1-1-2012

Modes and mechanisms of hfq mediated stress regulation in bacteria

Nilshad Nilam Salim
Wayne State University,

MODES AND MECHANISMS OF HFQ MEDIATED STRESS REGULATION IN BACTERIA

by

NILSHAD N. SALIM

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2012

MAJOR: CHEMISTRY (Biochemistry)

Approved by:

_____

Advisor Date

_____

_____

_____

DEDICATION

Dedicated to

My parents and My grandmother

Nilam Salim, Shanaaz Salim and Shaiba Omerdeen

And also to my Wife and Daughter

Sanofar Abdeen and Nuha Salim

ACKNOWLEDGEMENTS

I thank God for providing me with strength, endurance, good health and guidance throughout this period that enabled me to successfully complete my Ph.D. I have been fortunate to receive an overwhelming amount of support and corporation throughout these five years that was essential in completing my dissertation work. First of all I would like to thank, Prof. Andrew Feig for his insightful guidance and encouragement that was immensely helpful for me to develop as a research scientist and as a person. I am also thankful to my committee members, Prof. David Rueda, Prof. Mary Kay Plfum and Prof. Athar Ansari for their invaluable suggestions throughout. I would like thank Prof. Christine Chow and Prof. Tamara Hendrickson for their support whenever I needed them. I would also like acknowledge my lab members at the Feig Lab for providing me with support, encouragement and for their help in generating a pleasant working environment which was a key aspect to my productivity. I would also like thank Dr. Amy Kerzmann, Dr. Taewoo Lee and Dr. Tuhina Banerjee for their support during the initial stages of my research. My sincere gratitude also goes to Rajan Lamichhane and Rui Zhao with whom I collaborated during the Kissing hairpin project. Finally I would like thank my wife Sanofar Abdeen for all of her support and encouragement throughout these years and to all of friends at Wayne State University who were there for me whenever I needed them the most.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

sRNA	small RNA
UTR	Un-translated region
mRNA	Messenger RNA
snRNP	Small nuclear ribonucleoproteins
Sm-	Smith-*like*
SELEX	Systematic evolution of ligands by exponential enrichment
RNP	Ribonucleoprotein
IDP	Intrinsically disordered proteins
SRCD	Synchrotron radiation circular dichroism
OMP	Outer membrane protein
TIR	Translation initiation region
RBS	Ribosome binding site
PAP I	PolyA polymerase
PNPase	Polynucleotide Phosphorylase
F220	*fhlA*220
F53	*fhlA*53

SHAPE	Selective 2'-hydroxyl acylation analyzed by primer extension
NMIA	*N*-methylisatoic anhydride
SD	Shine Dalgano
SPR	Surface Plasmon Resonance
asRNA	Anti-sense RNA
GFP	Green fluorescent protein
smFRET	Single-Molecule Fluorescence Energy Transfer
DIS	Dimerization initiation site
ITC	Isothermal titration calorimetry
KC	Kissing complex
ED	Extended duplex
NN	Nearest neighbor
NCp7	Nucleocapsid protein 7

INSTRUMENTATION

Isothermal titration calorimetry, VP ITC, GE Healthcare-MicroCal

UV melting, Aviv 14DS UV-VIS spectrophotometer

Surface Plasmon resonance, Biacore 2000, GE Healthcare-Biacore

Capillary sequencer, CEQ8000, Beckman Coulter

Fluorometer/Plate reader, Tecan, GENios Plus multi label plate reader

Phosphorimager, Typhoon 9200, GE-Healthcare-Amersham Biosciences

CHAPTER 1

Introduction to Functions of Hfq During Stress Regulation in Bacteria

1.1 Abstract

To survive, bacterial systems must be able respond to the ever-changing environmental conditions. Hfq has been shown to be an integral component in facilitating sRNA-coupled post-transcriptional gene regulation in bacteria that involves a variety of stress-responsive networks. Recent advances in sequencing and co-immunoprecipitation techniques have helped identify a variety of novel regulatory pathways that depend on Hfq. These findings indicate that the mechanism of Hfq mediated gene regulation is more complex than previously anticipated. In this chapter, we review in detail the present understanding of bacterial stress regulation, facilitated by Hfq using experimental observations gathered by the field.

1.2 Introduction

In bacteria, sRNAs are known to effect a variety of regulatory functions such as transcriptional attenuation (6S RNA) (1), translational regulation (OxyS-*fhlA*) (2), RNA stability (SgrS-*ptsG*) (3) and DNA maintenance/stability (CRISPER RNAs) (4,5). As a whole these sRNAs can be categorized into two main classes named cis- and trans-acting RNAs. Cis-RNAs are encoded from the strand opposite to that of the mRNA it regulates, providing extended base pairing with its target. Most of these cis-RNAs are encoded in mobile genetic elements such as bacteriophages, plasmids and transposons to help maintain appropriate copy numbers (6). Another class of RNAs that act in cis- is the riboswitch where the 5'-UTR of the mRNA acts as a sensor with the ability to

regulate gene expression through transcription or translation attenuation (7). On the other hand, trans-acting RNAs are involved in a diverse variety of regulatory networks. These RNAs are transcribed from a different locus than the target mRNA, providing limited sequence complementarity often relying on RNA recognition through structural motifs such as loop-loop (kissing) or bulge-loops interactions. In bacterial systems these RNA recognition processes are often facilitated by the chaperone protein Hfq.

In bacteria, Hfq is an essential component of sRNA-mediated stress-regulation that helps bacteria to survive in non-optimal environments. The primary function of Hfq is to facilitate sRNA-mRNA duplex formation. Absence of Hfq triggers phenotypes such as slow growth and increased sensitivity to stress. Also it was demonstrated that reduced virulence was observed in the absence of Hfq in a variety of bacterial pathogens (8-10). Despite the presence of a range of regulatory networks that depend on Hfq, significant ambiguity still exists regarding the exact mechanism(s) of how Hfq propagates gene regulation. In this chapter we describe the functional roles, mechanisms and partnerships which Hfq forges during bacterial stress regulation.

1.3 Structure of Hfq

The classical donut-shaped structure of Hfq is shown in **Figure 1.1A**. This type of proteins that form hexameric or heptameric cores belong to the Sm/Lsm family that are ubiquitous among eukaryotes and archaea. Sm-proteins interact with snRNAs (U1, U2, U4 and U5 among others) that presents a Sm-RNA sequence (single stranded U-rich RNAs) to form snRNP complexes that are part of the human spliceosome (11). On the other hand the heteroheptameric Lsm-proteins in eukaryotes transiently bind to a wide variety of RNAs. Their protein subunits are named Lsm1-8 and the two rings vary by the

exchange of the Lsm1 and Lsm8 subunits. Lsm1-7 is important in RNA degradation in the cytoplasm while Lsm2-8 functions during RNA maturation in the nucleus (12,13).

The highly conserved monomeric unit of Hfq is shown in **Figure 1.1B.** Each Hfq subunit comprises of a single α-helical domain followed by five β-sheets that constitute the nucleic acid binding Sm cores (Sm1:β_{1-3} and Sm2:β_{4-5}). The Sm cores are folded in such a way that it exposes two independent RNA binding surfaces called the distal (Sm1, indicated in yellow) and proximal sites (Sm2, indicated in blue) that have distinct RNA sequence specificities as indicated below. Continuous antiparallel interactions between beta strands β4 and β5' (prime denotes a neighboring subunits) facilitates the formation of the exceptionally stable toroidal structure of Hfq.

1.3.1 RNA binding at the distal and proximal biding sites of Hfq and implications to regulatory function

Several high resolution crystal structures are available that provides a clear interpretation of biochemical and biophysical data presented for RNA specificities for Hfq (14-18). The *E. coli* Hfq co-crystallized with A_{15} RNA is shown in **Figure 1.1C** (14)**.** The ploy (A) RNAs threads the distal binding site of the Hfq hexamer while exposing a fair number of residues to surrounding solvents. It was observed that the poly A RNAs were shared among Hfq, sectioning into six tri-nucleotide repeats. Each tripartite unit was shown to be composed of a 5'-adenosine specific site (A), a purine binding site (R) and a non-specific RNA binding site (N). The adenosine at the A site interacts with the peptide backbone of Lys31. The R site is buried between the β-sheets of two neighboring subunits **(Figure 1.1C)** where the purine ring stacks against the side chain residues of Tyr25, Leu26', Ile30' and Leu32'. The N site however does not form any

interaction with the protein backbone of Hfq and is flipped-out and could potentially make contact with solvent molecules or other RNAs. This arrangement of ribonucleotides at the N site could then presumably act as an exit site for RNA exchange which is a natural process for Hfq recycling among target RNAs (19,20). These crystallographic data confirmed previous reports of such specificities for A-rich sequences for Hfq and at the distal binding site (21). Furthermore the existence of $(ARN)_x$ motifs in many Hfq binding mRNAs and the specific enrichment of A-rich sequences that were observed in SELEX approaches that was used to probe Hfq binding RNAs, validates these sequence preferences on the distal surface of Hfq.

A recent crystal structure of *bacillus subtilis* Hfq bound to $(AG)_3A$ RNA was solved by Someya et al. (18). This AG-rich RNA sequence was identified using a SELEX screen that was performed to enrich *B. subtills* Hfq (Hfq^{Bs}) binding RNAs. Interestingly, in vitro studies suggests that $(AG)_3A$ RNA has a higher binding affinity than A_{18} for the distal surface of Hfq^{Bs}. This difference was attributed to the modes of RNA recognition in Hfq^{Bs} as it was evident in the $(AG)_3A$-Hfq^{Bs} crystal structure (18). In Hfq^{Bs}, $(AG)_3A$ assumed a circular conformation compared to the wavy binding pattern observed in the A_{18}-Hfq^{Ec} complex (14). The tripartite A-R-N binding motif that was observed in Hfq^{Ec} was reduced to a A-R (AG repeat) recognition element in Hfq^{Bs} when bound to $(AG)_3A$ RNA. The purine rings in the A bases stacks against the side chains of F24 and F29' (Y25 and I30 in Hfq^{Ec}) while forms H-bonds with S60, T61, G28 and Q30 residues. This binding mode is similar to the R-site in Hfq^{Ec}. The G-base forms a stacking interaction with the side chain of Q30 (K31 in Hfq^{Ec}) while H-bonding with R32 (Q33 in Hfq^{Ec}) that holds the purine ring parallel to the distal surface of Hfq^{Bs}. The G-

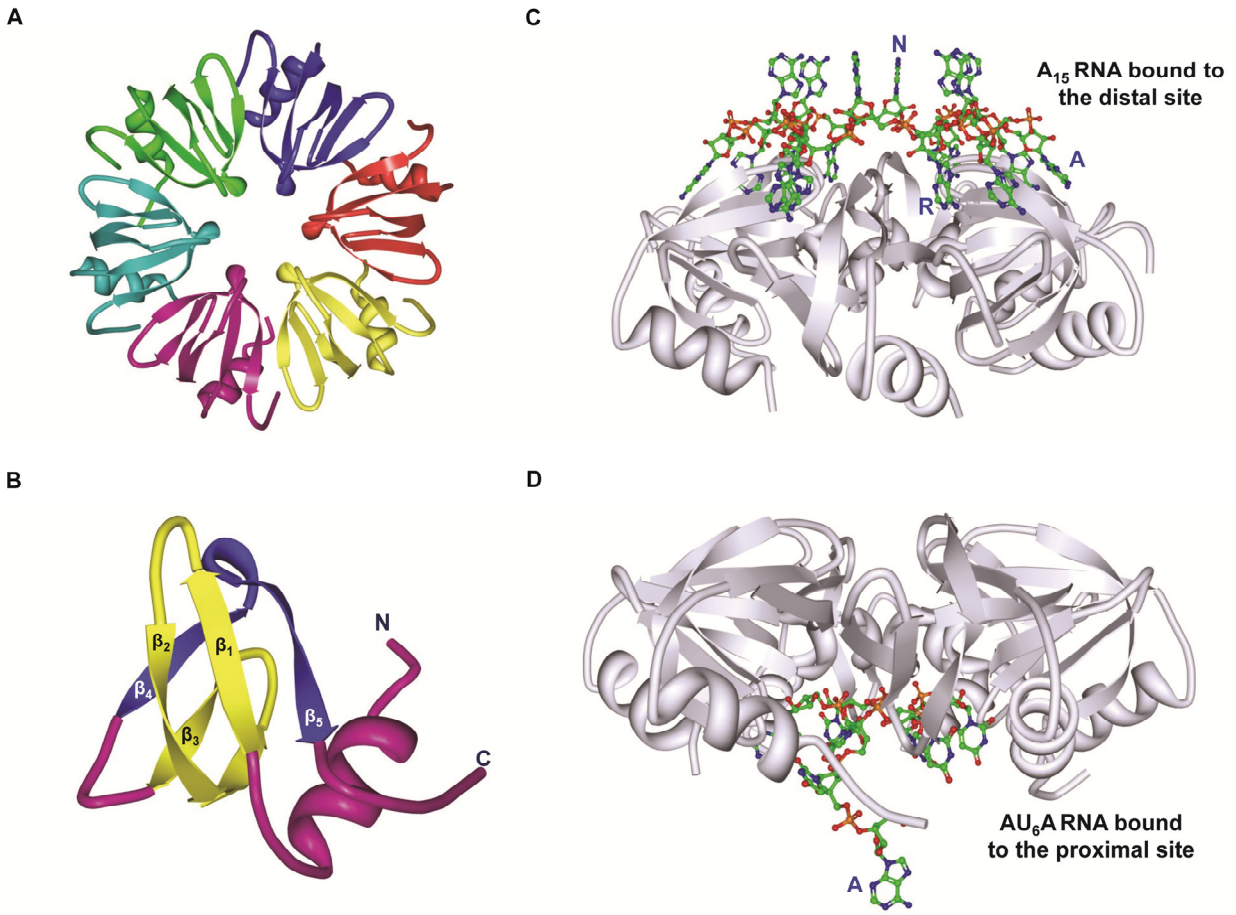
base site in Hfq^{Bs} was somewhat analogous to the A-site in Hfq^{Ec}. These observations imply the variation in both sequence and mode of RNA binding to Hfq based on differences in critical amino acid residues that are essential for nucleic acid binding (also see **Figure 1.1E)**.

RNA binding at the proximal site has been described using Hfq variants that belong to *staphylococcus aureus*, *E. coli* and *salmonella typhimurium* bound to AU_5G, AU_6A and U_6 RNAs respectively (16,17,22). To exemplify general RNA binding modes at the proximal site, the crystal structure of Hfq^{Ec}-AU_6A complex is shown in **Figure 1.1D** (17). The AU_6A RNA segment was derived as the Hfq binding sequence in DsrA (23). To better explain the binding modes of each nucleotide to Hfq^{Ec}, the RNA sequence AU_6A is renamed as $A_1U_{(1-5)}A_2$ that indicates the position of each nucleotide with respect to the 5'-end. As shown in **Figure 1.1D**, the 5'-A_1 is protruding outward of the proximal pore of Hfq. U_1 is inserted into a binding pocket that is adjacent to the N-terminal helix, that stacks between F42 and Q41' that also forms H-bonds with highly conserved D9 and K56 residues. U_2 and U_5, floats around the proximal face of Hfq with little or no direct interactions. The remaining U3, U4, U6 and A2 nucleotides forms hydrophobic staking and polar interactions with residues Q41 and F42 side chains that are buried in the central pore of Hfq. This mode of interaction was similar to the previously described AU_5G binding to Hfq^{Sa} (16). Overall, these observations indicate that the A_1, U_2 and U_5 nucleotides to be more accessible to solvent or RNA interactions. The proximal site binding of U_6 RNA to *S. typhi* Hfq was indicated by Sauer and Weichenrieder (22). Here a model for 3'-end recognition of RNAs with U-rich terminals is presented where the 3'-OH group specifically H-bonds to the δN of H57. Terminal U-

stretches are commonly found in many RNAs as a result of Rho-independent terminations.

The above observations indicate the diverse RNA sequence specificities and binding modes to Hfq. Nonetheless, the primary function of Hfq is to promote sRNA-mRNA base pair formation by using its distal and proximal RNA binding surfaces. In fact by superimposing the two crystal structures from *E. coli* Hfq bound to A_{15} RNA and S. aureus Hfq bound to AU_5G RNA it was shown that both RNAs can co-exist within Hfq with no significant sterical constraints. This ability of Hfq to accommodate two RNAs simultaneously and independently is important in understand the functional role of Hfq in gene regulation.

Much debate has focused on how recognition between mRNAs and sRNAs occurs if RNAs are bound on opposite sides of an Hfq hexamer. Initial models proposed based on gel shift data indicated a 2:1 (Hfq_6:RNA) stoichiometry between Hfq and RNA (24,25) implying involvement of multiple Hfq molecules during mRNA-sRNA base pairing. Furthermore, Wang et al., claims that in the AU_6A-Hfq^{Ec} complex mentioned above (**Figure 1.1D**) the 5'-Adenosine is accessed by the previously described A-site within the distal site of another Hfq hexamer sandwiching the RNA. However, it remains to be seen whether similar conformations are observed in the presence of competing RNAs that mimic physiologically relevant Hfq limiting environments (20,26,27). Isothermal titration calorimetry and recent mass spectrometric data however suggested a 1:1 stoichiometry for RNA binding to Hfq (21,28). Also recent studies has shown that



E

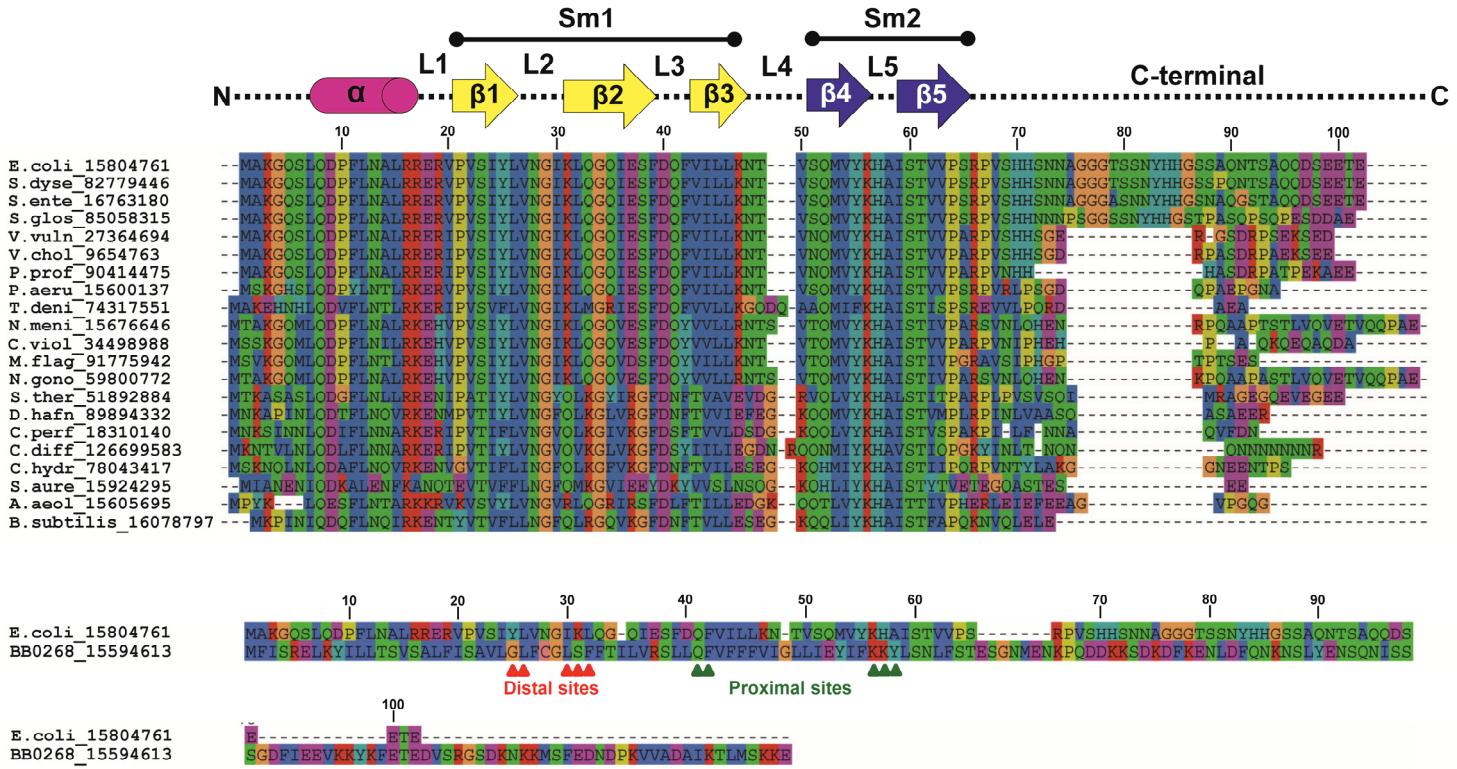


Figure 1.1. Structural properties and RNA biding modes to Hfq. (A) The functional hexameric core of Hfq (22). The assembly of 6-subunits to form the donut shaped functional core of Hfq. For clarity each subunit is colored differently. (B) An isolated monomeric unit from the Hfq core is used to illustrate the functional domains of Hfq. Each protomer consists of an α-helix (pick), five β-parallel sheets (yellow and blue) followed by the C-terminal. Also shown in the isolated subunit is the formation of the nucleic acid binding Sm domains. The β-sheets indicated in yellow represents the Sm1 motif that is comprised of β_1,β_2 and β_3. This forms the distal RNA binding site of Hfq. The β_4 and β_5 sheets form the proximal RNA binding site of Hfq (blue). (C) A_{15} RNA binding to the distal site in Sm1 motif of Hfq^{Ec} (14). One of the six tripartite binding repeats are indicated A-adenosine site, R-purine site and N-nonspecific site. (D) AU_6A RNA bound to the proximal biding site via the Sm2 motif of Hfq^{Ec} (17). (E) Multiple sequence alignment of 21 bacterial Hfq variants to show the sequence conservation (top). Both organism and accession numbers that represents each Hfq protein is also listed. The corresponding secondary structural motifs α-helix, β_{1-3} (Sm1), β_{4-5} (Sm2) and the C-terminal regions are indicated above the amino acid sequences. Species (from top): *escherichia coli, shigella dysenteriae, salmonella enteric, sodalis glossinidius, vibrio vulnificus, Vibrio cholerae , photobacterium profundum 3tck, pseudomonas aeruginosa, thiobacillus denitrificans, Neisseria meningitides, Chromobacterium violaceum, Methylobacillus flagellates, Neisseria gonorrhoeae, symbiobacterium thermophilum, desulfitobacterium hafniense, clostridium perfringens, Clostridium difficile, Carboxydothermus hydrogenoformans, Staphylococcus aureus, aquifex aeolicus and Bacillus subtilis.* (bottom) CLUSTALW alignment of *Borrelia burgdorferi* hypothetical protein BB0286 and Hfq^{Ec} to show sequence conservation between the two. Important amino acid resides that are essential for nucleic acid binding within the distal and proximal surfaces of Hfq^{Ec} is also indicated (29).

mRNAs such as *fhlA* and *glmS* can interact with both RNA binding surfaces of Hfq (30) and correlates with the structural evidences indicated above (14). Given the intense competition towards Hfq in vivo (26), it is beneficial if RNP complex formation can be driven by a single Hfq hexamer. However, the exact pathway towards RNP complex formation and the extent of Hfq participation in this process in unclear.

1.3.2 Conservation of Hfq among the bacterial kingdom

Hfq is widely conserved among the bacterial kingdom. Hfq variants from several bacterial species are aligned to show the sequence conservation (**Figure 1.1E** top panel). It is evident that the ~67 amino acids that comprise the Sm1 and Sm2 cores are fairly conserved implying similar nucleic acid binding modes and specificities throughout the kingdom. Also the recently identified non-canonical Hfq variant from *Borrelia burgdorferi* is aligned with Hfq^{Ec} to show the degree of conservation between the two variants (**Figure 1.1E** bottom). In this work by Lybecker et al., the BB0286 hypothetical protein in *B. burgdorferi* was identified as a functional homologue for Hfq (29). As indicated in **Figure 1.1E**, the Sm cores in Hfq^{Bb} was homologues to Hfq^{Ec} and other classical Hfq variants with some of the critical nucleic acid binding amino acids fairly conserved within the distal and proximal sites (29). Furthermore, Hfq^{Bb} was able to cross-compliment for Hfq^{Ec} in regulating $rpoS^{Ec}$ expression. This shows the ability of Hfq variants to recognize target RNAs across the bacterial kingdom despite the evolutionary distance between species. Similarly, Hfq mediated gene regulation was defined in the Gram-positive firmicute *Listeria monocytogenes* (31)

The C-terminal region of Hfq on the other hand has much less conservation across species. In general it has been observed that lengthy C-terminal tails are found

exclusively in γ- and β-proteobacteria while species belonging to the α-proteobacterial family lack longer extensions (32). The roles and the requirement of the C-terminal domain for gene regulation is a controversial topic in the field (15,33). Recent structural characterization of a full length Hfq^{Ec} shows that the C-terminal region protrudes away from the hexameric core, resembling features of intrinsically disordered proteins (IDP). IDPs are known to act as a landing pad for many proteins and nucleic acids that promotes intermolecular contacts (15). Beich-Frandsen et al. propose that the C-terminal acts as an RNA binding domain based on data obtained using synchrotron radiation circular dichroism (SRCD) spectroscopy. Conversely, as described in the following sections, Hfq has shown to interact with a variety of protein components that makeup the RNA degradosome. Since the absence of any indications for protein-protein interactions that occur at the Sm core, these protein interactions have long been predicted to take place through the C-terminal extensions. Nonetheless, significant experimental support for this hypothesis has not been materialized. Recent work on the GlmZ/GlmY-*glmS* system showed that the presence of the C-terminal was essential in assembling a RNA processing complex that involved a predicted NTPase YhbJ and PNPase (**Chapter 3**). Furthermore these data support a model where the requirement of the C-terminal region of Hfq was conditional, depending on the mode of regulation. In order to fully understand the functions of the C-terminal region, the potential RNA and or protein interaction domains needs to be identified.

1.4 Nature of Hfq dependent sRNA networks in bacteria

Recent advances in co-immunoprecipitation, microarray and deep-sequencing approaches have unveiled a collection of Hfq associated mRNAs and a number of

sRNAs to a lesser extent compared to the latter (34,35). The relative populations of mRNAs and sRNA identified in these global searches were consistent with the previous work that had been performed on sRNAs such as OxyS (36) and DsrA (37,38) where both sRNAs were shown to effect expression levels of multiple genes directly or indirectly. At present a plethora of sRNA networks has been identified that modulate the expression of multiple genes in response to cellular processes in *E. coli* and *Salmonella*.

A collection of Hfq dependent sRNA networks that operate in *E. coli*, *Salmonella* and *S. aureus* are shown in **Figure 1.2A**. A detailed review on each network is present elsewhere (39,40). However, to understand the nature of these regulatory nodes we can focus on one examples network that revolves around RybB sRNA which is essential in envelope homeostasis in *Salmonella*. As shown in **Figure 1.2A**, RybB modulates expression of multiple genes including a variety of outer membrane proteins (OMPs) and a few transport/channel proteins (FadL, PagC and Tsx) (41). This particular node has three apparent regulatory layers. The first layer operates during bacterial envelope stress that is caused by the mis-folding of OMPs in the periplasm where the alternative sigma factor σ^E mediates the synthesis of RybB that then limits further accumulation of OMPs in a larger scale. The obvious advantage visible in this sRNA network is the ability to rapidly control multiple genes that perform similar functions during stress by only synthesizing a single sRNA. Second, multiple sRNA networks can overlap to regulate identical genes only to have a differential level of modulation between the two sRNAs. For instance, the sRNA MicA also regulates OmpA under similar stress conditions but to a lesser extent relative to RybB. Third, RybB has been shown to

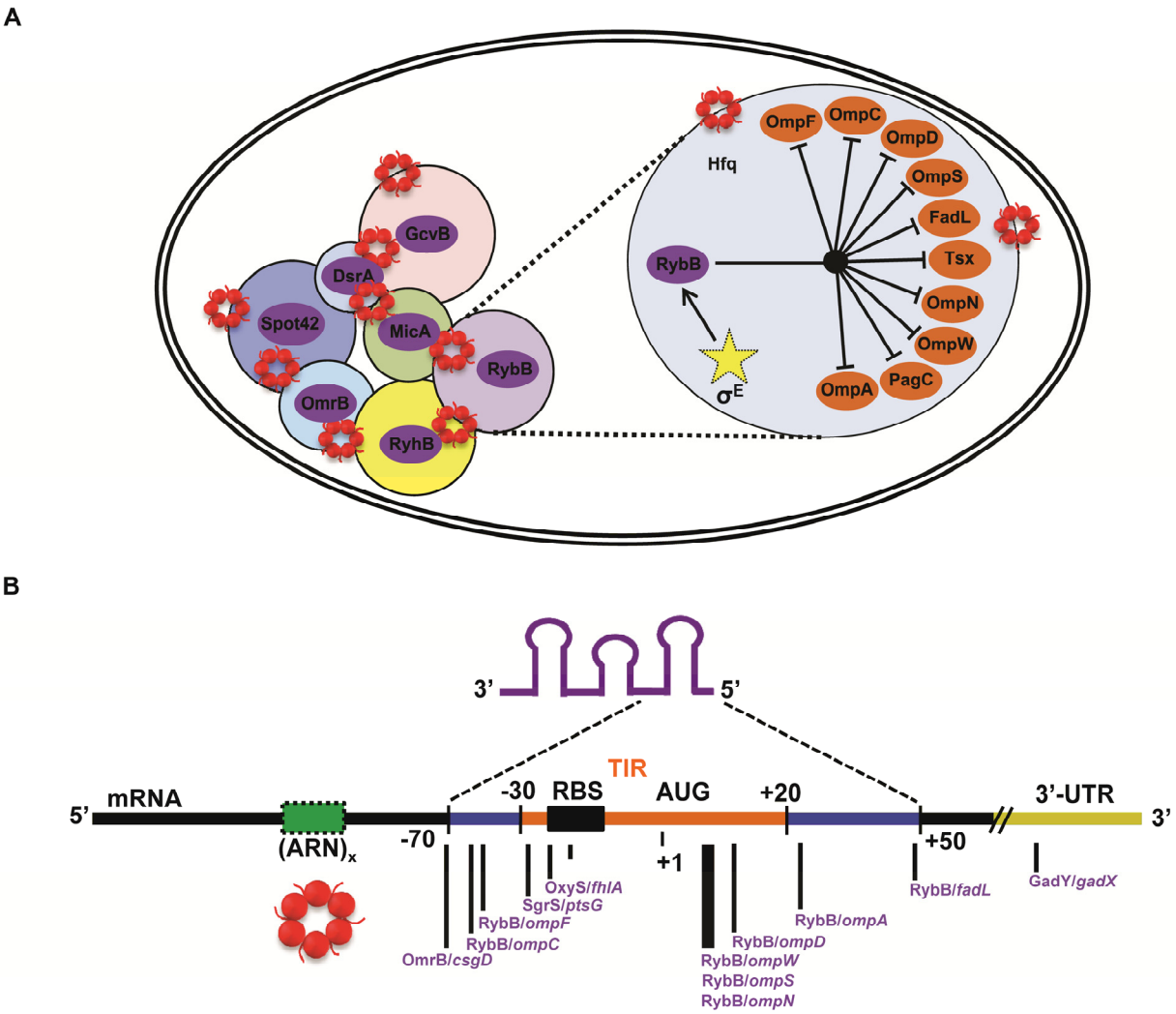


Figure 1.2 Features of sRNA-mRNA interactions and networks. (A) Representation of sRNA-mediated regulatory networks in bacteria. (B) The typical span of sRNA binding functional elements with mRNAs. sRNA binding region within the mRNA with respect to the translation start site according to experimental evidences. Often this this window spans between -70 to +50 (blue). The classical translation initiation region (TIR) is indicated in orange that generally spans though -30 to +20 according structural data. Although a majority of sRNA uses the TIR to prevent translation initiation, recent observations have widened the sRNA effecter window with mRNAs. Several sRNA-mRNA interactions are indicated with respect to the sRNA binding site within the mRNA as examples.

function as part of the surveillance system in un-stressed cells, where Δ*rybB* stains overproduce OMPs that massively activates σ^E (41,42). This implies that, separate form OMP-stress signals the levels of RybB RNAs are maintained as part of a housekeeping system.

The ability of sRNAs to control global gene expression depends on its ability to recognize diverse mRNA sequences using short stretchers of base complimentarity that either inhibits translation initiation and/or modulate mRNA stability. As indicated in **Figure 1.2B,** RybB interaction with target mRNAs spans from the far leader region (*ompC*, *ompF*) to the TIR (*tsx*, *ompW*, *ompS*, *ompN*, *ompD*) and into the deep coding sequence (*ompA*). In an intriguing study Papenfort et al. showed the presence of a 'seed' region at the 5'-end of RybB sRNA that interacts with mRNAs, a concept parallel to eukaryotic microRNAs (43). Furthermore, in the above mentioned study, fusion of the 5'-seed sequence to an unrelated sRNA was sufficient to induce regulation of RybB dependent mRNA targets. Conversely, sequence conservation in mRNAs was also observed where the RybB interaction potentially takes place implying the intertwined nature of these sRNA-mRNA interactions in these regulatory networks. Based on this evidence, one can imagine the sRNA target space to be much wider than anticipated due to the reduced complimentarity that is actually required for mRNA recognition. Intriguingly, in species such as *Burkholderia cenocepacia* multiple Hfq proteins have been identified that could assist regulation. Similarly, in *E. coli* an Hfq like protein ProQ was discovered that has a C-terminal domain structurally similar to Hfq (44). These evidences may suggest the potential magnitude of sRNA networks that operates in

bacterial systems and the ability to substitute for Hfq using Hfq-like proteins in multiple stress conditions.

1.5 Hfq mediated gene regulation: the processes and outcomes

A simplified model that indicates the processes that underlie Hfq mediated gene regulation is shown in **Figure 1.3.** A typical sequence of events that is often triggered by a stress signal includes:1) sRNA synthesis; 2) RNP complex formation followed by gene regulation and finally; 3) the recycling of Hfq. In this section, each of the aforementioned basic processes that take place within the lifetime of a stress response is described in detail.

1.5.1 sRNA-mRNAs interactions

Trans-acting sRNAs control and propagate a variety of gene regulatory networks compared to its cis-acting counterparts in bacteria due to their flexibility in target recognition, where it requires only a short stretch of complimentarity with a target mRNAs. Typical sRNA interaction sites with mRNAs often but not necessarily coincide the ribosome binding site (RBS) located within the translation initiation region (TIR). According to structural and biochemical data, the TIR region spans between -30 to +16 nt positions with respect to the translation start site a region which steric hindrance has been shown to prevent translation initiation (**Figure 1.2B**) (45,46). Trans acting sRNAs typically target the TIR in mRNAs to modulate protein expression, although this window of effecter sRNA binding regions within mRNAs is significantly wider than the classic definition given above. As indicated in **Figure 1.2B**, Hfq mediated sRNA recognition spans between -70 to ~+50 nts with respect to the translation start site within mRNAs (2,43,47,48) where the sRNA interaction also triggers rapid turnover of translationally

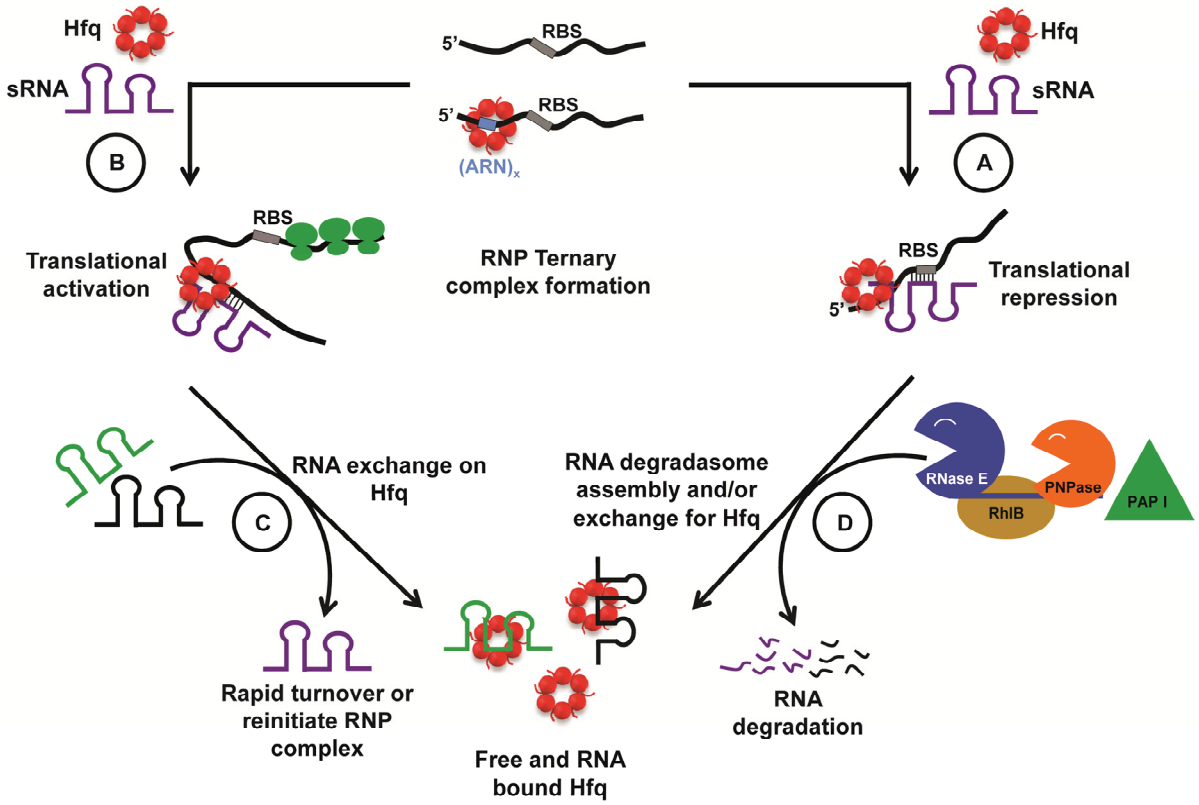


Figure 1.3 Outline of events and pathways during Hfq mediated gene regulation. Steps A/B: Hfq loading into RNAs to facilitate the sRNA-mRNA duplex formation. Prior to sRNA-mRNA base pairing RNAs need to bind Hfq. In sRNAs Hfq is presumed to bind to either AU-rich sequence or A/U stretches present at the 3'-end. mRNAs abundantly use upstream $(ARN)_x$ motifs to bind to Hfq. Often the sRNA-mRNA complex generates two possible outcomes which is translation activation (step B) or translation repression (step A) (also see text for details). Upon RNA duplex formation, in translationally repressive systems the RNA degradosome is often activated that degrades both the sRNA and mRNA to irreversibly regulate gene regulation (step D). In translationally active complexes the need for Hfq is diminished as the RNA duplex is formed. This enables competing RNAs to acquire Hfq from these complexes.

repressed sRNA-mRNA complexes (step **A** in **Figure 1.3)**. In certain cases the sRNA-mRNA complex can also serve as a deterrent for degradasome formation. For example, the sRNA GadY interacts at the 3'-UTR of the transcriptional regulator *gadX*, that increases the stability of the mRNA, up-regulating protein expression in the stationary phase that in turn stimulates synthesis of its effecter genes such as GadA and GadB that code for glutamate decarboxylases (49). sRNAs can also positively modulate protein synthesis. In these cases, sRNAs do not necessarily bind within the TIR of the mRNA, but often interact at leader regions that relieve structural motifs that inhibit ribosome accessibility to the mRNA (step **B** in **Figure 1.3)**. Examples that belong to this class include DsrA/*rpoS* (50), RprA/*rpoS* (51), and GlmZ/*glmS* (52) where the sRNA interaction in the presence of Hfq triggers structural changes in the mRNA that activates translation.

1.5.2 Hfq loading to RNAs and duplex formation

In order to perform regulation both the sRNA and mRNA need to interact with Hfq to yield a ternary complex. Questions that remain regarding this process is, whether a particular order is present in RNA loading during recognition and the nature of Hfq binding elements that are presented by either RNA to specifically interact with Hfq. Based on experimental evidences at least two potential models can be envisioned during RNP complex formation. First, a sRNA centric pathway indicates that the sRNA-Hfq complex is formed first that then seeks for its target mRNA(s). So far, a common Hfq binding motif in sRNAs is unknown. Usually, sRNA binding to Hfq has been implicated to AU rich sequences that are adjacent stem regions (23), and also to 3'-polyA and polyU regions that are shown to be present in certain sRNAs. PolyA

stretches are usually generated by polyA polymerase (PAP I) as part of the RNA turnover machinery (53,54) and 5 to 6 uridines are generally present in most sRNAs that results from Rho-independent terminations (55). Since both polyA and polyU sequences are shown to bind Hfq with high affinity (14,16), it is plausible that sRNAs can load Hfq at these regions. Given the dynamic nature of sRNAs and their shorter life times compared to mRNAs, Hfq binding to sRNAs also helps evade the RNA degradasome. However, given the broader sequence specificities found, it is unclear whether a common Hfq binding sequence motif can be presented in sRNAs that would both initiate gene regulation and help RNA stabilization.

The second model for RNA loading indicates formation of an mRNA-Hfq complex that recruits its cognate sRNA upon its synthesis. Hfq binding sequence elements in mRNAs are more defined than it is for sRNAs. It was recently shown in systems such as *rpoS, fhlA, rprA* and *glmS* that Hfq specifically interacts at $(ARN)_x$ sequence motifs that are present upstream of the ribosome binding site (RBS) (30,56,57). Biophysical characterization on these systems showed increased association rates for RNP complex formation, stimulated by Hfq binding to the upstream $(ARN)_x$ motifs (30,57). When $(ARN)_x$ regions were mutated in *rpoS* and *glmS,* Hfq was unable to regulate gene expression when cognate sRNAs (DsrA and GlmZ respectively) were introduced in vivo, implying the significance of these elements for regulation (56). The high incidence of these motifs has also been projected in a variety of bioinformatic screens as well as in SELEX experiments (14,58). Furthermore, in the OxyS-*fhlA* system, it was shown that *fhlA* wrapped around Hfq interacting with both RNA binding surfaces. Mutations at either distal (Y25A) or proximal (K56A) regions of Hfq, that has been shown to be detrimental

for nucleic acid binding, caused a ~18 fold reduction in the Hfq-*fhlA* complex (30), validating the 'wrap-around' mode of interaction with Hfq.

If mRNAs interact with Hfq using both its RNA binding faces, how does sRNA-mRNA base pairing occur? This question may be answered by considering the dynamic nature of these interactions. sRNAs are transient species that are synthesized upon stress signals where the copy numbers vary between ~10 to 1000 (59,60). In order to attain biologically relevant stress response times (~1-2 mins) (41,61,62), sRNAs must rapidly find its target mRNAs before encountering the RNA degradosome. However, if the mRNA-Hfq interaction is already pre-formed, it only requires the sRNA to complete the RNP complex providing the means for a rapid response. This would suggest that the pre-formed mRNA-Hfq complex would have a way to transiently sample incoming sRNAs to find its cognate partner. Previous work on *fhlA* has implicated a dynamic equilibrium between the mRNA-Hfq complex at both distal and proximal RNA binding sites, that could potentially help the entry of sRNAs into Hfq (30). Furthermore, in the case of *rpoS* it was shown that, Hfq binding to *rpoS* triggered a structural change in the mRNA that was required to favorably interact with DsrA (63). Therefore if Hfq "catalyses" an apparent induced-fit of the mRNA to its sRNA, it would be beneficial for the mRNA to load into Hfq prior to sampling sRNAs to promote 'rapid' RNA duplex formation.

On the other hand during stress, due to the higher concentration of sRNAs, the equilibrium of cellular Hfq may be greatly driven towards the sRNA-Hfq complex (64). This sRNA-Hfq complex can lead to several outcomes. First, since sRNA binding to Hfq is necessary for its stable accumulation, induction of one sRNA might reduce Hfq

access to other sRNAs. This would diminish gene regulation carried out by other sRNAs (26,27). Second the sRNA-Hfq complex may seek its cognate mRNA partner to promote gene regulation. In theory the sRNA-Hfq complex may encounter a variety of mRNAs before forming the relevant RNP complex. In addition to the abovementioned sRNA- and mRNA-centric models, one could also propose a pathway in which a combination of both modes can be utilized during RNP complex formation.

1.6 The aftermath of Hfq mediation

1.6.1 Hfq recycling

It is clear that recruiting Hfq into relevant RNAs is the "process limiting" step in successfully promoting gene regulation in bacteria. Therefore, it is crucial that the cellular machinery utilizes Hfq in an efficient manner. Previous studies have shown that cellular Hfq can be easily sequestered by over expressing Hfq binding RNAs, and the need for a coordinated expression of matching sRNA/mRNA partners for the proper functioning of the regulatory system (20,60). These observations imply the limited availability of Hfq for RNA binding, and the need to recycle Hfq after signal identification. In fact, it was shown that the affinity of Hfq for the DsrA-*rpoS* complex was much lower than for individual RNAs (24,57) implying that once the RNA duplex is formed the need for Hfq is minimal and can be easily driven away by competing RNAs (pathway C in **Figure 1.2**). An attractive model for Hfq recycling was recently proposed by Fender et al (19). In this model, RNA binding was attributed to a single or multitude of Hfq subunits that is constantly competed away by RNAs binding at the rest of the Hfq RNA binding sites. This intense competition for Hfq at a subunit level implies the propensity for Hfq exchange among RNAs depending on their cellular concentrations.

1.6.2 Hfq mediated RNA turnover

In many cases of Hfq mediated gene regulation, sRNA base pairing to its cognate mRNA is often followed by RNA decay (pathway D, **Figure 1.3)**. In *E. coli*, RNase E and PNPase is considered to be the primary components of the RNA degradosome that also includes the RNA helicase RhlB and enolase. Degradation of the RNA duplex insures irreversible gene regulation as well as provides means to recover Hfq from the RNP complex.

RNase E is a large multi-domain protein that consists of 1061 amino acids (65). The catalytic core of this enzyme is harbored at the N-terminal region while the C-terminal domain exists as a long stretch of unstructured protein (> 500 amino acids) with no enzymatic activity (66). Two RNA binding domains and three protein binding sites have been localized to the non-catalytic region of RNase E, and involved in the assembly of the RNA degradosome.

As described above, RNase E mediates RNA turnover of sRNAs that are uncoupled to a target mRNA (67-69) and the sRNA-mRNA duplexes that inhibits translation (70-74). A clear sequence specificity for RNase E cleavage has not been identified, although in a few example systems it was shown to cleave at AU-rich stretchers (75). It has been observed that the rate of RNase E cleavage is ~30 fold faster in the presence of 5'-monophosphorylated RNAs (76) and that the initial cleavage of an RNA is considered to be the rate limiting factor for RNA turnover (77). Therefore, any apparent mechanism that protects RNAs from RNase E would have to prevent or delay the initial transcript cleavage. Immunoprecipitation experiments have consistently shown that an Hfq-RNase E complex coexists as part of the RNA degradosome (74,78).

Intriguingly, the proposed Hfq binding region in RNase E maps to an RhlB binding domain (79). Deletion of this region in RNase E inhibits mRNA decay in response many sRNAs, implying an essential partnership with Hfq during RNA turnover.

A tentative model proposed for RNase E-mediated RNA turnover is shown in **Figure 1.4** (73,80). The Hfq mediation in RNA degradation is best studied with respect to the sRNA RyhB that influences the stability of ~18 mRNA transcripts (72,81,82) that primarily encodes a set of iron binding proteins. Recent work by Prévost et al. showed how the rate determining initial cleavage by RNase E occurs in RNA duplexes using the RyhB-*sodB* system (73). In this work a specific RNase E cleavage site was detected between +400 and +430 nts deep within the coding sequence of *sodB* and downstream of the RyhB pairing site. This cleavage site was located within an AU-rich sequence motif that was shown to be essential for cleavage by mutational analysis. This result once again emphasizes the propensity of RNAase E to cleave within AU-rich sequences. Similar results were observed for the RNase E degradation of *fumA* and *iscRSUA* transcripts where the initial cleavage site was located far downstream of the sRNA binding site, although the exact break point varied depending on the mRNA. This suggests that the presence of a distal RNase E cleavage site in mRNAs could be a common factor in these pathways.

