Biochemical & Biophysical Studies of an Hfq Homolog from a Deep-branching Bacterium

Jennifer Patterson West
East Stroudsburg, PA

B.S. Chemistry, Shippensburg University, 2009
B.S. Biology, Shippensburg University, 2009

A thesis presented to the Graduate Faculty of the University of Virginia in candidacy for the Degree of Doctor of Philosophy

Department of Chemistry

University of Virginia
May, 2014

To all those who came before me;
Thank you for paving the road.

Acknowledgements

Throughout my graduate studies there have been numerous people who have supported my research, both directly and indirectly. First and foremost, I would like to thank my graduate advisor, Cameron Mura, for his tutelage during my scientific development. His guidance, accessibility, and expertise were essential to my progress and his willingness to listen to new ideas allowed me to grow more than I ever imagined. I would also like the members of my dissertation committee – Robert Bryant, Linda Columbus, David Cafiso, Christopher Stroupe, and Melissa Kendall.

I would like to thank Peter Randolph, a fellow lab member, who assisted in developing experiments and provided technical support. Our many discussions often cemented key concepts and pinpointed the right directions to pursue. I would also like to thank all the members of the Mura laboratory, members of the Columbus laboratory, Dr. Linda Columbus, and Dr. Carol Price for their technical assistance and feedback on manuscripts, posters, and oral presentations. Additionally, I would like to thank the two undergraduates, Collin Conrad and Ellen Shleckman, who I had the opportunity to mentor during my doctoral studies; it was a privilege to see them grow as scientists.

I would also like to thank my family for their support throughout my academic career. My parents taught me to never give up on the things that matter in life, even when nothing is working in your favor. I would also like to thank my husband, Kenneth West, for his constant support, and for being my patient outlet.

Chapter 2 of this dissertation is an adapted pre-print of the following manuscript: *Discovery & Identification of Hfq-binding nanoRNAs in Escherichia coli* by Patterson, Cao, Fox, Columbus, Clardy & Mura. The following individuals were helpful in this work: Christina Puckett (JCSG) provided genetic material; Drs. Nicholas Sherman, Erin Jeffery, and John Shannon (all of

Appendix 1 of this dissertation is adapted from the following article: *Rapid Colorimetric Assays to Qualitatively Distinguish RNA and DNA in Biomolecular Samples* by Patterson & Mura *J. Vis. Exp.* (2013), 72, 1-10. I thank Linda Columbus, Kanishk Jain, and Peter Randolph for helpful discussions and critical reading of that manuscript.

Appendix 2 of this dissertation is a summary of the immunochemical methods used – namely, the purification and characterization studies that were performed on the polyclonal antibodies used in other parts of this dissertation. Initial training and background knowledge for

the optimization of immunoassays was provided by Antonia Silva and the *BioTrac-1:*

Biotechnology techniques course at the National Institute of Health (September 2012).

Abstract

RNA-based regulation enables exquisite control over the extent and timing of gene expression, enabling bacteria to rapidly respond to their environment. The bacterial host factor Hfq is involved in regulation of gene expression via its role as an RNA chaperone; Hfq binds to sRNAs and regulates their stability, and also interacts with sRNA and mRNA to facilitate their annealing. Hfq has been shown to preferentially bind adenosine– and uracil–rich RNAs. Here, we discovered that the Hfq from *Thermotoga maritima* (*Tma*) interacts with short uracil/cytosine-rich nanoRNAs when recombinantly expressed in *Escherichia coli*. These Hfq-binding nanoRNAs interact with nanomolar affinities and feature 5'-monophosphate and 3'-hydroxyl end chemistries.

Previous work by others revealed the functional form of Hfq in *Escherichia coli* and other bacterial species is a homohexamer. Biophysical and biochemical characterization of *Tma* Hfq supports the formation of both homohexamers and homododecamers. Isothermal calorimetry and semi-native Western blot analysis indicate that the hexameric and dodecameric states are in equilibrium. Fluorescence polarization assays show that both oligomeric states bind to adenosine- and uracil-rich RNAs with nanomolar affinities. To elucidate the atomic basis of *Tma* Hfq function, we determined the crystal structure of this protein to 2.1 Å resolution, which further demonstrated that *Tma* Hfq forms a homohexamer.

I also describe initial *in vivo* studies of the expression levels, cellular localization, and RNA binding partners of Hfq in *Tma*. Polyclonal antibodies to *Tma* Hfq were propagated, purified and characterized for downstream enrichment and detection assays. A quantitative Western blot was developed to determine the concentration of *Tma* Hfq in cell lysate; the concentrations determined via this assay, along with cell counts determined by flow cytometry, are being used

to determine the expression level of *Tma* Hfq at different phases in the growth of this bacterium. Initial images of immunogold-labeled *Tma* cells have been collected by electron microscopy to identify and localize Hfq *in vivo*. Most recently, co-immunoprecipitation assays are being pursued to isolate Hfq along with its native binding partners in *Tma*; downstream analyses of these data include identification of protein and RNA binding partners via mass spectrometry-based proteomics and next-generation sequencing.

.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction to the Hfq protein and its role in RNA

regulation

<u>**Overview**</u>

The central dogma of molecular biology defines the flow of information in living systems: genetic information is stored as DNA, which is transcribed into RNA and then translated into protein. This classical perspective [1] is a highly simplified model of the information networks actually present *in vivo*, which also include non-coding RNAs and numerous regulators, ranging from small molecules to RNAs to proteins. These complex networks enable the cell to finely control gene expression, which is essential for proper cellular function.

Proteins are generally the functional molecules in the cell and they have key roles in most cellular processes, including catalyzing metabolic reactions, mediating chemical signals, and synthesizing DNA. The ability of a cell to rapidly modify its protein expression patterns under various environmental conditions is critical for survival. Regulatory pathways can be triggered by many external factors including nutrients, temperature, and pH (Figure 1.1a). These pathways can cause either up-regulation or down-regulation of a specific protein through various mechanisms, including transcription initiation through a promoter (e.g. lac promoter [2]), targeting a protein for degradation (polyadenylation[3]), or inhibiting translation by blocking the ribosome binding site (small RNA (sRNA) mediated regulation [4,5,6,7]).

The past decade has revealed that regulatory pathways in bacteria which control protein expression often work at the post-transcriptional level via numerous non-coding sRNAs (Figure 1.1b-d) [7]. These bacterial sRNA regulatory pathways can be viewed as roughly analogous to the eukaryotic microRNA systems [8,9]. The sequences of these regulatory sRNAs are antisense to their target RNA and are either completely complementary (*cis*-encoded) or otherwise imperfect complements (*trans*-encoded) [7,10,11]. In order for *trans*-encoded sRNAs to anneal with their target mRNA and thereby achieve their physiological functions, an RNA chaperone known as "Hfq" is frequently required [12] (Figure 1.1d).

Hfq: A facultative global regulator

In the early 1970s, Hfq was discovered as a <u>h</u>ost <u>f</u>actor necessary for the replication of an *Escherichia coli* RNA virus known as 'Qβ' [13]. Since then, a gene encoding the Hfq protein has been bioinformatically identified in the genomes of approximately 50% of sequenced Gram-positive and Gram-negative bacteria [14]. Most bacteria encode a single copy of the *hfq* gene, but in recent years a few species have been identified that contain two distinct *hfq* genes. For instance, members of the *Burkholderia cepacia complex* encode two *hfq* genes that are differentially expressed. Hfq1 is a typical bacterial Hfq and is maximally expressed during exponential growth, whereas Hfq2 has an extended sequence and is maximally expressed during stationary phase; both genes are required for bacterial virulence [15,16]. The bacterium *Novosphingobium aromaticivorans* encodes a single protein that contains two Hfq sequences separated by twenty-three residues [17]. Any physiological or structural implication of the linkage between the two Hfq monomers in this bacterium remains unknown.

It was not until the 1990's that the physiological role of Hfq in gene expression began to be understood. The pleiotropic effects of a Hfq knockout mutant were first observed in *E. coli*, which exhibited decreased growth rates and increased susceptibility to environmental challenges, including ultraviolet light and oxidative stress [18]. The role of Hfq as a global transcriptional regulator has since been demonstrated for numerous Gram-negative and Gram-positive species, including *Borrelia burgdorferi* [19], *Salmonella enterica* [20,21]*, Agrobacterium tumefaciens* [22], and *Vibrio cholerae* [23]. These studies frequently identified Hfq related RNAs and pathways via genome wide approaches, such as deep sequencing (RNA-seq) [20,24] and microarray analysis [25]. The multiple phenotypic effects observed for these and numerous other Hfq mutants have suggested that Hfq is a central hub in many cellular pathways (Figure 1.2), including quorum sensing [26], bacterial virulence [5], and biofilm formation [27].

Despite the role of Hfq as a global translational regulator in Υ-proteobacteria, Hfq's role in some species is quite limited or even absent [28]. Thus far no Hfq homolog has been identified in any bacteria belonging to the class ε-proteobacteria or the order actinomycetales [29]. The absence of an Hfq homolog in these bacteria may be the result of limited sequence similarities (as observed for cyanobacteria [30]), the presence of a complementary protein, or loss of the Hfq protein due to dispensable function [29]. In some bacteria, Hfq is required for only a subset of the *trans*-encoded sRNAs. For instance, *V. cholerae* requires Hfq to facilitate sRNA-mediated degradation of *hapR* mRNA by the Qrr1-4 sRNAs, which is essential for proper gene expression at high cell density [5], whereas Hfq is dispensable for the down-regulation of *ompA* mRNA by the sRNA VrrA [31]. Hfq's limited role in these bacteria may indicate that the intrinsic thermodynamic barrier for formation of the sRNA-mRNA complex is lower in these bacteria or that another RNA-binding protein facilitates the interaction in the absence of Hfq [28,29]. The former explanation appears to be true for a subset of these Hfq-independent sRNA-mRNA complexes, as they have been shown to form in the absence of Hfq *in vitro*, without the addition of any complementary protein. For instance, RliB, RliE, and RliI sRNAs from *Listeria monocytogenes* have been shown to form a functional complex with their target mRNAs in the absence of Hfq *in vitro* [32]. This variability in the requirement of Hfq for sRNA-mediated translation by *trans*-encoded sRNA suggests that the role of Hfq is not completely passive in those complexes that are Hfq-dependent.

Expression and localization of Hfq

In *E. coli*, Hfq expression has been shown to depend on growth phase, with approximately 55,000 and 30,000 monomers expressed per cell during exponential phase and stationary phase, respectively [33,34]. Hfq has also been identified as one of the twelve most abundant proteins

in the nucleoid, which is where approximately 10-20% of cellular Hfq is localized [33,35]. Cellular imaging by electron microscopy (EM) indicates that Hfq with a C-terminal metallothionein tag localizes near the membrane of the cell in addition to the cytoplasm and nucleoid [36]. The portion of Hfq localized proximal to the membrane was estimated to be less than 50% [36].

The expression level of Hfq protein can be quite variable between different bacterial species and even between different strains of the same species. For instance, contradictory findings regarding the phenotypic alterations in a Hfq mutant of *Staphylococcus aureus* may be the result of variability in Hfq expression between the strains of *S. aureus* that were used in the two studies [37]. Bohn *et al*. observed no phenotypic alterations in Hfq mutants of *S. aureus* RN6390 and *S. aureus* COL, whereas Liu *et al.* observed phenotypic alterations, including increased pigmentation and decreased pathogenicity in a Hfq mutant of *S. aureus* 8325-4 [37,38]. The presence of Hfq protein was detected by Western blots for eight strains of *S. aureus*, including 8325-4, but was not detected in strains RN6390 and COL [37]. The level of *hfq* transcript did not vary between the different strains, as found by quantitative real-time polymerase extension reaction, suggesting post-transcriptional regulation is responsible for the variation in Hfq protein levels between the different strains of *S. aureus* [37].

Evidence for post-transcriptional regulation of the *hfq* mRNA has been found in other bacteria as well, including *E. coli* [39,40] and *Sinorhizobium meliloti* strain 2011 [41]. In *E. coli*, two distinct Hfq binding sites were identified in the 5' untranslated region (UTR) of the *hfq* transcript. One of these Hfq binding sites contains the ribosome binding site (RBS), thereby preventing translation of the mRNA upon Hfq binding [39]. Additionally, *hfq* transcript has been co-immunoprecipitated with Hfq protein in *Rhodobacter sphaeroides* [42] and *Thermotoga maritima* (Patterson & Mura, unpublished). Together these findings suggest that translation of

hfq transcripts is subject to an autoregulatory mechanism. One proposal is that maintaining a constant intracellular level of Hfq will allow sRNA concentrations -- not Hfq levels -- to dictate cellular riboregulatory processes [12,29]. Such a model is also plausible considering the lifetimes of sRNA (short) versus protein (long) species in the cell.

Alternatively, both the ribonuclease RNase E and CsrA (a protein involved in carbon storage regulation [43]) have been shown to post-transcriptionally regulate the level of Hfq in *E. coli* [18,44]. Regulation of Hfq levels by RNase E (encoded by the *rne* gene) was suggested by a three-fold increase in Hfq concentration in an *rne* mutant compared to the wild type strain [18]. CsrA has been shown to bind to a single site within the 5' UTR of the *hfq* transcript thereby competing with the ribosomal 30S subunit and preventing formation of the translation initiation complex [44]. Both of these examples illustrate how other RNA-related pathways may regulate intracellular levels of Hfq.

In *E. coli*, Hfq is located in the superoperon *amiB-mutL-miaA-hfq-hflX-hflK-hflC,* which contains several σ^{32}- and σ^{70}-dependent promoters [45]. The protein σ^{70} is the general housekeeping sigma factor, whereas σ^{32} is the heat shock sigma factor. The presence of a σ^{32}-dependent promoter upstream of the *hfq* gene may act to maintain constant *hfq* transcript levels under heat-shock conditions (50°C), since transcription from σ^{70}-dependent promoters will decrease under such conditions [45]. Hfq is located in a similar gene cluster in the *Salmonella* genome [21], but this gene cluster is only partially conserved, or completely divergent, in other sequenced bacteria. For instance, in *Neisseria* strains (except *N. gonorrhoeae* FA1090) the organization of the gene cluster encoding *hfq* is partially conserved but, unlike *E. coli*, the penicillin-binding protein 7 and a hypothetical ribosomal RNA (rRNA) methylase are encoded upstream and downstream of the *hfq* gene, respectively.

Hfq is an Sm protein: Structural features and implications for function

Since protein function arises from three-dimensional (3D) structure, we need to understand the various structural features of a protein to fully understand its physiological role. A milestone in our knowledge of Hfq structure was reached in 2002, when structural, biophysical, and bioinformatics analyses revealed that Hfq is the bacterial branch of the Sm protein family [17,46,47,48,49]. Specifically, these studies revealed that each monomer adopts an Sm fold, consisting of an N-terminal α-helix followed by five β-strands (Figure 1.3a) arranged in an antiparallel β5-β1-β2-β3-β4 topology [48]. Additionally, these studies indicated that Hfq forms a hexameric toroid structure with six-fold symmetry (Figure 1.3b) [46].   Strands β4 and β5 are the most conserved among bacterial Hfq sequences (Figure 1.4) and form the interface between adjacent subunits in the oligomer [50,51]. As with other Sm rings, the hydrogen bonds and other interatomic contacts between strands β4 and β5 stabilize the oligomer [52]. All N-terminal α-helices appear on one face of the hexamer, commonly referred to as the "proximal" face.

The β-rich, barrel-like Sm fold characterizes proteins belonging to the Sm/Sm-like (LSm) superfamily, which contains RNA-binding proteins across all domains of life (Bacteria, Archaea, Eukaryotes) [53]. Sm proteins are comprised of two conserved sequence motifs, referred to as Sm1 and Sm2, which contain the first three β-strands and the last two β-strands, respectively. The amino acid sequence of the Sm1 motif is well conserved across all three domains of life, whereas the Sm2 motif is divergent in bacteria. This divergence accounts for some of the initial difficulties in assigning Hfq to the Sm/LSm protein superfamily [14]. Despite low sequence conservation across the different domains of life, the characteristic Sm fold is structurally conserved, with a root mean squared deviation ranging from 0.85 Å to 1.3 Å when comparing bacterial Hfq to archaeal Sm and human Sm protein monomers [50].

Despite the overall conservation of the Sm fold, there are key structural features that distinguish different Sm proteins. The L4 loop, linking strands β3 and β4, is significantly shortened in the bacterial Hfq proteins, consisting of only a few residues in Hfq (3 residues in *E. coli*) compared to the extended L4 loop of eukaryotic Sm proteins (e.g., 11 residues in Human LSm8). The L4 loop is located on the "distal" face of Sm oligomers and contributes to differences in the surface electrostatic potential. For instance, the electrostatic potential of the distal face of an archaeal Hfq from *Methanocaldoccocus jannaschii* is predominantly negative [54], but is positive for the Hfq from *E. coli* [50]. Differences in electrostatic potential between various Sm proteins are illustrated in Figure 1.5. Because RNA is a polyanion, these differences would be expected to contribute to the different RNA-binding profiles of different Hfq rings. For instance, the negatively charged distal face of *M. jannaschii* Hfq would not be expected to interact favorably with RNA, whereas the distal face of *E. coli* Hfq has been shown to have nanomolar affinity for adenosine-rich (A-rich) RNAs [55].

In addition to key differences in the loops that link secondary structural elements, the quaternary structures of Sm proteins vary greatly between the different domains of life. In bacteria, which typically encode a single *hfq* gene, the Hfq protein forms homohexamers (Figure 1.6a). Archaea encode one to three distinct Sm-like archaeal proteins (SmAPs) and form homo-oligomers, which vary in number of subunits (Figure 1.6b). Specifically, SmAPs have been shown to form hexamers [56], heptamers [56,57,58,59,60], 14-mers [61], and octamers (Randolph & Mura, personal communication). Eukaryotes encode more than 20 distinct Sm paralogs (Figure 1.6c) [62]. Eukaryotic Sm proteins are subdivided into Sm and LSm subfamilies [63,64]. Heteroheptamers are formed by either seven distinct members of one of these subfamilies or a mixture of seven distinct members from both subfamilies. Heteroheptamers containing only members of the LSm subfamily can spontaneously oligomerize in the absence of RNA, whereas

rings containing the canonical small nuclear RNA (snRNA) Sm proteins require the presence of RNA, and potentially auxiliary proteins such as the survival of motor neurons (SMN) complex, to assemble [63,65].

Together, these structural differences result in different cellular functions of Sm proteins in the different domains of life. While Sm proteins from all three domains interact with RNA, the RNA-binding sequence specificity seems to be quite variable. The bacterial Hfq interacts with both A-rich mRNAs and uridine-rich (U-rich) sRNAs on opposite faces of the disc-shaped Hfq hexamer, referred to as the "distal" and "proximal" face, respectively [66]. The specific population of RNAs bound to Hfq can vary significantly between different homologs [12,28,67]. In terms of physiological effects, the Hfq protein decreases the rate of RNA turnover of bound RNAs and facilitates interactions between sRNAs and their target mRNAs.

In contrast, eukaryotic Sm proteins bind to snRNA during assembly of the Sm ring, which forms around the RNA (Figure 1.6c). The Sm rings serves as a scaffold [65] for other proteins to associate with in order to form the large, multi-component small nuclear ribonucleoproteins (snRNP) particles involved in RNA splicing [68,69,70,71]. The eukaryotic LSm subfamily bind snRNA, P RNA (RNA component of RNase P), ribosomal RNA, and transfer RNA (tRNA) and are associated with assembly, processing, nuclear localization, 3' and 5' end-processing, and splicing [62,72]. Unlike the eukarya and bacteria Sm paralogs, the RNA-binding properties of SmAPs remain largely uncharacterized. Thus far, work has suggested that archaeal Sm proteins have a role in RNA processing and possibly in RNA-RNA contacts. Structural data of *Archaeoglobus fulgidus* Sm1 with U_5 RNA indicated that it binds RNA similar to *S. aureus* Hfq [59]. Similarly, structural data of *Methanobacterium termautotrophicum* SmAP1 bound to uridine-5'-monophosphate showed conservation of the uridine-binding site [57]. Additionally, co-immunoprecipitation of the Sm protein from the archaeon *Haloferax volcanii* identified more

than 15 different sRNAs and 30 different proteins; most of these binding partners are associated with RNA metabolism and translational regulation, suggesting a similar role in post-transcriptional regulation as observed for bacterial Hfq [73,74].

Riboregulation: The role of Hfq in facilitating sRNA-mRNA interactions

Although Hfq is required for only in a subset of sRNA-mRNA complexes, many of these complexes are essential for bacterial stress response pathways. This fact underlines the important role of Hfq in facilitating phenotypic changes to environmental signals. The specific features of the sRNA-mRNA binding process that determine the extent of Hfq-dependence include the level of transcript expression, the extent of base pairing complementarity between the sRNA and mRNA target, and the secondary structural elements within the two RNA species [28,75,76].

Translational regulation by sRNAs of the *rpoS* mRNA, which encodes sigma S (σ^s), a central regulator of general bacterial stress response and stationary phase genes, is one of the most highly studied Hfq-dependent processes in *E. coli*. The three sRNAs involved in regulating *rpoS* modulate the level of translation in response to different environmental signals: DsrA, RprA, and OxyS are involved in temperature, envelope, and oxidative stress response, respectively. The promoter sequences of these sRNAs are what dictate the environmental cue responsible for up-regulating their transcription. RprA and OxyS are regulated by the *trans*-acting factors (Figure 1.1b) RcsC/RcsB and OxyR, respectively, whereas up-regulation of DsrA is the result of temperature dependent RNA polymerase-DNA interactions [77,78,79]. Although these sRNAs have the same mRNA target, their effects on translation vary; DsrA and OxyS down-regulate translation of *rpoS* by occluding the ribosome-binding site (Figure 1.7a); in contrast, RprA up-regulates translation by disrupting an otherwise inhibitory stem-loop in the 5' leader

sequence of *rpoS*, thereby releasing the ribosome binding site (Figure 1.7b). The physiological roles of other Hfq-binding sRNAs were recently reviewed and cataloged in Gottesman & Storz (2011) [6].

As alluded to above, Hfq's role in sRNA-mediated translation regulation is to facilitate the annealing between an sRNA and its target mRNA. Several models have been proposed, at various levels of detail to explain Hfq-facilitated RNA annealing. One model suggests that Hfq simultaneously binds sRNA and its target mRNA to form a ternary complex that bridges the two RNAs [66]. This mechanism is supported by the detection of ternary DsrA-Hfq-*rpoS* complexes in an electrophoretic mobility shift assays (EMSA) [66]. The crystal structure of $AU_6A\bullet Hfq\bullet A_7$ (PDB 4HT8) further supports that Hfq can form a ternary complex that bridges U-rich and A-rich RNA segments [80]. Additionally, solution nuclear magnetic resonance spectroscopy (NMR) of Hfq with AU_6A and AACAACAAG indicates that these RNA fragments can bind simultaneously to Hfq [80]. A ternary complex is further supported by mass spectrometry (MS) analysis of crosslinked complexes consisting of DsrA domain II, A_{18}, and Hfq, which was consistent with a 1:1:1 stoichiometry complex [81].

A second model suggests that RNA binds to two separate Hfq hexamers and interaction between the two hexamers facilitates annealing [67]. This model is partially supported by the crystal structure of AU_6A with Hfq, which revealed the AU_6A bound to the proximal face of one Hfq with the 5' adenosine docked to the distal face of another Hfq ring [82]. Hypothetically, interactions between Hfq hexamers could facilitate the annealing of an mRNA-sRNA duplex by providing a favorable orientation for the two RNAs to interact.

