* (Ec-mRNA and Ec-S15, Figure 1.6A, B, Portier 1990, Philippe 1990). Using mutagenesis and *in vitro* structure probing techniques it was found that in the absence of Ec-S15, the Ec-mRNA folds into two mutually-exclusive stem loops that overlap the ribosome binding site, start codon, and first four codons of *rpsO* (Portier 1990, Philippe 1990, Philippe 1995, Serganov 2002). The stem-loop structure allows ribosome-recognition and translation of the *rpsO* transcript. To autoregulate, DMS and other footprinting experiments show that Ec-S15 specifically binds a G•U/G-C motif that stabilizes the formation of a pseudoknotted mRNA structure (Figure 1.6A). Binding causes the two helices of the pseudoknot to co-axially stack. No base-specific interactions were found within the pseudoknot but its presence is required, indicating this region forms a secondary Ec-S15 binding site (Philippe 1990, Philippe 1995). Two adenines bridge the stacked helices, and they were protected by Ec-S15 from nuclease cleavage, suggesting they may also be important for Ec-S15 recognition. Deletion analysis showed the remainder of the sequence, which may fold into a stem loop, is not essential for binding Ec-S15 (Serganov 2002). However, the start codon and ribosome binding site are located in this loop. This

allows the 30S ribosomal subunit to co-dock on the mRNA at the same time as S15. Both bound to the mRNA entraps the small ribosomal subunit, prevents full ribosome assembly and ultimately prevents translation. Thus, autoregulation of the *rpsO* operon in *E. coli* occurs through an entrapment mechanism.

Through one *in vivo* regulation study the residues of Ec-S15 required for autoregulation with Ec-mRNA have become more clear. Ec-S15 site-directed mutagenesis shows that residues T21, G22, Q27, H41, D48, S51, and K58 are all required for autoregulation (Figure 1.6B, Mathy 2004). Many of these amino acids are highly conserved and known to be involved in rRNA recognition (Nikulin 2000). Both Ec-mRNA and rRNA contain a G•U/G-C motif, which suggests that Ec-S15 is recognizing this motif using the same amino acids, H41, D48, and S51. Because residues T21, G22, and Q27 were also found to be essential for autoregulation, it suggests that S15 recognizes and stabilizes the pseudoknot stem via these residues. There remain some noteworthy differences from the amino acids within Ec-S15 that are required for autoregulation as opposed to rRNA-binding. Intriguingly, several rRNA-specific binding residues are not required for autoregulation, including R64, Y68, and R71. These residues contact the 3HJ of the rRNA; this confirms there is only topological mimicry in Ec-mRNA, and no direct 3HJ mimic (Nikulin 2000). Finally, an autoregulation-specific residue was identified, K58. This amino acid is hypothesized to bind the adenine bridge, however, this has not been verified experimentally.

A second autoregulatory RNA structure was discovered in the species, *G. stearothermophilus* (Gs-mRNA and Gs-S15, Figure 1.6C, Scott 2001, Scott 2005). The sequence of Gs-mRNA is distinct from that of Ec-mRNA and rRNA. Additionally, *in*

vitro mutagenesis and Gs-S15 binding assays suggest that the Gs-mRNA secondary structure is distinct from Ec-mRNA. In these studies it has been hypothesized, yet not experimentally verified, that Gs-mRNA structure mimics portions of the rRNA, especially in the formation of a 3HJ, albeit with distinct nucleotide sequence.

Several residues essential for autoregulation between Gs-S15 and Gs-mRNA were identified in one *in vivo* mutagenesis study (Figure 1.6D, Scott 2005). Strikingly, those residues almost completely matched those found to be essential for binding rRNA (figure 1.2C). These include K7, D20, T21, Q27, R64, Y68, N71, and K72. This suggests these residues interact with a 3HJ in both RNAs. Additionally, residues known to interact with the G•U/G-C motif in rRNA were also found to be essential for autoregulation, including H41, D48, and S51, which suggests the Gs-mRNA may contain a structural mimic of the G•U/G-C motif and Gs-S15 is recognizing this motif in both the rRNA and Gs-mRNA via the same residues. *In vitro* cross-species binding assays were also performed with Gs-S15 and the RNA from *E. coli*, Ec-mRNA (Scott 2005). Gs-S15 was unable to bind Ec-mRNA, which suggest S15 homologs from these two species may utilize alternative motifs in their respective regulatory mRNA structures.

In 2003, a third *cis*-regulatory structure was reported for S15 in the species *T. thermophilus* (Tt-mRNA, Tt-S15, Figure 1.6E, Serganov 2003). *In vitro* characterization studies show that Tt-S15 recognizes Tt-mRNA primarily at a 3HJ whose sequence is nearly identical to that found in rRNA. In hydroxyl-radical and nuclease footprinting assays, Tt-S15 protected a second region of Tt-mRNA, a site distal to the 3HJ (Serganov 2003), though this RNA contains no obvious sequence mimic of the G•U/G-C. No *in vivo* regulation studies have been performed to validate the regulatory role of Tt-mRNA;

however, an *in vitro* cell-free translation system showed synthesis of products from the Tt-mRNA transcript decreased in a Tt-S15 dose-dependent manner (Serganov 2003). These authors go on to show using toeprinting experiments that Tt-S15 competes with the ribosome for binding Tt-mRNA, which is highly suggestive that regulation occurs via a displacement mechanism. At this time, the autoregulation-specific residues in Tt-S15 remain unknown.

To summarize, three distinct *cis*-regulatory RNA structures have been documented to control expression of the *rpsO* operon in the species *E. coli*, *G. stearothermophilus*, and *T. thermophilus*. Extensive structure probing of Ec-mRNA and limited structure probing of Gs-mRNA and Tt-mRNA does little to elucidate how such diversity of RNA structures have evolved to perform analogous function. A goal of this thesis work is to address whether the predicted secondary structures of these mRNAs conceal mimicry to rRNA, whether there are mRNA-specific binding motifs, or whether there are phyla-specific binding profiles for these mRNA regulators.

FIGURES & LEGENDS

Figure 1.1 Complexity of The Prokaryotic Ribosome

An overview of the structure of prokaryotic ribosomes. (A) The prokaryotic ribosome is made up of a small (30S) subunit and a large (50S) subunit. Three ribosomal RNAs (B) (gray) and over 50 ribosomal proteins (C) (various colors) come together in a coordinated fashion to form functional ribosomes. Figures were generated using PyMOL and crystal structure data from *T. thermophilus* (Agalarov 2000).

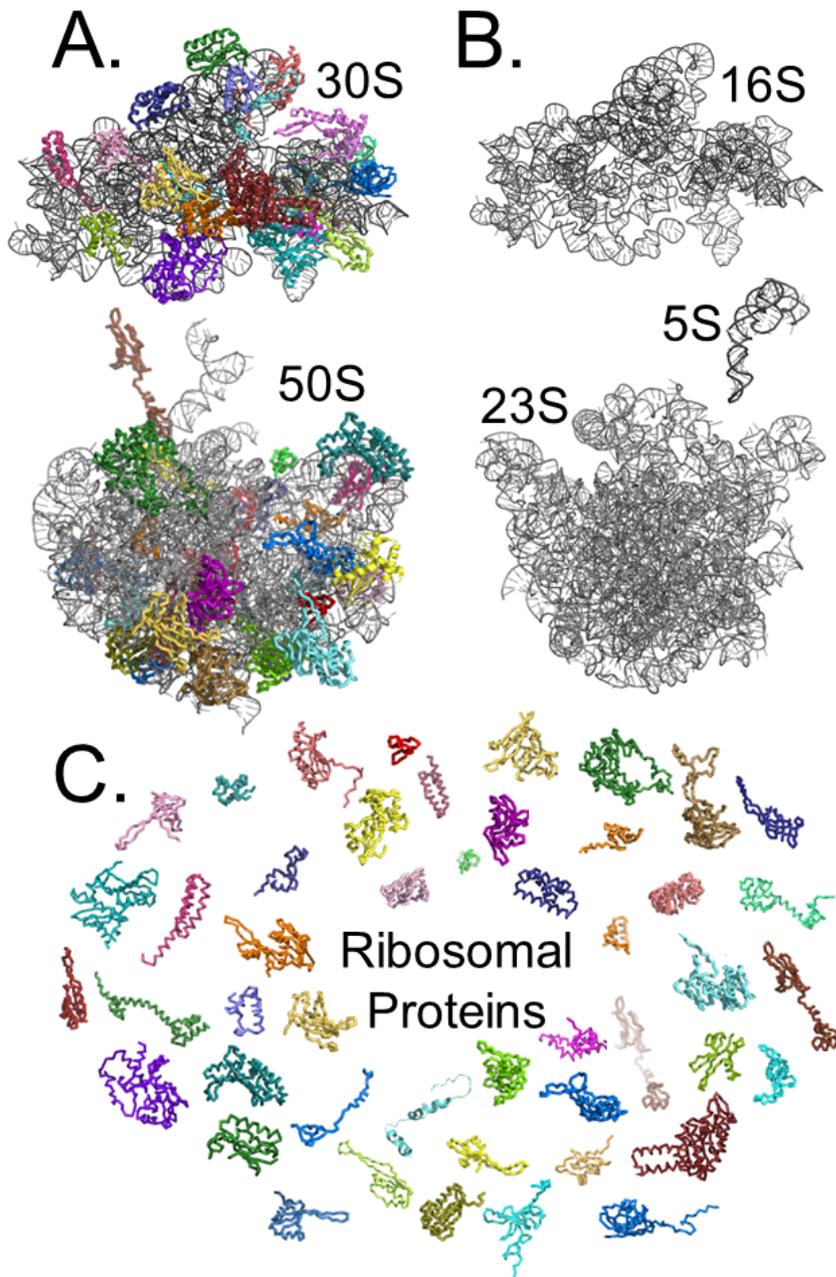


Figure 1.3 Conservation of Residues and Nucleotides Involved in S15-rRNA recognition

Conservation of the nucleotides and amino acids involved in the S15-rRNA interaction in bacteria (A) Region of the 16S rRNA that S15 binds (adapted from data at the Comparative RNA Website (Cannone 2002)), upper case red letters are conserved >98%, lower case letters 90-98%, closed circle 80-90%, and open circle <80% conserved (B) Conservation of individual amino acids (generated with Weblogo (Crooks 2004)) using three representative species from Firmicutes (*Geobacillus stearothermophilus*, *Geobacillus* sp. Y412MC61, and *Geobacillus kaustophilus*), Thermales (*Thermus thermophilus* HB8, *Thermus oshimai* JL-2, and *Thermus parvatiensis* strain RL) and Gammaproteobacteria (*Shigella sonnei*, *Escherichia coli* strain MRE600, and *Escherichia coli* strain SF-173). Highly conserved residues in red. (C) Diagram of S15 protein, alpha helices are indicated (α), and the regions that interact with the three helix junction (3HJ, red) or the G•U/G-C motif (green) are boxed. (D) Amino acid sequence comparison of the S15 residues from distantly related bacteria, *E. coli* (Ec-S15), *G. kaustophilus* (Gk-S15), and *T. thermophilus* (Tt-S15). These sequences were aligned using MultAlin (Corpet 1988). Black residues have high consensus, dark gray residues have low consensus, light gray residues have no consensus between the three species.

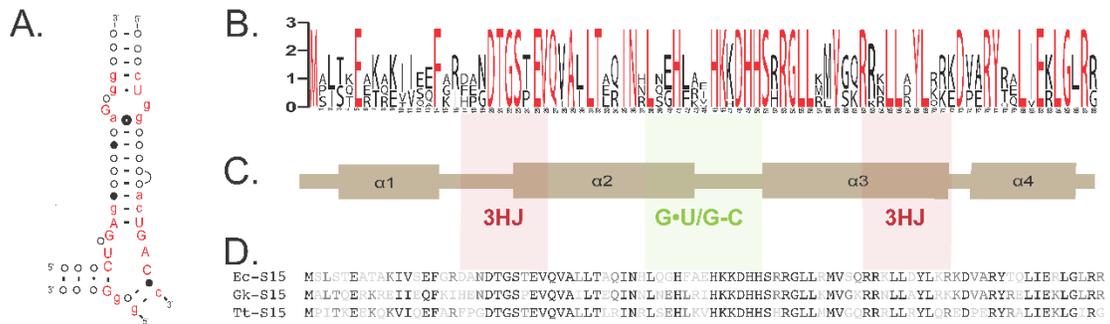


Figure 1.4 Autoregulation of R-Protein S15

Autoregulation of the *rpsO* operon by r-protein S15. When no additional S15 is needed for ribosome assembly, the protein binds a structure RNA element in the 5'-UTR of the *rpsO* transcript to prevent further expression of S15. In this manner, a bacterium can quickly respond to changing environmental conditions and stimuli, utilizing its resources economically.

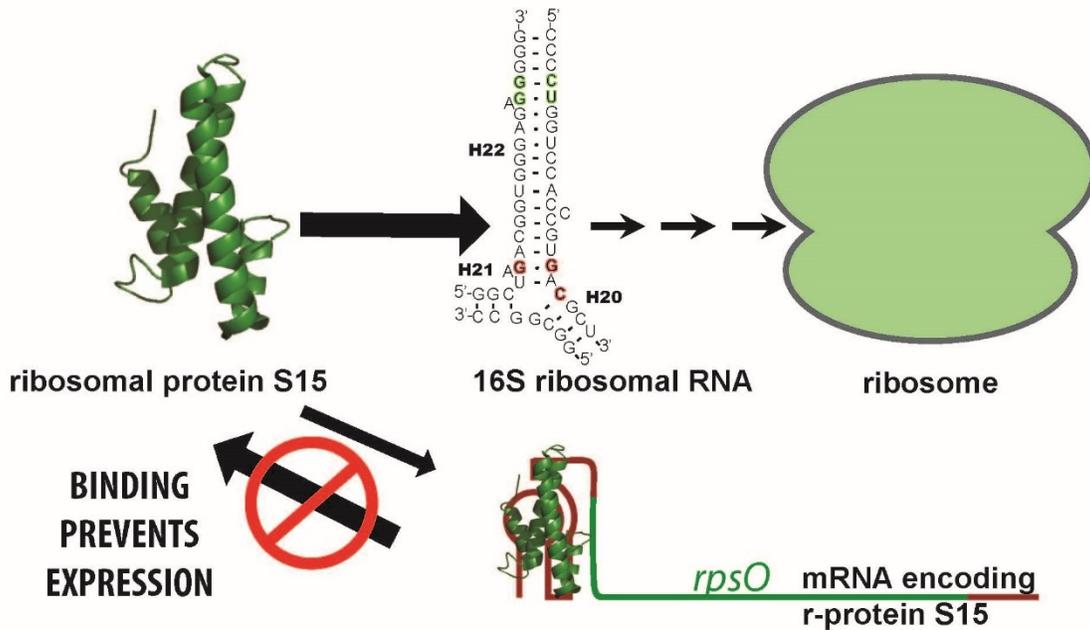
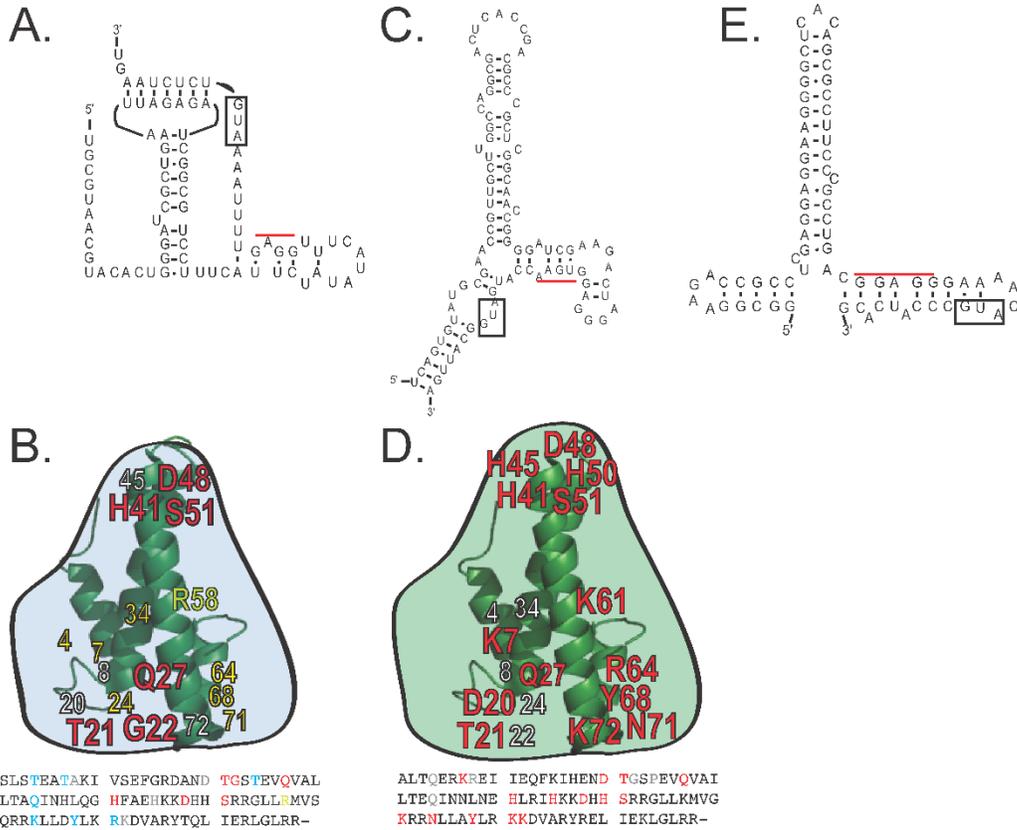


Figure 1.6 *rpsO* Regulatory RNAs

Three regulatory RNA structures have been previously reported for ribosomal protein S15. The RNA from (A) *Escherichia coli* that interacts with the indicated residues in (B) S15 from *Escherichia coli*. The RNA from (C) *Geobacillus kaustophilus* (close relative of *G. stearothermophilus*) that interacts with the residues indicated on the diagram of (D) S15 from *G. stearothermophilus*. And the RNA from (E) *Thermus thermophilus*. In each structure the *rpsO* start codon is boxed, and a bar is placed over the Shine Dalgarno sequence.



CHAPTER II

Discover and Validate Novel mRNA Regulatory Structures for S15 in Diverse Bacterial Phyla

The work in this Chapter is based up on the following published journal article. Text, figures, and tables from this publication are used throughout this Chapter without additional notice.

Slinger BL, Deiorio-Hagger K, Anthony J, Gilligan MM, Meyer MM: Discovery and validation of novel and distinct RNA regulators for ribosomal protein S15 BMC Genomics 15:657 (2014).

Authors' contributions:

BLS performed the vast majority of the experimental work including development of the GFP reporter system, the filter-binding assays, the structural probing, preliminary data analysis, and wrote the manuscript. KD-H performed the comparative genomic search to identify RNAs, manually curated the alignments, and edited the manuscript. JSA implemented the computational strategy to allow the comparative genomic search and edited the manuscript. MG performed filter-binding assays and GFP regulatory analysis. MMM conceived of the project, wrote the manuscript, and finalized data analysis for both experimental and comparative genomic sections. All authors read and approved the final manuscript.

INTRODUCTION

RNA-based autoregulation as a regulatory mechanism for ribosomal protein (r-protein) synthesis is widespread in bacteria. The narrow distribution of the regulatory RNAs for r-protein S15 coupled with the existence of alternative RNA structures that perform the same regulatory function strongly suggests that many similar such mechanisms remain to be discovered in other bacterial phyla. While this collection of RNA regulators already highlights RNA structural diversity, examination of their phylogenetic distributions indicates that most bacterial phyla have no previously described S15 regulation (Figure 1.5, Fu 2013, Deiorio-Haggar 2013).

We implement a framework for computational identification of structured RNAs in bacterial genomes, Genomic Analysis for Illuminating Structured RNA (GAISR, Figure 2.1). GAISR was applied to genomic regions proximal to the S15 coding region (*rpsO*) to assess the diversity of S15-interacting RNAs in bacteria. Our search resulted in many putative structured RNAs across different phyla of bacteria. Sequence alignments corresponding to several of these putative RNA structures were further examined to determine phylogenetic distributions and identify transcription start sites from available RNA-seq data.

To establish the biological relevance of our results, we experimentally demonstrate that one of these RNAs, originating from the alphaproteobacterium *Rhizobium radiobacter* (also called *Agrobacterium tumefaciens*), has the expected biological function. We validate specific interactions between the predicted RNA structure and the S15 protein using *in vitro* binding assays, and pinpoint regions of the RNA important for protein-interaction using mutagenesis and truncation. The secondary

structure is further confirmed using structural probing assays. Finally, we also demonstrate that the novel mRNAs regulate gene expression in response to their respective S15 homolog using an *E. coli* surrogate reporter system.

This work illustrates the importance of integrating comparative genomic and transcriptomic approaches during *de novo* ncRNA identification, revealing a plethora of distinct natural RNA regulators that can support analogous biological functions. Furthermore, this work indicates that a diversity of distinct RNA regulators is likely to exist within bacterial genomes and the plasticity of RNA structure allows distinct, and likely independently derived, solutions to the same biological problem.

RESULTS & DISCUSSION

Non-coding Regulatory RNA Discovery Using Comparative Genomics

To identify putative RNA structures associated with the coding region for ribosomal protein S15 (*rpsO*), we implemented a computational pipeline, GAISR (Genomic Analysis for Illuminating Structured RNA, Figure 2.1) for *de novo* ncRNA discovery and candidate refinement. GAISR is based on existing RNA discovery pipelines (Yao 2007) that have been very successful at identification of ncRNA candidates (Weinberg 2010, Weinberg 2007). GAISR utilizes several pre-existing tools, including CMfinder, a *de novo* ncRNA discovery tool (Yao 2005), and Infernal 1.1, an RNA homology search tool (Nawrocki 2013) to streamline sequence selection, identify potential ncRNAs, and efficiently detect additional homologues for putative RNA structures. We used GAISR to examine the genomic region corresponding to the 5'-

untranslated region of the gene encoding S15, *rpsO*, in fully sequenced bacterial genomes. From the initial search we identified 52 potential ncRNA sequences, originating from 16 initial phylogenetic sequence clusters.

From these initial sequences, we identified five promising RNA structures based on the number of representative species in the alignment and the predicted RNA structure based on those sequences. Among these structures were the two known RNAs that allow regulation of *rpsO* in Firmicutes and Gammaproteobacteria (Deiorio-Haggar 2013, Fu 2013). Of note, the RNA structure reported for *Thermus thermophilus* was not identified by our search, suggesting that more RNAs may be present that were not uncovered here. There are several potential reasons for this result including biases in sequence coverage (there were only 19 sequences derived from *Deinococcus/Thermus* available for analysis), and our use of a single RNA discovery tool for identification of RNA structures may limit our ability to identify putative RNA structures. No tool for RNA de novo discovery is designed to identify potential pseudoknotted structures, yet these are very common in biologically functional RNAs (Staple 2005). Because of this, the pseudoknotted structures we have identified (e.g. from Gammaproteobacteria) are typically identified as individual helices by CMfinder and manually merged during the curation process.

Alignments corresponding to the three promising novel structures were curated and additional examples identified using Infernal homology searches. In addition, the phylogenetic distribution of each putative ncRNA was examined, and each alignment was compared with existing RNA-seq data to identify regions likely to be within the *rpsO* transcript. Consensus diagrams of the three candidate ncRNAs are shown in Figure 2.2

(A-C) and the alignments that correspond to these structures may be found as Additional file 1, Additional file 2 and Additional file 3. RNA-secondary structures determined from analysis of large phylogenies are often well defined by co-varying nucleotide positions. However, individual sequences corresponding to the RNA structures we identified contain extensive variability including many non-canonical base-pairs and variable-length regions outside of the very well-conserved regions that are likely directly involved in protein-binding. Thus the secondary structure predictions in Figure 2.2 should be considered tentative. However, the degree of conservation observed here is consistent with that observed for other ribosomal protein-interacting regulatory RNAs that have been experimentally validated in the past (Fu 2013, Deiorio-Haggar 2013). Therefore, despite the sequence and structure variability, we believe that the RNAs we identified are likely to have a regulatory function.

RNAs Identified Are Diverse in Sequence and Secondary Structure

Our first RNA (Figure 2.2A, D) was identified in greater than 90% of species within the Alphaproteobacteria orders of Rhizobiales, Rhodobacterales, Rhodospirillales, Caulobacterales, and Sphingomonadales. However, only a single example of the RNA was found in a Rickettsiales species (from 58 genomes explored), potentially reflective of genome reduction in most Rickettsiales species (Merhej 2011). Our original putative RNA structure included three predicted pairing elements (H0-H2). In ~50% of examples there is also a long-linker region between H1 and H2 (up to 400 nt) that is typically base-paired, although the precise position of this base-pairing within the sequence does not appear to be well-conserved (see Additional file 1 for alignment). The most highly

conserved portion of the putative RNA is the H1 helix. This helix shows extensive evidence of co-variation and the loop region is highly conserved suggesting that it is important for protein binding. The H2 helix is less conserved, but typically encompasses a putative ribosome-binding site in the 3' portion. While H0 shows some co-variation, the loop region is not well conserved in sequence or length. In addition, transcriptomic analysis of RNA-seq data derived from *Rhodobacter spaeroides* (Giannoukos 2012) (Additional file 4, Figure 2.3A) suggests that the 5' portion of this pairing element is not transcribed (Figure 2.2D), thus we believe that the originally predicted H0 pairing element is likely not part of the biologically relevant RNA.

Our second RNA (Figure 2.2B) was identified mainly in the Actinomycetales order of Actinobacteria. The putative RNA structure contains a kissing-loop pseudoknotted structure that bears faint resemblance to the RNA structure originating from *E. coli* (Figure 1.6A), and there are weakly scoring homologs that appear in various Gammaproteobacteria (e.g. *Pseudomonas*) lacking the known *E. coli* S15 regulator (Fu 2013). However, the closing pseudoknot occurs prior to any potential regulatory sequences suggesting that the “entrapment” mechanism proposed for the *E. coli* RNA is not likely to play a role here (Serganov 2002, Philippe 1993). Like the RNA described above, a ribosome-binding site is apparent in the 3' portion of the H2 helix, suggesting a potential translational regulatory mechanism (see Additional file 2 for alignment). Analysis of RNA-seq data from *Mycobacterium tuberculosis* (Additional file 4, Figure 2.3B), suggests that the transcription start site for this RNA is approximately 10 nucleotides upstream from the start of the first predicted pairing element.

Our third RNA originates from Chlamydia, and is the one in which we have the least confidence, mainly due to the limited sequence diversity available for analysis (Figure 2.2C, F, see Additional file 3 for alignment). However, there is a very strongly conserved hairpin overlapping start codon of *rpsO* in approximately 30 sequenced strains of Chlamydia and a second potential short pairing element displaying some covariation and compatible mutations. In our original prediction, this hairpin was significantly extended (H0). However, pre-existing analysis of transcript start sites in *Chlamydia trachomatis* indicates that the transcript start site is just upstream of H1 (Figure 2.2F) (Albrecht 2010). Therefore, we believe that H0 is likely not part of the biologically relevant RNA. Notably, very few regulatory RNAs have been identified in Chlamydia. Only examples of the TPP and cobalamin riboswitches have been identified in this class of bacteria (Gardner 2011), and in these cases there appear to be only isolated sequences rather than elements that are conserved in many genomes.

The process of curating our original alignments, and in particular the incorporation of RNA-seq data, was critical for narrowing our focus to the portions of the predicted RNAs that are most likely to be biologically relevant. In two cases, transcriptomic data allowed us to determine that putative hairpins predicted through comparative genomics are unlikely to be part of the transcript. Our analysis exemplifies that in assessing the biological relevance of a given ncRNA candidate it is important to determine whether a putative RNA is actually transcribed as well as identify the transcription start site of the RNA candidate (Lu 2011). Thus, archives that consolidate RNA-seq data, and provide easily accessible read-depth information for many bacterial species are of great importance moving forward in RNA comparative genomics.

RNA from Alphaproteobacterium *Rhizobium radiobacter* Specifically Interacts with S15 Protein

To experimentally validate the biological relevance of our results, we further examined an example of the Alphaproteobacterial RNA originating from *Rhizobium radiobacter* (NC_003062, organism also known as *Agrobacterium fabrum* strain C58, and formerly known as *Agrobacterium tumefaciens* strain C58). The sequence from *R. radiobacter* conforms well to our consensus structure, containing the highly conserved H1, and the predicted H2 pairing element. In addition, this sequence is one where the region directly preceding our transcription start site has the potential to base-pair with the 5'-most portion of the RNA. We designated this helix H0 due to its position 5' of the predicted transcription start site.

We first tested the full-length version of the RNA (nucleotides -108 to +27) and called it Rra-RNA1 because it was the first RNA tested from this organism (Figure 2.4A). To examine whether this RNA interacts specifically with S15 protein from the same organism (Rra-S15) we utilized filter-binding assays (Hall 1999). These assays confirmed that Rra-RNA1 binds Rra-S15 with nanomolar affinity (Figure 2.4A, B, $K_D = 22.2 \pm 0.7$ nM). This value is similar to those reported for the interactions between S15 and the RNA structures originating from *G. stearothermophilus* (20 nM) (Scott 2001), and the *T. thermophilus* (5 nM) (Serganov 2003).

Truncation Analysis Suggests that Rra-S15 Minimal Binding Site Includes Both H1 and H2

To experimentally investigate the validity of the putative transcription start site we constructed several 5' truncations to the Rra-RNA1 sequence and tested their ability to bind Rra-S15. Based on the putative transcription start site derived from analysis of RNA-seq data from *Rhizobium sphaeroides* (at C-95 according to our alignment), the potential H0 helix predicted from comparative genomics in the absence of RNA-seq analysis (Figure 2.2A, D) is unlikely to be necessary for Rra-S15 binding. We performed 5'-RACE for this RNA to further identify the transcription start site. Although C-95 was one of the 5'-ends identified (Figure 2.5), this experiment provided multiple 5'-ends and was ultimately inconclusive. Truncations Rra-RNA2 (nucleotides -91 to +10) and Rra-RNA3 (nucleotides -79 to +10) appear to have negligible effects on Rra-S15 binding ($K_D = 14.5 \pm 6.1$ nM and 21 ± 4.8 nM, respectively) (Figure 2.4A, B). These results indicate that all bases upstream of nucleotide -79 are not required for binding Rra-S15, consistent with the putative transcription start site prior to this nucleotide at C-95. Binding was not significantly affected until the RNA was truncated to G-72, Rra-RNA4 ($K_D = 125 \pm 106.5$ nM) (Figure 2.4A, B). Collectively, these results suggest the entire H0 stem and loop are dispensable and the C-95 identified during analysis of RNA-seq data from *R. sphaeroides* likely represents the transcription start site.

To identify the minimal protein binding-site, we examined 3'-truncations to the Rra-RNA (Figure 2.4A, C). In the Gammaproteobacterial RNA, the initial amino acid encoding nucleotides of *rpsO* form an integral part of the RNA structure and function (Philippe 1993). However, removing the coding region of the Alphaproteobacterial RNA

(Rra-RNA5, nucleotides -108 to +5) has minimal effect on the binding affinity ($K_D = 16.6 \pm 10.8$ nM). Rra-RNA6 (nucleotides -108 to -6) was designed to remove all bases downstream of the predicted H2; again, this RNA binds Rra-S15 with an affinity better than that of Rra-RNA1 ($K_D = 11.9 \pm 1.8$ nM). The observed increase in binding affinity is likely due to removal of potential alternative competing structures, thus allowing a tighter interaction between the protein and the RNA. Rra-RNA7 (nucleotides -108 to -31) was designed to remove all of predicted hairpin H2 including the five uracils (U-26 to U-30) through the putative ribosome binding site (purine-rich sequence from A-8 to A-13), start codon and subsequent protein coding nucleotides. This truncation completely abolishes Rra-S15 binding ($K_D > 500$). To assess whether slippage along the predicted H2 might allow the five uracils (U-26 to U-30) to base-pair with the putative ribosome binding site (A-8 to A-13), we mutated the polyuridine to a purine-rich sequence to destabilize this alternative pairing (Rra-RNA8, Figure 2.4A). This mutant was able to bind Rra-S15 with a similar affinity to the full length Rra-RNA1 ($K_D = 12.5 \pm 2.9$ nM) suggesting that the pairing we have drawn is one that allows for protein binding. This mutant did slightly affect the maximum RNA fraction bound to Rra-S15, suggesting this rather larger nucleotide swap may result in some slight conformational alteration, yet does not affect the protein binding regions. Based on these data we predict the minimal RNA regulatory region includes nucleotides G-79 through U-6, which is fully encompassed by our predicted transcript, and includes both of the predicted pairing elements H1 and H2.

Mutation Analysis Suggests a Potential Binding-site

The most highly conserved portion of this RNA is the H1 stem loop (Figure 2.2A). Due to its high sequence conservation, we hypothesize that this region is essential for Rra-S15 binding. Mutations made in this stem, Rra-RNA9 and Rra-RNA10, both significantly inhibit binding (Figure 2.4A, D, K_D values of >500 and 221 ± 52.3 nM, respectively). The compensatory mutation, Rra-RNA11, was able to partially recover Rra-S15 binding ($K_D = 114 \pm 37$ nM). In this compensatory mutant, it is likely an alternative base-pair forms with usually unpaired A-41 and U-75, which may slightly alter the H1 stem structure. The dynamic equilibrium of the two RNA structures may allow, but does not enable complete restoration of Rra-S15 binding. In combination with the truncation experiments above, these results suggest that Rra-S15 binds its RNA regulator in the highly conserved stem-loop structure of H1 but that H2 is still required for binding.

The *R. radiobacter* RNA Allows Regulation in Response to S15 *in vivo*

To determine whether the Rra-RNA has regulatory activity in addition to S15-binding activity, we conducted *in vivo* reporter assays to assess regulation. To do this we used a GFP reporter to measure expression of the gene following the Rra-RNA in response to different levels of Rra-S15. The RNA sequence was cloned in-frame as a translational fusion with the GFP reporter under the control of the *ptrc* promoter. This construct included the *rpsO* start codon, Shine-Dalgarno sequence, and the first nine codons of the *rpsO* gene to form the *ptrc*-RNA-GFP fusion, called pBS1-RNA. On a second plasmid, the *R. radiobacter rpsO* coding sequence was placed under the control of

an L-arabinose inducible promoter. The pair of plasmids were co-transformed into an *E. coli* K12 strain CK1953. We chose to use a surrogate organism, *E. coli*, due to its ease of use and manipulation. Using this GFP *in vivo* reporter system, we assessed the ability of Rra-S15 to regulate gene expression by measuring the GFP levels in the cells in the presence and absence of induced Rra-S15. If the RNA interacts with Rra-S15 to regulate gene expression, we expect to see a decrease in GFP expression in cells expressing Rra-S15 compared to cells not expressing Rra-S15.

Cells co-transformed with plasmids containing full length Rra-RNA1-GFP, and Rra-S15 were grown in the presence and absence of L-arabinose. The cells grown in the presence of the sugar (induced Rra-S15) displayed a ~4-fold decrease in GFP-reporter expression (Figure 2.6). Because L-arabinose induces Rra-S15 production, the decrease in GFP reporter expression is likely due to an interaction between the RNA and Rra-S15. Next, to corroborate that our predicted transcription start site at the C-95 allows regulation, the sequence for Rra-RNA3 (nucleotides -78 to +27) was also tested in this system and behaved in a similar manner. These results indicate an RNA sequence starting at the transcription start site derived from *R. spaeroides* is sufficient to allow regulation *in vivo*.

We also examined whether mutations to H1 that abolish Rra-S15 binding would affect regulation. Cells containing either Rra-RNA9-GFP and Rra-RNA10-GFP did not display a significant difference in GFP reporter expression when grown in the presence and absence of L-arabinose. This is likely because both Rra-RNA9 and Rra-RNA10 do not interact specifically with S15 and are unable to regulate the expression of the reporter. However, it should be noted that GFP expression levels in the absence of

arabinose were also significantly lower than those observed with the Rra-RNA1 and Rra-RNA3 constructs.

To assess whether the partial compensation of binding observed for RraRNA11 *in vitro* represented a biologically relevant, functional compensation, Rra-RNA11 was also examined in this system. In this case, cells grown without L-arabinose displayed an increased GFP expression level compared to Rra-RNA1, and cells grown in the presence of L-arabinose had a ~10-fold decrease in relative GFP fluorescence (Figure 2.6). However, the increase in fold-change is solely due to increased RNA11-GFP expression levels and the repressed level of gene expression is comparable between the two RNA elements. Thus, the Rra-RNA11 compensatory mutation that partially restored the *in vitro* RNA-protein interaction also restored the regulatory interaction between the Rra-RNA and Rra-S15. The partial restoration of *in vitro* binding by the compensatory mutant Rra-RNA11 is likely due to the presence of several competing structures formed by the RNA under these conditions. However, the *in vivo* conditions enable the RNA to adopt a secondary structure that increases overall reporter expression and enables regulation in response to S15. Together, these assays indicate that not only does this RNA interact with Rra-S15 *in vitro*, but it is a biologically relevant regulatory element responding to S15.

Structural Probing Confirms Predicted Secondary Structure of Rra-RNA

To further examine the secondary structure of Rra-RNA in the absence of protein we used several structural probing methods in combination with a minimal RNA construct (Rra-RNA6) including nuclease cleavage assays (with RNase VI and RNase A),

and in-line probing. RNase VI cleaves double stranded RNA non-specifically, RNase A cleaves single stranded C's and U's, and in-line probing the RNA structure reveals the flexible regions of the RNA structure (and likely single-stranded regions) that are more prone to spontaneous self-cleavage.

Although the putative stem H2 is predicted in our alignment (Additional file 1), there are many sequences that contain short polypyrimidine sequences that are unpaired in our sequence alignment. These sequences may form alternative pairings with the ribosome-binding site (AG rich region ~8 nucleotides before the translation start site). Based on sequence data alone it is difficult to distinguish which bases are interacting with the ribosome-binding site. However, several lines of evidence indicate that we have identified the correct *in vitro* base-pairing conformation for our putative H2 in the *R. radiobacter* example of the RNA (Figure 2.7A, B). First, our mutagenesis and truncation analyses indicate that mutating the polyuridine (U-26 to U-30 in Rra-RNA8) does not alter protein-binding activity. This suggests that this region is unlikely to pair with the putative ribosome-binding site (-13 to -8). However, deleting this region and the following hairpin (Rrad-RNA7) abolishes protein binding indicating that H2 is important for protein binding. Second, RNase V1 cleavage occurs symmetrically in regions that are base-paired in our figure (-8 through -12 and -23 to -28), and RNase A cleavage occurs at C-17 as would be expected for a loop region. In addition, in-line probing shows that the entire 3' portion of the molecule is somewhat flexible, from bases -11 through -22 (region A). In conjunction with our mutagenesis results, this strongly suggests the correct pairing-element has been identified.

Consistent with our mutagenesis results, the highly conserved stem H1 (bases U-42 to A-65) is almost certainly double-stranded with a loop from C-49 to A-56. Bases U-42 through C-49 are shielded from in-line attack and there are strong cleavage bands in RNase VI probing for bases C-43, bases -59 to -61, and G-64. Also, there are no RNase A cleavage products for any of these uracils or cytosines, suggesting that these nucleotides are not single-stranded. There is also evidence for the predicted loop region in H1. Probing with RNase A results in cleavage products for C-53 and C-56 and in-line probing reveals that C-53 through A-57 (region D) are flexible. At the base of the H2 stem, we predict a bulged adenosine (A-41) and the highlighted region C from our in-line probing gel corresponds to this bulged base. These data corroborate our other evidence that the region essential for Rra-S15 binding in H1 forms a double-helix.

The nature of the junction between the two predicted helices is still unresolved. This region is not well-conserved so there is little phylogenetic evidence of structure, and in several cases the different assays give conflicting results, which may be the result of multiple folding conformations. The string of uridines from U-27 to U-30 does not appear to be flexible based on in-line probing, is cleaved by RNase VI, and is not cleaved by RNase A, indicating that the region is not single-stranded. However, there are also no clear binding partners for these nucleotides suggesting that they may be forming a constrained tertiary structure. The string of cytosines that follows this region, C-32 to C-35, do show strong RNase A cleavage suggesting they are single-stranded, and this is corroborated by the in-line cleavage at these positions (region B). However, these bases also display RNase VI cleavage indicating that they may sometimes adopt a double-stranded conformation. Nucleotides from -35 to -40 are not cleaved by either RNase VI

or RNase A, and appear to be structurally constrained. This suggests that they are not necessarily double-stranded, but may be participating in some tertiary structure. Nucleotides -75 to -80, which potentially could interact with these bases, also show conflicting results, cleaving with both RNase V1 and RNase A. We have included the possible base pairing of the nucleotides at the base of H1 in our structure figures (Figures 2.4 and 2.7), but these interactions are likely weak. Taken together, our data suggest H1 and H2 form, but the nature of the junction between these helices remains unclear presently.

CONCLUSION

We have discovered three additional regulatory RNA structures for ribosomal protein S15 that are narrowly distributed to Chlamydia, Actinobacteria and Alphaproteobacteria (Figure 2.8). We also present experimental evidence that an example from Alphaproteobacteria, in *R. radiobacter*, performs its predicted regulatory function *in vivo* using a distinct structure from those previously described.

This work demonstrates the premise that nature may invent many unique ways to solve a single biological problem. In the context of other forms of RNA-based regulation the diversity of distinct RNA structures allowing *cis*-regulation of the *rpsO* operon is nearly unmatched. The only similar example of such diversity in RNA regulators for a specific function are the SAM-binding riboswitches, where more than three completely distinct classes (Gilbert 2008, Montange 2006, Lu 2008), and several additional subclasses with re-arranged or modified secondary structure elements have been

characterized (Weinberg 2008, Poiata 2009). The S15 autoregulatory RNA structures we identified are quite diverse from one another, and from the existing known characterized S15 regulatory RNAs that originate from *E. coli*, *G. stearothermophilus*, and *T. thermophilus* (Philippe 1990, Scott 2001, Serganov 2003).

All of the previously characterized RNA structures encompass a predicted Shine-Dalgarno sequence, but beyond such regulatory features the RNAs appear to share very few common sequence features or patterns in secondary structure. While the S15-interacting RNAs potentially share some tertiary structure similarities that are not captured in the secondary structure diagrams, previous studies indicate that the *E. coli* S15 does not interact with the regulatory RNA originating from *G. stearothermophilus* (Scott 2005). This finding suggests that there may be no single conserved tertiary structure shared by the S15-binding mRNA structures. In the absence of structural data, it remains to be seen whether the structural diversity apparent in natural S15-interacting mRNA structures is a result of RNA's inherent ability to generate a similar tertiary structure from diverse arrangements of primary and secondary structure (Choonee 2007, Nevskaya 2005), or from differences between the S15 protein homologs that lead to distinct pools of potential RNA ligands. From the structures we describe here, it is clear that there are many ways to solve this particular biological problem. Based on the natural diversity of S15-interacting RNAs, we expect that this number is large, and that as more genomes are sequenced and the sensitivity of computational searches increases, additional structures with this function will be identified.

MATERIALS & METHODS

Computational Identification of Putative RNAs and Curation of RNA Alignments

rpsO was identified in the genomes of fully sequenced bacteria (refseq58-microbial) using tBLASTn (Nevskaya 2005). Sequences corresponding to the putative 5' non-coding regions (500 nucleotides 5' of the translation start, or the end of the previous gene) in addition to 25 nucleotides of the *rpsO* coding region was collected. Sequences containing >90% sequence identity over >70% of the sequence length were removed as redundant. The remaining sequences were clustered based on taxonomy into groups of 100 or fewer sequences. CMFinder was run on these clusters with the default parameters (Yao 2005). The resulting alignments were manually curated to identify the most promising RNA candidates.

Covariance models for each RNA alignment were constructed and calibrated using Infernal 1.1 (cmbuild, cmcalibrate), and homologues were identified for each alignment (cmsearch) (Nawrocki 2013). Cmsearch was performed against a custom sequence database described above using a lenient e-value cut-off of 1.0. Alignments were manually adjusted as necessary when sequences with variable-length helices and/or loops were added. The search process was repeated approximately 3-4 times per multiple sequence alignment, to expand sequence diversity. During the course of these searches, the alignments were extended at the 5' and 3' ends to encompass any potential flanking sequence and pseudoknotted or alternative structures were identified through curation of the alignment.

Transcription start sites were identified through examination of mapped read depths derived from RNA-seq data (Merhej 2011, Pruitt 2011) compiled at AREBA (An

(<https://github.com/UCanCompBio/AREBA>), or from previously assessed transcription start sites in the literature (Albrecht 2010). The counts for evolutionary diversity were calculated from the number of completed genomes within refseq58 based on the final alignments. Consensus secondary structure diagrams were created from the alignments using GSC-weighting in R2R (Altschul 1990).

RNA Preparation

DNA corresponding to the 5'-UTR of the *rpsO* gene with the T7-promoter appended was PCR amplified from *R. radiobacter* genomic DNA. Mutants 8-10 were generated through QuickChange mutagenesis on Rra-RNA1 template, then PCR amplified using Rra-RNA1 primer set. T7 RNA Polymerase (Milligan 1987) was used to transcribe RNA, and RNAs were purified by denaturing PAGE (6%), bands visualized using UV shadow, and RNA eluted from excised bands in 300 mM NaCl, 1 mM EDTA. Purified RNA was 5'-labeled with ³²P-ATP (Regulski 2008) and again purified as described above.

