ABSTRACT

Title of Dissertation: INVESTIGATING THE COMPETITION BETWEEN COMPONENTS OF DUAL-FUNCTION RNA

Jordan James Masuo Aoyama, Doctor of Philosophy, 2021

Dissertation Directed by: Dr. Gisela Storz
NIH Distinguished Investigator

Non-coding RNAs (ncRNAs) and small proteins have both emerged as important regulators of gene expression. Dual-function RNAs encode a small protein and have a separate function as a regulatory RNA. Although first discovered in bacteria, dual-function RNAs have now been identified and characterized in eukaryotes as well. These RNAs allow two activities of a single gene to regulate targets at multiple levels. The work described here explores how two novel and one synthetic dual-function RNA act and how competition between the components of a dual-function RNA impacts their functions. AzuCR is a 164-nucleotide *E. coli* RNA that was previously shown to encode a 28 amino acid protein (AzuC). This work demonstrates that the AzuC small protein impacts glycerol metabolism, with the small protein increasing activity of GlpD, an essential enzyme in glycerol catabolism, while the RNA base pairs with and represses *galE* mRNA, a gene essential for galactose metabolism. The second dual-function RNA studied in this work is Spot 42, a 109-nucleotide RNA known to base pair with and repress mRNAs encoding proteins involved in the metabolism of non-preferred carbon sources. Although Spot 42 is a well-characterized base pairing small RNA (sRNA) in *E. coli*, this work shows it also encodes a 15-amino acid protein, SpfP. SpfP was found to bind to cAMP receptor protein (CRP) and block activation of some target genes. For both AzuCR and Spot 42, the coding sequence for the small protein overlaps the base pairing region, and we have observed that translation

interferes with base pairing activity suggesting competition between the sRNA and mRNA activities. Finally, a synthetic dual-function RNA was constructed from the *Escherichia coli* sRNA MgrR and the mRNA for the small protein MgtS. Various versions of this hybrid molecule are used to probe how the organization of components is important for the proper functioning of a dual-function RNA. These three studies highlight the complexities of regulation by dual-function RNAs and provide insights into how these molecules coordinate two different activities to carry out regulatory roles in the cell.

INVESTIGATING THE COMPETITION BETWEEN COMPONENTS OF DUAL-FUNCTION RNA

By

Jordan James Masuo Aoyama

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
NIH Distinguished Investigator Gisela Storz, Chair
Associate Professor Jiqiang Ling
Professor Leslie Pick
Professor Wade C. Winkler
Associate Professor Jason D. Kahn, Dean's Representative

Acknowledgements

I would like to thank Dr. Storz for her critical input, guidance, and support; my committee for their critical feedback to improve my research; and the past and present members of the Storz Lab. Thank you, Katie, for your support and love. I could not have done any of this without you.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Many organisms employ noncoding RNA (ncRNA) to regulate diverse biological processes and respond to environmental stress. These ncRNAs vary in size and complexity to exert important regulatory roles in their respective organisms. The majority of ncRNA do not contain open reading frames (ORFs) and function through interactions with other target RNA molecules or proteins. The discovery of ncRNAs that encode functional small ORFs in addition to binding RNA targets illuminated how some RNAs break this mold (Raina, King et al. 2018). Since this discovery, only a handful of dual-function RNA have been characterized in bacteria and eukaryotes. This is perhaps due to a lack of organized searches for potential dual-function RNA throughout the genome and the inherent difficulties in studying these molecules. The reality is that these unique molecules have diverse regulatory capabilities and are likely much more prevalent than is currently recognized.

Small regulatory RNAs in bacteria

Small regulatory RNAs (sRNAs) are molecules that are transcribed from DNA but do not encode a protein and are involved many cellular processes. One of the first discovered sRNA that regulated gene expression was Spot 42 in *E. coli* (Ikemura and Dahlberg 1973). Since the initial discoveries, numerous other ncRNA have been found with a wide variety of functions. These regulatory molecules have been found to play an important role in attenuating gene expression in response to environmental stresses. For most sRNAs, base pairing with their target mRNA blocks access to the ribosome binding site (RBS) and prevents translation by preventing

initiation. For other sRNAs, base pairing with their target results in destabilization of the RNA target (Frohlich, Papenfort et al. 2013, Papenfort, Sun et al. 2013). Finally, in cases where a sRNA base pairs with the 5' untranslated region (UTR) of a target, an inhibitory secondary structure preventing translation of the mRNA can be relieved to permit ribosome binding and subsequent translation. The hexameric RNA chaperone Hfq assists in the base pairing between sRNA and mRNA and is required by most sRNAs for stability (Updegrove, Zhang et al. 2016).

Small proteins are poorly annotated and characterized

Some small proteins have been demonstrated to interact with and modulate a larger protein partner (Andrews and Rothnagel 2014, Storz, Wolf et al. 2014, Saghatelian and Couso 2015). An issue with studying small proteins is that many databases, such as GenBank, do not have widespread annotation of small ORFs. This problem is in part due to small ORFs often not encoding functional proteins. Biochemical techniques used to characterize proteins are often not optimized for small proteins (e.g., co-purification of small proteins may simply run off a gel). Despite these issues, in the past decade, small proteins that modulate cell division, enzymatic activity, transport across membranes, metabolism, and stress responses have been identified (Ramamurthi and Storz 2014, Storz, Wolf et al. 2014, Orr, Mao et al. 2020).

Regulatory RNAs may not be "non-coding"

Given the bias against characterizing and annotating small proteins there may be numerous ncRNAs that actually contain an open reading frame (ORF). Additionally, computational searches have predicted that many sRNAs in bacteria contain ORFs between 10-50 amino acids. Only a small number of sRNAs contain ORFs, and of these only a fraction have a documented

function. An RNA molecule that has RNA specific regulatory capability and also encodes a protein is called a dual-function RNA.

Definition of dual-function RNAs

Regulatory systems in eukaryotes and bacteria can have the same function while their origins and steps required for execution can be quite different. As an example, sRNA in bacteria and miRNA in eukaryotes can both regulate genes via direct base-pairing interactions, but their maturation and the steps involved in actually regulating an mRNA are very different. While only just beginning to be understood, in bacteria a dual-function RNA encodes a small protein and directly regulates mRNA via base pairing. In eukaryotes, there are many types of ncRNA including the 21-25 nt regulatory microRNA (miRNA) and the hundreds to thousands of nucleotides in length

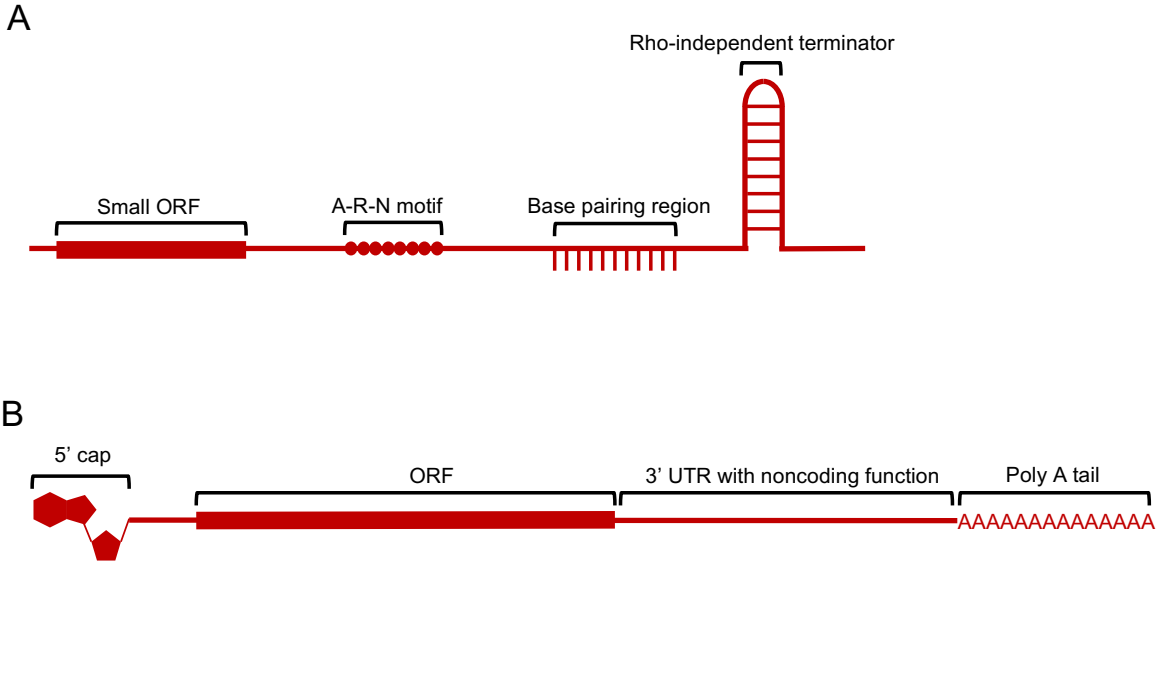


Figure 1. General features of dual-function RNA.
A Bacterial model of a dual-function RNA with general features that are important to many of these molecules. The important features indicated are the small ORF, ARN motif (where A = adenine, R = adenine or guanine, N = any nucleotide), base pairing region, and rho-independent terminator.
B Eukaryotic model of a dual-function RNA. Important features of the 5' cap, ORF, 5' untranslated region (UTR) with noncoding function, and poly-A tail are indicated.

long noncoding RNA (lncRNA). Eukaryotic dual-function RNA can have diverse functions including RNA that can serve a scaffolding function during development and also contain an ORF, or an mRNA that gains additional non-coding function upon mutation (Mayba, Gilbert et al. 2014). As is currently understood, the general features of dual-function RNA in bacteria and eukaryotes are summarized in Figure 1. Until now, only a limited number of dual-function RNAs have been identified and characterized because they tend to be overlooked due to the difficulty in distinguishing true translated ORFs from small ORFs that occur by chance.

Discovery of dual-function RNAs

The identification of most dual-function RNAs has been serendipitous. For example, human LINC00948 was thought be a lncRNA until the bioinformatic identification of a 46 amino acid ORF, myoregulin (Anderson, Anderson et al. 2015, Nelson, Makarewich et al. 2016). In *E. coli*, RNAIII was originally identified and characterized as an mRNA until a regulatory function was found for the RNA (Balaban and Novick 1995, Verdon, Girardin et al. 2009). In contrast, SgrS was originally understood to be a ncRNA until the identification of a small ORF encoding SgrT (Wadler and Vanderpool 2007). Systematic identification of potential dual-function RNAs would be a valuable and interesting addition to the field.

Roles of dual-function RNAs in bacteria

Once an RNA has been documented to have two functions, the biggest challenges are elucidating each function and determining how the functions are coordinated. As each dual-function RNA has two functional components that can each have unique regulatory targets, these molecules can control two separate pathways or regulate a single system using both components.

Dual-function RNAs affecting metabolism

SgrST is a well characterized dual-function RNA in enterobacteria (Bobrovskyy and Vanderpool 2014). This 227-nt sRNA was originally identified in a computational screen for sRNAs in *E. coli* and plays a role in the cellular response to glucose-phosphate stress (Wassarman, Repoila et al. 2001). SgrS counteracts the accumulation of sugar phosphates in the cell by reducing translation of sugar transporters and increasing dephosphorylation of sugar phosphates (Maki, Morita et al. 2010). Two of the targets of SgrS are *ptsG* and *manXYZ*, important sugar transporters in the cell. SgrS base pairs near the 5' end of *ptsG* RNA in a short region that also includes the ribosome binding site (RBS) and prevents translation of PtsG protein (Rice and Vanderpool 2011). In addition to functioning as a regulatory sRNA, SgrS also encodes the 43 amino acid small protein SgrT (Wadler and Vanderpool 2007). This small protein functions to inhibit the transport activity of the glucose permease PtsG. Superresolution microscopy demonstrated that SgrT binds the $EIIC^{Glc}$ domain of PtsG to inhibit the function of this transporter (Lloyd, Park et al. 2017). In summary, SgrS and the small protein SgrT work together to block sugar phosphate accumulation, and actively reduce intracellular glucose-6-phosphate.

Table 1: Dual-function RNA in bacteria.

Name	Organism	Protein Name	Protein Function	RNA Targets
RNAIII	*S. aureus*	*hld* (delta hemolysin)	Host cell lysis.	Inhibits *spa, coa, lytM, rot, sbi, sa1000, sa2353, sa2093,* and *ltaS*. Activates *hla, map,* and *mgrA*.
PSM-mec	*S. aureus*	PSMα	Lyse erythrocytes and neutrophils.	Inhibits and degrades *agrA* to decrease the high density agr response
SR1	*B. subtilis*	SR1P	Binds GapA to prevent *gapA* operon RNA degradation.	Inhibits *ahrC* to regulate arginine metabolism.
Pel RNA	*S. pyogenes*	Streptolysin S	Host cell lysis.	Activates *sic, nga,* and *emm*.
SgrS	*E. coli*	SgrT	Blocks activity of PtsG glucose transporter.	Inhibits *ptsG, manXYZ, asd, adiY, folE,* and *purR*. Activates *yigL*.

SgrS base pairs with target mRNAs to inhibit synthesis of sugar transporters, while SgrT inhibits the activity of translated PtsG transporter.

SgrST was the first dual-function RNA for which researchers explored the possibility of competition between the base pairing and small protein components. The base pairing region of SgrS is 15-nt downstream of the SgrT ORF making it possible that a translating ribosome could interfere with base pairing (Balasubramanian and Vanderpool 2013). Mutations to inhibit *sgrT* translation do not impede or enhance SgrS regulation of mRNA targets, but mutations that

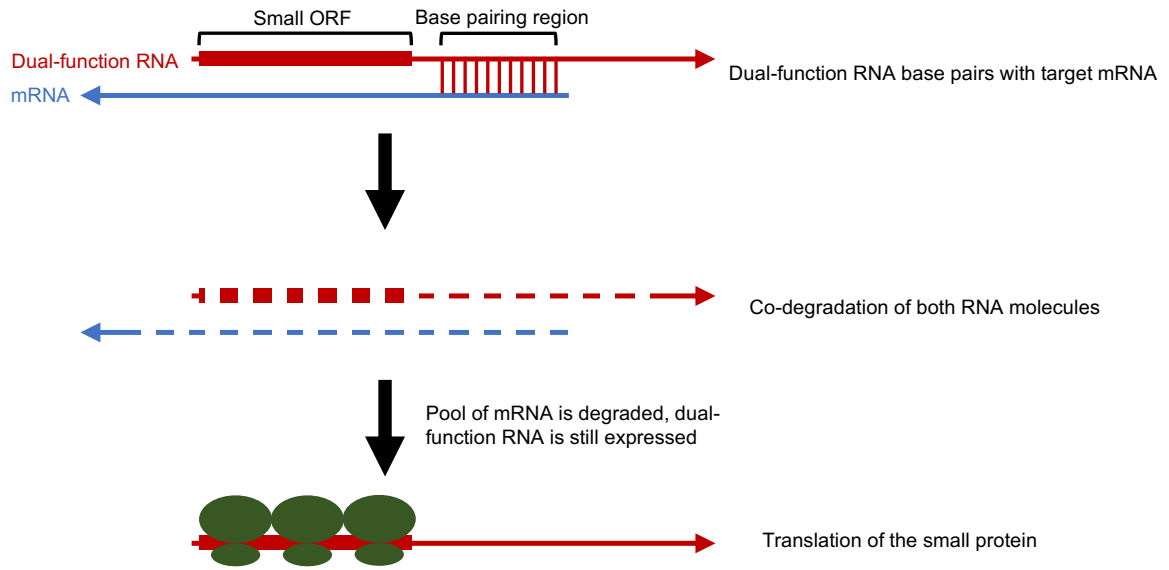


Figure 2. Stability can dictate competition between components of a dual-function RNA.
A candidate dual-function RNA base pairs and co-degrades with an mRNA target. Once the levels of the target have been sufficiently degraded the dual-function RNA is available to bind ribosomes and translate the small protein.

restrict base-pairing increase SgrT translation. Furthermore, translation of the small protein lags behind transcription of SgrS RNA by approximately 30 minutes (Balasubramanian and Vanderpool 2013). These observations suggest that the stability of SgrS RNA dictates competition between the activities. SgrS transcripts that base pair with target mRNAs are unavailable for use as an mRNA, as they are degraded. Once the pool of targeted mRNA has

been sufficiently degraded, SgrS is able to function as an mRNA (Balasubramanian and Vanderpool 2013).

SR1 is the first dual-function RNA identified in *B. subtilis*. It was discovered through a bioinformatic search of intergenic regions for sRNAs. This 205-nt RNA is expressed under gluconeogenic conditions and repressed under glycolytic conditions by CcpN, an RNA regulator involved in carbon catabolite repression (Licht, Preis et al. 2005). CcpN represses SR1 by binding to DNA and inducing a conformational change to reduce transcription. SR1 regulates arginine metabolism through base pairing interactions with *ahrC* mRNA, which encodes the activator of the *rocABC* and *rocDEF* arginine catabolism operons (Heidrich, Moll et al. 2007). The functional outcome of this base pairing is a structural change downstream of the *ahrC* start codon, which inhibits binding of the 30S ribosomal subunit thus blocking translation initiation. The decrease in *ahrC* mRNA expression leads to a decrease in *rocABC* and *rocDEF* RNA levels.

SR1 RNA was found to encode the 39 amino acid small protein SR1P after researchers mutated a potential AUG start codon and noticed that certain overexpression phenotypic effects were alleviated (Gimpel, Heidrich et al. 2010). This result was confirmed by detecting the product of a small potential ORF with a FLAG tag using immunoblot analysis. This small protein was found to directly interact with GapA, one of two glyceraldehyde-3-phosphate-dehydrogenases in *B. subtilis*. GapA binds RNase J1 and RNase Y to inhibit the degradation of *gapA* mRNA (Gimpel and Brantl 2016). SR1P binding enhances GapA binding to RNase J1, increasing RNase J1 activity. Both SR1 and SR1P are highly conserved, with 23 identified homologs in *Bacillus*, *Geobacillus*, *Anoxybacillus*, and *Brevibacillus* species (Gimpel, Preis et al. 2012). All of the identified homologs share high structural similarity to *B. subtilis* SR1 and all contain the SR1P ORF.

Dual-function RNAs affecting pathogenesis

In *S. aureus*, RNAIII is one of the primary effectors of the *agr* system, a system that is vital for pathogenesis. RNAIII was the first dual-function RNA discovered during studies to understand the function of the *agr* quorum-sensing system (Novick, Ross et al. 1993). Almost all *S. aureus* isolates from acute infections produce RNAIII. At the end of exponential growth, RNAIII RNA is expressed and begins repressing the translation of early virulence genes and increasing the production of secreted factors that are required for late infection (Novick, Ross et al. 1993). This 514-nt long dual-function RNA has fourteen stem-loop motifs and is implicated in the base pairing regulation of 12 different mRNA targets including coagulase (*coa*), fibrinogen-binding protein (SA1000), staphylococcal secretory antigen (SsaA), surface protein A (*spa*), lipoteichoic acid synthase (*ltaS*), the major cell wall autolysin (*lytM*), and immunoglobulin-binding protein (*sbi*). RNAIII indirectly controls transcription of many genes through repression of *rot* mRNA, encoding the transcription factor Rot (repressor of toxins) (Boisset, Geissmann et al. 2007).

RNAIII encodes a small ORF that is translated to produce the 26 amino acid cytotoxic peptide δ-hemolysin (*hld)*, which lyses host cells by targeting their membranes (Verdon, Girardin et al. 2009). δ-hemolysin is produced during late exponential phase but lags behind the increase in RNAIII levels by 1 hour (Balaban and Novick 1995). The ORF for δ-hemolysin is located between the 5' and 3' base pairing regions of RNAIII and may be sequestered in secondary structure that could be responsible for the lag in δ-hemolysin levels (Benito, Kolb et al. 2000). Consistent with this hypothesis, deletion of the 3' end of RNAIII prevents the delay between the increase in RNAIII levels and translation of δ-hemolysin (Balaban and Novick 1995).

Another dual-function RNA that affects pathogenesis is *Psm-mec*. The function of *Psm-mec* RNA was first uncovered when researchers were attempting to understand the genetic differences between community-acquired (CA) multi-drug resistant *S. aureus* (MRSA), which infects healthy people outside of hospitals, and hospital-associated (HA) MRSA, which infects the immunocompromised patients within a hospital. This study discovered that in HA-MRSA the staphylococcal cassette chromosome (*SCCmec*), the mobile genetic element that normally encodes *Psm-mec* and confers antibiotic resistance, was absent *Psm-mec* in CA-MRSA (Kaito, Saito et al. 2011). The 143-157-nt sRNA inhibits translation by base pairing with *agrA* mRNA, which encodes the AgrA transcription factor that upregulates the accessory gene regulator (*agr*) quorum-sensing system. *Psm-mec* also destabilizes *agrA* mRNA in a manner separate from the activity inhibiting translation (Kaito, Saito et al. 2013). The regulatory effects of *Psm-mec* have been found to be highly strain dependent (Chatterjee, Chen et al. 2011).

Psm-mec also contains a small ORF that translates the 22 amino acid phenol-soluble modulin, PSM-mec (Queck, Khan et al. 2009). This is a cytolytic toxin that contributes to *S. aureus* infection and aids in immune evasion of the host response. PSM-mec translation is upregulated by AgrA, a transcription factor that is a component of the quorum-sensing system (Chatterjee, Chen et al. 2011). Similar to *Psm-mec*, the expression of the small protein varies among MRSA strains.

Dual-function RNAs in eukaryotes

In eukaryotes, there are several different classes of ncRNAs have been defined: short (~20-30-nt), silencing RNAs (siRNAs) (Fire, Xu et al. 1998), microRNAs (miRNAs) (Gebert and MacRae 2019), PIWI-interacting RNAs (P-element induced wimpy testis or piwiRNA (Iwasaki,

Siomi et al. 2015)); small ncRNAs smaller than 200-nt; and long ncRNAs (greater than 200-nt) (Burenina, Oretskaya et al. 2017). Long non-coding RNAs (lncRNAs) is a term that broadly describes RNA transcripts that are several thousand nucleotides in length and include long intergenic ncRNAs (lincRNAs) and enhancer RNAs (eRNAs) (Lam, Li et al. 2014). Given the large number of ncRNA types in eukaryotes, it is likely that there will be more diversity in the organization and types of dual-function RNA than are present in bacteria.

Small ORFs in lncRNAs

Perhaps one of the most intriguing potential types of dual-function RNA in eukaryotes are the lncRNAs that encode small proteins. However, the regulatory function of lncRNAs has, in most cases, not been demonstrated (Mercer, Dinger et al. 2009). Using ribosome profiling and examining sequence conservation, some lncRNAs have been shown to contain small conserved ORFs (Andrews and Rothnagel 2014, Aspden, Eyre-Walker et al. 2014, Mackowiak, Zauber et al. 2015). However, despite evidence to confirm the synthesis of these small proteins their cellular relevance remains largely unexplored. An interesting example of one of the small proteins that has been well described is the 46 amino acid myoregulin protein which inhibits the Sarco/endoplasmic reticulum Ca^{2+} ATPase (SERCA) pump (Anderson, Anderson et al. 2015,

Table 2. Dual-function RNAs in eukaryotes.

Name	Organism	Protein Name	Protein Function	RNA function
Sqt	*D. rerio*	Squint	Signaling in mesendoderm patterning and specification of the neural tube.	The 3' UTR associates with scaffolding RNA binding proteins and initiation of dorsal axis specification.
Oskar	*D. melanogaster*	Oskar	Involved in late oogenesis.	The 3' UTR associates with scaffolding RNA binding proteins and regulates early oogenesis.
DMPK	*H. sapiens*	DM1 protein kinase	Necessary for maintenance of skeletal muscle structure and function.	The CUG repeat expansion in the 3' UTR forms a hairpin structure that sequesters the mRNA and MBNL protein.

Anderson, Makarewich et al. 2016, Nelson, Makarewich et al. 2016). Myoregulin forms a transmembrane alpha helix and colocalizes with the SERCA pump in the sarcoplasmic reticulum membrane to inhibit calcium reuptake (Anderson, Anderson et al. 2015). The Ca^{2+} pump has been demonstrated to be regulated by two other small proteins, phospholamban and sarcolipin, raising the possibility that some large transmembrane proteins may be the target of numerous smaller proteins that serve to alter their function (Asahi, Kurzydlowski et al. 2002, Asahi, Sugita et al. 2003, Shaikh, Sahoo et al. 2016). Given this example, lncRNAs provide an interesting pool of potential novel dual-function RNA that remain to be characterized.

Dual-function RNAs in animal development

Another type of dual-function RNA in eukaryotes are those RNAs that play a structural role in embryonic development. Early embryogenesis in many animals relies upon many maternal transcripts that are deposited in oocytes and guide the initial stages of development. These maternal RNAs are widely believed to be under tight post-transcriptional control in order to regulate the spatial and temporal expression of their protein (Bashirullah, Cooperstock et al. 2001, Tadros and Lipshitz 2005, Bettegowda and Smith 2007, Kugler and Lasko 2009). In the last decade, studies in *Drosophila* and zebrafish have suggested that some of the oocyte localized transcripts may have additional roles beyond coding for protein (Kloc, Wilk et al. 2005, Jenny, Hachet et al. 2006, Lim, Kumari et al. 2012). For example, *Oskar* (*osk*) was originally identified as a maternal-effect gene required for antero-posterior patterning in *Drosophila* embryogenesis (Lehmann and Nusslein-Volhard 1986). In early oogenesis, *osk* mRNA is transported to oocytes where it localizes to posterior poles and is translated to give the Osk protein, which is required for embryonic patterning and germ cell formation (Ephrussi, Dickinson et al. 1991). Females

with *osk* knockouts failed to lay eggs but could be complemented by nonsense mutant alleles that still expressed *osk* RNA. This led to a series of experiments that ultimately demonstrated that expression of *osk* 3'UTR alone is sufficient to rescue the eggless phenotype of *osk*-null mutants (Jenny, Hachet et al. 2006). The *osk* 3'UTR was subsequently discovered to include a motif recognized by RNA-binding proteins (RBP) that induce the formation of RBP scaffolds that are important for embryonic development (Ephrussi, Dickinson et al. 1991, Jenny, Hachet et al. 2006). The mechanisms underlying the non-coding functions of *osk* RNA are still being studied.

A second example is *Squint* (*sqt*) in zebrafish, which encodes a nodal-related signaling molecule from the transforming growth factor beta (TGFβ) superfamily with a well-studied role in mesendoderm patterning, specification of the ventral neural tube, and left-right axis development (Rebagliati, Toyama et al. 1998, Shen 2007, Schier 2009). The protein is also involved in embryonic development, playing an important part in germ layer patterning (Shen 2007). However, *sqt* RNA is localized to one or two cells by the 4-cell stage and these asymmetrically localized *sqt* transcripts play a role in dorsal axis specification independent of mRNA activity (Lim, Kumari et al. 2012). Depletion of *sqt* RNA using antisense oligonucleotides produces embryos with defective dorsal structures (Gore, Maegawa et al. 2005). These dorsal defects can be rescued by expressing the *sqt* 3'UTR. By using mutants of *sqt*, including one with a stop codon in the second codon and the other with a truncation to remove the 3' end of the RNA and the potential ORF, researchers suggest that the 3' UTR *sqt* RNA contributes to initiation of dorsal axis specification by binding and transporting unknown factors (Lim, Kumari et al. 2012). Thus, *sqt* mRNA is a potential dual-function RNA with an RNA function in embryogenesis that is separate from protein coding activity.

mRNAs in which disease-causing mutations lead to aberrant RNA function

A final potential category of dual-function RNA in eukaryotes are mRNAs where an aberrant non-coding function arises from acquired mutations. These mutant RNA molecules acquire an additional RNA specific function that undermines normal mRNA translation. The autosomal dominant inherited disease, myotonic dystrophy (DM), displays multi-system symptoms like muscle wasting, myotonia, cardiac defects, and reduced cognitive ability. The mutation associated with type 1 myotonic dystrophy is caused by a CUG repeat expansion in the 3'UTR of DM protein kinase (DMPK) that is expressed in tissues affected by DM (Mahadevan, Tsilfidis et al. 1992). Research has shown that the disease is caused by gain of function from the expanded CUG repeats. The repeats form a stable hairpin structure that causes the mutant *DMPK* mRNA to sequester in discrete foci and deplete actively translating *DMPK* mRNA (Michalowski, Miller et al. 1999). This is a scenario in which a mutation in an RNA transcript gives a normal mRNA a new function that alters localization and causes disease.

Open questions regarding dual-function RNAs

The origins of dual-function RNA and how these molecules internally regulate their separate components remain largely unanswered questions in both eukaryotes and bacteria. In bacteria, only SgrS has had the competition between its components carefully examined but no eukaryotic dual-function RNA has been studied in such depth. Furthermore, a better understanding of whether these molecules evolve *de novo*, from a progenitor RNA, or the combination of two separate RNA fragments would provide much information on what are the essential elements of a dual-function RNA.

Competition between activities of dual-function RNA

Although the study of dual-function molecules has advanced our knowledge about their targets and conservation, little is known about the interaction between their two regulatory functions. For many dual-function RNAs there is very little, if any, separation between the RNA binding site and the ORF. For some, the two components actually overlap. Furthermore, the levels of RNA and protein are sometimes independent of one another suggesting that there are conditions where the RNA functions and others where the protein is translated. Ribosomal occupancy of an mRNA would serve as an impediment for base pairing interactions. Similarly, the base pairing necessary for regulation may disrupt translation initiation. It is not difficult to imagine that the dual nature of these molecules must somehow be controlled to allow each component to function in situations where they are in competition. As of the writing of this dissertation, SgrS remains the only dual-function RNA to have the competition between components explored (Balasubramanian and Vanderpool 2013). The balance between translation and base pairing regulation within dual-function RNA permits for additional layers of regulation. The competition within a single molecule is a concept that should be addressed for every dual-function RNA that is identified and described. Fortunately, as more of these molecules are characterized our understanding of different types of competition will also expand.

Evolution of dual-function RNAs

Studying how these regulatory molecules evolve would provide important insights into whether they are generated *de novo* or develop from an existing RNA that gains additional functionality. Developing an understanding for the evolution of any RNA is greatly aided through phylogenetic analysis across species and strains. Evolution of dual-function RNAs is not well understood as

there are only a handful that have been characterized. Furthermore, most dual-function RNAs are first characterized as either a ncRNA or a small protein until the second function is uncovered. SgrS was first understood to be sRNAs until a functional ORF was discovered. However, RNAIII and Pel RNA were originally characterized as mRNAs until a base pairing function was identified for both. Does a ncRNA evolve to encode a small protein or does the small ORF develop base pairing capabilities? The identification of more dual-function RNA will facilitate the study of their evolution.

A dual-function RNA will presumably need to possess at least some of the characteristics of a regulatory ncRNA. The evolution of dual-function RNA would presumably have similar requirements with the evolution of regulatory ncRNAs. These requirements were recently reviewed and include an environmentally-sensitive promoter, a rho-independent terminator, and a relatively unstructured seed sequence region to permit base pairing (Updegrove, Shabalina et al. 2015). Each of these features levy restrictions on the sequence of the sRNA and thus the evolution of the sRNA. Interestingly, these restrictions are apparently not so onerous as to prevent significant changes in sRNA sequence (Gottesman and Storz 2011). An investigation into the conservation of dual-function RNA across bacterial species would shed much light on how these unique molecules develop. Further study of the evolution of dual-function RNA will contribute to our understanding of what features are important for their function.

Perspectives

This work seeks to add to the existing pool of characterized dual-function RNAs, explore the features that are important for their function, and further our understanding of how competition between components guides the regulatory activity of these molecules. As we have discussed,

examples of dual-function RNA are relatively rare in both bacteria and eukaryotes, but more are being discovered every year. Despite their scarcity, these molecules can have potent regulatory function. For example, *Oskar* RNA plays a critical role in early embryogenesis of *Drosophila,* and in *Staphylococcus aureus* RNAIII is the primary effector of the *agr* quorum-sensing system (Balaban and Novick 1995). This dissertation will describe the characterization of two new dual-function RNA and the creation of a synthetic dual-function RNA in order to more fully examine how competition between components impacts the regulation of these interesting RNA molecules.

The first dual-function RNA characterized in my dissertation is AzuCR, which was first identified in a bioinformatic search for novel sRNAs in *E. coli* (Chen, Lesnik et al. 2002). It was then discovered to be a potential dual-function RNA when translation of a 28 amino acid protein, AzuC, was detected (Hemm, Paul et al. 2008). AzuC levels are induced by low pH, high temperature, and hydrogen peroxide (Hemm, Paul et al. 2010). This work demonstrates that AzuC interacts with GlpD, the aerobic glycerol 3-phosphate dehydrogenase, leading to an increase in GlpD activity. *cadA* and *galE* were identified as base pairing targets of AzuCR. Interestingly, AzuC translation also disrupts the base pairing base pairing regulation of *cadA* and *galE*, another example of competition between components of a dual-function RNA.

The second dual-function RNA I will examine is Spot 42, one of the first sRNAs described in *E. coli*. Spot 42 is repressed by the transcription factor cAMP receptor protein (Crp) and regulates numerous targets involved in carbon catabolite repression (De Lay and Gottesman 2009, Beisel and Storz 2011). Spot 42 is found to encode a small ORF that translates a 15 amino acid protein, SpfP, that interacts with Crp to interfere with transcriptional activation of Crp targets. As the base pairing region and ORF overlap, the temperature dependent increase in

translation also changes the regulatory activity of the sRNA function of Spot 42. As the SpfP ORF overlaps the base pairing regions of Spot 42, we seek to understand how translation of the small protein both contributes to and interferes with the already characterized functions of this RNA.

Both AzuCR and Spot 42 are examples of dual-function RNA whose components overlap. To better study the organization and the competition between components I designed a synthetic dual-function RNA. The goal of constructing this synthetic dual-function RNA was to gain insight into the functional constraints that govern the organization and design of these unique RNA molecules. I used the small protein MgtS and the sRNA MgrR as the functional components of a synthetic dual-function RNA MgtSR. MgtS is a 31 amino acid small inner membrane protein that is required for the accumulation of MgtA, the Mg^{2+} transporter. MgrR is a 98 nt sRNA that negatively regulates *eptB* (phosphoethanolamine transferase). Various versions of this synthetic dual-function RNA were used to examine how organization of components is important for the proper function of both activities and eliminates competition between the two activities.

The mysteries behind the structure and function of dual-function RNA can only be understood through further study and identification. Searching bacterial and eukaryotic genomes for novel dual-function RNA will allow for a greater understanding of these molecules. This dissertation focuses on understanding the competition between activities of dual-function RNA. The question of when and why translation of the ORF or base pairing can be the dominant activity is a persistent and interesting question that I attempt to answer by characterizing AzuCR and Spot 42 and studying the construction of a synthetic dual-function RNA.

Chapter 2

AzuCR RNA modulates carbon metabolism as a dual-function RNA

Abstract

Bacteria have evolved small RNAs (sRNAs), to regulate numerous biological processes and stress responses. While sRNAs are generally considered to be "noncoding", a few have been found to also encode a small protein. Here we describe one such dual-function sRNA that modulates carbon utilization in *Escherichia coli*. The 164 nucleotide RNA was previously shown to encode a 28 amino acid protein (denoted AzuC). We discovered the membrane-associated AzuC protein interacts with GlpD, the aerobic glycerol 3-phosphate dehydrogenase, leading to increased GlpD activity. Overexpression of the RNA encoding AzuC leads to a growth defect in glycerol and galactose medium. The defect in galactose medium was still observed for a stop codon mutant, pointing to a potential regulatory role for the RNA. Consistent with this observation, we found that *cadA* and *galE* are repressed by base pairing with the RNA (denoted AzuCR). Interestingly, translation of AzuC interferes with the observed repression of *cadA* and *galE* by AzuCR and base pairing interferes with AzuC translation, showing that the translation and base pairing functions compete.

Introduction

Bacteria are exposed to rapidly changing environmental conditions, which include variations in carbon availability, pH, temperature and osmolarity, to name a few. To survive in these fluctuating conditions, bacterial cells utilize fast, flexible, and energy-efficient mechanisms to regulate protein amounts and activity. Changes in nutrient availability or stress detected by the cell are transduced into changes in the activation or repression of transcription, post-transcriptional changes to mRNA stability, the modulation of mRNA translation as well as the modification of protein stability or activity.

For carbon source availability in *E. coli* studied here, the sequence-specific DNA binding protein CRP is a key regulator of transcription. When the levels of the preferred carbon source glucose are low, the levels of the small molecule cAMP increase. cAMP binds and activates the highly conserved CRP transcription factor, which in turn activates genes for the uptake and utilization of alternative carbon sources in a process termed carbon catabolite repression (CCR) (reviewed in (Soberon-Chavez, Alcaraz et al. 2017)).

Small RNAs (sRNAs), which are commonly transcribed under specific environmental conditions and modulate the stability or translation of mRNAs through short base pairing interactions, are major post-transcriptional regulators in bacteria (reviewed in (Wagner and Romby 2015)). In many cases, these sRNAs require RNA chaperones such as Hfq and ProQ for their stability and for optimal base pairing with their target mRNAs (reviewed in (Updegrove, Zhang et al. 2016, Olejniczak and Storz 2017, Holmqvist, Li et al. 2018)). Several sRNAs including GlmY, GlmZ, ChiX and the CRP-repressed sRNAs Spot 42 and CyaR have been found to impact carbon metabolism in *E. coli* (reviewed in (Papenfort and Vogel 2014, Durica-Mitic, Gopel et al. 2018)).

The majority of the base-paring sRNAs are thought to not encode an open reading frame (ORF). However, a few sRNAs have been shown to be translated to give small proteins and thus are denoted "dual-function sRNAs" (reviewed in (Raina, King et al. 2018)). Computational analyses of the genomes of fourteen phylogenetically diverse bacteria predicted that a number of other sRNAs contain small ORFs (smORFs) that could encode proteins between 10-50 amino acids (Friedman, Kalkhof et al. 2017). However, translation of these smORFs has only been documented in a limited number of cases. Even fewer sRNA-encoded translation products have experimentally been demonstrated to have a function.

Small proteins of less than 50 amino acids generally have been long overlooked due to many challenges related to their annotation and biochemical detection. The few that have now been studied show that small proteins modulate diverse cellular functions ranging from morphogenesis and cell division to transport, enzymatic activities, regulatory networks, and stress responses by forming complexes with larger proteins (reviewed in (Storz, Wolf et al. 2014, Hemm, Weaver et al. 2020)).

To date, the only dual-function sRNA identified and characterized in *E. coli* is SgrS, which is involved in the regulation of carbon metabolism (Wadler and Vanderpool 2007). The SgrS sRNA was first found to protect cells against elevated levels of glucose phosphate by regulating the stability and translation of mRNAs encoding proteins involved in glucose transport and catabolism (Vanderpool and Gottesman 2004). The sRNA subsequently was shown to encode a 43-amino-acid protein, SgrT, which interacts with the glucose importer PtsG to block glucose transport and promote utilization of nonpreferred carbon sources to maintain growth during glucose-phosphate stress (Wadler and Vanderpool 2007, Lloyd, Park et al. 2017). Thus,

both the sRNA and its encoded small protein act together to repress glucose import to relieve glucose phosphate stress.

The 164-nt RNA initially denoted IS092 or IsrB (now denoted AzuCR) was first noted in a bioinformatic search to identify new sRNA genes in *E. coli* (Chen, Lesnik et al. 2002). Subsequently, this RNA was found to encode a 28 amino acid ORF (Hemm, Paul et al. 2008) (Fig 3A). Synthesis of the small protein was documented by the detection of a tagged derivative (Hemm, Paul et al. 2008) and is supported by data indicating ribosome binding to the RNA (Weaver, Mohammad et al. 2019) (Fig 3A). While the protein, denoted AzuC, is only conserved in a limited number of enteric bacteria (Fig 3B), expression of AzuC was found to be highly regulated. The levels of the tagged small protein were elevated for growth in glucose compared to glycerol due to CRP-mediated repression in the absence of glucose (Hemm, Paul et al. 2010). AzuC-SPA levels also were shown to be reduced under anaerobic conditions but induced upon exposure to low pH, high

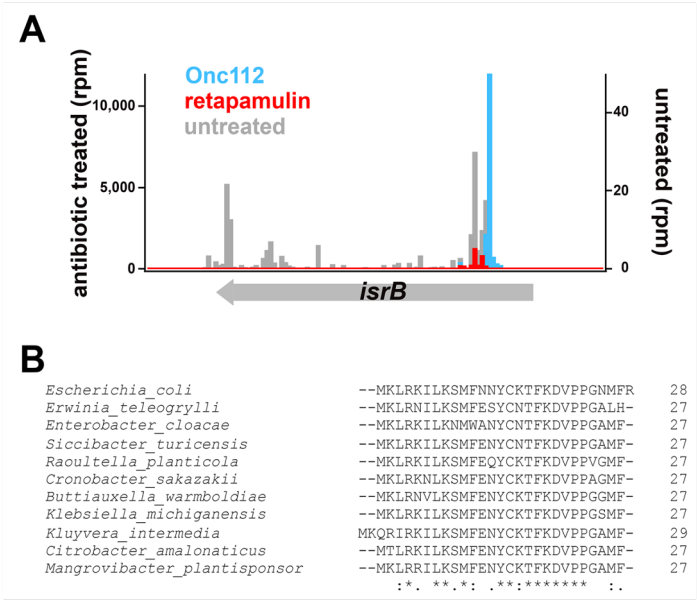


Figure 3. Ribosome binding to and conservation of *azuC*.

A) The *azuC* open reading frame lies within a region that was previously reported to encode the IsrB sRNA (Chen, Lesnik et al. 2002). Translation is detected by ribosome density on the *isrB* gene for an untreated control (gray) (Weaver, Mohammad et al. 2019), and cells treated with the translation inhibitors Onc112-treated (blue) (Weaver, Mohammad et al. 2019) or retapamulin-treated (red) (Meydan, Marks et al. 2019).

B) The AzuC amino acid sequences from *E. coli* K12 and other bacterial species aligned with ClustalW (Madeira, Park et al. 2019). "*" indicates the residues are identical in all sequences and ":" and "." respectively indicate that conserved and semi-conserved substitutions as defined by ClustalW.

temperature, and hydrogen peroxide suggesting an important role in cellular stress responses (Hemm, Paul et al. 2010).

Here we show that AzuC is associated with the membrane and binds GlpD, an essential enzyme required for glycerol catabolism, increasing GlpD activity in presence of glycerol. Additionally, we document that the transcript acts as a regulatory sRNA, denoted AzuR, repressing expression of *cadA*, a lysine decarboxylase involved in maintaining pH homeostasis, and *galE*, encoding UDP-glucose 4-epimerase, through direct base pairing. Thus, AzuCR has mRNA and sRNA activities in two different pathways, and we find that there is inherent competition between the two activities. Intriguingly, while the transcript base pairs with other mRNAs as a regulator, translation of AzuC itself is repressed by the FnrS sRNA, an sRNA that also represses GlpD synthesis.