An active mRNA cleavage model is described in **Figure 1.4A**. In the absence of stress, translating mRNAs are targeted minimally by the RNA degradosome due to polysome formation that masks the distal cleavage sites. However, upon sRNA synthesis, target sRNAs act upstream of the TIR, preventing further access to ribosomes. Once the already loaded ribosomes have cleared the mRNA, RNase E

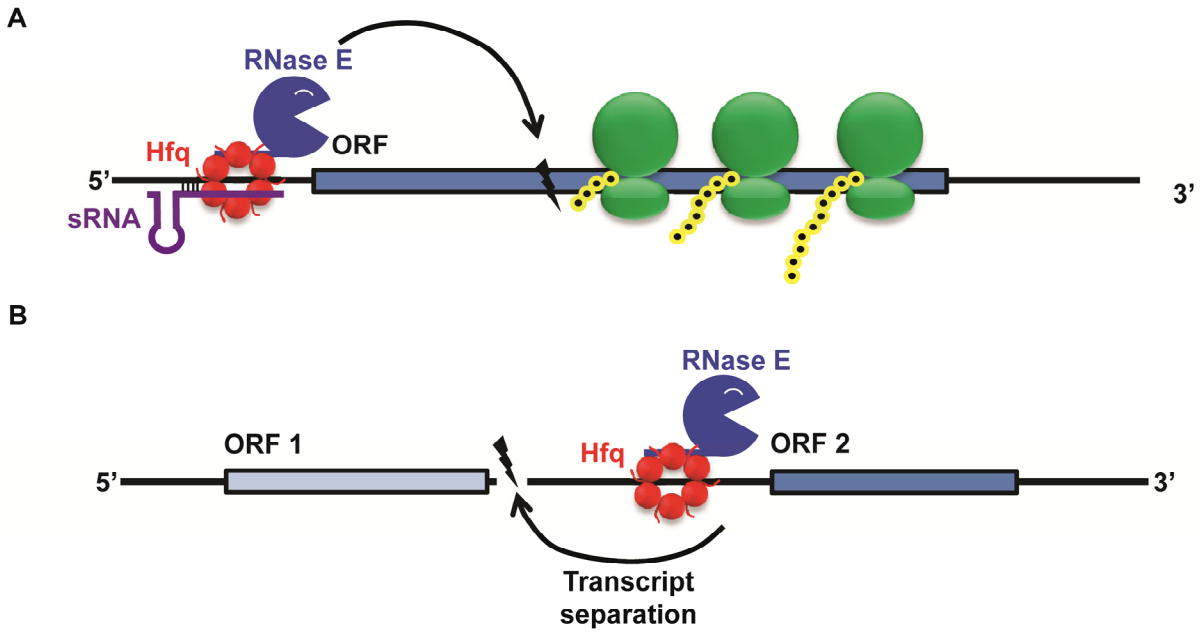


Figure 1.4 Model for RNase E mediated RNA cleavage in bacteria. (A) RNase E cleavage action at distal regions of the mRNA. The Hfq binds to the RNase E C-terminal region. In the presence of sRNA, the RNase E-Hfq-sRNA complex is formed at the sRNA-mRNA pairing site that prevents ribosome assembly. Once translation ribosomes are cleared RNase E directs cleavage of the mRNA. (B) Poly-cistronic transcript separation by RNase E action. Upon transaction of the polycistronic message Hfq is recruited to mRNAs through upstream $(ARN)_x$ elements that also partners with RNase E. This Hfq-RNase E complex then cleaves the upstream message (for more details read text).

cleaves at the distal site to promote rapid degradation of the transcript.

An important question regarding Hfq mediated RNA degradation is the exact role of Hfq in this complex. Since the Hfq-RNase E interaction is already established, and given that an Hfq-mRNA interaction is present at upstream $(ARN)_x$ regions, an upstream Hfq-mRNA-RNase E interaction can be envisioned. If this complex is present, what prevents RNase E from attacking the mRNA? This can be answered once again using the distal cleavage model. In the absence of sRNAs, the rate-limiting distal cleavage may be inhibited due to the aforementioned reasons. However, in the case of stalled translation (due to stalled ribosomes) this upstream Hfq-RNase E complex may target the transcript, if the tmRNA (SsrA) pathway fails to aid the recovery of translation (83). Nonetheless this would imply RNA cleavage in the absence of a sRNA. Whether this sort of surveillance complex can be formed in upstream mRNA reigns needs to be investigated. Another functionally important outcome that can be facilitated by an upstream RNase E-Hfq-mRNA complex is shown in **Figure 1.4B**. In this case RNase E mediated cleavage is used to separate polycistronic messages. This is exemplified nicely in the GlmU-GlmS locus (70). In this case transcription at these loci is governed by a single promoter that produces the *glmU-glmS* transcript. Upon transcription, RNase E cleaves *glmU-glmS* at the stop codon of *glmU* (UAA site). This separates the translational control of *glmS* from *glmU*. The post-cleaved *glmS* transcript is regulated by sRNAs, GlmZ and GlmY in an Hfq dependent manner. No sRNA has been implicated in this RNase E mediated transcript separation of *glmU* and *glmS* (84,85). The presence of Hfq binding $(ARN)_x$ regions in mRNAs may facilitate this mode of RNase E cleavage although this needs to be experimentally validated.

PNPase is another essential component in the RNA degradasome. PNPase is 3'-exoribonuclease that directly binds to the C-terminal region of RNase E (86,87) while interacting with Hfq (53). The principle function of PNPase is thought to be analogous to RNase II and RNase R that acts secondary to RNase E cleavage (65). Nonetheless in a recent study by De Lay and Gottesman, PNPase was shown to be essential for sRNA mediated gene regulation (88). In this work, strains with mutant *pnp* showed decreased half lives in sRNAs such as RyhB, SgrS and CyaR which is the opposite effect one would expect to see when an RNA degrading enzyme is inactivated. Here *pnp* mutants failed to carry out sRNA regulation since they turned over before they could base pair with their target mRNAs. A model that explains these observations is shown in **Figure 1.5A**. As indicated in this model PNPase is introduced as a component that helps RNA turnover in a controlled manner perhaps by providing RNase E limited access to RNAs. Another potential model could be the decreased catalytic activity of RNase E caused by an allosteric effect upon PNPase binding to the C-terminal domain of RNase E. Similar allosteric effects has been reported in the RNA degradosome where the interaction between RhlB and RNase E has shown to significantly stimulate RhlB activity (65). Apart from being a devoted possessive exoribonuclease, in certain cases PNPase has shown to catalyze polymerization using nucleoside diphosphates at low inorganic phosphate (P_i) concentrations (89,90). In *E. coli* addition of heteropolymeric tails in RNAs by PNPase has been mainly observed in absence of the main polyadenylating

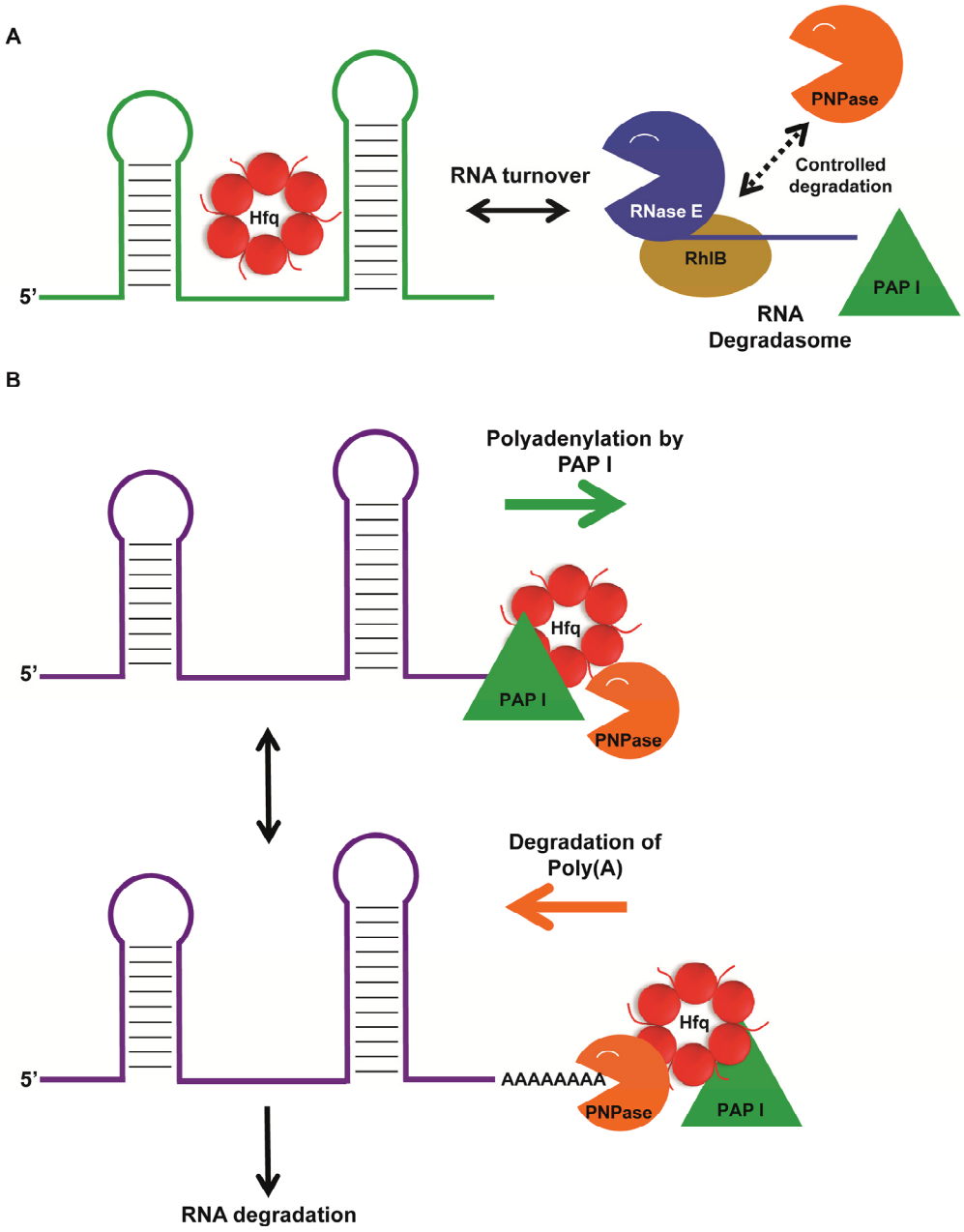


Figure 1.5 Role of PNPase and PAP I in RNA decay. (A) Function of PNPase in the RNA degradosome. (B) Model for PAP I facilitation of RNA degradation by PNPase.

enzyme PAP I (91,92) and in transcripts that are Rho-dependent terminators (93). Also transcripts that are cleaved at polycistronic junctions **Figure 1.4B** can also use this type of polymerization since these messages are akin to Rho-dependent terminations. Recruitment of PNPase to these junctions can once again be facilitated by the upstream Hfq-mRNA complex. However, the presence of such models for PNPase in Hfq mediated stress regulation needs to be experimentally validated.

Apart from the sRNA-coupled degradosome, the other common pathway employed in bacteria is the **polyA polymerase I** (PAP-I) dependent RNA turnover mechanism (**Figure 1.5B**). This mode of decay is widely employed in RNAs containing Rho-independent terminator stems where the single-stranded 3'->5' exonucleases are unable to invade the terminal helix. To overcome this difficulty, PAP I helps decay by adding polyA tails to the 3'-terminus of the RNA that is then degraded by PNPase (77,94-96). In the absence of Hfq, PAP I was unable to efficiently polyadenylate RNAs thereby increasing half-lives of transcripts (53). Whether Hfq facilitation is a result of structural remodeling the RNA or an allosteric interaction with PNPase as presumed above needs to be experimentally validated.

The RNA degradosome also contains a DEAD box RNA helicase **RhlB**. In vitro studies have shown that these proteins have RNA-dependent ATPase activity and ATP-dependent RNA unwinding activity. In *E. coli* five DEAD box proteins are identified out of which CsdA, DbpA and SrmB has been implicated in ribosome assembly (97-99). No functional role has been identified for RhlE. These proteins are generally known to unwind a few RNA base pairs that is often enough to compliment exoribonucleases such as PNPase to invade structured RNAs (65). A recent study indicated that both Hfq

and RhlB interact at an identical location in the C-terminal region of RNase E and that overexpression of RhlB prevented Hfq binding to RNase E that in turn decreased the half life of *ptsG* mRNA in the presence of SgrS (79). As described above, ATPase activity of RhlB is significantly stimulated upon the interaction with RNase E and one would think that this interaction is paramount for RhlB activity (100). Therefore it is unclear how both Hfq and RhlB can be part of the RNA degradosome while competing with each other to interact with RNase E.

This paves the way to a model where the presence of alternative forms of RNA degradosome complexes needs be considered. For instance there is evidence to suggest that the RNA degradosome components are modified during cold shock (101). The DEAD box helicase CsdA which is a cold shock protein that is highly expressed at lower temperatures is recruited into the RNA degradosome. Here, it was shown that CsdA can reconstitute RhlB function in vitro (101). Furthermore, the CsdA binding site in RNase E was shown to be distinct from that of RhlB, implying the possibility of either having two helicases or a combination that lets both Hfq and CsdA to be present as part of the RNA degradosome. This was confirmed by a recent study by Resch et al., where they showed the requirement of CsdA in regulating *rpoS*, which is considered to be the primary cold-shock stress response pathway in bacteria (102). In another study it was reported that the composition of the RNA degradosome in *E. coli* can be modulated depending on cellular conditions. In this work an RNase E pull down experiments were performed in the presence of RraA and RraB which inhibits RNase E by binding into the non-catalytic region. Both inhibitors diminished PNPase levels, while RraB specifically

reduced RhlB binding to degradosome (103). However it was observed that in the absence of RhlB (inhibition by RraB) the degradosome recruited CsdA as a substitute.

Based on observations described above it is clear that the RNA degradosome can modulate its composition based on the regulatory system or availability of proteins. Whether the RNA degradosome can be remodeled in response different stress or cellular requirements needs to be experimentally tested. Another question that needs to be addressed is the absolute role of Hfq as part of the RNA degradosome, specifically whether the RNase E-Hfq complex is distinct from the common RNA degradosome or its composition. This question becomes more relevant as Hfq binding sRNAs such GlmZ and GlmY that regulate the expression of *glmS* has been shown to be processed by a predicted NTPase **YhbJ** (52,104,105). In this case the stability of RNAs also depended on the presence of PNPase. Whether these enzymes are part of another alternate RNA processing complex that also includes Hfq needs to be tested as our data indicates that the C-terminal region of Hfq was essential in this pathway (Chapter 3).

The exact function of **enolase**, as part of degradosome in not clear. Enolase is a glycolytic enzyme that catalyses the reversible dehydration of 2-phosphoglycerate to phosphoenolpyruvate. It has been suggested that enolase could potentially act as a sensory system that can detect the energetic state of cells, which can be later linked to a degradation signal. Nonetheless, no clear experimental evidence is present that direct the need for enolase during RNA turnover.

1.7 Future directions

Based on the experimental evidence gathered so far on Hfq-mediated gene regulation, regulatory outcomes are triggered by the presence of the sRNA-mRNA duplex. Also it is clear, that although considered abundant in cells, Hfq is actually a limiting factor in assembly of these regulatory complexes. Extensive work has been performed to characterize the mode and sequence specificities of RNA binding to Hfq. Also it has been confirmed that RNA degradation takes place secondary to Hfq mediated RNA recognition to irreversibly regulate gene expression.

Despite the tremendous amount of insight that was gathered during the past decade to decipher the mechanism of stress regulation mediated by Hfq, a significant amount of questions still remain that requires to be answered in order to fully understand the functions of this protein. Several RNA binding modes have been put forth based on crystallographic observations that indicate how various RNA sequences can be accommodated at the functional core of Hfq. While these examples provides a great deal of insight into how Hfq operates, a unified model needs to be established that has to take into consideration the broad RNA sequence specificities Hfq possesses. This fact is further epitomized where studies show how Hfq variants can regulate gene expression in across evolutionally distant species (29).

Much of the effort has been directed to understand the nucleic acid binding properties of Hfq while little attempts has been taken to understand the functional roles performed by the C-terminal region of Hfq. Given the recently obtained structural insights on the C-terminal of Hfq (106) it is likely that many protein and RNA interactions potentially takes place during regulation. This would then help explain how some of the

partnerships with proteins such as RNase E, PNPase, PAP I and YhbJ are formed that are shown to be essential in carrying out gene regulation.

Recent advances in global sequencing and bioinformatic approaches have helped unveil a number of Hfq dependent systems that suggests that the number of cellular processes that depend on Hfq could be significantly higher than previously anticipated. Therefore, it is worthwhile to investigate how Hfq can fulfill its supply for demand. Regulation of Hfq is poorly understood in bacterial system despite its abundant usage in various cellular processes. Furthermore, an increasing number of reports indicate the presence of additional or similar chaperone proteins to Hfq that can also be useful in accommodating multiple stress responses.

Finally, Man et al. showed that Hfq can be successfully utilized to perform artificial gene regulation in vivo (107). In this work, an artificial sRNA was constructed that comprised of an mRNA base pairing region, an Hfq binding site and a Rho-independent terminator. This sRNA was not only able to base pair with its intended target but also invoked the RNA degradosome. In a similar study a set of RyhB variants were prepared to understand the sequence-function relationships between sRNA base pairing with its target mRNAs (108). Here they report the sRNA-mediated regulation as a ‘tunable property’ where the sRNA-mRNA interaction can be varied to modulate the degree of repression. These studies shows that, as far as the sRNA-mRNA interaction can be made specifically, Hfq mediated regulation can be constructed in vivo. This is an attractive strategy that can be used for gene silencing in bacteria.

CHAPTER 2‡

An Upstream Hfq Binding Site in the *fhlA* mRNA Leader Region Facilitates the OxyS-*fhlA* Interaction (30)

2.1 Abstract

To survive, bacteria must be able to adapt to environmental stresses. Small regulatory RNAs have been implicated as intermediates in a variety of stress-response pathways allowing dynamic gene regulation. The RNA binding protein Hfq facilitates this process in many cases, helping sRNAs base pair with their target mRNAs and initiate gene regulation. Although Hfq has been identified as a critical component in many RNPs, the manner by which Hfq controls these interactions is not known. To test the requirement of Hfq in these mRNA-sRNA complexes, the OxyS-*fhlA* system was used as a model. OxyS is induced in response to oxidative stress and down regulates the translation of *fhlA*, a gene encoding a transcriptional activator for formate metabolism. Biophysical characterization of this system previously used a minimal construct of the *fhlA* mRNA which inadvertently removed a critical element within the leader sequence of this mRNA that effected thermodynamics and kinetics for the interaction with Hfq. Herein, we report thermodynamic, kinetic and structural mapping studies during binary and ternary complex formation between Hfq, OxyS and *fhlA* mRNA. Hfq binds *fhlA* mRNA using both the proximal and distal surfaces and stimulates association kinetics between the sRNA and mRNA but remains bound to *fhlA* forming a ternary complex. The upstream Hfq binding element within *fhlA* is similar to $(ARN)_x$ elements recently

‡ Portions of this work were previous published in PLoS One (30) under a Creative Commons BY 3.0 license. All rights to this work are retained by the authors and any reuse requires permission of the authors.

identified in other mRNAs regulated by Hfq. This work leads to a kinetic model for the dynamics of these complexes and the regulation of gene expression by bacterial sRNAs.

2.2 Introduction

Small non-coding RNAs (sRNA) mediate gene regulation in both bacteria and eukaryotes (5,109,110). Bacteria commonly employ sRNAs during stress responses, allowing them to survive when exposed to suboptimal growth environments (111). Two main classes of sRNAs exist in bacteria, *cis* and *trans*-encoded variants. Cis RNAs derive from the same genetic locus as the regulated message but are transcribed from the antisense strand; thus exhibiting perfect complementarity with their target. These RNAs are known to control regulatory pathways such as transcriptional attenuation, RNA processing and decay, and translation initiation (6,112). Unlike cis-acting sRNAs, the trans-acting sRNAs are expressed from genetic loci different than their targets and interact using imperfect base pairing. These sRNAs often require accessory proteins such as Hfq for activity.

Hfq is a homohexamer that belongs to the Sm/LSm family of RNA binding proteins (16,113-115). It typically varies in length between 70 and 110 amino acids in bacteria, and is highly abundant where an estimated 10,000 hexamers present in the cytoplasmic fraction, often in association with ribosomes (113,116). Hfq is mostly conserved among the bacterial kingdom with more than 3000 homologs currently annotated in genomic databases. Mutational studies in *E. coli* and other organisms have shown that strains lacking Hfq exhibit pleiotropic effects such as decreased growth rates, increased stress sensitivity (UV, oxidative and cold shock), ineffective tRNA

maturation and mini-cell formation (117-119). In addition, it was demonstrated that reduced virulence was observed in the absence of Hfq for a variety of bacterial pathogens (8,9,120-122).

Hfq is known largely for its role in post-transcriptional gene regulation by facilitating pairing between sRNAs and mRNAs. A common feature in these pathways is the presence of overlapping networks of RNA interactions where one sRNA regulates multiple genes. For example the sRNA RyhB has been shown to regulate *sodB*, *ftnA*, *bfr*, *acnA* and *sdhC* and thus acts as a regulatory node allowing a complex and integrated response to a given growth condition, in this case low iron concentrations (40). Although Hfq has been identified as a critical component in these systems, a common mechanism as to how it facilitates complex formation is not clear.

To further understand the requirement of Hfq during sRNA:mRNA pairing, we have studied the OxyS-*fhlA* system (**Figure 2.1**). OxyS is a regulatory RNA expressed in response to oxidative stress. One of the mRNAs it interacts with is *fhlA,* a message encoding a transcriptional activator for formate metabolism (123). Interaction at two short pairing elements, one at the RBS and another within the coding region, are sufficient to prevent translation of the *fhlA* mRNA. Previous studies in vivo showed that, in the absence of Hfq, OxyS was unable to regulate the expression of *fhlA* (124). Most of the work on this regulatory network used a minimal *fhlA* mRNA construct that was sufficient to interact with its sRNA OxyS. A recent study by Soper et al., however, showed a marked difference in the *rpoS*-DsrA interaction with Hfq when the leader sequence of the mRNA was extended (57). This study showed that Hfq interacted with the *rpoS* mRNA at

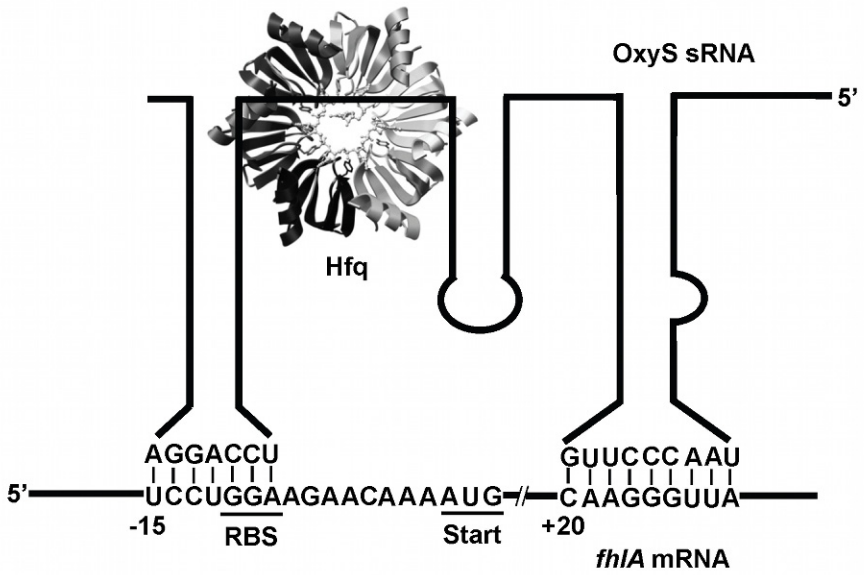


Figure 2.1. Regulation of *fhlA* by sRNA OxyS in the presence of Hfq. Interaction between *fhlA* mRNA and the sRNA OxyS is shown. *fhlA* encodes a transcription factor for formate metabolism. During oxidative stress the sRNA OxyS is expressed and in the presence of Hfq was proposed to form two kissing interactions (123) through the stem loops present in the mRNA and the sRNA. The interaction formed within the 5' leader region of *fhlA* sequesters the ribosome binding site preventing translation. The contact within the coding sequence was shown to be important for efficient gene regulation (123).

a novel $(ARN)_X$ sequence element (125).* To understand whether this recognition element was a commodity among regulatory networks involving Hfq, we tested this hypothesis in the OxyS-*fhlA* system. Here we show that the extension of the *fhlA* leader sequence enhances the kinetics and thermodynamics of *fhlA* associating to Hfq stimulating ternary complex formation with OxyS. Structural probing of the leader sequence confirmed Hfq-binding to the $(ARN)_x$ element at position -76 in the upstream region of the *fhlA* mRNA. Finally, we show that Hfq binds to the *fhlA* leader sequence using both its proximal- and distal-RNA binding surfaces which helps explain the kinetic and thermodynamic properties of Hfq association with mRNAs. These data lead us to revise the model for Hfq-dependent gene regulation to be less sRNA-centric and providing more importance to the manner in which Hfq interacts with a subset of bacterial messages that are subject to regulation by sRNAs.

2.3 Results

2.3.1 OxyS and Hfq interacts with *fhlA* leader construct to form a ternary complex

To test whether the upstream leader region of *fhlA* facilitates the interaction with OxyS and Hfq as previously described for the *rpoS*-DsrA system (125), the leader sequence of *fhlA* was extended from the previously characterized position -53 relative to the start codon, to -136 and -220 respectively. A putative upstream Hfq-binding element was absent in the minimal *fhlA*53 construct but is encoded at position -76 and thus is

* This motif was originally called an AAYAA element by Soper et al. [23] It was subsequently refered to as an $(ARE)_x$ element by Link and co-workers who showed based on crystallographic and biochemical studies the site has broader specificity than AAYAA [28]. Unfortunately, the acronym ARE has already been used for many years to refer to A/U-Rich Elements in eukaryotic mRNAs [24, 25]. We therefore propose calling this sequence motif by the name $(ARN)_x$ to distinguish it from AREs while still retaining the necessary information about the sequence specificity.

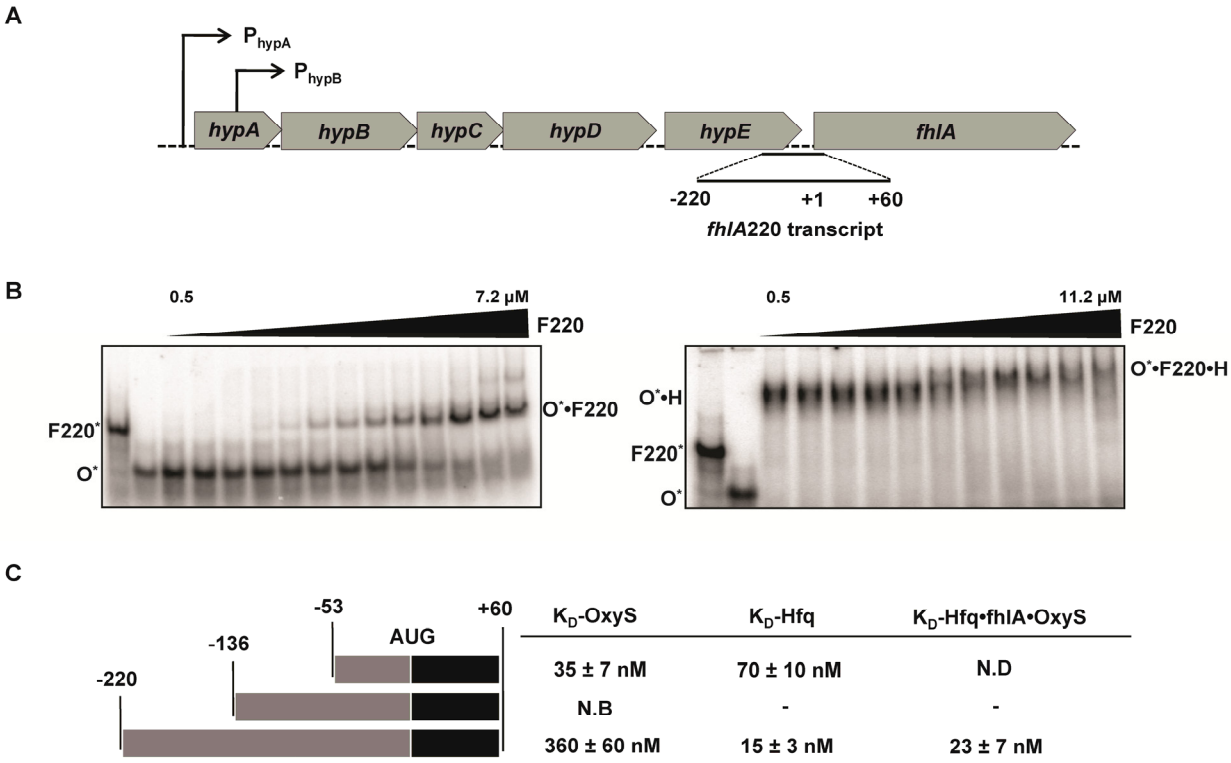


Figure 2.2. The *fhlA* mRNA leader constructs used in this study. (A) The *fhlA* locus. *fhlA* is transcribed by promoters that belong to the family of hydrogenase iso-enzymes specifically by promoters P_{hypA} and P_{hypB} (126). (B) Analysis of binary and ternary complexes of *fhlA*220 mRNA with OxyS and Hfq (also see **Figure 2.4**). Gel shift experiments showing binary complex formation between OxyS (O^*) and *fhlA*220 (F220) (left). UniformLy ^{32}P-labled OxyS (*) was titrated with varying concentrations of *fhlA*220 ranging from 0.5 nM-7.2 µM. Ternary complex formation between OxyS (O), Hfq (H) and *fhlA*220 (F220) (right). The O•H complex was pre-formed by incubating ~1 pmol of [5'-^{32}P] labeled OxyS with 1 µM Hfq. *fhlA*220 (F220) was titrated from 0.5 nM-11.2 µM. (C) *fhlA* mRNA leader constructs used during this study and their affinities. Three *fhlA* leader constructs were tested. *fhlA*53 was previously characterized by Argaman et al. and comparisons were made relative to this construct. A stable OxyS•*fhlA*136 complex was not detected in gel shift assays due to probable mis-folding of OxyS binding elements. The *fhlA*220 leader sequence showed activity, forming stable binary and ternary complexes with OxyS and Hfq.

present in both extended constructs. RNA transcripts are denoted according to the length of their leader sequence: *fhlA*220 has 220 nt upstream of its start codon. In its genomic context, *fhlA is* part of a polycistronic message, transcribed under the control of two promoters P_{hypA} and P_{hypB} (126) (**Figure 2.2A**).

To test the potential of *fhlA* mRNA leader constructs to undergo post-transcriptional regulation, gel mobility shift assays were used to detect binding to radiolabeled OxyS. Equilibrium experiments were performed with *fhlA*136, but neither binary complexes with OxyS nor ternary complexes with OxyS and Hfq were observable using EMSA. For *fhlA*220, on the other hand, both binary and ternary complexes with OxyS and Hfq were stable. The OxyS•*fhlA*220 (O•F220) complex formed with a K_D of 360 ± 60 nM (**Figure 2.2B** and **Figure 2.3A**). This affinity is ~10-fold less than that of the OxyS•*fhlA53* (O•F53) complex where a K_D of 35 ± 7 nM was measured (**Figure 2.2C**). This value compares favorably with the K_D previously reported for the O•F53 complex by Argaman et al. in a slightly different binding buffer (10 mM Tris, 60 mM KCl, 10 mM $MgCl_2$ and 1 mM DTT (pH 8.0)) (123). Reduced affinity between OxyS and *fhlA*220 was surprising, but this phenomenon was also observed for DsrA and RprA, both of which bound extended *rpoS* leader sequences 10 – 20-fold less tightly than to the truncated leader (125,127).

To detect whether *fhlA*220 interacts with Hfq, [5'-^{32}P] *fhlA*220 mRNA was incubated with Hfq at concentrations of 0 nM to 1.67 μM Hfq (hexamer). The Hfq•*fhlA*220 (H•F220) complex was resolved on native gels to provide a K_D of 15 ± 3 nM Hfq hexamer (**Figure 2.3B** and **Figure 2.4**). The F220•H complex formed with ~5 fold

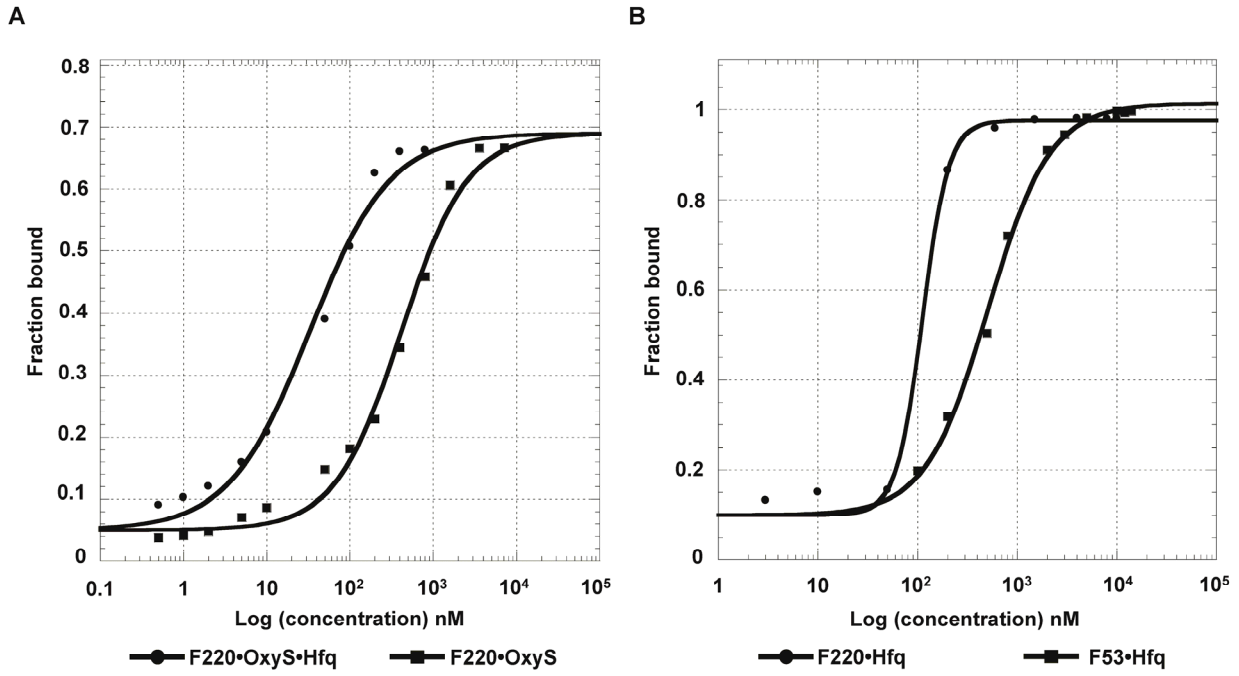


Figure 2.3. Quantitative analysis gel shift assays. (A) Analysis of gel shift assays shown in Fig. 2B. Binding of F220 to OxyS (closed squares) and ternary complex formation between F220, OxyS and Hfq (closed circles). (B) Quantization of thermodynamic constants for gel shifts for Hfq binding to F53 (closed squares) and F220 (closed circles). As described in materials and methods thermodynamic constants were determined by nonlinear least-square analysis fitted to a cooperative binding model.

tighter affinity than F53•H (K_D, 70 ± 10 nM) revealing that the extended structure may present extra Hfq binding elements or a more favorable Hfq-binding motif (**Figure 2.4B**). The OxyS•*fhlA*220•Hfq ternary complex (O•F220•H) was also resolved using gel shift assays when pre-formed [5'-^{32}P] OxyS•Hfq complex was titrated with 0.5 nM - 11.2 μM *fhlA*220 mRNA (**Figure 2.2B**), yielding a K_D of 23 ± 9 nM. Thus the presence of Hfq provides 15-fold tighter binding between OxyS and *fhlA*, a net stabilization of ~1.6 kcal/mol.

2.3.2 *fhlA*220 mRNA interacts with both proximal and distal RNA binding surfaces of Hfq

Hfq binds RNAs using two distinct surfaces – a proximal surface that favors U-rich sequences like those in many Hfq-binding sRNAs (such as DsrA and RybB) and the distal surface that binds polyA RNAs (**Figure 2.5A**) (14,25,116,127). Competitive binding studies were used to investigate *fhlA*220 mRNA binding to Hfq. [5'-^{32}P] H•F220 complex was prepared and then incubated with excess of either a proximal surface binder (DsrA), a distal surface binder (A_{18}) or both simultaneously. The complexes were then visualized by native gel electrophoresis (**Figure 2.5B**). As shown in **Figure 2.5B**, F220•H was not disrupted by addition of DsrA or A_{18} alone. A faster migrating species was observed on these gels with increasing concentration of competing RNAs. These complexes may result from conformational changes of the H•F220 complex when a competitor displaces F220 from one face of Hfq. To displace *fhlA*220 from Hfq entirely, both A_{18} and DsrA had to be present. This experiment clearly shows that *fhlA*220 mRNA interacts with Hfq using both the proximal and distal surfaces. Although such a binding

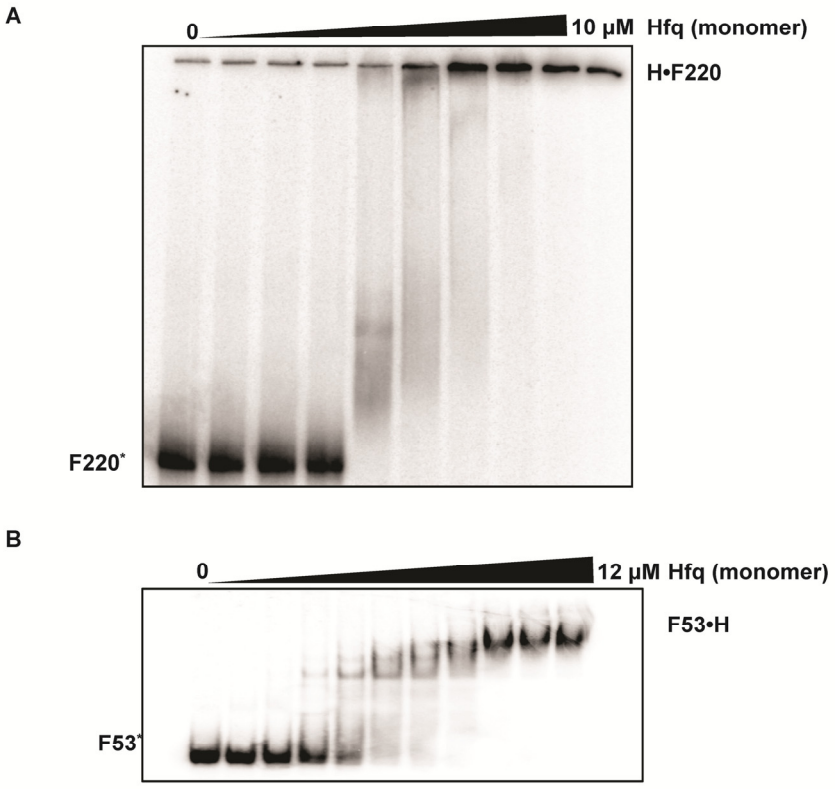


Figure 2.4. *fhlA* interaction with Hfq. (A) Gel shift assay wherein [5'-^{32}P]-*fhlA*220 mRNA was titrated with increasing concentration of Hfq in the range of 0 to 1.67 μM hexamer. (B) *fhlA*53 mRNA interaction with Hfq. Gel shift assay for Hfq•*fhlA*53 binary complex formation

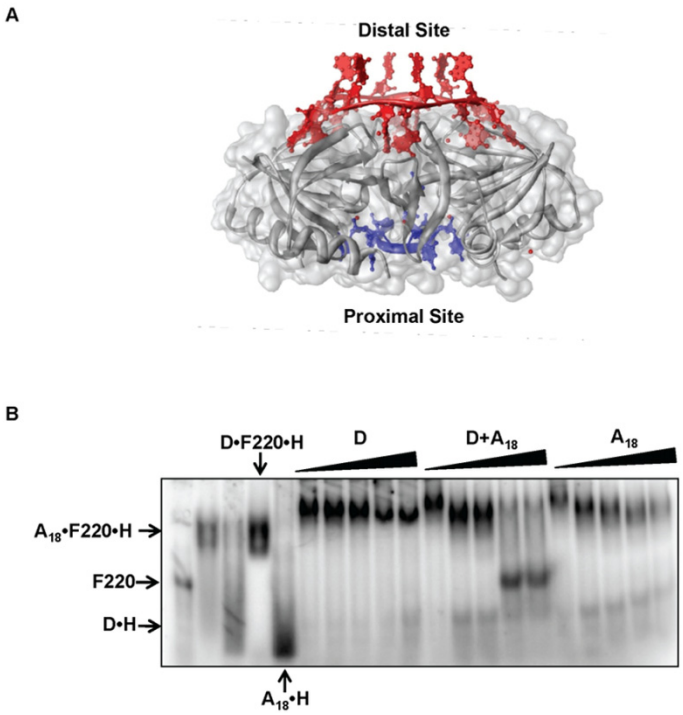


Figure 2.5 Competition binding experiments to determine the Hfq binding surface that interacts with *fhlA*220. (A) Superposition of two Hfq crystal structures that are crystallized in the presence of (AU_5G) RNA at the proximal site (1KQ2) and in A_{15} RNA bound at the distal site (3GIB) (14,16). Superposition of the two crystal structures were performed using UCSF Chimera software (128). (B) The complex between [5'-^{32}P] *fhlA*220 (F220) and Hfq (H) was pre-formed and incubated with increasing concentrations of A_{18} RNA (0 – 30 μM), DsrA (0 – 30 μM) or with both RNAs (0 – 30 μM). DsrA and A_{18} have been previously shown to specifically bind to proximal and distal RNA binding sites of Hfq respectively (129).

mode with Hfq has not been reported, this sort of behavior was previously predicted based on mutagenesis

2.3.3 Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) to probe for RNA secondary structure

SHAPE methodology depends on the sensitivity of the RNA ribose hydroxyl group to react with an electrophile such as *N*-methylisatoic anhydride (NMIA) to form a 2'-adduct. Reactivity at the 2'-ribose OH with the eletrophile depends on the local nucleotide flexibility since a conformationally constrained 3'- phosphodiester has shown to inhibit formation of the deprotonated oxyanion. Therefore by monitoring the relative reactivity of RNA with NMIA one can measure the local structure at every nucleotide position (130-132). A brief overview of a typical SHAPE experiment is shown in **Figure 2.6**. Shown in **Figure 2.6A** is the 'RNA structure cassette' where the RNA of interest is placed between. Each stem-loop in the structure cassette consist the UUCG tetraloop loop sequence that has been shown have robust folding (133) to its designated structure that would prevent the structure cassette from folding into the RNA of interest. The structure cassette is included to perform three specific functions. (1) Primer binding site (green): since the extent of RNA modification is measured by primer extension a common primer binding site needs to be used to initiate reverse transcription. (2) the 3'-linker nucleotides (blue) allow the reverse transcriptase become fully processive before reaching the RNA structure of interest. This linker also avoids non-templated primer extension products from masking structural information at the 3'-end of the RNA. (3) the 5'-linker (red) is used to separate the full length RT product if not would mask structural information at the 5' region of the RNA.

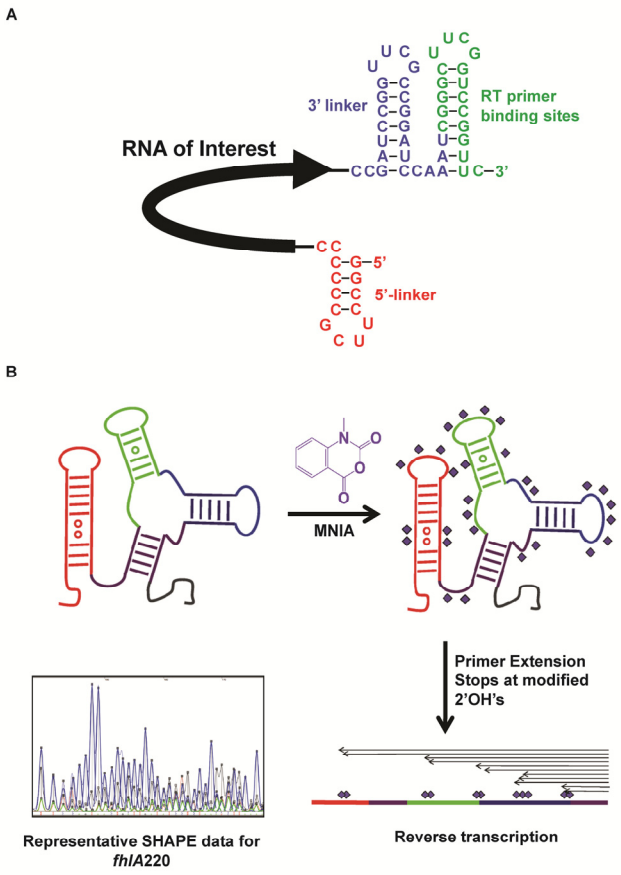


Figure 2.6. Basic outline of a SHAPE experiment. (A) Structure cassette used to position the RNA of interest. Three functional elements are present in the cassette used during SHAPE experiments. (1) Reverse transcription primer binding site (green) that could be used as a common primer binding site for primer extension. (2) 3'-linker (blue): additional nucleotides enable the reverse transcriptase to become fully processive before reaching the RNA interest. (3) 5'-linker (red): this linker is used to separate the RNA from the abundant full length reverse transcription product, where in the absence of such linker structural information on the 5'-end could be lost. (B) Basic steps during a SHAPE experiment. Prior to NMIA modification, the RNA is folded using an appropriate protocol. RNAs are then modified using NMIA. In addition to modified RNAs three additional reactions are revere transcribed. (1) Unmodified RNA, to account for natural RT stops and to control for reaction conditions. (2) and (3) Unmodified RNA samples reverse transcribed in the presence of a single ddNTP nucleotide to create sequencing reactions. Finally, NMIA modified reaction, control RNA, two-sequencing reactions are combined and separated in a DNA capillary sequencer. The data is processed using ShapeFinder® (134) to obtain relative reactivities at each base position. The relative reactivities are then exported into RNAStructure to model the RNA secondary structure (135).

Basic steps involving the SHAPE experiment is shown in **Figure 2.6B.** As indicated the first step involves the modification of the RNA using NMIA in DMSO. Prior to modification the RNAs is folded according to an appropriate folding protocol. The use of appropriate NMIA concentrations and incubation times for modification is extensively described elsewhere (130). Simultaneously, a control reaction is often setup to undergo identical processing but without the modifying agent NMIA. The modified/control reactions are then ethanol precipitated and reverse transcribed using dye labeled primers. Each SHAPE experiment consists of four unique dye labeled reactions. NMIA modified, control unmodified, two unmodified RNAs RT using ddNTPs for sequencing are combined and separated in a capillary DNA sequencer. The data is then analyzed using SHAPE finder®, that aligns, normalizes and measures the extent of modification. The processed data (136) is then used in RNAStructure software to model optimal RNA structures (135). Reports indicate that SHAPE chemistry can provide structural accuracy levels >95% for RNAs (137,138).

2.3.4 Secondary structure analysis of *fhlA*220 mRNA

SHAPE (132,139) was carried out on *fhlA*220 mRNA to investigate structural and functional elements that might be important for RNP complex formation with Hfq and OxyS. **Figure 2.7A** shows the experimental secondary structure for *fhlA*220 mRNA derived from SHAPE constraints, superimposed with the modification intensity data. In **Figure 2.7B**, the previously published secondary structure for *fhlA*53 derived from nuclease digestion data is shown for comparison (123). *fhlA*220 exhibits highly structured regions at both the 5'-region and 3' ends of the construct and a rather flexible region

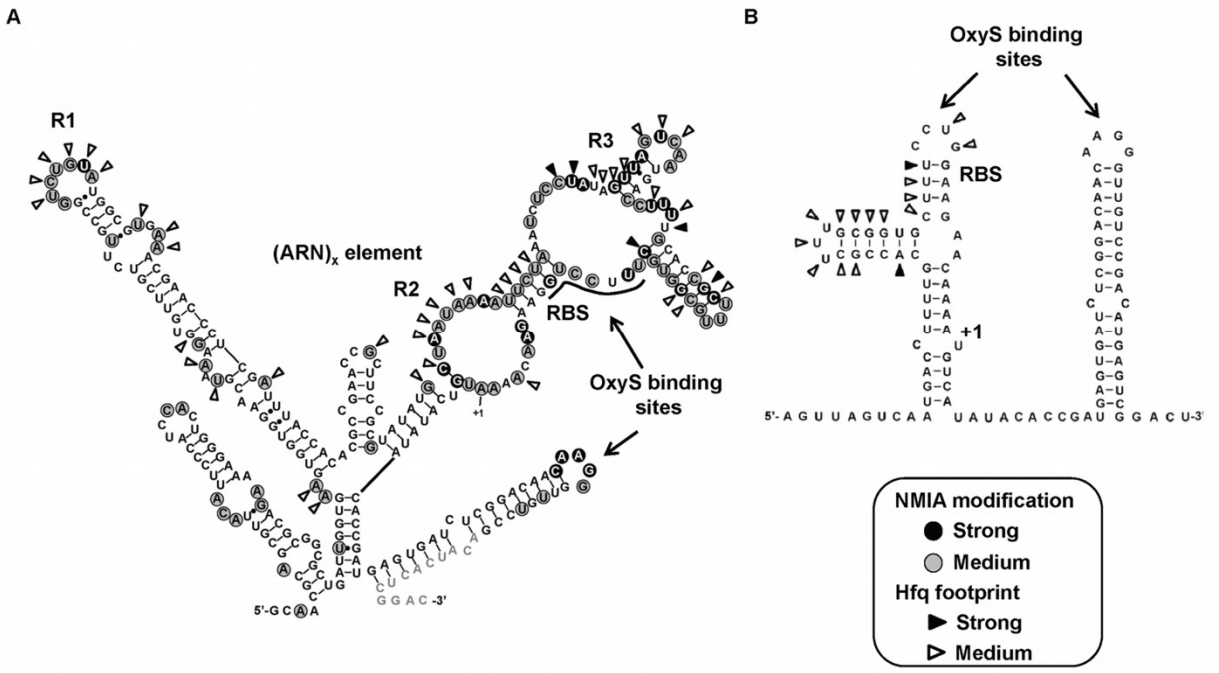


Figure 2.7. Proposed secondary structure models for *fhlA*220 and *fhlA*53. (A) SHAPE-derived secondary structure model for *fhlA*220 mRNA. The NMIA reactivities are depicted on each base position of the proposed structure. SHAPE reactivities above 0.7 are depicted by dark circles; reactivities 0.2 - to 0.7 are shown in gray circles and reactivities below 0.2 are un-circled. Bases shown in gray text annotations were not analyzed. Hfq footprinting data measured by SHAPE in the presence and absence of 1 μM Hfq are superimposed on the structure model for *fhlA*220. Closed wedges are base positions with strong Hfq footprints, where the relative reactivity was > 0.7 and open wedges represent relative reactivities between 0.3 and 0.7. OxyS binding sites, RBS and the $(ARN)_x$ elements are also shown. (B) Secondary structure for the previously characterized *fhlA*53. To identify Hfq binding sites of *fhlA*53, terbium mediated footprinting was performed. Strong Hfq footprints at base positions are shown by closed wedges and medium reactivities with Hfq are shown by open wedges.