Protein Preparation

The *R. radiobacter rpsO* ORF was cloned into pET-HT overexpression vector (Block 2011) and transformed into BL-21(DE3) cells (Invitrogen). Protein was over-expressed and cells lysed by sonication using S15 Resuspension Buffer (100 mM Tris-HCl, pH 8.0, 800 mM NaCl, 150 mM MgCl₂). S15 was soluble and was purified at 4°C using non-denaturing FPLC cation exchange chromatography with a linear salt gradient

(100 mM-1 M NaCl) (Regulski 2008). A second purification was performed under conditions previously described (Block 2011) using pH 8.0 and a linear salt gradient (20 mM – 1 M KCl) at 4°C by non-denaturing FPLC cation exchange chromatography. Proteins were concentrated, analyzed via SDS-PAGE, buffer exchanged for the S15 Storage Buffer (50 mM Tris-Acetate, pH 7.5, 20 mM Mg-Acetate, 270 mM KCl) and final protein concentration determined by Bradford assay and stored at 4°C.

Filter-Binding Assays

RNA binding capability was examined by filter binding assay (FBA). A fixed amount of 5'-labeled RNA (1000 cpm, <1 nM) was renatured 15 minutes 42°C, then incubated with serial dilution of S15 in Buffer A (50 mM Tris-Acetate, pH 7.5, 20 mM Mg-Acetate, 270 mM KCl, 5 mM dithiothreitol, 0.02% bovine serum albumin), for 30 minutes at 25°C. Nitrocellulose membrane (GE Healthcare) was used to collect RNA-S15 and nylon (GE Healthcare) to collect unbound RNA under suction. Membranes were air-dried 5 minutes and fraction bound quantified by imaging membranes on a phosphorimager screen. Radioactivity counts per sample per membrane were measured using GE Healthcare STORM 820 phosphorimager and ImageQuant. The fraction bound was calculated per individual protein concentration $F_b = (\text{counts nitrocellulose}) / (\text{counts total})$. Solver (Microsoft Excel) was used to fit the range of variables (Protein concentration vs. F_b) in order to find K_D .

Structural and Nuclease Probing Assays

The RNA-protein binding reaction described above was used for RNase probing assays. After incubation, 1 uL RNase A (1 ug/mL, Ambion) or VI (1:400 dilution of 0.1 U/uL, Ambion) was added and the reaction incubated 15 minutes at 25°C. The nuclease was inactivated with inactivation/precipitation buffer (Life Sciences) and RNA fragments recovered by ethanol precipitation. Precipitated RNAs were suspended in 10 uL Urea Loading solution (Life Sciences) and incubated 5 minutes 95°C. Five uL of each reaction was loaded on 10% denaturing Acrylamide/Bis-acrylamide gel. The gel was dried and examined using a GE Healthcare STORM 820 phosphorimager and ImageQuant software. Partial hydroxyl cleavage reactions were generated by incubating RNA in Reaction Buffer (50 mM Na₂CO₃ pH 9.0, 1 mM EDTA) at 95°C for 7 minutes. Denaturing T1 reaction was conducted according to manufacturer's protocol (Ambion). For in-line probing, 5'-labeled RNA was incubated 40 hours at 25°C in reaction buffer (20 mM MgCl₂, 100 mM KCl, 50 mM Tris-HCl pH 8.3). The reaction was stopped using Urea loading solution (10 M Urea, 1.5 mM EDTA).

GFP Reporter Plasmid Construction

The ptrc-RNA-GFP plasmid was constructed from pLac-thiMwt-tetA-gfpuv plasmid (kind gift from Yokobayashi, Muranaka 2009). To change the promoter, oligonucleotides encoding the ptrc IPTG-inducible promoter flanked by XhoI and EcoRI restriction sites at the 5' and 3' termini respectively (5'-gagctgttgacaattaatcatccggctcgtataatgtgtggaattgtgagcggataacaatt-3'), and its reverse complement were chemically synthesized (Eurofins MGW Operon), phosphorylated (T4 polynucleotide kinase, NEB), and annealed. The double-stranded DNA was inserted

between the XhoI and EcoRI sites of digested pLac-thiMwt-tetA-gfpuv plasmid, replacing the plac promoter.

DNA fragment containing the RNA leader sequence with restriction sites EcoRI and SalI on the 5' and 3' ends respectively was PCR amplified from genomic template using gene specific primers. The PCR product was digested with EcoRI and SalI and inserted into pLac-thiMwt-tetA-gfpuv plasmid digested by the same enzymes, replacing the thiMwt riboswitch sequence.

A DNA fragment encoding gfpuv replaced the existing gfpuv-tet reporter in this vector. In this translational fusion, the RNA sequence, including the first nine codons of the *rpsO* gene were placed in frame and directly upstream with the GFP. This replaced the thiamine responsive riboswitch and existing ribosome binding site.

S15 Expressing Plasmids

The S15 expression plasmids was constructed by amplifying the DNA fragment encoding the *rpsO* gene from *R. radiobacter* genomic template. On the 5' termini of the sequence, one primer contained a SacI restriction site, ribosome-binding site (RBS) and linker sequence that matched that of the *E. coli rpsO* RBS (5'-AGGAGGTTTTAAA), and an ATG start codon, and *rpsO* binding sequences. The second primer contained *rpsO* binding sequences and an XbaI site on the 3' termini (Table of Primers). PCR product was amplified using genomic DNA extracted from *R. radiobacter* (ATCC 23308). The PCR product was digested with SacI and XbaI enzymes and inserted into the pBAD33 expression vector (ATCC 87402) digested with the same enzymes.

***E. coli* GFP Regulatory Assays for *R. radiobacter* RNA-S15 Interaction**

E. coli (Strain K12, CGSC# 7154: strain CK1953, Yale University) were co-transformed with an RNA and protein plasmid (made competent using the Z-competent buffer system, Zymo Research). Overnight cultures were grown +/- L-arabinose (15 mM), then diluted the next day to OD=0.150 in fresh media (LB + 100 µg/mL AMP + 34 µg/mL CHL +/- 15 mM L-arabinose). At log phase IPTG (2 mM final) was added to induce GFP expression and cells grown an additional 5 hours. Cells were collected, washed with PBS, then stored in PBS overnight. GFP expression was measured using a SpectraMax M5 fluorimeter (excitation: 395 nm, emission: 508 nm, Molecular Devices). Fluorescence was calculated by normalizing GFP to cell density (GFP/OD600).

***R. radiobacter* RNA 5'-RACE**

Total RNA was extracted from log phase *R. radiobacter* cells grown in LB and 5'RACE performed using Invitrogen GeneRacer kit. Reverse transcription was conducted using a gene specific primer (623-RradS15M11R: 5'-atccttctgttttaaggaaaacggg), and the product PCR amplified with an oligo linker-specific primer (9-Forward5'RACE: 5'-gactggagcagaggacactga) and gene-specific primer (647-RradS15M13R: 5'-gacgcctagagccgggatgctgt). PCR product was cloned using TOPO-cloning kit (Invitrogen) and sequenced (Eton Biosciences) to identify the transcription initiation site of the *rpsO* transcript. The results of this analysis can be found in Figure 2.5.

FIGURES, & LEGENDS

Figure 2.1 Overview of Comparative Genomic Pipeline: Genomic Analysis for Illuminating Structured RNA

The process begins with completed microbial genomes, and identifies putative 5'-untranslated regions (5'-UTRs) for *rpsO*, clusters the obtained sequences by their taxonomic group. CMFinder is used to identify potential ncRNA motifs within these clusters. Following ncRNA identification, the RNA motifs are manually inspected and additional homologs are identified using Infernal 1.1. The genomic context of putative homologs is assessed, and they are incorporated into the alignment using *cmalign*. The alignment is typically then manually inspected to identify potential pseudoknots or other regulatory features and the curation process may be repeated several times. Then finally, transcriptomic data are sought to support the transcription, and in particular transcription start site of the putative ncRNA.

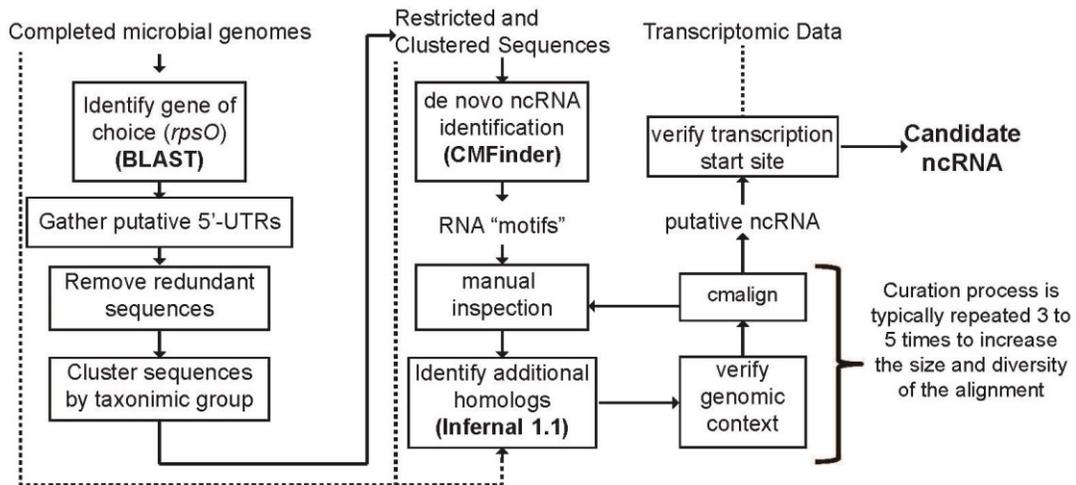


Figure 2.2 Consensus Diagrams of Novel Putative RNA Structures and Individual Examples Used for Transcriptomic Analysis

Novel regulatory RNAs we have identified upstream of the *rpsO* operon. H0 helices were originally predicted by comparative genomics, but not supported by transcriptomic analysis and therefore are unlikely to be biologically relevant. (A) RNA originating from Alphaproteobacteria, (B) RNA originating from Actinobacteria, and (C) RNA originating from Chlamydia. (D) RNA example originating from *Rhodobacter spaeroides* (NC_011963.1) showing putative transcription start site determined from analysis of RNA-seq reads (Figure 2.3A) (Giannoukos 2012). (E) RNA example originating from *Mycobacterium tuberculosis* (NC_000962.3) showing putative transcription start site determined from analysis of RNA-seq reads (Figure 2.3B). (F) RNA example originating from *Chlamydia trachomatis* with previously determined transcription start site (NC_010280.1/275170) (Albrecht 2010).

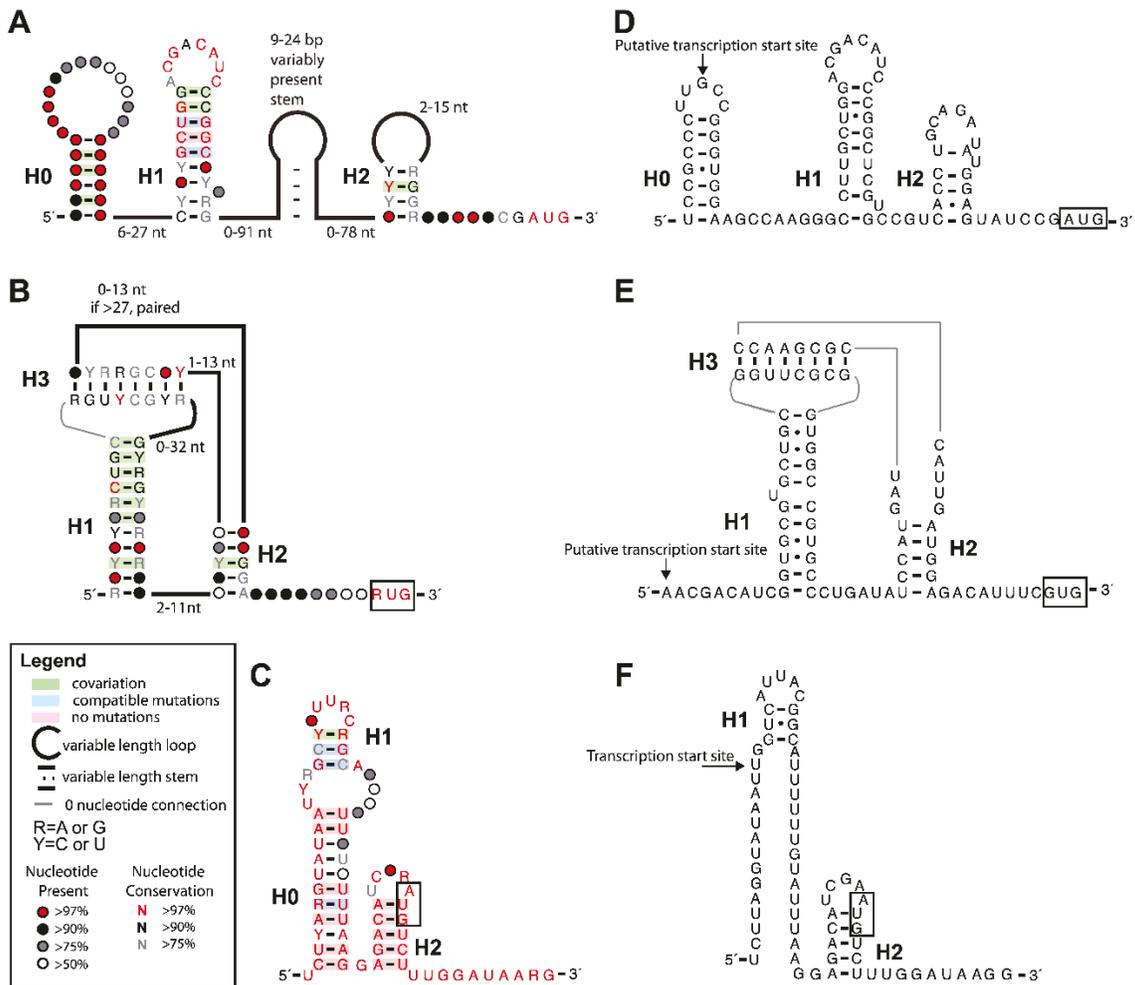


Figure 2.4 Nitrocellulose Binding Assays Identify Regions Important for Rra-S15 Binding

In vitro nitrocellulose filter binding assays confirm the transcription start site as well as indicate H1 as the region essential for Rra-S15 binding. (A) Truncation sites and specific mutation to the Rra-RNA. The start codon AUG is boxed, and a red bar is over the ribosome binding site. (B) 5' truncations, (C) 3' truncations, (D) Putative binding-site mutations. Each curve represents at least three independent replicates. For the purposes of comparison, the data from Rra-RNA1 was included in graphs B & D. Reported K_D measurements represent the protein concentration at which half of the maximum percentage of Rra-RNA is protein bound. Max% refers to the maximum percentage of Rra-RNA that interacts with Rra-S15 in this *in vitro* assay (see Methods for calculations).

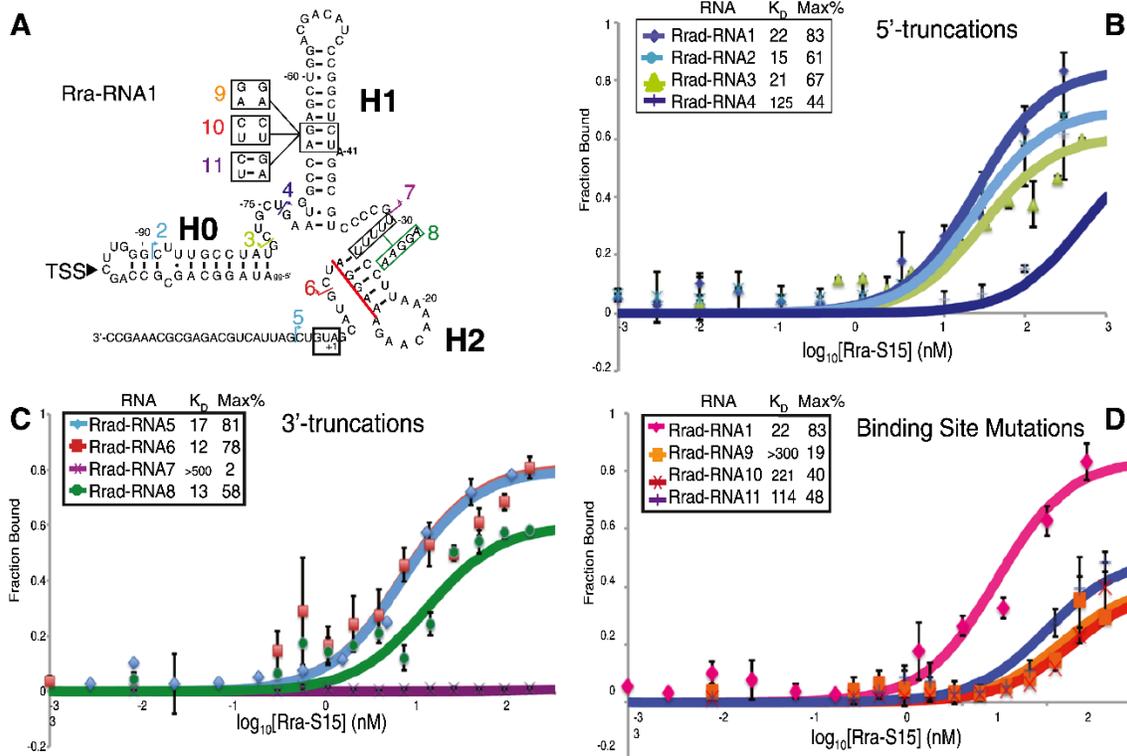


Figure 2.5 5'-RACE for *rpsO* Transcription Start Site in *R. radiobacter*

5'-RACE was performed to experimentally identify the transcription start site. 5'-RACE products were sequenced, then mapped to the *R. radiobacter rpsO* genomic region. The predicted mRNA regulatory sequence is underlined (Rra-RNA1 full length), protein coding nucleotides in gray, and individual nucleotides highlighted with colors based on the number of independent 5'-RACE products: 1 (yellow), 2 (red), or 4 (pink). The predicted transcription start site based on *R. sphaeroides* transcriptome data is indicated (large bold cytosine).

```
tcttcaattgcacccggaatagttatagcagcgcagCtgggcttgcctatgctgctgaatggccag
agctggacgacatcccggctctagcgtccccgttttccttaaacaagaaggatcgtacgATGTC
GATTACTGCAGAGCGCAAAGCCGCCCTCATCACGGAATATGCC
ACCAAGGCAGGCGACACCGGTTCTCCGGAAGTTCAGGTCGCAA
TCCTGACCGAGCGGATCAACAACCTGACCGGTCACTTCAAGGA
CCACAAGAAGGACAACCACTCCCGTCGTGGCCTTCTGACGCTC
GTTTCGAGCCGCCGTTTCGCTTCTCGACTATCTGAAGAAGAAGGA
CGAAGCCCGTTACACCAAGCTGATCGGTGCTCTCGGCATTTCGCC
GCTAA
```

Figure 2.6 *In vivo* Regulation Assays Validate Rra-RNA Regulates Gene Expression
GFP reporter assays validate the regulatory capacity of Rra-RNA in response to Rra-S15. Rra-RNA structure and sequence are the same as described in Figure 2.4. All relative fluorescence values were calculated by normalizing GFP/OD₆₀₀. All bars are 3+ independent experiments. * indicates p<0.01; ** indicates p<0.001.

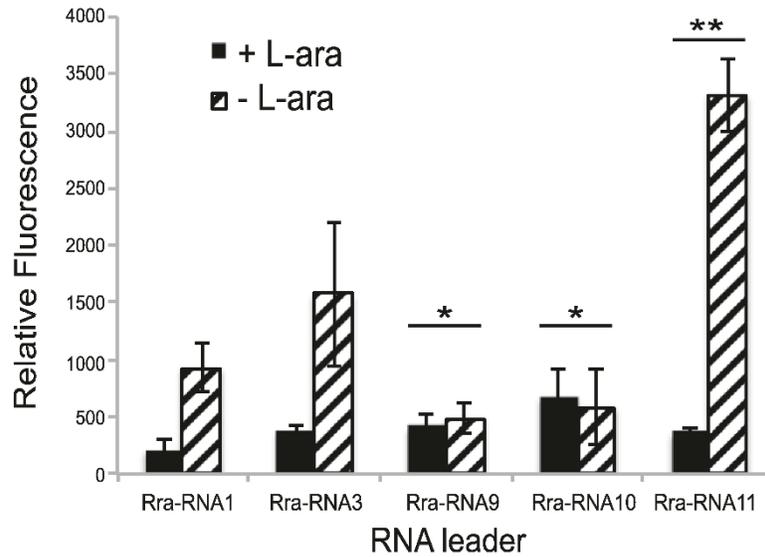
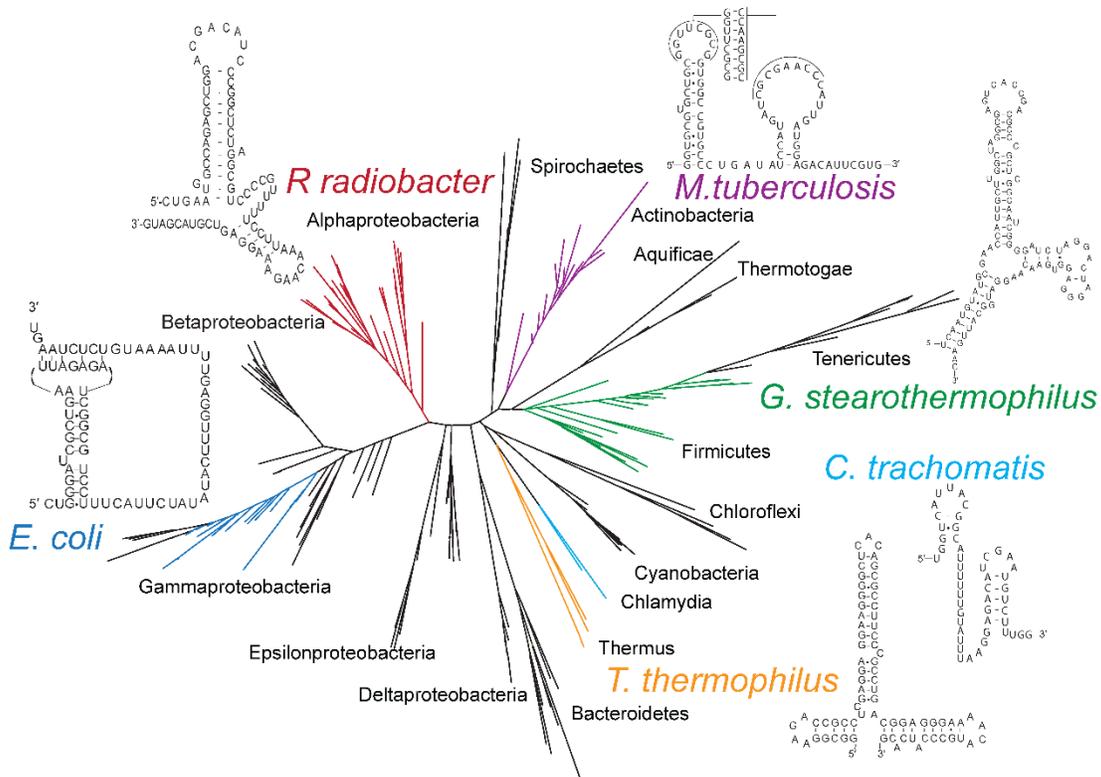


Figure 2.8 Updated Phylogenetic Tree for Regulatory RNAs of Ribosomal Protein S15

Figure 1.5 updated to include the three additional regulatory RNAs for ribosomal protein S15 identified in this study. *M. tuberculosis* is an example from Actinobacteria. *C. trachomatis* is an example from Chlamydia. *R. radiobacter* is an example from Alphaproteobacteria. Previously shown on Figure 1.5: *E. coli* is an example from Gammaproteobacteria, *T. thermophilus* is an example from Thermales, *G. stearothermophilus* is an example from Firmicutes.



Additional File 1: Alignment of Alphaproteobacterial RNA

Available at

<http://www.biomedcentral.com/content/supplementary/1471-2164-15-657-S1.txt>

Additional File 2: Alignment of Actinobacterial RNA

Available at

<http://www.biomedcentral.com/content/supplementary/1471-2164-15-657-S2.txt>

Additional File 3: Alignment of Chlamydia RNA

Available at

<http://www.biomedcentral.com/content/supplementary/1471-2164-15-657-S3.txt>

CHAPTER III

Co-evolution of Ribosomal Protein S15 with Diverse Regulatory RNA Structures

The work in this Chapter is based up on the following published journal article. Text, figures, and tables from this publication are used throughout this Chapter without additional notice.

Slinger BL, Newman H, Lee Y, Pei S, Meyer MM: Co-evolution of bacterial ribosomal protein S15 with diverse mRNA regulatory structures PLOS Genet. 2015 Dec 16;11(12):e1005720. doi: 10.1371/journal.pgen.1005720.

Author's contributions

Conceived and designed the experiments: BLS SP MMM. Performed the experiments: BLS HN YL. Analyzed the data: BLS HN YL SP MMM. Wrote the paper: BLS SP MMM

INTRODUCTION

RNA-protein interactions are vital to many cellular processes including ribosome assembly and gene regulation, yet how the evolution of one partner influences the evolution of the other partner remains unexplored. Describing the evolution of these partners is especially challenging when multiple regulatory RNA structures interact with homologous proteins to perform the same biological function. In contrast to the strikingly different regulatory RNA structures, S15's sequence and structure is highly conserved (Figure 1.3). In spite of this, previous studies and our own data demonstrate that a given S15 from one species does not interact with all mRNA regulatory structures. It is unknown what enables such a highly conserved protein to discriminate between these regulatory structures. We wished to better understand this species-specific regulatory interaction and pinpoint bases or residues implicated in this species-specific interaction.

Four RNA structures have been discovered and are narrowly distributed to Gammaproteobacteria, Thermales, Alphaproteobacteria, and Firmicute Phyla (Figure 1.5) (Chapter II, Slinger 2014, Serganov 2003, Scott 2005, Philippe 1990), Despite their structural diversity, all mRNA structures perform analogous *rpsO*-regulatory function by interacting with a conserved protein, S15. It is unknown how a conserved protein is able to recognize and regulate gene expression using such diversity of RNA structures. These natural mRNA-protein interactions provide us with an opportunity to explore whether the diverse RNA architectures present similar tertiary structure surfaces to the protein, or if the different S15 protein homologs have distinct RNA recognition profiles.

We improved upon our existing *in vivo* reporter assay to assess the RNA-S15 recognition among the different S15 homologs and their corresponding mRNA structures.

To summarize, we used a translationally fused β -galactosidase reporter to characterize biologically relevant regulatory interactions over a shorter time scale. Using this new assay as well as *in vitro* binding assays we assessed all cross-species interactions between the S15 homologs and RNAs from the four species in which there has been experimental validation. We find that the results of the regulatory assays and *in vitro* assays largely agree and together show that there are differences between S15 homologs that result in specific recognition of the diverse mRNA structures. Furthermore, we analyze the conservation of S15 amino acid sequences from species showing different recognition patterns and identify amino acid mutations responsible for these specificity changes. Together our results suggest that even highly conserved RNA-binding proteins may have distinct RNA recognition profiles, and that co-evolution has occurred between bacterial S15 homologs and their respective mRNA regulators.

RESULTS AND DISCUSSION

***LacZ in vivo* Reporter Assay Validates mRNA-S15 Regulatory Interactions**

To assess whether the S15 homologs are able to recognize and use the diverse RNA architectures present in these four different species, we improved our existing *in vivo* reporter assay system. The mRNA-GFP reporter plasmid and was sufficient for validation of the *R. radiobacter* mRNA-S15 interaction (Chapter II). However, to more accurately quantify an S15-mRNA interaction among different homologs and using mRNAs with differing expression levels, some improvements were implemented.

First, the GFP reporter gene displays low overall expression levels in the *E. coli* host organism, which ultimately affects the maximum fold-repression measurable using the GFP assay. To fix this we altered our reporter plasmid, swapping the GFP reporter gene with a *lacZ* reporter gene (endogenous to *E. coli*), forming an mRNA-*lacZ* fusion in the reporter plasmid. The measurable range of gene expression was much improved in comparison to GFP expression in our hands. Additionally, measuring enzymatic activity enables measurement over a larger dynamic range.

A second issue with the GFP assay was the long period of time between mRNA-GFP induction and subsequent fluorescence measurement. Our previous assay optimizations showed that the GFP fused to our RNA leader sequences required roughly 48 hours to measure maximal fluorescence in the absence of protein (data not shown). To account for this in GFP assay conditions, we measured GFP fluorescence after growing the bacteria to stationary phase then storing the bacteria at 4°C for 2 days (~48 hours) in PBS in a 96-well plate. There are many issues with this treatment of the cells that may also have ultimately affected our fluorescence measured. For example, the overall protein (both S15 or GFP) stability and degradation may have been impacted. Additionally, overall bacteria viability was likely affected due to subsisting such a long period of time under non-optimal stationary phase conditions. To fix this issue, we not only swapped the GFP reporter for a *lacZ* gene as mentioned previously, but we modified the mRNA-GFP reporter to be tightly inducible by swapping the *ptrc* promoter with a pLAC promoter. This change enabled us to alter the experimental design of the *in vivo* regulation assay to measure instantaneous reporter expression after a much shorter period of reporter

induction (30 minutes). This promoter swap made the new mRNA- *lacZ* reporter plasmid, named pBS2-RNA, IPTG-inducible.

A third issue with our GFP *in vivo* regulation assay was the strain used in the study. Subsequent analysis of the *E. coli* strain CK1953 used for the study did not confirm the strain was $\Delta rpsO$ as advertised. An *E. coli* $\Delta rpsO$ strain was highly desired to eliminate endogenous Ec-S15 expression, which would reduce background riboregulatory interactions with our RNA-*lacZ* reporter. To fix this issue, we endeavored to generate our own $\Delta rpsO$ *E. coli* strain. The methodology is outlined below; however, it was abandoned because we were able to obtain an *E. coli* $\Delta rpsO$ organism from the Culver Lab (University of Rochester, Rochester, NY).

We used λ -recombineering methodology (Datta 2006) to insert the kanamycin resistance gene (Kan^R) in place of both *rpsO* and its leader sequence in the *E. coli* K12 NCM strain (CGSC#8256). This strain was chosen for several genetic reasons. It lacks endogenous *lacZ*, therefore, *lacZ* could be used on our RNA reporter plasmid with little to no background expression. Second, the strain was $\Delta araD-araB$, ensuring any L-arabinose added to the media to induce S15 expression by the pBAD33 vector would not instead be metabolized by the bacterium. Finally, the NCM strain contains the *lacI*^Q allele of the lac operon repressor protein. This ensures our reporter plasmid, pBS2-RNA, would be IPTG-inducible.

We utilized a gene knockout technique used twice previously by other groups to successfully generate an *E. coli* $\Delta rpsO$ strain (Mathy 2004 and Bubunencko 2006), the λ -red recombineering. First, the pKD46 plasmid was transformed into our desired NCM strain. This plasmid contains and expresses the recombination machinery under the

control of the pBAD promoter. Cells were grown to early log phase, L-arabinose added to induce expression of recombination genes, then electroporation was performed to transform the bacterium with linear PCR product containing integration sequence of interest. In our case, the linear PCR product was designed to contain *rpsO* homology sequences flanking the Kan^R coding sequence. Transformants were plated on kanamycin to screen for Kan^R genomic integration, then single colonies screened via PCR. Several attempts were performed using variations of those conditions, different temperatures, different amounts of linear PCR product, and several different *rpsO* integration sites were designed into our Kan^R amplification primers. Six attempts were unsuccessful at generating an *E. coli rpsO*-knockout organism using this method. A final seventh attempt using single stranded linear PCR product was also unsuccessful.

Because we were unsuccessful in creating an S15-knockout organism, we hypothesized that the deletion is lethal. Swapping the native *Ec-rpsO* promoter with an inducible Tn10 tetR repressor operator sequence would create a conditional knockout organism (Grkovic 2002). It was this methodology that we next attempted using λ -red recombineering. We altered the design of the linear PCR product to contain Tn10-Kan^R flanked by *rpsO* integration sequences. When integrated, this product was designed to replace the mRNA leader and promoter sequence of the *rpsO* gene. Standard λ -red recombineering technique was applied to the NCM strain transformed with pKD46; however, three attempts using this new linear PCR product were unsuccessful. Our methodology and screening process need optimization before the *rpsO* gene of the *E. coli* NCM strain can be replaced with Kan^R under our laboratory procedures and conditions.

In the process of attempting to generate our own *rpsO* knockout organism, the Culver lab generously shared the $\Delta rpsO$ organism generated in their lab (*E. coli* $\Delta rpsO$, Bubunenko 2006). In their knockout organism the *rpsO* gene deletion was from the ATG start codon through the first two bases of the stop codon, inserting the Kan^R gene. We confirmed this via PCR analysis of the genomic locus (Figure 3.1A-D). However, this organism lacks the strong pLAC repressor allele, *lacI^Q*, rendering our reporter plasmid, pBS2-RNA, constitutively expressed. Additionally, this knockout organism contains a genomic copy of the *lacZ* gene. A final modification was made to the reporter plasmid by cloning the *lacI^Q* gene and promoter sequence, creating pBS3-RNA (Figure 3.2). This modification made the *lacZ* reporter IPTG-inducible in our newly acquired $\Delta rpsO$ K12 strain and greatly reduced background expression of *lacZ* off the genome (Figure 3.3). The pBS3-RNA plasmid, the K12: $\Delta rpsO$ strain, and the new experimental conditions were used for all subsequent *in vivo* regulation assays throughout this thesis.

Regulation Assays Confirm Native mRNA-S15 Interactions

To identify whether S15 homologs can specifically recognize different mRNA architectures to allow regulation within the cell, we utilized our updated β -galactosidase reporter assay. This functional assay directly tests the regulatory interaction between an mRNA and ribosomal protein S15 and enables the mRNA to fold into a biologically relevant structure. To recap, one plasmid contains an mRNA-*lacZ* fusion (pBS3-RNA) that was constructed by cloning the 5'-UTR through the first 5-9 codons of *rpsO* in-frame with *lacZ* and downstream of an IPTG-inducible promoter. A second plasmid (pS15) includes a full-length *rpsO* open reading frame (encoding S15) under the control of the

pBAD33 L-arabinose inducible promoter. The plasmids have compatible replication origins and different antibiotic markers allowing them to be stably maintained in the same bacterium. For regulatory assays the plasmids are co-transformed into an *E. coli* K12: $\Delta rpsO$ strain that lacks endogenous S15 (Bubunenko 2006) (Figure 3.1A-D).

Cells containing a pRNA and a pS15 are grown with and without L-arabinose, and at stationary phase the reporter is induced for 30 minutes with the addition of IPTG. Subsequently, the β -galactosidase activity within + and – L-arabinose cultures started from a single colony are compared to indicate whether a given mRNA structure enables S15-dependent regulation of *lacZ* expression. Given the short induction time during stationary phase, we did not observe any noticeable growth changes upon induction of individual mRNA reporter constructs. The four experimentally validated riboregulators and their respective S15 homologs from *Escherichia coli* (Ec-mRNA, Ec-S15), *Geobacillus kaustophilus* (Gk-mRNA, Gk-S15), *Thermus thermophilus* (Tt-mRNA, Tt-S15), and *Rhizobium radiobacter* (Rr-mRNA, Rr-S15) were each examined using the β -galactosidase reporter assay (Figure 3.4A-D). *G. kaustophilus* is a close relative of *G. stearothermophilus* and predicted to have the same RNA structure (Deiorio-Haggart 2013). Each of the S15 homologs complemented this strain, enabling much faster growth when protein expression was induced (Figure 3.4E).

We confirmed all native mRNA-S15 regulatory interactions (Figure 3.4F) by directly comparing fold repression of pS15 to pBAD33 with no insert (pEMPTY). In each case we find that the native regulatory interaction can be detected using our assay in the surrogate organism. However, the unregulated levels of β -galactosidase expression using each mRNA riboregulator affects the resulting fold-repression (Figure 3.5). The

Ec-mRNA showed the highest β -galactosidase activity (~5,000-10,000 Miller Units) whereas the remaining mRNAs tested were all within a similar range (~1000-2000 Miller Units). To further ensure the significance of our observed interactions, a mutation abolishing the native binding interaction was introduced into each mRNA. In each case repression was reduced, typically to levels comparable to that observed for pEMPTY (~2-fold), although the Tt-mRNA-M1 does retain some regulatory activity (Figure 3.4).

Regulation Assays Reveal Specific RNA Recognition Patterns

To determine whether the distinct mRNA architectures contain a shared tertiary structure or binding motifs, we examined all inter-species interactions using our regulatory assay. These results show that each mRNA structure has a specific set of S15 homologs to which it responds. For the mRNA regulator from *E. coli*, Ec-mRNA, both Rr-S15 and Tt-S15 successfully regulated β -galactosidase expression, yet do so more modestly than its native binding partner, Ec-S15 (Figure 3.6A). The mutation abolishing the native RNA-protein interaction (Ec-mRNA-M1, derived from Philippe 1995) deregulated reporter expression in response to both Rr-S15 and Tt-S15. Gk-S15 did not regulate the Ec-mRNA or its mutant. These results suggest that these three S15 homologs, Ec-S15, Tt-S15, and Rr-S15, interact with this mRNA in a similar fashion to regulate gene expression. The inability of this mRNA to respond to Gk-S15 suggests that this homolog requires a regulatory motif or structure not found in Ec-mRNA.

In contrast, the mRNA from *R. radiobacter*, Rr-mRNA, regulates gene expression in response to all the S15 homologs (Figure 3.6B). A mutation to Rr-mRNA in the main stem was sufficient to deregulate expression in response to Rr-S15 and Ec-S15 (Chapter

II, Figure 2.6). However, this mutation did not impact the convincing regulation observed in response to Tt-S15 and Gk-S15 (>10-fold repression observed). This suggests that the Ec-S15 and Rr-S15 homologs utilize similar determinants to recognize the mRNA, but that the Gk-S15 and Tt-S15 homologs may be recognizing alternative motifs that are not impacted by the mutation.

The mRNA from *T. thermophilus*, Tt-mRNA, displayed regulatory activity in response to all the S15 homologs (Figure 3.6C). A mutation to the three helix junction (3HJ) (derived from Serganov 2003) diminishes Tt-mRNA's response to Tt-S15, Ec-S15, and Gk-S15 homologs. However, this mutation does not completely abolish regulation in response to Tt-S15, Ec-S15 and Gk-S15, and had no effect on regulation in response to Rr-S15. These results have two potential interpretations. First, Tt-S15, Ec-S15, and Gk-S15 proteins may recognize Tt-mRNA in a different manner than Rr-S15, and therefore a mutation to the binding site for Tt-S15 at the three helix junction may not impact binding and regulation in response to Rr-S15. A second explanation is that the relatively modest 6-fold regulation observed for Rr-S15 is an artifact of our regulatory assay.

The mRNA from *G. kaustophilus*, Gk-mRNA, is also responsive to all S15 homologs tested (Figure 3.6D). Like the Rr-mRNA, the convincing regulatory responses to Ec-S15, Rr-S15, and Tt-S15 were not abolished by the mutation to Gk-mRNA (a truncation used during *in vitro* studies in Scott 2001 expected to disrupt the 3HJ), while regulation in response to Gk-S15 was abolished by this mutation. Like the Rr-mRNA, these data suggest that the binding determinants for Tt-S15, Ec-S15, and Rr-S15 on Gk-mRNA are different from those of Gk-S15, and that different S15 homologs may utilize distinct features to recognize the same mRNA.

Together, the regulatory assays show that there is extensive, but not universal cross-reactivity in the inter-species mRNA-S15 regulatory interactions. However, results obtained with mRNA mutants suggest that even mRNAs recognized by multiple S15 homologs may be recognized using different determinants. In particular, for both the Gk-mRNA and the Rr-mRNA, mutations that abolish native interactions have little or no impacts on interactions with other S15 homologs.

***In vitro* Binding Assays Show Distinct Recognition Profiles for S15 Homologs**

Given that many of our mutations that abolish native interactions still allowed regulation in response to other protein homologs, we also used *in vitro* nitrocellulose filter-binding assays to directly measure the strength of RNA-protein binding interactions to corroborate our findings. All four S15 homologs were purified and nitrocellulose filter binding assays were performed for all cross-species interactions. We find that the dissociation constants for native interactions are in the 2-20 nM range. However, the native interactions were not always the strongest interactions. For example, Gk-S15 bound Tt-mRNA with an affinity that was almost an order of magnitude smaller than Tt-S15 (0.35 nM vs. 2.11 nM).

We were unsuccessful in demonstrating Ec-mRNA interactions with any S15 homolog including its native binding partner; therefore, it was omitted from further study. The native Ec-mRNA interaction with Ec-S15 has been characterized *in vitro* in the past (K_D = 231 nM, Serganov 2002). Notably, this value is significantly higher than those that we measured for the other native interactions. Although a 3'-terminal ^{32}P -pCp has been previously shown to decrease the K_D four-fold in truncated versions of this RNA

Serganov 2002, we found that labeling the full-length mRNA with ^{32}P -pCp did not change our result. We did not explicitly test the truncated RNA since we are primarily interested in the wild-type interaction. Previously it has been shown that *E. coli* uses an entrapment mechanism where the pre-initiation complex of the ribosome binds the *rpsO* transcript simultaneously with Ec-S15 (Philippe 1993, Philippe 1994). These binding assays were performed in the absence of additional purified ribosome components.

Aside from our inability to measure interactions with Ec-mRNA, we find that our *in vitro* findings closely follow the results of the regulatory assays. The Rr-mRNA was able to interact with all S15 homologs *in vitro*, and all are relatively strong interactions with dissociation constants ranging from 1 to ~30 nM (Figure 3.7). The inactivating mutation (Rr-mRNA-M1) abolished interaction with Ec-S15, but had little impact on interactions with the Gk-S15 or Tt-S15 homologs. These data corroborate our results from the regulatory assay indicating that Gk-S15 and Tt-S15 interact with the Rr-mRNA-M1, and further indicates that Ec-S15, Gk-S15, and Tt-S15 homologs use distinct features to regulate gene expression using this mRNA.

Tt-mRNA binds strongly to both Tt-S15 and Gk-S15, which corroborates our *in vivo* regulation findings. Conversely, Ec-S15 and Rr-S15 both do not bind Tt-mRNA *in vitro*, which makes interpreting the regulatory assay results less clear. They both displayed modest regulatory activity *in vivo*. Mutating Tt-mRNA decreased the regulatory response to Ec-S15, yet did not significantly impact the response to Rr-S15 (Figure 3.7). However, neither Ec-S15 nor Rr-S15 were able to bind this mutant *in vitro*. In comparison to Gk-S15 and Tt-S15, the dissociation constants measured for Ec-S15 and Rr-S15 tend to be significantly higher for all measured S15-mRNA interactions,

indicating that perhaps these proteins behave less well *in vitro*. Alternatively, relatively high levels of noise in our regulatory assay (even empty vector controls typically display 2-3 fold repression) may bias our findings. This may mean that Ec-S15 is able to regulate gene expression using the Tt-mRNA structure for the regulatory activity decreased with the mutated mRNA. However, whether Rr-S15 interacts with the Tt-mRNA to allow regulation remains unclear. In addition, although regulation of the Tt-mRNA by Gk-S15 is significantly reduced by the Tt-mRNA-M1 mutation, Tt-mRNA-M1 is not sufficient to completely abolish *in vitro* binding of Gk-S15. However, the measured K_D is over two-orders of weaker (0.35 nM vs 76.3 nM), and the maximum fraction bound by the protein is <20%, indicating that the *in vitro* interaction may be non-specific. To summarize, Tt-mRNA only binds Tt-S15 and Gk-S15 *in vitro* and only showed appreciable regulatory activity in response to both Tt-S15 and Gk-S15 *in vivo*; this suggests these are the only regulatory interactions that occur.

Gk-mRNA interacted with all four S15 homologs *in vitro* (Figure 3.7). The strongest interaction was with Tt-S15, roughly an order of magnitude stronger than the native Gk-S15 interaction, and roughly three orders of magnitude stronger than with Ec-S15 and Rr-S15. In addition, the Ec-S15 and Tt-S15 homologs retain strong interactions with the Gk-mRNA-M1. This suggests that the retained regulation for this mutant in response to Ec-S15 and Tt-S15 is because these homologs still bind the mutant mRNA. In addition, while we do not measure any interaction between Rr-S15 and the Gk-mRNA-M1 (up to 250 nM Rr-S15), the interaction between Rr-S15 and Gk-mRNA is relatively weak in comparison to the other S15 homologs (K_D ~200 nM). Therefore, Rr-S15 may bind Gk-mRNA-M1 weakly, yet this interaction is sufficient to regulate reporter

expression. In conclusion, our *in vitro* results with the Gk-mRNA suggest that the regulatory interactions we observed between Ec-S15, Rr-S15, Tt-S15 and the Gk-mRNA and its mutant (Gk-mRNA-M1) are indeed due to differences in the way that the proteins interact with the mRNA.

In summary, we find that measuring cross-species interactions between S15 homologs and diverse mRNA structures using both regulatory assays and *in vitro* binding assays shows that the two approaches largely agree. While in isolation each type of assay is prone to various artifacts ranging from poor *in vitro* binding properties, to likely differences in protein expression levels in the surrogate organism, the large extent of agreement between our two assays significantly strengthens our conclusions. Overall, we find that Tt-S15 and Gk-S15 bind very tightly *in vitro*. This may be due to many factors including that the Tt-S15 and Gk-S15 homologs are both thermophiles and may be more stable resulting in better *in vitro* binding characteristics. We also assessed Tt- and Gk-S15 *in vitro* binding at 55° C and found that no significant differences were detected at the higher temperature.