A third model suggests that Hfq facilitates annealing by altering the secondary structure of one of the RNA transcripts, thus exposing nucleotides in a region previously occluded. This model is supported by RNase footprinting studies of *sodB* mRNA, which indicate that Hfq melts a

double-stranded region of *sodB*, resulting in a larger terminal loop that is predicted to contain a region complementary to RyhB sRNA [83]. This model is further supported by thermal denaturation studies of segments of DsrA RNA in the presence and absence of Hfq, which reveal that Hfq destabilizes the two hairpin duplex regions of DsrA [76]. Small-angle X-ray scattering, small-angle neutron scattering, and circular dichroism experiments of Hfq with RprA and OxyS collectively indicated a substantial change in the structure of the sRNAs upon complex formation, whereas the structure of Hfq remains unchanged [84]. The data were consistent with both RprA and OxyS losing double-stranded structural content upon binding to Hfq [84].

Finally, a fourth model suggests that Hfq binds to both sRNA and mRNA on opposite faces of the toroid, while additional binding sites on the lateral surface that forms the rim of the disc help orient the sRNA and mRNA for favorable annealing (Figure 1.8a). Binding studies of RhyB sRNA with 2'-3' cyclic phosphate end chemistry (RhyB-cP) revealed that sRNA can bind to an additional surface of Hfq (this end chemistry prevents binding to the proximal surface) [52]. Through a series of systematic mutations and binding studies, it was determined that RhyB-cP binds to a positively charged region on the lateral surface [52]. Gel filtration chromatography studies revealed that the Hfq hexamer can bind RhyB-cP and A_{20} RNA simultaneously indicating that an sRNA bound to the lateral face would not necessarily occlude Hfq from binding A-rich RNA on its distal face [52]. Collectively, these findings support a binding model where in the lateral surface orients the sRNA and mRNA for favorable annealing [85]. Additional support for this model was provided by Förster resonance energy transfer (FRET) and fluorescence anisotropy (FA) experiments, which revealed that complementary RNAs do not dissociate from a Hfq-mRNA-sRNA ternary complex as a duplex in the absence of this arginine-rich lateral surface [86]. Thus, the lateral rim is important for proper binding and unbinding of RNA. Furthermore, sequence alignment of 384 Hfq homologs revealed that arginines located in the lateral surface

are highly conserved [86].   Fluorescence quenching studies revealed that the lateral surface does not display any nucleotide sequence specificity, suggesting that it can interact with both mRNAs and sRNAs to orient them for annealing [86].

Notably, the models described above do not elucidate a key discrepancy between *in vitro* and *in vivo* studies of Hfq, which is the slow dissociation rate (>100 minutes) of the Hfq-RNA complex found *in vitro* versus the fast phenotypic changes (1-2 minutes) that occur *in vivo* [87,88,89]. An alternative mechanism resolving this discrepancy proposes that RNAs bound to Hfq can be actively displaced by a competitor RNA, thereby resulting in less stable ribonucleoproteins with faster dissociation rates [89].   Filter binding assays reveal that competitor RNA promotes the dissociation of Hfq-bound RNA in a concentration dependent manner, with the observed half-lives of the RNA-Hfq complex decreasing from approximately 150 minutes to 1 minute in the presence of submicromolar concentrations of competitor RNA [89]. Another study showed that nine distinct sRNAs are able to associate with *E. coli* Hfq and compete for binding [90]. The dissociation rates of all sRNAs were found to be faster in the presence of a competitor RNA; intriguingly, the performance of an sRNA as a competitor was dependent on the identity of the two RNAs, despite similar binding sites and affinities for various RNAs to Hfq [90].   This RNA displacement mechanism is also in agreement with a mathematical model that incorporates an RNA network consisting of ternary complexes, which are either productive (leading to duplex formation) or non-productive, into the model of Hfq-sRNA duplex formation [91]. In this cellular-scale mathematical model, RNA duplex formation is strongly influenced by the association kinetics and abundance of competing RNAs, as well as Hfq [91].

Hfq's association with DNA

Despite Hfq's association with mRNA and sRNA dominating the literature, Hfq has also been reported to bind other nucleic acids, including DNA [35,92,93,94] and tRNA [95]. In fact, in the early 1990's it was already recognized that approximately 10-20% of Hfq localizes to the bacterial nucleoid, suggesting a potential role for Hfq in the regulation of DNA activity or DNA topology [33]. It is worth reiterating that a proteomic analysis of the *E. coli* nucleoid found that Hfq is one of the twelve most predominant nucleoid-associated proteins [35].

Gel-shift assays revealed that Hfq preferentially binds curved DNA compared to non-curved DNA with a dissociation constant (K_d) of 128 nM and 250 nM, respectively [35]. The archaeal homolog of Hfq, SmAP, has also been shown to bind supercoiled DNA [57]. Taking into account Hfq's ability to bind both RNA and DNA, it was proposed that Hfq may play a role in transcription-translation coupling [35]. In addition to direct association of Hfq with DNA, Hfq has been shown to associate with S1-core RNA polymerase (RNAP) via direct interactions with the ribosomal protein S1 [96]. *In vitro* transcription reactions revealed that Hfq has a mild stimulatory effect on S1-core RNAP, resulting in an increase in the overall yield of RNA transcripts [96]. It is unclear if this stimulatory effect is due to Hfq-protein interactions or Hfq's interaction with either DNA or RNA. To further examine Hfq's potential role in transcription-translation coupling, Sukhodolets and Garges assayed luciferase production under varying conditions using the *E. coli* S30 extract system (e.g. an *in vitro* transcriptional/translational system) for circular DNA (e.g. plasmid) [96]. Briefly, it was observed that S30 extract had a significantly reduced ability to catalyze multiple rounds of translation in a Hfq-depleted sample [96]. However, S30 extract from a *hfq*⁻ deletion mutant produced full-length transcript more efficiently at earlier time points (albeit with a lower overall yield compared to *hfq*⁺ strains) [96].

The identification of genomic DNA fragments co-purifying with Hfq provides further support of a direct interaction between Hfq and DNA [93]. Sequencing data revealed twenty-four different DNA sequences that co-purified with Hfq expressed in *E. coli;* twenty of these originated from the *E. coli* genome and four from the expression plasmid used in the study [93]. Interestingly, thirteen of the identified DNA sequences are predicted to have helical curvature [93], which is in agreement with earlier findings that Hfq preferentially binds curved DNA [35]. Despite previous reports that Hfq binds DNA without sequence specificity, 65% of the *E. coli* DNA sequences were identified as encoding transport proteins or membrane proteins; this observation is unlikely to be due to chance alone based on the representation of membrane proteins in the *E. coli* genome (11.2%) [93]. Further analysis of the identified sequences revealed one to three copies of a consensus sequence, (A/T)T(A/G)TGCCG, in 80% of the *E. coli* Hfq-binding DNA sequences [93]. Systematic mutagenesis of Hfq revealed that the distal face and C-terminal domain affect DNA binding [93]. Competion assays further support the involvement of the distal face in DNA binding, as both A_{18} RNA and RprA (both bind the distal face) were shown to compete with DNA, whereas DsrA (which binds the proximal face) did not influence binding.

Fluorescence anisotropy experiments of Hfq with the DNA probes dA_{20}, a dA_{20}-dT_{20} duplex, and dC_{20} indicated dissociation constants of 180 ± 22 nM, 250 ± 14 nM, and 2.5 ± 0.5 μM, respectively [94]. Hfq's preference for A-rich DNA sequences notwithstanding, EM images of the pBR322 plasmid in the presence of Hfq revealed that Hfq covered and highly compacted it with respect to naked plasmid [94]. Fourier transform infrared spectroscopy of these systems indicated that Hfq partially opens the dA_{20}-dT_{20} duplex and binds to the dA_{20} strand, stabilizing the *N*-type ring pucker geometry, which has been shown to co-exist with the *S*-type sugar pucker geometry in AT-rich sequences of free DNA [94]. Despite the collective evidence from

multiple *in vitro* experiments for a link between Hfq and DNA functions or DNA topology, *in vivo* experimental evidence is unavailable to clarify this issue [29].

Hfq-protein complexes and their role in riboregulation

In addition to associating with nucleic acids, Hfq has been shown to co-purify with numerous proteins in *E. coli*. Of the 59 co-purifying proteins identified via proteomic studies, 34 were validated by sequential rounds of tagging and purification to interact with Hfq [97]. Most of the Hfq-associated proteins are involved in major steps of gene regulation and expression for instance transcription (RNA polymerase α, β, and β' subunits), translation (30S ribosomal subunit S13), RNA metabolism (RNase E), and protein folding (protein chaperones Hsp40 and Hsp70) [29,97]. Despite Hfq's connection to all steps of gene regulation, through such protein interactions, the link between Hfq and the function of most of these proteins remains unexplored.   The few protein-Hfq interactions that have been examined include poly(A) polymerase I (PAP I) [98,99], RNase E [26,97,100,101], and the transcription terminator Rho [102].

The average length and abundance of poly(A) tails in *E. coli* have been shown to decrease in a Hfq knockout mutant, suggesting that Hfq plays a role in the synthesis and/or stability of these tails [98,99,103]. Specifically, Hajnsdorf *et al.* found that poly(A) tails were approximately 20 adenosines long in a Hfq knockout mutant and that RNA from cells containing an active Hfq had tails exceeding 30 adenosines [98]. Prior to these findings, it was thought that poly(A) length depends solely on a balance between the activities of PAP I and exoribonucleases [104,105]. Therefore, it would be logical to predict that Hfq is either stimulating synthesis by PAP I or protecting the poly(A) extensions from endonucleases. Hfq has been shown to protect RNAs from degradation by RNase E [25,106] therefore the longer poly(A) tails in the presence of

Hfq may be due to Hfq binding to and protecting the extended tails from exoribonucleases. Contrary to this model, it was found that a Hfq deletion mutant further reduced the length of poly(A) tails in a series of exonuclease mutants, including a PNPase-7 knock-out, RNase II conditional mutant, and RNase E conditional mutant [25].

In vitro studies have verified that Hfq has a direct stimulatory effect on poly(A) synthesis by PAP I [98,103,107]. Purified PAP I extends an RNA primer by 200 to 900 adenosines in the presence of Hfq, but only by 20-150 adenosines in the absence of Hfq [98]. The lengths of these poly(A) extensions were dependent on the concentration of Hfq [98]; it should be noted that overexpression of Hfq in *E. coli* has not been shown to further stimulate poly(A) extension [99]. Furthermore, it was observed that poly(A) extension is biphasic in the presence of PAP I and Hfq, consistent with Hfq stimulating poly(A) extension more strongly once the extension exceeds 20 adenosines [98]. The biphasic nature of the reaction in the presence of Hfq is presumably due to Hfq having little to no affinity for RNAs containing short poly(A) tails, which is supported by EMSAs that show only week binding of Hfq to RNAs with poly(A) tails less than 18 adenosines [98]. Additionally, the observation of either rapidly elongated or non-elongated RNAs suggests that Hfq switches PAP I from being a distributive enzyme to a processive enzyme [98].

Co-immunoprecipitation studies have revealed that Hfq also interacts with components of the RNA "degradosome", including RNase E, polynucleotide phosphorylase (PNPase), and enolase [26,97,100,101]. Additional pull-down assays using an *E. coli* strain containing a FLAG-tagged RNase E truncated mutant revealed that the Hfq-RNase E interaction requires the C-terminal scaffold region of RNase E (residues 702 to 1061) [100]. Pull-down assays of cell extracts lacking PNPase or the adenosine triphosphate(ATP)-dependent RNA helicase RhlB showed that the Hfq-RNase E interaction is not mediated by either of these protein components

of the degradosome; additionally, a pull-down assay using cell extracts depleted of enolase revealed that the Hfq interaction is not mediated by this protein either [100]. Finally, a pull-down assay using cell extract treated with micrococcal nuclease revealed that Hfq still interacts with RNase E in the absence of RNA, further supporting a direct association between these two proteins [100]. Recent studies revealed that Hfq interacts with multiple sites in the C-terminal domain of RNase E [26]. Specifically, it was found that Hfq binds to residues 711-750, 801-844, and 845-4061, though only the interaction with residues 711-750 is required for sufficient recruitment of RNase E to target mRNAs .

Pull-down experiments using purified Rho factor and hexahistidine-tagged Hfq revealed that these two proteins also interact [26,102]. This interaction was not affected by pre-treatment with nucleases, suggesting that the interaction is at least partially due to a protein-protein interaction rather than an indirect result of binding the same RNAs [102]. Fluorescence quenching experiments using Cy3-labeled Hfq gave a K_d of 40 nM for the Hfq-Rho interaction [102]. *In vitro* transcription experiments using *E. coli* RNAP, *E. coli* Rho (*Eco* Rho), and a DNA template revealed that Hfq represses Rho-dependent termination in a dose dependent manner, independent of the order of addition of the reaction components [102]. Hfq did not suppress Rho from *Mycobacterium tuberculosis* (*Mtu* Rho)*,* which is a species that does not contain a detectable Hfq homologue [102]. The inability of Hfq to suppress *Mtu* Rho suggests that the *in vitro* repression of Rho-dependent termination is due to an interaction between Hfq and *Eco* Rho. Furthermore, Hfq was shown to inhibit Rho-dependent unwinding of a synthetic RNA-DNA construct, further supporting a direct interaction between Rho and Hfq [102]. Additionally, an ATP assay revealed that Hfq inhibits Rho's ATPase activity in a dose dependent manner [102]. Collectively, these results indicate that Hfq inhibits all enzymatic functions of Rho [102]. Transcription termination experiments in the presence of poly(A) RNA were shown to strongly

suppress Hfq-mediated antitermination, whereas poly(U) RNA had no effect. This result suggests that Rho inhibition is mediated by the "distal" face of Hfq, which is known to interact with A-rich RNAs [102]. In addition to these findings, poly(A) RNA was also shown to suppress Hfq's inhibition of the helicase and ATPase functions of Rho [102]. Furthermore, *in vivo* studies revealed that Hfq mediates transcription antitermination under conditions suboptimal for Rho function [102].

Objectives

The first objective of the work reported here was to biochemically characterize Hfq–RNA molecular interactions. This project initially focused on determining the binding affinity of a previously unstudied Hfq homolog from *Thermotoga maritima* (*Tma*) and fluorescently labeled RNA probes, thus enabling comparison of RNA-binding by *Tma* Hfq to Hfq homologs previously described in the literature. Although the 3D structures of numerous Hfq homologs indicate a high degree of global structural conservation (e.g. hexamers), there are multiple amino acid substitutions in the RNA-binding sites across different species; a comparative analysis of binding properties for a limited number of well-characterized RNA probes enabled us to elucidate the effects of amino acid variations on RNA binding. The expression and purification of *Tma* Hfq in *E. coli* led to the fortuitous discovery of novel, Hfq-binding nanoRNAs (Chapter 2). The Hfq-binding properties of these nanoRNAs were characterized using fluorescence polarization assays, which revealed nanomolar binding affinities. While identifying these nanoRNAs, a series of rapid colorimetric assays were implemented and optimized to differentiate between RNA and DNA (Appendix I).

The second aim of this thesis project was to biophysically characterize the oligomeric states of *Tma* Hfq, and to examine the RNA-binding properties of these various states in order to

explore the mechanistic implications of 'molecular coupling' between oligomerization and RNA binding (Chapter 3). The oligomeric states were characterized by crosslinking experiments, analytical size exclusion chromatography (AnSEC), isothermal calorimetry (ITC), and semi-native polyacrylamide gel electrophoresis (PAGE). This work showed that *Tma* Hfq forms a hexameric and dodecameric state in solution and that the two states are in equilibrium with the hexameric state being more predominant in the mM concentration range. Fluorescence polarization assays revealed that both oligomeric states interact with A-rich and U-rich RNA with nanomolar affinities.

The third aim of this thesis project was to determine the 3D structure of *Tma* Hfq by X-ray crystallography, motivated by the fact that biochemical function arises from biomolecular 3D structure (Chapter 4). Crystals grown in the presence and absence of RNA were used to collect diffraction data and ultimately determine the atomic resolution structure. This project was pursued in collaboration with Mr. Peter Randolph, who solved the 3D structure via molecular replacement.

Transitioning from the atomic to the cellular scale, the fourth aim of this project was to elucidate the expression level, localization, and binding partners of Hfq within *T. maritima* cells. Initial *in vivo* studies of the expression levels, cell localization, and binding partners of *T. maritima* Hfq are summarized in Chapter 5. Quantitative Western blots were optimized to determine the concentration of Hfq in a heterogeneous sample. In future studies, flow cytometry will be used to determine the number of *T. maritima* cells in a sample so the concentration of Hfq determine via quantitative Western blots can be used to estimate the concentration of Hfq expressed *in vivo*. *T. maritima* cells were visualized using EM with immunogold-labeling to determine the localization of Hfq *in vivo*. Future studies will focus on optimizing the fixation protocol so that that localization can be determined in intact *T. maritima*

cells. Binding partners were co-immunoprecipitated with *Tma* Hfq. The RNA and protein binding

partners will be identified by next generation sequencing and liquid chromatogram tandem MS

(LC-MS/MS), respectively.   This work required the production, purification, and characterization

of Anti-*Tma* Hfq polyclonal antibodies (pAb) for the *in vitro* and *in vivo* detection and isolation of

Tma Hfq (Appendix 2).

The main implications of this work are (1) a far deeper understanding of the molecular

and structural basis for Hfq's role in mediating sRNA-based assemblies and (2) a clearer and

higher resolution picture of Hfq's *in vivo* role in RNA-based cellular circuits in the otherwise well-

studied model organism, *T. maritima*.   Ultimately, exploring Hfq's role in an evolutionarily

ancient bacterial species, such as *Tma*, will broaden and deepen our understanding of the

evolution of the ubiquitous RNA-associated Sm protein family.

References

1. Voet D, Voet JG (2011) Biochemistry. Hoboken, N.J.: Wiley. xxv, 1428, 1453 p. p.

2. Perez-Martin J, Rojo F, de Lorenzo V (1994) Promoters responsive to DNA bending: a common
theme in prokaryotic gene expression. Microbiol Rev 58: 268-290.

3. Dougan DA, Mogk A, Bukau B (2002) Protein folding and degradation in bacteria: to degrade
or not to degrade? That is the question. Cell Mol Life Sci 59: 1607-1616.

4. Gripenland J, Netterling S, Loh E, Tiensuu T, Toledo-Arana A, et al. (2010) RNAs: regulators of
bacterial virulence. Nat Rev Microbiol 8: 857-866.

5. Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, et al. (2004) The small RNA chaperone
Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio
cholerae*. Cell 118: 69-82.

6. Gottesman S, Storz G (2011) Bacterial small RNA regulators: versatile roles and rapidly evolving variations. Cold Spring Harb Perspect Biol 3.

7. Storz G, Vogel J, Wassarman KM (2011) Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell 43: 880-891.

8. He L, Hannon GJ (2004) MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet 5: 522-531.

9. Filipowicz W, Bhattacharyya SN, Sonenberg N (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat Rev Genet 9: 102-114.

10. Liu JM, Camilli A (2010) A broadening world of bacterial small RNAs. Curr Opin Microbiol 13: 18-23.

11. Wagner EG (2009) Kill the messenger: bacterial antisense RNA promotes mRNA decay. Nat Struct Mol Biol 16: 804-806.

12. Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. Nature Reviews Microbiology 9: 578-589.

13. Franze de Fernandez MT, Eoyang L, August JT (1968) Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. Nature 219: 588-590.

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

15. Ramos CG, Sousa SA, Grilo AM, Feliciano JR, Leitao JH (2011) The second RNA chaperone, Hfq2, is also required for survival under stress and full virulence of *Burkholderia cenocepacia* J2315. J Bacteriol 193: 1515-1526.

16. Sousa SA, Ramos CG, Moreira LM, Leitao JH (2010) The hfq gene is required for stress resistance and full virulence of *Burkholderia cepacia* to the nematode *Caenorhabditis elegans*. Microbiology 156: 896-908.

17. Sun XG, Zhulin I, Wartell RM (2002) Predicted structure and phyletic distribution of the RNA-binding protein Hfq. Nucleic Acids Re 30: 3662-3671.

18. Tsui HC, Leung HC, Winkler ME (1994) Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in *Escherichia coli* K-12. Mol Microbiol 13: 35-49.

19. Lybecker MC, Abel CA, Feig AL, Samuels DS (2010) Identification and function of the RNA chaperone Hfq in the Lyme disease spirochete *Borrelia burgdorferi*. Mol Microbiol 78: 622-635.

20. Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, et al. (2008) Deep Sequencing Analysis of Small Noncoding RNA and mRNA Targets of the Global Post-Transcriptional Regulator, Hfq. Plos Genetics 4.

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

22. Wilms I, Moller P, Stock AM, Gurski R, Lai EM, et al. (2012) Hfq influences multiple transport systems and virulence in the plant pathogen *Agrobacterium tumefaciens*. J Bacteriol 194: 5209-5217.

23. Ding Y, Davis BM, Waldor MK (2004) Hfq is essential for *Vibrio cholerae* virulence and downregulates sigma expression. Mol Microbiol 53: 345-354.

24. Sharma CM, Vogel J (2009) Experimental approaches for the discovery and characterization of regulatory small RNA. Curr Opin Microbiol 12: 536-546.

25. Zhang A, Wassarman KM, Rosenow C, Tjaden BC, Storz G, et al. (2003) Global analysis of small RNA and mRNA targets of Hfq. Mol Microbiol 50: 1111-1124.

26. Ikeda Y, Yagi M, Morita T, 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.

27. Kulesus RR, Diaz-Perez K, Slechta ES, Eto DS, Mulvey MA (2008) Impact of the RNA chaperone Hfq on the fitness and virulence potential of uropathogenic *Escherichia coli*. Infect Immun 76: 3019-3026.

28. Jousselin A, Metzinger L, Felden B (2009) On the facultative requirement of the bacterial RNA chaperone, Hfq. Trends Microbiol 17: 399-405.

29. Sobrero P, Valverde C (2012) The bacterial protein Hfq: much more than a mere RNA-binding factor. Crit Rev Microbiol 38: 276-299.

30. Boggild A, Overgaard M, Valentin-Hansen P, Brodersen DE (2009) Cyanobacteria contain a structural homologue of the Hfq protein with altered RNA-binding properties. FEBS J 276: 3904-3915.

31. Hammer BK, Bassler BL (2007) Regulatory small RNAs circumvent the conventional quorum sensing pathway in pandemic *Vibrio cholerae*. Proc Natl Acad Sci U S A 104: 11145-11149.

32. Mandin P, Repoila F, Vergassola M, Geissmann T, Cossart P (2007) Identification of new noncoding RNAs in *Listeria monocytogenes* and prediction of mRNA targets. Nucleic Acids Res 35: 962-974.

33. Kajitani M, Kato A, Wada A, Inokuchi Y, Ishihama A (1994) Regulation of the *Escherichia coli* hfq gene encoding the host factor for phage Q beta. J Bacteriol 176: 531-534.

34. Azam TA, Iwata A, Nishimura A, Ueda S, Ishihama A (1999) Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. Journal of Bacteriology 181: 6361-6370.

35. Azam TA, Ishihama A (1999) Twelve species of the nucleoid-associated protein from *Escherichia coli* - Sequence recognition specificity and DNA binding affinity. J Biol Chem 274: 33105-33113.

36. Diestra E, Cayrol B, Arluison V, Risco C (2009) Cellular electron microscopy imaging reveals the localization of the Hfq protein close to the bacterial membrane. PLoS One 4: e8301.

37. Liu Y, Wu N, Dong J, Gao Y, Zhang X, et al. (2010) Hfq is a global regulator that controls the pathogenicity of *Staphylococcus aureus*. PLoS One 5.

38. Bohn C, Rigoulay C, Bouloc P (2007) No detectable effect of RNA-binding protein Hfq absence in *Staphylococcus aureus*. BMC Microbiol 7: 10.

39. Vecerek B, Moll I, Blasi U (2005) Translational autocontrol of the *Escherichia coli hfq* RNA chaperone gene. RNA 11: 976-984.

40. Ziolkowska K, Derreumaux P, Folichon M, Pellegrini O, Regnier P, et al. (2006) Hfq variant with altered RNA binding functions. Nucleic Acids Res 34: 709-720.

