Published in final edited form as:

Methods. 2013 September 15; 63(2): 144–159. doi:10.1016/j.ymeth.2013.04.023.

Identifying and Characterizing Hfq-RNA Interactions

M.A. Faner and A.L. Feig*

Department of Chemistry, Wayne State University, 5101 Cass Ave., Detroit, MI, United States

Abstract

To regulate stress responses and virulence, bacteria use small regulatory RNAs (sRNAs). These RNAs can up or down regulate target mRNAs through base pairing by influencing ribosomal access and RNA decay. A large class of these sRNAs, called *trans*-encoded sRNAs, requires the RNA binding protein Hfq to facilitate base pairing between the regulatory RNA and its target mRNA. The resulting network of regulation is best characterized in *E. coli* and *S. typhimurium*, but the importance of Hfq dependent sRNA regulation is recognized in a diverse population of bacteria. In this review we present the approaches and methods used to discover Hfq binding RNAs, characterize their interactions and elucidate their functions.

Keywords

Hfq; sRNA; ncRNA; Protein-RNA interactions

1. Introduction*

Bacteria employ small regulatory RNAs (sRNAs) to control gene expression [1, 2]. In bacteria, sRNAs are critical for bacterial survival under adverse conditions as well as to express virulence factors [3]. There are two main types of sRNAs in bacteria. *Cis*-encoded transcripts originate from the same locus as the genes or operons they regulate, and have 1:1 correspondence with them. This class includes riboswitches and natural antisense transcripts. Riboswitches are RNA motifs encoded within the mRNA that modulate gene expression through structural rearrangements in response to a regulatory signal [4]. Natural antisense transcripts are RNAs transcribed from the opposite strand of the gene and act by base pairing with perfect complementarity to their target [5]. Unlike the *cis*-encoded sRNAs, *trans*-encoded sRNAs, which are the focus of this article, are transcribed from a different locus than their targets and act through imperfect base pairing. In this way they often regulate multiple mRNAs, forming a web of regulatory activities that occur in response to

© 2013 Elsevier Inc. All rights reserved

*To whom correspondence should be addressed afeig@chem.wayne.edu Phone: 313-577-9229 Fax: 313-577-8822.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

* Abbreviations: sRNAs, small regulatory RNAs; (ARN)_x, where A is adenine, R is adenine or guanosine, and N is any nucleotide; SELEX, systematic evolution of ligands by exponential enrichment; co-IP, coimmunoprecipitation; HTS, high through-put sequencing; CLIP, cross-linking and immunoprecipitation; TEX, 5' phosphate dependent terminator endonuclease; qRT-PCR, quantitative reverse transcription polymerase chain reaction; MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight; LC-MS, liquid chromatography-mass spectrometry; EMSA, electrophoretic mobility gel shift assay; TBE, Tris/Borate/EDTA; RNP, ribonucleoprotein; SPR, surface plasmon resonance; ITC, isothermal titration calorimetry; SHAPE, selective 2'-hydroxyl acylation analyzed by primer extension; NMIA, N-methylisatoic anhydride; RT, reverse transcription; NAI, 2-methylnicotinic acid imidizolide; MEME, multiple em for motif elicitation; MAST, motif alignment and search tool; CLR, context likelihood of relatedness; 5'RACE, 5' rapid amplification of cDNA ends

the environment of the bacterium [1]. Often, *trans*-sRNAs act to positively or negatively regulate the translation of target mRNAs by freeing or blocking the ribosome binding site or targeting a message for degradation [6]. Interactions that occur between a *trans*-sRNA and its targets often require the RNA binding protein Hfq [1]. Hfq facilitates these interactions by stabilizing RNA-RNA duplex formation, aiding in structural rearrangements, increasing the rate of structural opening or by increasing the rate of annealing (Figure 1A) [7–10].

Hfq is widely conserved in bacteria and about half of all gram-positive and gram-negative bacteria express it [11, 12]. In the case of *hfq* mutant or deletion strains, regulatory effects of sRNAs fail to occur even though the sRNAs are transcribed in response to environmental cues. Phenotypes of these mutants typically include: slowed growth rates, increased cell size, and increased sensitivity to stress [13–15]. Hfq has also been recognized as a virulence factor in many bacteria including *Vibrio cholerae*, and *Salmonella typhimurium* where *hfq* deletion strains fail to colonize, regulate motility or regulate outer membrane protein expression [12, 16, 17].

Hfq forms a donut shaped homohexamer and has been shown to have two well characterized RNA binding sites (Figure 1). In *E. coli*, sequences that are A/U rich and typical of sRNAs bind to the proximal surface of Hfq while A rich sequences typical of mRNAs bind to the distal surface [18–21]. The proximal site was first characterized by a crystal structure of *S. aureus* Hfq bound to AU₅G RNA and showed that the RNA wrapped itself around the central pore of the protein in a circular manner (Figure 1B). Biochemical analyses later showed that Hfq binds to short A/U rich stretches that are preceded or followed by a stem-loop structure sometimes found in a central location of the RNA and more recently at the rho-independent terminator [22–24]. The binding of A-rich sequences to the distal face was first defined by a series of mutations to the distal face that led to disruption of polyA binding [19]. Some years later the specificity for the distal face interaction was further elucidated in a study of the interaction of Hfq with the mRNA *rpoS* as being an AAYAA motif, where Y is a C or a U [25]. These results were further verified by investigation of the interactions of Hfq with two more mRNAs, *fhfA* and *glmS*, genomic SELEX, as well as a crystal structure of *E. coli* Hfq bound to polyA RNA (Figure 1B). The motif has been described as AAYAA, (ARE)_x, and most recently as (ARN)_x [20, 25, 26]. The nomenclature for this motif has evolved to (ARN)_x as the binding site was found to be less specific than AAYAA and the acronym ARE was already in use to describe A/U-Rich Elements in eukaryotic mRNAs [20, 27]. Hfq binds to sRNAs and mRNAs with similar affinity. In vitro, the order of binding does not appear to matter with respect to formation of tertiary complexes of duplex annealing (Figure 1A). Crystal structures of Hfq from other organisms reveal some species specific RNA interactions. While *S. aureus* Hfq binds A/U rich sequences in common with *E. coli* it has recently been shown that the distal site binds an (RL) motif, similar to *B. subtilis*, and in contrast with the *E. coli* (ARN)_x motif [18, 20, 28]. The RL motif is a two nucleotide repeat where R is purine specific and L is a non-specific linker. A signature sequence difference in RL binders is a phenylalanine at position 30 [18]. This residue makes the R site selective for adenosine by creating a pocket and providing stacking stabilization. While the biological consequences of these differences in distal site binding specificity is not fully understood, conservation of key binding residues suggests delineation between gram-positive versus gram-negative organisms and the way sRNA/Hfq-dependent gene regulation is used in these organisms [18, 28]. Crystal structures and binding studies of two Hfq proteins from cyanobacteria suggest that the proximal site binding of these proteins is not specific for A/U rich RNAs as seen in other bacterial Hfqs. In addition to the well characterized proximal and distal surfaces the lateral surface and the C-terminal extension have also been shown to bind to RNA [29–31]. It has been proposed that the lateral surface binds to polyU tracts located in the body of an sRNA while the polyU tract at the 3' end of an sRNA anchors the transcript to the proximal face of Hfq [30]. The role of the C-terminal

domain in RNA binding remains murky but structural and biochemical studies suggest that it may bind to longer RNA molecules and/or increase interaction specificity by recognizing additional motifs within an RNA [29, 31].

Identification of Hfq binding RNAs, characterization of their structure and interactions with Hfq as well as unraveling their functions is fundamental to gaining an understanding of this complex regulatory network in bacteria. The complexity of the sRNA regulation has gradually come into focus over the last decade and now it is clear that this network is indeed vast. The ability of one sRNA to regulate multiple mRNAs and one mRNA to be regulated by multiple sRNAs as well as the regulation of mRNAs that serve as transcriptional regulators themselves add to the complexity of the network [32–34]. In *E. coli* and *S. typhimurium* ~30–35 Hfq-binding sRNAs have been discovered and approximately ~25% of all *S. typhimurium* mRNAs bind Hfq *in vivo*, making the number of potential RNA binding partners for Hfq in the cell very large [35–38]. Thus despite high levels of Hfq expression, it is believed that the availability of Hfq is often limiting in the cell [39, 40]. There is also evidence that Hfq and/or Hfq-RNA complexes may also engage in protein-protein interactions with RNaseE, PNPase, poly(A) polymerase, RNA polymerase, the degradosome and the ribosome which provides mechanistic insight but also implies additional complex biological roles [41–45]. While Hfq is abundant in the cell observations tell us that it is a limiting factor which is not surprising given the plethora of RNA and protein binding partners possible for Hfq [39, 40, 46, 47]. Still, Hfq is able to coordinate a rapid cellular response to stress, in 1–2 minutes [48–50]. How is Hfq able to successfully perform this job? While several plausible theories based on current evidence exist, many of which have been recently reviewed [51], it is critical to continue studying Hfq-RNA interactions at three different levels: discovery, biophysical characterization, and functional analysis. Finally, since so much of our understanding comes from a small set of organisms it is important to branch out into other bacterial species to increase our understanding of this complex and fascinating regulatory network.

The goal of this review is to provide a brief overview of some of the key techniques used to investigate and characterize Hfq-RNA interactions and to provide the reader with insight into the strengths of various methods and how they should optimally be applied. We have structured the article as if the reader were new to the field of Hfq-associated regulatory RNAs and needed to know what the fundamental questions are and how to go about answering them. In Section 2 the identification of binding partners will be discussed. The main question here is to whom does Hfq bind? There will also be insight into the function of the interaction. Section 3 focuses on the biophysical nature of Hfq-RNA interactions. Where do RNAs bind on Hfq and where does Hfq bind on RNAs? What is the effect of Hfq binding on RNA secondary structure and duplex formation? What are the relative contributions of thermodynamics versus kinetics in Hfq-RNA interactions? The last section focuses on questions surrounding the function of Hfq-RNA binding. What are the biologically relevant outcomes of Hfq-RNA interactions and how do they impact the fitness and virulence of bacteria?

2. Identification of binding partners

The first step in studying Hfq-RNA interactions and gaining insight into their regulatory outcomes is to identify the binding partners. Strong binding between Hfq and its sRNA or mRNA partners and the effects of Hfq on transcript and protein levels can be used to identify novel sRNAs and their targets. Three main methods will be discussed; co-immunoprecipitation of RNAs with Hfq, proteomics and transcriptomics in *hfq* knockout strains, and SELEX.

2.1 Co-immunoprecipitation

Hfq co-immunoprecipitation (co-IP) is one of the most common methods used to identify Hfq binding RNAs. The co-IP step can be performed by isolating Hfq bound transcripts using an Hfq specific antibody, an epitope tagged Hfq, or by incubating cellular extracts or purified RNA pools with an affinity tagged Hfq. Once the binding partners have been isolated there are several methods for determining which RNAs have been pulled down. Early work used microarrays, shot gun cloning, and enzymatic sequencing. More recently, the advent of inexpensive high-throughput sequencing (HTS) has altered the experimental landscape and is now the most common approach to deconvolute the pull-down [36, 38, 52, 53]. One of the best features of co-IP is the ability to directly identify Hfq-RNA interactions in a high-throughput fashion, but some limitations occur due to the potential for non-specific interactions. Another drawback is that the lengthy protocol can result in degradation of large mRNA transcripts.

A critical decision that the researcher has to make concerns the growth conditions of the bacteria. It is important because many transcripts are short lived or only expressed under specific growth conditions and thus may go undetected in one experiment while being highly abundant in another. In order to detect as many transcripts as possible several different conditions should be selected. Some researchers may wish to select a stress condition of particular importance in a pathogen of interest, or a growth phase that is known to exhibit significant expression changes in the absence of Hfq. Whatever the conditions, it is critical to recognize that it is most likely that many Hfq binding RNAs may not be present and thus go unnoticed.

Incorporation of a polyacrylamide gel electrophoresis size fractionation step is another key decision. It depends on whether the goal is to find sRNAs only or to also capture mRNA targets. The feasibility of sequencing a large number of isolated transcripts also plays into this equation. Size fractionation is helpful to enrich for sRNAs as well as to limit the size of the library that requires sequencing. The affordability of HTS makes the latter concern less relevant than in the past. It is ideal to use HTS without a size fractionation step so that both Hfq binding sRNAs and mRNAs are discovered simultaneously.

The choice to use an Hfq specific antibody or an affinity/epitope-tagged Hfq protein for the RNA pull down should be made with the following considerations. An Hfq specific antibody is available for *E. coli* but to use this technique in interesting pathogens, either the *E. coli* antibody must cross react with that organism's Hfq or a new antibody must be developed [54]. Sonnleitner *et al.* and Christiansen *et al.* have successfully developed antibodies in *P. aeruginosa* and *L. monocytogenes* for this purpose [52, 55]. The other option is to use an affinity or epitope-tagged Hfq which provides an excellent opportunity to perform this experiment without first preparing an antibody. Ramos *et al.* took advantage of the affinity tag method and discovered 24 novel sRNA in *B. cenocepacia* [56]. They purified a His-tagged Hfq protein that was subsequently incubated with an isolated RNA pool followed by capture of the Hfq-RNA complexes using Ni-NTA agarose magnetic beads. An epitope-tagged Hfq system was developed in Sittka *et al.* in which they created a chromosomal FLAG-tagged Hfq protein (Figure 2A) [36]. To obtain the Hfq bound RNAs they incubated a FLAG antibody with cell lysates and separated the bound from unbound RNAs using protein-A sepharose beads. One thing to keep in mind when using an epitope tag is that its presence may perturb RNA binding and therefore bias the results. The Hfq antibody or the FLAG-tag antibody detection directly from cell lysates provides the benefit of detecting transcripts that were bound *in vivo*.