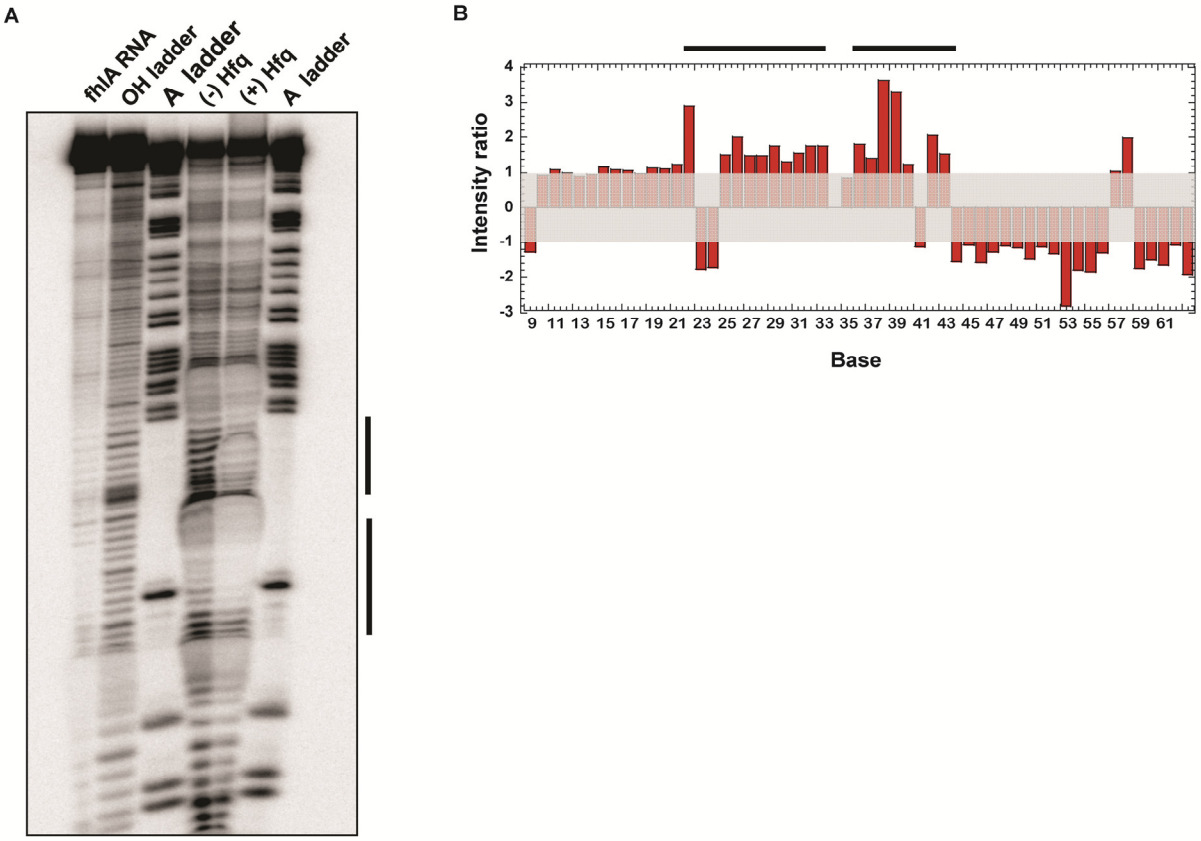


Figure 2.8. Terbium mediated footprinting of *fhlA*53 with Hfq. (A) Poly-acrylamide gel showing the effect of Hfq binding on Tb(III) mediated to ^{32}P-*fhlA*53. (B) Quantitative analysis of Hfq binding based on the gel in panel B. Data are represented as a ratio of the intensity of each band in the absence and presence of 1 μM Hfq hexamer. Values greater than 1 represent protection. Data between +1 and -1 were considered to insignificant.

between positions -76 to +3. It is apparent that *fhlA*220 has distinct functional elements beyond the limits of the previously characterized *fhlA*53 construct.

The two structures differ significantly between -53 and +60. The distinctions in this region arise mainly from bases upstream of -53 pairing with downstream sequences in *fhlA*220. The binding context of *fhlA*53 to OxyS was through two kissing complexes at the leader and coding regions of the mRNA. Only one of the two-stem loop structures remains in the context of *fhlA*220. The proposed stem loop within the upstream region of *fhlA*53 is a bulge in *fhlA*220 and thus remains accessible for OxyS binding. Furthermore the start site and the Shine Dalgano (SD) sequence are unhindered within the flexible region of *fhlA*220. Although the spatial requirement of OxyS to interact with the bulge and the kissing loop seems to be met, complex formation may be hindered by the tertiary structure of *fhlA*220. The 15-fold lower affinity between OxyS and *fhlA*220 relative to *fhlA*53 may be attributed to this structural complexity in the leader sequence and presumably the need to break tertiary contacts to accommodate the bimolecular interaction.

2.3.5 Hfq binding sites of *fhlA* mRNA

Two methods were used to determine the Hfq binding sites on the leader sequences of *fhlA* - terbium-mediated hydrolysis (*fhlA*53, **Figure 2.8**) and NMIA modification (*fhlA*220). Changes in reactivity at each site were categorized as strong or medium and are depicted in **Figure 2.7**, superimposed on the experimental secondary structures (also see **Figure 2.8**). Three main regions showed differential activity in the presence and absence of Hfq (labeled R1, R2 and R3 in **Figure 2.7A**). R2 contains a canonical $(ARN)_x$ motif like the one identified in *rpoS* mRNA and recently shown to bind

the distal face of Hfq (14,125). This site was predicted based on sequence analysis and is now a confirmed interaction site. R1 is comprised of a 7nt loop and an adjacent bulge. R3 lies within a highly flexible region adjacent to the upstream portion of the OxyS interaction sequence. Since the competition gel shift assay showed that *fhlA*220 interacts with both distal and proximal binding sites in Hfq, either R1 and or R3 might interact with the proximal face of Hfq while R2 interacts at the distal site.

2.3.6 Kinetics of *fhlA* association with Hfq

Kinetic analysis of the binding of Hfq to *fhlA*53 and *fhlA*220 was performed using Surface Plasmon Resonance (SPR). 5'-biotinylated mRNAs were bound to streptavidin coated sensor chips. Hfq was then allowed to bind under various conditions while monitoring the interaction. *fhlA*220 was already shown to interact with both the proximal and distal RNA binding surfaces of Hfq (14,116,129). The kinetic model that was used to fit the binding data for H•F220 is shown in **Figure 2.9A**, providing rate constants that correspond to proximal and distal surface interactions. Since, *fhlA*220 wraps around Hfq in order to interact with both surfaces, one could envision more complex models than that shown in **Figure 2.9A**, but this minimal model was sufficient to fit the data to obtain the magnitudes of the rates in which *fhlA* binds to Hfq.

An example kinetic trace fit to this model is shown in **Figure 2.9B**. Both open complexes ($H•F_D$ and $H•F_P$) then converge to a single closed $H•F_{DP}$ complex in which *fhlA*220 binds both surfaces of Hfq. This latter step has not been included in the data fitting shown in **Figure 2.9**. The SPR experiment does not appear to be particularly sensitive to this conformational change as it is an internal rearrangement with little

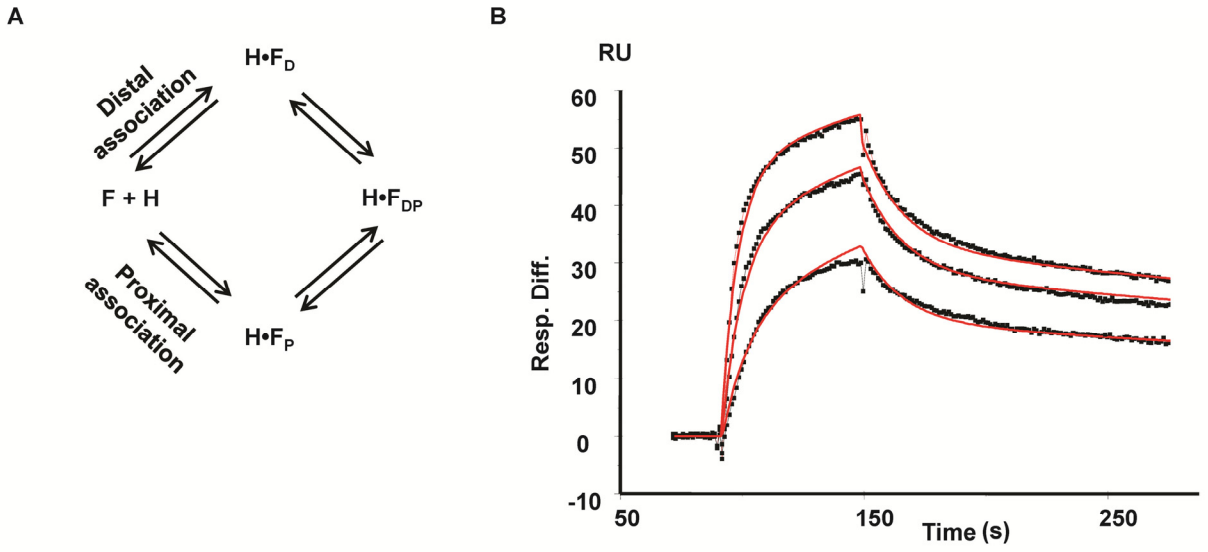


Figure 2.9. Kinetics of *fhlA* association to Hfq measured using Surface plasmon resonance. (A) The interacting model of Hfq to *fhlA* mRNA. Hfq interaction with *fhlA*220 was viewed as a parallel binding model since Hfq presents two distinct RNA binding surfaces (distal and proximal). The SPR data was fitted into a model that identifies two complex formations through 2-independent pathways reflecting RNA association to distal and proximal sites (line arrows). Formation of the closed complex (dashed arrows) leads to no significant change in the SPR response and was not used in data fitting. (B) Sensorgram for *fhlA*220 interacting with Hfq. Here varying concentrations of Hfq (5 – 15 nM) were titrated to surface immobilized *fhlA* mRNA. At the end of each injection (Hfq), the dissociation was monitored by flowing buffer over the sensor surface, allowing spontaneous dissociation in the absence of competitors. Solid lines represent the above-mentioned fitting models applied to the data sets (materials and methods).

Table 2.1 Kinetics and affinities for *fhlA* constructs associating to wt-Hfq, Y25A Hfq and K56A Hfq measured using the kinetic model shown in **Figure 2.9A**. $K_{D,agg}$ is the equilibrium constant based on the compound association and dissociation rate constants defined in **Eq 1.6**.

System	k_{a1} ($M^{-1} s^{-1}$)	k_{d1} (s^{-1})	K_{D1} (nM)	k_{a2} ($M^{-1} s^{-1}$)	k_{d2} (s^{-1})	K_{D2} (nM)	$K_{D,agg}$ (nM)
Hfq-*fhlA*220	$(1.60 \pm 0.06) \times 10^{6}$	$(1.40 \pm 0.02) \times 10^{-3}$	0.83 ± 0.04	$(4 \pm 1) \times 10^{6}$	$(5.6 \pm 0.8) \times 10^{-2}$	13 ± 4	10 ± 2
Hfq-*fhlA*53	$(2.0 \pm 0.2) \times 10^{6}$	$(1.0 \pm 0.1) \times 10^{-1}$	50 ± 10	$(3.7 \pm 0.7) \times 10^{5}$	$(1.9 \pm 0.2) \times 10^{-3}$	5 ± 1	43 ± 6
K56A Hfq-*fhlA*220	$(2 \pm 2) \times 10^{5}$	$(6.6 \pm 0.3) \times 10^{-4}$	3 ± 2	$(2 \pm 1) \times 10^{5}$	$(7 \pm 1) \times 10^{-2}$	400 ± 300	131 ± 60
Y25A Hfq-*fhlA*220	$(2 \pm 1) \times 10^{6}$	$(3.0 \pm 0.1) \times 10^{-1}$	160 ± 90	$(2.9 \pm 0.8) \times 10^{5}$	$(1.0 \pm 0.2) \times 10^{-3}$	4 ± 1	176 ± 100

change in refractive index in the interfacial zone. Rate data for Hfq interacting with *fhlA*220 and *fhlA*53 are collected in **Table 2.1**. Thermodynamic dissociation constants (K_Ds) were computed from kinetic data providing values similar to those measured by gel shift assays for *fhlA*220 (**Table 2.1**).

Looking at these data, one can draw several significant conclusions regarding the interactions between *fhlA* and Hfq. Two low-nanomolar K_Ds result from fitting to this model and one can also calculate an aggregate dissociation constant ($K_{D,agg}$) of 10 ± 2 nM for this interaction since the observed rates are a combination of the microscopic rate constants associated with distal and proximal site binding. This aggregate value was in good agreement with data obtained using gel shift assays. In contrast, for *fhlA*53 binding to Hfq, only one of the binding interactions produced a K_D comparable to that observed with *fhlA*220. $K_{D,agg}$ of 43 ± 6 nM was calculated for this interaction, once again in agreement with gel shift data. This phenomenon likely reflects the loss of the specific interaction with the distal surface due to the lack of the $(ARN)_x$ motif.

2.3.7 OxyS binding kinetics to *fhlA*220 and *fhlA*53

Surface plasmon resonance was also used to measure OxyS binding to *fhlA*220 and *fhlA*53 in the absence of Hfq (**Figure 2.10**). 5'-Biotin labeled *fhlA*53 was immobilized and OxyS was titrated at various concentrations. Data were fit to a Langmuir binding model to obtain kinetic parameters. The association rate constant for OxyS interacting with *fhlA*53 was (1.3 ± 0.5) x 10^4 $M^{-1}s^{-1}$ while the dissociation rate constant was (6.75 ± 0.08) x 10^{-4} s^{-1}, yielding a dissociation constant (K_D) of 50 ± 20 nM. This value is in good agreement with equilibrium data obtained using gel shift assays. To obtain kinetic data for *fhlA*220, the experiment had to be inverted such that

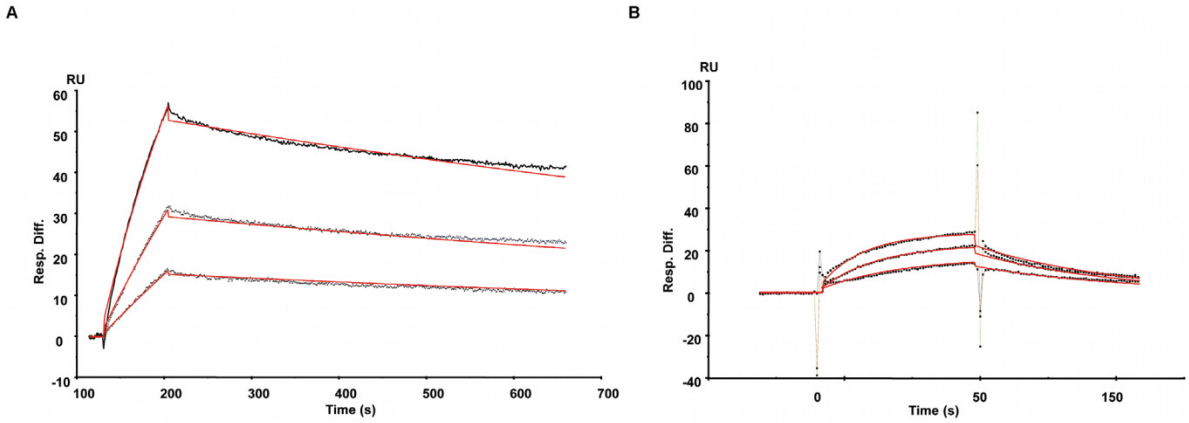


Figure 2.10. Kinetic analysis of OxyS interaction with *fhlA*220 and *fhlA*53. (A) Representative SPR sensorgram for OxyS•*fhlA*53 interaction is shown. 5'-Biotin labeled *fhlA*53 mRNA was immobilized and varying concentrations of OxyS was titrated (400, 200 and 100 nM). (B) SPR sensorgram for OxyS binding to *fhlA*220. Here the biotin label was added to OxyS sRNA and titrated with *fhlA*220 to monitor the interaction (1.5, 3 and 4.5 μM). For both interactions data were fitted into a Langmuir binding model to yield kinetic constants. The model is an over-simplification of a complex system as it ignores unimolecular RNA structural rearrangements that might be required prior to association, but the model sufficient to illustrate the interactions in the absence of Hfq and their approximate rates.

5'-biotin labeled OxyS was immobilized into the SPR sensor surface with subsequent addition of *fhlA*220. As predicted by gel shift assays (**Figure 2.2B**), the OxyS•F220 complex formed weakly with a K_D of ~1.96 ± 0.01 μM estimated from the SPR kinetic constants. This K_D is ~5 fold weaker than that measured using gel shift assays. The weak affinity resulted from slow association, ((9.7 ± 0.4) x 10^3 $M^{-1}s^{-1}$), likely due to the need for *fhlA*220 to rearrange, to make the interaction site accessible. These data imply that in the absence of Hfq, binary complexes still form but relatively slowly in both cases and with significantly lower affinity.

2.3.8 *fhlA*220 interaction with proximal and distal mutants of Hfq

To measure effects of the kinetics when one of the two RNA binding surfaces of Hfq were abrogated, proximal and distal mutants were used. Hfq mutants, Y25A and K56A were previously shown to disrupt RNA interactions at distal and proximal sites of Hfq respectively (129). Hfq mutants were allowed to interact with *fhlA*220 and kinetic parameters were measured using SPR (**Figure 2.11** and **Table 2.1**). For both Y25A and K56A Hfq, one of the two apparent affinities was dramatically altered while the other was essentially unchanged. $K_{D,agg}$s of 131 ± 60 nM and 176 ± 100 nM were estimated from kinetic data for Y25A and K56A Hfq mutants respectively. This destabilization can clearly be attributed to the partial loss of activity in Hfq caused by mutations essential for RNA binding. Also note that, the overall association rates for *fhlA*220 were diminished for both Hfq mutants compared to wt-Hfq, indicating that Hfq facilitates faster association rates, allowing it to capture the mRNA more effectively by using both RNA binding faces.

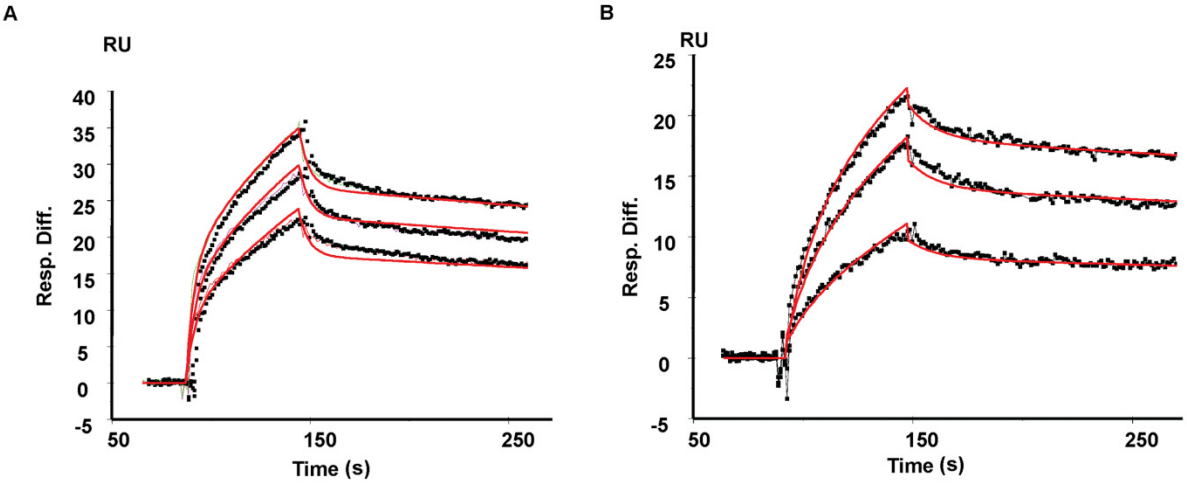


Figure 2.11. Kinetics of *fhlA*220 mRNA interaction with Y25A and K56A Hfq mutants. (A) Sensorgram for *fhlA*220 binding to Y25A Hfq. It was shown that Y25A mutation abrogated RNA binding at the distal site of Hfq (129). As described previously ~3 fmol of mRNA was immobilized in the SPR sensor surface and titrated with mutant Hfq at varying concentrations of 33 nM - to 67 nM hexamer. (B) Sensorgram for *fhlA*220 binding to K56A Hfq. K56A Hfq mutant has been shown to destabilized RNA binding at the proximal surface of Hfq. Hfq was titrated at concentrations of 17 nM - to 50 nM K56A Hfq hexamer. For both (A) and (B), the above-mentioned two-site parallel binding model was applied.

2.3.9 Competing Hfq from the *fhlA*220•Hfq complex using distal and proximal binding RNAs

In the previous section, SPR was used to study the direct dissociation of Hfq from *fhlA*220. Within a cell, a more common scenario might be an exchange reaction where Hfq passes from one RNA to another without ever being free in solution. To measure exchange rates, the H•F220 complex was formed and immediately titrated with DsrA, A_{18} or a mixture of DsrA and A_{18}. DsrA is a sRNA associated with cold shock rather than oxidative stress, and binds to the proximal site of Hfq while A_{18} represents a distal site binding RNA (129). SPR sensorgrams for these experiments are shown in **Figure 2.12**. In these experiments, 100 nM Hfq was injected to ~3 fmol of surface bound *fhlA*220. Once the injection was complete, DsrA and A_{18} (500 nM and 300 nM) or both RNAs were introduced (**Figure 2.12B-D**). The dissociation data were fit to a Langmuir dissociation model. Hfq dissociates with a rate constant of 0.07 ± 0.01 s^{-1} when competed with both proximal and distal RNAs, a rate up to 50-fold faster than simple dissociation (**Table 2.1**). Addition of A_{18} alone gave practically the same dissociation rate as the mixture (0.06 ± 0.01 s^{-1}) whereas the addition of DsrA alone competed for Hfq at a rate 3-fold slower (0.020 ± 0.005 s^{-1}) (**Figure 2.12**). Both types of competition, however, are faster than the direct dissociation. These results are consistent with the wrap-around model for Hfq binding to mRNAs presented in **Figure 2.12A**. The data are also consistent with an exchange mechanism when 2 RNAs compete for Hfq, but further experiments will be required to validate that such a process really occurs.

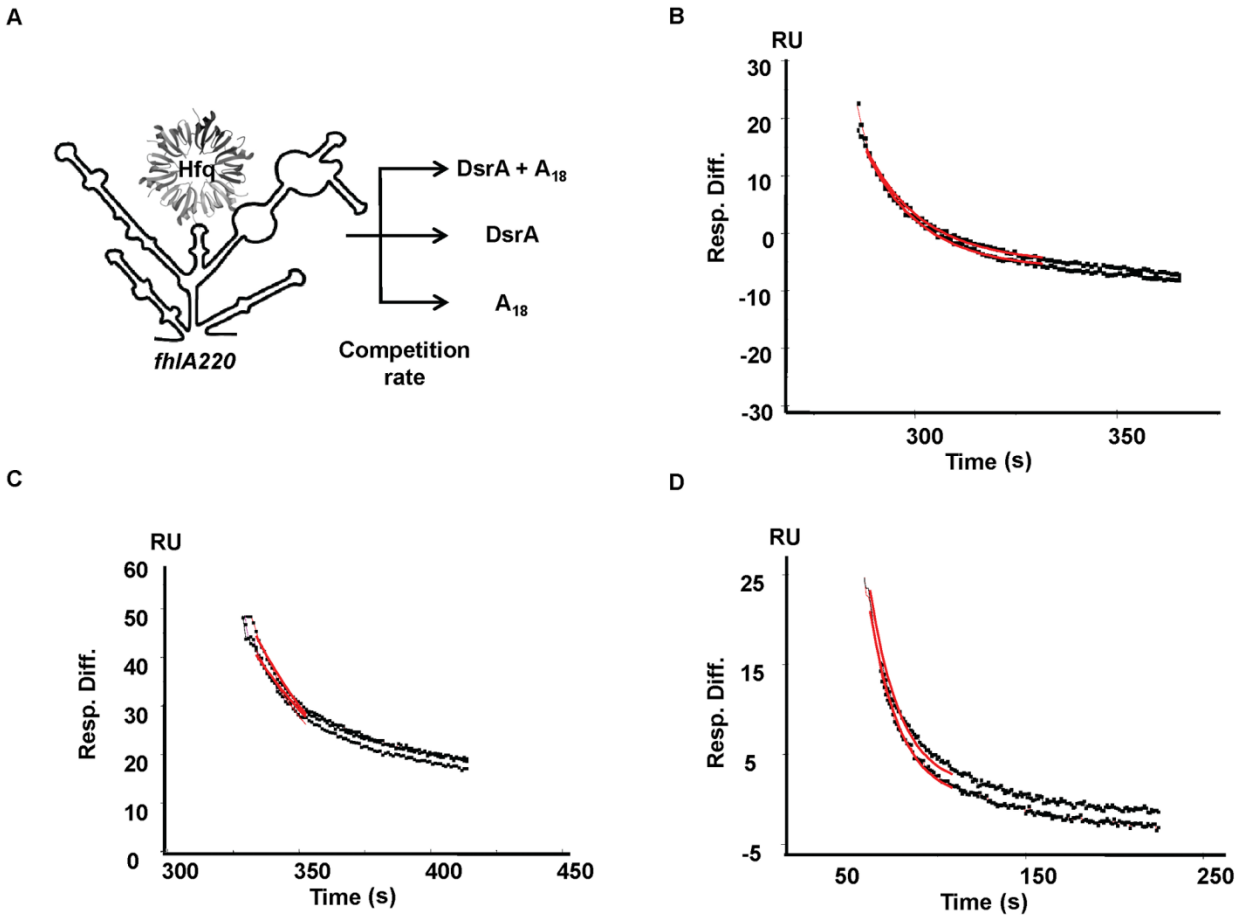


Figure 2.13. Handoff kinetics of Hfq from Hfq•*fhlA*220 complex. (A) Overview of handoff experiment. The Hfq•*fhlA*220 complex was pre-formed and dissociation kinetics of Hfq was monitored by titrating competing RNAs DsrA (proximal), A_{18} (distal) or both RNAs. (B) Sensorgram of Hfq dissociation from the Hfq•*fhlA*220 complex in the presence of 500 nM and 300 nM DsrA and A_{18}. (C) Sensorgram of Hfq dissociation from the Hfq•*fhlA*220 complex in the presence of 500 nM and 300 nM DsrA. (D) Sensorgram of Hfq dissociation from the Hfq•*fhlA*220 complex in the presence of 500 nM and 300 nM of A_{18}. All dissociation data were fitted in to Langmuir model.

2.4 Discussion

Almost all trans-acting sRNAs in *E. coli* and *Salmonella* require Hfq for their gene regulatory activity. In addition to its function of promoting base pairing of sRNAs to their target mRNAs, Hfq is also thought to engage ribosomes, poly A polymerase and RNase E and other enzymes that are involved in RNA transactions (116). Early models that investigated functional mechanisms for Hfq were sRNA centric however, where the Hfq-sRNA complex sought their mRNA targets within the cell. This hypothesis led most studies to use mRNA constructs that only included structural elements necessary for sRNA binding (24,123,124,140). These mRNA constructs may have lacked important upstream regulatory sequence elements such as the recently identified $(ARN)_x$, motifs that are essential for effective Hfq-mediated gene regulation (56,125,127).

One interesting consequence of extending the *fhlA* mRNA leader region was the ability of the mRNA to interact simultaneously with both RNA binding surfaces of Hfq. In the absence of the upstream portion of the leader sequence, *fhlA*53 interacts only with the proximal surface of Hfq. The more complex binding interaction of *fhlA*220 supports a model wherein mRNAs interact with Hfq even in the absence of regulatory sRNAs. This complex can essentially act as an Hfq tag. It can alter basal translation levels alone, while also marking the mRNA as one susceptible to regulation by a sRNA if and when the appropriate sRNA is transcribed. Such a model would enable the cell to respond to stress signals more readily since it streamLines the search for appropriate messages.

The SHAPE-derived secondary structure model proposed for *fhlA*220 provides insight into functional elements in *fhlA*220 mRNA. A region with weak base pairing contains most of the regulatory elements such as the OxyS binding site, RBS,

translation start site and the putative $(ARN)_x$ motif. This relatively floppy section is flanked on both 3' and 5' ends with highly structured regions. When these structures were inadvertently disrupted, such as in the *fhlA*136 species, the ability to bind Hfq was lost indicating the importance of these folds for presenting the Hfq binding motif. Since both RNA binding surfaces of Hfq interact with the *fhlA* leader, multiple footprints were expected. One point of contact was the $(ARN)_x$ sequence element, which should interact with the distal surface of Hfq (14). This leaves regions R1 and R3 as potential proximal binding sites. The A/U rich nature of R3 and its proximity to the $(ARN)_x$ motif lead us to propose that this site is more likely than R1 to be the natural proximal binding element in *fhlA*220.

The presence of the Hfq binding motif $(ARN)_x$ in the leader sequence was interesting. These A-rich stretches are now widely accepted as being present in mRNA leader sequences that are regulated by Hfq (125,141). In a recent study in *Salmonella,* it was projected that Hfq modulated the synthesis of ~20% of all proteins either directly or indirectly (122). A similar number was predicted by a brief bioinformatic survey of potential Hfq binding regions (AAYAA and ARN tracts) in upstream sequences within *E. coli* mRNAs (14). These sequence elements are quite degenerate and thus will appear with a high frequency by chance. Whether all of these AAYAA and $(ARN)_x$ elements represent Hfq binding sites requires further investigations as it is possible that an additional structural context is required to define an Hfq-dependent regulatory element within an mRNA.

Elongation of the *fhlA* leader from -53 to -220 significantly alters both the kinetics and thermodynamics of its interactions with Hfq and OxyS. Complex formation with

OxyS was much weaker for *fhlA*220 relative to *fhlA*53 in the absence of Hfq. Addition of Hfq, however reversed this trend and restored OxyS affinity for the Hfq•*fhlA*220 complex. This finding implies the requirement of Hfq to facilitate regulatory RNP complex formation as observed in vivo assays for this system (2). Similar observations were made by Soper *et al*. in the *rpoS* leader interaction with Hfq. Hfq binds more tightly to *fhlA*220 than to *fhlA*53. This extra stabilization results from the upstream binding element that enables *fhlA*220 mRNA to interact Hfq through both the proximal and distal surfaces simultaneously.

The kinetic model used to characterize *fhlA* binding to Hfq measured two compound rate constants. By using distal (Y25A) and proximal (K56A) face mutants, a large overall destabilization in affinities were observed for *fhlA* binding to Hfq. This result indicated that both binding surfaces of Hfq were in use and validated the competition gel shift experiments. Association of *fhlA*220 to Hfq was rapid, with similar magnitudes observed for both rate constants. The shorter *fhlA*53 construct on the other hand showed very different binding behavior with a 50-fold difference in the dissociation rates between the two apparent rates, explaining the propensity for *fhlA53* to binding to a single site (proximal).

Simple dissociation may not represent the natural behavior of Hfq in the cell however. With so much Hfq and so many Hfq-binding RNAs, an exchange from one RNA to another may be a more common behavior. We therefore exposed the Hfq•*fhlA*220 complex to other RNAs to determine their ability to induce exchange. This experiment showed faster exchange rates than were observed for simple dissociation with the most pronounced effect occurring when a distal binding RNA was introduced.

The competing RNA essentially traps one of the two Hfq binding sites facilitating complete dissociation of *fhlA220* from Hfq. This finding has functional importance since most sRNAs bind through the proximal site. Thus, if Hfq is bound to both proximal and distal sites of an mRNA in a closed complex, exposure to a sRNA will lead to formation of a ternary complex that retains Hfq contact with the mRNA through the $(ARN)_x$ motif and Hfq's distal surface. If complementarity between the RNAs is found within the lifetime of the ternary complex, the conformational changes responsible for gene regulation will ensue. Otherwise, the sRNA will simply dissociate leaving the Hfq-mRNA complex intact and unchanged. Hfq-RNA complexes challenged with the distal-binding A_{18} RNA, showed a three-fold kinetic advantage over dissociation with proximal binding DsrA. These findings imply that the Hfq-mRNA complex will be more resistant to a non-cognate sRNA than to an A-rich RNA that resembles an alternative mRNA. Whether these findings are consistent with other Hfq binding mRNAs besides *fhlA* remains to be validated.

In conclusion the data presented here supports the notion that Hfq-mRNA complexes are essential elements in sRNA mediated gene regulation. This work supports recent findings by the Woodson lab that sequence elements in upstream regions of mRNAs are important for Hfq binding and gene regulation in vivo (56). These $(ARN)_x$ motifs are widely dispersed in bacterial mRNAs and we are only now learning about the importance of such signals in bacterial genes. A recent report suggested an Hfq interaction with the RNA polymerase β-subunit (142) and might imply the potential to handoff of Hfq to nascent transcript marking them for subsequent regulation by an sRNA if necessary in response to an environmental signal. This would be an efficient

way for Hfq to locate target mRNAs and ensure that they are positioned properly to support sRNA-mediated gene regulation if required.

2.5 Materials and Methods

2.5.1 Plasmid construction for *fhlA*53 and OxyS

pNS10901 carries the *fhlA* 5'-end fragment from -53 upstream of the AUG initiation codon to +60. The construct was prepared by positioning the dsDNA of the FhlA fragment behind a T7 promoter sequence flanked by a EcoR I and a BamH I site using the following two primers 5'-ACGTACGAATTCTAATACGACTCACTATAGGCAG TTAGTCAATGACCTTTTGCAC CGCTTTGCGGTGCTTTCCTGGAAGAAC-3' and 5'-CGAGCTGGATCCAATATTTGTTGTCCGAGTGATGTCGAACAACCCTTGTTGTCCGA GATCACTCATCGGTGCATATGACATTTTGTTCTTCCAGGAAAGCACCGC-3'. The primer extension assay was performed using a standard procedure described previously (143). The resulting DNA was cloned into pUC19 and used to transform into XL-10 supercompetent *E. coli* cells. The resulting plasmid (pNS10901) was verified by sequencing. The plasmids (pNS10901) were isolated using a Giga-prep kit (Qiagen). Plasmid DNA were further purified using phenol-chloroform-isoamyl extraction and ethanol precipitated. The plasmid was prepared for runoff transcription by digesting with Ssp I. The OxyS fragment was amplified in XL-10 *E. coli* cells using primers OxyS F- 5'-GGAACAAGCTTTAATACGACTCACTATACCTTCGCCTAGGACCTCTAGG-3 and OxyS-R- 5'-CCGAGCGAATTCTTTAAAGCCTCGCCGTGGAG-3'. The primers were designed with a T7 RNA promoter in the OxyS-F primer with a flaking EcoR I and Hind III restriction sites. The dsDNA fragments were inserted into pUC19 and transformed

into XL-10 *E. coli* cells (pNS10092). The plasmids were verified using sequencing. The plasmid was prepared for runoff transcription by digesting using Dra I.

2.5.2 RNA preparation for SHAPE, in vitro binding and kinetic analysis

A_{18} RNA was purchased from Dharmacon Technologies and deprotected following the manufacturer's protocol. RNA quality was assessed using denaturing PAGE and gel purified. For in vitro binding and kinetic analysis, *fhlA*220 mRNA were transcribed using a DNA fragment that amplifies the FhlA gene from -220 to +60 in *E. coli* XL-10 cells using primers 5'-GGAACCGAATTCTAATACGACTCACTATAGCAGCG TTACATTCCCATCCACTGG-3' and 5'-CCGAGCGGATCCAATATTTGTCCGAGTGAT GTCGAACAACCC-3' and digested with Ssp I before transcribing. For SHAPE analysis, *fhlA*220 mRNA was in vitro transcribed using a DNA template that was amplified in XL-10 cells using primers 5'-GGAACCGAATTCTAATACGACTCACTATAGGCCTTCGG GCCAAGCAGCGTTACATTCCCATCCACTG-3' and 5'-CCGAGCGGATCCAATATT GAACCGGACCGAAGCCCGATTTGGATCCGGCGAACCGGATCGATGTCCGAGTGA TGTCGAACAACCC-3' that includes a structure cassette in the 5' and 3' regions as previously described (132). In vitro transcription was performed after digesting the amplified product with Ssp I. DsrA was obtained by runoff transcription of pBAU10301 that was digested by Ssp I (23).

2.5.3 Biotinylation of RNA

mRNAs *fhlA*220 and *fhlA*53 was 5'-labled with biotin to be used in surface plasmon resonance experiments. RNAs were first treated with Calf Intestinal Phosphatase (CIP) and phosphorylated using ATP-γ-S using the Ambion Kinase Max kit (Ambion, Inc). In brief, 1 nmol of RNA was treated with CIP (in 10X

dephosphorylation buffer, 0.5 units of CIP at 37 °C for 2 h). The reaction mixture was purified using the Phosphatase Removal Reagent (Kinase Max kit) as described by the product manual. Purified RNAs were phosphorylated with ATP-γ-S using T4 Polynucleotide kinase. Phosphorylated RNAs were purified using a G-25 spin column (GE Healthcare) and speed vacuumed to dryness. RNAs were then dissolved in 45 µL of 100 mM $KHPO_4$, pH 8.0, 5 µL of 20 mM N-iodoacetyl-N-biotinylhexylenediamine dissolved in DMF (EZ-Link Iodoacetyl-LC-Biotin, Thermo Scientific). The reaction was incubated at 45 °C for 1 h while shaking under dark conditions. The reaction was then ethanol precipitated and analyzed using PAGE.

2.5.4 Hfq expression and purification

Expression and purification of Hfq and its mutants Y25A and K56A was performed as previously described (129).

2.5.5 Electrophoretic mobility shift assays

All binding reactions were performed in 50 mM Tris-HCl pH 7.5,100 mM KCl and 10 mM $MgCl_2$ at room temperature. Prior to any interaction all RNAs in buffer were annealed at 90 °C for 3 min, cooled to room temperature for 30 min. For all reactions 8 µL aliquots were loaded after diluting with loading buffer (10% (w/v) sucrose, xylene cyanol, bromophenol blue) under a power of 5 W on native 5-8% polyacrylamide (37:1) gel in 1X TBE. Dried gels were visualized by phosphorimaging (Molecular Dynamics) using a Typhoon 9210 imaging system (Amersham/GE healthcare). Quantification was done using ImageQuant 5.1 (Molecular Dynamics) and Kaleidagraph 3.0 (Synergy). Data were fit using nonlinear least-square analysis to a cooperative binding model shown

below (**Eq 2.1**). Here, L is the ligand concentration and the cooperatively is indicated by n. Typical values for n ranged from 1.1 to 2.7.

$$Q_{Fraction} = \frac{(L)^n}{K_D + L^n} \quad (\mathbf{2.1})$$

In the case of A_{18}, DsrA competition assays, the *fhlA*220•Hfq complex was pre-formed and A_{18} and DsrA was titrated at varying concentrations from 0 to 30 μM.

2.5.6 Chemical SHAPE analysis

The secondary structure of *fhlA*220 was mapped using SHAPE chemistry as described previously (130,132). In brief 1 pmol of RNA was folded in buffer (50 mM Tris-HCl pH 7.5, 100 mM KCl, 10 mM $MgCl_2$) by heating to 95 °C and cooled to room temperature for 15 min. Then added N-methylisatoic anhydride (NMIA) in anhydrous DMSO to a final concentration of 3 mM. The reactions were incubated at 37 °C for 45 min. A control experiment without NMIA was performed for 1 pmol of RNA where instead of NMIA, DMSO was added. The reaction was then ethanol precipitated in the presence of a co-precipitant (20 μg, Glycogen).

The 2'-O-adducts were analyzed using primer extension. The modified/unmodified RNAs (1 pmol, 10 μL, in 0.5X TE) were heated to 95 °C for 3 min in a thin PCR tube and cooled in ice for 1 min. Fluorescently labeled primer (5'-F-GAACCGGACCGAAGCCCG) (3 μL) was added to (+ NMIA) (0.3 μM WellRED D4) and (- NMIA) (0.4 μM WellRED D3) reactions respectively. The primer template solutions were then incubated at 65 °C for 5 min and 37 °C for 15 min. Primer extension was initiated by adding enzyme mix (4 μL of Superscript III FS buffer, 1 μL 0.1 M DTT, 1 μL 10 mM dNTP mix) and incubating at 52 °C for 1 min. Then added Superscript III (1 μL) and incubated at 52 °C for 15 min. In addition to these two reactions two sequencing

reactions were performed to identify corresponding peaks. The sequencing reactions were assembled as mentioned above (RNA, 1 pmol), 3 µL primer (ddCTP, 1.2 µM LICOR IR 800) and (ddATP, 2 µM WellRED D2), 6 µL enzyme mix, ddNTP (1 µL, 0.25 mM) and Superscript IIII (1 µL) and performed RT. The four reactions (+ NMIA), (- NMIA) and sequencing reactions were combined and ethanol precipitated in the presence of glycogen. The pellets were washed twice with 70% ethanol and dried under vacuum. The pellets were re-suspended in SLS loading solution (Beckman). cDNA samples were separated on a Beckman CEQ 8000 DNA sequencer. The separation was performed using the following parameters (capillary temp: 60 °C, denature temp: 90 °C, time 150 s, injection voltage: 2 kV, time 7 s, separation voltage 3 kV and separation time 100 min).

The raw fluorescence intensities were analyzed using the software ShapeFinder (134). The quantitative shape data were normalized to a scale that falls between 0 for un-reactive sites and reactive bases would attain an average reactivity of 1. This method of normalizing was extensively described elsewhere (136). SHAPE reactivities were then imported to RNAstructure software where the intensities were converted to pseudo-free energy changes (144) that were obtained using triplicate experiments.

2.5.7 Chemical footprinting

To identify Hfq binding sites in *fhlA*220, SHAPE chemistry was performed in the presence and absence of Hfq as described above. 2 pmol of RNA was reacted with 1 µM Hfq for 30 min. The RNA was then modified using NMIA (3 mM) at 37 °C for 45 min, and treated with Proteinase K (1 µL, 20 mg/mL) at 37 °C for 30 min. The reaction was then phenol-chloroform-isoamyl alcohol extracted followed by ethanol precipitation. The

primer extensions were performed as previously described and the (+Hfq) base reactivities were compared to (-Hfq) reactions.

Tb(III) mediated footprinting was performed as previously described (119). In brief, 250K cpm of 5' ^{32}P-end labeled RNA was incubated in probing buffer (50 mM Tris-HCl at pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$) containing 0 and 1 µM Hfq hexamer for 30 min at room temperature. $TbCl_3$ was added to a concentration of 100 mM and incubated for 2h. The reaction was quenched by adding EDTA (50 mM) and SDS (0.1%). Samples were then treated with Proteinase K and incubated at 37 °C for 30 min. The mixture was then resolved on an 8% denaturing PAGE. The data was then analyzed using SAFA software to normalize, align and measure reactivities of base positions with reference to a ladder (145).

2.5.8 Surface Plasmon resonance

Kinetic experiments were performed on a Biacore 2000 instrument. Experiments were done on either a streptavidin-coated chip (SA chip, Biacore) or CM5 chip where 3000 RUs (response units) of streptavidin was coated using amine coupling. Streptavidin coating the CM5 chip involved, activating the carboxymethylated dextran (CM) sensor chip with 35 µL of 0.2 M EDC and 0.05 M NHS. A 35 µL solution of streptavidin (200 µg/mL in 10 mM sodium acetate, pH 4.8) was injected repeatedly to achieve the expected 3000 RUs of surface coverage. This was then followed by the injection of 35 µL of 1 M ethanolamine to quench residual NHS esters. The Immobilization was carried out at 25 °C at a flow rate of 5 µL/min.

All experiments were performed in the same reaction buffer (50 mM Tris-HCl, pH 7.5, 100 mM KCl and 10 mM $MgCl_2$). During all experiments ~3 fmol of 5' biotin labeled

RNAs were immobilized on the sensor chip. Immobilizations of RNAs were performed at a flow rate of 3 µL/min to make sure homogeneous surface coverage is attained. Experiments were carried out at 25 °C and at a flow rate of 30 µL/min.

To measure kinetics of OxyS binding to *fhlA*53, OxyS was titrated at varying concentrations of (400 nM, 200 nM and 100 nM). Kinetics of OxyS interacting with *fhlA*220 was measured by immobilizing ~3 fmols of 5'-Biotin labeled OxyS and titrating *fhlA*220 at concentrations (1 - 4.5) µM. Surface regeneration was performed by injecting 300 µL of regeneration buffer (50 mM Tris-HCl pH 7.5, 20 mM EDTA) at a flow rate of 100 µL/min. In case of Hfq binding to *fhlA*220 or *fhlA*53 Hfq was titrated at varying concentrations of (15 nM, 10 nM and 5 nM) hexamer. Hfq for this experiment was dialyzed with the reaction buffer using a Slide-A-Lyzer® mini dialysis kit (Thermo Scientific, 3500 MWCO) prior to the experiment. The regeneration was performed by injecting 60 µL of 500 nM DsrA and A_{18} RNA solution. For Y25A Hfq binding to *fhlA*220 Hfq hexamer concentrations at 67 nM, 50 nM and 33 nM were used. For K56A Hfq concentrations of 50 nM, 33 nM and 17 nM were used.

To measure the exchange kinetics between Hfq and *fhlA*220, the co-injection mode that was available in the Biacore 2000 control software was used. Here 17 nM wt-Hfq was titrated and immediately after the injection, either DsrA, A_{18} or both together at concentrations of 500 nM and 300 nM was introduced.

The data were analyzed globally by fitting both the dissociation and association (where applicable) phases simultaneously (BIA evaluation software version 4.1). A 1:1 (Langmuir) model (two fitting parameters) and a parallel reaction model (four fitting parameters) were used (**Figure 2.9A**). The binding model was constructed in BIA

evaluation software and the equation for the model is shown below, making the assumption that $R_{HF_D} \approx R_{HF_P}$ due to similar molecular weights for the two complexes.

$$R_{tot} = \chi_D R_{HF_D} + \chi_P R_{HF_P} + RI \quad (\mathbf{2.2})$$

Where,

$$\frac{d[HF_p]}{dt} = \left(k_{a1}[H][F] - k_{d1}[HF_P]\right) \quad (\mathbf{2.3})$$

$$\frac{d[HF_D]}{dt} = \left(k_{a2}[H][F] - k_{d2}[HF_D]\right) \quad (\mathbf{2.4})$$

$$\frac{d[F]}{dt} = -\left(k_{a1}[H][F] - k_{d1}[HF_P]\right) - \left(k_{a2}[H][F] - k_{d2}[HF_D]\right) \quad (\mathbf{2.5})$$

$$\frac{d[F]}{dt} = -\left(k_{a1} + k_{a2}\right)[H][F] + \left(k_{d1} + k_{d2}\right)\left([HF_P] + [HF_D]\right) \quad (\mathbf{2.6})$$

Total changes in response units for this model was due to contributions from the distal and proximally coordinated complexes (χ_D (R_{HF_D}) and χ_P (R_{HF_P}), respectively) and *RI,* which corresponds to the bulk refractive index contribution to the overall response (R_{tot}). Here species χ_D and χ_P represents the mole fractions of the Hfq complexes formed with the distal and the proximal sites respectively. Surface immobilized *fhlA* mRNA is represented as *F* and Hfq is depicted as *H* in above equations. Kinetic rates k_{a1} and k_{a2} define association phase parameters whereas dissociation rates are given by k_{d1} and k_{d2} for the two binding phases. BIA evaluation uses Marquardt-Lavenberg algorithm to optimize parameters in fits and assigns kinetic constants to the above described equation. The goodness of the fit was judged by the reduced chi-square (χ^2) values.

2.6 Acknowledgements.

The authors wish to acknowledge helpful discussions with Alain Laederach, Richard Federly and Anne-Cecile Duc on aspects of this work as well as other members of the Feig Lab.

CHAPTER 3

Requirement of Upstream Hfq Binding $(ARN)_x$ Elements in *glmS* and the Hfq C-Terminal Region for GlmS Up-regulation by sRNAs GlmZ and GlmY

3.1 Abstract

Hfq is an important RNA binding protein that helps bacteria adopt to stress. Primary function of Hfq is to promote *trans*-acting sRNA base pairing to target mRNAs using its distal and proximal RNA binding surfaces to form a regulatory RNP node. This RNP complex promotes post-transcriptional gene regulation that results in translation activation/repression and or RNA turnover. Moreover, Hfq facilitates multiples regulatory pathways simultaneously to construct integrated responses towards multiple stress pathways when required. A common mechanism as to how Hfq control and propagates concerted stress responses using multiple messages is not established. However recent identification of Hfq binding $(ARN)_x$ sequence motif in up-stream regions of *rpoS* and *fhlA* that was essential for regulation ask the question whether these elements are a common occurrence among other Hfq dependent mRNAs. Here in this work we confirm the presence of similar $(ARN)_x$ motifs in *glmS* RNA, a Hfq dependent regulatory node that requires sRNAs GlmZ and GlmY to activate translation. As previously observed in the OxyS-*fhlA* system, *glmS* interacts with both RNA binding surfaces of Hfq in the absence of sRNAs. Using a *glmS:gfp* fusion system, we identify the $(ARN)_x$ element that is essential for Hfq to propagate regulation. Furthermore, we measured the requirement of the less-conserved C-terminal region of Hfq for activation of GlmS expression in the presence of sRNAs. Here we find that for promoting GlmZ-*glmS* base-pairing the Hfq tail is mostly not essential. However for GlmY mediated up-regulation of

GlmS, at least the amino acid residues between 66 and 72 in the Hfq C-terminal were required. Hfq variants from *clostridium perfringens* and *clostridium difficile* that consist of homologues nucleic acid biding motifs to *E. coli* Hfq was able to facilitate GlmZ-*glmS* base pairing to activate translation of *glmS*. On the contrary, these Hfq variants were unable to cross-compliment for *E. coli* Hfq function during the GlmY sRNA pathway. This further validated the requirement of the Hfq C-terminal region for GlmY up-regulation of *glmS*.

