Optimization of a genetic assay for detection of expanded RNA-Hfq interactions

by

Hannah LeBlanc

A Paper Presented to the Faculty of Mount Holyoke College

in Partial Fulfillment of the Requirements

for the Degree of Bachelors of Arts with Honor

Program in Biochemistry,

Department of Chemistry

South Hadley, MA 01075

May 2020

This paper was prepared

under the direction of

Professor Katie Berry

for eight credits.

Acknowledgements

First and foremost, I want to thank my thesis advisor, Dr. Katie Berry, for all the support and everything you've taught me over the last several years. Thank you for welcoming me into your lab, for helping me discover my love of research, and for believing in me as a scientist. Thank you for all the opportunities I've had as part of your lab and for all the time you've spent mentoring and guiding me. I'm so grateful for the time I've been able to spend learning from you in class, in meetings, at conferences, and at the bench. I could not have asked for a better or more rewarding research experience and I am so grateful for everything you've given me.

Thank you to everyone I've shared the Berry Lab with – alumni who I learned from when I was a new member, those of you who I've been able to spend years working alongside, and new members who I'm excited to see grow from afar. Kelly, Lillia, Clara, Courtney, Smriti, Amaka, Chandra, Rachel, Anh, Oliver, Salina, Katherine, and Linh: thank you for being such thoughtful, intelligent, and fun people to work with. Thank you as well to the members of the Camp, Lijek, and McMenimen labs who have broadened my horizons by sharing their work, and allowed me to share my work with them. I can't wait to keep learning from all of you and watching you grow.

Thank you to the professors who have guided me during the last four years. While many professors who have played important roles in my time at Mount Holyoke, I would like in particular to thank Dr. Amy Camp as my advisor and a member of my thesis committee for her steady and reassuring mentorship, her thoughtful feedback on my research at lab meetings, and her enthusiasm for microbiology. I would also like to thank Dr. Nina Emery for being a member of my thesis committee, for encouraging my interest in philosophy, and for taking the time to be part of the culmination of my undergraduate research.

Thank you to my family – my mom, dad, and Hayley – for all their support and for letting me share my excitement about biochemistry with them. Thank you as well to my rowing family. Thank you Seth and Cara for your wisdom, your fantastic coaching, and for giving me a team where I could throw myself into both rowing and research. Thank you to my teammates for their support of my academic pursuits and the community they've given me, especially as my time on campus came to a close. To give the thanks they deserve to those I've rowed with would take an acknowledgements section at least as long as this thesis, but to thank a few people in particular: thank you to Liz for walking my research journey with me and for all the time in the 2x. Thank you Molly for your quiet strength and inspiring me with your science to write about my own. Thank you Lainie for teaching me how to have fun even when I'm working hard. Casey, thank you for the editing help, growth, and happiness. I love you all. Thank you so much.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

α-	N-terminal domain of α subunit of RNA polymerase
β-gal	β-galactosidase
B2H	bacterial two-hybrid
B3H	bacterial three-hybrid
DNA	deoxyribonucleic acid
K_d	dissociation constant
λCI	bacteriophage λCI protein
mRNA	messenger RNA
MS2^{CP}	bacteriophage MS2 coat protein
MS2^{hp}	21-nucleotide cognate RNA hairpin of MS2^{CP}
ORF	open reading frame
RBS	ribosome binding site
RNA	ribonucleic acid
RNAP	RNA polymerase
rRNA	ribosomal RNA
sRNA	small RNA
tmRNA	transfer-messenger RNA
tRNA	transfer RNA
UTR	untranslated region
Y2H	yeast two-hybrid
Y3H	yeast three-hybrid

Table 1. Amino acid residues, codes, and ionization states at physiological pH.

Amino Acid	Three-Letter Code	One-Letter Code	Side chain charge at pH 7.4
Alanine	Ala	A	Neutral
Arginine	Arg	R	Positive
Asparagine	Asn	N	Neutral
Aspartic Acid	Asp	D	Negative
Cysteine	Cys	C	Neutral
Glutamine	Glu	Q	Neutral
Glutamic Acid	Gln	E	Negative
Glycine	Gly	G	Neutral
Histidine	His	H	Neutral
Isoleucine	Ile	I	Neutral
Leucine	Leu	L	Neutral
Lysine	Lys	K	Positive
Methionine	Met	M	Neutral
Phenylalanine	Phe	F	Neutral
Proline	Pro	P	Neutral
Serine	Ser	S	Neutral
Threonine	Thr	T	Neutral
Tryptophan	Trp	W	Neutral
Tyrosine	Tyr	Y	Neutral
Valine	Val	V	Neutral

Abstract

Non-coding small RNAs (sRNAs) contribute to bacterial biofilm formation, antibiotic resistance, pathogenesis, and virulence by regulating gene expression. However, there is much we do not understand about their molecular mechanisms, which often involve facilitation by a chaperone protein. Hfq is the best-studied bacterial RNA chaperone protein and has become the paradigmatic example for how RNA-binding proteins facilitate sRNA-based gene regulation. An *in vivo* bacterial three-hybrid assay for genetic detection of RNA-protein interactions has previously been established using interactions using *Escherichia coli* Hfq and its sRNA substrates. In order to broaden the utility of the assay to provide a deep understanding of varied RNA-protein interactions, I have focused on expanding the B3H assay to detect mRNA-Hfq interactions and on understanding the energetic implications of the data provided by the assay. In this work, I demonstrate that the B3H assay is capable of detecting Hfq's interactions with sequence elements typical for mRNA-Hfq binding. Expansion from minimal binding elements to native sequences has involved exploration of the impacts of translation and turnover via mRNA surveillance pathways. While these factors complicate the detection of interactions with 5' UTRs and coding sequences, I have refined the B3H assay to detect the interaction between Hfq and its mRNA target *sodB*, demonstrating the potential of the assay to detect a wider variety of RNA-protein interactions. Just as it is useful to detect a wide variety of interactions using this method, it is also imperative to understand what the assay's data truly indicates about these interactions. I have therefore compared B3H and *in vitro* binding data for a set of Hfq-interacting RNAs and Hfq variants to work toward a model for the relationship between B3H data and binding energetics. My results suggest that high B3H signal typically reflects strong binding affinity, and that B3H optimization efforts have resulted in data that is increasingly consistent with binding energetics. The work I present here represents significant steps forward in both the assay's detection capacity and our understanding of that capacity, developing it as a more broadly useful and informative method for studying RNA-protein interactions.

Chapter I: Introduction

I-1. Gene expression and the central dogma of molecular biology

The process of gene expression – decoding information from genes into final molecular products – is central to all known forms of life. Storage of genetic information in DNA is universal across all domains of life, and the decoding of genes into RNA or proteins is necessary for the utilization of this information. The process known as the central dogma of molecular biology describes the basic steps necessary for decoding this information (Fig. 1). At the simplest level, there are two steps necessary for gene expression: transcription of DNA into RNA, and translation of RNA into protein (reviewed in 19).
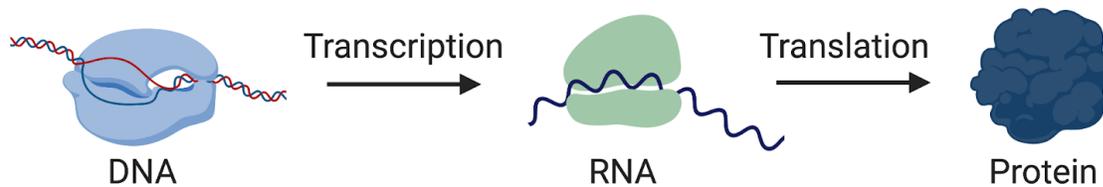
Figure 1. The central dogma of molecular biology. As described by the central dogma, genetic information is stored in DNA and expressed by transcription from DNA to RNA by the enzyme RNA polymerase, then translation from RNA to protein by the ribosome. Figure made in BioRender.

I-1-i. Transcription

Transcription is the process of building an RNA strand based on a DNA template with the enzyme RNA polymerase (RNAP), using complementary base pairing to determine the correct sequence of RNA nucleotides. At the 5' end of a gene, RNA polymerase binds at a sequence known as a promoter and constructs a strand of RNA

based on the sequence templated by DNA, adding ribonucleotides complementary to the DNA sequence to the 3' end of the new RNA strand. This process is depicted in Fig. 2. In bacteria, this processive transcription continues until RNAP encounters either a ρ-dependent terminator encoded by the template DNA in the presence of the protein ρ (Rho), or a ρ-independent terminator, at which point RNAP releases the DNA strand and the newly formed RNA transcript (reviewed in 19).
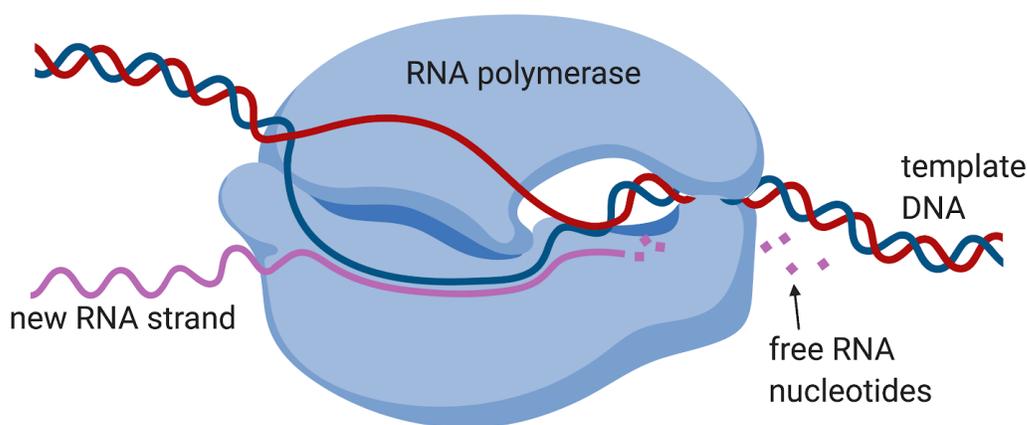
Figure 2. Transcription of DNA into RNA. RNA polymerase pairs complementary RNA nucleotides to a template DNA sequence to build an RNA transcript. Figure made in BioRender.

While the central dogma of molecular biology focuses on RNA that is translated into protein, there are a number of types of RNA, only one of which encodes protein sequence information. Messenger RNA, or mRNA, carries information from a protein-coding gene that will be translated into protein in the next step of the central dogma, as well as 5' and 3' untranslated regions (UTRs) that may play roles in regulating its translation. Two classical types of non-coding RNA play a role in translation: transfer RNA (tRNA), which decodes mRNA sequence into the appropriate amino acids for protein production, and ribosomal RNA (rRNA), which along with ribosomal proteins

makes up the ribosome (the cellular machine that carries out translation) (reviewed in 19). Additional non-coding RNA categories include short regulatory RNAs such as eukaryotic small interfering RNAs and microRNAs (siRNAs and miRNAs), prokaryotic small RNAs (sRNAs), and transfer-messenger RNA (tmRNA) which is involved in bacterial identification and degradation of defective mRNAs. Fig. 3 shows these various types of RNA.
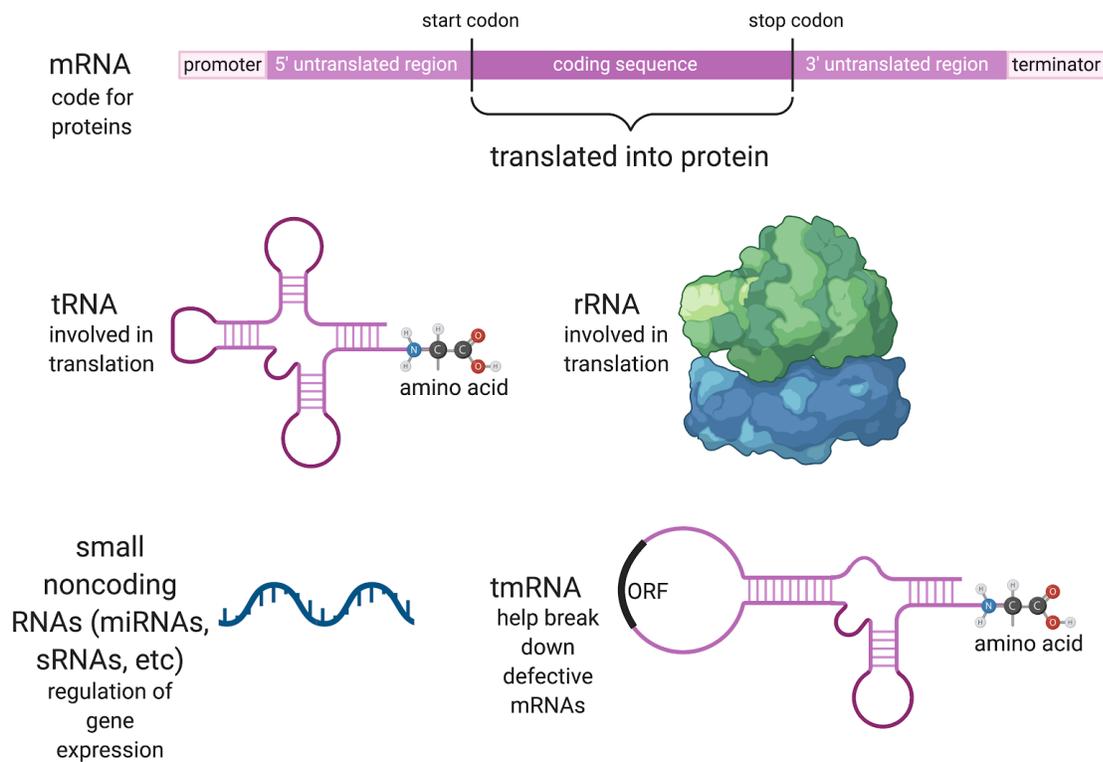
Figure 3. Types of RNA. Most RNAs discussed in this study are mRNAs, which encode protein sequences, and sRNAs, which are the bacterial version of small noncoding RNAs. Also significant are tRNA and rRNA, which are necessary for translation, and tmRNA, which contains both mRNA and tRNA components and is involved in the process of breaking down mRNA fragments. Figure made using BioRender.

I-1-ii. Translation

The second step in gene expression, translation, involves decoding the protein-coding sequence of an mRNA into a polypeptide. An overview of this process is shown in Fig. 4.
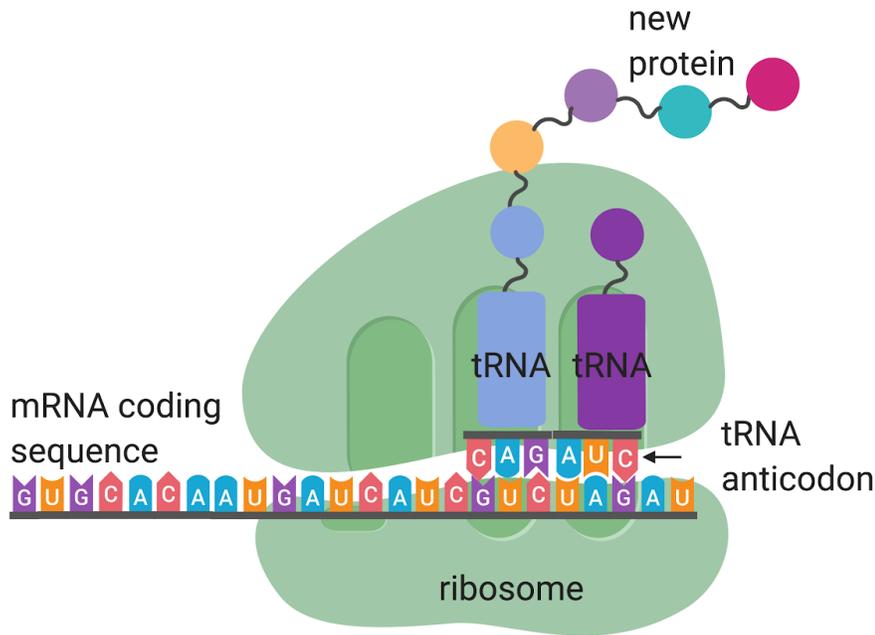
Figure 4. Translation of an mRNA into a protein. The ribosome, a large RNA-protein complex, carries out the steps of translation. It proceeds along an mRNA coding sequence, matching tRNA anticodons with their complementary codons and forming covalent bonds between the amino acids carried by each tRNA to build a new protein. Figure made using BioRender.

The coding section of the mRNA, known as the open reading frame (ORF), carries information about protein sequence in which each set of three nucleotides, known as a codon, denotes a single amino acid. To read these codons as a protein sequence, a tRNA molecule carries a particular amino acid and recognizes the mRNA codon for that amino acid. To begin the process of bacterial translation, the ribosome first binds to a site upstream of the ORF known as the ribosome binding site (RBS), which includes an

element known as the Shine-Dalgarno sequence and a start codon. This process is shown in Fig. 5.
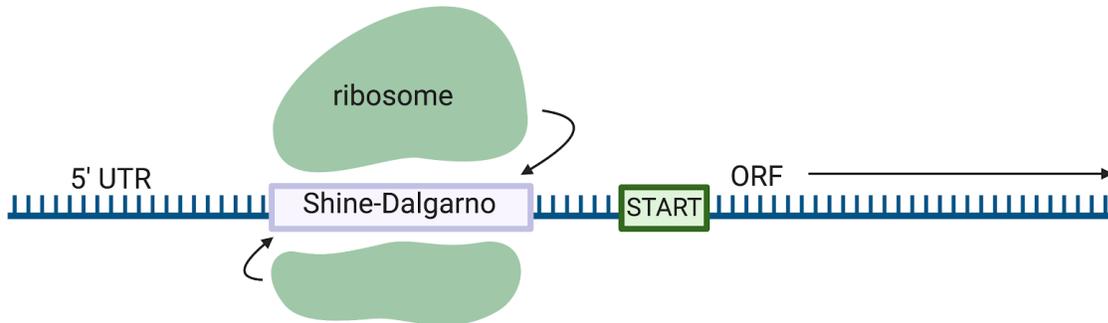
Figure 5. Translation initiation on a bacterial mRNA. The two subunits of the ribosome bind to the Shine-Dalgarno sequence upstream of the ORF and close to the start codon, before starting to translate the ORF into a polypeptide. Upstream of the RBS is the 5' UTR, which is not translated into protein. Figure made in BioRender.

The ribosome then proceeds along the ORF to each successive codon, decoding the mRNA into a sequence of amino acids and chaining those amino acids together to produce a polypeptide. When the ribosome reaches a stop codon, it disengages from the mRNA transcript and releases its polypeptide. The ribosome can then translate another mRNA molecule, and the mRNA transcript can be translated again. Most mRNA molecules are translated by multiple ribosomes at any one point in time, and bacterial translation can occur co-transcriptionally (that is, translation can occur while an mRNA is still being transcribed) (reviewed in 19).

I-2. Regulation and quality control of gene expression in bacteria

No organism uses all of its genes at once. At any given time, only a subset of an organism's genome is relevant to its environment or developmental stage, and expression of unnecessary genes can be detrimental by creating additional demands on the

organism's energy and resources. Therefore, all organisms must express the genes necessary for their current situation – for instance, to respond to stresses like a pathogen, lack of resources, or physical damage; or to adapt their metabolism appropriately to a changing environment – while avoiding expression of unneeded genes. There are a number of stages at which gene expression can be controlled: at transcription initiation; at translation initiation; via degradation of mRNA; via post-translational modification of proteins; during protein localization or transport; or via protein degradation (reviewed in 19). Here, transcriptional and regulated RNA turnover modes of regulation will be focused on.

I-2-i. Transcriptional control

Control of gene expression at the transcriptional level primarily (but not exclusively) involves regulating the extent to which transcription is initiated – that is, whether or not RNA polymerase binds and begins to transcribe a DNA sequence into RNA. If a gene is not transcribed, none of the further steps of expression can happen; if it is transcribed, it may go through the full expression process if expression is not prevented at a later step. In bacteria, much of this transcriptional initiation control is carried out by proteins known as σ factors, which act as a subunit of RNAP to direct binding and transcriptional initiation at a particular set of genes. Individual σ factors may govern a stress response or the set of genes necessary for a particular stage of bacterial growth, for example. In *E. coli*, one example of a σ factor is σ^S, also known as RpoS. This σ factor regulates a large number of genes involved in general stress response (15). In addition to

σ factors, transcriptional initiation can be modulated by proteins known as activators (which activate transcription) and repressors (which block transcription) (reviewed in 19).

I-2-ii. mRNA degradation and stability

The stability of an mRNA transcript – *i.e.*, how long a single mRNA molecule persists in the cell without being broken down – plays an important role in post-transcriptional regulation (reviewed in 31). In bacteria, mRNA transcripts may have half-lives between 40 seconds and an hour. The longer an mRNA persists, the more it can be translated into protein (reviewed in 23). When a transcript is degraded, no more protein can be produced from its ORF, so degradation is the most permanent type of post-transcriptional regulation. Due to the significant role of RNA degradation in controlling gene expression, there are multiple paths a cell may take to break down an mRNA molecule. One step in this pathway is endoribonucleolytic cleavage, often by the enzyme RNase E, in which enzymes cut RNA strands at specific locations within their sequence (reviewed in 23). mRNAs are additionally broken down into individual nucleotides by exoribonucleases, which remove nucleotides from one end of the RNA (reviewed in 23). These enzymes preferentially degrade single-stranded RNA sections: some are only capable of breaking down single-stranded sequences, while others have limited ability to melt secondary structure and can be aided in taking apart double-stranded RNA by the addition of a poly(A) tail at the 3' end of an mRNA

(reviewed in 23). Fig. 6 shows the way that endo- and exoribonucleolytic enzymes play differing roles in prokaryotic mRNA degradation.
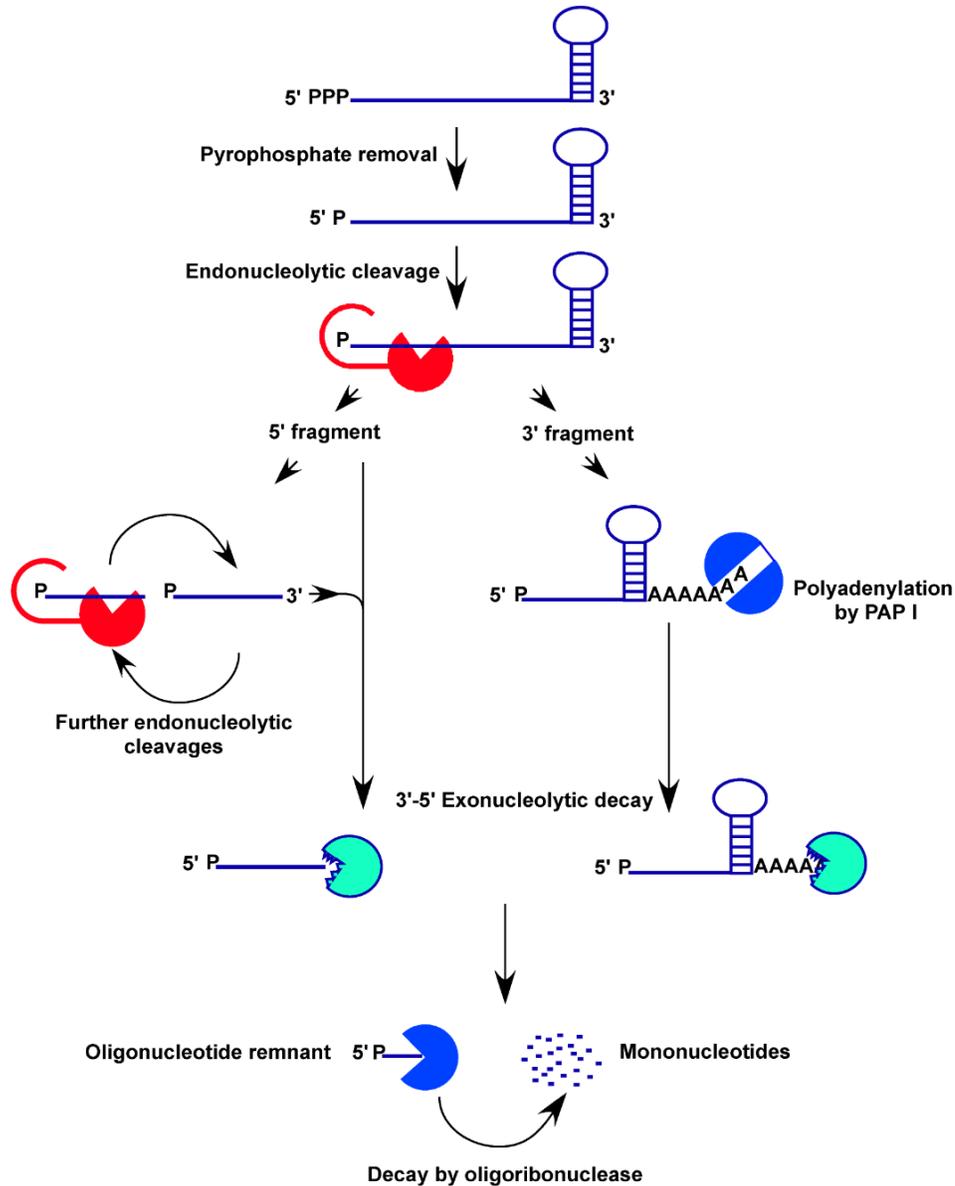
Figure 6. mRNA degradation pathway in *E. coli*. Endonucleases (RNAse E, red) break down RNAs at defined cleavage sites, creating fragments that can be further degraded into individual nucleotides by exonucleases. Polyadenylation by PAP I (blue) assists in degradation by marking mRNA fragments for breakdown and assisting exonucleases in melting secondary structures. Figure adapted from Richards *et al*. 2008 (23).