Results

AzuC protein and mRNA levels are discordant for cells grown in glucose and low pH glycerol

Previous analysis of chromosomally-encoded AzuC, which was C-terminally tagged with the sequential peptide affinity (SPA) tag, showed that AzuC-SPA protein levels were elevated in cells grown in minimal medium with glucose compared to glycerol as well as in pH 5.5

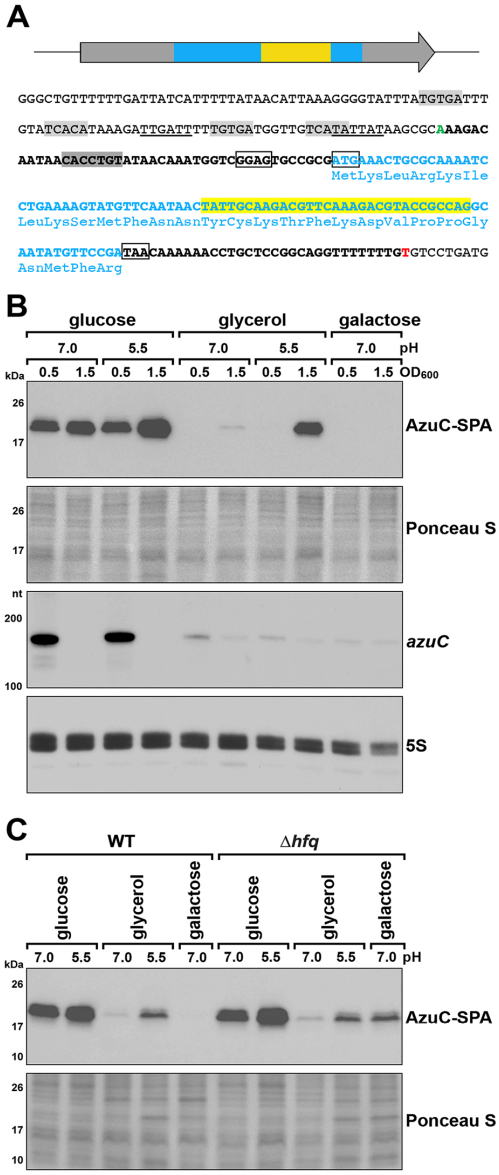


Figure 4. AzuC protein synthesis is regulated at a post-transcriptional level.

A) Diagram of the AzuC RNA and sequence of the *azuC* promoter and coding region. Boxes and text in light blue denote AzuC coding sequence and yellow box and highlighted text denote region of base pairing with target mRNAs. The AzuCR transcript is indicated in bold with the +1 site of transcription (position 1988001 of the *E. coli* K-12 genome) in green font and the 3′ end of the transcript in red font. The ribosome binding site and the start (ATG) and the stop codons (TAA) of the AzuC ORF are indicated by black boxes. Potential σ^{70} -10 and -35 sequences are underlined, the predicted Crp binding sites are highlighted in light gray (Hemm, Paul et al. 2010) and the region targeted by the FnrS sRNA is highlight in dark gray.

B) Immunoblot blot analysis of AzuC-SPA levels (top) and northern blot analysis of *azuC* mRNA levels (bottom) for cells grown in with different carbon sources. Cultures of the *azuC-SPA::kan* (GSO351) or unmarked (MG1655) strains were grown in M63 medium supplemented with glucose, glycerol, or galactose at pH 7.0 or 5.5. Samples were taken at to OD_{600} = 0.5 and 1.5. α-FLAG antibody was used to detect the SPA tag. The membrane was stained with Ponceau S stain to control for loading. The *azuC* mRNA and 5S RNA were detected by probes specific to each of these transcripts.

C) Immunoblot blot analysis of AzuC-SPA levels in *hfq*$^{+}$ (GSO351) and Δ*hfq* (GSO1007) cells grown in with different carbon sources. Strains were grown in the same media as in (B) and collected at OD_{600} = 1.5. Immunoblot blot analysis also was carried out as in (B).

compared to pH 7.5, and decreased under anerobic conditions (Hemm, Paul et al. 2010). The decreased levels in minimal glycerol medium and part of the pH-induction were attributed to Crp-mediated repression of *azuC* mRNA transcription.

To further evaluate the conditions under which AzuC-SPA and *azuC* mRNAs levels are highest, strains were cultured in M63 media supplemented with glucose or glycerol at pH 7.0 and 5.5, and in M63 galactose at pH 7.0. Cells were collected in exponential (OD_{600} = 0.5) and stationary (OD_{600} = 1.5) phase (Fig 4B). As observed previously, AzuC-SPA levels were significantly higher in glucose compared to glycerol and galactose. A notable exception was the elevated AzuC-SPA levels in cells grown to stationary phase in glycerol at pH 5.5. As expected for a Crp-regulated transcript, *azuC* mRNA levels were low for all conditions except for cells grown to exponential phase in glucose. The discordance between AzuC-SPA protein levels and *azuC* mRNA levels in glycerol at pH 5.5 raised the possibility that translation and/or RNA or protein stability is regulated and that the protein and RNA may have different roles.

Hfq is a key regulator of posttranscriptional regulation in many bacterial cells (reviewed in (Updegrove, Zhang et al. 2016, Holmqvist, Li et al. 2018)). To determine whether Hfq had any impact on AzuC, AzuC-SPA levels were compared in wild type and Δ*hfq* cells grown to stationary phase under the same conditions as above (Fig 4C). AzuC-SPA levels were elevated in the Δ*hfq* strain for the cells grown in M63 galactose. This observation is consistent with potential posttranscriptional repression by Hfq together with base-pairing sRNAs. However, before delving further into the regulation of AzuC expression, we wanted to learn more about the function of the 28 amino acid protein.

AzuC protein is localized to the membrane

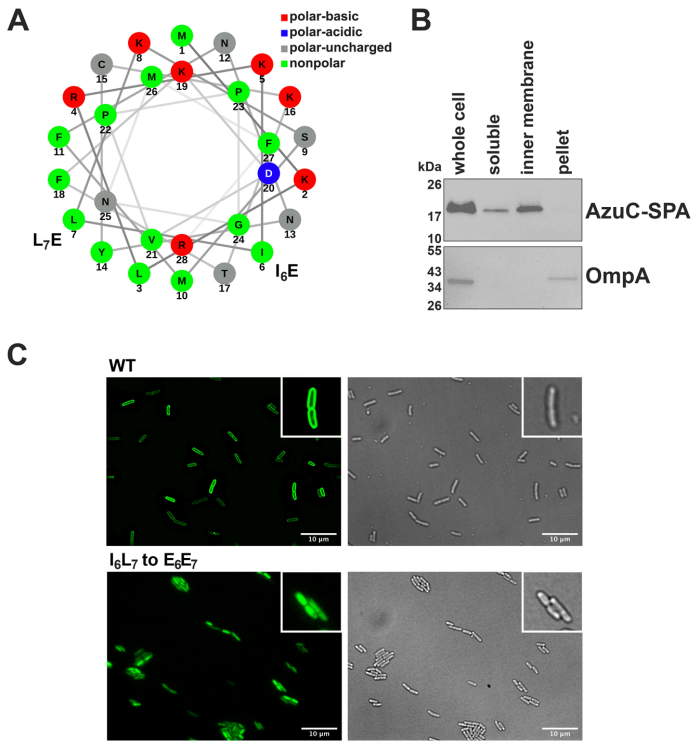


Figure 5. AzuC protein is membrane associated.

A) Helical wheel projection generated using NetWheels (Mól, Castro et al. 2019) showing amphipathic nature of AzuC. Mutations introduced in (C) are indicated.

B) Fractionation of AzuC-SPA strain. A culture expressing AzuC-SPA (GSO351) was grown in M63 medium supplemented with 0.2% glucose, and cells were fractionated into a soluble, inner membrane, and pellet fractions, which were compared to the whole cell lysate. The top panel shows AzuC-SPA as detected with α-FLAG antibody. The bottom panel shows the outer membrane OmpA control detected with α-OmpA antibody.

C) Microscopy of AzuC-GFP. AzuC-GFP (GSO1008) and AzuC-GFP IL>EE mutant (GSO1009) cells were grown in M63 glucose medium to OD_{600}=0.5 to observe membrane localization by fluorescent microscopy. Left panels are fluorescent images showing GFP labeled AzuC and the right panels are the corresponding brightfield images. Insets provide an enlargement of a few cells.

Information about the subcellular localization of proteins can give clues about possible interacting partners and functions in the cell. Secondary structural predictions suggested that AzuC has the potential to fold into an amphipathic helix (Fig 5A), indicating the protein might associate with the membrane. To test this, AzuC-SPA cells grown in M63 glucose to OD_{600} ~1.0 were lysed, and cell extracts were homogenized and fractionated into soluble, inner membrane and outer membrane fractions by sucrose cushion fractionation (Rhoads, Tai et al. 1984, Fontaine, Fuchs et al. 2011). Consistent with the secondary structure prediction, immunoblot analysis of the fractions showed that AzuC-SPA was enriched in the inner membrane fraction, while the OmpA control protein was enriched in the outer membrane fraction (Fig 5B). Similar fractionation of untagged AzuC, expressed from a plasmid and detected by α-AzuC antiserum, also showed enrichment in the membrane fraction (Fig 6A).

The localization of AzuC to the membrane was further confirmed by fluorescence microscopy imaging of chromosomally expressed AzuC C-terminally tagged by with GFP (Fig 5C). While wild type AzuC-GFP showed clear membrane localization, mutations replacing hydrophobic residues with charged residues (I_6L_7 to E_6E_7) eliminated the membrane localization. Taken together, these data support the hypothesis that AzuC is inserted into the membrane as an amphipathic helix.

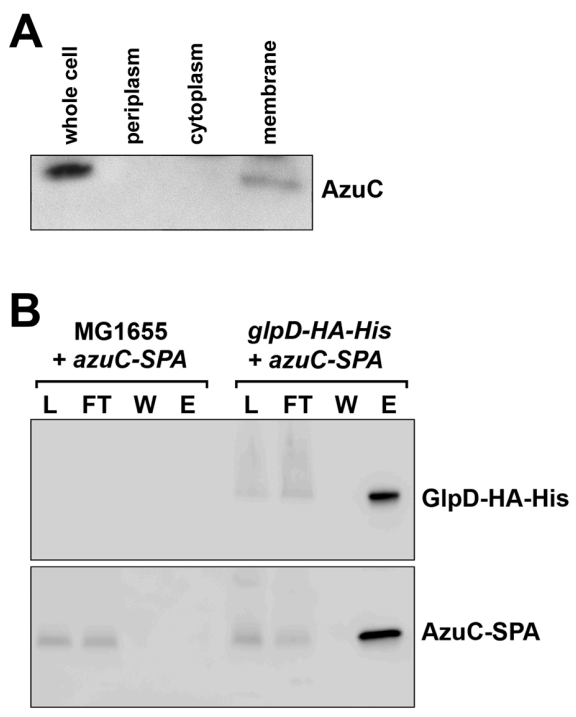


Figure 6. Fractionation showing subcellular localization of untagged AzuC and AzuC-SPA co-purification with GlpD-HA-His$_6$ compared to untagged control strain.

A) AzuC was overexpressed at low levels from the arabinose-inducible P_{BAD} promoter on the multicopy pAZ3 plasmid derivative of pBAD18 (Kawano, Reynolds et al. 2005). After induction with arabinose, cell extracts were fractionated into periplasmic, cytoplasmic, and membrane fractions. The fractions were then examined on immunoblots using polyclonal α-AzuC primary antibody followed by α-rabbit secondary antibody. AzuC expressed from the chromosome could not be detected by the polyclonal α-AzuC antibody.

B) AzuC-SPA cells grown in M63 glucose and MG1655 cells or GlpD-HA-His$_6$ cells grown in M63 glycerol, to OD_{600}~ 1 were mixed in a 1:1 ratio. The mixed cells were homogenized, cell lysates (L) were applied to anti-HA beads and the flow-through (FT) fractions were collected. The beads were washed (W), after which the bound proteins were eluted (E) and examined on immunoblots using either α-HA antibodies to detect GlpD-HA-His$_6$ (top panel) or α-FLAG antibodies to detect AzuC (bottom panel).

AzuC protein interacts with GlpD

To further investigate the AzuC role in the cell, we carried out co-purification assays to identify interacting proteins. Cells expressing chromosomally encoded AzuC-SPA or previously characterized AcrZ-SPA (Hobbs, Yin et al. 2012) as a control were grown in M63 glucose medium. Cell lysates prepared from exponentially growing cells were applied to calmodulin

beads, and the eluants from each column were separated by SDS-PAGE (Fig 7A). Unique bands from each of the elutions were sent for mass spectrometric analysis for identification. In two independent experiments, a prominent band of ~60 kDa observed only for the AzuC-SPA cells was identified as the aerobic glycerol 3-phosphate dehydrogenase (GlpD), which catalyzes the oxidation of glycerol 3-phosphate. The most prominent band in the AcrZ sample was AcrB, a known interactor (Hobbs, Yin et al. 2012).

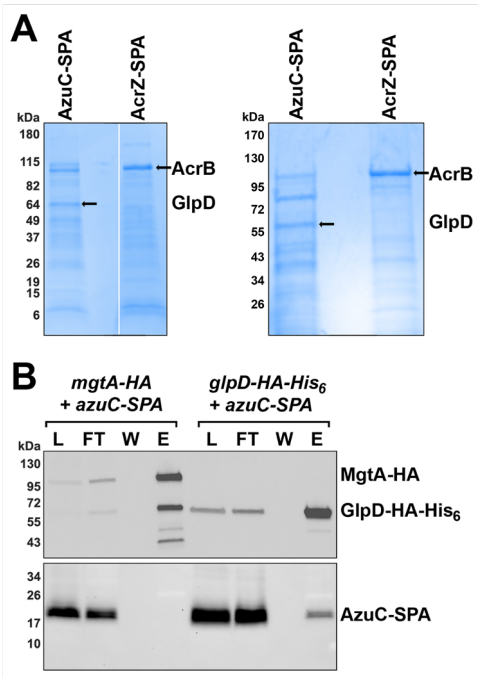


Figure 7. AzuC copurifies with GlpD.

A) GlpD co-purifies with AzuC-SPA. Cells expressing AzuC-SPA (GSO351) or AcrZ-SPA (GSO350) from the chromosome were grown in M63 glucose medium to OD_{600}~ 1 or in LB to OD_{600}~ 0.6, respectively. The cell lysates were split and passed over calmodulin beads. Eluants from each column were subjected to SDS/PAGE followed by Coomassie blue staining. The bands enriched in the eluant from the calmodulin beads and indicated by the arrows were excised from the gel and identified by mass spectrometry.

B) AzuC-SPA co-purifies with GlpD-HA-His_6. Cells expressing either AzuC-SPA (GSO351) or GlpD-HA-His_6 (GSO1011) from the chromosome were grown in M63 glucose or M63 glycerol media, respectively, to OD_{600}~ 1 and mixed in a 1:1 ratio. As a control, cells expressing MgtA-HA (GSO785) grown in N medium supplemented without added $MgSO_4$ to OD_{600}~ 0.5, were mixed with the AzuC-SPA (GSO351) cells in the same ratio. The mixed cells were homogenized, cell lysates (L) were applied to anti-HA beads and the flow-through (FT) fractions were collected. The beads were washed (W), after which the bound proteins were eluted (E) and examined on immunoblots using either α-HA antibodies to detect MgtA-HA or GlpD-HA-His_6 (top panel) or α-FLAG antibodies to detect AzuC (bottom panel).

We tested the interaction between AzuC and GlpD, by assessing reciprocal co-purification of AzuC-SPA with GlpD-HA-His_6. Cells with chromosomally encoded AzuC-SPA, grown to exponential phase in M63 glucose medium, were mixed with cells with chromosomally encoded GlpD-HA-His grown to exponential phase in M63 glycerol medium, a condition where GlpD is known to be expressed. The mixed cells were lysed and incubated with dodecyl β-D-maltoside (DDM) to facilitate mixing of the membrane fractions. The mixed lysate was then applied to α-HA beads, washed and eluted. The eluates were analyzed for the respective

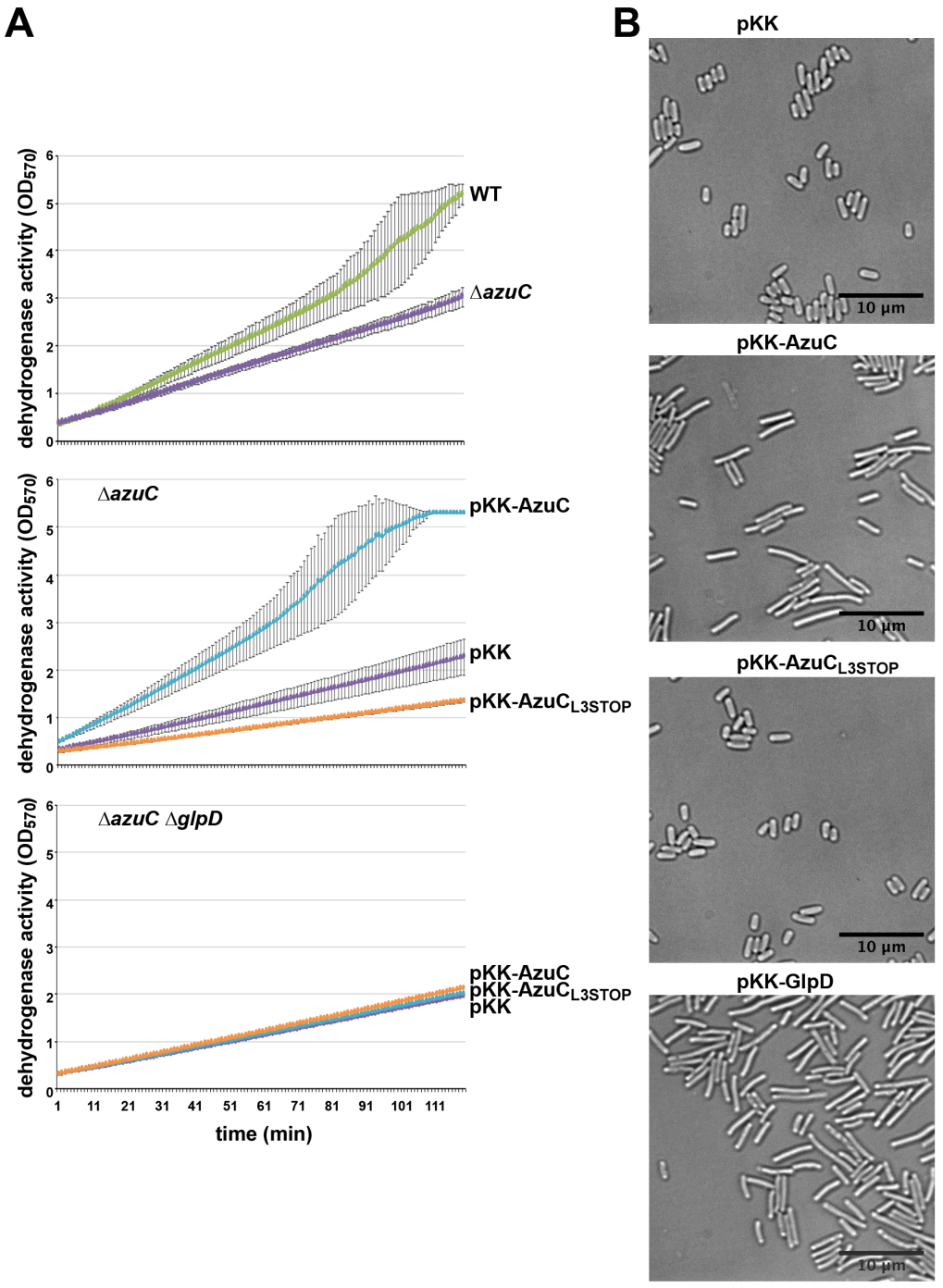


Figure 8. AzuC increases GlpD activity and affects cell shape.

A) Effect of AzuC overexpression on GlpD activity. WT or Δ*azuC::kan* cells (top panel), Δ*azuC::kan* (GSO193) (middle panel) or Δ*azuC* Δ*glpD::kan* (GSO1015) (bottom panel) cells transformed with pKK, pKK-AzuC, and pKK-AzuC$_{L3STOP}$ were grown in M63 glucose medium to OD$_{600}$~1.0. Cells were washed and resuspended in M63 glycerol medium, pH 5.5 for 3 h prior to measurement of dehydrogenase activity based on the reduction of MTT to formazan.

B) Effects of AzuC overexpression on *E. coli* cell morphology. Δ*azuC::kan* (GSO193) transformed with pKK, pKK-AzuC, pKK-AzuC$_{L3STOP}$, or pKK-GlpD were grown in M63 glucose medium to OD$_{600}$~1.0. Cells were washed and resuspended in M63 glycerol medium, pH 5.5 for 3 h prior to microscopy.

tagged proteins by carrying out immunoblot analysis by using α-FLAG and α-HA antibodies (Fig 7B). As controls, similar purifications were carried out by mixing the AzuC-SPA cells with cells lacking tagged proteins grown in M63 glycerol medium (Fig 6B) or cells expressing chromosomally encoded MgtA-HA grown in N medium without added $MgSO_4$ to induce MgtA expression (Fig 7B). Consistent with the first purification, AzuC-SPA co-purified with GlpD-HA-His_6 and not with MgtA-HA, supporting the conclusion that GlpD is an interacting partner of AzuC. As we predict for AzuC, GlpD has been reported to be a peripheral membrane protein which associates with the membrane through an amphipathic alpha helix (Walz, Demel et al. 2002).

AzuC protein increases GlpD activity

Binding of AzuC to GlpD could potentially impact the stability, localization or activity of the enzyme as has been found for other small proteins (Hobbs, Yin et al. 2012, Wang, Yin et al. 2017). To distinguish among these possibilities, we first examined the levels of chromosomally encoded GlpD-HA-His_6 in cells carrying pKK, pKK-AzuC or pKK-$AzuC_{L3STOP}$. In the latter two plasmids, the wild type or mutant (harboring a stop codon mutation of the third codon) *azuC* ORF was cloned downstream of the heterologous P_{tac} promoter and ribosome binding site on pKK. Cells were grown in M63 glucose medium to OD_{600}~1.0 and then transitioned to glycerol (pH 5.5) for 3 h given that chromosomally expressed AzuC-SPA levels are elevated under these conditions (Fig 4B). The GlpD-HA-His_6 levels were similar for all three strains grown under these conditions (Fig 9A).

To test whether AzuC affects GlpD activity, we employed a dehydrogenase activity assay. In this assay, cellular dehydrogenase activity in cell extracts is measured by the reduction

of 2-(4,5-dimethyl-2-thiazolyl)-3,5-diphenyl-2*H*-tetrazolium bromide (MTT) to formazan, which accumulates and is detected at OD_{570} (Wegener, Vogtmann et al. 2016). Dehydrogenase activity was found to be almost 2-fold lower in the absence of AzuC when extracts were made from a

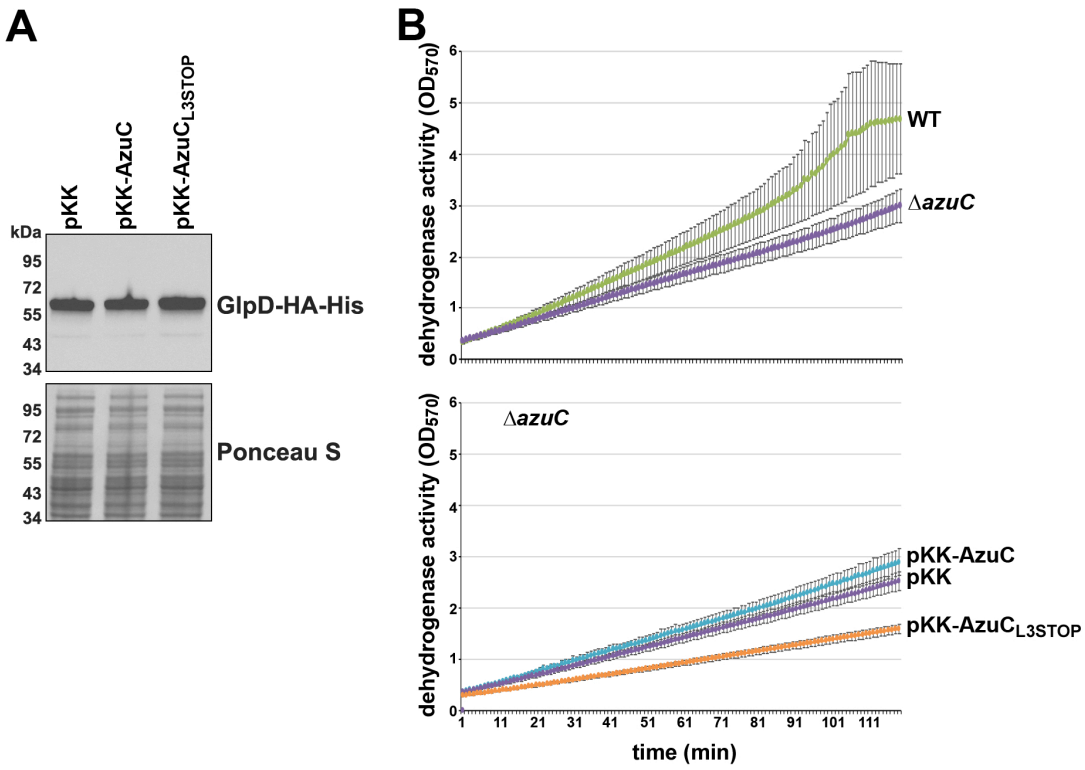


Figure 9. Effect of AzuC overexpression on GlpD levels at pH 5.5 and on GlpD activity at pH 7.0.
A) Δ*azuC* cells expressing GlpD-HA-His$_6$ (GSO1013) were transformed with pKK, pKK-AzuC, and pKK-AzuC$_{L3STOP}$, grown in M63 glycerol medium to OD_{600}~0.5 and examined on immunoblots using α-HA antibodies to detect GlpD-HA-His$_6$. The membrane was stained with Ponceau S stain to control for loading.
B) MG1655 or Δ*azuC* (GSO193) cells (top panel) or Δ*azuC* (GSO193) cells transformed with pKK, pKK-AzuC, or pKK-AzuC$_{L3STOP}$ (bottom panel) were grown in M63 glucose medium to OD_{600}~0.5 and then washed and resuspended in M63 glycerol medium at pH 7.0. Cells were allowed to grow for an additional 3 h at which point the dehydrogenase activity assay was performed.

WT and a Δ*azuC* strain grown in M63 glucose medium and shifted to glycerol (pH 5.5) for 3 h. In contrast, overexpression of wild type AzuC, but not AzuC$_{L3STOP}$, led to an increase in the dehydrogenase activity for the extracts (Fig 8A, middle panel). To further verify that this increase was GlpD dependent, the assay was repeated in a Δ*azuC* Δ*glpD* double mutant. The double mutant did not show the increase in dehydrogenase activity upon AzuC overexpression

(Fig 8A, bottom panel), indicating that the interaction of the small protein AzuC with GlpD increases the dehydrogenase activity of the larger protein. Interestingly, the pKK-AzuC$_{L3STOP}$ plasmid gave rise to a GlpD-dependent decrease in dehydrogenase activity. We think this may be due to regulatory activity of the RNA (see below). The dehydrogenase assay also was carried out with these strains shifted to M63 glycerol, pH 7, where we observed similar, albeit somewhat smaller, effects of Δ*azuC* and AzuC overexpression on activity (Fig 9B).

AzuC overexpression causes an increase in cell length

The substrate for GlpD, glycerol-3-phosphate, is a precursor for phospholipid biosynthesis. Thus, we wondered whether increasing the activity of GlpD by AzuC might bias the flow of glycerol-3-phosphate towards glycerol metabolism rather than phospholipid biosynthesis, which could impact cell morphology. To assess this, we carried out live-cell phase contrast microscopy of cells carrying the pKK vector control, pKK-AzuC or pKK-AzuC$_{L3STOP}$ (Fig 8B). We observed AzuC overexpressing cells, but not those carrying the vector or pKK-AzuC$_{L3STOP}$, had an elongated morphology. The elongated morphology was similar to the morphology observed for cells upon GlpD overexpression (Fig 8B) as well as in cells lacking phosphatidylethanolamine, which comprises ~75% of the membrane phospholipid (Rowlett, Mallampalli et al. 2017).

AzuC and GlpD protein levels are repressed by the FnrS small RNA

We previously reported that AzuC levels are higher under aerobic compared to anaerobic conditions (Hemm, Paul et al. 2010). Similarly, GlpD is required under aerobic conditions and is downregulated during anaerobic growth while a second glycerol dehydrogenase, GlpABC, is required under anaerobic conditions. Interestingly, interactions between the anaerobic-induced sRNA FnrS and both the *azuC* and *glpD* mRNAs were found in genome-wide assays of RNA-RNA interactions on the Hfq chaperone (Melamed, Peer et al. 2016, Melamed, Adams et al. 2020). We also could predict base pairing between the 5′ end of FnrS and *azuC* as well as *glpD* (Fig 10A and 10C). These observations suggested possible FnrS-mediated repression of AzuC and GlpD synthesis. Consistent with this hypothesis, we observed higher AzuC-SPA levels in a Δ*fnrS* strain (Fig 11A) and lower AzuC-SPA levels upon overexpression of WT FnrS and previously generated FnrS-I and FnrS-II mutants (Durand and Storz 2010) but not FnrS-III for which base pairing is predicted to be disrupted (Fig 10B). We similarly observed that WT

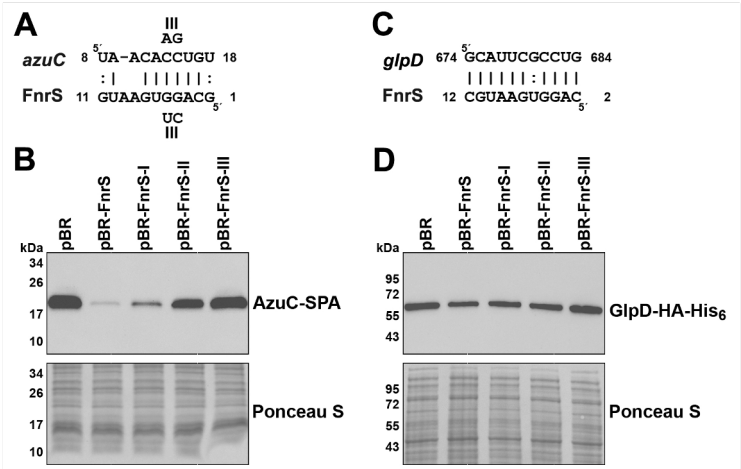


Figure 10. FnrS sRNA represses synthesis of both AzuC and GlpD.

A) Predicted base pairing between FnrS and *azuC*. The coordinates for both are relative to the +1 of the transcript.

B) Effect of FnrS overexpression on AzuC-SPA levels. Cultures of the *azuC-SPA::kan* (GSO351) strain carrying pBR, pBR-FnrS, pBR-FnrS-I, pBR-FnrS-II, or pBR-FnrS-III were grown in LB with 1 mM IPTG to OD_{600}~ 0.5. α-FLAG antibodies were used to detect the SPA tag.

C) Predicted base pairing between FnrS and *glpD*. The predicted region of pairing in *glpD* is within the coding sequence. The coordinates for FnrS are relative to the +1 of the transcript, while the coordinates for *glpD* are relative to the first nucleotide of the start codon.

D) Effect of FnrS overexpression on GlpD-HA-His_6 levels. Cultures of the *glpD-HA-His*$_6$ (GSO1011) strain carrying pBR, pBR-FnrS, pBR-FnrS-I, pBR-FnrS-II, or pBR-FnrS-III were grown in LB with 1 mM IPTG to OD_{600}~ 0.5. α-HA antibodies were used to detect GlpD-HA-His_6.

For (B) and (D), the membrane was stained with Ponceau S stain to control for loading.

FnrS, FnrS-I and FnrS-II, but not FnrS-III repressed an *azuC-lacZ* translational fusion expressed from the heterologous P_{BAD} promoter (Fig 11B). Repression was restored for the FnrS-III mutant but not WT FnrS, FnrS-I, and FnrS-II by compensatory mutations in the *azuC-lacZ*-III mutant fusion demonstrating direct base-pairing between FnrS and the *azuC* mRNA. We also observed

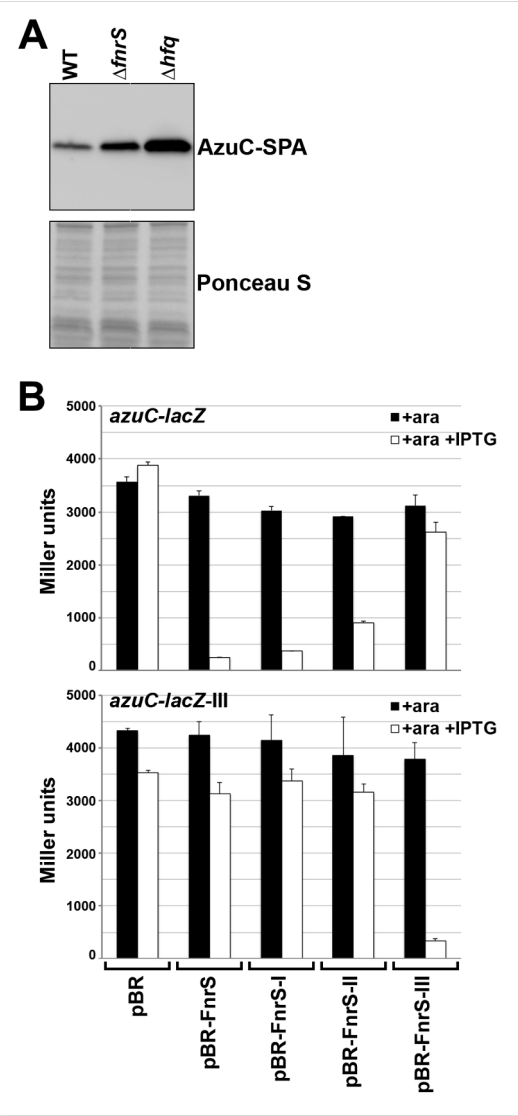


Figure 11. FnrS sRNA represses AzuC expression.

A) AzuC-SPA levels in strains lacking Hfq or FnrS. AzuC-SPA levels in WT (GSO351), Δ*hfq* (GSO1007) or Δ*fnrS* (GSO1023) strains grown in LB to OD_{600}~ 0.5. α-FLAG antibody was used to detect the SPA tag.

B) β-galactosidase activity was assayed for P_{BAD}-5′-UTR_{azuC}-*lacZ* (GSO1024) or P_{BAD}-5′-UTR_{azuC}-*lacZ* -III (GSO1025) cells carrying pBR, pBR-FnrS or pBR-FnrS mutants (pBR-FnrS I, II or III). Cells were grown to OD_{600}~0.4-0.5 and treated with either just arabinose (0.2%) (black bars) or arabinose (0.2%) and IPTG (1 mM) (white bars), and cells were grown another 40 min. The average of three independent trials is shown, and the error bars represent one SD.

slightly lower levels of the GlpD-HA-His_6 protein upon overexpression of WT FnrS, FnrS-I and FnrS-II, but not FnrS-III (Fig 10D). Together these results indicate that the 5′ end of FnrS base pairs with the *azuC* and *glpD* mRNAs to repress synthesis of AzuC and GlpD.

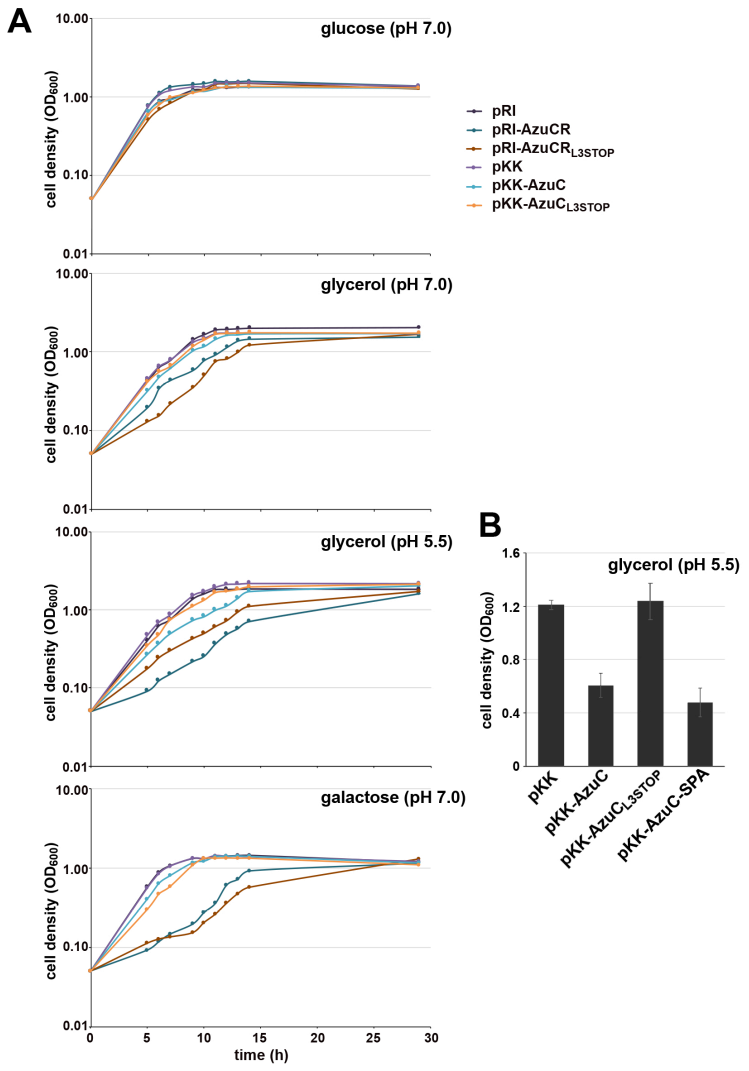


Figure 12. Growth curves for AzuC and AzuCR overexpression.

A) Δ*azuC::kan* strain (GSO193) transformed with pRI, pRI-AzuCR, pRI-AzuCR$_{L3STOP}$, pKK, pKK-AzuC, or pKK-AzuC$_{L3STOP}$ was grown in M63 medium with different carbon sources: glucose (pH 7.0), glycerol (pH 7.0 and 5.5), and galactose (pH 7.0) and growth was tracked by OD_{600} over 30 h.

B) Growth of the Δ*azuC::kan* strain (GSO193) transformed with pKK, pKK-AzuC, pKK-AzuC$_{L3STOP}$, or pKK-AzuC-SPA in M63 medium with glycerol pH 5.5 was measured at 16 h after dilution by OD_{600}.

AzuCR overexpression reduces growth in glycerol and galactose

Given the AzuC effect on GlpD together with the different AzuC levels for cells grown in the presence of different carbon sources, we examined the consequences of AzuC overexpression from the pKK vector for growth in glucose and glycerol at pH 7.0 and 5.5 and galactose at pH 7.0 (Fig 13 and Fig 12A). Consistent with the larger effect of AzuC on GlpD activity in M63 glycerol pH 5.5 compared to pH 7.0, we observed a significant growth defect for cells grown under these conditions for 16 h but not for the pKK vector control and pKK-AzuC$_{L3STOP}$ strain (Fig 13). A similar phenotype was observed for overexpression of AzuC-SPA indicating that the

tagged derivative of AzuC is functional (Fig 12B). Growth in minimal medium with either glucose or galactose was not significantly changed by the pKK-AzuC plasmid.

We also examined the effect of overexpressing the full-length *azuC* mRNA (pRI-AzuCR) without or with the L3STOP mutation (pRI-AzuCR$_{L3STOP}$) (Fig 13). Interestingly, we observed different effects on growth for these plasmids. While growth in minimal glucose was not affected, pRI-AzuCR led to a growth defect in glycerol pH 7.0 and even more so in pH 5.5 as well as in galactose pH 7.0. Contrary to the detrimental effect of the L3STOP mutation when only the *azuC* coding sequence was included, the L3STOP mutation in the full-length transcript still blocked growth and, in glycerol pH 7.0, actually exacerbated the growth defect. This observation suggested the transcript could have a second role as a regulatory RNA, which we have denoted AzuR.

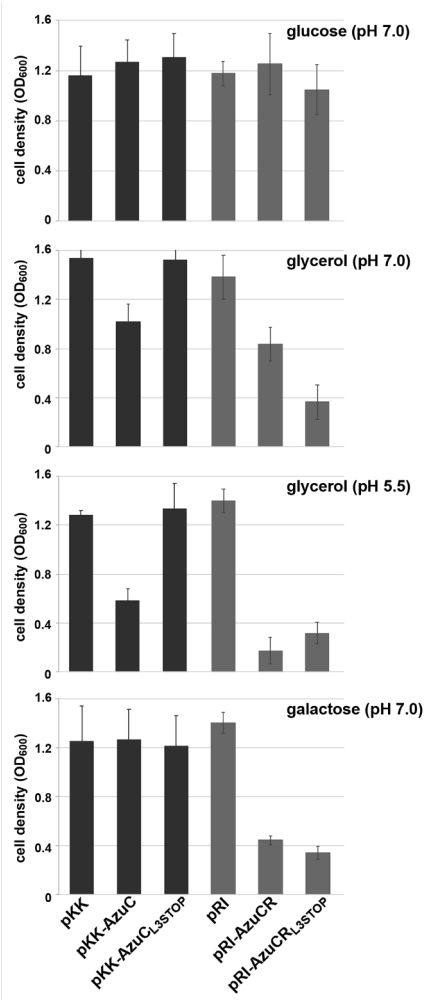


Figure 13. AzuC and AzuR overexpression leads to different growth phenotypes in different carbon sources.
Growth of the Δ*azuC::kan* strain (GSO193) transformed with pKK, pKK-AzuC, pKK-AzuC$_{L3STOP}$, pRI, pRI-AzuCR, or pRI-AzuCR$_{L3STOP}$ in M63 medium with different carbon sources, glucose (pH 7.0), glycerol (pH 7.0 and 5.5), and galactose (pH 7.0), was measured 16 h after dilution by OD$_{600}$.