S15 Homologs Recognize mRNAs Via Distinct Motifs

To combine our *in vitro* and regulatory results into a single determination of whether or not an interaction occurs, we consider all measureable dissociation constants as viable interactions. For regulatory interactions, we consider all interactions that are significantly reduced by a mutation to the RNA, or corroborated by *in vitro* data as viable interactions (Figure 3.8A-F). From our collected data it is clear that there is extensive cross-reactivity, but that S15 homologs often recognize mRNAs using different

characteristics, as demonstrated by the very divergent responses of different S15 homologs to the mutated mRNA structures.

Using this criterion there are two ambiguous interactions with Tt-mRNA and Ec-S15 and Rr-S15. For the Tt-mRNA and Rr-S15 pairing, the mutation to Tt-mRNA did not affect the regulatory interaction, which could either be due to an artifact of our regulation assay, or indicate that Rr-S15 recognizes a portion of the Tt-mRNA not affected by the mutation. However, there was no *in vitro* binding interaction between Tt-mRNA and Rr-S15. In addition, we did not detect an *in vitro* interaction between Ec-S15 and the Tt-mRNA, although regulation was observed for this pairing (~ 10 fold repression), and it is reduced by the Tt-mRNA-M1, suggesting that it is not an artifact. The ambiguous regulatory results using the Tt-mRNA may be due to our *in vivo* conditions. Because this RNA originates from a thermophilic organism (*T. thermophilus* thrives at 75°C), it may be that our assay conditions at 37°C are not optimal for folding this RNA. Because of the clear results from the binding assay where neither Ec-S15 nor Rr-S15 binds Tt-mRNA, we do not believe a regulatory interaction occurs between Tt-mRNA and either Ec-S15 or Rr-S15.

Using the *in vivo* regulation and the *in vitro* binding data, we can start to assess what RNA structural motifs result in the different recognition profiles. The rRNA binding site for S15 is bipartite, consisting of a three helix junction (3HJ) and a G•U/G-C motif approximately one helical turn away from the 3HJ (Figure 1.2B, Figure 3.8A). Previous studies have established that the *E. coli* mRNA (Ec-mRNA) mimics the G•U/G-C motif (Serganov 2002), and that the *T. thermophilus* mRNA (Tt-mRNA) mimics the G-G-C base-triple found in the 3HJ of the rRNA (Serganov 2003) (Figure 1.6A, E). However, in

both of these cases it is clear that while the mRNA is contacted at a second position consistent with bipartite binding, the second position bears limited resemblance to the rRNA. In the case of Ec-mRNA, the second binding site occurs within the co-axially stacked pseudoknot (Serganov 2002) (Figure 3.8B), and in the case of the Tt-mRNA, the long H2 stem is necessary for binding, but the G•U/G-C motif is replaced by a G•G mismatch (Serganov 2003) (Figure 3.8D). In contrast, the Gk-mRNA appears to contain mimics of both binding determinants. The 3HJ is mimicked in the multi-stem junction and a G•U/G-C motif is apparent approximately one helical turn away from this junction (Scott 2005) (Figure 1.6C, 3.8E). In the case of the Rr-mRNA, far less data exist concerning which bases are necessary for binding. However, a G•U/G-C motif is apparent in the most conserved portion of the Rr-mRNA (Figure 2.2A), and like the Tt-mRNA, the junction of the stems is important for retaining interaction with its native binding partner (Figure 3.8C).

Taking our results in conjunction with previously published results, the data suggest that each of the mRNA structures mimics a portion of the rRNA. Both the Ec-mRNA, and Tt-mRNA contain a direct mimic for a portion of the binding site, while the Gk- and Rr- mRNAs likely contain both sections. The inactivating mutations for each of the mRNAs target different portions of these rRNA binding sites. Rr-mRNA-M1 and Ec-mRNA-M1 both target putative G•U/G-C motifs, the Gk-mRNA-M1 is a truncation that presumably disrupts the three helix junction, and the Tt-mRNA-M1 also targets the 3 helix junction. This partial mimicry of the S15 rRNA binding site potentially explains the regulatory differences we observe for the S15 homologs.

Our observations suggest that Ec-S15 and Rr-S15 preferentially recognize the G•U/G-C motif. Regulatory interactions between both Ec-S15, and Rr-S15 and several mRNAs are significantly impacted when this region is mutated (Ec-mRNA-M1, and Rr-mRNA-M1). The regulatory interaction between Ec-S15 and Rr-S15 does not appear to be impacted by Gk-mRNA-M1, a mutant targeting the putative 3HJ. Tt-mRNA lacks the G•U/G-C motif, and while Ec-S15 appears to regulate gene expression using Tt-mRNA, this interaction could not be reproduced *in vitro*.

In contrast, the Gk-S15 appears to preferentially interact with a mimic of the three-dimensional motif formed at the helical junction to regulate gene expression. Gk-S15 does not interact with the Ec-mRNA (lacks a mimic of the junction), it is not impacted by the Rr-mRNA-M1 mutation that targets the G•U/G-C motif, and mutations that impact the junction result in lack of regulatory activity (Tt-mRNA-M1, and Gk-mRNA-M1).

Finally, the Tt-S15 appears to regulate gene expression with any mRNA structure that contains either portion of the rRNA binding site. The Ec-mRNA and Tt-mRNA each contain an obvious mimic for a single portion of the rRNA binding site, and mutations to these regions prevent gene regulation in response to Tt-S15. The Gk-mRNA and Rr-mRNA are presumed to contain mimics of the entire rRNA binding site, and mutations that impact only one of these regions do not affect the regulatory interaction with Tt-S15. In summary, we propose that the four S15 homologs preferentially recognize different sections of the naturally occurring mRNA regulators.

To test our model for S15 interaction we constructed a second mutation of Gk-mRNA targeting the putative G•U/G-C motif (Figure 3.9A). We hypothesized that this

mutant should abolish regulation and binding of Ec-S15 and Rr-S15 for these homologs appear to recognize RNAs that contain the G•U/G-C motif. The Tt-S15 homolog appears to recognize all RNA structures, and our results suggest Gk-S15 requires a three helix junction to regulate gene expression, so this mutation may not affect the regulatory interaction with the Tt-S15 and Gk-S15 homologs.

The interaction between this mutant mRNA and all four S15 homologs was assessed using both our regulatory assay and *in vitro* binding assay. We find that this mutation indeed abolishes regulation of β -galactosidase expression in response to Ec-S15 and Rr-S15, and reduces regulation in response to Tt-S15 (Figure 3.9B). In addition, this mutation abolishes *in vitro* interactions with each of these proteins (Figure 3.9C). Gk-S15 weakly binds this mutant (the dissociation constant is nearly two orders of magnitude higher than that for the native interaction), but displays significant regulatory activity. These results are consistent with our proposal that while the G•U/G-C motif alone is not sufficient to enable interaction between Gk-mRNA and its native binding partner, it is sufficient to allow interactions between Gk-mRNA and the other three S15 homologs. For the Tt-S15 homolog, these results suggest the protein primarily recognizes the G•U/G-C motif in these mRNAs; however, it remains able to recognize the three helix junction and must because its own Tt-mRNA lacks the G•U/G-C motif. The near perfect sequence similarity between the three helix junction in Tt-mRNA and the rRNA may have evolved to increase the affinity with Tt-S15 and its mRNA due to the lack of a G•U/G-C motif. In summary, our results indicate that homologous proteins, even those that recognize the same RNA structures, do so using different structural determinants.

S15 Homologs Interacting with Non-homologous mRNA Regulators Have Different Conservation Patterns

Our specificity data as well as existing studies indicate that the determinants for mRNA and rRNA binding are distinct (Scott 2001, Mathy 2004). We hypothesize that, depending on the RNA regulator present in the organism, the positions in S15 under strong selection are different. Such positions may be responsible for mRNA as opposed to rRNA recognition. To explore this hypothesis, we analyzed the *rpsO* coding sequences from sequenced microbial genomes containing each class of mRNA regulator. For the *E. coli*, *G. kaustophilus*, and *R. radiobacter* RNAs there are high-quality RNA alignments that provide a list of genomes containing each mRNA regulator (Fu 2013, Deiorio-Haggar 2013, Slinger 2014). For each class of RNA regulator, we constructed alignments of the corresponding S15 protein coding sequences, which we will refer to by their species type (e.g. alignment of S15 sequences from organisms containing homologs of the Ec-mRNA will be referred to as the Ec-alignment). S15 is typically well-conserved and the alignments contain few if any gapped regions. The Gk-alignment was the largest at 202 sequences; the Ec-alignment had 165 sequences, and the Rr-alignment 65 sequences. In the case of the *T. thermophilus* mRNA regulator, no RNA alignment exists, and a cursory BLAST search did not return hits to the mRNA outside the Thermus genus. Both the Rr-S15 and Tt-S15 were omitted from further analysis due to the limited sequence alignments that could be constructed for them. In addition, this choice allows us to focus on the differences between Gk- and Ec-S15, which display very different RNA interaction behaviors based on our data. Previous mutagenesis studies for both Ec-S15

and Gk-S15 suggest they use similar, but not exactly the same, residues in recognition of their mRNA and rRNA (Scott 2005, Serganov 2002, Mathy 2004) (Figure 1.2, 1.6).

To systematically assess which positions might be under selective pressure, we used the tool Rate4site to evaluate each of the alignments (Pupko 2002). Rate4Site returns a Z-score for each position indicating the extent of conservation. Statistical significance of the Z-score depends on the overall extent of conservation over the entire protein sequence. Therefore, due to the small size and the high degree of conservation in our alignments, no site had statistically significant Z-scores (even those that are completely conserved). However, the Z-score may be used as a rough indicator of conservation (Figure 3.10A). There are many positions that are strongly conserved (Z-score < -0.1), however most of these have the same amino acid conserved in both alignments (e.g. position 28, which is a strongly conserved glutamine) (Figure 3.10B). Positions 2, 40, 58, and 61 show evidence of strong conservation of different amino acids in the two alignments (e.g. at position 2 an alanine is conserved in the Gk-alignment, but a serine in the Ec alignment). Positions 9, 18, 71, 72, 73, and 79 are strongly conserved in one alignment, but highly variable (Z-score > 0.5) in the other (e.g. position 18 is a conserved histidine in the Gk-alignment, but quite variable in the Ec-alignment). Additionally, both the N- and C-termini of the proteins show high degrees of variability in both alignments compared with the central portion that is expected to make direct contacts with the RNA.

Mutated *G. kaustophilus* S15 Shows Altered Specificity

To determine whether the positions identified above contribute to our observed interaction specificity (Figure 3.9), we focused on the Ec- versus Gk-alignment differences, with the goal of identifying amino acid changes that would enable Gk-S15 to recognize and regulate gene expression of its 3HJ-mutant (Gk-mRNA-M1), or Ec-mRNA, both of which had no regulatory activity with Gk-S15. Several of the positions identified are not expected to contact the RNA based on structural data (positions 2, 4, 9, and 79) (Agalarov 2000), or are the same in the Ec-S15 and Gk-S15 sequences (position 73) (Figure 3.9C). Therefore, we assessed whether Gk-S15 carrying the sextuple mutation to positions H18D, N40Q, K58R, G61S, R71K, and K72R (Gk-S15-6MUT) would regulate gene regulation with Ec-mRNA, or Gk-mRNA-M1.

We find that Gk-S15-6MUT is capable of regulating gene expression with both Gk-mRNA-M1 and Ec-mRNA (Figure 3.10D, E). Furthermore, this interaction appears to be specific as it is abolished in the Ec-mRNA-M1. This result suggests that one or more of the altered positions are responsible for recognition of these mRNA structures (Figure 3.10D, E). We speculate these residues contribute to higher affinity recognition of the G•U/G-C motif or possibly play a role in stabilizing a secondary binding site on the mRNA, independent of the G•U/G-C. When tested with Gk-mRNA-M2, Gk-S15-6MUT retains significant regulatory activity, evocative of that displayed by the Tt-S15 (Figure 3.9B). Our results suggest that Gk-S15-6MUT still recognizes the 3HJ, and the presence of either motif is sufficient to allow gene regulation (Figure 3.10D, E).

To further assess whether the diversity present in the N- and C-termini of the protein play a significant role in recognition, we also created a series of chimeric proteins

for Gk-S15 and Gk-S15-6MUT where the N- and C-terminal sections were swapped from Gk-S15 to Ec-S15 (Figure 3.11A-C). From these studies we found that the N-terminal residues from Ec-S15 typically decreased the extent of regulation across the board (Figure 3.11D). This could be due to several factors including potential deleterious interactions between the N-terminus and other portions of the protein structure (the N-terminus represents 12 changes between Ec-S15 and Gk-S15), as well as differences in protein expression levels. The N-termini of protein coding sequences have been implicated in the past in determining expression levels (Plotkin 2010, Tuller 2010, Gu 2010, Bentele 2013). We also found that chimeras with swapped C-terminal portions behaved very similarly, likely due to the small number of amino acid changes (three) between the two sequences. In summary, the N- and C-terminal regions of the protein are unlikely to play a large role in mRNA recognition and gene regulation, but do impact the extent of regulation observed in our regulatory assay due to alterations in effective protein concentration.

CONCLUSION

The goal of this study was to assess how the differences between S15 homologs may contribute to the diversity of mRNA regulators that arise across different bacterial phyla to allow gene regulation. This work shows how the rRNA binding site for S15 may be partially mimicked in the four different mRNA regulators. We demonstrate that S15 homologs have distinct RNA binding profiles, and that even when recognizing the same RNA, different homologs may be using distinct sequence features. These results suggest

that either S15 has co-evolved with its mRNA regulators, or that differences between the ancestral S15 proteins lead to the development of a diverse array of RNA regulators that we observe in nature today.

MATERIALS & METHODS

S15 Knockout Plasmid Construction and Linear PCR Product Amplification

The pKD46 (ATCC) plasmid containing recombinase genes was transformed into *Z*-competent *E. coli* K12 NCM strain. The Kan^R with *rpsO*-homology sequences was PCR amplified using plasmid pKD4 template and Phusion Polymerase (Combinations of 1155F, 1173F, 1175F, 1156R, 1174R, 1176R, Table of Primers, Supplemental File S15 KO Primer Designs).

Tn10 primers contained sequences to integrate upstream of *rpsO* (726R) or assemble with KanR (1196F). Taq was used to amplify Tn10 from BL-21 (DE3) *E. coli* genomic DNA (NEB). Phusion was used to amplify Kan^R using pKD4 template and primers containing a Tn10-assembly sequence (1195R) or *rpsO*-integration sequences (1155F, 1173F, 1175F). Full-length Tn10-KanR linear PCR product was amplified using assembly PCR and Taq Polymerase.

CRISPR-Cas9 plasmids and linear PCR products. Plasmid pKD46-Cas9 (kind gift from Jiang 2013) was transformed into the *E. coli* K12 NCM strain made *Z*-competent. Plasmid pCRISPR was modified to contain spacer sequence matching the *rpsO* genomic locus (Addgene pDB129). Primers contained the correct CRISPR spacer sequences targeting the *rpsO* region as well as BsaI digest sites. T4 Polynucleotide Kinase was used to phosphorylate oligonucleotides. Quick Ligase was used to ligate annealed oligonucleotides with BsaI-digested pCRISPR. CRISPR editing sequence was generated by amplifying Tn10 sequence using Taq polymerase on *E. coli* Xl-1 genomic DNA. Primers contained homology flanking sites to CRISPR spacer sequences.

pBS3-RNA Plasmid Construction

The pRNA plasmid was constructed by modifying the reporter plasmid ptrc-Ec-mRNA-GFP from (Chapter II, Slinger 2014). First, the ptrc promoter was replaced with the plac promoter. Complementary oligonucleotides of the lac promoter sequence flanked by the cohesive ends corresponding to a XhoI site (5') and a EcoRI site (3') were phosphorylated using T4 Polynucleotide Kinase, annealed, then ligated into ptrc-EcmRNA-GFP digested with XhoI and EcoRI using Quick Ligase. Second, the *lacZ* gene was amplified from *E. coli* genomic DNA using Phusion DNA polymerase and primers containing restriction sites Sall and XbaI. The PCR product was digested and ligated into ptrc-RNA-GFP digested using the same enzymes (GFP was excised in this process). This new plasmid, pBS2-Ec-RNA, was sequence verified. Finally, the lac repressor coding sequence (*lacI^Q*) was cloned into pBS2-Ec-RNA at the XhoI site. The *lacI^Q* gene flanked by XhoI sites was amplified from *E. coli* genomic DNA (Strain NCM534, K12 derivative, Yale *E. coli* Genetic Stock Center #8256) using Taq DNA polymerase to generate pBS3-RNA. The plasmid sequence was verified by Sanger sequencing.

All mRNA sequences were cloned into the pBS3-RNA plasmid as a translational fusion with *lacZ* using primers containing EcoRI and Sall restriction sites (See Figure 3.2 for overview of plasmid, and Figure 3.12 for list of primers). Translational fusions were constructed such that the first 9 amino acids originating from *E. coli* or *R. radiobacter rpsO*, 5 amino acids from *T. thermophilus rpsO*, or 4 amino acids from *G. kaustophilus rpsO*, were appended to the N-terminus the *lacZ* sequence. The *lacZ* sequence requires a start codon from the fused *rpsO* sequence. All enzymes for molecular biology were purchased from New England Biolabs unless otherwise noted. Mutations to the mRNAs were constructed by site-directed mutagenesis (Figure 3.12).

pS15 protein expression plasmids were constructed by amplifying the *rpsO* open reading frame from genomic DNA with a forward primer containing SacI site plus a strong ribosome binding site that matched the *E. coli* ribosome binding site preceding *rpsO* and an 8 nucleotide linker (Figure 3.12) preceding the *rpsO* start site and subsequent codons. The native ribosome-binding sites preceding *rpsO* from both *G. kaustophilus* and *T. thermophilus* were tested, however, these did not allow sufficient protein production to complement the knockout *rpsO* strain and were consequently abandoned. The reverse primer contained an XbaI site. After digestion, the PCR product was cloned into the pBAD33 vector (ATCC 87402) digested with the same enzymes. All pS15 were sequence verified. The Gk-S15-6MUT sextuple mutant was created using site-directed mutagenesis with primers listed on Figure 3.13 and chimeras created by PCR assembly using pEc-S15, pGk-S15, or pGk-S15-6MUT as template DNA.

Growth Assay

K12: Δ *rpsO* *E. coli* cells were transformed with a pS15 and a single colony picked to grow cultures +/- 15 mM L-arabinose for ~16 hours in LB + 34 ug/mL chloramphenicol. Cultures were diluted to OD₆₀₀ = 0.01 in 0.5 mL of fresh medium 24-well plates, and OD₆₀₀ was measured for 27.5 hours. Each pS15 was performed 3+ replicates. Doubling times were calculated by taking the inverse of the slope of ln (OD₆₀₀) in exponential phase readings.

***LacZ* Regulatory Assays**

K12: $\Delta rpsO$ *E. coli* cells (kind gift from Gloria Culver, Bubunenko 2006) were co-transformed with pRNA and pS15 plasmid (made competent using the Z-competent buffer system, Zymo Research). Although this strain does contain a chromosomal copy of *lacZ*, we find that it is significantly repressed by the *lacI^Q* allele present on our reporter plasmid such that the background levels of beta-galactosidase expression from the native *lacZ* are < 10-20% of those that we observe from our reporter carried on a multi-copy plasmid (Figure 3.3, 3.5). However, no doubt some of the experimental variation and background that we observe is due to this additional copy. For our assays, a single colony was used to start overnight cultures, grown +/- L-arabinose (15 mM) at 37°C, then diluted the next day to OD₆₀₀ = 0.15 in fresh media (LB + 100 ug/mL ampicillin + 34 ug/mL chloramphenicol +/- 15 mM L-arabinose). At stationary phase (5 hours after dilution) 1 mM IPTG was added to induce β -galactosidase expression. After 30 minutes, 100 ug/mL spectinomycin was used to stop initiation of protein translation, and the cultures assayed immediately according to Miller (Miller 1992) to determine the levels of reporter expression. Fold repression = (Miller units of - L-arabinose)/(Miller units of + L-arabinose). All RNA/S15 combinations were examined with 3+ independent replicates. To determine the significance, all fold repression values were compared as indicated in Figure 3.13 (data on Figure 3.4 and 3.6) and Figure 3.14 (data on Figure 3.9 and 3.10) using a Welch's single-tailed T-test in Microsoft Excel. Regulation was considered biologically significant if greater than 3-fold repression was observed, and the fold-repression was significantly different (p<0.05) than that observed with an empty pBAD33 vector.

RNA Preparation

DNA corresponding to the 5'-UTR of the *rpsO* gene was PCR amplified using species-specific primers with the T7-promoter sequence added within the forward primer sequence. Genomic DNA extracted from each species was used as template. Indicated mutations were inserted to a DNA sequence using PCR primers containing the mutation. T7 RNA polymerase (Milligan 1987) was used to transcribe RNA and transcription reactions were purified by 6% denaturing PAGE. Bands were visualized using UV shadow, excised, and the RNA eluted (in 200 mM NaCl, 1 mM EDTA pH 8, 10 mM Tris-HCl pH 7.5) and ethanol precipitated. Purified RNA (10 pmol) was 5'-labeled with ³²P-ATP and purified as previously described (Regulski 2008). pCp labeling was performed using T4 RNA ligase with 50 pmol RNA and 50 pmol of ³²P-pCp. 3'-labeled RNA was isolated using Ambion MEGAclean kit.

Protein Preparation

The *rpsO* open reading frame was PCR amplified using whole genomic DNA and species-specific primers. It was cloned into pET-HT overexpression vector similarly to previously described (Block 2011). Sequence verified plasmid was transformed into chemically competent BL-21 cells (DE3). Protein expression and purification for all four S15 homologs was conducted as described previously (Chapter II, Slinger 2014).

Nitrocellulose Filter-Binding Assays

A fixed amount of 5'-³²P-labeled RNA (1000 cpm, <1 nM) was renatured for 15 minutes at 42°C, then incubated with serial dilution of S15 in Buffer A (50 mM Tris-Acetate, pH 7.5, 20 mM Mg-acetate, 270 mM KCl, 5 mM dithiothreitol, 0.02% bovine

serum albumin (Philippe 1994) for 30 minutes at 25°C. For RNAs originating from thermophilic organisms, assays were also conducted at 55°C, but these either did not yield a productive interaction, or the results were not significantly different from those observed at 25°C. Nitrocellulose membrane (GE Healthcare) was used to collect RNA-S15 complexes and positively charged nylon membrane (GE Healthcare) was used to collect unbound RNA under suction in a filter binding apparatus. Membranes were air-dried 5 minutes and the fraction bound quantified by imaging membranes on a phosphorimager screen. Radioactivity counts per sample on each membrane were measured using GE Healthcare STORM 820 phosphorimager and ImageQuant. For each sample the fraction bound (Fb) corresponds to the (counts nitrocellulose)/(counts nitrocellulose + counts nylon). To determine the K_D and the maximum fraction bound (Max%), the resulting values were fit to the equation: $Fb = (Max\% * [S15]) / ([S15] + K_D)$ where [S15] corresponds to the concentration of S15 in the reaction. The residuals were minimized using the Solver function in Microsoft Excel to find both the Max% and the K_D . K_D values given in Figure 3.7 represent the mean of 3 or more independent binding assays \pm the standard deviation.

S15 Sequence Analysis

Amino acid sequences corresponding to the *rpsO* open reading frame from all bacterial species carrying each mRNA regulator were gathered based on existing RNA alignments (Fu 2013, Deiorio-Haggard 2013, Slinger 2014, Chapter II). These sequences were aligned using ClustalW (Thompson 1994), and the alignments analyzed using Rate4site (Pupko 2002).

FIGURES, & LEGENDS

Figure 3.1 *E. coli* $\Delta rpsO$ Confirmation

Confirmation of *E. coli* $\Delta rpsO$. (A) Diagram depicting genomic region of *rpsO* in *E. coli*, flanked by genes *pnp* and *truB*. Arrows and numbers indicate primers and primer placement. (B) *rpsO*-specific primers used with either *E. coli* $\Delta rpsO$ (Δ), *E. coli* XI-1 (XI1), or no template (no), then products separated using 1% agar and visualized using ethidium bromide (C) PCR product was generated from $\Delta rpsO$ strain (Δ), *E. coli* XI-1 strain (XI1), or no template (no) using the primer sets indicated (D) Individual colonies of the *E. coli* $\Delta rpsO$ strain ($\Delta 1$ - $\Delta 8$) were PCR checked using primers 739+740 to confirm replacement of *rpsO* with *kanR*. *E. coli* strain XI-1 (XI1) and no template (no) were amplified at the same time for size and condition controls.

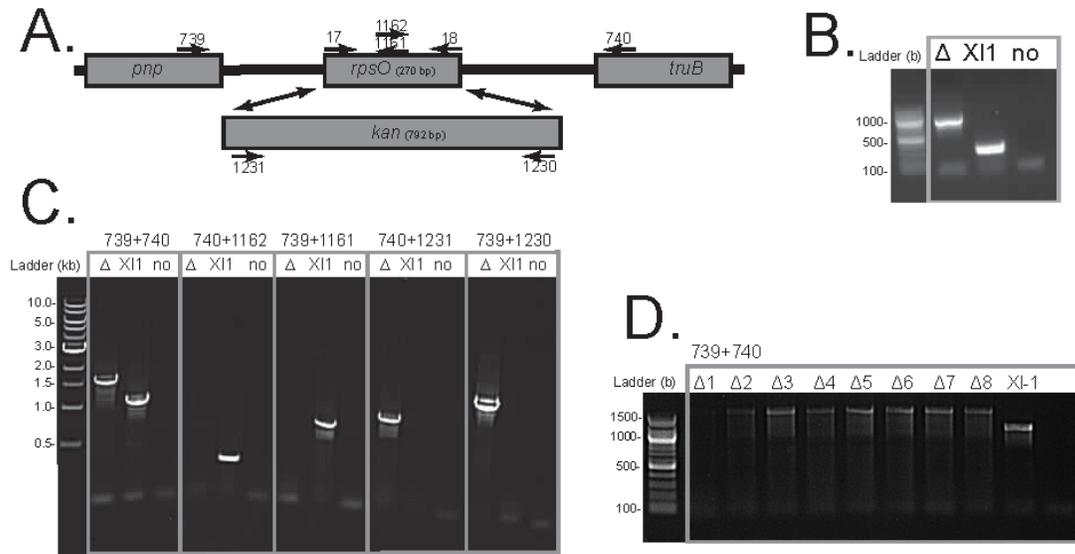


Figure 3.2 pBS3-RNA Plasmid Diagram
pBS3-RNA plasmid diagram (not drawn to scale).

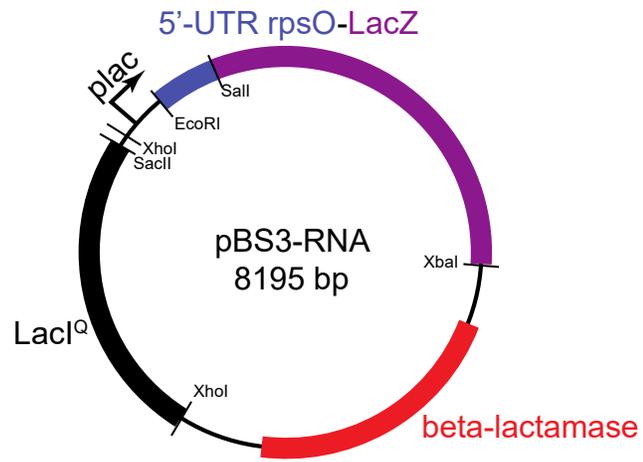


Figure 3.3 Background β -galactosidase Expression of the *E. coli* *ArpsO* Strain

Background Expression of β -galactosidase was assessed using Miller Assays performed under the same conditions used to assay regulator activity. (A) Cells that lack a pRNA reporter plasmid display ~600–1800 Miller Units. (B) Cells that contain a pRNA plasmid (carrying a *lacI^Q* allele) where the *lacZ* reporter gene was replaced with a GFP reporter gene (pBS4) display 6–250 Miller Units. This indicates that the *lacI^Q* carried by the high-copy pRNA plasmid significantly reduces endogenous *lacZ* expression. (C) Representative data from cells containing pBS3-RNA, a plasmid that contains both *lacI^Q* repressor and *lacZ* reporter gene, shows that the *lacZ* reporter produces significant β -galactosidase activity over the endogenous levels. Figure 3.5 shows the β -galactosidase expression with pBS3 containing all versions of the mRNAs tested.

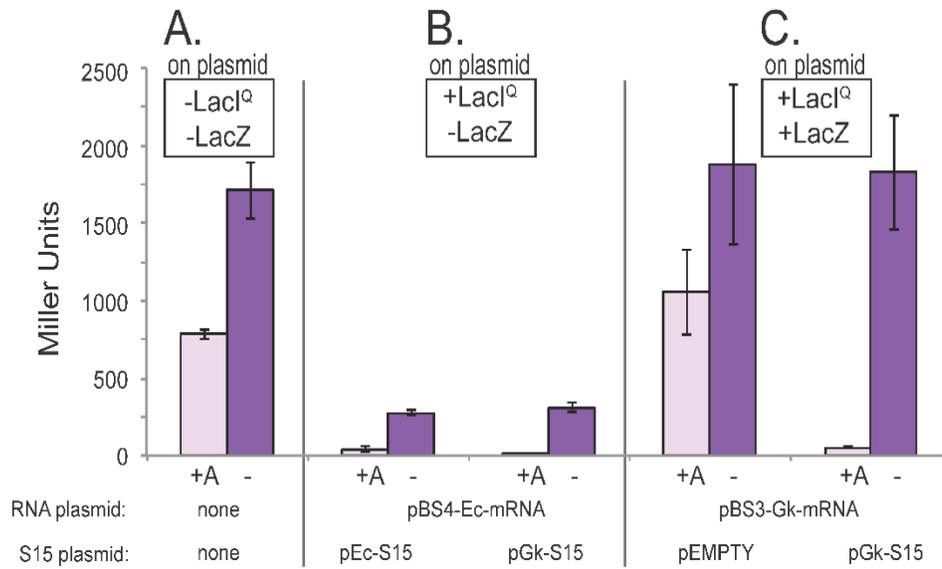


Figure 3.4 Native mRNA-S15 Regulatory RNA-S15 Interactions

Native mRNA-S15 regulation is observed for mRNA structures that interact with S15 homologs in different bacterial species. (A) Ec-mRNA from *E. coli* and Ec-mRNA-M1, (B) Rr-mRNA from *R. radiobacter* and Rr-mRNA-M1, (C) Tt-mRNA from *T. thermophilus* and Tt-mRNA-M1, (D) Gk-mRNA from *G. kaustophilus* and Gk-mRNA-M1, (E) Doubling times calculated during logarithmic phase growth for $\Delta rpsO$ strain carrying plasmids that express different S15 homologs (pEc-S15, pRr-S15, pTt-S15, pGk-S15) or the vector with no protein insert (pEMPTY) under conditions where protein is expressed (+arabinose) and not expressed (-arabinose). (F) Fold-repression for each mRNA with its native binding partner. Fold-repression corresponds to (β -galactosidase activity (+arabinose))/(β -galactosidase activity (-arabinose)). Each mRNA is compared to its own mutant (e.g. Ec-mRNA and Ec-mRNA-M1 are compared in the same set of bars). Error bars represent standard error across 3 or more biological replicates.

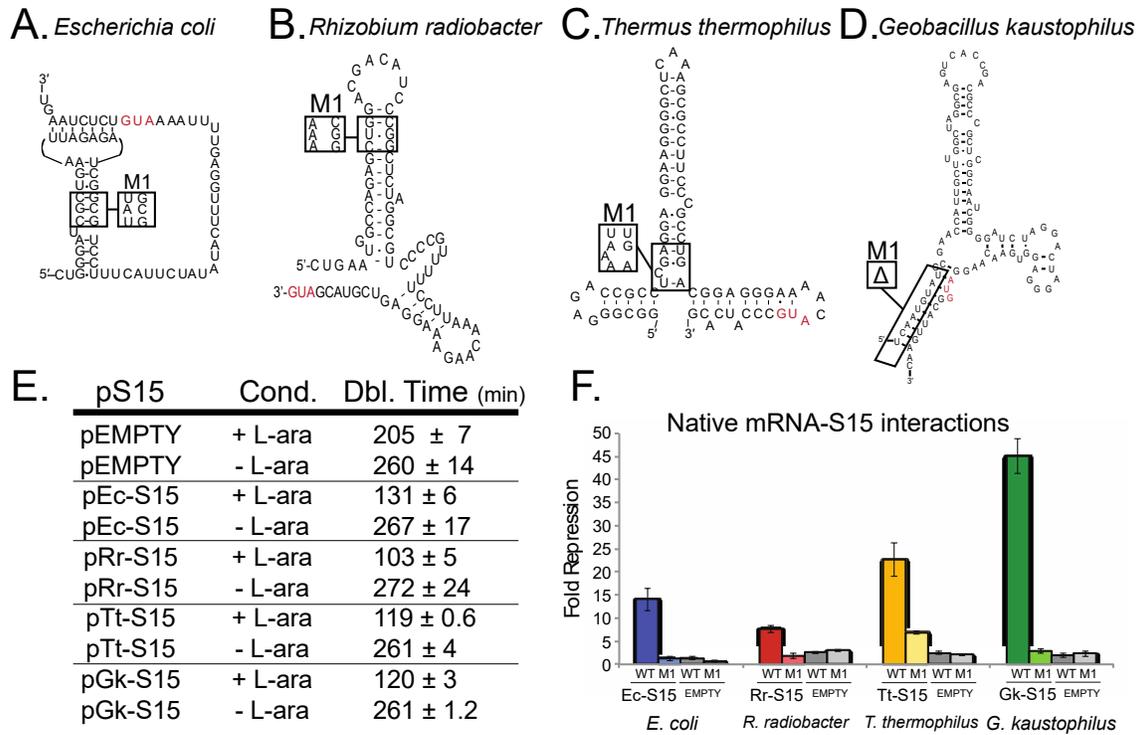


Figure 3.5 All + and – L-arabinose Miller Units

The Miller Units for +L-arabinose and –L-arabinose (protein induced and protein uninduced) conditions for each mRNA with exogenous protein expression (Ec-S15, Rr-S15, Tt-S15 and Gk-S15) and empty vector (EMPTY). (A) Ec-mRNA and Ec-mRNA-M1, (B) Rr-mRNA and Rr-mRNA-M1, (C) Tt-mRNA and Tt-mRNA-M1, (D) Gk-mRNA, Gk-mRNA-M1, and Gk-mRNA-M2. Solid bars are + arabinose, hatched bars are – arabinose. Dark gray bars are WT, white bars are M1, and light gray bars are Gk-mRNA-M2. Error bars represent the standard error of 3 or more independent replicates.

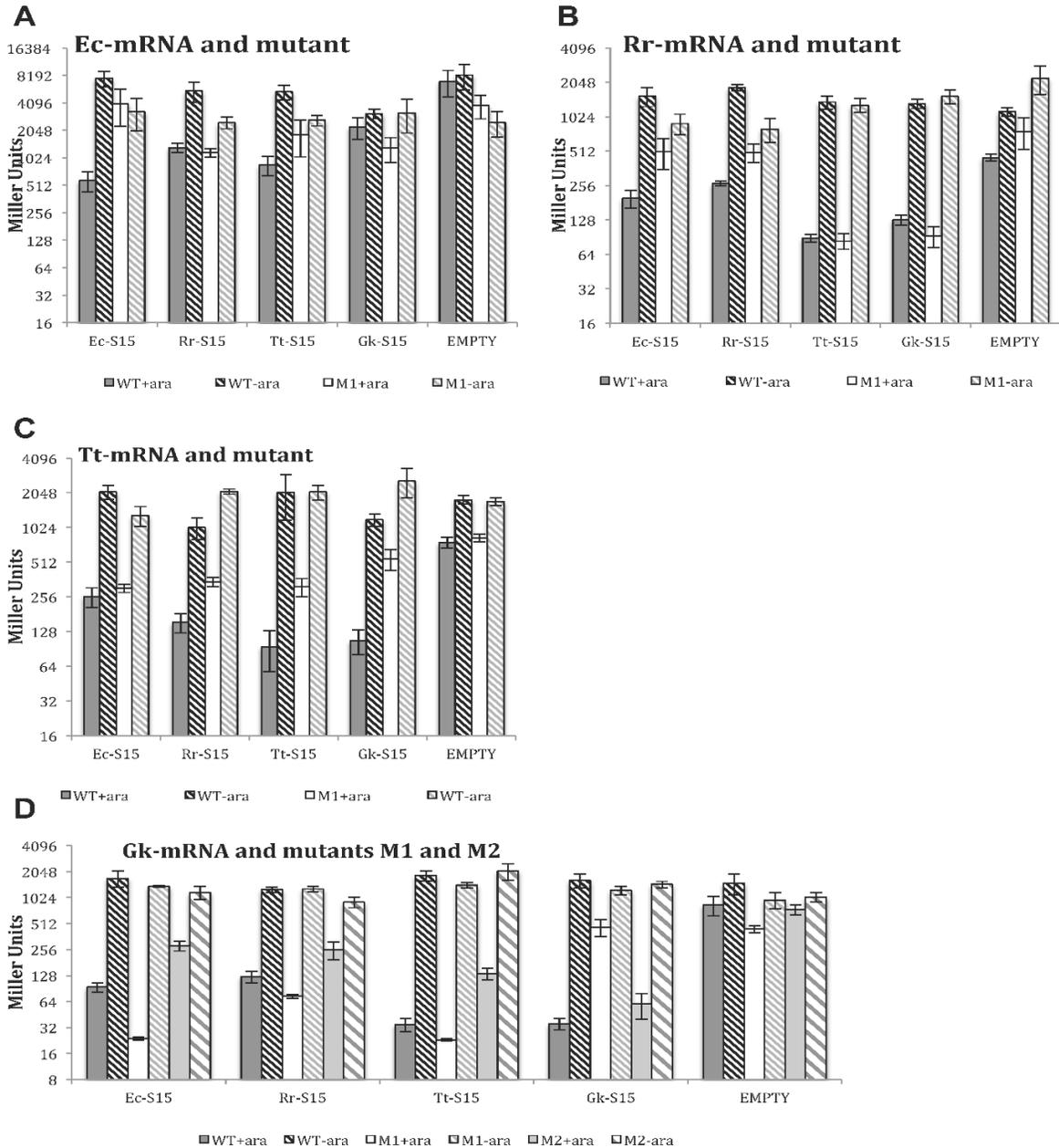


Figure 3.6 Non-native Regulatory RNA-S15 Interactions

Inter-species regulatory activity (fold-repression) of each mRNA in response to each S15 homolog, (A) Ec-mRNA, (B) Rr-mRNA, (C) Tt-mRNA, (D) Gk-mRNA. Each mRNA is compared to its mutant and to pEMPTY (see Figure 3.13). Error bars correspond to standard error for 3 or more replicates. Data corresponding to native interactions is replotted from Figure 3.4 for comparison.

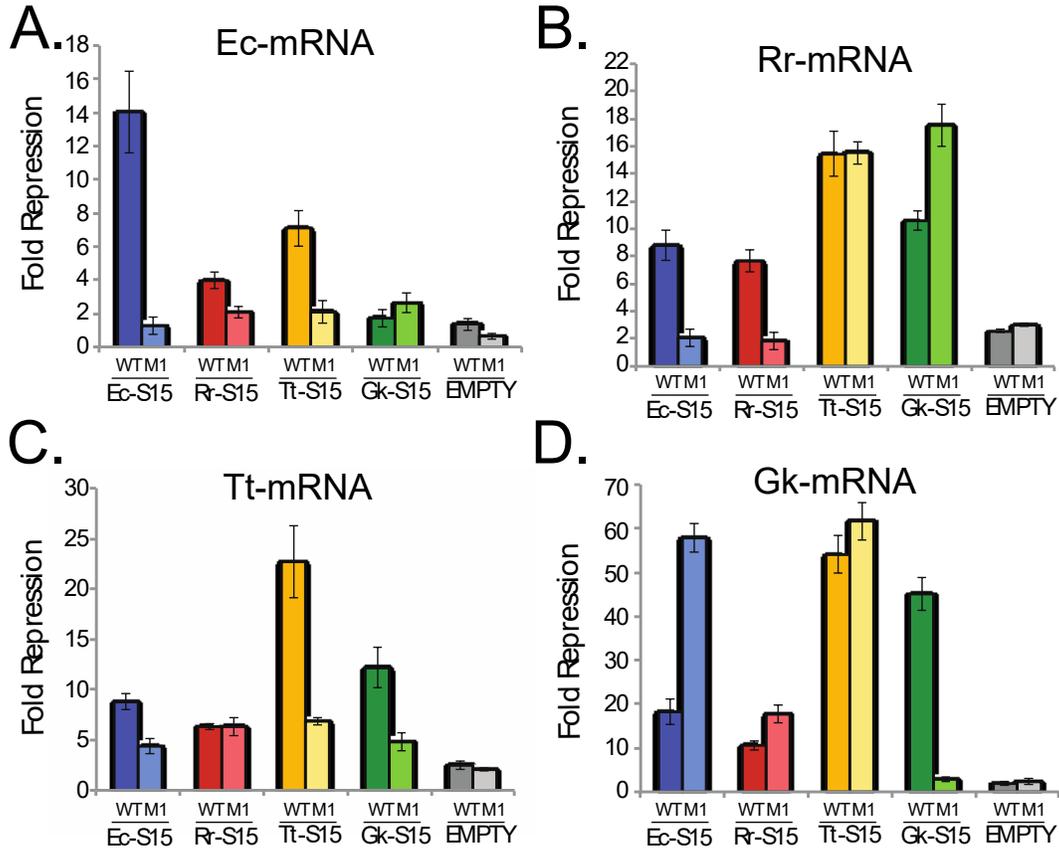


Figure 3.7 Table of Inter-Species Nitrocellulose Binding Assays

Binding assays were used to measure the strength of the mRNA-S15 interaction among all homologs tested. (A) Table of calculated K_D and protein concentration at maximal fraction bound (F_{MAX}). (B) Graph of curves used to calculate K_D . Each curve represents three replicates. The fraction bound was calculated per individual protein concentration $F_b = (\text{counts nitrocellulose}) / (\text{counts total})$. Dots represent average \pm standard error (error bars) fraction bound at each protein concentration. Solver (Microsoft Excel) was used to fit the range of variables (Protein concentration vs. F_b) in order to find K_D . The curve represents a line fit to each set of data points where $F_b = (F_{MAX} * \text{Protein concentration}) / (\text{Protein concentration} + K_D)$. *data reported from Serganov 2002

A.

PROTEIN mRNA	Ec-S15 K_D (nM)	Rr-S15 K_D (nM)	Tt-S15 K_D (nM)	Gk-S15 K_D (nM)
Ec-WT	231*	>1000	>500	>400
Ec-M1	n/a	n/a	n/a	n/a
Rr-WT	28.6 \pm 4.8	14.5 \pm 6.1	12.0 \pm 7.0	1.23 \pm 0.25
Rr-M1	>300	n/a	8.8 \pm 1.8	11.8 \pm 6.9
Tt-WT	>500	>500	2.11 \pm 0.25	0.35 \pm 0.23
Tt-M1	>500	>500	>500	76.3 \pm 5.7
Gk-WT	112 \pm 38	205 \pm 142	0.62 \pm 0.07	3.47 \pm 6.0
Gk-M1	57.5 \pm 19.7	>250	0.12 \pm 0.03	>2000

B.

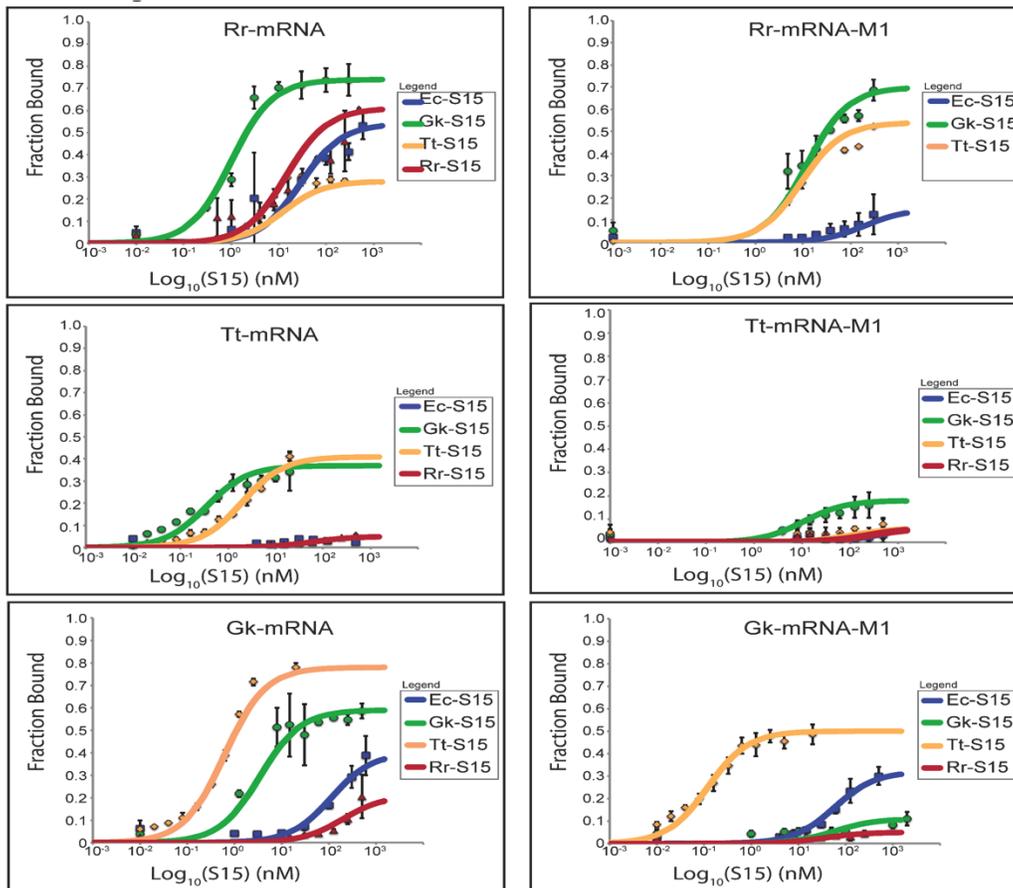


Figure 3.8 Cartoon Representation of S15-mRNA Binding Sites

Summary of S15-mRNA binding sites and cartoon representation of RNA binding sites for S15: (A) rRNA, (B) Ec-mRNA, (C) Rr-mRNA, (D) Tt-mRNA, (E) Gk-mRNA. Regions circled in green putatively correspond to rRNA G•U/G-C motif, regions circled in red putatively correspond to three helix junction, regions circled in blue correspond to phyla-specific S15-binding regions. Important aspects of the binding site as well as regulatory features such as Shine Dalgarno sequences, start codons, and the regions targeted by mutations are indicated. (F) Table summarizing results from both regulatory assays and *in vitro* binding assays. “R” indicates regulatory activity observed, “r” indicates ambiguous regulatory activity observed, “B” indicates *in vitro* binding observed, “b” indicates weak binding, “no” indicates no regulatory activity or no *in vitro* binding, and “-” indicates unmeasured.