Once the Hfq bound RNAs have been isolated they can be identified by microarrays, shot gun cloning, enzymatic sequencing, or high-throughput sequencing. A pioneering study used

direct detection of bound RNAs on genomic microarrays to detect 20 novel sRNAs as well as a number of mRNAs that interact with Hfq [38]. The sensitivity of this method was unparalleled at the time but required the use of an antibody specific for RNA:DNA hybrids as well as a species specific high density microarray. These features limit its use in other bacteria of interest. Co-IP has also been used in combination with enzymatic RNA sequencing and shotgun-cloning (RNomics [53]) to identify novel sRNAs in *L. monocytogenes* and *P. aeruginosa* respectively. The use of enzymatic sequencing was a success because it identified Hfq binding sRNAs in *L. monocytogenes* for the first time, but it required large amounts of RNA and time consuming sequencing gels making it unsuitable for large scale analyses. Similarly, shotgun cloning was able to identify new sRNAs on a small scale but the lengthy cloning and use of capillary electrophoresis make it sub-optimal for high-throughput investigations. That being said these approaches are successful and make use of techniques that are readily available and relatively low cost.

The advent and recent affordability of HTS has likely made it the ideal choice for identification of Hfq bound transcripts from co-IP. This method obtains sequence information for a large number of RNAs at one time making it more feasible to identify both mRNAs and sRNA in a wide variety of growth conditions. It does not have species specific requirements so it can be used regardless of sample origin. Also, the alignment of the cDNA clusters can often determine the 3' and 5' ends of sRNAs. This method was implemented by Sittka *et al.* in combination with FLAG-tag Hfq co-IP to identify 1,253 mRNAs that were bound to Hfq as well as large number of sRNAs [36]. However, this method as well as any other protocol involving cDNA synthesis may have a bias against sRNAs because of the restricted capability of reverse transcriptase to process through highly stable structures [57].

Classification of an Hfq bound RNA as an sRNA or an mRNA is the final critical step in the discovery process. Once the transcript has been identified and its location mapped to the genome several criteria can be used to make the determination. mRNAs are often already annotated in the genomes of sequenced bacteria so assignment as an mRNA is relatively simple. If the species is not annotated, one can look for the classic characteristics of an open reading frame (ORF), including; conserved regulatory sequences, a ribosome binding site, and reduced conservation of the third nucleotide of codons. For sRNAs there are no hard and fast rules for required features. One seemingly tried and true predictor of an sRNA is an orphan rho-independent terminator and many searches incorporate this criterion [52, 53, 58]. Although, it should be noted that there are examples in multiple bacterial species of regulatory RNAs that also code for short peptides, recently reviewed by Vanderpool *et al.* [59]. A length component is often incorporated as well. This criterion can be implemented during gel fractionation or when scoring sequencing results and commonly enforces a general size range of ~ 50–500 nucleotides [38, 55]. Genomic location is also typically considered because historically, sRNAs have been found to be transcribed as standalone transcripts in intergenic regions [36, 38]. This requirement is a good place to start in a novel organism, but the results will not be comprehensive. It has been observed that sRNAs can be derived from the 3' ends of transcripts and from genes coding for tRNAs [35, 58]. So, for an exhaustive search, one should not only look in intergenic regions. Conservation of sRNA candidates among closely related species can also be taken into consideration but can become difficult as the sequences rapidly become disparate. Often, sRNAs involved in metabolic processes will be well conserved among related species but sRNAs found in pathogenicity islands or in cryptic prophages are species specific [58]. Given that most of these rules apply to some but not all sRNAs it is advisable to combine them in a way that can help identify the sRNAs but will not be exclusionary to certain types.

While cross-linking has not been used to pull down Hfq associated RNAs thus far we would be remiss not to mention the cross-linking and immunoprecipitation (CLIP) assay due to its

success in identifying RNA-protein interactions [60–63]. This method uses UV cross-linking to create covalent bonds between RNA and protein that are in direct contact with one another. Cross-linking at 254 nm occurs due to the natural photoreactivity of nucleic acids as well as some amino acids at this wavelength. A typical CLIP experiment involves *in vivo* cross-linking followed by lysis and partial digestion of RNA. The RNP complexes are purified by co-IP to select for the protein of interest and the bound RNA fragments are identified using high-throughput sequencing. This technique identifies bound RNA and also provides information on the location of the interaction between the binding partners. Chi *et al.* used CLIP to identify miRNA and mRNA binding partners of Argonaute in the mouse brain [60]. Cross-linking provides advantages over co-IP alone. First, cross-linking directly reflects RNA-protein interactions *in vivo* because the bonds are formed in whole cells rather than lysates or purified RNA pools. This excludes the formation of unnatural complexes due to the lack of cellular components that result in the detection of biologically irrelevant interactions. Second, the RNAs obtained more accurately reflect direct targets because RNAs bound to a protein that associates with the bait protein are not pulled down. A disadvantage of CLIP, especially for potential use in the Hfq system, is the low cross-linking efficiency of purine bases. This caveat may limit the identification of mRNA binding partners. The success that cross-linking has had in the miRNA field makes the CLIP assay a logical candidate for use in the discovery of Hfq associated sRNAs and mRNAs as well as identification of Hfq binding motifs.

2.2 Transcriptomics and Proteomics

Transcriptomics and proteomics provide information on the effect of Hfq on transcription and translation which can lead to the identification of Hfq binding partners as well as insights into function. These methods do not provide evidence for a direct interaction between Hfq and RNA nor can it distinguish between primary and secondary effects, so interpretation must be performed with care. In addition, some changes in transcript/protein levels may only occur during specific conditions or may be too small to detect, so there is potential to miss or overlook important regulatory events.

Transcriptomics in wild type and *hfq* deletion strains can lead to detection of Hfq binding sRNAs and mRNAs. In these analyses wt and *hfq* cells are often grown under various conditions, followed by isolation of total RNA and detection using microarrays or high-throughput sequencing. This method identifies RNAs whose transcription or decay is significantly affected (directly or indirectly) by Hfq [36, 64, 65]. A direct effect occurs when Hfq acts on the transcription rate or decay rate through physical contact with the gene or mRNA. An indirect effect occurs when a change in transcript level occurs due the action of Hfq on some other DNA, RNA, or protein. Transcriptomic analysis only, cannot distinguish these mechanisms, so it is often coupled with another technique like Hfq co-immunoprecipitation [36]. The growth conditions can also be manipulated to disfavor the effects of transcriptional regulators that are known to be connected to Hfq [66]. Mapping the affected transcripts to the genome, identifies the genes and their functions, if annotated, can be suggested. For example, transcriptome profiling of *S. enterica*, *S. maltophilia*, and *Y. pestis* found that Hfq affects the levels of genes associated with stress response and virulence [36, 64, 65]. Published studies have demonstrated the effectiveness of microarrays to detect transcript levels but they require a high density oligonucleotide array to be available for the bacteria of interest [65]. Roscetto *et al.* have taken advantage of the increased affordability of HTS *in lieu* of microarrays to identify mRNAs that show changes in transcript levels due to Hfq, to predict novel sRNAs and to annotate transcription start sites in *S. maltophilia* [65]. They sequenced RNA from wild type and *hfq* mutants as well as in the presence and absence of 5' phosphate dependent terminator endonuclease (TEX). Comparison of RNAs from wild type and mutant strains scored changes in transcript levels

caused by Hfq, while the TEX treatment allowed them to annotate transcription start sites and identify potential sRNAs. Northern analysis and qRT-PCR can be used to validate observed changes in transcript abundance although it has been noted that the abundance levels measured by qRT-PCR are lower than those obtained in microarray results [64].

Proteomics can be used to characterize global control of gene expression at the post-transcriptional level by monitoring which proteins show significant expression changes in the presence and absence of Hfq. Examining protein levels can identify targets that are regulated translationally and would have been missed by transcriptome analysis. This approach often uses 2-D gel electrophoresis to identify proteins with differential expression but the technique resolves only a fraction of total protein, so proteins with low abundance, low solubility, or that co-migrate with another species may not be detected. These studies have been done with the intent to find Hfq-sRNA targets rather than both sRNAs and mRNAs. Using MALDI-TOF mass spectrometry, Barra-Bily *et al.* were able to identify a set of 55 proteins with expression differences in an *hfq* mutant in *S. meliloti* [67]. Proteomics alone cannot distinguish between transcriptional and post-transcriptional/translational regulation, but a sample-matched procedure that combines transcriptomics and proteomics can resolve this problem. This method was used in *S. Typhimurium* by using part of a culture for RNA isolation and microarray analysis and the other part for proteomics analysis using LC-MS [13]. Ansong *et al.* compared the change in transcript levels with the change in protein content to distinguish between the two types of regulation. They found that the majority of effects in *hfq* mutant strains were due to post-transcriptional events. Proteins from their results were validated by western blot analysis and agreed with previously published results. Another benefit of using simultaneous transcriptomics and/or proteomics approach is that no tagging or isolation of Hfq was required.

2.3 Genomic SELEX

A significant problem that plagues all of the techniques described above is that they require the RNA to be transcribed at detectable levels under the selected condition. While the use of HTS has made it easier to obtain data from multiple growth conditions, it is unreasonable to expect a researcher to assay all possible growth or stress conditions under which an sRNA could be expressed. A complimentary approach to discover Hfq binding RNAs avoids the issue of growth dependent expression by screening a genomic library for Hfq binding RNAs in a systematic evolution of ligands by exponential enrichment (SELEX) experiment (Figure 2B) [68].

The uncoupling of RNA detection and growth conditions occurs by creating a genomic library via random priming of all endogenous DNA sequences behind a T7 promoter. Transcription of the library yields a pool of RNAs that represents the entire genome of the bacteria; therefore, all possible transcripts are present regardless of growth condition. The caveat however, is that the RNAs start and stop at random genomic positions and do not correlate with actual transcription start sites or termination sites. Once the RNA pool is created, Hfq is added and allowed to bind. The bound and unbound transcripts are separated using filter binding and the enriched transcripts are reverse transcribed and amplified. The PCR products obtained are transcribed into a new RNA pool and the selection process is repeated. In their protocol, the whole cycle was repeated 9–10 times at which point the Hfq binding RNAs were sequenced and mapped to the genome. They also verified that the aptamers that were obtained bound to Hfq in a cellular environment by employing a yeast three hybrid assay. This technique was able to recover many known Hfq binding sRNAs and mRNAs although it missed some of the most well know and prolific species. This oversight may be due to a lower affinity of these for Hfq than the selected aptamers or to reverse transcription stops as a result of their highly structured nature. It is also possible that some transcripts were overlooked because they were misfolded or amplified in a manner that

altered Hfq affinity. A notable result from this study was the large number of aptamers that corresponded to the antisense strand of protein coding genes. This differs from the focus on *trans*-sRNAs as Hfq binders. The location of these *cis*-antisense transcripts near start codons and intervening sequences between genes in operons suggests a potential role for Hfq in translation regulation, gene processing and expression within polycistronic messages.

3. Folding and Interactions

Once the question of to whom Hfq binds has been answered, one may begin to consider the nature of the interactions. A large amount of biochemical and crystallographic data are now available to support the RNA binding surfaces on *E. coli* Hfq. It is generally accepted that A/U rich elements typical of sRNAs bind to the proximal surface and that (ARN) tracts typical of mRNAs bind to the distal surface [19–22, 25]. Existing evidence also supports a role for the lateral surface in binding U-rich sequences found in the body of sRNAs and for the C-terminal extension in binding longer RNA fragments [29–31]. Crystal structures in other organism including *S. aureus* and *B. subtilis* have shown that species specific Hfq-RNA binding occurs [18, 28]. With the discovery of sRNAs and Hfq in pathogenic bacteria as well as their link to virulence, the need to characterize the specificity of binding and the binding surfaces of these Hfq homologs is of particular interest. Crystallographic data provide tremendous insight into these questions but this review will focus on biochemical and biophysical techniques that are readily available to a wide variety of labs.

Another question is where does Hfq bind on mRNAs and sRNAs? This question is more difficult because binding sites that have been characterized often have unique features based on the specific RNA studied. This heterogeneity has prevented the formation of an exacting definition. A general trend seen in Hfq binding sites on sRNAs is the presence of single stranded A/U rich regions flanked on one or both sides by a stable stem loop structure [22, 69–72]. These motifs have been found in the body of the RNA as well as at the very 3' end of the RNA where it is part of the polyU stretch of the rho-independent terminator [22–24]. The importance of Hfq interactions with mRNAs did not become apparent until recently, so these sites have just started to be defined. However, several well-studied examples provide valuable insight and it has been established that in most bacteria the sequence of the binding site is (ARN)_x and is present in highly structured 5'UTRs of regulated messages [25, 26, 73]. All three validated sites lie to the 5' side and in close proximity to their sRNA binding sites. When the (ARN)_x motif of these messages is mutated it results in decreased ternary complex formation and dysfunctional regulation.

Duplex formation between an sRNA and mRNA is often central to the regulatory outcome desired in response to stress and environment. Hfq serves to aid in duplex formation by remodeling RNA, by increasing the local concentration of the two RNA molecules, or by increasing the rate of structural opening [7–10]. The effect that Hfq has on duplex formation is vital to understanding how a specific regulatory pair functions. Hfq is an RNA chaperone and it has been proposed to remodel RNAs into more favorable structures for duplex formation. This activity has been shown in some instances and not others; therefore, investigating this possibility in an RNA of interest can provide insight into how Hfq promotes duplex formation [22, 54, 74]. Elucidating the relative contribution of thermodynamics and kinetics to the Hfq RNA interactions is also important in understanding how a specific regulatory outcome is achieved. In a cellular environment Hfq is very abundant but it has been shown to be limiting and that RNAs have to compete with each other for binding [39, 40, 46, 47]. The ability of a regulatory pair to affect its outcome is dependent upon its ability to compete for Hfq. This competition is modulated by how tightly and how fast it associates and dissociates from the secondary and ternary complexes with Hfq.