AzuR functions as an sRNA to repress *cadA* and *galE*

Based on our findings that the AzuCR transcript could have a second role as an sRNA, we investigated its potential as a base-pairing sRNA

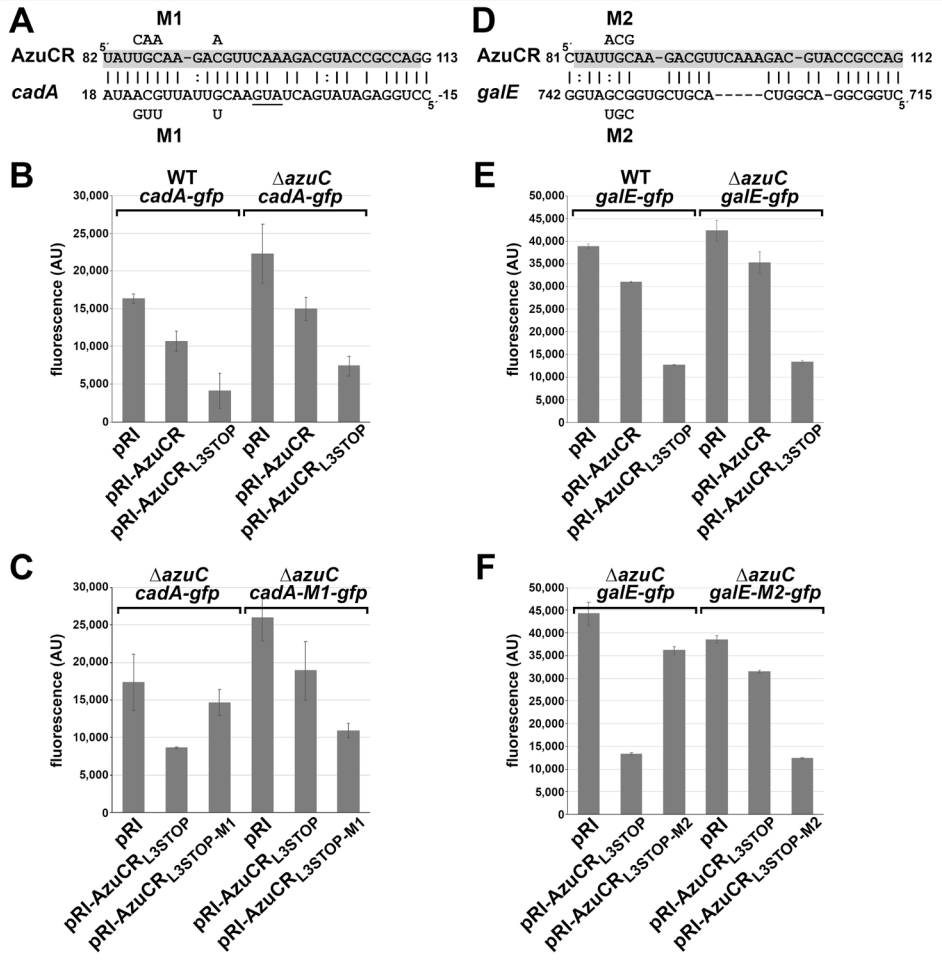


Figure 14. AzuR represses *cadA* and *galE* expression.

A) AzuCR-*cadA* base pairing predicted by TargetRNA2 (Kery, Feldman et al. 2014). Mutations introduced into AzuCR and *cadA* are indicated. The coordinates for AzuCR are relative to the +1 of the transcript, while the coordinates for *cadA* are relative to the first nucleotide of the start codon.

B) Effect of AzuCR and AzuCR$_{L3STOP}$ overexpression on a *cadA-gfp* fusion in wild type and Δ*azuC* backgrounds. WT and Δ*azuC::kan* (GSO193) cells were co-transformed with a reporter plasmid expressing a *cadA-gfp* translational fusion and either the empty pRI vector, AzuCR or AzuCR$_{L3STOP}$.

C) Test of AzuCR-*cadA* base pairing. Δ*azuC::kan* (GSO193) cells were cotransformed with the WT *cadA-gfp* translational fusion reporter plasmid or a M1 derivative with mutations in the predicted region of base pairing along with the empty pRI vector, AzuCR$_{L3STOP}$, or mutant AzuCR$_{L3STOP-M1}$. The mutations in the *cadA-gfp* translational fusion and AzuCR$_{L3STOP}$ are indicated (A).

D) AzuCR-*galE* base pairing predicted by IntaRNA (Mann, Wright et al. 2017). Mutations introduced into AzuCR and *galE* are indicated. The coordinates for AzuCR are relative to the +1 of the transcript, while the coordinates for *galE* are relative to the first nucleotide of the start codon.

E) Effect of AzuCR and AzuCR$_{L3STOP}$ overexpression on a *galE-gfp* fusion in wild type and Δ*azuC* backgrounds. WT and Δ*azuC::kan* (GSO193) cells were co-transformed with a reporter plasmid expressing a *galE-gfp* translational fusion and either the empty pRI vector, AzuCR or AzuCR$_{L3STOP}$.

F) Test of AzuCR-*galE* base pairing. Δ*azuC::kan* (GSO193) cells were cotransformed with the WT *galE-gfp* translational fusion reporter plasmid or a M2 derivative with mutations in the predicted region of base pairing along with the empty pRI vector, AzuCR$_{L3STOP}$, or mutant AzuCR$_{L3STOP-M2}$. The mutations in the *galE-gfp* translational fusion and AzuCR$_{L3STOP}$ are indicated in (D).

For (B), (C), (E) and (F), cells were grown in LB for 3 h before measuring the fluorescence corresponding to GFP expression. The the average of three independent trials is shown, and the error bars represent one SD.

by searching for possible base pairing targets using TargetRNA2 (Kery, Feldman et al. 2014) and IntaRNA (Fallmann, Will et al. 2017) prediction programs. Given the reduced growth associated with AzuCR overexpression in cells grown in galactose and low pH, we focused on potential targets encoding activities that might be important under these conditions. One predicted target with extensive potential base pairing was *cadA*, encoding lysine decarboxylase (Fig 14A). Consistent with AzuR-mediated regulation of *cadA*, we observe decreased expression of a *cadA-gfp* fusion upon AzuCR overexpression (Fig 14B). Interestingly, the derivative with a stop codon at the third codon of AzuC ($AzuCR_{L3STOP}$) had a stronger repressive effect than wild type AzuCR. Additionally, there were higher overall levels of *cadA-gfp* expression in the Δ*azuC* strain compared to the WT strain, suggesting that chromosomally encoded AzuC contributes to the repression. Consistent with the base pairing predicted in Fig 14A, the M1 mutations in $AzuCR_{L3STOP}$ reduced *cadA-gfp* repression, while regulation was restored when compensatory mutations were introduced in the *cadA-gfp* construct (Fig 14C). Another predicted target for base pairing with AzuR was *galE* (Fig 14D), the first gene in the *galETKM* galactose operon. The $AzuCR_{L3STOP}$ derivative also repressed a *galE-gfp* fusion in both the WT and Δ*azuC* backgrounds, with partial repression by AzuCR (Fig 14E). Again, there is direct base pairing between AzuCR and *galE*, as the M2 mutations in $AzuCR_{L3STOP}$ or *galE* alone reduced AzuCR-mediated *galE-gfp* repression, while repression was restored when both compensatory mutations were present (Fig 14F).

AzuCR RNA association with Hfq and ProQ is not required for *cadA* repression

Consistent with the observation that Δ*hfq* impacts AzuC protein levels (Fig 4B), we found that the AzuCR mRNA co-immunoprecipitates with Hfq (Fig 15A). Another RNA chaperone that has

been found to bind to sRNA-mRNA pairs in *E. coli* and impacts the stability of some RNAs is ProQ (Melamed, Adams et al. 2020). As with Hfq, AzuCR co-purifies with ProQ (Fig 15A). However, in contrast to the increased AzuCR RNA levels in the Δ*hfq* background, AzuCR RNA levels were decreased in the Δ*proQ* background. We wondered whether AzuCR functions as an sRNA repressor were mediated by Hfq or ProQ and examined repression of the *cadA-gfp* in Δ*hfq* and Δ*proQ* single as well as Δ*hfq* Δ*proQ* double mutant backgrounds. GFP activity levels overall were lower when Hfq was absent, but we observed *cadA-gfp* repression by $AzuCR_{L3STOP}$ overexpression in all backgrounds (Fig 15B). These observations indicate that although both Hfq and ProQ bind to AzuCR, the RNA chaperones are not required for the repression of *cadA* when $AzuCR_{L3STOP}$ is overexpressed, possibly due to the long region of potential base pairing.

AzuC translation and AzuR base pairing activity interfere

The region of base pairing between AzuR and *cadA* and *galE* (89-107-nt relative to the transcription start) overlaps the *azuC* coding sequence (40-126 nt relative to the transcription start) raising the question of whether the mRNA and base pairing activities of the AzuCR RNA interfere which each other. This supposition that translation interferes with base pairing is supported by the observations that $AzuCR_{L3STOP}$ was more effective at repressing the *cadA-gfp* and *galE-gfp* fusions than AzuCR (Fig 14B and 14E). To determine if base pairing activity also interferes with translation, we examined the levels of chromosomally encoded AzuC-SPA upon overexpression of the base pairing regions of *cadA* and *galE* along with control regions of these genes not predicted to base pair with AzuCR. Interestingly, no repression was observed for cells grown in M63 glucose. In contrast, the base pairing fragments, but not the control fragments, led to decreased AzuC-SPA levels for cells grown in M63 with galactose (Fig 15C). These

observations suggest that base pairing can interfere with translation, particularly when AzuC protein levels overall are low as is the case for cells grown in M63 galactose.

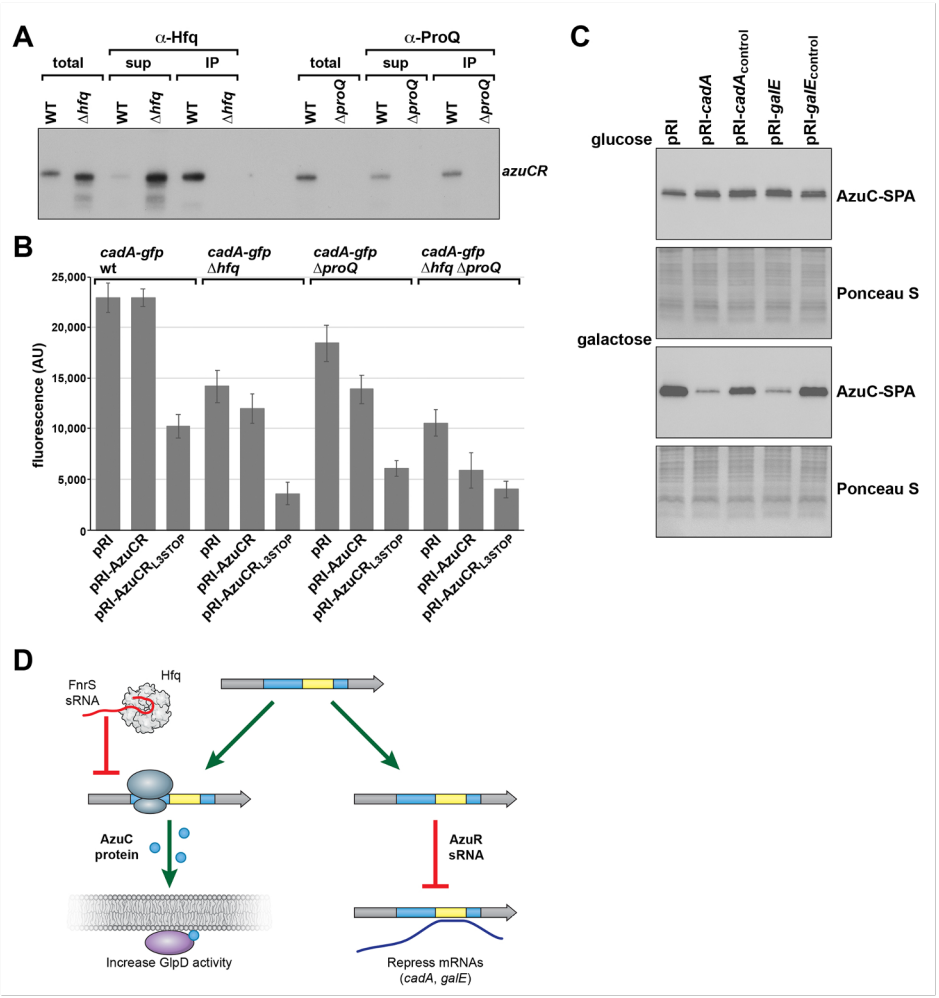


Figure 15. AzuCR mRNA and base pairing activities are differentially affected by Hfq and ProQ.

A) AzuCR co-immunoprecipitation with Hfq and ProQ. Co-IP was carried out on cellular extracts from MG1655 (GSO982), Δ*hfq-cat::sacB* (GSO954) and Δ*proQ::kan* (GSO956) cells using α-Hfq or α-ProQ antiserum. Total and RNA chaperone-bound RNA was extracted and subjected to Northern analysis using a primer specific for AzuCR.

B) Effect of Δ*hfq::kan* (GSO955), Δ*proQ::kan* (GSO956) and Δ*hfq* Δ*proQ::kan* (GSO959) double mutant on AzuCR repression of *cadA-gfp*. *cadA-gfp* expression from pXG10-SF in the presence of AzuCR or AzuCR$_{L3STOP}$ in WT, Δ*hfq* or Δ*proQ* backgrounds.The average of three independent trials is shown, and the error bars represent one SD.

C) Effect of *cadA*$_{base\ pairing}$, *cadA*$_{control}$, *galE*$_{base\ pairing}$ and *galE*$_{control}$ on AzuC-SPA levels in cells (GSO351) transformed with the respective overexpression plasmid and grown in M63 medium supplemented with glucose or galactose. Samples were taken at OD$_{600}$ ~ 0.5 and α-FLAG antibody was used to detect the SPA tag. The membranes stained with Ponceau S stain serves as a loading control.

D) Model for the different functions of the AzuCR RNA. For growth in M63 glycerol, pH 5.5, the RNA can be translated to give the 28 amino acid amphipathic AzuC protein, which activates the activity of GlpD glycerol dehydrogenase. Under anaerobic conditions this translation and the translation of GlpD is blocked by the FnrS sRNA. The RNA can also act as the AzuR base pairing sRNA to repress synthesis of CadA and GalE.

Discussion

Successful adaptation to varying environmental conditions requires regulation that can rapidly change metabolism. Along with transcription factors, sRNAs and small proteins are emerging as important regulators. While base-pairing sRNAs generally are not thought to be translated, a few have been reported to encode small proteins. Even fewer of these dual-function sRNAs have been characterized. Here we report that the 164-nt RNA previously reported to encode a 28 amino acid small protein (AzuC) (Hemm, Paul et al. 2010) also functions as a regulatory RNA (AzuR). The AzuC protein binds and increases activity of GlpD, an enzyme at the junction of respiration, glycolysis, and phospholipid biosynthesis (Fig 7), while the RNA base pairs with and represses expression of the *cadA* and *galETKM* mRNAs (Fig 14).

AzuCR is a unique dual-function sRNA

There are a number of features of AzuCR that are unique compared to other dual-function sRNAs. First, while the regions involved in base pairing and protein coding are separate for the well-characterized *E. coli* SgrS-SgrT and *Staphylococcus aureus* RNAIII dual-function sRNAs, the region of AzuCR involved in base pairing with the *cadA* and *galE* targets overlaps the *azuC* coding sequence (Fig 4A). Another notable feature of this dual-function RNA is that while the gene is not broadly conserved, the levels of the RNA and protein are highly regulated. In our previous study, we observed the AzuC protein accumulates in minimal glucose medium as well as in response to low pH, high temperature and hydrogen peroxide, while the levels are low under anaerobic conditions (Hemm, Paul et al. 2010). The regulation in response to glucose availability was shown to be at the transcriptional level via Crp derepression. In our current

study, we show that AzuC repression under anaerobic conditions is mediated by the Hfq-dependent sRNA FnrS, which base pairs near the AzuC ribosome binding site (Fig 10A). To the best of our knowledge, FnrS regulation of AzuCR is the first example of another sRNA regulating the translation of a dual-function sRNA. The observation that AzuC levels are higher in Δ*hfq* compared to Δ*fnrS* mutant cells (Fig 11A) suggests that still other Hfq-dependent sRNAs might repress AzuC synthesis. This is further emphasized by the discordance between RNA and protein levels under some conditions (Fig 4B).

Consisted with the discordant expression of the AzuCR RNA compared to the AzuC protein, we found that the small protein and base pairing activities modulate overlapping but distinct pathways; AzuC plays a role in glycerol metabolism (Fig 7 and 8) and AzuCR impacts galactose and glycerol metabolism (Fig 13). The regulation of different pathways by the two activities of AzuCR contrasts with SgrST RNA where the SgrT protein and the base-pairing SgrS RNA both down-regulate the PtsG glucose transporter activity.

AzuC stimulation of GlpD activity

Although the functions of only a few small proteins have been described, most are inhibitory. Thus, AzuC is unique in that it increases GlpD activity. GlpD, one of the key flavin-linked primary dehydrogenases of the respiratory electron transport chain, catalyzes the oxidation of glycerol-3-phosphate to dihydroxyacetone phosphate (DHAP) (Yeh, Chinte et al. 2008). GlpD exists in in both soluble and membrane-bound forms and is only fully active when the enzyme is associated with the cytoplasmic membrane through lipid-enzyme interactions or when reconstituted with phospholipids *in vitro* (Schryvers, Lohmeier et al. 1978, Robinson and Weiner 1980, Yeh, Chinte et al. 2008). Interestingly, GlpD activity was previously reported to be

increased by amphipaths (Robinson and Weiner 1980). Like GlpD, AzuC is also an amphipathic protein localized to the cytoplasmic membrane (Fig 5). Thus, it is possible that AzuC promotes GlpD binding to the cytoplasmic membrane. AzuC also could change the stability of GlpD, though we did not observe obvious differences in protein levels (Fig 9A). Alternatively, AzuC could increase GlpD by causing a conformational change of the protein upon binding. It is intriguing that expression of both AzuC and GlpD is repressed by the sRNA FnrS, which is expressed under anaerobic conditions. Under these conditions, GlpABC, the anaerobic glycerol-3-phosphate dehydrogenase, serves the same role as GlpD.

The physiological role of AzuC activation of GlpD, particularly at pH 5.5, is also an interesting question. We suggest AzuC binds GlpD under acidic conditions to modulate the levels of glycerol-3-phosphate. As a substrate for phospholipid biosynthesis, glycerol-3-phosphate undergoes two acylation steps to form phosphatidic acid, which in turn is converted to the intermediate CDP-diacylglycerol (CDP-DAG), the precursor for the phospholipids phosphatidylethanolamine (PE), phosphatidylglycerol (PG) and cardiolipin (CL). *E. coli* membranes are composed of ~75% PE, ~20% PG, and ~5% CL under normal growth conditions. Bacterial adaptation to environmental stress can be accompanied by cellular envelope remodeling, including changes in the LPS structure, and the protein content of the outer membrane and/or the inner membrane, as well as the phospholipid composition (Rowlett, Mallampalli et al. 2017). These changes in turn impact ell division, energy metabolism, osmoregulation as well as resistance to cationic antimicrobial peptides. In support of a hypothesis that AzuC activation of GlpD might affect membrane composition, we observed that cells overexpressing AzuC or GlpD grown in low pH showed increased cell length (Fig 8B) and

slow growth (Fig 13), phenotypes that have also been observed in cells lacking the phospholipids PE and CL (Rowlett, Mallampalli et al. 2017).

AzuR repression of the *cadA* and the *galETKM* mRNAs

We found that as a base pairing RNA, AzuR represses expression of CadA (Fig 11A, B and C), which is induced under acidic growth conditions and confers resistance to weak organic acids produced during carbohydrate fermentation under conditions of anaerobiosis and phosphate starvation. This regulation could partially explain the growth defect for cells growing in glycerol pH 5.5 observed upon AzuCR overexpression without and with a stop codon (Fig 13). We also identified *galETKM* mRNA as a direct AzuR target (Fig 14D, E and F). Consistent with this regulation, we see a drastic growth defect with galactose as the sole carbon source upon overexpression of AzuCR without and with a stop codon (Fig 13). While AzuR base pairs near the ribosome binding site of the *cadA* mRNA likely blocking ribosome binding, the base pairing with the *galETKM* mRNA interestingly is internal to the *galE* coding sequence. We suggest that for this mRNA, base pairing may lead to changes in mRNA stability or alternatively Rho-dependent transcription termination reported for the *galETKM* mRNA (Wang, Ji et al. 2014). Since we also observe an RNA-dependent growth phenotype for cells grown in glycerol pH 7, we suggest that AzuR might target other genes, particularly genes related to glycerol metabolism.

Competition between two AzuCR activities

Several of our experiments indicate that there is competition between the mRNA and base pairing activities of AzuCR. We observed that a stop codon blocking translation improves AzuR

base pairing activity (Fig 14B and 14E) and overexpression of fragments of the base pairing targets *cadA* and *galE* inhibits AzuC translation (Fig 15). The conflict between base pairing and translation raises intriguing questions about what activity predominates under different growth conditions, whether the RNA can transition from one function to the other, what factors determine which activity predominates, and how the two activities evolved. We suggest that there are a number of scenarios for how AzuCR could act. There may be conditions where AzuCR acts solely a riboregulator, other conditions where AzuCR is solely an mRNA. Finally, there may be conditions where there are two populations of AzuCR, some transcripts acting as an sRNA and others being translated. It is also possible that AzuCR first acts as an mRNA, but subsequently goes on to act as a riboregulator.

The factors that regulate the distribution of AzuCR between these regulatory roles are not fully understood but clearly depend on the levels of the sRNAs that repress AzuC translation, the levels of the mRNA targets of AzuCR, the levels of the Hfq and ProQ chaperones and likely other factors. The observed FnrS-dependent repression of AzuC synthesis is at least partially dependent on Hfq, while the stability of the RNA appears to depend on ProQ. The finding that AzuCR may only be a base-pairing RNA under specific conditions raises caveats for global approaches examining sRNA function by pulse overexpression or sRNA targets such as RIL-seq (Melamed, Peer et al. 2016). If these experiments are carried out under conditions where translation predominates, the effects of the base pairing activity may not be detected. These questions about AzuCR likely are relevant for other dual-function sRNAs and are an important direction for future research.

Material and Methods

Bacterial strains and plasmid construction

Bacterial strains, plasmids, and oligonucleotides used in this study are listed in Appendix Tables. E. coli strains are derivatives of wild-type MG1655 (F- lambda- ilvG- rfb-50 rph-1). Tagged strains were generated by λ Red–mediated recombineering (Yu, Ellis et al. 2000) using NM400 and the oligonucleotides listed in the Appendix. The chromosomal PBAD-5′UTRazuC-lacZ and PBAD-5′UTRazuC-lacZ III fusions (carrying the first 87-nt of the azuC mRNA fused to the seventh codon of the lacZ coding sequence) were created by carrying out PCR using primers listed in the Appendix to amplify the desired region of azuC followed by integration of the product into the chromosome of PM1205 (Mandin and Gottesman 2009). Alleles marked by antibiotic markers were moved between strains by P1 transduction. When necessary, kanamycin resistance cassettes were excised from the chromosome by FLP-mediated recombination (Cherepanov and Wackernagel 1995). All chromosomal mutations and fusions and plasmid inserts were confirmed by sequencing.

Bacterial growth

Cells were grown to the indicated OD600 in Luria-Bertani broth (LB) or M63 minimal media supplemented with 0.001% vitamin B1 and glucose, glycerol or galactose (0.2%, 0.4% or 0.2%, respectively) after a 1:100 dilution of the overnight culture. For some experiments, M63 medium was buffered to pH 5.5 with 100 mM MES. Where indicated, media contained antibiotics with the following concentrations: ampicillin (100 μg/ml), chloramphenicol (25 μg/ml) and kanamycin (30 μg/ml).

Immunoblot analysis

The cell pellet from 1 ml of cells grown in the indicated medium was resuspended in 1X PBS (KD Medical), and 10 µl were loaded on a Mini-PROTEAN TGX 5%–20% Tris-Glycine gel (Bio-Rad) and run in 1X Tris Glycine-SDS (KD Medical) buffer. The proteins were electro-transferred into nitrocellulose membranes (Invitrogen) for 1 h at 100 V. Membranes were blocked with 5% non-fat milk (BioRad) in 1X PBS with 0.1% of Tween 20 (PBS-T) for 1 h and probed with a 1:3,000 dilution of α-FLAG-HRP antiserum (Sigma), 1:1,000 dilution of α-AzuC antiserum (New England Peptide); 1:1,000 dilution of α-His-HRP antiserum (Qiagen), or 1:1,000 dilution of α-OmpA antiserum (Antibody Research Corporation) in the same PBS-T buffer with 5% milk for 1 h. After the incubation with the α-AzuC and α-OmpA antiserum, membranes were incubated with a 1:2,000 dilution of HRP-labelled anti-rabbit antibody (Life Technologies). All blots were washed 4X with PBS-T and then developed with an Amersham ECL Western Blotting Detection Kit (GE Healthcare).

Total RNA isolation

Cells corresponding to the equivalent of 10 OD600 were collected by centrifugation, and snap frozen in liquid nitrogen. RNA was extracted according to the standard TRIzol (Thermo Fisher Scientific) protocol. Briefly, 1 ml of room temperature TRIzol was add to cell pellets, resuspended thoroughly to homogenization, and incubated for 5 min at room temperature. After the addition of 200 µl of chloroform and thorough mixing by inversion, samples were incubated for 10 min at room temperature. After samples were centrifuged for 10 min at 4˚C on maximal speed, the upper phase (~0.6 ml) was transferred into a new tube and 500 µl of isopropanol was

added. Samples again were mixed thoroughly by inversion, incubated for 10 min at room temperature and centrifuged at maximal speed for 15 min at 4˚C. RNA pellets were washed twice with 75% ethanol and then dried at room temperature. RNA was resuspended in 20-50 µl of DEPC water and quantified using a NanoDrop (Thermo Fisher Scientific).

Northern analysis

Total RNA (5-10 µg per lane) was separated on denaturing 8% polyacrylamide gels containing 6 M urea (1:4 mix of Ureagel Complete to Ureagel-8 (National Diagnostics) with 0.08% ammonium persulfate) in 1X TBE buffer at 300 V for 90 min. The RNA was transferred to a Zeta-Probe GT membrane (Bio-Rad) at 20 V for 16 h in 0.5X TBE, UV-crosslinked, and probed with 32P-labeled oligonucleotides (Listed in Table 1) in ULTRAhyb-Oligo buffer (Ambion Inc.) at 45˚C. Membranes were washed twice with 2X SSC/0.1% SDS at room temperature, once with 0.2X SSC/0.1% SDS at room temperature, washed for 25 min with 0.2 × SSC/0.1% SDS at 45˚C, followed by a final wash with 0.2X SSC/0.1% SDS at room temperature before autoradiography was performed with HyBlot CL film (Denville Scientific Inc.).

Sub-cellular fractionation

Cells with chromosomally encoded AzuC-SPA were grown in the indicated medium at 37˚C to an OD600~0.3, centrifuged at 20,000 × g for 10 min at 4˚C, resuspended in 1/20 vol of 20% sucrose, 50 mM Tris pH 8, 1 mM EDTA, and 0.1 mg/ml lysozyme, followed by a 1 h incubation at 25˚C with gentle shaking. After the cells were centrifuged at 20,000 × g for 15 min at 4˚C, the top periplasmic fraction was removed. The pellet fraction was resuspended in water to lyse the spheroplasts. The resulting crude lysate was passed through a 30-gauge syringe needle 6X to

homogenize the sample and reduce viscosity. The lysate was then clarified by centrifugation at 20,000 × g for 5 min at 4˚C. This was repeated 3X. A 500 µl of the clarified lysate was layered on top of a 500 µl-sucrose cushion (5 mM EDTA and 1.4 M sucrose. Samples were centrifuged at 130,000 × g for 2 h at 4˚C in a TLA100.3 rotor (Beckman Optima TLX tabletop centrifuge). Following centrifugation, 425 µl was carefully removed from the top layer (soluble fraction). Then, the interface and remaining liquid were removed (inner membrane fraction). The pelleted material was resuspended in 500 µl of fractionation buffer (pellet fraction). SDS was added to all fractions (final concentration 1%) and the samples were incubated overnight at room temperature. Equal volumes of fractions were assayed by immunoblotting with α-FLAG-HRP and α-OmpA antibody.

Cells expressing AzuC from a plasmid were grown as above, collected by centrifuged at 4K RPM for 10 min at 4˚C, resuspended as above but incubated 10 min on ice. After the lysate was incubated as above the periplasmic fraction removed, the pellet was resuspended in 1 ml of 20% sucrose/50mM Tris pH 8 and sonicated with a Sonic Dismembrator Model 100 (Fisher Scientific) 3X for 5 sec at power setting 4. Samples were centrifuged 3X at 12,000 × g for 5 min at 4˚C to remove unlysed cells. The supernatant was then centrifuged at 56k rpm for 1 h at 4˚C in a Beckman TLA100.3 rotor. The supernatant containing the cytoplasmic fraction was removed and the pellet containing the membrane fraction was resuspended in 1 ml of 20%sucrose/10mM Tris pH 8 by sonication. Equal volumes of fractions were assayed by immunoblotting with polyclonal α-AzuC antibody.

Microscopy

Cells grown as indicated were harvested, resuspended in phosphate buffered saline (PBS) (KD Medical) and placed on lysine-coated glass bottom dish (Mattek Corporation). Cells were fixed by applying a 1% agarose pad on top of the sample with gentle pressure. Cells were viewed with a DeltaVision Core microscope system (Applied Precision) equipped with an environmental control chamber. Bright field and fluorescence images were captured with a Photometrics CoolSnap HQ2 camera. Seventeen planes were acquired every 0.2 μm at 22˚C, and the data were deconvolved using SoftWorx software (GE Healthcare).

Purification of chromosomally encoded SPA-tagged AzuC

Cells expressing AzuC-SPA (GSO351) cells grown in LB at 37˚C overnight culture were diluted 1:100 into 1 l of M63 glucose minimal media and incubated at 37˚C. At OD600 ~1, cells were collected by centrifugation (4,650 × g, 20 min). The pellet was resuspended in 20 ml of TNG buffer [10 mM Tris (pH 7.5), 100 mM NaCl, 10% glycerol] supplemented with Protease Inhibitor Cocktail (Roche). The cells were lysed using a microfluidizer processor (Microfluidics) at 20,000 psi, and the insoluble cellular debris was removed by centrifugation (20,000 × g, 30 min). The cleared lysate was incubated with 50 mM dodecyl β-D-maltoside (DDM) at 4˚C for 2 h. Next the lysate supplemented with DDM was split and added to either 500 μl of α-FLAG M2-agarose beads from mouse (Sigma) or calmodulin-sepharose beads (Amersham Biosciences) overnight at 4˚C. The lysate and beads were applied to a Bio-Spin disposable chromatography column (Bio-Rad Laboratories) and allowed to drain by gravity. The α-FLAG columns were washed with 15 mL of TNG buffer with 2 mM DDM while the calmodulin column was washed 15 mL of TNG buffer with 2 mM DDM, 5 mM β-ME, and 2 mM CaCl. Finally, proteins were eluted from the α-FLAG column with 1 ml of elution buffer (0.1M glycine pH 3.5, 100 mM

NaCl and 0.1% Triton X-100). The proteins from the calmodulin column were eluted in 1 ml TNG buffer supplemented with 4 mM EDTA, 5 mM β-ME, and 2.5% SDS. To analyze the protein samples, 7.5 µl of 2X Laemmli buffer was added to 21 µl of each sample. The samples were heated at 95˚C for 5 min, and aliquots were subjected to SDS/PAGE in a 10–20% Tris-glycine gel (Invitrogen) at 12 V/cm. Proteins were visualized with Coomassie Blue Stain. Bands of interest were excised from the gel and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). An identical purification was carried out for cells with chromosomal acrZ-SPA (GSO350) grown in 1 L of LB to OD600~0.6.

Purification of chromosomally encoded HA-His-tagged GlpD

MG1655 cells or cells expressing AzuC-SPA (GSO351), GlpD-HA-His (GSO1011) or the control MgtA-HA (GSO785) from the chromosome grown in LB at 37˚C for 16 h, were diluted 1:100 into 1 L of M63 glycerol minimal medium, M63 glycerol minimal medium, M63 glucose minimal medium or 1 L of N medium with 500 µM MgSO4, respectively and incubated at 37˚C. The WT strain and strains expressing GlpD-HA-His_6 and AzuC-SPA were grown to OD600~1. The strains expressing MgtA-HA were grown to OD600~0.4–0.6, collected, washed 2X in N medium without added MgSO4, resuspended in N medium without added MgSO4 and grown for another 2.5 h to induce MgtA-HA expression. For all cultures, cells were collected by centrifugation (4,650 × g, 20 min) and resuspended in 15 ml of TNG buffer supplemented with Protease Inhibitor Cocktail (Roche). Cells from the SPA tagged protein cultures were mixed with the control WT or HA-tagged protein cultures at a 1:1 ratio. To ensure thorough mixing, cells were shaken gently at 4˚C for 15 min. The cells were then homogenized as for the SPA-tagged protein purification and incubated with 50 mM DDM in 4˚C for 2 h. The insoluble cellular debris

was removed by centrifugation (20,000 × g, 20 min). Subsequently, the supernatant was applied to 100 µL of Pierce Anti-HA magnetic beads (Thermo Scientific) in a 50 ml tube and incubated overnight at 4˚C. Beads were collected with a MagneSphere technology magnetic separation stand (Promega) and resuspended in 1 ml of TNG buffer. The beads were washed with 1 ml of TNG buffer (10X). The beads were then resuspended in 1XPBS (50 µl) and 2X Laemmli buffer (50 µl) and heated at 95˚C for 5 min. Samples (15 µL) were analyzed on immunoblots using α-His or M2 α-FLAG antibodies.

Dehydrogenase activity assay

Cells were grown in M63 glucose minimal medium to OD600 ~1. Cells were pelleted and washed with M63 glycerol medium, pH 7.0 or pH 5.5. Cells were then resuspended in same volume of the same medium and grown at 37˚C for 3 h. Cells (500 µl) were pelleted and resuspended in 500 µl of lysis buffer (25 mM Tris-HCl, 10 mM NaCl and 0.4% Triton X100). Cells were lysed by adding 0.6 g of glass beads and vortexing 30 s followed by 30 s incubation on ice, repeated 5X. The cells were then centrifuged at 20,000 × g for 2 min at 4˚C, and the lysate was used to measure the dehydrogenase activity. A method monitoring MTT reduction to quantitate the dehydrogenase activity of GlpD (Yeh, Chinte et al. 2008) was modified as follows. Each 225 µl microcuvette contained the following: 25 mM Tris/HCl pH 7.4, 100 mM NaCl, 1 mM MTT (Sigma Aldrich), 3 mM phenazine methosulfate (PMS, Sigma Aldrich) and 100 µl of lysate. This was used as the blank, and the reaction was initiated by the addition of 3.7 mM DLglycerol-3-phosphate (Sigma Aldrich). The reduction of MTT at 570 nm was continuously monitored on a BMG LABTECH plate reader for 118 min at room temperature.

β-galactosidase assays

Cultures were grown in LB to OD_{600}~1.0 with arabinose (0.2%). 100 ul of cells were added to 700 ul of Z buffer(60 mM Na2HPO4z7H2O, 40 mM NaH2PO4zH2O, 10 mM KCl, 1 mMMgSO4z7H2O, 50 mMb-mercaptoethanol). After adding 15 ul of freshly prepared 0.1% SDS and 30 ul of chloroform to each sample the cells were vortexed for 30 s and then incubated at room temperature for 15 m to lyse the cells. At zero time, the assay is initiated by adding 100 ul of ONPG (4 mg/ml) to each sample in 10 s intervals. The samples are incubated at room temperature before the reaction is terminated by the addition of 500 ul of 1M Na_2CO_3. Then, the A420 and A550 values are determined with a spectrophotometer and the absorbance data is transferred to a Microsoft Excel spreadsheet to calculate Miller units.

Growth curves

Colonies of ΔazuC::kan (GSO193) transformed with pRI, pRI-AzuCR, pRI-AzuCRL3STOP, pKK, pKK-AzuC or pKK-AzuCL3STOP grown on LB plates were inoculated into glucose (pH 7.0), glycerol (pH 7.0 and 5.5), and galactose (pH 7.0) and allowed to grow overnight at 37˚C, at which point all cultures were in stationary phase. Cultures were diluted to OD600~0.05 (time 0) in 25 ml of the same media and grown at 37˚C. OD600 was measured at 16 h or growth was followed for 29 h.

GFP reporter assay

The GFP reporter assay was principally done as described previously (Urban and Vogel 2009, Corcoran, Podkaminski et al. 2012). ΔazuC::kan (GSO193) cells were transformed with a cadA-gfp, cadA-gfp-M1, galE-gfp or galE-gfp-M2 reporter plasmid and a pRI-AzuCR, pRI-

AzuCRL3STOP, pRI-AzuCRL3STOP-M1 or AzuCRL3STOP-M2 over-expressing plasmid or pRI as a control. Single colonies were grown overnight at 37˚C in LB supplemented with ampicillin and chloramphenicol. The cultures were diluted to OD600 = 0.05 in fresh medium and grown at 37˚C for 3 h in a 96 deep-well plate. An aliquot (1 ml) of each culture was centrifuged and the pellet was resuspended in 220 µl of 1X PBS. Fluorescence was measured using the CytoFLEX Flow Cytometer (Beckman Coulter). Three biological repeats were analyzed for every sample.

Hfq and ProQ co-immunoprecipitation assays

Cell extracts were prepared from MG1655 cells grown in M63 glucose medium to OD600~0.5. Cells corresponding to the equivalent of 20 OD600 were collected, and cell lysates were prepared by vortexing with 212-300 µm glass beads (Sigma-Aldrich) in a final volume of 1 ml lysis buffer (20 mM Tris-HC, pH 8.0, 150 mM KCl, 1 mM MgCl2, 1 mM DTT). Immunoprecipitations were carried according to (Zhang, Wassarman et al. 2002) using 100 µl of Hfq antiserum (Zhang, Wassarman et al. 2002) or 100 µl of ProQ antiserum (Melamed, Adams et al. 2020), 120 mg of protein-A-sepharose (Amersham Biosciences, Piscataway, NJ) and 950 µl of cell extract per immunoprecipitation reaction. Immunoprecipitated RNA was isolated from immunoprecipitated pellets by extraction with phenol:chloroform:isoamyl alcohol (25:24:1, pH 8), followed by ethanol precipitation. Total RNA was isolated from 50 µl of cell lysate by TRIzol (Thermo Fisher Scientific) extraction followed by chloroform extraction and isopropanol precipitation. Total and co-IP RNA samples were resuspended in 20 µl of DEPC H2O and 2 µg of total RNA or 200 ng of IP RNA was subjected to Northern analysis as described below.

Addendum

The AzuCR work has provided insights into the functions of the AzuC small protein and AzuR sRNA and revealed how the RNA and protein components of this dual-function RNA compete. There are still multiple interesting avenues of research to pursue for AzuCR. Given how transient membrane association of GlpD impacts activity of the protein, I think that further investigation of the subcellular localization of AzuC during interactions with GlpD could be intriguing. GFP and YFP tagged derivatives of AzuC and GlpD have been generated. I could move both proteins into the same background and examine their behavior before, during, and immediately after a shift from glucose to low pH (5.5) glycerol media. As AzuC translation is induced by this media I would expect to see changes in localization of either protein during these shifts.

I also think it would be interesting to more fully characterize how AzuC modulates GlpD activity. Alanine mutagenesis could be carried out for the AzuC protein and mutants could be overexpressed in Δ*azuC* cells which would be analyzed by the MTT dehydrogenase activity assay. This assay would identify which residues of AzuC are important for altering GlpD activity. If this method is successful, the co-precipitation assays that were used to verify the interaction of AzuC with GlpD could be repeated with a promising alanine mutant compared against wild type AzuC to determine if the mutated residue was important for interaction or modulation of GlpD activity.

This work did not fully explain the underlying mechanisms that control how these two overlapping components compete. The hypothesis that AzuCR interacts with ProQ as an sRNA and Hfq as an mRNA, based upon the co-immunoprecipitation data, could be pursued to answer this question. I think repeating the co-immunoprecipitation of AzuCR RNA with Hfq and ProQ

RNA binding proteins in different carbon sources could be valuable to identify conditions where AzuCR is potentially acting as an sRNA or an mRNA.

Another interesting avenue of research is how AzuCR is regulating mRNA targets. An interesting experiment would be to overexpress AzuCR RNA and the stop codon control and examine the behavior of AzuCR and *galE* RNA levels by northern analysis. If there is evidence that RNA molecules co-degrade then RNA stability could be a guiding factor in determining when AzuCR functions as a regulatory RNA or an mRNA. To search for additional mRNA targets, I suggest carrying out RIL-seq using a chromosomal stop codon mutant of AzuCR to prevent translation of the small protein from disrupting base-pairing interactions. This genomic search could be done in many conditions, but I would start in glycerol to search for RNA interactions that may also contribute to regulation of glycerol metabolism, as with the small protein. I believe that the regulation of AzuCR by the anaerobic sRNA FnrS may not be the only regulatory RNA involved in interactions with this dual-function RNA. The RIL-seq analysis would also have the potential to identify other regulatory RNAs that interact with AzuCR.

The work that has been undertaken on AzuCR has identified a novel dual-function RNA and characterized how it regulates targets using the RNA and protein functions. This project has potential to provide further understanding of the competition between activities of dual-function RNA and possibly identify new mechanisms for the control of this competition.

Chapter 3

Spot 42 small regulatory RNA encodes a 15-amino acid protein that blocks Crp-mediated transcription activation

Abstract

The 109-nucleotide Spot 42 RNA is one of the best characterized base-pairing small RNAs (sRNAs) in *E. coli*. Transcription of Spot 42 is repressed by the cAMP receptor protein (Crp). Consistent with high levels of Spot 42 levels in glucose-grown cells, the RNA blocks the expression of transporters and enzymes involved in the utilization of non-preferred carbon such as galactose. We now document that Spot 42 also encodes a 15-amino acid protein denoted SpfP. Previous studies showed overexpression of Spot 42 reduces growth in galactose and other non-preferred carbon sources. Overexpression of just the small protein from a Spot 42 derivative deficient in base pairing activity also prevented growth on galactose, revealing that the sRNA and protein impact the same pathway. Co-purification experiments showed that SpfP binds Crp. This binding blocks the ability of Crp to activate specific genes such as the *galETKM* operon, impacting the kinetics of induction when cells are shifted from glucose to galactose medium. Thus, the small protein reinforces the feedforward loop regulated by the base pairing activity.