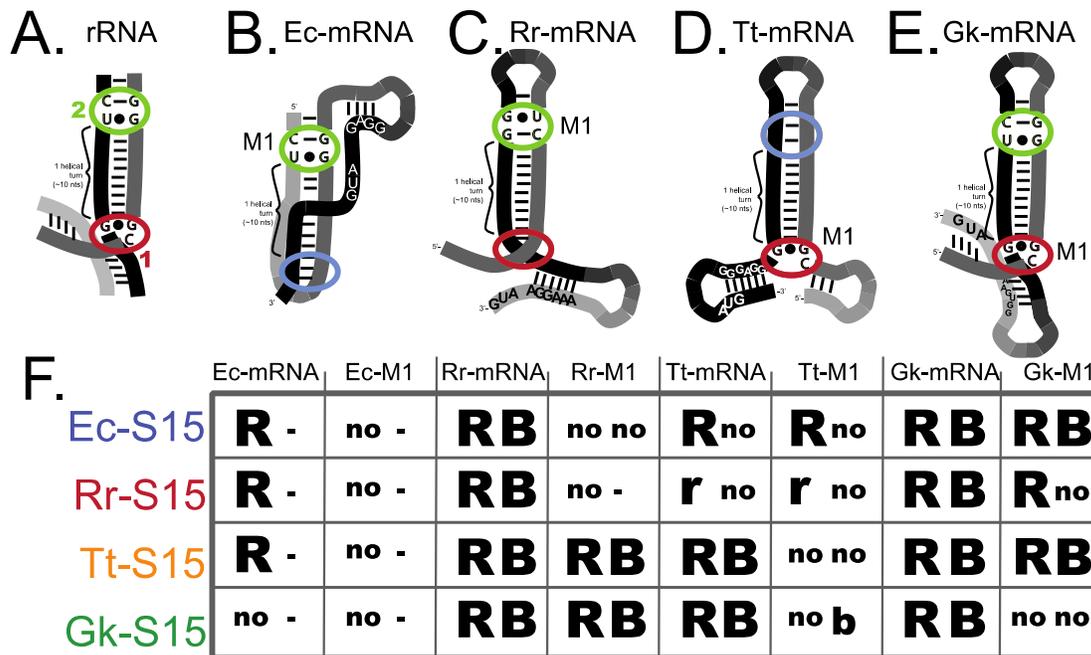


Figure 3.9 Mutation to Gk-mRNA-M2 Strengthens Model of Interaction

(A) Mutation Gk-mRNA-M2 disrupts putative G•U/G-C motif. (B) Fold-repression for Gk-mRNA (WT), Gk-mRNA-M1, and Gk-mRNA-M2 in response to each S15 homolog and the empty vector (pEMPTY). Data for Gk-mRNA and Gk-mRNA-M1 are re-plotted from Figure 3 for comparison. Error bars correspond to standard error for 3 or more replicates. (C) *In vitro* binding data for Gk-mRNA-M2 with each S15 homolog.

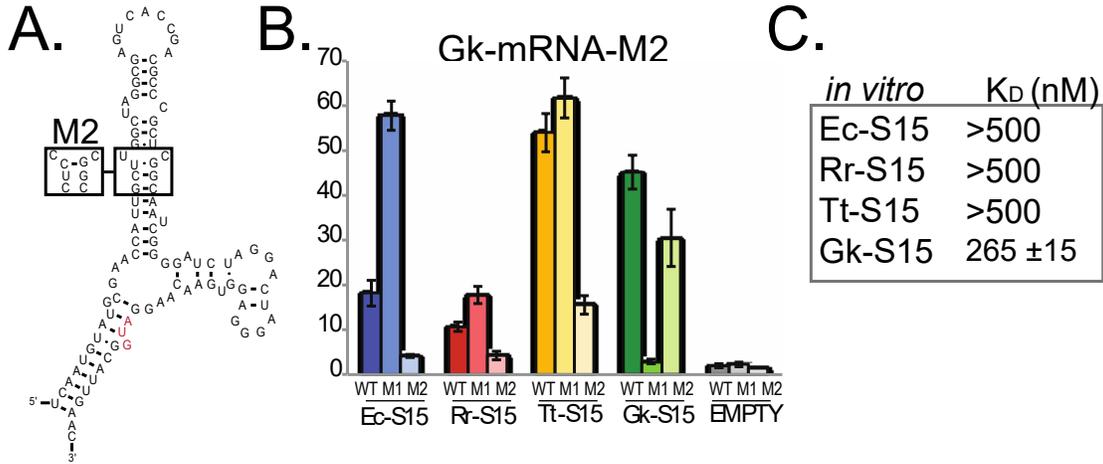


Figure 3.10 S15 Protein Conservation in Gammaproteobacteria and Firmicutes

(A) Rate4Site Z-value indicating degree of conservation for each alignment of S15 coding regions. Blue points correspond to values from the Ec-alignment, and green to values from the Gk-alignment. Lower values are more highly conserved positions. Solid arrows indicate positions that are conserved in both the alignments but have different amino acid identities. Open arrows indicate positions that are conserved in one alignment but not in others, red arrows indicate mutation present in Gk-S15-6MUT. (B) Conservation of individual amino acids within each alignment (generated with Weblogo (Crooks 2004). Residue actually present in the Gk-S15 sequence colored in green, residue actually present in the Ec-S15 sequence colored in blue. (C) Secondary structure diagram of S15, indicating looped or alpha-helix regions, and regions that interact with either the three helix junction (red) or the G•U/G-C motif of rRNA (green). (D) Regulation of Ec-mRNA and Ec-mRNA-M1. In contrast to Gk-S15, Gk-6MUT regulates Ec-mRNA, and this interaction is abolished in Ec-mRNA-M1. (E) Regulation of Gk-mRNA, Gk-mRNA-M1, and Gk-mRNA-M2. GK-6MUT regulates all three of the Gk-mRNA. Error bars represent standard error for three or more replicates. Data for pEMPTY, and Gk-S15 are repeated from Figure 3.4 and 3.6 for comparison.

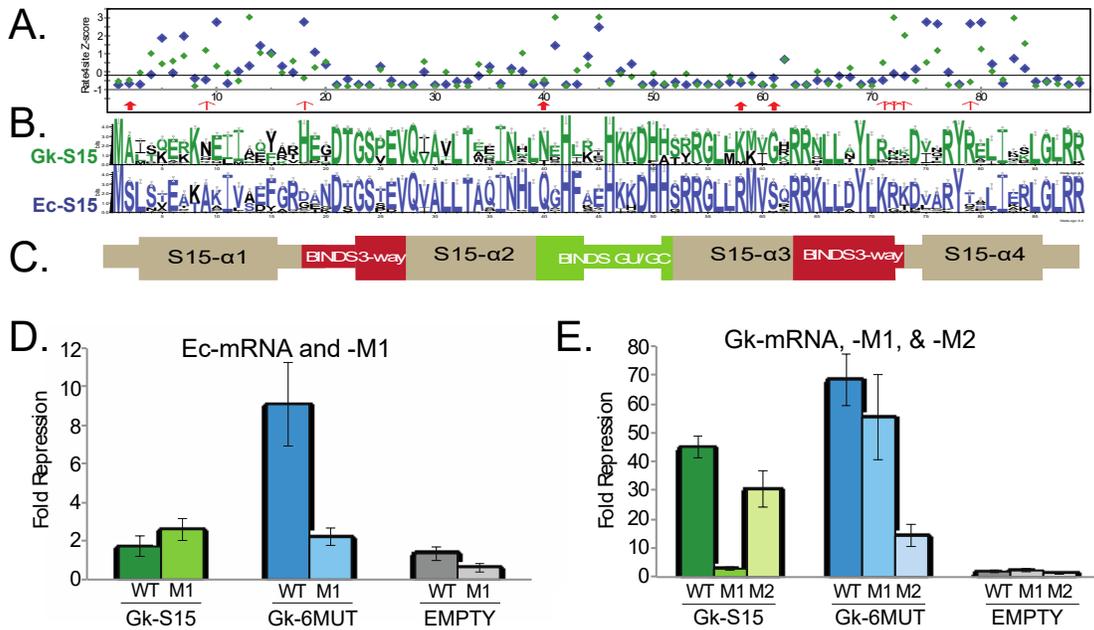


Figure 3.11 Chimeric *G. kaustophilus* S15 with *E. coli* S15

Chimeric Gk-Ec-S15 protein designs and results from regulatory assays. (A) Conservation of individual amino acids in the Firmicute phyla (Gk-S15) and the Gammaproteobacterial phyla (Ec-S15). The amino acid sequence used in all experiments for Gk-S15 is colored green, Ec-S15 colored blue (repeated from Figure 3.10 for clarity). (B) Diagram of S15, repeated from main text, indicating important rRNA-binding regions. (C) Design of chimeric proteins, green bars indicate the amino acid sequence matches Gk-S15, blue bars and letters indicates the amino acid sequence matches Ec-S15 for those regions of the protein. Black bars indicate the break point where amino acid sequences were swapped from one species to the other in constructing each chimera, position 18 and position 72. (D) Miller assay results for all chimeric proteins tested with Gk-mRNA, Gk-mRNA-M1, Gk-mRNA-M2, and Ec-mRNA.

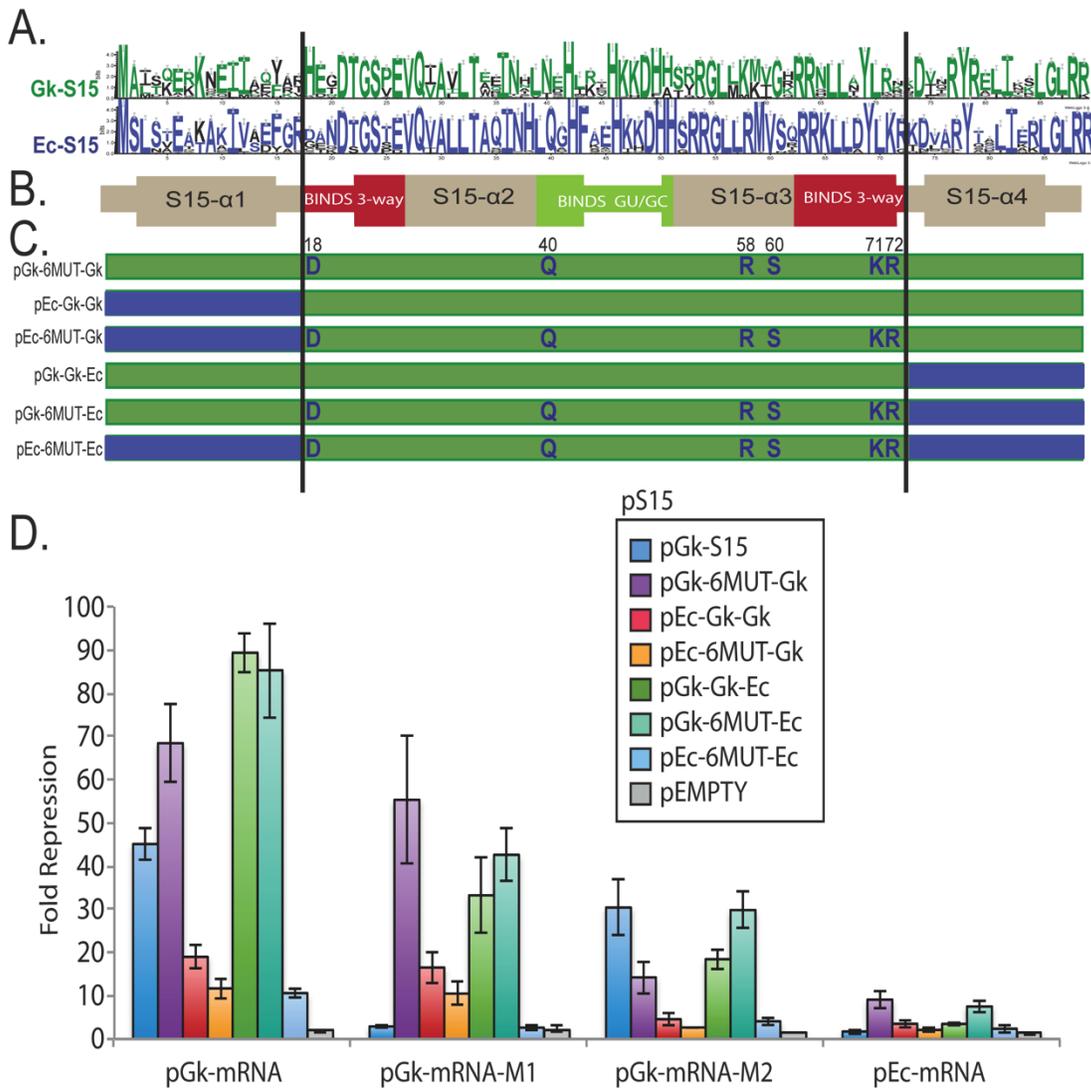


Figure 3.12 Sequences and Primers
 Primers to generate *E. coli* K12: Δ *rpsO* organism.

Primer Number	Sequence
1155	5' -GCTTAACGTCGCGTAAATTGTTTAACTTTGCGTAA CGTGTGTAGGCTGGAGCTGCTTC
1156	5' -TAAGGAGGATATTCATATGGTTACACCCAGCTCATCG AGCGCCTGGGTCTGCGTCGCTAA
1173	5' -GCGTTGCGCCTCGTCGCCTGGTGGTTGAATACCCGGC GTAATGTTAACCGTGTGTAGGCTGGAGCTGCTTC
1174	5' -GCCGTCAGCTTGAAAAAGGGGCCACTCAGGCCCCCT TTTCTGAACTCGCATATGAATATCCTCCTTA
1175	5' -AATTAGAGATCGGCGTCCTTTCATTCTATATACTTTG GAGTTTTAAATGTGTGTAGGCTGGAGCTGCTTC
1176	5' -CAGGCCCCCTTTTCTGAACTCGCAAGAATTAGCGAC GCAGACCCAGGCGCATATGAATATCCTCCTTA
1192	5' -GCAAGCTACCTGCTTCTCTTTG
1193	5' -CTAGAGAATAGGAACCTCGGAATAG
1194	5' -CGATTTTAGCTGTTGCTTCAGTACTTAGAGACATTC ACTTTTCTCTACTGATAGGG
1195	5' -TAATGTGAAAGTGGGTCTTAAAAGCAGCATCATATGA ATATCCTCCTTAGTTCCTATTC
1196	5' -GAATAGGAAC TAAGGAGGATATTCATATGATGCTGCT TTTAAGACCCACTTTCACATTTA
1210	5' -GATAACAGGTCGCTACGAGTAGAATACTGGCTTTTAA GACCCACTTTCACATTTAAGTTG
1211	5' -GTTGAATACCCGGCGTAATGTTAACCGTCGCTTTTAA GACCCACTTTCACATTTAAGTTG
1212	5' -GTCTACGAGGTTTTAGAGCTATGCTG
1222	5' -GATAACCGTATTACCGCCTTTGAGTGAG
1230	5' -ATCGGCGTCCTTTCATTCTATATACTTTGGAGTTTTA AAATGATTGAACAAGATGGATTG
1231	5' -GGGGCCACTCAGGCCCTTTTCTGAACTCGCAAGAA TCAGAAGAACTCGTCAAGAAG
1232	5' -GCGTTGCGCCTCGTCGCCTGGTGGTTGAATACCCGGC GTAATGTTAACCGTATTGAACAAGATGGATTG

Reporter Assay mRNA constructs for pBS3: Coding sequence is bolded, restriction sites in primers are underlined. Mutations to WT sequence are indicated in red.

Construct	Sequence	primers
EC m-RNA WT	ACTGGGATCGCTGAATTAGAGATCG GCGTCCTTTTCATTCTATATACTTTG GAGTTTTAAAAT GTCTCTAAGTACT GAAGCAACAGCT	5' - GGAATTCACTGGGATCGCTGAATTAGAGATC GG 5' - ACGCGT <u>CGACAGCT</u> GTTGCTTCAGTACTTAG AGACA
EC m-RNA M1	ACTGGGAT TATT GAAATTAGAGATCG GCGTCCTTTTCATTCTATATACTTTG GAGTTTTAAAAT GTCTCTAAGTACT GAAGCAACAGCT	5' - TCTGGAATTCACTGGGAT TATT GAAATTAGAG ATCGGC 3' -same as for Ec-mRNA WT
Rr-mRNA WT	ATAGGCAGCGCCAGCTTGGGCTTTG CCTATGCTGCTGAATGGCCAGAGCT GGACGACATCCCGGCTCTAGGCGTC CCCGTTTTTCCCTTAAAACAAGAAAG GATCGTACGAT GTCTGATTACTGCAG AGCGCAAAGCC	5' - CAAGAATTCATAGGCAGCGCCAGCTTGGG 5' - ACGCGT <u>CGACGGCT</u> TTCGCTCTGCAGTAAT CG
Rr-mRNA M1	ATAGGCAGCGCCAGCTTGGGCTTTG CCTATGCTGCTGAATGGCCAGAG AA AGACGACATCCCGGCTCTAGGCGTC CCCGTTTTTCCCTTAAAACAAGAAAG GATCGTACGAT GTCTGATTACTGCAG AGCGCAAAGCC	5' - GAATGGCCAGAG AAAGACGACATCCCGG 5' - CCGGGATGTCGTC TTT TCTCTGGCCATTC
Tt-mRNA WT	AGGCTTGGCGGGAGACCGCCTCGAG GAGGAAGGGGCTCAAAGCGCCTTCC CGCCTGACGGAGGGAAAAC ATGCC ATCACGAAG	5' - GGAATTCAGGCTTGGCGGGAGACCGCCTCGA GGAGGAAG 5' - ACGCGT <u>CGACTT</u> CGTGATGGGCATGTTTTCC CTCCGTCAGGCGGGAAG
Tt-mRNA M1	AGGCTTGGCGGGAGACCGCC AAATG GAGGAAGGGGCTCAAAGCGCCTTCC CGCCTGACGGAGGGAAAAC ATGCC ATCACGAAG	5' - ACGGAATTCAGGCTTGGCGGGAGACCGCC AA ATGGAGGAAG 3' -same as for Tt-mRNA WT
Gk-mRNA WT	TCAATGTATGCGAACCATTGCTTGG CTAGGCGAGTCACCGACGCCCGCTC GGCAATCGGGGATCTAGGACTAGGG AGGTGAACAAGGAT TGGCATTGAAC	5' - GGAATTCCTCAATGTATGCGAACCATTGC TTGGC 5' - ACGCGT <u>CGACGTT</u> CAATGCCATCCTTGT TCACCTCCCTAGT
Gk-mRNA M1	CGAACCATTGCTTGGCTAGGCGAGT CACCGACGCCCGCTCGGCAATCGGG GATCTAGGACTAGGGAGGTGAACAA GGAT TGGCATTGAAC	5' - GTGGCCGAAT <u>TCCGAACCATTGCTTGGC</u> TAGGCGA 3' -same as for Gk-mRNA WT
Gk-mRNA M2	TCAATGTATGCGAACCATT CTCCGG CTAGGCGAGTCACCGACGCCCGCTC GGCAATCGGGGATCTAGGACTAGGG AGGTGAACAAGGAT TGGCATTGAAC	5' - CCGGAATTCCTCAATGTATGCGAACCATT CTCCGGCTAGGCGAG 3' - same as for Gk-mRNA WT

S15 reporter constructs (pS15). Coding sequence is bolded, restriction sites in primers are underlined. Mutations to WT sequence are indicated in red.

S15 coding sequences

Sequence	Primers
pEC-S15 ATGTCTCTAAGTACTGAAGCAACAGCTAAAATCGTTTCT GAGTTTGGTTCGTGACGCAAACGACACCGGTTCTACCGAA GTTTCAGGTAGCACTGCTGACTGCACAGATCAACCACCTG CAGGGCCACTTTGCAGAGCACAAAAAGATCACCACAGC CGTCGTGGTCTGCTGCGCATGGTTTCTCAGCGTCGTAAA CTGCTCGACTACCTGAAACGTAAAGACGTAGCACGTTAC ACCCAGCTCATCGAGCGCCTGGGTCTGCGTCGCTAA	5' - CACGAGCTCAGGAGGTTTTAAAATGTCTC TAAGTACTGAAGCACAG 5' - GCTCTAGATTAGCGACGCAGACCCAGGCG C
pRr-S15 ATGTCGATTACTGCAGAGCGCAAAGCCGCCCTCATCACG GAATATGCCACCAAGGCAGGCGACACCGGTTCTCCGGAA GTTTCAGGTCGCAATCCTGACCGAGCGGATCAACAACCTG ACCGGTCACCTTCAAGGACCACAAGAAGGACAACCCTCC CGTCGTGGCCTTCTGACGCTCGTTTCGAGCCGCCGTTTCG CTTCTCGACTATCTGAAGAAGAAGGACGAAGCCCGTTAC ACCAAGCTGATCGGTGCTCTCGGCATTGCGCGCTAA	5' - CAAGAGCTCAGGAGGTTTTAAAATGTCTCGA TTACTGCAGAGCGCAAAG 5' - CAATCTAGATTAGCGGCGAATGCCGAGAG C
pTt-S15 ATGCCATCACGAAGGAAGAGAAGCAGAAGGTCATCCAG GAGTTCGCCCGCTTCCCCGGGACACGGGGAGCACCGGAG GTGCAGGTGGCGCTCCTTACCCTGAGGATCAACCAGGCTT TCCGAGCACCTCAAGGTCCACAAGAAGGACCACCCTCC CACCGCGGCTCCTGATGATGGTGGGCCAGCGCCGCGAGG CTCTCCGCTACCTCCAGCGGGAGGACCCCGAGCGGTAC CGGGCCCTTATTGAGAAGCTGGGCATCCGGGGTTAA	5' - CACGAGCTCAGGAGGTTTTAAAATGCCCA TCACGAAGGAAGAG 5' - GGCTCTAGATTAACCCCGGATGCCAGCT TCTCAATAAGGGCCCC
pGk-S15 ATGGCATTGACGCAGGAGCGCAAACGCGAAATCATCGAG CAGTTTTAAAATCCATGAGAACGACACTGGTTCTCCGGAA GTGCAAGTTGCGATCCTGACGGAGCAAATCAACAACCTTG AACGAGCATTGCGCATTCATAAAAAAGACCATCATTCA CGGCGCGGCTTGCTGAAAATGGTCGGGAAGCGCCGCAAC TTATTGGCCTACTTGCGCAAGAAAGATGTGGCGCGCTAC CGTGAATTGATTGAGAAACTTGGATTACGTCGATAA	5' - CACGAGCTCAGGAGGTTTTAAAATGGCAT TGACGCAGGAGCGC 5' - GCTCTAGATTATCGACGTAATCCAAGTTTT CTCAATC
pGK-6MUT ATGGCATTGACGCAGGAGCGCAAACGCGAAATCATCGAG CAGTTTTAAAATCGATGAGAACGACACTGGTTCTCCGGAA GTGCAAGTTGCGATCCTGACGGAGCAAATCAACAACCTTG CAAGAGCATTGCGCATTCATAAAAAAGACCATCATTCA CGGCGCGGCTTGCTGCGAATGGTCAGCAAGCGCCGCAAC TTATTGGCCTACTTGAAGGAAAGATGTGGCGCGCTAC CGTGAATTGATTGAGAAACTTGGATTACGTCGATAA	Gkrps0-H18D-F 5' - CATCGAGCAGTTTTAAAATCGATGAGAACG ACACTGGTTCTCCG Gkrps0-H18D-R 5' - CGGAGAACCAGTGTCTGTTCTCATCGATTT TAAACTGCTCGATG Gkrps0-N40Q-F 5' - CGGAGCAAATCAACAACCTTGCAAGAGCAT TTGCGCATTC Gkrps0-N40Q-R 5' - GAATGCGCAAATGCTCTTGCAAGTTGTTG ATTTGCTCCG Gkrps0-R71K-K72R-F 5' - CAACTTATTGGCCTACTTGAAGGAAAG ATGTGGCGCGC Gkrps0-R71K-K72R-R 5' - GCGCGCCACATCTTCTTTCAAGTAGG CCAATAAGTTG Gkrps0-K58R-G60S-F 5' - CGCGGCTTGCTGCGAATGGTCAGCAAGCG CCGCAAC

Primers for *in vitro* binding assays. Coding sequence is bolded. Mutations to WT sequence are indicated in red.

Primers for *in vitro* transcription of RNAs

RNA	Primer Sequences	Template
Rr-WT	5' - TAATACGACTCACTATAGGATAGGCAGCGCCAGCTTGG	pBS3-Rr-WT
	5' - GGCTTTGCGCTCTGCAGTAATC	
Rr-M1	5' - TAATACGACTCACTATAGGATAGGCAGCGCCAGCTTGG	pBS3-Rr-M1
	5' - GGCTTTGCGCTCTGCAGTAATC	
Tt-WT	5' - TAATACGACTCACTATAGGCTTGGCGGGAGACCGCCT	pBS3-Tt-WT
	5' - TTCGTGATGGGCATGTTTTCCCTCCGTCAGGCGGGAAGGCGCTTGAGCC	
Tt-M1	5' - TAATACGACTCACTATAGGCTTGGCGGGAGACCGCC AAAT GGAGGAAG	pBS3-Tt-WT
	5' - TTCGTGATGGGCATGTTTTCCCTCCGTCAGGCGGGAAGGCGCTTGAGCC	
Gk-WT	5' - TAATACGACTCACTATAGGTC AATGTATGCGAACCATTGC	pBS3-Gk-WT
	5' - TTCAATGCCATCCTTGTTTAC	
Gk-M1	5' - TAATACGACTCACTATAGGCGAACCATTGCTTGCTAGGCGA	pBS3-Gk-WT
	5' - TTCAATGCCATCCTTGTTTAC	
Gk-M2	5' - TAATACGACTCACTATAGGTC AATGTATGCGAACCATTGC	pBS3-Gk-M2
	5' - TTCAATGCCATCCTTGTTTAC	

Primers and reference sequences for chimeric proteins, *G. kaustophilus* S15 and *E. coli* S15.

Chimeric Sequences:

<p>Ec-Gk-Gk ATGTCTCTAAGTACTGAAGCAACAGCTAAAATCGTTTCT GAGTTTGGTCGTCATGAGAACGACACTGGTTCTCCGGAA GTGCAAGTTGCGATCCTGACGGAGCAAATCAACAACCTTG AACGAGCATTGCGCATTCATAAAAAAGACCATCATTCA CGGCGCGGCTTGCTGAAAATGGTCGGGAAGCGCCGCAAC TTATTGGCCTACTTGCACAAGAAAGATGTGGCGCGCTAC CGTGAATTGATTGAGAACTTGGATTACGTCGATAA</p>	<p>5' – GTTTCTGAGTTTGGTTCGTCATGAGAA CGACACTGG</p>
<p>Ec-6MUT-Gk ATGTCTCTAAGTACTGAAGCAACAGCTAAAATCGTTTCT GAGTTTGGTCGTCATGAGAACGACACTGGTTCTCCGGAA GTGCAAGTTGCGATCCTGACGGAGCAAATCAACAACCTTG CAAGAGCATTGCGCATTCATAAAAAAGACCATCATTCA CGGCGCGGCTTGCTGCGAATGGTCAGCAAGCGCCGCAAC TTATTGGCCTACTTGGAAAAGAAAGATGTGGCGCGCTAC CGTGAATTGATTGAGAACTTGGATTACGTCGATAA</p>	<p>5' – CCAGTGTCGTTTCTCATGACGACCAAA CTCAGAAAC</p>
<p>Gk-Gk-Ec ATGGCATTGACGCAGGAGCGCAAACGCGAAATCATCGAG CAGTTTAAAATCCATGAGAACGACACTGGTTCTCCGGAA GTGCAAGTTGCGATCCTGACGGAGCAAATCAACAACCTTG AACGAGCATTGCGCATTCATAAAAAAGACCATCATTCA CGGCGCGGCTTGCTGAAAATGGTCGGGAAGCGCCGCAAC TTATTGGCCTACTTGCACAAGAAAGACGTAGCACGTTAC ACCCAGCTCATCGAGCGCCTGGGTCTGCGTCGCTAA</p>	<p>5' – CAACTTATTGGCCTACTTGCACAAGA AAGACGTAGCACGTTACAC</p>
<p>Gk-6MUT-Ec ATGGCATTGACGCAGGAGCGCAAACGCGAAATCATCGAG CAGTTTAAAATCCATGAGAACGACACTGGTTCTCCGGAA GTGCAAGTTGCGATCCTGACGGAGCAAATCAACAACCTTG CAAGAGCATTGCGCATTCATAAAAAAGACCATCATTCA CGGCGCGGCTTGCTGCGAATGGTCAGCAAGCGCCGCAAC TTATTGGCCTACTTGGAAAAGAAAGACGTAGCACGTTAC ACCCAGCTCATCGAGCGCCTGGGTCTGCGTCGCTAA</p>	<p>5' – GTGTAACGTGCTACGTCTTTCTTGGC CAAGTAGGCCAATAAGTTG</p>
<p>Ec-6MUT-Ec ATGTCTCTAAGTACTGAAGCAACAGCTAAAATCGTTTCT GAGTTTGGTCGTCATGAGAACGACACTGGTTCTCCGGAA GTGCAAGTTGCGATCCTGACGGAGCAAATCAACAACCTTG CAAGAGCATTGCGCATTCATAAAAAAGACCATCATTCA CGGCGCGGCTTGCTGCGAATGGTCAGCAAGCGCCGCAAC TTATTGGCCTACTTGGAAAAGAAAGACGTAGCACGTTAC ACCCAGCTCATCGAGCGCCTGGGTCTGCGTCGCTAA</p>	<p>5' – CAACTTATTGGCCTACTTGGAAAAGGA AAGACGTAGCACGTTAC</p> <p>5' – GTAACGTGCTACGTCTTTCTTTTCA AGTAGGCCAATAAGTTG</p>

Figure 3.13 Statistics for Data: Native and Non-Native Interactions

Statistics for data on Figures 3.4 and 3.6. Interactions are considered significant if they display >3 fold-repression and have a p-value < 0.05 when compared to empty vector. For reference we have also compared the response of all mutant RNAs to both the response of the mutant with the empty vector, and the response of the unmutated RNA with in the presence of the same protein. Statistically significant results are bolded.

Figure 3.4 and 3.6 Statistics

RNA	S15	AVE	STE	P Value	pRNA/pS15
pEc-WT	pEc-S15	14.03	2.45	1.63E-03	v pEc-WT/pEMPTY
	pRr-S15	3.96	0.51	1.05E-03	v pEc-WT/pEMPTY
	pTt-S15	7.10	1.08	2.13E-03	v pEc-WT/pEMPTY
	pGk-S15	1.72	0.53	3.01E-01	v pEc-WT/pEMPTY
	pEMPTY	1.36	0.37		
pEc-M1	pEc-S15	1.26	0.51	1.48E-03	v pEc-WT/pEc-S15
				1.75E-01	v pEc-M1/pEMPTY
	pRr-S15	2.06	0.32	5.05E-03	v pEc-Wt/pRr-S15
				1.94E-03	v pEc-M1/pEMPTY
	pTt-S15	2.08	0.68	3.39E-03	v pEc-WT/pTt-S15
				5.98E-02	v pEc-M1/pEMPTY
	pGk-S15	2.61	0.58	1.51E-01	v pEc-WT/pGk-S15
			1.83E-02	v pEc-M1/pEMPTY	
	pEMPTY	0.64	0.17		
pRr-WT	pEc-S15	8.79	1.09	5.86E-04	v pRr-WT/pEMPTY
	pRr-S15	7.65	0.77	1.23E-03	v pRr-WT/pEMPTY
	pTt-S15	15.43	1.64	2.09E-03	v pRr-WT/pEMPTY
	pGk-S15	10.58	0.71	1.42E-04	v pRr-WT/pEMPTY
	pEMPTY	2.53	0.11		
pRr -M1	pEc-S15	2.06	0.63	2.60E-04	v pRr-WT/pEc-S15
				1.17E-01	v pRr-M1/pEMPTY
	pRr-S15	1.83	0.61	5.57E-04	v pRr-WT/pRr-S15
				9.75E-02	v pRr-M1/pEMPTY
	pTt-S15	15.54	0.83	4.79E-01	v pRr-WT/pTt-S15
				2.74E-04	v pRr-M1/pEMPTY
	pGk-S15	17.52	1.55	3.75E-03	v pRr-WT/pGk-S15
			3.49E-04	v pRr-M1/pEMPTY	
	pEMPTY	3.00	0.10		
pTt-WT	pEc-S15	8.79	0.76	2.36E-05	v pTt-WT/pEMPTY
	pRr-S15	6.32	0.34	9.28E-05	v pTt-WT/pEMPTY
	pTt-S15	22.69	3.62	1.47E-02	v pTt-WT/pEMPTY
	pGk-S15	12.16	2.01	1.80E-02	v pTt-WT/pEMPTY
	pEMPTY	2.49	0.41		
pTt -M1	pEc-S15	4.38	0.77	9.09E-04	v pTt-WT/pEc-S15
				1.46E-02	v pTt-M1/pEMPTY
	pRr-S15	6.33	0.90	4.97E-01	v pTt-WT/pRr-S15
				8.75E-03	v pTt-M1/pEMPTY
	pTt-S15	6.85	0.35	2.37E-02	v pTt-WT/pTt-S15
				2.33E-04	v pTt-M1/pEMPTY
	pGk-S15	4.80	0.92	2.46E-02	v pTt-WT/pGk-S15
			4.81E-02	v pTt-M1/pEMPTY	
	pEMPTY	2.07	0.09		
pGk-WT	pEc-S15	18.25	2.95	5.39E-03	v pGk-WT/pEMPTY
	pRr-S15	10.63	1.13	1.25E-03	v pGk-WT/pEMPTY

pGk -M1	pTt-S15	54.05	4.31	3.20E-03	v pGk-WT/pEMPTY
	pGk-S15	45.08	3.74	3.45E-03	v pGk-WT/pEMPTY
	pEMPTY	1.92	0.40		
	pEc-S15	57.84	3.31	4.18E-02	v pGk-WT/pEc-S15
				9.05E-03	v pGk-M1/pEMPTY
	pRr-S15	17.68	1.95	1.38E-02	v pGk-WT/pRr-S15
				1.26E-03	v pGk-M1/pEMPTY
	pTt-S15	61.67	4.38	1.36E-01	v pGk-WT/pTt-S15
				3.64E-04	v pGk-M1/pEMPTY
	pGk-S15	2.90	0.54	3.41E-03	v pGk-WT/pGk-S15
			2.54E-01	v pGk-M1/pEMPTY	
	pEMPTY	2.31	0.62		

Figure 3.14 Statistics for Gk-S15 Mutants Assays

Statistics for data on Figures 3.9 and 3.11. Interactions are considered significant if they display >3 fold-repression and have a p-value < 0.05 when compared to empty vector. For reference we have also compared the response of all mutant RNAs to both the response of the mutant with the empty vector, and the response of the non-mutated RNA with in the presence of the same protein. Statistically significant results are bolded.

Figure 3.9 Statistics

RNA	S15	AVE	STE	P Value	pRNA/pS15
pGk -M2	pEc-S15	4.12	0.42	8.09E-03	v pGk-WT/pEc-S15
				8.84E-03	v pGk-M1/pEc-S15
				4.70E-04	v pGk-M2/pEMPTY
	pRr-S15	4.29	1.03	3.06E-03	v pGk-WT/pRr-S15
				1.21E-03	v pGk-M1/pRr-S15
				3.38E-02	v pGk-M2/pEMPTY
	pTt-S15	15.57	2.20	1.97E-03	v pGk-WT/pTt-S15
				2.17E-04	v pGk-M1/pTt-S15
				3.75E-03	v pGk-M2/pEMPTY
	pGk-S15	30.40	6.34	5.36E-02	v pGk-WT/pGk-S15
				1.11E-02	v pGk-M1/pGk-S15
				9.82E-03	v pGk-M2/pEMPTY
pEMPTY	1.42	0.12			

Figure 3.10 and 3.11 Statistics

RNA	S15	fold	STE	P Value	pRNA/pS15
pGk-WT	pGk-6MUT-Gk	68.41	9.00	3.72E-02	v pGk-WT/pGk-S15
				3.97E-03	v pGk-WT/pEc-S15
				2.54E-03	v pGk-WT1/pEMPTY
	pEc-Gk-Gk	18.98	2.61	1.55E-02	v pGk-WT/pGk-S15
				4.30E-01	v pGk-WT/pEc-S15
				1.03E-02	v pGk-WT1/pEMPTY
	pEc-6MUT-Gk	11.69	2.45	1.39E-01	v pGk-WT/pGk-S15
				6.76E-02	v pGk-WT/pEc-S15
				7.67E-03	v pGk-WT1/pEMPTY
	pGk-Gk-Ec	89.34	4.49	9.26E-04	v pGk-WT/pGk-S15
				1.64E-04	v pGk-WT/pEc-S15
				1.23E-03	v pGk-WT1/pEMPTY
	pGk-6MUT-Ec	85.29	11.0	1.47E-02	v pGk-WT/pGk-S15
				3.29E-03	v pGk-WT/pEc-S15
				2.34E-03	v pGk-WT1/pEMPTY
pEc-6MUT-Ec	10.62	1.02	1.34E-01	v pGk-WT/pGk-S15	
			3.80E-02	v pGk-WT/pEc-S15	
			3.39E-03	v pGk-WT1/pEMPTY	
pGk -M1	pGk-6MUT-Gk	55.31	14.7	8.04E-03	v pGk-M1/pGk-S15
				4.36E-01	v pGk-M1/pEc-S15
				7.70E-03	v pGk-M1/pEMPTY
				2.35E-01	v pGk-WT/pGk-6MUT-Gk

	pEc-Gk-Gk	16.52	3.57	2.68E-02	v pGk-M1/pGk-S15
				3.47E-02	v pGk-M1/pEc-S15
				2.69E-02	v pGk-M1/pEMPTY
				3.06E-01	v pGk-WT/pEc-Gk-GK
	pEc-6MUT-Gk	10.50	2.55	3.87E-02	v pGk-M1/pGk-S15
				3.02E-02	v pGk-M1/pEc-S15
				4.16E-02	v pGk-M1/pEMPTY
				2.88E-01	v pGk-WT/pEc-6MUT-Gk
	pGk-Gk-Ec	33.22	8.23	9.23E-03	v pGk-M1/pGk-S15
				1.95E-02	v pGk-M1/pEc-S15
				8.57E-03	v pGk-M1/pEMPTY
				4.24E-04	v pGk-WT/pGk-Gk-Ec
	pGk-6MUT-Ec	42.58	6.01	1.31E-03	v pGk-M1/pGk-S15
				3.49E-02	v pGk-M1/pEc-S15
				1.22E-03	v pGk-M1/pEMPTY
				1.03E-02	v pGk-WT/pGk-6MUT-Ec
	pEc-6MUT-Ec	2.62	0.77	1.84E-01	v pGk-M1/pGk-S15
				9.26E-03	v pGk-M1/pEc-S15
				3.80E-01	v pGk-M1/pEMPTY
				1.22E-03	v pGk-WT/pEc-6MUT-Ec
pGk -M2	pGk-6MUT-Gk	14.24	3.80	4.00E-02	v pGk-M2/pGk-S15
				1.40E-02	v pGk-M2/pEMPTY
	pEc-Gk-Gk	4.55	1.38	1.40E-01	v pGk-M2/pGk-S15
				5.35E-02	v pGk-M2/pEMPTY
	pEc-6MUT-Gk	2.65	0.24	4.94E-01	v pGk-M2/pGk-S15
				3.63E-03	v pGk-M2/pEMPTY
	pGk-Gk-Ec	18.45	2.26	7.75E-02	v pGk-M2/pGk-S15
				8.14E-04	v pGk-M2/pEMPTY
	pGk-6MUT-Ec	29.81	4.12	4.71E-01	v pGk-M2/pGk-S15
				1.02E-02	v pGk-M2/pEMPTY
	pEc-6MUT-Ec	4.10	0.87	1.15E-01	v pGk-M2/pGk-S15
				1.79E-02	v pGk-M2/pEMPTY
pEc-WT	pGk-6MUT-Gk	9.09	2.19	7.24E-03	v pEc-WT/pGk-S15
				5.97E-03	v pEc-WT/pEMPTY
	pEc-Gk-Gk	3.51	0.82	1.71E-02	v pEc-WT/pGk-S15
				2.83E-02	v pEc-WT/pEMPTY
	pEc-6MUT-Gk	2.10	0.43	2.59E-02	v pEc-WT/pGk-S15
				1.14E-01	v pEc-WT/pEMPTY
	pGk-Gk-Ec	3.47	0.54	2.73E-02	v pEc-WT/pGk-S15
				7.08E-03	v pEc-WT/pEMPTY
	pGk-6MUT-Ec	7.50	1.30	7.48E-03	v pEc-WT/pGk-S15
				7.24E-03	v pEc-WT/pEMPTY
	pEc-6MUT-Ec	2.37	0.91	8.86E-02	v pEc-WT/pGk-S15
				1.70E-01	v pEc-WT/pEMPTY
	pGk-6MUT-Gk	2.20	0.46	9.70E-03	v pEc-Wt/pGk-6MUT-Gk
				1.18E-02	v pEc-M1/pEMPTY

CHAPTER IV

Synthetic *cis*-regulatory RNAs for Ribosomal Protein S15

The work in this Chapter is based up on the following unpublished data and observations, and is in preparation for publication in a scientific journal.

Slinger BL, Meyer MM: Synthetic *cis*-regulatory RNAs for Ribosomal Protein S15. *In preparation*.

Authors' contributions:

BLS performed all experimental work including implementation of *in vitro* selection, development of the *in vivo* reporter system, the filter-binding assays, the structural probing, preliminary data analysis, and wrote the manuscript. MMM conceived of the project, wrote the manuscript, and finalized data analysis.

INTRODUCTION

From data presented in Chapter III, it is clear that S15 homologs have distinct regulatory mRNA requirements (Figure 3.8). For example, the RNA from *E. coli* (Ec-mRNA) did not regulate gene expression in response to the S15 homolog from *G. kaustophilus* (Gk-S15). However, taking all the cross-species interactions into account, we find that the structure of all regulatory RNAs at least partially mimics the ribosomal RNA structure with which S15 binds, a three helix junction (3HJ) or a G•U/G-C motif (Figure 1.2, 3.8). Based on these observations it appears that Gk-S15 requires a 3HJ in its regulatory RNA; because Ec-mRNA lacks this motif, gene regulation does not occur between the two. Though the model explains the modern regulatory structures in nature, it does not explain how it came to be this way. The work presented in this chapter addresses how novel *cis*-regulatory RNAs arise and are transformed into regulatory elements.

We performed *in vitro* evolution experiments to identify novel, synthetic *cis*-regulatory RNA structures for ribosomal protein S15. RNA structures that bind a ligand of interest with high affinity and specificity (aptamers) can be efficiently engineered through repeated rounds of an *in vitro* selection strategy called SELEX (Systematic Evolution of Ligands by EXponential Enrichment, (Ellington 1990, Tuerk 1990, Szeitner 2014, Darmostuk 2015). Though it remains a challenge to optimize the selection process to identify synthetic aptamers that are functional *in vivo* (Weigand 2008, Filonov 2014), a powerful strategy to overcome this couples *in vivo* screens on the *in vitro* selected RNA pool (ex. neomycin, Weigand 2008). This has been performed with regulatory RNAs that interact with DNT (Davidson 2013), tetracycline (Hanson 2005, Wunnicke 2011), and

atrazine (Sinha 2010). In the current Chapter, we use this strategy to investigate how readily these important regulatory structural features in S15-interacting RNAs are recapitulated. We begin with a randomized RNA pool based upon Ec-mRNA, and select for RNA sequences that bind Gk-S15. Six individual sequences bound with high affinity *in vitro*, and four of these were able to regulate gene expression *in vivo* in response to Gk-S15. Footprinting experiments of both a regulatory RNA and a non-regulatory RNA elucidate RNA structural features essential for regulation by Gk-S15, as opposed to merely a binding interaction. Finally, mutagenesis of our synthetic regulator confirms that its secondary structure is distinct from the naturally-occurring regulatory RNAs (Figure 2.8).