3.2 Introduction

Non-coding RNAs (sRNAs) are used to modulate a wide variety of cellular responses in both bacterial and eukaryotic systems (5,39). In eukaryotes microRNAs and siRNAs (~21 nt) regulate endogenous processes and pathways such as transcription, heterochromatin formation, translation and RNA/DNA degradation and also defends the genome by invasive nucleic acids (146). Similarly in bacterial systems two main classes of sRNAs have been identified based on the origin of the sRNAs relative to its target. *Cis*-acting RNAs are typically transcribed from an anti-sense strand at the same loci of the intended target RNA except in riboswitchers where in these cases the 5'-UTR region act on cis- to attenuate mRNA translation (6,7). Anti-sense RNA (asRNAs) interactions with their mRNA partners are known to promote regulatory outcomes such as translation initiation, RNA processing and decay and transcriptional attenuation (6,112). In contrast, *trans*-acting RNAs are synthesized at a different locus to its target RNA and are widely used to modulate stress responses in bacteria. Unlike asRNAs, *trans*-acting RNAs can accommodate imperfect base pairing with its target RNAs where the RNA binding protein Hfq often facilitates complex formation.

Hfq was first identified as a host factor that was essential for the replication of phage Qβ RNA (114). Hfq belongs to the Sm/Lsm family of proteins and associates into a homo-hexameric structure to form the functional nucleic acid binding core. In general the nucleic acid binding Sm1 and Sm2 domains are highly conserved across bacterial species (7-66 amino acids) that form two distinctive distal and proximal RNA binding surfaces respectively (147). These two RNA binding surfaces are utilized to promote sRNA-mRNA base pairing by increasing the local concentration of the two binding partners. Loss of Hfq in bacteria has shown increased sensitivity to stress and reduced pathogenesis among many other phenotypes (8,9,120,122,148). One most intriguing feature of Hfq-mediated gene regulation is the network-based organization that allows a single sRNA to trigger a muti-dimensional response to stress (20). For instance, the sRNA RyhB regulates multiple genes such as *sodB*, *ftnA*, *bfR*, *acnA* and *sdhC* as a response to low iron concentrations in cells (40). While the abundant usage of Hfq in many other stress pathways such as *cold shock* (DsrA) (38), *oxidative stress* (OxyS) (36) and *sugar stress* (SgrS) (3) are well established, a common mechanism as to how Hfq controls these processes is lacking.

Recent discovery of Hfq binding sequence elements within upstream regions of mRNAs has provided a basis in elucidating the Hfq mediation process in stress regulation. For example, in *rpoS* and *fhlA* mRNAs, upstream $(ARN)_x$ Hfq binding elements are been shown be essential for regulation (30,56,57). Furthermore, structural data confirms $(ARN)_x$ sequence specificity towards the distal RNA binding surfaces of Hfq (14). If these upstream Hfq binding sequence motifs are a common feature in stress regulation in bacteria, this could help understand how Hfq identifies relevant mRNAs

during stress regulation. In theory, as described in the *fhlA* work, Hfq could pre-form complexes with relevant mRNAs, which could promote an efficient stress response upon sRNA synthesis if and when it is induced by environment stimuli (30). Furthermore, recent work on the DsrA-*rpoS* has shown that Hfq binding to *rpoS* re-folds the mRNA for a favorable interaction with the sRNA (63). These results and observations, although tested only on a few examples, help explain the complex manner by which Hfq facilitates and propagates through integrated responses to help bacterial systems overcome stress.

To further clarify the recruitment of $(ARN)_x$ motifs in upstream regions of mRNAs, the Hfq-dependent regulation of *glmS* was studied. *glmS* codes for an essential enzyme glucosamine-6-phosphate synthase (GlcN-6-P) (149). In many gram-positive bacteria such as *bacillus subtilis* the expression of GlmS is tightly regulated by a self-cleaving ribozyme located within *glmS* that catalyzes cleavage in response to high GlcN-6-P concentrations (150). However, in gram-negative species, GlmS expression is regulated by GlmZ and GlmY, a dual sRNA system that requires Hfq for activity (105). In addition to Hfq and RNase E, YhbJ, a predicted NTPase was shown to be involved in the regulatory network of GlmS in *E. coli* (52) (**Figure 3.1**). In brief, GlmS is encoded as a poly- cistronic message that also includes GlmU that codes for a bi-functional enzyme that produces UDP-GlcNAc, an intermediate component essential for bacterial cell wall synthesis. In order to individually regulate GlmS expression, RNase E cleaves the *glmU- glmS* transcript at the stop codon of GlmU to produce *glmS* that is translationally

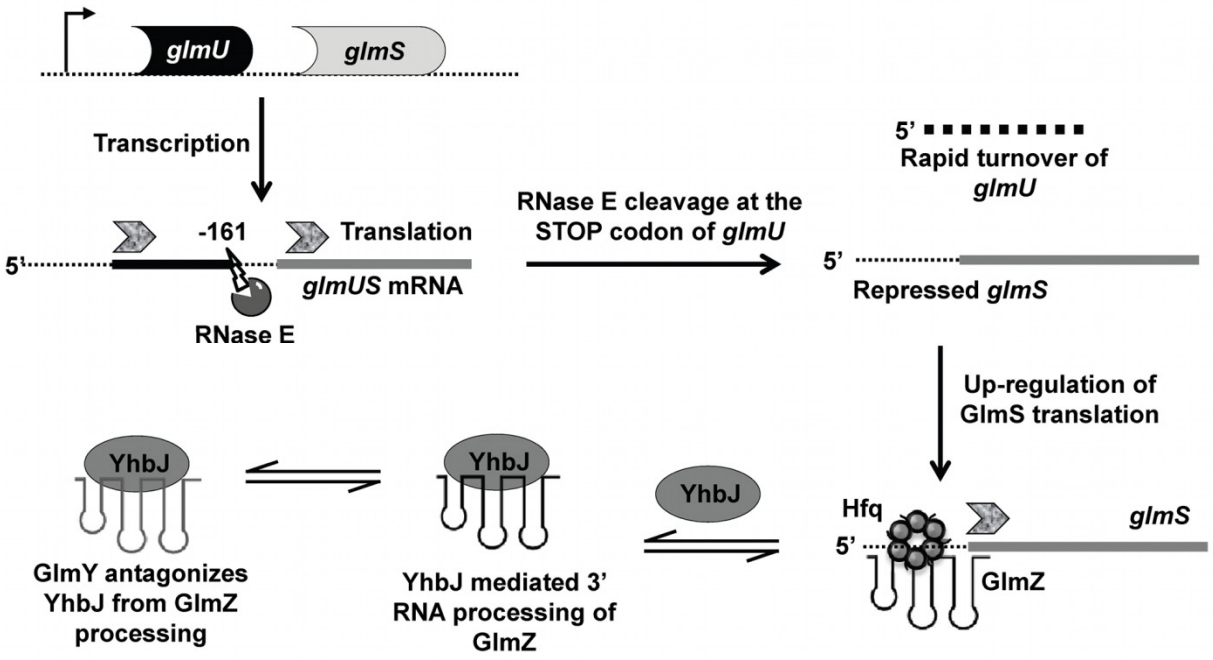


Figure 3.1. Mechanism of Hfq mediated regulation of *glmS* by sRNAs, GlmZ and GlmY (adopted from (52)). *glmS* is transcribed as a poly-cistronic message that also includes *glmU*. Upon transcription, RNase E cleaves at the stop codon of *glmU* (at -161 nt position with respect to the translation start site of *glmS*) that separates the translational control of GlmS and GlmU genes. The RNase E cleaved *glmS* transcript is repressed for translation by its 5'-UTR structure that masks the RBS. Upon synthesis of GlmZ sRNA, *glmS* translation is activated in an Hfq-dependent manner. Here the GlmZ-*glmS* interaction relieves the inhibitory structure of *glmS* for translation. YhbJ, a predicted NTPase negatively regulates GlmS expression by processing GlmZ sRNA. A second sRNA GlmY, which has structural and sequence homology to GlmZ antagonizes YhbJ from GlmZ that positively regulates the GlmS expression. Overall, the GlmS expression is activated by two homologues sRNAs GlmZ and GlmY in a coordinated manner in the presence of Hfq.

repressed by its 5'-UTR structure. The sRNA GlmZ interacts with *glmS* to un-mask the RBS to promote ribosome assembly for translation. On the other hand YhbJ has been shown to be involved in GlmZ turnover that negatively regulates GlmS expression (105). The second sRNA GlmY has significant sequence and structural homology to GlmZ that is sufficient to antagonize YhbJ from GlmZ (104).

In this work we show that the *glmS* interaction with Hfq is similar to what was previously reported with *fhlA* where both RNA binding surfaces of Hfq was involved in binding. Structural characterization of *glmS* revealed potential upstream $(ARN)_x$ motifs similar to *fhlA* and *rpoS*. Using a GFP fusion system we show that an upstream $(ARN)_x$ element is essential for Hfq-mediated up-regulation of *glmS* by sRNAs GlmZ and GlmY extending the number of systems that rely on this functional motif for Hfq mediated gene regulation. Furthermore, we look into the requirement of the variable C-terminal region of Hfq for regulation that has been a controversial topic in the field. We show that the Hfq C-terminal is not required for the classical role of Hfq mediation that involves promoting sRNA-mRNA base pairing. However, at least 7 residues of the C-terminal extension (~65-72 aa) were essential for GlmY to activate GlmS expression. This represents a non-classical facilitation role of Hfq where it potentially partners with processing enzymes such as YhbJ during regulation. This shows the conditional use of the Hfq C-terminal that is dependent on the sRNA pathway.

3.3 Results

3.3.1 *glmS* interaction with Hfq

Previous work performed on the *glmS* regulatory system in *E. coli* showed the requirement of Hfq for GlmZ- and GlmY-mediated activation of *glmS* (52,104). While the

basic outline of the regulatory network was established no molecular level understanding was developed. Based on our previous work with OxyS/*fhlA* (30) and DsrA/*rpoS* (57) we asked whether Hfq pre-binds the *glmS* mRNA and how similar the *glmS* interaction relates to better understood systems. The *glmS* mRNA used throughout this study was 182 nts in length. This fragment extends -161 nts into the upstream leader region which is the RNaseE processing site that separates the *glmU-glmS* transcripts (**Figure 3.1**) and extends +21 nts into the coding region.To test whether a similar binding model is present in *glmS* we measured the binding constant and interaction mode with Hfq (**Figure 3.2**).

To measure the binding constant between *glmS* and Hfq, [5'-^{32}P] *glmS* mRNA was incubated with varying Hfq concentrations from 0 nM to 2 μM (hexamer). The *glmS*-Hfq complex was resolved on native gels to provide a K_D of 30 ± 4 nM (**Figure 3.2A**). This value was comparable to the interaction between *fhlA*-Hfq and also to typical mRNA-Hfq interactions measured previously using gel shift assays (21,30,57).

To determine the *glmS* binding surfaces of Hfq a competition gel shift assay was carried out where the *glmS**-Hfq complex was pre-formed and competed against bona fide distal (A_{18} RNA) and proximal (DsrA) site binders of Hfq. As observed in **Figure 3.2B**, the *glmS**-Hfq complex could only be displaced by the simultaneous addition of both DsrA and A_{18} showing that GlmS interacts with both of the faces of Hfq similar to *fhlA* and *rpoS* (30).

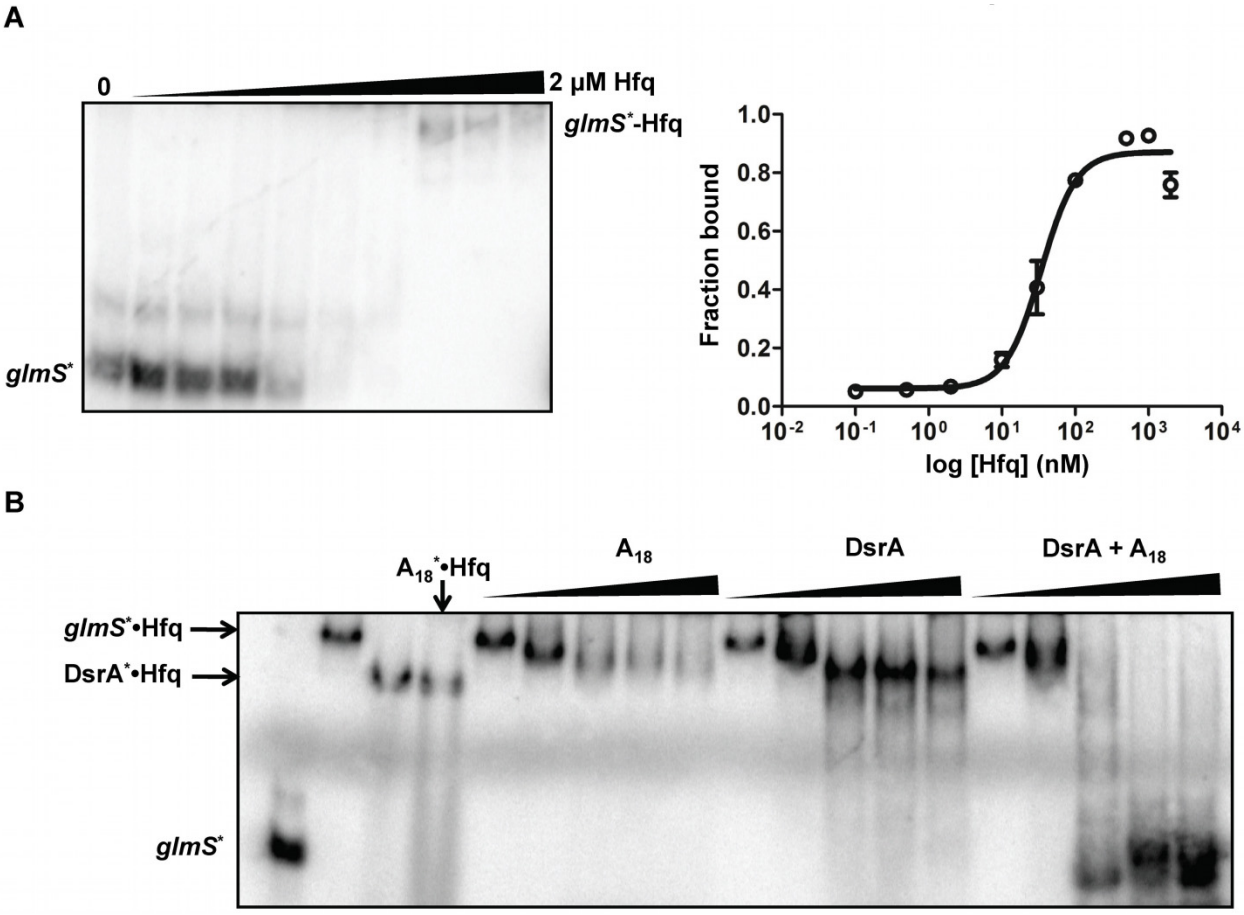


Figure 3.2. *glmS* interaction with Hfq. **(A)** Analysis of *glmS* interaction with Hfq (left). Gel shift experiment showing the binary complex between Hfq and *glmS* mRNA. ^{32}P-labled *glmS** was titrated with varying concentrations of Hfq ranging from 0 – 2 μM. Quantitative analysis of the *glmS*-Hfq gel (right). As described in materials and methods, thermodynamic constants were determined by non-linear least squares analysis fitted to a cooperative binding model. **(B)** Competition gel binding experiment to determine the Hfq binding surface to *glmS*. The *glmS**-Hfq complex was pre-formed and incubated with increasing amounts of either distal (A_{18} RNA), proximal (DsrA) or both (A_{18} and DsrA) binding RNAs (0 – 30 μM) to compete *glmS** away from Hfq.

3.3.2 The secondary structure of *glmS* mRNA and Hfq binding sites probes using SHAPE

SHAPE (as described in Chapter 2) was carried out to determine the secondary structure of *glmS* mRNA to identify functional elements. **Figure 3.3** depicts the secondary structure for *glmS* mRNA derived from SHAPE constraints, superimposed with the modification intensity data. The functional elements in *glmS* mRNA such as the potential $(ARN)_x$ motifs, GlmZ binding site, ribosome binding site (RBS) and translation start sites are also depicted in **Figure 3.3**. According to the proposed *glmS* structure, two potential $(ARN)_x$ motifs can be identified as candidates that could serve as upstream Hfq binding elements. To confirm Hfq binding regions in *glmS*, SHAPE footprinting was used as described previously (30) and the results are indicated in **Figure 3.3**. As expected, both $(ARN)_x$ regions footprinted with Hfq among other probable bindings sites indicated in **Figure 3.3**.

3.3.3 Measuring translational regulation of GlmS using a GFP fusion system

A GFP fusion system was used to measure the translational regulation of GlmS by sRNAs (GlmZ and GlmY) that is also dependent on Hfq. A schematic diagram indicating functional elements in the fusion system is shown in **Figure 3.4**. As described above, a *glmS* fragment that spans from -161 to +21 was N-terminally fused to GFP under the transcriptional control of a pBAD promoter. The fusion system is constructed in a pBAC vector backbone (kindly provided by the Philip Cunningham Lab, Wayne State University, Biological Sciences) which is based on the F-plasmid *ori2*/RepE replicon, that allows stable maintenance of large genomic fragments as very low or

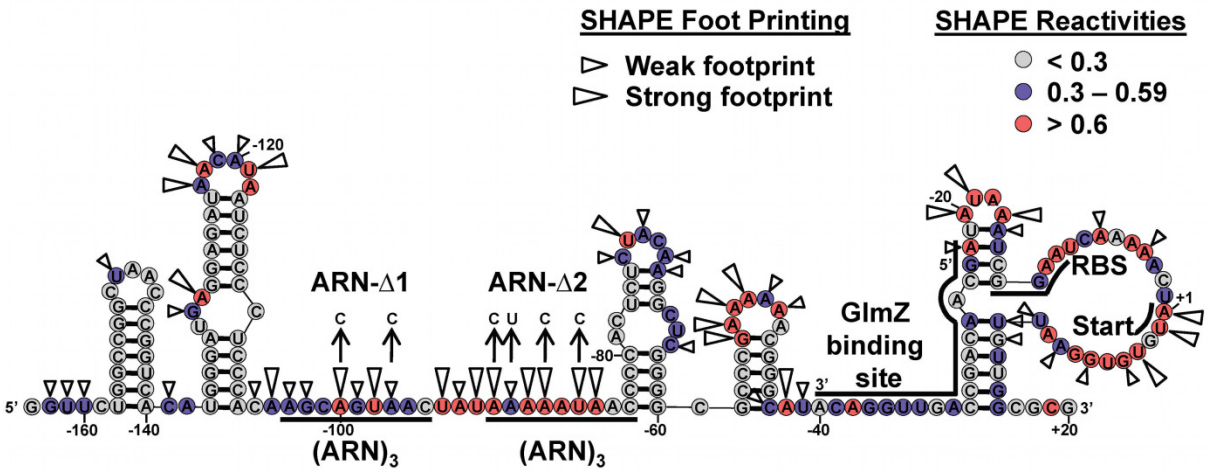


Figure 3.3. SHAPE-derived secondary structure model for *glmS* mRNA. The NMIA reactivity for each nucleotide correlates to a given color, nucleotides with reactivities greater than 0.6 are shown in red, from 0.3 – 0.59 in purple and less than 0.3 in grey. Hfq footprinting data in the presence and absence of 0.5 μM Hfq hexamer are superimposed on the structure. Large wedges indicate where there was a strong protection (greater than 0.6) from wild type Hfq was observed and small wedges where a weak protection (0.3-0.59) was observed. The $(ARN)_x$ sites, glmZ binding site, ribosome binding site and the start codon are also shown.

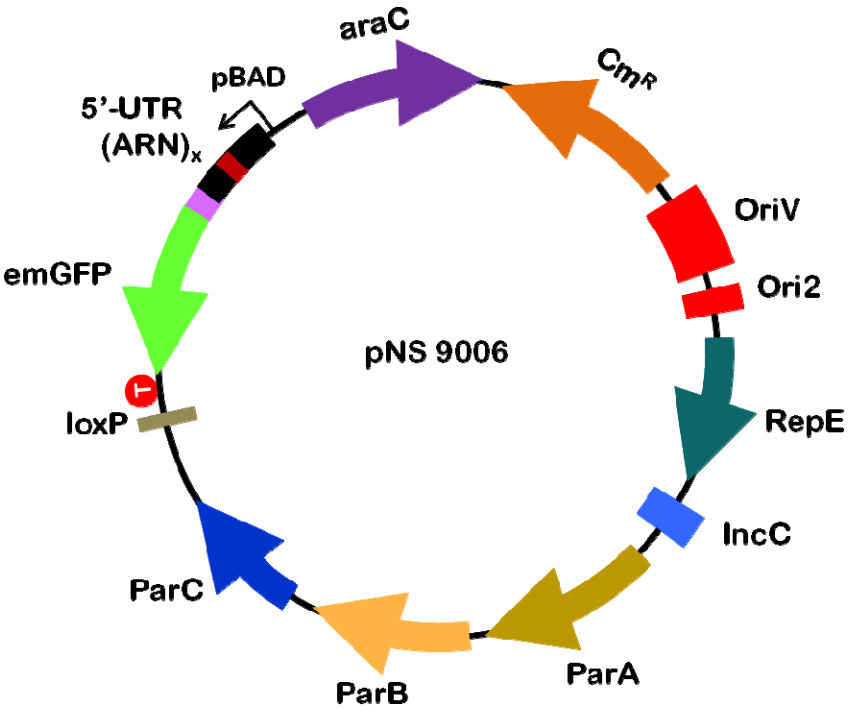


Figure 3.4. The *mRNA*:*gfp* fusion construct. A *glmS* mRNA fragment that spans from -161 to +21 relative to the translation start site was N-terminally fused to GFP that is under the transcriptional control of a pBAD promoter. The fusion system is constructed in a pBAC vector backbone that stably maintains low copies of plasmids using the *ori2*/RpeE replication origin. In addition, the vector backbone also consist an inducible replication origin *oriV* that can be conditionally induced to produce high copy numbers of the plasmid. However throughout this work low copies of fusion plasmids were maintained in cells.

single copy plasmids (151). In addition the vector also contains an *oriV* replication origin that can be induced by L-arabinise to produce high copy numbers of the plasmid in an *oriV* specific manner when expressed in a host strain that can supply the replication initiation protein TrfA in *trans* (152). However, throughout this study a low copy number of fusion plasmids were maintained.

To monitor expression of GlmS in the presence of sRNAs, GlmZ or GlmY were co-expressed together with the fusion plasmid. Here a vector pNM12 that is a derivative of pBAD plasmid series was used to express GlmZ and GlmY (kindly provided by Prof. Susan Gottesman) (38). Therefore co-expression of the GFP fusion and sRNAs were both under the control of a pBAD promoter induced with 0.1% arabinose. The coordinated transcription of mRNAs and its cognate sRNA were shown to be beneficial for the Hfq machinery for efficient regulation in vivo (20).

To confirm the proper functioning of the fusion construct and co-expression conditions, GlmS-GFP expression was measured to obtain previously validated observations on the GlmS regulatory system (52,104) and the results are shown in **Figure 3.5**. To measure GFP fluorescence, liquid cultures were grown from overgrown overnight cultures. The new cultures were then grown at 37 °C for 3 h while shaking before inducing with 0.1% L-arabinose. The cultures were left to grow for another 150 mins upon induction before the cells were harvested. The growth of each individual culture was monitored by its OD_{600} absorbance. Cells were consistently collected at their early stationary phase throughout (OD_{600}, ~1.4 ± 0.2). Lower dilutions from overnight cultures were made for slow growing strains such Δ*hfq* cells to reach early-stationary phase within the time frame of the assay. The fluorescence was

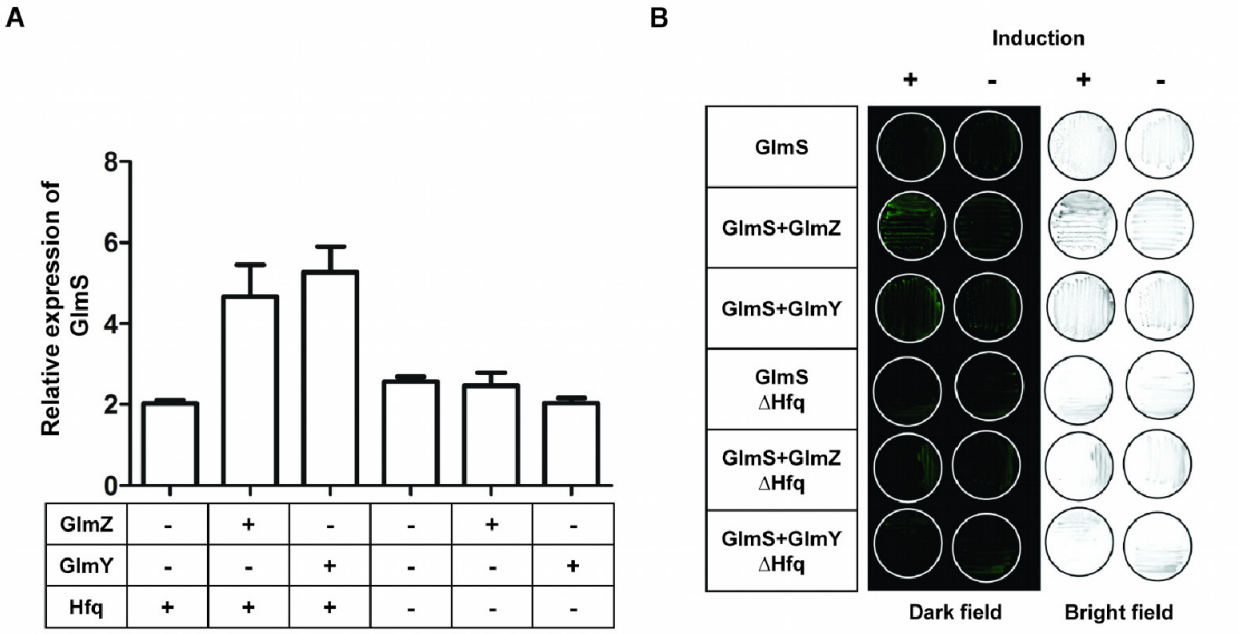


Figure 3.5. Expression of GlmS:GFP measured using GFP fluorescence. **(A)** Relative expression levels of GlmS are measured in response to sRNAs GlmZ, GlmY and Hfq. Fluoresce was measured from *e.coli* cells that were grown in liquid media according to the conditions described in material methods. The growth of cells was controlled and monitored using the cell density (OD_{600}) where all cultures were harvested upon reaching early-stationary phase (OD_{600} 1.4 ± 0.2). Cells were lysed before fluorescence was measured using a Tecan GENios Plus multi label plate reader. All fluorescence measurements were at least triplicated using independent inoculations. To calculate relative expression levels of GlmS, the fluorescence levels of each sample was compared to an identical un-induced culture to correct for cellular fluorescence. GlmS:GFP and sRNA synthesis was controlled using a pBAD promoter inducible by L-arabinose. To measured the Hfq dependence on GlmS expression experiments were carried out in a Δ*hfq* background. **(B)** Colony fluorescence of GFP-GlmS grown in LB-agar plates for cells that carry indicated plasmids and conditions. Fluorescence were measured in fluorescence mode using a Typhoon 9210 imaging system (Amersham) to indicate GFP expression in colonies and imaged in bright field (right panel) to show the colony morphology of these stains.

measured upon lysis of 3.0 mL of cell culture (Materials and methods). The fluorescence measured from liquid cultures is shown in **Figure 3.5A**, where GFP expression is normalized to an identical un-induced culture to correct for cellular fluorescence and used to calculate the fold expression of GlmS. Data indicates basal levels of GlmS expression in the absence of either sRNA but increases 5-to 6 fold in the presence of GlmZ or GlmY. The ability to up-regulate GlmS expression is lost in the absence of Hfq even in the presence of sRNAs. These data nicely correlate with what was previously observed regarding this regulatory system (52,104). Colony fluorescence of GlmS-GFP at identical conditions described in **Figure 3.5A** is shown in **Figure 3.5B**. These data show that the GFP fusion construct and assay conditions are able to capture the effects of sRNAs and Hfq in the GlmS regulatory system.

To test whether the translational fusion can recognize defective Hfq species that are incapable of regulating gene expression, proximal and distal mutants of Hfq were used to measure GlmS activation (**Figure 3.6**). Y25A Hfq, K56A Hfq mutants have been shown previously to diminish nucleic acid binding at distal and proximal RNA binding sites respectively (21,30) and the H57 proximal residue has been implicated to be crucial for hexamer organization in *P. aeruginosa* (153) and *S. typhimurium* (22) Hfq species. Data presented in **Figure 3.6**, indicates either proximal (K56A) or distal (Y25A) mutants are incapable of facilitating GlmS up-regulation in the presence of sRNAs. However, H57A Hfq shows a weak facilitation towards GlmS expression compared to wild type Hfq. These data indicate the proper functioning of the GlmS:GFP fusion system sensitive to sRNAs (GlmZ/GlmY) and Hfq.

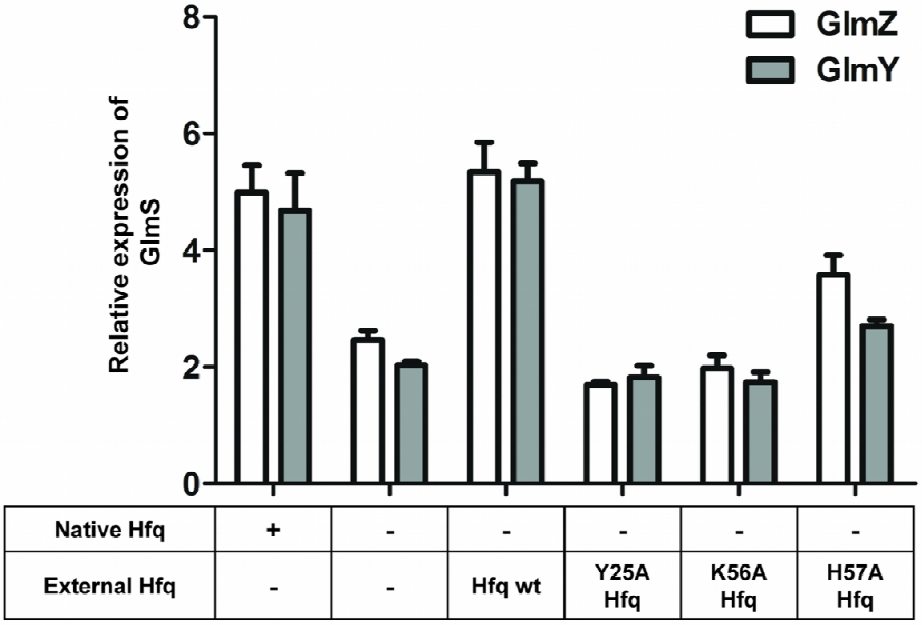


Figure 3.6. GlmS:GFP expression levels in the presence of mutant Hfq variants. To further characterize the functionality of the fusion system, GlmS expression levels were monitored in the presence of sRNA with mutant Hfq variants that were previously shown to deter Hfq function. External Hfq for these experiments was supplemented using a plasmid system where the synthesis of Hfq was under the control of a pTac promoter that is inducible by IPTG. Here Hfq expression was induced prior to GlmS:GFP and sRNA synthesis to allow efficient regulation.

3.3.4 Upstream Hfq binding $(ARN)_x$ elements are essential for regulation

$(ARN)_x$ motifs in upstream regions are essential for Hfq mediated gene regulation of *fhlA* (30) and *rpoS* (56,57). Structural evidence also indicates the preferential binding of $(ARN)_x$ motifs at the distal site of Hfq (14). Furthermore, a recent study showed the specific enrichment of these sequence elements in genomic SELEX experiments that were performed to identify Hfq binders (58). The above observations leads to a model where, the presence of upstream $(ARN)_x$ motifs could be a common occurrence in Hfq dependent mRNAs. The secondary structure model together with Hfq footprinting data for *glmS*, indicated two potential $(ARN)_x$ motifs being present that could be required for regulation (**Figure 3.3**). The GFP fusion system was used to identify $(ARN)_x$ element(s) that are essential for GlmS activation in vivo. Both potential (ARN)s were mutated serially (*glmS* Δ1 and *glmS* Δ2) and simultaneously (*glmS* Δ1Δ2) to identify the role of each in *glmS* regulation (**Figure 3.3**).

The expression of GlmS Δ1:GFP, GlmS Δ2:GFP and GlmS Δ1Δ2:GFP was monitored in the presence of GlmZ (**Figure 3.7A**) and GlmY (**Figure 3.7B**). Data in **Figure 3.7A** show that, sRNA GlmZ is unable to activate GlmS expression when *glmS* 5'-UTR lacks ARN Δ2. However GlmS expression is higher than wild type levels for the *glmS* Δ1 mutant. The over-activation caused by *glmS* Δ1 was absent when both ARNs were mutated (Δ1Δ2). To test whether *glmS* Δ1 activation is Hfq dependent the regulation was monitored in the absence of Hfq. Data from this experiment reveal that the Δ1 mutation on *glmS* did not change the Hfq dependence for up-regulation of GlmS by GlmZ. However, neither ARNs mutant' Δ1 nor Δ2 were able to up-regulate GlmS expression in the presence of GlmY (**Figure 3.7B**) implying the Δ1 mutation only affects

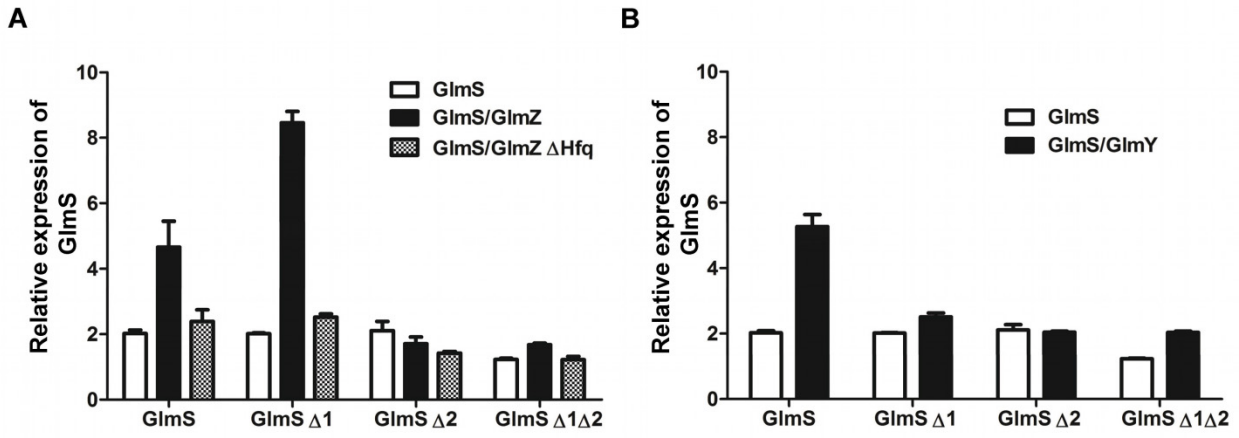


Figure 3.7. Effect of upstream Hfq binding $(ARN)_x$ elements in GlmS activation by GlmZ and GlmY. **(A)** Relative GlmS:GFP expression levels of *glmS* ARN mutants measured in the presence of GlmZ. Here mutations to *glmS* ARN elements were made as indicated in above **Figure 3.3**. Since two potential ARNs were detected both motifs were removed sequentially and simultaneously to produce *glmS* Δ1, *glmS* Δ2 and *glmS* Δ1Δ2 respectively. The expression level of GlmS was measured as indicated in material and methods. The relative up-regulation of GlmS was compared to expression levels when in the absence of GlmZ and Hfq. **(B)** Up-regulation of *glmS* ARN mutants were measured with GlmY and compared to wild type *glmS* expression levels.

the GlmZ pathway. These results indicate the requirement of $(ARN)_x$ elements for Hfq mediated regulation of GlmS and that the ARN Δ2 was the most essential Hfq binding motif for both GlmZ and GlmY pathways.

3.3.5 Requirement of the C-terminal domain of Hfq for regulation of GlmS by sRNAs GlmZ and GlmY

Hfq is present in most Gram–positive and Gram–negative bacteria. In all cases the nucleic acid binding Sm1 and Sm2 motifs are evolutionary conserved across species which spans generally from positions 7 through 66 amino acids. The greatest variability is observed at the C-terminus of Hfq where species such as *E.coli* and *salmonella enterica* have fairly long tails (~102 aa) while species such as *C. diff* or *C. perf* often have shorter C-terminals. The functional importance of the C-terminus of Hfq for regulation is not clear. Therefore, the necessity of the C-terminal region of Hfq for GlmS activation by sRNAs, GlmZ or GlmY was tested using the GFP fusion system.

Several C-terminal truncations of *E. coli* Hfq were co-expressed together with *glmS:gfp* fusion plasmid and sRNAs in a Δ*hfq* background to identify if the C-terminal amino acids that are required for regulation. In addition to wild-type Hfq that consisted of 102 amino acids, shortened C-terminal constructs of Hfq87 (87 amino acids), Hfq72 and Hfq65 were prepared. Here the Hfq65 variant can essentially be regarded as the tail-less version of Hfq since it only consist the nucleic acid binding Sm1 and Sm2 motifs. All C-terminal variants of Hfq were constructed using the ptac-KanL94P plasmid as the parent vector (154), that is under the transcriptional control of an IPTG inducible pTac promoter. The assay conditions to measure the GFP expression were similar to those described earlier, although Hfq induction was carried out using 1 mM IPTG, 1 h before

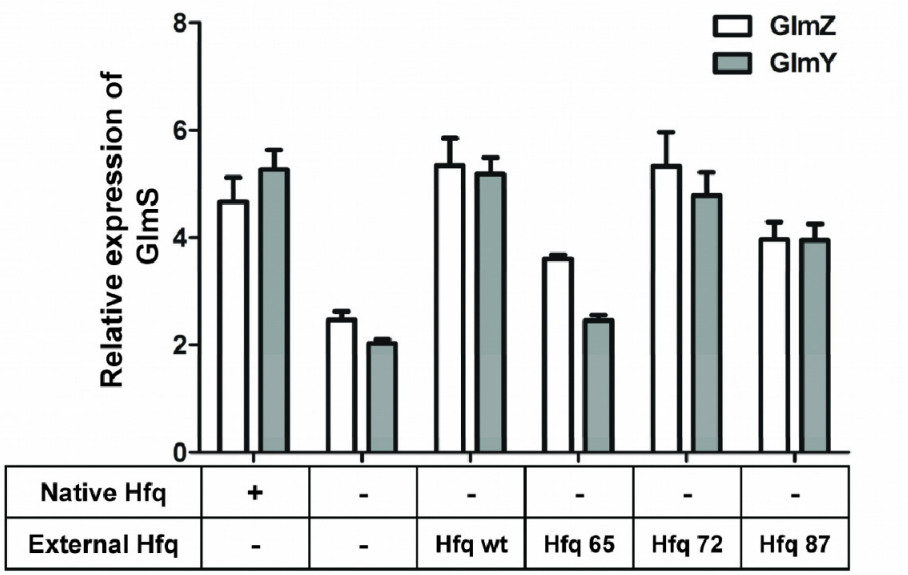


Figure 3.8. Requirement of Hfq C-terminal domain for GlmS:GFP up-regulation by GlmZ and GlmY. Activation of GlmS:GFP expression by sRNAs was measured in the presence of Hfq that consists of varying C-terminal extension lengths. In addition to Hfq wild type (102 aa), Hfq65 (65 aa), Hfq72, Hfq87 were tested for its ability to perform regulation. Hfq65 consists only the nucleic acid binding Sm1 and Sm2 domains while Hfq72 and Hfq87 are 7 aa and 22 aa extensions off the Sm cores. All externally supplemented *hfq* was under the transcriptional control of a pTac promoter that was induced using 1 mM IPTG. Co-expression and induction conditions of GlmS:GFP, sRNA and Hfq plasmid systems are described in material and methods.

arabinose induction to allow enough synthesis of Hfq prior to sRNA regulation. Relative expression levels of GlmS were monitored as the ability to up-regulated GlmS synthesis individually by sRNAs GlmZ and GlmY with varying C-terminal extensions of Hfq (**Figure 3.8**).

From data indicated in **Figure 3.8**, it is clear that plasmid borne and native Hfq (Hfq wt) regulates GlmS expression at comparable levels with both sRNAs in the assay conditions used. GlmS regulation by GlmZ was mostly independent of the length of the Hfq C-terminal extension as Hfq72 and Hfq87 showed identical levels of GlmS expression to wild type Hfq. Hfq65 on the other hand showed reduced levels of GlmS expression implying a possible effect caused by shortening the Hfq C-terminal. However, all Hfq C-terminal variants were better at up-regulating GlmS expression relative to a Δ*hfq* strain in the presence of GlmZ (**Figure 3.8**). Conversely, for GlmY mediated activation, Hfq65 was unable to up-regulate GlmS above basal levels while Hfq72 and Hfq87 were proficient in regulation. These data show that the absence of amino acid residues from at least 66-72 affected GlmYs ability to up-regulate GlmS in a manner distinct from GlmZ regulation. However, as described above GlmZ and GlmY act in a hierarchical manner during regulation. Therefore a loss of GlmYs ability to recruit YhbJ for self-processing would affect GlmZs ability to activate GlmS synthesis. These data provide sufficient evidence to suggest the importance of at least the 7 amino acid tale off the nucleic acid binding core.

3.3.6 *C. perf* and *C. diff* Hfq variants are proficient in GlmZ but not GlmY mediated regulation of GlmS in *E. coli*

To validate these observations that the 7-amino acid tail of Hfq was essential for the GlmY pathway, Hfq variants belonging to Gram-positive bacteria *Clostridium difficile* and *Clostridium perfringens* were used to substitute for *E. coli* Hfq to activate GlmS synthesis. Both *C. diff* and *C. perf* Hfq (Hfq^{Cd} and Hfq^{Cp}) variants have similar nucleic acid binding regions (Sm1/Sm2) compared to *E. coli*. However little sequence homology is observed beyond the Sm regions (**Figure 3.9A**). Therefore, essentially the nucleic acid binding properties should be conserved in Hfq^{Cd} and Hfq^{Cp} variants to promote GlmZ-*glmS* base pairing. However, if there is a need for the C-terminal region or its *E. coli* identity to facilitate regulation, Hfq^{Cp} and Hfq^{Cd} may not be able to cross compliment GlmY native functions.

GlmS activation was monitored in the presence of Hfq^{Cp} and Hfq^{Cd} variants with either GlmZ or GlmY sRNAs in a Δ*hfq* background. The results are shown in **Figure 3.9B**. Hfq from both Hfq^{Cp} and Hfq^{Cd} successfully promote GlmS up-regulation in the GlmZ-dependent pathway. In fact, an increased up-regulation of GlmS is observed for Hfq^{cp} facilitation in the presence of GlmZ while, Hfq^{Cd} mediation was comparable to wild type levels. These data fundamentally confirm that the GlmZ-*glmS* interaction which revolves around the ability of Hfq to promote base-paring between the sRNA and mRNA can be achieved as long as the core nucleic acid binding domain is presented. However, Hfq^{Cd} and Hfq^{Cp} were unable to up-regulate GlmS expression via the GlmY pathway. Since the only noticeable difference between the three Hfq variants are at the C-terminal region, a potential role at the C-terminus is implied in the GlmY pathway.

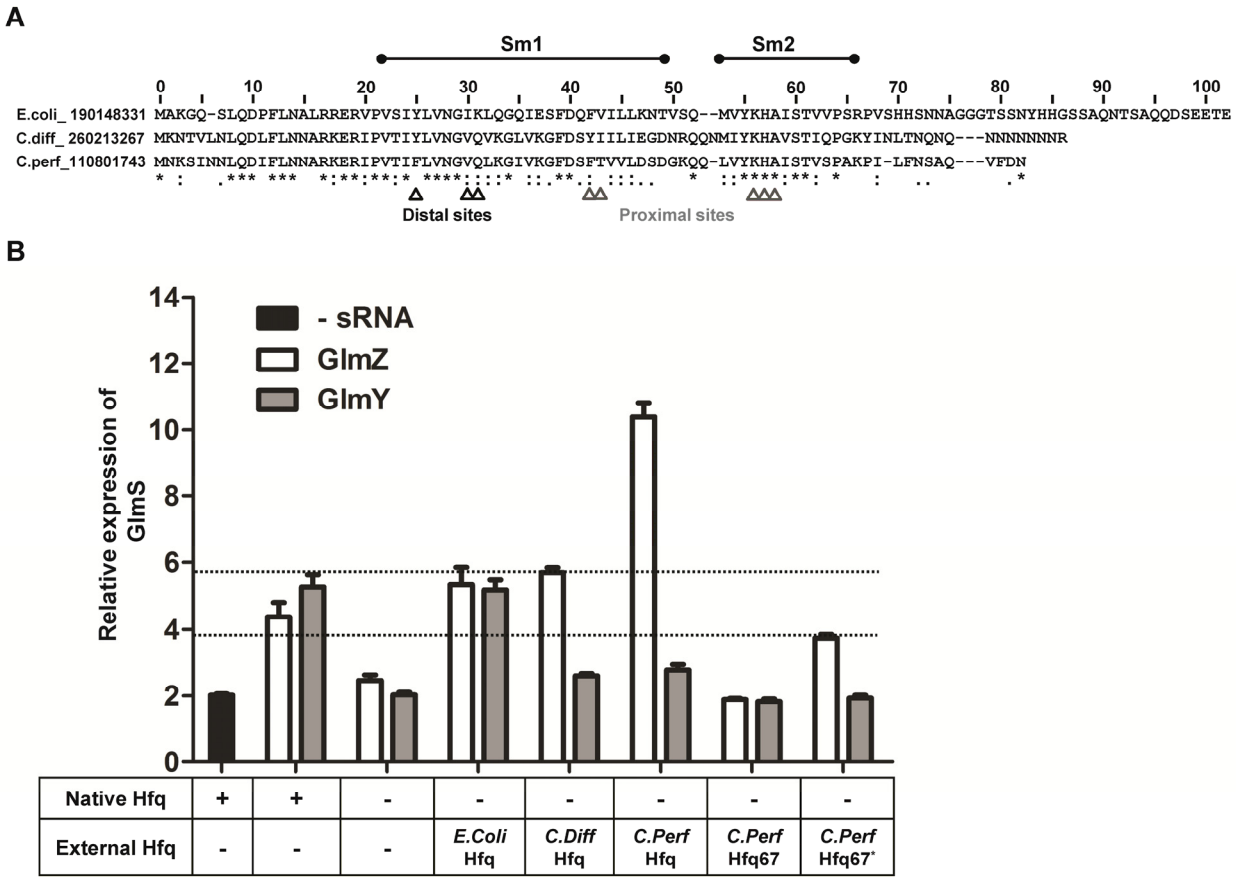


Figure 3.9. GlmS:GFP up-regulation using Hfq^{Cp} and Hfq^{Cd} variants by GlmZ and GlmY. **(A)** Sequence homology between *E.coli*, *C.diff* and *C.perf* Hfq variants. Nucleic acid binding Sm1/Sm2 domains and important amino acid residues in proximal and distal RNA binding surfaces required for regulation are indicated. **(B)** Cross-complementation for Hfq^{Ec} using Hfq^{Cd}, Hfq^{Cp}, Hfq^{Cp}_{67} and $Hfq^{Cp}_{67}{}^{*}$ (amino acids in the *E. coli* Hfq C-terminal in positions between 66-72 were fused to *C. perf* Hfq67) homologues for GlmS up-regulation. The ability to activate GlmS expression by GlmZ and GlmY was measured in the presence of $Hfq^{Cd}$and Hfq^{Cp} variants in *E.coli* Δ*hfq* cells. Dotted lines indicate wild type expression levels of GlmS in the presence of sRNAs.

To further investigate whether, recognition of the Hfq^{Ec} C-terminal amino acids between 66 and 72 (RPVSHHS) is sufficient for GlmS activation by GlmY, this peptide sequence was C-terminally fused to $Hfq^{Cp}{}_{67}$ to create $Hfq^{Cp}{}_{67}{}^{*}$. Here $Hfq^{Cp}{}_{67}$ comprises only the native nucleic acid binding Sm cores. Relative expression levels of GlmS in the presence of both sRNAs, $Hfq^{Cp}{}_{67}$ and $Hfq^{Cp}{}_{67}{}^{*}$ was measured and are shown in **Figure 3.9B**. According to data presented here, removal of the C-terminal amino acids in Hfq^{Cp} severely hindered its ability to activate GlmS expression with either sRNAs. $Hfq^{Cp}{}_{67}{}^{*}$ on the other hand, was able to up-regulate GlmS expression only in the presence of GlmZ where Hfq activity was restored by the addition of the *E. coli* tale peptide sequence. However, $Hfq^{Cp}{}_{67}{}^{*}$ was still unable to activate GlmS synthesis via the GlmY pathway implying that the peptide recognition alone is insufficient to recruit Hfq. Therefore, it is clear that additional factors are in play here in terms of the requirement of the C-terminal region of Hfq for GlmY mediated sequestering of GlmZ from YhbJ.