Introduction

Carbon catabolite repression (CCR) is a mechanism used by bacteria to promote the use of a preferred carbon source in an environment where non-preferred substrates are also available (reviewed in (Bruckner and Titgemeyer 2002)). In gram-negative bacteria, like *Escherichia coli,* this is generally achieved by preventing expression of genes involved in catabolism of non-preferred carbon sources when the preferred carbon source such as glucose is available. In *E. coli*, the key components of the CCR pathway are the IIA component of the glucose-specific phosphotransferase system (PTS) [$EIIA^{Glc}$; also called catabolite repression resistance (Crr) or EIIACrr], adenylate cyclase, cAMP, and the cyclic AMP (cAMP) receptor protein (Crp) (reviewed in (Kolb, Busby et al. 1993, Gorke and Stulke 2008). When glucose is limiting, the PTS proteins, including $EIIA^{Glc}$, are predominantly phosphorylated. In this form, P-$EIIA^{Glc}$ binds to adenylate cyclase and activates cAMP synthesis. The high cAMP levels are sensed by Crp, which in turn binds and regulates hundreds of promoters of catabolic genes such as the *galETKM* operon for galactose metabolism and the *malEFG* operon for maltose metabolism (Zheng, Constantinidou et al. 2004).

Crp is a bifunctional regulator (reviewed in (Kolb, Busby et al. 1993)). As a transcriptional activator, cAMP-bound Crp binds to a sequence located upstream from (class I activation) or close to (class II activation) promoter DNA and participates in protein-protein interactions with RNA polymerase leading to transcription initiation (reviewed in (Busby and Ebright 1999)). Crp has also been shown to negatively regulate expression of several genes by inhibiting transcription initiation (Mallick and Herrlich 1979, Aiba 1983).

In addition to regulating protein-coding genes, *E. coli* Crp modulates the transcription of small regulatory RNAs (sRNAs); repressing Spot 42 expression and activating CyaR expression

(Polayes, Rice et al. 1988, De Lay and Gottesman 2009). Spot 42 is a 109 sRNA encoded by the *spf* (spot forty-two) gene found in α-proteobacteria (reviewed in (Baekkedal and Haugen 2015)). Spot 42 is highly expressed in glucose and its transcription is inhibited by the cAMP-Crp complex during growth in non-preferred carbon sources (Polayes, Rice et al. 1988). Many years after its first characterization (Ikemura and Dahlberg 1973, Sahagan and Dahlberg 1979), Spot 42 was shown to bind the RNA chaperone Hfq and base pair with mRNA targets to regulate their expression (Moller, Franch et al. 2002). The third gene in the galactose operon, *galK* encoding galactokinase, was the first identified Spot 42 target; down regulation of this gene leads to discordant expression of the galactose operon (Moller, Franch et al. 2002). Later, Spot 42 was shown to base pair with and repress expression from a number of mRNAs encoding proteins involved in the uptake and catabolism of non-preferred carbon sources (Beisel and Storz 2011, Beisel, Updegrove et al. 2012). Consistent with this regulation, Spot 42 overexpression negatively impacts growth on a number of non-preferred carbon sources (Beisel and Storz 2011).

Apart from being an sRNA, Spot 42 RNA resembles a short mRNA. It contains a ribosome binding site followed by an AUG start codon, a 15 amino acid open reading frame (ORF) and a UGA stop codon (Fig 17A) (Sahagan and Dahlberg 1979). However, an early study examining the affinity between Spot 42 and the 70S ribosome showed that although the RNA bound to ribosomes, it did so inefficiently and nonproductively (Rice, Polayes et al. 1987). A fusion between the Spot 42 ORF to *lacZ* also did not support the synthesis of β-galactosidase. These studies led to the conclusion that Spot 42 does not function as a mRNA (Rice, Polayes et al. 1987).

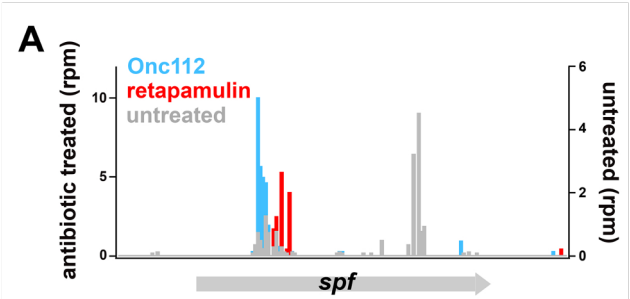
Here we report that in fact, Spot 42 encodes a functional small protein (SpfP) in addition to acting as a regulatory sRNA and thus should be classified as a dual-function RNA. The small

protein, whose levels are elevated at higher temperature, functions by binding to Crp. Binding of SpfP to Crp modulates Crp activity by blocking Crp-dependent activation thereby aiding in CCR by inhibiting the function of Crp in the presence of glucose and reinforcing the feedforward loop regulated by the base pairing activity of Spot 42.

Results

Spot 42 regulatory sRNA encodes a 15 amino acid protein

Ribosome profiling in the presence of translation inhibitors (Onc112 and retapamulin) that trap ribosomes on start codons suggested the 15 amino acid ORF encoded by Spot 42 RNA might be translated (Fig 16) (Weaver, Mohammad et al. 2019). To directly test whether the protein



B

```
Escherichia coli                MFYLSDLLLHVIGFG  15
Erwinia billingiae              MFYLSNLLLHVIGFG  15
Salmonella enterica             MFYLSDLLLHVIGFG  15
Brenneria goodwinii             MFYLSDLLLHVIGFG  15
Hafnia alvei                    MFYLSDLLLHVIGFG  15
Buttiauxella agrestis           MFYLSDLLLHVIGFG  15
Pantoea ananatis                MFYLSDLLLHVIGFG  15
Raoultella ornithinolytica      MFYLSDLLLHVIGFG  15
Cedecea davisae                 MFYLSDLLLHVIGFG  15
Cronobacter sakazakii           MFYLSDLLLHVIGFG  15
Plesiomonas shigelloides        MFYLSDLLLHVIGFG  15
Pluralibacter gergoviae         MFYLSDLLLHVIGFG  15
Klebsiella pneumoniae           MFYLSDLLLHVIGFG  15
Serratia fonticola              MFYLSDLLLHVIGFG  15
Lonsdalea quercina              MFYLSDLLLHVIGFG  15
Tatumella morbirosei            MFYLSDLLLHVIGFG  15
Kluyvera ascorbata              MFYLSDLLLHVIGFG  15
Citrobacter koseri              MFYLSDLLLHVIGFG  15
Shigella dysenteriae            MFYLSDLLLHVIGFG  15
Enterobacter aerogenes          MFYLSDLLLHVIGFG  15
Candidatus sodalis              MFYLSDLLLHVIGFG  15
Leclercia adecarboxylata        MFYLSDLLLHVIGFG  15
Edwardsiella ictaluri           MFYLSDLLLHVIGFG  15
Sodalis praecaptivus            MFYLSDLLLHVIGFG  15
Trabulsiella guamensis          MFYLSDLLLHVIGFG  15
Yokenella regensburgei          MFYLSDLLLHVIGFG  15
Rahnella aquatilis              MFYLSDLLLHVIGFG  15
Chania multitudinisentens       MFYLSDLLLHVIGFG  15
Shimwellia blattae              MFYLSDLLLHVIGFG  15
Enterobacter cloacae            MFYLSDLLLHVIGFG  15
Yersinia mollaretii             MFYLSDLLLHVIGFG  15
Morganella morganii             MFYLSDLLLHVIGFS  15
Photorhabdus luminescens        MFYLSDLLLHVIGFS  15
Proteus penneri                 MFYLSDLLLHVIGFS  15
Moellerella wisconsensis        MFYLSDLLLHVIGFS  15
Xenorhabdus beddingii           MFYLSDLLLHVIGFS  15
Providencia burhodogranariea    MFYLSDLLLHVIGFS  15
Enterovibrio coralii            MFYLSDLLFHVIGFS  15
Grimontia hollisae              MFYLSDLLFHVIGFS  15
Vibrio cholerae                 MFYLSDLLFHVIGLG  15
Photobacterium aphoticum        MFYLSDLLFHVIGFG  15
Prevotella dentalis             MFY--DDLLHVIGFE  13
Hallella seregens               MFY--DDLLHVIGFE  13
Rhodopirellula maiorica         MFYL-GLLLQVIGLA  14
                                ***  . *::**
```

Figure 16. Ribosome binding to and conservation of *spf*.

A) Ribosome density on the *spf* gene detected by ribosome profiling in the presence of the translation inhibitors Onc112-treated (blue) (Weaver, Mohammad et al. 2019), retapamulin-treated (red) (Meydan, Marks et al. 2019) and an untreated control sample (gray) (Weaver, Mohammad et al. 2019).

B) The SpfP amino acid sequences from *E. coli* K12 and other bacterial species aligned with ClustalW. "*" indicates the residues are identical in all sequences, and ":" and "."indicate that conserved and semi-conserved substitutions, respectively, as defined by ClustalW.

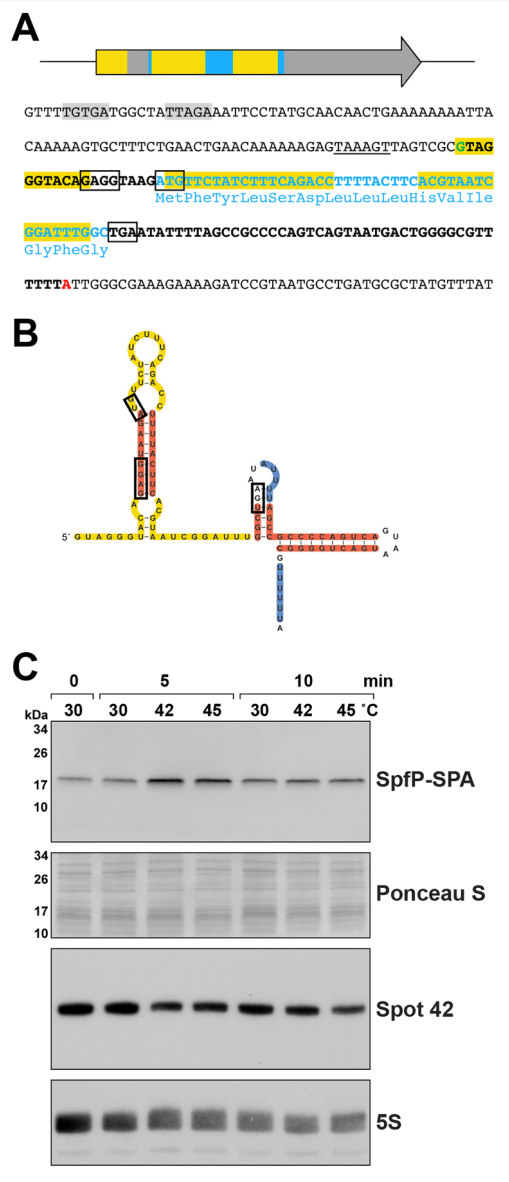


Figure 17. Spot 42 is a dual-function sRNA.

A) Diagram of the Spot 42 RNA and sequence of the *spf* promoter and coding region. Boxes and text in light blue denote SpfP coding sequence and yellow box and highlighted text denote region of Spot 42 base pairing with target mRNAs. The Spot 42 transcript is indicated in bold with the +1 site of transcription (position 1988001 of the *E. coli* K-12 genome) in green font and the 3′ end of the transcript in red font. The ribosome binding site and the start (ATG) and the stop codons (TAA) of the AzuC ORF are indicated by black boxes. A potential σ^{70} -10 sequence is underlined and a predicted Crp binding site (Keseler, Mackie et al. 2017) is highlighted in light gray.

B) Diagram of Spot 42 secondary structure (taken from (Updegrove, Shabalina et al. 2015)) with the Shine-Dalgarno sequence, start, and stop codon boxed. Single-stranded regions involved in base pairing with mRNAs are highlighted in yellow. The Hfq binding regions are indicated in dark blue and double stranded regions are indicated in red.

C) Immunoblot analysis of SpfP-SPA cells (GSO1037) grown at 30˚C in LB and transferred to 42˚C and 45˚C. Samples collected before heat shock at OD_{600}~0.4 and 5- and 10-min following heat shock. α-FLAG antibody was used to detect the SPA tag. The Ponceau S stain documents approximately equal loading of the samples. For total RNA isolated for WT cells grown under the same conditions, the Spot 42 and 5S RNAs were detected with probes specific to each of these transcripts.

encoded by the predicted ORF is synthesized, we integrated the sequence for the sequential peptide affinity (SPA) tag onto the chromosome just upstream of the predicted stop codon. We then carried out immunoblot analysis with α-FLAG antibodies to assay for the protein. Given that the potential RBS and start codon might by sequestered by a stem-loop in the Spot 42 structure, we hypothesized that higher temperatures might resolve the stem-loop promoting

translation (Fig 17B). Cells with the chromosomal SPA fusion grown to exponential phase in LB at 30˚C thus were shifted to 30˚C, 42˚C or 45˚C. The tagged protein was detected at all temperatures (Fig. 17C) indicating that Spot 42 is also an mRNA encoding a 15 amino acid protein, hereafter referred to as SpfP. Interestingly however, compared to the 30˚C samples, the levels of SpfP-SPA were higher at 42˚C and 45˚C, particularly at the 5 min time point, while the levels of the Spot 42 RNA assayed for cells grown under the same conditions were somewhat lower at 42˚C and 45˚C (Fig 17C).

Spot 42 base pairing and protein coding activities can be genetically separated

Upon documenting SpfP expression, we wondered whether SpfP might be responsible for some of the previously observed phenotypes attributed to Spot 42 (Beisel and Storz 2011). To separate the protein coding activity from the base pairing activities of *spf*, we constructed an overexpression plasmid encoding only the *spfP* ORF in which the sequence of two of the Spot 42 base pairing regions were scrambled while maintaining the proper amino acid sequence of SpfP (pKK-SpfP-scram) (Fig 18A).

To check whether the scrambled construct functioned as a regulatory sRNA, we assayed the effect of SpfP-scram overexpression on *lacZ* fusions to three mRNAs known to be direct targets of Spot 42 (Beisel and Storz 2011). β-galactosidase activity was assayed in cells expressing *nanC-lacZ*, *srlA-lacZ*, and *glpF-lacZ* translational fusions transformed with vector controls or the corresponding plasmids overexpressing either Spot 42 (from pRI, a derivative of pKK177-3 without the ribosome binding site) or SpfP-scram (from pKK177-3). Regulation of the targets was only observed when the full-length Spot 42 transcript was overexpressed

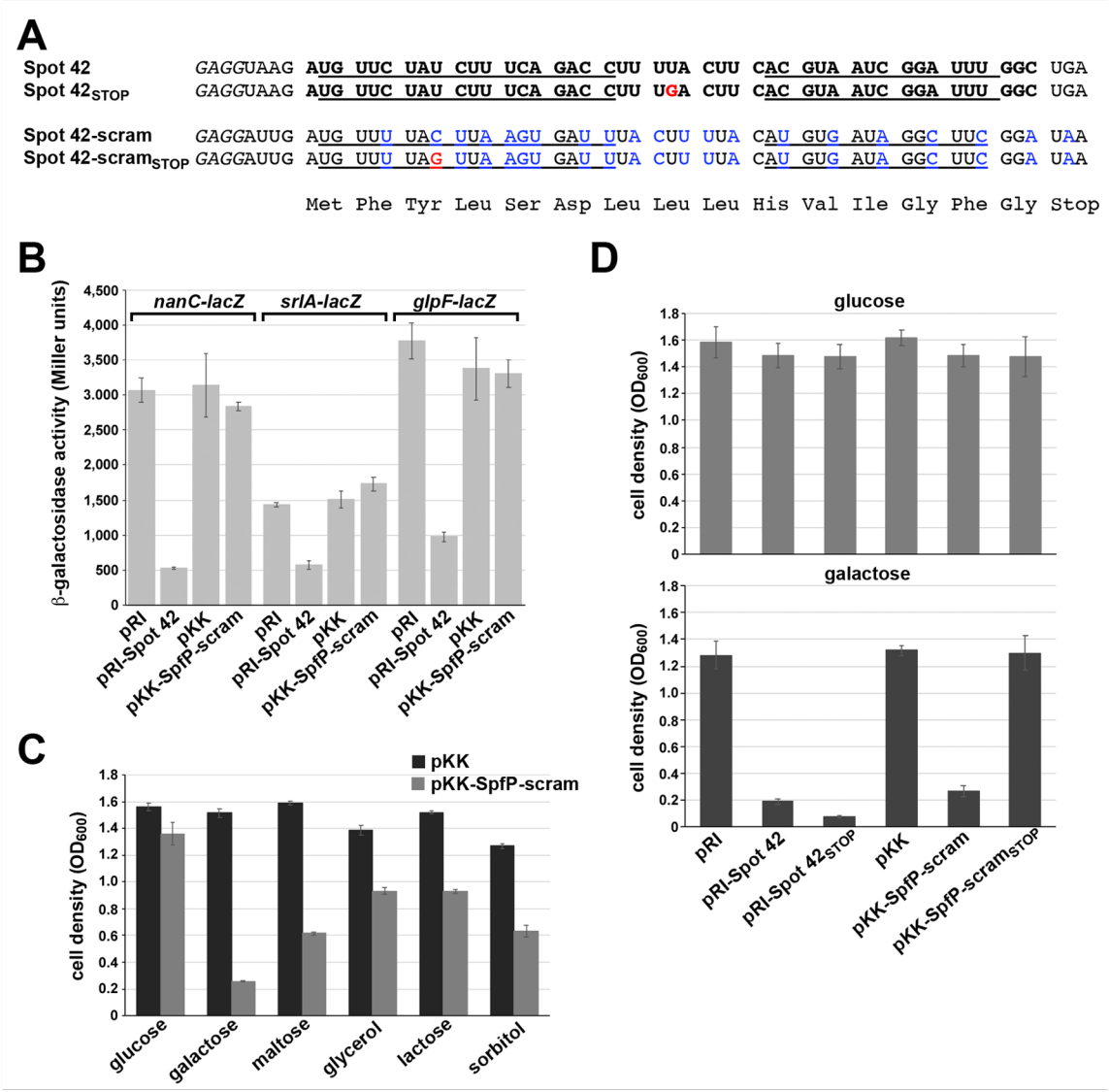


Figure 18. SpfP expression impacts growth on galactose.

A) Sequence of region encoding 15 amino acid SpfP. Nucleotides changed in the scrambled derivative are indicated in blue, and nucleotides changed in the STOP mutants are indicated in red.

B) β-galactosidase assay of Δ*spf nanC-lacZ* (GSO440), Δ*spf srlA-lacZ* (GSO441), and Δ*spf glpF-lacZ* (GSO519) expressing cells transformed with pRI, pRI-Spot 42, pKK, or pKK-SpfP-scram. The cells were grown to OD_{600}~1.0 in LB supplemented with 0.2% arabinose. Bars depict average of three biological replicates and error bars correspond to the standard deviation.

C) Growth assays of Δ*spf::kan* cells (GSO433) transformed with pKK or pKK-SpfP-scram in M63 minimal medium supplemented with the indicated carbon sources all at 0.2% except for glycerol, which was at 0.4%.

D) Growth assays of Δ*spf::kan* cells (GSO433) transformed with pRI, pRI-Spot42, pRI-Spot42$_{STOP}$, pKK, pKK-SpfP-scram or pKK-SpfP-scram$_{STOP}$ were grown in M63 minimal medium supplemented with either glucose (top panel) or galactose (bottom panel).

For (C) and (D), cells grown overnight in LB with ampicillin were diluted to OD_{600}~0.05 in M63 minimal medium with the indicated carbon sources and grown for 16 h, at which point OD_{600} was measured.

For (B), (C) and (D), the average of three replicates is showed together with the standard deviation of the mean.

suggesting that these targets are regulated by Spot 42 but not by the scrambled construct (Fig 18B). These data show that SpfP-scram does not function as a base pairing sRNA (Fig 18B).

To test whether SpfP contributes to Spot 42 phenotypes, we examined how expression of SpfP impacts growth on non-preferred carbon sources as previously tested for Spot 42 (Beisel and Storz 2011). Δ*spf* cells were transformed with the overexpression plasmids and OD_{600} was measured after 16 h of growth in M63 minimal media supplemented with either glucose or other non-preferred carbon sources. Interestingly, as observed for Spot 42, overexpression of SpfP led to a growth defect in minimal medium with most of the non-preferred carbon sources, with galactose giving the greatest defect (Fig 18C).

We also constructed stop mutants of Spot 42 and SpfP to further examine the effects of both the RNA and the small protein on growth on galactose. We observed that overexpression of full-length Spot 42 produces a growth defect in galactose that is not relieved by the stop codon control (Fig 18D). This indicates that overexpression of the sRNA alone is able to produce a growth defect. In contrast, a STOP codon mutation eliminated the growth phenotype associated with SpfP-scram overexpression. Taken together, these data demonstrate that *spf* functions as both an sRNA and mRNA.

Histidine 10 is critical for SpfP function

To determine which SpfP residues are important for its observed phenotype in galactose, we carried out alanine scan mutagenesis of the SpfP coding sequence in the scrambled-sequence context. The mutants were overexpressed in a Δ*spf* background and growth in M63 galactose was measured after 16 h. Cell density was approximately 4-fold less upon overexpression of the scrambled small protein (pKK-SpfP-scram) when compared to empty vector. The F2A and S5A

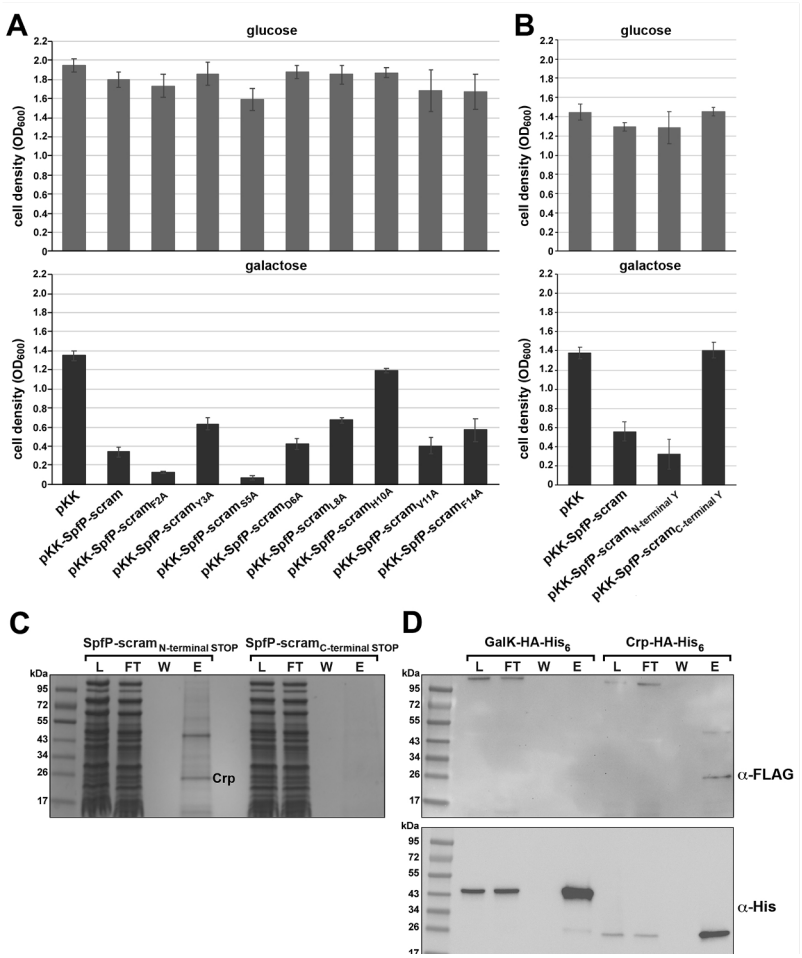


Figure 19. SpfP copurifies with Crp.

A) Growth assays of Δ*spf::kan* cells (GSO433) expressing alanine-substitution mutants compared to WT pKK-SpfP-scram in M63 minimal media supplemented with either glucose (top panel) or galactose (bottom panel).

B) Growth assays of Δ*spf::kan* cells (GSO433) transformed expressing pKK-SpfP-scram derivatives with either an additional N-terminal Y or C-terminal Y in M63 minimal media supplemented with either glucose (top panel) or galactose (bottom panel).

For (A) and (B), cells were assayed as in Fig 18, and the average of three replicates is showed together with the standard deviation of the mean.

C) Crp copurifies with biotin-tagged SpfP. Δ*spf::kan* cells (GSO433) with the SpfP-scram$_{\text{N-terminal STOP}}$ or SpfP-scram$_{\text{C-terminal STOP}}$ plasmids and the orthogonal tRNA and amino-acyl tRNA pair were grown in M63 glucose medium with 1mM *p*-azido phenylalanine to OD_{600}~ 0.5 and induced with 0.2% arabinose for 3 h. Cell lysates were treated with biotin-PEGG-alkyne to biotinylate SpfP and passed over streptavidin beads. Fractions from the lysate (L), flow-through (FT), wash (W) and eluants (E) for each sample were subjected to SDS/PAGE followed by Coomassie blue staining. The unique ~25 kDa band enriched in the eluant from the SpfP expressing cells was excised from the gel and identified by mass spectrometry.

D) N-terminally FLAG tagged SpfP copurifies with Crp-HA-His$_6$ but not GalK-HA-His$_6$. Δ*spf* cells expressing Crp-HA-His$_6$ or GalK-HA-His$_6$ from the chromosome (GSO1061 and GSO1060, respectively) were transformed with pKK-SpfP-scram$_{\text{N-FLAG}}$ and grown in LB until OD_{600}~ 0.5. The cell lysate was incubated overnight with 50 µl of Pierce α-HA magnetic beads. The beads were collected using a magnet and proteins were eluted with Laemmeli buffer. Immunoblots of fractions from the lysate (L), flow-through (FT), wash (W) and eluants (E) separated by SDS/PAGE were probed α-FLAG (top panel) and α-His (bottom panel) antibodies. A cross-reacting band of high molecular weight is detected with α-FLAG antibodies.

derivatives produced a more severe effect on cell density when compared to pKK-SpfP-scram (Fig 19A). Conversely, Y3A, L8A, and F14A slightly relieved the growth defect while H10A almost completely relieved the effect of SpfP-scram overexpression (Fig 19A).

SpfP associates with Crp

To gain insight into the role of SpfP, we sought to identify co-associating proteins. As the SPA tag is almost three times the size of SpfP, we set out to generate a SpfP derivative with a non-native amino acid azido-modified phenylalanine amino acid (*p*-AzF) that could be modified by biotin and is structurally somewhat similar to tyrosine. We first tested the possible effect of introducing *p*-AzF on SpfP activity by inserting a tyrosine at either the N-terminus of SpfP between the first and second amino acid or at the C-terminus just upstream of the stop codon, again in the context of the scrambled-sequence construct. The growth assay indicated that the N-terminally tagged construct (pKK-SpfP-scram$_{\text{N-terminal Y}}$) has wild type activity, while the C-terminally tagged construct (pKK-SpfP-scram$_{\text{C-terminal Y}}$) does not and could serve as a negative control in our subsequent co-purification experiments (Fig 19B).

To purify SpfP, we introduced a stop codon between the first and second amino acid or just upstream of the stop codon as a control. Each of the stop codon derivatives of SpfP was then expressed under conditions where *p*-AzF was incorporated at the stop codon by utilizing an engineered, orthogonal tRNA and aminoacyl-tRNA synthetase pair (pEVOL-*p*-AzF) (Chin, Santoro et al. 2002). The SpfP derivatives bearing *p*-AzF were biotinylated and purified on streptavidin beads. The eluant from each column was separated by SDS-PAGE. A prominent band of ~25 kDa was observed for the tagged protein expressed from the SpfP-scram$_{\text{N-terminal STOP}}$ construct but not the SpfP-scram$_{\text{C-terminal STOP}}$ construct (Fig 19C). In an independent experiment,

we also found this ~25 kDa protein enriched in the SpfP-scram$_{N\text{-terminal STOP}}$ sample compared to the sample for another small protein, YoaK$_{N\text{-terminal STOP}}$, purified in the same way (Fig 20A). Mass spectrometric analysis revealed that Crp (23.6 kDa) was enriched in this ~25 kDa band in both experiments. We noted that an ~45 kDa band co-purified with the protein expressed from the SpfP-scram$_{N\text{-terminal STOP}}$ construct in both experiments but focused on Crp in this study given the connection to carbon metabolism.

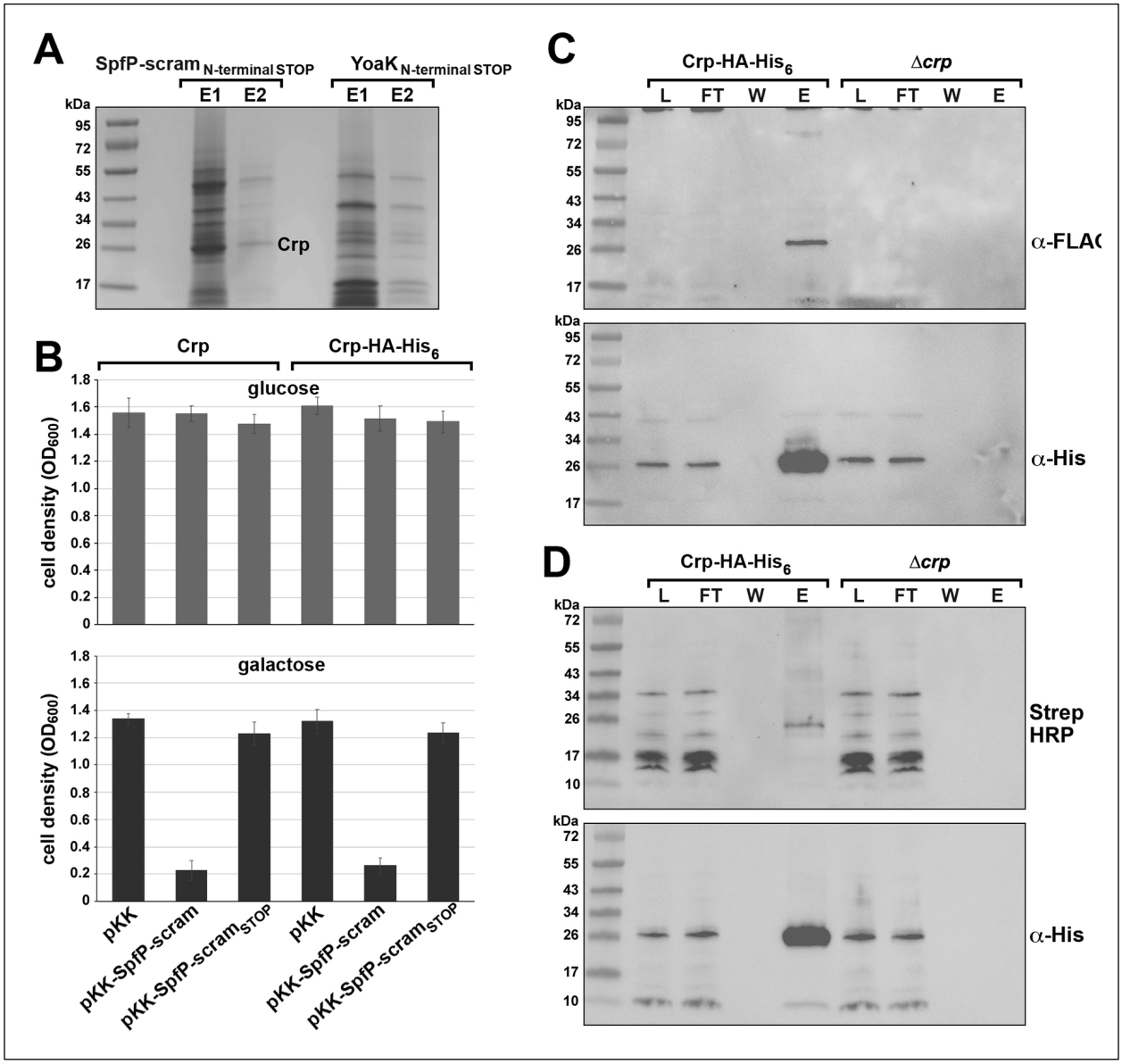


Figure 20. Lack of SpfP-FLAG association in Δ*crp* mutant background.

A) Crp copurifies with biotin-tagged SpfP but not YoaK. WT cells (GSO433) carrying the SpfP-scram$_{\text{N-terminal STOP}}$ (left), or YoaK$_{\text{N-terminal STOP}}$ (right) plasmids were both transformed with pEVOL-p-AzF, a plasmid that expresses an orthologous aminoacyl-tRNA and tRNA pair. Cells grown in LB to OD_{600}~ 0.5 were lysed and applied to streptavidin beads. After the beads were washed, the bound protein was eluted in two steps (E1 and E2). The samples were separated by SDS-PAGE and stained with Coomassie blue. Unique bands in the eluate were sent for mass spectrometric analysis and Crp was identified in the indicated band.

B) Same repressive effect of SpfP is observed in *crp*$^{+}$ and *crp-HA-His$_6$* backgrounds. Growth assays of Δ*spf::kan* (GSO433) and cells Δ*spf crp-HA-His$_6$::kan* (GSO1061) transformed with pKK, pKK-SpfP-scram or pKK-SpfP-scram$_{\text{STOP}}$ were grown in M63 minimal medium supplemented with either glucose (top panel) or galactose (bottom panel). Cells were grown overnight in LB with ampicillin were diluted to OD_{600}~0.05 in M63 minimal medium with the indicated carbon sources and grown for 16 h, at which point OD_{600} was measured. The average of three replicates is showed together with the standard deviation of the mean.

C) N-terminally FLAG tagged SpfP associates with ~25 kDa band in Crp-HA-His$_6$ but not Δ*crp* background. Δ*spf crp-HA-His$_6$::kan* (GSO1061) or Δ*spf::kan* Δ*crp::cm* (GSO1063) cells expressing N-terminally FLAG-tagged SpfP-scram from the pKK plasmid were grown in LB to OD_{600}~ 0.5. Cells were lysed (L) and applied to anti-HA beads, and the flow through (FT) fraction was collected. Beads were washed (W) and the bound protein was eluted (E). The samples were examined on immunoblots using α-FLAG antibodies to detect SpfP-scram$_{\text{NFLAG}}$ (top) or a-His antibodies to detect CRP-HA-His$_6$ (bottom panel). A cross-reacting band of high molecular weight is detected with the α-FLAG antibodies, and a cross-reacting band of the same size as CRP is detected with the α-His antibodies.

D) Synthetic SpfP associates with ~25 kDa band in Crp-HA-His$_6$ but not Δ*crp* background. Extracts prepared from Δ*spf crp-HA-His$_6$::kan* (GSO1061) or Δ*spf::kan* Δ*crp::cm* (GSO1063) cells grown LB to OD_{600}~ 0.5 were incubated with N-terminally biotinylated SpfP (biotin-MFYLSDLLLHVIGFG-COOH, Thermo Fisher Scientific) for 2 h. The lysates (L) were applied to anti-HA beads, and the flow through (FT) fraction was collected. Beads were washed (W) and the bound protein was eluted (E). The samples were examined on immunoblots using streptavidin-HRP to detect biotin-SpfP (top panel) or α-His antibodies to detect CRP-HA-His$_6$ (bottom panel). A number of cross-reacting bands are detected with streptavidin-HRP, and a cross-reacting band of the same size as CRP is detected with the α-His antibodies.

To test for an interaction between SpfP and Crp, we carried out a reciprocal co-purification experiment in which N-terminally FLAG-tagged SpfP was expressed in a Δ*spf* background with chromosomally tagged Crp-HA-His$_6$ or GalK-HA-His$_6$ as a cytosolic protein control. The inhibitory effect of SpfP on growth in minimal galactose medium was not altered in the Crp-HA-His$_6$ strain (Fig 20B). The cells expressing the HA-His$_6$-tagged proteins were grown in LB media to OD_{600}~0.5, lysed and incubated with α-HA magnetic beads. Immunoblot analysis of the fractions from this purification was carried out using α-FLAG and α-His antibodies (Fig 19D). The immunoblot probed with the α-His antiserum showed that the tagged derivatives of GalK and Crp were clearly enriched by incubation with the α-HA beads. The immunoblot probed with α-FLAG showed one prominent band of ~25 kDa. FLAG-tagged SpfP should be of ~ 2

kDa. We hypothesized that this ~25 kDa cross-reacting band was FLAG-SpfP complexed with Crp. To test this hypothesis, we carried out a similar co-purification with the chromosomally tagged Crp-HA-His_6 compared to a Δ*crp* control strain (Fig 20C). Again FLAG-tagged SpfP migrated as a ~25 kDa band, the same molecular weight as Crp-HA-His_6, while the band was not detected for the Δ*crp* strain. As a further assay of the SpfP-Crp interaction, we mixed synthetic SpfP peptide biotinylated at the N-terminus with extracts from Crp-HA-His_6 Δ*spf* and Δ*crp* Δ*spf* cells and incubated the extracts with α-HA magnetic beads to isolate Crp-HA-His_6 with associated proteins (Fig 20D). When we probed an immunoblot of this fractionation with streptavidin, we again observed a band that specifically co-migrated with Crp-HA-His_6. Together these results indicate that SpfP associates with Crp and this association is sufficiently tight that the 15 amino acid SpfP protein remains bound to Crp during SDS-PAGE. Interestingly, despite being non-functional in the growth assay, the H10A derivative of the FLAG-SpfP still co-purified with Crp (Fig 20E).

SpfP overexpression blocks activation of some Crp-dependent operons

We hypothesized that SpfP binding to Crp could affect the activity of this transcription regulator. To test this possibility, we examined the effects of pKK-SpfP-scram or pKK-SpfP-$scram_{STOP}$, the stop mutant control, on the levels of proteins encoded by various Crp-activated or -repressed operons. The strains with chromosomally tagged operons were grown with carbon sources that activated the expression of the specific operons. These experiments showed that pKK-SpfP-scram, but not pKK nor pKK-SpfP-$scram_{STOP}$, led to decreased levels of GalM-SPA and GalK-HA-His_6 for cells grown in M63 with galactose (Fig 21A) and decreased levels of MalE-SPA and MalK-SPA for cells grown in M63 with maltose (Fig 21B). Consistent with the reduced

growth defect observed for the SpfP-scram$_{H10A}$ mutant, this derivative did not reduce GalM-SPA

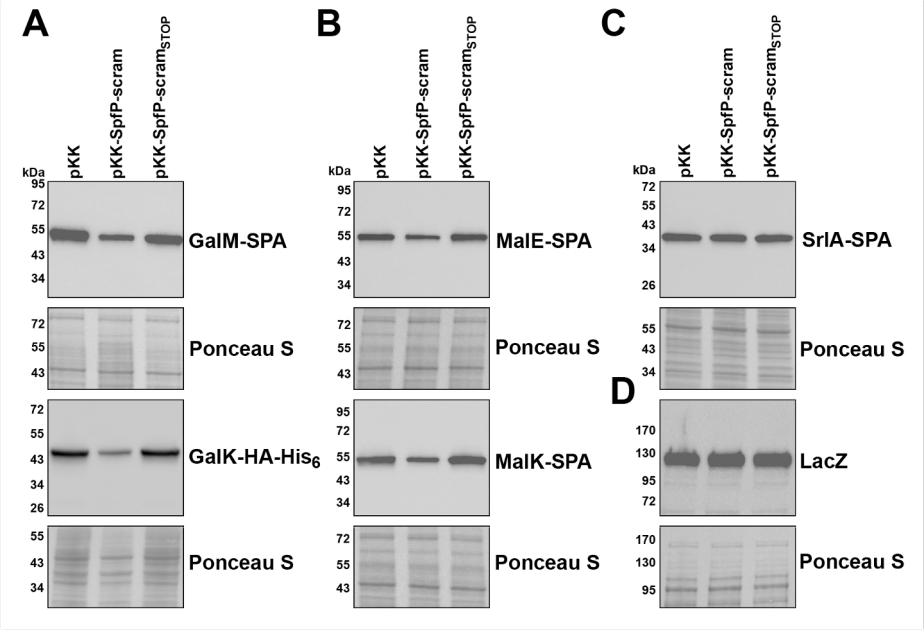


Figure 21. SpfP overexpression leads to downregulation of Crp-activated genes.

A) Immunoblot blot analysis of GalM-SPA and GalK-HA-His$_6$ levels in Δ*spf* strains with chromosomal *galM-SPA::kan* or *galK-HA-His$_6$::kan* (GSO1066 and GSO106, respectively) transformed with pKK, pKK-SpfP-scram, or pKK-SpfP-scram$_{STOP}$. Cells were grown in M63 galactose to OD_{600} ~ 0.5. α-FLAG (top panel) or α-His antibody (bottom panel) were used to detect GalM-SPA or GalK-HA-His$_6$, respectively.

B) Immunoblot blot analysis of MalE-SPA and MalK-SPA levels in Δ*spf* strains with chromosomal *malE-SPA::kan* or *malK-SPA::kan* (GSO1067 and GSO1068, respectively) transformed with pKK, pKK-SpfP-scram, or pKK-SpfP-scram$_{STOP}$. Cells were grown in M63 maltose to OD_{600} ~ 0.5. α-FLAG antibody were used to detect both proteins.

C) Immunoblot blot analysis of SrlA-SPA levels in Δ*spf* strains with chromosomal *srlA-SPA::kan* (GSO449) transformed with pKK, pKK-SpfP-scram, or pKK-SpfP-scram$_{STOP}$. Cells were grown in M63 sorbital to OD_{600} ~ 0.5. α-FLAG antibody was used to detect the protein.

D) Immunoblot blot analysis of LacZ levels in the Δ*spf* strain (GSO1059) transformed with pKK, pKK-SpfP-scram, or pKK-SpfP-scram$_{STOP}$. Cells were grown in M63 lactose to OD_{600} ~ 0.5. α-β-galactosidase antibody was used to detect the protein.

For all panels, the Ponceau S stained membrane documents approximately equal loading of the samples.

and MalE-SPA levels (Fig 22A). In contrast, the levels of SrlA-SPA encoded in another Crp-activated operons were not affected by overexpression of SpfP-scram for cells grown in M63 medium with sorbitol (Fig 21C), nor did SpfP overexpression lead to reduced β-galactosidase levels from *lacZ* for cells grown in M63 with lactose (Fig 21D). We also did not detect differences in the levels of the AzuC-SPA small protein (Fig 22B) whose expression is repressed

by Crp or in the levels of Crp itself (Fig 22C) whose expression is both positively and negatively autoregulated.