41. Sobrero P, Valverde C (2011) Evidences of autoregulation of hfq expression in *Sinorhizobium meliloti* strain 2011. Arch Microbiol 193: 629-639.

42. Berghoff BA, Glaeser J, Sharma CM, Zobawa M, Lottspeich F, et al. (2011) Contribution of Hfq to photooxidative stress resistance and global regulation in *Rhodobacter sphaeroides*. Mol Microbiol 80: 1479-1495.

43. Timmermans J, Van Melderen L (2010) Post-transcriptional global regulation by CsrA in bacteria. Cell Mol Life Sci 67: 2897-2908.

44. Baker CS, Eory LA, Yakhnin H, Mercante J, Romeo T, et al. (2007) CsrA inhibits translation initiation of *Escherichia coli* hfq by binding to a single site overlapping the Shine-Dalgarno sequence. J Bacteriol 189: 5472-5481.

45. Tsui HC, Feng G, Winkler ME (1996) Transcription of the *mutL* repair, *miaA* tRNA modification, *hfq* pleiotropic regulator, and *hflA* region protease genes of *Escherichia coli* K-12 from clustered Esigma32-specific promoters during heat shock. J Bacteriol 178: 5719-5731.

46. Zhang A, Wassarman KM, Ortega J, Steven AC, Storz G (2002) The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. Mol Cell 9: 11-22.

47. Arluison V, Derreumaux P, Allemand F, Folichon M, Hajnsdorf E, et al. (2002) Structural Modelling of the Sm-like Protein Hfq from *Escherichia coli*. J Mol Biol 320: 705-712.

48. Schumacher MA, Pearson RF, Moller T, Valentin-Hansen P, Brennan RG (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. EMBO J 21: 3546-3556.

49. Moller T, Franch T, Hojrup P, Keene DR, Bachinger HP, et al. (2002) Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. Mol Cell 9: 23-30.

50. Sauter C, Basquin J, Suck D (2003) Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from Escherichia coli. Nucleic Acids Res 31: 4091-4098.

51. Nikulin A, Stolboushkina E, Perederina A, Vassilieva I, Blaesi U, et al. (2005) Structure of *Pseudomonas aeruginosa* Hfq protein. Acta Crystallogr D Biol Crystallogr 61: 141-146.

52. Sauer E, Schmidt S, Weichenrieder O (2012) Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. Proc Natl Acad Sci U S A 109: 9396-9401.

53. Murina VN, Nikulin AD (2011) RNA-binding Sm-like proteins of bacteria and archaea. similarity and difference in structure and function. Biochemistry (Mosc) 76: 1434-1449.

54. Nielsen JS, Boggild A, Andersen CB, Nielsen G, Boysen A, et al. (2007) An Hfq-like protein in archaea: crystal structure and functional characterization of the Sm protein from *Methanococcus jannaschii*. RNA 13: 2213-2223.

55. Sun XG, Wartell RM (2006) *Escherichia coli* Hfq binds A(18) and DsrA domain II with similar 2:1 Hfq(6)/RNA stoichiometry using different surface sites. Biochemistry 45: 4875-4887.

56. Toro I, Basquin J, Teo-Dreher H, Suck D (2002) Archaeal Sm proteins form heptameric and hexameric complexes: crystal structures of the Sm1 and Sm2 proteins from the hyperthermophile *Archaeoglobus fulgidus*. J Mol Biol 320: 129-142.

57. Mura C, Kozhukhovsky A, Gingery M, Phillips M, Eisenberg D (2003) The oligomerization and ligand-binding properties of Sm-like archaeal proteins (SmAPs). Protein Sci 12: 832-847.

58. Collins BM, Harrop SJ, Kornfeld GD, Dawes IW, Curmi PM, et al. (2001) Crystal structure of a heptameric Sm-like protein complex from archaea: implications for the structure and evolution of snRNPs. J Mol Biol 309: 915-923.

59. Toro I, Thore S, Mayer C, Basquin J, Seraphin B, et al. (2001) RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. EMBO J 20: 2293-2303.

60. Mura C, Cascio D, Sawaya MR, Eisenberg DS (2001) The crystal structure of a heptameric archaeal Sm protein: Implications for the eukaryotic snRNP core. Proc Natl Acad Sci U S A 98: 5532-5537.

61. Mura C, Phillips M, Kozhukhovsky A, Eisenberg D (2003) Structure and assembly of an augmented Sm-like archaeal protein 14-mer. Proc Natl Acad Sci U S A 100: 4539-4544.

62. Wilusz CJ, Wilusz J (2005) Eukaryotic Lsm proteins: lessons from bacteria. Nat Struct Mol Biol 12: 1031-1036.

63. Wilusz CJ, Wilusz J (2013) Lsm proteins and Hfq: Life at the 3' end. RNA Biol 10: 592-601.

64. Scofield DG, Lynch M (2008) Evolutionary diversification of the Sm family of RNA-associated proteins. Mol Biol Evol 25: 2255-2267.

65. Mura C, Randolph PS, Patterson J, Cozen AE (2013) Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function. RNA Biol 10: 636-651.

66. Mikulecky PJ, Kaw MK, Brescia CC, Takach JC, Sledjeski DD, et al. (2004) *Escherichia coli* Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. Nat Struct Mol Biol 11: 1206-1214.

67. Sittka A, Sharma CM, Rolle K, 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.

68. Kambach C, Walke S, Young R, Avis JM, de la Fortelle E, et al. (1999) Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. Cell 96: 375-387.

69. Stark H, Dube P, Luhrmann R, Kastner B (2001) Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. Nature 409: 539-542.

70. Achsel T, Stark H, Luhrmann R (2001) The Sm domain is an ancient RNA-binding motif with oligo(U) specificity. Proc Natl Acad Sci U S A 98: 3685-3689.

71. Will CL, Luhrmann R (2011) Spliceosome structure and function. Cold Spring Harb Perspect Biol 3.

72. Moll JM, Sobti M, Mabbutt BC (2011) The Lsm Proteins: Ring Architectures for RNA capture. In: Grabowski PP, editor. RNA Processing: InTech.

73. Fischer S, Benz J, Spath B, Jellen-Ritter A, Heyer R, et al. (2011) Regulatory RNAs in *Haloferax volcanii*. Biochem Soc Trans 39: 159-162.

74. Fischer S, Benz J, Spath B, Maier LK, Straub J, et al. (2010) The archaeal Lsm protein binds to small RNAs. J Biol Chem 285: 34429-34438.

75. Sledjeski DD, Whitman C, Zhang A (2001) Hfq is necessary for regulation by the untranslated RNA DsrA. J Bacteriol 183: 1997-2005.

76. Jain K, Updegrove TB, Wartell RM (2011) A Thermodynamic Perspective of sRNA-mRNA Interactions and the Role of Hfq. Frontiers in Nucleic Acid.

77. Sledjeski DD, Gupta A, 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.

78. Majdalani N, Hernandez D, Gottesman S (2002) Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. Mol Microbiol 46: 813-826.

79. Aslund F, Zheng M, Beckwith J, Storz G (1999) Regulation of the OxyR transcription factor by hydrogen peroxide and the cellular thiol-disulfide status. Proc Natl Acad Sci U S A 96: 6161-6165.

80. Wang W, Wang L, Wu J, Gong Q, Shi Y (2013) Hfq-bridged ternary complex is important for translation activation of rpoS by DsrA. Nucleic Acids Res 41: 5938-5948.

81. Updegrove TB, Correia JJ, Chen YF, Terry C, Wartell RM (2011) The stoichiometry of the *Escherichia coli* Hfq protein bound to RNA. RNA 17: 489-500.

82. Wang W, Wang L, Zou Y, Zhang J, Gong Q, et al. (2011) Cooperation of *Escherichia coli* Hfq hexamers in DsrA binding. Genes Dev 25: 2106-2117.

83. Geissmann TA, Touati D (2004) Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. EMBO J 23: 396-405.

84. Henderson CA, Vincent HA, Casamento A, Stone CM, Phillips JO, et al. (2013) Hfq binding changes the structure of *Escherichia coli* small noncoding RNAs OxyS and RprA, which are involved in the riboregulation of rpoS. RNA 19: 1089-1104.

85. Sauer E (2013) Structure and RNA-binding properties of the bacterial LSm protein Hfq. RNA Biol 10: 610-618.

86. Panja S, Schu DJ, Woodson SA (2013) Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. Nucleic Acids Res 41: 7536-7546.

87. Lease RA, Woodson SA (2004) Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. J Mol Biol 344: 1211-1223.

88. Wagner EG (2013) Cycling of RNAs on Hfq. RNA Biol 10: 619-626.

89. Fender A, Elf J, Hampel K, Zimmermann B, Wagner EG (2010) RNAs actively cycle on the Sm-like protein Hfq. Genes Dev 24: 2621-2626.

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

91. Adamson DN, Lim HN (2011) Essential requirements for robust signaling in Hfq dependent small RNA networks. PLoS Comput Biol 7: e1002138.

92. Takada A, Wachi M, Kaidow A, Takamura M, Nagai K (1997) DNA binding properties of the hfq gene product of *Escherichia coli*. Biochem Biophys Res Commun 236: 576-579.

93. Updegrove TB, Correia JJ, Galletto R, Bujalowski W, Wartell RM (2010) *E. coli* DNA associated with isolated Hfq interacts with Hfq's distal surface and C-terminal domain. Biochim Biophys Acta 1799: 588-596.

94. Geinguenaud F, Calandrini V, Teixeira J, Mayer C, Liquier J, et al. (2011) Conformational transition of DNA bound to Hfq probed by infrared spectroscopy. Phys Chem Chem Phys 13: 1222-1229.

95. Lee T, Feig AL (2008) The RNA binding protein Hfq interacts specifically with tRNAs. RNA 14: 514-523.

96. Sukhodolets MV, Garges S (2003) Interaction of *Escherichia coli* RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. Biochemistry 42: 8022-8034.

97. Butland G, Peregrin-Alvarez JM, Li J, Yang W, Yang X, et al. (2005) Interaction network containing conserved and essential protein complexes in *Escherichia coli*. Nature 433: 531-537.

98. Hajnsdorf E, 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.

99. Mohanty BK, Maples VF, Kushner SR (2004) The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in Escherichia coli. Mol Microbiol 54: 905-920.

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

101. Worrall JA, Gorna M, Crump NT, Phillips LG, Tuck AC, et al. (2008) Reconstitution and analysis of the multienzyme *Escherichia coli* RNA degradosome. J Mol Biol 382: 870-883.

102. Rabhi M, Espeli O, Schwartz A, Cayrol B, Rahmouni AR, et al. (2011) The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. EMBO J 30: 2805-2816.

103. Le Derout J, Folichon M, Briani F, Deho G, Regnier P, et al. (2003) Hfq affects the length and the frequency of short oligo(A) tails at the 3' end of *Escherichia coli* rpsO mRNAs. Nucleic Acids Res 31: 4017-4023.

104. Hajnsdorf E, Braun F, Haugel-Nielsen J, Le Derout J, Regnier P (1996) Multiple degradation pathways of the rpsO mRNA of *Escherichia coli*. RNase E interacts with the 5' and 3' extremities of the primary transcript. Biochimie 78: 416-424.

105. Coburn GA, Mackie GA (1999) Degradation of mRNA in *Escherichia coli*: an old problem with some new twists. Prog Nucleic Acid Res Mol Biol 62: 55-108.

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

107. Hankins JS, Denroche H, Mackie GA (2010) Interactions of the RNA-binding protein Hfq with cspA mRNA, encoding the major cold shock protein. J Bacteriol 192: 2482-2490.

108. Ansong C, Yoon H, Porwollik S, Mottaz-Brewer H, Petritis BO, et al. (2009) Global systems-level analysis of Hfq and SmpB deletion mutants in *Salmonella*: implications for virulence and global protein translation. PLoS One 4: e4809.

109. Castro SL, Nelman-Gonzalez M, Nickerson CA, Ott CM (2011) Induction of attachment-independent biofilm formation and repression of Hfq expression by low-fluid-shear culture of *Staphylococcus aureus*. Appl Environ Microbiol 77: 6368-6378.

110. Wilson JW, Ott CM, Honer zu Bentrup K, Ramamurthy R, Quick L, et al. (2007) Space flight alters bacterial gene expression and virulence and reveals a role for global regulator Hfq. Proc Natl Acad Sci U S A 104: 16299-16304.

111. Guisbert E, Rhodius VA, Ahuja N, Witkin E, Gross CA (2007) Hfq modulates the sigmaE-mediated envelope stress response and the sigma32-mediated cytoplasmic stress response in *Escherichia coli*. J Bacteriol 189: 1963-1973.

112. Geng J, Song Y, Yang L, Feng Y, Qiu Y, et al. (2009) Involvement of the post-transcriptional regulator Hfq in *Yersinia pestis* virulence. PLoS One 4: e6213.

113. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A 98: 10037-10041.

114. Mura C, McCrimmon CM, Vertrees J, Sawaya MR (2010) An introduction to biomolecular graphics. PLoS Comput Biol 6.

115. Udekwu KI, Darfeuille F, Vogel J, Reimegard J, Holmqvist E, et al. (2005) Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. Gen Dev 19: 2355-2366.

116. Updegrove T, Wilf N, Sun XG, Wartell RM (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.
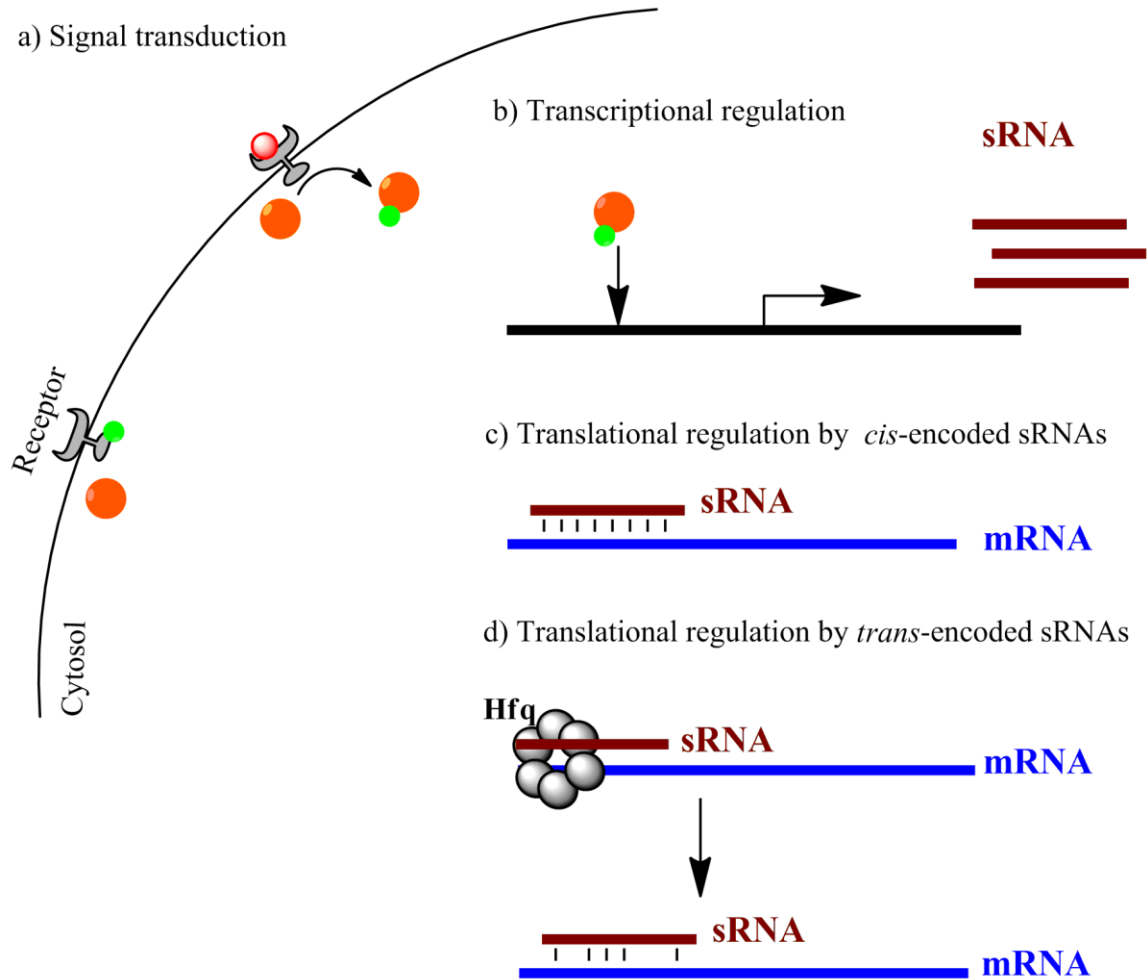
Figure 1.1 Regulation of gene expression by small non-coding RNAs (sRNA). The cell can detect environmental changes via receptors in the outer membrane (a), which allow the cell to adapt its gene expression depending on what nutrients are available or what stress conditions the cell must withstand. The environmental signal activates transcription of an sRNA, which is either located on the antisense strand within a gene (*cis*-encoded sRNA) or in an intergenic region (*trans*-encoded sRNA) (b). *Cis*-encoded sRNA can then anneal with its target mRNA with perfect complementarity (c), whereas *trans*-encoded sRNA only has partial complementarity with its target mRNA (d) and requires the RNA chaperone, Hfq, to facilitate annealing [30].
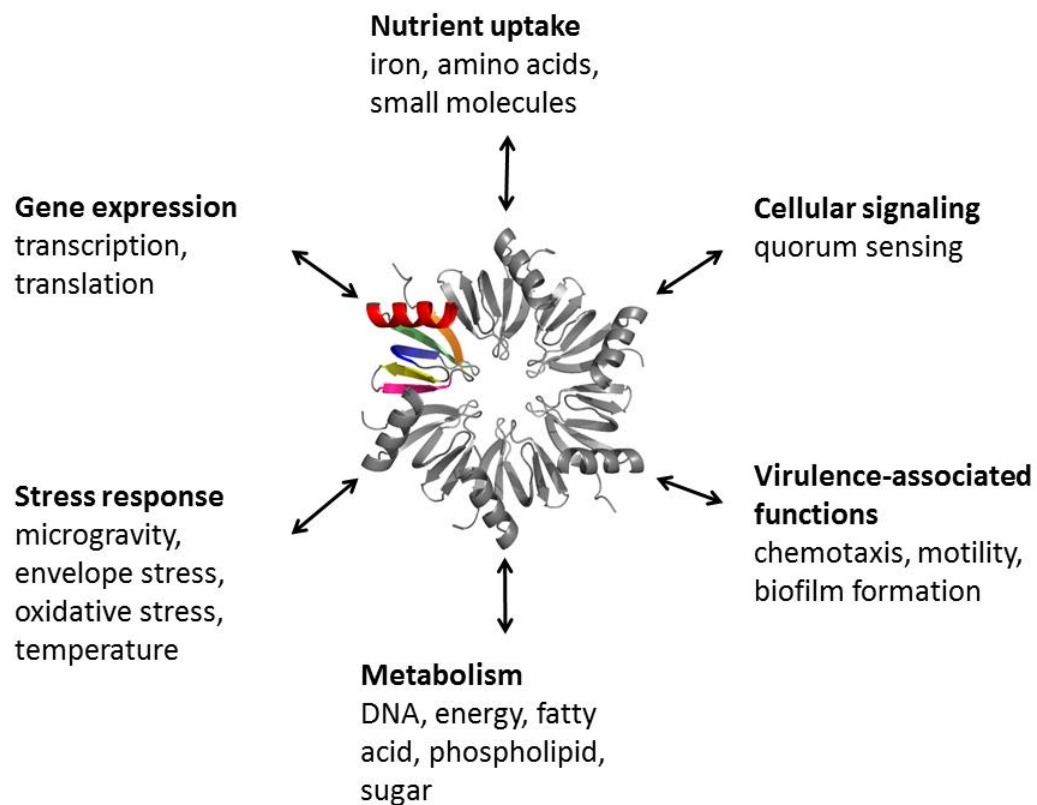
Figure 1.2 Hfq-dependent pathways. Hfq is now known to regulate multiple bacterial physiological pathways, via its interactions with RNA. Shown here is a subset of the Hfq-dependent pathways identified by functional and genomic studies in *Salmonella enterica* [20,108], *Staphylococcus aureus* [109], *Salmonella typhimurium[110]* , *Escherichia coli* [111], *Yersinia pestis[112]* , and *Agrobacterium tumefaciens* [22].
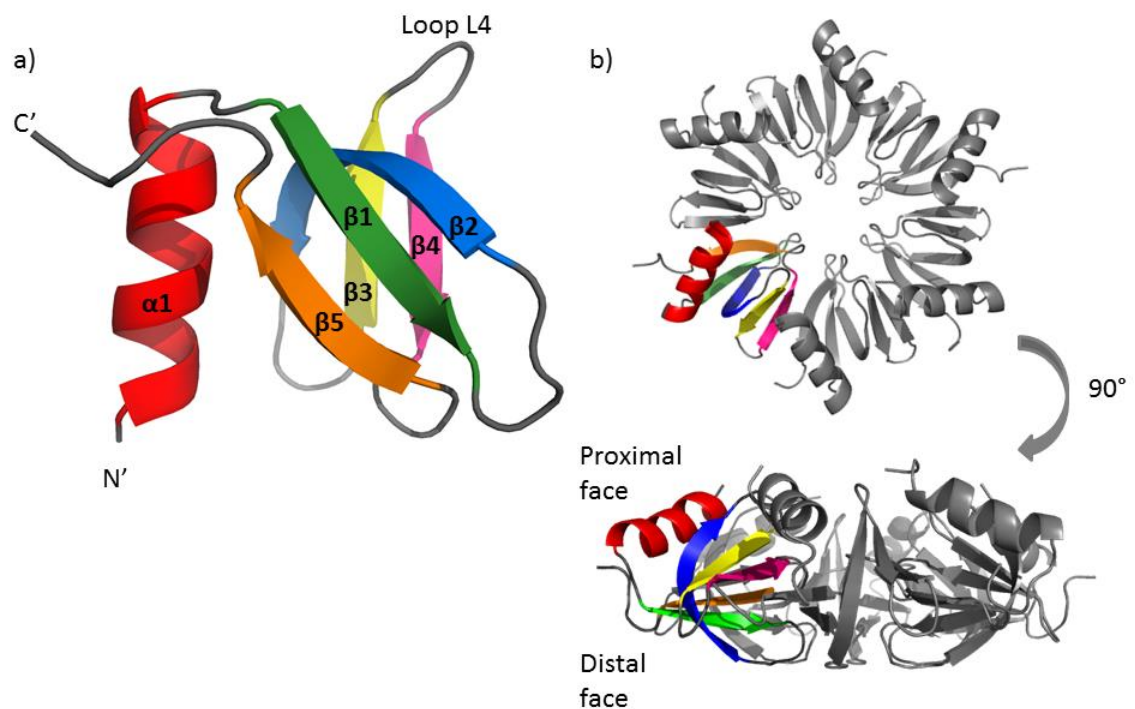
Figure 1.3 Tertiary (a) and quaternary (b) structure of Hfq depicted as a cartoon. (a) The Hfq monomer (3GIB) exhibits the Sm fold, which consists of the Sm 1 and Sm 2 motifs containing three (β1-β3) and two (β4-β5) consecutive β-strands, respectively. In addition to the Sm fold the Hfq protein contains an N-terminal α-helix that is present in all Sm proteins. (b) Crystal structures of Hfq have shown that Hfq forms a homohexameric toroid structure that is approximately 70-80 Å in diameter with an 8-10 Å central hole [48,50,51], shown here from both a top and side view.

Sequence Logo alignment figure.