3.4 Discussion

Being able to adapt to environmental stress is vital for the survival and propagation of bacteria. The RNA binding protein Hfq plays a key role in modulating stress responses in bacteria by regulating gene expression in a sRNA-dependent manner. Hfq functions primarily in promoting sRNA-mRNA base pairing by increasing the local concentrations of sRNAs and target mRNAs through its distal and proximal RNA binding surfaces. In addition, Hfq acts as a chaperone to trigger structural transitions in RNAs to initiate recognition between cognate interactions (63), or partners directly or indirectly with processing enzymes such as RNase E (155,156), Poly A polymerase (157), PNPase (155), RNA helicases (158,159) and YhbJ (105) to aid

stringent gene regulation. These diverse functions enable Hfq to be involved in a cascade of regulatory nodes that control multiples stress responsive pathways. Since all of Hfq networks involve sRNAs to initiate gene regulation at the post-transcriptional level, the sRNA-mRNA interaction can be regarded as the main trigger for down-stream regulatory outcomes.

A number of recent reports have indicated that upstream Hfq binding $(ARN)_x$ sequence motifs to be a common denominator in Hfq dependent mRNAs that help initiate RNP complex formation. Presence of Hfq binding $(ARN)_x$ elements in upstream regions of mRNAs has shown to be important for OxyS-*fhlA*, DsrA-*rpoS* and RprA-*rpoS* RNP complex formations both in vitro and in vivo (30,56,127). According to structural evidence, $(ARN)_x$ elements interact at the distal RNA binding site of Hfq while the proximal site harbors sRNAs during regulation (14). In this work $(ARN)_x$ motifs were structurally characterized to be present in *glmS* that recruits sRNAs GlmZ and GlmY to regulate gene expression in a Hfq dependent manner. The interaction model here between Hfq and *glmS* was consistent with previous proposals to *fhlA,* where the mRNA used both RNA binding surfaces to interact with Hfq in the absence of sRNAs. This mode of interaction between mRNAs and Hfq was attributed to a model wherein, the complexes could serve as a pre-cursor RNP complex that enables a rapid response to stress in bacteria.

Examining the proposed secondary structure of *glmS* and Hfq footprinting data obtained using SHAPE, it was apparent that two potential ARNs (ARN-1 and ARN-2) were present upstream of the RBS (**Figure 3.3**). To identify ARN element(s) that was essential for the regulation, *glmS* mRNA was fused N-terminally to GFP to measure

GlmS expression levels. GlmS translation was up-regulated in the presence of either GlmZ or GlmY in an Hfq dependent manner confirming the proper functioning of the fusion system. Out of the two potential ARNs in *glmS*, ΔARN-2 had the greatest effect in up-regulating translation in the presence of either sRNA. However, the loss of ARN-1 caused a hyper-activation in translational levels of GlmS only in the presence of GlmZ. The up-regulation at the *glmS* ΔARN-1 was nonetheless dependent on Hfq implying that the regulatory node was still intact despite the changes in *glmS* functional elements. However it is clear that the essential $(ARN)_x$ motif in *glmS* that is required for Hfq mediated regulation is the ARN-2 element. While we do not completely understand why the ARN-1 mutant led to over-activation of GlmS expression, it might be a result of a structural change or localization of Hfq at the ARN-2 site that may have caused preferable circumstances for GlmZ mediated regulation. Overall, as observed in previous cases such as in *rpoS, rprA* (56) and *fhlA* (30), Hfq mediated regulation of *glmS* is dependent on the presence of upstream $(ARN)_x$ elements, adding to the list of networks that rely on such sequence motifs for function. In addition to this experimental evidence, usage of ARN_x elements by Hfq at a global scale has been projected by RNA co-immunoprecipitation and SELEX experiments (34,35,58).

Hfq binding at upstream regions of target mRNAs elements could be important in understanding the mechanism of stress regulation in bacteria. As described above we propose that, this interaction could serve to pre-form an Hfq•mRNA complex that identifies messages for regulation. While we have no definitive data to conform this hypothesis, this model would explain Hfq's ability to handle multiple regulatory networks using multiple sRNAs. Also recent evidence suggests that Hfq can be a limiting factor in

cells and that a co-ordinate expression of sRNA and mRNAs is required for the proper functioning of regulatory nodes (20). This implies the intense competition between RNAs to acquire Hfq that in turn explains the need for already expressed mRNAs to capture Hfq to facilitate its own regulation. Other functions of this interaction includes causing structural rearrangements in the mRNA to facilitate the interaction with sRNA which was observed in the case of DsrA and *rpoS* (63). However, the structural transitions that are triggered by the mRNA-Hfq interaction alone are not sufficient to effect basal- or regulate translation levels. The *glmS*-Hfq interaction alone had no effect in basal translation levels (data not shown). Similar observation were reported for *sodB, rpoS*, *ompC* and *fhlA*, all having only minimal effects on translation upon binding to Hfq (20).

The role of Hfq in stress regulation throughout the bacterial kingdom is well established. Comparison of Hfq variants across species has confirmed the presence of an evolutionary conserved nucleic acid binding core (Sm1 and Sm2) and a variable C-terminal region that is less well preserved. While the requirement of the Sm cores for Hfq function is clear, involvement of the C-terminal region of Hfq in regulation is controversial (15,33). Recent structural data compares the Hfq C-terminal regions to intrinsically disordered proteins (IDPs) that could potentially facilitate intermolecular interactions with proteins or RNAs (15,160). While it is widely accepted that proteins such as RNaseE and PAP I act downstream of Hfq mediated regulation, the mode of recruitment of these enzymes are not fully understood (65). Also Hfq pull down experiments that were carried out to mine proteins that interact with Hfq have produced a wide variety of regulatory proteins in bacteria (Taewoo lee and Andrew Feig, Thesis

data). Furthermore, the C-terminal has been implicated in Hfq hexamer association and stabilization (148). However, on one hand a possible involvement of the C-terminal region for Hfq function can be imagined as some bacterial species have evolutionary retained lengthy tails in Hfq (*Sphingopyxis alaskensis,* 196 amino acids) (161) that could imply a beneficent role to the organism. On the other hand contrasting observations are been put forth regarding the requirement of the Hfq C-terminal for gene regulation and stabilization of RNAs (33,162) that has created significant disparities in the field.

To understand the function of the C-terminal region of Hfq for regulation, the GlmZ/GlmY-*glmS* system was used. A clear advantage to using this system to solve this problem stems from the presence of two regulatory layers that are directed by GlmZ and GlmY in an Hfq dependent manner. Here, Hfq on one hand facilitates the GlmZ-*glmS* interaction which is the classical mode of Hfq mediation and on the other hand, may play an indirect role in sequestering GlmZ processing by YhbJ by diverting the degradasome to GlmY. Here the latter function of Hfq could be considered as a non-canonical facilitation by Hfq that could be attributed to C-terminus of Hfq.

When the requirement of the C-terminal region of Hfq for GlmS activation was measured, data indicated that Hfq72 and Hfq87 were proficient in up-regulating GlmS in both sRNA pathways. These two Hfq variants consists 7- to 22 amino acid extensions from the nucleic acid binding Sm cores respectively. However, when only the Sm cores of Hfq (Hfq65) was presented, only GlmZ was able to significantly up-regulate GlmS expression relative to a Δ*hfq* strain. The inability of GlmY to use Hfq65 to up-regulate GlmS implies that the Sm core alone is insufficient to aid its mode of function. This was further validated when it was observed that Hfq^{Cp} and Hfq^{Cd} variants that had mostly

identical Sm regions to *E. coli* could only cross compliment GlmZ but not GlmY mediated regulation of GlmS. First it was envisioned that, since Hfq72 was successfully able to function in the presence of GlmY, the peptide sequence between 66 and 72 could be used during regulation. This was tested by fusing the above peptide sequence into a shortened Hfq^{Cp}_{67} that comprised only the native Sm cores. This hybrid Hfq variant also failed to up-regulate GlmY showing that the *E. coli* Hfq tale (66-72 aa) alone is insufficient to restore *E. coli* wild type activity in Hfq^{Cp}. While these data show that the C-terminal is required for the proper functioning of the GlmS regulatory system the absolute role is still unclear. However, since previous work shows the involvement of processing enzymes such as YhbJ and PNPase in the GlmY pathway, one possibility to these observations could be an interaction with these proteins at the C-terminal of Hfq. Further investigations are underway to delineate this conditional requirement of the C-terminal amino acids during non-classical Hfq mediated gene regulation.

Overall, this work further points out the requirement and the presence of upstream Hfq binding $(ARN)_x$ motifs in mRNAs. The usage of a two-layered regulatory system that was driven by sRNAs GlmZ and GlmY to up-regulate GlmS expression was exploited to understand the requirement of the Hfq C-terminal region for regulation. We show that the C-terminal amino acids that span between positions 66- and 72 are essential for GlmY mediated regulation of GlmS that also involves projected NTPase YhbJ, implying a conditional need for the C-terminal region of Hfq depending on the mode of sRNA recruitment.

3.5 Materials and Methods

3.5.1 DNA oligonucleotides

The complete list of DNA oligonucleotides used for cloning is provided as Supplementary **Table 3.1**.

3.5.2 Bacterial strains, media and growth conditions

The *E. coli* strain Top10 (Invitrogen) was used for cloning, GFP fusion expression analysis and in all that involved co-expression of GFP fusions with sRNAs and Hfq variants. The Δ*hfq* strains were also derived from Top10 cells and were generated as described previously by replacing *hfq* with chloramphenicol or kanamycin resistance cassettes using the Quick and Easy Conditional Knockout Kit (Gene Bridges) (119). In order to comply with the usage of antibiotics, the chloramphenicol resistance cassette was removed from the *hfq::Cam*^{R} strain as described in the Quick and Easy Conditional Knockout Kit (Gene Bridges) operational manual. The successful removal of the *Cam*^{R} cassette was verified by antibiotic screening followed by PCR.

Growth in Luria–Bertani (LB) broth or on LB plates at 37 $^{\circ}$C was used throughout this study. Antibiotics were applied at the following concentrations: 100 µg/mL ampicillin, 100 µg/mL spectinomycin, 34 mg/mL chloramphenicol and 30 mg/mL kanamycin.

3.5.3 Plasmids

A list of plasmids used in this study is listed in **Table 3.2**

Fusion plasmids: All GFP fusion systems were constructed using the plasmid pBacEmGH (Cunningham Lab to be described elsewhere). In brief, a *glmS* mRNA fragment from -162 to +21 relative to translation start site is cloned behind a pBAD promoter to prepare pNS9006. Primers NS20 and NS21 were used to amplify the *glmS*

Table 3.1. List of DNA oligonucleotides (sequences are given in 5'→ 3').

Code	DNA primer name	Sequence
NS 20	010111 pBAC glmS +21 Rev	GGAACCGCTAGCCGCGCCAACAATTCCACAC
NS 21	010111 pBAC glmS -162 For	TACATCCGAGCAGCGGCCGCTTCTGGCCGGCTAACCC
NS 22	GlmS Mutation Amp For	GGAACC GCTAGC CGCGC CAACAATTCCACACATAG
NS 23	GlmS Mut Top1 For	CAACAATTCCACACATAGTTTTTGATTCCGATTTATATCGTTGTCGGTCAA CCTGTATGCCCGTTTTTTTCGGGCGCCCCGAGCCTTGTAGAG
NS 24	GlmS Mut Bot X'Y	GAGGAGATAACATAATCTC CCTCCCACAAGCCGTCACTATAAAAATAACCCCACTCTCTACAAGGCTCG G
NS 25	GlmS Mut Bot XY'	GAGGA GATAACATAATCTC CCTCCCACAAGCAGTAACTATCUACCTAACCCCACTCTCTACAAGGCTCG G
NS 26	GlmS Mut Bot X'Y'	GAGGAGATAACATAATCTC CCTCCCACAAGCCGTCACTATCUACCTAACCCCACTCTCTACAAGGCTC GG
NS 27	GlmS Bot Modified 2	TTCTGGCCGGCTAACCCGGTCACATGGGATGAGGAGATAACATAATC
NS 28	pBAC For (sequencing primer)	CGTTCACCGACAAACAACAG
NS 29	pBAC Rev (sequencing primer)	TTGATTATTTGCACGGCGTC
NS 30	pBAC Rev2 (sequencing primer)	AACTTGTGGCCGTTTACGTC
NS 31	010111 Shape GlmS-For	GGAACCGAATTCTAATACGACTCACTATAGGCCTTCGGGCCAATTCTGG CCGGCTAACC
NS 32	010111 Shape GlmS-rev	CCGAGCGGATCCAATATTGAACCGGACCGAAGCCCGATTTGG ATCCGGCGAACCGGATCGA CGCGCCAACAATTCCAC
NS 33	022111 T7glmS for	GGAACCGAATTCTAATACGACTCACTATAGGTTCTGGCCGGCTAACC
NS 34	010111 glmS Rev	CCGAGCAAGCTTAATATTCGCGCCAACAATTCCAC
NS 35	GlmY pNM12 For	GGAACCTGGCCAAGTGGCTCATTCACCGAC
NS 36	GlmY pNM12 Rev	CCGAGCGAATTCAAAAACAAAGCCGGGAATTAC
NS 37	GlmZ pNM12 For	GGAACC TGGCCA GTAGATGCTCATTCCATCTCTTATG
NS 38	GlmZ pNM12 Rev	CCGAGC GAATTC AAAAAAACGCCTGCTCTTATTACG
NS 39	pNM12 secR (sequencing primer)	CTCATCCGCCAAAACAG
NS 40	pNM12 secR v2 (sequencing primer)	ATTTGATGCCTGGCAGTTCC
NS 41	pNM12 secF (sequencing primer)	ACGGCGTCACACTTTGCTATG
NS 42	Hfq For BamHI	GGAACCGGATCCGATGGCTAAGGGGCAATC
NS 43	Hfq Wt rev HindIII	CCGAGC AAGCTT TTC TCA TTATTCGGTTTCTTCGCTG
NS 44	Hfq65 rev Hind III	CCGAGC AAGCTT TTC TCA TTA AGACGGGACAACAGTAG
NS 45	Hfq72 rev Hind III	CCGAGC AAGCTT TTC TCA TTA ACT GTG ATG AGA AAC CGG
NS 46	Hfq 87 rev Hind III	CCAAGC AAGCTT TTC TCA TTA GCTACCATGATGGTAGTTAC
NS 47	Hfq C perf For BamHI	GGAACC GGATCC G ATGAATAAGTCAATCAATAACC
NS 48	Hfq C perf rev HindIII	CCAAGC AAGCTT TTC TCA TTAATTATCAAAAACTTGCGC
NS 49	Hfq C diff for BamHI	GGAACC GGATCC G ATGAAAAATACAGTTTTAAATTTAC
NS 50	Hfq C diff rev HindIII	CCGAGC AAGCTT TTC TCA CTATCTGTTGTTATTATTATTGTTG
NS 51	Ptac-KanL94P secF	GGCTGTGCAGGTCGTAAATC
NS 52	Ptac-KanL94P secR	CAGCCAAACTATCAGGTCAAGTC
JP01	Hfq C perf short Rev HindIII	CCAAGC AAGCTT TTC TCA AGCTGGTGATACTGTGC
JP02	Hfq C perf Short + Ecoli66-72 Rev HindIII	CCAAGC AAGCTT TTC TCA ACTGTGATGAGAAACCGGGCG AGCTGGTGATACTGTGC

fragment from genomic DNA and inserted into pBacEmGH using NotI and NheI restrictions endonucleases. The plasmid was then verified using DNA sequencing. Preparation of ARN mutants in *glmS* was performed by synthesizing *glmS* using primer extension where mutations were introduced by the primers that were used. To construct pNS 9009, that consists the ARN Δ1 mutation in *glmS*, five cycles of PCR was performed using primers NS23, NS24 and NS27 where the desired mutation was introduced at NS24. The PCR amplification was performed using Taq polymerase according to standard procedure. After the first five cycles, the PCR product was diluted 100-fold into a fresh PCR to amplify the 5'- and 3'- regions of *glmS* using primers NS22 and NS21. The amplified PCR fragment was then subjected to NotI/NheI restriction digestion before cloning into pBacEmGH. Similarly pNS9010 (ARN Δ2, NS23, NS25 and NS27) and pNS9011 (ARN Δ1 Δ2, NS23, NS26 and NS27) were constructed. Successful construction of the plasmid was validated by DNA sequencing.

sRNA plasmids: sRNA plasmids, pNS9007 (GlmZ) and pNS9008 (GlmY) were based on pNM12 (50). sRNA fragments for GlmZ (NS37/NS38) and GlmY (NS35/NS36) were amplified using TOP10 *E. coli* cells. The amplified fragments were then inserted into pNM12 after restriction digestion using MscI/EcoRI. The plasmids were verified by DNA sequencing.

Hfq plasmids: All Hfq constructs were based on the plasmid ptac-KanL94P (Addgene plasmid 15927) (154). Hfq wild type (pNS9012, NS42/NS43), Hfq65 (pNS9013,NS42/NS44), Hfq72 (pNS9014, NS42/NS45), Hfq87 (pNS9015, NS42/NS46) were amplified using Top10 *E. coli* genomic DNA. Hfq mutants Y25A Hfq (pNS9017),

Table 3.2. Description of plasmid constructs used in this study

.	Description
pNS9006	*glmS* (-161-to +21) in pBAC EmGFP (CAM)
pNS9007	GlmZ in pNM12 (AMP)
pNS9008	GlmY in pNM12 (AMP)
pNS9009	*glmS* ΔARN-1 in pBAC EmGFP (CAM)
pNS9010	*glmS* ΔARN-2 in pBAC EmGFP (CAM)
pNS9011	*glmS* ΔARN-1/ΔARN-2 in pBAC EmGFP (CAM)
pNS9012	Wild type Hfq in pTac-KanL94P (SPN)
pNS9013	Hfq65 in pTac-KanL94P (SPN)
pNS9014	Hfq72 in pTac-KanL94P (SPN)
pNS9015	Hfq87 in pTac-KanL94P (SPN)
pNS9016	H57A Hfq in in pTac-KanL94P (SPN)
pNS9017	Y25A Hfq in pTac-KanL94P (SPN)
pNS9018	*C.perf* Hfq in in pTac-KanL94P (SPN)
pNS9019	*C.diff* Hfq in pTac-KanL94P (SPN)
pNS9020	K56A Hfq in pTac-KanL94P (SPN)
pJP9002	*C.perf* Hfq67 in pTac-KanL94P (SPN)
pJP9002	*C.perf* Hfq + *E.coli* Hfq tale (66-72) in pTac-KanL94P (SPN)

K56A Hfq (pNS9020) and H57A Hfq (pNS9016) were amplified using a plasmid as template that was previously described (21). *C.perf* Hfq (pNS9018, NS47/NS48), *C. diff* Hfq (pNS9019, NS49/NS50), *C. perf* Hfq67 (pJP9002, NS47/JP01) and *C. perf* Hfq67* (pJP9003, NS47/JP02) was amplified using *C. perf* and *C. diff* genomic DNA. The PCR fragments were then restriction digested with BamHI and HindIII and inserted into ptac-KanL94P. The plasmids were once again verified by DNA sequencing. A list of strains that were used throughout this segment of study is given in **Table 3.3**

3.5.4 Fluorescence data collection

Fluorescence was measured using a Tecan GENios Plus multi label plate reader.

GFP expression of fusion constructs was measured after lyses of cells that were grown to early stationary phase. All cultures were grown at 37 °C in LB media with appropriate antibiotics at concentrations listed above. Bacterial cultures were generated by diluting overgrown overnight cultures. After 3h of growth, the cultures were induced with 0.1% of arabinose. Cells were then let to grow for another 2.5 h. For Hfq complementation studies using the ptac-plasmid system, Hfq expression was induced using 1 mM IPTG after 1h of growth of diluted cultures before inducing the expression of GlmS-GFP fusion proteins at 3h. The cells were let to grow for another additional 2.5h, before the absorbance was measured at 600 nM to monitor the growth of cells. 3.0 mL of cells were then harvested by centrifuging at 10K rpm for 4 mins. The cells were then re-suspended in 200 µL of lysis buffer (50 mM Tris-HCl pH 7.5, 25 mM NaCl, 2 mM EDTA). To lyse cells 15 µL of lysozyme (20 mg/mL, Fisher), 30 µL of protease inhibitor (complete EDTA-free, Roche) and 30 µL of 1% TritonX-100 was added followed by an

Table 3.3. List of strains used in this study

Strain	Genotype
NS1001	*glmS:gfp*, TOP10
NS1002	*glmS:gfp,* GlmZ, TOP10
NS1003	*glmS:gfp,* GlmY, TOP10
NS1005	*glmS:gfp,* GlmZ, Δ*hfq::kan*
NS1006	*glmS:gfp*, GlmY, Δ*hfq::kan*
NS1009	*glmS* Δ*ARN-1:gfp*
NS1010	*glmS* Δ*ARN-2:gfp*
NS1011	*glmS*Δ*ARN-1/*Δ*ARN-2:gfp*
NS1012	*glmS* Δ*ARN-1:gfp,* GlmZ
NS1013	*glmS* Δ*ARN-2:gfp*, GlmZ
NS1014	*glmS*Δ*ARN-1/*Δ*ARN-2:gfp*, GlmZ
NS1015	*glmS*Δ*ARN-1:gfp*, GlmY
NS1016	*glmS* Δ*ARN-2:gfp*, GlmY
NS1017	*glmS*Δ*ARN-1/*Δ*ARN-2:gfp*, GlmY
NS1018	*glmS* Δ*ARN-1:gfp,* GlmZ, Δ*hfq*
NS1019	*glmS*Δ*ARN-2:gfp*, GlmZ, Δ*hfq*
NS1020	*glmS*Δ*ARN-1/*Δ*ARN-2:gfp*, GlmZ, Δ*hfq*
NS1021	Δ*hfq* (CAM cassette deletion) (STR)
NS1022	*glmS:gfp,* Δ*hfq*
NS1023	*glmS:gfp,* GlmZ, Δ*hfq*
NS1024	*glmS:gfp,* GlmZ, Δ*hfq, hfq-wt*
NS1025	*glmS:gfp,* GlmZ, Δ*hfq, hfq65*
NS1026	*glmS:gfp,* GlmZ, Δ*hfq, hfq72*
NS1027	*glmS:gfp,* GlmZ, Δ*hfq, hfq87*
NS1028	*glmS:gfp,* GlmZ, Δ*hfq,* Y25A *hfq*
NS1029	*glmS:gfp,* GlmZ, Δ*hfq,* H57A *hfq*
NS1030	*glmS:gfp,* GlmZ, Δ*hfq, C,perf hfq*
NS1031	*glmS:gfp,* GlmY, Δ*hfq, C.diff hfq*
NS1032	*glmS:gfp,* GlmY, Δ*hfq*
NS1033	*glmS:gfp,* GlmY, Δ*hfq, hfq-wt*
NS1034	*glmS:gfp,* GlmY, Δ*hfq, hfq65*
NS1035	*glmS:gfp,* GlmY, Δ*hfq, hfq72*
NS1036	*glmS:gfp,* GlmY, Δ*hfq, hfq87*
NS1037	*glmS:gfp,* GlmY, Δ*hfq, C.perf hfq*
NS1038	*glmS:gfp,* GlmY, Δ*hfq, C.diff hfq*
NS1039	*glmS:gfp,* GlmY, Δ*hfq,* Y25A *hfq*
NS1040	*glmS:gfp,* GlmY, Δ*hfq,* H57A *hfq*
NS1041	*glmS:gfp,* GlmZ, Δ*hfq,* K56A *hfq*
NS1042	*glmS:gfp,* GlmY, Δ*hfq,* K56A *hfq*
JP1001	*glmS:gfp,* GlmZ, Δ*hfq, C.perf hfq67*
JP1002	*glmS:gfp,* GlmZ, Δ*hfq, C.perf hfq67**
JP1003	*glmS:gfp,* GlmY, Δ*hfq, C.perf hfq67*
JP1004	*glmS:gfp,* GlmY, Δ*hfq, C.perf hfq67**

incubation of 30 mins at 37 °C while shaking. The lysed cells were then centrifuged briefly. 200 µL of the supernatant was then withdrawn to measure the fluorescence intensities of GFP in a 96-well flat bottom black plate (Coaster®). The fluorescent measurements were made at an excitation wavelength of 485 nM and at an emission wavelength of 535 nM. The detector gain values were consistently set between 95 and 103. All fluorescent measurements were normalized to an identical culture where the expression of GFP was not induced by arabinose to account for cellular fluorescence. All measurements were at least triplicated and error values for each experiment were based on the standard deviation between independent trials.

3.5.5 RNA preparation for SHAPE and gel shift analysis

A_{18} RNA was purchased from Thermo Scientific Dharmacon Technologies (Lafayette, CO) and deprotected following the manufacturer's protocol. RNA quality was assessed using denaturing PAGE and gel purified. For in vitro binding analysis, *glmS* mRNA was transcribed using a DNA fragment that amplifies the *glmS* from -161 to +21 in *E. coli* XL-10 cells using primers NS33 and NS34 and digested with Ssp I before transcribing. For SHAPE analysis, *glmS* mRNA was in vitro transcribed using a DNA template that was amplified in XL-10 cells using primers NS31 and NS32, that includes a structure cassette in the 5' and 3' regions as previously described (132). In vitro transcription was performed after digesting the amplified product with Ssp I. DsrA was obtained by runoff transcription of pBAU10301 that was digested by Ssp I (23).

3.5.6 Electrophoretic mobility shift assays

All binding reactions were performed in 50 mM Tris-HCl pH 7.5,100 mM KCl and 10 mM $MgCl_2$ at room temperature. Prior to any interaction all RNAs in buffer were annealed at

90 °C for 3 min, cooled to room temperature for 30 min. For all reactions 10 µL aliquots were loaded after diluting with loading buffer (10% (w/v) sucrose, xylene cyanol, bromophenol blue) under a power of 5 W on native 5-7% polyacrylamide (37:1) gel in 1X TBE. Dried gels were visualized by phosphorimaging (Molecular Dynamics) using a Typhoon 9210 imaging system (Amersham). Quantification was done using ImageQuant 5.1 (Molecular Dynamics) and Kaleidagraph 3.0 (Synergy). Data were fit using nonlinear least-square analysis to a cooperative binding model shown below (**Eq 3.1**). Here, L is the ligand concentration and the cooperatively is indicated by n. Typical values for n ranged from 1.5 to 2.1.

$$Q_{Fraction} = \frac{(L)^n}{K_D + L^n} \qquad (\mathbf{3.1})$$

In the case of A_{18}, DsrA competition assays, the *glmS*•Hfq complex was pre-formed and A_{18} and DsrA was titrated at varying concentrations from 0 to 30 µM.

3.5.7 Chemical SHAPE analysis

SHAPE, as described previously(132,136), was performed to determine the secondary structure of *glmS*. To fold the RNA 1 pmol was heated to 95°C in a buffer containing 10 mM Tris-HCl pH 7.5 and 60 mM KCl. The mixture was then cooled to RT for 5 min. and 10 mM $MgCl_2$ was added. The RNA was then incubated at 37°C for 20 min. After folding, N-methylisatoic anhydride (NMIA) in anhydrous DMSO was added to a final concentration of 3 mM. The reaction (NMIA+) was incubated at 37°C for 45 min. along with a control sample that was folded in an identical way but had only DMSO added in the absence of NMIA (NMIA-). After incubation the RNA was precipitated in the presence of 20 µg glycogen.

Primer extension was performed to analyze adduct formation. The NMIA+ and NMIA- RNAs were heated in a PCR tube to 95°C for 1 min. in 1X TE, total volume 10 µL, and then cooled to 4°C for 1 min. Then 3 µL of a fluorescently labeled primer (5′-F-GAACCGGACCGAAGCCCG) complementary to the 3' end of the RNA was then annealed at 65°C for 5 min. followed by cooling to 37°C for 15 min. Primers with the WellRED D4 (0.4 µM) fluorophore was added to the NMIA+ reaction and primer with the WellRED D3 (0.6 µM) fluorophore was added to the NMIA- reaction. After primer annealing, enzyme mixe was added to the reaction (4 µL of Superscript III FS buffer, 1 µL 0.1 M DTT, 1 µL 10 mM dNTP mix) and incubated at 52°C for 1 min. followed by addition of 0.5 µL SuperScriptIII RT. Primer extension was carried out at 52°C for 30 min. In addition to NMIA+/- reactions two sequencing reactions were performed in order to be able to align reactivity peaks with the sequence of the RNA. The reverse transcription for these was performed using NMIA- RNA in a similar manner as mentioned above but with different fluorophores, 2.2 µM WellRED D2 and 2.2 µM WellRED D2 and the addition of 1 µL, 2.0 mM ddCTP and 1 µL, 1.5 mM ddGTP respectively. After incubation all four reaction were then combined together and precipitated in the presence of 20 µg glycogen. The sample was then re-dissolved in SLS (Beckman) and separated by capillary electrophoresis on a Beckman CEQ 8000 DNA sequencer. The following parameters were used for separation (capillary temp: 60°C, denature temp: 90°C, time 150 s, injection voltage: 5 kV, time 20 s, separation voltage 3 kV and separation time 100 min).

The data obtained was analyzed using ShapeFinder(134). The SHAPE values were normalized to a scale of 0-1 where 0 is unreactive and the most reactive sites have an

average value of 1. Details of the normalization are described elsewhere(136). The reactivities were then input into the RNAstructure software where they were converted to pseudo-free energy changes(163).

3.5.8 Chemical SHAPE footprinting

To determine the Hfq binding sites on *glmS* mRNA, SHAPE was performed in the presence and absence of Hfq and Hfq mutants (Y25A, K56A). 2 pmol *glmS* mRNA was folded as described above and then incubated with and without Hfq and Hfq mutants at RT for 30 min. to allow binding. The RNA was modified as described above and the NMIA was quenched by added 1 volume of 250 mM DTT followed by digestion with proteinase K (10 µL, 20 mg/mL) for 30 min. at 37°C. Phenol chloroform extraction was performed and the RNA was precipitated. Primer extension and ShapeFinder analysis was performed as described previously.

3.6 Acknowledgements

I would like to thank Martha Faner for help in solving the *glmS* secondary structure using SHAPE.

CHAPTER 4

Thermodynamic and Kinetic Analysis of an RNA Kissing Interaction and Its Resolution into an Extended Duplex

4.1 Abstract

Kissing hairpin interactions form when the loop residues of two hairpins have Watson-Crick complementarity. In a unimolecular context, kissing interactions are important for tertiary folding and pseudoknot formation, whereas in a bimolecular context, they provide a basis for molecular recognition. In some cases, kissing complexes can be a prelude to strand displacement reactions where the two hairpins resolve to form a stable extended intermolecular duplex. The kinetics and thermodynamics of kissing complex formation and the subsequent strand displacement reactions are poorly understood. Here, biophysical techniques including isothermal titration calorimetry, surface plasmon resonance and UV melting have been employed to probe the factors that govern the stability of kissing complexes and their subsequent structural rearrangements. We show that the general understanding of RNA duplex formation can be extended to kissing complexes but that kissing complexes display an unusual level of stability relative to simple duplexes of the same sequence. Also we compare the thermodynamic parameters between RNA kissing and duplex formation to understand the partitioning of the energetics during complex formation. Furthermore, activation energies were calculated using the temperature dependence of kissing complex kinetics. With a collaborative study with the David Rueda Lab, smFRET experiments were performed on kissing complex formations to understand the molecular basis of these interactions. Overall, this work leads us to understand the energetics of kissing complex formation and the structural transitions associated with it.

4.2 Introduction

In recent years, reports of riboregulation have become commonplace, dramatically changing our view of the central dogma and the way in which gene regulation is controlled (164,165). Regulation may occur co-transcriptionally as in the case of most metabolite sensing riboswitches (166,167) or post-transcriptionally as in the case of the small bacterial regulatory RNAs (168,169) or eukaryotic microRNAs (170,171). Riboswitches are structures that allow small molecule binding to an aptamer domain to control a unimolecular rearrangement, allowing the RNA to adopt either a termination or an anti-termination structure, and hence affect the transcription of downstream genes (166,167). Post-transcriptional regulation typically requires two RNAs to recognize one another (a sRNA and its cognate mRNA target), sometimes on their own, but more commonly with the assistance of RNA binding proteins (168,172). The resulting complex can then be recognized and acted upon by cellular machinery. In its simplest form, evident in the biology of bacterial sRNAs, up-regulation occurs because refolding exposes regulatory sequences in the mRNA such as the ribosome binding site, or conversely, the translation is repressed by hiding such sequences (173). More complicated patterns of regulation can be induced by recruitment (or displacement) of RNA binding proteins to either degrade the message or localize it within the cell. All of these processes depend integrally upon RNA structural rearrangements and must balance issues of stability and specificity of the RNA-RNA and RNA-protein contacts that stabilize the various structural states.

Stability of these complexes must be finely tuned. If the initial contact yields a species that is too stable, selectivity is lost since release of the non-cognate partner

becomes slow. Conversely, if the initial complex interacts too weakly, its lifetime is too short to allow stable pairing interactions to form and the benefits of co-localization are lost. Thus, the initial contact must fall within a relatively narrow range of lifetimes to be beneficial.

In the case of bacterial and viral systems, loop-loop contacts (kissing interactions) are particularly prevalent in regulatory complexes. Numerous examples of rearrangements are known that depend upon initial formation of loop-loop interactions. Complexes between CopA and CopT RNAs and between RNAI and RNAII were among the earliest members of this class of RNAs to be studied (174,175). Both systems control plasmid replication and copy number. In the biology of HIV, the dimerization initiation site sequences (DIS) form kissing complexes and ultimately resolve into an extended duplex that dimerizes the HIV genome for packaging into capsids (176-179). A plethora of examples from the biology of bacterial stress responses are also known. For instance, during oxidative stress, the sRNA OxyS regulates *fhlA* mRNA among other targets (30,123). This process was shown to involve the formation of loop-loop contacts between the sRNA and its mRNA partner (2,60). The kissing interaction alone is sufficient to induce translational repression of *fhlA*. In other cases, like that of the sRNA DsrA involved in cold shock response, the regulatory RNA recognizes its mRNA target *rpoS* through an initial RNA-RNA contact with the help of Hfq followed by an intermolecular strand displacement reaction to form an extended region of base pairing (38,180,181). Proteins such as Hfq are essential for these types regulatory RNA-RNA interactions that act either by increasing the local concentration of binding partners and/or triggering structural changes of RNAs for efficient target recognition (63).

Kissing interactions occur when complementary sequences in apical loops of two hairpins can form Watson-Crick base-pairing (182). The bases in these hairpin loops are often more accessible than duplex regions and are thus particularly well adapted to initiate sequence-dependent molecular recognition. The stability of these kissing complexes are primarily based on loop complementarity, but factors such as the orientation of the loops, internal loop structure, the loop closing base pair, and the sequence of the stems adjacent to the loops are all known to contribute to the stability of these interactions (174,183,184). Despite the prevalence of these interactions, information on the thermodynamic and kinetic stabilities of these kissing hairpins is relatively scarce (185).

Multi-step structural rearrangements involving transient metastable intermediates (i.e. kissing complexes that resolve into long-lived, thermodynamically stable duplexes), help to balance the needs for stability and selectivity in biological signaling networks (40). The paradigm for a kissing complex that resolves to an extended duplex is HIV DIS but several other systems are also known. As illustrated in **Figure 4.1**, the strand displacement reaction can initiate through two possible mechanisms at modest temperatures: (a) nucleation through the 3'/5' termini resulting in formation of a zipper intermediate, or (b) nucleation through the complementary hairpin loops resulting in a kissing complex (186,187). At elevated temperatures where small amounts of the

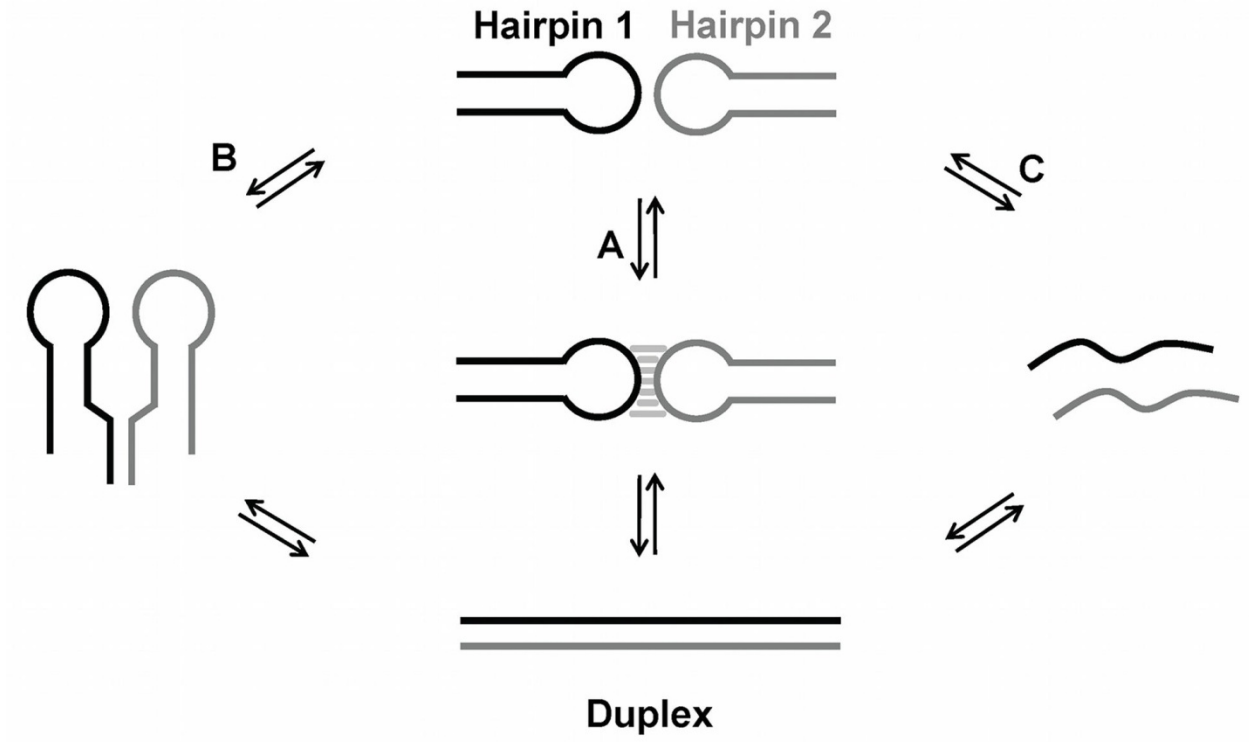


Figure 4.1. Possible pathways for strand displacement reactions. Strand displacement can be nucleated by formation of a kissing complex (A), by initiation at the termini (B), or through complete unfolding of the hairpins (C).

unfolded state might be present, or in the presence of helicases that can actively unfold the RNAs, an unfolding and annealing pathway (Pathway C in **Figure 4.1**) is also accessible.

In this study, we have used ITC, surface plasmon resonance (SPR), UV melting and single-molecule Förster Resonance Energy Transfer (smFRET) (the latter in collaboration with the David Rueda lab) to study the kinetic and thermodynamic properties of a series of RNA hairpins that can form kissing complexes and their subsequent rearrangement to an extended duplex – a classical strand displacement reaction such as might occur during a riboregulation event. To deconstruct the steps along the reaction pathway and probe the energy landscape, the reaction can be forced to stop at the kissing complex by removing sequence complementarity within the stem regions. Thus, we can isolate the individual steps as the reaction proceeds from independent hairpins, to kissing complexes and finally to extended duplexes. For simplicity, in this study all proteins have been removed from the system, although we recognize that in certain cases, reactions such as these would be facilitated by RNA chaperones such as Hfq or NCp7.

Our results show that the thermodynamic trends for kissing complex formation is similar to RNA duplexes although KCs were consistently more stable. By measuring the kinetic profiles using SPR and smFRET (in collaboration with the David Rueda lab) in a variety of KCs, we show that the recognition between two RNAs are often under kinetic control and that the final fate of the complex is dictated by the dissociation phase. We also show the need for chaperone proteins such as Hfq to trigger structural transitions beyond the initial recognition of the two RNAs. Overall our work describes a potential

thermodynamic and kinetic model for kissing complex formation and associated structural changes.

4.3 Results

4.3.1 Design of RNA hairpins

In the present study kissing interactions have been probed between the complementary loops of short hairpins to measure the thermodynamics for these processes. Transformation of the kissing complex (KC) into extended duplex (ED) conformation was also measured to understand the energetic landscape faced by regulatory interactions that rely on this type of structural dynamics. The resolution of a pair of hairpins into an extended duplex described here is the extreme case of structured single-stranded RNAs undergoing a transition to a bimolecular duplex.

The hairpin constructs used in the study are shown in **Figure 4.2**. They derive from two parental species we call HP1 and HP2 that have been studied previously (188). The parent hairpins are complementary in both their loop and stem regions, enabling them to form a KC and also to resolve into the ED state. HP3 derives from HP1, but lacks complimentarity with HP2 within the stem region arresting the reaction at the level of the KC. To assess the effects of loop size and sequence on the overall energetics, mutations were introduced as shown in **Figure 4.2**.

Each of these hairpins was subjected to thermal melting analysis prior to initiating biophysical analysis on these systems. The thermodynamic parameters for each hairpin are shown in **Table 4.1**. In addition to the isolated hairpins, the bi-molecular extended

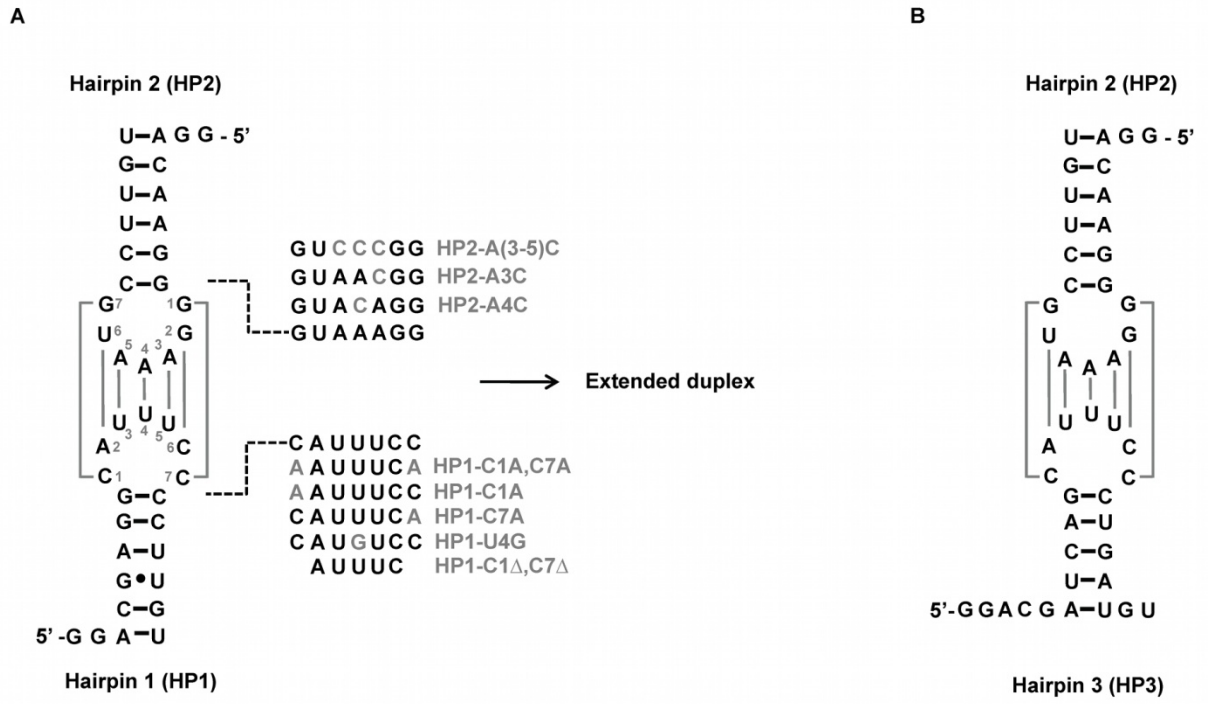


Figure 4.2. Schematic diagram of the hairpin constructs used in this study. (A) Hairpins capable of rearranging into extended duplex complexes. (B) Hairpin sequences incapable of further rearrangements.

Table 4.1. Summary of thermodynamic parameters obtained from UV melting experiments.

Construct	1 M NaCl		10 mM $MgCl_2$	
	T_M (°C)	ΔH^{UV} (ΔH) (kcal/mol)	T_M (°C)	ΔH^{UV} (ΔH) (kcal/mol)
Hairpins				
HP3	60.6	-47	62.5	-57
HP1-C1A	63.9	-51	76.2	-64
HP1	60.8	-49	66.3	-38
HP2	71.5	-41	76.1	-42
HP2-A3C,A4C,A5C	72.1	-28	81.9	-47
HP1-U4G	61.7	-55	67.8	-47
HP2-A4C	71.2	71	76	51
Duplexes				
HP1:HP2	84.7	-110	84.3	-154
HP1:HP2-A3C,A4C,A5C	74.6	-52	67.8	-43

[a] The estimated error for ΔH^{UV} (ΔH) for all individual data points is ± 1 kcal/mol and T_M is ± 0.4°C, based on replicate measurements.

duplexes of HP1::HP2 and HP1::HP2-A(3-5)C were also assayed. The melting temperatures (T_M) of the hairpins were independent of sample concentration showing that the observed transitions were uni-molecular and thus represent the hairpins and not the dimeric forms of the molecules.

4.3.2 Thermodynamic analysis of kissing interactions by ITC

Thermodynamics of the kissing interactions were probed using isothermal titration calorimetry (ITC). ITC is a powerful method for measuring the thermodynamics of molecular interactions of bio-molecules (189-192). Since calorimetry uses the heat of binding rather than hyperchromicity for detection, it is typically much more sensitive to kissing interactions than thermal melting studies. In addition, it allows one to probe temperature dependence of the reactions to look at their detailed thermal profiles. Since the formation of stably folded structures requires the compensation of the negative charges of backbone phosphates (whose electrostatic repulsion inhibits close packing), titrations were performed in the presence of either 1 M NaCl or 10 mM $MgCl_2$.

HP2 and HP3 can form up to 7 base pairs within the kissing complex of which 4 are AU pairs (**Figure 4.2**). Native gel electrophoresis was used to confirm that these sequences folded into single stable hairpins without contaminating duplex structures after annealing (**Figure 4.3**). Optical melting studies could not detect the kissing interactions because of the small change in hyperchromicity between the free and bound states (data not shown) but the interaction was clearly evident by ITC. A representative titration for the binding of HP2 to HP3 with formation of a KC is shown in **Figure 4.4A**. Thermodynamic parameters obtained for this interaction under various conditions are summarized in **Table 4.2**.

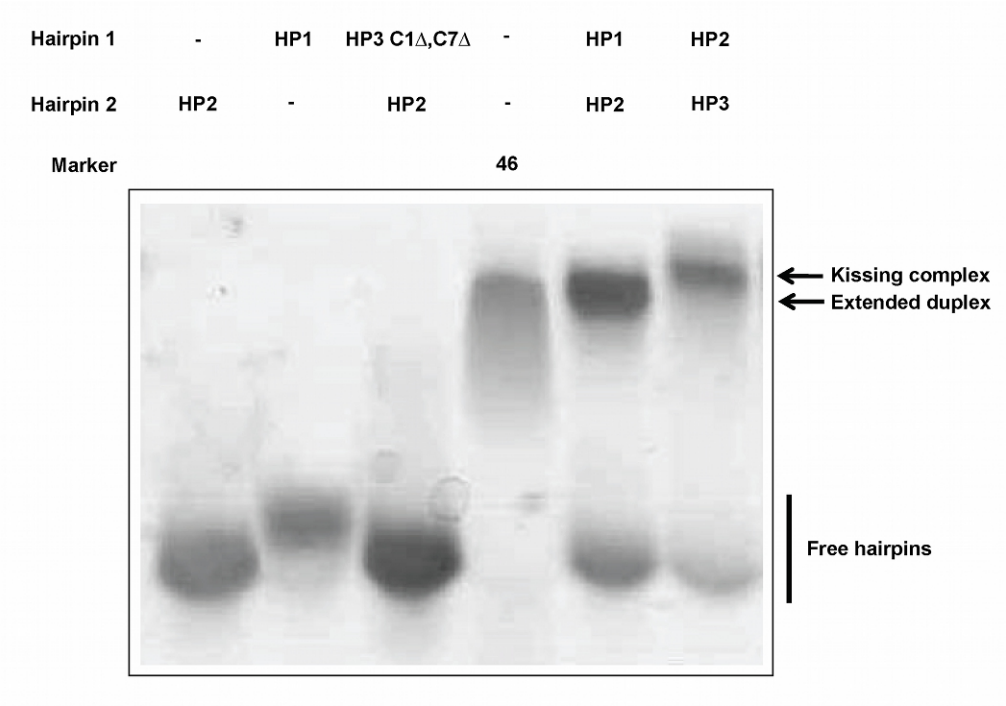


Figure 4.3. Native gel analysis of kissing and extended duplex formation. Lanes 1 and 2 consist of hairpins HP2 and HP1. In lane 3, the interaction between HP3-C1Δ,C7Δ::HP2 is probed. Lanes 5 and 6 probes for extended duplex formation between HP1 and HP2 and the kissing complex between HP1 and HP3. Lane 4 comprises a 46 size marker.