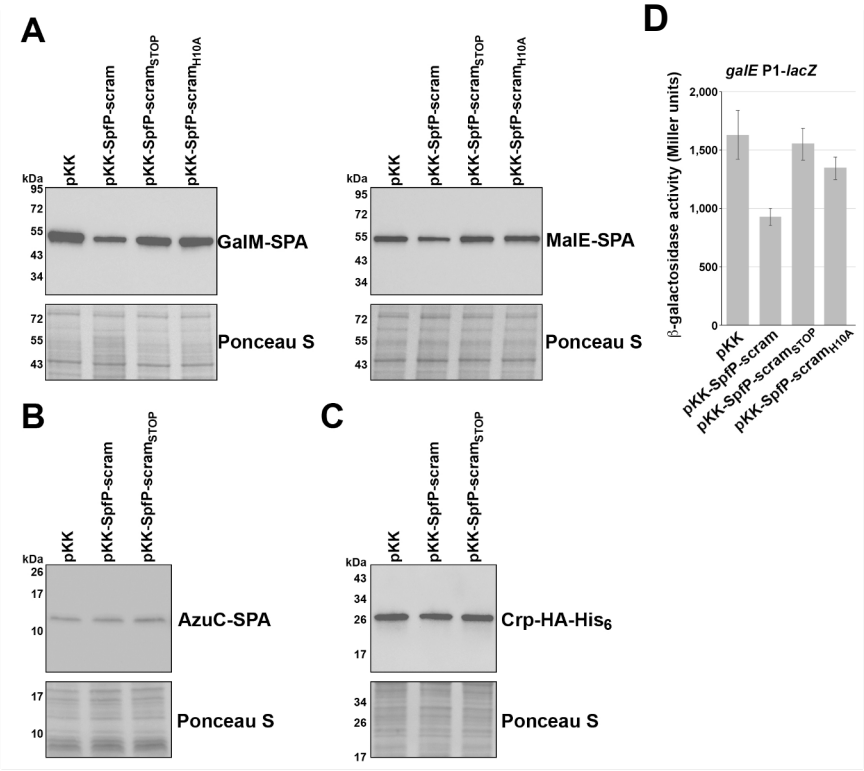


Figure 22. No effect of SpfP overexpression on AzuC or Crp levels and no effect of SpfP$_{H10A}$ overexpression.

A) SpfP-scram$_{H10A}$ is less effective at repressing GalM-SPA and MalE-SPA. Panels show same blots depicted in Fig 4A and 4C but with an additional lane for cells transformed with pKK-SpfP-scram$_{H10A}$.

B) Immunoblot blot analysis of expression of chromosomally tagged AzuC-SPA (GSO351) from cells transformed with pKK, pKK-SpfP-scram, or pKK-SpfP-scram$_{STOP}$ and grown in LB to OD$_{600}$~0.5. α-FLAG antibody was used to detect the SPA tag.

C) Immunoblot blot analysis of expression of chromosomally-tagged Crp-HA-His$_6$ in Δ*spf crp-HA-His$_6$::kan* (GSO1061) cells transformed with pKK, pKK-SpfP-scram, or pKK-SpfP-scram$_{STOP}$. Cells were grown in M63 galactose to OD$_{600}$~0.5. α-His antibody was used to detect Crp.

D) SpfP-scram$_{H10A}$ is less effective at repressing *galE* P1-*lacZ*. β-galactosidase activity assay of *galE* P1-*lacZ* Δ*spf* (GSO1071) cells transformed with pKK, pKK-SpfP-scram, pKK-SpfP-scram$_{STOP}$, and pKK-SpfP-scram$_{H10A}$. β-galactosidase assays were performed for cells grown to OD$_{600}$~0.5 in LB. Bars depict average of three biological replicates and error bars correspond to the standard deviation.

These assays showed that SpfP blocks the activation of some but not all Crp-dependent genes. Since SpfP did not affect all Crp-target genes equally, we hypothesize SpfP does not affect either the cAMP effector binding to Crp or Crp binding to DNA, but rather modulates the Crp interaction with RNA polymerase at specific promoters. Given that *malE* and *galE*, Class I

and Class II-activated genes, respectively, are both repressed by SpfP, while *lacZ* and *srlA*, also Class I and Class II-activated genes, respectively, are both unaffected by SpfP, the regulation is not based on the position of Crp binding relative to RNA polymerase.

SpfP blocks Crp-dependent transcription from the *galE* P_1 promoter

To further test whether SpfP modulated Crp-mediated transcription activation, we assayed a transcriptional fusion of the P1 promoter of the galactose operon to *lacZ* (DM0021) (Geanacopoulos, Vasmatzis et al. 1999) for cells with pKK, pKK-SpfP-scram, or pKK-SpfP-$scram_{STOP}$. This β-galactosidase activity assay showed that SpfP-scram but not SpfP-$scram_{STOP}$ down regulates transcription from the P_1 promoter (Fig 23). SpfP-scram did not repress the levels of β-galactosidase activity for a *galE* P_1-*lacZ* fusion with a C to G mutation at the -35 position in the Crp binding site, though, as expected. β-galactosidase was significantly lower for this derivative. These observations indicate that SpfP affects the ability of Crp to activate transcription of the galactose operon. As seen for GalM-SPA and MalE-SPA, the SpfP-$scram_{H10A}$ mutant did not repress the *galE* P_1-*lacZ* fusion (Fig 22E)

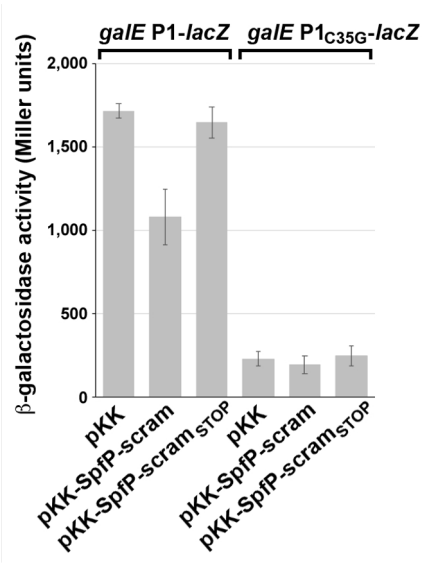


Figure 23. SpfP blocks Crp-dependent transcriptional activation.
β-galactosidase activity assay of transcriptional fusion of the P1 promoter of the galactose operon to *lacZ* (*galE* P1-*lacZ*) (GSO1069) and the same fusion with a C35G mutation (GSO1070). Both fusion strains carrying Δ*spf* (GSO1071 and GSO1072, respectively) were transformed with pKK, pKK-SpfP-scram, and pKK-SpfP-$scram_{STOP}$, and β-galactosidase assays were performed for cells grown to OD_{600}~0.5 in LB. Bars depict average of three biological replicates and error bars correspond to the standard deviation.

Relative importance of Spot 42 base pairing activity and SpfP protein activity varies with temperature

Given the somewhat higher levels of SpfP at 42˚C and 45˚C (Fig 17C), we wondered whether the relative importance of the sRNA and mRNA activities of *spf* varied depending on the growth temperature. To test this possibility, we repeated the β-galactosidase assay of the *srlA-lacZ* fusion for cells grown at 30˚C, 37˚C, and 42˚C. This assay showed that Spot 42 overexpression resulted in less repression at 42˚C compared to the lower temperatures (Fig 24). In contrast,

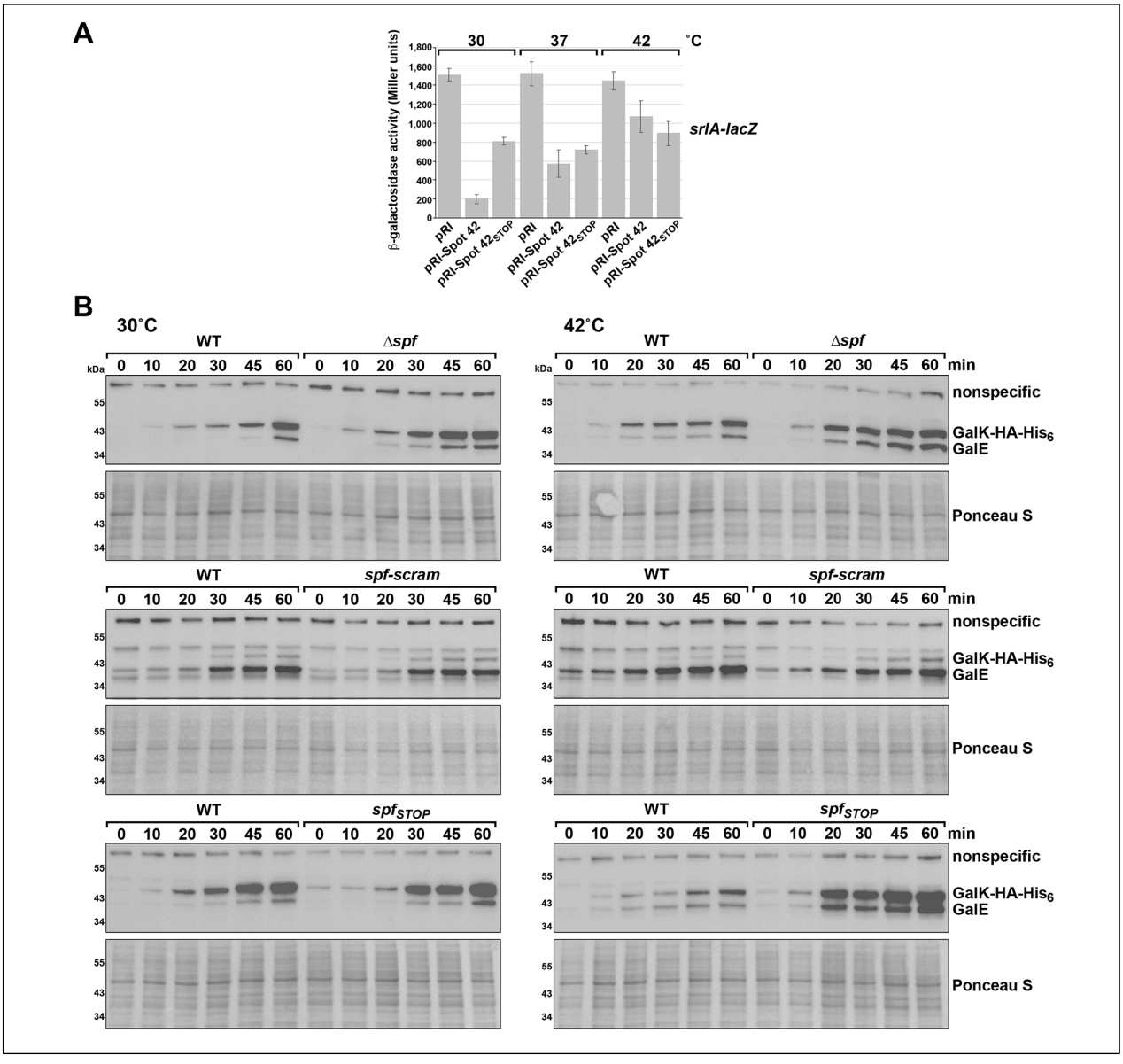


Figure 24. Spot 42 and Spot 42$_{STOP}$ have different effects at different temperatures.

A) β-galactosidase assay of Δ*spf::kan srlA-lacZ* expressing cells (GSO441) transformed with pRI, pRI-Spot 42, and pRI-Spot 42$_{STOP}$. The cells were grown to OD_{600}~1.0 at 30˚C, 37˚C, and 42˚C in LB supplemented with 0.2% arabinose. Bars depict average of three biological replicates and error bars correspond to the standard deviation.

B) Immunoblot analysis of GalE levels in *galK-HA-His$_6$::kan* (GSO1057), Δ*spf galK-HA-His$_6$::kan* (GSO1060), *spf-scram galK-HA-His$_6$::kan* (GSO1077), and *spf$_{STOP}$::kan galK-HA-His$_6$* (GSO1075). Blots in 25C were stripped and probed with α-GalE polyclonal antiserum. The stripping did not completely remove the GalK-HA-His6 signal, and some nonspecific bands are detected with the α-GalE antibodies. The Ponceau S stain documents approximately equal loading of the samples.

Spot 42$_{STOP}$ overexpression led to the same repression at all temperatures resulting in less repression than WT Spot 42 at 30˚C and more repression than WT Spot 42 at 42˚C.

We also examined how the same constructs affected GalK-HA-His$_6$ levels at the different temperatures. Consistent with the β-galactosidase activity assay, we observed that full length Spot 42 was less effective at reducing GalK-HA-His$_6$ levels at 42˚C, while Spot 42$_{STOP}$ overexpression led to strong repression at all temperatures (Fig 25A). Overexpression of SpfP-scram led to similar, but slightly less GalK-HA-His$_6$ repression at all temperatures, while SpfP-scram$_{STOP}$ had no effect on GalK-HA-His$_6$ levels (Fig 25B). We interpret these results to mean that Spot 42 levels are higher or base pairing is more effective at repression at lower temperatures and that translation of Spot 42 can interfere with base pairing activity at higher temperatures.

Previous studies showed that Spot 42-repressed targets were induced more rapidly in Δ*spf* cells compared to WT cells upon a shift from glucose to a non-preferred carbon, source such that cells were primed to utilize glucose for as long as possible (Beisel and Storz 2011). We repeated this experiment with a shift from glucose to galactose at 30˚C and 42˚C and similarly observed a more rapid induction of GalK-HA-His$_6$ (Fig 25C) and GalE (Fig 24B) in the Δ*spf* background when comparing the same point between the WT and Δ*spf* strain at both temperatures. A similar comparison between WT cells and cells in which the chromosomal version of *spf* carried the scrambled sequence (*spf-scram*), and thus only expressed SpfP, showed less of a difference

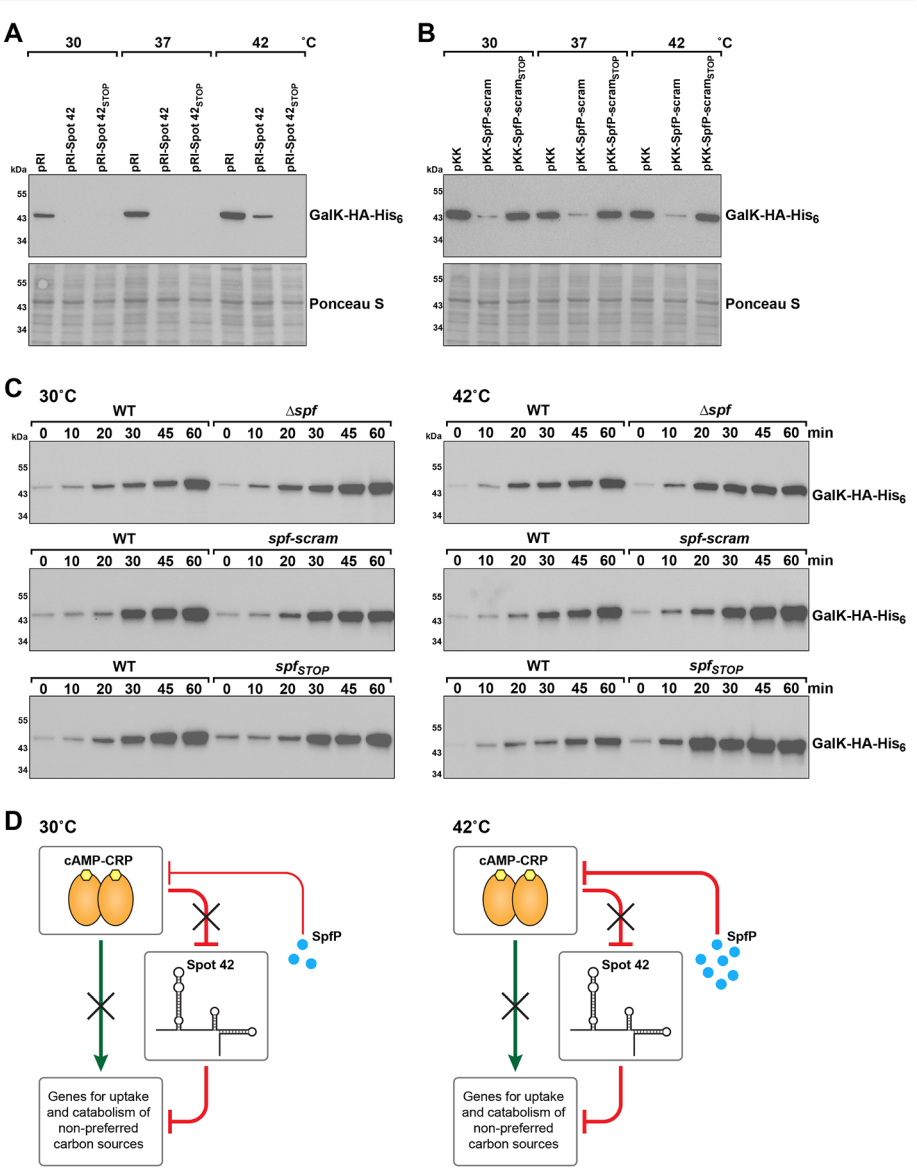


Figure 25. Spot 42 and SpfP have different effects at different temperatures.

A) Immunoblot analysis of GalK-HA-His$_6$ levels in a Δ*spf galK-HA-His$_6$::kan* strain (GSO1060) transformed with pRI, pRI-Spot 42, and pRI-Spot 42$_{STOP}$ and grown in LB at 30˚C, 37˚C, and 42˚C.

B) Immunoblot analysis of GalK-HA-His$_6$ levels in a Δ*spf galK-HA-His$_6$::kan* strain (GSO1060) transformed with pKK, pKK-SpfP-scram, and pKK-SpfP-scram$_{STOP}$ and grown in LB at 30˚C, 37˚C, and 42˚C.

For (A) and (B), samples were collected at OD$_{600}$~0.5. α-His antibody was used to detect the HA-His$_6$ tag. The Ponceau S stained membrane documents approximately equal loading of the samples.

C) Immunoblot analysis of GalK-HA-His$_6$ levels in *galK-HA-His$_6$::kan* (GSO1057), Δ*spf galK-HA-His$_6$::kan* (GSO1060), *spf-scram galK-HA-His$_6$::kan* (GSO1077), and *spf$_{STOP}$::kan galK-HA-His$_6$* (GSO1075) grown in M63 glucose to OD$_{600}$~0.4 at 30˚C or 42˚C, collected and then resuspended in M63 galactose at 30˚C or 42˚C, respectively with samples collected at the indicated times.

D) The small protein SpfP reinforces the multi-output feedforward loop between Crp and Spot 42. For cells grown in the absence of glucose, Crp directly increases transcription of targets and represses Spot 42. In medium with glucose, Spot 42 represses Crp-activated targets through base pairing and the small protein SpfP blocks Crp-dependent activation, particularly at high temperature.

indicating that the small protein is functioning to repress transcription at 30˚C and 42˚C. It is worth noting that due to the scramble sequence, the *spfP* ribosome binding site and start codon likely are no longer in a stem-loop structure, probably allowing equal translation at all temperatures. When the chromosomal version of *spf* had the same stop codon mutation as in Spot 42_{STOP}, we also observed very little difference between the WT and spf_{STOP} strain at 30˚C, probably due to Spot 42_{STOP} function as an sRNA. However, we detected more rapid induction for the spf_{STOP} strain compared to WT at 42˚C. This observation shows that the lack of small protein expression from the chromosome leads to less repression at high temperature.

Discussion

While bacterial cells encounter environments with mixed nutrients, they selectively metabolize the most energetically favorable carbon sources in order to maximize growth. To accomplish this, bacteria employ transcription factors, regulatory sRNAs, as well as small proteins to control gene expression to specifically synthesize the appropriate transporters and enzymes. In this study, we have demonstrated that the base pairing sRNA Spot 42 is also translated to produce a 15 amino acid protein, SpfP, that plays a role in catabolite repression. Previous work showed that the Crp-repressed Spot 42 RNA represses numerous targets, including *galK*, *srlA*, *fucI,* and *nanC*, involved in secondary metabolism and the uptake and use of non-preferred carbon sources (Beisel and Storz 2011, Beisel, Updegrove et al. 2012). Through copurification analysis we showed that SpfP interacts with Crp. We suggest that this interaction functions to block Crp-mediated activation of gene expression. Our discovery that Spot 42 can act as an mRNA demonstrates that this sRNA is a dual-function sRNA whose base pairing and small protein both impact carbon source utilization.

Impact of dual-function RNAs on carbon metabolism

The findings reported here add another layer to the Spot 42 modulation of catabolite repression. It was already shown that Spot 42 coordinates with Crp in a multioutput feedforward loop to repress genes from auxiliary metabolic pathways when the preferred carbon source glucose is available delaying induction of these genes until the preferred carbon source disappears and thus prioritizing use of glucose as the most effective carbon source (Beisel and Storz 2011). As shown in Fig 25D, we suggest that SpfP reinforces this regulatory loop by directly blocking Crp-mediated activation when cells are growing in glucose, particularly at high temperatures.

Interestingly, we found the regions of Spot 42 involved in base pairing overlap the SpfP coding sequence raising the question of whether the sRNA or mRNA activity predominates under specific conditions. Given the potential to open the secondary structure overlapping the SpfP ribosome binding site and the slightly higher levels of SpfP at 42˚C, we hypothesized that SpfP-mediated effects may be more predominant at higher temperatures when base pairing may be less efficient. Consistent with this hypothesis, we found Spot 42 repression of the *srlA-lacZ* translation fusion was less effective at higher temperatures. If Spot 42 base pairing activity is decreased at temperatures where SpfP translation increases, perhaps SpfP continues the regulatory role of Spot 42 in carbon metabolism when Spot 42 is no longer functioning as effectivity as an sRNA.

Given that only a handful of dual-function sRNAs have been characterized in bacteria, it is noteworthy that all of the dual-function sRNAs studied thus far in *E. coli* impact carbon utilization. The SgrST sRNA is specifically activated by the SgrR transcription factor when cells encounter glucose phosphate stress (Vanderpool and Gottesman 2004). As with Spot 42, where both the base pairing function and small protein target the same pathway, the sRNA activity and small protein encoded by SgrS both regulate glucose phosphate utilization metabolism by targeting the PtsG transporter (Vanderpool and Gottesman 2004, Lloyd, Park et al. 2017). In a parallel study, we characterized another dual-function sRNA with connections to carbon metabolism. This 164-nt RNA, AzuCR, whose transcription also is repressed by Crp, base pairs with *galE*, a gene that contributes to galactose metabolism in *E. coli* (Raina, Aoyama et al. 2020). The RNA is also translated to give the 28 amino acid small protein AzuC which binds to and increases the activity of GlpD, a protein essential for glycerol catabolism during growth in stationary phase. Thus, in contrast to SgrS-SgrT and Spot 42-SpfP, AzuR-AzuC modulates

different aspects of metabolism and glycerol metabolism. While the small protein and base pairing activities are encoded in separate parts of the 227-nt long SgrST RNA, for both *spfP* and *azuC* the ORF overlaps the base pairing region, leading to competition between the activities of these dual-function sRNAs. The perceived enrichment for dual-function sRNAs that modulate carbon metabolism may be a reflection of the fact that they are relatively abundant and thus readily detected. Alternatively, the strong selection imposed by metabolism may be a driving force for the evolution of dual-function sRNA regulators.

SpfP interaction with Crp

Another interesting question is the mechanism of how SpfP impacts Crp activity, a global transcription factor with ~300 regulated genes (Shimada, Fujita et al. 2011). The functions of only a limited number of small proteins less than 50 amino acids have been identified in *E. coli*. The majority of these characterized small proteins have been found to be localized to the inner membrane (reviewed in (Hemm, Weaver et al. 2020)). At 15 amino acids, SpfP is among the smallest proteins to be characterized. We suggest the hydrophobic protein acts as a small molecular inhibitor to affect the ability of Crp to bind RNA polymerase, thus blocking Crp-dependent transcription activation at specific promoters. The mutational studies indicate that the H10 residue of SpfP is needed for this activity, which is interesting given that two of the activation regions (AR1 and AR2) of Crp have histidine residues required for the interaction with RNA polymerase (Evangelista, Dong et al. 2019). Given how extensively Crp has been studied, it is intriguing that SpfP was not detected previously. Fortunately, the existing tools and approaches to study Crp should help facilitate further mechanistic studies.

SpfP is also unique among the small proteins characterized in *E. coli* because it binds and modulates a transcription factor. Intriguingly, a few other very small proteins or peptides have been found to modulate transcription factor activity. For example, the small signaling peptides of 5-10 amino acids that are secreted by *Bacillus* and *Streptococcus* are recognized by sensory receptors from the RRNPP protein family (named for the representative proteins Rap, Rgg, NprR, PlcR, and PrgX) (reviewed in (Neiditch, Capodagli et al. 2017)). Binding of these peptides to the RRNPP proteins induces conformational changes that favor oligomerization, DNA binding and transcriptional activation of target genes. Similarly, the 6-10 amino acid Arbitrium peptide released by phages modulates the lysis-lysogeny decision by interacting with the AimR transcription factor (Erez, Steinberger-Levy et al. 2017). We predict that other very small proteins likely will be found to target other transcription factors and possibly other components of the transcriptional machinery. It is worth considering whether derivatives of these small proteins, which are easy to synthesize, could be exploited to control bacterial cell growth in specific ways.

Material and Methods

Bacterial strains and plasmids

Bacterial strains, plasmids, and oligonucleotides used in this study are listed in Appendix Tables. E. coli strains are derivatives of wild-type MG1655 (F- lambda- ilvG- rfb-50 rph-1). Tagged strains were generated by λ Red–mediated recombineering (Yu, Ellis et al. 2000) using NM400 and the oligonucleotides listed in the Appendix**.** Alleles marked by antibiotic markers were moved between strains by P1 transduction. When necessary, kanamycin resistance cassettes were excised from the chromosome by FLP-mediated recombination (Cherepanov and Wackernagel 1995). All mutations and fusions were confirmed by sequencing. Descriptions of plasmids used in this study are listed in the Appendix. All SpfP derivatives were overexpressed from pKK177-3 (a derivative of pKK223-3 lacking the region between at BamHI at 256 and PvuII at 1,945), while Spot 42 WT and the Spot 42_{STOP} mutant were overexpressed from the pRI derivative of pKK177-3 in which an EcoRI site was introduced at the P_{tac} +1 site (Opdyke, Kang et al. 2004). The Spot 42 and SpfP derivatives were ordered as primers listed in Table S3, annealed and cloned into pKK177-3 or pRI digested with the EcoRI and HindIII restriction enzymes. All inserts were confirmed by sequencing.

Bacterial growth

Unless indicated otherwise, bacterial strains were grown with shaking at 250 rpm at 37˚C in either LB rich or M63 minimal medium supplemented with 0.001% vitamin B1 and glucose, galactose or maltose at 0.2%, and sorbitol at 0.4%. Where indicated, media contained antibiotics with the following concentrations: ampicillin (100 μg/ml), chloramphenicol (25 μg/ml), and kanamycin (30 μg/ml).

Immunoblot analysis

Aliquots (1 ml) of cells were grown in the indicated medium were collected and the cell pellet was resuspended in 1XPBS and 10 ul were loaded on a Mini-PROTEAN TGX 5%–20% Tris-Glycine gel (Bio-Rad) and run in 1X Tris Glycine SDS buffer (KD biomedical). The gel was electro transferred into nitrocellulose membranes (Invitrogen) for 1 h at 100 volts. Membranes were blocked with 5% non-fat milk (Bio-rad,) in 1X PBS 0.1% of Tween 20 (PBS-T) for 1 h and probed with a 1:3,000 dilution of a-FLAG antibody (Sigma), 1:1,000 dilution of α-His antibody (Qiagen), 1:1,000 dilution of α-β-galactosidase antibody (Abcam) or 1:1,000 dilution of a-GalE antibody (kind gift of S. Adhya) in PBS-T for 1 h. Incubation with the a-GalE antibodies was followed by an incubation with a 1:20,000 dilution of peroxidase-labelled α-rabbit antibodies. After three washes with PBS-T, the membrane was developed with an Amersham ECL Western Blotting Detection Kit (GE Healthcare).

RNA isolation

Cells corresponding to the equivalent of 10 OD_{600} were collected by centrifugation, and snap frozen in liquid nitrogen. RNA was extracted according to the standard TRIzol (Thermo Fisher Scientific) protocol. Briefly, 1 ml of room temperature TRIzol was added to cell pellets, resuspended thoroughly to homogenization, and incubated for 5 min at room temperature. After the addition of 200 µl of chloroform and thorough mixing by inversion, samples were incubated for 10 min at room temperature. After samples were centrifuged for 10 min at 4˚C on maximal

speed, the upper phase (~0.6 ml) was transferred into a new tube and 500 µl of isopropanol was added. Samples again were mixed thoroughly by inversion, incubated for 10 min at room temperature and centrifuged at maximal speed for 15 min at 4˚C. RNA pellets were washed twice with 75% ethanol. After the second wash, the ethanol was aspirated, and the RNA pellet was left to dry at room temperature. RNA was resuspended in 20-50 µl of DEPC water and quantified using a NanoDrop (Thermo Fisher Scientific).

Northern analysis

Total RNA (5- 10 µg per lane) was separated on denaturing 8% polyacrylamide gels, urea gels containing 6 M urea (1:4 mix of Ureagel Complete to Ureagel-8 (National Diagnostics) with 0.08% ammonium persulfate) in 1X TBE buffer at 300V for 90 min. The RNA was transferred to a Zeta-Probe GT membrane (Bio-Rad) at 20 V for 16 h in 0.5X TBE, UV-crosslinked, and probed with 32P-labeled oligonucleotide (Listed in Appendix Tables) in ULTRAhyb-Oligo buffer (Ambion Inc.) at 45˚C. Membranes were washed twice with 2X SSC/0.1% SDS at room temperature, once with 0.2X SSC/0.1% SDS at room temperature, washed for 25 min with 0.2 × SSC/0.1% SDS at 45˚C, followed by a final rinse with 0.2X SSC/0.1% SDS at room temperature before autoradiography was performed with HyBlot CL film (Denville Scientific Inc.).

Growth assay

Colonies of Δ*spf::kan* (GSO433) transformed with indicated plasmids were grown overnight in LB with ampicillin and diluted to OD_{600}=0.05 in M63 minimal medium supplemented with 0.001% vitamin B1 and either 0.2% glucose or 0.2% galactose. OD_{600} was measured after 16h.

Purification of scrambled SpfP

The Δ*spf::kan* strain (GSO433) transformed with pKK-SpfP-$scram_{N\text{-terminal tag}}$ or pKK-SpfP-$scram_{C\text{-terminal tag}}$ and pEVOL-*p*-AzF was grown in LB at 37˚C overnight. The overnight culture was diluted 1:100 into 1 l of M63 glucose minimal media and incubated at 37˚C. At OD_{600} ~0.5, 0.2% arabinose was added to induce the orthologous aaRS and tRNA pair along with 1 mM *p*-azido phenylalanine (Chem-Impex Int'l. Inc.) and incubated for another 3 h. Cells were collected by centrifugation (4,650 × g, 20 min) and the pellet was resuspended in 30 ml of TNG buffer [10 mM Tris (pH 7.5), 100 mM NaCl, 10% glycerol] supplemented with Protease Inhibitor Cocktail (Roche). The cells were lysed using a microfluidizer processor (Microfluidics) at 20,000 psi, and the insoluble cellular debris was removed by centrifugation (20,000 × g, 30 min). The cleared lysate was incubated with 5 µM $CuCl_2$, 1 mM ascorbic acid (pH adjusted to 7 by NaOH) and 5 µM of Biotin-PEG4-Alkyne (Click Chemistry Tools) dissolved in DMSO at room temperature by nutating for 30 min. The lysate was then supplemented with 25 mM DDM and incubated overnight at 4˚C with 500 ul of streptavidin agarose beads (Thermo Scientific). The lysate and beads were applied to a Bio-Spin disposable chromatography column (Bio-Rad Laboratories) and allowed to drain by gravity. The streptavidin column was washed with 15 ml TNG buffer with 2 mM DDM. Finally, proteins were eluted from the column with 1 ml of elution buffer (0.1 M glycine pH 3.5, 100 mM NaCl and 0.1% Triton X-100) and TCA precipitation was carried out

to concentrate the protein samples. The pellet was resuspended in 30 µl of 2X Laemmli buffer. The samples were heated at 95˚C for 10 min, and aliquots were subjected to SDS/PAGE in a 10–20% Tris-glycine gel (Invitrogen) at 12 V/cm. Proteins were visualized with Coomassie Blue Stain. Bands of interest were excised from the gel and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS).

Reciprocal purification

Δ*spf* cells expressing Crp-HA-His_6 or GalK-HA-His_6 from the chromosome (GSO1061 and GSO1060, respectively) or Δ*spf::kan* Δ*crp::cm* cells GSO1063 were transformed with pKK-SpfP-$scram_{N\text{-terminal FLAG}}$ and grown in LB at 37˚C for 16h, diluted 1:100 into 1L of LB and incubated at 37˚C. At ~OD=0.5 cells were collected by centrifugation (10,000 x g, 20min) and resuspended in 15 ml of TNG buffer supplemented with Protease Inhibitor cocktail (Roche). The cells were homogenized and incubated with Pierce α-HA magnetic beads (Thermo Scientific) in a 50 ml tube and incubated overnight at 4˚C. Beads and the bound protein were collected using a MagneSphere magnetic separation stand (Promega) and washed with 1 ml of TNG buffer (10X). The beads were resuspended in 1X PBS (25 ul) and 2X Laemmli buffer (25 ul) and heated at 95˚C for 5 min. Samples were analyzed on immunoblots using α-His and α-FLAG antibodies. Synthetic N-terminally biotin tagged SpfP (produced by Thermo Scientific) was added to lysates of Δ*spf* cells expressing Crp-HA-His_6 or Δ*spf* Δ*crp* cells prior to incubation with Pierce α-HA magnetic beads (Thermo Scientific).

β-galactosidase assays

Cultures were grown in LB to OD_{600}~1.0 with arabinose (0.2%). 100 ul of cells were added to 700 ul of Z buffer (60 mM Na2HPO4z7H2O, 40 mM NaH2PO4zH2O, 10 mM KCl, 1 mMMgSO4z7H2O, 50 mM β-mercaptoethanol). After adding 15 ul of freshly prepared 0.1% SDS and 30 ul of chloroform to each sample the cells were vortexed for 30 s and then incubated at room temperature for 15 m to lyse the cells. At zero time, the assay is initiated by adding 100 ul of ONPG (4 mg/ml) to each sample in 10 s intervals. The samples are incubated at room temperature before the reaction is terminated by the addition of 500 ul of 1M Na_2CO_3. Then, the A420 and A550 values are determined with a spectrophotometer and the absorbance data is transferred to a Microsoft Excel spreadsheet to calculate Miller units.

Addendum

While expanding the known function of the well-studied Spot 42, this study illuminates some interesting avenues for additional research. I think the most interesting question is the mechanism by which SpfP affects Crp. SpfP does not alter Crp levels, so SpfP must be affecting how Crp is acting

One possibility is that SpfP is altering the effects of Crp on transcription. My work on Spot 42 suggests that SpfP interferes with activation of some, but not all, Crp regulated genes. Using an *in vitro* transcription assay with purified RNAP, Crp, and the promoter DNA from differentially regulated genes, *lacZ* and *galE*, as template I could assess if SpfP is modulating how Crp increases RNAP transcription. This assay could also be used to determine how the H10A mutant is impacting the function of SpfP.

My co-purification experiments have demonstrated that SpfP interacts with Crp, but the actual nature of this interaction has not been uncovered. I could gain further insights into the interaction by characterizing additional SpfP and Crp mutants to define the interaction surfaces. One method is to take advantage of the growth phenotypes on nonpreferred carbon sources when SpfP is overexpressed. I would use error prone PCR on a plasmid expressing Crp in a Δ*crp* background and grow cells on galactose while simultaneously overexpressing SpfP. Mutants would suppress the growth defect, and I could sequence both SpfP and Crp from their respective plasmids to attempt to identify mutations that are important for the functional interaction between these proteins. I could also use chemical mutagenesis to achieve a similar result. Additionally, I could carefully purify Crp and SpfP and utilize nuclear magnetic resonance (NMR) to assess the structural interaction between the two proteins. This would identify how SpfP is interacting with Crp. As Crp has been well studied and the structure has been solved,

visualizing the interaction with SpfP could also provide important insights into how SpfP affects Crp.

Finally, I have demonstrated that temperature is important for modulating the regulatory activities of Spot 42. Using structure probing experiments such as enzymatic cleavage we could examine how temperature is able to disrupt base pairing interactions. It would also be interesting to search for other methods by which the Spot 42 structure could be resolved to permit translation of the SpfP ORF. I would like to MS2 tag the Spot 42 RNA and then conduct an immunoprecipitation in an attempt to identify other proteins that may be interacting with the Spot 42 RNA to increase translation of SpfP. I would expect to identify proteins and RNA molecules that are capable of altering the Spot 42 structure to permit SpfP translation under conditions where temperature is not responsible for the same effect. We could then repeat the structure probing experiment suggested above to test how these proteins and RNAs are altering Spot 42 structure.

The finding that Spot 42 encodes a 15 amino acid protein has added a new layer of regulation to a well-studied network controlled by an RNA regulator. The translation of the small protein SpfP reinforces the effects of Spot 42 on carbon catabolite repression. The identification of SpfP has provided an interesting opportunity to study how regulation by Spot 42 changes under conditions where translation of the small protein is more prevalent.

Chapter 4

Critical features of a synthetic dual-function RNA

Abstract

Small base pairing RNAs (sRNAs) and small proteins have emerged as components of numerous regulatory networks. Both of these classes of regulators allow bacterial cells to rapidly adapt to a wide variety of growth conditions through post-transcriptional regulation. Dual-function RNAs regulate mRNA expression by base pairing and also encode a small protein. Thus, the products of a single gene can regulate targets at the level of mRNA stability or translation as well as at the level of protein activity or stability. While an increasing number of these molecules are being identified and characterized, little is known about the interplay between their two regulatory functions. To investigate the competition between the two activities of dual-function RNAs, we constructed hybrid RNAs encoding the *Escherichia coli* sRNA MgrR and small protein MgtS, hereafter referred to as MgtSR or MgtRS. Expression of MgrR and MgtS is normally induced by low magnesium (Mg^{2+}) by the PhoQP two-component system. MgrR is a 98-nt sRNA that negatively regulates *ygdQ* (unknown function) and *eptB* (phosphoethanolamine transferase). MgtS is a 31 amino acid small inner membrane protein that is required for the accumulation of MgtA, the Mg^{2+} transporter. By generating various versions of this synthetic dual-function RNA, we probed how the organization of components is important for proper function of both activities of a dual-function RNA and how the organization affects competition between the two activities. By understanding these features of synthetic and natural dual-function sRNA, future synthetic molecules can be constructed to maximize their regulatory impact.

Introduction

Small RNAs (sRNAs) play an important role in shaping and regulating the cellular responses to changing environmental conditions. Hundreds of sRNAs have been identified through various deep sequencing methodologies. Most of these 50-300 nucleotide molecules regulate target mRNAs through base pairing (reviewed in (Storz, Vogel et al. 2011) (Adams and Storz 2020)). In many bacteria, sRNAs depend on the RNA binding protein Hfq (host factor for bacteriophage Qβ) for stability and base pairing with target mRNAs (Updegrove, Zhang et al. 2016). The functional outcomes of sRNA base pairing with the target mRNA are an inhibition of translation, a change in mRNA degradation or an increase in translation.

While generally assumed to be noncoding, some sRNAs have been found to contain small ORFs and for a small subset of the RNAs, this ORF has been shown to be translated. These sRNAs capable of base pairing regulation while also encoding a protein are called dual-function RNAs (reviewed in (Raina, King et al. 2018)). The limited number of dual-function RNAs that have been identified and characterized is because they tend to be overlooked due to the difficulty in distinguishing a translated ORF from small ORFs that are randomly present in the genome. Furthermore, traditional biochemical methods for identifying and purifying proteins commonly miss proteins that are 10-50 amino acids (Storz, Wolf et al. 2014).

Despite the difficulties inherent in the study of dual-function RNAs, several *E. coli* dual function sRNAs have been examined in sufficient depth to clearly describe both the base pairing and protein coding activities: SgrST, Spot 42-SpfP and AzuCR. All three of these RNAs are involved in regulating carbon metabolism.

The 227-nt SgrS sRNA was originally identified in a computational screen for sRNAs in *E. coli* (Wassarman, Repoila et al. 2001). Subsequently, SgrS was studied for the role it plays in

the cellular response to glucose-phosphate stress and found to base pair with the *ptsG* and *manXYZ* mRNAs encoding important sugar transporters (Rice and Vanderpool 2011). In the course of these studies, it was found that SgrS is translated to give the 43 amino acid small protein SgrT, which inhibits the transport activity of the glucose permease PtsG (Wadler and Vanderpool 2007). Together SgrS and SgrT counteract the accumulation of sugar phosphates in the cell by reducing the translation and activity of sugar transporters and increasing dephosphorylation of sugar phosphates.

The 109-nt Spot 42 RNA identified in 1973 is one of the first sRNA described (Ikemura and Dahlberg 1973) and represses the synthesis of many transporters and enzymes for the utilization of non-preferred carbon sources. Many of the mRNA targets repressed by Spot 42 are activated by the transcription factor cAMP receptor protein (Crp) (Beisel and Storz 2011). Recently, it was shown that Spot 42 also encodes a small ORF that is translated to give a 15 amino acid protein, SpfP. Translation of SpfP is increased following heat shock, and the small protein has been demonstrated to interact with Crp to disrupt transcriptional activation of some Crp targets. This additional function means Spot 42 is a dual-function RNA with the small protein reinforcing Spot 42 RNA control over carbon catabolite repression (CCR).

Another recently identified dual-function RNA is AzuCR. This 164-nt RNA was initially identified in a bioinformatic search for novel sRNA genes but then was reclassified as an mRNA when it was found to encode a 28 amino acid protein, AzuC (Chen, Lesnik et al. 2002, Hemm, Paul et al. 2010). AzuC, interacts with aerobic glycerol 3-phosphate dehydrogenase (GlpD) to increase its activity. However, overexpression of AzuCR produces growth defects in galactose and glycerol medium that persist even when an AzuCR derivative carrying a stop codon mutation in the small ORF is expressed, suggesting that AzuCR also functions as an sRNA.

Consistent with this observation, AzuCR is capable of direct base pairing regulation of *cadA*, lysine decarboxylase, and *galE*, a gene involved in galactose metabolism. Together, the sRNA and mRNA activities contribute to the modulation of carbon utilization in *E. coli*.

One question that arises for dual-function RNAs is whether the two functions can be carried out simultaneously by the same molecule. The base pairing region of SgrS is 15-nucleotides downstream of the SgrT ORF. Mutations to inhibit *sgrT* translation do not impede SgrS regulation of mRNA targets, but mutations that restrict base-pairing increase SgrT translation indicating that the base-pairing function interferes with translation. Furthermore, translation of the small protein lags behind transcription of SgrS RNA by approximately 30 minutes (Balasubramanian and Vanderpool 2013). Based on these observations, it was suggested that the base pairing function predominates, and SgrST molecules that base pair with target mRNAs are unavailable for use as an mRNA as they are degraded. Once the pool of targeted mRNA has been sufficiently degraded, SgrST accumulates and is translated.