Secondary structure annotation (top): α-helix, β1, β2, β3, β4, β5

Bits scale: 4, 2, 0

Position markers: 1, 10, 20, 30, 40, 50, 60, 70

```
Sequence Logo     [logo: I QDPFLNALRK ERYPVSIYLV NGIKLQGQIE SFDQFVLLK NTVSQMVK HAISTIVD]

A. aeolicus       MPYKE     RESFINTAAK  KRVKVSVYLV  NGVRLQGRIR  SFDLFTIIME  DGKQQTLVVK  HAITTIVPHE
A. tumerfaciens   MAERSQNIE QDLFINTVVK  QKISLTIFLI  NGVKLTGVVT  SFDNFCVIMR  RDGHSQIVVK  HAISTIMPGQ
B. anthracis*     MYLEHIE   QDELYKQIKE  EKGIMTIFLK  SEVRIVSEIV  AIDKFTVEML  VDGKQQLIMK  QAITSRIMK
B. mallei*        MSNKGQLIE QDPFINALKK  EHVPVSIYHV  NGIKLQGNIE  SFDQYVVIMR  NT-VTQMVVK  HAISTVVPAR
B. subtilis       MKPIN I   ODIFINQISR  ENTYMTVFIL  NGFQLRHQVK  GFDNFTVIME  SEGKQQLIME  HAISFAFABQK
C. acetobutylicum MNNKSTNNE ODIFINSARK  NKIPMAIHIT  NGFQMRGSVK  GFDSFTVIIE  SDGKQMMIMK  HAVSTITELR
E. coli           MAKGQSIE  ODPFINALER  ERVPMSIYHV  NGIKLQGQIE  SFDQFVIIMK  NT-VSQMVMK  HAISTVVESR
H. influenzae     MAERSQNIE ODLFINSVVK  ERIPLSIYHV  NGIKLQGQIE  SFDQFVIIMK  NT-VNQMVMK  HAISTIMPSQ
M. loti           MGNSHAGHNV ODPFINTLKR SKNPLTIFHI  NGVKLTGVVT  SFDNFCVIMR  RDGHSQIVMK  HAISTIMESR
M. marinus        MGVKGQLIE ODPFINILKK  EKVPMTVFHV  NGIKLQGMIT  SFDNYCLIMK  NS-VTQIVF   HAISNIVMER
N. europaea       MTAKGQMIE ODPFINALKK  ERIPLSIYHV  NGIKLQGQID  SFDQYVVIMR  NS-VTQMVMK  HAISTIVPAK
N. meningitidis   MSKGHSIE  ODPYINTLKK  EHVPVSIYHV  NGIKLQGQVE  SFDQYVVIMR  NTSVTQMVMK  HAISTIVPAR
P. aeruginosa     MSNKGQLIE ODPFINALRK  ERVPLSIYHV  NGIKLQGNIE  SFDQYVimMR  NT-VTQMVMK  HAISTVVESR
R. solanacearum   MIANENI   ODKALENFKA  NQTEMTVFFL  NGFQMKGVIE  EYDKYVVSIN  SQGKQHLIME  HAISTYTVET
S. aureus         MAKGQSIE  ODPFINALRV  ERVPMSIYHV  NGIKLQGQIE  SFDQFVIIMK  NT-VSQMVMK  HAISTVVESR
S. enterica       MAERSQNIE ODLFINTVVK  QKISLTIFLI  NGVKLTGVVT  SFDNFCVIMR  RDGHSQLVMK  HAISTIMPGQ
S. meliloti       MALAEKFNIE DRFINHLRV  NKIEMKVYHV  NGFQTKGFIR  SFDSYTVIME  SGNQQSLIMK  HAISTIIESS
T. maritima       MAKGQSIE  ODPFINALRR  ERIPLSIYHV  NGIKLQGQIE  SFDQFVIIMK  NT-VNQMVMK  HAISTVVEAR
V. cholerae       MAKGQSIE  ODPFINALR   ERVPMSIYHV  NGIKLQGQVE  SFDQFVIiMK  NT-VSQMVMK  HAISTVVESR
Y. pestis                   ODPFINALR   ERVPVSIYHV  NGIKLQGQVE  SFDQFVIIMK  NT-VSQMVMK  HAISTVVESR
```

Figure 1.4 Sequence alignment of Hfq proteins from 20 bacterial species. Sequences were aligned using Geneious 7.0.3 (Biomatters) with a Blosum62 scoring matrix and no end gap penalty. Secondary structural elements based on the *E. coli* Hfq crystal structure (2YHT) are shown above the aligned sequences: cylinder, α helix; arrow, β strand. The 85% consensus sequence is highlighted based on polarity: hydrophobic nonpolar residues (AFGILMPVW) are orange; uncharged polar residues (CNSTY) are green; acidic polar residues (DE) are red; basic polar residues (HKR) are blue. Accession numbers and full species names are as follows: *Aquifex aeolicus* (O66512); *Agrobacterium tumefaciens* (EHH08904); *Bacillus anthracis* (AAP25589); *Bacillus mallei* (AAU47553); *Bacillus subtilis* (O31796); *Clostridium acetobutylicum* (AAK23721); *Escherichia coli* (AGW11210); *Haemophilus influenza*e (AAC22070); *Mesorhizobium loti* (BAB47991); *Magnetococcus marinus* (WP_011711976); *Nitrosomonas europaea* (WP_011111865); *Neisseria meningitidis* (ADY93426); *Pseudomonas aeruginosa* (3QUI_A); *Ralstonia solanacearum* (AGH83762); *Staphylococcus aureus* (AGU61371); *Salmonella enterica* (AGK69916); *Sinorhizobium meliloti* (ADE20386); *Thermotoga maritima* (Q9WYX6); *Vibrio Cholerae* (ERP70544); *Yersinia pestis* (Q8ZIW2). Species marked with an asterisk (*) appear to encode two Hfq homologs in their genomes.
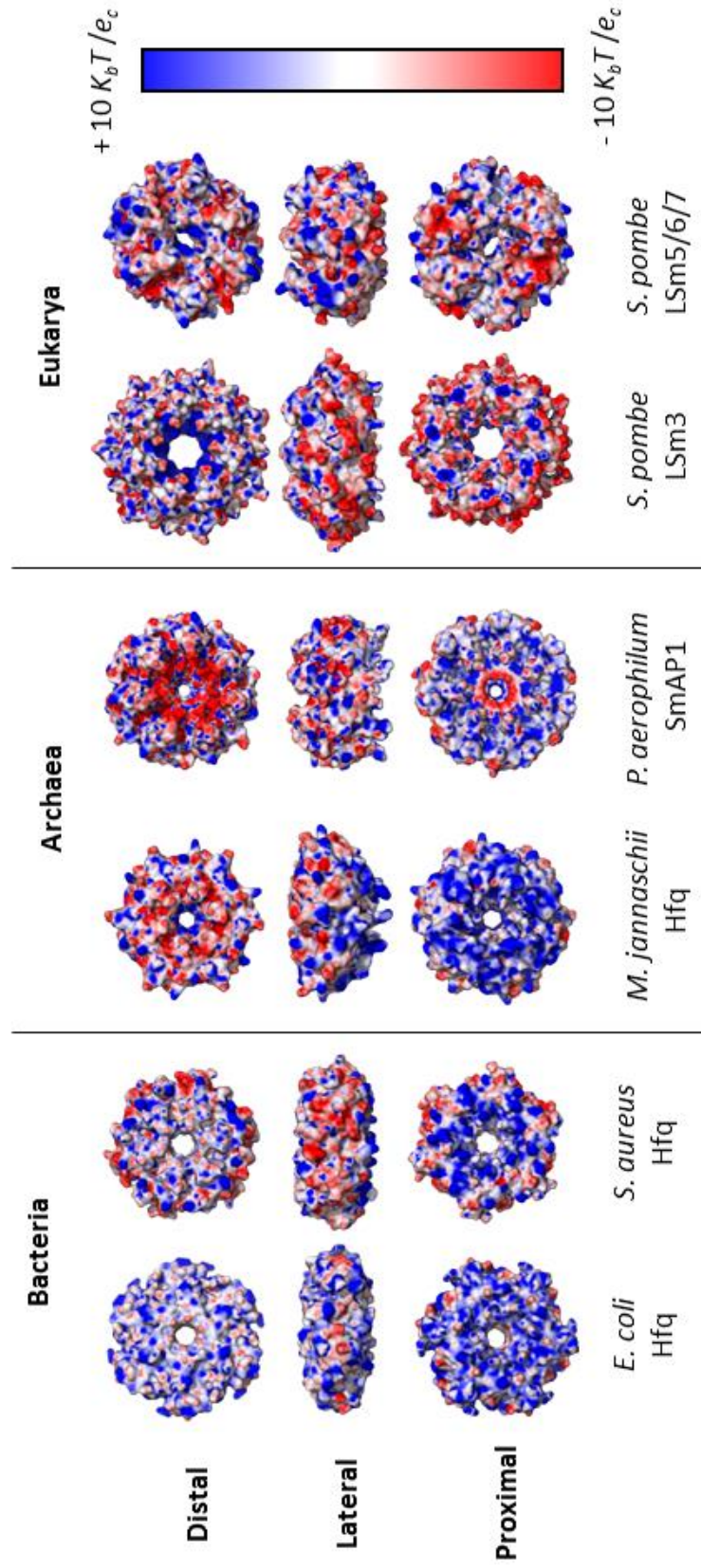
Figure 1.5 Surface electrostatic potential of the distinct Sm proteins from the three domains of life. Three-dimensional structures were visualized using PYMOl and surface electrostatic potential calculations were performed using the APBS (Adaptive Poisson-Boltzmann Solver [113,114]) plugin. The following PDB (protein data base) entries were used for these calculations: *E. coli* Hfq (2YHT); *S. aureus* Hfq (1KQ2); *M. jannaschii* Hfq (2QTC); *Pyrobaculum aerophilum* SmAP1 (1LNX); and *Schizosaccharomyces pombe* LSm3 (4EMG); *Schizosaccharomyces pombe* LSm5/6/7 (4EMK). For each molecular structure, potentials were graded as per the scale bar to the right.

a) Bacteria

b) Archaea

c) Eukarya

E. coli
Hfq

A. fulgidus
Sm2

P. aerophilum
SmAP1

P. aerophilum
SmAP3

Human U4 snRNP
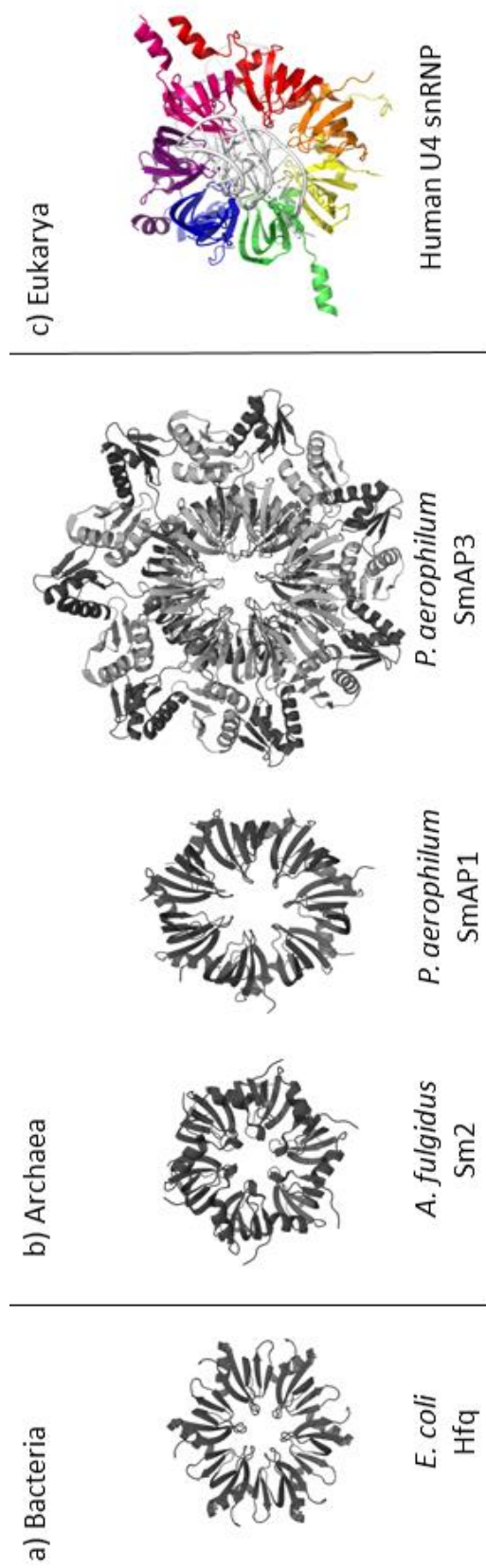
Figure 1.6 Oligomeric plasticity of the Sm/LSm ring across the different domains of life. (a)   A representative example of the homohexameric state adopted by the bacterial Hfq protein is illustrated by the structure of *E. coli* Hfq. (b) Sm proteins in archaea (SmAPs) assemble into various homo-oligomeric states.   Shown here are three examples of the known oligomeric states of SmAPs: hexamer, heptamer, and 14-mer, from left to right.   (c) Eukaryotic Sm proteins form heteroheptameric rings in the presence of RNA. The structure shown in this panel is the human U4 small nuclear ribonucleoprotein (snRNP) complex, which consists of Sm B/B′, Sm D1, Sm D2, Sm D3, Sm E, SmF, Sm G, and U4 snRNA (grey).   The following PDB entries were used to generate these images in PyMOL were 2YHKT (*E. coli* Hfq), 1LJO (*A. fulgidus* Sm2), 1LNX (*P. Aerophilum* SmAP1), 1M5Q (*P. Aerophilum* SmAP3), and 4EMK (human U4 snRNP).

a) Translational repression
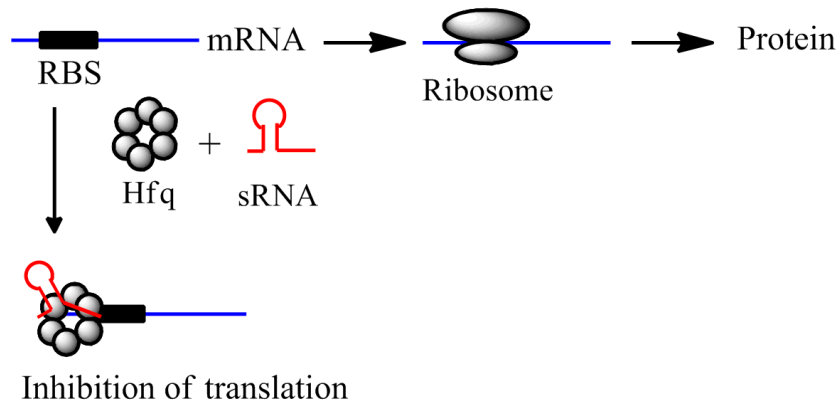
b) Translational activation
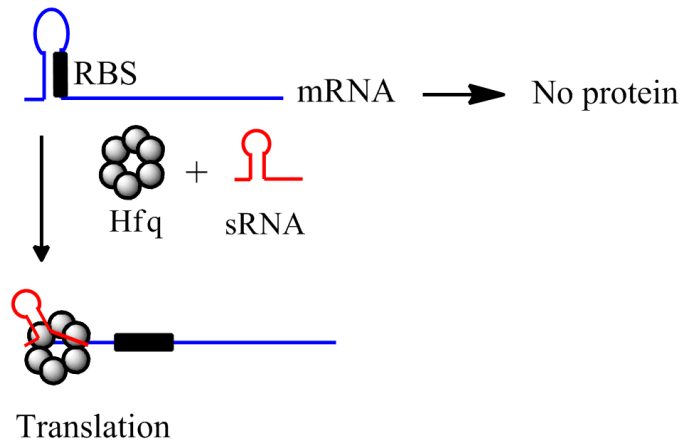
Figure 1.7 Hfq-mediated post-transcriptional regulation by sRNA. Translational repression (a) occurs when a sRNA anneals to a region near or overlapping the ribosome binding site (RBS), preventing the ribosome from binding to the mRNA and translating protein. Examples of sRNA/mRNA complexes that result in translation repression are RprA/*rpoS* and MicA/*ompA* [115,116]. Translational activation (b) occurs when the sRNA binds to the mRNA target sequence causing secondary structural changes that unmask the RBS. Examples of sRNA/mRNA complexes that result in translation activation are DsrA/*rpoS* and RyaA/*ptsG* .

a) Directed binding through the lateral surface
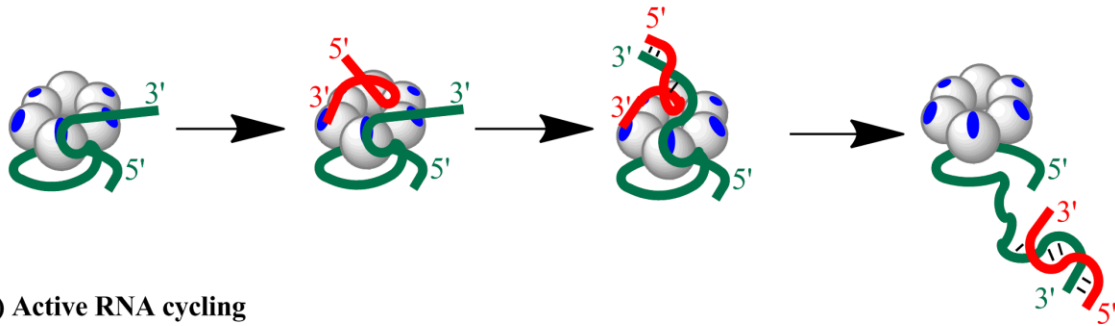
b) Active RNA cycling
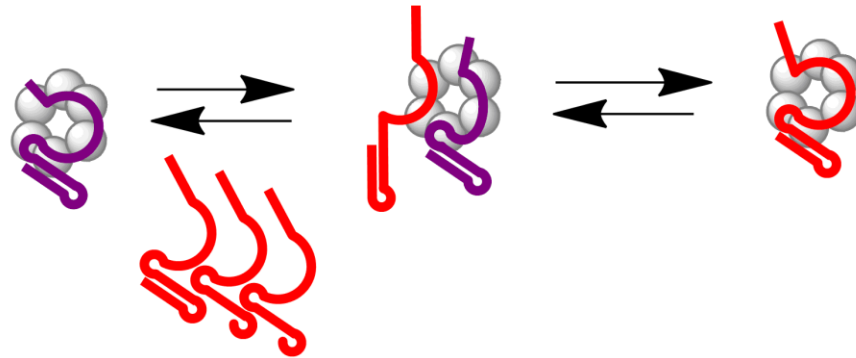
Figure 1.8 Current models for Hfq-directed annealing through the lateral face (a) and RNA cycling (b). The model of Hfq-directed annealing through the lateral face (a) proposes that the sRNA binds to the proximal face via a 3' anchor and interacts with an additional binding surface on the lateral face, or rim, of the protein which orient it with mRNA to promote annealing. The illustration above shows the Hfq hexamer in **light grey** with arginine-rich patches shown in **blue**. The mRNA is shown in **green** and the sRNA is shown in **red**. The RNA cycling model (b) proposes that RNA bound to Hfq can be actively displaced by a competitor RNA thereby resulting in less stable ribonucleoproteins with faster dissociation rates. The Hfq hexamer is shown in **light grey** with the bound sRNA in **purple** and the competing RNA in **red**. This model predicts that dissociation of the annealed mRNA-sRNA complex or an unpaired sRNA requires active dissociation by a competitor to free the Hfq for additional sRNAs and to achieve a quick phenotypic change.

Chapter 2

Discovery & Identification of Hfq-binding nanoRNAs in

Escherichia coli

Introduction

RNA-based regulatory pathways, and their effects on gene expression, enable bacteria to rapidly adapt to various environments that may be rapidly changing [1]. A bacterium's level of protein production for a given gene is not constant in time, but instead is regulated by the rate of translation and the availability of the messenger RNA (mRNA) transcript, which in turn depends on the rates of both transcription and degradation [2]. Multiple regulatory pathways have evolved to allow intricate control over expression levels depending on environmental signals, including available nutrients [3], cell density [4], and temperature [5].

Cellular RNA decay can occur via multiple mechanisms, each involving a subset of the nucleases encoded in the bacterial genome [2,6]. RNA degradation is commonly initiated by endonucleolytic cleavage at susceptible sites within a transcript, resulting in particular end chemistries (e.g., 5' monophosphate or 3' hydroxyl) that are required for subsequent degradation steps [2,7]. Exonucleases will then progressively degrade the RNA transcript into mononucleotides and release a small fragment of RNA (~ 2 to 5 nucleotides (nts) long). These small RNAs can then be further degraded by an oligoribonuclease.

Oligoribonuclease (*orn*) is an essential gene in *Escherichia coli;* inactivation of *orn* results in an accumulation of small oligonucleotides and termination of cell growth [8]. The term "nanoRNA" has been coined to refer to these ~ 2 to 4 nts RNA transcripts that accumulate in the absence of a functional oligoribonuclease [9,10]. Unlike *E. coli, Bacillus subtilis* encodes multiple enzymes with oligoribonuclease activity; this redundancy in oligoribonuclease function enables the production of viable knockout strains. Studies with knockout strains indicate that nanoRNAs regulate transcriptional start site selection in a growth phase-dependent manner [11,12]. The nanoRNAs anneal with the 5' regulatory region of a gene, within the -3 to 0 position, serving as a primer for transcription; notably, these transcripts do not have the 5' triphosphate end

chemistry characteristic of *de novo* transcription but instead carry either a 5' hydroxyl or 5' monophosphate [11].

The rate of degradation of a particular cellular RNA depends on the intrinsic thermodynamic stability of the transcript [13], the RNA's resistance to nucleases [7], as well as the presence or absence of other factors, such as small non-coding RNAs (sRNA) which modulate both gene expression and transcript stability [6]. These sRNAs often require an RNA chaperone known as Hfq in order to anneal to their target mRNAs. Hfq was first identified as a host factor essential for the replication of RNA phage Qβ in *E. coli* [14]. Hfq is the bacterial branch of the "Sm" superfamily of proteins that are found in all three domains of life (Bacteria, Archaea, Eukarya); these proteins generally assemble into cyclic oligomers that associate with RNAs [15,16]. Hfq exhibits two distinct RNA-binding surfaces, commonly referred to as the "distal" and "proximal" faces, which interact with adenosine-rich and uracil-rich sequences, respectively [17]. Hfq has been shown to play an important role in RNA metabolic stability [18,19,20]. Specifically, Hfq can impede RNase E activity in the absence of sRNA by binding to a site on the target mRNA near the RNase E cleavage site; when the sRNA is present, Hfq facilitates annealing, either allowing for degradation by a nuclease within the double stranded region (i.e. RNase III) [21] or promoting cleavage by an RNase upstream of the regulatory region [22].

Here we investigate whether nanoRNAs, which we take to also include ~5-6 nt RNA species, interact with Hfq. We have discovered, and now identified, U/C-rich nanoRNAs that co-purify with *Thermotoga maritima* (*Tma*) Hfq when heterologously expressed in *E. coli*; these experiments made heavy use of liquid chromatography – tandem mass spectrometry (LC-MS/MS). In addition, we demonstrate, in carefully controlled *in vitro* experiments, that these

nanoRNAs interact with Hfq with nanomolar affinities. Notably, the affinities vary based on the position and frequency of cytosine in the pentauridine or hexauridine sequences.

Methods

Cloning and Expression of *Tma* Hfq. The coding sequence for *Tma* Hfq (geneID = 897578) was cloned into the pET-28b(+) expression vector (Novagen), between the *NdeI* and *XhoI* restriction sites, using the polymerase incomplete primer extension (PIPE) method [23]. The vector and insert were transformed into chemically competent *E. coli* TOP10 cells (Invitrogen) for the purpose of ligation. Recombinant plasmid was isolated and purified from TOP10 cells using the QIAprep Spin Miniprep Kit (QIAGEN) and successful cloning was verified by DNA sequencing (Genewiz). For an expression strain, purified plasmid was transformed into competent *E. coli* BL21 (DE3) cells that were subsequently grown on lysogeny broth supplemented with 50 μg/mL Kanamycin (LB-Kan) for selection of successfully transformed cells. The transformed cells from one colony were grown at 37 °C in LB-Kan and induced with 1 mM IPTG at an OD_{600} between 0.8 and 1.0 for 3.5 to 4 hrs. Cells were harvested by centrifugation at 15,000*g* for 5 min at 4 °C and cell pellets were stored overnight at -20 °C.