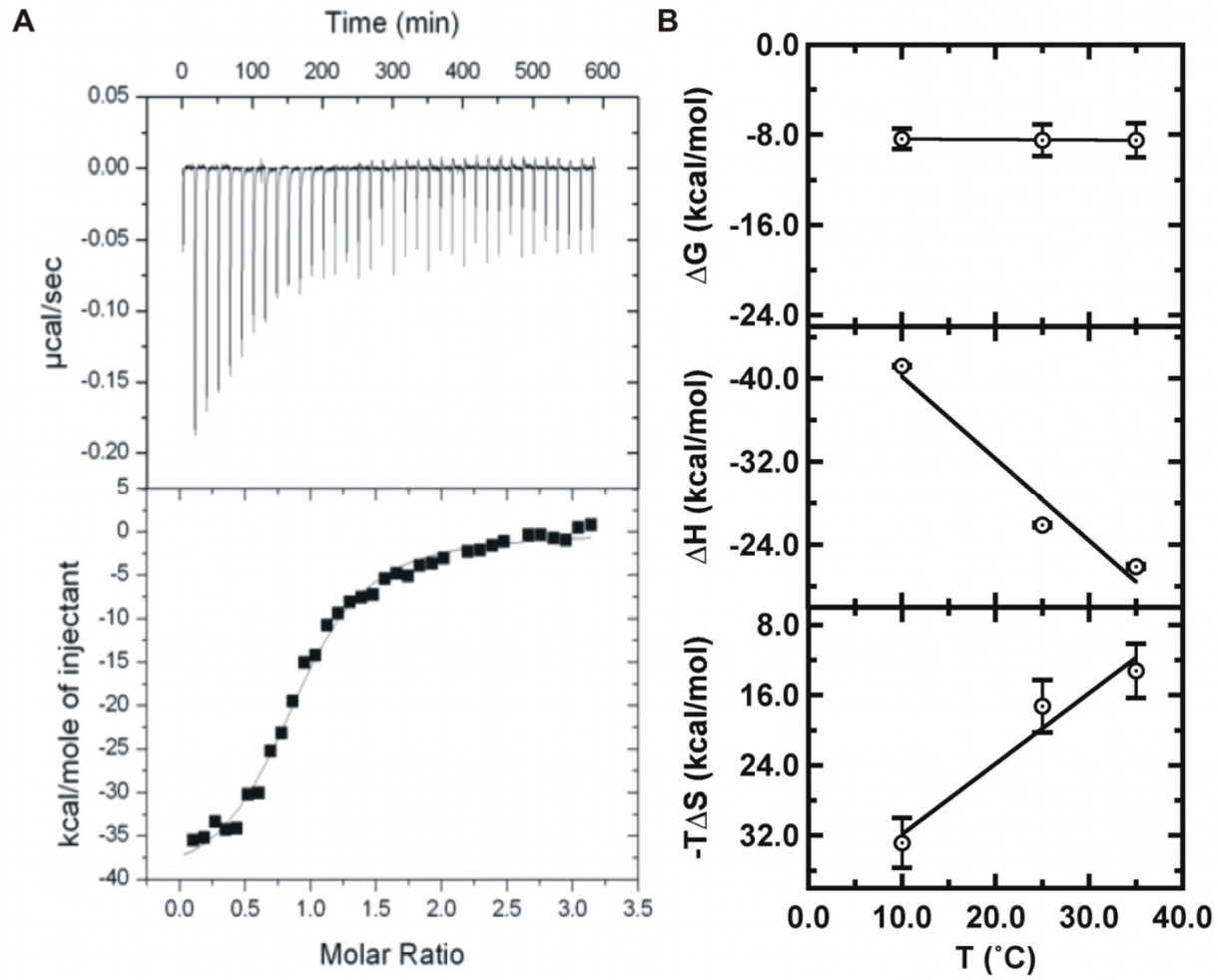


Figure 4.4. Representative ITC data. (A) Thermograms showing formation of the HP2::HP3 kissing interaction. HP3 solution (60 µM) was titrated into 1.4 mL HP2 (3.5 µM) at 10 °C in sodium binding buffer. Integrated data (solid squares, lower panel) was fit to a single-site binding model (solid line), yielding: ΔH = -41.2 kcal mol^{-1}; $-T\Delta S$ = 32.8 kcal mol^{-1} K^{-1}; K_A = 2.8 x 10^6 M^{-1}; n =1.0. (B) Thermodynamic energies ΔG (top), ΔH (middle) and $-T\Delta S$ (bottom) as a function of temperature.

Table 4.2. Thermodynamic parameters obtained by ITC for the kissing interaction of HP1::HP3 and HP2-A4C::HP3-U4G in 10 mM NaHEPES, pH 7.5. N.B. – no binding; n.d. – not determined.

System	Ionic Condition	T (˚C)	ΔH kcal/mol	ΔS cal/mol•K	ΔG kcal/mol	K_a x 10^6 (M^{-1})	n
Kissing complex (HP2::HP3)	1 M NaCl	10	-41.2 ± 0.1	-116 ± 10	-8.4 ± 0.9	2.8 ± 0.3	0.9
		25	-25.9 ± 0.2	-58 ± 10	-9 ± 1	1.7 ± 0.3	0.9
		35	-21.7 ± 0.2	-43 ± 10	-9 ± 2	1.1 ± 0.2	0.9
		45	N.B.	-	-	-	-
	10 mM $MgCl_2$	10	-14.9 ± 0.3	-19 ± 10	-10 ± 1	38 ± 1	0.9
		25	-29.9 ± 0.7	-69 ± 10	-9 ± 2	8 ± 1	0.9
		35	n.d.	-	-	-	-
		45	-29.9 ± 0.2	-67 ± 10	-9 ± 1	1.0 ± 0.1	0.9
Kissing Complex (HP2-4C)::(HP3-4G)	1 M NaCl	10	-26 ± 1	-56 ± 9	-9.8 ± 0.1	37 ± 1	0.9
		15	-28 ± 3	-64 ± 12	-9.9 ± 0.1	37 ± 6	0.9
		20	-39 ± 1	-99 ± 8	-10.3 ± 0.9	130 ± 90	0.8
		30	-35 ± 8	-79 ± 26	-10.7 ± 0.5	690 ± 40	0.9
		40	-23 ± 1	-38 ± 10	-10.7 ± 0.5	31 ± 4	1.0
	10 mM $MgCl_2$	15	-34.4 ± 1	-83 ± 9	-10.5 ± 0.5	90 ± 1	0.9
		20	-39.4 ± 1	-99 ± 10	-10.4 ± 0.5	55 ± 1	0.9
		25	-39.7 ± 1	-98 ± 8	-10.5 ± 0.5	50 ± 9	0.8
		30	-46.0 ± 1	-119 ± 13	-10.2 ± 0.5	23 ± 4	0.9
		40	-37.6 ± 1	-86 ± 10	-10.5 ± 0.5	51 ± 2	1.0

Thermodynamics for the kissing interaction between HP2 and HP3 measured in 1 M NaCl are plotted as a function of temperature in **Figure 4.4B**, exemplifying the results. The heats of binding for the kissing interaction was measureable by ITC between 10 – 35 °C, but unmeasurable at 45 °C indicating a melting transition between 40 – 45 °C. As shown in **Figure 4.4B**, ΔG of the kissing interaction was essentially unchanged over this range, but analysis of ΔH and $T\Delta S$ shows significant enthalpy-entropy compensations, potentially originating from the reorganization of hairpin loops as a function of temperature. Significant ion-dependent effects were observed for these reactions as well, consistent with this interpretation. In 1 M NaCl, KC formation displayed a positive heat capacity change (ΔC_P) of 0.8 ± 0.2 kcal/mol K^{-1}. ΔH was most favorable at low temperatures. On the other hand, in 10 mM $MgCl_2$ the trends were reversed yielding a ΔC_P of -0.3 ± 0.2 kcal/mol K^{-1}. The reaction was entropically opposed at all temperatures but more so at lower temperatures. Furthermore, the overall free energy for KC formation was stabilized by ~1 kcal/mol in 10 mM $MgCl_2$ relative to 1 M NaCl (**Table 4.2**). ITC experiments cannot determine whether the observed differences derive from changes in the ground state (conformation of the loop nucleotides) or product state (altered co-axial stacking in the KC) although we believe loop confirmation may be the more likely of the two explanations based on the ion-dependent behavior.

To analyze sequence effects on the energetics of KCs, mutations were introduced within the loops. First the central adenosine (A4) in HP2 was mutated to cytosine (HP2-A4C) together with its counterpart within HP3 to form HP3-U4G – changing the central A•U base pair of the kissing interaction into a G•C. Thermodynamics were measured using ITC and are shown in **Table 4.2**. An extra

stabilization of ~2 kcal/mol in 1 M NaCl and ~1 kcal/mol in 10 mM $MgCl_2$ was observed for this mutant relative to the parental species, presumably originating from the G•C

base pair. Heat capacity changes (ΔC_P) of -0.5 ± 0.1 and -0.7 ± 0.1 kcal/mol K were measured in NaCl and $MgCl_2$ respectively for this KC. In contrast to HP2::HP3, the entropy-enthalpy compensation here was counter ion independent, implying that structure within the uni-molecular loop is responsible for ion dependent ΔC_P effects, similar to previous observations on duplex systems (193,194).

Mutagenesis of the loops indicated that the stability of this kissing interaction is near the edge of the viable range for ITC analysis. Removal of 2 bases from the loop (HP3-C1Δ,C7Δ) greatly destabilized the interaction with HP2. HP1-C1Δ,C7Δ::HP2 complex was neither stable on native gels at 4°C (**Figure 4.3**) nor was it detectable by ITC at temperatures even at the lowest accessible temperatures (5 °C) in 10 mM $MgCl_2$. The origin of this instability could be from diminished base pairing or from a reduction in the loop size. To distinguish between the two possibilities, the hairpin HP1-C1A,C7A was constructed that retained the 7 nt loop of HP3 but contained only 5 nts of complementarity. This construct also failed to form a stable kissing complex. Thus, the loss of the G•C base pairs at either end of this kissing interaction leads to significant destabilization.

4.3.3 Thermodynamics of kissing complex vs. duplex formation

We next compared the thermodynamics of the KCs to simple duplexes of the same sequences. This comparison allows us to determine if the sequence- and loop-

Table 4.3. Thermodynamic parameters measured using ITC in 1M NaCl for RNA duplexes with sequences associated with kissing hairpins.

System	Temperature ° C	ΔH (kcal/mol)	ΔS (cal/mol K)	ΔG (kcal/mol)
	5	-39 ± 1	-108 ± 3	-9.6 ± 0.8
	10	-39 ± 3	-104 ± 11	-9.5 ± 0.1
	15	-43 ± 2	-117 ± 6	-9.1 ± 0.1
	15	-37 ± 4	-93 ± 12	-10.4 ± 0.2
	20	-50 ± 6	-136 ± 18	-10.3 ± 0.4
	25	-61 ± 1	-171 ± 1	-10.2 ± 0.1
	5	-15 ± 3	-27 ± 10	-8.1 ± 0.5
	10	**No binding in ITC**		
	15			
		**No binding in ITC**		

size effects within the kissing interaction originated from the structural context of the hairpin loops or were inherent to sequences themselves. Thus, RNA2::RNA3 is equivalent to HP2::HP3, just removed from the constraint of the kissing complex hairpins. Thermodynamic values measured in 1 M NaCl are provided in **Table 4.3**. Similar trends in stabilities were observed for duplexes and KCs of the identical sequence. Stable duplexes formed in the case of the wt sequence and the central GC mutant (A4C::U4G). The duplex with terminal GA mismatches (HP2::HP3-C1A,C7A) was detected at 5 °C in ITC but the shortened HP3 with 5' and 3' overhangs was not sufficiently stable to be observed. These results suggests that the inherent sequence- and loop-size effects observed in the ΔG of the kissing KCs simply recapitulated stability issues of the underlying duplexes.

Several interesting characteristics were revealed in the ITC however. Enthalpic and entropic components for duplex and kissing complex formation vary significantly from each other. In the context of a simple duplex, the energetics were more enthalpically favorable and more entropically opposed relative to their related KC. Duplex formation was accompanied by a sizeable ΔC_P, -0.4 kcal $mol^{-1}K^{-1}$ for the parental sequence and a much larger -2.4 kcal $mol^{-1}K^{-1}$ for the duplex with the central GC pair (RNA2-A4C::RNA3-U4G). The -0.4 kcal mol^{-1} K^{-1} value is typical of simple duplexes which have an average of -42 cal mol^{-1} K^{-1} (195), while the larger sequence is more indicative of a species with a latent structure in the single stranded state. The observed trends in ΔC_P are consistent between the kissing complexes and duplexes. These data suggests that thermodynamics principles that govern duplex formations can essentially be extrapolated into kissing complexes.

4.3.4 Thermodynamic analysis of strand displacement reactions

In the experiments described above, hairpins could only form kissing complexes and could not proceed further because the stems were non-complementary. Next the sequences were allowed to resolve into the ED conformation. Using ITC, HP1 was titrated into HP2 to determine the energetics of the strand displacement reaction as a function of temperatures in either 1 M NaCl or 10 mM $MgCl_2$ (**Table 4.4**). Thermodynamic parameters for these reactions are quite different from those associated with kissing complex formation indicating that the reaction proceeds rapidly to the ED state under these conditions. Similar to the formation of the kissing complexes, there is a significant ΔC_P associated with these transitions and the magnitude is highly dependent on the ionic strength of the solution. Using the approximation that ΔC_P is temperature independent, values of -1.1 ± 0.1 and -1.8 ± 0.1 kcal mol^{-1} K^{-1} in 1 M NaCl and 10 mM $MgCl_2$ were observed for the rearrangement, respectively. The more negative heat capacity change was once again associated with the $MgCl_2$ reaction conditions.

4.3.5 Kissing loop interaction is not essential for subsequent extended duplex formation

If ED formation were dependent on formation of an intermediate kissing complex, mutations that reduce or abolish loop complementarity should prevent the reaction. To determine the effect of destabilizing the kissing interaction on strand displacement, all 3 adenosines in the loop of HP2 were substituted with cytosines (HP2-A(3-5)C). Both native gel analysis and ITC at 5 °C failed to show the formation of a stable kissing interaction, as expected, but strand displacement still proceeded rapidly to form ED

Table 4.4. Thermodynamic parameters obtained for strand displacement reaction of HP1::HP2 and HP1::HP2-A(3-5)C to form the extended duplex in 10 mM Na HEPES, pH 7.5.

Interaction	Ionic Condition	T ˚C	ΔH (kcal/mol)	ΔS[a] (cal/mol K)	ΔG (kcal/mol)	K_a x 10^6 (M^{-1})	n
HP1::HP2	1 M NaCl	15	-37.3 ± 0.9	-99	-8.8 ± 0.9	3.8 ± 0.4	1.0
		25	-52.3 ± 0.3	-146	-9 ± 1	2.6 ± 0.3	0.9
		35	-62.3 ± 0.2	-171	-9.6 ± 0.6	7.7 ± 0.5	1.0
		45	-75.9 ± 0.2	-208	-10 ± 1	4.3 ± 0.5	1.0
	10 mM $MgCl_2$	15	-32.7 ± 0.6	-80	-10 ± 1	17 ± 2	1.0
		25	-52.8 ± 0.2	-147	-9 ± 2	3.8 ± 0.7	0.8
		35	-76 ± 1	-214	-9.8 ± 0.4	8.1 ± 0.3	0.9
		45	-84 ± 1	-234	-10 ± 1	5.7 ± 0.8	0.9
HP1::HP2-A(3-5)C	1 M NaCl	25	-34.2 ± 0.6	-86	-9 ± 1	1.9 ± 0.2	0.9
		45	-70.4 ± 0.3	-192	-9 ± 1	2.7 ± 0.5	0.9

[a] The estimated error for ΔS for all individual data points is ± 10 cal/mol K.

(**Table 4.4**). Since these RNAs cannot kiss to any appreciable extent, these data indicate that formation of a stable kissing complex is not required for strand displacement. Under these conditions the fraying pathway (pathway B in **Figure 4.1**) is accessible. Calorimetry showed relatively small changes in ΔG between KC and ED states at these temperatures. The enthalpic and entropic components of the ED state are, however, quite different from the KC - with much greater enthalpic stabilization in the case of the extended duplex. The kinetic behaviors of the two species are also dramatically different as described below.

4.3.6 Kinetic analysis of kissing complex formation

To measure the kinetics of kissing complex formation surface Plasmon resonance (SPR) and smFRET were used. For SPR, 5'-biotinylated hairpins were bound to streptavidin coated sensor chips. The complimentary hairpin was then introduced at varying concentrations to monitor the interaction in 10 mM $MgCl_2$. Sensorgram traces were fit to a Langmuir binding model to obtain association and dissociation kinetics. A representative SPR sensorgram for HP2::HP3 kissing complex formation is shown in **Figure 4.5A**. Kinetic data obtained for HP2::HP3 and HP2-A4C::HP3-U4G kissing complex formation are collected in **Table 4.5**. Both wild type and HP2-A4C::HP3-U4G associate with moderate on-rates that were essentially temperature independent. On the other hand, off-rates for HP2::HP3 were faster at higher temperatures but changed minimally for HP2-A4C::HP3-U4G. Stability constants for kissing complex formation were calculated using kinetic data obtained from SPR (**Table 4.5**) and are comparable to values measured by ITC (**Table 4.2**).

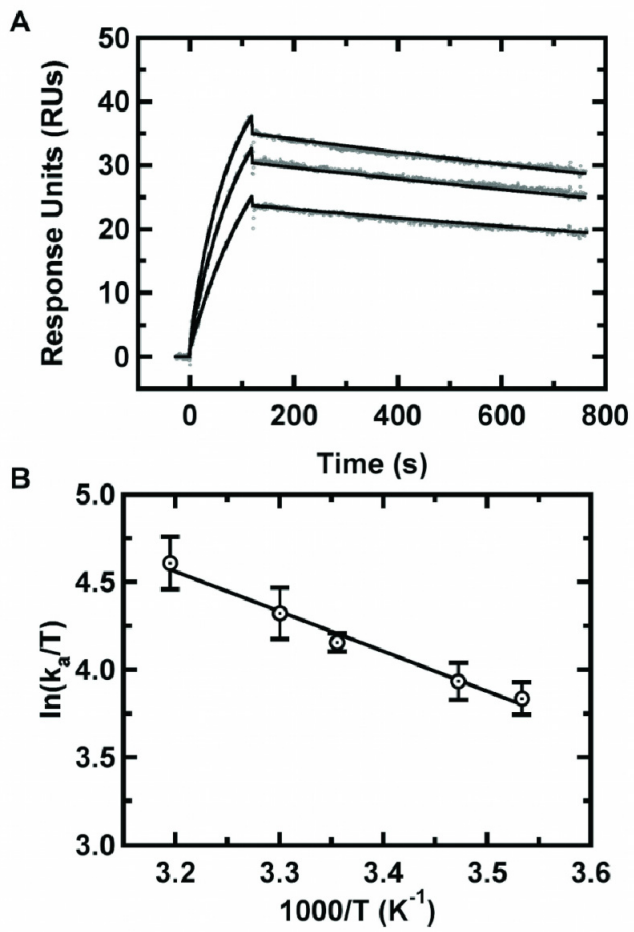


Figure 4.5. Kinetics of HP2::HP3 kissing complex formation measured by SPR. (A) Representative sensorgrams as a function of HP3 concentration (400–800 nM) with surface immobilized HP2 RNA monitored at 15 °C. Association and dissociation data were fit into a Langmuir binding model. (B) Eyring plot of HP2::HP3 association rates studied between 10–40 °C.

4.3.7 Activation energetics for kissing complex formation

As depicted in **Figure 4.1**, kissing interactions can serve as key intermediates to structural transitions such as extended duplex formation. To understand the transition state thermodynamics of kissing complexes relative to ED formation, the temperature dependence of the rates for association and dissociation were probed. The measured rates were then subjected to Eyring analysis to provide activation parameters $\Delta H^{\ddagger}$, $\Delta S^{\ddagger}$, and $\Delta G^{\ddagger}_{298}$. A representative Eyring plot for HP2::HP3 KC is shown in **Figure 4.5B** with clean linear Eyring behavior.

Activation parameters for HP2::HP3 and HP1-U4G::HP2:A4C complexes are collected in **Table 4.5**. Similar activation energies were observed for the association phase of both constructs. In contrast, activation parameters for dissociation varied significantly between the parent and G•C mutant. Both entropic and enthalpic components were much more favorable for the dissociation of HP2::HP3 compared to HP2-U4G::HP3-A4C. These data suggest that the stabilities of these two complexes are dictated mainly by their dissociation transition states while the energy landscapes for association are similar. From equilibrium measurements it was observed that, HP2-U4G::HP3-A4C formed a relatively stronger complex with a $\Delta\Delta G_{298}$ of ~1.1 kcal/mol consistent with ITC data.

4.3.8 Kinetics of kissing complex formation using smFRET[‡]

smFRET was also used to study the kinetics of these kissing complexes. The main difference between the two methodologies is, in SPR the physical observables represents the average ensemble of the system, whereas in smFRET, changes at the molecular level can be measured. Kinetic rates measured in smFRET experiments were consistently ~100 fold faster for both association and dissociation rate constants. For instance association and dissociation rates for the kissing interaction between HP2 and HP3 were, k_{on} = 3.4 ± 0.3 x 10^6 $M^{-1}s^{-1}$ and k_{off} = 0.17 ± 0.02 s^{-1} respectively. Similar observations were reported when smFRET and SPR was used to measure the kinetics of nucleotides binding to DNA polymerase (Lou Romano and David Rueda, unpublished results). Differences in surface passivation in smFRET slides and SPR sensor chips and also issues related to mass transport into the boundary layer may contribute to these kinetic effects. The origins to these anomalies are currently under investigation. However, ΔG values estimated from the rate constants measured by smFERT was in excellent agreement with the ensemble averaged ITC data (ΔG = -9.9 ± 0.2 kcal/mol). The agreement between the single molecule study and bulk measurement indicates that the immobilization approach does not affect the dynamic behavior of the kissing hairpins or their equilibrium.

smFRET was also used to measure the kinetics of kissing interactions of mutated hairpins described above (**Figure 4.6)**. KCs were detected for HP1-U4G::HP2-A4C, HP1-C7A::HP2 and HP1-C1A::HP2. Here the association rate constant was shown to be invariant at ~ 4x10^6 M^{-1} s^{-1} for all hairpins. In contrast, the dissociation rate constants

[‡] These experiments were performed in collaboration with Dr. Rajan Lamichhane and Dr. Rui Zhao of the Rueda lab.

Table 4.5. Kinetic parameters measured using surface plasmon resonance (SPR) for kissing complex formation in 10 mM $MgCl_2$. From kinetic measurements obtained using SPR, ΔG was calculated. Activation energies were computed using the temperature dependence of the kinetics according to the Eyring's equation (Materials and Methods).

					Association			Dissociation		
System	**T (°C)**	**k_a x 10^4 ($M^{-1}s^{-1}$)**	**k_d x 10^{-4} (s^{-1})**	**ΔG (kcal/mol)**	**$\Delta G^{\ddagger}_{298}$ (kcal/mol)**	**$\Delta H^{\ddagger}$ (kcal/mol)**	**$\Delta S^{\ddagger}$ (cal/mol K)**	**$\Delta G^{\ddagger}_{298}$ (kcal/mol)**	**$\Delta H^{\ddagger}$ (kcal/mol)**	**$\Delta S^{\ddagger}$ (cal/mol K)**
HP2::HP3	10	1.3 ± 0.1	2.3 ± 0.2	-10.0 ± 0.1						
	15	1.5 ± 0.2	3.0 ± 0.2	-10.2 ± 0.1						
	25	1.9 ± 0.1	15 ± 1	-9.7 ± 0.1	11 ± 1	5 ± 1	-24 ± 1	21 ± 1	26 ± 1	16 ± 1
	30	3.8 ± 0.2	38 ± 2	-9.7 ± 0.1						
	40	3.1 ± 0.5	180 ± 20	-9.0 ± 0.1						
HP2-A4C::HP3-U4G	10	2.2 ± 0.2	2.30 ± 0.08	-10.3 ± 0.1						
	15	3.0 ± 0.3	1.2 ± 0.4	-11.0 ± 0.2						
	25	3.6 ± 0.8	0.9 ± 0.2	-11.7 ± 0.2	11 ± 1	4 ± 1	-24 ± 1	23 ± 2	32 ± 4	33 ± 5
	30	4.7 ± 0.2	3.0 ± 0.3	-11.3 ± 0.1						
	35	3.4 ± 0.1	8.2 ± 0.4	-10.7 ± 0.1						

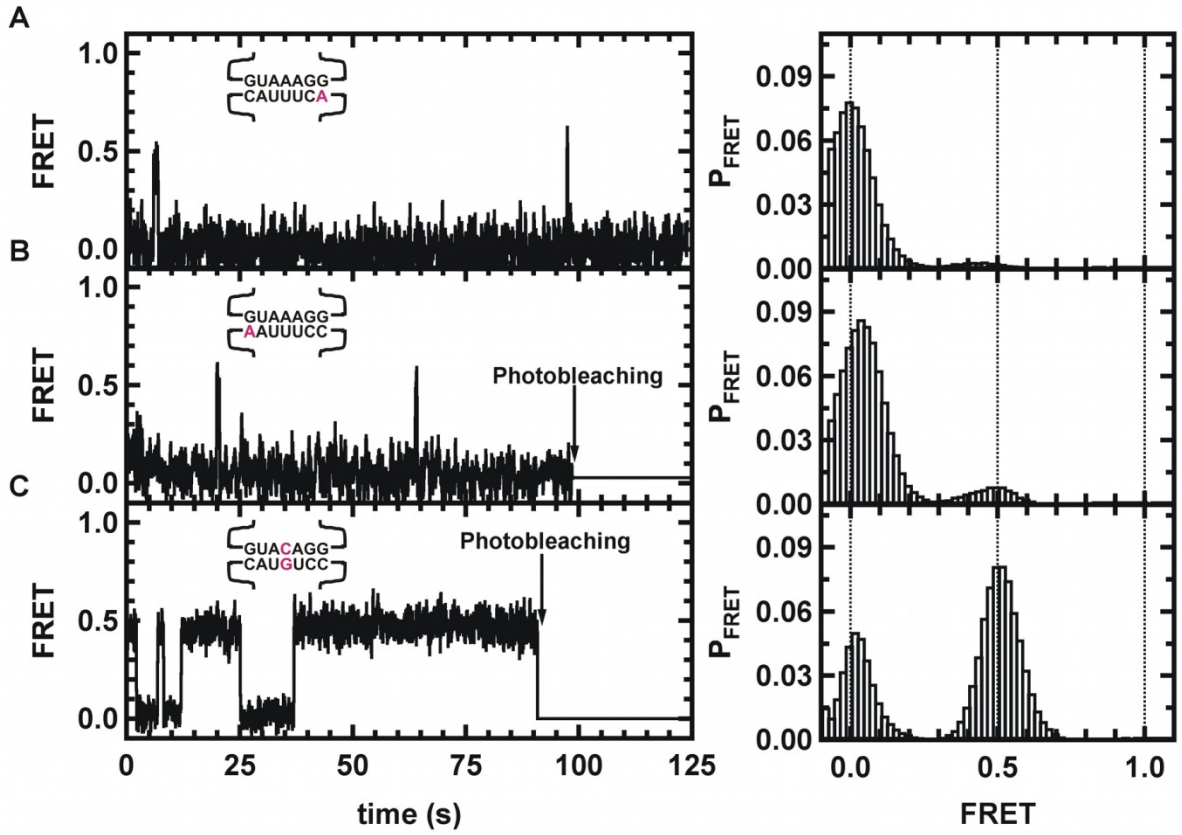


Figure 4.6. Representative single-molecule FRET trajectories (left) and histograms (right) from selected mutants. (A) HP1-C7A::HP2, and (B) HP1-C1A::HP2, and (C) HP2-A4C::HP1-U4G, which allowed the formation of GC base pair by replacing AU base pair. Arrow lines indicating the photobleaching point for each trace.

differ dramatically depending on the loop sequence. For HP1::HP2 and HP1-U4G::HP2:A4C, stable kissing complexes were observed using ITC, SPR and smFRET. Compared to the parent complex, the G•C mutant has one net extra hydrogen bond and slightly altered stacking potential, which decreased k_{off} by ~3 fold and resulted in a 0.6 kcal/mol stabilization. Attesting to the power of smFRET, transient associations of HP1-C7A::HP2 (k_{off} = 10 ± 1 s^{-1}) and HP1-C1A::HP2 (k_{off} = 6 ± 1 s^{-1}) was detected, neither of which was observed by ITC or SPR. Overall, kinetic data on mutated hairpin systems obtained by smFRET indicated that the stability of kissing complexes was primarily dictated by the dissociation rates.

4.3.9 Energy Landscapes for kissing complexes and their rearrangement to extended duplex structures

Using the thermodynamics (ΔG) and kinetic ($\Delta G^{\ddagger}$) data derived from the study of HP1::HP2 and HP2::HP3, a potential energy surface for the rearrangement from hairpins to extended duplexes can be proposed (**Figure 4.7**). This surface combines the insights derived from each of the biophysical approaches described above, including UV melting, ITC, SPR and smFRET. Random coiled strands (A) and free hairpins (B) could be considered reactant states, while the kissing complex (C) and extended duplex (D) forms represented the low energy products. These have been placed in coordinates of increasing intra-molecular and inter-molecular base pair interactions. Thermodynamics of transitions AD and AB were determined using UV melting analysis while BC and BD were measured using ITC and SPR. Transition state energetics for BC and CB was computed from SPR data. Energy wells have been drawn as symmetric Gaussian functions. That is a significant simplification as the actual energy landscape is expected

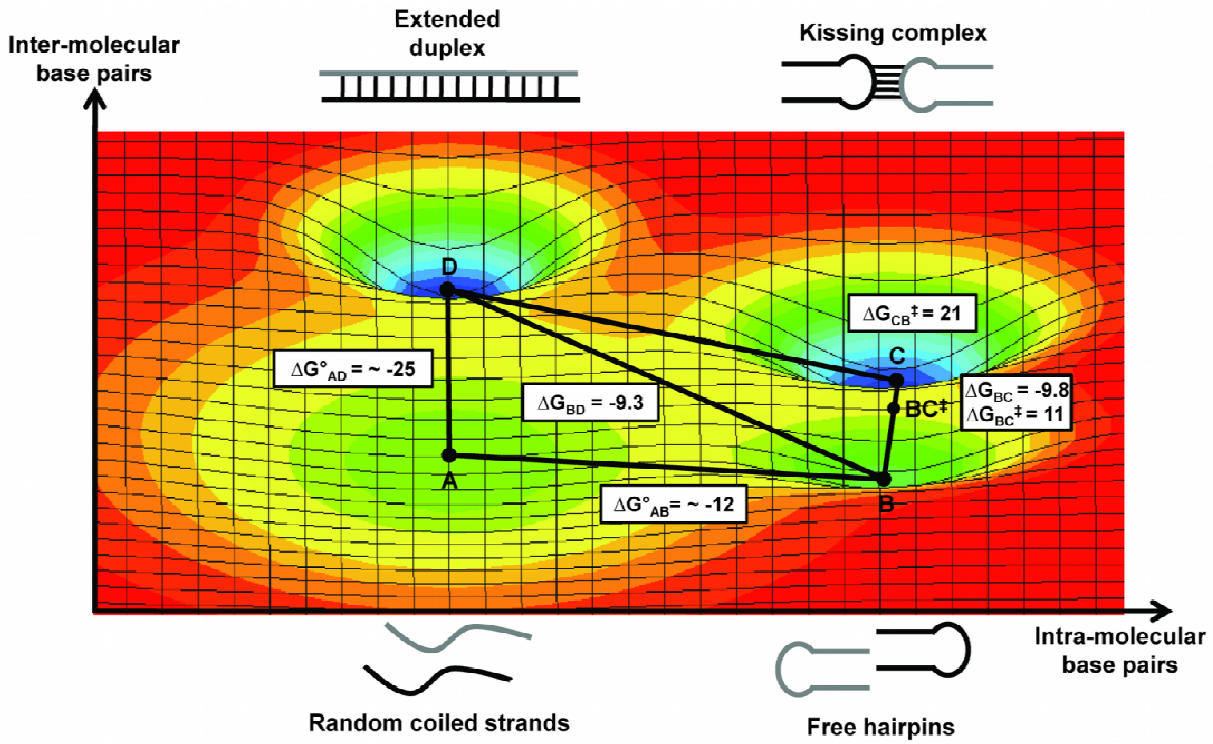


Figure 4.7. Putative potential energy surface for kissing and extended duplex formation. Thermodynamic (ΔG_{IJ}) and activation energies ($\Delta G^{\ddagger}_{IJ}$) for pathways associated with kissing complex formation and their conversion to extended duplexes. Thermodynamic data were obtained using ITC, UV meting and smFRET, and activation parameters were derived from kinetic data obtained by SPR and smFRET. Reaction coordinates are defined by the number of intra-molecular (X) and inter-molecular (Y) base pairs present in the molecular ensemble of a particular state. Stabilization energies are plotted along the Z-axis. Four different state shown: unfolded strands (A), free hairpins (B), kissing complex (C) and extended duplex (D). The energy of (A) has been used as the reference state ($\Delta G_A = 0$).

to be much more rugged than depicted here. Nonetheless, significant insight can derive from viewing the data in this way.

Kissing complexes and extended duplex states lie in the lowest energy wells on this surface while random coiled strands and free hairpins occupy higher energy regions. Data obtained from independent measurements when viewed as an ensemble correlate nicely within error in a thermodynamic cycle representing all possible transitions. Kissing complexes and extended duplexes lie in comparable energy wells (**Table 4.2**) despite the differences in the two structural contexts from which they were measured. Free hairpins associate to form a KC that crosses a ~11 kcal/mol energy barrier, whereas ~21 kcal/mol energy is required for dissociation. These energetics yield a net stabilization of ~10 kcal/mol, identical to ITC and smFRET observations. The smFRET data reveals that spontaneous rearrangement from kissing complexes to extended duplex occurs with a probability of only ~0.005 (out of 600 trajectories) within 2.5 min, suggesting a large activation barrier for this transition (196). Thus, one of the roles proteins should play in these rearrangements is to destabilize kissing complexes. By using RNA-protein binding energy to break the co-axial stacking of the kissing complex, it will facilitate the rearrangement by providing a driving force conversion to the kinetically inert ED state.

4.4 Discussion

While kissing complexes are well-known in RNA biochemistry (177,182,197-200), relatively few have been studied in great detail, leaving a significant gap in the understanding of these interactions. Work on HIV-DIS showed that its stability depended significantly on the sequences adjacent to the Watson-Crick interaction

including any flanking unpaired bases and the closing base pair of the stem (201,202). These results attest to the fact that the intricacies of a loop's structure contribute to the overall free energy of the subsequent kissing complex due to relative stabilization or destabilization of the ground state. Most of these studies have used thermal scanning techniques, however, that only allow measurement of the energetics near the transition temperature. In this study, ITC, SPR and single-molecule techniques were used to observe the dynamic behavior of a pair of designed hairpins that can be constrained to form only the kissing complex or allowed to proceed through the strand displacement reaction to form an extended duplex.

Energetically speaking, how similar are these kissing interactions to simple duplexes of the same sequence? Can one simply use a nearest neighbor (NN) formalism (derived from simple duplexes) to predict the energies of the kissing complexes? To test this possibility, NN parameters (203-205) were used to calculate free energies for RNA duplexes mentioned above and then compared to the measured values for the relevant simple duplexes and kissing interactions (**Table 4.6**). As expected, NN rules predicted simple duplex stability quite well for these short RNAs with $\Delta\Delta G_{298}$ less than 1 kcal/mol. NN predictions fared less well with the kissing hairpins, where the kissing complexes were >2 kcal/mol more stable than predicted, likely due to additional stacking interactions from co-axial alignment of the helices, electrostatic interactions or possible metal binding events not accounted for in the NN model. Thermodynamic analysis on a more extensive collection of sequences will be required to establish a suitable correction value to account for the offset.

Table 4.6. Comparison between stabilities predicted by nearest neighbor parameters (NN) and experimentally measured values for RNA duplex and kissing complex formation (203-205).

System	Context	ΔG_{298}^{NN} kcal/mol	ΔG_{298}^{Exp} Kcal/mol	$\Delta\Delta G_{298}^{Exp-NN}$ kcal/mol	$\Delta\Delta G_{298}^{Exp(kissing)-NN (duplex)}$ Kcal/mol
RNA2:RNA3	Duplex	-6.7	-6.8	0.1	
HP2:HP3	Kissing	N/A	-8.3	N/A	-1.6
RNA2-A4C:RNA3-U4G	Duplex	-9.2	-8.0	1.2	
HP2-A4C:HP3-U4G	Kissing	N/A	-11.2	N/A	-2

Calculated ΔG_{298} values for RNA2::RNA3 and RNA2-A4C::RNA3-U4G from NN were -6.7 kcal/mol and -9.2 kcal/mol respectively. The stability (ΔG_{298}) measured from experimental data for the RNA2::RNA3 duplex was -6.8 kcal/mol, giving a $\Delta\Delta G$ of 0.1 kcal/mol, well within the error for such a prediction. For RNA2-A4C::RNA3-U4G, the calculated ΔG_{298} was ~ -8.0 kcal/mol which was approximately 1.2 kcal/mol less stable than the NN predicted stability. However, the comparison was less effective in the context of the kissing complexes. Calculated stabilities from experimental data for HP2::HP3 predict ΔG_{298} ~ -8.3 kcal/mol and -11.2 kcal/mol for HP2-A4C::HP3-U4G. These two experimental stabilities for KC formation were ~2 kcal/mol more stable than what was predicted by the NN method for the identical RNA duplex. Furthermore, for duplexes RNA2::RNA3-C1Δ,C7Δ and HP2::HP3-C1A,C7A free energies of ~-2.6 and ~-4 kcal/mol were predicted by NN methods. This was in good agreement with what was observed in duplex and kissing contexts where stable complexes were not detectable by ITC. Above calculations predict that kissing complexes are slightly more stable than its RNA duplex. More interestingly these data puts forth the notion that at least for these example systems, kissing complex stabilities can be approximated using a slight offset predicted by the NN method for the identical RNA duplex. Further investigations are required using a larger sequence space for kissing interactions to fully characterize the factors that govern the offset between NN parameters for RNA duplexes and kissing complexes.

A significant difference between duplexes and kissing complexes was internal energetics (the separation of the free energy into enthalpic and entropic components) that was readily measured by ITC (**Table 4.2** and **Table 4.3**). For duplex formation a

higher enthalpic contribution was observed than for KCs. This was in part due to the expenditure of energy in reorganizing the internal structure of the loop that is required to interact with its complimentary hairpin. In duplex formation, the reorganization energy is much smaller relative to KCs (206). On the other hand the entropic penalty was greater in duplexes due to the higher loss in degrees of freedom from the random coil state of single stranded RNAs relative to the structured hairpins. Once again this exemplifies how ITC was able to unveil significant differences in the energetic profiles between duplexes and kissing complexes. Similar results were also obtained for the tetraloop-tetraloop receptor interactions (207). Similar data will be eventually required to accurately model thermodynamics kissing complexes.

When thermodynamics for kissing complexes were measured using ITC (**Table 4.2** and **Figure 4.4**), one interesting finding was the salt sensitivity of the enthalpic and entropic parameters. For both HP2::HP3 and HP2-A4C::HP3-U4G, enthalpy-entropy compensation was observed that led to little net $\Delta\Delta G$ within the temperature range studied, despite large changes for ΔH and ΔS terms. As expected, the energetics were more stabilized by the addition of $MgCl_2$ as compared to NaCl.

A significant ΔC_P was observed for kissing complex formation. Work from a number of laboratories has shown that changes in the internal structure of RNA and DNA in the single-stranded state contribute to the observed ΔC_P for duplex formation (208-210). This occurs because the extent of these partially-folded intermediate states is highly dependent on the solution conditions and temperature. A stable hairpin can be viewed as the extreme case of a structured single-stranded state and thus one would expect large ΔC_Ps to be associated with formation of kissing complexes, a prediction

that was borne out in these studies. While modest differences as a function of ionic conditions were expected based on prior studies (211), the large magnitude of these changes were somewhat surprising. More negative ΔC_P values were consistently observed in 10 mM $MgCl_2$ solutions relative to 1 M NaCl implying differences in salt stabilizations of reactant and product species. ΔC_P contributions towards the overall energies of these motifs are significant and understanding the origin of these energies in nucleic acid folding is important if one wishes to accurately predict the stability of RNA structures. In fact a recent study showed that incorporating experimentally determined ΔC_Ps values in conjunction with nearest neighbor models helped overcome systematic deviations between predicted and experimentally determined thermodynamics in DNA duplexes (212).

Thermodynamic parameters were also monitored for the refolding of the kissing hairpins into extended duplexes. Under calorimetric conditions, this reaction is highly favorable and occurs quantitatively within the timeframe of each ITC injection as represented by the difference in thermodynamic parameters relative to what was observed for kissing complexes. Thus, if the kissing complex does form under these conditions, its existence is transient and the reaction proceeds forward to the extended duplex state. The molecules become kinetically trapped in the ED state due to their slow dissociation rate. This result was in contrast to smFRET studies in which transitions from KC to ED states were extremely rare. The kissing interaction was made and broken many times at room temperature before committing to a slow, forward transition to the strand-displaced form that under these conditions was essentially irreversible. Thus, at room temperature, extended duplex formation may or may not involve a kissing

complex as an intermediate structure. The big difference in experimental parameters between these two experiments is the concentration of the individual RNAs. ITC is performed at high strand concentration so that the reaction generates sufficient heat to accurately measure the reaction whereas the smFRET experiments are performed at very low dilutions to ensure special localization of individual molecules. Thus the effect of local concentration can be seen as paramount in the expected outcome. This may represent one of the primary roles of accessory proteins in strand displacement reactions in vivo, improving the likelihood of a forward progression which then can be kinetically trapped by the slow off rate for the ED state.

The heat capacity changes ED formation show similar trends to that of the kissing complexes although the magnitude is more negative. From a mechanistic perspective, the experiments with HP1 and HP3 binding to HP2-A(3-5)C suggests an important fact with respect to the mechanism of strand displacement. In this case, the three central AU pairs have been mutated to A•C mismatches and as predicted, the kissing complex is unstable. Strand displacement on the other hand proceeds effectively, albeit with less favorable free energy due to the central mismatches in the product duplex. Thus, the forward reaction to the extended duplex does not need to proceed through intermediate kissing complex in this case. In fact, these results, together with the single-molecule data discussed above, may go so far as to imply that the kissing complex is a low energy state that is off pathway with respect to strand displacement and that the fraying pathway depicted as route B in **Figure 4.1** dominates by providing a mechanism to stabilize the off pathway intermediates. Furthermore, there may be kinetic benefits through increased local concentration of the RNAs as well.

This conclusion is similar to recent discussions on the mechanism of strand displacement for HIV's DIS. There has been much discussion regarding the proposal that the kissing complex proceeds through a two-step rearrangement where the first step involves pairing across the loop and the second step results in co-axial alignment of the three helical elements (186,213). While this latter step may be critical to the stability of a kissing complex, it likely inhibits the rapid expansion of the intermolecular base pairing at the expense of intramolecular interactions, as that pathway requires destacking of the helical elements at each step of the helix expansion.

Finally, the potential energy surface constructed using data obtained by ITC, SPR, UV melt and smFTRT helps to visualize the pathways associated with kissing and extended duplex formation in a quantitative perspective. A large energy barrier is assumed for the extended duplex formation via the kissing complex due to the significantly low number of events observed in smFRET assays although the transition was readily observed in ITC. This difference may have originated through the lesser number of molecular collisions that are present in smFRET than in bulk experimental approaches. In nature, proteins often participate in the conversion between kissing complexes and extended duplexes. In the case of DIS, NCp7 is the chaperone that facilitates the reaction (179,186,214,215) whereas for sRNAs, binding to their mRNA targets typically involves the protein Hfq (216-218). The fact that the initial contacts through apical loop to apical loop or hairpin to internal bulge have trouble progressing forward to an extended duplex may help to explain how these proteins function. By stabilizing a ternary complex that lacks the co-axially stacking of the kissing complex, one would naturally favor the forward progression to the extended duplex. Furthermore,

the co-localization that such a complex would provide could dramatically reduce the entropic cost during strand exchange. On-going experiments are looking at further defining the energetic landscape of these species and the role of protein cofactors in facilitating the refolding of RNAs such as these model hairpins to their extended duplex conformation.

4.5 Materials and methods

4.5.1 RNA constructs used

Figure 4.2 summarizes the RNA hairpins used in this study. Nucleotide sequences (loop nucleotides are underlined) are: HP1 has the sequence 5′-GGACGAGGCAUUUCCCCUUGU-3′. HP3 has a RNA sequence of 5′-GGACGAUCAGCAUUUCCCUGAUGU-3′, HP3-C1A,C7A 5′-GGACGAUCAGAAUUUCACUGAUGU-3′, HP3-C1A 5’- GGACGAUCAGAAUUUCCCUGA UGU-3′, HP3-C7A 5’-GGACGAUCAGCAUUUCACUGAUGU-3′, HP3-C1Δ,C7Δ 5′-GGACGAUCAGAUUUCCUGAUGU-3′, HP3-U4G 5’- GGACGAUCAGCAUGUCCCUGAUGU-3′. HP2 5′-GGACAAGGGGAAAUGCCUUGU-3′ and the HP2-A(3-5)C is 5′-GGACAAGGGGCCCGGCCUUGU-3′, HP2-A3C 5’- GGACAAGGGGCAAUGCCUUGU-3’, HP2-A4C 5’- GGAC AAG GGGACAUGCCUUGU-3’. RNA oligomers were synthesized by Dharmacon/Thermo Scientific (Lafayette, CO). RNAs were deprotected using the procedures specified by the manufacturer and were purified by denaturing PAGE, eluted into 0.5 M ammonium acetate, ethanol precipitated, and resuspended in water. Concentrations of RNA stock solutions were determined by absorbance at 260 nm (219).

4.5.2 UV melting studies

Absorbance versus temperature profiles (melting curves) were measured with an Aviv 14DS UV-VIS spectrophotometer with a five-cuvette thermoelectric controller as previously described (220-222). Custom-manufactured microcuvettes (Hellma Cells) with 0.1 and 0.2 path lengths (60 and 120*μL* volumes, respectively) were used. Oligonucleotides were dissolved in 10 mM NaHEPES, pH 7.5 containing either 1.0 M NaCl or 10 mM $MgCl_2$. Samples were annealed and degassed by raising the temperature to 85 °C for 5 min and then cooling to -1.6 °C over a period of 25 min just prior to a melting experiment. Care was taken not to allow the total absorbance to rise above 2.0. While at 85 °C, the absorbance was measured at 260 nm for later calculation of oligonucleotide concentration using the extinction coefficients. Thermodynamic parameters for hairpin and duplex formation were derived from the thermal melting data using the program MELTWIN v.3.0 (221). This program uses the Marquardt-Levenberg non-linear least squares method to solve for ΔH and ΔS and upper and lower baselines from absorbance versus temperature profiles. The above method assumes a two-state model and a $\Delta C_P = 0$ for the transition equilibrium. Thus, these parameters are most accurate at the melting temperature of the system.

4.5.3 Isothermal Titration Calorimetry (ITC)

A VP-ITC titration calorimeter (MicroCal, Inc.) was used for all measurements.

Samples were prepared by diluting a small volume of stock into 10 mM NaHEPES, pH 7.5 and 1.0 M added NaCl. The syringe and sample cell RNAs were prepared in matched buffers to minimize background heats of dilution. All buffers were prepared from stock solutions on the day of use and extensively degassed under vacuum.

Samples were heated at 95 °C for 3 min and cooled at room temperature for 30 min prior to data collection. After an initial 2 *μL* injection to counteract backlash in the auto-titrator (223), ITC experiment consisted of 40 injections (at 7 *μL* per injection) of a 60 *μ*M oligonucleotide into 1.4 mL of the complementary strand at 3.5 *μ*M. Sample stirring was set at 310 rpm for all measurements. Titration data were collected at several different temperatures. ITC data were analyzed with ORIGIN software (MicroCal Inc., ver. 7.0). Raw thermograms were integrated and normalized, resulting in a plot of Δ*H* (mol of injectant)$^{-1}$ versus molar ratio. In each experiment, a long upper baseline was collected after the binding transition was saturated. The terminal 10 points (i.e., the clearly linear portion) of the upper baseline in each experiment were fit to a straight line, which was subsequently subtracted from the entire data set to remove contributions from background heats of mixing and dilution. A total Δ*H* for each reaction was obtained by nonlinear least-squares fitting of the plot of Δ*H* (mol of injectant)$^{-1}$ versus molar ratio to a single site binding model (224) to obtain Δ*H*, K_A, and the reaction stoichiometry, *n* under each condition. ΔG was obtained from K_A and ΔS was subsequently derived based on the Gibbs equation at each temperature and errors were estimated based on standard propagation (225).

Temperature dependence of ΔH was quite significant for both kissing complex formation as well as conversion of the hairpins to extended duplexes, indicating a non-zero heat capacity change (ΔC_P) for these transitions (206,209). The reaction enthalpy was plotted as a function of temperature to obtain ΔC_P based on **Eq 4.1** using the approximation that ΔC_P is temperature independent (206). Deviations from linearity begin to occur about 15-20 °C below the melting temperature of the product (kissing

hairpin or extended duplex depending on the experiment) due to incomplete formation of reaction products at equilibrium. Only the linear portion of the data was used to compute ΔC_P.

$$\Delta C_P = d(\Delta H) / dT \qquad (4.1)$$

4.5.4 Native polyacrylamide gel electrophoresis

Complex formation between a pairs of hairpins was verified using native gel electrophoresis on 20% polyacrylamide gels. Samples were prepared in 50 mM Tris-acetate buffer pH 7.5 containing 10 mM $MgCl_2$. After allowing them to equilibrate, 2 µL of 40% (wt/vol) sucrose was added and the RNA was loaded immediately onto the native gels while applying a potential of 300 V. Species were resolved under constant power (5 W for 6 – 12 hrs) at 4 °C. The running buffer for these gels was 50 mM Tris-acetate at pH 7.5 containing 10 mM $MgCl_2$. Bands were visualized by staining with StainsAll.

4.5.5 Surface Plasmon resonance

Kinetic experiments were performed on a Biacore 2000 instrument. Experiments were done on a streptavidin-coated chip (SA chip, Biacore). All experiments were performed in the same reaction buffer (50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$). During all experiments ~15 fmol of 5' biotin labeled RNAs were immobilized on the sensor chip. Immobilizations of RNAs were performed at a flow rate of 5 µL/min to make sure homogeneous surface coverage is attained. Experiments were carried out at temperatures from 10 – 40 °C and at flow rates of 30 - 50 µL/min. Change in flow rates

did not affect the rate constants observed in the experiments implying the absence of mass transfer effects in the conditions used.