I-2-ii-a. mRNA surveillance and non-stop decay

In addition to more general pathways for degradation of mRNA transcripts, organisms have specific ways to break down defective mRNAs. Here, one type of defective mRNA transcript will be focused on: molecules that contain a ribosomal binding site but lack a stop codon. This type of mRNA, called a non-stop transcript, is capable of loading the ribosome and beginning translation, but since it lacks a stop codon, the ribosome will reach the 3' end of the transcript without disengaging from the mRNA. In this situation, the ribosome stalls and is unable to leave the mRNA without assistance, which sequesters some of the cell's translational machinery in an unproductive situation (reviewed in 31). However, degradation of actively translated mRNAs is challenging for the typical mRNA breakdown pathways, as ribosomes bound to an RNA block endoribonucleases from cleaving the transcript, thus increasing the defective mRNA's stability (reviewed in 31). To overcome the challenges presented by non-stop mRNAs, all bacterial species use a degradation pathway involving transfer-messenger RNA (tmRNA). An overview of this process is shown in Fig. 6. tmRNA is an RNA that includes both tRNA- and mRNA-like components (reviewed in 23 and 31). It has both an amino acid acceptor stem that carries an alanine, similar to a regular alanine tRNA molecule, and an mRNA-like ORF that encodes (in *E. coli*) the protein degradation tag ANDENYALAA followed by a stop codon (reviewed in 23). tmRNA makes use of the protein small protein B (SmpB) to assist in recognizing and interacting with stalled ribosomes (reviewed in 23).

The pathway by which tmRNA recognizes stalled ribosomes, rescues them, and promotes decay of the defective mRNA and protein is known as *trans*-translation. In this process, SmpB and tmRNA recognize a stalled ribosome, and tmRNA is charged with an alanine molecule. tmRNA binds to the ribosome as if it were the next tRNA to be decoded from the mRNA, and the nascent polypeptide is transferred to tmRNA's tRNA-like domain and the alanine it carries. The stalled ribosome then disengages from the original mRNA molecule and binds to the mRNA-like domain of tmRNA, which it translates until reaching the tmRNA stop codon. The ribosome can then disengage from the tmRNA and releases its degradation-tagged polypeptide, and the SmpB-tmRNA complex facilitates degradation of the non-stop mRNA (reviewed in 23). This process is shown in Fig. 7.
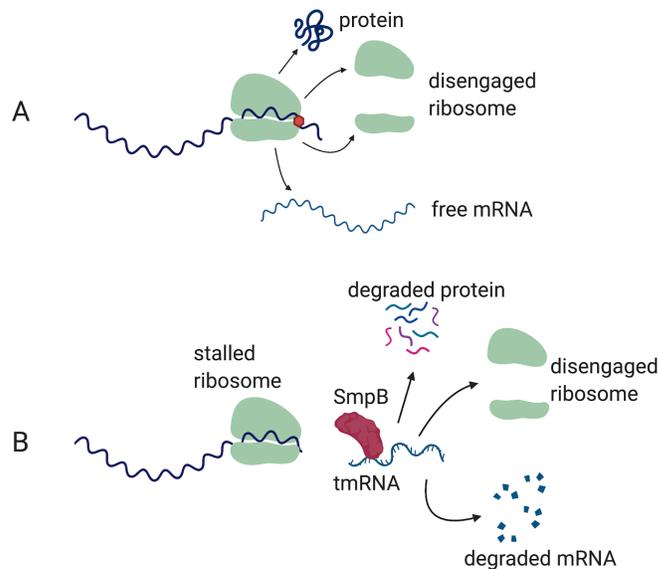
Figure 7. Normal translation termination and *trans*-translation of non-stop mRNAs. (A) A translating ribosome reaches the end of an mRNA coding sequence, encounters a stop codon (red octagon), and disengages from the mRNA, releasing a newly formed polypeptide. (B) A translating ribosome reaches the end of an mRNA transcript without encountering a stop codon and stalls. The stalled ribosome is recognized by SmpB and tmRNA, which transfer the ribosome to the tmRNA's open reading frame and carry out *trans*-translation. The defective mRNA and nascent polypeptide are degraded and the ribosome is able to disengage from the RNA using the tmRNA's stop codon. Figure made using BioRender.

I-2-iii. Translational control

Just as gene expression can be regulated based on initiation of transcription and the extent to which the mRNA persists in the cell, it can also be controlled at translation initiation. The amount of ribosome binding to an mRNA transcript will directly affect the amount of protein produced, and since a single mRNA transcript may be translated multiple times, regulation at this stage adds a further level of control in addition to regulation of transcription initiation. Furthermore, since translation is only one step removed from a protein product, regulating at this stage allows an organism to respond to a signal quickly. Translation initiation is often determined by the accessibility of the RBS – translation can only occur if the ribosome is able to bind to the beginning of the ORF. This accessibility can be controlled by the secondary structure of the 5' UTR of the mRNA transcript, by interactions with regulatory proteins or RNAs, or a combination of several of these factors. The impacts of some regulatory RNAs and proteins will be explored in more detail below.

I-2-iii-a. Bacterial small RNAs

One type of control that plays a significant role in bacterial post-transcriptional regulation is carried out by small RNAs. sRNAs are non-coding RNA transcripts that are significantly shorter than a typical mRNA and act in a regulatory capacity by base-pairing with target mRNAs to modulate their translation or stability (reviewed in 32). There are two categories of sRNAs: *cis*-encoded sRNAs that are encoded on the same genetic locus as the mRNA they regulate, which are fully complementary to their

mRNA target; and *trans*-encoded sRNAs that are found elsewhere in the genome and typically have partial complementarity to their regulatory target (reviewed in 32). While both types of sRNAs play regulatory roles, *trans*-encoded sRNAs will be focused on here. Regulation by this type of sRNA is often negative (*i.e.* results in decreased expression of the target mRNA) (reviewed in 32). Negative sRNA-mediated regulation may use one or both of two mechanisms: the sRNA typically base-pairs to the 5' UTR of the mRNA target to block the ribosome binding site and prevent translation, and sRNA-mRNA base pairing may also facilitate degradation by the enzyme RNAse E (reviewed in 32). In cases where sRNA activity increases translation rather than preventing it, base pairing between the sRNA and its mRNA target disrupts a secondary structure of the mRNA that blocks the ribosome binding site to make it accessible for translation initiation (reviewed in 32). In many cases, sRNA regulation of mRNA expression is facilitated by an RNA chaperone protein, the best-known of which – Hfq – will be discussed below.

I-2-iv. Hfq, a bacterial RNA chaperone protein

Hfq (host factor Qβ), originally discovered as a protein necessary for replication of the bacteriophage Qβ (6), is a homohexameric bacterial RNA chaperone protein. It is homologous to the Sm and LSm family of proteins found in eukaryotes, which are involved in eukaryotic RNA splicing and decay (reviewed in 32).

Broadly speaking, Hfq facilitates base-pairing interactions between a *trans*-encoded sRNA and its mRNA target, although Hfq is also known to regulate

translation for at least one mRNA independent of any sRNA activity (7). There are two models by which Hfq may regulate gene expression, which are not mutually exclusive. First, Hfq may increase local concentrations of an sRNA and mRNA, which facilitates base pairing when the two RNAs have complementarity (24,32). It also can actively melt or remodel secondary structure, changing the sequence available both for sRNA base pairing and for ribosome binding (24,32). In fact, it is this structural remodeling that earned Hfq its label as a chaperone protein (5,17). As part of its role as in negative regulation, Hfq may recruit degradation machinery, although the mechanism by which this occurs is not well understood (1).

As with sRNA regulation in general, Hfq-mediated regulation is typically negative (Fig. 8), although a notable exception to this trend is regulation of the *rpoS* mRNA. Translation of this gene's protein product, the stationary phase sigma factor σ^S, requires Hfq for translation activation. Deletion of Hfq was found to closely mimic deletion of σ^S before Hfq's regulatory activity was fully understood (reviewed in 30). In the case of *rpoS* and other positively regulated mRNAs, 5' UTRs usually contain inhibitory secondary structures that are remodeled only upon binding of Hfq and sRNAs to make the ribosome binding site accessible (22). Due to Hfq's wide range of mRNA targets, its ability to regulate more than multiple mRNAs with one sRNA, and its impacts on global regulators like σ^S, deletion of Hfq has widespread effects (5). As Hfq plays a role in regulating a variety of targets, it is significant in multiple aspects of bacterial physiology including areas relevant to human health such as virulence, biofilm formation, and drug resistance (reviewed in 24,31).
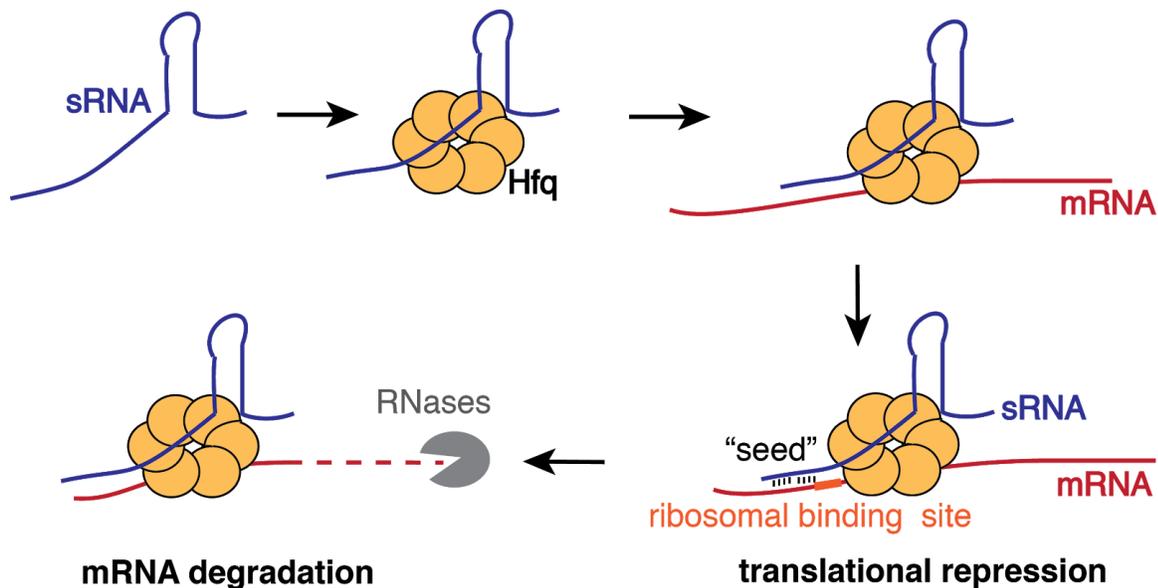
Figure 8. Pathway of Hfq-mediated negative regulation by an sRNA. Hfq binds both an sRNA and its mRNA target, facilitating base pairing between the two RNAs that blocks the ribosomal binding site and/or leads to degradation of the mRNA. Both processes prevent translation and therefore reduce protein production. Figure made by Katie Berry.

I-2-iv-a. Hfq-RNA interaction mechanisms and Hfq structure

In order to bind two RNAs – an sRNA and an mRNA – simultaneously, Hfq has multiple regions that act as RNA-binding faces. Hfq's primary binding surfaces are known as the proximal face, the distal face, and the rim or lateral face (reviewed in 29). Additionally, Hfq has a disordered C-terminal tail that may act in an autoregulatory capacity (25). Figure 8 shows structural models of Hfq's three binding faces in *E. coli* and *Staphylococcus aureus* (5).
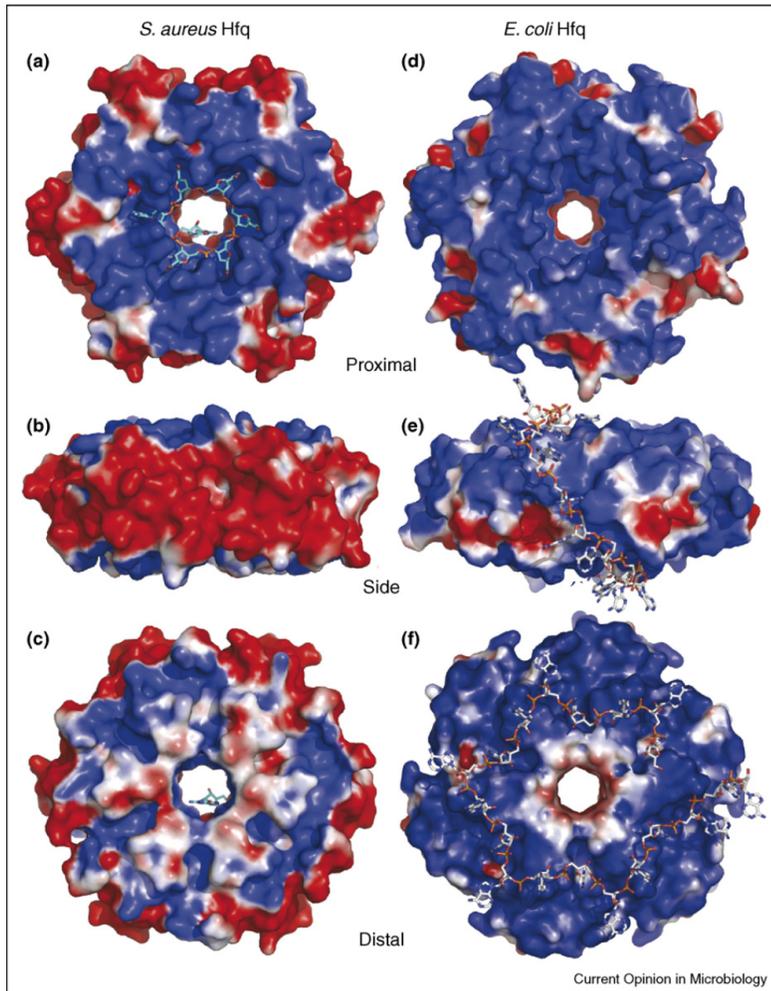
Figure 9. Binding surfaces of *S. aureus* and *E. coli* Hfq. Structures (a) and (d) show the proximal face of Hfq, and (a) demonstrates the way the proximal face binds RNAs close to Hfq's central pore. Structures (b) and (e) show the rim or lateral face, and shows how an RNA might bind to both the proximal and distal face by crossing the rim. Structures (c) and (f) show the distal face of Hfq, and (d) shows the binding of an RNA to the distal face in three-nucleotide repeats dependent on hydrophobic adenosine-binding pockets. Coloring represents electrostatics: blue indicates positive charge, red indicates negative charge, and gray regions are electrostatically neutral. Figure adapted from Brennan and Link 2007 (5).

Each of Hfq's faces binds different nucleotides or motifs in its target RNAs, in keeping with Hfq's ability to bind multiple types of RNA molecules. While the binding profiles of Hfq varies between bacterial species, *E. coli* Hfq will be focused on here. The proximal face of Hfq recognizes uridine-rich sequences near stem-loop structures (17) This sequence-structure combination is found in the intrinsic (ρ-independent) terminators

28

of sRNAs, making the proximal face the classical sRNA-binding face of Hfq. As shown in Fig. 9A, the proximal face can bind six nucleotides in a ring close to the central pore of the protein (5).

The distal face of Hfq is commonly thought of as the mRNA-binding face, and this face in fact plays a role in two different types of Hfq-RNA binding: binding to 5'UTRs of mRNA transcripts, and binding to poly(A) tails in the standard RNA degradation pathway (see Section I-2-ii). Hfq uses one mechanism to bind each of these types of sequences: it binds three-nucleotide repeats primarily via specificity for an adenine in the first position of the repeat (17,26). The RNA sequences bound by the distal face have been described as $(A-R-N)_n$ motifs, in which the A is an adenine, the R is a purine (adenine or guanine) and the N is any nucleotide (17,26). The first A in the $(A-R-N)_n$ motif is bound in a hydrophobic pocket formed by (among others) residues Tyr25 and Ile30 shown in Figure 10 (5).
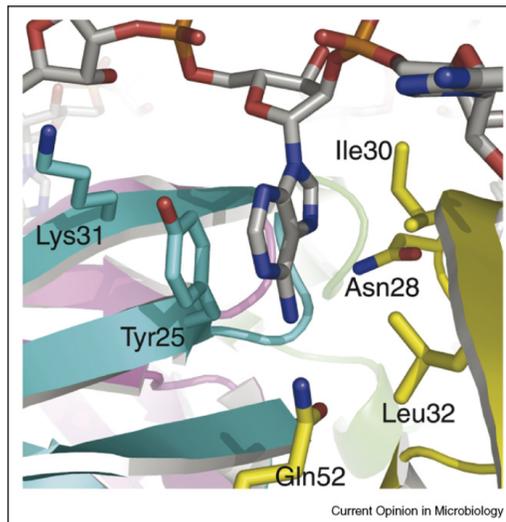
Figure 10. Adenine-binding pocket of *E. coli* Hfq's distal face. Structure shows a hypothetical interaction between the distal face of Hfq and the first adenine of an $(A-R-N)_n$ motif in a poly(A) RNA sequence. The adenine base is believed to interact with Tyr25 via π stacking interactions, while Lys31 interacts with the negatively charged RNA backbone electrostatically. Figure adapted from Brennan and Link 2007.

The second nucleotide in the $(A-R-N)_n$ motif has been theoretically proposed to fit into a binding site that may accommodate an adenine or a guanine nucleotide, as the structure of the binding pocket appears to be able to fit either base. However, experimental evidence has only supported adenine binding as being physiologically relevant (24). Thus, Hfq may be more accurately described as preferentially binding an $(A-A-N)_n$ motif rather than $(A-R-N)_n$. The third nucleotide in the motif may be any of the four nucleotides, as the binding region for this part of the sequence (known as the entrance/exit site) can accommodate any nucleotide (18,26). In poly(A) tails, the three-nucleotide repeats bound by the distal face are AAA triplets; $(A-R-N)_n/(A-A-N)_n$ motifs are found in the 5' UTRs of mRNAs targeted for Hfq-mediated regulation.

The third and final RNA-binding surface of Hfq, the rim or lateral face, binds to UA-rich sequences and may be involved in chaperone and RNA-base pairing activity carried out by Hfq (26). This region of the protein has a patch of positively charged residues – three arginines at positions 16, 17, and 19 – that are involved in interactions with rim-binding RNA sequences (reviewed in 29). While the proximal face is often considered the primary sRNA-binding face, and the distal face the primary mRNA-binding face, the rim also plays a role in Hfq-RNA binding, and RNAs may bind to surface other than their canonical binding region. The variation in RNA-Hfq binding is described in the Class I/II model, in which most sRNAs can be divided into one of two classes (although some fall into an intermediate category) (26). Class I sRNAs utilize classical proximal face binding and depend secondarily on the rim surface (26). In *E. coli*, most sRNAs fall into this category, and sRNAs that bind in this way target mRNAs

that bind to the distal face via $(A\text{-}R\text{-}N)_n/(A\text{-}A\text{-}N)_n$ motifs (26). Class II sRNAs bind to the proximal and distal face, and contain $(A\text{-}R\text{-}N)_n/(A\text{-}A\text{-}N)_n$ motifs in single-stranded regions to support their distal face binding (26). As these sRNAs take up the distal face when they bind, their target mRNAs must bind elsewhere, and indeed the target mRNAs of Class II sRNAs tend to lack the $(A\text{-}R\text{-}N)_n$ sequences found in their Class I-targeted counterparts (26). These motifs and their binding surfaces in Hfq are summarized in Table 2.

Table 2. Summary of Hfq-binding motifs in RNA. RNA types in bold indicate the type of RNA considered to bind primarily to that surface.

Hfq surface	Binds to	Binding motif/nucleotides
Proximal	**sRNAs** (Class I and Class II)	U-rich sequences, especially U tail of sRNA terminators
Distal	**mRNAs**, sRNAs (Class II)	$(A\text{-}A\text{-}N)_n/(A\text{-}R\text{-}N)_n$ motifs
Rim	sRNAs (Class I), mRNAs	A- and U-rich sequences

I-3. Traditional methods for studying RNA-protein interactions

I-3-i. Molecular recognition

As is clear from discussions of biomolecular interactions thus far, noncovalent interactions between molecules are central to important life processes. These interactions fall under the umbrella of molecular recognition, which includes interactions between biological macromolecules and with small molecules that have a high degree of affinity and specificity (reviewed in 8). Molecular recognition is involved in all of the processes discussed previously: transcription, translation, and various modes of gene regulation all involve specific interactions between various biomolecules. Here, we'll consider the case

of a protein and another molecule. The molecule a protein recognizes, called its ligand, may be a small molecule, DNA, RNA, or another protein (reviewed in 8).