Purification of *Tma* Hfq. The cell pellet was subjected to a freeze/thaw cycle prior to resuspension in lysis buffer (50 mM Tris pH 7.50, 750 mM NaCl, 0.4 mM phenylmethylsulfonyl fluoride (PMSF), 0.01 μg/mL hen egg white lysozyme (Fisher)). Resuspended cells were incubated at 37 °C for 30 min followed by mechanical lysis using a microfluidizer (Microfluidics). Lysate was centrifuged at 35,000*g* for 20 min at 10 °C. As an initial purification step for the presumably thermostable [24,25] *Tma* Hfq, the clarified supernatant was incubated at 85 °C for 20 minutes. This was followed by centrifugation at 35,000*g* for 20 min at 10 °C. The protein sample was denatured by addition of guanidinium hydrochloride (Gnd-HCl) to a final

47

concentration of 6 M. The supernatant was then loaded onto a Ni^{2+}–charged affinity column (HiTrap) at room temperature (RT) using high pressure liquid chromatography (HPLC), followed by 10 column volumes of wash buffer (50 mM Tris pH 7.50, 150 mM NaCl, 6 M Gnd-HCl, 10 mM imidazole) and eluted with elution buffer (50 mM Tris pH 7.50, 150 mM NaCl, 6 M Gnd-HCl, 600 mM imidazole) using a 0-100% gradient over the course of five column volumes. The elution fractions were analyzed for purity and proper molecular weight (MW) with sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Gnd-HCl was removed via dialysis against 25 mM Tris pH 8.00, 1M arginine, and 0.2 mM PMSF overnight at RT. The sample was then dialyzed into digestion buffer (50 mM Tris pH 8.00, 150 mM NaCl, 12.5 mM ethylenediaminetetraacetic acid (EDTA) pH 8.00). Protein was digested with thrombin (1:600 molar ratio) overnight at 37 °C and cleavage was verified by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). After digestion, the recombinant *Tma* Hfq protein sample was subjected to a benzamidine column to remove thrombin. The protein was subsequently run on a Hiload 26/60 Superdex 75 prep-grade column with digestion buffer at a flow rate of 1 mL/min to ensure that the protein was folded and oligomerized.

Oligonucleotide purification. The above protein purification protocol was followed through the 85 °C heat-cut step. The resulting sample was loaded onto a Ni^{2+}–charged affinity column (HiTrap) using an HPLC, followed by 10 column volumes of wash buffer (50 mM Tris pH 7.50, 150 mM NaCl, 10 mM imidazole) and eluted with elution buffer (50 mM Tris pH 7.50, 150 mM NaCl, 600 mM imidazole) using a 0-100% gradient over the course of five column volumes. The elution fractions were analyzed for purity and proper MW with SDS-PAGE then dialyzed into digestion buffer (50 mM Tris pH 8.00, 150 mM NaCl, 12.5 mM EDTA pH 8.00). Protein was digested with thrombin (1:600 molar ratio) overnight at 37 °C, and the sample was then applied

to a benzamidine column to remove thrombin. The proteolyzed sample was then diluted with three volumes of buffer A (25 mM Tris pH 8.50, 50 mM NaCl) per volume of sample and loaded onto a quaternary amine anion exchange column (HiTrap). The column was washed with 10 column volumes of buffer A and eluted with buffer B (25 mM Tris pH 8.50, 2 M NaCl) using a three step gradient over 34 column volumes. Those fractions that eluted at 20% buffer B (440 mM NaCl) were combined and concentrated using a Millipore centrifugal filter unit with a 3350 Da molecular weight cut-off. The oligonucleotide components of the sample were then purified from recombinant *Tma* Hfq by phenol-chloroform extraction [26]. Then, an equal volume of 100% ethanol was added to the aqueous phase from the phenol-chloroform extraction and the resulting solution was incubated at -80 °C for at least 2 hours, followed by centrifugation at 9,300g (Eppendorf fixed-angle rotor) for 20 minutes at 4 °C. The supernatant from this step was removed and the remaining pellet was dried at RT then stored at -20 °C.

Colorimetric sugar assays. Dried samples from the previous step were resuspended in distilled and deionized water. The Benedict's assay was performed to test for free reducing sugar; 10 μL of Benedict's Reagent (943 mM anhydrous sodium carbonate, 588 mM sodium citrate tribasic dehydrate, 68.6 mM cupric sulfate pentahydrate) was added to 50 μL of sample in a PCR tube. The Bial's orcinol assay was performed to test for pentose sugar; 50 μL of orcinol reagent (24.2 mM orcinol monohydrate, 0.025% w/v ferric chloride hexahydrate, 6 M HCl) was added to a PCR tube containing 50 μL of sample. The Dische's diphenylamine assay was performed to test for deoxyribose; 50 μL of Dische's diphenylamine reagent (58.12 mM diphenylamine, 0.66% v/v ethanol, 11.4 M glacial acetic acid, 17.7 mM sulphuric acid) was added to a PCR tube followed by 50 μL of sample. The samples for each assay were sealed and incubated for 20 minutes in boiling water. These assays are further described in Patterson & Mura [27] and Appendix 1.

The controls used throughout these assays included 0.15 mg/mL ribose (Sigma), 7.5 mg/mL RNA from Baker's yeast (Sigma), 0.45 mg/mL calf thymus DNA (Sigma), 0.45 mg/mL bovine serum albumin (BSA, Sigma), and water.

Crosslinking assays. *Tma* Hfq was dialyzed into 25mM HEPES pH 8.00 and 200 mM NaCl prior to crosslinking with formaldehyde using an 'indirect' method [28]. Briefly, the experimental setup for this indirect method consisted of a Linbro plate with a microbridge and coverslip upon which 40 μL of the crosslinking agent and 15 μL of *Tma* Hfq were aliquoted, respectively. The chamber was sealed with vacuum grease. The concentration of crosslinking agent and *Tma* Hfq were 25% v/v and 1 mg/mL, respectively. For spiked samples, nucleic acid purified using the oligonucleotide purification above was added directly to the protein sample prior to crosslinking. The crosslinking agent was acidified with 122 mM HCl immediately before incubation. Samples were incubated at 37 °C for 40 minutes and the reaction was stopped by the addition of 5 μL 1 M Tris pH 8.00. Salts and crosslinking reagents were removed from samples using a C_4 zip tip (Millipore) [29], and then the crosslinking reactivities were assessed by MALDI-TOF MS.

MALDI-TOF MS. MALDI-TOF MS was performed on a Bruker Microflex MALDI instrument. Proteins of approximately 1 mg/mL were diluted 1:4 with 0.01% trifluoroacetic acid (TFA); samples purified using C_4 zip tips were not diluted. In each case the protein sample was spotted onto an MSP 96 target ground steel sample plate (Bruker Daltonics), with an equal volume of SA matrix (15 mg/mL sinapinic acid in 50% acetonitrile and 0.05% TFA) and allowed to air dry [30]. The instrument was calibrated by the close external method [31] using a series of low MW (Insulin, Cytochrome C, Ubiquitin I, and Myoglobin) or high MW (Protein A, Trypsinogen, Protein A, and Bovine Albumin) protein calibrants. Spectra were obtained by averaging approximately

50 laser shots with the following settings: positive ion detection mode, linear mode, grid voltage 40-75%, and an m/z range of 4,000-20,000 or 20,000-100,000 Da.

Oligonucleotide samples from the purification described above were spotted on the ground steel plate with an equal volume of THAP matrix (8.9 mg/mL 2',4',6'-trihydroxyacetophenone and 5.6 mg/mL diammonium citrate in 44% v/v acetonitrile) and allowed to air dry [32]. The MS instrument was calibrated with oligonucleotide calibration standard LMW (Bruker) by the close external method [31]. Spectra were obtained by averaging approximately 100 laser shots with the following settings: negative ion detection mode, linear mode, grid voltage 40-75%, and an m/z range of 180-2000 Da.

Nuclear magnetic resonance (NMR) spectroscopy of putative binding partners. Samples were prepared for NMR studies by dissolving approximately 857 μg of nucleic acid from the oligonucleotide purification above into 500 μL of 90%/10% D_2O/H_2O. The amount of nucleic acid binding partner in each sample was estimated based on absorbance at 260 nm, using the MW and extinction coefficient for pure 5'Phos-rUrUrUrCrU, which is 1547.9 g/mol and 46,300 $M^{-1}cm^{-1}$, respectively. All NMR spectra were recorded at 40 ˚C. ^{1}H spectra were recorded on a Bruker AVANCE spectrometer operating at a proton frequency of 600 MHz equipped with a Bruker 5 mm TXI cryoprobe. The ^{1}H data were acquired and processed using TopSpin 3.0 software package (Bruker Biospin, Rheinstetten, Germany). One-dimensional proton spectra were recorded with water suppression by excitation sculpting. ^{31}P spectra were recorded on Varian Mercury spectrometer operating at a ^{31}P frequency of 121.3 MHz. The data were processed using Vnmr 6.1c. All data were analyzed using MestReNova suite (Mestrelab Research, USA).

HPLC and mass spectrometry. Dried oligonucleotide samples from the purification described above were further purified using an Xbridge OST C18 column (10 mm × 50 mm, 2.5 μm

particles; Waters, Milford, MA, USA). Samples were resolved using 60 mL, 20 mL, and 40 mL of solvent A (0.1 mM triethylammonium acetate (TEAA) in water) and solvent B (0.1 mM TEAA in 93.25% water/ 6.75% acetonitrile) at 15%, 35%, and 100% solvent B, respectively. The system was run at 2.0 mL/min and absorbance was monitored at 260 nm.

High-resolution mass spectrometry and electrospray ionization-tandem mass spectrometry (ESI-MS/MS) of baseline resolved fractions were performed by oligonucleotide LC/MS analysis services (Novatia, LLC, Princeton, NJ).

Fluorescence polarization measurements. *Tma* Hfq was dialyzed into FP binding buffer (25mM Tris pH 8.00, 350 mM NaCl) prior to binding studies. Fluorescence polarization measurements were performed with a PHERAstar microplate reader. 5-Carboxyfluorescein labeled U_6 (FAM-U_6) RNA purchased from Integrated DNA Technologies was used to probe the proximal binding site of *Tma* Hfq for RNA binding. Samples were excited at 490 nm and emission was measured at 522 nm [33]. To determine the apparent equilibrium dissociation constant (K_d) for FAM-U_6, *Tma* Hfq was serially diluted in 96-well black polystyrene assay plates (Costar) in the presence of 5 nM FAM-U_6; the final volume in each well was 150 μL. The highest and lowest concentrations of Hfq monomer were 180 μM and 2.4 pM, respectively. The binding assay samples were incubated in the dark for 45 minutes at RT prior to measurements to ensure equilibrium binding. Data were collected in four independent replicates.

Competitive binding assay. The oligonucleotides used in this study include the pentamers 5'Phos-rUrUrUrUrU, 5'Phos-rUrUrUrCrU, 5'Phos-rUrUrUrUrC, and 5'Phos-rCrUrUrCrU and the hexamers 5'Phos-rUrUrUrUrCrU and 5'Phos-rCrUrUrCrUrU. Custom oligonucleotides were ordered from Integrated DNA Technologies; all oligonucleotides were purified by RNase-free HPLC after synthesis. Fluorescence polarization measurements were collected using the workflow described above. Samples contained 5 nM FAM-U_6 and 300 nM Hfq hexamer in FP

binding buffer. The highest and lowest concentrations of oligonucleotide used in all competitive

assays were 5 µM and 0.071 fM, respectively. Samples were mixed by pipetting and then

incubated in the dark for 45 minutes at RT prior to measurements. Data were collected in three

independent replicates.

Analysis of fluorescence polarization data. Fluorescence data were fit to a model that assumed

that a 1:1 complex formed between *Tma* Hfq$_6$ and FAM-U$_6$ under experimental conditions in

which Hfq exists only as a hexamer [15,34].   The model involves fitting the data to a sigmoidal

Boltzmann function, which is related to the Hill equation [35,36], and which can be rearranged

to read

$$y = \frac{(A_1 - A_2)}{1 + e^{\frac{(x - x_0)}{dx}}} + A_2 \qquad \text{(Equation 1)}$$

where x_0 is the inflection point sigmoidal curve, *dx* is the width of the transition and A$_1$ and A$_2$

are the fluorescence polarization intensities of the initial and final states, respectively

[15,37,38].   Nonlinear least-squares fits of the equation to the data were performed in

OriginPro7.5.

Results

Tma Hfq co-purifies with nucleic acids

Throughout our *Tma* Hfq recombinant protein purification efforts, spectrophotometric readings

indicated that the ratio of absorbance at 260 nm to 280 nm (A$_{260}$/A$_{280}$) was approximately 1.80;

this value is generally an indication that nucleobases are abundant in the sample [39].   To

further assess whether *Tma* Hfq was co-purifying with nucleic acid, samples were fractionated

on an anion exchange column. *Tma* Hfq has a predicted isoelectric point (pI) of 7.03 and the

chromatography buffer was at pH 8.50. The resulting chromatograms (Figure 2.1) exhibited

three 'peaks,' with A_{260}/A_{280} ratios of approximately 0.9-1.7, 1.7-1.9, and 2.0, in that order of elution.

Fractions from the first and second elution peaks were further analyzed by MALDI-TOF MS (Figure 2.2). These fractions were found to contain two peaks corresponding to the +1 and +2 charge stated of recombinant *Tma* Hfq protein. The third elution peak did not contain protein, at least within the detection limit of MALDI-TOF MS, thus implying a non-proteinaceous binding partner.

Operating under the assumption that the protein samples were co-purifying with nucleic acid, the co-purifying binding partners (BP) from peak 1 were isolated by phenol-chloroform extraction and then concentrated by ethanol precipitation. In order to determine the sugar moiety (e.g. ribose or deoxyribose) present in the co-purifying BPs, the isolated BP samples were subjected to a series of colorimetric assays to determine the nature of any sugar components. The binding partner did not produce a colored product for the Benedict's assay or the Dische diphenylamine assay (Figure 2.3a, c), whereas a distinctly blue product was observed for the Bial's orcinol assay (Figure 2.3b).

The isolated BP samples were further characterized by ^{1}H and ^{31}P NMR spectroscopy. The ^{1}H spectra (Figure 2.4a) contained peaks corresponding to both the sugar and base protons present in a ribonucleotide, but a specific base was not apparent due to the complexity of the spectra in the 7.5-8.5 ppm and 5-6 ppm regions. The complexity of the spectra and broadness of the peaks suggested that the BP sample contained multiple components. The ^{31}P spectra (Figure 2.4b) exhibited three peaks that can be ascribed to a phosphodiester at ~0 ppm, phosphate monoester at ~4 ppm, and phosphonate at ~21 ppm.

Identification of Hfq-binding nanoRNAs

In order to elucidate the chemical structure of the oligonucleotides in the co-eluting sample, the components were fractionated using an Xbridge OST C18 column. The chromatogram indicated that the BP sample contained more than 20 components (Figure 2.5).

Baseline resolved fractions from the HPLC column were analyzed by MS to determine their MW and sequence. The MWs of the intact BPs were determined using high-resolution mass spectrometry (table 2.1). The oligonucleotide sequences of the fractionated BPs were determined by LC-MS/MS (Table 2.1). All six identified oligonucleotides were found to be either 5 or 6 nts long with a 5' monophosphate and 3' hydroxyl.

Binding partners can be cross-linked to the Hfq hexamer

Using formaldehyde as a crosslinker, Hfq oligomers were chemically linked via an indirect method relying on vapor diffusion as a means to slowly introduce the crosslinking agent. During analysis, the MALDI-TOF MS spectra of formaldehyde-crosslinked *Tma* Hfq (black spectra in Figure 2.6) contained two peaks in the 64 kDa to 68 kDa region of the MS spectra that were not baseline resolved. The peak at ~ 65 kDa is consistent with the expected MW for a *Tma* Hfq hexamer, whereas the peak at ~ 67 kDa is inconsistent with an oligomeric state of *Tma* Hfq.

Operating under the hypothesis that the peak around 67 kDa is the *Tma* Hfq hexamer crosslinked to co-purifying RNA, *Tma* Hfq was spiked with fractions (f) 13 and 23 from the HPLC purification prior to crosslinking with formaldehyde (Figure 2.5). The relative intensity of the peak at 67 kDa was increased in the spectra of *Tma* Hfq spiked with f13 (Figure 2.6a). The spectra of *Tma* Hfq spiked with f23 had 3 peaks that are not baseline resolved in the 66 kDa to 77 kDa region of the spectra (Figure 2.6).

Hfq binds to nanoRNAs with high affinity

The binding properties of FAM-U$_6$ and *Tma* Hfq were determined by measuring fluorescence polarization at various *Tma* Hfq concentrations (Figure 2.7). The apparent K_d was determined to be 61 ± 10 nM (in terms of the Hfq monomer) by fitting the data to the modified Hill's equation given by Equation 1.

The concentration of the Hfq-binding nanoRNAs required to displace 50% of FAM-U$_6$ (IC50) was determined by measuring fluorescence polarization at varying nanoRNA concentrations (Figure 2.7). The inhibition constant for each nanoRNA (Table 2.1) was calculated using the equation,

$$K_i = \frac{IC50}{1+\frac{[S]}{Kd}} \qquad \text{(Equation 2)}$$

where [S] is the concentration of FAM-U$_6$ and K_d is the apparent binding constant for FAM-U$_6$ under experimental conditions. The binding properties of each of the nanoRNAs characterized in these studies are summarized in Table 2.1.

Discussion

Previous studies have shown that Hfq preferentially interacts with A/U-rich RNA sequences [40,41,42,43,44]. Crystal structures have revealed that A-rich and U-rich sequences bind on opposite sides of the disc-shaped Hfq hexamer, referred to as the distal and proximal faces, respectively [15,45]. In this study, we have discovered and identified U/C-rich nanoRNA sequences, 5 to 6 nucleotides long that co-purify with *Tma* Hfq when heterologously expressed in *E. coli*.

A high A$_{260}$/A$_{280}$ ratio during spectrophotometric measurements of the protein sample indicated that Hfq was purifying with nucleic acid. The measured A$_{260}$/A$_{280}$ ratio throughout

purification was approximately 1.80; accepted A_{260}/A_{280} ratio for "pure" DNA, RNA, and protein are 1.8, 2, and 0.57, respectively [39,46,47,48]. A series of colorimetric sugar assay were performed in order to determine the sugar component of the nucleic acid [27]. A colored product was not observed in the Benedict's reaction, indicating that the sample does not contain free reducing sugar (Figure 2.3a). The production of a blue-green product in the Bial's orcinol reaction identified the presence of a pentose sugar (figure 2.3b); no colored product was observed in the Dische's diphenylamine reaction indicating that the pentose sugar is not deoxyribose (Figure 2.3c). These data further support that the high-A_{260} specimen(s) co-purifying with *Tma* Hfq are ribonucleic acids.

The predominant phosphodiester peak at 0 ppm in the ^{31}P NMR spectra indicated that the nucleic acids in the sample were oligonucleotides (Figure 2.4b). Two additional peaks were observed in the ^{31}P NMR spectra at ~4 ppm and ~21 ppm can be attributed to the presence of phosphate monoester and either phosphonate or cyclic phosphate functional groups in the sample, respectively. Phosphate monoesters are commonly found on the 5' end of ribonucleic acids that are being targeted for RNA metabolism by cellular nuclease. The 5' end of RNA is a determinant of the transcripts stability since RNase E as well as other nucleases preferentially bind sequences with a 5' monophosphate [7,49,50].

The ^{1}H NMR spectra verified the presence of nucleotides in the sample, but was inconclusive in identifying the bases present (Figure 2.4a). Peaks present in the 3.8-4.2 ppm region of the ^{1}H NMR spectra can be attribute to the hydrogens at the C2', C3', C4' and C5' positions in ribose [51]. Peaks present in the 5-6 ppm region of the ^{1}H NMR spectra can be attributed to the hydrogen at the C1' position in ribose as well as possible contributions from hydrogen in the C3' position of a pyrimidine base (e.g. uracil or cytosine) [51]. Peaks in the 7.5-8.5 ppm region of the ^{1}H NMR spectra can be attributed to additional contributes form the

nucleotide bases present in the sample although a specific base cannot be specified since hydrogens from all four bases are expected in this regions [51]. The overall complexity of the ^{1}H NMR spectra and broadness of the peaks suggested that the sample contained multiple chemical components.

The binding partners were ultimately successfully separated by HPLC using an Xbridge OST C18 column (Figure 2.5); the molecular weight and sequences of the oligonucleotides in the baseline resolved HPLC fractions were determined using high-resolution mass spectrometry and LC-MS/MS, respectively (Table 2.1). All six oligonucleotides feature a 5' monophosphate and a 3' hydroxyl, suggesting that this end chemistry is conductive to *Tma* Hfq binding. This is consistent with the high resolution crystal structure of C-terminally truncated *Salmonella typhimurium* Hfq (*Sty* Hfq) bound to U_6 RNA, which suggested that a hydrogen bond could form between H57 and the 3' hydroxyl of RNA thereby stabilizing the interaction [40]. The H57 position is highly conserved among bacterial Hfq homologs including *Tma* Hfq suggesting that this mechanism of 3' hydroxyl RNA recognition may be conserved (Figure 2.9). The importance of a 3' hydroxyl end was further supported by *in vitro* binding studies that indicated that the addition of a propyl-phosphate group to the 3' end of U_6 RNA significantly hindered binding with *Sty* Hfq [40].   In addition, RNA termination with 2'-3' cyclic phosphate was also shown to impede RNA binding at the proximal site of *Sty* Hfq [40]. A direct interaction between the 3' end of RNA and Hfq could protecting the RNA from degradation by cellular nucleases as has been previously suggested [52], thereby regulate RNA turnover. Further studies could focus on systematically examining the effects of other end chemistries (e.g. 5' hydroxyl or 3' phosphonate) on the affinity of nanoRNAs for *Tma* Hfq to determine if 3' end chemistry selection is a general feature of bacterial Hfq.

The two peaks in the 64 kDa to 68 kDa region of the MALDI-TOF MS spectrum of formaldehyde-crosslinked *Tma* Hfq in the presence of BP are consistent with Hfq in its hexamer form in the absence and presence of nucleic acid (Figure 2.6, black trace). To verify that the binding partners that were isolated by HPLC are the components of the sample that were crosslinking with the Hfq hexamer, a *Tma* Hfq sample stripped of nucleic acid via the protein purification described above was spiked with the HPLC-purified fractions prior to crosslinking the sample with formaldehyde. The relative intensity of the peak at 67 kDa did increase in the MALDI-TOF MS spectrum of formaldehyde-crosslinked *Tma* Hfq spiked with f13 (Figure 2.6a); these data further supports a 1:1 binding stoichiometry between *Tma* Hfq_6 and CU_2CU (f13). Similar results were observed for HPLC-purified fractions f17, f19, f21, and f27 (same peak as f28). Interestingly, the spectrum of formaldehyde-crosslinked *Tma* Hfq spiked with f23 exhibited 3 peaks that are not baseline resolved in the 66 kDa to 77 kDa region of the spectra, suggesting that the binding stoichiometry of *Tma* Hfq_6 to U_5 can be 1:1, 1:2, or 1:3 (Figure 2.6b).

For carefully controlled *in vitro* binding studies, the nanoRNAs that we identified (as *in vivo* binding partners) were chemically synthesized and purified, insuring well-defined nucleotide sequences and end-chemistries. The IC50 and K_i for these nanoRNAs are reported in Table 2.1. The pattern of variance in the binding affinities suggests that Hfq preferentially interacts with uracil, as is known from many previous *in vitro* studies [40,53,54,55,56]. More specifically, the data indicate that a uracil base at the 3' end of the nanoRNA is important for high affinity interactions with Hfq, while a cytosine in the fifth position of the nanoRNA significantly decreases the affinity between *Tma* Hfq and the nanoRNA.