For Spot 42 and AzuCR, there is direct overlap between the base pairing region and small ORF such that one function necessarily interferes with the other. In both of these dual-function RNAs, the introduction of a stop codon, which presumably reduces ribosomal occupancy of the transcript increases base pairing activity. For Spot 42, the start of the encoded ORF is sequestered in secondary structure, and translation of the small protein increases at high temperatures that may partially resolve the secondary structure and may coincide with a temperature-dependent reduction in base pairing regulation. For AzuCR, translation of the small protein is repressed by the anaerobic-growth induced sRNA FnrS suggesting sRNAs are another mechanism by which the competition between the base pairing and protein coding activities can be controlled.

To further explore the features of dual-function RNAs that impact what activity predominates, we chose to construct synthetic dual-function RNAs using parts of the transcripts encoding the sRNA MgrR and small protein MgtS. The sRNA component, MgrR, was identified through Hfq co-immunoprecipitation followed by genome-wide RNA detection on microarrays. The sRNA negatively regulates *eptB*, a lipopolysaccharide (LPS) modifying enzyme, and *ygdQ*, a protein of unknown function (Moon and Gottesman 2009). The small protein MgtS was first identified as a potential ORF using microarray analysis (Wassarman, Repoila et al. 2001). Further study confirmed that MgtS is a 31 amino acid small protein that interacts with MgtA, a P-type ATPase Mg^{2+} importer (Hemm, Paul et al. 2008, Wang, Yin et al. 2017). This interaction with MgtA serves to increase levels of the importer and increase intracellular concentration of Mg^{2+}. Expression of both components, MgrR and MgtS, is dependent on the two-component system, PhoQ/PhoP. In low Mg^{2+} conditions, the sensor kinase PhoQ becomes activated. Once active, PhoQ phosphorylates PhoP which in turn activates transcription of a number of different genes including Mg^{2+} transporters and proteins that modify the LPS (Prost and Miller 2008).

In this study, we demonstrate that it is possible to create a synthetic dual-function RNA from existing separate components. Additionally, we discover that the composition of the sequence between the functional elements impact their activities. Interestingly, despite construction of a functional dual function MgtSR RNA from two separate genes, reversing the order of *mgrR* and *mgtS* did not produce a functional regulator.

Results

Construction of a synthetic dual-function RNA

We set out to design a synthetic dual-function RNA built from two components that could be easily tested for their regulatory activity. We chose MgtS as the small protein component and MgrR as the sRNA component. The corresponding two genes are convergently transcribed in a PhoPQ-dependent manner in response to low Mg^{2+} conditions (Fig 26A). The design of this first synthetic dual-function RNA denoted MgtSR was modeled after the characterized dual-function RNA, SgrS, whose small ORF precedes the base pairing region by 15-nt (Raina and Storz 2017). Our design included the 96-nt of the MgtS ORF, from the start codon to the stop codon,

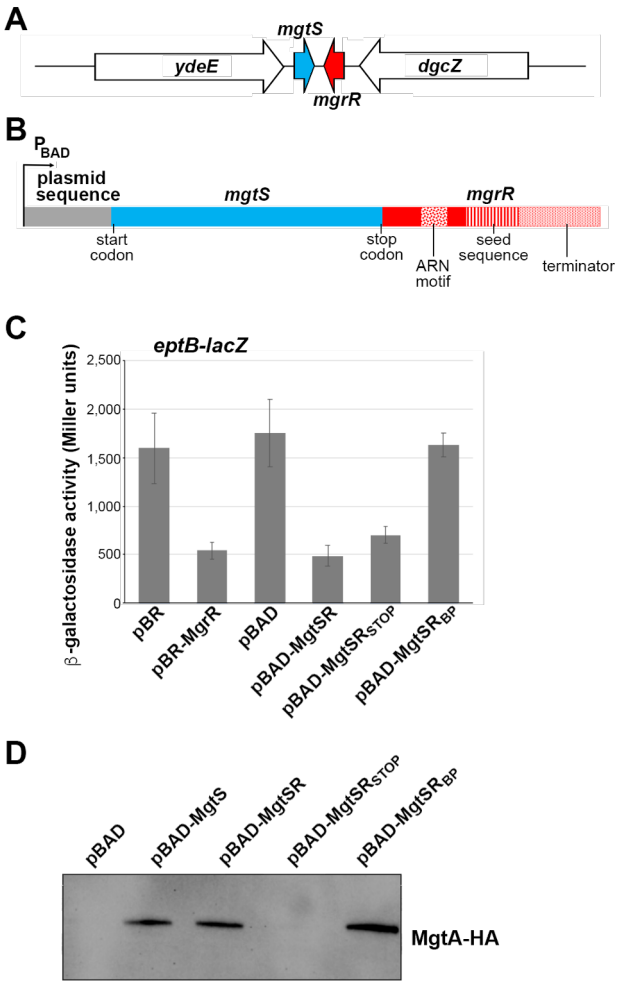


Figure 26. Regulatory activity of the synthetic dual-function RNA.

A) Diagram of the endogenous location of *mgtS* (blue) and *mgrR* (red).

B) Schematic of the hybrid dual-function RNA, MgtSR. Plasmid sequences are in gray, *mgtS* sequences are in blue and *mgrR* sequences are in red. The ARN, seed, and terminator sequences are indicated by distinct patterns as indicated.

C) Effect of pBR, pBR-MgrR, pBAD, pBAD-MgtSR, pBAD-$MgtSR_{STOP}$, and pBAD-$MgtSR_{BP}$ overexpression on an *eptB-lacZ* fusion in Δ*mgrR* background. Data represents triplicate biological replicates with error bars of standard deviation.

D) Immunoblot analysis of MgtA-HA cells in a Δ*mgtS* transformed with pBAD, pBAD-MgtSR, pBAD-$MgtSR_{STOP}$, and pBAD-$MgtSR_{BP}$ grown to OD_{600} 0.4 in N-media supplemented with 500 uM Mg^{2+}, washed, and then resuspended and grown in N-media supplemented with arabinose but without Mg^{2+}. Samples collected after 20 minutes and visualized with α-HA antibodies to detect MgtA-HA.

immediately followed by 77-nt of native *mgrR* sequence. The pBAD24 plasmid carrying the construct provided the ribosome binding site (RBS) for MgtS. The sequence from *mgrR* includes a 9-nt Hfq A-R-N (where A = adenine, R = adenine or guanine, N = any nucleotide) binding motif, the seed sequence for base pairing interactions with MgrR targets, and an intrinsic terminator with a polyU stretch also bound by Hfq (Updegrove, Zhang et al. 2016, Kwiatkowska, Wroblewska et al. 2018). Thus, the MgtS ORF and the seed sequence of MgrR are separated by 30-nt of native sequence from MgrR (Fig 26B). The hybrid gene was cloned behind an arabinose inducible promoter to allow for controllable expression.

We also designed controls to be able to specifically eliminate the activity of one component or the other. To this end we constructed a stop codon control, pBAD-MgtSR$_{STOP}$, to prevent translation of the small protein component while maintaining the integrity of the base pairing region, and pBAD-MgtSR$_{BP}$, a construct with base pairing mutations to prevent sRNA activity while maintaining translation of the small protein.

MgtSR is a functional dual-function RNA

To establish whether each component of MgtSR maintained regulatory activity, we first assayed the effect of MgtSR overexpression on expression of an *eptB-lacZ* fusion. β-galactosidase activity was assayed in Δ*mgrR* cells expressing *eptB-lacZ* translational fusions transformed with pBR322, pBR-MgrR, pBAD, pBAD-MgtSR, pBAD-MgtSR$_{STOP}$, and pBAD-MgtSR$_{BP}$. Regulation of *eptB-lacZ* was observed upon overexpression of MgtSR to a similar extent as overexpression of MgrR (Fig 26C). Furthermore, the pBAD-MgtSR$_{STOP}$ derivative was still able to regulate, but pBAD-MgtSR$_{BP}$ no longer repressed *eptB-lacZ* expression. This result demonstrates that the MgrSR RNA has the capability to act as a base pairing regulatory RNA.

We similarly tested for functional expression of the MgtS component by carrying out immunoblot analysis to examine the levels of MgtA, a membrane protein that is stabilized by MgtS (Wang, Yin et al. 2017). Δ*mgtS* cells expressing chromosomally expressed HA-tagged MgtA were transformed with pBR, pBR-MgrR, pBAD, pBAD-MgtSR, pBAD-MgtSR$_{STOP}$, and pBAD-MgtSR$_{BP}$. Overexpression of MgtSR stabilizes MgtA-HA in conditions of Mg^{2+} limitation similar to MgtS (Fig 26D). As expected, pBAD-MgtSR$_{STOP}$ did not stabilize levels of MgtA-HA while pBAD-MgtSR$_{BP}$ did, demonstrating that a functional small protein is translated from *mgtSR* (Fig 26D). The β-galactosidase data taken together with the immunoblot analysis indicate that we have successfully constructed a synthetic dual-function RNA that can express a functional small protein and also act as a base pairing sRNA.

Translation of MgtS is the dominant activity in the absence of an intervening sequence

To examine how the intervening sequence impacts the activities of the two MgtSR components, we removed the 30-nt of intervening sequence to generate the pBAD-MgtSR-no spacer construct (Fig 27A). We hypothesized that in the absence of the spacer, the two functions may interfere with each other resulting in the loss of one or both of the activities. To test this hypothesis, we assayed the *eptB-lacZ* fusion and MgtA-HA levels for cells transformed with pBR, pBR-MgrR, pBAD, pBAD-MgtSR, and pBAD-MgtSR-no spacer. We discovered that removing the intervening sequence between the components of MgtSR eliminates the regulation of the *eptB-lacZ* fusion (Fig 27B) but not stabilization of MgtA (Fig 27C). These results suggest that translation of MgtS predominates and may sterically block MgrSR base pairing activity (Takyar, Hickerson et al. 2005). To test this prediction, we introduced a stop codon into the pBAD-MgtSR-no spacer construct. This construct, pBAD-MgtSR-no spacer$_{STOP}$, rescued *eptB-lacZ*

regulation consistent with translation of the small protein MgtS disrupting base pairing regulation by MgtSR (Fig 27B). It is interesting to note, that pBAD-MgtSR-no spacer$_{STOP}$ is still functional despite the upstream Hfq binding site being removed and that the levels of this sRNA are higher than MgtSR (Fig 28). As expected, MgtSR-no spacer$_{STOP}$, which should not be translated, was unable to stabilize the larger protein. These results demonstrate that for dual-function RNAs with non-overlapping components, the presence of a 30-nt intervening sequence helps to maintain the independent function of each component.

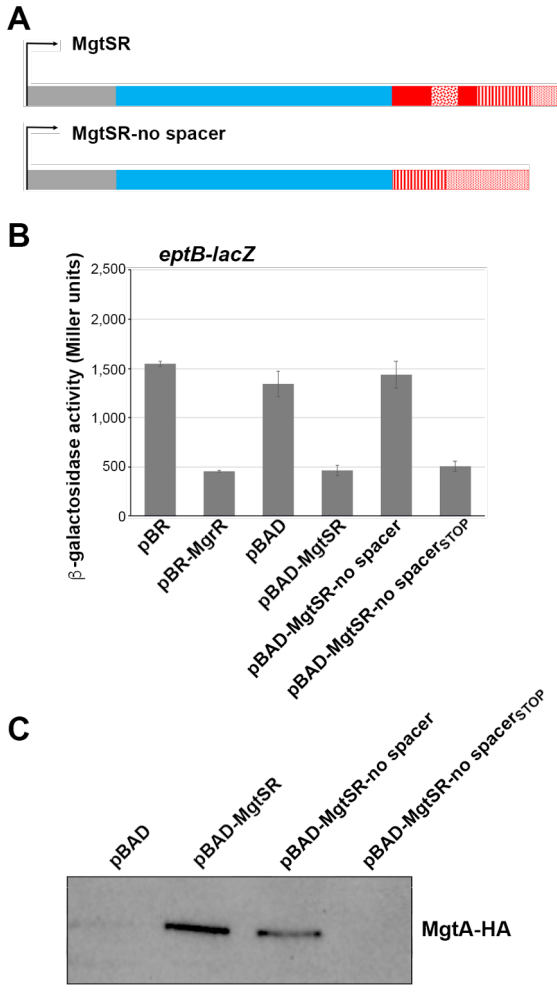


Figure 27. MgtS translation is the dominant activity absent the intervening sequence.

A Schematic of the hybrid dual-function RNA with the intervening sequence removed. Labeling of plasmid, *mgtS*, *mgrR*, ARN, seed and terminator sequences are indicated as in Fig 26B.

B Effect of pBR, pBR-MgrR, pBAD, pBAD-MgtSR, pBAD-MgtSR$_{NO\ SPACE}$, and pBAD-MgtSR$_{BP}$ overexpression on an *eptB-lacZ* fusion in a Δ*mgrR* background. Data represents triplicate biological replicates with error bars of standard deviation.

C Immunoblot analysis of MgtA-HA cells in a Δ*mgtS* transformed with pBAD, pBAD-MgtSR, pBAD-MgtSR$_{NO\ SPACE}$, and pBAD-MgtSR$_{NO\ SPACE\ MUT}$ grown to OD_{600} 0.4 in N-media supplemented with 500uM Mg^{2+}, washed, and then resuspended and grown in N-media supplemented with arabinose but without Mg^{2+}. Samples collected after 20 minutes and visualized with α-HA antibodies to detect MgtA-HA.

Truncating the intervening sequence disrupts regulation by each component

To better understand how the composition of the intervening sequence contributes to regulation by each component, we next examined the effects of truncations from either end of the intervening sequence to determine if the 5′ or 3′ end played a more prominent role in maintaining the functions of the synthetic dual-function RNA. We thus constructed derivatives with five and ten nucleotide truncations from either the 5′ or 3′ ends of the synthetic dual-function RNA: pBAD-MgtSR-5′Δ5nt, pBAD-MgtSR-5′Δ10nt, pBAD-MgtSR-3′Δ5nt, and pBAD-MgtSR-3′Δ10nt (Fig 29A). Expression of these MgtSR deletion derivatives is similar to the full length MgtSR RNA (Fig. 28). We carried out β-galactosidase assays for *eptB-lacZ* strains carrying each of these plasmids or the control pBAD and pBAD-MgtSR plasmids. Interestingly, the derivatives with truncations near the 3′ end of the RNA completely lost the ability to regulate *eptB-lacZ*, while the constructs with truncations near the 5′ end retained some ability to repress *eptB-lacZ* (Fig 29B). When stop codons were introduced into each of the truncation constructs, the RNAs all repressed *eptB-lacZ* demonstrating that the truncated constructs still have the potential to base pair and translation of the small ORF is responsible for disrupting regulatory activity.

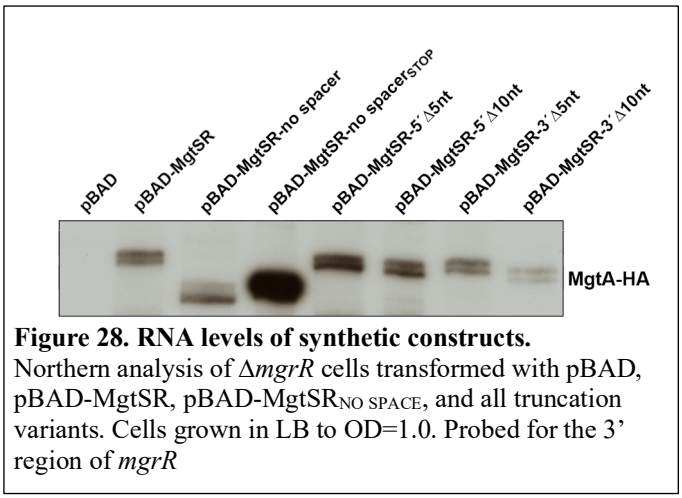


Figure 28. RNA levels of synthetic constructs. Northern analysis of Δ*mgrR* cells transformed with pBAD, pBAD-MgtSR, pBAD-MgtSR$_{NO\ SPACE}$, and all truncation variants. Cells grown in LB to OD=1.0. Probed for the 3' region of *mgrR*

In contrast, the truncations towards the 5′ end of the intervening sequence prevented stabilization of MgtA-HA, while the 3′ truncation derivatives still were still capable of increasing MgtA-HA levels similar to MgtSR (Fig 29C). Thus, while removal of portions of the

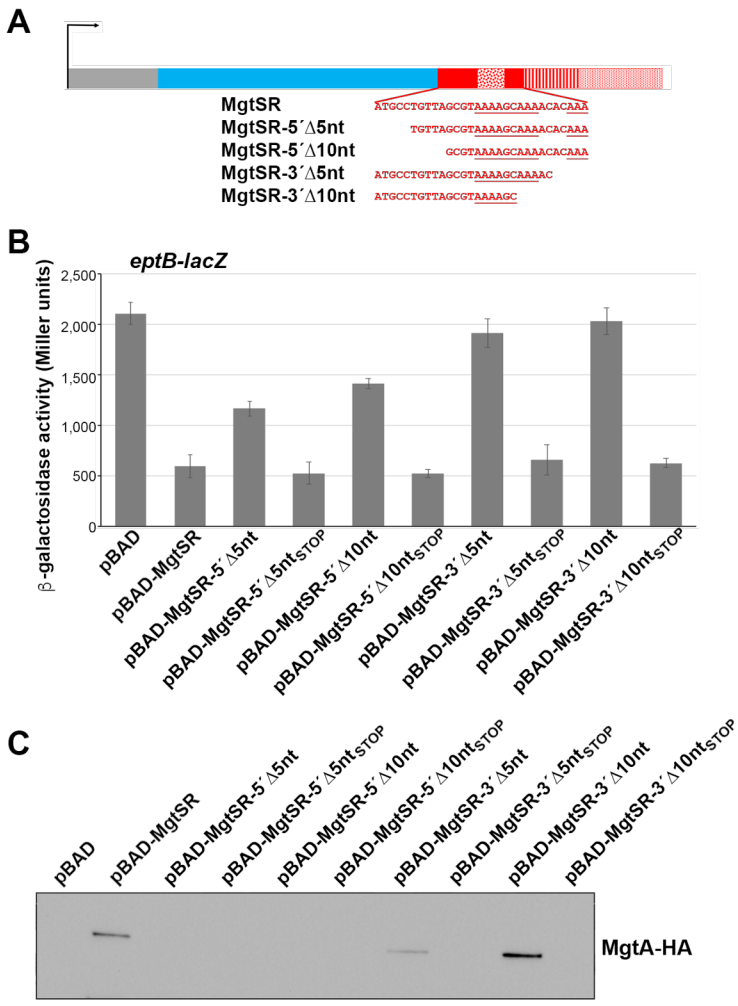


Figure 29. Analysis of the regulatory contributions of the intervening sequence.

A) Schematic of the 5-nt and 10-nt truncation mutants of MgtSR. Labeling of plasmid, *mgtS*, *mgrR*, ARN, seed and terminator sequences are indicated as in Fig 26B.

B) Effect of pBAD, pBAD-MgtSR, pBAD-MgtSR truncations and truncation stop construct overexpression on an *eptB-lacZ* fusion in a Δ*mgrR* background. Data represents triplicate biological replicates with error bars of standard deviation.

C) Immunoblot analysis of MgtA-HA cells in a Δ*mgtS* transformed with pBAD, pBAD-MgtSR, pBAD-MgtSR$_{\text{NO SPACE}}$, and pBAD-MgtSR$_{\text{NO SPACE MUT}}$ grown to OD_{600} 0.4 in N-media supplemented with 500uM Mg^{2+}, washed, and then resuspended and grown in N-media supplemented with arabinose but without Mg^{2+}. Samples collected after 20 minutes and visualized with α-HA antibodies to detect MgtA-HA.

intervening sequence from the 5′ end did not affect the base pairing activity, these deletions prevented stabilization of MgtA-HA. In contrast, the 3′ end truncation derivatives which lacked base pairing activity still stabilized MgtA-HA indicating translation of MgtS. It is possible that Hfq binding proximal to the ORF in the 5′Δ5nt and 5′Δ10nt derivatives has a negative effect on translation by competing with ribosome binding (Moon and Gottesman 2009). Another possibility is that the truncations at either the 5′ or 3′ end produce changes in RNA structure that disrupt translation or base pairing, respectively. These results highlight show that the specific sequence between the ORF and seed region has an impact on the functions of the chimeric RNA.

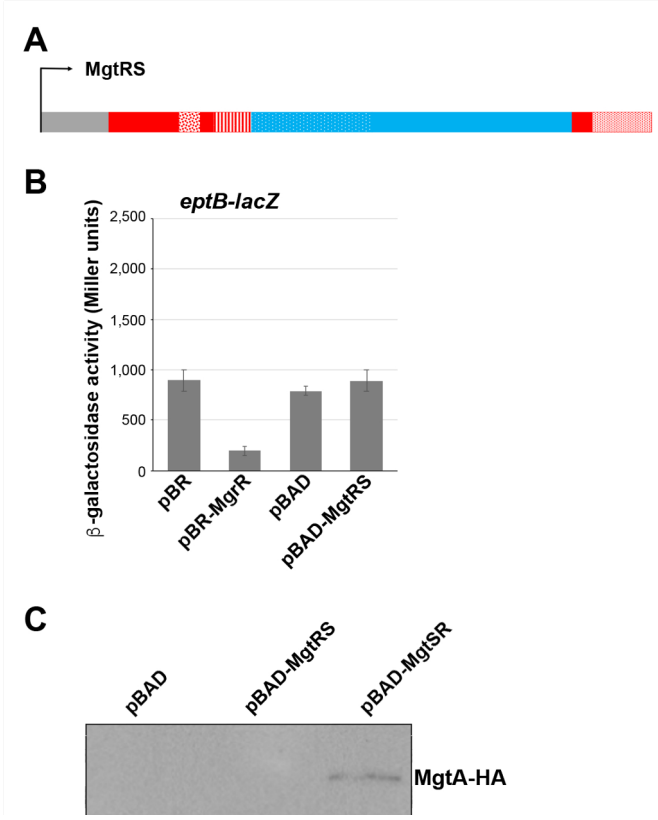


Figure 30. Reversing the organization of the components produces a nonviable synthetic dual-function RNA.

A Schematic of the inverse hybrid construct, MgtRS. Labeling of plasmid, *mgtS*, *mgrR*, ARN, seed and terminator sequences are indicated as in Fig 26B. The stippled blue region corresponds to *mgtS* sequence outside of the coding region.

B Effect of pBR, pBR-MgrR, pBAD, and pBAD-MgtRS overexpression on a *eptB-lacZ* fusion in a Δ*mgrR* background. Data represents triplicate biological replicates with error bars representing standard deviation.

C Immunoblot analysis of MgtA-HA cells in a Δ*mgtS* transformed with pBAD, pBAD-MgtRS, and pBAD-MgtSR grown to OD_{600} 0.4 in N-media supplemented with 500uM Mg^{2+}, washed, and then resuspended and grown in N-media supplemented with arabinose but without Mg^{2+}. Samples collected after 20 minutes and visualized with α-HA antibodies to detect MgtA-HA.

Reversing the organization of the components of MgtSR disrupts regulatory capability

To test whether the order of the encoded functions of the synthetic dual-function RNA impacted the activities, we placed the coding sequence of *mgtS* together with its RBS and 30-nt of intervening sequence between the base pairing region and terminator sequence of *mgrR* (Fig 30A). However, we found that this did not result in a functional construct. When cells expressing an *eptB-lacZ* fusion were transformed with pBR, pBR-MgrR, pBAD, and pBAD MgtRS and then assayed for β-galactosidase activity, we did not observe regulation (Fig 30B). Furthermore, pBAD-MgtRS was incapable of stabilizing MgtA-HA when compared against pBAD-MgtSR and a vector control (Fig 30C). We have several hypotheses to explain why the reverse construct displays no regulatory activity from either component, while the original MgtSR construct regulates targets effectively. First, it is possible that the reverse construct produces secondary

structure that sequester the RBS and seed sequence to prevent regulation. Another explanation is that the process of reversing the organization of the components has moved the ARN motif of MgrR too distant from the terminator to disrupt interactions with the RNA binding protein, Hfq. Finally, as can occur with mutations in any RNA, it is very possible that by altering the original construct we have destabilized the RNA to the point that is unable to function as either an sRNA or an mRNA. Taken together, these data suggest that the organization of a dual-function RNA is critical to function.

Discussion

Dual-function RNAs are an intriguing yet largely unexplored class of RNA regulatory molecules. Given the underidentification of small proteins, these RNA transcripts are likely more prevalent than has currently been reported. To better understand the organization and competition between the components of a dual-function RNA, we designed and constructed a synthetic dual-function RNA from the MgrR sRNA and MgtS small protein (Fig 26B). This hybrid RNA is capable of regulating by base pairing and is translated to give rise to a functional MgtS protein. We used derivatives of this synthetic construct to examine the how the organization of the components affects the regulatory activities of this RNA molecule.

Elements necessary for dual function activity

The MgtSR chimera and its derivatives point to some of the elements necessary to construct a dual-function RNA. First, translation of the small ORF likely requires an unobstructed ribosome binding site. For sRNA function, required elements include an unobstructed seed sequence for base pairing along with Hfq binding sites including a Rho-independent terminator. Where these elements are relative to one another has a significant impact on the two activities. While the MgtSR construct was completely functional, reversing the order of the two genetic elements eliminated activity (Fig 30). In the MgtRS construct, when the sRNA sequence for MgrR was moved to the 5' end the ARN motif it contained was displaced from the Rho-independent terminator, potentially eliminating Hfq activity.

The length of the sequence that genetically separates the components of dual-function RNA also is important for maintaining the independent regulatory activity of each component. Most deletions that reduced the sequence separating the smORF and the seed sequence,

eliminated base pairing activity (Fig 27 and 29). The observation that all of these derivatives regained base pairing activity upon the introduction of a stop codon suggests the main block to base pairing is occlusion by ribosome binding although some of the deletions might also affect Hfq binding. For most of these deletion constructs, the MgtS activity predominated, possibly because this activity is encoded first. Interestingly however, with a limited distance between the two genetic elements, base pairing can also interfere with translation as illustrated for the 5′Δ5-nt and 5′Δ10-nt derivatives.

In other dual-function RNA such as AzuCR or Spot 42, the overlap of the ORF and base pairing regions causes competition between the activities. Mutations that prevent translation of the ORF have been found to enhance regulation by the sRNA component. In contrast, for SgrS, where the components are separated by 15-nt, the introduction of mutations to disrupt translation of the encoded small protein, SgrT, do not change base pairing regulation of mRNA targets by SgrS (Balasubramanian and Vanderpool 2013). Furthermore, mutants with disrupted base pairing actually led to an increase in SgrT levels (Balasubramanian and Vanderpool 2013). This is consistent with the model that the stability of SgrS RNA dictates that translation will occur after base pairing regulation is completed. Overall, however, it is expected that separation of the parts of a dual-function RNA by an intervening sequence gives each component more autonomy despite existing on the same RNA molecule.

Evolution of dual-function RNAs

In thinking about the components of dual-function sRNAs, it is interesting to consider how these regulatory molecules evolve (Dutcher and Raghavan 2018) and whether there are more barriers to evolution of base pairing or a smORF as the second feature. Phylogenetic analyses of SgrST

and the RNAIII dual-function RNA from *Staphylococcus aureus* suggests that evolution in both directions is possible. For example, all SgrST homologs across enteric species carry the highly conserved 13-nt seed sequence complementary to the *ptsG* RNA near the 3' end of the RNA, while some species with a SgrS homolog, such as *Yersinia*, lack a functional SgrT ORF suggesting that the sRNA served as the precursor RNA which ultimately developed an mRNA function. In contrast, the 26 amino acid δ-hemolysin encoded by RNAIII appears to be more broadly conserved than the base pairing regions (Verdon, Girardin et al. 2009). For Spot 42-SpfP, the overlapping base pairing and SpfP-encoding regions are so well conserved that it is difficult to determine which function is more broadly conserved.

Possibility of exploiting dual-function RNAs

Dual-function RNA have the potential to be a useful tool for synthetic biology. Given that these molecules can regulate both the translation and activity of a protein, they are capable of efficient control of a specific process. Furthermore, if the two activities regulate different pathways, a single dual-function RNA can provide a regulatory link between the two pathways. By modulating when each component is active, these molecules could be used to regulate two pathways under mutually exclusive conditions. It is hoped that the observations gained from this study will guide future investigations into dual-function RNA function and construction. Ultimately, there is much to be explored about the regulatory potential of natural and synthetic dual-function RNAs.

Material and Methods

Bacterial strains and plasmid construction

Bacterial strains, plasmids, and oligonucleotides used in this study are listed in Appendix Tables. E. coli strains are derivatives of wild-type MG1655 (F- lambda- ilvG- rfb-50 rph-1). Alleles marked by antibiotic markers were moved between strains by P1 transduction. When necessary, kanamycin resistance cassettes were excised from the chromosome by FLP-mediated recombination (Cherepanov and Wackernagel 1995). All mutations and fusions were confirmed by sequencing.

Bacterial growth

Cells were grown to the indicated OD600 in Luria-Bertani broth overnight (LB) or N media after a 1:100 dilution of the overnight culture (Nelson and Kennedy 1971). Where indicated, media contained antibiotics with the following concentrations: ampicillin (100 μg/ml), chloramphenicol (25 μg/ml) and kanamycin (30 μg/ml).

Immunoblot analysis

The cell pellet from 1 ml of cells grown in the indicated medium was resuspended in 1X PBS (KD Medical) and 10 μl were loaded on a Mini-PROTEAN TGX 5%–20% Tris-Glycine gel (Bio-Rad) and run in 1X Tris Glycine-SDS (KD Medical) buffer. The proteins were electro-transferred into nitrocellulose membranes (Invitrogen) for 1 h at 100 V. Membranes were blocked with 5% non-fat milk (BioRad) in 1X PBS with 0.1% of Tween 20 (PBS-T) for 1 h and probed with a 1:1,000 dilution of α-HA antiserum (Abcam) in the same PBS-T buffer with 5%

milk for 1 h. After the incubation with the α-HA antiserum, membranes were incubated with a 1:20,000 dilution of HRP-labelled goat α-rabbit antibody (Pierce). All blots were washed 4X with PBS-T and then developed with an Amersham ECL Western Blotting Detection Kit (GE Healthcare).

Total RNA isolation

Cells corresponding to the equivalent of 10 OD600 were collected by centrifugation, and snap frozen in liquid nitrogen. RNA was extracted according to the standard TRIzol (Thermo Fisher Scientific) protocol. Briefly, 1 ml of room temperature TRIzol was add to cell pellets, resuspended thoroughly to homogenization, and incubated for 5 min at room temperature. After the addition of 200 µl of chloroform and thorough mixing by inversion, samples were incubated for 10 min at room temperature. After samples were centrifuged for 10 min at 4˚C on maximal speed, the upper phase (~0.6 ml) was transferred into a new tube and 500 µl of isopropanol was added. Samples again were mixed thoroughly by inversion, incubated for 10 min at room temperature and centrifuged at maximal speed for 15 min at 4˚C. RNA pellets were washed twice with 75% ethanol and then dried at room temperature. RNA was resuspended in 20-50 µl of DEPC water and quantified using a NanoDrop (Thermo Fisher Scientific).

Northern analysis

Total RNA (5-10 µg per lane) was separated on denaturing 8% polyacrylamide gels containing 6 M urea (1:4 mix of Ureagel Complete to Ureagel-8 (National Diagnostics) with 0.08% ammonium persulfate) in 1X TBE buffer at 300 V for 90 min. The RNA was transferred to a Zeta-Probe GT membrane (Bio-Rad) at 20 V for 16 h in 0.5X TBE, UV-crosslinked, and probed

with 32P-labeled oligonucleotides (Listed in Table 1) in ULTRAhyb-Oligo buffer (Ambion Inc.) at 45˚C. Membranes were washed twice with 2X SSC/0.1% SDS at room temperature, once with 0.2X SSC/0.1% SDS at room temperature, washed for 25 min with 0.2 × SSC/0.1% SDS at 45˚C, followed by a final wash with 0.2X SSC/0.1% SDS at room temperature before autoradiography was performed with HyBlot CL film (Denville Scientific Inc.).

β-galactosidase assays

Cultures were grown in LB to OD_{600}~1.0 with arabinose (0.2%). 100 ul of cells were added to 700 ul of Z buffer (60 mM Na2HPO4z7H2O, 40 mM NaH2PO4zH2O, 10 mM KCl, 1 mMMgSO4z7H2O, 50 mM β-mercaptoethanol). After adding 15 ul of freshly prepared 0.1% SDS and 30 ul of chloroform to each sample the cells were vortexed for 30 s and then incubated at room temperature for 15 m to lyse the cells. At zero time, the assay is initiated by adding 100 ul of ONPG (4 mg/ml) to each sample in 10 s intervals. The samples are incubated at room temperature before the reaction is terminated by the addition of 500 ul of 1M Na_2CO_3. Then, the A420 and A550 values are determined with a spectrophotometer and the absorbance data is transferred to a Microsoft Excel spreadsheet to calculate Miller units.

Addendum

My assays of MgtSR has shown that it is possible to construct functional synthetic dual-function RNAs and that these artificial constructs can be used to successfully probe how dual-function RNA work. However, there are other interesting aspects of dual-function RNA that this construct could help to study. The successful creation of the reverse construct with the small ORF 3' of the base-pairing region would be very informative. I think that issue with the first attempt could have been due to moving the entire MgrR sequence to the 5' end and using native MgtS sequence to separate the two components. This may have resulted in too much separation between the two elements in the original construct, the A-R-N motif important for binding Hfq, and the terminator. I would like to replace the original A-R-N sequence with one from ChiX, another sRNA with a very strong Hfq binding site, and then assess both regulatory capability of the construct as well as RNA levels using northern analysis. This would test if alterations to the Hfq binding site within the intervening sequence contribute to regulation of targets as well as the inherent stability of the molecule. I predict that stronger Hfq binding sites would assist in regulation and stability.

Another subject I would like to address with MgtSR is if translation of the smORF is altering the RNA structure. To test this, I would like to create different constructs with stop codons placed throughout the ORF and use enzymatic cleavage to probe the structure of these different constructs. Structural analysis of constructs that were defective for base-pairing and small protein activity should also be undertaken. MgtSR is not particularly large, and it is possible that the truncations and mutations may have disrupted the original structure to inhibit translation of the smORF or base-pairing.

The experiments that I have proposed to further explore this synthetic construct and dual-function RNA at large could certainly be expanded beyond what I have created. What happens if a new ORF or base pairing sRNA are incorporated? The creation of this initial construct will hopefully inspire other researchers to pursue the study of dual-function RNA through synthetic biology.

Chapter 5

Discussion

Prior to the current study only five dual-function RNAs had been described in bacteria. The characterization of two additional dual-function RNAs and a synthetic dual-function RNA greatly contributes to our understanding of these regulatory molecules.

Roles of dual-function RNAs in regulating carbon metabolism

Bacteria maximize growth by selecting the most energetically viable carbon source to metabolize that is available in the environment. My work has demonstrated that the base pairing sRNA Spot 42 also encodes a 15 amino acid protein, SpfP, that interacts with Crp to block Crp mediated activation of gene expression. I have also characterized the novel dual-function RNA, AzuCR. AzuCR base pairing regulates galactose metabolism while the translated small protein impacts glycerol metabolism. My work suggests that AzuCR and Spot 42 use both sRNA and mRNA activity to exert control over carbon catabolite repression. It is intriguing to note that in addition to Spot 42 and AzuCR, two other dual-function RNAs, *E. coli* SgrST and *B. subtilis* SR1, have been found to play a role in carbon metabolism. SgrS mediates the cellular response to glucose-phosphate stress in enterobacteria by base pairing with target mRNAs to reduce accumulation of sugar phosphates while the translated small protein specifically inhibits the transport activity of the major glucose permease PtsG (Wadler and Vanderpool 2007). SR1 base pairs with mRNA targets to inhibit arginine metabolism and is translated to give the small protein SR1P that stabilizes *gapA* mRNA and modulates the ability of GapA protein to recruit RNases (Gimpel, Heidrich et al. 2010).

The question of whether there is some evolutionary pressure that encourages the use of dual-function RNA to regulate carbon metabolism is intriguing. A base pairing RNA uses less energy and less complicated machinery to regulate mRNA and protein targets than is required by translation of a protein to fulfill the same role. However, translated proteins can withstand reversible modifications to alter their own function. Perhaps it is desirable for an energy generating pathway to be regulated in an efficient and flexible manner by molecules that can translate a protein and also riboregulate other mRNAs. The flexibility provided by having two distinct regulatory mechanisms within a single molecule allows for diverse combinations of regulation that may be beneficial for such a complicated system.

Competition between base pairing and translation

The competition between components has previously been documented for SgrS and briefly explored for RNAIII. The work presented in this dissertation characterizes the competition inherent in two additional dual-function RNAs, AzuCR and Spot 42. My work explores the features that are important for controlling competition using a synthetic construct.

The base pairing region of SgrS is 15 nucleotides downstream of the SgrT ORF. Mutations that inhibit *sgrT* translation do not impede SgrS regulation of mRNA targets, but mutations that restrict base-pairing increase SgrT translation, indicating that the base-pairing function interferes with translation (Balasubramanian and Vanderpool 2013). Furthermore, translation of the small protein lags behind transcription of SgrS RNA by approximately 30 minutes (Balasubramanian and Vanderpool 2013). Based on these observations, it was suggested that the base pairing function predominates and that SgrST molecules that base pair with target mRNAs are unavailable for use as an mRNA as they are degraded. Once the pool of targeted

mRNA has been sufficiently degraded, SgrST accumulates and is translated. Given that co-degradation of a sRNA with mRNA targets has been shown for other sRNAs, this type of competition may be present in other dual-function RNAs.

The primary effector molecule of the *S. aureus agr* quorum sensing system is the 514-nt RNAIII. This dual-function RNA contains fourteen stem-loop motifs and has been demonstrated to directly regulate 12 mRNA targets. RNAIII also contains an ORF for the 26 amino acid δ-hemolysin (*hld*) which targets host cell membranes to lyse cells (Verdon, Girardin et al. 2009). Interestingly, the production of *hld* is delayed by 1 hour following RNAIII transcription (Balaban and Novick 1995). Analysis of RNAIII structure demonstrated that the 5' and 3' ends of the

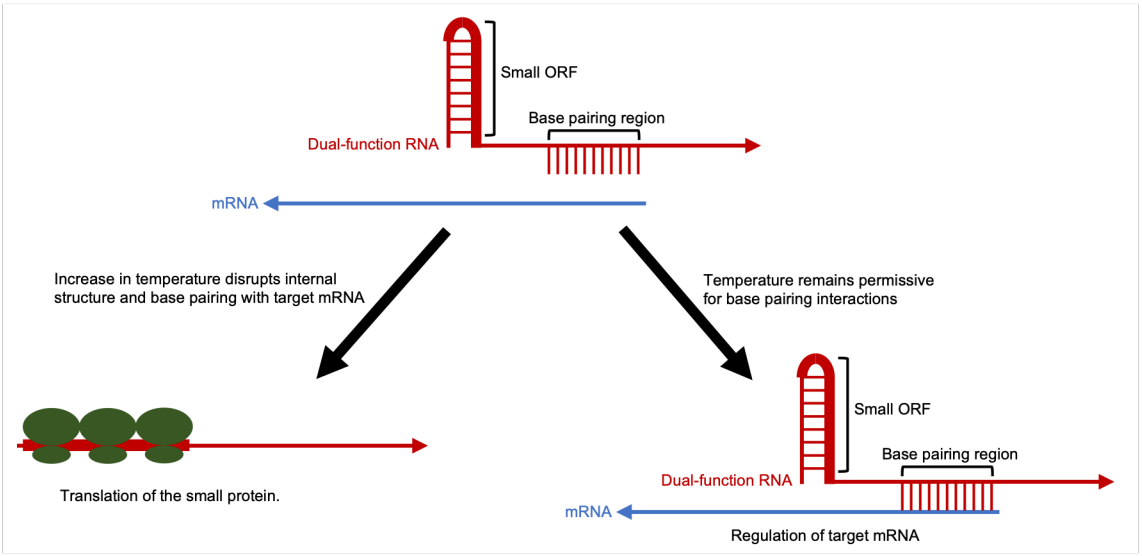


Figure 31. Temperature can control competition between elements of a dual-function RNA. The dual-function RNA (red) contains a secondary structure that sequesters a small ORF. At a permissive temperature the dual-function RNA regulates target mRNA (blue) via base pairing interactions. An increase in temperature can disrupt base pairing and internal structure to permit and translation of the small protein.

RNA are in close proximity, perhaps suggesting that the delay in δ-hemolysin levels are due to this interaction partially occluding the RBS of *hld* (Benito, Kolb et al. 2000). In support of this hypothesis, it has been shown that deletion of the 3' end of RNAIII eliminates the delay between *hld* protein levels and RNAIII RNA levels (Balaban and Novick 1995). The method by which

RNAIII mediates a conformational change in structure to permit δ-hemolysin translation are unknown.

The work on Spot 42-SpfP and AzuCR has provided additional mechanisms for how competition can control the activities of the two components of a dual-function RNA. For Spot 42-SpfP and AzuCR, there is direct overlap between the base pairing region and smORF which causes the activities to interfere with one another. In both dual-function RNAs, the introduction of a stop codon to reduce ribosomal occupancy of the transcript increases base pairing activity and exacerbates growth defects caused by overexpression of the full-length RNA.

For Spot 42, the ribosome binding site and smORF are sequestered in secondary structure. Levels of the small protein increase following heat shock, perhaps due to partially resolving this structure. Additionally, the increase in levels of SpfP coincide with a temperature dependent reduction in base pairing regulation. Taken together, the data demonstrated sRNA activity of Spot 42 is inhibited by the increase in temperature. I hypothesize that the temperature dependent increase in SpfP levels reinforces the repression of Crp targets by Spot 42 RNA. As many studied sRNA have some form of secondary structure it is possible that additional undiscovered dual-function RNAs are subject to temperature guided competition between their components as illustrated in Figure 7.

The work on the dual-function RNA AzuCR has uncovered many hints as to how this molecule regulates the activity of its components. AzuCR encodes the small protein AzuC which specifically interacts with glycerol-3-phosphate dehydrogenase, GlpD, to increase activity. This work also demonstrated that the anaerobically expressed sRNA FnrS represses both AzuC and GlpD, AzuCR RNA co-immunoprecipitates with Hfq, and SPA tagged AzuC levels were elevated in a Δ*hfq* strain. Finally, introduction of a stop codon to prevent translation of the

smORF within AzuCR produces a stronger repressive effect on target mRNAs than the *wild type* transcript alone. The processes that control the regulatory roles of AzuCR are not fully understood, but they clearly depend on the levels of sRNAs that target the dual-function RNA (FnrS), the levels of RNA chaperones (Hfq), and the levels of mRNA targets.