Protein-ligand binding equilibria can be described in a way similar to a chemical reaction:

$$P + L \rightleftharpoons PL$$

where P represents the protein, L the ligand, and PL the protein-ligand complex. The rates at which this binding occurs in the forward and reverse directions are described by the rate constants k_{on} and k_{off}, which are values describing binding and dissociation rates of the protein-ligand complex (reviewed in 8). The binding affinity of an interaction can be described in terms of these constants using binding or dissociation constants. The following relationships are true for protein-ligand interactions in general:

$$K_a = \frac{k_{on}}{k_{off}} = \frac{[PL]}{[P][L]} = \frac{1}{K_d}$$

where K_a is the association constant, which describes the forward binding interaction and has units of concentration^{-1}. K_d refers to the dissociation constant, which describes the unbinding between the protein and the ligand and has units of concentration. The value of the dissociation constant describes the concentration of ligand necessary for 50% of a given amount of protein to form a protein-ligand complex (Fig. 11), and is frequently used to describe the binding affinity for the complex. A high-affinity interaction has a low K_d (and therefore a high K_a), requiring only a small amount ligand for a substantial amount of interaction to occur; lower affinity interactions have greater K_d values and therefore require a higher concentration of ligand for a substantial number of protein-ligand complexes to form.
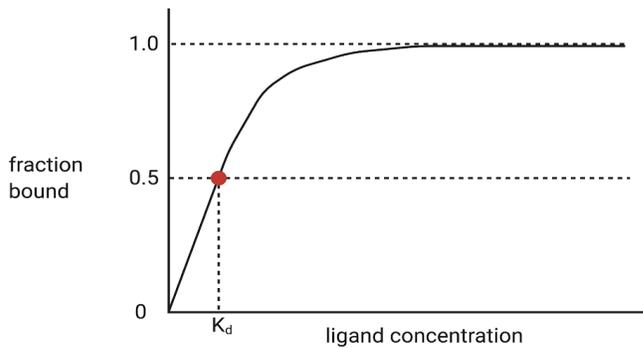
Figure 11. Hypothetical protein binding curve used to determine K_d. A protein-binding ligand is titrated into a constant protein concentration and the fraction of total protein bound to the ligand is determined at each ligand concentration. A standard protein-ligand binding curve is hyperbolic. The dissociation constant for the protein-ligand binding is determined by finding the ligand concentration at which half the total protein is bound to the ligand (fraction bound = 0.5). Figure made using BioRender.

Association and dissociation constants provide information about the energetics of a protein-ligand interaction. Energetics in biological systems are generally described in terms of Gibbs free energy, a thermodynamic value that describes a system's ability to do work at constant temperature and pressure (reviewed in 8). In terms of biomolecular interactions, the relative Gibbs free energy difference between unbound and bound states is important in determining binding affinity. An interaction is only favorable if it has a lower Gibbs free energy in its bound state than its unbound state. In this case, the interaction has a negative ΔG. The lower the ΔG, the more favorable the interaction; an interaction with a positive ΔG is energetically unfavorable and will not occur spontaneously. ΔG and the dissociation constant for a given equilibrium are related by the following expression, which includes temperature (T) and the ideal gas constant (R):

$$\Delta G = RT\,lnK_d$$

Gibbs free energy is composed of contributions from enthalpy (ΔH) which roughly describes the energetic favorability of noncovalent interactions in biomolecules,

and entropy (ΔS) which represents disorder (reviewed in 8). These values are related to one another by the following expression:

$$\Delta G \,=\, \Delta H \,-\, T \Delta S$$

While these factors are largely determined by the specific interaction between the protein and its ligand, interactions do not happen in isolation – they occur in a molecular environment, which may include the solvent in an *in vitro* experiment or in the much more complex intracellular environment in a living organism. Methods used to study protein-ligand interactions – specifically, RNA-protein interactions – will be discussed below.

I-3-ii. Classical in vitro *methods for studying RNA-protein interactions*

Traditional biochemical methods for studying individual RNA-protein interactions involve purifying both components – the RNA and the protein – from cells and examining their interaction in isolation. Two of these *in vitro* methods to determine the dissociation constant for an interaction between proteins and nucleic acids are electrophoretic mobility shift assays (EMSAs) and filter binding assays. While these methods will be described here for RNA-protein interactions, they may also be applied to other protein-ligand interactions. Both methods generally use radioactively labeled (radiolabeled) RNA to identify nucleic acid-protein complexes, and make use of changes in the behavior of molecules when they participate in an interaction versus when they are unbound.

In filter-binding assays, the RNA and protein are combined in different concentrations, allowed to come to binding equilibrium, and are passed through a nitrocellulose filter. Nitrocellulose is negatively charged; since RNA is negatively charged and proteins have regions of positive charge, the filter will bind proteins but not free RNA. When RNA is bound to a protein, the entire complex will bind to the filter, and will be identifiable by the RNA's radioactive labelling (28). Following quantification of RNA bound to the filter, the fraction bound (amount of the initial RNA concentration bound to the protein) can be plotted against protein concentration (Figure 12A) and dissociation constant can be determined from the protein concentration at which fraction bound is 50%. As seen in Figure 12A, a lower-affinity interaction (right) requires a greater protein concentration to reach 50% saturation than a higher-affinity interaction (left) (28).

EMSAs use a similar principle to filter binding assays – radiolabeled RNA behaves differently when it is free than when it is bound to a protein – but use gel electrophoresis to elucidate these differences rather than a charged filter. Multiple RNA-protein mixtures are run on a gel, where the concentration of RNA is held constant and the concentration of the protein is varied. At low protein concentrations, the radiolabeled RNA is unbound by protein and travels through the gel quickly; as protein concentration increases, RNA begins to bind to the protein and runs more slowly on the gel. This difference is apparent in the distance traveled by the free RNA versus the RNA-protein complex, and the amount of RNA bound to protein or free can be quantified via the amount of radiolabeled RNA in the free RNA band and in the RNA-protein

complex band. These two bands are shown in Figure 12B. Like with filter-binding assays, K_d may be determined based on the protein concentration at which half the radiolabeled RNA is bound (12).
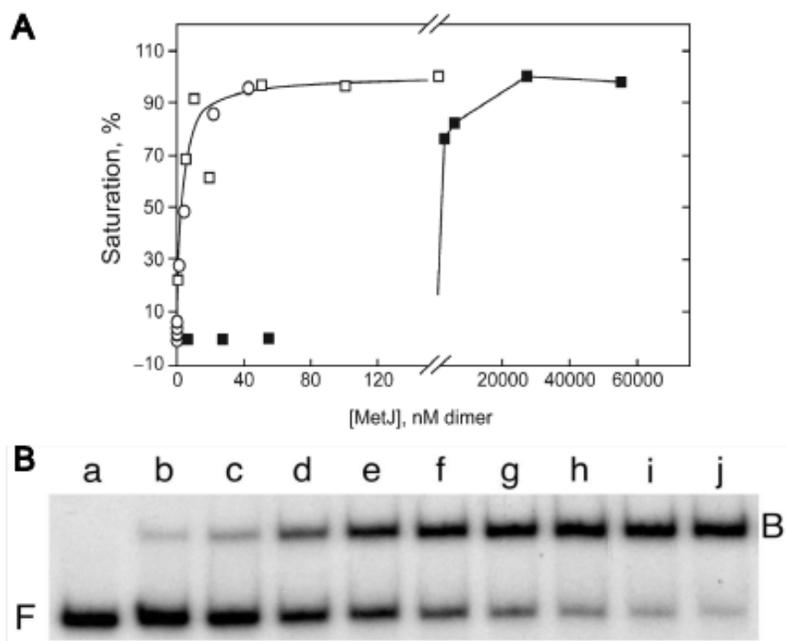
Figure 12. *In vitro* **RNA-protein interactions via filter binding and EMSA.** Panel A shows data collected using a filter binding assay with the protein MetJ, where the fraction bound of radiolabeled DNA is plotted against protein concentration. Dissociation constant can be determined from the protein concentration at a fractional saturation of 50%. Adapted from Stockley 2009 (28). Panel B shows EMSA data for the *E. coli* CAP protein with a radiolabeled *lac* promoter DNA fragment. Protein is titrated in from left to right, with no protein in lane a and the highest protein concentration in lane j. F refers to free DNA and B refers to bound DNA. Dissociation constant can be determined from the protein concentration at which half the initial DNA concentration is bound by the protein. Adapted from Hellman and Fried 2007 (12).

I-4. Genetic methods for studying macromolecular interactions

Thus far, only *in vitro* approaches for studying biomolecular recognition, which involve purifying molecules of interest and studying them outside of their biological system of origin, have been discussed here. An additional category of *in vivo* methods

exists which studies interactions within living organisms, but independent from those interactions' biological functions. These approaches are known as two- and three-hybrid assays (referred to collectively as 'n'-hybrid assays). They have a multitude of applications in bacteria and yeast, including using forward and reverse genetic studies to study interactions and to discover unknown participants in biomolecular interactions. Since purification of the molecules involved in the interaction is not necessary, n-hybrid assays are higher-throughput than many alternative methods.

I-4-i. Two-hybrid assays: bacteria and yeast

The first type of 'n'-hybrid assay developed was the yeast two-hybrid assay, used to study protein-protein interactions in *Saccharomyces cerevisiae* (10). A yeast two-hybrid system uses the transcriptional activator protein GAL4 split into two separate proteins – its DNA-binding domain and its transcriptional activation domain – each of which is fused to one of the proteins of interest (known as the "bait" and "prey" proteins) (10). These fusion proteins are encoded on plasmids that are transformed into yeast cells, enabling them to express each of the proteins to be studied fused to the appropriate domain of GAL4. An interaction between the bait and prey proteins results in transcriptional activation of a reporter gene such as β-galactosidase (β-gal), allowing indirect measurement of the interaction via β-gal activity (10). The mechanism of this assay is shown in Fig. 13A.

While the yeast two-hybrid assay is useful for studying eukaryotic protein-protein interactions, it is less applicable to the study of prokaryotic proteins. Bacterial two-hybrid

assays are a type of analogous assay that use bacterial transcriptional machinery in bacterial cells, rather than eukaryotic machinery in eukaryotic cells. In a bacterial two-hybrid assay, one protein of interest is fused directly to the α subunit of RNA polymerase and the other is fused to the DNA-binding protein λCI (9). Figure 13B shows this transcription activation-based mechanism.
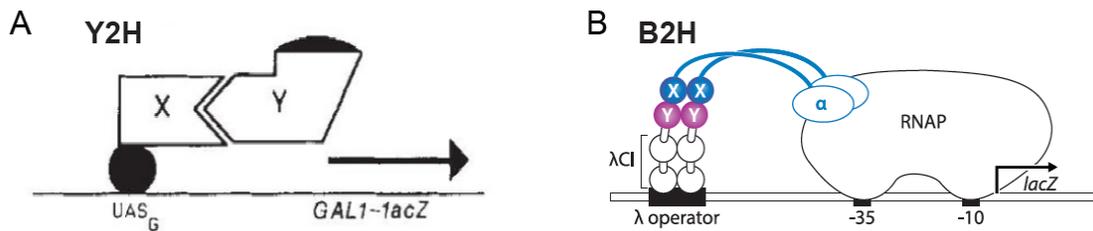
Figure 13. Yeast and bacterial two-hybrid assays. (A) Yeast two-hybrid assay. Two fusion proteins are used: the X protein fused to the DNA-binding UAS_G GAL4 domain, and the Y protein fused to the transcription activation domain of GAL4. When X and Y interact, GAL4 activates transcription of the reporter gene (*lacZ*). Figure adapted from Fields and Song 1989 (10). (B) Bacterial two-hybrid assay. The blue protein X is fused to the α subunit of RNA polymerase, while the purple protein Y is fused to the λCI DNA-binding protein. λCI binds to DNA upstream of the reporter gene promoter, and when X and Y interact the reporter gene *lacZ* is expressed. Figure adapted from Berry and Hochschild 2018 (3).

I-4-ii. The yeast three-hybrid assay

Yeast three-hybrid assays are similar in mechanism to two-hybrid assays, but use RNA-protein interactions to activate reporter gene transcription rather than protein-protein recognition. In a yeast three-hybrid assay, a DNA-binding protein binds upstream of the reporter gene promoter, but rather than being fused to a protein, it is fused to an RNA-binding protein (Fig. 14, 2). The bacteriophage MS2 coat protein (MS2^{CP}) is often used for this purpose in three-hybrid assays. To anchor the RNA sequence of interest upstream of the reporter gene promoter, it is transcribed as a hybrid RNA with the MS2 hairpin (MS2^{hp}) sequence. The MS2^{CP} and MS2^{hp} interact with high

affinity and specificity, allowing the RNA sequence to be reliably attached to DNA upstream of the reporter gene (Fig. 14). The MS2^{hp} may be added to the RNA sequence of interest at either its 5' or 3' end – while using a downstream MS2^{hp} tends to give higher signal than an upstream one, the optimal position is the one least likely to perturb the RNA's natural secondary structure based on structural prediction (2).
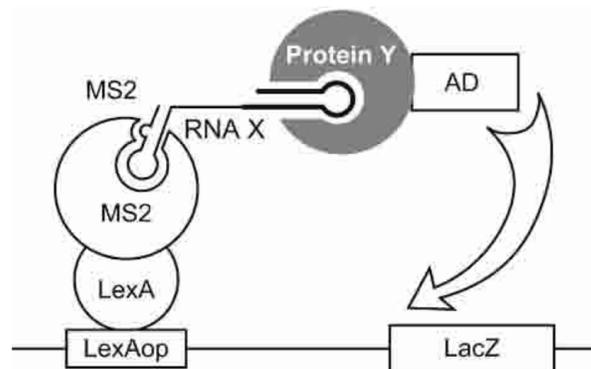
Figure 14. The yeast three-hybrid system. The protein LexA is used as the DNA-binding protein in this setup and binds to the LexA operator upstream of the LacZ reporter gene. LexA is fused to an MS2 coat protein, which binds to the MS2 hairpin on RNA X. RNA X interacts with Protein Y, which is fused to an activation domain (AD) that activates transcription of the LacZ gene. Figure adapted from Hook *et al.* 2005 (13).

The yeast three-hybrid system has been described and optimized by Marvin Wickens' lab at University of Wisconsin-Madison (2,13,27). Several factors have been demonstrated to be important in reliable and accurate detection of yeast three-hybrid signal. One of these factors is RNA abundance – high amounts of RNA result in higher signal as shown by using RNA vectors with varying copy numbers. The Y3H system also has an optimal RNA length of 150-200 nucleotides, shorter RNAs tend to be preferable to longer sequences (2). The yeast three-hybrid system produces signal such that the logarithm of the Y3H signal is linearly related to *in vitro* binding affinity for

RNA-protein interactions within a K_d range of 10 to 80 nM (13). Figure 15 shows this relationship.
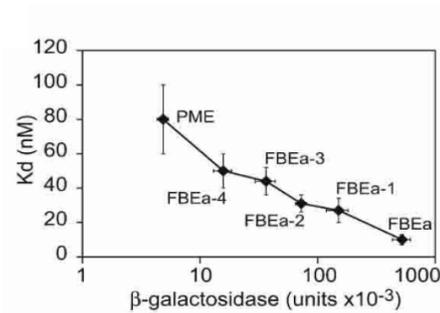
Figure 15. Linear relationship between log of β-galactosidase activity and K_d in the yeast three-hybrid assay. As binding affinity becomes stronger (lower K_d), β-gal activity increases. Adapted from Hook *et al*. 2005 (13).

I-4-iii. The bacterial three-hybrid assay

Similarly to how a bacterial equivalent of the yeast two-hybrid assay exists in *E. coli*, an analogous bacterial three-hybrid (B3H) assay for prokaryotic RNA-protein interactions has been developed in *Escherichia coli* by Katherine Berry and Ann Hochschild (3). Figure 16A shows the construction of the B3H system.
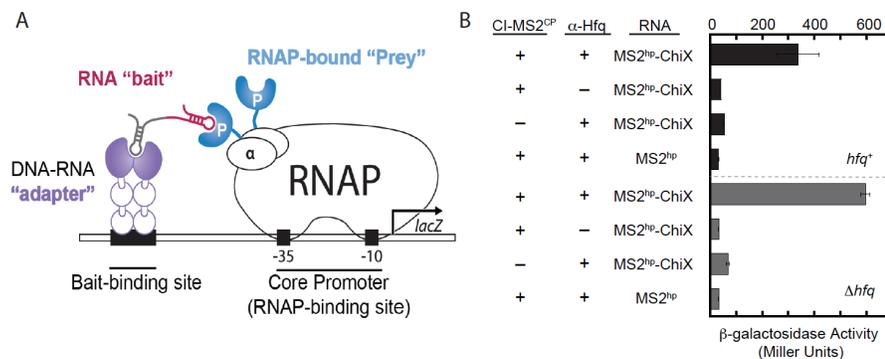
Figure 16. The bacterial three-hybrid assay. Panel A: The DNA-RNA adapter protein is composed of the λ CI protein fused to the MS2^{CP}, which binds to the DNA at the bait-binding O_L2 site. The RNA bait is attached to an MS2^{hp} sequence, which binds to the MS2^{CP} component of the adapter protein. The prey protein is fused to α subunit of RNA polymerase (RNAP). Interaction between the bait RNA and the prey protein results in stabilization of RNAP at the core promoter, leading to *lacZ* expression. Panel B: β-galactosidase activity for interactions between Hfq and ChiX using the B3H assay. Experimental conditions (all three components present) and negative controls (missing an individual component) are shown for *hfq+* and *Δhfq* reporter strains. Adapted from Berry and Hochschild 2018 (3).

The bacterial three-hybrid assay uses a set of components that is similar to, but not entirely the same as, the yeast three-hybrid system. Both systems use a DNA-RNA binding fusion protein that includes the MS2^{CP} as its RNA-binding element, and the B3H uses a λCI protein and an O$_L$2 site on DNA to anchor the protein upstream of the reporter gene promoter (3). The λCI-MS2^{CP} fusion protein is referred to as the DNA-RNA adapter. The RNA bait includes the RNA sequence of interest and an MS2^{hp} sequence to bind to the MS2^{CP} (3). Rather than fusing the prey protein to a transcriptional activator protein as in the yeast three-hybrid system, the prey protein is fused to the α subunit of RNA polymerase (similar to bacterial two-hybrids) (3). As in other n-hybrid assays, the interaction between the bait and the prey results in reporter gene expression. In the B3H assay, this expression is quantified by measuring β-gal activity. Data is typically presented as fold interaction; *i.e.* the experimental amount of β-gal activity divided by the highest negative control amount of β-gal activity. Negative control measurements are made by carrying out the assay with one of the three components missing (the DNA-RNA adapter, the bait RNA, or the prey protein) to disrupt the interaction (Figure 16B, 3). In negative controls, expression of the reporter gene is not promoted by an RNA-protein interaction, but a small amount of transcription still occurs, producing background noise that the experimental amount of β-gal activity must be normalized to (FIgure 16B, 3). The B3H assay has been demonstrated to detect interactions between *E. coli* Hfq and a number of its target sRNAs (3) as well as *E. coli* ProQ and several sRNAs and mRNAs (21).

I-4-iii-a. Limitations of the bacterial three-hybrid assay

While a number of RNA-protein interactions can be detected reliably in the bacterial three-hybrid assay, many known interactions still do not produce a significant or interpretable level of signal. As of publication of the B3H assay in 2018, the assay had only been tested with sRNA-Hfq interactions (3). Hfq, as a protein that regulates gene expression at the post-transcriptional level, binds to both regulatory sRNAs and the mRNAs that it regulates – thus, both types of interactions are necessary for a full picture of Hfq's function. In order to be broadly useful in studying both Hfq and other bacterial RNA-binding proteins, the B3H must be applicable with both regulatory RNAs and mRNAs, and mRNA-Hfq interactions therefore represented a gap in the B3H's established capabilities.

Furthermore, in addition to the inability to study mRNA-protein interactions using the B3H system at the time of its publication, the assay could only reliably detect Hfq-sRNA interactions for a limited set of sRNAs. Figure 17 shows B3H signal for some of these sRNAs; other sRNAs not shown in this figure (*e.g.* DsrA) generally produce a ≤2 fold interaction (at or below the threshold for a robust level of interaction).
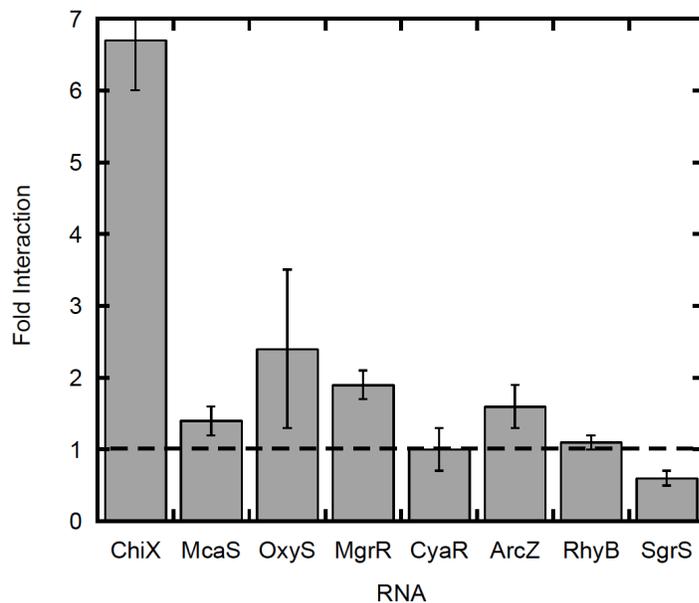
Figure 17. sRNA-Hfq interactions in the bacterial three-hybrid assay. A fold interaction greater than 2 is considered a robust interaction. All of the sRNAs shown here are known to interact with Hfq; some (such as ChiX) show substantial fold interaction, while others (such as MgrR and ArcZ) are close to the threshold for significance and do not reliably present a robust fold interaction under the conditions described in Berry and Hochschild 2018. Data collected by Clara Wang MHC '19.

There are a number of potential reasons why a known Hfq-interacting sRNA may not produce a substantial fold interaction in the B3H system. These reasons include, but are not limited to, misfolding of the RNA from its native secondary structure due to the addition of exogenous sequence (i.e. the MS2^{hp}) in ways that block Hfq binding; degradation of the hybrid RNA producing shortened MS2^{hp}-containing fragments that act as competitive inhibitors for full-length RNA binding to the MS2^{CP} (Supplementary Fig. 1); or interactions that are physiologically important but nevertheless have a weaker binding affinity than the B3H system is capable of detecting.

This last potential reason for inability to detect known interactions suggests an additional gap in knowledge of the B3H system: the relationship between B3H signal and binding affinity. Traditional approaches to studying RNA-protein interactions generally

involve determining the interaction's K_d *in vitro* to understand the strength of the interaction. While these methods are informative about the binding energetics of molecules in isolation, they are measured outside a cellular environment and therefore may not reflect the factors involved in biomolecular binding in living cells. *In vivo* approaches like n-hybrid assays are therefore complementary to *in vitro* studies of interaction strength. However, to place the relatively novel B3H assay in the broader context of classical methods for studying RNA-protein interactions, and to fully understand the information collected from it, it is important to understand the relationship between *in vitro* K_d and B3H signal. Intuitively, the B3H may be expected to produce greater signal for stronger interactions and potentially to have a direct relationship between *in vitro* K_d and *in vivo* fold interaction. However, the actual relationship between these two types of data was not clear at the time of the assay's initial publication. This unknown relationship limited our understanding of the information that could be obtained from the B3H system.