These findings are consistent with previous structural and binding studies of the proximal RNA-binding site of Hfq [15,40,57]. High resolution crystal structures of *Staphylococcus aureus* Hfq (Sau Hfq) with AU_5G RNA (PDB 1KQ2), *Sty* Hfq with U_6 RNA (PDB

2YLC), and *Escherichia coli Hfq* (*Eco* Hfq) with A_7 and AU_6A RNA (PDB 4HT9) indicated that the amino acid residues Q8, K/Q41, F/Y42, K56, and H57 (number based on the *Eco* Hfq protein sequence) present in the proximal RNA-binding site are directly involved in nucleotide-binding and selection [15,40,57]. A sequence alignment of *Tma* Hfq, which the protein sequences for *Sau* Hfq, *Sty* Hfq, and *Eco* Hfq, indicates that the proximal RNA binding site is conserved with the exception of a single amino variance of Q41S in *Tma* Hfq (Figure 2.9).

Furthermore, the binding properties of the nanoRNAs reported here are consistent with previous work which examined single nucleotides substitutions in a hexauridine substrate at either the first or sixth position, and concluded that cytosine in the sixth position has little to no effect on the binding affinity of Hfq [40]. Specifically, the K_d of *Sty* Hfq with U_6 and U_5C were shown to be 13 ± 2 nM and 20 ± 3 nM, respectively. Nanomolar binding affinities were also observed for *Eco* Hfq and *Sau* Hfq with U-rich oligonucleotides as well; specifically the K_d for *Eco* Hfq and *Sau* Hfq with U_6 were shown to be 766.0 ± 73.7 nM and 69.8 ± 7.0 nM, respectively [54,58].

The crystal structure of *Sau* Hfq with AU_5G RNA (PDB 1KQ2) suggested that the nucleotide-binding specificity of the proximal RNA binding site of Hfq was for uracil [15]. Additionally, it was proposed that together K41 and K47 discriminate against cytosine-binding, whereas Q8 sterically hinders guanine-binding [15]. Interestingly, both *Tma* Hfq and *Sty* Hfq have an amino acid variance in the K41 position of K41S and K41Q, respectively (Figure 2.9). Future studies could focus on high-resolution crystal structure of *Tma* Hfq with the nanoRNAs reported here to determine the amino acid residues involved in their recognition. In addition, future studies could characterize the binding properties of a *Tma* Hfq S43K mutant to determine if lysine in this position discriminates against cytosine-binding.

Based on the findings reported here, we suggest that a novel class of U/C-rich nanoRNAs can bind to *Tma* Hfq with high (nanomolar) affinity. These findings have a potential impact on the turn-over and regulation of nanoRNAs, which have recently been shown to prime RNA transcription in *E. coli* and thereby control 5' transcriptional start site selection in a growth phase dependent manner [11]. Future studies could use RNA-primed quantitative real-time polymerase chain reaction in the presence and absence of Hfq to explore the effect of Hfq on nanoRNA-priming.

It is widely believed that nanoRNAs are produced from a number of RNA degradation pathways, which Hfq has been linked to through its role in regulating sRNA and mRNA stability [45,59,60,61]; Hfq has been explicitly implicated in interactions with nucleases, such as RNase E [18,19,21,22,62]. Ultimately, nanoRNAs are hydrolyzed into mononucleotides by oligoribonuclease [8,63]. Future studies could focus on characterizing the effect of Hfq on the activity of oligoribonuclease.

In addition to the possible role of Hfq in nanoRNA metabolism and transcriptional start site selection, the results reported here may have implication on Hfq-mediated sRNA-mRNA annealing. Cellular nanoRNAs could potentially regulated sRNA binding to the proximal binding site of Hfq by directly compete with the sRNA. Direct competition between a sRNA and nanoRNAs would lower the apparent K_d of Hfq for the sRNA and thereby increasing the rate of sRNA-mRNA annealing. This role of nanoRNAs in Hfq-mediated sRNA-mRNA annealing proposed here is consistent with a recently proposed mechanism that suggests that RNAs are cycled through the proximal binding site of Hfq by competing RNAs, which facilitate the rapid sRNA-mRNA annealing observed *in vivo* [64]. Future studies could explore the role of nanoRNAs in post-transcriptional regulation by sRNA through a systematic *in vivo* study that examines the effect of nanoRNA depletion and accumulation on Hfq-mediate and Hfq-independent post-

transcriptional regulation by sRNAs. Additionally, future studies could use biophysical *in vitro* studies to carefully characterize the effect of nanoRNAs on Hfq-mediated sRNA-mRNA annealing.

References

1. Gripenland J, Netterling S, Loh E, Tiensuu T, Toledo-Arana A, et al. (2010) RNAs: regulators of bacterial virulence. Nat Rev Microbiol 8: 857-866.

2. Silva IJ, Saramago M, Dressaire C, Domingues S, Viegas SC, et al. (2011) Importance and key events of prokaryotic RNA decay: the ultimate fate of an RNA molecule. Wiley Interdiscip Rev RNA 2: 818-836.

3. Shemesh M, Tam A, Steinberg D (2007) Expression of biofilm-associated genes of Streptococcus mutans in response to glucose and sucrose. J Med Microbiol 56: 1528-1535.

4. Rutherford ST, Bassler BL (2012) Bacterial quorum sensing: its role in virulence and possibilities for its control. Cold Spring Harb Perspect Med 2.

5. Segal G, Ron EZ (1998) Regulation of heat-shock response in bacteria. Ann N Y Acad Sci 851: 147-151.

6. Lalaouna D, Simoneau-Roy M, Lafontaine D, Masse E (2013) Regulatory RNAs and target mRNA decay in prokaryotes. Biochim Biophys Acta 1829: 742-747.

7. Evguenieva-Hackenberg E, Klug G (2011) New aspects of RNA processing in prokaryotes. Curr Opin Microbiol 14: 587-592.

8. Ghosh S, Deutscher MP (1999) Oligoribonuclease is an essential component of the mRNA decay pathway. Proc Natl Acad Sci U S A 96: 4372-4377.

9. Mechold U, Fang G, Ngo S, Ogryzko V, Danchin A (2007) YtqI from Bacillus subtilis has both oligoribonuclease and pAp-phosphatase activity. Nucleic Acids Res 35: 4552-4561.

10. Nickels BE, Dove SL (2011) NanoRNAs: A Class of Small RNAs That Can Prime Transcription Initiation in Bacteria. Journal of Molecular Biology 412: 772-781.

11. Vvedenskaya IO, Sharp JS, Goldman SR, Kanabar PN, Livny J, et al. (2012) Growth phase-dependent control of transcription start site selection and gene expression by nanoRNAs. Genes & Development: 1498-1507.

12. Goldman SR, Sharp JS, Vvedenskaya IO, Livny J, Dove SL, et al. (2011) NanoRNAs Prime Transcription Initiation In Vivo. Molecular Cell 42: 817-825.

13. Turner DH (2013) Fundamental interactions in RNA: Questions answered and remaining. Biopolymers 99: 1097-1104.

14. Franze de Fernandez M, August JT, Hayward WS (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: 824-831.

15. Schumacher MA, Pearson RF, Moller T, Valentin-Hansen P, Brennan RG (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. Embo Journal 21: 3546-3556.

16. Mura C, Randolph PS, Patterson J, Cozen AE (2013) Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function. RNA Biol 10: 636-651.

17. Mikulecky PJ, Kaw MK, Brescia CC, Takach JC, Sledjeski DD, et al. (2004) Escherichia coli Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. Nat Struct Mol Biol 11: 1206-1214.

18. Morita T, 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.

19. Ikeda Y, Yagi M, Morita T, 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.

20. Jain K, Updegrove TB, Wartell RM (2011) A Thermodynamic Perspective of sRNA-mRNA Interactions and the Role of Hfq. ACS Symposium Series. pp. 111-131.

21. Afonyushkin T, Vecerek B, Moll I, Blasi U, Kaberdin VR (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.

22. Bandyra KJ, Said N, Pfeiffer V, Gorna MW, Vogel J, et al. (2012) The Seed Region of a Small RNA Drives the Controlled Destruction of the Target mRNA by the Endoribonuclease RNase E. Mol Cell 47: 943-953.

23. Klock HE, Koesema EJ, Knuth MW, Lesley SA (2008) Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. Proteins 71: 982-994.

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

25. Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. Nat Rev Microbiol 9: 578-589.

26. Chomczynski P, Sacchi N (2006) The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. Nat Protoc 1: 581-585.

27. Patterson J, Mura C (2012) Rapid Colorimetric Assays to Qualitatively Distinguish RNA and DNA in Biomolecular Samples. Journal of Visualized Experiments.

28. Fadouloglou VE, Kokkinidis M, Glykos NM (2008) Determination of protein oligomerization state: two approaches based on glutaraldehyde crosslinking. Anal Biochem 373: 404-406.

29. Warren ME, Brockman AH, Orlando R (1998) On-probe solid-phase extraction/MALDI-MS using ion-pairing interactions for the cleanup of peptides and proteins. Anal Chem 70: 3757-3761.

30. Frecklington D (2007) General Method for MALDI-MS Analysis of Proteins and Peptides. CSH Protoc 2007: pdb prot4679.

31. Fenyo D, Wang Q, DeGrasse JA, Padovan JC, Cadene M, et al. (2007) MALDI sample preparation: the ultra thin layer method. J Vis Exp: 192.

32. Lavanant H, Lange C (2002) Sodium-tolerant matrix for matrix-assisted laser desorption/ionization mass spectrometry and post-source decay of oligonucleotides. Rapid Commun Mass Spectrom 16: 1928-1933.

33. Oefner PJ, Huber CG, Umlauft F, Berti GN, Stimpfl E, et al. (1994) High-resolution liquid chromatography of fluorescent dye-labeled nucleic acids. Anal Biochem 223: 39-46.

34. Sun X, Wartell RM (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.

35. Gesztelyi R, Zsuga J, Kemeny-Beke A, Varga B, Juhasz B, et al. (2012) The Hill equation and the origin of quantitative pharmacology. Arch Hist Exact Sci 66: 427-438.

36. Seber GAF, Wild CF (1989) Nonlinear Regression: John Wiley & Sons, Inc.

37. Hunke C, Antosch M, Muller V, Gruber G (2011) Binding of subunit E into the A-B interface of the A(1)A(O) ATP synthase. Biochim Biophys Acta 1808: 2111-2118.

38. Hanes MS, Ratcliff K, Marqusee S, Handel TM (2010) Protein-protein binding affinities by pulse proteolysis: application to TEM-1/BLIP protein complexes. Protein Sci 19: 1996-2000.

39. Warburg O, Christian W (1942) Isolation and crystallization of enolase. Biochem Z: 384-421.

40. Sauer E, Weichenrieder O (2011) Structural basis for RNA 3 '-end recognition by Hfq. Proceedings of the National Academy of Sciences of the United States of America 108: 13065-13070.

41. Geissmann TA, Touati D (2004) Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. Embo Journal 23: 396-405.

42. Brescia CC, Mikulecky PJ, Feig AL, Sledjeski DD (2003) Identification of the Hfq-binding site on DsrA RNA: Hfq binds without altering DsrA secondary structure. Rna-a Publication of the Rna Society 9: 33-43.

43. Vecerek B, Moll I, Blasi U (2005) Translational autocontrol of the Escherichia coli hfq RNA chaperone gene. Rna-a Publication of the Rna Society 11: 976-984.

44. Moll I, Leitsch D, Steinhauser T, Blasi U (2003) RNA chaperone activity of the Sm-like Hfq protein. Embo Reports 4: 284-289.

45. Link TM, Valentin-Hansen P, Brennan RG (2009) Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. Proceedings of the National Academy of Sciences of the United States of America 106: 19286-19291.

46. Chomczynski P, Mackey K, Drews R, Wilfinger W (1997) DNAzol: a reagent for the rapid isolation of genomic DNA. Biotechniques 22: 550-553.

47. Manning K (1991) Isolation of nucleic acids from plants by differential solvent precipitation. Anal Biochem 195: 45-50.

48. Gallagher SR (2011) Quantitation of DNA and RNA with absorption and fluorescence spectroscopy. Curr Protoc Neurosci Appendix 1: Appendix 1K.

49. Deana A, Celesnik H, Belasco JG (2008) The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. Nature 451: 355-358.

50. Celesnik H, Deana A, Belasco JG (2007) Initiation of RNA decay in Escherichia coli by 5' pyrophosphate removal. Mol Cell 27: 79-90.

51. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, et al. (2013) HMDB 3.0--The Human Metabolome Database in 2013. Nucleic Acids Res 41: D801-807.

52. Sauer E, Schmidt S, Weichenrieder O (2012) Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. Proc Natl Acad Sci U S A 109: 9396-9401.

53. Wang W, Wang L, Zou Y, Zhang J, Gong Q, et al. (2011) Cooperation of Escherichia coli Hfq hexamers in DsrA binding. Genes Dev 25: 2106-2117.

54. Robinson KE, Orans J, Kovach AR, Link TM, Brennan RG (2014) Mapping Hfq-RNA interaction surfaces using tryptophan fluorescence quenching. Nucleic Acids Res 42: 2736-2749.

55. Ishikawa H, Otaka H, Maki K, Morita T, Aiba H (2012) The functional Hfq-binding module of bacterial sRNAs consists of a double or single hairpin preceded by a U-rich sequence and followed by a 3' poly(U) tail. RNA 18: 1062-1074.

56. Otaka H, Ishikawa H, Morita T, 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.

57. Wang W, Wang L, Wu J, Gong Q, Shi Y (2013) Hfq-bridged ternary complex is important for translation activation of rpoS by DsrA. Nucleic Acids Res 41: 5938-5948.

58. Horstmann N, Orans J, Valentin-Hansen P, Shelburne SA, 3rd, Brennan RG (2012) Structural mechanism of Staphylococcus aureus Hfq binding to an RNA A-tract. Nucleic Acids Res 40: 11023-11035.

59. Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, et al. (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. PLoS Genet 4: e1000163.

60. Someya T, Baba S, Fujimoto M, Kawai G, Kumasaka T, et al. (2012) 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 40: 1856-1867.

61. Papenfort K, Bouvier M, Mika F, Sharma CM, 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.

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

63. Yu D, Deutscher MP (1995) Oligoribonuclease is distinct from the other known exoribonucleases of Escherichia coli. J Bacteriol 177: 4137-4139.

64. Wagner EG (2013) Cycling of RNAs on Hfq. RNA Biol 10: 619-626.
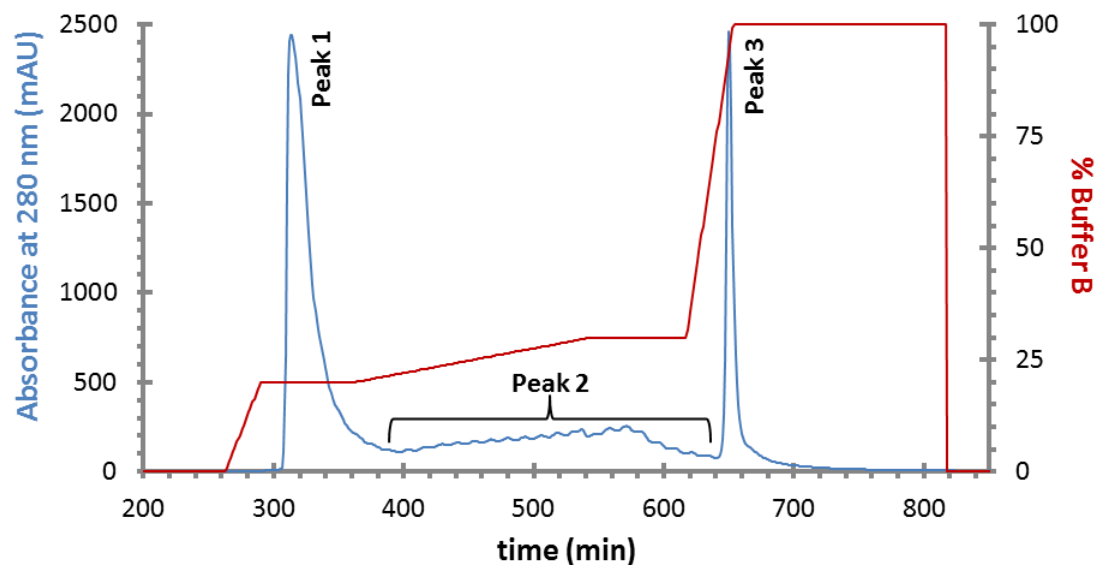
Figure 2.1 Representative chromatogram of *Tma* Hfq being separated from a subset of Hfq-associated RNAs, via a quaternary amine anion-exchange column. The absorbance trace (**blue**) indicates two distinct peaks eluting at different salt concentrations, which increases with buffer B (**red**). An intermediate "peak" is observed that is extremely broad and rippled. The two discrete peaks, which elute at approximately 20% and 100% buffer B, correspond to Hfq bound to oligonucleotides (peak 1) and isolate RNA species (peak 3), respectively.
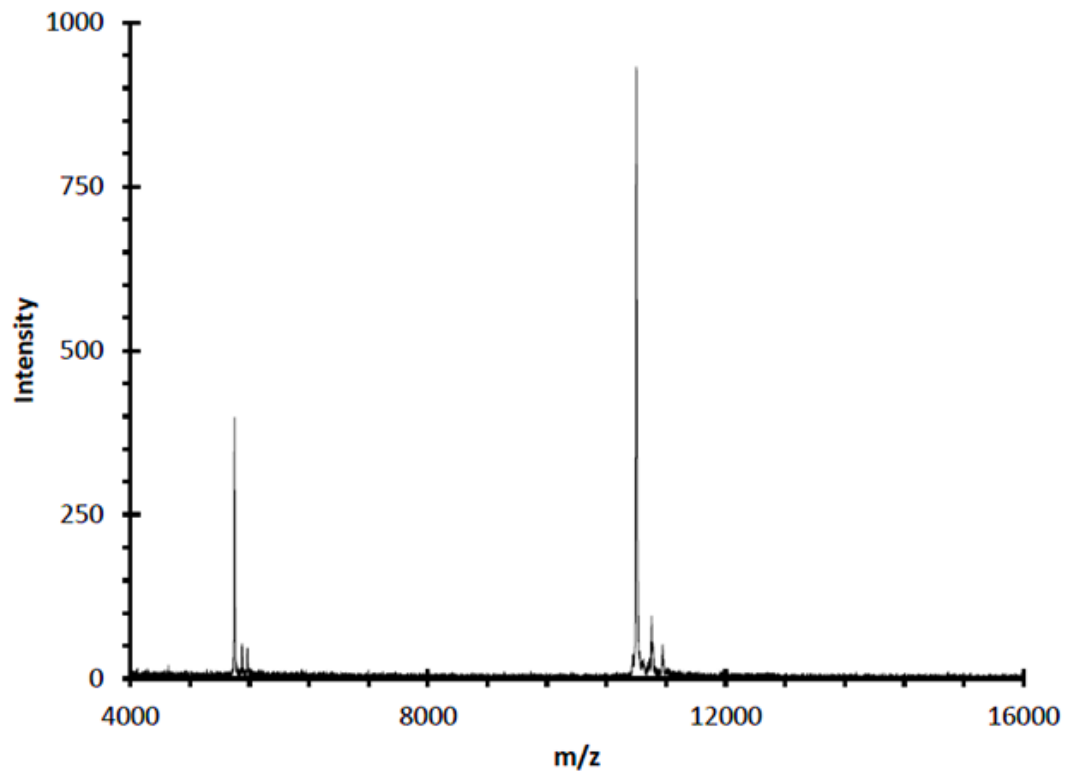
Figure 2.2 MALDI-TOF MS spectra of purified recombinant _Tma_ Hfq. The two peaks correspond

to the +1 and +2 charge states of a recombinant Hfq monomer. The $z=+1$ charge state is within