As suggested in the work on AzuCR, a mechanism by which the activities of a dual-function RNA can be regulated is through an external factor, either an RNA binding protein or a regulatory RNA. In this scenario, the factor could block or expose an RBS or affect access to the base pairing region or binding of an RNA chaperone. This control over translation could also be achieved if the dual-function RNA contained a riboswitch that would alter levels of translation in response to binding with a metabolite or small molecule. Finally, it is possible that a dual-function RNA could undergo a maturation or activation phase where the transcript is processed so that only one or the other activity is functional.

Generation of the synthetic dual-function RNA, MgtSR, has also provided some insight into the competition that is inherent to these molecules. The work with MgtSR focused on understanding how the sequence between separate components contributes to the regulatory integrity of the entire molecule. Different permutations of the construct were made to study this intervening sequence. This study concluded that for MgtSR the separating sequence prevented the translation of the smORF from disrupting riboregulation by the base pairing region. In addition, this intervening sequence was found to play an important role in the proper function of both activities. The work demonstrated how the creation of a dual-function RNA is both possible and has the potential to aid in the research of these molecules.

This dissertation has identified a new method by which dual-function RNA regulate their components and attempted to provide additional insights into the inherent competition within

these molecules by constructing and studying a synthetic dual-function RNA. We hypothesize that there are other types of competition that can impact translation of the smORF as a method to control when sRNA or mRNA activity is predominate. As more of these molecules are identified we anticipate a more comprehensive understanding of how translation and RNA function compete within dual-function RNA.

Where do we go from here?

Application of what has been learned about dual-function RNAs in E. coli to other bacteria

The majority of currently identified dual-function RNA are in bacteria but have not been examined in terms of how their separate components compete. All of the dual-function RNA presented in this dissertation have been studied in sufficient depth to begin describing how the RNA and protein functions are coordinated. The Spot 42 project described how at higher temperatures the smORF that is sequestered in a hairpin is translated more effectively and at lower temperatures base pairing regulation is more efficient. The analysis of competition between components should be undertaken for all future characterized dual-function RNA as it is the core mechanism that dictates how and when these molecules exert their regulatory impact. Additionally, the methods we have used to study the molecules characterized in this dissertation can be utilized to fully understand the RNA and protein functions of future dual-function RNA and how these elements compete.

Application of what has been learned about dual-function RNAs in E. coli to eukaryotes

Although the work in this dissertation has studied dual-function RNA in bacteria, the conclusions and discoveries could potentially be applied to eukaryotes. As previously explained, the many

different types of ncRNA in eukaryotes undergo a variety of different maturation steps, are of a wide range of sizes, and regulate a large number of different cellular process. Given the variety of ncRNA in eukaryotes it is reasonable to assume that there are many more dual-function RNA species that have functions outside what has been identified in bacteria.

As mentioned in the introduction, one subset of dual-function RNA in eukaryotes are those whose RNA perform a role during embryonic development. The examples that we have previously discussed are from different organisms, *Oskar* in *Drosophila* and *Squint* in zebrafish (Kloc, Wilk et al. 2005, Jenny, Hachet et al. 2006, Lim, Kumari et al. 2012). All three of these RNAs assist in cell patterning during early embryogenesis. The early embryo is a unique and short-lived regulatory environment. What are the advantages of dual-function RNA regulators in this unique cellular ecosystem? Many mRNA are translationally inhibited when they are localized to oocytes in order to translate their protein product at a specific interval important for proper development (Evans and Hunter 2005, Lasko 2009). Perhaps RNA that can have additional functions while translationally dormant are selected for in this early developmental environment. By having a translation independent RNA function while mRNA activity is suppressed these molecules would serve to increase the regulatory capacity of the early embryo while reducing the need to deliver additional regulatory RNA molecules.

Another potential dual-function RNA type consists of those that could arise from lncRNAs. lncRNAs are defined as RNA molecules that are longer than 200-nt and do not encode a protein. More than 16,000 lncRNAs have been annotated with some estimates being even higher (Derrien, Johnson et al. 2012). Despite the enormous numbers of these large RNA molecules less that 1% have actually been experimentally validated (Quek, Thomson et al. 2015). Given the vast amount of uncharacterized RNA it stands to reason that some of these may

contain viable ORFs, have a ncRNA function, or both. We have already discussed the example of the small protein, sarcolipin, that was found to be translated from a formerly characterized lncRNA (Singh, Dalton et al. 2019). I believe that further examination of eukaryotic genomes to search for ORFs located in what are currently classified as lncRNAs may uncover many novel dual-function RNAs.

The search for more dual-function RNAs

Multiple computational searches and genetic screens have been carried out to identify potential ncRNAs in bacteria and eukaryotes. These experiments have provided evidence of function for these RNA molecules and represent an important first step in identifying novel dual-function RNA. In bacteria, RNA interaction by ligation and sequencing (RIL-seq) has been used to identify Hfq-bound sRNA:mRNA pairs (Melamed, Peer et al. 2016, Melamed, Adams et al. 2020). In eukaryotes, Argonaute binding information can be generated from crosslinking and immunoprecipitation followed by sequencing (CLIP-seq) and has greatly increased the resolution of potential miRNA interactions (Wessels, Lebedeva et al. 2019). Both of these methods leverage next generation sequencing to produce large numbers of potential RNA:RNA interactions which can then be examined in greater detail.

Perhaps one of the most significant obstacles to the search for dual-function RNAs is the difficulty in identifying smORFs that are translated. A powerful approach for the identification of small protein-coding regions is ribosome profiling. This method uses deep sequencing of ribosome protected mRNA fragments to globally monitor ribosome binding *in vivo*. In bacteria, ribosome profiling has been employed to identify potential small proteins (Weaver, Mohammad et al. 2019). In this method, antibiotics were used to arrest initiating 70S ribosomes over start

codons followed by ribosome profiling to produce an accurate database of initiation sites throughout the *E. coli* genome (Meydan, Marks et al. 2019). In eukaryotes, ribosome profiling has been leveraged in numerous ways to identify protein variants from alternative start codons and small proteins translated from RNAs that were previously considered noncoding (Brar and Weissman 2015). Determining what smORFs are translated will be valuable for identifying potential dual-function RNAs in bacteria and eukaryotes.

By combining data sets of potential ncRNAs with potential small proteins it should be possible to create databases of potential dual-function RNAs for bacteria and eukaryotes. Based on what is currently understood about dual-function RNA, we suggest that the search for novel molecules be constrained to below 600 nucleotides and the minimum size for the smORF should be set to 10 amino acids in order to prevent the inclusion of nonviable ORFs in this analysis. This database could then be examined for dual-function RNAs whose components are non-overlapping, as with SgrS, or whose components overlap, as with AzuCR or Spot 42.

Once databases of potential dual-function RNAs have been established, a critical step will the documentation of both functions. The gold standard for demonstrating direct base pairing interaction is mutational analysis of the putative base pairing region. Some regulatory RNAs interact with target proteins. An example of this is the sRNA CsrB binding the carbon storage regulator, CsrA, to inhibit CsrA-mediated acceleration of *glg* mRNA degradation (Liu, Gui et al. 1997). Verification of this interaction can be assessed using a co-purification for a tagged derivative of the potential protein and then identifying the interacting RNA using northern analysis. This purification can also be done using mutants to identify residues that are important for the interaction. Structure probing can also be used to demonstrate that the protein makes relevant contacts with the RNA. The documentation of protein function can be achieved

through copurification assays to determine interacting proteins followed by mass spectrometric analysis. Often an initial purification with a tagged derivative of the small protein of interest is conducted to identify potential large interacting proteins. After a possible protein partner for the small protein of interest is determined, the larger protein is purified to attempt to identify the original small protein of interest to confirm the result. Following identification of a potential interaction an in vitro analysis of purified components can provide further evidence of this interaction.

Generation of synthetic dual-function RNAs to regulate desired processes

Dual-function RNAs represent an interesting opportunity for synthetic biology applications. As we have demonstrated in this work, the creation of an artificial dual-function RNA from existing ncRNA and small protein components can provide valuable information on how these molecules are constructed and how they behave. Using specific truncation mutants, we were able to more carefully describe how alterations to the intervening sequence between components contribute to the regulatory integrity of the entire molecule. However, synthetic constructs could be employed to examine other aspects of dual-function RNA. For example, one could introduce an adenylate-uridylate-rich element (ARE) into the 3' UTR of a eukaryotic synthetic dual-function RNA to destabilize the molecule in order to test how reduced RNA half-life impacts the competition between components of the molecule. In contrast, an RNA chaperone binding site could be introduced to greatly increase the stability of the molecule to determine if this predisposes the molecule towards the regulatory RNA function. By controlling when each element of a dual-function RNA is active, these molecules have the potential to regulate two separate pathways under separate conditions making them a versatile regulatory tool.

Finally, understanding how to design dual-function RNAs could provide an efficient method to regulate RNA and protein using the same molecule. As these molecules can regulate both mRNA and protein, they could be employed to specifically control a single process. Another interesting use would be to design a dual-function RNA to regulate two different pathways at the same time. Ultimately, dual-function RNAs represent an opportunity to design constructs that can effectively control gene expression using two different regulatory mechanisms.

Conclusions

Despite the work in bacteria and eukaryotes to characterize and understand dual-function RNAs, it is evident that much remains to be learned about these RNA molecules. As we continue to study these RNA transcripts, we will uncover more about their functions, competition, and evolution. Expanding the numbers of characterized dual-function RNA would be a significant advancement in understanding these unique molecules.

Appendices

Appendix 1: Strain List

Name	Genotype	Source
MG1655	E. coli F- λ- *ilvG- rfb-50 rph-1*	lab stock
TOP10	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) *endA1 λ*$^-$	Invitrogen
NM400	MG1655 *mini-λ:cat*	N. Majdalani
PM1205		Mandin and Gottesman, 2009
GSO351	MG1655 *azuC-SPA::kan*	Hemm et al. 2008
GSO955	MG1655 Δ*hfq*::*kan*	Melamed et al. 2020
GSO1006	MG1655 *azuC-SPA*	This study
GSO1007	MG1655 *azuC-SPA* Δ*hfq*::*kan*	This study
GSO1008	MG1655 *azuC-gfp::kan*	This study
GSO1009	MG1655 $azuC_{IL6,7EE}$-*gfp::kan*	This study
GSO1010	MG1655/pAZ3-AzuC	This study
GSO350	MG1655 *acrZ-SPA::kan*	Hemm et al. 2008
GSO1011	MG1655 *glpD-HA-*His_6*::kan*	This study
GSO193	MG1655 Δ*azuC::kan*	Hobbs et al. 2010
GSO1012	MG1655 Δ*azuC*	This study
GSO1013	MG1655 Δ*azuC glpD-HA-*His_6*::kan*	This study
GSO785	MG1655 *mgtA-HA*::*kan* Δ*mgtS*	Wang et al. 2017
JW3389	Δ*glpD*::*kan*	Keio collection
GSO1014	MG1655 Δ*glpD*::*kan*	This study
GSO1015	MG1655 Δ*azuC* Δ*glpD::kan*	This study
GSO1016	MG1655 Δ*azuC::kan*/pKK177-2	This study
GSO1017	MG1655 Δ*azuC::kan*/pKK-AzuC	This study
GSO1018	MG1655 Δ*azuC::kan*/pKK-$AzuC_{L3STOP}$	This study
GSO1019	MG1655 Δ*azuC::kan*/pKK-GlpD	This study
GSO1020	MG1655 Δ*azuC::kan*/pRI	This study
GSO1021	MG1655 Δ*azuC::kan*/pRI-AzuCR	This study
GSO1022	MG1655 Δ*azuC::kan*/pRI-$AzuCR_{L3STOP}$	This study
GSO402	MG1655 Δ*fnrS::kan*	Durand and Storz, 2010
GSO1023	MG1655 *azuC-SPA* Δ*fnrS::kan*	This study
GSO1024	PM1205 *azuC::lacZ*	This study
GSO1025	PM1205 *azuC-III::lacZ*	This study
GSO1026	MG1655 Δ*azuC::kan*/pXG10-SF-*cadA-gfp*	This study
GSO1027	MG1655 Δ*azuC::kan*/pXG10-SF-*cadA-M1-gfp*	This study
GSO1028	MG1655 Δ*azuC::kan*/pR1-$AzuCR_{L3STOP-M1}$	This study

GSO1029	MG1655 Δ*azuC::kan*/pXG10-SF-*galE-gfp*	This study
GSO1030	MG1655 Δ*azuC::kan*/pXG10-SF-*galE-M2-gfp*	This study
GSO1031	MG1655 Δ*azuC::kan*/pR1-AzuCR$_{L3STOP-M2}$	This study
GSO1032	MG1655 *azuC-SPA::kan*/pRI-*cadA*	This study
GSO1033	MG1655 *azuC-SPA::kan*/pRI-*cadA*$_{control}$	This study
GSO1034	MG1655 *azuC-SPA::kan*/pRI-*galE*	This study
GSO1035	MG1655 *azuC-SPA::kan*/pRI-*galE*$_{control}$	This study
GSO982	MG1655 (crl+)	Melamed et al. 2020
GSO954	MG1655 Δ*hfq-cat::sacB*	Melamed et al. 2020
GSO956	MG1655 Δ*proQ*::*kan*	Melamed et al. 2020
GSO959	MG1655 Δ*hfq* Δ*proQ*::*kan*	Melamed et al. 2020
GSO1037	MG1655 *spfP-SPA::kan*	This study
GSO441	MG1655 *srlA-lacZ* Δ*spf::kan*	Beisel and Storz, 2011
GSO440	MG1655 *nanC-lacZ* Δ*spf::kan*	Beisel and Storz, 2011
GSO519	MG1655 *glpF-lacZ* Δ*spf::kan*	Beisel and Storz, 2011
GSO433	MG1655 Δ*spf::kan*	Beisel and Storz, 2011
GSO1038	MG1655 Δ*spf::kan*/pRI	This study
GSO1039	MG1655 Δ*spf::kan*/pRI-Spot 42	This study
GSO1040	*MG1655* Δ*spf::kan*/pRI-Spot 42$_{STOP}$	This study
GSO1041	MG1655 Δ*spf::kan*/pKK177-3	This study
GSO1042	MG1655 Δ*spf::kan*/pKK177-3-SpfP-scram	This study
GSO1043	*MG1655* Δ*spf::kan*/pKK177-3-SpfP-scram$_{STOP}$	This study
GSO1044	MG1655 Δ*spf::kan*/pKK177-3-SpfP-scram$_{F2A}$	This study
GSO1045	MG1655 Δ*spf::kan*/pKK177-3-SpfP-scram$_{Y3A}$	This study
GSO1046	MG1655 Δ*spf::kan*/pKK177-3-SpfP-scram$_{S5A}$	This study
GSO1047	MG1655 Δ*spf::kan*/pKK177-3-SpfP-scram$_{D6A}$	This study
GSO1048	MG1655 Δ*spf::kan*/pKK177-3-SpfP-scram$_{L8A}$	This study
GSO1049	MG1655 Δ*spf::kan*/pKK177-3-SpfP-scram$_{H10A}$	This study
GSO1050	MG1655 Δ*spf::kan*/pKK177-3-SpfP-scram$_{V11A}$	This study
GSO1051	MG1655 Δ*spf::kan*/pKK177-3-SpfP-scram$_{F14A}$	This study
GSO1052	MG1655 Δ*spf::kan*/pKK177-3-SpfP-scram$_{N\text{-terminal Y}}$	This study
GSO1053	MG1655 Δ*spf::kan*/pKK177-3-SpfP-scram$_{C\text{-terminal Y}}$	This study
GSO1054	MG1655 Δ*spf::kan*/pKK177-3-SpfP-scram$_{N\text{-terminal STOP}}$	This study
GSO1055	MG1655 Δ*spf::kan*/pKK177-3-SpfP-scram$_{C\text{-terminal STOP}}$	This study
GSO1056	MG1655 Δ*spf::kan*/pKK177-3-SpfP-scram$_{N\text{-terminal FLAG}}$	This study
GSO1057	MG1655 *galK-HA-His$_6$::kan*	This study
GSO1058	MG1655 *crp-HA-His$_6$::kan*	This study
GSO1059	MG1655 Δ*spf*	This study
GSO1060	MG1655 Δ*spf galK-HA-His$_6$::kan*	This study

GSO1061	MG1655 Δ*spf crp-HA-His$_6$::kan*	This study
GSO1062	MG1655 Δ*spf::kan*/pKK177-3-YoaK$_{\text{N-terminal STOP}}$	This study
GSO388	MG1655 Δ*crp::cm*	Durand and Storz, 2010
GSO1063	MG1655 Δ*spf::kan* Δ*crp::cm*	This study
GSO1064	*galM-SPA::kan*	SPA collection
GSO1065	*malE-SPA::kan*	SPA collection
GSO1066	MG1655 Δ*spf galM-SPA::kan*	This study
GSO1067	MG1655 Δ*spf malE-SPA::kan*	This study
GSO1068	MG1655 Δ*spf malK-SPA::kan*	This study
GSO449	MG1655 Δ*spf srlA-SPA::kan*	Beisel and Storz, 2011
GSO351	MG1655 *azuC-SPA::kan*	Hemm et al.
GSO1069	PM1205 *galE* P1-*lacZ*	This study
GSO1070	PM1205 *galE* P1$_{C35G}$-*lacZ*	This study
GSO1071	PM1205 *galE* P1-*lacZ* Δ*spf*	This study
GSO1072	PM1205 *galE* P1$_{C35G}$-*lacZ* Δ*spf*	This study
GSO1073	MG1655 Δ*spf galK-HA-His$_6$*	This study
GSO1074	NM400 *spf$_{STOP}$::kan*	This study
GSO1075	MG1655 *spf$_{STOP}$::kan galK-HA-His$_6$*	This study
GSO1076	NM400 *spf-scram::kan*	This study
GSO1077	MG1655 *spf-scram::kan galK-HA-His$_6$*	This study
KM100	*mal*::*lacI* q, Δ*araBAD*, *leu* +, *araC*+, *araE*	Moon et. al., 2013
KM132	Δ*mgrR ::kan*	Moon et. al., 2014
KM238	*eptB-lacZ* Δ*mgrR::kan* positions -150 to +94 of *eptB* promoter	Moon et. al., 2015
GSO229	MG1655 Δ*mgtS::kan*	Wang et. al., 2017
GSO786	MG1655 MgtA-HA Δ*mgtS::kan*	Wang et. al., 2017

Appendix 2: Plasmid List

Name	Relevant features	Source
pAZ3	pBAD18 with an EcoRI site at +1, Cm^R	Kawano et al. 2005
pRI	pKK177-3 with an EcoRI site at transcription start site, Amp^R	PMID:15466020
pKK177-3	pRI with a ribosome binding site	PMID: 6390428
pBR	Derivative of the pBR322-derived pBRplac vector	PMID: 16359331
pJL148	Template for amplification of SPA tag together with the adjacent kanamycin-resistance cassette flanked by FRT sites	PMID: 15253427
pCP20	Temperature-sensitive vector that expresses FLP recombinase. Used to excise antibiotic resistance cassettes flanked by FRT sites	PMID: 7789817
pXG-0	Superfolder GFP plasmid negative control	PMID: 17264113
pXG10-SF	Superfolder GFP plasmid	PMID: 22458297
pAZ3-AzuCR	pAZ3 with AzuCR cloned into EcoRI site	This study
pRI-AzuCR	pRI with AzuCR cloned into EcoRI and HindIII, Amp^R	This study
pRI-AzuCR$_{STOP}$	pRI-AzuCR with the third codon mutated to TAG, Amp^R	This study
pKK-AzuC	pKK with AzuC open reading frame cloned into EcoRI and HindIII, Amp^R	This study
pKK-AzuC$_{STOP}$	pKK-AzuC with the third codon mutated to TAG, Amp^R	This study
pKK-GlpD	GlpD coding region cloned into EcoRI and HindIII, Amp^R	This study
pBR-FnrS	FnrS sequence cloned in pBR322	Durand and Storz
pBR-MicA	MicA sequence cloned in pBR322	PMID: 25030700
pBR-FnrS-I	FnrS sequence cloned in pBR322 containing truncation as described in Durand and Storz	Durand and Storz
pBR-FnrS-II	FnrS sequence cloned in pBR322 containing truncation as described in Durand and Storz	Durand and Storz
pBR-FnrS-III	FnrS sequence cloned in pBR322 containing truncation as described in Durand and Storz	Durand and Storz
pXG10-SF-cadA	*cadA* sequence from the transcription start to 20 codons into the open reading frame cloned between NheI and NsiI	This study
pXG10-SF-cadA-M1	pXG10-SF-CadA sequence mutated from CGTTATTGCA to ttTgATTttg	This study
pXG10-SF-galE	galE sequence cloned into PXG10-SF	This study
pXG10-SF-galE-M2	galE sequence with mutations to prevent AzuCR base pairing	This study
pRI-AzuCR$_{STOP}$-M1	pRI-AzuCRSTOP sequence mutated from TATTGCAAGACGTTC to GATCGCGACACA	This study
pRI-AzuCR$_{STOP}$-M2	pRI-AzuCRSTOP sequence mutated from TGCAAGACG to caaAAGACa	This study
pRI-cadA frag	cadA coding sequence from nucleotide 1 to 46 with 50 nucleotide upstream cloned into pRI between EcoRI and HindIII, AmpR	This study

pRI-cadA-frag control	cadA coding sequence from nucleotide 1965 to 2061 cloned into pRI between EcoRI and HindIII, AmpR	This study
pRI-galE frag	galE coding sequence from nucleotide 679 to 778 cloned into pRI between EcoRI and HindIII, AmpR	This study
pRI-galE-frag control	galE coding sequence from nucleotide 881 to 964 cloned into pRI between EcoRI and HindIII, AmpR	This study
pRI-galR-frag	GalR coding sequence from nucleotide 209 to 290 cloned into pRI between EcoRI and HindIII, AmpR	This study
pXG10-SF-CadA-M3	pXG10-SF-CadA sequence mutated from CGTTATTGCA to tGTTATTttg	This study
pEVOL-p-AzF	Orthologous aminoacyl-tRNA and tRNA pair	PMID: 12148987
pRI-Spot 42	pRI with *spf* cloned into EcoRI and HindIII, Amp^R	This study
pRI-Spot 42_{STOP}	pRI-Spot 42 with T27G mutation, Amp^R	This study
pKK-SpfP-scram	pKK with SpfP open reading frame scrambled and cloned into EcoRI and HindIII, Amp^R	This study
pKK-SpfP-$scram_{STOP}$	pKK-SpfPscram with the third codon mutated to TAG, Amp^R	This study
pKK-SpfP-$scram_{F2A}$	pKK-SpfPscram with the second codon mutated to GCG, Amp^R	This study
pKK-SpfP-$scram_{Y3A}$	pKK-SpfPscram with the third codon mutated to GCG, Amp^R	This study
pKK-SpfP-$scram_{S5A}$	pKK-SpfPscram with the fifth codon mutated to GCG, Amp^R	This study
pKK-SpfP-$scram_{D6A}$	pKK-SpfPscram with the sixth codon mutated to GCG, Amp^R	This study
pKK-SpfP-$scram_{L8A}$	pKK-SpfPscram with the eighth codon mutated to GCG, Amp^R	This study
pKK-SpfP-$scram_{H10A}$	pKK-SpfPscram with the tenth codon mutated to GCG, Amp^R	This study
pKK-SpfP-$scram_{V11A}$	pKK-SpfPscram with the eleventh codon mutated to GCG, Amp^R	This study
pKK-SpfP-$scram_{F14A}$	pKK-SpfPscram with the fourteenth codon mutated to GCG, Amp^R	This study
pKK-SpfP-$scram_{NY}$	pKK-SpfP-scram with tyrosine (TAC) introduced after the start codon, Amp^R	This study
pKK-SpfP-$scram_{CY}$	pKK-SpfP-scram with tyrosine (TAC) introduced before the stop codon, Amp^R	This study
pKK-SpfP-$scram_{NSTOP}$	pKK-SpfP-scram with stop (TAG) introduced after the SpfP start codon, Amp^R	This study
pKK-SpfP-$scram_{CSTOP}$	pKK-SpfP-scram with stop (TAG) introduced before the SpfP stop codon, Amp^R	This study
pKK-N-FLAG-SpfP	pKK-SpfP with an N terminal 3X FLAG tag	This study
pBR-MgrR	pBR with *mgrR* cloned into EcoRI and HindIII, Amp^R	Moon, et. al., 2013
pBAD-MgtSR	Initial construct with MgtS ORF 30-nt upstream of *mgrR*	This study
pBAD-$MgtSR_{STOP}$	pBAD-MgtSR with TAG stop mutant as second amino acid	This study
pBAD-$MgtSR_{BP}$	pBAD-MgtSR with GCAA>GCTT mutation to disrupt base-pairing	This study

pBAD-MgtSR$_{\text{NO SPACE}}$	pBAD-MgtSR with 30-nt intervening sequence removed	This study
pBAD-MgtSR$_{\text{NO SPACE MUT}}$	pBAD-MgtSR$_{\text{NO SPACE MUT}}$ with TAG stop mutant as second amino acid	This study
pBAD-MgtSR$_{5'\Delta 5nt}$	pBAD-MgtSR with 5-nt removed from 5' end of intervening sequence	This study
pBAD-MgtSR$_{5'\Delta 10nt}$	pBAD-MgtSR with 10-nt removed from 5' end of intervening sequence	This study
pBAD-MgtSR$_{3'\Delta 5nt}$	pBAD-MgtSR with 5-nt removed from 3' end of intervening sequence	This study
pBAD-MgtSR$_{3'\Delta 10nt}$	pBAD-MgtSR with 10-nt removed from 3' end of intervening sequence	This study
pBAD-MgtSR$_{5'\Delta 5nt\ STOP}$	pBAD-MgtSR with 5-nt removed from 5' end of intervening sequence and stop mutant	This study
pBAD-MgtSR$_{5'\Delta 10nt\ STOP}$	pBAD-MgtSR with 10-nt removed from 5' end of intervening sequence and stop mutant	This study
pBAD-MgtSR$_{3'\Delta 5nt\ STOP}$	pBAD-MgtSR with 5-nt removed from 3' end of intervening sequence and stop mutant	This study
pBAD-MgtSR$_{3'\Delta 10nt\ STOP}$	pBAD-MgtSR with 10-nt removed from 3' end of intervening sequence and stop mutant	This study
pBAD-MgtRS	Reverse construct	This study

Appendix 3: Northern Probes

Name	Sequence
AzuCR	TGCCTGGCGGTACGTCTTTGAACGTCTTGC
5S	CGGCGCTACGGCGTTTCACTTCTG
Spot 42	GGTCTGAAAGATAGAACATCTTACCTCTGT
MgtSR	GTCATCCCATTTGTGGCTGAAATACGCGGCCAG

Appendix 4: Oligonucleotides for Plasmid Construction

Name	Sequence
pRI-AzuCR$_{STOP}$-M1F	GAAAAGTATGTTCAATAACGATCGCGACACATTCAAAGACGTACCGCCAG
pRI-AzuCR$_{STOP}$-M1R	CTGGCGGTACGTCTTTGAATGTGTCGCGATCGTTATTGAACATACTTTTC
pRI-AzuCR$_{STOP}$-M3F	GTATGTTCAATAACTATCAAAAGACATTCAAAGACGTACCGCCAGGC
pRI-AzuCR$_{STOP}$-M3R	GCCTGGCGGTACGTCTTTGAATGTCTTTTGATAGTTATTGAACATAC
pKK-AzuCF	CCGGAATTCATGAAATAGCGCAAAATCCTGAAAAGTATGTTCAATAACTATTGCAAGACGTTCAAAGACGTACCGCCAGGCAATATGTTCCGATAAAAGCTTGGG
pKK-AzuCR	CCCAAGCTTTTATCGGAACATATTGCCTGGCGGTACGTCTTTGAACGTCTTGCAATAGTTATTGAACATACTTTTCAGGATTTTGCGCAGTTTCATGAATTCCGG
pKK-AzuC$_{STOP}$F	CCGGAATTCATGAAATAGCGCAAAATCCTGAAAAGTATGTTCAATAACTATTGCAAGACGTTCAAAGACGTACCGCCAGGCAATATGTTCCGATAAAAGCTTGGG
pKK-AzuC$_{STOP}$R	CCCAAGCTTTTATCGGAACATATTGCCTGGCGGTACGTCTTTGAACGTCTTGCAATAGTTATTGAACATACTTTTCAGGATTTTGCGCTATTTCATGAATTCCGG
pKK-GlpDF	TTTTTGGGCTAGCAGGAGATGGAAACCAAAGATCTGATTGTGATAGG
pKK-GlpDR	CTCATCCGCCAAAACAGCCAAGCTTTTACGACGCCAGCGATAACCTC
pXG10-SF-CadAF	GTTTTTTATGCATCTCCTTCGAGCTGGCAGGTA
pXG10-SF-CadAR	GTTTTTTGCTAGCACGGATGGGTTCTTCTTTAA
pXG10-SF-CadA-M1F	CCTGGAGATATGACTATGAATGTGATCGCGATCTTGAATCACATGGGGG
pXG10-SF-CadA-M1R	CCCCCATGTGATTCAAGATCGCGATCACATTCATAGTCATATCTCCAGG
pXG10-SF-CadA-M3F	CCTGGAGATATGACTATGAATGTTATTTTGATATTGAATCACATGGGGG
pXG10-SF-CadA-M3R	CCCCCATGTGATTCAATATCAAAATAACATTCATAGTCATATCTCCAGG
pXG10-SF-GalEF	GTTTTTTATGCAT ATCGCGCATAAAAAACGGCT
pXG10-SF-GalER	GTTTTTTGCTAGC GCCTACGCCAGCGCCGAGGT
cadA frag F	AATTCCATTTTGTCCCATGTGTTGGGAGGGGCCTTTTTTACCTGGAGATATGACTatgAACGTTATTGCAATATTGAATCACATGGGGGTTTATTTTAAAGA
cadA frag R	agcttCTTTAAAATAAACCCCCATGTGATTCAATATTGCAATAACGTTcatAGTCATATCTCCAGGTAAAAAAGGCCCCTCCCAACACATGGGACAAAATGG
cadA ctrl frag F	AATTCAGTTCCTCTGGTAATGCCGGGTGAAATGATCACCGAAGAAAGCCGTCCGGTTCTGGAGTTCCTGCAGATGCTGTGTGAAATCGGCGCTCACTATCCGA
cadA ctrl frag R	agcttCGGATAGTGAGCGCCGATTTCACACAGCATCTGCAGGAACTCCAGAACCGGACGGCTTTCTTCGGTGATCATTTCACCCGGCATTACCAGAGGAACTG
galE frag F	AATTCGGTACTGGCGTACGCGATTACATCCACGTAATGGATCTGGCGGACGGTCACGTCGTGGCGATGGAAAAACTGGCGAACAAGCCAGGCGTACACATCTACAA
galE frag R	agcttTGTAGATGTGTACGCCTGGCTTGTTCGCCAGTTTTTCCATCGCCACGACGTGACCGTCCGCCAGATCCATTACGTGGATGTAATCGCGTACGCCAGTACCG
galE ctrl frag F	aattcGCGACCTTCCGGCCTACTGGGCGGACGCCAGCAAAGCCGACCGTGAACTGAACTGGCGCGTAACGCGCACACTCGATGAAATGGa
galE ctrl frag R	agcttCCATTTCATCGAGTGTGCGCGTTACGCGCCAGTTCAGTTCACGGTCGGCTTTGCTGGCGTCCGCCCAGTAGGCCGGAAGGTCGCg
galR frag F	AATTCccgatccgtttttcggtgcaatggtgaaagcggtcgaacaggtggcttatcacaccggtaatttttttattgattggcaccggA

galR frag R	agcttccgttgccaatcaataaaaaattaccggtgtgataagccacctgttcgaccgctttcaccattgcaccgaaaaac ggatcggG
pRI FnrS F	AATTCGCAGGTGAATGCAACGTCAAGCGATGGGCGTTGCGCTCCATATT GTCTTACTTCCTTTTTTGAATTACTGCATAGCACAATTGATTCGTACGAC GCCGACTTTGATGAGTCGGCTTTTTA
pRI FnrS R	agcttAAAAAGCCGACTCATCAAAGTCGGCGTCGTACGAATCAATTGTGCT ATGCAGTAATTCAAAAAAGGAAGTAAGACAATATGGAGCGCAACGCCC ATCGCTTGACGTTGCATTCACCTGCG
cadA-LacZ F	ACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATCTCCTTCGAG CTGGCAGGTA
cadA-LacZ R	TAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACACGGATGG GTTCTTCTTTAA
pRI-Spot 42F	AATTCGTAGGGTACAGAGGTAAGATGTTCTATCTTTCAGACCTTTTACTT CACGTAATCGGATTTGGCTGAATATTTTAGCCGCCCCAGTCAGTAATGA CTGGGGCGTTTTTTAA
pRI-Spot 42R	AGCTTTAAAAAACGCCCCAGTCATTACTGACTGGGGCGGCTAAAATATT CAGCCAAATCCGATTACGTGAAGTAAAAGGTCTGAAAGATAGAACATC TTACCTCTGTACCCTACG
pRI-Spot 42$_{STOP}$F	CCGGAATTCGTAGGGTACAGAGGTAAGATGTTCTAGCTTTCAGACCTTT TACTTCACGTAATCGGATTTGGCTGAATATTTTAGCCGCCCCAGTCAGT AATGACTGGGGCGTTTTTTAAAGCTTG
pRI-Spot 42$_{STOP}$R	CAAGCTTTAAAAAACGCCCCAGTCATTACTGACTGGGGCGGCTAAAAT ATTCAGCCAAATCCGATTACGTGAAGTAAAAGGTCTGAAAGCTAGACA TCTTACCTCTGTACCCTACGAATTCCGG
pKK-SpfP-scramF	AATTCATGTTTTACTTAAGTGATTTACTTTTACATGTGATAGGCTTCGGA TAAA
pKK-SpfP-scramR	AGCTTTTATCCGAAGCCTATCACATGTAAAAGTAAATCACTTAAGTAAA ACATG
pKK-SpfP-scram$_{STOP}$F	AATTCATGTTTTAGTTAAGTGATTTACTTTTACATGTGATAGGCTTCGGA TAAA
pKK-SpfP-scram$_{STOP}$R	AGCTTTTATCCGAAGCCTATCACATGTAAAAGTAAATCACTTAACTAAA ACATG
pKK-SpfP-scram$_{F2A}$F	AATTCATGGCGTACTTAAGTGATTTACTTTTACATGTGATAGGCTTCGG ATAAA
pKK-SpfP-scram$_{F2A}$R	AGCTTTTATCCGAAGCCTATCACATGTAAAAGTAAATCACTTAAGTACG CCATG
pKK-SpfP-scram$_{Y3A}$F	AATTCATGTTTGCGTTAAGTGATTTACTTTTACATGTGATAGGCTTCGGA TAAA
pKK-SpfP-scram$_{Y3A}$R	AGCTTTTATCCGAAGCCTATCACATGTAAAAGTAAATCACTTAACGCAA ACATG
pKK-SpfP-scram$_{S5A}$F	AATTCATGTTTTACTTAGCGGATTTACTTTTACATGTGATAGGCTTCGGA TAAA
pKK-SpfP-scram$_{S5A}$R	AGCTTTTATCCGAAGCCTATCACATGTAAAAGTAAATCCGCTAAGTAAA ACATG
pKK-SpfP-scram$_{D6A}$F	AATTCATGTTTTACTTAAGTGCGTTACTTTTACATGTGATAGGCTTCGGA TAAA
pKK-SpfP-scram$_{D6A}$R	AGCTTTTATCCGAAGCCTATCACATGTAAAAGTAACGCACTTAAGTAAA ACATG
pKK-SpfP-scram$_{L8A}$F	AATTCATGTTTTACTTAAGTGATTTAGCGTTACATGTGATAGGCTTCGGA TAAA
pKK-SpfP-scram$_{L8A}$R	AGCTTTTATCCGAAGCCTATCACATGTAACGCTAAATCACTTAAGTAAA ACATG
pKK-SpfP-scram$_{H10A}$F	AATTCATGTTTTACTTAAGTGATTTACTTTTAGCGGTGATAGGCTTCGGA TAAA

pKK-SpfP-scram$_{H10A}$R	AGCTTTTATCCGAAGCCTATCACCGCTAAAAGTAAATCACTTAAGTAAA ACATG
pKK-SpfP-scram$_{V11A}$F	AATTCATGTTTTACTTAAGTGATTTACTTTTACATGCGATAGGCTTCGGA TAAA
pKK-SpfP-scram$_{V11A}$R	AGCTTTTATCCGAAGCCTATCGCATGTAAAAGTAAATCACTTAAGTAAA ACATG
pKK-SpfP-scram$_{F14A}$F	AATTCATGTTTTACTTAAGTGATTTACTTTTACATGTGATAGGCGCGGGA TAAA
pKK-SpfP-scram$_{F14A}$R	AGCTTTTATCCCGCGCCTATCACATGTAAAAGTAAATCACTTAAGTAAA ACATG
pKK-SpfP-scram$_{NY}$	CCGGAATTCATGTACTTTTACTTAAGTGATTTACTTTTACATGTGATAGG CTTCGGATAAAAGCTTGGG
pKK-SpfP-scram$_{CY}$	CCCAAGCTTTTATCCGAAGCCTATCACATGTAAAAGTAAATCACTTAAG TAAAAGTACATGAATTCCGG
pKK-SpfP-scram$_{NSTOP}$	CCGGAATTCATGTAGTTTTACTTAAGTGATTTACTTTTACATGTGATAGG CTTCGGATAAAAGCTTGGG
pKK-SpfP-scram$_{CSTOP}$	CCCAAGCTTTTATCCGAAGCCTATCACATGTAAAAGTAAATCACTTAAG TAAAACTACATGAATTCCGG
P1galElacZ F	GCGGATCCTACCTGACGCTTTTTAT
P1galElacZ R	GGGTAACGCCAGGGTTTTCCCAGTC
MgtSR gblock	CTAAGGAATTCatgCTGGGTAATATGAATGTTTTTATGGCCGTACTGGGA ATAATTTTATTTTCTGGTTTTCTGGCCGCGTATTTCAGCCACAAATGGGA TGACtaaATGCCTGTTAGCGTAAAAGCAAAACACAAATCTATCCATGCAA GCATTCACCGCCGGTTTACTGGCGGTTTTTTTTCAAGCTTTGTGG
MgtRS gblock	CTCCTGGAATTCGATTCGTTATCAGTGCAGGAAAATGCCTGTTAGCGTA AAAGCAAAACACAAATCTATCCATGCAAGCATTTTGAGATGAAAATTA AGGTAAGCGAGGAAACACACCACACCATAAACGGAGGCAAATAatgCTG GGTAATATGAATGTTTTTATGGCCGTACTGGGAATAATTTTATTTTCTGG TTTTCTGGCCGCGTATTTCAGCCACAAATGGGATGACtaaTGAACGGAGA CACCGCCGGTTTACTGGCGGTTTTTTTTCAAGCTTCAGCCA
MgtSR stop f	GGGCTAGCAGGAGGAATTCatgtaGGGTAATATGAATGTTTTTATG
MgtSR stop r	CATAAAAACATTCATATTACCCtacatGAATTCCTCCTGCTAGCCC
MgtSR_bp_f	CAAAACACAAATCTATCCATcgttGCATTCACCGCCGGTTTACTG
MgtSR_bp_r	CAGTAAACCGGCGGTGAATGCaacgATGGATAGATTTGTGTTTTG
MgtSR_nospace_f	AATTCatgCTGGGTAATATGAATGTTTTTATGGCCGTACTGGGAATAATTT TATTTTCTGGTTTTCTGGCCGCGTATTTCAGCCACAAATGGGATGACtaaT CTATCCATGCAAGCATTCACCGCCGGTTTACTGGCGGTTTTTTTTCA
MgtSR_nospace_r	AGCTTGAAAAAAAACCGCCAGTAAACCGGCGGTGAATGCTTGCATGGA TAGAttaGTCATCCCATTTGTGGCTGAAATACGCGGCCAGAAAACCAGAA AATAAAATTATTCCCAGTACGGCCATAAAAACATTCATATTACCCAGcat G
MgtSR_5pdel_-5_f	TGTTAGCGTAAAAGCAAAACA
MgtSR_5pdel_-10_f	GCGTAAAAGCAAAACACAAAT
MgtSR_5pdel_inv_r	ttaGTCATCCCATTTGTGGCT
MgtSR_3pdel_-5_r	GTTTTGCTTTTACGCTAACAG
MgtSR_3pdel_-10_r	GCTTTTACGCTAACAGGCA
MgtSR_3pdel_inv_f	TCTATCCATGCAAGCATTCA
Mgtrs_bp_f	GTAAAAGCAAAACACAAATCTATCGTACGTTGCATTTTGAGATG
Mgtrs_bp_r	CTTAATTTTCATCTCAAAATGCAACGTACGATAGATTTGTGTTTTGC
Mgtrs_stop_f	CATAAACGGAGGCAAATAatgTAGGGTAATATGAATGTTTTTATG
Mgtrs_stop_r	GTACGGCCATAAAAACATTCATATTACCCTAcatTATTTGCCTC