3.1 Electrophoretic Mobility Gel Shift Assay (EMSA)

EMSA is a very common, easy and adaptable assay that can be used to answer a wide variety of questions regarding Hfq RNA interactions. The technique is based on the change in migration of RNA upon binding of a protein. Use of P^{32} labeling and phosphorimaging allows for accurate quantitation. The assay can be used qualitatively to determine whether or not an RNA binds Hfq or quantitatively to allow the determination of thermodynamic and kinetic parameters.

An EMSA should be performed on a native polyacrylamide gel poured at the percentage optimal for migration of the bound complex into the gel and for resolution of the free and bound RNA complex, which is dependent on the size of the RNA. Typical gel percentages are 4–8% and may also contain 3–5% glycerol which can improve complex resolution. The acrylamide:bisacrylamide ratio used is typically 29:1 to accommodate the large size of the Hfq-RNA complex and gels are typically run in 0.5–1X TBE buffer often at 4°C to stabilize the complex during resolution. While the use of EDTA in the running buffer deviates from the conditions used in RNA conformational studies, we have found it is not detrimental and simplifies the experiment by eliminating the need for buffer recirculation and long running times.

It is important to obtain a homogeneously folded RNA population, but due to the complex structure of some sRNA and mRNAs this can be difficult. Multiple folding conditions can be evaluated by changing monovalent salt conditions, magnesium ion concentrations and annealing conditions. Typical monovalent salt concentrations are from 100 mM to 500 mM and magnesium concentrations are from 1 mM to 10 mM. sRNAs with regions of self-complementarity have exhibited the tendency to form homodimers which must be avoided [19, 75]. This tendency can be exacerbated by high magnesium ion concentrations. The RNA should be annealed prior to binding by heating to 75–95°C followed by a period of cooling. The temperatures and durations vary between labs but we have found that 1 minute at 90°C in the absence of magnesium followed by slow cooling at room temperature for 30 minutes works for many RNAs.

Binding specificity can be influenced by the salt concentration as well as addition of competitor RNA. It is well known that Hfq can interact non-specifically with RNAs mainly due to electrostatic interactions between the negatively charged RNA and the overall positive charge of Hfq. This non-specific interaction is stronger at low salt concentration. While non-specific binding is highly dependent on salt, specific RNA-protein interactions are less dependent on ionic strength. This is due to the added stabilization provided by the favorable free energy associated with the specific contacts made. The general outcome is that as salt concentration is increased the interaction becomes more specific and the affinity decreases. This effect has been observed with Hfq as it has been shown that high salt concentrations will decrease the affinity of some sRNAs for Hfq [76]. We have found that the salt conditions used for folding provide a good balance between specific and non-specific interactions. It is common in RNA-protein binding assays to add a competitor RNA to reduce non-specific binding. This addition should be considered carefully in the case of an Hfq-RNA binding reaction, as Hfq has been shown to specifically bind tRNA and poly(A) RNA which may inadvertently alter the measured binding constants [77].

Once the assay conditions are selected, the goal of the experiment should be chosen from several options: the presence or absence of an interaction between the RNA and Hfq can be determined, the effect of Hfq on duplex formation can be assessed or thermodynamic and kinetic parameters can be obtained. If thermodynamics is the focus, equilibrium dissociation constants (K_d) can be determined by titrating an RNA with increasing amounts of Hfq so that a range of free and bound complexes is present. The concentration of P^{32} labeled RNA

in this reaction should be trace, at least 10–100 fold below the K_d . The Hfq concentrations should cover two orders of magnitude above and below the K_d and should maximize the number of data points in the binding transition region. Trace conditions simplify the K_d calculation by allowing one to assume that the free protein concentration changes insignificantly. When determining a K_d the binding reaction must be incubated long enough to achieve equilibrium, typically 5–30 minutes at room temperature. For very tight binders longer incubation may be required due to slow off rates. The binding reactions are then combined with loading buffer containing glycerol or sucrose, loading dyes of choice and then resolved. It should be recognized that loading dyes may affect migration of RNP complexes and can be omitted to avoid problems. A drawback of EMSA is that the gel may need to be run for several hours to adequately resolve the complexes.

To obtain thermodynamic parameters the free and bound bands are quantified from the phosphorimage. The percent bound RNA is determined and then plotted versus the log Hfq concentration. These points are then fit by a nonlinear least-squares analysis to a cooperative binding model (for Hfq the cooperativity values typically fall between 2 and 3). Multiple binding events may occur because one RNA may bind multiple Hfq hexamers. This effect can be observed in the case of Hfq binding DsrA (Figure 3) as well as with other RNAs. This case may be dealt with by using a partition function for two sites or by simplifying the data to consider only the K_1 events in which case all shifted bands are summed to yield a “bound” state. The analysis of the gel shown in Figure 3A demonstrates the two site fitting method based on equations 1–3.

$$f_{DH\ I} = \text{cpm}_{DH\ I} / \text{cmp}_{\text{total}} = ([\text{Hfq}] / K_1)^n / Q_{DH} \quad (1)$$

$$f_{DH\ II} = \text{cpm}_{DH\ II} / \text{cmp}_{\text{total}} = ([\text{Hfq}]^2 / K_1 K_2)^n / Q_{DH} \quad (2)$$

$$Q_{DH} = 1 + ([\text{Hfq}] / K_1)^n + ([\text{Hfq}]^2 / K_1 K_2)^n \quad (3)$$

where $f_{DH\ I}$ and $f_{DH\ II}$ are the fractions of DsrA in complexes D•HI or D•HII, the Hfq concentration is of monomers, K_1 and K_2 are binding constants for the first and second site, n is the Hill coefficient and the function Q_{DH} in (3) is the sum of the terms for each bound state. To obtain a binding constant for each state one simultaneously fits equations (1–3) while if only the first binding constant is desired an equation like (4) can be used.

$$Q_{DH} = [\text{Hfq}]^n / (K_d)^n + [\text{Hfq}]^n \quad (4)$$

The K_d determined in equation (4) should have a value similar to the K_1 value obtained from the partition functions in equations (1–3). Some labs use the dual binding fit whereas others use the single site. The choice of which fitting to use is based on perceptions of physiological relevance. The decision is not straight forward as the topic is still debated. The binding of two Hfq hexamers by one RNA may be an effect only observed *in vitro* due to the trace conditions of RNA and the large concentrations of Hfq. This condition may not exist in the cell because of competition for Hfq. A recent study by the Weichenrieder group particularly calls the biological relevance of multiple Hfq binding into question because they found strong evidence that an sRNA binds both the proximal and lateral surfaces of Hfq [30]. If this is the case it is unlikely that one sRNA could bind multiple Hfq protein except under *in vitro* conditions of trace RNA.

A variation of this technique, called a competition assay, can be used to assess which face of Hfq an RNA is binding as well as its ability to bind compared to other RNAs (Figure 3B). This method is particularly useful for determining binding of RNAs by Hfq homologs whose binding specificities have not yet been determined. This approach uses a preformed Hfq-RNA complex which is then challenged with increasing concentrations of a competitor of interest. The ability of a competitor to promote dissociation of the RNA from the preformed complex can then be assessed by monitoring an increase of the free RNA. To use this assay to gain information about the Hfq binding face specific for an RNA of interest, the preformed complex should contain Hfq and an RNA for which the binding face is known. The ability of a competitor molecule to remove the known RNA from Hfq indicates that they bind the same face and a lack of ability to compete indicates binding on a different area of Hfq. One specific application of this technique was used by Mikulecky *et al.* to determine the RNA binding sites on *E. coli* Hfq (Figure 3B). In that experiment a DsrA-Hfq complex was preformed and unlabeled A₂₇ was added at increasing concentrations. As the amount of A₂₇ increased the formation of DsrA-Hfq-A₂₇ occurred, indicating that Hfq binds DsrA and A₂₇ on different faces and that they act independently.

EMSA can also be used in a straightforward experiment to evaluate the effect that Hfq has on duplex formation of a regulatory RNA pair. The Aiba group studied the duplex formation of SgrS and *ptsG* over time in the presence and absence of Hfq to investigate if Hfq could enhance the rate of duplex formation [78]. To explore the effect of Hfq they added Hfq to the binding reaction and then extracted Hfq with phenol before loading the reaction onto the gel. Before treatment with phenol it is advisable to first digest with proteinaseK to prevent the RNA from transferring to the organic phase along with Hfq. They found that within one minute a significant amount of duplex had formed in the presence of Hfq, suggesting that Hfq strongly enhances the rate of duplex formation. Rapid duplex formation in the presence of Hfq highlights a limitation of EMSA. The time it takes to prepare the samples may exceed the time it takes for the complex to form so one may not be able to quantify fast events, although quench-flow techniques can resolve this issue.

Kinetic parameters can also be determined using EMSA by following binding reactions over a time course. The fraction of each complex can be plotted against time and fit to rate equations. This application was used to demonstrate the effect of Hfq on the rate of DsrA and *rpoS* annealing [25]. In this experiment the two RNAs were monitored over time for the formation of duplex, in the presence and absence of Hfq. Using this technique Soper and Woodson were able to show that Hfq increased the rate of duplex formation ~ 30 fold [25].

The use of EMSA to evaluate the binding of Hfq to truncated RNA constructs has been used to identify the portions of RNA that are necessary for binding. This approach was used to identify the lengths of the 5'UTRs of *fhfA* and *rpoS* required for Hfq binding and sRNA-mRNA duplex formation [25, 26]. In both cases constructs with varying 5'UTRs were made and assayed for their ability to form a duplex. Salim and Feig, as well as Soper and Woodson, were able to determine the relevant 5'UTR length for optimal duplex formation using this approach.

3.2 Filter Binding Assays

Filter binding assays allow for the measurement of both thermodynamic and kinetic properties of Hfq-RNA binding [76]. Unlike EMSA where complexes are separated in a gel matrix this assay employs a double filter to separate the bound from unbound RNAs. The top membrane is nitrocellulose and binds the RNA-protein complexes and the bottom membrane is a charged membrane that binds free RNA. The two membranes are seated in a dot blot apparatus and samples are drawn through by applying a vacuum. Quantitation of the RNA and RNA-protein complexes is performed using phosphorimaging. Some particular

benefits to this assay are the ability run on high-throughput 96 well plates, to manipulate the volume of the reactions to obtain optimal detection, high sensitivity and low cost. This method is particularly useful for determining fast kinetics of binding due the rapid rate of complex separation [76]. One potential drawback however, is that you can no longer resolve multiple binding events; as discussed earlier those events may or may not be relevant in a given study. Equilibrium dissociation constants can be obtained by titrating the RNA with increasing amounts of Hfq and fitting the data to standard binding isotherms. Kinetic parameters can also be determined by keeping the RNA and protein concentrations constant and varying time. Control experiments in the absence of protein should be performed to account for non-specific nucleic acid binding to the nitrocellulose membrane. This technique was implemented to investigate and compare the binding properties of nine different sRNAs [76]. Olejniczak found that the sRNAs had similar affinities for Hfq but varied in their ability to compete for Hfq binding. The binding properties determined using the filter binding assay agreed with those obtained using other methods under the same conditions.

3.3 Surface Plasmon Resonance

Surface plasmon resonance (SPR) has been used to study the thermodynamics and kinetics of Hfq binding to both sRNAs and mRNAs. SPR monitors changes in the refractive index near the surface of a sensor that occur due to binding events. One strength of this technique is the simultaneous, real time measurements of both kinetic and thermodynamic parameters. In this technique, one binding partner is immobilized on the sensor surface and the other is continuously flowed in. When a binding event occurs, the refractive index increases and when the complex dissociates, the refractive index decreases. The results are plotted as response units versus time and are most commonly fit to a simple 1:1 Langmuir binding model to obtain k_{on} and k_{off} values. The K_d can then be calculated by dividing the k_{off} by k_{on} .

There are several steps that must be taken in order to execute a successful SPR experiment investigating an Hfq-RNA interaction. Though theoretically it shouldn't matter, it is most typical to immobilize the RNA on the surface of the sensor and flow in Hfq. The larger size of Hfq provides a greater change in response when the two molecules interact [79]. Also, the negative character of the chip surface can repel the RNA if it is chosen as the analyte [80]. For a high affinity interaction like that of Hfq with an RNA, the RNA should not be immobilized at too high of a concentration or problems associated with mass transfer could arise [79]. We have found that ~ 3 fmol works well in the case of *fhfA*. To prepare the RNAs for SPR, they are biotinylated at the 5' end and purified using a spin column. It is critical that the samples are very pure as the presence of contaminants could affect the SPR signal or interact with the analyte and impact binding. A benefit of SPR is that it is a label free approach but it does require immobilization which could lead to changes in binding. Unfortunately, this technique is not suitable for high-throughput analysis as only a few samples can be analyzed at a time and each analysis requires 5–15 minutes.

This approach has been used to analyze the kinetics and thermodynamics of Hfq binding to the mRNAs *fhfA* and *ompA* and to the sRNA MicA [7, 26]. The Wagner group used SPR in addition to EMSA and filter binding to obtain K_d values, association and dissociation rates for *ompA* and MicA [7]. In both cases the values obtained were similar between the three techniques demonstrating the value of each in obtaining reliable data. In the study of *fhfA*-Hfq, SPR was used to demonstrate that the (ARN)_x motif is important for distal site interaction and to support a wrap-around model for *fhfA* binding. This model suggests that the RNA binds to both surfaces of Hfq at once [26]. To investigate the importance of the (ARN)_x site contact with Hfq, the ability of constructs with or without the site to interact with Hfq were compared. It had previously been shown that *fhfA* interacts with both Hfq surfaces, so the data was fit to a parallel binding model where *fhfA* can interact with either

side of Hfq independently before forming the complex where both sites are bound. The step where both sites are bound was omitted from the fitting because the technique cannot register that type of unimolecular rearrangement. The inability of SPR to detect internal rearrangements of this type is its shortcoming. The *fhfA* construct that contained both the proximal and distal site had two low nanomolar K_{ds} whereas the construct with only the proximal binding site had only one, indicating that the $(ARN)_x$ site is important for distal surface binding. Salim and Feig also performed a competition experiment by pre-binding Hfq to *fhfA* and then flowing in either A₁₈, DsrA, or both. All three scenarios led to faster than direct dissociation rates which suggests that *fhfA* binds in a wrap-around fashion. These experiments highlight the use of SPR to obtain information beyond thermodynamic and kinetic parameters.