To our knowledge this is the first synthetic *cis*-regulatory RNA that responds to a single protein. The diversity of RNA structures that survived the selection underscores the plasticity of the mRNA:S15 interaction. Novel regulatory RNAs for S15 appear to be readily made, for four surviving sequences regulate gene expression using distinct sequence. Additionally, our results underline the importance of *cis*-regulation using the rRNA-binding face of S15.

RESULTS

In vitro Selection of RNA Aptamers for Gk-S15

We carried out successive rounds of SELEX on a randomized RNA pool to isolate RNAs that bind S15 from *G. kaustophilus* (Gk-S15) with a high affinity (Figure 4.1A). Because there is no regulatory interaction between the mRNA regulator from *E. coli* (Ec-mRNA, Chapter III) and Gk-S15, the randomized RNA pool was based upon Ec-

mRNA sequence: 5'-TGCGTAACGTACACT-N₃₀-TCATTCTATATACTTTGGAGTTTTAAATGTCTCTAAGTACTGAAGCAACAGC

I, where primer binding regions are underlined, and N₃₀ denotes a randomized region of 30 nucleotides. First, a negative selection using nitrocellulose membrane alone was carried out in the absence of Gk-S15. Non-filter binding RNAs were then incubated with Gk-S15 and nitrocellulose membrane was used to capture Gk-S15 and any RNA bound to Gk-S15. RNA was isolated from the protein-bound filter, and was reverse transcribed. To begin the next round of selection, mutagenic PCR was used to both amplify cDNA of the population as well as increase sequence diversity of the non-primer regions.

Over the eleven rounds of selection we decreased the concentration of Gk-S15 while increasing the overall population binding affinity for the protein (Figure 4.1B, C). The population binding affinity dramatically increased from a K_D of $> 1 \mu\text{M}$ in the unselected population to 150 nM at the final round. To assess the affinity of individual sequences in the Round 11 pool, we isolated and sequenced six individuals from this population. The sequences of these individuals were diverse from one another, containing no common sequence or motif in the randomized region. Additionally, these sequences were predicted to fold into unique secondary structures in RNAfold of the Vienna RNA Package (Lorenz 2011, Figure 4.2). Nitrocellulose filter binding assays were performed using Gk-S15 and 5'-end labeled RNA for each of the six sequences (Figure 4.3A, B). All of the RNAs were able to bind Gk-S15, although the range of binding affinities spans several orders of magnitude. We identified one sequence, 11-1, which has a binding affinity that rivals that of the native mRNA interaction for Gk-S15 (Gk-mRNA, ~ 0.9 nM). Four of the remaining sequences still strongly bind (K_D 8.5-20.7 nM), whereas

sequence 11-6 has a relatively weak binding affinity (289 nM). These results suggest that despite the sequence diversity of the final pool, our SELEX experiment was successful in selecting for RNA aptamers for Gk-S15.

Four Synthetic RNAs Regulate Gene Expression with Gk-S15

To assess whether any of the RNAs allowed regulation, we used an *in vivo* regulation assay to screen potential riboregulators from the survivor pool. This method is our two plasmid system that was described in Chapter III that assesses whether over-expressed Gk-S15 interacts with an RNA to regulate β -galactosidase expression in the cell. One plasmid contains an RNA-*lacZ* reporter where a synthetic RNA sequence is cloned upstream and in-frame with *lacZ* (Slinger 2015) and under the control of an IPTG-inducible plac promoter. A second plasmid carries the *G. kaustophilus rpsO* coding sequence under the control of an L-arabinose inducible promoter, pBAD. The plasmids are co-transformed into an *E. coli* K12: $\Delta rpsO$ strain (Bubunenko 2006). The regulatory assay itself is performed with cultures grown with and without L-arabinose (to induce S15 expression). We performed the assay on stationary phase cells, as during log phase no S15-based regulation with any RNA was apparent, likely because over-expressed S15 rapidly assembles on the rRNA during log phase. At stationary phase (OD₆₀₀ ~1.5) we performed a 30 minute mRNA-*lacZ* induction (inducing expression with IPTG).

The regulatory capacity for all six individual RNAs isolated from the Round 11 pool was assessed using this assay. Strikingly, four of the six surviving sequences enabled a range of regulatory responses to Gk-S15, while two showed no ability to regulate gene expression (Figure 4.4A). Not surprisingly, the maximal amount of reporter

expression allowed by each aptamer differed in the absence of L-arabinose (~1000-5000 Miller Units, Figure 4.5A, B). This not only affects the measurable fold-repression using this assay, it also suggests that these four synthetic regulators are behaving as genetic “OFF” switches. In other words, reporter expression is “ON” in the absence of protein, and Gk-S15 binding results in altering RNA secondary structure so as to turn “OFF” reporter expression.

The strongest binder, RNA 11-1, enabled the strongest gene regulatory response, (30.4 fold-repression). 11-4 and 11-5 have modest binding affinities, yet both regulate reporter expression in response to Gk-S15 (23.1, and 8.9 fold-repression, respectively). Finally, 11-6 has the weakest binding affinity of the six individuals, yet shows a strong regulatory response to Gk-S15 (16.3 fold-repression). Thus, binding strength did not correlate with regulatory capability for RNAs 11-2 and 11-3 bind Gk-S15 strongly *in vitro*, yet neither was able to regulate gene expression *in vivo*. These results suggest the six RNA sequences examined, and likely the sequence pool itself, are folding into distinct secondary structures that are recognized by Gk-S15 *in vitro*, yet something beyond the binding interaction is required to perform gene regulation.

RNAs Regulate Gene Expression with S15 Homolog from *T. thermophilus* but not *E. coli*

We tested the regulatory capacity of three of our most promising regulatory RNAs, 11-1, 11-4, and 11-5, to see if they would respond to homologs of S15 from distant bacterial phyla. By doing this we hoped to gain information about the potential protein binding motifs within these RNAs. Previously, we have shown that S15 homologs

originating from diverse bacterial phyla require distinct structural motifs for recognition of their respective RNA regulators. Briefly, homologs from *G. kaustophilus* and *T. thermophilus* require a 3 helix junction (3HJ), whereas the homolog from *E. coli* requires a G•U/G-C motif (Figure 3.8). Additionally, different mechanisms control the expression of the *rpsO* operon in different bacterial phyla. We find that RNAs 11-1, 11-4, 11-6 all regulate in response to not only Gk-S15, but also the S15 homolog from *T. thermophilus* (Tt-S15) (Figure 4.4B). The homolog from *E. coli* (Ec-S15) exhibited low regulatory capacity with these three synthetic regulatory RNAs. This result is especially striking because the original RNA pool was based on Ec-mRNA; however, it strongly indicates that the selected sequences do not contain the regulatory motif Ec-S15 requires (i.e. G•U/G-C motif). Moreover, these results suggest Gk-S15 and Tt-S15 recognize the three RNAs in a similar fashion.

Elucidation of the Gk-S15 Binding Face

To better understand why some RNA sequences enable Gk-S15-based gene regulation, whereas others do not, we closely examined our two most tightly binding RNAs, one of which is a regulator (11-1), and the other of which is not (11-2). Previous studies have shown that the naturally-occurring mRNA regulators of the *rpsO* operon as well as the 16s rRNA interact with a conserved set of amino acids in the S15 protein, all of which fall on the same side of the globular S15 protein (Mathy 2004, Scott 2005, Slinger 2014). We performed several experiments to assess whether these RNAs are recognizing the same face of Gk-S15 as Gk-mRNA. First, an *in vitro* competition experiment was performed using a fixed amount of 5'-end labelled RNA, a fixed amount

of Gk-S15 and an increasing amount of non-labeled competitor RNA. We find that both 11-1 and 11-2 displace Gk-mRNA from Gk-S15 (Figure 4.6A, B). This suggests that both RNAs bind the same face of Gk-S15. Second, nitrocellulose binding assays were performed with both 11-1 and 11-2 RNAs and the S15 homologs from *T. thermophilus* (Tt-S15) and *E. coli* (Ec-S15). We find that both RNAs are only able to bind Tt-S15, not Ec-S15 (Figure 4.7A-C).

We further examined the 11-1 RNA-protein recognition in our cell-based assay using several Gk-S15 mutants (derived from Scott 2005). We were unable to examine 11-2 in this manner because it is not a functional regulatory RNA. Mutations to the binding face of Gk-S15 (Y68A and D48L) prevented RNA recognition and subsequent gene regulation (Figure 4.6C, Figure 4.5D). These individual amino acids were also found to be essential for autoregulation with the native Gk-S15 regulatory Gk-mRNA. However, mutations to the non-binding face of Gk-S15 (E40L and E79L) do not prevent Gk-S15 from regulating gene expression in response to 11-1. Taken together, these data show 11-1 not only binds the same face of Gk-S15 as its native RNA regulator, but it also may utilize similar amino acids.

Footprinting Experiments Elucidate Nucleotides Important in 11-1 Binding Gk-S15

RNA-S15 binding must occur to regulate gene expression; however, only some of the surviving sequences were able to regulate gene expression. We further investigated the binding interaction between Gk-S15 and the best performing regulatory RNA, 11-1, to more clearly establish how Gk-S15 recognizes the RNA to enable regulation. RNA footprinting experiments were performed to elucidate the secondary structure features in

11-1 that may be essential for regulation in response to Gk-S15 (Figure 4.8A-E). Using 5'-labeled 11-1 in the presence and absence of Gk-S15, RNA secondary structure was probed using RNase VI (VI-, cleaves double stranded regions, not base-specific), RNase A (A-, cleaves single-stranded cytosines and uracils), in-line probing (IL-, cleaves flexible, and likely single-stranded regions, not base-specific), and lead(II) probing (Pb-, cleaves flexible regions, not base-specific).

Using these structure probing data, we have drawn our predicted structure for 11-1 when bound to Gk-S15 (Figure 4.8A). Overall, 11-1 appears relatively unstructured in the absence of protein. This is especially apparent with the number of IL- and Pb-cleavage products in the absence of protein (Figure 4.8B, E). Then upon Gk-S15 binding the RNA locks into its secondary structure. The intensity of VI-cleavage products corresponding with U10, C12, U19, U20, and U25 decreases in the protein-bound RNA, suggesting this region is shielded from RNase cleavage by Gk-S15, or is becoming single stranded (Figure 4.8D). Additionally, in the protein-bound RNA, there is clear shielding from Pb- cleavage in nucleotides C13 through G22, as well as in the G30 through G32 nucleotides, suggesting this region is double stranded in nature (Figure 4.8E). The RNA sequence spanning nucleotide U49 through U67 is also likely to be involved in Gk-S15 recognition. In the presence of Gk-S15, there is shielding from Pb-cleavage, spanning the entire region from U52 through G65 (Figure 4.8E). Also, the VI-cleavage product intensity for U50 through U52 decreases as Gk-S15 concentration is increased (Figure 4.8D). There is strong VI-protection of nucleotides U60, U61, and U68, and general shielding of the remaining nucleotides 62 through 67, suggesting Gk-S15 binding and

shielding of this region. All of these results suggest Gk-S15 binds this portion of the RNA, and much of it is double stranded.

The footprinting data suggest that the central part of the RNA sequences, G36 through C47, folds into a hairpin. As Gk-S15 concentration is increased, there is increased VI cleavage product intensity of nucleotides G36, A37, C39, C47, indicating that Gk-S15 does not protect this region, and it is double stranded (Figure 4.8D). Also, in the presence of Gk-S15, the Pb-cleavage of the 11-1 RNA increases for nucleotide C41 and U43, there are IL-cleavage products for A40 through U43 (Figure 4.8B) as well as A-cleavage for C41 (Figure 4.8C). This all suggests the formation of a stem loop region. Taken together, nucleotides G36 through C47 likely fold into a hairpin that does not directly interact with Gk-S15 upon protein binding.

Mutagenesis Experiments Confirm Gk-S15 Binding Regions in 11-1

To confirm our secondary structure model, a variety of mutations to the 11-1 RNA sequence were designed and the ability for these to bind Gk-S15 was tested using a filter binding assay procedure identical to those used in the selection process (Figure 4.9A, B). Mutations to the 5'-region of the aptamer were first assessed. A 5' truncation of 11 nucleotides (11-1-M1), completely abolishes Gk-S15 binding, which suggests this region is critical for Gk-S15 binding. When taking our footprinting results into account, the RNA structure prediction program, RNAfold (Lorenz 2011), suggests this region of the RNA folds into a small hairpin. Therefore, we created a mutation to this region that prevents the putative double helix formation (11-1-M2) and does abolish Gk-S15 recognition. Furthermore, the compensatory mutation (11-1-M3) successfully restored

Gk-S15 binding. This strongly suggests U1 through A11 fold into a hairpin. Together these results suggest nucleotides U1 through A11 fold into a hairpin whose presence is required for Gk-S15 binding.

Truncations to the 3' end of the RNA sequence confirm that many of these nucleotides are not required for binding Gk-S15 (Figure 4.9). A 22 nucleotide 3'-truncation (11-1-M4) only slightly affected protein binding but a 29 nucleotide 3'-truncation (11-1-M5) abolishes binding. This suggests Gk-S15 does not require the 3' deleted by M4 to bind the RNA. This finding supports our footprinting assays that suggest this region remains unstructured. We also confirmed the putative stem loop region spanning G33 through U49 suggested by the Pb²⁺ and VI probing data is not required for binding Gk-S15. Replacement of this entire region with a GUAA sequence (11-1-M6) did not affect recognition by the protein.

Mutations to the central core of the putative RNA 11-1 structure drastically affect Gk-S15 binding (Figure 4.9). Because of the protection from nuclease cleavage we observed in the footprinting assays, we created a mutation to helix 2 (11-1-M7) that abolished Gk-S15 binding. There was also decreased cleavage of G53 through G65 in the presence of Gk-S15 (protection from both VI and Pb). When we mutated the opposite side of the helix, 11-1-M8, protein binding was only slightly affected. The compensatory mutant did not compensate for the RNA secondary structure, no S15 binding was apparent with 11-1-M9. Testing an alternative binding partners for the nucleotides C14-U15, 11-1-M10, also did not compensate the RNA structure (for all mutants tested see Figure 4.10, for alternative structure diagram see Figure 4.12, and see Figure 4.13 for footprinting data mapped to this alternative structure). Therefore, we may have mutated a

nucleotide-specific interaction for Gk-S15 in the C14-U15 of RNA 11-1. Based on these results and the footprinting data, we believe we have drawn the correct structure of RNA 11-1 upon Gk-S15 binding.

Footprinting Experiments of Non-regulatory RNA 11-2

RNA 11-2 was one of two synthetic RNA aptamers that survived the eleven rounds of selection, yet did not regulate gene expression *in vivo*. To better understand what about this interaction allows it to be relatively strong *in vitro* but non-functional *in vivo*, we performed footprinting assays on this RNA. Again, we used RNase VI, RNase A, RNase T1, and in-line probing on 5'-end labelled RNA sequence in the presence and absence of Gk-S15 (Figure 4.14A-D). Overall, the RNA appears unstructured in the absence of Gk-S15, which is especially evident in the Pb-cleavage footprint without Gk-S15 (Figure 4.14D) and the distinct number of strong IL-cleavage products (U21, C22, U25, U67, U68) (Figure 4.14B).

Protein binding does little to affect the RNA's secondary structure and our data suggest that Gk-S15 does not bind any part of the sequence that may be considered important for regulation. The most striking evidence for this is the Pb-footprint (Figure 4.14D). In particular, we see clear shielding of nucleotides 26 through 28 and of the region 49 through 55. This strongly suggests that the protein-bound structure of the RNA forces these regions into double stranded secondary structure. Additionally, there is protection of C51 from VI-cleavage (Figure 4.14C), which corroborates the Pb-footprint and suggests this region is important for binding Gk-S15. There is no apparent shielding or changes in cleavage patterns to the 3'-region of the RNA where transcription and

translational elements are located. More specifically, up to 500 nM Gk-S15 does little to affect the A- nor VI-cleavage patterns (Figure 4.14C). To corroborate our putative model for Gk-S15 binding to 11-2 in the C51 region, we designed and tested a mutation to this region of the RNA (11-2-M1). This mutation was sufficient to prevent binding by Gk-S15 ($K_D > 300$, $F_{MAX} 0.068 \pm 0.044$, Figure 4.15).

DISCUSSION

Our goal was to use *in vitro* selection to identify RNA aptamers that could be used *in vivo* to regulate gene expression in response to Gk-S15. We have successfully identified four RNA aptamers that regulate gene expression. Additionally, we have identified structural features that enable a surviving sequence to regulate as opposed to merely bind Gk-S15. More broadly, this regulatory aptamer may be useful in the design of novel synthetic genetic circuits and understanding the evolution of the regulatory RNA repertoire for S15 in diverse bacterial phyla.

An important choice during the *in vitro* selection experiment was to not proceed until a single sequence dominated the selected pool. We believe this increased the opportunity to find *in vivo* regulators. RNAs with modest binding affinity (such as 11-6) can regulate gene expression. Thus, selections that proceed until one high-affinity sequence dominates the pool may remove potentially functional *in vivo* regulators. Our data indicate that *in vitro* affinity does not necessarily need to be exceedingly high. Coupling a low-stringency *in vitro* selection with the use of an *in vivo* regulation screen may be a powerful way to identify regulatory RNA aptamers of interest in the future.

In the selection of aptamers for a protein of interest, future selection experiments may also benefit from an altered negative selection. Our regulatory RNAs all bind the same face of the protein, and likely interact with similar amino acids as the native regulatory RNA (i.e. Gk-mRNA). Additionally, many of our regulatory aptamers interact with some but not all homologs of S15. A negative selection using a protein mutant, or a distant homolog of the protein may improve isolation of regulatory aptamers of interest that respond to a protein of interest. Or a co-SELEX of both the RNA and the protein at the same time could isolate a completely novel, synthetic RNA:protein interaction if desired.

RNAs resulting from the selection that also found a way to occlude important regulatory features, including the putative Shine Dalgarno (SD) sequence, ATG start codon, and protein-coding nucleotides, are likely to have regulatory function. All of our synthetic regulators behave as genetic OFF switches, and, more specifically, our footprinting and site-directed mutagenesis experiments show regulatory RNA 11-1 occludes the SD sequence, whereas non-regulatory RNA 11-2 structure does not appear to interact with regulatory features. In the rational design or evolution of synthetic RNA regulators this must be taken into account: protein binding must occur in the 5'-most region of the RNA and result in occlusion or presentation of translation elements.

The S15:mRNA regulatory interaction is plastic, and many naturally-occurring regulatory RNA structures have been described that all perform analogous regulatory functions. It is not so surprising that we were able to isolate novel structures that also perform gene regulation in response to S15. What is surprising about our findings is that our synthetic regulatory RNA does not overtly share any of the features we know to be

important for gene regulation in the natural regulators (e.g. Gk-mRNA). The natural RNAs all share some identifiable partial mimicry with the rRNA binding site for S15. Our synthetic regulatory RNA, 11-1, does not share any obvious mimicry with a three helix junction, or with the G•U/G-C motifs found to be important for gene regulation in the natural regulatory RNAs (Slinger 2015, Chapter III). Our results suggest that novel *cis*-regulatory structures for S15 are readily derived, and that there are many structural solutions to this biological problem.

Our choice of the S15 homolog from *G. kaustophilus* likely affected our selected Round 11 RNA pool and our results suggest the mechanism of regulation Gk-S15 utilizes with these synthetic RNAs. Furthermore, the S15 homolog from *T. thermophilus* (Tt-S15) is able to regulate three of our synthetic regulatory RNAs, whereas the homolog from *E. coli* (Ec-S15) did not. Both Gk-S15 and Tt-S15 use a “displacement” mechanism where S15 directly competes for binding the mRNA transcript with the ribosome. Ec-S15 uses an “entrapment” mechanism to regulate expression of the *rpsO* operon in which both Ec-S15 and the pre-initiation complex of the ribosome bind the same mRNA simultaneously, which ultimately prevents full ribosome assembly and halts translation. Selecting for this type of regulatory mechanism *in vitro* is impossible, but biophysical modeling has shown this mechanism may allow lower affinity interactions to still regulate efficiently (Draper 1983) Additionally, our results show that Gk-S15 binding almost certainly leads to occlusion of the SD sequence, which highly suggests the mechanism of gene regulation occurs by a displacement mechanism.

MATERIALS & METHODS

Protein Overexpression and Purification

The *rpsO* open reading frame was PCR amplified using whole genomic DNA (see Figure 4.15) and cloned into pET-HT overexpression vector similarly to previously described (Block 2011). Sequence verified plasmid was transformed into chemically competent BL-21 cells (DE3). Protein expression and purification was conducted as described previously (Slinger 2014).

RNA Preparation and SELEX

RNA selection experiments proceeded using the template 5'-TAATACGACTCACTATAGGTGCGTAACGTACACT-N₃₀-TCATTCTATATACTTTGGAGTTTTAAAATGTCTCTAAGTACTGAAGCAACAGC T where N₃₀ represents 30 random nucleotides per position, and T7 RNA polymerase promoter sequence is underlined). Transcription reactions were performed using T7 polymerase (Milligan 1987), then purified by 6% denaturing PAGE. Bands were visualized using UV shadow, excised, and the RNA eluted (in 200 mM NaCl, 1 mM EDTA pH 8, 10 mM Tris-HCl pH 7.5) and ethanol precipitated.

Aptamers that bind Gk-S15 were obtained after 11 rounds of selection: 300 pmol of RNA were renatured 42°C 15 minutes, then filtered through 0.45 µM nitrocellulose to remove non-Gk-S15 binders. Surviving RNAs were incubated with Gk-S15 in Binding Buffer A (50 mM Tris-Acetate, pH 7.5, 20 mM Mg-Acetate, 270 mM KCl, 5 mM dithiothreitol, 0.02% bovine serum albumin) 25°C for 30 minutes then RNA-Gk-S15 complexes isolated by filtering with nitrocellulose. After 2 washes the bound RNAs were

eluted from the filter (7 M Urea, 100 mM Na₃C₆H₅O₇, 3 mM EDTA pH 8.0) and purified using isopropanol. The RNA aptamers were reverse transcribed using M-MuLV, cDNA amplified using Mutagenic PCR (Cadwell 2006), then this pool used to transcribe RNA for the next round of selection. cDNA was ligated into the pCR 2.1 vector (making pCR-RNA) to sequence individual survivors and to be used as template to test individual sequences (*in vitro* and *in vivo*).

Binding and Competition Assay

DNA corresponding to the 5'-UTR of the *rpsO* gene was PCR amplified using species-specific primers with the T7-promoter sequence added within the forward primer sequence (Figure S13). Genomic DNA extracted from the species was used as template. For all synthetic sequences the TOPO cloned RNA sequence (pCR-RNA) was used as template to amplify DNA. T7 RNA polymerase (Milligan 1987) was used to transcribe RNA and transcription reactions were purified and eluted as described in SELEX experiment. Purified RNA (10 pmol) was 5'-labeled with ³²P-ATP and purified as previously described (Regulski 2008). Binding assays were performed and 2+ replicates quantitated as previously described (Slinger 2014) using nitrocellulose and nylon membranes (GE Healthcare). Mutations to the mRNAs were constructed by site-directed mutagenesis (Figure 4.15).

Footprinting assays

The RNA-protein binding reaction described above was used for RNase probing assays. After incubation, 1 μL RNase A (1 ug/mL, Ambion) or VI (1:400 dilution of 0.1

U/uL, Ambion) was added and the reaction incubated 15 minutes at 25°C. The nuclease was inactivated with inactivation/precipitation buffer (Life Sciences) and RNA fragments recovered by ethanol precipitation. Precipitated RNAs were suspended in 10 uL Urea Loading solution (Life Sciences) and incubated 5 minutes 95°C. Reactions were loaded on 10% denaturing Acrylamide/Bis-acrylamide gel. The gel was dried and examined using a GE Healthcare STORM 820 phosphorimager and ImageQuant software. Partial hydroxyl cleavage reactions were generated by incubating RNA in Reaction Buffer (50 mM Na₂CO₃ pH 9.0, 1 mM EDTA) at 95°C for 7 minutes. Denaturing T1 reaction (1:10 dilution) was conducted according to the manufacture's protocol (Ambion). For in-line probing, 5'-labeled RNA was incubated 40 hours at 25°C in reaction buffer (20 mM MgCl₂, 100 mM KCl, 50 mM Tris-HCl pH 8.3). The reaction was stopped using Urea loading solution (10 M Urea, 1.5 mM EDTA).

Plasmid Construction

All synthetic sequences were cloned into the pBS3-RNA plasmid as a translational fusion with *lacZ* using primers containing EcoRI and SalI restriction sites and template from TOPO 2.1 cloned PCR product. The *lacZ* sequence requires a start codon from the fused *rpsO* sequence. All enzymes for molecular biology were purchased from New England Biolabs unless otherwise noted (Figure S13).

pS15 protein expression plasmids were constructed by amplifying the open reading frame from genomic DNA with a forward primer containing SacI site plus a strong ribosome binding site that was native to the organism or matching *E. coli* ribosome binding site and an 8 nucleotide linker (Figure 4.15) preceding the ATG start

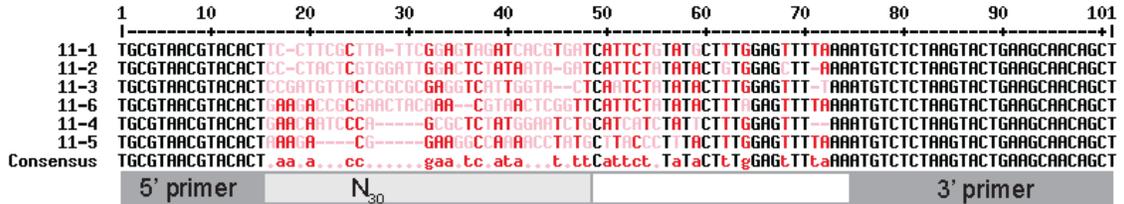
site and subsequent codons. The reverse primer contained an XbaI site. After digestion, the PCR product was cloned into the pBAD33 vector (ATCC 87402) digested with the same enzymes. All pS15 were sequence verified. S15 mutants were constructed using site-directed mutagenesis (Figure 4.15).

***LacZ* Regulatory Assays**

K12: $\Delta rpsO$ *E. coli* cells were co-transformed with pRNA and pS15 plasmid (made competent using the Z-competent buffer system, Zymo Research). A single colony was used to start overnight cultures, grown +/- L-arabinose (15 mM) at 37°C, then diluted the next day to OD₆₀₀ = 0.15 in fresh media (LB + 100 ug/mL ampicillin + 34 ug/mL chloramphenicol +/- 15 mM L-arabinose). At stationary phase (5 hours after dilution) 1 mM IPTG was added to induce β -galactosidase expression. After 30 minutes, 100 ug/mL spectinomycin was used to stop initiation of protein translation, and the cultures assayed immediately according to Miller (Miller 1992) to determine the levels of reporter expression. Fold repression = (Miller units of - L-arabinose)/(Miller units of + L-arabinose). All RNA/S15 combinations were examined with 3+ independent replicates. To determine the significance, all fold repression values were compared as indicated in Figure 4.16 using a Welch's single-tailed T-test in Microsoft Excel. Regulation was considered biologically significant if greater than 2.5-fold repression was observed, and the fold-repression was significantly different ($p < 0.05$) than that observed with an empty pBAD33 vector.

Figure 4.2 Individual RNAs Isolated from Round 11 Have Diverse Sequence and Predicted Structure

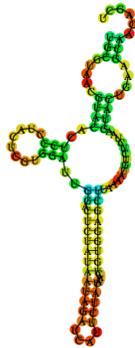
(Top) Six sequences were isolated from the Round 11, aligned below using MultAlin (Corpet 1988). High consensus (black, e.g. the 5' and 3' primers), low consensus (red), no consensus (pink). Primer regions and N₃₀ region indicated. (Bottom) The predicted structure (minimum free energy) for all sequences was analyzed using RNAfold (Lorenz 2011) and is shown.



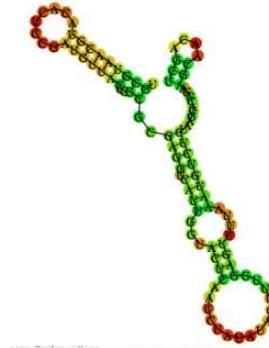
>11-1



>11-2



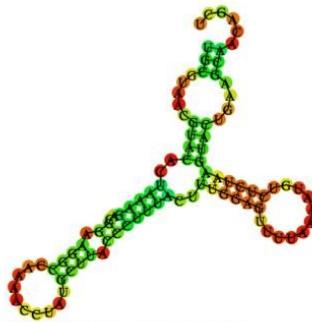
>11-3



>11-4



>11-5



>11-6



Figure 4.3 RNAs from Round 11 Bind Gk-S15

(A) Binding affinity (K_D), Standard Error (STE), and Maximum Fraction Bound (F_{MAX}) for six Round 11 RNAs with Gk-S15. (B) Binding curves for Gk-S15 binding assays with RNAs Gk-mRNA (black), Ec-mRNA (blue), 11-1 (pink), 11-2 (orange), 11-3 (red), 11-4 (yellow), 11-5 (green), and 11-6 (purple). See Materials & Methods for details on the calculations for K_D and F_{MAX} .

A.

RNA	K_D (nM)	STE	F_{MAX} (%)	STE
Gk-mRNA	0.7	0.02	66	5.7E-03
Ec-mRNA	>100	n/a	32	1.0E-02
11-1	0.9	0.02	70	4.9E-02
11-2	9.7	1.6	79	3.8E-03
11-3	20.7	15.2	40	4.3E-02
11-4	8.5	2.0	21	1.2E-02
11-5	10	1.4	44	1.6E-02
11-6	289	121	55	4.0E-02

B.

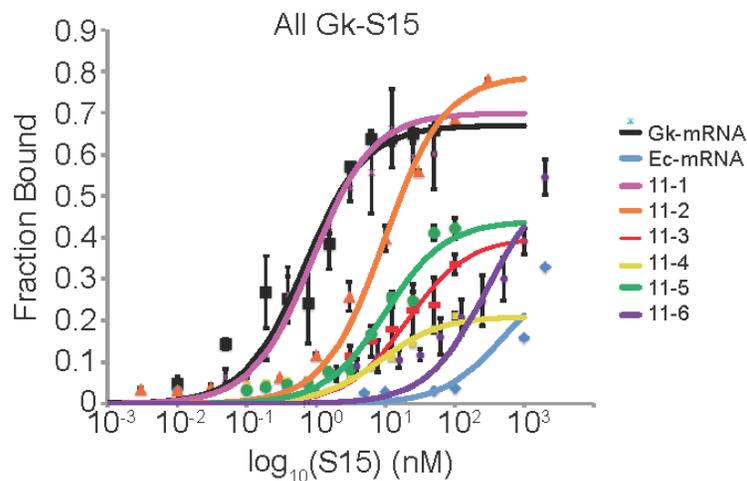


Figure 4.4 *In vivo* Regulation Assays for Synthetic RNAs

In vivo regulation assays (fold-repression) using *lacZ* reporter and *E. coli* host cells containing the indicated pRNA and the pS15. (A) The six individual sequences isolated from the round 11 pool were assessed *in vivo* in response to Gk-S15 (pGk-S15, dark gray). Each RNA is compared to pEMPTY (see Figure 4.5C, 4.16). (B) The regulatory response of RNAs 11-1, 11-4, and 11-6 with S15 homologs from *T. thermophilus* (pTt-S15, light gray) and *E. coli* (pEc-S15, white). Data corresponding to pGk-S15 and pEMPTY interactions is re-plotted from Figure 4.4A for comparison. All error bars correspond to standard error for 3 or more replicates.

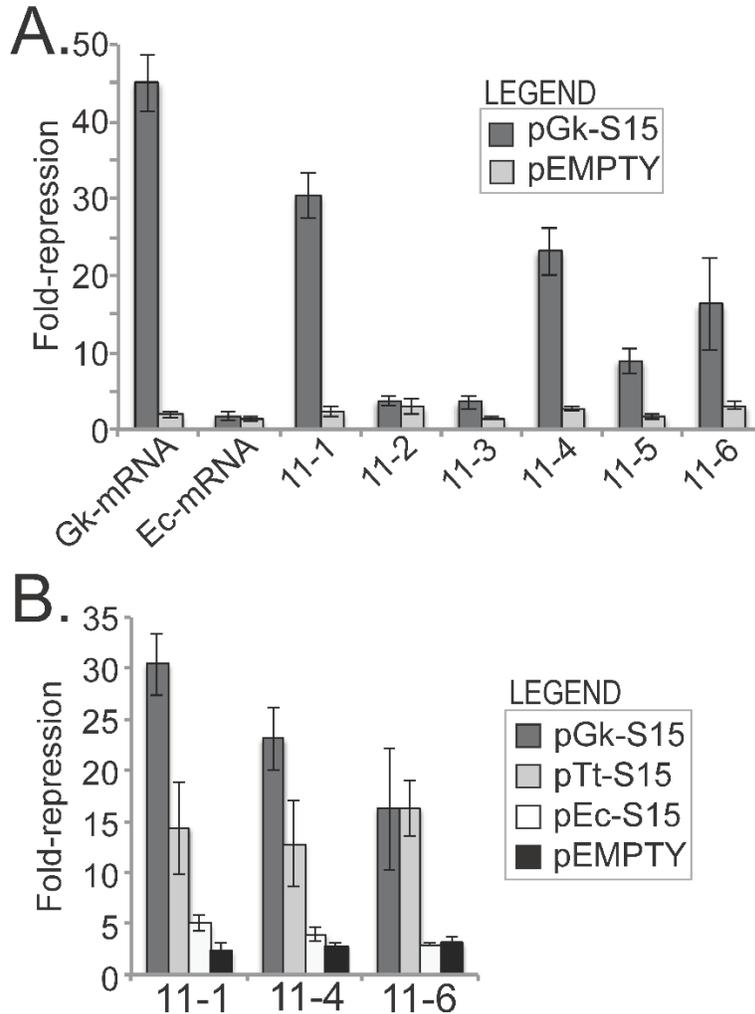


Figure 4.5 All Miller Units, + and - L-arabinose

All Miller Units, + and - L-arabinose, from *in vivo* regulation assays that were used to calculate fold-repression values. For all, solid bars indicate + L-arabinose conditions, striped bars indicate -L-arabinose conditions. (A) pGk-S15 and pEMPTY interactions with Gk-mRNA, Ec-mRNA, RNA 11-1, and RNA 11-2, dark gray bars are pGk-S15 and white/light gray bars are pEMPTY (B) pGk-S15 and pEMPTY interactions with RNA 11-3, RNA 11-4, RNA 11-5, and RNA 11-6, dark gray bars are pGk-S15 and white/light gray bars are pEMPTY (C) pTt-S15 and pEc-S15 interactions with RNA 11-1, RNA 11-4, and RNA 11-6, light gray bars are pTt-S15, dark gray bars are pEc-S15 (D) RNA 11-1 interactions with pGk-S15-E40L, pGk-S15-D48L, pGk-S15-Y68A, and pGk-S15-E79L.

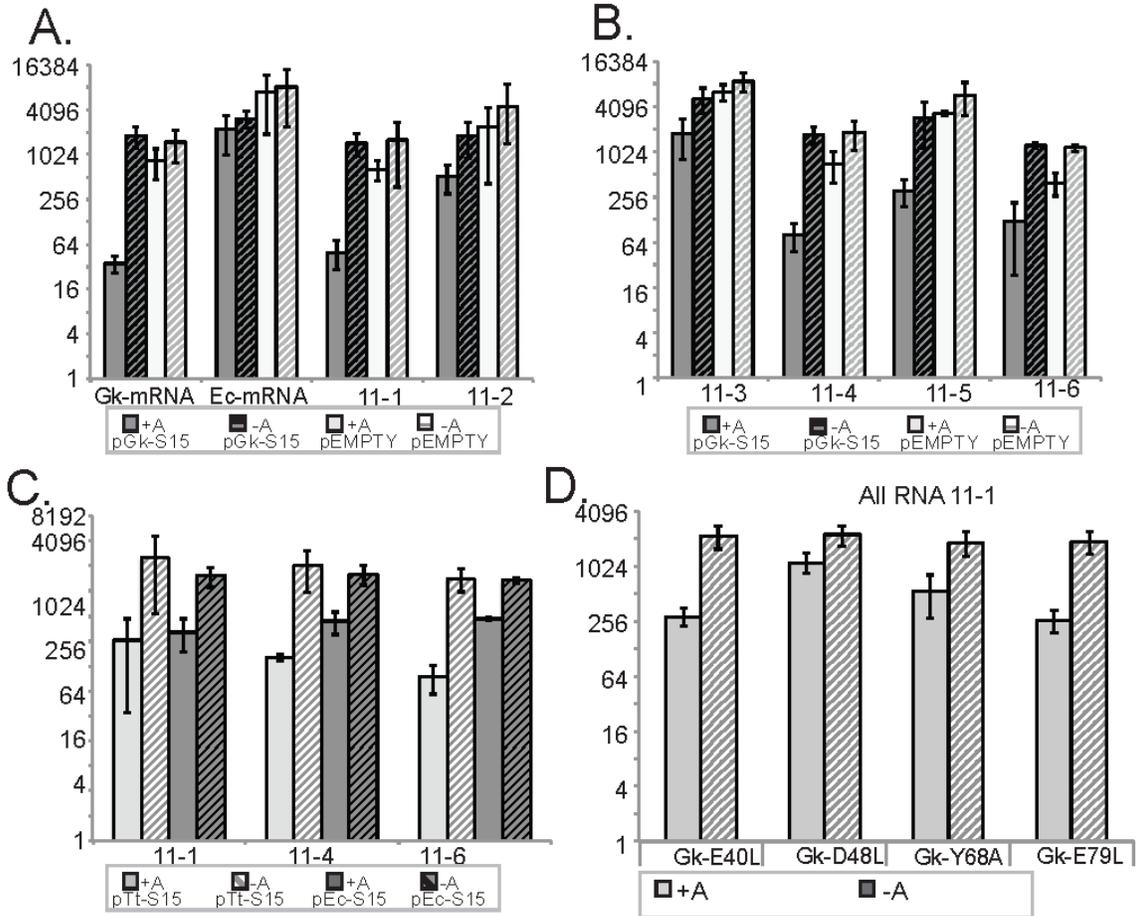


Figure 4.6 RNA 11-1 Interacts with Same Face of Gk-S15 as Gk-mRNA

In vitro competition binding data and *in vivo* regulation data to assess how synthetic RNAs 11-1 and 11-2 interact with Gk-S15. (A) Titration of unlabeled competitor, “Inh”: Gk-mRNA or 11-1, with 32 P-labeled 11-1 and Gk-S15. (B) Titration of unlabeled competitor, “Inh”: Gk-mRNA or 11-1, with 32 P-labeled 11-2 and Gk-S15. (C) *In vivo* regulation assay for RNA 11-1 with pGk-S15-Mutants indicated. Data corresponding to pGk-S15 and pEMPTY is re-plotted from Figure 4.4A for comparison. All error bars correspond to standard error for 3 or more replicates.

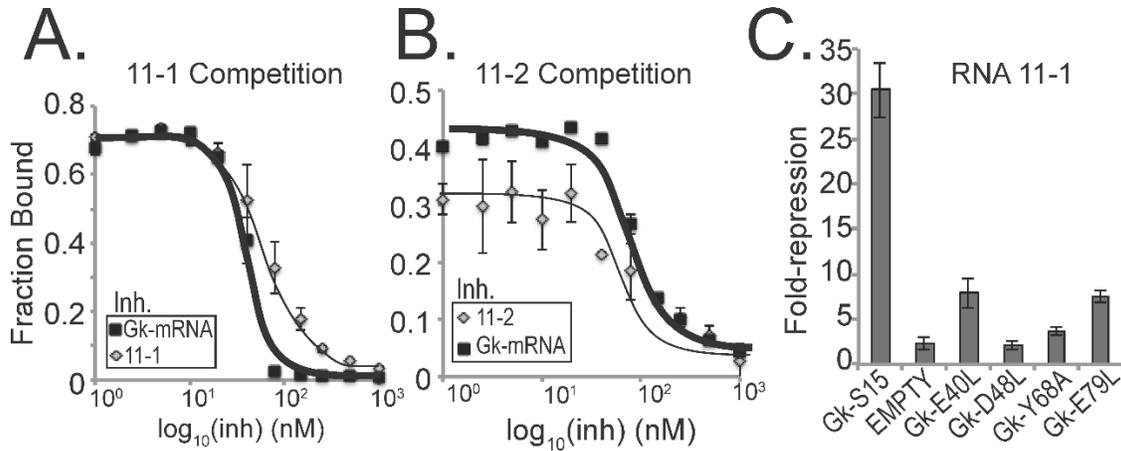


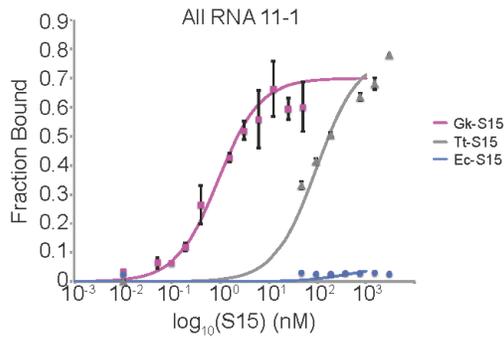
Figure 4.7 RNAs 11-1 and 11-2 bind Tt-S15, not Ec-S15

In vitro binding interactions with RNA 11-1, 11-2 and the S15 homologs from *T. thermophilus* (Tt-S15) and *E. coli* (Ec-S15). (A) Table of binding affinity (K_D), Standard Error (STE), and Maximum Fraction Bound (F_{MAX}). Binding curves for RNA 11-1 (B) and 11-2 (C). Data corresponding to Gk-S15 is re-plotted from Figure 4.3B for comparison.

A.

RNA	Protein	K_D (nM)	STE	F_{MAX}	STE
11-1	Tt-S15	94.9	22.5	0.78	3.7E-03
11-1	Ec-S15	>3000	n/a	n/a	n/a
11-2	Tt-S15	39	6.27	0.25	6.5E00
11-2	Ec-S15	>300	n/a	n/a	n/a

B.



C.

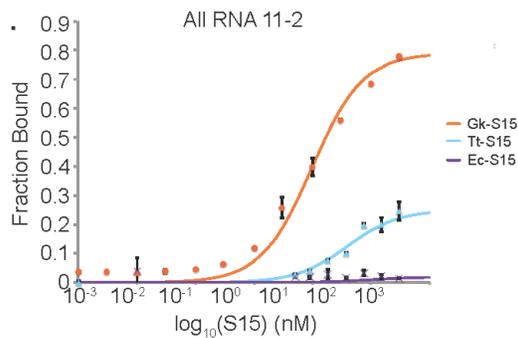


Figure 4.8 Structure Probing Elucidates the Secondary Structure of RNA 11-1

For all individual gels, no reaction (N), hydroxyl cleavage (OH), and denaturing RNase T1 (T1), all cleavage products have been separated by denaturing 10% PAGE. (A) Predicted RNA 11-1 structure with all footprinting data mapped to the structure, (B) Two independent replicates of in-line probing reactions (IL), (C) RNase VI (V1), RNase (A) in the absence of Gk-S15, (D) Titration of Gk-S15 with RNase VI, where protein concentration (nM) is indicated (E) Lead(II)-probing reactions (Pb²⁺) in the presence and absence of 200 nM Gk-S15.