I-4-iii-b. Approach to B3H optimization

There are two ways in which the B3H system may be made into a more useful tool within the context of many other existing methods for studying RNA-protein interactions. The first of these ways is the expansion of the detection capabilities of the assay to a wider range of RNAs. This includes developing ways in which mRNA-protein interactions may be detected in the assay so that the full range of regulatory RNA-binding protein activity can be detected; and determining how to detect sRNA interactions that

currently present non-substantial levels of signal in the B3H assay. As mRNA-Hfq interactions had not been attempted in the B3H system as of its initial publication (3), attempting to detect this category of interactions begins simply with putting Hfq-interacting mRNA sequences into the system as the RNA bait. Attempts to detect interactions with low-signal sRNAs includes addressing potential reasons why interactions may not be currently detectable, such as secondary structure misfolding.

The second way in which the B3H's utility can be improved is in gaining a better understanding of its relationship with other types of data on RNA-protein interactions; namely, with *in vitro* dissociation constants. K_d is the standard way that protein-ligand affinity is described; therefore, to fully understand the information provided by the B3H assay, it is necessary to understand how B3H signal relates to K_d. Intuitively, the B3H may be expected to produce higher signal for higher-affinity interactions. However, the complexity of the intracellular environment and the nuances of the B3H assay mean that this may not be the case, or that the relationship between K_d and B3H fold interaction may not be linear. To understand this relationship, data from the B3H and from an *in vitro* method for the same interaction can be compared, and this type of comparison across a range of interaction strengths can offer insight into the relationship between these types of data.

I-5. Statement of purpose

Here we demonstrate several ways to improve the utility of the bacterial three-hybrid assay. To show that the B3H assay is capable of detecting $(A\text{-}A\text{-}N)_n$ motif interactions with the distal face of Hfq, I developed minimal distal face-binding RNA sequences and determined that a subset of them produced detectable B3H signal. I then detected the assay's first interaction between Hfq and a native mRNA 5' untranslated region, showing that physiologically relevant interactions that depend on the $(A\text{-}A\text{-}N)_n$-distal face mechanism can be measured.. In addition to demonstrating the assay's capacity to detect interactions with artificial RNAs and native mRNA sequences, I also used comparisons with *in vitro* binding affinity to show that efforts to optimize the B3H assay via changes in CI-MS2^{CP} protein level improve our ability to detect interactions accurately, and to propose a general relationship between B3H signal and binding affinity for both sRNAs and artificial Hfq-binding RNAs. The results presented here demonstrate that the B3H assay can be used to detect a wider range of interactions than established previously and provide insight into progress on expanding and optimizing the capabilities of the assay. These results and conclusions also suggest directions for further improvement of the assay that will let the assay be a more useful tool for understanding bacterial gene regulation, RNA-protein interaction mechanisms, and the potential discovery of novel interactions.

Chapter II: Materials and Methods

II-1. Bacterial strains

    E. coli strains used in this study are listed in Table 3. NEB5α is the recipient strain for all cloned plasmids and was purchased from New England Biolabs. KB473, KB483, and KB532A are the reporter strains used in liquid and plate-based β-galactosidase (β-gal) assays. Strains contain antibiotic-resistant genes as listed in Table 2, abbreviated as follows: TetR (tetracycline), KanR (kanamycin), StrR (streptomycin), SpecR (spectinomycin), AmpR (ampicillin and carbenicillin, which may be used interchangeably), CmR (chloramphenicol). All strains are stored as glycerol stocks at -80ºC.

Table 3. *E. coli* **strains used in this study.**

Strain	Genotype	Antibiotic resistance	Source
NEB5α-F'I^q Competent *E. coli*	Host strain for plasmid construction: F' *proA$^+$B$^+$ lacI^q Δ(lacZ)M15 zzf::Tn10* (TetR)/*fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	TetR	New England Biolabs
KB473	FW102 Δ*hfq*::FRT containing F' kan bearing test promoter (*placO$_L$2-62)* fused to *lacZ* and kanamycin resistance gene	KanR; StrR	3
KB483	FW102 Δ*hfq*::*kan* containing F' bearing test promoter (*placO$_L$2-62)* fused to *lacZ*	KanR; TetR; StrR	21
KB532A	FW102 Δ*hfq*::*kan* containing F' bearing test promoter (*placO$_L$2-62)* fused to *lacZ*	KanR; TetR; StrR	unpublished
MG1655	Strain K-12 F$^-$ lambda$^-$ *ivlG$^-$ rfb*-50 *rph-1*	–	4

II-2. Plasmid construction

II-2-i. Plasmids

Plasmids used in this study are listed in Table 2 below. Each plasmid confers antibiotic resistance, described in Table 4 using the same abbreviations as above. Plasmids beginning with the prefix pKB, pCH, and pCW were made by Dr. Katie Berry, Courtney Hegner (MHC '19) and Clara Wang (MHC '19), respectively (Table 2).

II-2-ii. Oligo-restriction enzyme cloning

Plasmids pHL2 and pHL3, each coding for $(A-R-N)_n$ motif repeats, were cloned by annealing inserts from single-stranded oligonucleotides (oligos) and ligating these inserts into restriction enzyme-digested plasmids. Fig. 18 shows a schematic of this process.
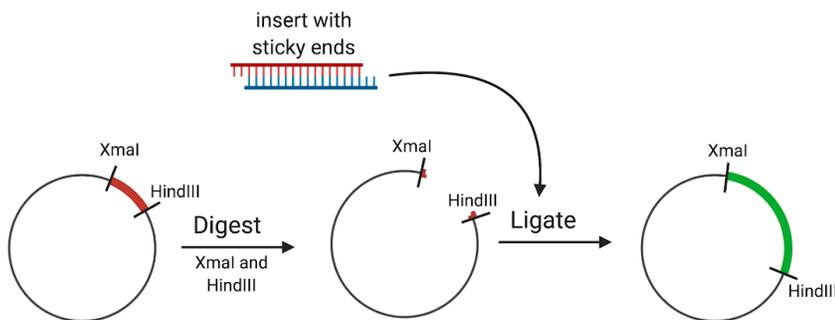
Figure 18. Cloning with restriction enzymes. For oligo-restriction enzyme cloning (Section II-2-ii), the insert with sticky ends is annealed from two single-stranded DNA oligonucleotides that have been synthesized with specific insert and sticky end sequences. For PCR-restriction enzyme cloning (Section II-2-iii), the insert is produced by PCR amplification and given sticky ends by restriction enzyme digestion. Figure made using BioRender.

Complementary oligos oHL3 and 4, and oHL5 and 6, were designed to constitute the full desired insert with sticky ends matching the restriction enzymes XmaI and HindIII when

annealed and ordered from Eurofins Genomics. To anneal, a complementary oligo pair was combined with 10X T4 PNK buffer (New England Biolabs) and rATP, heated briefly at 95ºC, and allowed to cool to room temperature. The 5' ends of the annealed insert were then phosphorylated by treatment with T4 polynucleotide kinase (New England Biolabs). The vector for pHL2 and pHL3, pCH6, was digested by combining the plasmid with 10X CutSmart Buffer (New England Biolabs) and the restriction enzymes XmaI and HindIII (New England Biolabs). Following digestion, the 5' ends of the digested vector were dephosphorylated using calf intestinal phosphatase or Antarctic phosphatase (New England Biolabs). The digested, dephosphorylated vector was then gel purified by running on 1% agarose and purifying an excised band using a Zymoclean Gel DNA Recovery Kit (Zymo Research). The annealed insert and the purified vector were then ligated by combining with T4 DNA ligase and 10X T4 DNA ligase buffer (New England Biolabs). Ligated plasmids were transformed into NEB5α competent *E. coli* cells (New England Biolabs, Table 2) and grown overnight on LB-agar plates containing the appropriate antibiotic (listed in Table 4) at 37ºC. Colonies were screened prior to sequencing using colony PCR using the primers oKB799 and oKB800 (Table 5) and visualization of PCR products in 1% or 2% agarose gel. Colonies that appeared to contain inserts of the right length were sent to Eurofins Genomics for sequencing. For each plasmid, a colony containing a plasmid that aligned correctly with the target sequence was inoculated and grown overnight in LB containing the appropriate antibiotic. This culture was then combined with 80% glycerol in a 1:1 ratio and stored at -80ºC for long-term storage.

II-2-iii. PCR-restriction enzyme cloning

Plasmids containing inserts derived from the *E. coli* genome (pHL28-32, Table 2) were cloned using PCR amplification and restriction enzyme site-based ligation. Fig. 10 demonstrates this process. Primers (oHL38-47) were designed to be partially complementary to the beginning and end of the desired endogenous sequence and to add the desired flanking restriction enzyme sites when used in PCR. These primers were ordered from Eurofins Genomics and used in PCR with lysed MG1655 *E. coli* cells (Table 2) to amplify the desired inserts from genomic *E. coli* DNA. Following PCR, products were visualized on a 1% agarose gel to ensure PCR success before moving to the next step. PCR products were then cleaned using a Zymo DNA Clean & Concentrator Kit (Zymo Research), digested using XmaI and HindIII (New England Biolabs) restriction enzymes, and cleaned again before being ligated into a digested vector (pCH1 or pHL6) as described above. Ligation products were then transformed, checked, sequenced, and stored as described above.

II-2-iv. Q5 site-directed mutagenesis

Q5-site directed mutagenesis introduces changes (mutations, insertions, or deletions) by using PCR to amplify an entire plasmid. Q5 was used to clone plasmids that only required small mutations, included sequences that do not exist endogenously in the *E. coli* genome, and/or involved sequences that are easily altered using restriction enzyme methods. The plasmids pHL1, pHL23-27, and pHL34-71 were cloned using Q5 mutagenesis. Q5 primers are designed to align back-to-back such that at least one of the

primers introduces the intended mutation during a PCR reaction and that the PCR product is the full, modified plasmid in linear form with blunt ends where the forward and reverse primers meet. Q5 primers were designed by hand or using New England Biolabs' NEBaseChanger tool and were used in PCR with the template plasmid to perform mutagenesis, either with New England Biolabs' Q5 2X Master Mix or Phusion 2X Master Mix. Following PCR, the reaction's success was checked by running PCR products on a 1% agarose gel. PCR products were then "KLDed": kinased (phosphorylation of 5' ends), ligated, and DpnI-treated, which was either carried out using New England Biolabs' KLD enzyme mix or by combining T4 polynucleotide kinase, T4 DNA ligase, and DpnI with the PCR products. NEB5α competent cells were then transformed with the KLD mixture, and sequenced and stored as described above. The colony PCR screening step was typically omitted for this type of cloning, as the difference between template and correct PCR product lengths was often not large enough to be informative.

Table 4. Plasmids used in this study.

Plasmid	Description	Antibiotic resistance	Source
pBr-α	Encodes residues 1-248 of the alpha subunit of RNA polymerase under the control of *lpp* and *lacUV5* promoters	AmpR	9
pACλCI	Encodes full-length λCI under the control of the *lacUV5* promoter	CmR	9
pKB816	pACλCI-Hfq; Encodes residues 1-236 of λCI fused via three alanine residues to full-length wild-type *E. coli* Hfq	CmR	3
pKB817	pBr-α-Hfq; Encodes residues 1-248 of the alpha subunit of RNA polymerase fused via three alanine residues to full-length wild-type *E. coli* Hfq	AmpR	3
pKB845	pCDF-pBAD-2xMS2^{hp}-XmaI-HindIII; two MS2 RNA hairpins (2xMS2^{hp}) and an XmaI site inserted into pKB822 CDF origin	SpecR	3

	vector between BamHI and HindIII sites		
pKB871	pBr-α-Hfq-Q8A; pKB817 with Q8A point mutation	AmpR	3
pKB872	pBr-α-Hfq-Y55A; pBK817 with Y55A point mutation	AmpR	Cloned by Katie Berry (pKB817, oKB1117 + oKB1118)
pKB873	pBr-α-Hfq-K56A; pKB817 with K56A point mutation	AmpR	Cloned by Katie Berry (pKB817, oKB1119 + oKB1120)
pKB874	pBr-α-Hfq-R16A; pKB817 with R16A point mutation	AmpR	Cloned by Katie Berry (pKB817, oKB1121 + oKB1122)
pKB903	pBr-α-Hfq-K31A; pKB817 with K31A point mutation	AmpR	Cloned by Katie Berry (pKB817, oKB1184 + oKB1185)
pKB905	pBr-α-Hfq-Y25D; pKB817 with Y25D point mutation	AmpR	Cloned by Katie Berry (pKB817, oKB1186 + oKB1187)
pKB907	pBr-α-Hfq-R17A; pKB817 with R17A point mutation	AmpR	Cloned by Katie Berry (pKB817, oKB1188 + oKB1189)
pKB941	pCDF-pBAD-2xMS2^{hp}-DsrA; pKB845 with DsrA inserted between XmaI and HindIII sites	SpecR	Cloned by Katie Berry (pKB845, oKB1209 + oKB1210)
pKB989	pACλCI-MS2^{CP}; encodes residues 1-236 of CI fused to MS2 coat protein (MS2^{CP}) via three alanine residues with V30I and A81G mutations and deletion of positions 68-80	CmR	3
pCW17	pAC-CI-MS2-(-35)-ACGATA-(-10)-TATAGT; CI-MS2^{CP} plasmid with a constitutive promoter, Δ*lacO* site	CmR	21

35u4	pAC-CI-MS2-(-35)-ACAGAT-(-10)-TATAGT; CI-MS2^{CP} plasmid with a constitutive promoter, $\Delta lacO$ site	CmR	21
pCH1	pCDF-pBAD-1xMS2^{hp}-XmaI-HindIII; one MS2 RNA hairpins (1xMS2^{hp}) and an XmaI site inserted into pKB822 CDF origin vector between BamHI and HindIII sites	SpecR	21
pCH6	pCDF-pBAD-1xMS2^{hp}-ChiX; pCH1 with ChiX inserted between XmaI and HindIII sites	SpecR	21
pCH9	pCDF-pBAD-1xMS2^{hp}-OxyS; pCH1 with OxyS inserted between XmaI and HindIII sites	SpecR	21
pHL1	pCDF-pBAD-1xMS2^{hp}-(AAC)$_{13}$; pCH1 with (AAC)$_{13}$ inserted between XmaI and HindIII sites	SpecR	This study (pCH1, oHL13 + oHL14)
pHL2	pCDF-pBAD-1xMS2^{hp}-(AAU)$_{17}$; pCH1 with (AAC)$_{17}$ inserted between XmaI and HindIII sites	SpecR	This study (pCH1, oHL3 + oHL4)
pHL4	pCDF-pBAD-1xMS2^{hp}-(GAC)$_{13}$; pCH1 with (GAC)$_{13}$ inserted between XmaI and HindIII sites	SpecR	This study (pCH1, oHL7 + oHL8)
pHL5	pCDF-pBAD-1xMS2^{hp}-(AACC)$_{13}$; pCH1 with (AACC)$_{13}$ inserted between XmaI and HindIII sites	SpecR	This study (pCH1, oHL9 + oHL 10)
pHL6	pCDF-pBAD-1xMS2^{hp}-T$trpA$; pCH1 with $trpA$ terminator downstream of HindIII site	SpecR	21
pHL13	pCDF-pBAD-1xMS2^{hp}-(AGC)$_{17}$; pCH1 with (AGC)$_{13}$ inserted between XmaI and HindIII sites	SpecR	This study (pKB845, oHL20 + oHL37)
pHL23	pBr-α-Hfq-I30D; pKB817 with I30D point mutation	SpecR	This study (pKB817, oHL52 + oHL58)
pHL24	pBr-α-Hfq-I30D-Y55A; pKB817 with I30D and Y55A point mutations	SpecR	This study (pKB817, oHL52 + oHL58)
pHL26	pCDF-pBAD-1xMS2^{hp}-A$_{27}$-T$trpA$; pHL6 with 27 adenosine residues between XmaI and HindIII sites based on sequence used by Mikulecky $et\ al.$ (18).	SpecR	This study (pHL6, oHL54 + oHL55)

pHL27	pCDF-pBAD-1xMS2^{hp}-A$_{29}$-T$trpA$; pHL6 with 29 adenosine residues between XmaI and HindIII sites based on sequence used by Mikulecky *et al*. (18)	SpecR	This study (pHL6,
pHL29	pCDF-pBAD-1xMS2^{hp}-*eptB*5'UTR-T$trpA$; pHL6 with the 5' UTR of *eptB* (-106 to +27) inserted between XmaI and HindIII sites	SpecR	This study (pHL6, oHL40 + oHL41)
pHL30	pCDF-pBAD-1xMS2^{hp}-*sodB*5'UTR-T$trpA$; pHL6 with the 5' UTR of *sodB* (-55 to +27) inserted between XmaI and HindIII sites	SpecR	This study (pHL6, oHL42 + oHL43)
pHL32	pCDF-pBAD-1xMS2^{hp}-*mutS*5'UTR-T$trpA$; pHL6 with the 5' UTR of *mutS* (-74 to +27) inserted between XmaI and HindIII sites	SpecR	This study and (pHL6, oHL46 + oHL47)
pHL34	pCDF-pBAD-2xMS2^{hp}-ΔXmaI-DsrA; pKB941 with XmaI site deleted	SpecR	This study (pKB941, oHL62 + oHL63)
pHL35	pCDF-pBAD-2xMS2^{hp}-A$_{6}$::XmaI-DsrA; pKB941 with XmaI site replaced with 6 adenosines	SpecR	This study (pKB941, oHL63 + oHL64)
pHL37	pCDF-pBAD-1xMS2^{hp}-*eptB*5'UTR-TAA-T$trpA$; pHL29 with stop codon (TAA) inserted between 5' UTR and HindIII site	SpecR	This study (pHL29, oHL66 + oHL68)
pHL38	pCDF-pBAD-1xMS2^{hp}-*sodB*5'UTR-TAA-T$trpA$; pHL30 with stop codon (TAA) inserted between 5' UTR and HindIII site	SpecR	This study (pHL30, oHL66 + oHL69)
pHL40	pCDF-pBAD-1xMS2^{hp}-*mutS*5'UTR-TAA-T$trpA$; pHL32 with stop codon (TAA) inserted between 5' UTR and HindIII site	SpecR	This study (pHL32, oHL66 + oHL70)
pHL54	pBr-α-Hfq-D9A; pKB817 with D9A point mutation	AmpR	This study (pKB817, oHL88 + oHL89)
pHL58	pBr-α-Hfq-Q41A; pKB817 with Q41A point mutation	AmpR	This study (pKB817, oHL96 + oHL97)
pHL60	pBr-α-Hfq-Y55W; pKB817 with Y55W point mutation	AmpR	This study (pKB817,

			oHL100 + oHL101)
pHL61	pBr-α-Hfq-Y25D-Y55A; pKB817 with Y25D and Y55A point mutations	AmpR	This study (pKB905, oHL102 + oHL103)
pHL63	pBr-α-Hfq-I30D-Y55W; pKB817 with I30D and Y55W point mutations	AmpR	This study (pKB817, oHL100 + oHL101)
pHL64	pBr-α-Hfq-Y25D-Y55W; pKB817 with Y25D and Y55W point mutations	AmpR	This study (pKB905, oHL100 + oHL101)
pHL66	pCDF-pBAD-1xMS2^{hp}-*sodB*5'UTR-ATAA-T*trpA*; pHL38 with an adenosine residue immediately 5' of TAA stop codon	SpecR	This study (pHL38, oHL110 + oHL69)
pHL67	pCDF-pBAD-1xMS2^{hp}-*sodB*5'UTR-AATAA-T*trpA*; pHL38 with two adenosine residues immediately 5' of TAA stop codon	SpecR	This study (pHL38, oHL111 + oHL69)
pHL68	pCDF-pBAD-1xMS2^{hp}-*sodB*5'UTR-AAT-T*trpA*; pHL38 with stop codon residues rearranged from TAA to AAT	SpecR	This study (pHL38, oHL112 + oHL113)
pHL69	pCDF-pBAD-1xMS2^{hp}-*sodB*5'UTR-ATA-T*trpA*; pHL38 with stop codon residues rearranged from TAA to ATA	SpecR	This study (pHL38, oHL114 + oHL113)
pHL70	pCDF-pBAD-1xMS2^{hp}-*sodB*5'UTR-TAG-T*trpA*; pHL38 with TAA stop codon replaced with TAG stop codon	SpecR	This study (pHL38, oHL115 + oHL116)
pHL71	pCDF-pBAD-1xMS2^{hp}-*sodB*5'UTR-TGA-T*trpA*; pHL38 with TAA stop codon replaced with TGA stop codon	SpecR	This study (pHL38, oHL117 + oHL118)

Table 5. Oligonucleotides used in this study. [1]Cloned by Katie Berry. [2]Oligo cloning. [3]Q5 PCR. [4]PCR for restriction enzyme cloning.