3.4 Other Biophysical Methods

Isothermal titration calorimetry (ITC) is most widely recognized in studying DNA-Protein interactions and protein biophysics but has also been successfully used to obtain thermodynamic information and binding stoichiometry of an RNA and protein interaction [19]. ITC directly measures the heat released or absorbed during a chemical reaction by monitoring the power consumption required to keep a sample cell and a control cell at the same temperature over the course of a reaction [81]. Direct measurement allows for a more accurate determination of thermodynamic data than a gel shift. Some issues that have limited its usefulness are the need for a large sample size as well as the inability to deconvolute the energy from the multiple binding interactions or from structural rearrangements.

Fluorescence anisotropy measures the change in polarized light emitted from a fluorophore in solution during a binding event [82]. This change is a result of decreased tumbling of the labeled molecule upon binding of a larger molecule. This phenomenon allows for the evaluation of a molecule's binding properties by providing a direct measure of the bound to free ligand ratio. Fluorescence is a safer option than radiolabeling but it is less sensitive and much larger which may affect binding. A benefit of this approach is that it is solution based which omits a separation step and therefore may more accurately reflect true equilibrium binding. This approach can be applied to Hfq-RNA systems by labeling the RNA molecule with 6-carboxyfluorescein, titrating it with increasing amounts of Hfq and observing the change in anisotropy [83–86]. The data are plotted as anisotropy versus time and fit by a nonlinear least-squares analysis to a two step binding model.

Fluorescence anisotropy was used to investigate the RNA binding surfaces on Hfq in a similar fashion as EMSA. Sun and Wartell assessed a variety of Hfq mutants followed by binding studies with RNA substrates [84]. In agreement with previous studies, they found that DsrA binds to the proximal surface and that A₁₈ binds to the distal surface of Hfq. They also used the fluorescence anisotropy data to determine reaction stoichiometries which led to some ambiguities regarding the binding ratio of the A₁₈ Hfq interaction. Uncertainty in the amount of Hfq required to saturate binding of the labeled RNA caused an underestimation in the amount of bound RNA, leading to a discrepancy with ITC data. This incongruity was later resolved by allowing flexibility in the variable that accounts for the fraction of bound RNA [85]. Determining an accurate binding model of other than two state reactions can be challenging using fluorescence anisotropy if the anisotropy change between the two states is not well defined and/or if there is cooperative binding. EMSA is better suited for this because of the added information provided from visualization of discrete bands that represent different complexes. This observation can provide the binding stoichiometry and guide the correct selection of a binding model.

3.5 Chemical and Enzymatic RNA Modification

The use of chemical and enzymatic analysis of RNAs can be employed to determine the secondary structure of an RNA, the Hfq footprint and structural changes upon Hfq binding. Additionally, some techniques allow structure determination and protein interaction mapping *in vivo*. One approach uses a complementary set of enzymatic and/or chemical modifications that react with the nucleotides in different ways to provide a complete assessment of each nucleotides environment. To determine Hfq binding sites on the RNAs, the probes can be used in the presence and absence of Hfq. In the presence of Hfq some nucleotides will become protected, indicative of a binding site. In addition to seeing protection, some nucleotides may become more reactive, indicative of secondary structure rearrangements. Two different methods can detect the cleavages or modifications. One route uses reverse transcription with an end labeled primer to detect both scissions and modification for RNAs of any length (by using multiple primers). The other technique uses end labeled RNA for direct detection but can only be used for shorter molecules, typically less than 300 nucleotides in length. The fragments obtained from these methods are then separated on denaturing polyacrylamide gels along with one or two ladders that assign the site of cleavage/modification on the RNA. Efforts to obtain a uniformly folded RNA as well as the selection of binding buffer conditions should be taken as discussed for EMSA.

One illustrative example of the use of nucleases was the determination of the effect of Hfq on the *sodB* mRNA and its regulatory partner, RyhB sRNA [87]. Geissmann *et al.* used a combination of RNaseA which cleaves 3' to single stranded cytosines and uracils, RNase T1 which is specific for single stranded guanines, RNase I which cleaves any single stranded residue and RnaseV1 that is specific for double stranded regions and provides positive evidence for helical regions. This probe combination allows for sufficient coverage of the RNA to provide an accurate secondary structure. RNases are large and therefore show signs of steric hindrance and care must be taken optimize enzyme concentration and incubation time as the presence of secondary cleavage events can lead to misinterpretation of the data. The data obtained allowed for the accurate determination of secondary structures of the two RNAs as well as footprints pinpointing the Hfq binding site(s). Also, the occurrence of enhanced cleavage at certain residues in the presence of Hfq can show a loop opening event or other rearrangements such as in the case of Hfq binding to *sodB* mRNA.

Another useful probe is the Tb^{3+} or Pb^{2+} ion which cleaves single stranded RNA in a sequence independent manner. The small size of these ions avoids the steric hindrance issues that RNases have, which allows for detection of subtle structural changes upon Hfq binding. The Masse lab used this method to detect an enhanced interaction between the regulatory pair, RyhB and *iscS*, in the presence of Hfq and the Hfq binding site on the *iscS* mRNA [88]. The concentration of the ion must be optimized to obtain conditions where less than one cleavage occurs per RNA molecule. The reaction can be quenched at the optimized time with addition of EDTA. Lead(II) has also been used to determine secondary structures *in vivo* and could potentially be used to map Hfq interaction sites *in vivo* in the future [89].

Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) is a chemical modification based technique that takes advantage of the ability of the hydroxyl selective electrophile, N-methylisatoic anhydride (NMIA), to react without sequence specificity with more flexible/accessible nucleotides [90]. The use of SHAPE to determine secondary structures and footprinting provides the advantage of only having to use one chemical modification technique to obtain necessary structural information. Modifications are revealed by reverse transcription and resolution on denaturing gels or by capillary electrophoresis. Capillary electrophoresis analysis allows for a significant increase in throughput and software to analyze the raw data and obtain reactivities for each nucleotide is available [91, 92]. Our lab has successfully used SHAPE in combination with capillary

electrophoresis to determine the secondary structures and Hfq footprinting of *glmS* and *fhfA* mRNAs [26, 73].

Several considerations are important to successfully implement SHAPE to study RNA-Hfq interactions including: RNA design, RNA folding, RNA modification and primer extension conditions. To detect adduct formation reverse transcription (RT) is used. RT can lead to a loss of information due to pausing at the 3' end during the initiation phase and at the 5' end because of an intense band equivalent to the full length extension product. To avoid this loss of information, the RNA can be inserted into a structured cassette, first described by the Weeks lab, where the RNA is flanked by highly structured hairpins and also an RT primer binding site on the 3' end [90]. While the cassette improves this problem it may still interfere with mapping Hfq binding sites at the polyU tract of the rho-independent terminator since the 3' end is unnatural in these constructs. The stability of the hairpins ensures that the cassette structure does not interfere with the folding of the RNA of interest. To facilitate analysis of many RNAs we have created a modified pUC19 vector containing the cassette behind a T7 promoter so that any RNA of interest can be cloned into the vector and transcribed. The RNA must be renatured prior to modification as described in the previous techniques.

To modify the RNA, NMIA is added at a concentration and time that must be optimized to obtain only one adduct formation per molecule. NMIA +/- reactions are run in parallel so that natural RT stops can be accounted for in the data analysis. In order to obtain footprinting data Hfq +/- reactions can be run as well. Hfq is added to a final concentration of 1 μ M hexamer and allowed to bind at room temperature for 30 minutes before reaction with NMIA. After NMIA reaction the RNA is ethanol precipitated or in the case of Hfq + reactions it is first proteinaseK digested and then phenol-chloroform extracted. The primer extension reaction is performed using RNA Superscript III in four separate reactions; NMIA +, NMIA -, and two sequencing ladders created by including ddNTPs into the reaction mixture. Each reaction contains an RT primer with a unique fluorophore that allows identification of the different reactions in the capillary electrophoresis readout.

The reaction is then separated by capillary electrophoresis. Reactivities for each nucleotide are determined by analyzing the raw data with ShapeFinder (Figure 4A) [93]. Data for Hfq + reactivities are obtained from a unique set of reactions that can be run in a parallel lane. The reactivities for the Hfq +/- reactions can then be compared to determine where protection has occurred (Figure 4). The resulting data is used to determine the fold of the RNA using RNAstructure and the Hfq protection can be mapped (Figure 4B) [94]. The structures for *fhfA* and *glmS* that we determined using this approach added to the evidence for an important Hfq binding interaction at (ARN)_x sites in the 5'UTRs of regulated messages [26, 73]. This method provides accurate, high-throughput structure determination and footprinting. The cost of fluorophore labeled primers is high but the use of a universal RNA cassette makes it a worthwhile one-time investment.

New developments in SHAPE that describe high-throughput analysis and *in vivo* structure mapping have recently been published [95, 96]. These techniques have not yet been applied to bacterial sRNA systems but hold promise for investigating Hfq-RNA interactions. Lucks *et al.* recently described high-throughput SHAPE analysis that is able to obtain structural information from an *in vitro* pool of RNAs that are distinguished from one another using bar-codes [95]. This method is able to obtain quantitative high resolution structure information for hundreds of RNAs in a single experiment. It is important to study RNA *in vivo* because the biologically relevant structure may exist only in the cellular environment. In addition, RNA-protein interactions are represented in the data. Chang *et al.* designed two new electrophiles, 2-methyl-3-furoic acid imidizolide (FAI) and 2-methylnicotinic acid

imidizolide (NAI), which maintain the selective reactivity to hydroxyl groups but are non-toxic and have a sufficient half life in cells to modify the RNAs. They found that NAI had a higher reactivity and chose to use it to validate their technique by probing the 5S rRNA in mouse embryonic stem cells and in yeast. When SHAPE data was overlaid with the crystal structure, they found that NAI had modified the RNA at the predicted nucleotides in as little as 1 minute. Comparison of *in vitro* and *in vivo* SHAPE structures for the 5S rRNA led to the identification of important contact sites with other RNA and proteins. One can easily imagine this technique being used to map the structures of mRNAs and sRNAs that interact with Hfq and to determine Hfq binding sites *in vivo*. Some potential complexity lies in separating the effect of protein binding and structural changes on the reactivities and declaring the identity of the protein binding partner.

3.6 Isoenergetic Microarray Mapping

Microarray mapping is a unique approach to secondary structure determination, Hfq binding site identification and Hfq derived structure change. The technique is based on the ability of single-stranded RNA regions to hybridize with complementary oligonucleotide probes in contrast to double stranded RNAs [71]. A microarray with probes specific for the RNA of interest is created to match the probe specifications required for the particular target. The structure of the RNA is determined alone and then various complexes can be studied by comparing the hybridization intensity in the presence of other complex components. The incorporation of locked nucleic acid and modified nucleotides are incorporated where necessary to account for varying thermodynamic stabilities of the probes due to the specific sequence. This technique can be used with a broader set of conditions than with chemical and enzymatic assays that often require specific conditions for reactivity. The method is limited by the thermodynamic stability of the target molecule structure and the stability of its interaction with other biomolecules. This approach has been used to determine the structure of DsrA in complex with Hfq and *rpoS*, and OxyS in complex with Hfq and *rpoS* or *fhfA* [97]. Fratzak *et al.* obtained structures for both sRNAs that agreed with previous data and identified previously suggested Hfq binding sites. They were also able to confirm that the DsrA secondary structure is not altered upon Hfq binding and that Hfq facilitates sRNA-mRNA duplex formation. The broad application of this technology has been minimized because of the large amount of effort that must be invested to create a unique microarray for each RNA of interest.

4. Functional Characterization

The number of Hfq dependent sRNAs identified in various bacterial species is large but only a small set have well defined biological functions. Bacterial sRNAs are not easily grouped into categories that indicate their functions and because of this, the function of these sRNA regulators often have to be elucidated on an individual basis. Many of the techniques discussed in Section 2 to identify Hfq binding mRNAs also give some information about function if the gene has been annotated. In addition to those techniques we will present approaches that allow for the identification of the RNA binding partner, given an Hfq associated sRNA or mRNA of interest (Table 1). Binding partner identification is often the first step after an initial discovery technique. After identifying potential RNA partners it is necessary to validate that the interaction is direct and that there is a real biological effect.

4.1 Bioinformatics

Due to the availability of many bacterial genomes, bioinformatics approaches for the discovery and analysis of sRNAs have flourished. There are many ways that computational tools can be employed to help elucidate the functions of Hfq binding sRNAs and mRNAs, specifically by aiding in the prediction of an RNA binding partner of a given sRNA or

mRNA. The most useful aspect of these approaches is the ability to guide lab work to obtain results in a more efficient manner. This guidance saves time and money in the lab. Computational approaches are often not sufficient on their own due to false positives and false negative feedback and therefore must be validated experimentally. In addition, prior information about the system to be studied is necessary to create a useful tool. These tools have been successful in organisms where Hfq binding sRNAs and mRNAs are well characterized and have the potential to be modified easily to accommodate species specific characteristics of the network. Bioinformatics can be a particularly useful tool when studying pathogens or bacteria that are hard to grow and manipulate in the lab. These studies could be facilitated by an initial computational analysis followed by experiments in a model bacterium.