0.01% of the expected MW for recombinant _Tma_ Hfq based on its amino acid sequence.
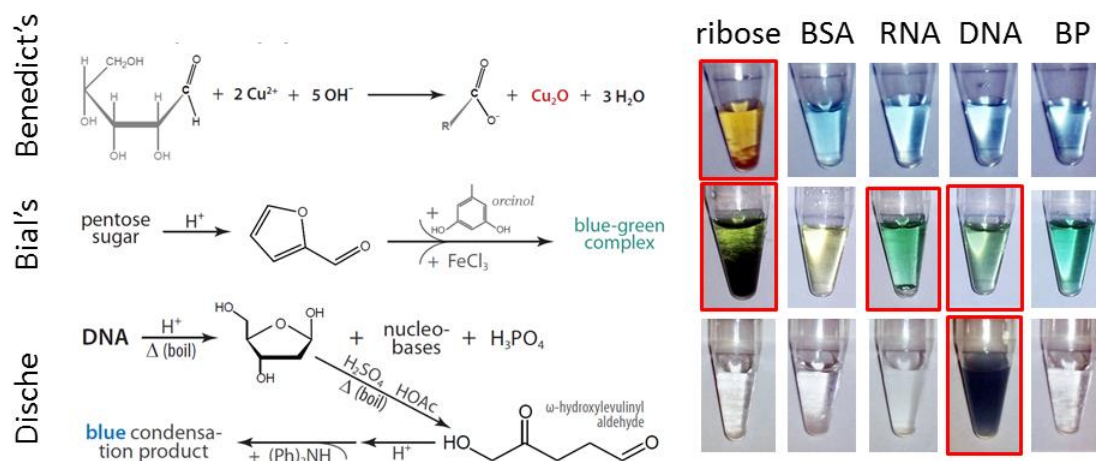
Figure 2.3 Colorimetric assays indicate that the binding partner contains a pentose sugar. The Benedict's reaction shows that the sample does not contain free reducing sugar. The Bial's reaction indicates the presence of a pentose sugar, and the Dische's diphenylamine reaction indicates that the pentose sugar is not deoxyribose. The reactions utilized in each assay are shown to the left of the results. The controls used throughout these assays include 0.15 mg/mL ribose, 7.5 mg/mL RNA from Baker's yeast, 0.45 mg/mL DNA from calf thymus, 0.45 mg/mL bovine serum albumin (BSA), and water. Positive controls for each assay are highlighted in **red** and the reactions containing binding partner are labeled "BP".
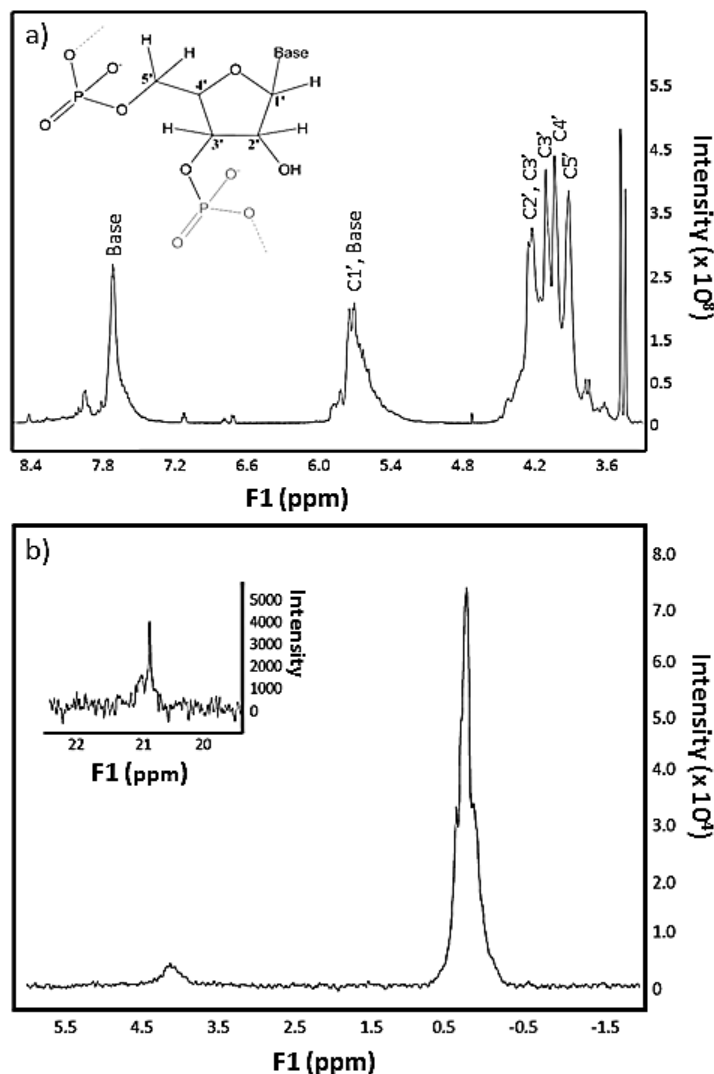
Figure 2.4 ^{1}H spectrum (a) and ^{31}P spectrum (b) of binding partner in D$_2$O. The peaks in the ^{1}H NMR spectrum (a) are consistent with a sample containing a complex mixture of nucleotides or oligonucleotides. Peaks are assigned using the labeling scheme shown on the ribonucleotide structure inset into the spectra. The complexity of the spectra in regions 7.5-8.5 ppm and 5-6 ppm impede the identification of the base or bases present in the BP sample [51]. The ^{31}P NMR spectrum (b) has three peaks that can be attributed to phosphate diester at ~0 ppm, phosphate monoester at ~4 ppm, and phosphonate or cyclic phosphate at ~21 ppm. The third peak at ~21 ppm is shown in the insert due to its significantly lower intensity compared to the other two peaks.
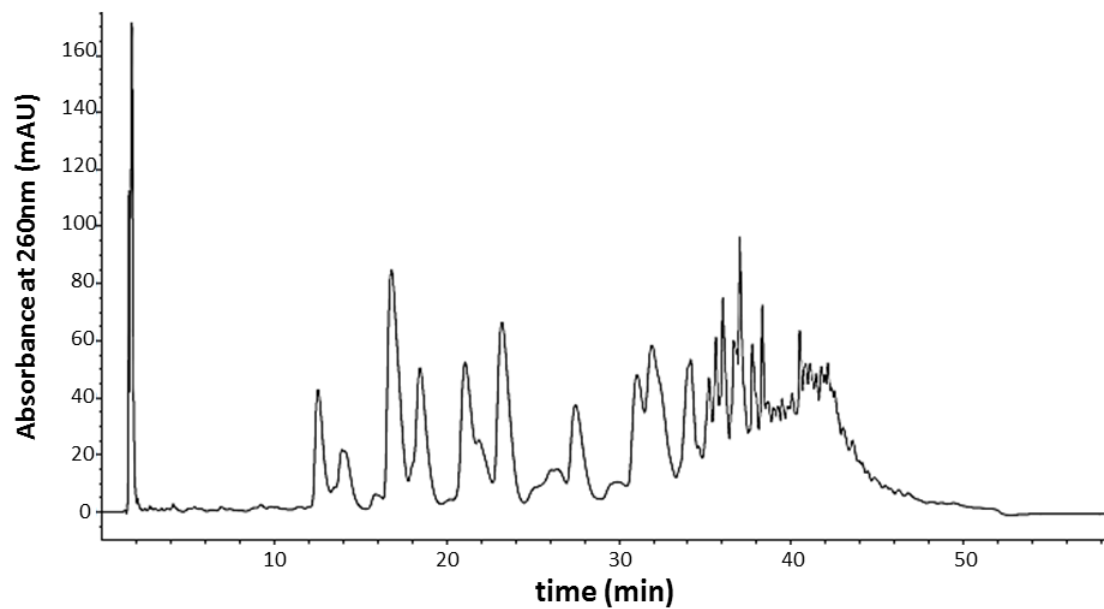
Figure 2.5 Representative chromatogram of co-purifying BP samples fractionated using an Xbridge OST C18 column. The absorbance trace indicates that the BP samples contain multiple RNA components.
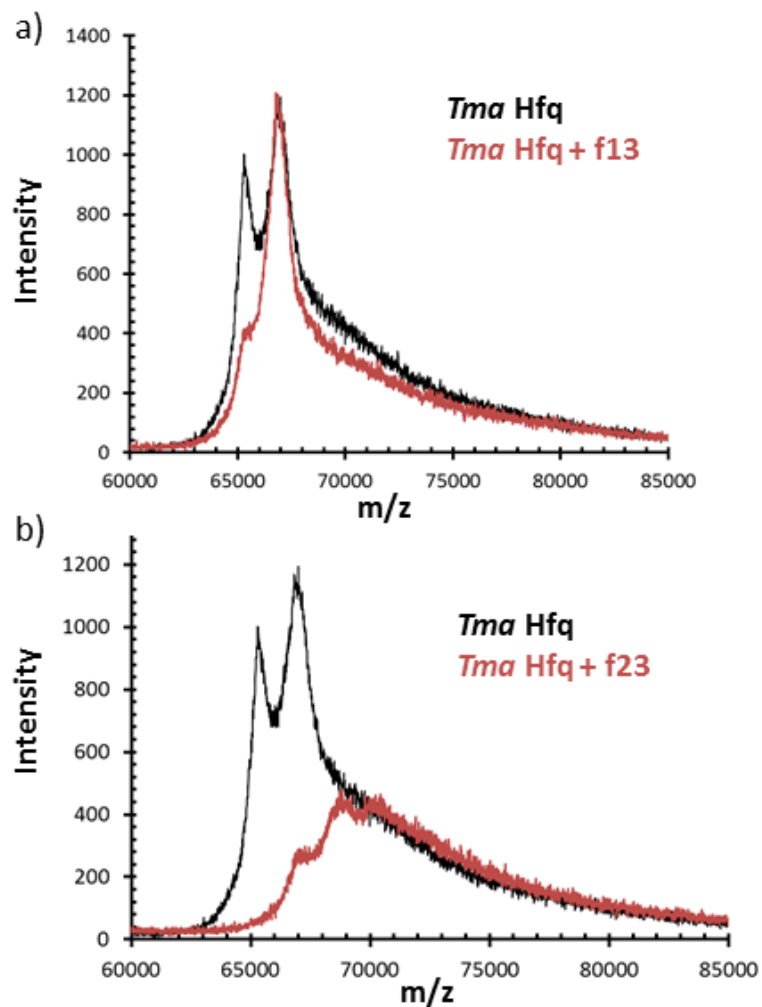
Figure 2.6 MALDI-TOF MS Spectra of *Tma* Hfq crosslinked indirectly with formaldehyde. The spectra shown in **black** are of *Tma* Hfq purified under non-denaturing conditions. The peak around 65 kDa is consistent with the expected MW for a *Tma* Hfq hexamer, whereas the peak around 67 kDa is not consistent with an oligomeric state of *Tma* Hfq. In part (a), Hfq was spiked with f13 (**red**), CU_2CU, indicated a 1:1 stoichiometry between *Tma* Hfq and CU_2CU. In part (b), Hfq spiked with f23 (**red**), U_5, suggests that stoichiometry of Hfq to U_5 can be 1:1, 1:2, or 1:3. The shift in MW upon addition of isolated binding partner suggests that the peak at 67 kDa is due to crosslinking the *Tma* Hfq hexamer to oligonucleotides that co-purify with the protein.
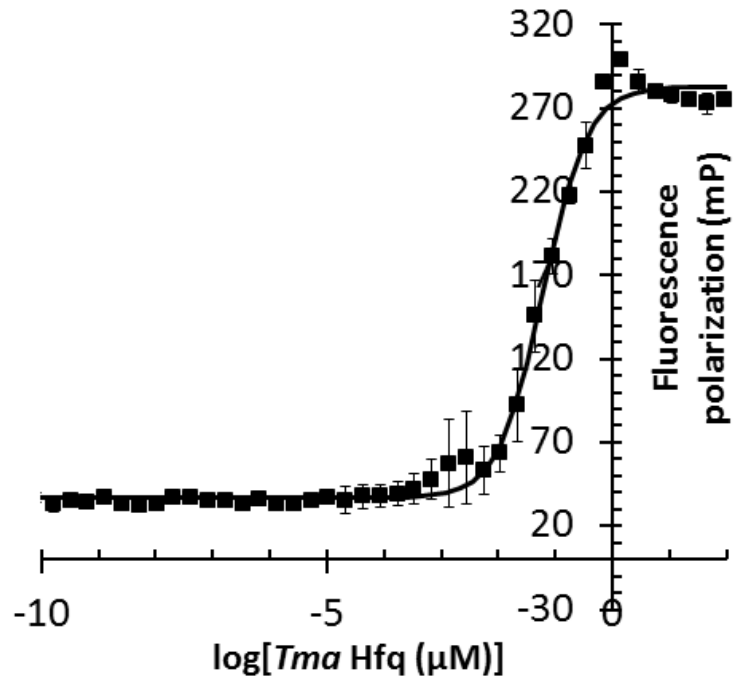
Figure 2.7 Binding curve for *Tma* Hfq with FAM-U$_6$. Binding was measured by fluorescence polarization in 25 mM Tris pH 8.00 and 350 mM NaCl at RT, using 5 nM FAM-U$_6$ and 180 µM – 2.4 pM *Tma* Hfq monomer. Fluorescence polarization from four independent experiments were averaged and plotted against Hfq concentration. The data was fit with a modified Hill's equation (Equation 1) yielding a K_d of 61 ± 10 nM.
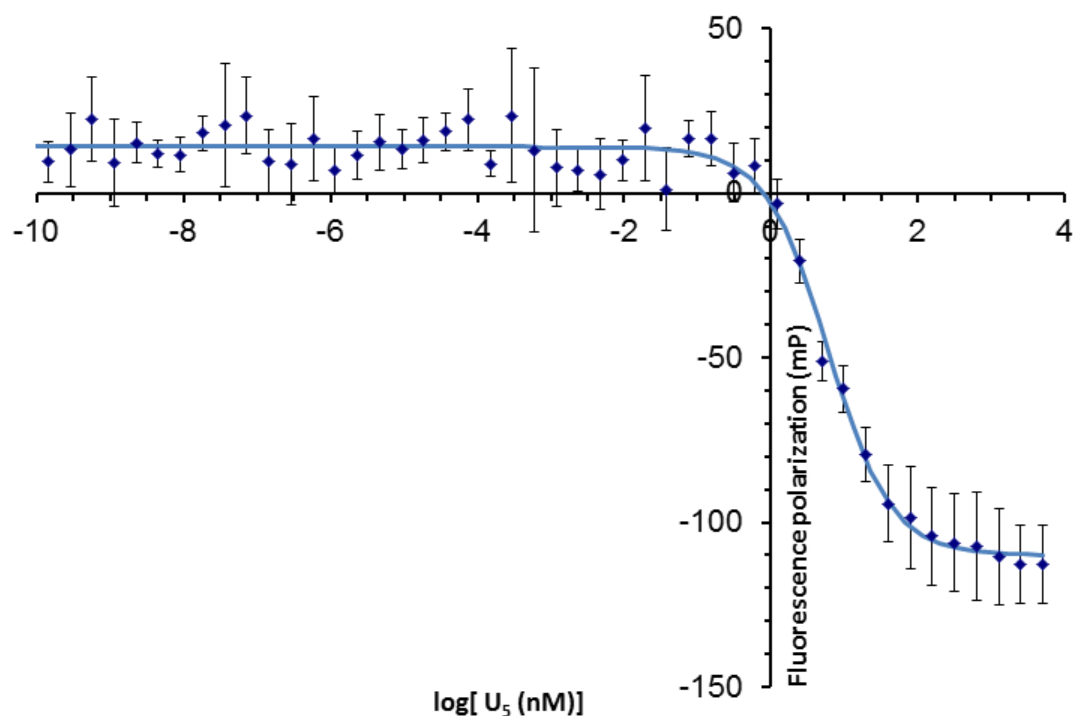
Figure 2.8 Binding curve for Tma Hfq with FAM-U_6 and 5'Phos-rUrUrUrUrU (U_5). Binding was measured by fluorescence polarization in 25 mM Tris pH 8.00 and 350 mM NaCl at RT, using 5 nM FAM-U_6, 300 nM Hfq monomer, and 5 μM – 71 pM U_5. Fluorescence polarization from four independent experiments were averaged and plotted against U_5 concentration. The data was fit with a modified Hill's equation (Equation 1) yielding an IC50 of 6 nM.
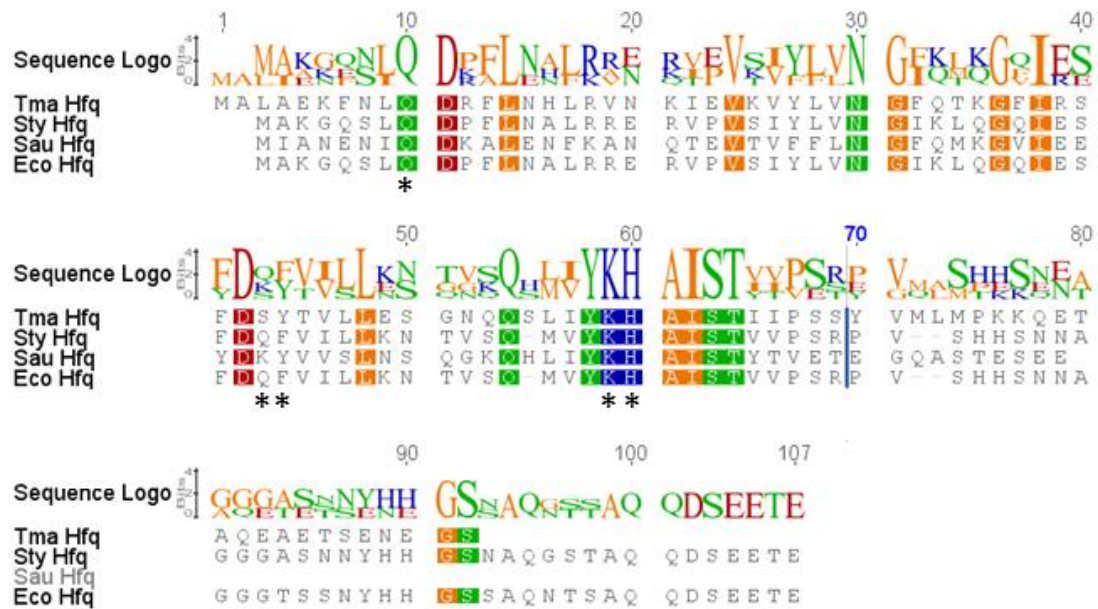
Figure 2.9 Hfq sequence alignment. Alignment of *Tma* Hfq with Hfq homologs for which the binding of short U-rich oligonucleotides has been characterized by both RNA-binding and structural studies, *Salmonella typhimurium* Hfq (*Sty* Hfq), *Staphylococcus aureus* Hfq (*Sau* Hfq) and *Escherichia coli* Hfq (*Eco* Hfq). Residues that are at least 85% conserved are highlighted based on polarity: hydrophobic nonpolar residues (AFGILMPVW) are **orange**; uncharged polar residues (CNSTY) are **green**; acidic polar residues (DE) are **red**; basic polar residues (HKR) are **blue**. Residues located in the proximal RNA-binding site of all three co-crystallization structures, *Sau* Hfq with AU_5G RNA (PDB 1KQ2), *Sty* Hfq with U_6 RNA (PDB 2YLC), and *Eco* Hfq with A_7 and AU_6A RNA (PDB 4HT9), are indicated by a star.

Table 2.1 Properties of the nanoRNA oligonucleotides that co-purified with *Tma* Hfq

Oligo name	Sequence	Chromatogram fraction	MW (g/mol)	N	IC50 (nM)	K_i (nM)
U_5	5'Phos-rUrUrUrUrU-OH 3'	F23	1548.9	4	6 ± 1	6 ± 1
U_3CU	5'Phos-rUrUrUrCrU-OH 3'	F17	1547.9	4	13 ± 1	12 ± 1
CU_2CU	5'Phos-rCrUrUrCrU-OH 3'	F13	1546.9	4	24 ± 1	22 ± 1
U_4C	5'Phos-rUrUrUrUrC-OH 3'	F19	1547.9	4	25 ± 1	23 ± 1
CU_2CU_2	5'Phos-rCrUrUrCrUrU-OH 3'	F21	1853.1	5	4 ± 1	4 ± 1
U_4CU	5'Phos-rUrUrUrUrCrU-OH 3'	F28	1854.0	4	28 ± 1	26 ±1

N is the number of independent replicates the competitive binding assay was performed.

Chapter 3

Characterization of *Thermotoga maritima* Hfq self-assembly: A regulatory mechanism involving two distinct oligomeric states?*

* This chapter is an adapted pre-print of the following manuscript:

Patterson J. and Mura C. (2014) "Characterization of *Thermotoga maritima* Hfq self-assembly: A regulatory mechanism involving two distinct oligomeric states?" [In preparation]

Introduction

Hfq was first identified as a <u>h</u>ost <u>f</u>actor essential for the replication of RNA phage Qβ in *Escherichia coli* (*Eco*) [1]. The function of Hfq has since been linked to quorum sensing [2], bacterial virulence [3], and biofilm formation [4] through the stabilization of various mRNA···sRNA complexes that are important in these various pathways. The key regulatory role of Hfq in numerous cellular pathways results in a pleiotropic phenotype when Hfq is disrupted in *E. coli* either by mutation or by knockout. These phenotypes include increased sensitivity to low pH, UV light and high temperatures, as well as a decreased growth rate [5]. These phenotypic changes can be linked to the stress response pathways that require Hfq to modify gene expression. For example, sRNA-regulated translation of the RNA polymerase subunit σ^s [6], which is involved in environmental stress response, requires Hfq to facilitate the annealing of the trans-encoded regulatory sRNAs (DsrA, RprA, or ArcZ) to their target mRNA transcript (*rpoS*) [7,8,9,10,11]. The precise molecular mechanism underlying RNA stabilization/destabilization by Hfq, and Hfq's mediation and modulation of RNA···RNA interactions, remains a challenging question. There exist a few hypothetical models, but due to gaps in our knowledge a single model has yet to be generally accepted for the molecular basis of Hfq activity [12]. It is worth noting that the overwhelming majority of *in vitro* biochemical studies have utilized only a single homolog of Hfq (*E. coli* Hfq [7,13,14,15]); therefore, data from other Hfq systems would broaden our understanding of this RNA regulatory system.

In order to dissect the molecular mechanism of Hfq activity, we must understand the structure and dynamics of Hfq. The Hfq protein backbone adopts a three-dimensional shape known as the 'Sm' fold, which occurs in RNA-binding proteins across all domains of life (Bacteria, Archaea, Eukaryotes) [16]. The Sm fold consists of five consecutive beta strands; β4 and β5 are the most conserved and form the interface between adjacent subunits in the protein oligomer,

contributing to its overall thermostability [17,18]. In addition to the β-sheet, the Hfq protein

contains an N-terminal α-helix that is present in all Sm-like proteins. Structural data have shown

that Hfq self-assembles into a homohexameric toroid structure that is approximately 70-80 Å in

diameter with an 8-10 Å central hole (Figure 1.3) [17,18,19]. The α-helices lie on one side of the

doughnut-shaped ring referred to as the 'proximal' face, whereas the opposite side is referred

to as the 'distal' face. The homohexamer has been shown be a functional form of Hfq in *E. coli*

and other bacterial species using multiple physical techniques such as analytical

ultracentrifugation [1], gel filtration chromatography [13], and electron microscopy [13].

Interestingly, some studies have suggested the occurrence of higher-order oligomeric states

[19,20,21], but the potential functional roles of such states remain elusive.

In this study, we characterize the oligomerization of *Thermotoga maritima* (*Tma*) Hfq,

and we establish that a dodecamer is in equilibrium with the hexameric state using a

combination of biophysical techniques including chemical cross-linking, analytical size exclusion

chromatography (AnSEC), 'semi-native' Western blots, and isothermal titration calorimetry (ITC).

The two oligomeric states that we found in a crosslinked sample of *Tma* Hfq were separated by

AnSEC. In order to assess the RNA-binding properties of these two oligomeric states, a series of

binding assays was conducted, measuring *Tma* Hfq's ability to bind both 5-Carboxyfluorescein

labeled poly-adenosine (FAM-A_{18}) and poly-uracil (FAM-U_6) RNAs. In contrast to previous studies

with Hfq from other bacterial species, we find that both the hexameric and dodecameric states

of *Tma* Hfq are able to interact with U-rich and A-rich RNA. Interesting, the presence of U_6 RNA

shifts the equilibrium between the hexamer and dodecamer toward the hexameric state,

whereas A_{18} RNA does not alter the equilibrium between these states.

Methods

Cloning, expression, and purification. The cloning, over-expression, and purification protocols described in Chapter 2 were used in this work in order to obtain pure samples of recombinant *Tma* Hfq.

Chemical cross-linking. Purified recombinant *Tma* Hfq was dialyzed into 25 mM HEPES pH 8.0 and 200 mM NaCl prior to cross-linking. Samples were crosslinked with either formaldehyde or glutaraldehyde using the indirect cross-linking method described in Chapter 2 [22]. After incubation with the cross-linking reagent, samples were purified from salts and cross-linker using a C_4 ZipTip® (Millipore) [23]. The molecular weight of formaldehyde-crosslinked *Tma* Hfq oligomers were determined via matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) using the protocol described in Chapter 2. Data were collected in three independent replicates for both the formaldehyde and glutaraldehyde experiments.

Tma Hfq was also crosslinked with 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) after dialysis into 25 mM HEPES pH 8.0 and 200 mM NaCl. The reaction mixture consisted of 67 mM EDC, 167 mM N-Hydroxysulfosuccinimide sodium salt (sulfo-NHS), and 1 mg/mL purified recombinant *Tma* Hfq. The reaction was incubated at room temperature (RT) for 4 hours then stopped by the addition of β-mercaptoethanol (BME) to a final concentration of approximately 18 mM. Cross-linking was assessed by MALDI-TOF MS using the protocol described in Chapter 2. Data were collected in three independent replicates.

Analytical size-exclusion chromatography (AnSEC). Recombinant *Tma* Hfq was dialyzed into 25 mM Tris pH 8.0 and 1 M NaCl. The chromatography system consisted of a Superdex 200 10/300GL column (spherical composite of crosslinked agarose and dextran matrix) and a Biologic DuoFlow™ system (BioRad) at 4°C. *Tma* Hfq and molecular weight standards were injected and eluted with approximately seven column volumes of 25 mM Tris pH 8.0 containing

1 M NaCl at 0.4 mL/min; absorbance was monitored at 280 nm throughout each run. The molecular weight standards used to generate a standard curve were blue dextran (2,000 kDa), β-amylase from sweet potato (200 kDa), alcohol dehydrogenase from yeast (150 kDa), albumin from bovine serum (66 kDa), carbonic anhydrase from bovine erythrocytes (29 kDa), and cytochrome C from horse heart (12.4 kDa) each one at a concentration 25 mg/mL.

Purification and characterization of polyclonal antibodies. The purification and characterization protocols described in Appendix 2 were used in this study.