Name	Sequence (5' to 3')	Description
oKB330s	CGGTAACCCCGCTTATTAAAAGC	pCDF-pBAD sequencing
oKB799	CACACTTTGCTATGCCATAGC	F pCDF-pBAD colony PCR
oKB800	TATCAGACCGCTTCTGCGTTC	R pCDF-pBAD colony PCR
oKB1077s	GAACAGCGTACCGACCTGG	pBr-α sequencing
oKB1117	GTCAGCCAGATGGTTGCCAAGCACGCGATTTC	F pKB872[1]
oKB1118	GAAATCGCGTGCTTGGCAACCATCTGGCTGAC	R pKB872[1]
oKB1119	CAGCCAGATGGTTTACGCGCACGCGATTTCTAC	F pKB873[1]
oKB1120	GTAGAAATCGCGTGCGCGTAAACCATCTGGCTG	R pKB873[1]
oKB1121	GTTCCTGAACGCACTGGCTCGGGAACGTGTTCC	F pKB874[1]
oKB1122	GGAACACGTTCCCGAGCCAGTGCGTTCAGGAAC	R pKB874[1]
oKB1184	CTATTTATTTGGTGAATGGTATTGCGCTGCAAGGGCAAATCG	F pKB903[1]
oKB1185	CGATTTGCCCTTGCAGCGCAATACCATTCACCAAATAAATAG	R pKB903[1]
oKB1186	CGTGTTCCAGTTTCTATTGATTTGGTGAATGGTATTAAGC	F pKB905[1]
oKB1187	GCTTAATACCATTCACCAAATCAATAGAAACTGGAACACG	R pKB905[1]
oKB1188	CGTTCCTGAACGCACTGCGTGCGGAACGTGTTCCAGTTTC	F pKB907[1]
oKB1189	GAAACTGGAACACGTTCCGCACGCAGTGCGTTCAGGAACG	R pKB907[1]
oKB1209	TCCCCCCGGGAACACATCAGATTTCCTGGTGTAAC	F pKB941[1]
oKB1210	CCGGCCAAGCTTAAAAAAAATCCCGACCCTGAGGG	R pKB941[1]
oHL3	CCGGGAATAATAATAATAATAATAATAATAATAATAATAATA ATAATAATAATA	F pHL2[2]
oHL4	AGCTTATTATTATTATTATTATTATTATTATTATTATTATTATTA TTATTATTC	R pHL2[2]
oHL5	CCGGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG CAGCAGCAGCAGCAGCA	F pHL3[2]
oHL6	TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT GCTGCTGCTCCCGG	R pHL3[2]

oHL7	CCGGGGACGACGACGACGACGACGACGACGACGACGA CGACGACGACGACGACA	F pHL4[2]
oHL8	TGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC GTCGTCGTCCCCGG	R pHL4[2]
oHL9	CCGGGAACCAACCAACCAACCAACCAACCAACCAACCAACC AACCAACCAACCAACCA	F pHL5[2]
oHL10	TGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG TTGGTTGGTTCCCGG	R pHL5[2]
oHL13	AACAACAACAACAACAACAAGCTTGGCTGTTTTGGCGGATG AG	F pHL1[3]
oHL14	GTTGTTGTTGTTGTTGTTCCCGGGCTGCAGACATGGG	R pHL1[3]
oHL15	GACGACGACGACGACGACAAGCTTGGCTGTTTTGGCGGATG AG	F pHL4[3]
oHL16	GTCGTCGTCGTCGTCGTCCCCGGGCTGCAGACATGGG	R pHL4[3]
oHL17	ACCAACCAACCAACCAACCAAGCTTGGCTGTTTTGGCGGATG AG	F pHL5[3]
oHL18	TGGTTGGTTGGTTGGTTCCCGGGCTGCAGACATGGG	R pHL5[3]
oHL20	CATTAGGCGGGCTAAGCTTGCATGCCTGCAGGTCCCGGGC	R pHL6, pHL7[3]
oHL37	AGCGGGCTTTTTTTTGGCTGTTTTGGCGGATGAG	R pHL6/pHL7[3]
oHL40	GGCCGGCCCGGGGCGCGTGTAGATTTTACTTATCTGAC	F pHL29[4]
oHL41	GGCCGGAAGCTTCTGTGTAATCGATTTGATGTATCTCATG	R pHL29[4]
oHL42	GGCCGGCCCGGGATACGCACAATAAGGCTATTGTACG	F pHL30[4]
oHL43	GGCCGGAAGCTTTGGTAGTGCAGGTAATTCGAATGAC	R pHL30[4]
oHL46	GGCCGGCCCGGGTGCGCCTTATGTGATTACAACGAAAATA	F pHL32[4]
oHL47	GGCCGGAAGCTTGGCGTCGAAATTTTCTATTGCACTC	R pHL32[4]
oHL52	GGTGAATGGTGACAAGCTGCAAG	F pHL23, pHL24, pHL25[3]
oHL54	AAAAAAAAAAAAAAAAAAAAAGCTTAGCCCGCCTAATGAG	F pHL26[3]
oHL55	TTTTTTTTCCCGGGCTGCAGACATGG	R pHL26[3]
oHL56	AAAAAAAAAAAAAAAAAAAAAAAGCTTGGCTGTTTTGGCG	F pHL27[3]
oHL57	TTTTTTCCCGGGCTGCAGACATG	F pHL27[3]

oHL58	AAATAAATAGAAACTGGAACAC	R pHL23, pHL24[3]
oHL62	AACACATCAGATTTCCTGG	F pHL34[3]
oHL63	CTGCAGACATGGGTGATC	R pHL34, pHL35[3]
oHL64	AAAAAAAACACATCAGATTTCCTG	F pHL35 (Q5 PCR)
oHL66	TAAAAGCTTAGCCCGCCTAAT	F pHL37, pHL38, pHL40[3]
oHL68	CTGTGTAATCGATTTGATGTATCTC	R pHL37[3]
oHL69	TGGTAGTGCAGGTAATTCG	R pHL38, pHL66, pHL67, pHL72[3]
oHL71	GGCGTCGAAATTTTCTATTGC	R pHL40[3]
oHL88	TCTTTACAAGCTCCGTTCCTGAACGCAC	F pHL54[3]
oHL89	TTGCCCCTTAGCTGCGGC	R pHL54[3]
oHL96	GTCTTTTGATGCGTTCGTGATCCTGTTGAAAAAC	F pHL58[3]
oHL97	TCGATTTGCCCTTGCAGC	R pHL58[3]
oHL100	CCAGATGGTTTGGAAGCACGCGA	F pHL60, pHL63, pHL64[3]
oHL101	CTGACCGTGTTTTTCAAC	R pHL60, pHL63, pHL64[3]
oHL102	CCAGATGGTTGCCAAGCACGCGA	F pHL61[3]
oHL103	CTGACCGTGTTTTTCAACAG	R pHL61[3]
oHL110	ATAAAAGCTTAGCCCGCCT	F pHL66[3]
oHL111	AATAAAAGCTTAGCCCGCCT	F pHL67[3]
oHL112	TGCACTACCAAATAAGCTTAGCC	F pHL68[3]
oHL113	GGTAATTCGAATGACATTG	R pHL68, pHL69[3]
oHL114	TGCACTACCAATAAAGCTTAGCC	F pHL69[3]
oHL115	CACTACCATAGAAGCTTAGCC	F pHL70[3]
oHL116	CAGGTAATTCGAATGACATTG	R pHL70[3]
oHL117	GCACTACCATGAAAGCTTAGC	F pHL71[3]
oHL118	AGGTAATTCGAATGACATTG	R pHL71[3]

II-3. β-galactosidase assays

II-3-i. Liquid-based bacterial three- and two-hybrid assays

For liquid B3H assays, reporter cells (KB473, KB483, or KB532A) (Table 3) were transformed with three plasmids: a plasmid expressing the CI-MS2^{CP} fusion protein; a pCDF-pBAD plasmid expressing an MS2^{hp} fusion RNA; and a pBr-α plasmid expressing α-Hfq (Table 4). Negative controls were also transformed with pBr-α, pACλCI, and pCDF-pBAD-MS2^{hp} plasmids that lacked Hfq, MS2^{CP}, or an experimental RNA sequence, respectively. For B2H assays, cells were transformed only with CI-fusion and α-fusion protein-expressing plasmids. Transformations were grown overnight on LB-agar plates containing chloramphenicol, carbenicillin or ampicillin, spectinomycin for B3H assays, and either kanamycin (for KB473) or tetracycline (KB483 and KB532A) (Gold Bio). Following overnight growth, three colonies (two for select experiments) were picked and inoculated in 1 mL LB containing the same set of antibiotics as the LB plates as well as 0.2% arabinose for B3H assays and IPTG (Gold Bio; 0, 5, 10, or 50 μM for most experiments) in a 96-well deep well plate (VWR). Inoculated deep well plates were then covered with a breathable film (VWR) and shaken at 800-900 rpm (VWR or Benchmark Scientific) at 37ºC overnight. 5-10 μM of cultures grown in deep well plates were back diluted the following day into 200 μL of LB media containing the same antibiotics and arabinose concentration as the overnight cultures, with IPTG concentrations that varied between experiments, in optically clear sterile 96-well plates (Olympus Plastics). These back-dilutions were covered and grown to mid-log (measured OD_{600} = 0.1-0.3, actual OD_{600} = 0.3-0.9, measured with a Molecular Devices SpectraMax

plate reader) at 37ºC while shaking at 800-900 rpm. Once at mid-log, cells were transferred to a new 96-well plate containing a mixture of rLysozyme and PopCulture reagent (EMD Millipore) and lysed for at least 30 minutes. Lysed cells were then transferred to a 96-well plate containing a mixture of Z-buffer, 2-nitrophenyl-β-D-galactopyranoside (ONPG), and β-mercaptoethanol, and β-galactosidase activity was measured by taking OD_{420} measurements at 1-minute intervals over an hour at 28ºC. OD_{420} values were then normalized to OD_{600} measurements at the time of lysis to determine β-galactosidase activity in Miller units. β-galactosidase activity was averaged across the three wells for each transformation, and averages for transformations containing all three experimental plasmids were divided by the highest average negative control to determine fold interaction. Error bars represent propagation of error from standard deviations of experimental and negative control averages.

II-3-ii. Plate-based blue-white bacterial three- and two-hybrid assays

For B3H and B2H assays using blue-white differences to quantify RNA-Hfq or hexamerization interactions, reporter cells were transformed with sets of plasmids as described above. In experiments in which a liquid assay was not also being carried out, transformations were performed in duplicate in 96-well PCR plates and grown overnight at 37ºC in deep well 96-well plates with shaking at 900 rpm in 1 mL LB media formulated in the same manner as described above. Transformations were directly inoculated in deep well plates from PCR plates, rather than being first plated on LB-agar

plates. Following overnight growth, cultures were plated on large LB-agar plates containing an appropriate set of antibiotics, 40 μg/mL X-gal, 70 or 125 μM TPEG, and 1.5 μM IPTG (Gold Bio). In some experiments, overnight cultures were diluted 1:100 prior to plating; in experiments where a liquid assay was simultaneously being performed, mid-log cultures grown from back-dilutions were plated rather than plating the overnight cultures themselves. Plates were incubated overnight at 37ºC and photographed the day after plating and/or after several days of storage at 4ºC.

Chapter III: Results

III-1. Expansion of the bacterial three-hybrid assay to interactions between Hfq's distal face and mRNA-like sequences

While the bacterial three-hybrid assay has previously been established as an effective way to detect and study a number of sRNA-protein interactions (3) and some interactions between proteins and mRNA 3' UTRs (21), it has not previously been shown to be capable of detecting 5' UTR-protein interactions. Since interactions with mRNA 5' UTRs are an important component of regulation by RNA-binding proteins, our inability to measure this type of interaction represented a significant gap in the assay's utility. To investigate the assay's capacity to detect mRNA 5' UTR-protein interactions, I started with a proof-of-principle approach focused on interactions between artificial mRNA-like sequences and the distal face of Hfq, before proceeding to ask how native mRNA sequence interactions with Hfq may be detected.

III-1-i. Creation of artificial distal-face-binding hybrid RNAs with $(A\text{-}R\text{-}N)_n$ motifs

To determine whether the B3H assay could detect mRNA-Hfq interactions, we started by focusing on the distal face – the region of Hfq where most mRNA 5' UTR binding occurs. Given that Hfq's distal face has been found to bind to $(A\text{-}A\text{-}N)_n$ motifs in mRNA 5' UTRs (and potentially to the more general $(A\text{-}R\text{-}N)_n$ motif) (17,34,26) we designed a set of RNA constructs to be expressed in B3H reporter cells as the bait RNA according to these motifs. These constructed RNAs were intended to be minimal distal

face binders – incorporating the aspects of mRNA 5' UTRs that allow binding to Hfq's distal face, without any other components of a native mRNA sequence. Thus, they lack other components of an mRNA, such as a 5' UTR-coding sequence-3' UTR architecture or translation sequence elements like a ribosome binding site or stop codon (see Introduction). They also lack intrinsic terminators with poly(U) tails, which are present in all sRNAs used in the B3H assay prior to this work and are an important part of sRNA-Hfq binding (see Section I-2-iv-a). Two of these constructs conform to the $(A-A-N)_n$ motif and consist of repeats of AAC and AAU, respectively. These RNAs will be referred to as $(AAC)_{13}$ and $(AAU)_{17}$.[1] One RNA construct follows the $(A-R-N)_n$ motif and consists of 17 repeats of AGC $((AGC)_{17})$. Finally, two RNAs were designed as negative controls. One, a GAC repeat $((GAC)_{13})$, has the same nucleotide content as the AGC RNA but in an order that does not result in $(A-R-N)_n$ motifs. The other, an AACC repeat, was created to contain an $(A-A-N)_n$ motif, but in a four-nucleotide repeat rather than a three-nucleotide repeat with an additional nucleotide acting as a spacer. This RNA will be referred to as $(AACC)_{13}$.[2] The layout of the RNAs expressed from these constructs is shown in Fig. 19 and includes an MS2[hp] and 13 or 17 repeats of each construct's respective motif.

---

[1] Due to obstacles in cloning, the $(AAC)_{13}$ and $(GAC)_{13}$ RNAs have 13 repeats of their respective motifs while the $(AAU)_{17}$ and $(AGC)_{17}$ RNAs have 17 repeats. While the importance of the number of repeats has not been tested, we expect no major differences as both numbers of repeats exceed the six $(A-R-N)_n$ motifs the distal face is capable of binding at any given time.

[2] $(AACC)_{13}$ has the same number of repeats as $(AAC)_{13}$ and $(GAC)_{13}$ but is roughly the same length as $(AAU)_{17}$ and $(AGC)_{17}$. Since it has a different number of nucleotides in its repeated motif, its length and/or number of repeats must vary from the three-nucleotide-repeat RNAs.
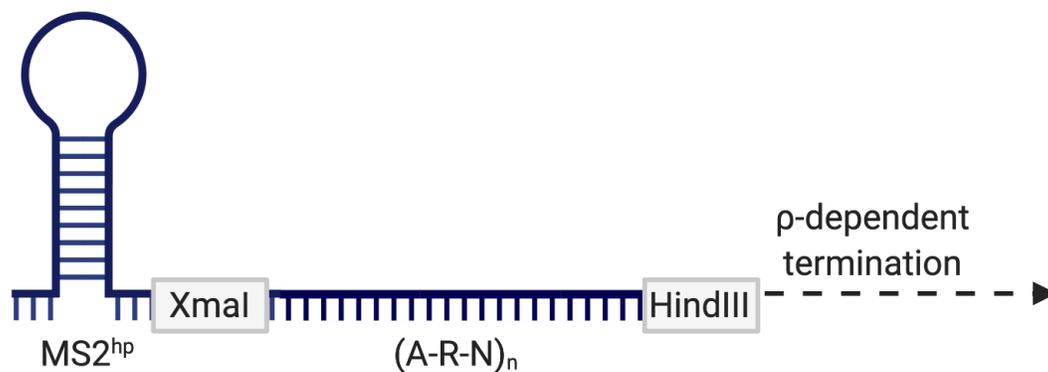
Figure 19. Layout of (A-R-N)$_n$ motif-containing minimal distal face binding RNAs. In (A-R-N)$_n$, n represents the number of repeats of the motif; constructs used here contained either 13 or 17 repeats depending on sequence characteristics (3-nt or 4-nt repeats) and the cloning method used. Downstream of the HindIII site is a ρ-dependent terminator, which terminates transcription at a point that is not clearly defined and may differ between individual RNA transcripts. Figure made using BioRender.

We hypothesized that the three RNAs following the (A-R-N)$_n$ motif were likely to interact with the distal face of Hfq and therefore produce B3H signal; however, we anticipated that the (GAC)$_{13}$ and (AACC)$_{13}$ constructs would not produce high signal, since they lacked the requisite order (GAC) or number of nucleotides per repeat (AACC) and are likely to be incompatible with the structure of Hfq's distal face. These RNAs and our hypotheses about their interactions with Hfq are summarized in Table 6.

Table 6. Summary of simple (A-R-N)$_n$ motif RNAs and their hypothesized interactions with Hfq.

RNA	Type of motif	Expected B3H signal
(AAC)$_{13}$	(A-A-N)$_n$	High
(AAU)$_{17}$	(A-A-N)$_n$	High
(AGC)$_{17}$	(A-R-N)$_n$	Medium
(GAC)$_{13}$	None (nucleotides out of order)	None
(AACC)$_{13}$	None (4-nucleotide repeat)	None

III-1-ii. Artificial (A-A-N)$_n$-containing sequences interact with the distal face of

Hfq

To determine whether the B3H assay could detect interactions between Hfq and minimal distal face-binding RNAs, we tested these RNAs for their interaction with Hfq in the B3H system. As predicted, the two RNAs consisting of (A-A-N)$_n$ motifs produced the greatest B3H signal in their interactions with Hfq: the (AAC)$_{13}$ RNA produced a roughly 2-fold interaction, while the (AAU)$_{17}$ RNA gave the greatest signal (Fig. 20). The (GAC)$_{13}$ and (AACC)$_{13}$ RNAs also conformed to our expectations, as their B3H signal (approximately or less than 1-fold) is indistinguishable from background noise (Fig. 20). The most surprising result of this experiment was the signal from the (AGC)$_{17}$ RNA-Hfq interaction, which consistently produced low (<1-fold) signal (Fig. 20). The lack of an observed interaction may reflect differences in RNA structure (see Sections IV-1 and IV-4).
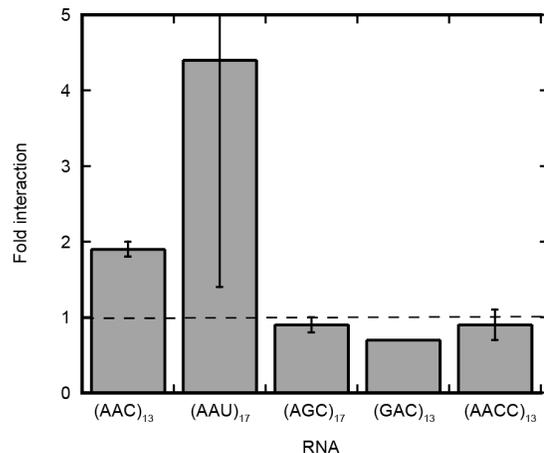
Figure 20. Artificial (A-A-N)$_n$ motif-containing RNA sequences interact with Hfq in the B3H assay. Results collected using a liquid β-galactosidase assay in Δ*hfq E. coli* reporter cells with a -62 O$_L$2 site (KB473). Cells were transformed with plasmids expressing α-Hfq, an MS2^{hp}-bait RNA, and CI-MS2^{CP}. An IPTG-inducible promoter (pKB989) was used to express CI-MS2^{CP}. Cells were grown in LB media + CAKS + 0.2% arabinose. Results shown here are an average of two independent experiments done in triplicate with standard deviation. GAC lacks a standard deviation due to lack of available data from one of the two experiments. Figure made using Kaleidagraph (16).

Having observed that the $(AAU)_{17}$ RNA had the most consistent and highest-signal Hfq interactor of this set of artificial RNAs, we were interested in determining whether this signal was indeed dependent on the distal-face of Hfq. We tested the interaction of the $(AAU)_{17}$ RNA with Hfq mutants containing substitutions at key interaction residues on each face. The sRNA ChiX was used as a positive control, which we expected to depend on both proximal and distal face residues. Our expectations for the impact of each substitution on the $(AAU)_{17}$ interaction are summarized in Table 7.

Table 7. Expected results of interactions between $(AAU)_{17}$ RNA and Hfq variants. While both Y25D and K31A are located on the distal face, Y25 plays a more important role in $(A-R-N)_n$ specificity than K31.

Hfq variant	Location of substitution	Expected impact on $(AAU)_{17}$ binding
Wild type (WT)	n/a	n/a
K56A	Proximal face	None
Y25D	Distal face	Substantially decreased binding
K31A	Distal face	Minor or no decrease in binding
R16A	Rim	None
R17A	Rim	None

Crucially and in support of our hypothesis, we observed that a substitution of Tyr25 on the distal face with an aspartate residue eliminated the interaction between $(AAU)_{17}$ and Hfq (Fig. 21). This residue is key for binding of the first A in $(A-R-N)_n$ motifs (see Section I-2-iv-a). Other than Y25D, each of the Hfq variants used here produces high signal with at least one of the RNAs, indicating no expression or hexamerization defects (Fig. 21); data presented later (see Sections III-3-i, III-3-ii-c, and Appendix) demonstrates a lack of expression or hexamerization issues for Y25D either

through B2H testing or by interactions with an sRNA.[3] With expression and hexamerization defects ruled out as reasons for low signal, these results reflect a disruption in binding caused by the amino acid substitution rather than a limitation of the protein prey. This data thus demonstrates that the $(AAU)_{17}$ RNA depends on the distal face for binding, and more specifically depends on the adenine-binding pocket implicated in $(A-R-N)_n$ motif binding.
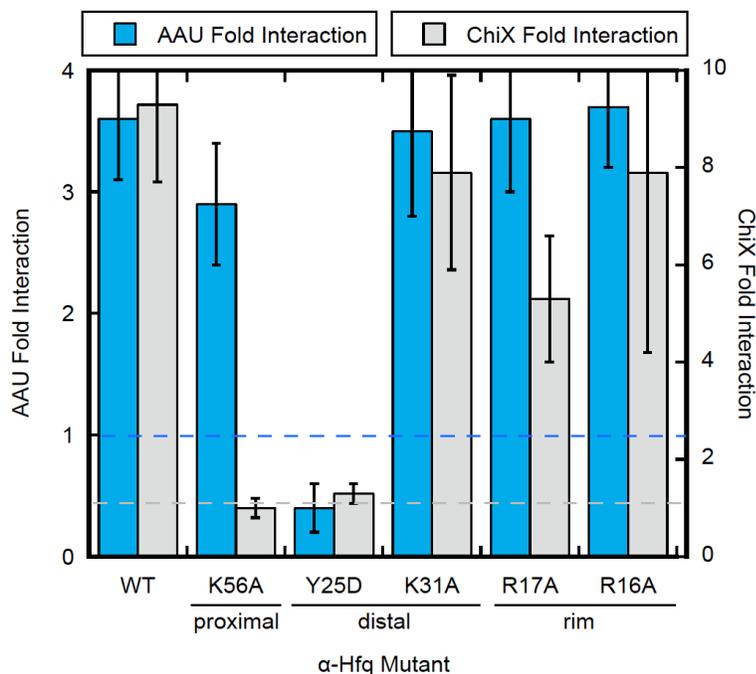
Figure 21. An RNA with a repeated AAU motif binds to the distal face of Hfq. Results collected using a liquid β-galactosidase assay in Δ*hfq E. coli* reporter cells with a -62 O_L2 site (KB473). Cells were transformed with plasmids expressing wild-type (WT) α-Hfq or α-Hfq with a single amino acid substitution, $(AAU)_{17}$ or ChiX as the MS2[hp]-bait RNA, and CI-MS2[CP]. An IPTG-inducible promoter (pKB989) was used to express CI-MS2[CP]. Cells were grown in LB media + CAKS + 0.2% arabinose. Figure made using Kaleidagraph (16).

---

[3] See Figs. 26 and 30 for Y25D B2H results and Fig. 30 and Supplementary Table 1 for interactions with DsrA.

III-2. Detection of native mRNA 5' UTR-Hfq interactions

With interactions between simple $(A-A-N)_n$ motifs and the distal face of Hfq established, we asked whether these detection capabilities carried over to native mRNA 5' UTR sequences. While the minimal distal face-binding constructs described this far were intended to mimic the way in which mRNAs bind to Hfq, the use of more complex and biologically relevant sequences potentially presented additional challenges. Artificial $(A-R-N)_n$ motif-containing RNAs used previously contained only $(A-R-N)_n$ motifs, while native RNAs are more sequentially and structurally complex, introducing the possibility of misfolding that could disrupt normal Hfq binding when inserted into the B3H system. Additionally, the fact that the sequences are native to *E. coli* introduced the potential for the RNAs to be subject to endogenous processes such as Hfq-dependent degradation, mRNA surveillance, and translation that could affect their abundance or availability for Hfq binding. Thus, we could not take for granted that native mRNA-Hfq interactions would be detectable simply because interactions had been detected in preliminary tests with artificial RNAs.