One of the earliest examples of employing bioinformatics to identify a target mRNA was a simple blastn search to identify a 16 nucleotide region of complementarity between MicC and *ompC* [98]. These searches are useful in identifying interactions that have long regions of continuous complementarity which is unfortunately a minority of Hfq-dependent sRNAs. Despite this limitation, Jrgensen *et al.* have very recently used blastn to identify an mRNA target of McaS after a proteomics/transcriptomics approach failed, demonstrating its utility and use as a starting point for RNA binding partner identification [99]. This approach is also useful because it requires no prior knowledge besides the requirement of complementarity between the two RNAs to guide the search. Another relatively simple bioinformatics approach is to look for the presence of a transcription factor binding motif. The transcription of some sRNAs is controlled by transcription factors [49, 100]. By identifying the transcription factor that controls expression of the sRNA the function may be apparent based on the role of the transcription factor. For instance, Papenfort *et al.* were able to identify two E-dependent sRNAs involved in *omp* mRNA regulation using this method [49].

Once a set of targets for a given sRNA have been validated, the knowledge of those interactions can guide a computational search for new mRNA targets [101]. Sharma *et al.* first defined a binding motif for GcvB based on 16 known target binding sites using the MEME (multiple em for motif elicitation) software [102]. By providing the sequences of the known targets the program was able to identify an 8 nucleotide long motif that was present in all but 2 of the mRNAs. To identify previously unknown targets of GcvB the motif was used to search the -70 to +30 regions of all annotated *Salmonella* protein coding genes using a MAST (motif alignment and search tool) [102]. Frequently, sRNAs interact with mRNAs in this region of the 5'UTR, but this parameter should be chosen based on the known targets of the specific sRNA of interest. Widening this criterion may lead to more false positives. The annotated transcription start sites of the genes should also be taken into consideration. If the interaction was found from -60 to -70 but the RNA is transcribed starting at -50 then the putative interaction is likely irrelevant. The genes that showed a significant match to the motif were then input into *TargetRNA* [103] to identify the targets that had the strongest base-pairing with GcvB. Overall they obtained 42 potential mRNAs that passed all of the bioinformatic criteria, 4 of the 5 that they chose to validate showed regulation. This technique successfully identified known and new targets that were missed by a transcriptomics approach and demonstrated the utility of a combining bioinformatics with other experimental approaches. A drawback of the method is that a large amount of previous knowledge is needed to train the computational queries. This limits its use in finding interactions for sRNAs that have few known targets or in organisms where sRNAs are not well characterized.

In addition to designing your own unique search strategy there are many accessible programs that have been designed to allow researchers to easily perform bioinformatics studies without designing their own algorithms. These programs and their detailed methods

have been recently reviewed by Li *et al.* so we will just provide a brief overview of a selection of these here [104]. *TargetRNA* was designed to identify potential mRNA partners given the sequence of an sRNA and the genome of interest. mRNA-sRNA interactions are scored based on the hybridization between the two RNAs without considering intramolecular base pairing or pseudoknots. The omission of the secondary structure of the RNAs is a limitation because the presence of these structures can significantly affect the likelihood of an interaction. The program also provides parameters that can be specified by the user such as seed length and the location of the interaction site relative to the promoter. Overall their approach was able to identify ~ 70% of the RNAs used in the training set. *TargetRNA* was one of the first programs developed to predict sRNA targets in bacteria. It was designed using a limited amount of known information which may make it less useful than some of the newer programs. That being said it has been successfully incorporated into several recent studies [101, 105, 106].

Many other programs have become available to aid in the identification of mRNA-sRNA interactions. sRNATarget was developed by Zhao *et al.* by incorporating 35 positive (validated interactions) and 86 negative targets into its training set. Unlike *TargetRNA* this approach also considers the secondary structure of the RNAs [107]. They were able to obtain a greater accuracy rate for predicting the training set than *TargetRNA*. The program IntaRNA evaluates RNA-RNA interactions using a complex algorithm based on hybridization and accessibility of the target site [108]. This program is effective but is computationally demanding, whereas an alternate server called RNAPredator achieves similar accuracy in less time [109]. The program sTarPicker has been shown to outperform the above methods in target prediction and accuracy of binding site prediction. The key difference in their approach is a two-step hybridization model that first picks targets based on stable seed interactions and then on extended hybridization of the entire binding site [110]. All of these available tools can aid in the discovery of Hfq-associated RNA binding partners when combined with other techniques and therefore contribute to the determination of their biological functions. The current searching methods will continue to evolve as new information about the sRNA regulatory network is learned. Some current insights that may improve predictions are to include requirements for Hfq binding sites in the mRNAs and to focus on known binding motifs for particular sRNAs.

The Collins group took a unique approach to defining the functions of bacterial sRNAs by using a network biology approach that takes advantage of existing microarray data to elucidate the functions of sRNAs [111]. Knowledge of sRNA interactions can often lead to clues about the function of the sRNA. This is the first program to take advantage of the large body of known interactions to make functional predictions for the whole sRNA network [111, 112]. First they applied a Context Likelihood of Relatedness (CLR) algorithm to a compilation of existing microarray expression profiles that were obtained under various conditions [113]. This algorithm identifies regulatory relationships using an inference approach and identified 459 potential targets. They were then able to identify functional enrichment in seven sRNA subnetworks by assigning gene functions to the putative targets. They validated the functional implications of three of these sRNAs. This technique is useful because there are several sRNAs known to regulate multiple mRNAs who all function in a similar physiological process [101, 114]. The identification of that process allows new targets to be inferred based on their involvement in that pathway. This approach was based on microarray data and therefore does not distinguish between direct and indirect interactions. The incorporation of proteomics data would be an improvement. In general, the more data the analysis includes the greater the predictive power. This method can easily be adapted to other organisms with profiling information.

4.2 Manipulation of sRNA Expression

A widely used approach for defining the function of an sRNA is to manipulate its expression. Many variations including over expression, pulse expression and knockouts have been used to identify the mRNA targets of an sRNA or to identify the sRNA regulator of a given mRNA or phenotype. The basic concept behind these experiments is that changing the expression of an sRNA will lead to detectable changes in transcript levels, protein levels or changes in phenotype.

Creating an sRNA over expression strain involves cloning the RNAs into a high copy plasmid behind an inducible or constitutive promoter. It is necessary to place the transcription of the sRNA under control of an alternative promoter because some sRNAs will not be highly expressed under the natural promoter even when present in a high copy number plasmid. The high copy number expression provides for a high level of sRNA transcription which minimizes the effect of any chromosomally derived sRNAs. They should be inserted such that transcription begins at the natural transcription site which if not known can be determined using a technique such as 5'RACE [115]. This approach allows for the study of sRNAs that may be poorly expressed naturally or are toxic to the bacteria. A caveat of sRNA over expression is the potential to cause inadvertent consequences by disrupting the balance of the natural sRNA network and can lead to confusing results.

Given an sRNA of unknown function, a good way to begin characterizing it is to determine the identity of proteins that show changes in expression when it is over expressed. The Wagner group used this approach to identify the regulation of *ompA* by MicA [116]. They observed differences in protein expression from strains with high, normal or low MicA expression using two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). Proteins that showed changes of greater than 2.5 fold between the strains were subsequently identified using MALDI-TOF. The protein, OmpA, showed the greatest change in abundance and was subsequently validated as a MicA target. This method was also used by Frohlich *et al.* to classify SdsR as a regulator of *ompD* [100]. In this case a significant change in OmpD expression was identified simply from a 1D-PAGE analysis due to its characteristic size, and then verified by northern and western blots. Proteome analysis does suffer from the inability to differentiate between direct or indirect effects, and mRNA stability or translational regulation as the mechanism of control.

In organisms where the majority of sRNAs have been discovered, an sRNA over expression library can be created to screen the effects of a large number of sRNAs on a given mRNA or phenotype. The utility of this approach was demonstrated by the identification of an additional sRNA that regulates *rpoS* [117]. An sRNA library with 26 Hfq binding sRNAs was created and co-transformed with an *rpoS-lacZ* fusion. The β -galactosidase output was monitored for significant increases or decreases and led to the identification of four sRNAs previously unrecognized to regulate *rpoS*. By observing the ability of two of the putative sRNAs to act on *rpoS* in strains where the positively acting sRNAs were deleted, they were able to determine that the effects produced in the original screen were indirect. In the deletion strains the down regulation of *rpoS* previously seen no longer occurred. This observation illustrates the need to be aware of effects caused by artificially titrating Hfq from natural sRNAs and target mRNAs, as this can be an issue during sRNA over expression. Mandin and Gottesman went on to further characterize the regulation of *rpoS* by the sRNA ArcZ as *bona fide*. A useful feature of this approach is that once an sRNA library has been created it is easy to rapidly screen any target mRNA of interest by simply cloning it into a fusion vector.

The Gottesman lab further used the sRNA over expression library to identify sRNAs involved in cell motility by inducing the sRNAs and observing their behavior on motility

plates [118]. They identified 8 sRNAs that had a recognizable effect on bacterial motility. Once sRNAs associated with motility were identified they hypothesized which mRNAs would be logical targets based on a relationship to the phenotype studied. They reasoned that the most efficient means of regulating motility would be the genes at the top of the cascade. They tested this hypothesis by examining the effect of the sRNAs on mRNA-lacZ fusions for the genes of interest and found several legitimate regulatory pairs that they further characterized.

sRNA over expression libraries are a useful way to rapidly screen potential targets for direct interactions with sRNAs but they require prior knowledge of the majority of sRNAs in an organism. The Gottesman lab has used an approach that uses a multicopy plasmid library of the whole genome that negates the need to know the sRNAs in an organism [119]. Using this approach they identified sRNA regulators for two genes involved in antibiotic resistance that were found to bind Hfq. In this approach they constructed a *lacZ* fusion of their gene of interest, *dpiB*. Into this strain they transformed a pBR322-based *E. coli* genomic library under the assumption that any genomic element in this library that regulates *dpiB* should produce white (Lac-) colonies on MacConkey Lactose plates containing arabinose [120]. They observed several colonies with this phenotype and sequenced the plasmids that they contained. They found several fragments of protein coding genes as well as two known sRNAs. If the sRNAs are unknown in the organism of study than one could deduce that the fragment may be an sRNA based on the typical characteristics of sRNAs discussed previously. The relevance of the sRNA-mRNA predicted is then further characterized and validated.

Pulse expression is a technique that makes use of an inducible promoter to briefly over express an sRNA in a strain where that sRNA of interest is deleted, followed by transcriptome analysis [48, 101, 121]. By over expressing the sRNA for a short time the differences between direct and indirect effects are more discernible. In this experiment sRNA expression is induced and total RNA is extracted at a specific time point, usually 10–15 min, and analyzed on a microarray. High-throughput sequencing or qPCR could also be used to analyze the RNA pool. The pulse duration should be optimized to successfully distinguish between indirect and direct effects as different systems may work faster than others. For example, it was found that a direct target of RyhB was degraded in 3 minutes and the mRNA of an indirect target followed closely behind at 7 minutes, therefore if analysis was done at 10 minutes you would not be able to distinguish between the two effects [48].

This technique has identified additional targets of both RyhB and GcvB [48, 101]. The GcvB study incorporated an additional feature that allowed identification of mRNAs that bind to a specific motif of the sRNA [101]. Several sRNAs have been shown to possess a conserved mRNA binding motif that binds to multiple targets [70, 72, 122, 123]. By performing pulse expression with a wild type strain in addition to a strain where the binding motif is ablated can help identify mRNAs that have expression changes due to interaction with that site. Because this technique monitors transcript levels it is important to recognize that only targets who show a decrease in mRNA levels will be detected. sRNAs that function primarily by translation inhibition will yield negative results.

sRNA knockout libraries can also be used to establish the targets of sRNAs and the phenotypes associated with them. One way to create the knockout strains is to disrupt the chromosomal sRNA genes by insertion of a drug resistance cassette facilitated by phage recombination. The mutation can be confirmed by PCR amplification using primers flanking the recombination site. One of the most significant problems facing this technique is the potential to disrupt neighboring genetic elements which could muddle the interpretation of the observed effect. To combat this problem a cassette with a transcriptional terminator to

prevent read through into downstream genes can be used or homologous regions to the flanking genes can be incorporated into the cassette ends. Another way to create a library is to incorporate bar-codes into the insertion cassettes so that individual strains can be pooled for phenotype studies and the sensitivity or resistance of each strain can be identified using microarray analysis. This approach allows rapid phenotyping of a large number of strains. An issue that one could encounter when using knockout libraries is the inability to resolve changes in target expression due to naturally low sRNA expression levels. In addition, many different growth conditions may need to be tested in order to observe a regulatory event. The success of this approach for sRNAs that also encode a peptide, like SgrS, may also be limited because it would be difficult to associate the outcome with the sRNA rather than the peptide [124].

An sRNA knockout library was successfully implemented to identify a regulator of *ompX* in *S. typhimurium* by Papenfort *et al.* [121]. They observed that *ompX* was associated with Hfq in two different studies (one in *E. coli* and one in *S. typhimurium*) which led them to suggest that it is regulated by sRNAs. Also, previous work that indicated the conservation of 35 *E. coli* sRNAs in *S. typhimurium* led them to include the homologs in their library [16, 38, 125]. Incorporation of homologs could prove useful for other bacteria where sRNA have not yet been identified but computational methods have predicted homologs of known sRNAs. To screen the library for sRNAs that affect the expression of OmpX they grew the strains and performed western blots to compare the amount of protein present in the knockouts compared to the wild type strain. They easily observed a significant increase of expression in one knockout strain indicating a specific regulatory effect. A useful validation experiment is to complement the knockout strain with a plasmid carrying the sRNA to observe the return of normal regulation. Overall they were able to find an sRNA regulator of *ompX* in *S. typhimurium* and further characterize it in their study.