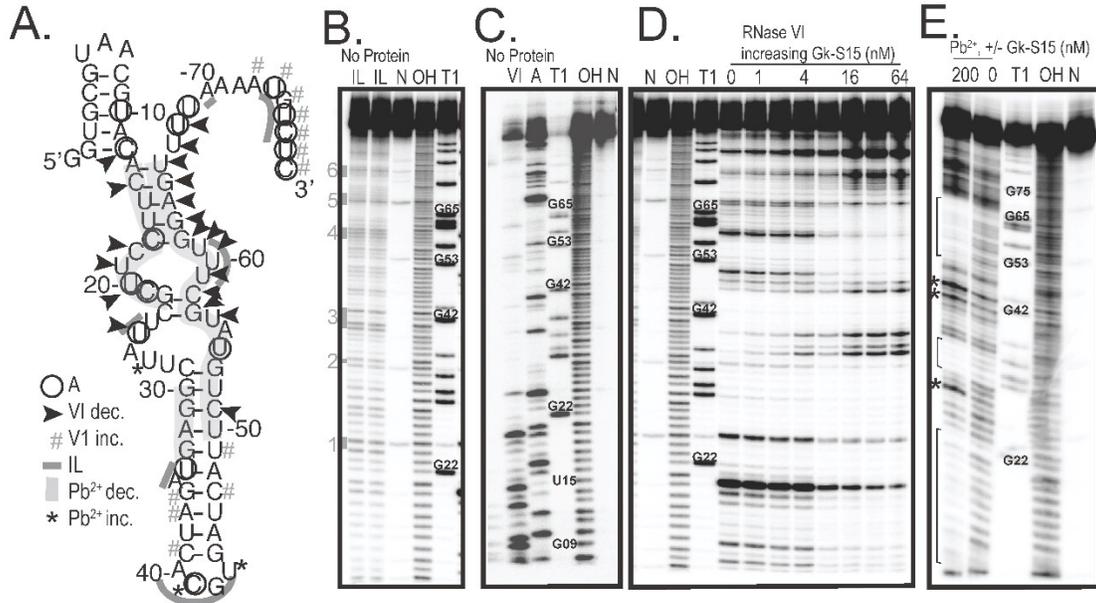
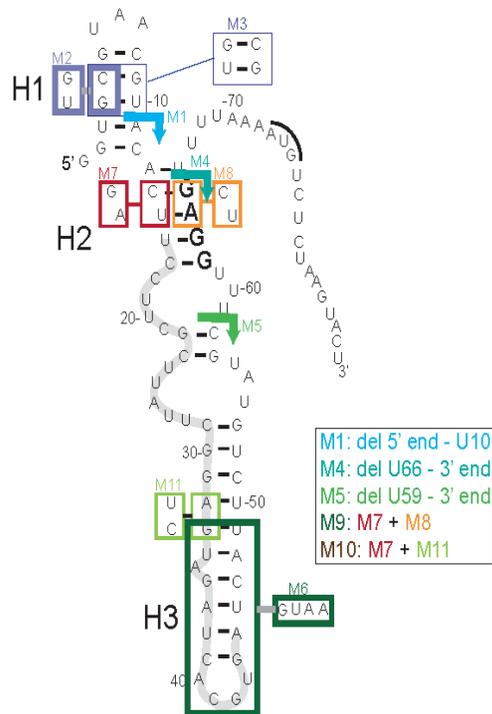


Figure 4.9 *In vitro* Binding Assays and Mutagenesis Confirms the Predicted RNA 11-1 Structure

In vitro binding assays were performed with Gk-S15 and mutant versions of RNA 11-1. These results largely confirm our footprinting results. (A) Truncation sites and specific mutations to 11-1 are shown. The Shine Dalgarno sequence is bolded, a bar is placed over the AUG start codon, and putative helices H1, H2, and H3 are indicated. The resulting N₃₀ region in RNA 11-1 is highlighted in gray, U16-A45. (B) Table of values determined from binding assays, binding affinity (K_D), standard error (STE), and maximum fraction bound (F_{MAX}). See Figure 4.11 for individual binding curves used to calculate these values.

A.



B.

RNA	K _D (NM)	STE	F _{MAX}	STE
11-1	0.9	0.02	0.70	4.9E-02
M1	>100	n/a	n/a	n/a
M2	56.5	5.74	0.56	2.0E-02
M3	20.9	6.3	0.85	3.7E-03
M4	7.7	5.89	0.36	5.8E-03
M5	>100	n/a	n/a	n/a
M6	4.63	0.8	0.69	2.3E-02
M7	>300	n/a	n/a	n/a
M8	84.5	14.5	0.71	1.4E-02
M9	>300	n/a	n/a	n/a
M10	>300	n/a	n/a	n/a

Figure 4.10 All RNA 11-1 Mutant Binding Assays with Gk-S15

Table of values for *in vitro* binding assays for all RNA 11-1 mutants with Gk-S15. See Figure 4.11 for binding curves.

RNA	K _D (nM)	STE	F _{MAX}	STE	Notes
11-1	0.9	2.0E-02	0.70	2.1E-03	
11-1-M1	>100	n/a	n/a	n/a	Del 5' end – U10
11-1-M2	56.5	5.74	0.56	2.0E-02	G2-C3 → U2-G3
11-1-M3	20.9	6.3	0.85	3.7E-03	11-1-M2 + G9-U10 → C9-G10
11-1-M4	7.7	5.89	0.36	5.8E-03	Del U66 – 3' end
11-1-M5	32	34	0.064	4.3E-02	Del U59 – 3' end
11-1-M6	4.63	0.8	0.69	2.3E-02	Del G33-U49 → Ins GUAA
11-1-M7	>300	n/a	n/a	n/a	C14-U15 → G14-A15
11-1-M8	84.5	14.5	0.71	1.4E-02	A64-G65 → U64-C65
11-1-M9	>300	n/a	n/a	n/a	M7 + M8
11-1-M10	>300	n/a	n/a	n/a	M7 + M11
11-1-M11	>300	n/a	n/a	n/a	A32-G33 → U32-C33
11-1-M12	0.6	8.6E-02	0.42	1.0E-02	Del G83 – 3' end
11-1-M13	0.81	2.5E-01	0.52	1.5E-02	Del U78 – 3' end
11-1-M14	>300	n/a	n/a	n/a	Del C14-U28
11-1-M15	>300	n/a	n/a	n/a	Del G42-U59
11-1-M16	4.74	2.3	0.77	1.0E-02	A40-U43 → GGGA
11-1-M17	3.94	0.50	0.59	4.0E-02	A37-U46 → del
11-1-M18	10.9	2.7	0.61	3.0E-02	U10 → C10
11-1-M19	68.8	20.1	0.69	1.7E-02	M13 + M6

Figure 4.11 Binding Curves for All RNA 11-1 Mutants with Gk-S15

(A) Mutants 1-5 (B) Mutants 6-10 (C) Mutants 11-15 (D) Mutants 16-19. See Materials and Methods for calculation of fraction bound.

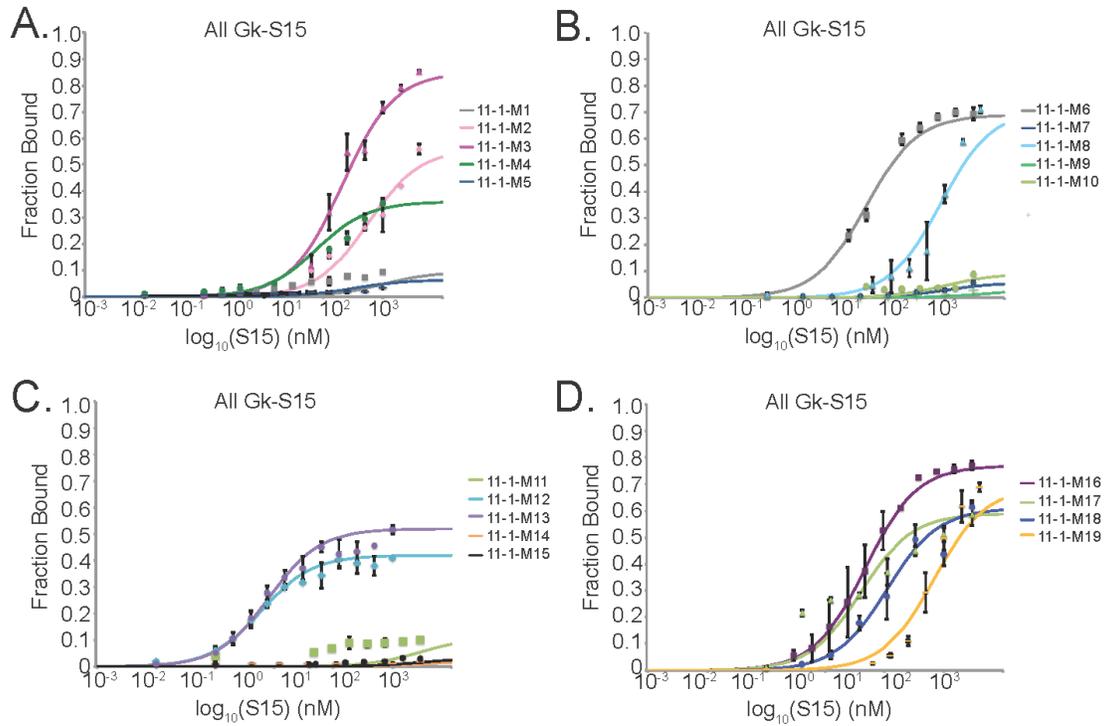


Figure 4.12 *In vitro* Mutagenesis Assays with Gk-S15 Plotted to an Alternative Potential Secondary Structure for RNA 11-1

Truncation sites and specific mutations to 11-1 are shown with an arrow. The Shine Dalgarno sequence is bolded, a bar is placed over the AUG start codon.

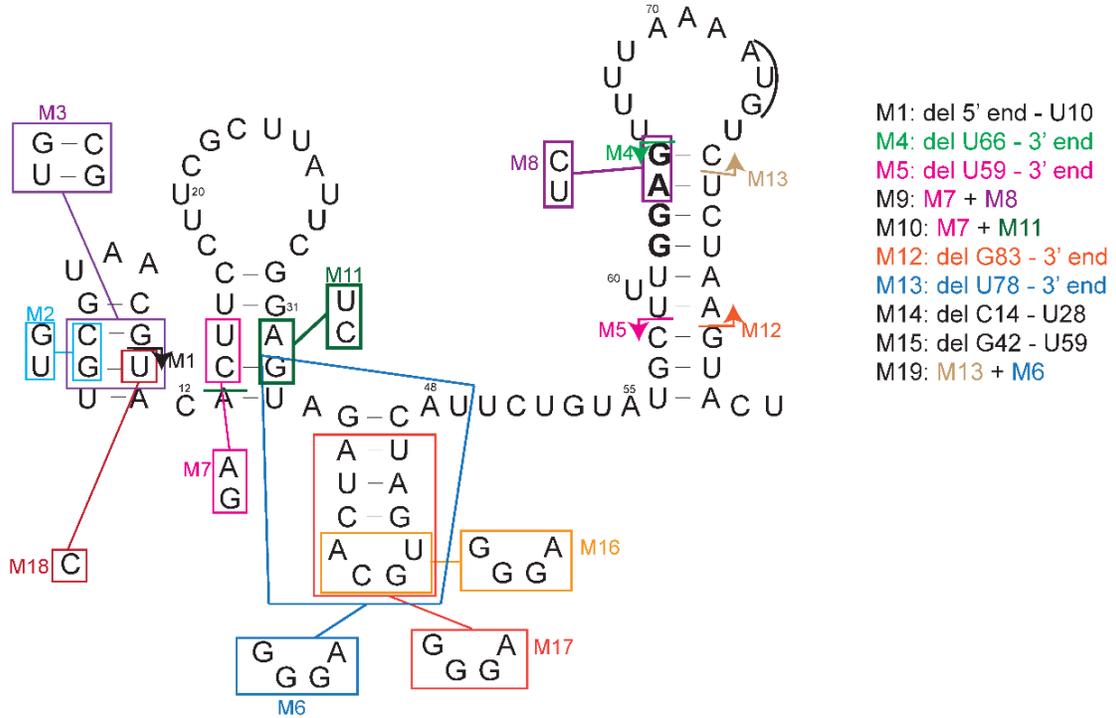


Figure 4.14 Structure Probing Assays Elucidate the Secondary Structure of RNA 11-2 in the Presence and Absence of Gk-S15

For all individual gels, no reaction (N), hydroxyl cleavage (OH), and denaturing RNase T1 ($T1^{SEQ}$), all cleavage products have been separated by denaturing 10% PAGE. (A) RNase T1 (T1), RNase A (A), and RNase VI (VI) cleavage products in the absence of Gk-S15. (B) Two independent replicates of in-line probing reactions (IL), (C) RNase VI (VI), RNase (A) in the presence and absence of 200 nM Gk-S15, (D) Lead(II)-probing reactions (Pb^{2+}) in the presence and absence of 200 nM Gk-S15.

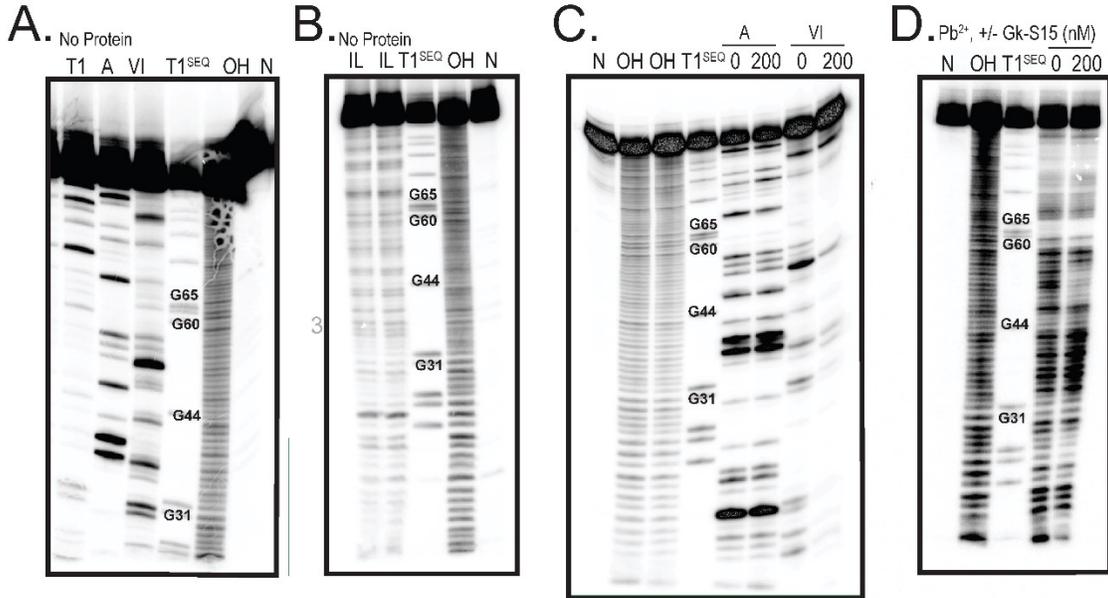


Figure 4.15 Binding curves for Gk-S15 binding assays with RNA 11-2-M1
Data for RNA 11-2 is re-plotted from Figure 4.3B for comparison.

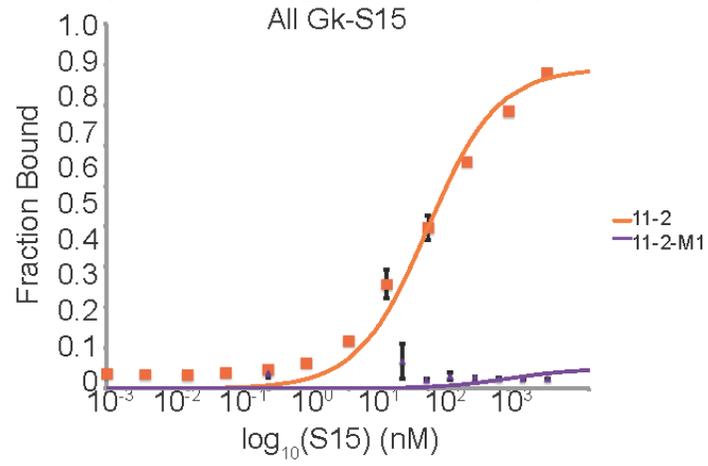


Figure 4.15 Table of Primers and Reference Sequences

(A) To amplify SELEX population (B) For cloning and sequences from pCR.2.1 TOPO vector (C) Site-directed mutagenesis of RNA 11-1 (D) Site-directed mutagenesis of RNA 11-2 (E) For cloning synthetic RNAs into pBS3 vector (F) For construction of pS15 vector (G) Site-directed mutagenesis of pGk-S15

A. To amplify SELEX population

Name	Sequence
361	5' -TAATACGACTCACTATAGGTGCGTAACGTACACT
509	5' -AGCTGTTGCTTCAGTACTTAGAGACATT

B. For cloning and sequencing from pCR.2.1 TOPO vector

Name	Sequence
34	5' -GTTTTCCCAGTCACGACGTTGTA
35	5' -CAGGAAACAGCTATGAC

C. RNA 11-1 site directed mutagenesis

Name	Sequence
836-M1F	5' -CAAGTAATACGACTCACTATAGGACACTTCCTTCGCTTATTCGGAGTAG
1398-m2F	5' -TAATACGACTCACTATAGGTtgGTAACGTACACT
1399-M3F	5' -TAATACGACTCACTATAGGTtgGTAACcgACACT
861-M4R	5' -CTCCAAAGCATAACAGAAATGATCACG
862-M5R	5' -GCATACAGAATGATCACGTGATCTAC
1105-M6F	5' -GCTTATTCG GAG TAA TCT GTA TGC TTTGGAG
1106-M6R	5' -CTCCAAAGCATAACAGATTACTCCGAATAAGC
1252-M7F	5' -TAATACGACTCACTATAGGTGCGTAACGTACAGATCCTTCG
1253-M8R	5' -AGTACTTAGAGACATTTTAAAAGACCAAAGC
11-1-M9	1253R ON M7 TEMPLATE
11-1-M10	1252 ON M11 TEMPLATE
1465-M11F	5' -CGCTTATTCGGTCTAGATCACG
1466-M11R	5' -CGTGATCTAGACCGAATAAGCG
859-M12R	5' -TTAGAGACATTTTAAAAGCATAACAG
860-M13R	5' -GACATTTTAAAAGCATAACAGAAATG
1083-m14F	5' -tgcgtaacgtacacttctagatcacgtgatcattctgtatgctttggagtttt
1084-m15F	5' -tgcgtaacgtacacttcttctcgttattcggagtagatcacttggagtttt
1101-M16F	5' -GAG TAG ATC GGG AGA TCA TTC TGT ATG
1102-M16R	5' -CATACAGAATGATCTCCCGATCTACTC
1103-M17F	5' -CTT ATT CGG AGT AGC ATT CTG TAT GC
1104-M17R	5' -GCATACAGAATGCTACTCCGAATAAG
1250-M18F	5' -TAATACGACTCACTATAGGTGCGTAACGCACACT
11-1-M19	860R ON M6 TEMPLATE

D. RNA 11-2 site directed mutagenesis

Name	Sequence
1463-11-2-M1F	5' -AATAGATCATTCGGGATACTGTGGAGC
1464-11-2-M1R	5' -GCTCCACAGTATCCCGAATGATCTATT

E. To clone surviving sequences in vivo:

Name	Sequence
673F	5'-caagaattcTGCGTAACGTACACT
52R	5'-ACGCGTCGACAGCTGTTGCTTCAGTACTTAGAGACA

F. To clone pS15

Plasmid	Primer	Primer Sequence	Reference Sequence
pEc-S15	411F	5'-CACGAGCTCAGGAGGTTTTAAA ATGTCTCTAAGTACTGAAGCACAG	Atgtctctaagtactgaagcaaca gctAaaatcgtttctgagtttggg cgtgacgcaaacgacaccggttct accgaagttcaggtagcactgctg actgcacagatcaaccacctgcag ggcactttgcagagcacaacaaaa gatcaccacagccgctcgtggctg ctgcgcatggtttctcagcgtcgt aaactgctcgactacctgaaacgt aaagacgtagcagttacaccag ctcatcgagcgcttgggtctgcgt cgtaa
	18R	5'-GCTCTAGATTAGCGACGCAGACCC AGGCGC	
pTt-S15	1372F	5'-CACGAGCTCAGGAGGTTTTAAAAtgc ccatcacgaaggaagag	Atgcccacacgaaggaagagaag Cagaaggtcatccaggagtccgcc Cgcttccccggggacacggggagc Accgaggtgcaggtggcgctcctt Accctgaggatcaaccggctttcc Gagcacctcaaggtccacaagaag Gaccaccactcccaccgcggtctc Ctgatgatggtgggcccagcgccgc Aggctcctccgctacctccagcgg Gaggaccccgagcggtaaccgggccc Cttattgagaagctgggcatccgg ggttaa
	565R	5'-cacgagctcggagggaaaacatgccca tcacgaaggaagag	
pGk-S15	1371F	5'-CACGAGCTCAGGAGGTTTTAAAAtggcattga cgcaggagcgc	Atggcattgacgcaggagcgcaaa Cgcgaaatcatcgagcagtttaa Atccatgagaacgacactggttct Cgggaagtgcagttgcatcctg Acggagcaaatcaacaacttgaac Gagcatttgcgcattcataaaaa Gaccatcattcacggcgcggttg Ctgaatggtcgggaagcgccgc Aacttattggcctacttgcgcaag Aaagatgtggcgcgctaccgtgaa Ttgattgagaaacttgattacgt cgataa
	23R	5'-GCTCTAGATTATCGACGTAATCCAAG TTTCTCAATC	

G. pGk-S15 site-directed mutagenesis

Name	Sequence
789-GKrpsOE40L-F	5' -GAGCAAATCAACAACCTGAACCTGCATTTGCGCATTC
790-Gk-rpsOE40L-R	5' -GAATGCGCAAATGCAGGTTCAAGTTGTTGATTTGCTC
791-GKrpsOD48L-F	5' -gcgcatcataaaaaactccatcattcacggcgcggc
792-GKrpsOD48L-R	5' -gccgcgccgtgaatgatggagttttttatgaatgcg
793-GKrpsOY68A-F	5' -cgcaacttattggccgccttgcgcaagaaagatgtg
794-GkrpsOY68A-R	5' -cacatctttcttgcgcaaggcgccaataagttg
795-GkrpsOE79L-F	5' -GTGGCGCGCTACCGTCGTCTTTGATTGAGAACT
796-GkrpsoE79L-R	5' -AGTTTCTCAATCAAAGACGACGGTAGCGGCCAC

Figure 4.16 Statistics for Data on Figure 4.4

P-value calculations using data indicated from Figure 4.5, as well as old-repression values (Fold) and standard error (STE) used to construct bar graphs.

RNA	pS15	FOLD	STE	P Value	pRNA/pS15
Gk-mRNA	pGk-S15	45.08	3.74	3.45E-03	v. Gk-mRNA/pEMPTY
	pEMPTY	1.92	0.40		
Ec-mRNA	pGk-S15	1.72	0.53	3.23E-03	v. Gk-mRNA/Gk-S15
	pEMPTY	1.36	0.37	3.01E-01	v. Ec-mRNA/pEMPTY
11-1	pGk-S15	30.44	3.05	1.63E-02	v. Gk-mRNA/Gk-S15
	pTt-S15	14.39	4.53	1.62E-02	v. 11-1/pEMPTY
	pEc-S15	3.75	1.57	2.92E-02	v. 11-1/pEMPTY
	pGk-E40L	7.91	1.68	1.83E-02	v. 11-1/pEMPTY
	pGk-D48L	2.05	0.42	3.78E-01	v. 11-1/pEMPTY
	pGk-Y68A	3.68	0.54	9.09E-02	v. 11-1/pEMPTY
	pGk-E79L	7.51	0.65	6.35E-04	v. 11-1/pEMPTY
	pEMPTY	2.30	0.72	2.60E-04	v. 11-1/pEMPTY
11-2	pGk-S15	3.69	0.62	3.41E-03	v. Gk-mRNA/Gk-S15
	pEMPTY	2.91	0.92	2.63E-01	v. 11-2/pEMPTY
11-3	pGk-S15	3.50	0.93	3.12E-03	v. Gk-mRNA/Gk-S15
	pEMPTY	1.42	0.21	1.98E-02	v. 11-3/pEMPTY
11-4	pGk-S15	23.13	3.07	5.71E-03	v. Gk-mRNA/Gk-S15
	pTt-S15	12.83	4.19	6.84E-02	v. 11-4/pEMPTY
	pEc-S15	3.16	1.04	9.14E-02	v. 11-4/pEMPTY
	pEMPTY	2.72	0.28	1.06E-02	v. 11-4/pEMPTY
11-5	pGk-S15	8.92	1.65	2.03E-03	v. Gk-mRNA/Gk-S15
	pEMPTY	1.69	0.41	9.36E-03	v. 11-5/pEMPTY
11-6	pGk-S15	16.27	6.00	2.27E-03	v. Gk-mRNA/Gk-S15
	pTt-S15	16.31	2.72	7.72E-03	v. 11-6/pEMPTY
	pEc-S15	2.91	0.10	3.26E-01	v. 11-6/pEMPTY
	pEMPTY	3.13	0.49	3.51E-02	v. 11-6/pEMPTY

CHAPTER V

Discussion

A Summary: Ribosomal Protein S15 Co-evolved with its Regulatory RNAs

This thesis advances our understanding of the co-evolution between ribosomal protein S15 and its regulatory RNAs in diverse bacterial phyla. The synthesis of ribosomal components is tightly regulated by bacteria to maintain the correct stoichiometric levels and to ensure economical use of energy and nutrients (Figure 1.4). For many r-proteins, there is mimicry between the rRNA binding site and its regulatory RNAs. However, the structure of the S15 regulators does not overtly mimic the bipartite binding motif in the 16S rRNA consisting of a three helix junction (3HJ) and a G•U/G-C motif (Figure 1.2, 1.3).

Though the method of autoregulation of the *rpsO* operon has been understood in *E. coli* for over 40 years, we are only now beginning to understand the diversity of regulatory RNA structures with which S15 interacts and how they evolved. At the onset of this thesis work, three different structured regulatory RNAs for S15 had been described in *E. coli*, *T. thermophilus*, and *G. kaustophilus* (Figure 1.6). We have shown that these three structures are narrowly distributed to the Phyla Gammaproteobacteria, Thermales, and Firmicutes, respectively (Figure 1.5). With the work presented in Chapter II of this thesis, we broaden our understanding of the number of different structures that regulate *rpsO* expression. We have identified three additional regulatory RNAs in the Phyla Alphaproteobacteria, Actinobacteria, and Chlamydia (Figure 2.2). Additionally, in Chapter II, we experimentally validate the regulatory function for an example of the regulatory from Alphaproteobacteria, found in *R. radiobacter* (Figure 2.6), and show it performs gene regulation using a novel structure (Figures 2.4, 2.7). Furthermore, we find these different structures to be narrowly distributed to their respective bacterial phyla

(Figure 2.8). The work presented suggests there are a variety of regulatory RNA structures capable of interacting with a conserved protein to perform analogous biological function.

The observation that an S15 homolog from one species is not able to regulate using the mRNA from another species calls into question how exactly S15 recognizes its mRNA—do the mRNAs conceal mimicry to rRNA, or are there phyla-specific binding profiles? To answer these questions, in Chapter III, “Co-evolution of Ribosomal Protein S15 with Diverse Regulatory RNA Structures”, I provide evidence as to how these diverse regulatory RNA structures co-evolved with protein S15. Based on evidence gained from *in vivo* and *in vitro* cross-species RNA-protein interactions, we find that despite their shared RNA binding function in the rRNA, S15 homologs have distinct RNA recognition profiles (Figure 3.6, 3.7). In addition, the same RNA may be recognized in a different manner by the S15 homologs, such as the RNA from *G. kaustophilus* (Figure 3.6, 3.7). I go on to present a model for mRNA recognition and regulation with ribosomal protein S15 in diverse bacterial phyla. We find each regulatory RNA at least partially mimics the rRNA, and which portion the mRNA mimics determines which S15 homologs with which it interacts (Figure 3.8). Additionally, we find that specific amino acids are conserved in different phyla, and this contributes to the differences we see in RNA recognition (Figure 3.10). This work shows that the differences between S15 homologs determines the RNAs it can recognize and may contribute to the diversity of regulatory RNAs that arise across different bacterial phyla.

Our model explains the behavior of the modern versions of the RNA and S15 homologs, yet an open question is how novel ligand-binding RNAs arise and become

regulatory elements. In Chapter IV, “Synthetic *cis*-regulatory RNAs for Ribosomal Protein S15”, I investigate the derivation of novel *cis*-regulatory RNA structures for S15 using *in vitro* evolution. I find the ability to bind S15 is distinct from the ability to regulate gene expression, for many different RNA sequences bind S15 (Figure 4.3). Structure probing experiments of an *in vivo*-functioning, synthetic RNA elucidate these requirements, and show its structure appears distinct from those previously described (Figure 4.8, 4.9). To regulate gene expression, an RNA must sequester important translation features upon protein binding and interact with the rRNA-binding face of the S15 (Figure 4.6). The work presented in this chapter underscores the ease of evolving novel structures that regulate gene expression in response to S15. Furthermore, our results suggest that regulation of the *rpsO* operon has arisen several different times in bacteria.

Based on the natural diversity of S15-interacting RNAs, and our ability to evolve four novel *cis*-regulatory RNAs, it is clear that there are many structural solutions that allow S15 recognition and subsequent regulation. This thesis work indicates that a diversity of distinct regulatory RNAs are likely to exist within bacterial genomes and the plasticity of RNA structure allows distinct, and likely independently-derived, solutions to the same biological problem. Despite its overall amino acid conservation, there are differences between the S15 homologs from diverse bacterial phyla and these differences contribute to the differences in mRNA recognition. This thesis work suggests that either S15 has co-evolved with its mRNA regulators, or that differences between the ancestral S15 proteins lead to the development of a diverse array of RNA regulators that we observe in nature today.

Evolution of Distinct Regulatory RNAs for Ribosomal Protein S15

This thesis has explored how r-protein S15 regulates expression of its *rpsO* operon in diverse bacterial phyla. At the onset of this work, three different regulatory RNA structures had been described in *E. coli*, *G. stearothermophilus*, and *T. thermophilus*, which were found to be narrowly distributed to their respective phyla, Gammaproteobacteria, Firmicutes, and Thermales (Figure 1.5). With work presented in this thesis, we now have a better understanding how such diverse regulatory RNA structures evolved yet are able to perform analogous function in different bacteria. We think part of the answer is that the plastic nature of RNA enables the evolution of distinct regulatory RNA structures that conceal at least partial mimicry to rRNA. However, the S15 homologs themselves certainly influence the evolution of their regulatory RNAs, as does the mechanism of regulation, in conjunction with the environment in which the bacterium lives.

No crystal structures have been solved for any of the regulatory RNAs with which S15 from any species interacts; however, the regulatory RNA from *E. coli* (Ec-mRNA and Ec-S15) was the best characterized at the onset of this thesis work. A variety of *in vitro* binding, mutagenesis, and structure probing assays have elucidated the pseudoknotted nature of the Ec-mRNA as well as established the importance of the G•U/G-C motif to S15 recognition (Portier 1990, Philippe 1990, Philippe 1995, Serganov 2002). Our results confirm the importance of this motif to gene regulation in *E. coli* (Chapter III), for Ec-S15 is only able to regulate gene expression in response to mRNAs that contain a G•U/G-C motif (i.e. not Tt-mRNA, Figure 3.8). We solidify this hypothesis with a second mutation to the RNA from *G. kaustophilus* (a close relative to

G. stearothermophilus). When we mutate solely this region of the RNA from this species (Gk-mRNA-M2) gene expression is de-regulated (Figure 3.9). G•U/G-C motif aside, an open question is how the overall pseudoknot architecture itself evolved? The answer may be tied to the mechanism *E. coli* utilizes to regulate the *rpsO* operon, an entrapment mechanism (Chapter I). The pseudoknot structure uses nucleotides that are also protein-encoding (Figure 1.6). This not only pressures those nucleotides (and resulting amino acids) to remain ones that will base pair and form the pseudoknot, but also leaves the important translational elements (ATG start codon, and Shine Dalgarno sequence) available for the ribosome to partially assemble upon. In addition, the promoter for the *rpsO* operon in *E. coli* is rather strong and the affinity between Ec-S15 and its RNA is rather poor (~231 nM, Serganov 2002). Therefore, the pseudoknot may have evolved to lower the affinity between Ec-S15 and its regulatory RNA so that Ec-S15 assembles into the ribosome primarily. Then, once the protein levels are depleted and more needed for ribosome assembly, the ribosome is already partially assembled on the mRNA transcript and ready to produce more Ec-S15.

Our results suggest the regulatory RNA from Firmicutes, such as *G. kaustophilus* (Gk-mRNA and Gk-S15), is almost certainly a complete mimic of the rRNA containing an intact G•U/G-C motif and three helix junction with a GGC base-triple, yet with distinct overall RNA sequence. All S15 homologs respond to this RNA *in vivo* and *in vitro* (Figure 3.6, 3.7, 3.8). Additionally, we find that the S15 homolog, Gk-S15, requires the three helix junction in its regulatory RNAs. This is supported by work by Scott & Williamson (Scott 2001) where the minimum RNA binding fragment for Gk-S15 was determined to contain the three helix junction. Additionally, we show that Gk-S15 does

not respond to Ec-mRNA (lacks a 3HJ mimic) or Gk-mRNA-M1 (mutates this region, Figure 3.6) in a regulatory capacity. The fact that the Gk-S15 homolog is the sole homolog to require a 3HJ is somewhat surprising; however, this may partially explain how Gk-mRNA evolved to look this way. Previous work has shown that the 3HJ is the primary binding site for S15 within the rRNA (studies using *E. coli* rRNA and protein, Serganov 2002). Because Gk-mRNA is a complete mimic of the rRNA, and because Gk-S15 is unable to regulate using RNA structures that lack a 3HJ, it is likely that Gk-S15 primarily recognizes the 3HJ in its regulatory RNA and the interaction is stabilized by binding the G•U/G-C motif. The mimicry between the Gk-mRNA and rRNA also reflects the mechanism of regulation. Our results strongly suggest Gk-mRNA and Gk-S15 use a displacement mechanism to regulate *rpsO* expression, though this remains to be experimentally verified. The residues important for autoregulation almost completely coincide with those required for binding rRNA (Figure 1.6). The fact that Gk-mRNA completely mimics the rRNA strongly suggests it utilizes the same residues in Gk-S15 as the rRNA and that the two compete for binding. Therefore, to regulate *rpsO* expression in *G. kaustophilus*, the Gk-mRNA must completely mimic the rRNA to give Gk-S15 a 3HJ to primarily bind and prevent ribosome access to Gk-mRNA.

The most interesting regulatory RNA that has evolved for S15 is that from Phylum Thermales, and characterized in *T. thermophilus* (Tt-mRNA and Tt-S15). Though the pseudoknotted architecture of the Ec-mRNA is overall the most unique, the Tt-mRNA is the only regulatory RNA for S15 that lacks a G•U/G-C motif, and instead contains a mimic of the 3HJ of rRNA with almost exact nucleotide similarity. And in spite of this, the S15 homolog from *T. thermophilus*, Tt-S15, retains the ability to regulate

gene expression in response to all regulatory RNA structures, including those that lack a 3HJ mimic. Previously it has been shown that Tt-S15 uses a displacement mechanism to regulate expression of the *rpsO* operon. So, in a similar fashion to the Gk-mRNA 3HJ-mimic, perhaps Tt-mRNA mimics the 3HJ in order to compete with the ribosome for binding the *rpsO* transcript. Though Tt-mRNA lacks an exact G•U/G-C mimic, a secondary binding site within a stem for Tt-mRNA was identified containing a G•G non-canonical base pair. Because Tt-S15 uses a displacement mechanism, this mismatch may have evolved to decrease Tt-S15 affinity for Tt-mRNA to ensure it primarily binds the rRNA. This G-G mismatch is surrounded by G-C base pairs (Figure 1.6), the strongest base pairs with three hydrogen bonds. This may be a product of the extreme hot springs environment within *T. thermophilus* thrives, suggesting a double helix is required at this location to stabilize Tt-S15 binding its regulatory RNA, yet a G•U/G-C motif is not tolerated.

Work presented in this thesis characterized a fourth regulatory RNA for S15, found in the Phylum Alphaproteobacteria, using an example from the species, *R. radiobacter* (Rr-mRNA and Rr-S15) (Chapter II, Figure 2.4, 2.6). Prior to this, nothing was known about the structure of this regulatory RNA or the S15 homolog's requirements for gene regulation. Our results suggest the Rr-mRNA completely mimics the rRNA, as all S15 homologs respond to this RNA in a regulatory fashion (Figure 3.6). While our structure probing and site-directed mutagenesis results strongly supported the existence of a G•U/G-C motif, much less was clear regarding the junction of the helices (Figure 2.4, 2.7). With work presented in Chapter III, in particular the fact that Gk-S15 responds to the G•U/G-C mutant version of this RNA (Rr-mRNA-M1), strongly suggests

the junction somehow mimics the 3HJ of the rRNA, Tt-mRNA and/or Gk-mRNA. While the Rr-mRNA was able to respond to all S15 homologs in a regulatory fashion, its Rr-S15 counterpart was much more discriminating in the regulatory RNAs with which it interacts. Taking our cross-species *in vivo* regulation results in conjunction with our *in vitro* binding assays, these data suggest that Rr-S15 requires a G•U/G-C motif in its regulatory RNA. Though it remains yet to be characterized, several lines of evidence suggest Rr-S15 may use a displacement mechanism to regulate gene expression. The strong binding affinity between Rr-mRNA and Rr-S15 (Figure 2.4), the double stranded nature of the Shine Dalgarno sequence (Figure 2.7), and the ability for Gk-S15 to regulate using this RNA (presumably at a 3HJ mimic) all suggest this species uses a displacement mechanism to regulate the *rpsO* operon. This would also suggest that Rr-S15 uses similar amino acids to recognize Rr-mRNA as it uses to recognize rRNA; however, the residues required for autoregulation remain to be elucidated. Taken together, the data presented in this thesis suggest the Rr-mRNA structure may have evolved for a similar reason as the Gk-mRNA structure, yet contain the opposite motif requirement in its regulatory RNA.

The variety of sequences and structures that regulate expression of the *rpsO* operon in different bacterial phyla showcase the plasticity and likely independent-derivation of RNA structure. Despite its overall amino acid conservation, there are differences among the S15 homologs from diverse bacterial phyla and these differences do contribute to differences in mRNA recognition. Because of the differences in regulatory RNA recognition (Figure 3.8), the results we present here strongly suggest the S15 homologs may have played a strong role in the evolution of their cognate regulatory RNAs, for the protein must be able to perform its primary role in the ribosome first, and

then regulate expression of itself when needed. This work suggests that differences between the ancestral S15 proteins lead to the development of a diverse array of regulatory RNAs.

Ribosomal Protein Biosynthesis as a Novel Antibacterial Target

As detailed in Chapter I, the role S15 plays in the prokaryotic ribosome has been delineated through the solving of the crystal structure as well as numerous *in vivo* and *in vitro* characterization studies. S15 plays an important scaffolding role as one of the first proteins recruited to the small ribosomal subunit where it binds and stabilizes the 16S rRNA and allows additional recruitment of r-proteins. Additionally, it plays a role bridging the small ribosomal subunit to the large. A goal of this work was not to further characterize the role S15 plays in the ribosome, which is very well characterized presently, but to better understand the regulatory role of S15. With results presented in this thesis, we have expanded our understanding of S15's regulatory role and propose that this regulatory interaction in bacteria may be an excellent, novel target for antibacterial pharmaceuticals.

Antimicrobial resistance is one of the greatest challenges facing our society today. One of the most successful targets for antibacterials has been the prokaryotic ribosome and the process of protein translation. The ribosome function is essential to bacteria and because of the importance and complexity of the ribosome the development of resistance is challenging to bacteria. Unfortunately, bacteria have well-developed resistance mechanisms for not only those compounds targeting the ribosome, but most other

commonly targeted cellular processes of bacteria (Davies 2010). Coupled with the stagnation of FDA approval for novel first-generation pharmaceuticals and the emerging need for novel antibacterial structural scaffolds, there is a dire need for novel bacterial processes to target.

Work presented in this thesis may offer not only an alternative target for novel antibacterials that is related to ribosome biogenesis, but one that has the potential to target a specific subset of bacterial species. Targeting the regulatory RNAs that control the expression of ribosome components, such as S15, may be attractive in the development of new drugs. The widespread distribution of autoregulation in bacteria for most ribosomal proteins underscores the importance of the regulation of ribosome components to overall bacterial viability. Our lab has shown that overexpression of ribosome components, such as the S6:S18 dimer, L20 and S8 (data not shown) drastically affects growth at permissive temperatures. Knockout organisms of certain ribosomal proteins are not even possible (L20, Guillier 2005), and those that are possible, such as the *rpsO* operon, have slow-growth phenotypes and cold-sensitivity (Guthrie 1969, Dammel 1993). Developing novel antibacterials that target the RNAs responsible for regulating the expression of these important ribosome components offers a new strategy to combat bacterial infections.

While the targeting of autoregulatory RNA structures in bacteria may overall be an excellent target for novel antibacterials, targeting the RNA structures with which S15 interacts offers several advantages. For one, S15 interacts with diverse regulatory RNAs in different bacterial phyla. It is increasingly appreciated how individual bacterial species function within a larger microbial community consisting of other bacterial, viral, and

fungal species. A healthy human contains billions of species of these microscopic organisms, the vast majority are benign, many of which even helpful to human health. Broad-spectrum antibiotics can affect both the good and bad actors of these communities and it is vital we find a way to target a specific phyla or even specific species of infectious bacteria, leaving the remaining species unhindered. This thesis demonstrates the diversity of RNA structures with which S15 interacts in a regulatory manner. In conjunction, our data suggest S15 homologs from different species have different regulatory RNA motif requirements and that S15 from one species does not interact with the regulatory RNA from another species. This suggests that novel antibacterial compounds would affect only the bacteria of a given phyla that contain said RNA structure, leaving the remaining friendly bacteria free to proliferate.

The use of bactericidal compounds puts selective pressure on bacteria to survive, which in turn leads to drug-resistant phenotypes. A second advantage of targeting the regulatory RNAs for S15 is that it may be less likely to pressure the bacteria to develop resistance. Our data show that overexpression of any S15 homolog in an *E. coli* host organism is not lethal, presumably because all S15 homologs assemble with ribosomal RNA to form ribosomes. However, the viable *E. coli* K12: $\Delta rpsO$ organism displays a severe growth defect. Targeting the regulatory RNA of a specific phyla would decrease expression of S15, which in turn may decrease both the proliferation and infectivity of a subset of infectious bacteria. Concurrently, the slow growth of the pathogen may be enough to allow other members of the community to out-compete them for resources as well as the human immune system to combat said infection. Less toxic compounds may be less prone to the appearance of antibacterial resistance.

Targeting bacterial *cis*-regulatory RNAs is a relatively new concept, yet there has been limited recent success in the development of compounds targeting riboswitches. Lysine analogs that target the *lysC* (lysine) riboswitch in *Bacillus subtilis* (a model organism for the pathogen *Bacillus anthracis*) (Blount 2006), and a riboflavin analog, 5FDQD, which targets the flavin mononucleotide (FMN) riboswitch in *Clostridium difficile* (Blount 2015) have been described to have bactericidal activity. These recent successes suggest compounds targeting the regulatory RNAs for ribosomal protein biosynthesis may be a viable, potentially lethal antimicrobial target.

As our arsenal of effective antibiotics dwindles, and the threat of multi-drug resistant bacteria to public health increases, it is becoming urgent to identify additional pharmaceuticals and pharmaceutical targets in bacteria. A novel strategy may be to target the regulatory RNAs that modulate the expression of ribosome components. The data presented in this thesis suggest that the diverse regulatory RNA structures that interact with S15 offer a unique way to target and slow the growth of a subset of specific bacterial phyla. Developing compounds that bind the regulatory RNA itself to prevent S15 expression or developing compounds that bind S15 itself with a higher affinity than that of the rRNA to prevent proper ribosome assembly are two different targets our data suggest may be fruitful in the development of novel antibacterial compounds.

A Broader Context: Many Structural Solutions to this Biological Problem

In terms of number of different structures that perform the same regulatory function in response to the same ligand, the regulatory repertoire for S15 is currently

unmatched. With work presented in this thesis we identified three additional naturally-occurring regulatory RNAs (Chapter II) and we evolved four synthetic sequences with relative ease (Chapter IV). Collectively, this brings the total to six distinct naturally-occurring structure and four distinct synthetic RNA sequences that respond to one ligand, S15, to regulate gene expression. In the realm of *cis*-regulatory RNAs found in bacteria, this represents the most diverse and structurally distinct group (see Chapter I). This strongly suggests additional regulatory RNA structures remain to be discovered and characterized in additional bacterial phyla.

Though the S15 protein itself is small and globular, the answer to this diversity of structures is not merely that it is a simple protein to find RNAs with which it binds. We identified several RNA structures that bind S15, RNA 11-2 in particular, yet did not regulate gene expression (Figure 4.4). Something more is required in these RNA structures to regulate gene expression. Our data suggest that the regulatory RNAs partially mimic the rRNA (Figure 3.8), which may be part of the answer. However, perhaps more importantly, we also find that the regulatory RNAs interact with the rRNA-binding face of the protein (Figure Chapter I, Figure 1.6, Figure 4.6) and upon protein-binding the RNA occludes elements important for ribosome binding, such as the Shine Dalgarno sequence (Figure 4.8). Evidence to support this is in the RNA structure probing of synthetic regulatory RNA 11-1, which does not appear to mimic the rRNA in any way. (Figure 4.8, 4.9). Therefore, in a functional regulatory RNA there must be a direct competition between both the mRNA and rRNA for S15, and also the S15 and rRNA for the mRNA transcript. When both of these are satisfied, a given mRNA structure is able to regulate gene expression in response to S15.

One simple way to ensure the mRNA binds the correct face of S15 to prevent its assembling into the ribosome is to mimic a portion of the rRNA. Data presented here and by others strongly suggest all naturally-occurring regulatory RNAs for S15 do this. In conjunction, the protein is under strong selective pressure to conserve its rRNA-binding amino acid sequence to form functional ribosomes (Figure 1.3). This additionally pressures the regulatory RNAs to interact with a conserved set of residues and limits the number of possible regulatory RNA structural solutions. Thus, the resulting regulatory RNA structures all bind the r-protein and prevent its binding within the ribosome, while concurrently preventing ribosome assembly upon itself.

In a larger biological context, the results presented in this thesis offer several insights. First, even though an RNA-protein interaction may be conserved among different species, the way one recognizes the other may be different in different species. We demonstrate this in Chapter III, where the Gk-mRNA was recognized by all S15 homologs, however the homologs utilize distinct recognition motifs (Figure 3.6, 3.7, 3.9). Therefore, though an RNA-protein interaction may be conserved, the specifics within the interaction may not be completely conserved in different species.