III-2-i. Design of RNA constructs for native 5' UTRs

To design constructs for testing mRNA 5' UTR-Hfq interactions, we began with a similar basic layout as for minimal distal face binders: an MS2^{hp} sequence, followed by the RNA sequence that would interact with Hfq. However, in cloning native *E. coli* sequences, important questions came into play that did not previously apply to our minimal distal face binders – namely, what sections of the mRNA are necessary for

detection of interactions? An entire mRNA is too long to insert into the B3H system, and unnecessary, since Hfq mediates regulation via 5' UTR interactions. Thus, we included the 5' UTR of each RNA, and since we were not able to narrow down what sections of this region are necessary for Hfq interaction, we included the entire region from the promoter to the ribosome binding site (RBS). Additionally, it was necessary to include the RBS itself and the start codon, since sRNA-mediated regulation often involves binding around this sequence to sterically block translation. To ensure that all potential interaction sequence was included, we included 8 codons (27 nucleotides) of the open reading frame. We therefore refer to these constructs as 5' UTR+8 codon RNAs. Finally, while we had been able to detect interactions between Hfq and minimal distal face-binding RNAs terminated by ρ-dependent termination, we were concerned that the indeterminate amount of additional transcription at the 3' end of these constructs may hinder our ability to detect these interactions. We thus included a transcriptional terminator sequence from the *trpA* operon (referred to here as the *trpA* terminator) so constructs would end at a definite and consistent point.[4] The layout of these constructs is shown in Fig. 22.

---

[4] Due to time constraints, we were unable to test if ρ-dependent or -independent termination is more optimal for detecting mRNA 5' UTR-Hfq interactions. Data presented in Section 3-III-ii-a suggests that inclusion of a *trpA* terminator may improve measurement capabilities for other RNAs.
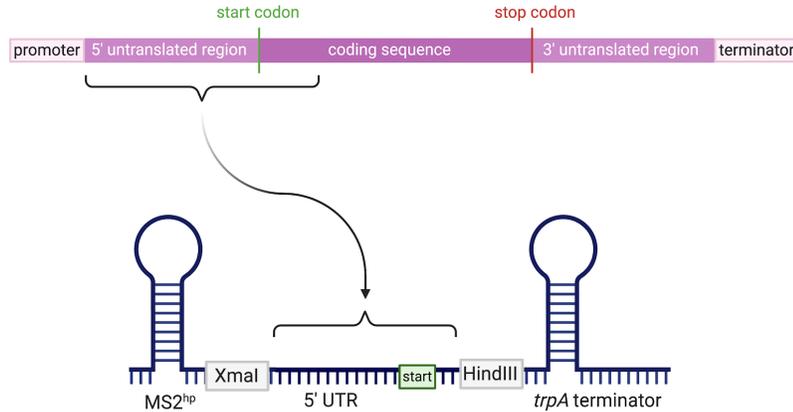
Figure 22. Construction of mRNA 5' UTR+8c codon RNAs. The "5' UTR" region of the construct between the XmaI and HindIII sites contains the mRNA's entire 5' untranslated region, its ribosome binding site, and 8 codons of the coding sequence/open reading frame. Figure made in BioRender.

III-2-ii. A stop codon is necessary but not sufficient for detection of Hfq-mRNA 5' UTR+8 codon interactions

Having designed constructs to test mRNA 5' UTR-Hfq interactions in the B3H assay, we tested these RNAs for their interactions with Hfq. In our initial tests of these constructs, we detected no signal above background noise (Fig. 23).
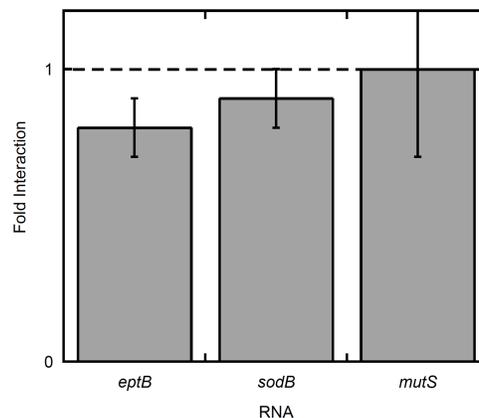
Figure 23. Initial tests of mRNA 5' UTR+8 codon-Hfq interactions produce no signal. Results collected using a liquid β-galactosidase assay in Δ*hfq E. coli* reporter cells with a -62 O_L2 site (KB473). ChiX was used as a positive control. Cells were transformed with plasmids expressing α-Hfq, an MS2^{hp}-bait RNA, and CI-MS2^{CP}. A constitutive promoter (35u4) was used to express CI-MS2^{CP}. Cells were grown in LB media + CCTS + 0.2% arabinose + 5 μM IPTG. Figure made using Kaleidagraph (16).

70

After detecting no signal in our first tests of mRNA 5' UTR-Hfq interactions, we realized that while the constructs included the ribosome binding site (the Shine-Dalgarno sequence and start codon) and a section of the open reading frame, they did not contain a stop codon. These RNAs thus may be subject to degradation related to ribosomal stalling at the end of a non-stop mRNA transcript. To account for this possibility, we created constructs identical to the original mRNA 5' UTR plasmids (Fig. 24A) with a stop codon (UAA) inserted immediately 3' of the open reading frame (Fig 24B). While this adjustment did not impact B3H signal for 2 of the 3 mRNA 5' UTRs tested, it did result in a consistently robust fold interaction for the binding of the *sodB* 5' UTR to Hfq (Fig. 24C). As with minimal distal face binding RNAs, the difference in signal between different RNAs may be a result of secondary structure (see Section IV-5).
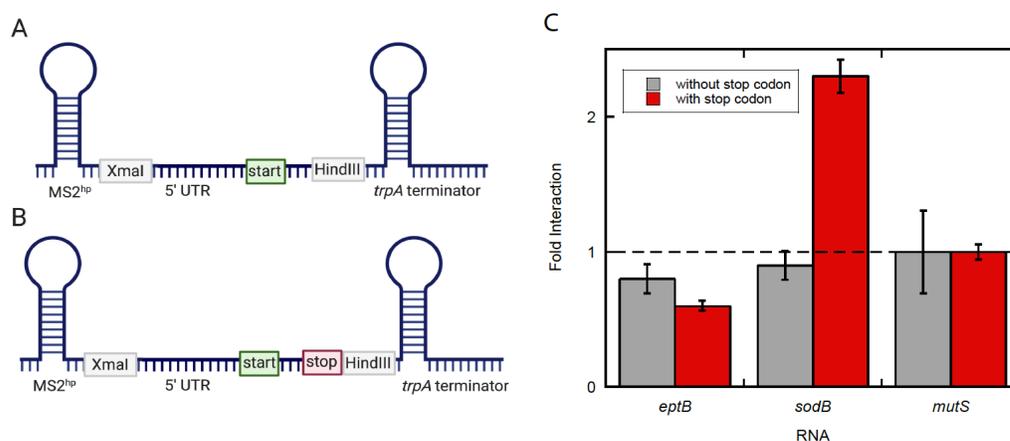
Figure 24. Determining the necessity of a stop codon in mRNA 5' UTR+8 codon constructs. (A) and (B) Layout of mRNA 5' untranslated region B3H constructs. 5' UTRs included native sequence obtained from EcoCyc starting from the mRNA's transcriptional promoter through the first 8 codons of the open reading frame (15). Construct (A) does not include a stop codon, and construct (B) has an UAA stop codon between the 5' UTR sequence and HindIII site. Figures made using BioRender. (C) mRNA 5' UTR-Hfq interactions with and without stop codons. The presence of a stop codon enables detection of the *sodB*-Hfq interaction, but not interactions with *eptB* or *mutS*. Results collected using a liquid β-galactosidase assay in Δ*hfq E. coli* reporter cells with a -62 O_L2 site (KB473). Cells were transformed with plasmids expressing α-Hfq, an MS2^{hp}-bait RNA, and CI-MS2^{CP}. A constitutive promoter (35u4) was used to express CI-MS2^{CP}. Cells were grown in LB media + CCTS + 0.2% arabinose + 5 μM IPTG. Predicted secondary structures (14) with and without MS2^{hp} and *trpA* terminator are shown in Supplementary Fig. 2 (Appendix). Figure made using Kaleidagraph (13) and BioRender.

III-2-iii. Only functional stop codons enable Hfq-mRNA 5' UTR+8 codon

interaction detection

We hypothesized that the reason that addition of a stop codon to the *sodB* 5' UTR bait RNA enabled detection of its interaction with Hfq is related to ribosomal stalling and non-stop decay mediated by tmRNA. If this explanation were true, the increased *sodB* 5' UTR-Hfq signal would depend on the presence of a functional, in-frame stop codon at the end of the open reading frame – either the UAA stop codon used thus far, or one of the other two stop codons (UAG or UGA). We imagined that other explanations for the increased signal were possible, however. For instance, the UAA codon is extremely similar to the AAU motif found previously in this study to produce substantial signal in the B3H assay when interacting with Hfq. We therefore wanted to rule out that the increased *sodB* 5' UTR-Hfq signal was a result of an interaction between the stop codon itself and Hfq and confirm that translation termination was necessary for detection of this interaction.

To differentiate between these hypotheses, we designed a set of constructs with the *sodB* 5' UTR+8 codon RNA in which the stop codon was modified in one of several ways. These modifications are listed in Table 8 and otherwise follow the architecture for a stop-codon containing mRNA 5' UTR bait plasmid presented in Fig. 24B.

Table 8. Stop codon modifications and expected impact on B3H signal.

Stop codon	Category	Expected signal
None	-	Low (Fig. 13)
UAA	Functional	High (Fig. 13)
UAG		High
UGA		High
xUAA	Out of frame	Low
xxUAA		Low
AAU	Scrambled	Low
AUA		Low

These modifications can be broken down into three categories: functional stop codons, out-of-frame stop codons, and scrambled UAA stop codons. Of these three categories, we predicted that only constructs containing functional stop codons would enable detection of robust signal in the B3H assay; the other modifications would not terminate translation and thus would not eliminate any ribosomal stalling-dependent degradation processes. However, if the increase in signal was instead the result of direct interaction between this codon and Hfq, we anticipated that all or some of the out-of-frame and rearranged stop codons would still enable robust signal detection, as the same set of nucleotides are still present.

When we tested these constructs for their interaction with Hfq in the B3H assay, the results we obtained conformed to our expectations. As we predicted, the two functional stop codons we had not previously tested produced comparable fold interactions to the original UAA construct (Fig. 25). Scrambled and out-of-frame stop

codons gave signal close to background (approximately 1-fold) and comparable with that of the construct lacking a stop codon (Fig. 25).
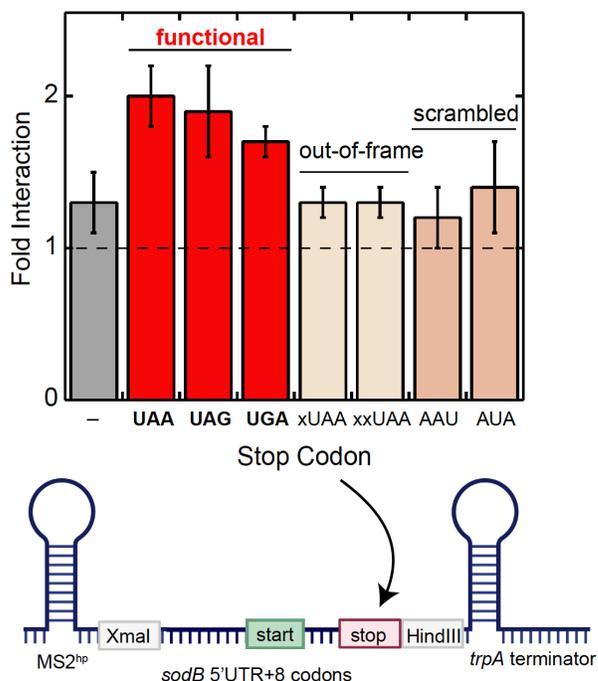
Figure 25. Comparison of stop codon modifications for *sodB* 5' UTR-Hfq interactions. Results collected using a liquid β-galactosidase assay in Δ*hfq E. coli* reporter cells with a -62 O_L2 site (KB483 or KB532A). Cells were transformed with plasmids expressing α-Hfq, an MS2^{hp}-bait RNA, and CI-MS2^{CP}. A constitutive promoter (35u4) was used to express CI-MS2^{CP}. Cells were grown in LB media + CCTS + 0.2% arabinose + 5 μM IPTG. Results shown here are an average of three independent measures, each done in triplicate, with standard deviation. Figure made collaboratively with Katie Berry using Kaleidagraph (16) and BioRender.

III-3. Defining the relationship between B3H signal and binding affinity

In addition to enabling detection of specific RNA-protein interactions as has been presented thus far, optimization of the B3H assay has also involved efforts to improve the assay globally by increasing fold interaction across a wide range of interactions. Work by Clara Wang and Rachel Mansky, seeking to find the optimal amount of the DNA-RNA adapter fusion protein, has produced improvements to expression of this protein – both making its expression constitutive rather than IPTG-dependent and honing in on the

optimum amount of adapter protein expression. These efforts began with the original

adapter plasmid (pAdapter), pKB989, which produced the protein using an

IPTG-inducible promoter (3). Expression of the α-Hfq protein in the B3H system is also

IPTG-inducible, and overexpression of CI-MS2^{CP} (but not α-Hfq) is detrimental to

interaction detection. Constitutive promoters, designed both rationally based on predicted

promoter strength and identified through a forward genetic screen, were created to

replace the IPTG-inducible pKB989 promoter, resulting in the plasmids pCW17 and

35u4 which have been demonstrated to improve B3H signal as compared to pKB989

when used as the pAdapter plasmid.[5]

III-3-i. Comparisons with in vitro binding affinity demonstrate assay improvement via optimized DNA-RNA adapter levels

Having established that the promoters in pCW17 and 35u4 increased B3H signal

over the original IPTG inducible promoter for a number of interactions, we wondered

how B3H data collected using each of these plasmids compared to binding affinity

determined *in vitro* (expressed as dissociation constant, K_d). While increased signal

indicates that these plasmids are promising ways to improve the B3H assay, optimization

ideally increases B3H signal in a way that is consistent with binding affinity.

We first asked how B3H signal obtained using each version of the pAdapter

plasmid compared to published K_d values for interactions between the *E. coli* sRNA

OxyS and a set of three Hfq mutants, as well as wild type Hfq. As expected, fold

---

[5] A manuscript presenting this data has not yet been published but is in preparation.

interaction for the interaction between OxyS and wild type Hfq, the strongest of the four interactions, increased when pCW17 or 35u4 was used as compared to pKB989 (Fig. 26A). pCW17 demonstrates a sharp drop in fold interactions even for strong (<10 nM) interactions, with no capability to distinguish between a wide range of interactions from relatively strong to very weak (Fig. 26A). While 35u4 produces slightly lower fold interactions for the strongest interactions (within the margin of error), it is the only one of the three pAdapter plasmids that gives signal greater than background levels across the full range of affinities tested here (Fig. 26A). To confirm that differences in fold interaction obtained here are due to differences in interaction rather than expression of the Hfq mutants, we used a B2H assay to measure expression levels of the α-fusion Hfq mutants used to collect OxyS-Hfq interaction data (Fig. 26B). The mutants used in this experiment produced fold interactions equal to or greater than wild type, demonstrating no expression defects for these mutants (Fig. 26B).
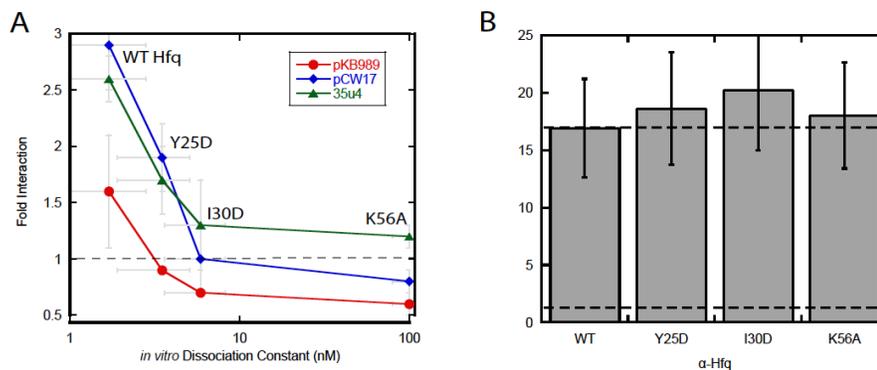
Figure 26. Comparison of OxyS-Hfq *in vivo* B3H signal and *in vitro* binding energetics. (A) B3H data plotted against *in vitro* K_d (20). B3H results collected using a liquid β-galactosidase assay in Δ*hfq E. coli* reporter cells with a -62 O_L2 site (KB483 or KB532A). Cells were transformed with plasmids expressing α-Hfq, an MS2hp-bait RNA, and CI-MS2CP. A constitutive promoter (35u4) was used to express CI-MS2CP. (35u4) was used to express CI-MS2CP. Cells were grown in LB media + CCTS + 0.2% arabinose + 5 μM IPTG. Results shown here are an average of two independent measures done in triplicate with standard deviation. *In vitro* results collected by filter binding assay (20). (B) B2H data collected using a liquid β-galactosidase assay in Δ*hfq E. coli* reporter with a -62 O_L2 site (KB483). Cells were transformed with plasmids expressing α-Hfq and CI-Hfq. Cells were grown in LB media + CCT + 5 μM IPTG. Figures made using Kaleidagraph (16).

III-3-ii. Comparing A₍₂₇₎ and DsrA interactions in the B3H assay to in vitro *data*

With 35u4 chosen as our most promising iteration of the pAdapter plasmid, we were interested in examining the relationship between K_d and fold interaction across a larger set of data points. Most studies of *in vitro* Hfq-RNA interactions report binding affinities for only a small number of mutants, but a study by Mikulecky *et al.* (18) included interactions with a large number of Hfq variants.[6] However, the specific RNAs used in this study – a poly(A) sequence (A_{27}), the sRNA DsrA, and the 5' UTR of the mRNA *rpoS* – presented an obstacle (18). Despite the advances in detection of Hfq-mRNA interactions presented above, we are not yet able to measure signal from the *rpoS* 5' UTR-Hfq interaction. DsrA and A_{27} were more promising, but before we could use them in comparisons with *in vitro* data, we first had to determine how to get the most reliable and useful interactions.

III-3-ii-a. Optimization of A₍₂₇₎ RNA-Hfq interactions for comparison with in vitro *K_d values*

With the A_{27} RNA as one of the two key interactors for comparison with *in vitro* K_d values, we asked how the A_{27} RNA might be optimally used in the B3H assay. We had previously looked at two sets of RNAs with different terminators: minimal distal face binders (see Section III-1-i), which lacked an intrinsic terminator and mRNA fragments,

---

[6] While a large dataset for comparison could in principle be put together from multiple studies, differences in approach (*i.e.* filter binding vs. EMSA vs. other methods) and in protocols mean that K_d values even for the same interaction can vary substantially from one paper to another. We therefore limited our comparisons to K_d values collected and published as part of the same study to ensure our comparisons were meaningful.

which had the intrinsic terminator sequence from the *trpA* operon added to their 3' ends (see Section III-2-i). We had been able to detect interactions both without a ρ-independent terminator and with such a terminator, but it was not yet clear if one of these options would provide more accurate results. We expected that a ρ-independent terminator may be ideal, since it would eliminate an undetermined amount of additional sequence that would be transcribed before ρ-dependent termination. This additional sequence may change the secondary structure of the bait RNA, sterically hinder interactions with Hfq, or interact with Hfq itself in unpredictable ways. However, since ρ-independent terminators interact with Hfq in many sRNA-Hfq interactions, introducing this type of terminator to the end of bait RNA constructs had the possibility of creating misleading B3H signal stemming from a terminator-Hfq interaction rather than an interaction with the bait RNA.

To attempt to differentiate between these options and determine RNA construct to use, we tested poly(A) constructs both with and without *trpA* terminator sequences.[7] We looked at each construct's interactions with the set of Hfq variants we would use for determining the relationship between B3H signal and K_d, and asked whether either construct showed greater consistency with *in vitro* data. We expected that if either type of construct was significantly affected by its type of termination, we would see B3H signal that was dramatically inconsistent with *in vitro* results and different from the other construct's results. The results of this test are shown in Figure 27.

---

[7] Due to obstacles in cloning, the poly(A) sequence without a *trpA* terminator contains 29 adenosines rather than the 27 adenosines used by Mikulecky *et al.* (18) and in our construct with a *trpA* terminator. Since the purpose of the construct is to be poly(A) rather than to have a specific number of adenosines, we expect this difference to be unimportant in comparing the two constructs.
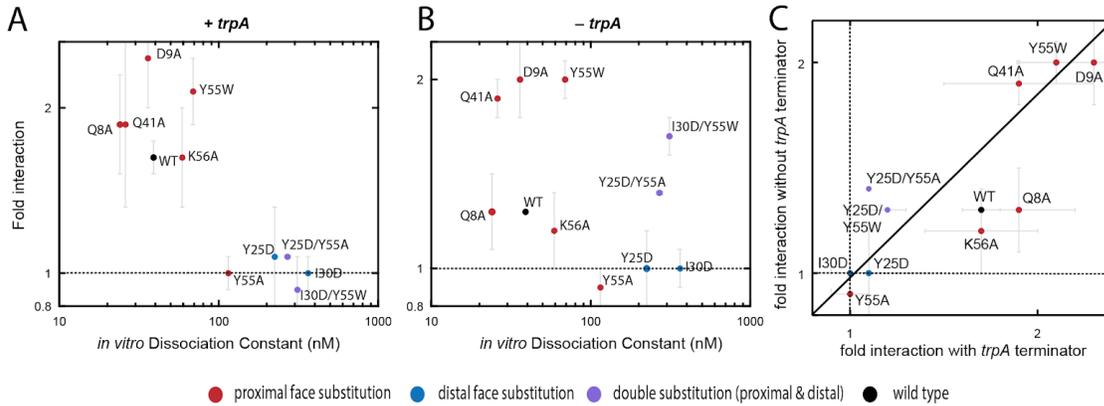
Figure 27. Comparison of A_{27}-Hfq interaction with and without a *trpA* terminator. B3H data plotted against *in vitro* K_d (18). B3H results collected using a liquid β-galactosidase assay in Δ*hfq E. coli* reporter cells with a -62 O_L2 site (KB483). Cells were transformed with plasmids expressing α-Hfq, an MS2^{hp}-bait RNA, and CI-MS2^{CP}. A constitutive promoter (35u4) was used to express CI-MS2^{CP}. Cells were grown in LB media + CCTS + 0.2% arabinose + 5 μM IPTG. *In vitro* results collected by electrophoretic mobility shift assay (18). (A) B3H signal vs. K_d for A_{27} construct with a *trpA* terminator sequence. (B) B3H signal vs. K_d for A_{27} construct without a *trpA* terminator sequence (termination carried out by ρ-dependent termination). (C) Comparison of + and − *trpA* poly(A) RNAs. Data from panels (A) and (B) are plotted against one another. Black line indicates where data points would fall for identical interactions with and without the *trpA* terminator. In all panels, red indicates a proximal face substitution, blue indicates a distal face substitution, and purple indicates a double substitution of residues on both the proximal and distal face. Wild type Hfq is indicated in black.