Jin *et al.* used a similar approach to identify the sRNA regulator associated with a specific phenotype [126]. In this case they compared the ability of the mutants to recover from acid stress to find out which sRNAs were associated with the pathway. A valuable control that they performed was to delete genes adjacent to the sRNA that showed a sensitive phenotype to verify that it was the sRNA deletion, not disruption of adjacent genes that caused the effect. Next they sought to determine the target of the identified sRNA, GcvB. Neither of the previously known GcvB targets played a role in the phenotype but they did find evidence using an *rpoS-lacZ* fusion construct that suggested that GcvB up regulated *rpoS*. Curiously they were not able to identify any complementarity between the two RNAs. This result highlights the fact that interpretation of phenotype screens can be precarious as phenotypes may arise due to any number of regulatory events not necessarily canonical sRNA mediated effects.

The use of bar-coded deletion libraries have been used to assess protein coding genes in *E. coli*, and genes in yeast but Hobbs *et al.* was the first to tailor this approach to sRNAs [127]. Using this technique they identified the Hfq-dependent sRNAs, RybB and MicA as being important in cell envelope stress. An advantage of this approach is the ability to phenotype a large number of deletion strains simultaneously. Homologous recombination was used to insert an antibiotic resistance gene in the place of the sRNA. Uniquely, they incorporated bar-codes distinct for each deletion so that they could be identified and quantified by microarray. They also incorporated common primer binding sites to be used for amplification before microarray analysis. To identify which genes were associated with cell envelope stress phenotypes they grew overnight cultures of all of the strains individually and then combined them for stress challenge. The genomic DNA was then purified and the barcodes amplified followed by hybridization to a commercially available microarray. Signals must be corrected to account for the nonlinear relationship between actual bar-code

concentration and fluorescence signal to obtain the relative abundance [127]. Relative abundances that are significantly increased compared non-stressed cells indicate a resistance to stress and signal less than indicate sensitivity to stress. The strains that exhibited phenotypes are then assessed in a one on one stress challenge and complementation experiments are performed to verify the results for a select group.

4.5 Validation Using Reporter Gene Fusions

The techniques presented so far in this section have aimed to elucidate the functions of Hfq binding RNAs by determining their RNA binding partner in a relatively high-throughput fashion. The nature of these techniques can lead to false positives or to identify regulation that is occurring by a mechanism other than base-pairing with Hfq-dependent sRNAs. It is therefore necessary to perform any number of validation techniques on an individual basis to determine if there is a direct base pairing interaction between the RNAs, if regulation occurs by affecting mRNA stability or by blocking translation and if the process is Hfq dependent. Several classic approaches can answer these questions, including, northern blot analysis, toe-printing, structure mapping and mutational analysis [99–101, 116, 117, 119, 121]. Reporter gene fusions have become a popular way to validate sRNA-mRNA interactions as well as biological significance and will be the focus of our discussion below.

Fusions of mRNAs with *lacZ* and *gfp* allow monitoring of direct effects of a sRNA on the target regardless of the natural transcription level of the mRNA in a given condition. The constructs usually include the 5'UTR starting from the annotated transcription start site through approximately 10 codons. The region is incorporated in frame with the reporter gene and is under the control of an inducible promoter [128]. It should be noted that interactions further into the coding sequence have been observed. If a previously seen regulatory event does not occur the length of coding sequence included in the fusion may need to be extended [100]. An inducible promoter is preferable to the natural target gene promoter so that the effect on translation instead of transcription does not have to be verified in additional experiments. Any other non-sRNA related regulatory regions of the mRNA should be removed or otherwise accounted for in order to draw clear conclusions about the regulatory outcome caused by the sRNA. The fusion can be created chromosomally or in a plasmid. The chromosomal fusion more accurately reflects natural gene expression but is more time consuming. A low copy plasmid fusion can provide a similar effect and is a simpler and less time consuming strategy.

To assess the effect of an sRNA on the reporter gene a plasmid containing the sRNA of interest is transformed into the strain harboring the fusion. Transcription of the sRNA can be under control of a constitutive promoter or an inducible promoter but it is important that both mRNA and sRNA are expressed at the same time. It has been seen that uncoordinated transcription of regulatory partners can disrupt regulation by sequestering Hfq. Translation output of *lacZ* constructs can be quantified using a β -galactosidase assay to determine the activity of the enzyme. This assay is somewhat more time consuming than measuring fluorescence in the GFP assays. GFP expression can be monitored by colony fluorescence, western blots with an anti-GFP antibody, cell lysates, and in whole cell liquid cultures. Measuring by colony fluorescence is the easiest but is less sensitive and not as quantitative. Measuring from whole cell cultures omits a lysis step and can save time but measuring from a lysate can increase reproducibility and doesn't require a flow cytometer. The western blot is the most time consuming but can detect fusion protein levels even if fluorescence signal is low.

The *lacZ* reporter is the most traditional option and it has been successfully used to monitor Hfq-dependent sRNA regulation in many cases [116, 119, 129]. In the studies discussed below the constructs are used to verify regulation observed in a large screen, to definitively

show a direct interaction between the two RNAs and to validate the binding site of the RNAs. Mandin *et al.* created a chromosomal *dpiB-lacZ* fusion to reproduce regulation by RybC that was observed in a genome wide screen that they performed [119]. They went on to verify the computationally predicted RNA-RNA binding site by performing mutational analysis. Three point mutations were incorporated into the predicted RNA binding site of the plasmid borne sRNA and its ability to affect the β -galactosidase activity of the mRNA fusion was monitored. They observed that the fusion was no longer regulated and were able to restore regulation by introducing compensatory mutations into the mRNA fusion. This assay unequivocally demonstrates that a direct interaction between the two RNAs is responsible for regulation and defines important residues involved. Udekwu *et al.* performed a similar compensatory mutational analysis but ablated six nucleotides thought to take place in RNA binding [116]. Depending on the RNA pair the number of nucleotide mutations necessary to destroy regulation may differ but care should be taken to ensure that the mutations do not cause significant secondary structure changes that could contribute to the observed outcome.

The use of GFP fusions has become popular more recently to study the regulatory outcomes of Hfq dependent sRNA-mRNA pairs. The Vogel lab established this approach by studying several regulatory pairs in a uniform manner using the two plasmid system and was able to reproduce all previously observed regulation [128]. Since then it has been used in several applications. Papenfort *et al.* performed a compensatory mutation analysis to pin point the interaction site between *ompX* and CyaR [121]. The assay has also been used to assess the ability of a series of mRNA constructs of different lengths to be regulated by the sRNA partner [100]. This experiment allows one to obtain preliminary information about the location of the RNA-RNA binding site. Nine potential targets obtained in a pulse expression experiment were confirmed in a more high-throughput manner using GFP fusions as well [101].

Our lab has also incorporated this technique into the detailed characterization of *glmS* regulation (Figure 5) [73]. After characterizing the interaction of *glmS* and Hfq using EMSA and determining the *glmS* secondary structure using SHAPE, the GFP assay was employed to further investigate the importance of the Hfq binding site on *glmS*. To do this we incorporated point mutations into the predicted Hfq binding site on *glmS* and monitored the ability of Hfq and GlnY/GlnZ to regulate the message. We were able to show based on GFP expression that the Hfq binding site was critical for regulation of *glmS*. The assay that we used involved co-expression of the *glmS-gfp* fusion and the sRNA plasmids upon addition of arabinose. Arabinose conditions were optimized to obtain optimal GFP expression and observable regulatory changes. The experiment was performed in triplicate using independently grown overnight cultures that are diluted on the following day to start the assay. Cells were harvested in early stationary phase and lysed using TritonX-100 in the presence of protease inhibitor and lysozyme. The lysate was centrifuged and the supernatant was assayed for GFP signal using a multi-label plate reader. The fluorescence signal was normalized to an identical culture where GFP was not induced.

5. Summary

In this review we have discussed various methods to discover Hfq binding RNAs, to characterize their interactions and to investigate their functions with the goal of serving as a guide to which technique is best suited to an individual researchers system and questions. New techniques such as high-throughput sequencing, CLIP, *in vivo* and high-throughput RNA structure probing promise new discoveries on the horizon. The existing knowledge from model systems can help pave the way to investigation of sRNAs and Hfq in pathogens that could serve as potential therapeutic targets. Whether you are starting from square one in

an organism where Hfq and sRNAs have not yet been characterized or you are interested in a specific regulatory pair in a well-known system that you wish to understand better there are many tools to guide your query. The significance of the role that Hfq-associated sRNAs play in coordinating gene regulation has never been more obvious and there is no doubt that as we continue to study these fascinating systems we will be greeted with even more surprising features and roles.

Acknowledgments

Funding for this work was supported in part by a grant from NIH (GM-075068 to A.L.F) and support from Wayne State University Office for the Vice President for Research and the Department of Chemistry. Many thanks to Rebecca Swett for assistance preparing Figure 1B.

References

- [1]. Storz G, Vogel J, Wassarman KM. Regulation by small RNAs in bacteria: expanding frontiers. *Mol Cell*. 2011; 43:880–891. [PubMed: 21925377]
- [2]. Berezikov E. Evolution of microRNA diversity and regulation in animals. *Nat Rev Genet*. 2011; 12:846–860. [PubMed: 22094948]
- [3]. Waters LS, Storz G. Regulatory RNAs in bacteria. *Cell*. 2009; 136:615–628. [PubMed: 19239884]
- [4]. Grundy FJ, Henkin TM. From ribosome to riboswitch: control of gene expression in bacteria by RNA structural rearrangements. *Crit Rev Biochem Mol Biol*. 2006; 41:329–338. [PubMed: 17092822]
- [5]. Brantl S. Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr Opin Microbiol*. 2007; 10:102–109. [PubMed: 17387036]
- [6]. Aiba H. Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr Opin Microbiol*. 2007; 10:134–139. [PubMed: 17383928]
- [7]. Fender A, Elf J, Hampel K, Zimmermann B, Wagner EG. RNAs actively cycle on the Sm-like protein Hfq. *Genes Dev*. 2010; 24:2621–2626. [PubMed: 21123649]
- [8]. Maki K, Morita T, Otaka H, Aiba H. A minimal base-pairing region of a bacterial small RNA SgrS required for translational repression of ptsG mRNA. *Mol Microbiol*. 2010; 76:782–792. [PubMed: 20345651]
- [9]. Soper T, Mandin P, Majdalani N, Gottesman S, Woodson SA. Positive regulation by small RNAs and the role of Hfq. *Proc Natl Acad Sci U S A*. 2010; 107:9602–9607. [PubMed: 20457943]
- [10]. Doetsch M, Stampfl S, Furtig B, Beich-Frandsen M, Saxena K, Lybecker M, Schroeder R. Study of *E. coli* Hfq's RNA annealing acceleration and duplex destabilization activities using substrates with different GC-contents. *Nucleic Acids Res*. 2013; 41:487–497. [PubMed: 23104381]
- [11]. Lee, T.; Feig, A. The RNA-Protein Complexes of *E. coli* Hfq: Form and Function. In: Walter, N.; Woodson, S.; Batey, R., editors. *Non-Protein Coding RNAs* Elsevier. New York: 2008. p. 249-272.
- [12]. Chao Y, Vogel J. The role of Hfq in bacterial pathogens. *Curr Opin Microbiol*. 2012; 13:24–33. [PubMed: 20080057]
- [13]. Ansong C, Yoon H, Porwollik S, Mottaz-Brewer H, Petritis BO, Jaitly N, Adkins JN, McClelland M, Heffron F, Smith RD. Global systems-level analysis of Hfq and SmpB deletion mutants in *Salmonella*: implications for virulence and global protein translation. *PLoS One*. 2009; 4:e4809. [PubMed: 19277208]
- [14]. Christiansen JK, Larsen MH, Ingmer H, Sogaard-Andersen L, Kallipolitis BH. The RNA-binding protein Hfq of *Listeria monocytogenes*: role in stress tolerance and virulence. *J Bacteriol*. 2004; 186:3355–3362. [PubMed: 15150220]
- [15]. Tsui HC, Leung HC, Winkler ME. Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in *Escherichia coli* K-12. *Mol Microbiol*. 1994; 13:35–49. [PubMed: 7984093]
- [16]. Sittka A, Pfeiffer V, Tedin K, Vogel J. The RNA chaperone Hfq is essential for the virulence of *Salmonella typhimurium*. *Mol Microbiol*. 2007; 63:193–217. [PubMed: 17163975]