Appendix 5: Oligonucleotides for Chromosomal Tagging

Name	Sequence
MG1655 *azuC-GFP* 5' primer	TATTGCAAGACGTTCAAAGACGTACCGCCAGGCAATATGTTCCGA atgagtaaaggagaagaacttttca
MG1655 *azuC-GFP* 3' primer	CACCGTCATCAGGACACAAAAAAACCTGCCGGAGCAGGTTTTTTG attccggggatccgtcgacc
MG1655 *azuC-GFP*$_{IL6,7EE}$ 5' primer	ACCTGTATAACAAATGGTCGGAGTGCCGCGATGAAACTGCGCAA Agaagaaaaaagtatgttcaataacta
PM1205 *azuC::lacZ* 5' primer	ACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATaaagacaata acacctgtata
PM1205 *azuC::lacZ* 3' primer	TAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACtcggaaca tattgcctggcgg
PM1205 *azuC-III::lacZ* 5' primer	ACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATaaagacaata acaagtgtataacaaatggtcggag
GlpD-HA-HisF	GCTGGTGGAGTATACGCAGCAGAGGTTATCGCTGGCGTCGTACCC ATACGATGTTCCAGA
GlpD-HA-HisR	CAGGCCAGATTGAAATCTGACCTGATCACCTTACGTTAATCATAT GAATATCCTCCTTAG
SpfP-SPA F	CTATCTTTCAGACCTTTTACTTCACGTAATCGGATTTGGCTCCATG GAAAAGAGAAGATG
SpfP-SPA R	CAATAAAAAACGCCCCAGTCATTACTGACTGGGGCGGCTAAAAT ATATATGAATATCCTCCTTAG
GalK-HA-His6 F	GAGACTTTTTACGTTTGTAAACCATCACAAGGAGCAGGACAGTGC TACCCATACGATGTTCCAGA
GalK-HA-His6 R	CAGTCGGTACGGCTGACCATCGGGTGCCAGTGCGGGAGTTTCGTC ATATGAATATCCTCCTTAG
GalK-HA-His6 SeqF	TGACCGGCGGCGGATTTGGCGG
GalM-SPA F	GGCGAAGAGTATTCCAGCCTGACGGAATATCAGTTTATTGCTGAG TCCATGGAAAAGAGAAGATG
GalM-SPA R	GGTGATTTGAACAATATGAGATAAAGCCCTCATGACGAGGGCGT AACACATATGAATATCCTCCTTAG
GalM-SPA SeqF	GGTCAGCAGATGAAAAATTG
Crp-HA-His$_6$ F	CTCCGCACACGGTAAAACCATCGTCGTTTACGGCACTCGTTACCC ATACGATGTTCCAGA
Crp-HA-His$_6$ R	CAAAATGGCGCGCTACCAGGTAACGCGCCACTCCGACGGGACAT ATGAATATCCTCCTTA
Crp-HA-His$_6$ SeqF	CTGGCAAAACAACCAGACGCTATG
Spot42-stop-chromosome F	AAAGAGTAAAGTTAGTCGCgtaGGGTACAGAGGTAAGatgttctatctttcag acctttgACTTCACGTAATCGGATTTGGCTGAGAAGTTCCTATACTTTC TAG
Scramble-chromosome F	GCTTTCTGAACTGAACAAAAAAGAGTAAAGTTAGTCGCgAGAGGT AAGatgttttacttaagtgatttacttttacatgtgataggcttcggtaaGAAGTTCCTATACTTT CTAG
Chromosome R	GGCCTGATAAACATAGCGCATCAGGCATTACGGATCTTTTTCAGA AGAACTCGTCAAGAA

Appendix 6: Oligonucleotides for Sequencing

Name	Sequence
lacZ F	GCAATGACCATTGAACAGGCAGC
lacZ R	GCGATCGGTGCGGGCCTCTTCGCTA
pRI Seq F	CTGGCAAATATTCTGAAATGAGCTG
pKK Seq R	CACTACCATCGGCGCTACGGCGTTT
pBAD f	AAACCAGAAAATAAAATTAT
pBAD r	GTTTTTATGGCCGTACTGGG
MgtSR seq f	GTCACACTTTGCTATGCCAT
MgtSR seq r	ACCCCACACTACCATCGGCG
pXG10-SSeqF	TGGGATATATCAACGGTGGT

Appendix 7: Sequences of Synthetic Constructs

pBAD-MgtSR	ACCCGTTTTTTTGGGCTAGCAGGAGGAATTCatgCTGGGTAATATGAA TGTTTTTATGGCCGTACTGGGAATAATTTTATTTTCTGGTTTTCTGGCC GCGTATTTCAGCCACAAATGGGATGACtaaATGCCTGTTAGCGTAAA AGCAAAACACAAATCTATCCATGCAAGCATTCACCGCCGGTTTACTG GCGGTTTTTTTTC
pBAD-MgtSR$_{STOP}$	ACCCGTTTTTTTGGGCTAGCAGGAGGAATTCatgTAGGGTAATATGAA TGTTTTTATGGCCGTACTGGGAATAATTTTATTTTCTGGTTTTCTGGCC GCGTATTTCAGCCACAAATGGGATGACtaaATGCCTGTTAGCGTAAA AGCAAAACACAAATCTATCCATGCAAGCATTCACCGCCGGTTTACTG GCGGTTTTTTTTC
pBAD-MgtSR$_{BP}$	ACCCGTTTTTTTGGGCTAGCAGGAGGAATTCatgCTGGGTAATATGAA TGTTTTTATGGCCGTACTGGGAATAATTTTATTTTCTGGTTTTCTGGCC GCGTATTTCAGCCACAAATGGGATGACtaaATGCCTGTTAGCGTAAA AGCAAAACACAAATCTATCCATCGTTGCATTCACCGCCGGTTTACTGG CGGTTTTTTTTC
pBAD-MgtSR$_{5'\Delta 5nt}$	ACCCGTTTTTTTGGGCTAGCAGGAGGAATTCatgCTGGGTAATATGAA TGTTTTTATGGCCGTACTGGGAATAATTTTATTTTCTGGTTTTCTGGCC GCGTATTTCAGCCACAAATGGGATGACtaaTGTTAGCGTAAAAGCAA AACACAAATCTATCCATGCAAGCATTCACCGCCGGTTTACTGGCGGT TTTTTTTC
pBAD-MgtSR$_{5'\Delta 10nt}$	ACCCGTTTTTTTGGGCTAGCAGGAGGAATTCatgCTGGGTAATATGAA TGTTTTTATGGCCGTACTGGGAATAATTTTATTTTCTGGTTTTCTGGCC GCGTATTTCAGCCACAAATGGGATGACtaaGCGTAAAAGCAAAACAC AAATCTATCCATGCAAGCATTCACCGCCGGTTTACTGGCGGTTTTTTT TC
pBAD-MgtSR$_{3'\Delta 5nt}$	ACCCGTTTTTTTGGGCTAGCAGGAGGAATTCatgCTGGGTAATATGAA TGTTTTTATGGCCGTACTGGGAATAATTTTATTTTCTGGTTTTCTGGCC GCGTATTTCAGCCACAAATGGGATGACtaaATGCCTGTTAGCGTAAA AGCAAAACTCTATCCATGCAAGCATTCACCGCCGGTTTACTGGCGGT TTTTTTTC
pBAD-MgtSR$_{3'\Delta 10nt}$	ACCCGTTTTTTTGGGCTAGCAGGAGGAATTCatgCTGGGTAATATGAA TGTTTTTATGGCCGTACTGGGAATAATTTTATTTTCTGGTTTTCTGGCC GCGTATTTCAGCCACAAATGGGATGACtaaATGCCTGTTAGCGTAAA AGCTCTATCCATGCAAGCATTCACCGCCGGTTTACTGGCGGTTTTTTT TC
pBAD-MgtSR$_{5'\Delta 5nt\ STOP}$	ACCCGTTTTTTTGGGCTAGCAGGAGGAATTCatgTAGGGTAATATGAA TGTTTTTATGGCCGTACTGGGAATAATTTTATTTTCTGGTTTTCTGGCC GCGTATTTCAGCCACAAATGGGATGACtaaTGTTAGCGTAAAAGCAA AACACAAATCTATCCATGCAAGCATTCACCGCCGGTTTACTGGCGGT TTTTTTTC

pBAD-MgtSR$_{5'\Delta 10nt\ STOP}$	ACCCGTTTTTTTGGGCTAGCAGGAGGAATTCatgTAGGGTAATATGAA TGTTTTTATGGCCGTACTGGGAATAATTTTATTTTCTGGTTTTCTGGCC GCGTATTTCAGCCACAAATGGGATGACtaaGCGTAAAAGCAAAACAC AAATCTATCCATGCAAGCATTCACCGCCGGTTTACTGGCGGTTTTTTT TC
pBAD-MgtSR$_{3'\Delta 5nt\ STOP}$	ACCCGTTTTTTTGGGCTAGCAGGAGGAATTCatgTAGGGTAATATGAA TGTTTTTATGGCCGTACTGGGAATAATTTTATTTTCTGGTTTTCTGGCC GCGTATTTCAGCCACAAATGGGATGACtaaATGCCTGTTAGCGTAAA AGCAAAACTCTATCCATGCAAGCATTCACCGCCGGTTTACTGGCGGT TTTTTTTC
pBAD-MgtSR$_{3'\Delta 10nt\ STOP}$	ACCCGTTTTTTTGGGCTAGCAGGAGGAATTCatgTAGGGTAATATGAA TGTTTTTATGGCCGTACTGGGAATAATTTTATTTTCTGGTTTTCTGGCC GCGTATTTCAGCCACAAATGGGATGACtaaATGCCTGTTAGCGTAAA AGCTCTATCCATGCAAGCATTCACCGCCGGTTTACTGGCGGTTTTTTT TC
pBAD-MgtSR$_{NO\ SPACE}$	ACCCGTTTTTTTGGGCTAGCAGGAGGAATTCatgCTGGGTAATATGAA TGTTTTTATGGCCGTACTGGGAATAATTTTATTTTCTGGTTTTCTGGCC GCGTATTTCAGCCACAAATGGGATGACtaaTCTATCCATGCAAGCATT CACCGCCGGTTTACTGGCGGTTTTTTTTC
pBAD-MgtSR$_{NO\ SPACE\ STOP}$	ACCCGTTTTTTTGGGCTAGCAGGAGGAATTCatgTAGGGTAATATGAA TGTTTTTATGGCCGTACTGGGAATAATTTTATTTTCTGGTTTTCTGGCC GCGTATTTCAGCCACAAATGGGATGACtaaTCTATCCATGCAAGCATT CACCGCCGGTTTACTGGCGGTTTTTTTTC
pBAD-MgtRS	ACCCGTTTTTTTGGGCTAGCAGGAGGAATTCGATTCGTTATCAGTGC AGGAAAATGCCTGTTAGCGTAAAAGCAAAACACAAATCTATCCATGC AAGCATTTTGAGATGAAAATTAAGGTAAGCGAGGAAACACACCACA CCATAAACGGAGGAAAATAatgCTGGGTAATATGAATGTTTTTATGG CCGTACTGGGAATAATTTTATTTTCTGGTTTTCTGGCCGCGTATTTCA GCCACAAATGGGATGACtaaTGAACGGAGACACCGCCGGTTTACTGG CGGTTTTTTTT
pBAD-MgtRS$_{STOP}$	ACCCGTTTTTTTGGGCTAGCAGGAGGAATTCGATTCGTTATCAGTGC AGGAAAATGCCTGTTAGCGTAAAAGCAAAACACAAATCTATCCATGC AAGCATTTTGAGATGAAAATTAAGGTAAGCGAGGAAACACACCACA CCATAAACGGAGGAAAATAatgTAGGGTAATATGAATGTTTTTATGG CCGTACTGGGAATAATTTTATTTTCTGGTTTTCTGGCCGCGTATTTCA GCCACAAATGGGATGACtaaTGAACGGAGACACCGCCGGTTTACTGG CGGTTTTTTTT

References

Adams, P. P. and G. Storz (2020). "Prevalence of small base-pairing RNAs derived from diverse genomic loci." Biochim Biophys Acta Gene Regul Mech **1863**(7): 194524.

Aiba, H. (1983). "Autoregulation of the Escherichia coli crp gene: CRP is a transcriptional repressor for its own gene." Cell **32**(1): 141-149.

Anderson, D. M., K. M. Anderson, C. L. Chang, C. A. Makarewich, B. R. Nelson, J. R. McAnally, P. Kasaragod, J. M. Shelton, J. Liou, R. Bassel-Duby and E. N. Olson (2015). "A micropeptide encoded by a putative long noncoding RNA regulates muscle performance." Cell **160**(4): 595-606.

Anderson, D. M., C. A. Makarewich, K. M. Anderson, J. M. Shelton, S. Bezprozvannaya, R. Bassel-Duby and E. N. Olson (2016). "Widespread control of calcium signaling by a family of SERCA-inhibiting micropeptides." Sci Signal **9**(457): ra119.

Andrews, S. J. and J. A. Rothnagel (2014). "Emerging evidence for functional peptides encoded by short open reading frames." Nat Rev Genet **15**(3): 193-204.

Asahi, M., K. Kurzydlowski, M. Tada and D. H. MacLennan (2002). "Sarcolipin inhibits polymerization of phospholamban to induce superinhibition of sarco(endo)plasmic reticulum Ca2+-ATPases (SERCAs)." J Biol Chem **277**(30): 26725-26728.

Asahi, M., Y. Sugita, K. Kurzydlowski, S. De Leon, M. Tada, C. Toyoshima and D. H. MacLennan (2003). "Sarcolipin regulates sarco(endo)plasmic reticulum Ca2+-ATPase (SERCA) by binding to transmembrane helices alone or in association with phospholamban." Proc Natl Acad Sci U S A **100**(9): 5040-5045.

Aspden, J. L., Y. C. Eyre-Walker, R. J. Phillips, U. Amin, M. A. Mumtaz, M. Brocard and J. P. Couso (2014). "Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq." Elife **3**: e03528.

Baekkedal, C. and P. Haugen (2015). "The Spot 42 RNA: A regulatory small RNA with roles in the central metabolism." RNA Biol **12**(10): 1071-1077.

Balaban, N. and R. P. Novick (1995). "Translation of RNAIII, the Staphylococcus aureus agr regulatory RNA molecule, can be activated by a 3'-end deletion." FEMS Microbiol Lett **133**(1-2): 155-161.

Balasubramanian, D. and C. K. Vanderpool (2013). "Deciphering the interplay between two independent functions of the small RNA regulator SgrS in Salmonella." J Bacteriol **195**(20): 4620-4630.

Bashirullah, A., R. L. Cooperstock and H. D. Lipshitz (2001). "Spatial and temporal control of RNA stability." Proc Natl Acad Sci U S A **98**(13): 7025-7028.

Beisel, C. L. and G. Storz (2011). "The base-pairing RNA spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in Escherichia coli." Mol Cell **41**(3): 286-297.

Beisel, C. L., T. B. Updegrove, B. J. Janson and G. Storz (2012). "Multiple factors dictate target selection by Hfq-binding small RNAs." EMBO J **31**(8): 1961-1974.

Benito, Y., F. A. Kolb, P. Romby, G. Lina, J. Etienne and F. Vandenesch (2000). "Probing the structure of RNAIII, the Staphylococcus aureus agr regulatory RNA, and identification of the RNA domain involved in repression of protein A expression." RNA **6**(5): 668-679.

Bettegowda, A. and G. W. Smith (2007). "Mechanisms of maternal mRNA regulation: implications for mammalian early embryonic development." Front Biosci **12**: 3713-3726.

Bobrovskyy, M. and C. K. Vanderpool (2014). "The small RNA SgrS: roles in metabolism and pathogenesis of enteric bacteria." Front Cell Infect Microbiol **4**: 61.
Boisset, S., T. Geissmann, E. Huntzinger, P. Fechter, N. Bendridi, M. Possedko, C. Chevalier, A. C. Helfer, Y. Benito, A. Jacquier, C. Gaspin, F. Vandenesch and P. Romby (2007). "Staphylococcus aureus RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism." Genes Dev **21**(11): 1353-1366.
Brar, G. A. and J. S. Weissman (2015). "Ribosome profiling reveals the what, when, where and how of protein synthesis." Nat Rev Mol Cell Biol **16**(11): 651-664.
Bruckner, R. and F. Titgemeyer (2002). "Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization." FEMS Microbiol Lett **209**(2): 141-148.
Burenina, O. Y., T. S. Oretskaya and E. A. Kubareva (2017). "Non-Coding RNAs As Transcriptional Regulators In Eukaryotes." Acta Naturae **9**(4): 13-25.
Busby, S. and R. H. Ebright (1999). "Transcription activation by catabolite activator protein (CAP)." J Mol Biol **293**(2): 199-213.
Chatterjee, S. S., L. Chen, H. S. Joo, G. Y. Cheung, B. N. Kreiswirth and M. Otto (2011). "Distribution and regulation of the mobile genetic element-encoded phenol-soluble modulin PSM-mec in methicillin-resistant Staphylococcus aureus." PLoS One **6**(12): e28781.
Chen, S., E. A. Lesnik, T. A. Hall, R. Sampath, R. H. Griffey, D. J. Ecker and L. B. Blyn (2002). "A bioinformatics based approach to discover small RNA genes in the Escherichia coli genome." Biosystems **65**(2-3): 157-177.
Chin, J. W., S. W. Santoro, A. B. Martin, D. S. King, L. Wang and P. G. Schultz (2002). "Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli." J Am Chem Soc **124**(31): 9026-9027.
Corcoran, C. P., D. Podkaminski, K. Papenfort, J. H. Urban, J. C. Hinton and J. Vogel (2012). "Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA." Mol. Microbiol. **84**(3): 428-445.
De Lay, N. and S. Gottesman (2009). "The Crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior." J Bacteriol **191**(2): 461-476.
Derrien, T., R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D. G. Knowles, J. Lagarde, L. Veeravalli, X. Ruan, Y. Ruan, T. Lassmann, P. Carninci, J. B. Brown, L. Lipovich, J. M. Gonzalez, M. Thomas, C. A. Davis, R. Shiekhattar, T. R. Gingeras, T. J. Hubbard, C. Notredame, J. Harrow and R. Guigo (2012). "The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression." Genome Res **22**(9): 1775-1789.
Durand, S. and G. Storz (2010). "Reprogramming of anaerobic metabolism by the FnrS small RNA." Mol Microbiol **75**(5): 1215-1231.
Durica-Mitic, S., Y. Gopel and B. Gorke (2018). "Carbohydrate Utilization in Bacteria: Making the Most Out of Sugars with the Help of Small Regulatory RNAs." Microbiol Spectr **6**(2).
Dutcher, H. A. and R. Raghavan (2018). "Origin, Evolution, and Loss of Bacterial Small RNAs." Microbiol Spectr **6**(2).
Ephrussi, A., L. K. Dickinson and R. Lehmann (1991). "Oskar organizes the germ plasm and directs localization of the posterior determinant nanos." Cell **66**(1): 37-50.
Erez, Z., I. Steinberger-Levy, M. Shamir, S. Doron, A. Stokar-Avihail, Y. Peleg, S. Melamed, A. Leavitt, A. Savidor, S. Albeck, G. Amitai and R. Sorek (2017). "Communication between viruses guides lysis-lysogeny decisions." Nature **541**(7638): 488-493.

Evangelista, W., A. Dong, M. A. White, J. Li and J. C. Lee (2019). "Differential modulation of energy landscapes of cyclic AMP receptor protein (CRP) as a regulatory mechanism for class II CRP-dependent promoters." J Biol Chem **294**(42): 15544-15556.
Evans, T. C. and C. P. Hunter (2005). "Translational control of maternal RNAs." WormBook: 1-11.
Fallmann, J., S. Will, J. Engelhardt, B. Gruning, R. Backofen and P. F. Stadler (2017). "Recent advances in RNA folding." J Biotechnol **261**: 97-104.
Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver and C. C. Mello (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." Nature **391**(6669): 806-811.
Fontaine, F., R. T. Fuchs and G. Storz (2011). "Membrane localization of small proteins in Escherichia coli." J Biol Chem **286**(37): 32464-32474.
Friedman, R. C., S. Kalkhof, O. Doppelt-Azeroual, S. A. Mueller, M. Chovancova, M. von Bergen and B. Schwikowski (2017). "Common and phylogenetically widespread coding for peptides by bacterial small RNAs." BMC Genomics **18**(1): 553.
Frohlich, K. S., K. Papenfort, A. Fekete and J. Vogel (2013). "A small RNA activates CFA synthase by isoform-specific mRNA stabilization." EMBO J **32**(22): 2963-2979.
Geanacopoulos, M., G. Vasmatzis, D. E. Lewis, S. Roy, B. Lee and S. Adhya (1999). "GalR mutants defective in repressosome formation." Genes Dev **13**(10): 1251-1262.
Gebert, L. F. R. and I. J. MacRae (2019). "Regulation of microRNA function in animals." Nat Rev Mol Cell Biol **20**(1): 21-37.
Gimpel, M. and S. Brantl (2016). "Dual-function sRNA encoded peptide SR1P modulates moonlighting activity of B. subtilis GapA." RNA Biol **13**(9): 916-926.
Gimpel, M., N. Heidrich, U. Mader, H. Krugel and S. Brantl (2010). "A dual-function sRNA from B. subtilis: SR1 acts as a peptide encoding mRNA on the gapA operon." Mol Microbiol **76**(4): 990-1009.
Gimpel, M., H. Preis, E. Barth, L. Gramzow and S. Brantl (2012). "SR1--a small RNA with two remarkably conserved functions." Nucleic Acids Res **40**(22): 11659-11672.
Gore, A. V., S. Maegawa, A. Cheong, P. C. Gilligan, E. S. Weinberg and K. Sampath (2005). "The zebrafish dorsal axis is apparent at the four-cell stage." Nature **438**(7070): 1030-1035.
Gorke, B. and J. Stulke (2008). "Carbon catabolite repression in bacteria: many ways to make the most out of nutrients." Nat Rev Microbiol **6**(8): 613-624.
Gottesman, S. and G. Storz (2011). "Bacterial small RNA regulators: versatile roles and rapidly evolving variations." Cold Spring Harb Perspect Biol **3**(12).
Heidrich, N., I. Moll and S. Brantl (2007). "In vitro analysis of the interaction between the small RNA SR1 and its primary target ahrC mRNA." Nucleic Acids Res **35**(13): 4331-4346.
Hemm, M. R., B. J. Paul, J. Miranda-Rios, A. Zhang, N. Soltanzad and G. Storz (2010). "Small stress response proteins in Escherichia coli: proteins missed by classical proteomic studies." J Bacteriol **192**(1): 46-58.
Hemm, M. R., B. J. Paul, T. D. Schneider, G. Storz and K. E. Rudd (2008). "Small membrane proteins found by comparative genomics and ribosome binding site models." Mol Microbiol **70**(6): 1487-1501.
Hemm, M. R., J. Weaver and G. Storz (2020). "Escherichia coli Small Proteome." EcoSal Plus **9**(1).

Hobbs, E. C., X. Yin, B. J. Paul, J. L. Astarita and G. Storz (2012). "Conserved small protein associates with the multidrug efflux pump AcrB and differentially affects antibiotic resistance." Proc Natl Acad Sci U S A **109**(41): 16696-16701.
Holmqvist, E., L. Li, T. Bischler, L. Barquist and J. Vogel (2018). "Global Maps of ProQ Binding In Vivo Reveal Target Recognition via RNA Structure and Stability Control at mRNA 3' Ends." Mol Cell **70**(5): 971-982 e976.
Ikemura, T. and J. E. Dahlberg (1973). "Small ribonucleic acids of Escherichia coli. I. Characterization by polyacrylamide gel electrophoresis and fingerprint analysis." J Biol Chem **248**(14): 5024-5032.
Iwasaki, Y. W., M. C. Siomi and H. Siomi (2015). "PIWI-Interacting RNA: Its Biogenesis and Functions." Annu Rev Biochem **84**: 405-433.
Jenny, A., O. Hachet, P. Zavorszky, A. Cyrklaff, M. D. Weston, D. S. Johnston, M. Erdelyi and A. Ephrussi (2006). "A translation-independent role of oskar RNA in early Drosophila oogenesis." Development **133**(15): 2827-2833.
Kaito, C., Y. Saito, M. Ikuo, Y. Omae, H. Mao, G. Nagano, T. Fujiyuki, S. Numata, X. Han, K. Obata, S. Hasegawa, H. Yamaguchi, K. Inokuchi, T. Ito, K. Hiramatsu and K. Sekimizu (2013). "Mobile genetic element SCCmec-encoded psm-mec RNA suppresses translation of agrA and attenuates MRSA virulence." PLoS Pathog **9**(4): e1003269.
Kaito, C., Y. Saito, G. Nagano, M. Ikuo, Y. Omae, Y. Hanada, X. Han, K. Kuwahara-Arai, T. Hishinuma, T. Baba, T. Ito, K. Hiramatsu and K. Sekimizu (2011). "Transcription and translation products of the cytolysin gene psm-mec on the mobile genetic element SCCmec regulate Staphylococcus aureus virulence." PLoS Pathog **7**(2): e1001267.
Kawano, M., A. A. Reynolds, J. Miranda-Rios and G. Storz (2005). "Detection of 5'- and 3'-UTR-derived small RNAs and cis-encoded antisense RNAs in Escherichia coli." Nucleic Acids Res **33**(3): 1040-1050.
Kery, M. B., M. Feldman, J. Livny and B. Tjaden (2014). "TargetRNA2: identifying targets of small regulatory RNAs in bacteria." Nucleic Acids Res **42**(Web Server issue): W124-129.
Keseler, I. M., A. Mackie, A. Santos-Zavaleta, R. Billington, C. Bonavides-Martinez, R. Caspi, C. Fulcher, S. Gama-Castro, A. Kothari, M. Krummenacker, M. Latendresse, L. Muniz-Rascado, Q. Ong, S. Paley, M. Peralta-Gil, P. Subhraveti, D. A. Velazquez-Ramirez, D. Weaver, J. Collado-Vides, I. Paulsen and P. D. Karp (2017). "The EcoCyc database: reflecting new knowledge about Escherichia coli K-12." Nucleic Acids Res **45**(D1): D543-D550.
Kloc, M., K. Wilk, D. Vargas, Y. Shirato, S. Bilinski and L. D. Etkin (2005). "Potential structural role of non-coding and coding RNAs in the organization of the cytoskeleton at the vegetal cortex of Xenopus oocytes." Development **132**(15): 3445-3457.
Kolb, A., S. Busby, H. Buc, S. Garges and S. Adhya (1993). "Transcriptional regulation by cAMP and its receptor protein." Annu Rev Biochem **62**: 749-795.
Kugler, J. M. and P. Lasko (2009). "Localization, anchoring and translational control of oskar, gurken, bicoid and nanos mRNA during Drosophila oogenesis." Fly (Austin) **3**(1): 15-28.
Kwiatkowska, J., Z. Wroblewska, K. A. Johnson and M. Olejniczak (2018). "The binding of Class II sRNA MgrR to two different sites on matchmaker protein Hfq enables efficient competition for Hfq and annealing to regulated mRNAs." RNA **24**(12): 1761-1784.
Lam, M. T., W. Li, M. G. Rosenfeld and C. K. Glass (2014). "Enhancer RNAs and regulated transcriptional programs." Trends Biochem Sci **39**(4): 170-182.
Lasko, P. (2009). "Translational control during early development." Prog Mol Biol Transl Sci **90**: 211-254.

Lehmann, R. and C. Nusslein-Volhard (1986). "Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of oskar, a maternal gene in Drosophila." Cell **47**(1): 141-152.
Licht, A., S. Preis and S. Brantl (2005). "Implication of CcpN in the regulation of a novel untranslated RNA (SR1) in Bacillus subtilis." Mol Microbiol **58**(1): 189-206.
Lim, S., P. Kumari, P. Gilligan, H. N. Quach, S. Mathavan and K. Sampath (2012). "Dorsal activity of maternal squint is mediated by a non-coding function of the RNA." Development **139**(16): 2903-2915.
Liu, M. Y., G. Gui, B. Wei, J. F. Preston, 3rd, L. Oakford, U. Yuksel, D. P. Giedroc and T. Romeo (1997). "The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in Escherichia coli." J Biol Chem **272**(28): 17502-17510.
Lloyd, C. R., S. Park, J. Fei and C. K. Vanderpool (2017). "The Small Protein SgrT Controls Transport Activity of the Glucose-Specific Phosphotransferase System." J Bacteriol **199**(11).
Mackowiak, S. D., H. Zauber, C. Bielow, D. Thiel, K. Kutz, L. Calviello, G. Mastrobuoni, N. Rajewsky, S. Kempa, M. Selbach and B. Obermayer (2015). "Extensive identification and analysis of conserved small ORFs in animals." Genome Biol **16**: 179.
Madeira, F., Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. N. Tivey, S. C. Potter, R. D. Finn and R. Lopez (2019). "The EMBL-EBI search and sequence analysis tools APIs in 2019." Nucleic Acids Res **47**(W1): W636-W641.
Mahadevan, M., C. Tsilfidis, L. Sabourin, G. Shutler, C. Amemiya, G. Jansen, C. Neville, M. Narang, J. Barcelo, K. O'Hoy and et al. (1992). "Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene." Science **255**(5049): 1253-1255.
Maki, K., T. Morita, H. Otaka and H. Aiba (2010). "A minimal base-pairing region of a bacterial small RNA SgrS required for translational repression of ptsG mRNA." Mol Microbiol **76**(3): 782-792.
Mallick, U. and P. Herrlich (1979). "Regulation of synthesis of a major outer membrane protein: cyclic AMP represses Escherichia coli protein III synthesis." Proc Natl Acad Sci U S A **76**(11): 5520-5523.
Mandin, P. and S. Gottesman (2009). "A genetic approach for finding small RNAs regulators of genes of interest identifies RybC as regulating the DpiA/DpiB two-component system." Mol. Microbiol. **72**(3): 551-565.
Mann, M., P. R. Wright and R. Backofen (2017). "IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions." Nucleic Acid Res. **45**(W1): W435-W439.
Mayba, O., H. N. Gilbert, J. Liu, P. M. Haverty, S. Jhunjhunwala, Z. Jiang, C. Watanabe and Z. Zhang (2014). "MBASED: allele-specific expression detection in cancer tissues and cell lines." Genome Biol **15**(8): 405.
Melamed, S., P. P. Adams, A. Zhang, H. Zhang and G. Storz (2020). "RNA-RNA Interactomes of ProQ and Hfq Reveal Overlapping and Competing Roles." Mol Cell **77**(2): 411-425 e417.
Melamed, S., A. Peer, R. Faigenbaum-Romm, Y. E. Gatt, N. Reiss, A. Bar, Y. Altuvia, L. Argaman and H. Margalit (2016). "Global Mapping of Small RNA-Target Interactions in Bacteria." Mol Cell **63**(5): 884-897.
Mercer, T. R., M. E. Dinger and J. S. Mattick (2009). "Long non-coding RNAs: insights into functions." Nat Rev Genet **10**(3): 155-159.
Meydan, S., J. Marks, D. Klepacki, V. Sharma, P. V. Baranov, A. E. Firth, T. Margus, A. Kefi, N. Vazquez-Laslop and A. S. Mankin (2019). "Retapamulin-Assisted Ribosome Profiling Reveals the Alternative Bacterial Proteome." Mol Cell **74**(3): 481-493 e486.

Michalowski, S., J. W. Miller, C. R. Urbinati, M. Paliouras, M. S. Swanson and J. Griffith (1999). "Visualization of double-stranded RNAs from the myotonic dystrophy protein kinase gene and interactions with CUG-binding protein." Nucleic Acids Res **27**(17): 3534-3542.
Mól, A. R., M. Castro, S. and W. Fontes (2019). "NetWheels: A web application to create high quality peptide helical wheel and net projections. ." bioRxiv **416347**: doi: https://doi.org/10.1101/416347
Moller, T., T. Franch, C. Udesen, K. Gerdes and P. Valentin-Hansen (2002). "Spot 42 RNA mediates discoordinate expression of the E. coli galactose operon." Genes Dev **16**(13): 1696-1706.
Moon, K. and S. Gottesman (2009). "A PhoQ/P-regulated small RNA regulates sensitivity of Escherichia coli to antimicrobial peptides." Mol Microbiol **74**(6): 1314-1330.
Neiditch, M. B., G. C. Capodagli, G. Prehna and M. J. Federle (2017). "Genetic and Structural Analyses of RRNPP Intercellular Peptide Signaling of Gram-Positive Bacteria." Annu Rev Genet **51**: 311-333.
Nelson, B. R., C. A. Makarewich, D. M. Anderson, B. R. Winders, C. D. Troupes, F. Wu, A. L. Reese, J. R. McAnally, X. Chen, E. T. Kavalali, S. C. Cannon, S. R. Houser, R. Bassel-Duby and E. N. Olson (2016). "A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle." Science **351**(6270): 271-275.
Nelson, D. L. and E. P. Kennedy (1971). "Magnesium transport in Escherichia coli. Inhibition by cobaltous ion." J Biol Chem **246**(9): 3042-3049.
Novick, R. P., H. F. Ross, S. J. Projan, J. Kornblum, B. Kreiswirth and S. Moghazeh (1993). "Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule." EMBO J **12**(10): 3967-3975.
Olejniczak, M. and G. Storz (2017). "ProQ/FinO-domain proteins: another ubiquitous family of RNA matchmakers?" Molecular Microbiology **104**(6): 905-915.
Opdyke, J. A., J. G. Kang and G. Storz (2004). "GadY, a small-RNA regulator of acid response genes in Escherichia coli." J Bacteriol **186**(20): 6698-6705.
Orr, M. W., Y. Mao, G. Storz and S. B. Qian (2020). "Alternative ORFs and small ORFs: shedding light on the dark proteome." Nucleic Acids Res **48**(3): 1029-1042.
Papenfort, K., Y. Sun, M. Miyakoshi, C. K. Vanderpool and J. Vogel (2013). "Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis." Cell **153**(2): 426-437.
Papenfort, K. and J. Vogel (2014). "Small RNA functions in carbon metabolism and virulence of enteric pathogens." Front Cell Infect Microbiol **4**: 91.
Polayes, D. A., P. W. Rice, M. M. Garner and J. E. Dahlberg (1988). "Cyclic AMP-cyclic AMP receptor protein as a repressor of transcription of the spf gene of Escherichia coli." J Bacteriol **170**(7): 3110-3114.
Prost, L. R. and S. I. Miller (2008). "The Salmonellae PhoQ sensor: mechanisms of detection of phagosome signals." Cell Microbiol **10**(3): 576-582.
Queck, S. Y., B. A. Khan, R. Wang, T. H. Bach, D. Kretschmer, L. Chen, B. N. Kreiswirth, A. Peschel, F. R. Deleo and M. Otto (2009). "Mobile genetic element-encoded cytolysin connects virulence to methicillin resistance in MRSA." PLoS Pathog **5**(7): e1000533.
Quek, X. C., D. W. Thomson, J. L. Maag, N. Bartonicek, B. Signal, M. B. Clark, B. S. Gloss and M. E. Dinger (2015). "lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs." Nucleic Acids Res **43**(Database issue): D168-173.

Raina, M., J. Aoyama, S. Bhatt, B. J. Paul and G. Storz (2020). "AzuCR RNA modulates carbon metabolism as both an mRNA encoding a 28-amino acid protein and as a base pairing small RNA." submitted.
Raina, M., A. King, C. Bianco and C. K. Vanderpool (2018). "Dual-Function RNAs." Microbiol Spectr **6**(5).
Raina, M. and G. Storz (2017). "SgrT, a Small Protein That Packs a Sweet Punch." J Bacteriol **199**(11).
Ramamurthi, K. S. and G. Storz (2014). "The small protein floodgates are opening; now the functional analysis begins." BMC Biol **12**: 96.
Rebagliati, M. R., R. Toyama, C. Fricke, P. Haffter and I. B. Dawid (1998). "Zebrafish nodal-related genes are implicated in axial patterning and establishing left-right asymmetry." Dev Biol **199**(2): 261-272.
Rhoads, D. B., P. C. Tai and B. D. Davis (1984). "Energy-requiring translocation of the OmpA protein and alkaline phosphatase of Escherichia coli into inner membrane vesicles." J Bacteriol **159**(1): 63-70.
Rice, J. B. and C. K. Vanderpool (2011). "The small RNA SgrS controls sugar-phosphate accumulation by regulating multiple PTS genes." Nucleic Acids Res **39**(9): 3806-3819.
Rice, P. W., D. A. Polayes and J. E. Dahlberg (1987). "Spot 42 RNA of Escherichia coli is not an mRNA." J Bacteriol **169**(8): 3850-3852.
Robinson, J. J. and J. H. Weiner (1980). "The effect of amphipaths on the flavin-linked aerobic glycerol-3-phosphate dehydrogenase from Escherichia coli." Can J Biochem **58**(10): 1172-1178.
Rowlett, V. W., V. Mallampalli, A. Karlstaedt, W. Dowhan, H. Taegtmeyer, W. Margolin and H. Vitrac (2017). "Impact of Membrane Phospholipid Alterations in Escherichia coli on Cellular Function and Bacterial Stress Adaptation." J Bacteriol **199**(13).
Saghatelian, A. and J. P. Couso (2015). "Discovery and characterization of smORF-encoded bioactive polypeptides." Nat Chem Biol **11**(12): 909-916.
Sahagan, B. G. and J. E. Dahlberg (1979). "A small, unstable RNA molecule of Escherichia coli: spot 42 RNA. II. Accumulation and distribution." J Mol Biol **131**(3): 593-605.
Schier, A. F. (2009). "Nodal morphogens." Cold Spring Harb Perspect Biol **1**(5): a003459.
Schryvers, A., E. Lohmeier and J. H. Weiner (1978). "Chemical and functional properties of the native and reconstituted forms of the membrane-bound, aerobic glycerol-3-phosphate dehydrogenase of Escherichia coli." J Biol Chem **253**(3): 783-788.
Shaikh, S. A., S. K. Sahoo and M. Periasamy (2016). "Phospholamban and sarcolipin: Are they functionally redundant or distinct regulators of the Sarco(Endo)Plasmic Reticulum Calcium ATPase?" J Mol Cell Cardiol **91**: 81-91.
Shen, M. M. (2007). "Nodal signaling: developmental roles and regulation." Development **134**(6): 1023-1034.
Shimada, T., N. Fujita, K. Yamamoto and A. Ishihama (2011). "Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources." PLoS One **6**(6): e20081.
Singh, D. R., M. P. Dalton, E. E. Cho, M. P. Pribadi, T. J. Zak, J. Seflova, C. A. Makarewich, E. N. Olson and S. L. Robia (2019). "Newly Discovered Micropeptide Regulators of SERCA Form Oligomers but Bind to the Pump as Monomers." J Mol Biol **431**(22): 4429-4443.
Soberon-Chavez, G., L. D. Alcaraz, E. Morales, G. Y. Ponce-Soto and L. Servin-Gonzalez (2017). "The Transcriptional Regulators of the CRP Family Regulate Different Essential Bacterial Functions and Can Be Inherited Vertically and Horizontally." Front Microbiol **8**: 959.

Storz, G., J. Vogel and K. M. Wassarman (2011). "Regulation by small RNAs in bacteria: expanding frontiers." Mol Cell **43**(6): 880-891.
Storz, G., Y. I. Wolf and K. S. Ramamurthi (2014). "Small proteins can no longer be ignored." Annu Rev Biochem **83**: 753-777.
Tadros, W. and H. D. Lipshitz (2005). "Setting the stage for development: mRNA translation and stability during oocyte maturation and egg activation in Drosophila." Dev Dyn **232**(3): 593-608.
Takyar, S., R. P. Hickerson and H. F. Noller (2005). "mRNA helicase activity of the ribosome." Cell **120**(1): 49-58.
Updegrove, T. B., S. A. Shabalina and G. Storz (2015). "How do base-pairing small RNAs evolve?" FEMS Microbiol Rev **39**(3): 379-391.
Updegrove, T. B., A. Zhang and G. Storz (2016). "Hfq: the flexible RNA matchmaker." Curr Opin Microbiol **30**: 133-138.
Urban, J. H. and J. Vogel (2009). "A green fluorescent protein (GFP)-based plasmid system to study post-transcriptional control of gene expression in vivo." Methods Mol. Biol. **540**: 301-319.
Vanderpool, C. K. and S. Gottesman (2004). "Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system." Mol Microbiol **54**(4): 1076-1089.
Verdon, J., N. Girardin, C. Lacombe, J. M. Berjeaud and Y. Hechard (2009). "delta-hemolysin, an update on a membrane-interacting peptide." Peptides **30**(4): 817-823.
Wadler, C. S. and C. K. Vanderpool (2007). "A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide." Proc Natl Acad Sci U S A **104**(51): 20454-20459.
Wagner, E. G. H. and P. Romby (2015). "Small RNAs in bacteria and archaea: who they are, what they do, and how they do it." Adv Genet **90**: 133-208.
Walz, A. C., R. A. Demel, B. de Kruijff and R. Mutzel (2002). "Aerobic sn-glycerol-3-phosphate dehydrogenase from Escherichia coli binds to the cytoplasmic membrane through an amphipathic alpha-helix." Biochem J **365**(Pt 2): 471-479.
Wang, H., X. Yin, M. Wu Orr, M. Dambach, R. Curtis and G. Storz (2017). "Increasing intracellular magnesium levels with the 31-amino acid MgtS protein." Proc Natl Acad Sci U S A **114**(22): 5689-5694.
Wang, X., S. C. Ji, S. H. Yun, H. J. Jeon, S. W. Kim and H. M. Lim (2014). "Expression of each cistron in the gal operon can be regulated by transcription termination and generation of a galk-specific mRNA, mK2." J Bacteriol **196**(14): 2598-2606.
Wassarman, K. M., F. Repoila, C. Rosenow, G. Storz and S. Gottesman (2001). "Identification of novel small RNAs using comparative genomics and microarrays." Genes Dev **15**(13): 1637-1651.
Weaver, J., F. Mohammad, A. R. Buskirk and G. Storz (2019). "Identifying Small Proteins by Ribosome Profiling with Stalled Initiation Complexes." mBio **10**(2).
Wegener, M., K. Vogtmann, M. Huber, S. Laass and J. Soppa (2016). "The glpD gene is a novel reporter gene for E. coli that is superior to established reporter genes like lacZ and gusA." J Microbiol Methods **131**: 181-187.
Wessels, H. H., S. Lebedeva, A. Hirsekorn, R. Wurmus, A. Akalin, N. Mukherjee and U. Ohler (2019). "Global identification of functional microRNA-mRNA interactions in Drosophila." Nat Commun **10**(1): 1626.

Yeh, J. I., U. Chinte and S. Du (2008). "Structure of glycerol-3-phosphate dehydrogenase, an essential monotopic membrane enzyme involved in respiration and metabolism." Proc Natl Acad Sci U S A **105**(9): 3280-3285.
Yu, D., H. M. Ellis, E. C. Lee, N. A. Jenkins, N. G. Copeland and D. L. Court (2000). "An efficient recombination system for chromosome engineering in Escherichia coli." Proc Natl Acad Sci U S A **97**(11): 5978-5983.
Yu, D., H. M. Ellis, E. C. Lee, N. A. Jenkins, N. G. Copeland and D. L. Court (2000). "An efficient recombination system for chromosome engineering in *Escherichia coli*." Proc. Natl. Acad. Sci. USA **97**(11): 5978-5983.
Zhang, A., K. M. Wassarman, J. Ortega, A. C. Steven and G. Storz (2002). "The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs." Mol. Cell **9**(1): 11-22.
Zheng, D., C. Constantinidou, J. L. Hobman and S. D. Minchin (2004). "Identification of the CRP regulon using in vitro and in vivo transcriptional profiling." Nucleic Acids Res **32**(19): 5874-5893.