We observed a general trend that with intrinsic termination (*i.e.* with a *trpA* terminator) stronger interactions had higher signal, and weaker interactions had lower signal (Fig. 27A). This trend was not apparent in the absence of a *trpA* terminator – in this case, some strong interactions produced low signal, and weaker interactions had relatively high signal (Fig. 27B). Furthermore, when fold interactions for each construct were compared, the data did not indicate that the *trpA* terminator added artificial proximal face binding (Fig. 26C). Only a few regions had noticeable differences between the constructs: some interactions with proximal face mutants and wild type Hfq produced higher signal in the + *trpA* construct, and proximal/distal double mutants produced higher signal when the *trpA* terminator was not present (Fig. 27C). If the + *trpA* construct depended on the

proximal face for signal due to terminator-Hfq binding, we would expect to see decreased proximal face signal from the $+$ *trpA* construct $-$ the opposite of what we observed. Furthermore, the increased signal from double mutant interactions in the $-$ *trpA* poly(A) construct was concerning, since those Hfq variants have some of the weakest interactions with poly(A) RNAs (Fig. 27C, Supplementary Table 1). Since the $+$ *trpA* terminator poly(A) sequence seemed most consistent with *in vitro* binding affinity, and misleading interactions between the terminator-Hfq interactions did not seem to occur, we chose to use this construct in subsequent experiments for comparisons with K_d.

III-3-ii-b. Optimization of DsrA-Hfq interactions for comparison with in vitro K_d values

While we had consistently detected $\geq$ 2-fold interactions between the A_{27} RNA and Hfq that we considered sufficient for comparisons with K_d values[8], tests of DsrA-Hfq interactions typically gave fold interactions equal to or less than 2-fold. These fold interactions were not robust enough to make comparisons between stronger and weaker binding, so we asked what factors may be limiting our ability to detect DsrA-Hfq interactions. One factor that stood out in particular was secondary structure: when the MS2^{hp} sequence is added to DsrA, a section of the RNA is predicted to fold differently than its native conformation (Fig. 28A and B). We predicted that this misfolding may contribute to our limited ability to detect DsrA-Hfq interactions, and sought to find ways

---

[8] Most of the "high" interactions in Fig. 27 are at or below 2-fold. The results from this experiment give the appearance that A_{27}-WT Hfq interactions may also not be consistently robust; however, the majority of experiments with A_{27} (some of which are not presented here) have produced greater fold interactions. We therefore did not find it necessary to further optimize A_{27} beyond studying the impacts of transcription termination.

to modify the constructs such that the MS2^{hp} and DsrA sequences would be unchanged, but DsrA could fold into its native conformation.

We hypothesized that the XmaI site was important in causing DsrA to misfold. It was located close in the construct's sequence to the misfolded section of the RNA and is GC-rich, which gave it the potential to cause unintended base pairing and therefore disruption of the native secondary structure. Since the XmaI site is used for restriction enzyme cloning and is unimportant in detecting interactions, we wondered if we could modify it in ways that would prevent DsrA misfolding. We designed two new constructs: one with the XmaI site deleted (ΔXmaI), and one where the XmaI site is replaced with six adenosines (XmaI::A$_6$) (Fig. 28C and D). Each of these constructs is predicted to retain DsrA's native conformation (Fig. 28C and D).
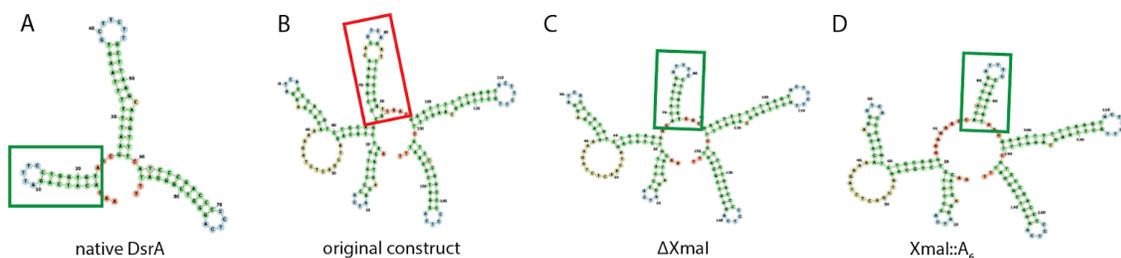
Figure 28. Predicted secondary structures of DsrA constructs. Structural prediction done using FORNA (14). (A) Predicted structure of native DsrA. Hairpin of interest is indicated in the green box. (B) Predicted structure of 2xMS2^{hp}-DsrA. Misfolded hairpin indicated in the red box. (C) 2xMS2^{hp}-DsrA with XmaI site between MS2^{hp} and DsrA sequence deleted. Correct folding of the hairpin of interest indicated in the green box. (D) 2xMS2^{hp}-DsrA with XmaI site replaced with six adenosine residues. Correct folding of the hairpin of interest indicated in the green box.

Having designed these constructs and predicted that they would produce more robust signal than the original DsrA construct, we tested their interactions with Hfq in the B3H assay. These results are shown in Fig. 29. We observed a slight increase in fold interaction from the original construct to the ΔXmaI construct and a substantially greater

amount of signal from the XmaI::A_6 construct. While the level of signal from the XmaI::A_6 construct was intriguing and a source of inspiration for future experiments (see Section IV-6), we chose to use the ΔXmaI construct in subsequent comparisons to K_d, as we do not know the reason for the robustness of the XmaI::A_6 construct's interaction with Hfq and because the ΔXmaI construct's sequence is most consistent with that used in the collection of the *in vitro* data we are comparing our data against (18).
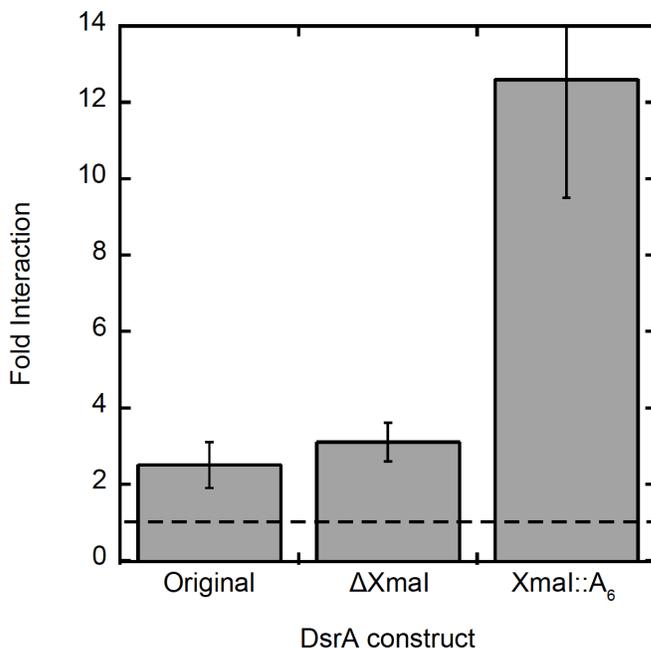
Figure 29. Comparison of original and modified DsrA constructs. Original refers to the first DsrA construct used by the lab, with a CCCGGG XmaI sequence. In the ΔXmaI construct, the XmaI site has been removed; in the XmaI::A_6 construct, the XmaI site's CCCGGG is replaced by AAAAAA.

III-3-ii-c. In vitro-in vivo comparisons show a sigmoidal relationship between dissociation constant and fold interaction

With 35u4 chosen as our most promising iteration of the pAdapter plasmid and our two RNAs optimized for reliable interactions in the B3H assay, we were interested in

examining the relationship between K_d and fold interaction across a larger set of data points. We recreated a number of the interactions published by Mikulecky *et al.* (18) between Hfq and the interacting RNAs A_{27} and DsrA in the B3H assay, and the comparison between the B3H signal we detected and published K_d values are shown in Figure 30A. All mutants used in these experiments were tested for expression level in a B2H assay, and none showed significant expression defects (Fig. 30B).
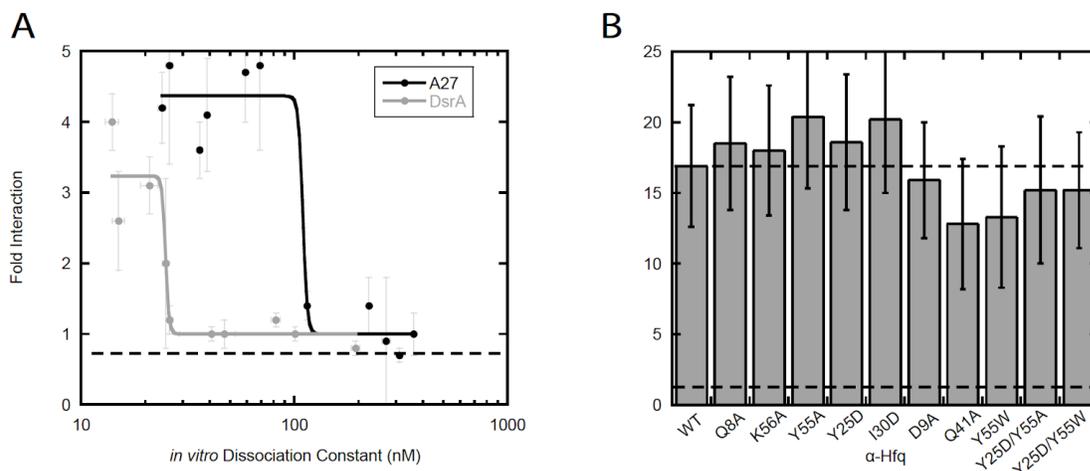
Figure 30. Comparison of A_{27} and DsrA B3H signal and *in vitro* binding energetics. (A) B3H data plotted against *in vitro* K_d (18). B3H results collected using a liquid β-galactosidase assay in Δhfq *E. coli* reporter cells with a -62 $O_L 2$ site (KB483). Cells were transformed with plasmids expressing α-Hfq, an MS2^{hp}-bait RNA, and CI-MS2^{CP}. A constitutive promoter (35u4) was used to express CI-MS2^{CP}. Cells were grown in LB media + CCTS + 0.2% arabinose + 5 μM IPTG. *In vitro* results collected by electrophoretic mobility shift assay (18). (B) B2H data collected using a liquid β-galactosidase assay in Δhfq *E. coli* reporter with a -62 $O_L 2$ site (KB483). Cells were transformed with plasmids expressing α-Hfq and CI-Hfq. Cells were grown in LB media + CCT + 5 μM IPTG. Figures made using Kaleidagraph (16).

Testing these interactions in the B3H assay and comparing their signal to *in vitro* K_ds led us to observe a sigmoidal relationship between B3H signal and K_d on a logarithmic scale (Fig. 30A). Both A_{27}-Hfq and DsrA-Hfq interactions have a range of strong interactions for which fold interaction is roughly consistent and relatively high (Fig. 30A). Each RNA then has a K_d threshold at which interactions drop sharply to

approximately 1-fold (Fig. 30A). These trends – the fold interactions at stronger and weaker interactions, and the drop at the threshold K_d – can be approximately modeled using a sigmoidal curve (Fig. 30A).

In addition to the liquid-assay approach used to collect data described thus far, the B3H system can also be used with a plate-based assay that provides qualitative information about RNA-protein interactions. We asked whether the relationship between B3H signal and K_d observed in a plate-based assay was consistent with that observed in liquid assays and performed a plate-based assay for A_{27} interactions with the same set of Hfq mutants as used in liquid assays (Fig. 30A and 31). As with liquid assays, our plate-based assay showed a set of strong interactions with relatively high signal: interactions Q8A through Y55W, where the experimental bacterial growth is significantly bluer than negative controls (Fig. 31). At the same K_d value threshold observed in liquid assays (between 69 nM and 115 nM, the K_d values for Y55W and Y55A interactions) the blueness of the experimental bacterial growth becomes equivalent to or lighter than the respective negative controls, which is indicative of the same detection threshold observed in liquid assays (Fig. 30A and 31).
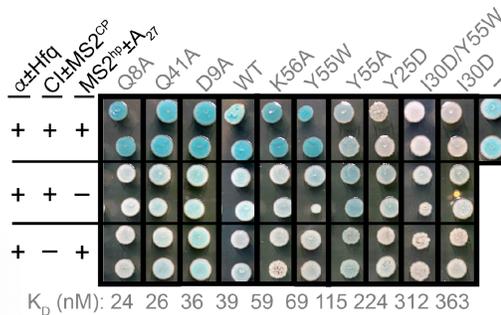
Figure 31. Plate-based comparison of A_{27} B3H signal and *in vitro* binding energetics. B3H results obtained using Δhfq *E. coli* reporter cells with a -62 O_L2 site (KB483) plated on LB-agar media + CCTS + 0.2% arabinose + 1.5 μM IPTG + 40 μg/mL X-gal + 250 μM TPEG. Cells were transformed with plasmids expressing α-Hfq, an MS2^hp-bait RNA, and CI-MS2^CP. A constitutive promoter (35u4) was used to express CI-MS2^CP. *In vitro* results collected by electrophoretic mobility shift assay (18).

Chapter IV: Discussion

The results presented in this study represent several advances in the RNA-protein interaction detection capabilities of the bacterial three-hybrid assay. To establish detection of mRNA-Hfq interactions, we designed artificial RNAs to interact with Hfq's distal face and confirmed both that they interact with Hfq and that those interactions occur in a way similar to native mRNA-Hfq interactions. Furthermore, we detected one interaction between a native mRNA sequence and Hfq, demonstrating in the process of detecting it that it is necessary to avoid ribosomal stalling at the end of a B3H RNA transcript. Finally, we showed that efforts at optimizing the B3H assay have led to greater consistency with *in vitro* binding energetics data and increased sensitivity, and observed a trend in the relationship between B3H signal and *in vitro* data for two RNA-Hfq interactions using the most up-to-date B3H conditions at the time of writing.

IV-1. Single-stranded $(A\text{-}A\text{-}N)_n$ motif-containing RNAs interact detectably with Hfq in the B3H assay

When the work described here began, the B3H assay had not yet been used to study mRNA-protein interactions – it had solely been established as a tool capable of detecting sRNA-Hfq interactions. While we expected that, in principle, it should be possible to study any of a wide range of interactions using this assay, it was necessary for us to determine its mRNA-protein detection capabilities. We therefore designed a set of potential distal face binding RNAs based on the $(A\text{-}R\text{-}N)_n/(A\text{-}A\text{-}N)_n$ motif that has been

proposed as a primary Hfq-binding sequence in mRNA 5' UTRs. Our goal with these RNAs was to demonstrate that the B3H assay was capable of detecting interactions that used the same mechanism as real mRNA-Hfq interactions without the potential complicating factors of native mRNA sequences.

We found that some, but not all, of the distal face-binding RNA sequences produced detectable B3H signal. The greatest signal was detected from an $(AAU)_{17}$ RNA interacting with Hfq, with lower but reliable signal from an $(AAC)_{13}$ RNA. We anticipated that both of these RNAs would interact with Hfq, since each RNA possessed a repeated $(A-A-N)_n$ motif, and our results supported this expectation. Furthermore, we did not detect interactions for $(GAC)_{13}$ or $(AACC)_{13}$ RNAs, also following our expectations – each of these RNAs did not follow the $(A-R-N)_n/(A-A-N)_n$ motif. The $(GAC)_{13}$ RNA simply did not have nucleotides in the correct order to bind to Hfq, and while the $(AACC)_{13}$ did contain an (AAC) sequence, it contained 4-nucleotide repeats rather than repeats of 3 nucleotides. This inconsistency with the 3-nucleotide $(A-A-N)_n$ motif is the most plausible explanation for why the other (AAC)-containing sequence, the $(AAC)_{13}$ RNA, produced a substantial fold interaction while $(AACC)_{13}$ did not.

The most surprising result was the lack of interaction detected from an $(AGC)_{17}$ repeat RNA. There are two plausible reasons for this lack of interaction, both of which may be true. The most likely reason for the lack of $(AGC)_{17}$-Hfq interaction detection is the $(AGC)_{17}$ repeat's secondary structure. The RNA is predicted to form a hairpin due to base pairing between guanines and cytosines in the three-nucleotide motif (Supplementary Figure 1). Hfq preferentially binds single-stranded RNAs and base

pairing causes nucleotide bases to be inaccessible to distal face binding sites. A substantial amount of secondary structure formation may therefore prevent interactions with RNA sequences that would, were they single stranded, interact with Hfq. Hairpin formation in the $(AGC)_{17}$ repeat RNA may therefore explain the lack of $(AGC)_{17}$-Hfq interaction signal in the B3H assay. Additionally, while the $(AGC)_{17}$ repeat follows the $(A-R-N)_n$ motif, the purine flexibility in the R position is supported by theoretical evidence only, and experimental results have not yet demonstrated binding of guanine residues in the R position. It is therefore possible that binding at an affinity detectable by the B3H assay only occurs when an adenine is present in both the A and R positions of the motif. While the presence of a G in the R position of the motif may prevent interaction, the high likelihood of hairpin formation in this RNA makes it difficult to ascertain whether an interaction would occur if the RNA was single-stranded. These data therefore do not constitute substantive evidence in favor of or against Hfq binding to $(A-G-N)_n$ RNA sequences. A closer examination of $(A-G-N)_n$-RNA binding would require designing $(A-R-N)_n$ constructs with G in the R position and no predicted hairpin formation, such as an $(A-G-G)_n$ construct.

In addition to demonstrating the capability of the B3H assay to detect single-stranded $(A-A-N)_n$ motif interactions with Hfq, we wanted to confirm that the signal we detect from these interactions depends on the distal face of Hfq. Notably, we saw a substantial decrease in B3H signal for the $(AAU)_{17}$-Y25D Hfq interaction as compared to the $(AAU)_{17}$ interaction with wild type Hfq. Tyr25 is a residue in the adenine binding pocket on the distal face of Hfq that is proposed to perform π stacking

87

interactions with adenine bases in the A position of the $(A-R-N)_n$ motif (see Section I-2-iv-a). Thus, dependence on Tyr25 for B3H signal is indicative of binding to $(A-R-N)_n$ binding sites on the distal face of Hfq. This result suggests that the $(AAU)_{17}$-Hfq interaction does indeed occur via these binding pockets on the distal face of Hfq. $(AAU)_{17}$-Hfq and ChiX-Hfq interactions do not appear to depend on Lys31, another distal face residue tested in this assay. However, this result is not strong evidence against $(AAU)_{17}$ distal face dependence. The Tyr25 residue's proposed interaction with the $(A-R-N)_n$ motif provides affinity for the first adenine in the motif, while the Lys31 residue acts only by a non-specific electrostatic interaction. It is therefore plausible that a substitution of Lys31 with alanine would not be sufficient to eliminate the $(AAU)_{17}$-Hfq interaction, but a Tyr25 to aspartate substitution would be. Thus, we conclude that the B3H assay is capable of detecting interactions between single-stranded $(A-A-N)_n$ motif-containing RNAs and the distal face of Hfq, and that interactions between these two molecules occur via the same mechanism expected for native mRNA binding to Hfq.

IV-2. In-frame stop codons allow for detection of mRNA fragment-Hfq interactions

Having determined that binding between RNA $(A-A-N)_n$ motifs and Hfq's $(A-A-N)_n$ binding pocket could be detected using the B3H assay, we sought to establish detection of interactions with native mRNA sequences that use this interaction mechanism. We were initially unable to detect any interactions between Hfq and our mRNA 5' UTR constructs; however, we noticed a key difference between the mRNA fragments we tested in the B3H assay and actual mRNA transcripts. Our mRNA 5' UTR

constructs included the 5' UTR, the Shine-Dalgarno sequence, the start codon, and several codons of the open reading frame. However, these transcripts lacked a stop codon; therefore, while all of the sequence elements necessary for ribosome loading are present in these constructs, the element necessary for ribosome dissociation is not. This introduces the potential for ribosomes to stall at the 3' end of these transcripts, causing B3H hybrid RNAs to be targeted by tmRNA for degradation as part of normal cellular surveillance for damaged mRNA via the non-stop decay pathway. If this ribosomal loading and stalling indeed occurs, and non-stop decay is rapid enough to substantially decrease cellular hybrid RNA abundance, targeting by this pathway would present an obstacle to detecting interactions between Hfq and our original mRNA 5' UTR hybrid RNAs. Additionally, buildup of stalled ribosomes on not-yet-degraded hybrid RNAs could sterically hinder interactions with Hfq, blocking interactions even if cellular RNA abundance is sufficient for interaction detection.

We hypothesized that tmRNA-mediated degradation prevented us from detecting mRNA-Hfq interactions, and created modified mRNA constructs with UAA stop codons at the end of their open reading frames. When tested in the B3H assay, introduction of a stop codon allowed us to detect Hfq's interaction with one RNA, the 5' UTR of *sodB*. To confirm that the impact of the stop codon was truly translational, and not due to a direct interaction between the stop codon and Hfq, we tested a set of constructs with modified stop codons. Our goal was to determine if translation termination was necessary for detection of the *sodB* 5' UTR-Hfq interaction, or if the addition of the UAA nucleotides

to the construct was sufficient to produce an interaction even in the absence of termination.

Our results showed that B3H signal increases for the *sodB*-Hfq interaction when an in-frame stop codon is present at the end of the fragment's open reading frame, but not when the codon is scrambled or moved out of frame. We were therefore able to conclude that functional stop codons are necessary for detection of this *sodB* construct's interaction with Hfq, and that rearrangements of the same nucleotides are not adequate for interaction detection. These results support our hypothesis that 5' UTR fragments containing ribosome binding sites and open reading frames are subject to translation (and non-stop decay and/or steric hindrance) that can be eliminated by the addition of a stop codon, and that accounting for this process in our construct design improves our ability to measure mRNA sequence fragment interactions with Hfq. Additionally, this provides evidence against the hypothesis that the increased signal is due to a direct stop codon-Hfq interaction, since the same sequence or set of nucleotides does not increase interaction when not part of a in-frame stop codon. The addition of a stop codon is effective for detecting interactions with the *sodB* 5' UTR as we have constructed it for the B3H, it points to the importance of overcoming least two barriers to detecting interactions: steric hindrance from binding of other complexes, and degradation by endogenous cellular pathways such as mRNA surveillance.

IV-3. B3H assay optimization improves consistency with *in vitro* binding energetics

   With an eye toward global optimization of the B3H assay for many interactions in addition to detection of mRNA 5' UTRs, we have also investigated optimized iterations of the DNA-RNA adapter plasmid. We sought to confirm if a low level of constant expression, independent of signaling from an inducer, of the adapter protein in two plasmids (pCW17 and 35u4) improved B3H signal over the original IPTG-dependent promoter for several interactions, and if increased signal was consistent with *in vitro* measures of binding affinity for interactions between the sRNA OxyS and a set of Hfq variants. We concluded that use of constitutively active pCW17 and 35u4 improves B3H signal over the original IPTG-inducible pKB989. Furthermore, we found that 35u4 allows detection of and differentiation between the widest range of K_d values of the three plasmids.