- [17]. Ding Y, Davis BM, Waldor MK. Hfq is essential for *Vibrio cholerae* virulence and downregulates sigma expression. *Mol Microbiol.* 2004; 53:345–354. [PubMed: 15225327]
- [18]. Horstmann N, Orans J, Valentin-Hansen P, Shelburne SA 3rd, Brennan RG. Structural mechanism of *Staphylococcus aureus* Hfq binding to an RNA A-tract. *Nucleic Acids Res.* 2012; 40:11023–11035. [PubMed: 22965117]
- [19]. Mikulecky PJ, Kaw MK, Brescia CC, Takach JC, Sledjeski DD, Feig AL. *Escherichia coli* Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nat Struct Mol Biol.* 2004; 11:1206–1214. [PubMed: 15531892]
- [20]. Link TM, Valentin-Hansen P, Brennan RG. Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA. *Proc Natl Acad Sci U S A.* 2009; 106:19292–19297. [PubMed: 19889981]
- [21]. Schumacher MA, Pearson RF, Moller T, Valentin-Hansen P, Brennan RG. Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J.* 2002; 21:3546–3556. [PubMed: 12093755]
- [22]. Brescia CC, Mikulecky PJ, Feig AL, Sledjeski DD. Identification of the Hfq-binding site on DsrA RNA: Hfq binds without altering DsrA secondary structure. *RNA.* 2003; 9:33–43. [PubMed: 12554874]
- [23]. Otaka H, Ishikawa H, Morita T, Aiba H. PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proc Natl Acad Sci U S A.* 2011; 108:13059–13064. [PubMed: 21788484]
- [24]. Sauer E, Weichenrieder O. Structural basis for RNA 3'-end recognition by Hfq. *Proc Natl Acad Sci U S A.* 2011; 108:13065–13070. [PubMed: 21737752]
- [25]. Soper TJ, Woodson SA. The rpoS mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. *RNA.* 2008; 14:1907–1917. [PubMed: 18658123]
- [26]. Salim NN, Feig AL. An upstream Hfq binding site in the fhlA mRNA leader region facilitates the OxyS-fhlA interaction. *PLoS One.* 2010; 5:e13028. [PubMed: 20927406]
- [27]. Chen CY, Shyu AB. AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem Sci.* 1995; 20:465–470. [PubMed: 8578590]
- [28]. Someya T, Baba S, Fujimoto M, Kawai G, Kumasaka T, Nakamura K. Crystal structure of Hfq from *Bacillus subtilis* in complex with SELEX-derived RNA aptamer: insight into RNA-binding properties of bacterial Hfq. *Nucleic Acids Res.* 2012; 40:1856–1867. [PubMed: 22053080]
- [29]. Beich-Frandsen M, Vecerek B, Konarev PV, Sjoblom B, Kloiber K, Hammerle H, Rajkowitsch L, Miles AJ, Kontaxis G, Wallace BA, Svergun DI, Konrat R, Blasi U, Djinovic-Carugo K. Structural insights into the dynamics and function of the C-terminus of the *E. coli* RNA chaperone Hfq. *Nucleic Acids Res.* 2011; 39:4900–4915. [PubMed: 21330354]
- [30]. Sauer E, Schmidt S, Weichenrieder O. Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. *Proc Natl Acad Sci U S A.* 2012; 109:9396–9401. [PubMed: 22645344]
- [31]. Vincent HA, Henderson CA, Ragan TJ, Garza-Garcia A, Cary PD, Gowers DM, Malfois M, Driscoll PC, Sobott F, Callaghan AJ. Characterization of *Vibrio cholerae* Hfq provides novel insights into the role of the Hfq C-terminal region. *J Mol Biol.* 2012; 420:56–69. [PubMed: 22484176]
- [32]. Beisel CL, Storz G. Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiol Rev.* 2010; 34:866–882. [PubMed: 20662934]
- [33]. Papenfort K, Vogel J. Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level. *Res Microbiol.* 2009; 160:278–287. [PubMed: 19366629]
- [34]. Adamson DN, Lim HN. Essential requirements for robust signaling in Hfq dependent small RNA networks. *PLoS Comput Biol.* 2011; 7:e1002138. [PubMed: 21876666]
- [35]. Chao Y, Papenfort K, Reinhardt R, Sharma CM, Vogel J. An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J.* 2012; 31:4005–4019. [PubMed: 22922465]

- [36]. Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TT, Hinton JC, Vogel J. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet.* 2008; 4:e1000163. [PubMed: 18725932]
- [37]. Majdalani N, Vanderpool CK, Gottesman S. Bacterial small RNA regulators. *Crit Rev Biochem Mol Biol.* 2005; 40:93–113. [PubMed: 15814430]
- [38]. Zhang A, Wassarman KM, Rosenow C, Tjaden BC, Storz G, Gottesman S. Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol.* 2003; 50:1111–1124. [PubMed: 14622403]
- [39]. Hussein R, Lim HN. Disruption of small RNA signaling caused by competition for Hfq. *Proc Natl Acad Sci U S A.* 2011; 108:1110–1115. [PubMed: 21189298]
- [40]. Moon K, Gottesman S. Competition among Hfq-binding small RNAs in *Escherichia coli*. *Mol Microbiol.* 2011; 82:1545–1562. [PubMed: 22040174]
- [41]. Butland G, Peregrin-Alvarez JM, Li J, Yang W, Yang X, Canadien V, Starostine A, Richards D, Beattie B, Krogan N, Davey M, Parkinson J, Greenblatt J, Emili A. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature.* 2005; 433:531–537. [PubMed: 15690043]
- [42]. Carabetta VJ, Silhavy TJ, Cristea IM. The response regulator SprE (RssB) is required for maintaining poly(A) polymerase I-degradosome association during stationary phase. *J Bacteriol.* 2010; 192:3713–3721. [PubMed: 20472786]
- [43]. Hu P, Janga SC, Babu M, Diaz-Mejia JJ, Butland G, Yang W, Pogoutse O, Guo X, Phanse S, Wong P, Chandran S, Christopoulos C, Nazarians-Armavil A, Nasser NK, Musso G, Ali M, Nazemof N, Eroukova V, Golshani A, Paccanaro A, Greenblatt JF, Moreno-Hagelsieb G, Emili A. Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. *PLoS Biol.* 2009; 7:e96. [PubMed: 19402753]
- [44]. Mohanty BK, Maples VF, Kushner SR. The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in *Escherichia coli*. *Mol Microbiol.* 2004; 54:905–920. [PubMed: 15522076]
- [45]. Morita T, Maki K, Aiba H. RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev.* 2005; 19:2176–2186. [PubMed: 16166379]
- [46]. Carmichael GG, Weber K, Niveleau A, Wahba AJ. The host factor required for RNA phage Qbeta RNA replication in vitro. Intracellular location, quantitation, and purification by polyadenylate-cellulose chromatography. *J Biol Chem.* 1975; 250:3607–3612. [PubMed: 805130]
- [47]. Kajitani M, Kato A, Wada A, Inokuchi Y, Ishihama A. Regulation of the *Escherichia coli* hfq gene encoding the host factor for phage Q beta. *J Bacteriol.* 1994; 176:531–534. [PubMed: 8288550]
- [48]. Masse E, Vanderpool CK, Gottesman S. Effect of RyhB small RNA on global iron use in *Escherichia coli*. *J Bacteriol.* 2005; 187:6962–6971. [PubMed: 16199566]
- [49]. Papenfort K, Pfeiffer V, Mika F, Lucchini S, Hinton JC, Vogel J. SigmaE-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global omp mRNA decay. *Mol Microbiol.* 2006; 62:1674–1688. [PubMed: 17427289]
- [50]. Vanderpool CK, Gottesman S. Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Mol Microbiol.* 2004; 54:1076–1089. [PubMed: 15522088]
- [51]. Vogel J, Luisi BF. Hfq and its constellation of RNA. *Nat Rev Microbiol.* 2011; 9:578–589. [PubMed: 21760622]
- [52]. Christiansen JK, Nielsen JS, Ebersbach T, Valentin-Hansen P, Sogaard-Andersen L, Kallipolitis BH. Identification of small Hfq-binding RNAs in *Listeria monocytogenes*. *RNA.* 2006; 12:1383–1396. [PubMed: 16682563]
- [53]. Vogel J, Bartels V, Tang TH, Churakov G, Slagter-Jager JG, Huttenhofer A, Wagner EG. RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.* 2003; 31:6435–6443. [PubMed: 14602901]

- [54]. Zhang A, Wassarman KM, Ortega J, Steven AC, Storz G. The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol Cell*. 2002; 9:11–22. [PubMed: 11804582]
- [55]. Sonnleitner E, Sorger-Domenigg T, Madej MJ, Findeiss S, Hackermuller J, Huttenhofer A, Stadler PF, Blasi U, Moll I. Detection of small RNAs in *Pseudomonas aeruginosa* by RNomics and structure-based bioinformatic tools. *Microbiology*. 2008; 154:3175–3187. [PubMed: 18832323]
- [56]. Ramos CG, Grilo AM, da Costa PJ, Leitao JH. Experimental identification of small non-coding regulatory RNAs in the opportunistic human pathogen *Burkholderia cenocepacia* J2315. *Genomics*. 2012:1–10.
- [57]. Messer LI, Currey KM, O'Neill BJ, Maizel JV Jr, Levin JG, Gerwin BI. Functional analysis of reverse transcription by a frameshift pol mutant of murine leukemia virus. *Virology*. 1985; 146:146–152. [PubMed: 2412343]
- [58]. Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S. Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev*. 2001; 15:1637–1651. [PubMed: 11445539]
- [59]. Vanderpool CK, Balasubramanian D, Lloyd CR. Dual-function RNA regulators in bacteria. *Biochimie*. 2011; 93:1943–1949. [PubMed: 21816203]
- [60]. Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*. 2009; 460:479–486. [PubMed: 19536157]
- [61]. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*. 2010; 141:129–141. [PubMed: 20371350]
- [62]. Milek M, Wyler E, Landthaler M. Transcriptome-wide analysis of protein-RNA interactions using high-throughput sequencing. *Semin Cell Dev Biol*. 2012; 23:206–212. [PubMed: 22212136]
- [63]. Ule J, Jensen K, Mele A, Darnell RB. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods*. 2005; 37:376–386. [PubMed: 16314267]
- [64]. Geng J, Song Y, Yang L, Feng Y, Qiu Y, Li G, Guo J, Bi Y, Qu Y, Wang W, Wang X, Guo Z, Yang R, Han Y. Involvement of the post-transcriptional regulator Hfq in *Yersinia pestis* virulence. *PLoS One*. 2009; 4:e6213. [PubMed: 19593436]
- [65]. Roscetto E, Angrisano T, Costa V, Casalino M, Forstner KU, Sharma CM, Di Nocera PP, De Gregorio E. Functional Characterization of the RNA Chaperone Hfq in the Opportunistic Human Pathogen *Stenotrophomonas maltophilia*. *J Bacteriol*. 2012; 194:5864–5874. [PubMed: 22923593]
- [66]. Guisbert E, Rhodius VA, Ahuja N, Witkin E, Gross CA. Hfq modulates the sigmaE-mediated envelope stress response and the sigma32-mediated cytoplasmic stress response in *Escherichia coli*. *J Bacteriol*. 2007; 189:1963–1973. [PubMed: 17158661]
- [67]. Barra-Bily L, Fontenelle C, Jan G, Flechard M, Trautwetter A, Pandey SP, Walker GC, Blanco C. Proteomic alterations explain phenotypic changes in *Sinorhizobium meliloti* lacking the RNA chaperone Hfq. *J Bacteriol*. 2010; 192:1719–1729. [PubMed: 20081032]
- [68]. Lorenz C, Gesell T, Zimmermann B, Schoeberl U, Bilusic I, Rajkowitsch L, Waldsich C, von Haeseler A, Schroeder R. Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts. *Nucleic Acids Res*. 2010; 38:3794–3808. [PubMed: 20348540]
- [69]. Altuvia S, Weinstein-Fischer D, Zhang A, Postow L, Storz G. A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell*. 1997; 90:43–53. [PubMed: 9230301]
- [70]. Balbontin R, Fiorini F, Figueroa-Bossi N, Casades J, Bossi L. Recognition of heptameric seed sequence underlies multi-target regulation by RybB small RNA in *Salmonella enterica*. *Mol Microbiol*. 2010; 78:380–394. [PubMed: 20979336]
- [71]. Kierzek E, Kierzek R, Turner DH, Catrina IE. Facilitating RNA structure prediction with microarrays. *Biochemistry*. 2006; 45:581–593. [PubMed: 16401087]

- [72]. Papenfort K, Bouvier M, Mika F, Sharma CM, Vogel J. Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA. *Proc Natl Acad Sci U S A*. 2010; 107:20435–20440. [PubMed: 21059903]
- [73]. Salim NN, Faner MA, Philip JA, Feig AL. Requirement of upstream Hfq-binding (ARN)_x elements in glmS and the Hfq C-terminal region for GlmS upregulation by sRNAs GlmZ and GlmY. *Nucleic Acids Res*. 2012; 40:8021–8032. [PubMed: 22661574]
- [74]. Soper TJ, Doxzen K, Woodson SA. Major role for mRNA binding and restructuring in sRNA recruitment by Hfq. *RNA*. 2011; 17:1544–1550. [PubMed: 21705431]
- [75]. Lease RA, Woodson SA. Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. *J Mol Biol*. 2004; 344:1211–1223. [PubMed: 15561140]
- [76]. Olejniczak M. Despite similar binding to the Hfq protein regulatory RNAs widely differ in their competition performance. *Biochemistry*. 2011; 50:4427–4440. [PubMed: 21510661]
- [77]. Lee T, Feig AL. The RNA binding protein Hfq interacts specifically with tRNAs. *RNA*. 2008; 14:514–523. [PubMed: 18230766]
- [78]. Kawamoto H, Koide Y, Morita T, Aiba H. Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Mol Microbiol*. 2006; 61:1013–1022. [PubMed: 16859494]
- [79]. Jing M, Bowser MT. Methods for measuring aptamer-protein equilibria: a review. *Anal Chim Acta*. 686:9–18. [PubMed: 21237304]
- [80]. Katsamba PS, Park S, Laird-Offringa IA. Kinetic studies of RNA-protein interactions using surface plasmon resonance. *Methods*. 2002; 26:95–104. [PubMed: 12054886]
- [81]. Feig AL. Studying RNA-RNA and RNA-protein interactions by isothermal titration calorimetry. *Methods Enzymol*. 2009; 468:409–422. [PubMed: 20946780]
- [82]. Perrin. Polarisation de la lumiere de fluorescence vie moyenne des molecules dans l'etat excite. *journal de physique et le radium*. 1926; 7:390–401.
- [83]. Arluison V, Mutyam SK, Mura C, Marco S, Sukhodolets MV. Sm-like protein Hfq: location of the ATP-binding site and the effect of ATP on Hfq-- RNA complexes. *Protein Sci*. 2007; 16:1830–1841. [PubMed: 17660259]
- [84]. Sun X, Wartell RM. Escherichia coli Hfq binds A18 and DsrA domain II with similar 2:1 Hfq6/ RNA stoichiometry using different surface sites. *Biochemistry*. 2006; 45:4875–4887. [PubMed: 16605255]
- [85]. Updegrove TB, Correia JJ, Chen Y, Terry C, Wartell RM. The stoichiometry of the Escherichia coli Hfq protein bound to RNA. *RNA*. 2011; 17:489–500. [PubMed: 21205841]
- [86]. Wang W, Wang L, Zou Y, Zhang J, Gong Q, Wu J, Shi Y. Cooperation of Escherichia coli Hfq hexamers in DsrA binding. *Genes Dev*. 2011; 25:2106–2117. [PubMed: 21979921]
- [87]. Geissmann TA, Touati D. Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *EMBO J*. 2004; 23:396–405. [PubMed: 14739933]
- [88]. Desnoyers G, Morissette A, Prevost K, Masse E. Small RNA-induced differential degradation of the polycistronic mRNA iscRSUA. *EMBO J*. 2009; 28:1551–1561. [PubMed: 19407815]
- [89]. Lindell M, Romby P, Wagner EG. Lead(II) as a probe for investigating RNA structure in vivo. *RNA*. 2002; 8:534–541. [PubMed: 11991646]
- [90]. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc*. 2005; 127:4223–4231. [PubMed: 15783204]
- [91]. Mitra S, Shcherbakova IV, Altman RB, Brenowitz M, Laederach A. High-throughput single-nucleotide structural mapping by capillary automated footprinting analysis. *Nucleic Acids Res*. 2008; 36:e63. [PubMed: 18477638]
- [92]. Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews DH, Giddings MC, Weeks KM. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol*. 2008; 6:e96. [PubMed: 18447581]
- [93]. Vasa SM, Guex N, Wilkinson KA, Weeks KM, Giddings MC. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA*. 2008; 14:1979–1990. [PubMed: 18772246]