To measure kinetics of kissing complex formation, varying concentrations of hairpins (400 – 800) nM were titrated into the surface immobilized complimentary hairpin. At the end of each injection (1 min injection) the dissociation was observed in the presence of binding buffer. Surface regeneration was performed by injecting 300 µL of regeneration buffer (50 mM Tris-HCl pH 7.5, 20 mM EDTA) at a flow rate of 100 µL/min. The data were analyzed globally by fitting both the dissociation and association (where applicable) phases simultaneously (BIA evaluation software version 4.1) using a 1:1 (Langmuir) model (two fitting parameters). BIA evaluation uses Marquardt-Lavenberg algorithm to optimize parameters in fits and assigns kinetic constants to the binding models. The goodness of the fit was judged by the reduced chi-square (χ^2) values.

4.5.6 RNA biotinylation

HP3 and HP3-U4G was 5'-labled with biotin to be used in surface plasmon resonance experiments. RNAs were first treated with Calf Intestinal Phosphatase (CIP) and phosphorylated using ATP-γ-S using the Ambion Kinase Max kit (Ambion, Inc). In brief, 1 nmol of RNA was treated with CIP (in 10X dephosphorylation buffer, 0.5 units of CIP at 37 °C for 2 h). The reaction mixture was purified using the Phosphatase Removal Reagent as described by the product manual. Purified RNAs were phosphorylated with ATP-γ-S using T4 Polynucleotide kinase. Phosphorylated RNAs were purified using a G-25 spin column (GE Healthcare) and speed vacuumed to dryness. RNAs were then dissolved in 45 µL of 100 mM $KHPO_4$, pH 8.0, 5 µL of 20 mM N-iodoacetyl-N-

biotinylhexylenediamine dissolved in DMF (EZ-Link Iodoacetyl-LC-Biotin, Thermo Scientific). The reaction was incubated at 45 °C for 1 h while shaking under dark conditions. The reaction was then ethanol precipitated and analyzed using PAGE.

4.5.7 Calculation of activation energetics

The temperature dependence of k_a and k_d was analyzed according to the Eyring's equation shown below.

$$ln\frac{k}{T} = -\frac{\Delta H^{\ddagger}}{RT} + ln\frac{k_B}{h} + \frac{\Delta S^{\ddagger}}{R} \quad (4.2)$$

In a plot of ln (k/T) vs.1/T (**Figure 4.5B**), activation enthalpy ($\Delta H^{\ddagger}$) derives from the slope and activation entropy ($\Delta S^{\ddagger}$) is obtained from the intercept. The activation energy ($\Delta G^{\ddagger}_{298}$) was calculated by using the Van't Hoff relationship.

4.6 Acknowledgments

This work was in collaboration with Prof. David Rueda's lab in the Chemistry department at Wayne State University. I would like to thank Prof. Rueda, Tuhina Banerjee, Rajan Lamichhane and Rui Zhao and Jane Philip for directions and help with experiments to complete this project successfully.

APPENDIX*

Isothermal Titration Calorimetry of RNA (192)

A1.1 Abstract

Isothermal titration calorimetry (ITC) is a fast and robust method to study the physical basis of molecular interactions. A single well-designed experiment can provide complete thermodynamic characterization of a binding reaction, including K_a, ΔG, ΔH, ΔS and reaction stoichiometry (n). Repeating the experiment at different temperatures allows determination of the heat capacity change (ΔC_P) of the interaction. Modern calorimeters are sensitive enough to probe even weak biological interactions making ITC a very popular method among biochemists. Although ITC has been applied to protein studies for many years, it is becoming widely applicable in RNA biochemistry as well, especially in studies which involve RNA folding and RNA-interactions with small molecules, proteins and with other RNAs. This review focuses on best practices for planning, designing, and executing effective ITC experiments when one or more of the reactants is an RNA.

A1.2 Introduction

Isothermal titration calorimetry (ITC) is a powerful and versatile method to study the physical basis of molecular interactions. In an ITC experiment, one physically measures the heat generated or absorbed during a binding reaction (ΔH_{obs}). The typical experiment involves addition of one binding partner (titrant) into the other binding partner (titrate) over time using one or more individual injections. Depending on the type

* This chapter was previously published in *Methods*.

of instrument being used, heat is measured either as a change in temperature or as the change in power required to maintain temperature between the sample and a reference cell. This energy is then converted into a binding enthalpy based on knowledge of the cell volume and the concentration of the reactants. The observed enthalpy measured in an experiment includes not only the heat of binding between the molecules, but also any additional heat sources associated with the reaction including solvent effects, molecular reorganization and conformational changes, heats of dilution and purely mechanical artifacts such as heat resulting from sample stirring. Thus, careful preparation of solutions and measurement of appropriate background heats are required to obtain thermodynamic parameters that accurately reflect the event(s) of interest.

The experiment can be performed as ether a continuous or sequential titration. The heat produced during each injection is proportional to the amount of complex formed. Thus, one expects maximal enthalpies at the earliest points in a titration with a decrease in intensity as free titrate is consumed. The change in heat over the course of the titration will give rise to the binding stoichiometry and affinity constant. Since K_a is related to the Gibbs free energy (ΔG) of binding and ΔH was measured directly, reaction entropies (ΔS) can be readily calculated from a single measurement without relying on van't Hoff approximations. By repeating the experiment over multiple temperatures, heat capacity changes (ΔC_P) for the reaction can also be easily measured.

The ability to obtain complete thermodynamic characterization of a reaction quickly and accurately from a single ITC measurement has made the technique very popular. Over the past two decades, calorimeters have been improved immensely in terms of

sensitivity and ease of use (226), which makes modern ITC versatile enough to explore interactions in biological systems. The sensitivity of the modern instruments has tremendously expanded the range of affinities over which measurements can be made, spanning from the low millimolar range for weak binders to sub-nanomolar interactions on the tight end. In certain cases one can extend the range of accessible affinities even lower if there are families of compounds that can be used when competition methods are employed (discussed in more detail below) (227).

In general, three types of ITC experiments have been performed on RNAs based on the types of binding interactions being analyzed. These include RNAs binding to: a) small molecules or ions; b) proteins; or c) other RNAs. For the most part we will treat these equivalently as the experimental design considerations are pretty similar. Several recent reviews provide broad description of modern ITC instrumentation and their application in biochemistry and biophysics (228-233). This review will focus on more practical aspects of performing effective ITC experiments on RNAs with only a limited discussion of the theory and instrumentation.

A1.3 Examples of RNA ITC experiments

Calorimetric analysis of RNAs is a growing field due to ready availability of the instrumentation in core facilities and reduced sample requirements of modern calorimeters. Of the three types of RNA interactions found in the ITC literature, small molecule interactions are among the most prevalent studies due to their relative simplicity from a conceptual perspective. Analysis of the specificity of riboswitches lends itself nicely to characterization by calorimetric methods, illustrated by the work from the

Batey Lab among others (234-236). Aminoglycoside binding to a variety of RNAs has also been prominent in the ITC literature (237-240) and unraveled thermodynamic relationships that could be important in characterizing and optimizing the drug-target interactions for this class of antibiotics.

A variety of RNA-protein interactions have been studied recently using ITC. The Puglisi's group used ITC to probe how structural features of the natural RNAs such as bulges and loops affect RNA binding to the double-stranded RNA binding domains (dsRBDs) of PKR, a double stranded RNA-dependent protein kinase (241). This particular study revealed using HIV TAR RNA that bulges and loops are essential for high affinity binding to the dsRBD domains. In another study, ITC was used to look at the interaction between the *E. coli* protein Hfq and one of its RNA binding partners, DsrA, a non coding RNA induced in response to cold shock (21,242) ITC was particularly important in this study for assaying the stoichiometry of the binding . ITC has also been used to study the thermodynamics of the assembly of individual subunits in the 30S ribosomal subunit in bacteria . In this work ITC was used to show that S8 and S11 are thermodynamically independent of the other proteins, while S15 had a cooperative effect on the S6/S8 heterodimer. ITC was also able to probe the sum of individual protein binding steps measured in isolation, proving the aggregate value was comparable to the binding that was performed using all components in a single injection, supporting the deconstructive approach in studying ribosomal assembly.

Understanding of fundamental RNA folding thermodynamics has also benefited from the application of ITC. Nearest-neighbor models which derived predominantly from UV-melting studies have been used extensively with great accuracy in RNA secondary

structure prediction (244,245). The main advantage that ITC provides is a direct measurement under a specific temperature condition that reveals non-two state behaviors lost into the baseline corrections of UV-melting traces. One such parameter is the heat capacity change (ΔC_P) that significantly affects enthalpic and entropic contributions of RNA and DNA folding but is often neglected due to its more modest effect on ΔG, a result of entropy-enthalpy compensation (206,246,247) Furthermore, RNA tertiary rearrangements that exhibit little UV-hyperchromicity like kissing hairpin interactions or junction folding can be observed by ITC. These studies are allowing the thermodynamic dissection of more complicated RNAs and beginning the process of parameterizing non-duplex contacts in RNA structures. The ability of ITC to effectively analyze non-two-state systems and to probe interactions that are unobservable by other methodologies makes it a powerful tool for the on-going characterization of RNA folding.

A1.4 Instrument description and methodology of ITCs

The most widely available ITCs used for microcalorimetry of biological samples are based on the power compensation methodology. Both the MicroCal VP-ITC and the Nano-ITC formerly produced by Calorimetry Sciences Corp. (purchased in 2007 by TA Instruments, a subsidiaryfrom the Waters Corporation) use this technology. The advantages of this style of instrument include its fast response time, high sensitivity and modest sample requirements. A schematic diagram of a typical ITC is shown in **Figure**

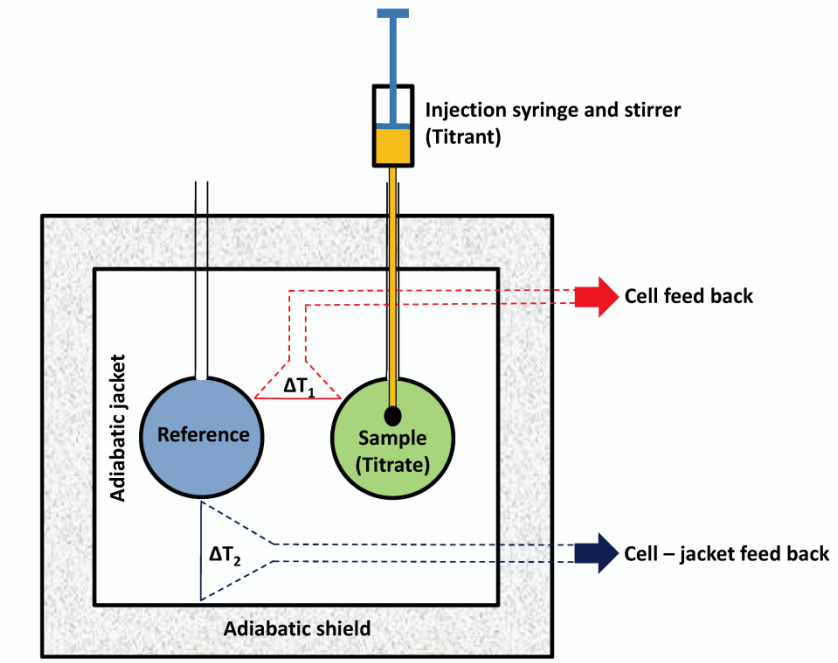


Figure A1.. A schematic diagram of a typical energy compensation isothermal titration calorimeter.

A1. Two equivalent lollipop shaped cells are suspended in an adiabatic jacket. The temperature of each cell is monitored and maintained at a constant temperature through an electronic feedback loop that controls thermoelectric heaters located adjacent to each cell. During the experiment the titrant, which is in the injection syringe, is added tothe sample cell. If the reaction is exothermic, heat evolves and ΔT_1 increases. The feedback loop responds by reducing the power to the resistive heater around the sample cell to restore ΔT_1 to zero. Thus, in a power versus time curve, the injection results in a negative deflection from the baseline, the integrated area of which corresponds to the total enthalpy released as a function of the injection (**Figure A2**). An endothermic reaction would result in a corresponding positive peak. At saturation when no reaction is taking place, the baseline reflects a constant power consumption proportional to the power needed to maintain the temperature differential between the cell and the adiabatic jacket (ΔT_2). Small peaks are often still observed due to effects such as buffer mismatch between the titrant and titrate or non-specific binding. After integration and background correction, one obtains a thermogram (bottom panel in Figure 2) that can be fit to a specified binding model.

The example ITC data shown in **Figure A2** is from an experiment in which the energetics of RNA duplex formation was measured. The titrate was a 7-nt RNA (5'-CGUUCCC-3') and the titrant was its complement. Both samples were prepared by dilution of a concentrated RNA stock solution in water into buffer (50 mM HEPES, pH 7.5, 1 M NaCl) and degassed prior to use. In this case the dilution method was required rather than dialysis due to the small size of these RNAs. Small amounts of residual salt in the RNA stock solutions from ethanol precipitation would be only a minor issue here

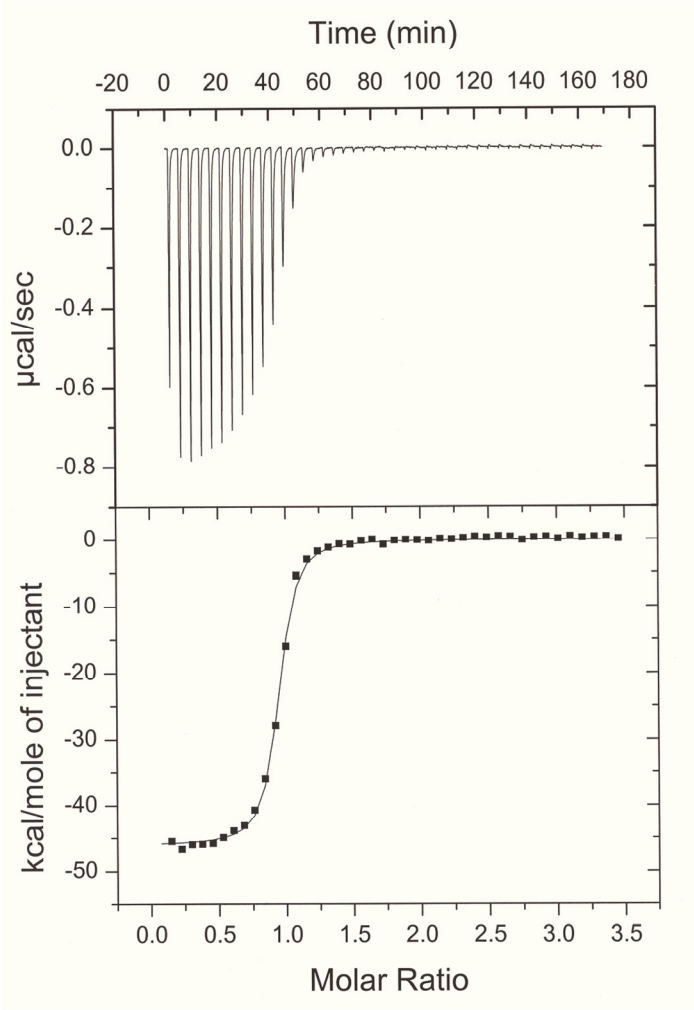


Figure A2.. Example ITC data from an experiment measuring the energetics of RNA duplex formation. A 7-nucleotide single-stranded RNA was titrated into its complement in 50 mM HEPES pH 7.5, 1 M NaCl at 25 ˚C. Top Panel: Power versus time curve. Bottom Panel: Thermogram of the integrated peak intensities plotted against the molar ratio of the two strands. Fitting of this titration yielded a K_a = 4.3 x 10^7 M^{-1}, ΔH = -46.1 ± 0.2 kcal/mol and n = 0.91 ± 0.01. Data reproduced with permission from ref.(247) (Copyright, American Chemical Society).

due to the high NaCl concentration of the binding buffer. Data were collected during forty 7 µL injections of 75 µM titrant into 5 µM titrate at 25 °C with stirring at 310 rpm and 240 s injection spacing. Equivalent data were collected when the titrant RNA was used as the titrate. The raw data were integrated and normalized using ORIGIN (MicroCal Inc., ver. 7.0) resulting in a plot of ΔH per mole of injectant versus molar ratio. This plot was fit to a simple binding model.

Measurement of background heats is a critical aspect of the ITC experiment. There are two standard ways in which to perform this correction and one must decide in advance which method is to be used as it impacts the experimental design. The most common method is to use a series of blank injections. This involves repeating the exact same injection protocol but with buffer in the cell rather than titrate. Thus, one sees the effects of diluting the titrant and any heats involved in that process. The disadvantage of this method is that it is time consuming since the entire titration protocol is typically repeated. The second method is to extend the titration to collect extra data at the end of the experiment under saturation conditions. This involves having additional titrant available to collect 5-10 extra injections after all of the titrate has been bound. One then prepares a linear extrapolation of the heats from these final injections to apply as a correction over the course of the entire experiment. This latter method of background correction was used in the ITC experiment shown in **Figure A2** which is why so many additional data points were collected after saturation was achieved. Both methods provide equivalent results in our hands so long as there is little non-specific binding of the titrant to the titrate.

A1.5 Planning an ITC experiment

The basic protocol for ITC experiments is similar for the three classes of ligands described above. Thus, unless explicitly stated, the protocols should be regarded as common to all three categories of interactions. A flow chart illustrating the basic design of an ITC experiment is depicted in **Figure A3** and will serve as the focal point of this discussion. Where there are differences between the classes of interactions, they typically arise due to issues such as differential solubility of proteins or small molecules and the solution conditions under which such species are stable. Also, issues relating to non-specific binding and stoichiometry must also be recognized and addressed during the experimental design. This section will describe the planning that needs to be done to maximize the likelihood of performing the experiment successfully.

A1.5.1 Preliminary binding studies

Even with the development of microcalorimeters, the samples required to perform ITC analysis are large relative to many other methodologies (i.e. UV melting, gel shift assays, surface plasmon resonance, etc). With that in mind, it is prudent to perform preliminary binding studies on a new system before filling the ITC cell with an expensive RNA.

The first consideration has to do with whether or not the RNA is well behaved. Does the material exhibit alternative conformations on native gels or are there problems with monomer-dimer equilibria for instance? These types of equilibria will make it next to impossible to analyze and interpret the ITC data and should be resolved prior to initiating ITC studies. While methods of resolving these issues are relatively standard,

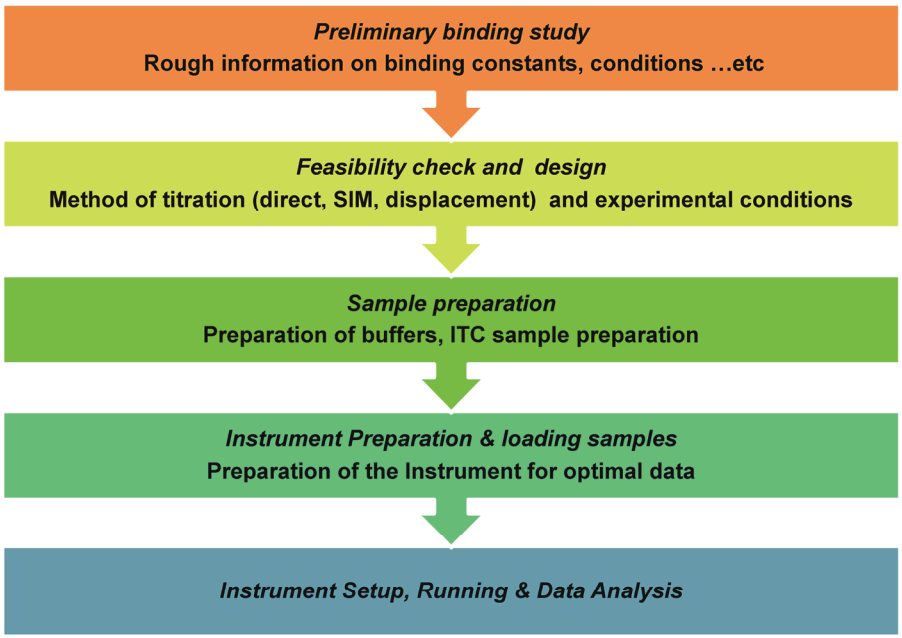


Figure A3. Overview of an ITC experiment.

one must remember that the concentrations used in the calorimeter are much higher than for many other types of experiments. Monomer-dimer equilibria are dependent on the concentration and one should perform small scale refolding studies at the target concentration you will use in the ITC. Hairpins that are well behaved at low concentrations used with ^{32}P-labeled material may not be equally well behaved when annealed at micromolar concentrations. Furthermore, one should evaluate sample homogeneity under the highest and lowest salt or $MgCl_2$ conditions one plans on using in the calorimeter to ensure that the calorimetry studies do not simply reflect changes in alternative folding of the reactants.

For RNA-RNA and RNA-protein interactions, gel shift assays will often be the best choice for preliminary studies. They require minute samples and allow evaluation of binding affinities. Furthermore, because gel shift assays provide spatial separation of the products, this preliminary experiment will also alert the user to potential problems associated with product heterogeneity. Is there a single product forming under a given condition or are there multiple bands present representing the formation of higher order complexes that might complicate ITC data analysis?

For small molecule interactions, preliminary binding studies are often not as simple as they are for macromolecular interactions. Rarely do the small molecule interactions provide large mobility changes upon binding making gel shift assays less useful. Several biophysical approaches can still be used however, depending on the system. A generic methodology that is reasonably effective is to look at ΔT_M in UV melting analysis upon binding (250). Since the method requires no extrinsic labels or significant structural changes upon binding, it is quite versatile. Other approaches such as circular

dichroism (250) or fluorescence anisotropy (251) can be used, but their utility is commonly dependent on the exact nature of the RNA and the binding partner being studied.

Since the intent of these preliminary studies is preparation for ITC analysis, we need to make a quick statement regarding buffers at this point. When selecting experimental buffer conditions for the ITC, most common salts are acceptable. The one notable exception is DTT, however, which one periodically finds as a reductant in protein storage buffers. DTT concentrations as low as 1 mM can cause severe baseline artifacts due to background oxidation during the titration. If possible, we recommend removing all reducing agents by dialysis prior to performing ITC analysis. If they must be present for stabilization of a reactant, one should consider purging the system with argon or nitrogen prior to degassing to avoid problems with autoxidation.

A1.5.2 Considerations that effect sample concentration

Once preliminary binding studies are complete and the system shows good homogeneous behavior, the next step is to determine exact conditions that will be used in the ITC. For most people embarking on an ITC study for the first time, they are most interested in knowing, how much material they need to prepare and at what concentration? While ITC instruments are much more sensitive than they used to be, these studies still require large samples. The knee jerk answer to this question for our instrument (a VP-ITC from MicroCal) is that one typically needs about 1.8 – 2 mL of titrate solution in the low micromolar range (typically 1 – 2 μM) while the titrant should be ~400 μL of a solution 15-20 times the concentration of titrate. Other instruments

have different cell volumes, so inquire regarding the specific ITC you plan on using prior to preparing your samples. There are more precise ways to determine sample concentration. To make that determination, however, one needs to know a little more about how a thermogram is analyzed and a parameter called the c-value (226).

A typical ITC thermogram was shown in the top panel of **Figure A2**. As explained above each deflection from the base line is a result of the change in the differential power that is required to maintain the sample and the reference cell at the same temperature. As the amount of heat absorbed and emitted are equal to the change in the power, the heats of each injection (q_i) can be obtained by integrating the respective peak and mapping it to the fractional saturation of the binding reaction (F) based on the stoichiometry (n), the total concentration of the titrate (M_T), the binding enthalpy (ΔH), and the cell volume (V), when the concentration of the titrant (B_T) is known (223,226,231,252) (**Eqs.A1-A3**).

$$q_i = nFM_T\Delta HV \qquad \textbf{(A1)}$$

$$F^2 - F\left\{1 + \frac{B_T}{nM_T} + \frac{1}{nK_BM_T}\right\} + \left(\frac{B_T}{nM_T}\right) = 0 \qquad \textbf{(A2)}$$

$$q_i = nM_T\Delta H\left(\frac{V}{2}\right)\left\{X - \sqrt{\left(X^2 - \frac{4B_T}{nM_T}\right)}\right\} \qquad \textbf{(A3)}$$

Where

$$X = 1 + \frac{B_T}{nM_T} + \frac{1}{nK_BM_T}$$

From **Eq A1** one can solve for *F* and substitute into **Eq A2** to obtain a quadratic relationship, which then can be solved for q_i given **Eq A3**. Since B_T is known, one can now fit the solution into a non linear least squares fitting to solve for thermodynamic parameters such as ΔH, K_B, *n* where K_B is the binding constant of the reaction. Once K_B is obtained ΔG can be calculated using **Eq A4** and ΔS from **Eq A5**.

$$\Delta G = -RT\ln(K_B) \qquad (\mathbf{A4})$$

$$\Delta G = \Delta H - T\Delta S \qquad (\mathbf{A5})$$

In the solution for X in **Eq A3**, one sees the product $n*K_B*M_T$ which was named the c-value. Wiseman *et al.* showed that to obtain useful information from the titration, 1< c < 1000 and ideally c would be between 10-100. Therefore, for weak interactions, while one might be tempted to increase the titrant concentration, this may lead to the loss of information as the whole reaction will occur within a single titration point and K_d would be poorly defined. On the other hand for stronger binding systems lower concentrations are preferred. Here the limitation will be the amount of heat given off by the reaction. When this signal becomes too small, the transition is hard to resolve above the background and the effective error in the resulting thermodynamic parameters increases (253). So, if you have a ballpark affinity and a notion regarding the stoichiometry, one can simply calculate M_T for c = 10 and 100. Tellinghuisen showed based on a theoretical analysis that one should optimally set the terminal ratio of B_T/M_T based on **Eq A6**.

$$R_m = \left(\frac{B_T}{M_T}\right) = \frac{6.4}{c^{0.2}} + \frac{13}{c} \qquad (\mathbf{A6})$$

In practice, for a typical titration involving addition of ~280 μL titrant into 1.4 mL titrate with a 1:1 stoichiometry, this yields a 30-fold concentration ratio to achieve c = 10 and 15-fold ratio for c = 100.

A1.5.3 Who should be titrated into whom?

When designing an experiment in ITC, one of the first questions is often who should be titrated into whom? This is one of the cases where one has to distinguish between the 3 common types of RNA experiments. For RNA-RNA interactions, solubility is rarely an issue and to a great extent it does not typically matter which species is in the cell and which is in the syringe. In fact, one commonly repeats the experiment by flipping the two to ensure that there are no differences which would reflect problems with alternative conformations during annealing at high concentrations. The situation is not so clear-cut for protein or small molecule interactions. Here, the primary concern is often one of solubility. The experiment is often easiest if the least soluble material is used as the titrate as this prevents precipitation or aggregation that might skew the data.

A1.5.4 Sample Preparation

As with all thermodynamic measurements on RNA, sample quality is of utmost importance. It makes little difference whether the material derives from *in vitro* transcription or chemical synthesis. Typically RNAs will be either gel purified or HPLC purified prior to use. The critical issue is that the material is as homogeneous as possible and folds into a single conformation. The preliminary studies described above

will have shown that the annealing conditions are appropriate for the desired experiment.

Buffer match between the titrate and titrant solutions is a critical aspect of sample preparation. How one achieves this will be dependent upon the size of the RNAs in question and the ligands with which they will be interacting. If both species are sufficiently large, the solution is simply to dialyze them in a common reservoir. This will ensure that the two samples are rigorously identical. Perform the dialysis at high RNA concentration so that some dilution during dialysis is not a problem and then adjust the final concentration of the RNAs by diluting them with dialysate. If one or both of the RNAs is too short to be retained during dialysis, one can also prepare them by precipitation, followed by resolubilization in the appropriate buffer. The potential problem here is that one must be careful to wash the RNA pellet very effectively as any residual salt from the precipitation will induce a buffer mismatch. While background correction will be used to account for small discrepancies, it is easy to incorporate significant errors as a result of this method. For small molecule ligands, it is common to dissolve them in the dialysate from the RNA buffer exchange. Once the samples are prepared they need to be degassed before being loaded into the ITC instruments. Typically we degas at a few torr partial pressure for 5 minutes in the thermovac instrument supplied by MicroCal. Inadequate degassing will produce large spikes in the ITC baseline due to the formation of air bubbles during the experiment, which adversely affects data quality. One should be consistent with this degassing step as it has the potential to alter sample concentrations. If this is a concern, one can re-measure the concentration after loading

the sample into the syringe/cell using the excess material leftover from filling the instrument.

It was mentioned above that DTT and other reductants should be avoided during ITC experiments. Another issue to consider carefully is sample pH. Remember that ITC measures total heat (ΔH_{obs}), which comprises the binding event, solvent dilutions, and as well as ionization/deionization enthalpies. If changes in protonation state accompany binding or folding, a buffer dependent contribution will be superimposed on the data the magnitude of which is dictated by the protonation enthalpy of the buffer and the solution pH. This was illustrated nicely in a study of aminoglycoside binding to 16S rRNA where pH and buffer sensitivity was observed (239,254,255). By repeating the ITC experiments under several buffer conditions the contribution from the buffer component can be factored out (231). One should routinely use control experiments to assess potential buffer linkage, by repeating the experiment in a different buffer and/or at a different pH.

For experiments involving nucleic acids, metal ions are needed for proper folding. The metal ion-RNA interactions and their rearrangement as a function of the binding reaction is an integral component of these systems. For this reason, one should expect a certain degree of linkage between the ionic conditions used in an experiment and the binding thermodynamics. Whether or not this contribution is significant will be highly dependent on the specific interaction being studied.

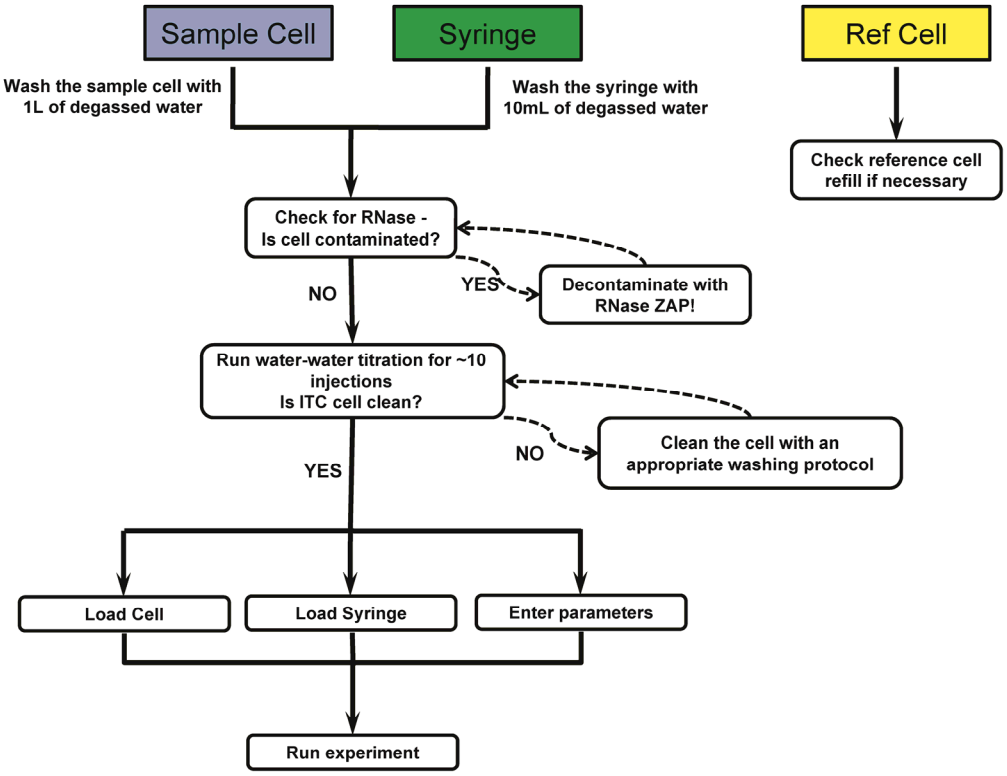


Figure A4.. Preparation of the ITC instrument for an experiment involving one or more RNA substrates.

A1.5.5 Instrument decontamination

Figure A4 provides a flow chart of the steps leading to ITC data collection. Cleaning the ITC is an important part of this process. This step is of particular importance to RNA users and is a step that should not be underestimated or taken lightly, especially if the instrument is part of a shared facility and used by non-RNA laboratories. The critical issue here is one of nuclease contamination. As a consequence of early work on power-compensation microcalorimeters, it has become standard in the field to use the binding of CMP to RNaseA as a standardization/quality control protocol and for training new users of VP-ITC instruments. This is a particularly dismal choice for those of us whose samples are readily degraded by this protein. For anyone considering using a shared ITC instrument, I would immediately inquire with the facility manager regarding this practice and urge them to change to a more benign standard as decontamination is a serious challenge. In our laboratory, for instance, we typically use the interaction between Ni(II) and histidine as our test system (256).

A typical cleaning protocol for the ITC cell entails soaking the cell with commercial detergents like 10% TopJob or Mr. Clean at room temperature or at 65 °C followed by extensive flushing with deionized water. Stringent cleaning involves a harsher detergent (5-20% Contrad-70), also followed by flushing with water. Unfortunately, neither of these treatments adequately removes nuclease contamination. If RNase has ever been used in the system or if non-RNA users have potentially contaminated the cell and syringe, one **MUST** decontaminate the instrument prior to use. Plan well in advance as this process can take as much as a week to complete depending on the severity of the contamination. Typically, we soak the ITC cell and syringe in neat RNase *Zap*®

(Ambion) overnight at room temperature and then rinse the cell with 1L of water. After this treatment, the ITC cell can be tested for nuclease contamination by letting an RNA sample stand in the cell for several hours before removing it and assaying it for degradation by PAGE. This process can be repeated several times and interspersed with treatment of the cell with 1% SDS at 65 °C or 0.1% pepsin at room temperature. Do not proceed until you sample is stable against degradation overnight.

Separate from RNase contamination that leads to wholesale degradation of the sample, a dirty ITC cell will lead to poor data quality. The common signature of this problem is significant baseline drift during ITC runs. Although minor drifts in the baseline (upward or downward) can be expected and corrected during data analysis, severe drift, often with non-linearity, is difficult to fit. Another indication of contamination in the cell is the reference power at the start of a titration. During an experiment the raw ITC data is plotted as the differential power (μCal/s) vs. time (s). The user sets the reference power at the beginning of a titration. If one sets the reference power to 30 μCal/s this setting means that 30 μCal/s is continuously supplied to the heater around the reference cell. If the ITC is behaving well and the cell is clean, the baseline in the raw differential power plot should be close to this value - most often 1-2 μCal/s lower than the set point. If the initial baseline is considerably lower that the reference power one should clean the cell prior to filling it with a precious RNA sample.

Once the user is confident that the cell is clean and free of nuclease contamination, the samples can be loaded into the cell and syringe respectively. This is a step that requires a certain amount of practice and dexterity to keep the sample bubble free. Training new users with water or buffer is highly encouraged before moving

on to more expensive RNA samples. Because buffer mismatch is of critical concern, it is important to rinse the cell and the syringe with the binding buffer prior to loading experimental samples. Both the cell and syringe should be thoroughly emptied to avoid diluting the experimental samples although a tiny amount of dilution is inevitable. Even though the titrant syringe only holds ~280 µL, plan on having ~400 µL of titrant solution to load the instrument easily and keep bubbles from forming inside the barrel.

A1.5.6 Instrument settings for data collection

There are a modestly large number of experimental parameters that need to be set by the user at the start of an ITC experiment. John Tellinghuisen has studied the systematic errors associated with ITC experiments and the manner in which these settings affect the data quality (253,257,258). He has put forth some excellent guidelines regarding certain parameters that have particularly significant impact on the data. Three simple rules include: a) minimize the number of total number of injections b) maximize the starting concentration of the titrate in the cell (M_T), and c) use **Eq. A6** to set the titrant concentration to keep the c-value between 10 and 100. These are good guidelines to follow when starting out, but they do require preliminary knowledge of the approximate association constant as discussed above. All of the adjustable parameters for a VP-ITC are listed in **Table A1** together with some initial recommendations for the ITC novice. In addition, the table includes some notes regarding artifacts or issues that might arise if a given parameter is set inappropriately.

The injection volumes and the delay time between injections are two of the most asked about variables. It is very common if you look at ITC data that the first peak

appears aberrantly small. Mizoue et al. have shown that this first injection anomaly results from a volumetric error due to the backlash in the motorized screw used to drive the syringe plunger (223). Most people set the first injection to a very small volume (typically 1-2 μL) and simply discard this point prior to data analysis to fix the issue. Alternatively, you can use the software to drive the syringe forward a couple of microliters prior to loading it into the cell for the start of the reaction. If you do this, be sure to wipe off the tip of the syringe to remove that material prior to inserting the needle into the ITC cell. The volume of subsequent injections needs to be set so that: a) a significant amount of heat is being produced by each injection; and b) a reasonable number of points are situated in the transition region of the titration curve. This will be highly dependent on the system being studied. If cell concentrations are low or the reaction enthalpy is small, larger injections will be required to obtain a suitable signal. Most people use constant volumes for all injections after the first one. This is more of an aesthetic issue than a scientific one. The user also needs to set the spacing time between injections. The critical issue here is that the system must return to a good baseline after every injection. Failure to achieve this will adversely affect the data quality due to poor integration of the peak area. A good starting point is usually around 300 s. This time can be shortened if reaction returns to baseline rapidly and It can be lengthened if a flat baseline between injections is not achieved. Setting the injection spacing arbitrarily high can adversely affect your data due to baseline drift and also makes your experiment unnecessarily long which is important if you are being charged by the hour for use of a facility instrument. For most of the studies a stirring speed in the range of about 260-310 rpm works well and provides a stable baseline. For binding

Table A1. Typical experimental parameters for an ITC experiment [a]

Experimental Parameter	Description	Setting	Comments
Cell Temperature	The desired reaction temperature	2-80 °C [b]	Cooling the sample below the experimental temperature reduces ITC equilibration time, when *No check temp* option is off
No of Injections	Total number of injection made throughout the experiment	25-35 [c]	The number of injections depend on the concentration the user sets, having considerable amounts of points after saturation is beneficial to obtain background heat information
Injection Volume	The injection volume to be injected into the ITC cell in a single injection	1st Inj: 1-2 µL Rest of Inj: 7-12 µL	The first injection is set to a lesser value to account for the backlash effect.
Injection Spacing	Time between two injections	250-500 s	The time between injections should be large enough to allow the DT signal to return to the baseline. The spacing also depends on the size of the peak
Titrate/Titrant Concentration	The concentration of the titrate/titrant that is input by the user which is also used for data analysis	Cell : 1.8 µM Syringe : 42 µM	This example concentration was used for a 1:1 RNA-RNA interaction where the $\Delta H \sim 40$ kCal/mol [d]
ITC Equilibration Options	Equilibration options determine the modes of equilibration steps the VP-ITC needs to perform before an injection. The available mode are **No Check Temp, Fast Equilibrium and Automatic**	Automatic	(a) No check temp: The experiment starts at the current cell temperature as the entered value for the cell temperature is ignored (b) Fast Equilibration: This setting will avoid the pre-stirring equilibration (c) Automatic : This will perform all equilibrations
Reference power	Amount of power that is supplied to the heater in the reference cell.	25-30 µCal/sec	Generally setting a much higher value than is required may affect the sensitivity of the instrument
Initial Delay	The delay between the start of data collection	60-100 s	A stable baseline needs to be developed prior to injections

	(baseline) and the first injection		
Stirring Speed	An optimum stirring speed is a value between efficient mixing and providing a stable a baseline	270-310 rpm	Higher stirring rates may cause high frequency noise in ITC data, whereas lower stirring rates may cause broader peaks
Feed Back Mode	The relative response times are indicated, **Fast** for faster response times, **None** mode will have the slowest response time but more sensitive for heat changes	Fast	For slow processes **None** mode is more suitable
Filter period	The data sampled from the cell feedback circuit is averaged over this period to produce a single data point	2 s	For slow reactions the filter period can be increased for better resolution of data

[a]Based on use with a VP-ITC from MicroCal. Specific parameter settings might need to be adjusted for use with other instruments.
[b]At higher temperatures the user must consider structural transitions of the titrate and titrant. Effects on the energetics of a reaction can be observed as much as 20 °C below the unfolding transition and these effects become increasingly significant as one gets closer to the T_M of any of the reactants or products.
[c]A terminal ratio of about 3 is obtained for a 1:1 interaction (i.e. The final excess ratio of the titrant over the titrate). In a separate study it was indicated that when the number of injections were limited to ~10 the errors in ΔH and K was significantly lower (257)
[d]The reaction was performed in 10mM HEPES, 10mM $MgCl_2$, at 25 °C.

systems that have c-values greater than 1000 a higher stirring speed needs to be maintained. (NOTE: under these conditions, the reaction enthalpy can be measured but the value for the association constant will not be accurate). If the sample is viscous stirring speeds will have to be adjusted so that a suitable baseline is obtained.

The last experimental parameter that needs to be set is the reference power. This value is essentially the amount of power that is used to heat the reference cell at the resting state and dictates the maximum heat that can be accommodated during a single injection. Setting the value too high can lead to a loss in instrument sensitivity and setting it too low leads to artifacts from truncated peaks. In practice this parameter can be set to 25 – 30 μCal/s, a value that has been suitable for all of the RNA transitions we have studied. If one had particularly large injection volumes or an exceptionally exothermic reaction, one might alter this value, but one could just as easily use smaller injections or more dilute solutions instead.

A1.5.7 Data Analysis

Having followed the advice above, at this point you hopefully have a suitable power curve for the titration of your system. Now it is time to convert that power curve into a thermogram. ΔH_{obs} represents an aggregate measurement of all heat sources in the cell including some that are of interest and some that are not. The user needs to recognize and remove the unwanted background contributions during data analysis. Two types of background corrections are common. The first is a separate background run in which using the exact same ITC parameters one injects titrate into the ITC containing buffer only. The second method of performing a background correction involves including

several injections at the end of the run in which no reaction occurs. There are pros and cons of each method. The independent run takes more time and material but is not subject to issues of non-specific binding effects. The use of terminal injections at the end of the titration ensures that the materials are exactly the same as those that were used to collect the primary data and that there are no inadvertent minor changes in the preparation of the buffers (such as from pipetting errors). Our primary rationale for using the terminal injections is to be certain that users collect the background data and that the background data cannot be separated from the ITC thermograms themselves.

In general, for a well-behaved system, the background heats are quite small. On occasion, however, the background injections can identify significant experimental artifacts. Salt and buffer discrepancies show up in these traces as do monomer-dimerequilibria (or higher aggregates) since upon dilution the aggregates may dissociate. If one suspects such behavior from one of the binding partners it can be confirmed by performing the ITC experiment at different concentrations or by switching the titrate and titrant to lower the concentration of the offending species.

Once raw data is corrected for background heats by either method above, the data are fit to a model that is appropriate to the system of interest. Most commercial ITC manufactures provide data analysis software containing the most common binding models. The software normalizes the heat of binding as a function of ligand concentration, automatically sets the baseline, performs a volume correction based on the dilution of the titrate during each injection and integrates each peak from the baseline before the binding model is selected by the user. Although most of the time the automated functions set the baseline appropriately, there may be instances where the

user may need to do this manually if the data collection was non-optimal (e.g. due to poor baselines between injections resulting from slow binding kinetics). The integrated data then is used to determine n, K_B, ΔH, by least-squares minimization of **Eq A3**.

The other aspect of data analysis has to do with fitting parameters. For a run-of-the- mill experiment, one will typically fit all three general parameters, n, K_B and ΔH. This provides a good level of quality control feedback on the experiment. If one is looking at an interaction that should be 1:1 but the stoichiometry is coming out to 0.7 - 0.8 (or 1.2 – 1.3), there is a reasonable likelihood that there is an experimental problem involving inaccurate concentration determination of one of the samples or an issue with alternative conformations that make a sub-population of the material unreactive.

In certain circumstances, one can fix parameters that are not accurately determined by the experiment. For instance, when the c-value is very high due to a very tight K_d (a situation that would be encountered if looking at nucleic acid secondary structure formation for instance), one can use ITC to obtain the reaction enthalpy but not the equilibrium constant. The software will allow you to lock parameters during fitting. Fixing the K_d at an independently determined value (such as from a thermal melting study) can get around the problem of an indeterminate value due to high c-values. Before doing this, however, one should be aware that there is an inherent danger as the analysis will only be as good as the independent measurement and only as reliable as the equivalence between the experimental conditions of the two studies.

A1.5.8 Special cases. What do I do if my K_d is too tight?

In practice, the range of accessible interactions is not usually a problem for studies with most RNAs. Because RNAs are highly charged, the affinity for small molecules and proteins often can be shifted by altering the ionic conditions. For the rare case that might push the envelope, however, there are two approaches to extend the range of affinities accessible to ITC analysis. The first involves a single injection method where the titrant in the syringe is added to the cell in one injection long injection (259). This process provides better definition of the transition for tight interactions and allows K_d*s* as low as 10 pM to be measured. The second approach involves a displacement titration methodology where a weak binding ligand is pre-bound to the macromolecule and the titration follows ligand exchange (227). This technique requires that the user know the binding constants of the weakly binding ligand in advance from a separate experiment. The range of affinities accessible using this method is entirely dependent on the competitive binding ligands available for use.

A1.6 Concluding remarks

The advantage that ITC holds on obtaining thermodynamics for binding interactions is that heat is the primary observable parameter that is a universal property of binding. Most other methodologies require modifications or labeling of the sample. One other advantage an ITC user enjoys over alternate binding methods is the ability to obtain the entire thermodynamic profile of a reaction from a single experiment without resorting to a concentration series (UV melting) or a temperature series for a van't Hoff treatment. As modern ITC instruments continue to become more sensitive and versatile, sample

requirements – currently the biggest disadvantage of ITC studies - will also decrease further. The ability to detect the once neglected heat capacity changes in RNA conformational changes and to probe the thermodynamic contribution arising from complex molecular systems makes ITC an important tool for RNA biochemistry and biophysics.

A1.7 Acknowledgements

This work was supported by NIH grants GM-065439 and GM-075068. Andrew L. Feig is a Cottrell Scholar of the Research Corporation.

REFERENCES

1. Wassarman, K.M. (2007) 6S RNA: a small RNA regulator of transcription. *Curr Opin Microbiol*, **10**, 164-168.
2. Altuvia, S., Zhang, A., Argaman, L., Tiwari, A. and Storz, G. (1998) The Escherichia coli OxyS regulatory RNA represses fhlA translation by blocking ribosome binding. *EMBO J*, **17**, 6069-6075.
3. Kawamoto, H., Morita, T., Shimizu, A., Inada, T. and Aiba, H. (2005) Implication of membrane localization of target mRNA in the action of a small RNA: mechanism of post-transcriptional regulation of glucose transporter in Escherichia coli. *Genes Dev*, **19**, 328-338.
4. Stork, M., Di Lorenzo, M., Welch, T.J. and Crosa, J.H. (2007) Transcription termination within the iron transport-biosynthesis operon of Vibrio anguillarum requires an antisense RNA. *J Bacteriol*, **189**, 3479-3488.
5. Waters, L.S. and Storz, G. (2009) Regulatory RNAs in bacteria. *Cell*, **136**, 615-628.
6. Brantl, S. (2007) Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr Opin Microbiol*, **10**, 102-109.
7. Breaker, R.R. (2011) Prospects for riboswitch discovery and analysis. *Mol Cell*, **43**, 867-879.
8. Ding, Y., Davis, B.M. and Waldor, M.K. (2004) Hfq is essential for Vibrio cholerae virulence and downregulates sigma expression. *Mol Microbiol*, **53**, 345-354.

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

10. Simonsen, K.T., Nielsen, G., Bjerrum, J.V., Kruse, T., Kallipolitis, B.H. and Moller-Jensen, J. (2011) A role for the RNA chaperone Hfq in controlling adherent-invasive Escherichia coli colonization and virulence. *PloS one*, **6**, e16387.

11. Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J. and Nagai, K. (2009) Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. *Nature*, **458**, 475-480.

12. He, W. and Parker, R. (2000) Functions of Lsm proteins in mRNA degradation and splicing. *Current opinion in cell biology*, **12**, 346-350.

13. Kufel, J., Allmang, C., Petfalski, E., Beggs, J. and Tollervey, D. (2003) Lsm Proteins are required for normal processing and stability of ribosomal RNAs. *The Journal of biological chemistry*, **278**, 2147-2156.

14. Link, T.M., Valentin-Hansen, P. and Brennan, R.G. (2009) Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. *Proc Natl Acad Sci U S A*, **106**, 19292-19297.

15. Beich-Frandsen, M., Vecerek, B., Konarev, P.V., Sjoblom, B., Kloiber, K., Hammerle, H., Rajkowitsch, L., Miles, A.J., Kontaxis, G., Wallace, B.A. *et al.* (2011) Structural insights into the dynamics and function of the C-terminus of the E. coli RNA chaperone Hfq. *Nucleic Acids Res*, **39**, 4900-4915.

16. Schumacher, M.A., Pearson, R.F., Moller, T., Valentin-Hansen, P. and Brennan, R.G. (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J*, **21**, 3546-3556.

17. Wang, W., Wang, L., Zou, Y., Zhang, J., Gong, Q., Wu, J. and Shi, Y. (2011) Cooperation of Escherichia coli Hfq hexamers in DsrA binding. *Genes & development*, **25**, 2106-2117.

18. Someya, T., Baba, S., Fujimoto, M., Kawai, G., Kumasaka, T. and Nakamura, K. (2011) 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*.

19. Fender, A., Elf, J., Hampel, K., Zimmermann, B. and Wagner, E.G. (2010) RNAs actively cycle on the Sm-like protein Hfq. *Genes Dev*, **24**, 2621-2626.