For the autoregulation of all ribosomal protein encoding bacterial operons, the regulatory RNA structure cannot solely be a structure that binds the protein. These *cis*-regulatory RNA structures must be multifaceted, both in the ability to bind a specific face of a given r-protein to prevent its ribosome assembly, but also changing their structure so as to prevent its own ribosome assembly and translation. This may be a theme for r-protein interacting autoregulatory RNAs, as more are described beyond model organism, *E. coli*.

Our results suggest binding strength does not correlate with the ability to regulate gene expression; this may speak to the functional capacity of RNA-protein interactions in a larger context. Our synthetic RNA 11-2, which had the second strongest binding affinity for Gk-S15, was unable to regulate gene expression, whereas RNA 11-6 displayed the lowest binding affinity for Gk-S15, yet was able to regulate gene expression (Figure 4.3, 4.4). We see this with the naturally-occurring regulatory RNAs, as well. For example, the Ec-mRNA had a weak binding affinity, yet was able to regulate gene expression (Figure 3.6, 3.7). This may be important in the engineering of novel regulatory RNAs for synthetic genetic circuits or in the general understanding of a given RNA-protein interaction. While a binding interaction must occur between a given RNA and protein, the strength of that binding interaction may not be as important or may change depending on the biological function of a given RNA-protein interaction.

Tracking the evolutionary history and trajectory for RNAs such as the ones presented in this thesis work, and others where distinct structures perform analogous function remains difficult. This is exacerbated when both the regulatory and the ligand, a protein in this case, can co-evolve over time. Our evolution experiment in Chapter IV began with a regulatory RNA pool based upon Ec-mRNA, and resulted in a variety of sequences that can regulate gene expression in response to Gk-S15 and Tt-S15, yet not with Ec-S15. Performing the same experiment with a different S15 homolog, such as Ec-S15, would have resulted in completely different regulatory RNA structures-very likely ones that contain a G•U/G-C mimic of which the Ec-S15 requires. In light of the common occurrence of horizontal gene transfer in bacteria, in conjunction with their smaller genomes (in comparison to eukaryotes) and their fast generation time,

deconvolution of what the regulatory RNA structure found in the last universal common ancestor looked like and how it evolved into the modern versions remains challenging.

Results presented in this thesis have elucidated the evolution of the regulatory RNA structures for S15. We present and discuss how those RNAs interacting with S15 may have come to be this way, and it is clear that both the RNA and the ligand with which it interacts plays a role in the regulatory RNA structures that have evolved. When the forces of evolution can act upon both the RNA and S15 independently, when the evolution of one can influence the evolution of the other, and when many structural solutions are possible, the result is the diverse RNA structural solutions we see in nature presently.

APPENDIX

CHAPTER AI

Complete RNA Inverse Folding: Computational Design of Functional Hammerhead Ribozymes

The work in this Chapter is based up on the following published journal article. Text, figures, and tables from this publication are used throughout this Chapter without additional notice.

*Dotu I, *Garcia-Martin JA & *Slinger **BL**, Mechery V, Meyer MM, Clote P: Complete RNA inverse folding: computational design of functional hammerhead ribozymes. *Nucleic Acids Research* 42:11752-11762 (2014). [*denotes co-first author]

Authors' contributions:

Conceived and designed the experiments: ID JAG-M BLS VM MMM PC. Performed computational development and analysis: ID JAG-M MMM PC Performed the experiments: BLS VM MMM Analyzed the data: ID JAG-M BLS VM MMM PC. Wrote the paper: ID MMM PC.

INTRODUCTION

In response to the increased appreciation and understand of the integral role RNA plays in biology, many recent efforts, including our own (Chapter IV), have focused on the creation of synthetic RNA-based gene regulatory elements. In the last decade, many groups have successfully developed synthetic RNA sequences capable of self-cleaving, sensing small molecules *in vivo* and *in vitro*, as well as regulating gene expression (Isaacs 2006, Collins 2012). Selection-based approaches, including the use of Systematic Evolution of Ligands by Exponential Enrichment (SELEX, Tuerk 1990, Ellington 1990), are a powerful way to generate synthetic sequences that perform a desired function, such as *cis*-regulation or self-cleaving (Piganeau 2009, Sinha 2010, Goldfless 2012, Belmont 2010). Design-based approaches have also been used successfully to create RNAs elements with engineered functions (Schultes 2000, Isaacs 2004, Bayer 2005). This current work concerns the development of a novel computational algorithm, RNAiFold, to design synthetic ribonucleic acid enzymes (ribozymes), and provides subsequent experimental validation of the synthetic RNA's predicted function.

The use of computational methods in the design of synthetic RNA elements is becoming more important as the complexity of synthetic genetic circuitry increases. The program RNAiFold is designed to take a given target RNA structure and determine all possible RNA sequences that fold into that target structure. RNAiFold uses constraint programming and can take into account several RNA sequence design constraints that may be necessary for a more biologically relevant result, such as GC content, fixing certain base pairs, specifying nucleotide frequencies, or requiring a specific nucleotide to be present at a specific position. In the current work, we selected the type III

hammerhead ribozyme Peach Latent Mosaic Viroid (PLMVd) as a target structure as numerous biochemical and structural studies have pinpointed key nucleotides required for catalysis in the ribozyme (Blount 2005, Martick 2006, Nelson 2008).

Using RNAiFold, we design ten *cis*-cleaving hammerhead ribozymes, all are shown to be functional by a cleavage assay. We additionally use RNAiFold to design a functional *cis*-cleaving hammerhead as a modular unit of a synthetic larger RNA. Analysis of kinetics on this small set of hammerheads suggests that cleavage rate of computationally designed ribozymes may be correlated with positional entropy, ensemble defect, structural flexibility/rigidity and related measures. Artificial ribozymes have been designed in the past either manually or by SELEX (Systematic Evolution of Ligands by Exponential Enrichment); however, this appears to be the first purely computational design and experimental validation of novel functional ribozymes.

RESULTS

RNAiFold Determines All RNA Sequences for the Target Structure, PLMVd

Hammerhead Ribozyme

As a target structure for our computationally designed type III hammerheads, we chose the secondary structure of the plus polarity strand of Peach Latent Mosaic Viroid (PLMVd) (isolate LS35, variant ls16b) from Rfam family RF00008, having accession code AJ005312.1/282-335 (Figure A1.1). Given the target Rfam consensus structure S of PLMVd AJ005312.1/282-335, which is identical with the MFE secondary structure using RNAfold 1.8.5, 16 highly conserved nucleotides were taken as constraints in the

generation of over 1 million sequences solving the inverse folding problem, as determined by RNAiFold 1.8.5. Sequence identity exceeds 96% for the 15 positions 6–8, 22–25, 27–29, 44–49; therefore, in running the software RNAiFold, sequence constraints were imposed for those positions. An additional constraint at position 8 was implemented based on experimental data of the hammerhead cleavage site. It is well-known that hammerhead cleavage sites are of the form NUH (e.g. GUH and CUH) (Pan 2003, Gonzalez-Carmona 2006), and for PLMVd, cleavage occurs immediately after the cytidine at position 8. Therefore, IUPAC code H (i.e. not G) was given as an additional constraint for RNAiFold. The remaining 38 positions were constrained to be distinct from those of PLMVd. To summarize, RNAiFold was used to solve the inverse folding problem using the constraints outlined above and the consensus structure of PLMVd used as target.

Using distance measures of *dissimilarity* of low energy structures to the MFE structure (positional entropy, ensemble defect, structural diversity, etc.) together with measures of molecular structural flexibility/rigidity, 10 putative hammerhead sequences were selected for *in vitro* validation using a cleavage assay, HH1-HH10. The selected sequences and selection criteria are given in Figure A1.2. To summarize, the measures used for sequence selection concern either structural diversity or regional structural flexibility/rigidity.

Hammerhead Candidate Sequences Self-Cleave

A cleavage reaction under mild conditions was used to assess the functionality of the 10 hammerhead candidates, HH1-HH10. All 10 hammerhead candidates were shown

to be functional at self-cleavage (Figure A1.3). A no magnesium (-Mg²⁺) reaction was used as a control for RNA folding with each candidate hammerhead, and no cleavage products were visible for any of the candidates under these conditions.

It is known from literature (Pan 2003, Gonzales-Carmona 2006) that hammerhead cleavage sites are of the form NUH (e.g. GUH and CUH, but not GUG). Indeed, Carbonell et al. (Carbonell 2006) suggest that G8 would pair with C22 (in our numbering) and impede its role in the catalytic pocket. Figure A1.3 shows that the H8G mutant of each designed sequence HH1–HH10 does not cleave under mild denaturing conditions that suffice for cleavage of HH1–HH10. Together these data strongly suggest that the designed sequences HH1–HH10 behave in a manner consistent with the expected mechanism for hammerhead ribozymes.

Rate of Cleavage Differs for the 10 Computationally Designed Hammerheads

We next assessed the efficacy in each of the different selection criteria used for choosing our hammerhead candidate sequences and their effect on functionality (Figure A1.2). To do this we measured the rate of cleavage for each of the 10 computationally designed hammerheads, HH1 through HH10 (Figure A1.4). A time series for cleavage fraction and kinetics curves for a typical designed hammerhead ribozyme (HH1) and the fastest designed ribozyme (HH7) are shown in Figure A1.5, while similar figures for the remaining designed hammerheads appear in Figure A1.6. Kinetics for the designed hammerheads should be compared with wild-type hammerhead kinetics, where under standard conditions of 10 mM MgCl₂, pH 7.5 and 25°C, cleavage rates between 0.5 and 2 per minute have been observed for at least 20 different hammerheads (Clouet-d'Orval

1997). It follows that kinetics of the computationally designed hammerheads described in this paper are slower than wild-type hammerheads approximately by a factor of 10.

Pearson correlation coefficient was determined between cleavage rate K_{obs} , obtained by fitting equation (1) with data from three to five technical replicates, and 21 measures, including average positional entropy, GC-content, MFE, etc. See Supplementary Information for all correlation values. The most pronounced correlations were observed between K_{obs} and (full) average structural positional entropy, ensemble defect, and expected base pair distance discrepancy for ‘conserved site’ with values respectively of -0.461 , -0.370 , -0.438 ; i.e. cleavage is faster when these measures are smaller.

Designed Hammerhead Functions Within a Larger Rationally Designed RNA

It has been observed that aptamers, hammerheads and other functional non-coding RNAs constitute modules, capable of function even when engineered to form part of a larger RNA molecule (Wieland 2008, Saragliadis 2013). We rationally designed a 166 nucleotide guanine-activated riboswitch with a putative type III hammerhead module. The target secondary structure contained the xanthine phosphoribosyltransferase (XPT) riboswitch, whereby the terminator loop was replaced by the type III hammerhead structure (Figure A1.7). Sequence constraints were chosen to be the highly conserved nucleotides of the consensus structures for the purine riboswitch and the hammerhead. An additional constraint was the hammerhead cleavage site (NUH) was required to be fully sequestered within a base-paired region (Positions 60-118).

Several experiments were performed to assess the functionality of our synthetic riboswitch-ribozyme RNA. First, a cleavage assay was performed in the presence and absence of Mg^{2+} (Figure A1.8A). Roughly 40% of our synthetic sequence rapidly cleaves, and only in the presence of Mg^{2+} , with a rate of 1.3/min with an F_{max} of 0.47 and MSE of 0.0026 (Figure A1.8D). Two mutants were designed that should inactivate hammerhead activity, C116G (mutates the GUC site of cleavage) and G142U (mutates a distal section of the ribozyme, known to be required for cleavage, the CUGAUGA sequence). Both mutations to our synthetic modular RNA prevent self-cleavage (Figure 1.8B, C). From these data, it is evident that cleavage only occurs for the wild-type sequence, and when Mg^{2+} is present. To confirm cleavage occurs at the expected site, C116, we used T1-RNase structure probing to map the sequence of the cleavage products (Figure A1.9A, B). These results confirm cleavage does occur at C116. Finally, we measured the rate of cleavage in the presence and absence of guanine (Figure A1.9C, D). These results indicate that addition of 1 mM guanine has no significant effect on either the K_{obs} or the F_{max} (i.e. the designed riboswitch was constitutively on).

Taken together, all of our results show that the cleavage is Mg^{2+} -dependent (Figure A1.8A), and the hammerhead appears to cleave rapidly within seconds (Figure A1.8D) at the expected nucleotide (Figure A1.9). Neither of the mutant sequences displays any cleavage under the same conditions, even with significantly longer incubation times (Figure A1.8B, C). Kinetics for the 166 nt synthetic ribozyme are comparable with those of wild-type hammerheads, with an observed cleavage rate K_{obs} of 1.3/min and F_{max} of 0.47 (Figure A1.8D).

DISCUSSION

In this paper, we have demonstrated the success of a purely computational approach for the rational design of artificial type III hammerhead ribozymes. Figure A1.3 clearly shows the Mg^{2+} -dependent cleavage of each designed sequence HH1-HH10, as well as the non-cleavage of the 8G mutant of each sequence, strongly suggesting that cleavage is due to the usual hammerhead mechanism. Cleavage time series data for three to five technical replicates for each of the 10 computationally designed hammerheads, displayed in Figure A1.4 and A1.5 lead to observed cleavage rates varying 100-fold from 0.0027 min^{-1} for HH3, to 0.25 min^{-1} for HH7. The relatively fast cleavage rate of HH7, selected from over 1 million sequences returned by RNAiFold solely on the criteria of minimizing ensemble defect, *with* the additional requirement of having GUC at the cleavage site, is slower only by a factor of 10 from wild-type hammerhead cleavage rates (recall that wild-type cleavage rates vary between 0.5 and 2 per minute (Clouet-d'Orval 1997)). In contrast, HH8 had an observed cleavage rate of 0.02 min^{-1} , although it was selected solely on the criteria of minimizing ensemble defect—*without* the additional requirement of having GUC at the cleavage site. This experimental result suggests that cleavage kinetics may be the underlying reason that cytidine is present at cleavage position 8 in 95% of the 84 sequences in the Rfam seed alignment of family RF00008.

Among more than 20 computational features, the features found to be most highly correlated with cleavage rate K_{obs} for HH1-HH10 were (full) average structural positional entropy, ensemble defect and expected base pair distance discrepancy for 'conserved site' with values respectively of -0.461 , -0.370 , -0.438 . However, this result is based on a tiny set of data and can only be taken as a suggestive first step toward a more systematic

determination of which measures of structural diversity/flexibility/rigidity might best predict ribozyme activity.

In addition to computationally designing the functional hammerheads HH1–HH10, we have designed the 166 nt sequence, in which a synthetic hammerhead is embedded within the terminal stem-loop of the structure depicted in Figure A1.7. The sequence does self-cleave at the expected GUC cleavage site 114–116 (Figure A1.9). Moreover, as shown in Figure A1.8D, cleavage kinetics for this 166 nt artificial ribozyme ($K_{\text{obs}} = 1.3/\text{min}$) are as fast as those of wild-type hammerheads, although the cleavage amount ($F_{\text{max}} = 0.47$) is quite poor compared with our other designed ribozymes HH1–HH10. By utilizing two mutants, one at the cleavage site position 116, and one further downstream at position 142 in the CUGUAGA segment necessary for catalysis of cleavage, we show effectively that cleavage in the synthetic wild-type, designed construct is due to the usual hammerhead mechanism (Figure A1.6B, C). Additionally, we have demonstrated Mg^{2+} -dependence, necessary for the cleavage mechanism, through the complete absence of 5'- and 3'-cleavage products when incubated for an extended period of time of 24 h in buffer lacking Mg^{2+} .

The software RNAiFold solves the inverse folding problem, not only for a target secondary structure, but as well when the target is the hybridization of two RNA secondary structures. Since RNAiFold uses constraint programming, it can perform a complete search of the space of compatible sequences, and thus return *all* sequences, whose MFE structure [resp. MFE hybridization] is a given target structure [resp. hybridization], or can certify that no such solution exists. Our results show that RNAiFold can be successfully used to rationally design functional non-coding RNA.

In this paper, by employing our constraint programming solution RNAiFold (Garcia-Martin 2013a, Garcia-Martin 2013b) to generate >1 million sequences, that agree with PLMVd AJ005312.1/282-335 at the 15 nucleotides having >96% conservation in Rfam RF00008 seed alignment, and have MFE structure identical to that of the Rfam consensus secondary structure of PLMVd. Ten candidate hammerheads, which were selected using criteria that measure either *structural diversity* or regional *structural flexibility/rigidity*, were shown to be functional, with varying kinetics, by an *in vitro* cleavage assay. This appears to be the first purely computational design and experimental validation of novel functional ribozymes. Moreover, by computationally designing a 166 nt synthetic RNA, whose terminal stem-loop harbors a functional computationally designed hammerhead, we show that *in silico* design and placement of artificial hammerheads is possible.

Since RNAiFold supports user-defined sequence constraints, as well as structural compatibility and incompatibility constraints, our method should be able to rationally design hammerheads that reside within larger RNAs, which meet user-defined sequence and structure constraints.

MATERIALS & METHODS

Computational Methods

RNAiFold returns sequences whose MFE structure is a given target structure, whereby the user may choose to use the free energy parameters from either Vienna RNA Package 1.8.5 (Turner 1999 parameters) or Vienna RNA Package 2.0.7 (Turner 2004 parameters) (Turner 2010). By abuse of notation, let RNAiFold 1.8.5 [resp. 2.0.7] denote the program RNAiFold with energy parameters from the corresponding version of Vienna RNA Package.

As target structure for our computationally designed type III hammerheads, we selected the secondary structure of a portion of the plus polarity strand of Peach Latent Mosaic Viroid (PLMVd) (isolate LS35, variant ls16b) from Rfam family RF00008 (Gardner 2011) having accession code [AJ005312.1/282-335](#).

RNAiFold was run four times, each time additionally constraining GC content to be within a specified range. Altogether, over one million solutions of RNA inverse folding were returned before memory exhaustion (using the 32 bit version of run-time system COMET): 200 072 with GC-content 30-39%, 352 924 with GC-content 40-49%, 349 325 with GC-content 50-59%, 366 323 with GC-content 60-69%, constituting a total of 1 268 644 sequences.

For additional information regarding computational methods and criteria used to select sequences, please see our publication:

*Dotu I, *Garcia-Martin JA & *Slinger BL, Mechery V, Meyer MM, Clote P: Complete RNA inverse folding: computational design of functional hammerhead ribozymes. *Nucleic Acids Research* 42:11752-11762 (2014). [*denotes co-first author].

Design of Modular Hammerhead Within Another Structure

With the intent of designing a guanine-activated riboswitch with a modular hammerhead, we followed the following steps in rationally designing a synthetic 166 nt RNA, with putative type III hammerhead module. Target secondary structure S was taken to be the structure of the gene OFF xanthine phosphoribosyltransferase (XPT) riboswitch, depicted in Figure 1A of (Serganov 2004), whereby the terminator loop (expression platform) was replaced by the Rfam consensus structure for a type III hammerhead. Sequence constraints were chosen to be the highly conserved nucleotides of the Rfam consensus structures for the purine riboswitch (RF00167 seqcons view of consensus structure) and for type III hammerhead (RF00008 seqcons view of consensus structure). Figure A1.7 displays the target structure S for computational design of a modular hammerhead within the terminal stem-loop of a structure similar to the XPT riboswitch. We gave RNAiFold an additional compatibility constraint, whereby returned sequences were required to be compatible to a second structure S' , in which the hammerhead cleavage site (NUH) is fully sequestered within a base-paired region. Positions 60–118 of S' are given as follows:

```
5'-ACUAYNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGUHHN-3'  
5'-.....(((((((((((.....)))))).....)))))))-3'
```

while all positions in S' outside of 60–118 (i.e. from 1–59 and 119–166) are unpaired. We filtered sequences output by RNAiFold, by applying RNAbor (Freyhult 2007), and its faster sequel, FFTbor (Senter 2012). Given reference structure S , RNAbor and FFTbor return the *density of states* with respect to S , which depicts the Boltzmann probability $p(k)=Z_k/Z$ for secondary structures to have base pair distance k from S . Additionally,

RNAbor computes, for each k , the MFE_k -structure; i.e. that structure having MFE over all structures whose base pair distance from the reference structure S is exactly k .

From a partial output of 3000 sequences from RNAiFold, only one sequence s satisfied the following two properties, when applying RNAbor with input s and reference MFE structure S : (i) The density of states figure has a pronounced peak at $k = 0$, corresponding to the location of the MFE structure S ; (ii) There was another pronounced peak for value $k \gg 0$, corresponding to a structure T containing the base pairs in S' , which thus should sequester the ribozyme cleavage site NUH, located at position 114–116.

The final, selected sequence 166 nt s is given as follows: GCCGC GUAUA AGGGC UGCGA UAAGG GCAGU CCGUU UCUAC GGGCG GCCGU AAACC GCCCA CUACG CGGCG UGGUU AAGCC GGAAA GGAGA CCGGC AGGAG GGUAA UGGGC CGCGU CGCGG GAGCG CGCCG CCUGA UGAGU CCGUG AGGAC GAAAC GCGGCC.

Experimental Validation

Complementary DNA oligonucleotides, corresponding to the DNA sequence of the designed RNAs preceded by a T7 RNA polymerase promoter, were purchased from MWG Operon. The 10 hammerhead candidate sequences HH1–HH10, extended 2 nt on the left by GG and 2 nt on the right by CC for transcriptional efficiency, and the 166 nt sequence, harboring a candidate hammerhead in the rightmost stem-loop were constructed using primer extension and PCR amplified (5 U Taq polymerase (New England Biolabs), 2.5 mM each NTP, 1x NEB Thermopol buffer). For each of the 10 designed hammerhead sequences, the H8G mutant was constructed in a similar manner, using alternative oligonucleotides containing the mutation. Similarly, C116G (analogous to H8G) and G142U mutations were constructed for the 166 nt designed ribozyme. The resulting PCR products were TOPO-cloned (Invitrogen), and the designed and mutant

sequences were verified by sequencing plasmids containing full-length PCR products. These plasmids were subsequently used as templates for PCR reactions to generate template for *in vitro* transcription.

To generate the RNA, *in vitro* transcription was performed using T7 RNA polymerase (400 U T7 polymerase, 80 mM HEPES-KOH pH 7.5, 24 mM MgCl₂, 2 mM spermidine, 40 mM DTT, 2 mM each NTP) with the addition of 10 μCi of α-³²P-GTP for transcriptions to generate body-labeled RNA when necessary. To prevent premature cleavage during transcription, 100 μM of oligonucleotides complementary to nucleotides 17–35 (numbering starts after the leading GG) were added to each reaction. Full-length RNAs were purified using denaturing polyacrylamide gel electrophoresis (PAGE) (20% acrylamide).

To assess self-cleavage of designed hammerhead sequences, RNA was incubated for 1 h in cleavage buffer (5 mM MgCl₂, 50 mM tris pH 7.5) at 25°C. Subsequently, 1 volume of 2x gel-loading buffer (16 M urea (supersaturated), 10 mM ethylenediaminetetraacetic acid (EDTA), 20% sucrose, 0.1% sodium dodecyl sulphate (SDS), 100 mM tris pH 8.0, 100 mM borate, 0.05% bromophenol blue) was added to quench the reaction with final urea and EDTA concentrations of 8 M and 5 mM respectively. The reaction was placed on ice until gel loading.

Samples lacking Mg⁺⁺ were incubated in 50 mM tris pH 7.5 for 1 h at 25°C. For the 166 nt RNA, cleavage experiments were conducted under similar conditions but reactions were incubated for a few seconds (0 h), 30 min, 5 h and 24 h, and samples lacking Mg⁺⁺ were incubated in 50 mM tris pH 7.5 for 24 h at 25°C. Cleavage products were separated by denaturing PAGE (10% acrylamide), and the gels dried prior to

exposure to phosphorimager plates (GE Healthcare) for 18 h. The gels were imaged using a STORM 820 phosphorimager (GE Healthcare).

Kinetics

To determine the cleavage rates for designed hammerhead sequences, body-labeled RNA was incubated in cleavage assays as described above for varying amounts of time. Cleavage products were separated and gels imaged as described above. The cleavage products were quantified using ImageQuant software (GE Healthcare). To calculate the fraction cleaved at time t , $F(t)$, the sum of the quantified counts for 5' and 3' cleavage product bands was divided by the total quantified counts for the entire reaction (uncleaved, 5' and 3' cleavage products).

The observed cleavage rate K_{obs} was computed by using the Matlab function `nlinfit` with constant error model to fit cleavage time series data using the equation $F_{\text{max}} - F(t) = (F_{\text{max}} - F(0)) \cdot \exp(-K_{\text{obs}} \cdot t)$ where $F(t)$ denotes the amount of cleavage product measured at time t , and F_{max} the maximal fraction cleaved. The 95% confidence interval of this fit was calculated from the resulting residuals and variance-covariance matrix using the Matlab function `nlpredci`.

T1-RNase Cleavage

To confirm the cleavage site of the 166 nt combined guanine riboswitch hammerhead, we mapped the sequences of the 5' and 3' cleavage products using a T1-RNase digest. RNA was transcribed *in vitro* as described in the experimental methods in the absence of blocking peptide. The cleavage products were purified by denaturing 10%

PAGE. To generate 5' ³²P-labeled RNA, 10 pmol RNA was de-phosphorylated (alkaline phosphatase, Roche) according to the manufacturer's instructions, the phosphate was heat inactivated, and the RNA labeled using polynucleotide kinase (NEB) and 40 μ Ci of γ -³²P-ATP. 5'-labeled RNA was incubated with T1-RNase (Roche) in 25 mM sodium citrate pH 5.0, 8 M urea, 1 mM EDTA, 10% sucrose for 20 minutes at 55°C and place on ice until gel-loading. A partial alkaline digest of the 5' labeled RNA was performed by incubating the reaction in 50 mM NaHCO₃, pH 9.2 at 95°C for 5 minutes. This reaction was quenched using 2x loading buffer (16 M urea, 10 mM EDTA, 20% sucrose, 0.1% SDS, 100 mM tris pH 8.0, 100 mM borate, 0.05% bromophenol blue), and stored on ice until gel loading. The digestion products were separated by denaturing PAGE (10% acrylamide). The gels were dried prior to exposure to a phosphorimager screen overnight (GE Healthcare). The phosphorimager screen was scanned using a Storm 820 phosphorimager scanner to produce the images.

TABLES, FIGURES & LEGENDS

Figure A1.1 RNAiFold Sequence Target PLMVd Hammerhead Ribozyme

(Left) Sequence conservation for the 56 nt consensus sequence for type III hammerhead ribozymes from version 11.0 of the Rfam database (Gardner 2011), image from <http://rfam.sanger.ac.uk/family/RF00008#tabview=tab3>. (Left) Sequence logo of conservation at positions aligned with the 54 nt Peach Latent Mosaic Viroid (PLMVd) AJ005312.1/282-335 from the hammerhead ribozyme type III seed alignment sequences from Rfam family RF00008. In-house program used to determine frequencies of positions aligned to those of PLMVd, sequence logo generated with WebLogo (Crooks 2004) (web server at <http://weblogo.berkeley.edu/>). The 15 positions 6–7, 22–25, 27–29, 44–49 of PLMVd had sequence conservation in excess of 96%, while cleavage site C at position 8, adjacent to region 6–8, was conserved in 94.9367% of RF00008 seed alignment sequences. RNAiFold was subsequently used to solve the inverse folding problem with consensus structure of PLMVd used as target, with sequence constraints at positions 6–8, 22–25, 27–29, 44–49, as explained in text. Resulting from this analysis, the sequence constraints for RNAiFold were defined to be HBVHBGUHVVH VHDVBBHDBD BCUGAVGAGV DVBVHBBBVH BHBCGAAACV DBVB. (Right) Sequence constraints for RNAiFold with indicated target secondary structure. The 15 positions 6-7, 22–25, 27–29, 44-49 having over 96% sequence conservation in the seed alignment of RF00008 were constrained to be those in PLMVd AJ005312.1/282-335, and the cleavage site 8 was constrained to be H (not G). All 38 remaining positions were constrained to be distinct from the corresponding nucleotides in PLMVd.

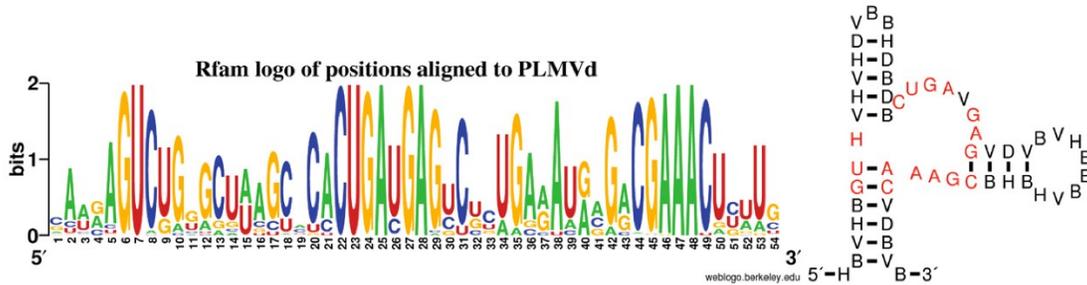


Figure A1.2 Table of HH Sequences

The sequences and selection criteria for the 10 hammerhead candidates selected, HH1 through HH10.

Table 1. Hammerhead candidates selected and selection criteria used

ID	Sequence	Selection criteria
HH1	UUAUUGUAGAGCGAUUCGULCCUGAAGAGCUAUAALUUCUUA GC GAAACAUUAI	GC-content 30–39%, $P(S_0, s) \geq 40\%$, smallest (binary) entropy distance for conserved site
HH2	UUAUUGUAGCGCGAUUCGCGCCUGAAGAGAU'GCGUUU'AAAC U'CGAAACAGUAI	GC-content 40–49%, $P(S_0, s) \geq 40\%$, smallest (binary) entropy distance for conserved site
HH3	CUA'UGUAGCGCGAUUCGCGCCUGAAGAGAU'UGUU'UAUGA U'CGAAACAGUAI	GC-content 40–49%, $P(S_0, s) \geq 40\%$, second smallest (binary) entropy distance for conserved site
HH4	UGGAU'GUAGCGCGAUUCGCGCCUGAAGAGCGGLCA'CCAUC GC GAAACAUUCU	GC-content 50–59%, $P(S_0, s) \geq 40\%$, smallest (binary) entropy distance for conserved site
HH5	CUCAGGUAGCGCGAUUCGCGCCUGAGGAGGGGUCUGGU'AUCC CCGAAACCU'GAU	GC-content 60–69%, $P(S_0, s) \geq 40\%$, smallest (binary) entropy distance for conserved site
HH6	UGGCG'GAGCGCGAUUCGCGCCUGAAGAGGGU'AAACGGGUCC CCGAAACCGUCU	GC-content 30–39%, $P(S_0, s) \geq 40\%$, <i>largest</i> (binary) entropy distance for conserved site
HH7	UCA'UGUAGCGCGGAUUCGCGCCUGAAGAGAU'GGAAU'U'AAAC U'CGAAACAUUGU	GUC in positions 6–8, smallest ensemble defect
HH8	UCA'UGUAGCGCGGAUUCGCGCCUGAAGAGAU'GGAAU'U'AAAC U'CGAAACAUUGU	smallest ensemble defect
HH9	UUA'AU'GCGCGCGAUUCGCGCCUGAAGAGAU'CGACUL'CU'GA U'CGAAACAUUAI	$P(S_0, s) \leq 20\%$, smallest (binary) entropy distance for conserved site
HH10	UUAAGGUCGCGCGAUUCGCGCCUGACGAGCUAUAU'UU'AUUA GC GAAACCU'AU	smallest (binary) entropy distance for conserved site

Note that, subject to presence or absence of additional constraint CS, HH7 and HH8 had also the largest probability of structure, the smallest full structural positional entropy, the smallest (Morgan-Tiggs and Vienna) structural diversity and smallest expected base pair distance.

Figure A1.3 Cleavage Assays of 10 Hammerhead Ribozymes

Cleavage assay results for hammerhead designs, HH1-HH10. Each designed hammerhead RNA was incubated under mild conditions for 1 h as described in the 'Materials and Methods' section to assess cleavage. As negative controls, a no magnesium ($-Mg^{2+}$) and a 0-h reaction were also conducted for each RNA. Additionally, the 8G mutation (mut), predicted to be incompatible with the hammerhead structure (see 'Materials and Methods' section), was constructed for each designed sequence and examined under equivalent conditions to confirm that self-cleavage occurs using the expected hammerhead mechanism.

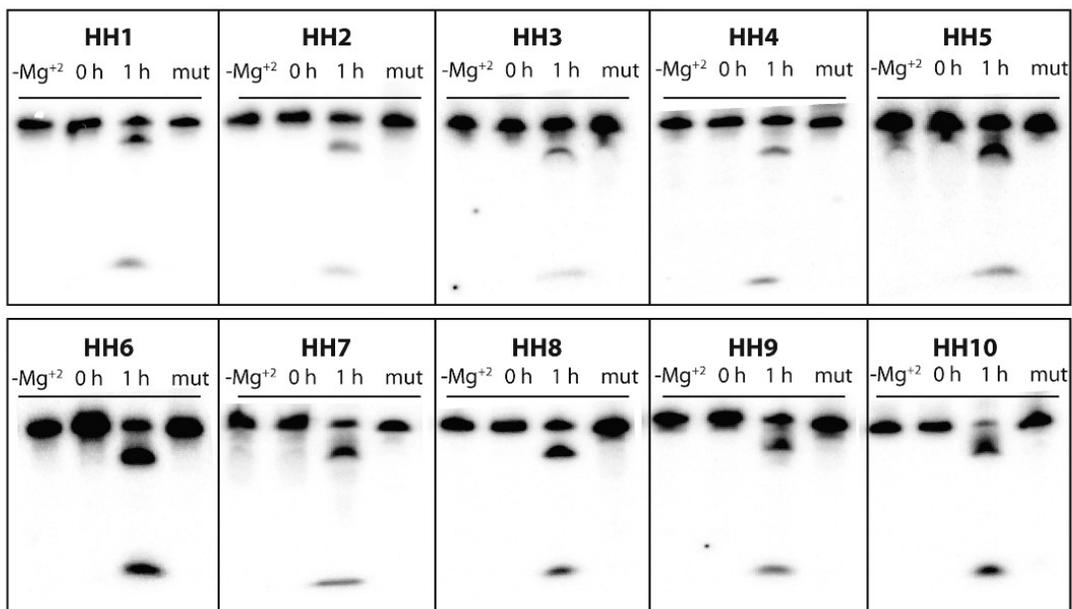


Figure A1.4 Table of Kinetics

Kinetics of cleavage for 10 computationally designed hammerheads, HH1-HH10 and correlation with several measures. Cleavage rate K_{obs} (min^{-1}), maximum percent cleavage F_{max} , mean squared error MSE, (full) structural positional entropy Pos ent, ensemble defect Ens def and expected base pair distance discrepancy for the ‘conserved (or active) site’ EBPD dis act. The Pearson correlation between cleavage rate and Pos ent, Ens def, EBPD dis act is respectively -0.461 , -0.370 , -0.438 , i.e. cleavage rate is faster when these secondary structure deviation values are smaller. Other measures, such as structural diversity, had smaller correlation, while measures such as GC-content and MFE had almost no correlation with cleavage rate.

Table 2. Kinetics of cleavage for 10 computationally designed hammerheads and correlation with several measures

ID	K_{obs}	F_{max}	MSE	Pos ent	Ens def	EBPD dis act
HH1	0.037	0.79	0.0029	0.270882	4.167687	0.0501207
HH2	0.0057	0.74	0.003	0.287235	4.552053	0.0386253
HH3	0.0027	0.65	0.0039	0.259577	4.121914	0.0410984
HH4	0.0127	0.55	0.0048	0.403846	6.755976	0.0354213
I015	0.0085	0.52	0.0066	0.382235	6.240083	0.033132
HH6	0.102	0.73	0.0047	0.414872	8.138131	0.059864
HH7	0.25	0.74	0.0107	0.119159	2.383671	0.0406728
HH8	0.02	0.68	0.0124	0.078518	1.45179	0.0662421
HH9	0.025	0.76	0.0015	0.247886	4.525597	0.0328018
I0110	0.14	0.77	0.01	0.286425	4.975979	0.0269354

Figure A1.5 HH1 and HH7 Cleavage Time Series

(Left) HH1: typical cleavage time series curve with good error parameters (standard deviation <10% of mean, with mean squared error (MSE) = 0.0029). Solid line represents fitted line, and dotted lines indicate 95% confidence interval. Different datasets represented by filled and unfilled squares, triangles, etc. (Right) HH7: fastest hammerhead cleavage rate, though determined with considerable error (MSE = 0.01). In data from the first experiments for HH7, indicated by filled squares, cleavage had been measured at times when maximum cleavage had nearly occurred (these points appear in the flat part of the fitted curve). Subsequent datasets have focused on shorter time periods. This curve was fitted using five datasets. Time series curves for cleavage data for the remaining eight designed hammerheads HH2-HH6 and HH8-HH10 are shown in Figure A1.6.

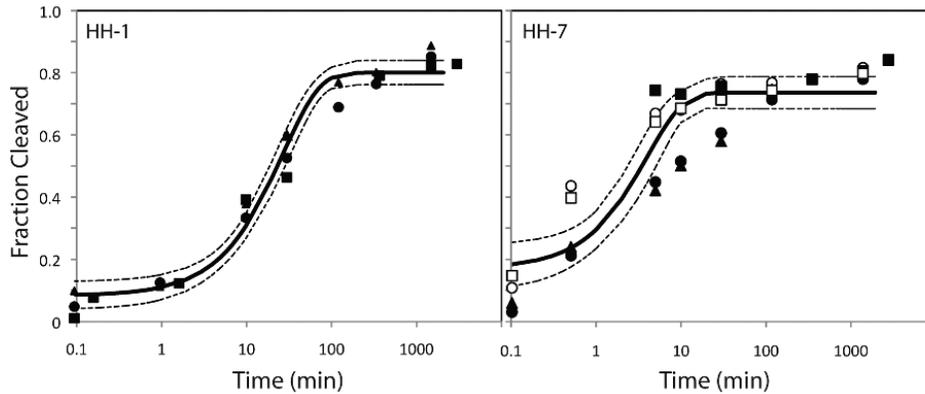


Figure A1.6 Cleavage Time Series for HH2-6, and HH8-10.

Best-fit kinetics curves for designed hammerhead sequences HH2-6 and HH8-10, see Figure A1.5 for HH1 and HH7. From 3-5 independent replicates of the time series were conducted for each designed hammerhead sequence. Each series is represented by a marker of different shape (e.g. closed square, closed circle). The solid line represents the best-fit curve, and the dotted lines represent the 95% confidence interval (see Methods for details on the calculation).

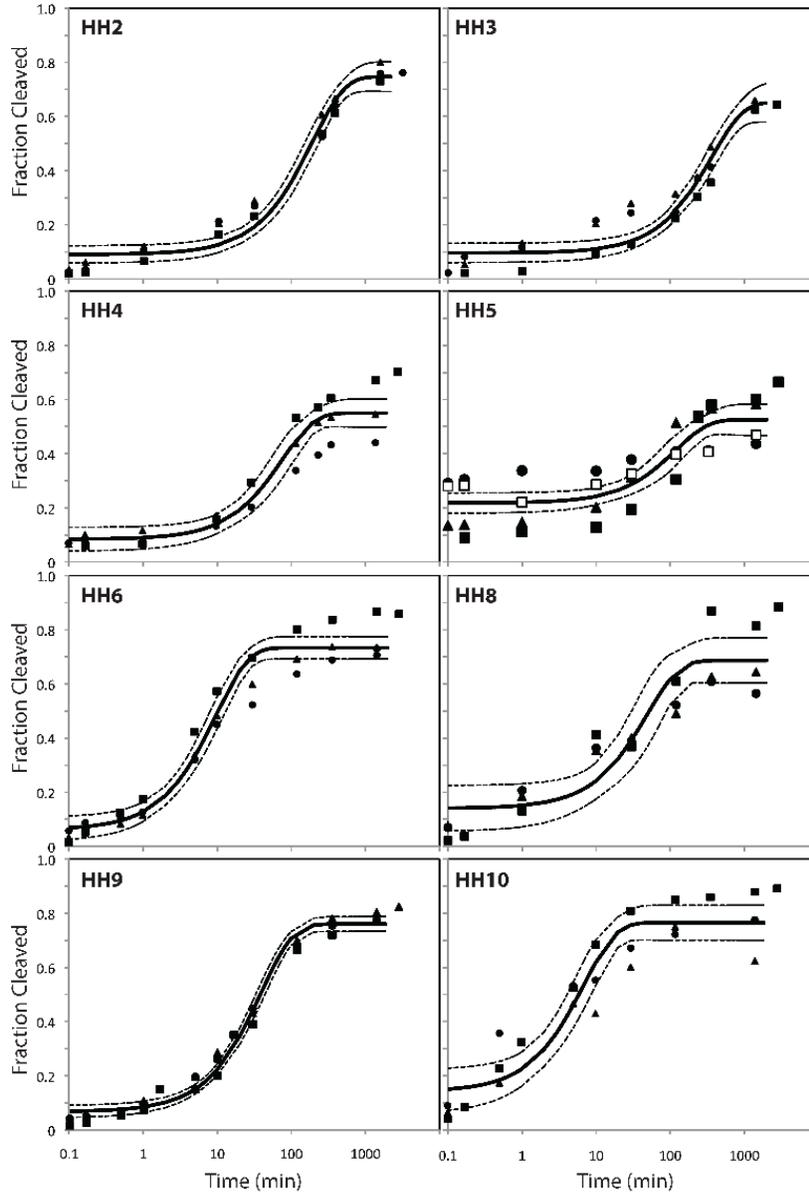


Figure A1.7 Modular Placement of XPT-Riboswitch with Hammerhead Ribozyme
 Target secondary structure for modular placement of artificial hammerhead within larger RNA molecule. (A) The structure and highly conserved nucleotides (sequence constraints) of the XPT-riboswitch appear on the left, while the structure and highly conserved nucleotides of the type III hammerhead ribozyme appear on the right. (B) Output sequence returned by RNAiFold that respects the sequence constraints and whose minimum free energy structure is the target structure.

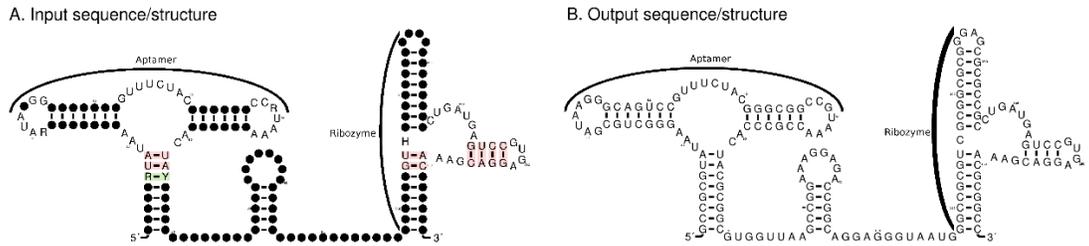


Figure A1.8 Cleavage Assay and Kinetics Data for Modular RNA containing XPT-Riboswitch and Hammerhead

(Left) Cleavage assay reactions (A, B, C) of designed hammerhead (wild-type), mutant C116G and mutant G142U. For the wild-type (A), mutant C116G (B), and mutant G142U (C) gel images, lane 1 is the undigested RNA (full-length, FT), lanes 2–5 are reactions in cleavage buffer (50 mM Tris pH 7.5, 5 mM MgCl₂) at the 0 s, 30 min, 5 h and 24 h time points respectively (5' and 3' cleavage products indicated). For the wild-type (A), lane 6 is a reaction lacking Mg²⁺ (50 mM tris pH 7.5) incubated for 24 h. Cleavage only occurs in the “wild type” sequence, and when Mg²⁺ is present. (Right) Cleavage time series curve (D) for the 166 nt designed hammerhead, with observed cleavage rate of 1.3/min with an F_{\max} of 0.47 and MSE of 0.0026.