- [94]. Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*. 2010; 11:129. [PubMed: 20230624]
- [95]. Lucks JB, Mortimer SA, Trapnell C, Luo S, Aviran S, Schroth GP, Pachter L, Doudna JA, Arkin AP. Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc Natl Acad Sci U S A*. 108:11063–11068. [PubMed: 21642531]
- [96]. Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, Chang HY. RNA SHAPE analysis in living cells. *Nat Chem Biol*. 2013; 9:18–20. [PubMed: 23178934]
- [97]. Fratzek A, Kierzek R, Kierzek E. Isoenergetic microarrays to study the structure and interactions of DsrA and OxyS RNAs in two- and three-component complexes. *Biochemistry*. 2011; 50:7647–7665. [PubMed: 21793590]
- [98]. Chen S, Zhang A, Blyn LB, Storz G. MicC, a second small-RNA regulator of Omp protein expression in *Escherichia coli*. *J Bacteriol*. 2004; 186:6689–6697. [PubMed: 15466019]
- [99]. Jorgensen MG, Nielsen JS, Boysen A, Franch T, Moller-Jensen J, Valentin-Hansen P. Small regulatory RNAs control the multi-cellular adhesive lifestyle of *Escherichia coli*. *Mol Microbiol*. 2012; 84:36–50. [PubMed: 22250746]
- [100]. Frohlich KS, Papenfort K, Berger AA, Vogel J. A conserved RpoS-dependent small RNA controls the synthesis of major porin OmpD. *Nucleic Acids Res*. 2012; 40:3623–3640. [PubMed: 22180532]
- [101]. Sharma CM, Papenfort K, Pernitzsch SR, Mollenkopf HJ, Hinton JC, Vogel J. Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol Microbiol*. 2011; 81:1144–1165. [PubMed: 21696468]
- [102]. Bailey TL, Gribskov M. Methods and statistics for combining motif match scores. *J Comput Biol*. 1998; 5:211–221. [PubMed: 9672829]
- [103]. Tjaden B. TargetRNA: a tool for predicting targets of small RNA action in bacteria. *Nucleic Acids Res*. 2008; 36:W109–113. [PubMed: 18477632]
- [104]. Li W, Ying X, Lu Q, Chen L. Predicting sRNAs and Their Targets in Bacteria. *Genomics Proteomics Bioinformatics*. 2012; 10:276–284. [PubMed: 23200137]
- [105]. Hebrard M, Kroger C, Srikumar S, Colgan A, Handler K, Hinton JC. sRNAs and the virulence of *Salmonella enterica* serovar Typhimurium. *RNA Biol*. 2012; 9:437–445. [PubMed: 22546935]
- [106]. Postic G, Dubail I, Frapy E, Dupuis M, Dieppedale J, Charbit A, Meibom KL. Identification of a novel small RNA modulating *Francisella tularensis* pathogenicity. *PLoS One*. 2012; 7:e41999. [PubMed: 22848684]
- [107]. Zhao Y, Li H, Hou Y, Cha L, Cao Y, Wang L, Ying X, Li W. Construction of two mathematical models for prediction of bacterial sRNA targets. *Biochem Biophys Res Commun*. 2008; 372:346–350. [PubMed: 18501192]
- [108]. Busch A, Richter AS, Backofen R. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*. 2008; 24:2849–2856. [PubMed: 18940824]
- [109]. Eggenhofer F, Tafer H, Stadler PF, Hofacker IL. RNApredator: fast accessibility-based prediction of sRNA targets. *Nucleic Acids Res*. 2011; 39:W149–154. [PubMed: 21672960]
- [110]. Ying X, Cao Y, Wu J, Liu Q, Cha L, Li W. sTarPicker: a method for efficient prediction of bacterial sRNA targets based on a two-step model for hybridization. *PLoS One*. 2011; 6:e22705. [PubMed: 21799937]
- [111]. Modi SR, Camacho DM, Kohanski MA, Walker GC, Collins JJ. Functional characterization of bacterial sRNAs using a network biology approach. *Proc Natl Acad Sci U S A*. 2011; 108:15522–15527. [PubMed: 21876160]
- [112]. Mika F, Busse S, Possling A, Berkholz J, Tschowri N, Sommerfeldt N, Pruteanu M, Hengge R. Targeting of *csgD* by the small regulatory RNA RprA links stationary phase, biofilm formation and cell envelope stress in *Escherichia coli*. *Mol Microbiol*. 2012; 84:51–65. [PubMed: 22356413]
- [113]. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol*. 2007; 5:e8. [PubMed: 17214507]

- [114]. Masse E, Gottesman S. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A*. 2002; 99:4620–4625. [PubMed: 11917098]
- [115]. Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EG, Margalit H, Altuvia S. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol*. 2001; 11:941–950. [PubMed: 11448770]
- [116]. Udekwu KI, Darfeuille F, Vogel J, Reimegard J, Holmqvist E, Wagner EG. Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. *Genes Dev*. 2005; 19:2355–2366. [PubMed: 16204185]
- [117]. Mandin P, Gottesman S. Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. *EMBO J*. 2010; 29:3094–3107. [PubMed: 20683441]
- [118]. De Lay N, Gottesman S. A complex network of small non-coding RNAs regulate motility in *Escherichia coli*. *Mol Microbiol*. 2012; 86:524–538. [PubMed: 22925049]
- [119]. Mandin P, Gottesman S. A genetic approach for finding small RNAs regulators of genes of interest identifies RybC as regulating the DpiA/DpiB two-component system. *Mol Microbiol*. 2009; 72:551–565. [PubMed: 19426207]
- [120]. Ulbrandt ND, Newitt JA, Bernstein HD. The *E. coli* signal recognition particle is required for the insertion of a subset of inner membrane proteins. *Cell*. 1997; 88:187–196. [PubMed: 9008159]
- [121]. Papenfort K, Pfeiffer V, Lucchini S, Sonawane A, Hinton JC, Vogel J. Systematic deletion of *Salmonella* small RNA genes identifies CyaR, a conserved CRP-dependent riboregulator of OmpX synthesis. *Mol Microbiol*. 2008; 68:890–906. [PubMed: 18399940]
- [122]. Bouvier M, Sharma CM, Mika F, Nierhaus KH, Vogel J. Small RNA binding to 5' mRNA coding region inhibits translational initiation. *Mol Cell*. 2008; 32:827–837. [PubMed: 19111662]
- [123]. Guillier M, Gottesman S. The 5' end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator. *Nucleic Acids Res*. 2008; 36:6781–6794. [PubMed: 18953042]
- [124]. Wadler CS, Vanderpool CK. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *Proc Natl Acad Sci U S A*. 2007; 104:20454–20459. [PubMed: 18042713]
- [125]. Hershberg R, Altuvia S, Margalit H. A survey of small RNA-encoding genes in *Escherichia coli*. *Nucleic Acids Res*. 2003; 31:1813–1820. [PubMed: 12654996]
- [126]. Jin Y, Watt RM, Danchin A, Huang JD. Small noncoding RNA GcvB is a novel regulator of acid resistance in *Escherichia coli*. *BMC Genomics*. 2009; 10:165. [PubMed: 19379489]
- [127]. Hobbs EC, Astarita JL, Storz G. Small RNAs and small proteins involved in resistance to cell envelope stress and acid shock in *Escherichia coli*: analysis of a bar-coded mutant collection. *J Bacteriol*. 2010; 192:59–67. [PubMed: 19734312]
- [128]. Urban JH, Vogel J. Translational control and target recognition by *Escherichia coli* small RNAs in vivo. *Nucleic Acids Res*. 2007; 35:1018–1037. [PubMed: 17264113]
- [129]. Majdalani N, Cunnig C, Sledjeski D, Elliott T, Gottesman S. DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proc Natl Acad Sci U S A*. 1998; 95:12462–12467. [PubMed: 9770508]
- [130]. Zimmermann B, Bilusic I, Lorenz C, Schroeder R. Genomic SELEX: a discovery tool for genomic aptamers. *Methods*. 2010; 52:125–132. [PubMed: 20541015]

Highlights

- Techniques relevant to the discovery and investigation of Hfq-RNA interactions are discussed
- The strengths and optimal application of the approaches are considered
- Issues unique to studying Hfq-RNA systems are addressed

Figure 1. Hfq-RNA Complex Formation and Hfq Binding Faces

A. Hfq binds sRNAs and mRNAs with similar affinity. Hfq may bind either the mRNA or sRNA first before forming the ternary complex. **B.** Crystal structures of Hfq-AU₅G (PDB ID: 1KQ1) and Hfq-polyA (PDB ID: 1HK9) superimposed. AU₅G binds the proximal face and polyA binds the distal of the homo-hexamer.

Figure 2. Discovery of Hfq Binding RNAs

A. Co-IP of Hfq bound RNAs using chromosomally FLAG tagged Hfq (adapted from Sittka *et al.*) [36]. Cellular extracts from *hfg*^{FLAG} cells are prepared and co-IP with an anti-FLAG antibody is performed. RNAs are extracted and modified to incorporate a polyA tail and 5'phosphate. A 5' adapter is ligated followed by cDNA synthesis and high-throughput sequencing to identify the bound RNAs. **B.** Genomic SELEX (adapted from Lorenz *et al.* and Zimmermann *et al.*) [68, 130]. A genomic library is created by random priming using primers that incorporate a T7 promoter and primer binding sites for reverse transcription and PCR. The library is transcribed to RNA and a binding reaction with Hfq is performed. Bound complexes are selected using filter binding. Bound RNAs are recovered from the filter followed by RT-PCR. The cycle is repeated multiple times followed by sequencing to identify the aptamers.

Figure 3. Electrophoretic Mobility Shift Assays

A. Two site fitting used for DsrA and Hfq (image from Lease and Woodson) [75]. ^{32}P labeled DsrA was titrated with 0–12.5 μM Hfq monomer. The D•H(I) band is DsrA bound to one Hfq hexamer and the D•H(II) band is DsrA bound to two Hfq hexamers. Plot of the fraction of DsrA bound (f_B) versus Hfq concentration showing the fitting of each site independently and combined. **B.** Competition assay of A₂₇ with pre-formed D•H(II) complex (image from Mikulecky *et al.*) [19]. Titration of increasing concentrations of A₂₇ (0–3 μM) leads to the formation of a ternary complex.

Figure 4. SHAPE Derived Structure and Hfq Footprinting of *glmS*

A. Normalized reactivities for each nucleotide in the presence and absence of Hfq (image from Salim *et al.*) [73]. Double stranded residues are indicated by P1, P2, etc..... and Hfq binding (ARN)_x sites are indicated. **B.** Schematic of the SHAPE derived secondary structure with reactivities and Hfq footprint superimposed. RBS is the ribosome binding site and GlmZ binding site is the binding site of the regulatory RNA. Footprints were deemed weak if Hfq binding resulted in a reactivity change between 0.3–0.59 and strong if the change was > 0.6.

Techniques for Finding RNA Binding Partners

Technique	Strengths	Weaknesses
Bioinformatics	<ul style="list-style-type: none"> • Increase efficiency by guiding lab work • Adaptable to species specific characteristics 	<ul style="list-style-type: none"> • High error rate • Requires experimental validation • Prior knowledge required
sRNA Over Expression	<ul style="list-style-type: none"> • Study sRNAs that normally have low expression • High throughput screening 	<ul style="list-style-type: none"> • Can disrupt the balance of the sRNA network • sRNAs not present in library will be missed • Does not distinguish direct vs. indirect effects
sRNA Pulse Expression	<ul style="list-style-type: none"> • Help discern direct vs. indirect effects 	<ul style="list-style-type: none"> • Can disrupt the balance of the sRNA network
sRNA Knockouts	<ul style="list-style-type: none"> • High throughput phenotyping using bar-codes • High throughput screening for mRNA regulators 	<ul style="list-style-type: none"> • Pleiotropic effects from gene knockout • Resolving changes in target expression