   Having established that 35u4 was the most promising pAdapter plasmid of the three we tested, we asked what the relationship between *in vitro* data and B3H signal collected with 35u4 looked like for a larger set of interactions. We tested a panel of Hfq mutants for their interactions with two RNAs, a poly(A) A_{27} transcript and the *E. coli* sRNA DsrA and compared B3H signal to published K_d values. In the process of optimizing the assay to detect these two interactions, we determined that the addition of an ρ-independent *trpA* terminator to a poly(A) sequence lacking its own terminator may improve consistency with *in vitro* data. We also found that modifying RNA constructs to prevent misfolding can improve B3H signal, and that including a poly(A) stretch between the MS2^{hp} and Hfq-interacting RNA sequence results in a substantial increase in fold

interaction (although the reason for the difference in signal between this type of construct and other attempts to improve folding is not clear – see Section IV-5). When plotting these two types of data against each other, we found that a sigmoidal curve was a useful way to represent the trends we observed in both RNAs. While the curves used here to trace these trends are not rigorous mathematical models of the B3H signal-K_d relationship, they do summarize the overall trend observed for interactions with both RNAs. In each case, we observed a set of strong interactions with high B3H signal and low K_d, and a set of weak interactions with low B3H signal and high K_d. In between these two sets is a threshold K_d value at which B3H signal appears to drop sharply. While A_{27} and DsrA have different K_d values at which they reach their detection threshold, both follow a generally sigmoidal shape. This indicates that while different RNAs demonstrate different trends, each individual RNA may in general display a sigmoidal relationship between B3H signal and *in vitro* K_d for interactions with Hfq when 35u4 is used as the pAdapter plasmid. In addition to demonstrating this relationship using fold interaction in a liquid assay, we also qualitatively observed a similar relationship between binding affinity and B3H signal for A_{27}-Hfq interactions using a plate-based assay. Similar to liquid assays, this experiment produced a set of strong interactions that showed greater signal via blueness of bacterial growth, and a set of weaker interactions with white, low-signal growth. The divide between these two sets of interactions occurred at the same location as in liquid assays, suggesting that plate-based assays, while qualitative, can produce results consistent with liquid assay data.

We can imagine several scenarios accounting for this trend. It is possible that even a fully optimized assay would have a threshold interaction strength at which the RNA-protein interaction can no longer produce signal, although work published by the Wickens lab on the relationship between Y3H signal and K_d suggests this is not the case (see Section 1-4-ii). A potentially more likely explanation is that at a certain weakness of interaction, limiting factors present in the current iteration of the assay overwhelm the ability of the system to activate *lacZ* transcription with that binding event. As limiting factors are optimized or removed, this threshold might shift right (*i.e.* to weaker interactions) even if fold interaction decreases, allowing detection of weaker interactions with a linear B3H signal-K_d relationship. The difference in detection threshold seen here between A_{27} and DsrA may support this explanation, as it is likely that different RNAs are affected by various limiting factors to different degrees.

IV-4. Applications

The applications of the work presented here apply broadly to many uses of the B3H, as they improve multiple aspects of the assay's utility. While Hfq-mRNA interactions have been relatively well-studied, there are still aspects of their mechanisms that are not yet understood, and adding the B3H assay to the set of approaches available to study them creates an additional avenue for understanding. Hfq is not the only protein that binds mRNAs, and this work presents the opportunity to study the interactions between similarly constructed mRNA fragments and other RNA-binding proteins. Indeed, the results presented here have provided a jumping-off point for studies of

mRNA 5' UTR interactions with the *E. coli* protein ProQ with the B3H assay that are already underway. In addition to its use in studying specific RNA-protein interactions, an advantage of the B3H assay over other approaches is its accessibility to genetics. Additionally, n-hybrid assays have previously been used to discover new RNA-binding proteins via genetic means, and adding mRNA 5' UTRs to the set of usable RNAs expands the assay's utility in discovering new interactions.

In addition to advancing the study of mRNA 5' UTRs in the B3H assay, results shown here support ongoing work in optimizing the assay globally. We have presented a framework for studying the B3H assay itself and the data it presents in the context of *in vitro* binding energetics. Furthermore, we have established a baseline for comparison – data collected with the original pAdapter plasmid, pKB989 – and compared it to several improved plasmids. This comparison has demonstrated that these plasmids are indeed improvements over the original pAdapter. We have also collected information on how B3H data collected using 35u4 behaves over a wide range of binding affinities. Not only does this work demonstrate that attempts at optimization are producing useful results, it also creates a body of data for future improvements (both to the pAdapter plasmid and other aspects of the assay) to be compared to, and a process by which those improvements may be evaluated. In collecting this data, we have supported the ongoing development of the B3H assay, making it a more useful method for studying an increasingly wide range of RNA-protein interactions.

IV-5. Limitations

While the results presented here are informative, a number of limitations were encountered in the process of collecting this data that restrict its usefulness. One major obstacle, alluded to by uncertainty in our inability to detect an $(AGC)_{17}$-Hfq interaction, is RNA secondary structure. The way an RNA folds is typically significant to its interactions with proteins – either the protein binds to a specific structure or (in Hfq's case) to specific unpaired nucleotides, and if those structures do not form or nucleotides are sequestered in a non-native secondary structure, the interaction will be obstructed. The hairpin formation that we believe is preventing detection of an $(AGC)_{17}$-Hfq interaction could be considered an oversight in our initial design of distal face-binding RNAs. However, secondary structure is a large issue for native RNAs that, endogenously, fold in particular ways to enable (or prevent) interactions with Hfq. When sections of these RNAs are removed from their natural context and inserted into a hybrid RNA with alien sequence (the MS2^{hp} and/or *trpA* terminator), they are often at high risk of misfolding in ways that prevent native interactions. Indeed, we hypothesize that this may be a primary obstacle to detecting *eptB* and *mutS* 5' UTR interactions with Hfq, or any other as-yet-undetectable mRNA-protein interaction. Structural predictions suggest that of the three mRNAs tested here, only the *sodB* 5' UTR retains its native conformation when the MS2^{hp} and *trpA* terminator are added (Supplementary Figure 2). Issues of this kind with folding may be remedied by adjusting the sequence of the MS2^{hp} or terminator in ways that do not affect the sequence's endogenous structure, but it is often unclear what modifications would accomplish this.

To further complicate this problem, it is difficult to know with certainty how an RNA will fold. Structure predictors such as FORNA (14) are useful ways to determine how a sequence is likely to fold based on thermodynamic parameters, but programs of this kind are limited in a number of ways. For instance, they typically do not take into account co-transcriptional folding, are often unable to model complex RNA structural components or those that occur between distant sequence elements, and provide a single snapshot of a structure that is likely to be dynamic and vary between individual molecules in the cell. There are many potential reasons for these predictions to be inaccurate, and even when they are accurate, they are oversimplifications. For these reasons, they have not been relied on heavily for interpretation of our results. However, despite their many limitations, there is dearth of useful alternatives – actual structural determination is highly impractical for the scope of B3H studies. Thus, when structural explanations are used here, they are based on computational predictions, but have an element of uncertainty due to problems with this type of approach.

An additional limitation of the work presented here is in modeling of relationships between B3H signal and *in vitro* data. While we present a sigmoidal trend as a representation of this relationship, it is not a rigorous mathematical model. Significantly more statistical work and expertise would likely need to be applied to this data than has been thus far to produce an accurate or useful mathematical expression of the relationships we are studying. We hypothesize that additional sets of RNA-protein interactions would follow a similar trend if studied in a similar way, but without more data, our ability to confidently make this prediction is limited. There are several reasons

why we are unable to present a larger set of data: one is a lack of large sets of published K_d values for individual RNA-protein interactions. While we were able to compare A_{27} and DsrA data for a large set of mutants, most published *in vitro* studies use a fraction of that number of protein variants. Comparisons to K_d values across different studies is uninformative, as each study's methodology differs and thus their data differs as well. Additionally, some published *in vitro* data sets that would be useful for comparison cannot yet be studied in the B3H assay – for instance, data is available for *rpoS* interactions with the same set of Hfq variants used here for A_{27} and DsrA, but *rpoS* 5' UTR-Hfq binding is not yet detectable in the B3H system. While it may be possible in the future to collect *in vitro* data for a specific desired set of interactions in the Berry lab for more targeted comparisons, the skill and labor required to collect this data is outside the scope of the work described here.

IV-6. Future directions

As the results presented here represent initial steps toward expanding the B3H assay in several directions, they present a variety of future experimental directions. One clear area for next steps is in mRNA-protein interaction detection – while we have indicated how some of these interactions have been detected thus far, it is clear that there is much more work to be done to be able to detect a large number of mRNA 5' UTR interactions with proteins. Some of these steps are already underway: for instance, members of the Berry lab have begun testing mRNA 5' UTR interactions in the B3H assay with the protein ProQ, working toward expanding this type of detection to proteins

other than Hfq. Other ongoing and future work can address the concerns about misfolding raised previously in Section IV-5. By manipulating the MS2^{hp} and terminator sequences added to a native mRNA sequence, it may be possible to prevent misfolding of 5' UTRs – for instance, by moving the MS2^{hp} to the 3' end of the mRNA insert, by modifying the MS2^{hp} so it can act as a transcriptional terminator, and by modifying these sequences to change their stability and thus their impact on the folding of the 5' UTR insert.

Another avenue for detecting mRNA-protein interactions not explored here is the impact of normal regulatory protein activity on detection. Since most Hfq-mediated regulation involves degradation of the target RNA, it is possible that the very interactions we are attempting to detect are responsible for lowering the abundance of the hybrid RNA. One potential solution to this issue would be testing mRNA-Hfq interactions in *E. coli* strains where targeting sRNAs have been deleted; while this approach is labor-intensive, it would eliminate a possible avenue for degradation of the mRNAs we are attempting to study.

Endogenous bacterial degradation affects our ability to detect signal in another way: cleavage of bait RNAs resulting in free MS2^{hp} fragments that can interact with the MS2^{CP} and competitively inhibit binding of full-length MS2^{hp}-RNA. Studies of hybrid RNA abundance have indicated the presence of these fragments, which have been separated from the Hfq-interacting RNA sequences they were transcribed with (Supplementary Fig. 1). When they interact with the MS2^{CP}, they prevent binding of a full RNA capable of Hfq (or other protein) interactions (Supplementary Fig. 1). While thus far it has been unclear how to avoid this issue, results of DsrA-Hfq interaction

optimization presented here serendipitously suggest a potential way to harness cellular RNA degradation signals to remove unwanted MS2^{hp} fragments. Introduction of a poly(A) sequence between the MS2^{hp} and Hfq-interacting RNA sequence causes a substantial increase in fold interaction. It is possible that when a poly(A) sequence is present near the site of cleavage between the MS2^{hp} and the rest of the RNA, that poly(A) sequence (which is used in normal bacterial processes as a flag marking RNA for degradation) may facilitate the degradation of MS2^{hp} fragments, preventing them from competitively inhibiting full-length RNA binding to the MS2^{CP}. While this mechanism is speculative, it suggests making similar modifications to other low-interacting RNAs as have been made to DsrA in this study to determine if such poly(A) stretches will improve detection of other RNA-protein interactions, as well as doing targeted studies of RNA levels in cells used for B3H assays to determine if competitive inhibitor MS2^{hp} fragments are less abundant when poly(A)-tailed.

In addition to future directions in the study of mRNA-protein interactions in the B3H assay, the work presented here also suggests ways in which assay optimization and assessment of that optimization may proceed in the future. One way in which the B3H assay can be better understood is to collect more data for *in vivo-in vitro* comparisons – for instance, by working to establish a reliable *rpoS* 5' UTR-Hfq interaction that can be used in comparisons with published *rpoS* interaction data (available from the same source as the A$_{27}$ and DsrA data used in this study.). The most useful data for these comparisons may in the future be collected by members of the Berry lab – while *in vitro* experiments are a significant amount of additional labor, measuring our own K$_d$s would allow us to

99

use mutants that are well-expressed in the B3H assay and find interactions with a useful range of binding affinities for comparison, and to look at specific RNAs of interest to our work. Additionally, work using this type of *in vivo-in vitro* comparison can be done alongside future optimization – of the CI-MS2^{CP} promoter, as shown here, or of other aspects of the assay. This type of comparison will, as it has here, complement and add information to efforts to make the B3H assay more accurate and useful.

Appendix

Supplementary Table 1. *In vitro* dissociation constant values and B3H fold interactions. [1] K_d values measured by electrophoretic mobility shift assay and reported in Mikulecky *et al.* (2004). [2] K_d values measured by filter binding assay and reported in Olejniczak (2011). – indicates no value was measured for that interaction.

Hfq variant	Dissociation Constant (K_d, nM)			B3H Fold Interaction				
	A_{27}[1]	DsrA[1]	OxyS[2]	A_{27}	DsrA	OxyS		
						pKB989	pCW17	35u4
WT	39 ± 1	32 ± 1	1.7 ± 1.1	4.1 ± 0.8	3.1 ± 0.4	1.6 ± 0.5	2.9 ± 0.4	2.6 ± 0.2
Q8A	24 ± 1	19 ± 1	–	4.2 ± 0.5	–	–	–	–
D9A	36 ± 1	14 ± 1	–	3.6 ± 0.4	4 ± 0.4	–	–	–
Y25D	224 ± 2	15 ± 1	3.5 ± 1.6	1.4 ± 0.4	2.6 ± 0.7	0.9 ± 0	1.9 ± 0.3	1.7 ± 0.3
I30D	363 ± 3	26 ± 1	5.9 ± 2.3	1.0 ± 0.3	1.2 ± 0.2	0.7 ± 0	1.0 ± 0.3	1.3 ± 0.4
Q41A	26 ± 1	25 ± 1	–	4.8 ± 1.4	2 ± 1.2	–	–	–
Y55A	115 ± 2	82 ± 4	–	1.4 ± 0.2	1.2 ± 0.1	–	–	–
Y55W	69 ± 1	47 ± 1	–	4.8 ± 1.2	1 ± 0.2	–	–	–
K56A	59 ± 1	101 ± 2	100 ± 20	4.7 ± 0.7	1 ± 0.1	0.6 ± 0.1	0.8 ± 0.1	1.2 ± 0.1
I30D/ Y55W	312 ± 9	41 ± 12	–	0.9 ± 0.1	1 ± 0.1	–	–	–
Y25D/ Y55A	270 ± 12	194 ± 9	–	0.9 ± 0.1	0.8 ± 0.1	–	–	–
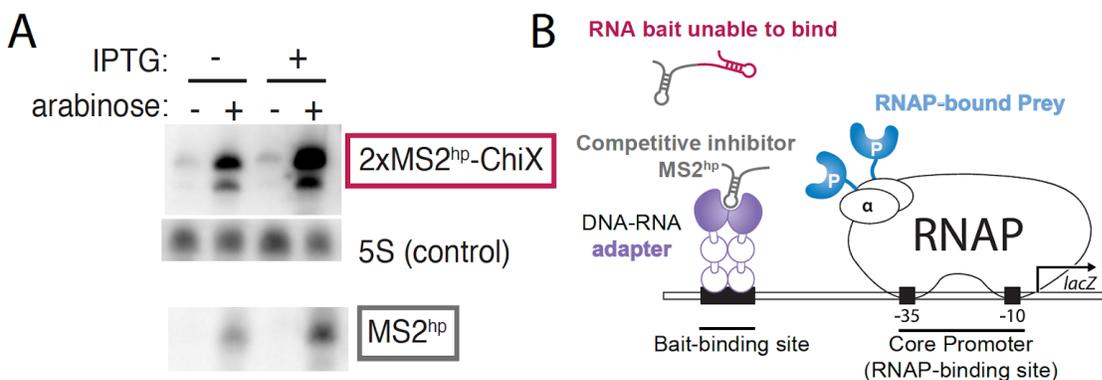
Supplementary Figure 1. Formation of MS2^{hp} degradation products and putative competitive inhibition of MS2^{CP} binding. (A) Northern blot of 2xMS2^{hp}-ChiX levels using a probe complementary to the MS2^{hp} sequence. Notable bands indicate presence of both full-length 2xMS2^{hp}-ChiX (red box) and MS2^{hp}-only fragments (gray box). Data collected by Katie Berry. (B) Putative mechanism of competitive inhibition by MS2^{hp}-only fragments. Binding of an MS2^{hp} fragment blocks binding of full-length RNA bait, preventing activation of *lacZ* transcription. Figure adapted from Berry and Hochschild 2018 (3).
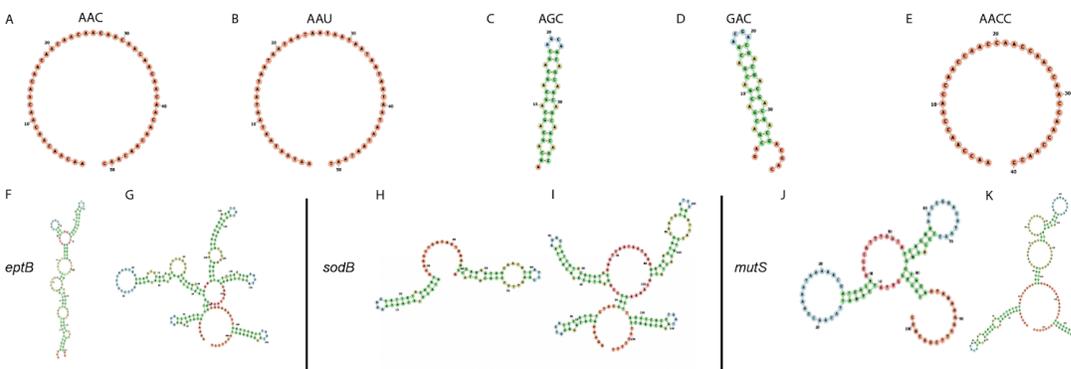
Supplementary Figure 2. Predicted RNA secondary structures. Predictions made using FORNA (14). (A) AAC insert structure (pHL1). (B) AAU insert structure (pHL2). (C) AGC insert structure (pHL13). (D) GAC insert structure (pHL4). (E) AACC insert structure (pHL5). (F) *eptB* 5' UTR insert structure (pHL37). (G) *eptB* 5' UTR + MS2^{hp} + stop codon + *trpA* terminator structure (pHL37). (H) *sodB* 5' UTR insert structure (pHL38). (I) *sodB* 5' UTR + MS2^{hp} + stop codon + *trpA* terminator structure (pHL38). (J) *mutS* 5' UTR insert structure (pHL40). (K) *mutS* 5' UTR + MS2^{hp} + stop codon + *trpA* terminator structure (pHL40).

References

1. Aiba H. 2007. Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr Opin Microbiol.* 10: 134-139.

2. Bernstein DS *et al.* 2002. Analyzing mRNA-protein complexes using a yeast three-hybrid system. *Methods*. 26: 123-141.

3. Berry KE and Hochschild A. 2018. A bacterial three-hybrid assay detects *E. coli* Hfq-sRNA interactions *in vivo*. *Nucleic Acids Res*. 46(2): e12.

4. Blattner FR *et al.* 1997. The complete genome sequence of *Escherichia coli* K-12. *Science*. 277: 1453-1462.

5. Brennan RG and Link TM. 2007. Hfq structure, function, and ligand binding. *Curr Opin Microbiol.* 10: 125-133.

6. Carmichael GC *et al.* 1975. The Host Factor Required for RNA Phage Qβ RNA Replication *in Vitro*. *J Biol Chem.* 250(10): 3607-3612.

7. Chen J and Gottesman S. 2017. Hfq links translation repression to stress-induced mutagenesis in *E. coli*. *Genes Dev*. 31: 1382-1395.

8. Du X *et al.* 2016. Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *Int J Mol Sci.* 17(144).

9. Dove SL, Joung K, and Hochschild A. 1997. Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature*. 386: 627-630.

10. Fields S and Song O. 1989. A novel genetic system to detect protein-protein interactions. *Nature.* 340: 245-245.

11. Gottesman S and Storz S. 2011. Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations. *Cold Spring Harb Perspect Biol*. 3:a003798.

12. Hellman LM and Fried MG. 2007. Electrophoretic Mobility Shift Assay (EMSA) for Detecting Protein-Nucleic Acid Interactions. *Nat Protoc.* 2(8): 1849-1861.

13. Hook B *et al.* 2005. RNA-protein interactions in the yeast three-hybrid system: Affinity, sensitivity, and enhanced library screening. *RNA*. 11: 227-233.

14. Kerpedjiev P, Hammer S, and Hofacker IL. 2015. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics.* 31(20): 3377-3379.

15. Keseler, IM *et al.* 2016. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res.* 45(D1): D543-D550.

16. Kirsch PD and Ekerdt JG. 2000. Kaleidagraph: Graphing and Data Analysis. Synergy Software. *J Am Chem Soc.* 122(47): 11755-11755.

17. Link TM, Valentin-Hansen P, and Brennan RG. 2009. Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA. *PNAS.* 106(46); 19292-19297.

18. Mikulecky PJ *et al.* 2004. *Escherichia coli* Hfq has distinct interaction surfaces for DsrA, *rpoS*, and poly(A) RNAs. *Nat Struct Mol Biol.* 11(12): 1206-1214.

19. Nelson DL and Cox MM. 2012. Lehninger Principles of Biochemistry, 6th ed.; W.H. Freeman, New York, NY. ISBN 978-1429234146.

20. Olejniczak M. 2011. Despite Similar Binding to the Hfq Protein Regulatory RNAs Widely Differ in Their Competition Performance. *Biochemistry.* 50: 4427-4440.

21. Pandey S *et al*. 2020. Genetic identification of the functional surface for RNA binding by *Escherichia coli* ProQ. *Nucleic Acids Res.* gkaa144.

22. Peng Y, Soper TJ and Woodson SA. 2013. Positional Effects of AAN Motifs in *rpoS* Regulation by sRNAs and Hfq. *J Mol Biol.* 426: 275-285.

23. Richards J *et al*. 2008. Quality Control of Bacterial mRNA Decoding and Decay. *Biochim Biophys Acta.* 1779(9): 574-582.

24. Robinson KE *et al.* 2014. Mapping Hfq-RNA interaction surfaces using tryptophan fluorescence quenching. *Nucleic Acids Res*. 42(4): 2736-2749.

25. Santiago-Frangos A *et al.* 2017. Acidic C-terminal domains autoregulate the RNA chaperone Hfq. *eLife*. 6: e27049.

26. Schu DJ *et al.* 2015. Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. *EMBO.* 34: 2557-2573.

27. SenGupta DJ *et al.* 1996. A three-hybrid system to detect RNA-protein interactions *in vivo*. *Proc Natl Acad Sci.* 93: 8496-8501.

28. Stockley, PG. 2009. Filter-Binding Assays. *Methods Mol Biol. DNA-Protein Interactions.* 543.

29. Updegrove TB, Zhang A, and Storz G. 2016. Hfq: the flexible RNA matchmaker. *Curr Opin Microbiol.* 30: 133-138.

30. Valentin-Hansen P, Eriksen M, and Udesen C. 2004. The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Mol Microbiol.* 51(6): 1525-1533.

31. Venkataraman K, Guja KE, Garcia-Diaz M, and Karzai AW. 2014. Non-stop mRNA decay: a special attribute of *trans*-translation mediated ribosome rescue. *Front Microbiol*. 5(93).

32. Waters LS and Storz G. 2009. Regulatory RNAs in Bacteria. *Cell*. 136: 615-628.