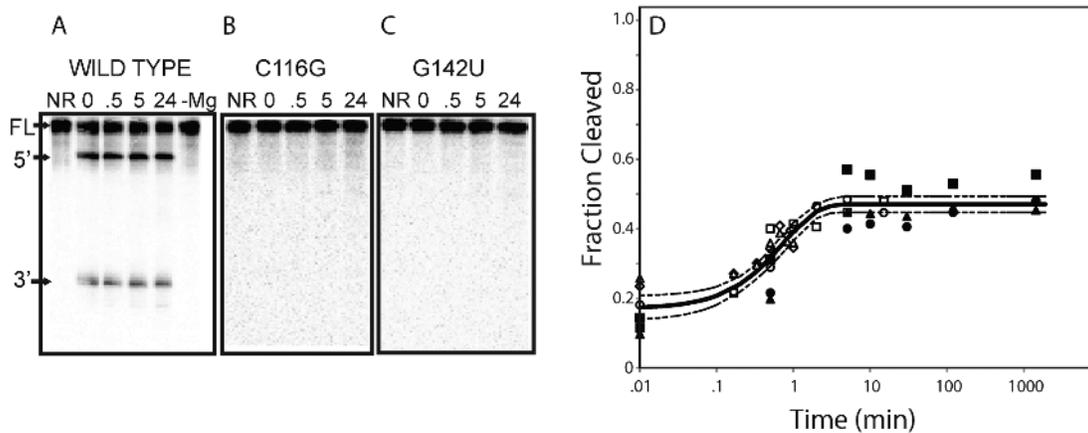
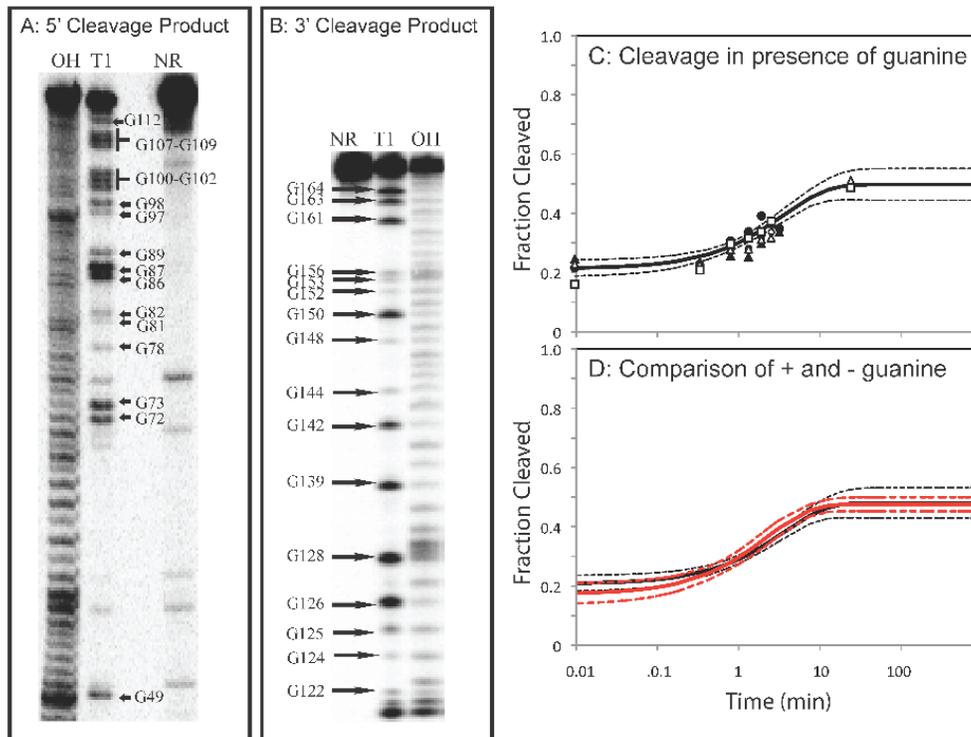


Figure A1.9 Structure Probing Confirms Modular RNA Cleaves at Expected Site

(A) T1-RNase mapping of the 5' cleavage product under denaturing conditions. Lane 1 is the partial alkaline digest (OH), followed by the RNase T1 digest (T1), and the undigested RNA (NR). (B) T1-RNase mapping of the 3' cleavage product under denaturing conditions. Lane 1 is the undigested product, followed by the RNase T1 digest (T1), and the partial alkaline digest (OH). The T1 digest of both the 5' cleavage product and the 3' cleavage product are consistent with the predicted cleavage at position C116. (C) Cleavage kinetics for 166 nt RNA in the presence of 1 mM guanine. Five independent time courses were conducted and the best-fit curve and the 95% confidence interval (dashed lines) determined as described in the Methods. (D) The best-fit curves and 95% confidence interval of the cleavage in the presence (black) and absence (red) of guanine plotted on the same set of axes. Data for the cleavage of 166 nt RNA in the absence of guanine is re-plotted from Figure A1.8 for comparison.



ABBREVIATIONS & DEFINITIONS

Ec-mRNA: regulatory mRNA from *E. coli* that responds to S15

Ec-S15: S15 homolog from *E. coli*

GAISR: Genomic Analysis for Illuminating Structure RNA

Gk-mRNA: regulatory mRNA from *G. kaustophilus* that responds to S15

Gk-S15: S15 homolog from *G. kaustophilus*

Gs-mRNA: regulatory mRNA from *G. stearothermophilus* that responds to S15

Gs-S15: S15 homolog from *G. stearothermophilus*

Kan^R: kanamycin resistance gene

pEMPTY: pBAD33 vector that does not contain an *rpsO* gene

PLMVd: Peach Latent Mosaic Viroid

ncRNA: non-coding ribonucleic acid

r-proteins: ribosomal proteins

ribozyme: ribonucleic acid enzyme

rpsO: gene encoding S15 by interacting with a structured RNA in its 5'-UTR

Rra-RNA: regulatory mRNA from *R. radiobacter* that responds to S15, a.k.a. Rr-mRNA

Rr-mRNA: regulatory mRNA from *R. radiobacter* that responds to S15, a.k.a Rra-RNA

Rr-S15: S15 homolog from *R. radiobacter*

SD: Shine Dalgarno sequence, part of the ribosome binding site on mRNA transcripts

SELEX: Systematic Evolution of Ligands by Exponential Enrichment

Tt-mRNA: regulatory mRNA from *T. thermophilus* that responds to S15

Tt-S15: S15 homolog from *T. thermophilus*

3HJ: three-helix junction

REFERENCES

- Agalarov, S. C., Sridhar Prasad, G., Funke, P. M., Stout, C. D., & Williamson, J. R. (2000). Structure of the S15,S6,S18-rRNA Complex: Assembly of the 30S Ribosome Central Domain. *Science*, *288*, 1–7.
- Albrecht, M., Sharma, C. M., Reinhardt, R., Vogel, J., & Rudel, T. (2010). Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome. *Nucleic Acids Research*, *38*(3), 868–877. <http://doi.org/10.1093/nar/gkp1032>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, *215*(3):403-10.
- Amaral, P. P., & Mattick, J. S. (2008). Noncoding RNA in development. *Mammalian Genome*, *19*(7-8), 454–492. <http://doi.org/10.1007/s00335-008-9136-7>
- Arnvig, K. B., Comas, I., Thomson, N. R., Houghton, J., Boshoff, H. I., Croucher, N. J., et al. (2011). Sequence-Based Analysis Uncovers an Abundance of Non-Coding RNA in the Total Transcriptome of *Mycobacterium tuberculosis*. *PLoS Pathogens*, *7*(11), e1002342. <http://doi.org/10.1371/journal.ppat.1002342.s012>
- Ban, N., Beckmann, R., Cate, J. I. H., Dinman, J. D., Dragon, F., Ellis, S. R., et al. (2014). A new system for naming ribosomal proteins. *Current Opinion in Structural Biology*, *24*, 165–169. <http://doi.org/10.1016/j.sbi.2014.01.002>
- Barrick, J. E., & Breaker, R. R. (2007). The power of riboswitches. *Scientific American*, *296*(1), 50–57. Retrieved from <http://www.nature.com/scientificamerican/journal/v296/n1/pdf/scientificamerican0107-50.pdf>
- Batey, R. T., & Williamson, J. R. (1996a). Interaction of the *Bacillus stearothermophilus* Ribosomal Protein S15 with 16 S rRNA: I. Defining the Minimal RNA Site, 1–14.
- Batey, R. T., & Williamson, J. R. (1996b). Interaction of the *Bacillus stearothermophilus* Ribosomal Protein S15 with 16 S rRNA:II. Specificity Determinants of RNA-Protein Recognition, 1–18.
- Batey, R. T., Gilbert, S. D., & Montange, R. K. (2004). Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, *432*(7015), 411–415. <http://doi.org/10.1038/nature03037>
- Bayer, T. S., & Smolke, C. D. (2005). Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nature Biotechnology*, *23*(3), 337–343. <http://doi.org/10.1038/nbt1069>
- Belmont, B. J., & Niles, J. C. (2010). Engineering a Direct and Inducible Protein–RNA Interaction To Regulate RNA Biology. *ACS Chemical Biology*, *5*(9), 851–861. <http://doi.org/10.1021/cb100070j>
- Benard, L., Mathy, N., Grunberg-Manago, M., Ehresmann, B., Ehresmann, C., & Portier, C. (1998). Identification in a pseudoknot of a U.G motif essential for the regulation of the expression of ribosomal protein S15. *Proceedings of the National Academy of Sciences*, *95*(5), 2564–2567.
- Benner, S. A. (1989). Modern Metabolism as a Palimpsest of the RNA World. *Proceedings of the National Academy of Sciences*, *86*(18), 7054–7058. <http://doi.org/10.1073/pnas.86.18.7054>
- Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., & Thgen, N. B. U. (2013). Efficient translation initiation dictates codon usage at gene start. *Molecular Systems Biology*, *9*, 1–10. <http://doi.org/10.1038/msb.2013.32>

- Block, K. F., Puerta-Fernandez, E., Wallace, J. G., & Breaker, R. R. (2010). Association of OLE RNA with bacterial membranes via an RNA-protein interaction. *Molecular Microbiology*, 79(1), 21–34. <http://doi.org/10.1111/j.1365-2958.2010.07439.x>
- Blount, K. F., & Uhlenbeck, O. C. (2005). The Structure-Function Dilemma of the Hammerhead Ribozyme. *Annual Review of Biophysics and Biomolecular Structure*, 34(1), 415–440. <http://doi.org/10.1146/annurev.biophys.34.122004.184428>
- Breaker, R. R. (2012). Riboswitches and the RNA World. *Cold Spring Harbor Perspectives in Biology*, 4(2), a003566–a003566. <http://doi.org/10.1101/cshperspect.a003566>
- Brooks, K. M., & Hampel, K. J. (2009). A Rate-Limiting Conformational Step in the Catalytic Pathway of the glmS Ribozyme. *Biochemistry*, 48(24), 5669–5678. <http://doi.org/10.1021/bi900183r>
- Bubunencko, M., Korepanov, A., Court, D. L., Jagannathan, I., Dickinson, D., Chaudhuri, B. R., et al. (2006). 30S ribosomal subunits can be assembled in vivo without primary binding ribosomal protein S15. *RNA (New York, N.Y.)*, 12(7), 1229–1239. <http://doi.org/10.1261/rna.2262106>
- Burge, S. W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E. P., et al. (2012). Rfam 11.0: 10 years of RNA families. *Nucleic Acids Research*, 41(D1), D226–D232. <http://doi.org/10.1093/nar/gks1005>
- Cadwell, R. C., & Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *Genome Research*, 2(1), 28–33. <http://doi.org/10.1101/gr.2.1.28>
- Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., et al. (2002). The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*.
- Carbonell, A., la Pena, De, M., Flores, R., & Gago, S. (2006). Effects of the trinucleotide preceding the self-cleavage site on eggplant latent viroid hammerheads: differences in co- and post-transcriptional self-cleavage may explain the lack of trinucleotide AUC in most natural hammerheads. *Nucleic Acids Research*, 34(19), 5613–5622. <http://doi.org/10.1093/nar/gk1717>
- Charpentier, E., & Hess, W. E. (2015). Editorial: RNA in bacteria: biogenesis, regulatory mechanisms and functions. *FEMS Microbiology Reviews*, 39(3), 277–279. <http://doi.org/10.1093/femsre/fuv025>
- Choonee, N., Even, S., Zig, L., & Putzer, H. (2007). Ribosomal protein L20 controls expression of the Bacillus subtilis infC operon via a transcription attenuation mechanism. *Nucleic Acids Research*, 35(5), 1578–1588. <http://doi.org/10.1093/nar/gkm011>
- Cluot-d'Orval, B., & Uhlenbeck, O. C. (1997). Hammerhead Ribozymes with a Faster Cleavage Rate. *Biochemistry*, 1–6.
- Collins, J. J. (2012). Bits and pieces come to life. *Nature*, 1–3.
- Corbino, K. A., Barrick, J. E., Lim, J., Welz, R. U. D., Tucker, B. J., Puskarz, I. (2005). Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. *Genome Biology*, 6(8), R70. <http://doi.org/10.1186/gb-2005-6-8-r70>
- Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research*, 16, 1–10.

- Croft, M. T., Moulin, M., Webb, M. E., & Smith, A. G. (2007). Thiamine biosynthesis in algae is regulated by riboswitches. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(52), 20770–20775. <http://doi.org/10.1073/pnas.0705786105>
- Cromie, M. J., Shi, Y., Latifi, T., & Groisman, E. A. (2006). An RNA sensor for intracellular Mg(2+). *Cell*, *125*(1), 71–84. <http://doi.org/10.1016/j.cell.2006.01.043>
- Crooks, G. E., Hon, G., Chandonia, J.-M., & Brenner, S. E. (2004). WebLogo: A Sequence Logo Generator. *Genome Research*, 1–4.
- Dambach, M. D., & Winkler, W. C. (2009). Expanding roles for metabolite-sensing regulatory RNAs. *Current Opinion in Microbiology*, *12*(2), 161–169. <http://doi.org/10.1016/j.mib.2009.01.012>
- Dammel, C. S., & Noller, H. F. (1993). A cold-sensitive mutation in 16S rRNA provides evidence for helical switching in ribosome assembly. *Genes & Development*, 1–12.
- Dann, C. E., III, Wakeman, C. A., Sieling, C. L., Baker, S. C., Irnov, I., & Winkler, W. C. (2007). Structure and Mechanism of a Metal-Sensing Regulatory RNA. *Cell*, *130*(5), 878–892. <http://doi.org/10.1016/j.cell.2007.06.051>
- Darmostuk, M., Rimpelova, S., Gbelcova, H., & Ruml, T. (2015). Biotechnology Advances. *Biotechnology Advances*, *33*(P2), 1141–1161. <http://doi.org/10.1016/j.biotechadv.2015.02.008>
- Datta, S., Costantino, N., & Court, D. L. (2006). A set of recombineering plasmids for gram-negative bacteria. *Gene*, *379*, 109–115. <http://doi.org/10.1016/j.gene.2006.04.018>
- Davidson, M. E., Harbaugh, S. V., Chushak, Y. G., Stone, M. O., & Kelley-Loughnane, N. (2013). Development of a 2,4-Dinitrotoluene-Responsive Synthetic Riboswitch in E. coli Cells. *ACS Chemical Biology*, *8*(1), 234–241. <http://doi.org/10.1021/cb300274g>
- Deiorio-Haggar, K., Anthony, J., & Meyer, M. M. (2013). RNA structures regulating ribosomal protein biosynthesis in bacilli. *RNA Biology*, *10*(7), 1180–1184. <http://doi.org/10.4161/rna.24151>
- Eichhorn, C. D., Kang, M., & Feigon, J. (2014). Structure and function of preQ1 riboswitches. *Biochimica Et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1839*(10), 939–950. <http://doi.org/10.1016/j.bbagr.2014.04.019>
- Ellington, A. D., & Szostak, J. W. (2002). In vitro selection of RNA molecules that bind specific ligands, 1–5.
- Filonov, G. S., Moon, J. D., Svensen, N., & Jaffrey, S. R. (2014). Broccoli: Rapid Selection of an RNA Mimic of Green Fluorescent Protein by Fluorescence-Based Selection and Directed Evolution. *Journal of the American Chemical Society*, *136*(46), 16299–16308. <http://doi.org/10.1021/ja508478x>
- Fox, G. E. (2010). Origin and Evolution of the Ribosome. *Cold Spring Harbor Perspectives in Biology*, *2*(9), a003483–a003483. <http://doi.org/10.1101/cshperspect.a003483>
- Fu, Y., Deiorio-Haggar, K., Anthony, J., & Meyer, M. M. (2013). Most RNAs regulating ribosomal protein biosynthesis in Escherichia coli are narrowly distributed to Gammaproteobacteria. *Nucleic Acids Research*, *41*(6), 3491–3503. <http://doi.org/10.1093/nar/gkt055>
- Fuchs, R. T., Grundy, F. J., & Henkin, T. M. (2006). The SMK box is a new SAM-

- binding RNA for translational regulation of SAM synthetase. *Nature Structural & Molecular Biology*, 13(3), 226–233. <http://doi.org/10.1038/nsmb1059>
- GARCIA-MARTIN, J. A., CLOTE, P., & DOTU, I. (2013). RNAiFOLD: A CONSTRAINT PROGRAMMING ALGORITHM FOR RNA INVERSE FOLDING AND MOLECULAR DESIGN. *Journal of Bioinformatics and Computational Biology*, 11(02), 1350001. <http://doi.org/10.1142/S0219720013500017>
- Garcia-Martin, J. A., Clote, P., & Dotu, I. (2013). RNAiFold: a web server for RNA inverse folding and molecular design. *Nucleic Acids Research*, 1–6. <http://doi.org/10.1093/nar/gkt280/-/DC1>
- Gardner, P. P., Daub, J., Tate, J., Moore, B. L., Osuch, I. H., Griffiths-Jones, S., et al. (2011). Rfam: Wikipedia, clans and the “decimal” release. *Nucleic Acids Research*, 39(Database), D141–D145. <http://doi.org/10.1093/nar/gkq1129>
- Garst, A. D., Heroux, A., Rambo, R. P., & Batey, R. T. (2008). Crystal structure of the lysine riboswitch regulatory mRNA element. *The Journal of Biological Chemistry*, 283(33), 22347–22351. <http://doi.org/10.1074/jbc.C800120200>
- Gelfand, M. S. (2006). Bacterial cis-regulatory RNA structures. *Molecular Biology*, 40(4), 541–550. <http://doi.org/10.1134/S0026893306040066>
- Giannoukos, G., Ciulla, D. M., Huang, K., Haas, B. J., Izard, J., Levin, J. Z., et al. (2012). Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biology*, 13(3), r23.
- Gilbert, S. D., Rambo, R. P., Van Tyne, D., & Batey, R. T. (2008). Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nature Structural Molecular Biology*, 15(2), 177–182. <http://doi.org/10.1038/nsmb.1371>
- Goldfless, S. J., Belmont, B. J., de Paz, A. M., Liu, J. F., & Niles, J. C. (2012). Direct and specific chemical control of eukaryotic translation with a synthetic RNA-protein interaction. *Nucleic Acids Research*, 40(9), e64–e64. <http://doi.org/10.1093/nar/gks028>
- Gonzalez-Carmona, M.-A., Schüssler, S., Serwe, M., Alt, M., Ludwig, J., Sproat, B. S., et al. (2006). Hammerhead ribozymes with cleavage site specificity for NUH and NCH display significant anti-hepatitis C viral effect in vitro and in recombinant HepG2 and CCL13 cells. *Journal of Hepatology*, 44(6), 1017–1025. <http://doi.org/10.1016/j.jhep.2005.10.022>
- Grkovic, S., Brown, M. H., & Skurray, R. A. (2002). Regulation of Bacterial Drug Export Systems. *Microbiology and Molecular Biology Reviews : MMBR*, 66(4), 671–701. <http://doi.org/10.1128/MMBR.66.4.671-701.2002>
- Grundy, F. J., & Henkin, T. (1992). Characterization of the Bacillus subtilis rpsD Regulatory Target Site. *Journal of Bacteriology*, 174.
- Grundy, F. J., & Henkin, T. M. (1998). The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria. *Molecular Microbiology*, 30(4), 737–749. <http://doi.org/10.1046/j.1365-2958.1998.01105.x>
- Gu, W., Zhou, T., & Wilke, C. O. (2010). A Universal Trend of Reduced mRNA Stability near the Translation-Initiation Site in Prokaryotes and Eukaryotes. *PLoS Computational Biology*, 6(2), e1000664. <http://doi.org/10.1371/journal.pcbi.1000664.s015>
- Guillier, M., Allemand, F., Dardel, F., Royer, C. A., Springer, M., & Chiaruttini, C.

- (2005). Double molecular mimicry in *Escherichia coli*: binding of ribosomal protein L20 to its two sites in mRNA is similar to its binding to 23S rRNA. *Molecular Microbiology*, 56(6), 1441–1456. <http://doi.org/10.1111/j.1365-2958.2005.04644.x>
- Gusarov, I., & Nudler, E. (1999). The Mechanism of Intrinsic Transcription Termination. *Molecular Cell*, 3, 495.
- Guthrie, C., Nashimoto, H., & Nomura, M. (1969). Structure and Function of *E. coli* Ribosomes, VIII. Cold-sensitive Mutants Defective in Ribosome Assembly. *Proc. Natl. Acad. Sci. USA*, 63, 1–8.
- Gutierrez-Preciado, A., Henkin, T. M., Grundy, F. J., Yanofsky, C., & Merino, E. (2009). Biochemical Features and Functional Implications of the RNA-Based T-Box Regulatory Mechanism. *Microbiology and Molecular Biology Reviews : MMBR*, 73(1), 36–61. <http://doi.org/10.1128/MMBR.00026-08>
- Hall, K.B., Kranz, J.K. (1999). Nitrocellulose filter binding for determination of dissociation constants. *Methods Mol. Biol.*, 118:105-14.
- Hanson, S., Bauer, G., Fink, B., & Suess, B. (2005). Molecular analysis of a synthetic tetracycline-binding riboswitch. *RNA (New York, N.Y.)*, 11(4), 503–511. <http://doi.org/10.1261/rna.7251305>
- Harms, J., Schlutzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., et al. (2001). High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell*, 107(5), 679–688.
- Harvey, R. J. (1970). Regulation of ribosomal protein synthesis in *Escherichia coli*. *Journal of Bacteriology*, 101(2), 574–583.
- Held, W. A., Ballou, B., Mizushima, S., & Nomura, M. (1974). Assembly Mapping of 30S Ribosomal Proteins from *Escherichia coli*. *The Journal of Biological Chemistry*, 249, 3103–3111.
- Isaacs, F. J., Dwyer, D. J., & Collins, J. J. (2006). RNA synthetic biology. *Nature Biotechnology*, 24(5), 545–554. <http://doi.org/10.1038/nbt1208>
- Isaacs, F. J., Dwyer, D. J., Ding, C., Pervouchine, D. D., Cantor, C. R., & Collins, J. J. (2004). Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology*, 22(7), 841–847. <http://doi.org/10.1038/nbt986>
- Jagannathan, I., & Culver, G. M. (2003). Assembly of the Central Domain of the 30S Ribosomal Subunit: Roles for the Primary Binding Ribosomal Proteins S15 and S8. *Journal of Molecular Biology*, 330(2), 373–383. [http://doi.org/10.1016/S0022-2836\(03\)00586-2](http://doi.org/10.1016/S0022-2836(03)00586-2)
- Johansson, J., & Cossart, P. (2003). RNA-mediated control of virulence gene expression in bacterial pathogens. *Trends in Microbiology*, 11(6), 280–285. [http://doi.org/10.1016/S0966-842X\(03\)00118-5](http://doi.org/10.1016/S0966-842X(03)00118-5)
- Kim, J. N., Roth, A., & Breaker, R. R. (2007). Guanine riboswitch variants from *Mesoplasma florum* selectively recognize 2-deoxyguanosine. *Proc. Natl. Acad. Sci. USA*, 104, 16092–16097.
- Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., & Ban, N. (2012). Atomic structures of the eukaryotic ribosome. *Trends in Biochemical Sciences*, 37(5), 189–198. <http://doi.org/10.1016/j.tibs.2012.02.007>
- Korostelev, A., Trakhanov, S., Laurberg, M., & Noller, H. F. (2006). Crystal Structure of a 70S Ribosome-tRNA Complex Reveals Functional Interactions and Rearrangements. *Cell*, 126(6), 1065–1077. <http://doi.org/10.1016/j.cell.2006.08.032>

- Lindahl, L., & Zengel, J. M. (1986). Ribosomal Genes in *Escherichia coli*. *Annual Review of Genetics*, 1–30.
- Lindgreen, S., Umu, S. U., Lai, A. S.-W., Eldai, H., Liu, W., McGimpsey, S., et al. (2014). Robust Identification of Noncoding RNA from Transcriptomes Requires Phylogenetically-Informed Sampling. *PLoS Computational Biology*, 10(10), e1003907. <http://doi.org/10.1371/journal.pcbi.1003907.s003>
- Livny, J., & Waldor, M. K. (2007). Identification of small RNAs in diverse bacterial species. *Current Opinion in Microbiology*, 10(2), 96–101. <http://doi.org/10.1016/j.mib.2007.03.005>
- Lorenz, R., Bernhart, S. H., Honer zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). Vienna RNA Package 2.0. *Algorithms for Molecular Biology*, 1–14.
- Lu, C., Smith, A. M., Fuchs, R. T., Ding, F., Rajashankar, K., Henkin, T. M., & Ke, A. (2008). Crystal structures of the SAM-III/SMK riboswitch reveal the SAM-dependent translation inhibition mechanism. *Nature Structural & Molecular Biology*, 15(10), 1076–1083. <http://doi.org/10.1038/nsmb.1494>
- Lu, X., Goodrich-Blair, H., & Tjaden, B. (2011). Assessing computational tools for the discovery of small RNA genes in bacteria. *RNA (New York, N.Y.)*, 17(9), 1635–1647. <http://doi.org/10.1261/rna.2689811>
- Lundrigan, M. D., Koster, W., & Kadner, R. J. (1991). Transcribed sequences of the *Escherichia coli* *btuB* gene control its expression and regulation by vitamin B12. *Genetics*, 88, 1479–1483.
- Mandal, M., & Breaker, R. R. (2003). Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature Structural & Molecular Biology*, 11(1), 29–35. <http://doi.org/10.1038/nsmb710>
- Mandal, M., Lee, M., Barrick, J. E., Weinberg, Z., Emilsson, G. M., Ruzzo, W. L., & Breaker, R. R. (2004). A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science (New York, N.Y.)*, 306(5694), 275–279. <http://doi.org/10.1126/science.1100829>
- Martick, M., & Scott, W. G. (2006). Tertiary Contacts Distant from the Active Site Prime a Ribozyme for Catalysis. *Cell*, 126(2), 309–320. <http://doi.org/10.1016/j.cell.2006.06.036>
- Mathy, N., Pellegrini, O., Serganov, A., Patel, D. J., Ehresmann, C., & Portier, C. (2004). Specific recognition of *rpsO* mRNA and 16S rRNA by *Escherichia coli* ribosomal protein S15 relies on both mimicry and site differentiation. *Molecular Microbiology*, 52(3), 661–675. <http://doi.org/10.1111/j.1365-2958.2004.04005.x>
- McCown, P. J., Liang, J. J., Weinberg, Z., & Breaker, R. R. (2014). Structural, Functional, and Taxonomic Diversity of Three PreQ1. *Chemistry & Biology*, 21(7), 880–889. <http://doi.org/10.1016/j.chembiol.2014.05.015>
- Merhej, V., & Raoult, D. (2011). Rickettsial evolution in the light of comparative genomics. *Biological Reviews*, 86(2), 379–405. <http://doi.org/10.1111/j.1469-185X.2010.00151.x>
- Meyer, M. M., Roth, A., Chervin, S. M., Garcia, G. A., & Breaker, R. R. (2008). Confirmation of a second natural preQ1 aptamer class in Streptococcaceae bacteria. *Rna New York Ny*, 14(4), 685–695.
- Milligan, J. F., Groebe, D., Witherell, G., & Uhlenbeck, O. C. (1987).

- Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Research*, 1–16.
- Montange, R. K., & Batey, R. T. (2006). Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature*, 441(7097), 1172–1175. <http://doi.org/10.1038/nature04819>
- Mougel, M., Philippe, C., Ebel, J. P., Ehresmann, B., & Ehresmann, C. (1988). The E. coli 16S rRNA binding site of ribosomal protein S15 : higher-order structure in the absence and presence of the protein. *Nucleic Acids Research*, 16, 1–15.
- Mulder, A. M., Yoshioka, C., Beck, A. H., Bunner, A. E., Milligan, R. A., Potter, C. S., et al. (2010). Visualizing ribosome biogenesis: parallel assembly pathways for the 30S subunit. *Science (New York, N.Y.)*, 330(6004), 673–677. <http://doi.org/10.1126/science.1193220>
- Muranaka, N., Sharma, V., Nomura, Y., & Yokobayashi, Y. (2009). An efficient platform for genetic selection and screening of gene switches in Escherichia coli. *Nucleic Acids Research*, 37(5), e39. <http://doi.org/10.1093/nar/gkp039>
- Nahvi, A., Barrick, J. E., & Breaker, R. R. (2004). Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Research*, 32(1), 143–150. <http://doi.org/10.1093/nar/gkh167>
- Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, 29(22), 2933–2935. <http://doi.org/10.1093/bioinformatics/btt509>
- Nelson, J. A., & Uhlenbeck, O. C. (2008). Hammerhead redux: Does the new structure fit the old biochemical data? *RNA (New York, N.Y.)*, 14(4), 605–615. <http://doi.org/10.1261/rna.912608>
- Nevskaya, N., Tishchenko, S., Gabdoulkhakov, A., Nikonova, E., Nikonov, O., Nikulin, A., et al. (2005). Ribosomal protein L1 recognizes the same specific structural motif in its target sites on the autoregulatory mRNA and 23S rRNA. *Nucleic Acids Research*, 33(2), 478–485. <http://doi.org/10.1093/nar/gki194>
- Nikulin, A., Serganov, A., Ennifar, E., Tishchenko, S., Nevskaya, N., Shepard, W., et al. (2000). Crystal structure of the S15-rRNA complex. *Nature Structural Biology*, 7(4), 273–277. <http://doi.org/10.1038/74028>
- Nissen, P., Hansen, J., Ban, N., Moore, P. B., & Steitz, T. A. (2000). The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science*, 1–12.
- Pan, W. H., Xin, P., Vui, V., Clawson, G. A. (2003). Rapid identification of efficient target cleavage sites using a hammerhead ribozyme library in an iterative manner. *Mol. Ther.*, 7, 129-139
- Piganeau, N. (2009). In vitro selection of allosteric ribozymes. *Methods Mol. Biol.*, 535, 45-57.
- Philippe, C., Benard, L., Eyermann, F., Cachia, C., Kirillov, S. V., Portier, C. (1994). Structural elements of rps0 mRNA involved in the modulation of translational initiation and regulation of E. coli ribosomal protein S15. *Nucleic Acids Research*, 22(13), 2538–2546.
- Philippe, C., Benard, L., Portier, C., Westhof, E., Ehresmann, B., & Ehresmann, C. (1995). Molecular dissection of the pseudoknot governing the translational regulation of Escherichia coli ribosomal protein S15. *Nucleic Acids Research*, 23(1), 18–28.
- Philippe, C., Eyermann, F., Benard, L., Portier, C., Ehresmann, B., & Ehresmann, C.

- (1993). Ribosomal protein S15 from *Escherichia coli* modulates its own translation by trapping the ribosome on the mRNA initiation loading site. *Proc. Natl. Acad. Sci. USA Biochemistry*, *90*, 4394–4398.
- Philippe, C., Portier, C., Mougel, M., Grunberg-Manago, M., Ebel, J. P., Ehresmann, B., & Ehresmann, C. (1990). Target site of *Escherichia coli* ribosomal protein S15 on its messenger RNA. Conformation and interaction with the protein. *Journal of Molecular Biology*, *211*(2), 415–426. Retrieved from http://ac.els-cdn.com/002228369090362P/1-s2.0-002228369090362P-main.pdf?_tid=1c2a9cea-0043-11e3-b672-00000aab0f6c&acdnat=1375977579_f3104d0020ea19cfc87047707e347eee
- Plotkin, J. B., & Kudla, G. (2010). Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics*, *12*(1), 32–42. <http://doi.org/10.1038/nrg2899>
- Poiata, E., Meyer, M. M., Ames, T. D., & Breaker, R. R. (2009). A variant riboswitch aptamer class for S-adenosylmethionine common in marine bacteria. *RNA (New York, N.Y.)*, *15*(11), 2046–2056. <http://doi.org/10.1261/rna.1824209>
- Portier, C., Dondon, L., & Grunberg-Manago, M. (1990). Translational autocontrol of the *Escherichia coli* ribosomal protein S15. *Journal of Molecular Biology*, *211*(2), 407–414. [http://doi.org/10.1016/0022-2836\(90\)90361-O](http://doi.org/10.1016/0022-2836(90)90361-O)
- Powers, T., & Noller, H. F. (1995). Hydroxyl radical footprinting of ribosomal proteins on 16S rRNA. *RNA (New York, N.Y.)*, *1*(2), 194–209.
- Pruitt, K. D., Tatusova, T., Brown, G. R., & Maglott, D. R. (2011). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Research*, *40*(D1), D130–D135. <http://doi.org/10.1093/nar/gkr1079>
- Pupko, T., Mayrose, I., Glaser, F., & Ben-Tal, N. (2002). Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*, 1–7.
- Ray, D., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation, 1–6. <http://doi.org/10.1038/nature12311>
- Recht, M. I., & Williamson, J. R. (2001). Central domain assembly: thermodynamics and kinetics of S6 and S18 binding to an S15-RNA complex. *Journal of Molecular Biology*, *313*(1), 35–48. <http://doi.org/10.1006/jmbi.2001.5018>
- Regulski, E. E., & Breaker, R. R. (2008). In-line probing analysis of riboswitches. *Methods in Molecular Biology (Clifton, N.J.)*, *419*(5), 53–67. http://doi.org/10.1007/978-1-59745-033-1_4
- Root-Bernstein, M., & Root-Bernstein, R. (2015). The ribosome as a missing link in the evolution of life. *Journal of Theoretical Biology*, *367*(C), 130–158. <http://doi.org/10.1016/j.jtbi.2014.11.025>
- Roth, A., Winkler, W. C., Regulski, E. E., Lee, B. W. K., Lim, J., Jona, I. (2007). A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nature Structural & Molecular Biology*, *14*(4), 308–317. <http://doi.org/10.1038/nsmb1224>
- Ruff, K. M., & Strobel, S. A. (2014). Ligand binding by the tandem glycine riboswitch depends on aptamer dimerization but not double ligand occupancy. *RNA*, *20*(11), 1775–1788. <http://doi.org/10.1261/rna.047266.114>

- Saragliadis, A., Krajewski, S. S., Rehm, C., Narberhaus, F., & Hartig, J. S. (2014). Thermozyymes. *RNA Biology*, *10*(6), 1009–1016. <http://doi.org/10.4161/rna.24482>
- Schaffer, M. F., Choudhary, P. K., & Sigel, R. K. O. (2014). The AdoCbl–Riboswitch Interaction Investigated by In-Line Probing and Surface Plasmon Resonance Spectroscopy (SPR). *Riboswitch Discovery, Structure and Function* (1st ed., Vol. 549, pp. 467–488). Elsevier Inc. <http://doi.org/10.1016/B978-0-12-801122-5.00020-9>
- Schmeing, T. M., & Ramakrishnan, V. (2009). What recent ribosome structures have revealed about the mechanism of translation. *Nature*, *461*(7268), 1234–1242. <http://doi.org/10.1038/nature08403>
- Schultes, E. A., & Bartel, D. P. (2000). One Sequence, Two Ribozymes: Implications for the Emergence of New Ribozyme Folds. *Science*, 1–6.
- Scott, L. G., & Williamson, J. R. (2001). Interaction of the *Bacillus stearothermophilus* ribosomal protein S15 with its 5'-translational operator mRNA. *Journal of Molecular Biology*, *314*(3), 413–422. <http://doi.org/10.1006/jmbi.2001.5165>
- Scott, L. G., & Williamson, J. R. (2005). The binding interface between *Bacillus stearothermophilus* ribosomal protein S15 and its 5'-translational operator mRNA. *Journal of Molecular Biology*, *351*(2), 280–290. <http://doi.org/10.1016/j.jmb.2005.06.030>
- Senter, E., Sheikh, S., DOTU, I., Ponty, Y., & CLOTE, P. (2012). Using the Fast Fourier Transform to Accelerate the Computational Search for RNA Conformational Switches. *PloS One*, *7*(12), e50506. <http://doi.org/10.1371/journal.pone.0050506.s001>
- Serganov, A. A., Masquida, B., Westhof, E., Cachia, C., Portier, C., Garber, M., et al. (1996). The 16S rRNA binding site of *Thermus thermophilus* ribosomal protein S15: comparison with *Escherichia coli* S15, minimum site and structure. *RNA (New York, N.Y.)*, *2*(11), 1124–1138.
- Serganov, A., Benard, L., Portier, C., Ennifar, E., Garber, M., Ehresmann, B., & Ehresmann, C. (2001). Role of conserved nucleotides in building the 16 S rRNA binding site for ribosomal protein S15. *Journal of Molecular Biology*, *305*(4), 785–803. <http://doi.org/10.1006/jmbi.2000.4354>
- Serganov, A., Ennifar, E., Portier, C., Ehresmann, B., & Ehresmann, C. (2002). Do mRNA and rRNA Binding Sites of *E. coli* Ribosomal Protein S15 Share Common Structural Determinants? *Journal of Molecular Biology*, *320*(5), 963–978. [http://doi.org/10.1016/S0022-2836\(02\)00553-3](http://doi.org/10.1016/S0022-2836(02)00553-3)
- Serganov, A., Polonskaia, A., Ehresmann, B., Ehresmann, C., & Patel, D. J. (2003). Ribosomal protein S15 represses its own translation via adaptation of an rRNA-like fold within its mRNA. *The EMBO Journal*, *22*(8), 1898–1908. <http://doi.org/10.1093/emboj/cdg170>
- Serganov, A., Yuan, Y.-R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A. T., et al. (2004). Structural Basis for Discriminative Regulation of Gene Expression by Adenine- and Guanine-Sensing mRNAs. *Chemistry & Biology*, *11*(12), 1729–1741. <http://doi.org/10.1016/j.chembiol.2004.11.018>
- Sinha, J., Reyes, S. J., & Gallivan, J. P. (2010). Reprogramming bacteria to seek and destroy an herbicide. *Nature Chemical Biology*, *6*(6), 464–470. <http://doi.org/10.1038/nchembio.369>
- Slinger, B. L., Deiorio-Haggar, K., Anthony, J. S., Gilligan, M. M., & Meyer, M. M.

- (2014). Discovery and validation of novel and distinct RNA regulators for ribosomal protein S15 in diverse bacterial phyla. *BMC Genomics*, *15*(1), 1–12. <http://doi.org/10.1186/1471-2164-15-657>
- Slinger, B., Newman, H., Lee, Y., Pei, S., & Meyer, M. M. (2015). Co-evolution of Bacterial Ribosomal Protein S15 with Diverse mRNA Regulatory Structures. *PLoS Genetics*, 1–21. <http://doi.org/10.1371/journal.pgen.1005720>
- Smith, A. M., Fuchs, R. T., Grundy, F. J., & Henkin, T. (2014). Riboswitch RNAs: Regulation of gene expression by direct monitoring of a physiological signal. *RNA Biology*, *7*(1), 104–110. <http://doi.org/10.4161/rna.7.1.10757>
- Staple, D. W., & Butcher, S. E. (2005). Pseudoknots: RNA Structures with Diverse Functions. *PLoS Biology*, 1–4. <http://doi.org/10.1371/journal.pbio.0030213.g001>
- Sudarsan, N., Barrick, J. E., & Breaker, R. R. (2003). Metabolite-binding RNA domains are present in the genes of eukaryotes. *Rna*, *9*(6), 644–647. <http://doi.org/10.1261/rna.5090103>
- Sudarsan, N., Cohen-Chalamish, S., Nakamura, S., Emilsson, G. M., & Breaker, R. R. (2005). Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine. *Chemistry & Biology*, *12*(12), 1325–1335. <http://doi.org/10.1016/j.chembiol.2005.10.007>
- Sudarsan, N., Hammond, M. C., Block, K. F., Welz, R., Barrick, J. E., & Breaker, R. R. (2006). Tandem Riboswitch Architectures Exhibit Complex Gene Control Functions. *Science (New York, N.Y.)*, *314*(5797), 298–300. <http://doi.org/10.1126/science.1131000>
- Sudarsan, N., Lee, E. R., Weinberg, Z., Moy, R. H., Kim, J. N., Link, K. H., & Breaker, R. R. (2008). Riboswitches in Eubacteria Sense the Second Messenger Cyclic Di-GMP. *Science (New York, N.Y.)*, *321*(5887), 411–413. <http://doi.org/10.1126/science.1159519>
- Sykes, M. T., Shajani, Z., Sperling, E., Beck, A. H., & Williamson, J. R. (2010). Quantitative Proteomic Analysis of Ribosome Assembly and Turnover In Vivo. *Journal of Molecular Biology*, *403*(3), 331–345. <http://doi.org/10.1016/j.jmb.2010.08.005>
- Szeitner, Z., András, J., Gyurcsányi, R. E., & Mészáros, T. (2014). Journal of Pharmaceutical and Biomedical Analysis. *Journal of Pharmaceutical and Biomedical Analysis*, *101*, 58–65. <http://doi.org/10.1016/j.jpba.2014.04.018>
- Tang, C. K., & Draper, D. E. (1989). Unusual mRNA Pseudoknot Structure Is Recognized by a Protein Translational Repressor. *Cell*, *57*.
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 1–8.
- Tuerk, C., & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (New York, N.Y.)*, *249*(4968), 505–510.
- Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., et al. (2010). An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. *Cell*, *141*(2), 344–354. <http://doi.org/10.1016/j.cell.2010.03.031>
- Turner, D. H., & Mathews, D. H. (2009). NNDB: the nearest neighbor parameter

- database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Research*, 38(Database), D280–D282. <http://doi.org/10.1093/nar/gkp892>
- Valley, C. T., Porter, D. F., Qiu, C., Campbell, Z. T., Tanaka Hall, T. M., & Wickens, M. (2012). Patterns and plasticity in RNA-protein interactions enable recruitment of multiple proteins through a single site. *Proc. Natl. Acad. Sci. USA*, 109, 1–6. <http://doi.org/10.1073/pnas.1200521109/-/DCSupplemental/pnas.201200521SI.pdf>
- Wang, J. X., & Breaker, R. R. (2008). Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine. *Biochemistry and Cell Biology*, 86(2), 157–168. <http://doi.org/10.1139/O08-008>
- Waters, L. S., & Storz, G. (2009). Regulatory RNAs in Bacteria. *Cell*, 136(4), 615–628. <http://doi.org/10.1016/j.cell.2009.01.043>
- Weigand, J. E., Sanchez, M., Gunnesch, E. B., Zeiher, S., Schroeder, R., & Suess, B. (2008). Screening for engineered neomycin riboswitches that control translation initiation. *RNA (New York, N.Y.)*, 14(1), 89–97. <http://doi.org/10.1261/rna.772408>
- Weinberg, Z., Barrick, J. E., Yao, Z., Roth, A., Kim, J. N., Gore, J., et al. (2007). Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Research*, 35(14), 4809–4819. <http://doi.org/10.1093/nar/gkm487>
- Weinberg, Z., Regulski, E. E., Hammond, M. C., Barrick, J. E., Yao, Z., Ruzzo, W. L., & Breaker, R. R. (2008). The aptamer core of SAM-IV riboswitches mimics the ligand-binding site of SAM-I riboswitches. *RNA (New York, N.Y.)*, 14(5), 822–828. <http://doi.org/10.1261/rna.988608>
- Weinberg, Z., Wang, J. X., Bogue, J., Yang, J., Corbino, K. A., Moy, R. H., & Breaker, R. R. (2010). Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biology*, 1–17.
- Welz, R., & Breaker, R. (2007). Ligand binding and gene control characteristics of tandem riboswitches in *Bacillus anthracis*. *RNA*, 1–11.
- White, H. B. (2012). Coenzymes as fossils of an earlier metabolic state. *Journal of Molecular Evolution*, 7(2), 101–104. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1263263>
- Wickiser, J. K., Winkler, W. C., Breaker, R. R., & Crothers, D. M. (2005). The Speed of RNA Transcription and Metabolite Binding Kinetics Operate an FMN Riboswitch. *Molecular Cell*, 18(1), 49–60. <http://doi.org/10.1016/j.molcel.2005.02.032>
- Wieland, M., & Hartig, J. O. R. S. (2008). Improved aptazyme design and in vivo screening enable riboswitching in bacteria. *Angewandte Chemie (International Ed. in English)*, 47(14), 2604–2607. <http://doi.org/10.1002/anie.200703700>
- Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, A. P., Vornrhein, C., et al. (2000). Structure of the 30S ribosomal subunit. *Nature*, 407(6802), 327–339. <http://doi.org/10.1038/35030006>
- Winkler, W. C., Nahvi, A., Roth, A., Collins, J. J., & Breaker, R. (2004). Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, 1–6.
- Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E., & Breaker, R. R. (2003). An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nature Structural Biology*, 10(9), 701–707. <http://doi.org/10.1038/nsb967>
- Wunnicke, D., Strohbach, D., Weigand, J. E., Appel, B., Feresin, E., Suess, B., et al. (2010). Ligand-induced conformational capture of a synthetic tetracycline riboswitch

- revealed by pulse EPR. *RNA (New York, N.Y.)*, 17(1), 182–188.
<http://doi.org/10.1261/rna.2222811>
- Yadav, S., Swati, D., & Chandrasekharan, H. (2015). Thiamine Pyrophosphate Riboswitch in Some Representative Plant Species: A Bioinformatics Study. *Journal of Computational Biology*, 22(1), 1–9. <http://doi.org/10.1089/cmb.2014.0169>
- Yao, Z., Barrick, J., Weinberg, Z., Neph, S., Breaker, R., Tompa, M., & Ruzzo, W. L. (2007). A computational pipeline for high- throughput discovery of cis-regulatory noncoding RNA in prokaryotes. *PLoS Computational Biology*, 3(7), e126.
<http://doi.org/10.1371/journal.pcbi.0030126>
- Yao, Z., Weinberg, Z., & Ruzzo, W. L. (2006). CMfinder--a covariance model based RNA motif finding algorithm. *Bioinformatics*, 22(4), 445–452.
<http://doi.org/10.1093/bioinformatics/btk008>
- Yarnell, W. S., & Roberts, J. W. (1999). Mechanism of Intrinsic Transcription Termination and Antitermination. *Science*, 611–615.
- Yusupov, M. M., Yusopov, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. I. H., & Noller, H. F. (2001). Crystal Structure of the Ribosome at 5.5 Å Resolution. *Science*, 292, 1–14.
- Zengel, J. M., & Lindahl, L. (1994). Diverse Mechanisms for Regulating Ribosomal Protein Synthesis in Escherichia coli. *Progress in Nucleic Acid Research and Molecular Biology*, (47), 331–370.