20. Hussein, R. and Lim, H.N. (2011) Disruption of small RNA signaling caused by competition for Hfq. *Proc Natl Acad Sci U S A*, **108**, 1110-1115.

21. Mikulecky, P.J., Kaw, M.K., Brescia, C.C., Takach, J.C., Sledjeski, D.D. and Feig, A.L. (2004) Escherichia coli Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nat Struct Mol Biol*, **11**, 1206-1214.

22. Sauer, E. and Weichenrieder, O. (2011) Structural basis for RNA 3'-end recognition by Hfq. *Proc Natl Acad Sci U S A*, **108**, 13065-13070.

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

24. Lease, R.A. and Woodson, S.A. (2004) Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. *J Mol Biol*, **344**, 1211-1223.

25. Sun, X. and Wartell, R.M. (2006) Escherichia coli Hfq binds A18 and DsrA domain II with similar 2:1 Hfq6/RNA stoichiometry using different surface sites. *Biochemistry*, **45**, 4875-4887.

26. Adamson, D.N. and Lim, H.N. (2011) Essential requirements for robust signaling in Hfq dependent small RNA networks. *PLoS computational biology*, **7**, e1002138.

27. Moon, K. and Gottesman, S. (2011) Competition among Hfq-binding small RNAs in Escherichia coli. *Mol Microbiol*.

28. Updegrove, T.B., Correia, J.J., Chen, Y., Terry, C. and Wartell, R.M. (2011) The stoichiometry of the Escherichia coli Hfq protein bound to RNA. *RNA*, **17**, 489-500.

29. Lybecker, M.C., Abel, C.A., Feig, A.L. and Samuels, D.S. (2010) Identification and function of the RNA chaperone Hfq in the Lyme disease spirochete Borrelia burgdorferi. *Mol Microbiol*, **78**, 622-635.

30. Salim, N.N. and Feig, A.L. (2010) An upstream Hfq binding site in the fhlA mRNA leader region facilitates the OxyS-fhlA interaction. *PLoS One*, **5**.

31. Nielsen, J.S., Lei, L.K., Ebersbach, T., Olsen, A.S., Klitgaard, J.K., Valentin-Hansen, P. and Kallipolitis, B.H. (2010) Defining a role for Hfq in Gram-positive bacteria: evidence for Hfq-dependent antisense regulation in Listeria monocytogenes. *Nucleic Acids Res*, **38**, 907-919.

32. Arluison, V., Folichon, M., Marco, S., Derreumaux, P., Pellegrini, O., Seguin, J., Hajnsdorf, E. and Regnier, P. (2004) The C-terminal domain of Escherichia coli Hfq increases the stability of the hexamer. *Eur J Biochem*, **271**, 1258-1265.

33. Olsen, A.S., Moller-Jensen, J., Brennan, R.G. and Valentin-Hansen, P. (2010) C-terminally truncated derivatives of Escherichia coli Hfq are proficient in riboregulation. *J Mol Biol*, **404**, 173-182.

34. Sittka, A., Lucchini, S., Papenfort, K., Sharma, C.M., Rolle, K., Binnewies, T.T., Hinton, J.C. and Vogel, J. (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet*, **4**, e1000163.

35. Sittka, A., Sharma, C.M., Rolle, K. and Vogel, J. (2009) Deep sequencing of Salmonella RNA associated with heterologous Hfq proteins in vivo reveals small RNAs as a major target class and identifies RNA processing phenotypes. *RNA Biol*, **6**, 266-275.

36. Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L. and Storz, G. (1997) A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell*, **90**, 43-53.

37. Lease, R.A., Cusick, M.E. and Belfort, M. (1998) Riboregulation in Escherichia coli: DsrA RNA acts by RNA:RNA interactions at multiple loci. *Proc Natl Acad Sci U S A*, **95**, 12456-12461.

38. Majdalani, N., Cunning, C., Sledjeski, D., Elliott, T. and Gottesman, S. (1998) 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*, **95**, 12462-12467.

39. Storz, G., Vogel, J. and Wassarman, K.M. (2011) Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Mol Cell*, **43**, 880-891.

40. Papenfort, K. and Vogel, J. (2009) Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level. *Res Microbiol*, **160**, 278-287.

41. Papenfort, K., Pfeiffer, V., Mika, F., Lucchini, S., Hinton, J.C. and Vogel, J. (2006) SigmaE-dependent small RNAs of Salmonella respond to membrane stress by accelerating global omp mRNA decay. *Mol Microbiol*, **62**, 1674-1688.

42. Thompson, K.M., Rhodius, V.A. and Gottesman, S. (2007) SigmaE regulates and is regulated by a small RNA in Escherichia coli. *J Bacteriol*, **189**, 4243-4256.

43. Papenfort, K., Bouvier, M., Mika, F., Sharma, C.M. and Vogel, J. (2010) Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA. *Proc Natl Acad Sci U S A*, **107**, 20435-20440.

44. Chaulk, S.G., Smith-Frieday, M.N., Arthur, D.C., Culham, D.E., Edwards, R.A., Soo, P., Frost, L.S., Keates, R.A.B., Glover, J.N.M. and Wood, J.M. (2011) ProQ Is an RNA Chaperone that Controls ProP Levels in Escherichia coli. *Biochemistry*, **50**, 3095-3106.

45. Huttenhofer, A. and Noller, H.F. (1994) Footprinting mRNA-ribosome complexes with chemical probes. *EMBO J*, **13**, 3892-3901.

46. Beyer, D., Skripkin, E., Wadzack, J. and Nierhaus, K.H. (1994) How the ribosome moves along the mRNA during protein synthesis. *J Biol Chem*, **269**, 30713-30717.

47. Bouvier, M., Sharma, C.M., Mika, F., Nierhaus, K.H. and Vogel, J. (2008) Small RNA binding to 5' mRNA coding region inhibits translational initiation. *Mol Cell*, **32**, 827-837.

48. Holmqvist, E., Reimegard, J., Sterk, M., Grantcharova, N., Romling, U. and Wagner, E.G. (2010) Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. *EMBO J*, **29**, 1840-1850.

49. Opdyke, J.A., Kang, J.-G. and Storz, G. (2004) GadY, a small-RNA regulator of acid response genes in Escherichia coli. *J Bacteriol*, **186**, 6698-6705.

50. Majdalani, N., Cunning, C., Sledjeski, D., Elliott, T. and Gottesman, S. (1998) 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*, **95**, 12462-12467.

51. Majdalani, N., Chen, S., Murrow, J., St John, K. and Gottesman, S. (2001) Regulation of RpoS by a novel small RNA: the characterization of RprA. *Mol Microbiol*, **39**, 1382-1394.

52. Urban, J.H. and Vogel, J. (2008) Two seemingly homologous noncoding RNAs act hierarchically to activate glmS mRNA translation. *PLoS Biol*, **6**, e64.

53. Mohanty, B.K., Maples, V.F. and Kushner, S.R. (2004) The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in Escherichia coli. *Mol. Microbiol.*, **54**, 905-920.

54. Hajnsdorf, E. and Regnier, P. (2000) Host factor Hfq of Escherichia coli stimulates elongation of poly(A) tails by poly(A) polymerase I. *Proc Natl Acad Sci U S A*, **97**, 1501-1505.

55. Otaka, H., Ishikawa, H., Morita, T. and Aiba, H. (2011) PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proc Natl Acad Sci U S A*, **108**, 13059-13064.

56. Soper, T., Mandin, P., Majdalani, N., Gottesman, S. and Woodson, S.A. (2010) Positive regulation by small RNAs and the role of Hfq. *Proc Natl Acad Sci U S A*, **107**, 9602-9607.

57. Soper, T.J. and Woodson, S.A. (2008) The rpoS mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. *RNA*, **14**, 1907-1917.

58. Lorenz, C., Gesell, T., Zimmermann, B., Schoeberl, U., Bilusic, I., Rajkowitsch, L., Waldsich, C., von Haeseler, A. and Schroeder, R. (2010) Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts. *Nucleic Acids Res*, **38**, 3794-3808.

59. Palazzo, A.F., Springer, M., Shibata, Y., Lee, C.S., Dias, A.P. and Rapoport, T.A. (2007) The signal sequence coding region promotes nuclear export of mRNA. *PLoS Biol*, **5**, e322.

60. Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L. and Storz, G. (1997) A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell*, **90**, 43-53.

61. Vanderpool, C.K. and Gottesman, S. (2004) Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the

glucose phosphoenolpyruvate phosphotransferase system. *Mol Microbiol*, **54**, 1076-1089.

62. Maki, K., Morita, T., Otaka, H. and Aiba, H. (2010) A minimal base-pairing region of a bacterial small RNA SgrS required for translational repression of ptsG mRNA. *Mol. Microbiol.*, **76**, 782-792.

63. Soper, T.J., Doxzen, K. and Woodson, S.A. (2011) Major role for mRNA binding and restructuring in sRNA recruitment by Hfq. *RNA*, **17**, 1544-1550.

64. Olejniczak, M. (2011) Despite similar binding to the Hfq protein regulatory RNAs widely differ in their competition performance. *Biochemistry*, **50**, 4427-4440.

65. Carpousis, A.J. (2007) The RNA degradosome of Escherichia coli: an mRNA-degrading machine assembled on RNase E. *Annu Rev Microbiol*, **61**, 71-87.

66. Callaghan, A.J., Marcaida, M.J., Stead, J.A., McDowall, K.J., Scott, W.G. and Luisi, B.F. (2005) Structure of Escherichia coli RNase E catalytic domain and implications for RNA turnover. *Nature*, **437**, 1187-1191.

67. Madhugiri, R., Basineni, S.R. and Klug, G. (2010) Turn-over of the small non-coding RNA RprA in E. coli is influenced by osmolarity. *Mol. Genet. Genomics*, **284**, 307-318.

68. Moll, I., Afonyushkin, T., Vytvytska, O., Kaberdin, V.R. and Blaesi, U. (2003) Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs. *RNA*, **9**, 1308-1314.

69. Viegas, S.C., Silva, I.J., Saramago, M., Domingues, S. and Arraiano, C.M. (2011) Regulation of the small regulatory RNA MicA by ribonuclease III: a target-dependent pathway. *Nucleic Acids Res*, **39**, 2918-2930.

70. Kalamorz, F., Reichenbach, B., Maerz, W., Rak, B. and Goerke, B. (2007) Feedback control of glucosamine-6-phosphate synthase GlmS expression depends on the small RNA GlmZ and involves the novel protein YhbJ in Escherichia coli. *Mol. Microbiol.*, **65**, 1518-1533.

71. Afonyushkin, T., Vecerek, B., Moll, I., Blasi, U. and Kaberdin, V.R. (2005) Both RNase E and RNase III control the stability of sodB mRNA upon translational inhibition by the small regulatory RNA RyhB. *Nucleic Acids Res*, **33**, 1678-1689.

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

73. Prevost, K., Desnoyers, G., Jacques, J.-F., Lavoie, F. and Masse, E. (2011) Small RNA-induced mRNA degradation achieved through both translation block and activated cleavage. *Genes Dev.*, **25**, 385-396.

74. Morita, T., Maki, K. and Aiba, H. (2005) RNase E-based ribonucleoprotein complexes: Mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev.*, **19**, 2176-2186.

75. Kaberdin, V.R. (2003) Probing the substrate specificity of Escherichia coli RNase E using a novel oligonucleotide-based assay. *Nucleic Acids Res*, **31**, 4710-4716.

76. Mackie, G.A. (1998) Ribonuclease E is a 5'-end-dependent endonuclease. *Nature*, **395**, 720-723.

77. Kushner, S.R. (2002) mRNA decay in Escherichia coli comes of age. *J Bacteriol*, **184**, 4658-4665; discussion 4657.

78. Carabetta, V.J., Mohanty, B.K., Kushner, S.R. and Silhavy, T.J. (2009) The response regulator SprE (RssB) modulates polyadenylation and mRNA stability in Escherichia coli. *J. Bacteriol.*, **191**, 6812-6821.

79. Ikeda, Y., Yagi, M., Morita, T. and Aiba, H. (2011) Hfq binding at RhlB-recognition region of RNase E is crucial for the rapid degradation of target mRNAs mediated by sRNAs in Escherichia coli. *Mol. Microbiol.*, **79**, 419-432.

80. Morita, T. and Aiba, H. (2011) RNase E action at a distance: degradation of target mRNAs mediated by an Hfq-binding small RNA in bacteria. *Genes & development*, **25**, 294-298.

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

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

83. Keiler, K.C., Waller, P.R. and Sauer, R.T. (1996) Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science*, **271**, 990-993.

84. Goerke, B. and Vogel, J. (2008) Noncoding RNA control of the making and breaking of sugars. *Genes Dev.*, **22**, 2914-2925.

85. Urban, J.H., Papenfort, K., Thomsen, J., Schmitz, R.A. and Vogel, J. (2007) A conserved small RNA promotes discoordinate expression of the glmUS operon mRNA to activate GlmS synthesis. *J Mol Biol*, **373**, 521-528.

86. Carpousis, A.J., Van Houwe, G., Ehretsmann, C. and Krisch, H.M. (1994) Copurification of E. coli RNAase E and PNPase: evidence for a specific association between two enzymes important in RNA processing and degradation. *Cell*, **76**, 889-900.

87. Py, B., Causton, H., Mudd, E.A. and Higgins, C.F. (1994) A protein complex mediating mRNA degradation in Escherichia coli. *Mol Microbiol*, **14**, 717-729.

88. De, L.N. and Gottesman, S. (2011) Role of polynucleotide phosphorylase in sRNA function in Escherichia coli. *RNA*, **17**, 1172-1189.

89. Godefroy, T., Cohn, M. and Grunberg-Manago, M. (1970) Kinetics of polymerization and phosphorolysis reactions of E. coli polynucleotide phosphorylase. Role of oligonucleotides in polymerization. *European journal of biochemistry / FEBS*, **12**, 236-249.

90. Sulewski, M., Marchese-Ragona, S.P., Johnson, K.A. and Benkovic, S.J. (1989) Mechanism of polynucleotide phosphorylase. *Biochemistry*, **28**, 5855-5864.

91. Mohanty, B.K. and Kushner, S.R. (2000) Polynucleotide phosphorylase, RNase II and RNase E play different roles in the in vivo modulation of polyadenylation in Escherichia coli. *Mol Microbiol*, **36**, 982-994.

92. Slomovic, S., Portnoy, V., Yehudai-Resheff, S., Bronshtein, E. and Schuster, G. (2008) Polynucleotide phosphorylase and the archaeal exosome as poly(A)-polymerases. *Biochimica et biophysica acta*, **1779**, 247-255.

93. Mohanty, B.K. and Kushner, S.R. (2006) The majority of Escherichia coli mRNAs undergo post-transcriptional modification in exponentially growing cells. *Nucleic Acids Res*, **34**, 5695-5704.

94. O'Hara, E.B., Chekanova, J.A., Ingle, C.A., Kushner, Z.R., Peters, E. and Kushner, S.R. (1995) Polyadenylylation helps regulate mRNA decay in Escherichia coli. *Proc Natl Acad Sci U S A*, **92**, 1807-1811.

95. Mohanty, B.K. and Kushner, S.R. (1999) Residual polyadenylation in poly(A) polymerase I (pcnB) mutants of Escherichia coli does not result from the activity encoded by the f310 gene. *Mol Microbiol*, **34**, 1109-1119.

96. Mohanty, B.K. and Kushner, S.R. (1999) Analysis of the function of Escherichia coli poly(A) polymerase I in RNA metabolism. *Mol Microbiol*, **34**, 1094-1108.

97. Charollais, J., Dreyfus, M. and Iost, I. (2004) CsdA, a cold-shock RNA helicase from Escherichia coli, is involved in the biogenesis of 50S ribosomal subunit. *Nucleic Acids Res*, **32**, 2751-2759.

98. Charollais, J., Pflieger, D., Vinh, J., Dreyfus, M. and Iost, I. (2003) The DEAD-box RNA helicase SrmB is involved in the assembly of 50S ribosomal subunits in Escherichia coli. *Mol Microbiol*, **48**, 1253-1265.

99. Karginov, F.V. and Uhlenbeck, O.C. (2004) Interaction of Escherichia coli DbpA with 23S rRNA in different functional states of the enzyme. *Nucleic Acids Res*, **32**, 3028-3032.

100. Coburn, G.A., Miao, X., Briant, D.J. and Mackie, G.A. (1999) Reconstitution of a minimal RNA degradosome demonstrates functional coordination between a 3' exonuclease and a DEAD-box RNA helicase. *Genes & development*, **13**, 2594-2603.

101. Prud'homme-Genereux, A., Beran, R.K., Iost, I., Ramey, C.S., Mackie, G.A. and Simons, R.W. (2004) Physical and functional interactions among RNase E,

polynucleotide phosphorylase and the cold-shock protein, CsdA: evidence for a 'cold shock degradosome'. *Mol Microbiol*, **54**, 1409-1421.

102. Resch, A., Vecerek, B., Palavra, K. and Blasi, U. (2010) Requirement of the CsdA DEAD-box helicase for low temperature riboregulation of rpoS mRNA. *RNA Biol*, **7**, 796-802.

103. Gao, J., Lee, K., Zhao, M., Qiu, J., Zhan, X., Saxena, A., Moore, C.J., Cohen, S.N. and Georgiou, G. (2006) Differential modulation of E. coli mRNA abundance by inhibitory proteins that alter the composition of the degradosome. *Mol Microbiol*, **61**, 394-406.

104. Urban, J.H., Papenfort, K., Thomsen, J., Schmitz, R.A. and Vogel, J. (2007) A conserved small RNA promotes discoordinate expression of the glmUS operon mRNA to activate GlmS synthesis. *J Mol Biol*, **373**, 521-528.

105. Kalamorz, F., Reichenbach, B., Marz, W., Rak, B. and Gorke, B. (2007) Feedback control of glucosamine-6-phosphate synthase GlmS expression depends on the small RNA GlmZ and involves the novel protein YhbJ in Escherichia coli. *Mol Microbiol*, **65**, 1518-1533.

106. Beich-Frandsen, M., Vecerek, B., Konarev, P.V., Sjoeblom, B., Kloiber, K., Haemmerle, H., Rajkowitsch, L., Miles, A.J., Kontaxis, G., Wallace, B.A. *et al.* (2011) Structural insights into the dynamics and function of the C-terminus of the E. coli RNA chaperone Hfq. *Nucleic Acids Res.*, **39**, 4900-4915.

107. Man, S., Cheng, R., Miao, C., Gong, Q., Gu, Y., Lu, X., Han, F. and Yu, W. (2011) Artificial trans-encoded small non-coding RNAs specifically silence the selected gene expression in bacteria. *Nucleic Acids Res.*, **39**, e50.

108. Hao, Y., Zhang, Z.J., Erickson, D.W., Huang, M., Huang, Y., Li, J., Hwa, T. and Shi, H. (2011) Quantifying the sequence-function relation in gene silencing by bacterial small RNAs. *Proc Natl Acad Sci U S A*, **108**, 12473-12478.

109. Zhang, C. (2009) Novel functions for small RNA molecules. *Curr Opin Mol Ther*, **11**, 641-651.

110. Aiba, H. (2007) Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr Opin Microbiol*, **10**, 134-139.

111. Vogel, J. (2009) A rough guide to the non-coding RNA world of Salmonella. *Mol Microbiol*, **71**, 1-11.

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

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

114. Franze de Fernandez, M.T., Hayward, W.S. and August, J.T. (1972) Bacterial proteins required for replication of phage Q ribonucleic acid. Pruification and properties of host factor I, a ribonucleic acid-binding protein. *J Biol Chem*, **247**, 824-831.

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

116. Brennan, R.G. and Link, T.M. (2007) Hfq structure, function and ligand binding. *Curr Opin Microbiol*, **10**, 125-133.

117. Tsui, H.C., Leung, H.C. and Winkler, M.E. (1994) Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in Escherichia coli K-12. *Mol Microbiol*, **13**, 35-49.

118. Takada, A., Wachi, M. and Nagai, K. (1999) Negative regulatory role of the Escherichia coli hfq gene in cell division. *Biochem Biophys Res Commun*, **266**, 579-583.

119. Lee, T. and Feig, A.L. (2008) The RNA binding protein Hfq interacts specifically with tRNAs. *RNA*, **14**, 514-523.

120. Sittka, A., Pfeiffer, V., Tedin, K. and Vogel, J. (2007) The RNA chaperone Hfq is essential for the virulence of Salmonella typhimurium. *Mol Microbiol*, **63**, 193-217.

121. Sonnleitner, E., Hagens, S., Rosenau, F., Wilhelm, S., Habel, A., Jager, K.-E. and Blasi, U. (2003) Reduced virulence of a hfq mutant of Pseudomonas aeruginosa O1. *Microb Pathog*, **35**, 217-228.

122. Ansong, C., Yoon, H., Porwollik, S., Mottaz-Brewer, H., Petritis, B.O., Jaitly, N., Adkins, J.N., McClelland, M., Heffron, F. and Smith, R.D. (2009) Global systems-level analysis of Hfq and SmpB deletion mutants in Salmonella: implications for virulence and global protein translation. *PLoS One*, **4**, e4809.

123. Argaman, L. and Altuvia, S. (2000) fhlA repression by OxyS RNA: kissing complex formation at two sites results in a stable antisense-target RNA complex. *J Mol Biol*, **300**, 1101-1112.

124. Zhang, A., Wassarman, K.M., Ortega, J., Steven, A.C. and Storz, G. (2002) The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol Cell*, **9**, 11-22.

125. Soper, T.J. and Woodson, S.A. (2008) The rpoS mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. *RNA*, **14**, 1907-1917.

126. Lutz, S., Jacobi, A., Schlensog, V., Bohm, R., Sawers, G. and Bock, A. (1991) Molecular characterization of an operon (hyp) necessary for the activity of the three hydrogenase isoenzymes in Escherichia coli. *Mol Microbiol*, **5**, 123-135.

127. Updegrove, T., Wilf, N., Sun, X. and Wartell, R.M. (2008) Effect of Hfq on RprA-rpoS mRNA pairing: Hfq-RNA binding and the influence of the 5' rpoS mRNA leader region. *Biochemistry*, **47**, 11184-11195.

128. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, **25**, 1605-1612.

129. Mikulecky, P.J., Kaw, M.K., Brescia, C.C., Takach, J.C., Sledjeski, D.D. and Feig, A.L. (2004) Escherichia coli Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nat Struct Mol Biol*, **11**, 1206-1214.

130. Wilkinson, K.A., Merino, E.J. and Weeks, K.M. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc*, **1**, 1610-1616.

131. Wilkinson, K.A., Merino, E.J. and Weeks, K.M. (2005) RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts. *J Am Chem Soc*, **127**, 4659-4667.

132. Merino, E.J., Wilkinson, K.A., Coughlan, J.L. and Weeks, K.M. (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc*, **127**, 4223-4231.

133. Tuerk, C., Gauss, P., Thermes, C., Groebe, D.R., Gayle, M., Guild, N., Stormo, G., d'Aubenton-Carafa, Y., Uhlenbeck, O.C., Tinoco, I., Jr. *et al.* (1988) CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc Natl Acad Sci U S A*, **85**, 1364-1368.

134. Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M. and Giddings, M.C. (2008) ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA*, **14**, 1979-1990.

135. Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M. and Turner, D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A*, **101**, 7287-7292.

136. Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C. and Weeks, K.M. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol*, **6**, e96.

137. Mortimer, S.A. and Weeks, K.M. (2009) Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat Protoc*, **4**, 1413-1421.

138. Deigan, K.E., Li, T.W., Mathews, D.H. and Weeks, K.M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci U S A*, **106**, 97-102.

139. Wilkinson, K.A., Merino, E.J. and Weeks, K.M. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc*, **1**, 1610-1616.

140. Zhang, A., Wassarman, K.M., Rosenow, C., Tjaden, B.C., Storz, G. and Gottesman, S. (2003) Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol*, **50**, 1111-1124.

141. Sharma, C.M., Darfeuille, F., Plantinga, T.H. and Vogel, J. (2007) A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev*, **21**, 2804-2817.

142. Windbichler, N., von Pelchrzim, F., Mayer, O., Csaszar, E. and Schroeder, R. (2008) Isolation of small RNA-binding proteins from E. coli: evidence for frequent interaction of RNAs with RNA polymerase. *RNA Biol*, **5**, 30-40.

143. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1989). New York : Cold Spring Harbor Laboratory Press.

144. Mathews, D.H. (2006) RNA secondary structure analysis using RNAstructure. *Curr Protoc Bioinformatics*, **Chapter 12**, Unit 12 16.

145. Das, R., Laederach, A., Pearlman, S.M., Herschlag, D. and Altman, R.B. (2005) SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA*, **11**, 344-354.

146. Carthew, R.W. and Sontheimer, E.J. (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell*, **136**, 642-655.

147. Lee, T. and Feig, A.L. (2009) In Walter, N. G., Woodson, S. A. and Batey, R. T. (eds.), *Non-Protein Coding RNAs*. Springer Berlin Heidelberg, Vol. 13, pp. 249-271.

148. Sonnleitner, E., Napetschnig, J., Afonyushkin, T., Ecker, K., Vecerek, B., Moll, I., Kaberdin, V.R. and Blasi, U. (2004) Functional effects of variants of the RNA chaperone Hfq. *Biochem Biophys Res Commun*, **323**, 1017-1023.

149. Kobayashi, K., Ehrlich, S.D., Albertini, A., Amati, G., Andersen, K.K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S., Bessieres, P. *et al.* (2003) Essential Bacillus subtilis genes. *Proc Natl Acad Sci U S A*, **100**, 4678-4683.

150. Klein, D.J. and Ferre-D'Amare, A.R. (2006) Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science*, **313**, 1752-1756.

151. Wild, J. and Szybalski, W. (2004) Copy-control pBAC/oriV vectors for genomic cloning. *Methods Mol Biol*, **267**, 145-154.

152. Wild, J., Hradecna, Z. and Szybalski, W. (2002) Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. *Genome Res*, **12**, 1434-1444.

153. Moskaleva, O., Melnik, B., Gabdulkhakov, A., Garber, M., Nikonov, S., Stolboushkina, E. and Nikulin, A. (2010) The structures of mutant forms of Hfq from Pseudomonas aeruginosa reveal the importance of the conserved His57 for the protein hexamer organization. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, **66**, 760-764.

154. Besmer, E., Market, E. and Papavasiliou, F.N. (2006) The transcription elongation complex directs activation-induced cytidine deaminase-mediated DNA deamination. *Mol Cell Biol*, **26**, 4378-4385.

155. Folichon, M., Arluison, V., Pellegrini, O., Huntzinger, E., Regnier, P. and Hajnsdorf, E. (2003) The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation. *Nucleic Acids Res*, **31**, 7302-7310.

156. Moll, I., Afonyushkin, T., Vytvytska, O., Kaberdin, V.R. and Blasi, U. (2003) Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs. *RNA*, **9**, 1308-1314.

157. Mohanty, B.K., Maples, V.F. and Kushner, S.R. (2004) The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in Escherichia coli. *Mol Microbiol*, **54**, 905-920.

158. Resch, A., Vecerek, B., Palavra, K. and Blasi, U. (2010) Requirement of the CsdA DEAD-box helicase for low temperature riboregulation of rpoS mRNA. *RNA Biol*, **7**, 796-802.

159. Ikeda, Y., Yagi, M., Morita, T. and Aiba, H. (2011) Hfq binding at RhlB-recognition region of RNase E is crucial for the rapid degradation of target mRNAs mediated by sRNAs in Escherichia coli. *Mol Microbiol*, **79**, 419-432.

160. Beich-Frandsen, M., Vecerek, B., Sjoblom, B., Blasi, U. and Djinovic-Carugo, K. (2011) Structural analysis of full-length Hfq from Escherichia coli. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, **67**, 536-540.

161. Lauro, F.M., McDougald, D., Thomas, T., Williams, T.J., Egan, S., Rice, S., DeMaere, M.Z., Ting, L., Ertan, H., Johnson, J. *et al.* (2009) The genomic basis

of trophic strategy in marine bacteria. *Proc Natl Acad Sci U S A*, **106**, 15527-15533.

162. Vecerek, B., Rajkowitsch, L., Sonnleitner, E., Schroeder, R. and Blasi, U. (2008) The C-terminal domain of Escherichia coli Hfq is required for regulation. *Nucleic Acids Res*, **36**, 133-143.

163. Reuter, J.S. and Mathews, D.H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, **11**, 129.

164. Hannon, G.J., Rivas, F.V., Murchison, E.P. and Steitz, J.A. (2006) The expanding universe of noncoding RNAs. *Cold Spring Harb Symp Quant Biol*, **71**, 551-564.

165. Mattick, J.S. and Makunin, I.V. (2006) Non-coding RNA. *Hum Mol Genet*, **15 Spec No 1**, R17-29.

166. Gilbert, S.D. and Batey, R.T. (2006) Riboswitches: fold and function. *Chem Biol*, **13**, 805-807.

167. Coppins, R.L., Hall, K.B. and Groisman, E.A. (2007) The intricate world of riboswitches. *Curr Opin Microbiol*, **10**, 176-181.

168. Storz, G., Altuvia, S. and Wassarman, K.M. (2005) An abundance of RNA regulators. *Annu Rev Biochem*, **74**, 199-217.

169. Storz, G. and Haas, D. (2007) A guide to small RNAs in microorganisms. *Curr Opin Microbiol*, **10**, 93-95.

170. Rana, T.M. (2007) Illuminating the silence: understanding the structure and function of small RNAs. *Nat Rev Mol Cell Biol*, **8**, 23-36.

171. Mello, C.C. and Conte, D., Jr. (2004) Revealing the world of RNA interference. *Nature*, **431**, 338-342.

172. Brennan, R.G. and Link, T.M. (2007) Hfq structure, function and ligand binding. *Curr Opin Microbiol*, **10**, 125-133.

173. Majdalani, N., Vanderpool, C.K. and Gottesman, S. (2005) Bacterial small RNA regulators. *Crit Rev Biochem Mol Biol*, **40**, 93-113.

174. Eguchi, Y. and Tomizawa, J. (1990) Complex formed by complementary RNA stem-loops and its stabilization by a protein: function of CoIE1 Rom protein. *Cell*, **60**, 199-209.

175. Gerhart, E., Wagner, H. and Brantl, S. (1998) Kissing and RNA stability in antisense control of plasmid replication. *Trends Biochem Sci*, **23**, 451-454.

176. Paillart, J.C., Marquet, R., Skripkin, E., Ehresmann, C. and Ehresmann, B. (1996) Dimerization of retroviral genomic RNAs: structural and functional implications. *Biochimie*, **78**, 639-653.

177. Jossinet, F., Paillart, J.C., Westhof, E., Hermann, T., Skripkin, E., Lodmell, J.S., Ehresmann, C., Ehresmann, B. and Marquet, R. (1999) Dimerization of HIV-1 genomic RNA of subtypes A and B: RNA loop structure and magnesium binding. *Rna*, **5**, 1222-1234.

178. Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C. and Dumas, P. (2001) Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat Struct Biol*, **8**, 1064-1068.

179. Rist, M.J. and Marino, J.P. (2002) Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. *Biochemistry*, **41**, 14762-14770.

180. Sledjeski, D.D., Gupta, A. and Gottesman, S. (1996) The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in *Escherichia coli*. *EMBO*, **15**, 3993-4000.

181. Lease, R.A. and Belfort, M. (2000) A trans-acting RNA as a control switch in *Escherichia coli*: DsrA modulates function by forming alternative structures. *Proc Natl Acad Sci U S A*, **97**, 9919-9924.

182. Brunel, C., Marquet, R., Romby, P. and Ehresmann, C. (2002) RNA loop-loop interactions as dynamic functional motifs. *Biochimie*, **84**, 925-944.

183. Gregorian, R.S., Jr. and Crothers, D.M. (1995) Determinants of RNA hairpin loop-loop complex stability. *J Mol Biol*, **248**, 968-984.

184. Paillart, J.C., Westhof, E., Ehresmann, C., Ehresmann, B. and Marquet, R. (1997) Non-canonical interactions in a kissing loop complex: the dimerization initiation site of HIV-1 genomic RNA. *J Mol Biol*, **270**, 36-49.

185. Tinoco, I., Jr. and Bustamante, C. (1999) How RNA folds. *J. Mol. Biol.*, **293**, 271-281.

186. Liu, H.W., Cosa, G., Landes, C.F., Zeng, Y., Kovaleski, B.J., Mullen, D.G., Barany, G., Musier-Forsyth, K. and Barbara, P.F. (2005) Single-molecule FRET studies of important intermediates in the nucleocapsid-protein-chaperoned minus-strand transfer step in HIV-1 reverse transcription. *Biophys J*, **89**, 3470-3479.

187. Bernacchi, S., Ennifar, E., Toth, K., Walter, P., Langowski, J. and Dumas, P. (2005) Mechanism of hairpin-duplex conversion for the HIV-1 dimerization initiation site. *J Biol Chem*, **280**, 40112-40121.

188. Rueda, D., Zhao, R., Marshall, M., Aleman, E.A., Lamichhane, R. and Feig, A. (2010) Laser-Assisted Single-Molecule Refolding (LASR). *Biophysical Journal*, **99**, 1925-1931.

189. Plum, G.E. and Breslauer, K.J. (1995) Calorimetry of proteins and nucleic acids. *Curr Opin Struct Biol*, **5**, 682-690.

190. Jelesarov, I. and Bosshard, H.R. (1999) Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J Mol Recognit*, **12**, 3-18.

191. Feig, A.L. (2007) Applications of isothermal titration calorimetry in RNA biochemistry and biophysics. *Biopolymers*.

192. Salim, N.N. and Feig, A.L. (2009) Isothermal titration calorimetry of RNA. *Methods*, **47**, 198-205.

193. Mikulecky, P.J. and Feig, A.L. (2006) Heat capacity changes associated with nucleic acid folding. *Biopolymers*, **82**, 38-58.

194. Mikulecky, P.J. and Feig, A.L. (2006) Heat capacity changes associated with DNA duplex formation: salt- and sequence-dependent effects. *Biochemistry*, **45**, 604-616.

195. Hughesman, C.B., Turner, R.F. and Haynes, C.A. (2011) Role of the heat capacity change in understanding and modeling melting thermodynamics of

complementary duplexes containing standard and nucleobase-modified LNA. *Biochemistry*, **50**, 5354-5368.

196. Nilshad Salim*, R.L., Rui Zhao, Tuhina Banerjee, Jane Philip, David Rueda‡ and Andrew L Feig. (2011) Thermodynamic and Kinetic Analysis of an RNA Kissing Interaction and Its Resolution into an Extended Duplex. *In prep*.

197. Verheije, M.H., Olsthoorn, R.C., Kroese, M.V., Rottier, P.J. and Meulenberg, J.J. (2002) Kissing interaction between 3' noncoding and coding sequences is essential for porcine arterivirus RNA replication. *J Virol*, **76**, 1521-1526.

198. Guenther, R.H., Sit, T.L., Gracz, H.S., Dolan, M.A., Townsend, H.L., Liu, G., Newman, W.H., Agris, P.F. and Lommel, S.A. (2004) Structural characterization of an intermolecular RNA-RNA interaction involved in the transcription regulation element of a bipartite plant virus. *Nucleic Acids Res*, **32**, 2819-2828.

199. Rakotondrafara, A.M., Polacek, C., Harris, E. and Miller, W.A. (2006) Oscillating kissing stem-loop interactions mediate 5' scanning-dependent translation by a viral 3'-cap-independent translation element. *Rna*, **12**, 1893-1906.

200. Zoll, J., Heus, H.A., van Kuppeveld, F.J. and Melchers, W.J. (2008) The structure-function relationship of the enterovirus 3'-UTR. *Virus Res*.

201. Weixlbaumer, A., Werner, A., Flamm, C., Westhof, E. and Schroeder, R. (2004) Determination of thermodynamic parameters for HIV DIS type loop-loop kissing complexes. *Nucleic Acids Res*, **32**, 5126-5133.

202. Lorenz, C., Piganeau, N. and Schroeder, R. (2006) Stabilities of HIV-1 DIS type RNA loop-loop interactions in vitro and in vivo. *Nucleic Acids Res*, **34**, 334-342.

203. Xia, T., SantaLucia, J., Jr., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C. and Turner, D.H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, **37**, 14719-14735.

204. Freier, S.M., Kierzek, R., Caruthers, M.H., Neilson, T. and Turner, D.H. (1986) Free energy contributions of G.U and other terminal mismatches to helix stability. *Biochemistry*, **25**, 3209-3213.

205. SantaLucia, J., Jr. (2007) Physical principles and visual-OMP software for optimal PCR design. *Methods Mol Biol*, **402**, 3-34.

206. Mikulecky, P.J. and Feig, A.L. (2006) Heat capacity changes associated with nucleic acid folding. *Biopolymers*, **82**, 38-58.

207. Fiore, J.L., Kraemer, B., Koberling, F., Edmann, R. and Nesbitt, D.J. (2009) Enthalpy-driven RNA folding: single-molecule thermodynamics of tetraloop-receptor tertiary interaction. *Biochemistry*, **48**, 2550-2558.

208. Holbrook, J.A., Capp, M.W., Saecker, R.M. and Record, M.T., Jr. (1999) Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: interpretation in terms of coupled processes of formation and association of single-stranded helices. *Biochemistry*, **38**, 8409-8422.

209. Mikulecky, P.J. and Feig, A.L. (2006) Heat capacity changes associated with DNA duplex formation: salt- and sequence-dependent effects. *Biochemistry*, **45**, 604-616.

210. Vander Meulen, K.A., Davis, J.H., Foster, T.R., Record, M.T., Jr. and Butcher, S.E. (2008) Thermodynamics and Folding Pathway of Tetraloop Receptor-Mediated RNA Helical Packing. *J Mol Biol*.

211. Takach, J.C., Mikulecky, P.J. and Feig, A.L. (2004) Salt-dependent heat capacity changes for RNA duplex formation. *J Am Chem Soc*, **126**, 6530-6531.

212. Hughesman, C.B., Turner, R.F. and Haynes, C. (2011) Correcting for Heat Capacity and 5'-TA Type Terminal Nearest Neighbors Improves Prediction of DNA Melting Temperatures Using Nearest-Neighbor Thermodynamic Models. *Biochemistry*, **50**, 2642-2649.

213. Rist, M. and Marino, J. (2001) Association of an RNA kissing complex analyzed using 2-aminopurine fluorescence. *Nucleic Acids Res*, **29**, 2401-2408.

214. Williams, M.C., Rouzina, I., Wenner, J.R., Gorelick, R.J., Musier-Forsyth, K. and Bloomfield, V.A. (2001) Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching. *Proc Natl Acad Sci U S A*, **8**, 8.

215. Cruceanu, M., Gorelick, R.J., Musier-Forsyth, K., Rouzina, I. and Williams, M.C. (2006) Rapid kinetics of protein-nucleic acid interaction is a major component of HIV-1 nucleocapsid protein's nucleic acid chaperone function. *J Mol Biol*, **363**, 867-877.

216. Gottesman, S. (2004) The small RNA regulators of Escherichia coli: roles and mechanisms*. *Annu Rev Microbiol*, **58**, 303-328.

217. Storz, G., Opdyke, J.A. and Zhang, A. (2004) Controlling mRNA stability and translation with small, noncoding RNAs. *Curr Opin Microbiol*, **7**, 140-144.

218. Lee, T. and Feig, A.L. (2008) In Batey, R. T., Walter, N. G. and Woodson, S. A. (eds.), *Non-Protein Coding RNAs*. Elsevier, New York, pp. In Press.

219. Fasman, G.D. (1976) *CRC Handbook of Biochemistry and Molecular Biology*. 3rd ed. CRC Press LLC, Boca Raton, FL.

220. SantaLucia, J., Jr., Allawi, H.T. and Seneviratne, P.A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, **35**, 3555-3562.

221. McDowell, J.A. and Turner, D.H. (1996) Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: Solution structure of r(GAGGUCUC)2 by 2-D NMR and simulated annealing. *Biochemistry*, **35**, 14077-14089.

222. SantaLucia, J., Jr. (2000) In Gore, M. (ed.), *Spectrophotometry and Spectrofluorimetry*. University Press, Oxford, NY, pp. 329-354.

223. Mizoue, L.S. and Tellinghuisen, J. (2004) The role of backlash in the "first injection anomaly" in isothermal titration calorimetry. *Anal Biochem*, **326**, 125-127.

224. Wiseman, T., Williston, S., Brandts, J. and Lin, L. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem*, **179**, 131-135.

225. Skoog, D.A., West, D.M., Holler, F.J. and Crouch, S.R. (2004) *Fundamentals of Analytical Chemistry*. 8th Edition ed. Brooks/Cole, Florence, KY.

226. Wiseman, T., Williston, S., Brandts, J.F. and Lin, L.N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem*, **179**, 131-137.

227. Sigurskjold, B.W. (2000) Exact analysis of competition ligand binding by displacement isothermal titration calorimetry. *Anal Biochem*, **277**, 260-266.

228. Ababou, A. and Ladbury, J.E. (2006) Survey of the year 2004: literature on applications of isothermal titration calorimetry. *J Mol Recognit*, **19**, 79-89.

229. Ababou, A. and Ladbury, J.E. (2007) Survey of the year 2005: literature on applications of isothermal titration calorimetry. *J Mol Recognit*, **20**, 4-14.

230. Buurma, N.J. and Haq, I. (2007) Advances in the analysis of isothermal titration calorimetry data for ligand-DNA interactions. *Methods*, **42**, 162-172.

231. Feig, A.L. (2007) Applications of isothermal titration calorimetry in RNA biochemistry and biophysics. *Biopolymers*, **87**, 293-301.

232. Haq, I. (2002) Thermodynamics of drug-DNA interactions. *Arch Biochem Biophys*, **403**, 1-15.

233. Okhrimenko, O. and Jelesarov, I. (2008) A survey of the year 2006 literature on applications of isothermal titration calorimetry. *J Mol Recognit*, **21**, 1-19.

234. Gilbert, S.D., Mediatore, S.J. and Batey, R.T. (2006) Modified pyrimidines specifically bind the purine riboswitch. *J Am Chem Soc*, **128**, 14214-14215.

235. Gilbert, S.D., Stoddard, C.D., Wise, S.J. and Batey, R.T. (2006) Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *J Mol Biol*, **359**, 754-768.

236. Muller, M., Weigand, J.E., Weichenrieder, O. and Suess, B. (2006) Thermodynamic characterization of an engineered tetracycline-binding riboswitch. *Nucleic Acids Res*, **34**, 2607-2617.

237. Hermann, T. and Westhof, E. (1998) Aminoglycoside binding to the hammerhead ribozyme: a general model for the interaction of cationic antibiotics with RNA. *J Mol Biol*, **276**, 903-912.

238. Lynch, S.R. and Puglisi, J.D. (2001) Structural origins of aminoglycoside specificity for prokaryotic ribosomes. *J Mol Biol*, **306**, 1037-1058.

239. Pilch, D.S., Kaul, M., Barbieri, C.M. and Kerrigan, J.E. (2003) Thermodynamics of aminoglycoside-rRNA recognition. *Biopolymers*, **70**, 58-79.

240. von Ahsen, U., Davies, J. and Schroeder, R. (1991) Antibiotic inhibition of group I ribozyme function. *Nature*, **353**, 368-370.

241. Kim, I., Liu, C.W. and Puglisi, J.D. (2006) Specific recognition of HIV TAR RNA by the dsRNA binding domains (dsRBD1-dsRBD2) of PKR. *J Mol Biol*, **358**, 430-442.

242. Sledjeski, D.D., Gupta, A. and Gottesman, S. (1996) The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in Escherichia coli. *EMBO J*, **15**, 3993-4000.

243. Recht, M.I. and Williamson, J.R. (2004) RNA tertiary structure and cooperative assembly of a large ribonucleoprotein complex. *J Mol Biol*, **344**, 395-407.

244. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, **31**, 3406-3415.

245. Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol*, **288**, 911-940.

246. Holbrook, J.A., Capp, M.W., Saecker, R.M. and Record, M.T., Jr. (1999) Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: interpretation in terms of coupled processes of formation and association of single-stranded helices. *Biochemistry*, **38**, 8409-8422.

247. Takach, J.C., Mikulecky, P.J. and Feig, A.L. (2004) Salt-dependent heat capacity changes for RNA duplex formation. *J Am Chem Soc*, **126**, 6530-6531.

248. Diamond, J.M., Turner, D.H. and Mathews, D.H. (2001) Thermodynamics of three-way multibranch loops in RNA. *Biochemistry*, **40**, 6971-6981.

249. Mikulecky, P.J., Takach, J.C. and Feig, A.L. (2004) Entropy-driven folding of an RNA helical junction: an isothermal titration calorimetric analysis of the hammerhead ribozyme. *Biochemistry*, **43**, 5870-5881.

250. McPike, M.P., Sullivan, J.M., Goodisman, J. and Dabrowiak, J.C. (2002) Footprinting, circular dichroism and UV melting studies on neomycin B binding to the packaging region of human immunodeficiency virus type-1 RNA. *Nucleic Acids Res*, **30**, 2825-2831.

251. Wickiser, J.K., Cheah, M.T., Breaker, R.R. and Crothers, D.M. (2005) The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry*, **44**, 13404-13414.

252. Indyk, L. and Fisher, H.F. (1998) Theoretical aspects of isothermal titration calorimetry. *Methods Enzymol*, **295**, 350-364.

253. Tellinghuisen, J. (2005) Statistical error in isothermal titration calorimetry: variance function estimation from generalized least squares. *Anal Biochem*, **343**, 106-115.

254. Barbieri, C.M. and Pilch, D.S. (2006) Complete thermodynamic characterization of the multiple protonation equilibria of the aminoglycoside antibiotic paromomycin: a calorimetric and natural abundance 15N NMR study. *Biophys J*, **90**, 1338-1349.

255. Kaul, M., Barbieri, C.M., Srinivasan, A.R. and Pilch, D.S. (2007) Molecular determinants of antibiotic recognition and resistance by aminoglycoside phosphotransferase (3')-IIIa: a calorimetric and mutational analysis. *J Mol Biol*, **369**, 142-156.

256. Zhang, Y., Akilesh, S. and Wilcox, D.E. (2000) Isothermal titration calorimetry measurements of Ni(II) and Cu(II) binding to His, GlyGlyHis, HisGlyHis, and bovine serum albumin: a critical evaluation. *Inorg Chem*, **39**, 3057-3064.

257. Tellinghuisen, J. (2005) Optimizing experimental parameters in isothermal titration calorimetry. *J Phys Chem B*, **109**, 20027-20035.

258. Tellinghuisen, J. (2006) Van't Hoff analysis of K degrees (T): how good...or bad? *Biophys Chem*, **120**, 114-120.

259. Markova, N. and Hallen, D. (2004) The development of a continuous isothermal titration calorimetric method for equilibrium studies. *Anal Biochem*, **331**, 77-88.

ABSTRACT

MODES AND MECHANISMS OF HFQ MEDIATED STRESS REGULATION IN BACTERIA

by

NILSHAD N. SALIM

May 2012

Advisor: Professor Andrew Feig

Major: Chemistry (biochemistry)

Degree: Doctor of Philosophy

To survive bacteria must be able to respond to its ever-changing environmental conditions. sRNAs have been implicated in a variety of stress-response pathways that help bacterial systems modulate gene expression. The RNA binding protein Hfq facilities this process by, helping sRNA to base pair with its target mRNAs to initiate gene regulation. A common feature of Hfq-mediated gene regulation is the network-based organization where a single sRNA can control multiple messages to promote integrated response to stress. Current mechanistic models that are present to describe Hfq functions cannot explain the complexity at which Hfq performs gene regulation. In this work we have used a variety of biophysical, biochemical and biological approaches to understand the nature of Hfq interactions with target mRNAs.

AUTOBIOGRAPHICAL STATEMENT

NILSHAD NILAM SALIM

Education:

B.Sc. Chemistry,
University of Colombo, Colombo, Sri Lanka

Ph.D. Chemistry - Major in Biochemistry
Wayne State University, Detroit, Michigan, USA

Publications:

1. Salim, N N., Faner M, and Feig, A. L. (2011) Requirement of Upstream Hfq Binding $(ARN)_x$ Elements in *glmS* and the Hfq C-Terminal Region for GlmS Up-regulation by sRNAs GlmZ and GlmY, *to be submitted to Nucleic acids research*.

2. Salim, N N., Banerjee, T., Lamichhane, R., Zhao, R., Philip, J., Rueda, D and Feig, A. L. (2011) Thermodynamic and Kinetic Analysis of RNA Kissing Interactions and Its Resolution into Extended Duplexes, *Accepted* to the *Biophysical Journal*

3. Salim, N. N., and Feig, A. L. (2010) An upstream Hfq binding site in the fhlA mRNA leader region facilitates the OxyS-fhlA interaction, *PLoS One 5*. (DOI-10.1371/journal.pone.0013028)

4. Salim, N. N., and Feig, A. L. (2009) Isothermal titration calorimetry of RNA, *Methods 47*, 198-205 (DOI-S1046-2023(08) 00151-5 [pii]10.1016/ j.ymeth .2008.09.003)

Presentations:

(Oral)

1. Salim, N. N., Feig, A. L.; "An Upstream Hfq Binding Site in the *fhlA* mRNA Leader Region Facilitates the OxyS-*fhlA* Interaction" The 14th Annual Meeting of the RNA Society, Madison, WI, May 2009 – oral presentation

(Poster)

1. Salim, N N; Banerjee, T.; Lamichhane, R.; Zhao, R.; Philip, J.; Rueda, D and Feig, A. L.; "Energy Landscapes of RNA Kissing Complexes" The 14th Annual Meeting of the RNA Society, Madison, WI, May 2009 – poster presentation

2. Salim, N N; Banerjee, T.; Lamichhane, R.; Zhao, R.; Philip, J.; Rueda, D and Feig, A. L.; "Energy Landscapes of RNA Kissing Complexes" The Michigan RNA Society Meeting, Albion College, Albion, MI, April 2009 – poster presentation