Semi-native Western blot analysis. Recombinant *Tma* Hfq was dialyzed into 25 mM Tris pH 8.00, 350 mM NaCl and serially diluted from 20 µM to 78 nM across nine samples. The diluted samples were incubated at RT for 30 minutes and then 4x 'semi-native' loading buffer was added to each sample (final 1x working concentrations: 50 mM Tris pH 6.8, 0.5% sodium dodecyl sulfate, 10% glycerol, 12.5mM ethylenediaminetetraacetic acid (EDTA), and 0.02% bromophenol blue). Protein samples were separated on a 7.5% w/v TGX gel (Bio-Rad) at RT using 1x Tris-glycine sodium dodecyl sulfate (SDS) running buffer (200V, 28 min), and then transferred to a nitrocellulose membrane using the Trans-Blot Turbo transfer system (Bio-Rad) with a Trans-Blot Turbo Transfer pack (Bio-Rad) at 2.5 A (up to 25 V) for 3 minutes. The Odyssey one-color protein molecular weight marker (Li-Cor), which fluoresces in the 700nm channel of any of the Odyssey Imaging Systems, was run in parallel to the protein samples. The membrane was blocked with 5% w/v dry milk and probed with rabbit anti-*Tma* Hfq polyclonal (pAb) antibodies (Covance; see Appendix 2). Goat anti-rabbit IgG IRDye 800CW was used as the secondary antibody (Li-Cor) for visualization using an Odyssey Li-Cor imaging system.   The signal intensity of each band was quantified using the Image Studio software. The signal was then normalized by the total signal intensity of each lane in order to determine the fraction of the sample in each oligomeric state at a specific concentration (i.e. in a particular lane).

Isothermal titration calorimetry (ITC). Measurements of heat changes associated with the oligomerization of *Tma* Hfq were made on a MicroCal VP^{TM}-ITC system at 25˚C. Recombinant *Tma* Hfq was dialyzed against 25 mM Tris pH 8.00, 350 mM NaCl, diluted with dialysis buffer to 22.5 µM and then degassed at 25 ˚C before loading into the ITC syringe. Dialysis buffer was filtered through a 0.22 µm filter membrane (Millipore) and degassed at 25˚C in a ThermoVac (MicroCal) before loading the sample cell (1.44 mL volume). Injections were set for 10 µL of injectant (22.5µM *Tma* Hfq), with a 2 minute spacing interval between injections. Raw data was collected as thermal power (µcal/s) over time (min).

Analysis of ITC data. Each titration peak in the thermograph (thermal power (µcal/s) as a function of time (min)) was integrated using MicroCal software (MicroCal LLC) to determine the thermal heat per injection (µcal/injection). To generate a binding curve, the thermal heat per injection was converted to the thermal heat per mole of injectant (µcal/mole or injectant) and plotted against the log of the concentration of *Tma* Hfq. The binding curve was then fit to a sigmoidal Boltzmann function, which is related to the Hill equation [24], and which can be rearranged to read

$$y = \frac{(A_1 - A_2)}{1 + e^{\frac{(x - x_0)}{\partial x}}} + A_2 \qquad \text{(Equation 1)}$$

where x_0 is the inflection point of the sigmoidal curve, dx is the width of the transition, and A_1 and A_2 are equal to the enthalpy (in kcal/mol) of the initial and final state, respectively [24,25]. Nonlinear least-squares fits of the equation to the data were performed in OriginPro7.5.

Fluorescence polarization. Fluorescence polarization measurements were collected using the protocol describe in Chapter 2. In addition to 5'-fluoresceinated (FAM) labeled U_6 RNA, fluorescence polarization measurements were performed using FAM-A_{18} in the presence of purified recombinant *Tma* Hfq. FAM-A_{18} was incubated at 85°C for 3 minutes and then placed on ice for 10 minutes prior to the binding assay [26]. Fluorescence polarization measurements

were analyzed using the protocol described in Chapter 2. FAM-labeled oligonucleotides were purchased from Integrated DNA Technologies.

Results

Tma Hfq assembles into two distinct oligomeric states

The purity of *Tma* Hfq was assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE, Figure 3.1a) and MALDI-TOF MS (Figure 3.1b), which revealed only two peaks corresponding to the +1 and +2 charge states of *Tma* Hfq. In general, only covalently bound molecular complexes are detected by MALDI-TOF MS because of the ionization process [27,28,29]; therefore, to covalently link the Hfq monomer subunits within any potential oligomers, *Tma* Hfq was chemically treated with a panel of crosslinking reagents. These series of experiments included EDC, formaldehyde, and glutaraldehyde. Formaldehyde and glutaraldehyde couple primary amines, whereas EDC cross-links carboxyl groups to primary amines [27,30,31]. The mass spectra of the three crosslinked samples contained two predominant peaks in the 20,000 to 200,000 mass-to-charge (m/z) ratio range (Figure 3.2). These peaks roughly correspond to the expected molecular weight for the Hfq hexamer (64.8 kDa) and dodecamer (129.6 kDa). The specific molecular weights and oligomeric states measured by MALDI-TOF MS for each cross-linking reagent are summarized in Table 3.1.

Additional peaks were observed that correspond to the +2 charge state of the protein hexamer as well as intermediate oligomeric states presumably due to incomplete cross-linking of the sample (Figure 3.2a,b). For the glutaraldehyde sample, a low intensity, broad peak corresponding to approximately 183 kDa was observed after cross-linking for 60 minutes. A secondary peak at a molecular weight approximately 1500 Da greater than the Hfq hexamer peak was found with in the formaldehyde treated sample (Figure 3.2b).

86

In order to analyze *Tma* Hfq oligomeric states under more native-like conditions (i.e. without cross-linking), AnSEC was used to determine molecular weights of the *Tma* Hfq oligomers present in solution. Based on the standard curve (Figure 3.3), the molecular weight of *Tma* Hfq in solution is approximately 98.4 kDa (Figure 3.4, red trace), which corresponds to a nine-mer.   To determine if the discrepancy between the oligomeric states determined by AnSEC and by cross-linking/MALDI-TOF MS is due to the shape of the Hfq hexamer, versus a possible equilibrium between the hexamer and dodecamer, EDC-crosslinked *Tma* Hfq was run on the AnSEC column. The chromatogram of the crosslinked sample (Figure 3.4, blue trace) exhibited two peaks corresponding to molecular weights of approximately 72.8 kDa and 187.9 kDa, which can be attributed to a *Tma* Hfq hexamer (AnSEC pk 2) and 17-mer (AnSEC pk1). The peaks from AnSEC were analyzed by MALDI-TOF MS (Figure 3.5). The MALDI-TOF MS spectra for the AnSEC pk2 sample contained a peak corresponding to the Hfq hexamer along with a significantly smaller peak corresponding to the expected molecular weight of an Hfq dodecamer; lower molecular weight peaks were also observe suggesting that cross-linking was not complete (Figure 3.5b). The MALDI-TOF MS spectra for the AnSEC pk1 sample (Figure 3.5a) contained a peak corresponding to the Hfq hexamer as well as a small peak corresponding to a dodecamer, but no peak corresponding to a 17-mer was observed. Thus, we can unambiguously assign the peaks in the AnSEC chromatogram to these Hfq assembly states.

Hfq is in equilibrium between hexamer and dodecamer states

To further elucidate a possible equilibrium between the hexamer and dodecamer states of Hfq, oligomer dissociation upon dilution was monitored by ITC (Figure 3.6). As the concentration of *Tma* Hfq increases in the sample cell, the thermal heat released upon injection of *Tma* Hfq decreased until the sample cell reached a concentration of *Tma* Hfq at which the dodecamer no longer dissociated into hexamer (Figure 3.6a); therefore, the forward direction of the hexamer

to dodecamer is being monitor in the ITC measurements. This experiment indicated that the association constant (K_a) and the enthalpy of association (ΔH_a) are 0.5 ± 0.1µM and 36.7 ±0.6 kcal/mol (N=4) at 25°C, respectively. The Gibbs free energy change was calculated to be −7.8 ± 0.1 kcal/mol (N=4) for the association of hexamers to dodecamer, indicating that association is favorable.

To verify that the oligomeric state transition being monitored by ITC (in the range of ~183 nM to 4 µM *Tma* Hfq) is genuinely the dodecamer to hexamer transition, Semi-native Western blots were performed. The semi-native Western blot indicated that the two predominant oligomeric states present in the concentration range (20 µM to 78 nM *Tma* Hfq) are the hexamer and dodecamer (Figure 3.7). The absence of the Hfq monomer in this concentration range was not due to a preference of the rabbit anti-*Tma* Hfq pAb for the hexamer (Figure 3.8). The fluorescence signal intensity of each band was used to estimate the fraction of hexamer and dodecamer at each protein concentration, so that binding curves could be generated for approximating the K_a (Table 3.2).

Both of the oligomeric states have nanomolar affinities for poly-A and poly-U RNA

In order to examine the RNA-binding properties of the hexamer and dodecamer state, the binding of FAM-A$_{18}$ and FAM-U$_6$ RNA to *Tma* Hfq was monitored by fluorescence polarization assays. The various states of *Tma* Hfq used in these experiments include: native *Tma* Hfq ("native"), *Tma* Hfq crosslinked with EDC ("EDC"), crosslinked dodecamer ("AnSEC pk1") and crosslinked hexamer ("AnSEC pk2"). Fluorescence polarization data from four independent experiments were averaged and plotted against Hfq monomer concentration (Figure 3.9). The data were fit with the modified Hill equation (Equation 1), yielding dissociation constants (K_d) for each state of *Tma* Hfq with FAM-A$_{18}$ and FAM-U$_6$ (Table 3.3). The K_d for FAM-A$_{18}$ was approximately equivalent for the native and EDC states of *Tma* Hfq, whereas the K_d for AnSEC

pk1 and AnSEC pk2 are significantly different (P<0.1, based on unpaired t-test, not assuming the same standard deviation [32]).

Poly-U RNA shifts the equilibrium between the oligomeric states

Semi-native Western blots of *Tma* Hfq were performed in the presence of stoichiometric equivalents of each U-rich and A-rich RNAs (1:1 Hfq to RNA). The semi-native Western blots indicated that in the presence of A_{18} the binding constant was unaffected (Table 3.3), whereas in the presence of U_6 the hexamer to dodecamer K_d increased by roughly two-fold (from 1.0 ± 0.2 µM to 1.8 ± 0.4 µM).

Discussion

Numerous studies have demonstrated that Hfq hexamers are a functional form of the protein [7,11,12,13,14,15]. Alternatively, some studies have indicated the presence of higher order oligomeric states *in vitro*, though presently any *in vivo* physiological roles of these higher order oligomeric states remains elusive [14,15,19,20,21]. A recent study of the oligomerization of Hfq from *E. coli* found that Hfq adopts multiple oligomeric states at µM concentrations [15]. Furthermore, this study indicated that the midpoints of the monomer-to-hexamer and the hexamer-to-multimer equilibria are 0.8 µM and 4.9 µM, respectively [15]. The midpoints of these two transitions suggest that at intracellular concentrations (1 µM in *E. coli* [33]) the predominant oligomeric states are the monomer and hexamer [15].

Since only covalently bound species are observed by MALDI-TOF MS, spectra of chemically crosslinked *Tma* Hfq were collected (Figure 3.2). For all three cross-linking reagents (formaldehyde, glutaraldehyde, EDC), peaks corresponding to a hexamer and dodecamer were observed (Table 3.1). The relative size of the dodecamer peak with respect to the hexamer peak increases as the cross-linker length increases (EDC<formaldehyde<glutaraldehyde). EDC is the

cleanest and most well-defined cross-linking agents used in this study, forming a covalent bond between a carboxyl group and primary amine in the protein [6]. Specifically, EDC forms an unstable, active intermediate with a carboxyl group in the protein sample, which subsequently reacts with the catalyst sulfo-NHS to form a stable, amine-reactive intermediate that reacts with a primary amine in the protein forming a direct bond between the carboxyl group and the primary amine. Although EDC is considered a "zero-length" cross-linker, the length of the crosslinked side chains between the α-carbons is over 10 Å [34].

To determine if both the hexamer and dodecamer states are present under native conditions, the sample was separated on a size exclusion column, which suggested that *Tma* Hfq exists as a single oligomeric state (Figure 3.4). When the elution volume of the sample was compared to a standard curve (Figure 3.3), the apparent oligomeric state was determine to be a 9-mer. The formation of a 9-mer disagrees with current literature regarding the oligomeric state of Hfq in other bacterial organism [7,11,12,13,14,15,17,20] as well as the MALDI-TOF MS results reported here. To resolve this discrepancy, EDC-crosslinked *Tma* Hfq was separated by size exclusion chromatography (SEC), which showed the presence of two oligomeric states. When the elution volumes of both peaks were compared to a standard curve (Figure 3.3), the apparent oligomeric states were determined to be a hexamer (6.75 ± 0.09 subunits) and a 17-mer (17.4 ± 0.6 subunits). The presence of a hexamer in the crosslinked sample agrees with the results obtained by MALD-TOF MS analysis of the same sample and suggests that the shape of the protein is not the sole factor causing a higher apparent molecular weight in the native sample. The apparent 17-mer observed in the size exclusion chromatogram is inconsistent with the dodecamer observed by MALDI-TOF MS. The 17-mer peak isolated by SEC was therefore analyzed by MALDI-TOF MS confirming the presence of hexamer and dodecamer in the sample (Figure 3.5a). Together these results suggest that the overall shape of the dodecamer

significantly contributes to its migration through the size exclusion column. The orientation of the hexamers or spacing between the hexamers in the dodecamer could potentially correspond to a large hydrodynamic radius resulting in a higher apparent molecular weight. Further studies are needed to determine the three-dimensional structure of the putative dodecamer.

The differences between elution volumes observed for the native and crosslinked *Tma* Hfq samples suggest that the hexamer and higher order oligomeric state are in (relatively) rapid equilibrium. ITC studies of the dodecamer-to-hexamer dissociation were used to further elucidate the thermodynamic properties and equilibrium constant of the hexamer to dodecamer equilibrium. The K_a for this equilibrium was determined to be 0.5 ± 0.1μM, which is 10-fold lower than what was found for *Eco* Hfq [15]. Semi-native Western blot verified that the two predominant oligomeric states in the concentration range being monitored by ITC (20 μM to 78 nM *Tma* Hfq) are indeed the hexamer and dodecamer (Figure 3.7).

There exist many known examples of proteins from thermophilic bacteria that adopt higher-order oligomeric states versus their mesophilic homologs [35,36]. These higher-order states tend to be associated with enhanced thermostability; formation of the proper oligomeric state is critical for function. To elucidate which of the oligomeric states are functional in terms of RNA-binding, a series of binding assays was performed using FAM-A_{18} and FAM-U_6 probes to evaluate the RNA-binding properties of the two oligomeric states (Table 3.2). These two probes were specifically selected to assess the binding properties of the distal and proximal face of Hfq, which bind to poly-A and poly-U RNA, respectively [14,37,38,39]. The K_d for FAM-U_6 for the native and EDC crosslinked samples were statistically different (P<0.01, based on unpaired t-test with Welch's correlation [32]), suggesting that the populations of dodecamer and hexamer were different in the two samples. Additionally, for FAM-U_6 the K_d was similar for the native and hexameric state, but differed significantly between the dodecamer and native state. In contrast,

the K_d for FAM-A$_{18}$ was approximately equivalent for the native and EDC states of *Tma* Hfq, whereas the K_d for crosslinked dodecamer and crosslinked hexamer are significantly different (P<0.01, based on unpaired t-test with Welch's correlation [32]). Specifically, FAM-A$_{18}$ exhibits a 2-fold higher affinity for the *Tma* Hfq dodecamer. Cumulatively, these finding suggest that the relative population of the two oligomeric states is altered by the presence of poly-U RNA, which shifts the equilibrium toward the hexameric state.

Semi-native Western blots of *Tma* Hfq in the presence of RNA support that U$_6$ RNA shifts the equilibrium toward the hexameric state, whereas A$_{18}$ RNA does not alter the equilibrium between the two oligomeric states. These findings suggest that the proximal faces form the interface between the two hexamers in the dodecameric state (Figure 3.10). This proposed model is supported both by the RNA-binding properties of the two oligomeric states and by the effect of poly-U RNA on the equilibrium between the two oligomeric states.

However, the nanomolar binding affinity of the dodecamer for FAM-U$_6$ is not consistent with an occluded binding site. A possible explanation for this inconsistency is the presence of an additional binding site on the Hfq toroid structure. Recently, the lateral surface of the *Eco* Hfq hexamer has been shown to play an important role in Hfq-RNA interactions [38,39]. This lateral surface of *Eco* Hfq has no apparent nucleic acid sequence specificity and was inefficient for U$_6$ RNA-binding in a proximal site mutant [38,39]. If the lateral surface of the *Tma* Hfq dodecamer binds U$_6$ RNA, this would suggest that the amino acid variation in the lateral surface of *Tma* Hfq (versus *Eco* Hfq and other Hfq homologs) is responsible for the different RNA-binding properties of *Tma* Hfq reported here than those reported for *E. coli* Hfq. Specifically, R16, R17, E18, and R19 were shown to be involved in RNA interactions at the lateral surface of *E. coli* Hfq [38], but only R16 is conserved in *Tma* Hfq. Instead of the arginine-rich rim present in *E. coli* Hfq, the equivalent surface in *Tma* Hfq consists of R18, V19, N20, and K21. Future studies could

concentrate on characterizing the lateral binding site through the use of *Tma* Hfq mutants and strategically designed RNA-probes.

An alternative explanation for the nanomolar binding affinity of U_6 RNA is that the two hexameric rings are not stacked in perfect alignment, but instead staggered, allowing for poly-U RNA to bind to the dodecameric state in a manner that decreases the stability of the oligomer and results in dissociation of the two hexamers. To determine the mode of binding, future experiments could examine the orientation of the two hexamers in the dodecamer, as well as the location of poly-U binding on the dodecamer.

References

1. Franze de Fernande MT, August JT, Hayward WS (1972) Bacterial Proteins Required for Replication of Phage Qbeta Ribonucleic-Acid - Purification and Properties of Host Factor I, a Ribonucleic Acid Binding Protein. Journal of Biological Chemistry 247: 824-&.

2. Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, et al. (2004) The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. Cell 118: 69-82.

3. Kulesus RR, Diaz-Perez K, Slechta ES, Eto DS, Mulvey MA (2008) Impact of the RNA chaperone Hfq on the fitness and virulence potential of uropathogenic *Escherichia coli*. Infection and Immunity 76: 3019-3026.

4. Kint G, De Coster D, Marchal K, Vanderleyden J, De Keersmaecker SCJ (2010) The small regulatory RNA molecule MicA is involved in *Salmonella enterica* serovar Typhimurium biofilm formation. Bmc Microbiology 10.

5. Tsui HC, Leung HC, Winkler ME (1994) Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in *Escherichia coli* K-12. Mol Microbiol 13: 35-49.

6. El-Shafey A, Tolic N, Young MM, Sale K, Smith RD, et al. (2006) "Zero-length" cross-linking in solid state as an approach for analysis of protein-protein interactions. Protein Sci 15: 429-440.

7. Brescia CC, Mikulecky PJ, Feig AL, Sledjeski DD (2003) Identification of the Hfq-binding site on DsrA RNA: Hfq binds without altering DsrA secondary structure. RNA 9: 33-43.

8. McCullen CA, Benhammou JN, Majdalani N, Gottesman S (2010) Mechanism of positive regulation by DsrA and RprA small noncoding RNAs: pairing increases translation and protects rpoS mRNA from degradation. J Bacteriol 192: 5559-5571.

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

10. Soper TJ, Woodson SA (2008) The rpoS mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. RNA 14: 1907-1917.

11. Mikulecky PJ, Kaw MK, Brescia CC, Takach JC, Sledjeski DD, et al. (2004) *Escherichia coli* Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. Nat Struct Mol Biol 11: 1206-1214.

12. Sharma CM, Papenfort K, Pernitzsch SR, Mollenkopf HJ, Hinton JC, et al. (2011) Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. Mol Microbiol 81: 1144-1165.

13. Zhang A, Wassarman KM, Ortega J, Steven AC, Storz G (2002) The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. Mol Cell 9: 11-22.

14. Updegrove TB, Correia JJ, Chen Y, Terry C, Wartell RM (2011) The stoichiometry of the *Escherichia coli* Hfq protein bound to RNA. RNA 17: 489-500.

15. Panja S, Woodson SA (2012) Hexamer to monomer equilibrium of *E. coli* Hfq in solution and its impact on RNA annealing. J Mol Biol 417: 406-412.

16. Murina VN, Nikulin AD (2011) RNA-binding Sm-like proteins of bacteria and archaea. similarity and difference in structure and function. Biochemistry (Mosc) 76: 1434-1449.

17. Sauter C, Basquin J, Suck D (2003) Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*. Nucleic Acids Research 31: 4091-4098.

18. Nikulin A, Stolboushkina E, Perederina A, Vassilieva I, Blaesi U, et al. (2005) Structure of *Pseudomonas aeruginosa* Hfq protein. Acta Crystallogr D Biol Crystallogr 61: 141-146.

19. Schumacher MA, Pearson RF, Moller T, Valentin-Hansen P, Brennan RG (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. EMBO J 21: 3546-3556.

20. Arluison V, Derreumaux P, Allemand F, Folichon M, Hajnsdorf E, et al. (2002) Structural Modelling of the Sm-like Protein Hfq from *Escherichia coli*. Journal of Molecular Biology 320: 705-712.

21. Arluison V, Mura C, Guzman MR, Liquier J, Pellegrini O, et al. (2006) Three-dimensional structures of fibrillar Sm proteins: Hfq and other Sm-like proteins. Journal of Molecular Biology 356: 86-96.

22. Fadouloglou VE, Kokkinidis M, Glykos NM (2008) Determination of protein oligomerization state: two approaches based on glutaraldehyde crosslinking. Anal Biochem 373: 404-406.

23. Warren ME, Brockman AH, Orlando R (1998) On-probe solid-phase extraction/MALDI-MS using ion-pairing interactions for the cleanup of peptides and proteins. Anal Chem 70: 3757-3761.

24. Gesztelyi R, Zsuga J, Kemeny-Beke A, Varga B, Juhasz B, et al. (2012) The Hill equation and the origin of quantitative pharmacology. Arch Hist Exact Sci 66: 427-438.

25. Seber GAF (1989) Nonlinear Regression: John Wiley & Sons, Inc.

26. Sun X, Wartell RM (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.

27. Sinz A (2003) Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. J Mass Spectrom 38: 1225-1237.

28. Siegel MM (1996) Determination of loading values and distributions for drugs conjugated to proteins and antibodies by MALDI-MS and ESI-MS. Methods Mol Biol 61: 211-226.

29. Farmer TB, Caprioli RM (1998) Determination of protein-protein interactions by matrix-assisted laser desorption/ionization mass spectrometry. J Mass Spectrom 33: 697-704.

30. Metz B, Kersten GF, Hoogerhout P, Brugghe HF, Timmermans HA, et al. (2004) Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. J Biol Chem 279: 6235-6243.

31. Avrameas S, Ternynck T (1969) The cross-linking of proteins with glutaraldehyde and its use for the preparation of immunoadsorbents. Immunochemistry 6: 53-66.

32. Glantz SA (2012) The Special Case of Two Groups: The t Test. In: Diedrich C, Pancotti R, editors. Primer of Biostatistics. New York: McGraw-Hill Companies, Inc. pp. 49-72.

33. Kajitani M, Kato A, Wada A, Inokuchi Y, Ishihama A (1994) Regulation of the *Escherichia coli* hfq gene encoding the host factor for phage Q beta. J Bacteriol 176: 531-534.

34. Rappsilber J (2011) The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. Journal of Structural Biology 173: 530-540.

35. Thoma R, Hennig M, Sterner R, Kirschner K (2000) Structure and function of mutationally generated monomers of dimeric phosphoribosylanthranilate isomerase from *Thermotoga maritima*. Structure 8: 265-276.

36. Vieille C, Zeikus GJ (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol Mol Biol Rev 65: 1-43.

37. Hopkins JF, Panja S, McNeil SA, Woodson SA (2009) Effect of salt and RNA structure on annealing and strand displacement by Hfq. Nucleic Acids Res 37: 6205-6213.

38. Panja S, Schu DJ, Woodson SA (2013) Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. Nucleic Acids Res 41: 7536-7546.

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

40. Velazquez-Campoy A, Leavitt SA, Freire E (2004) Characterization of protein-protein interactions by isothermal titration calorimetry. Methods Mol Biol 261: 35-54.
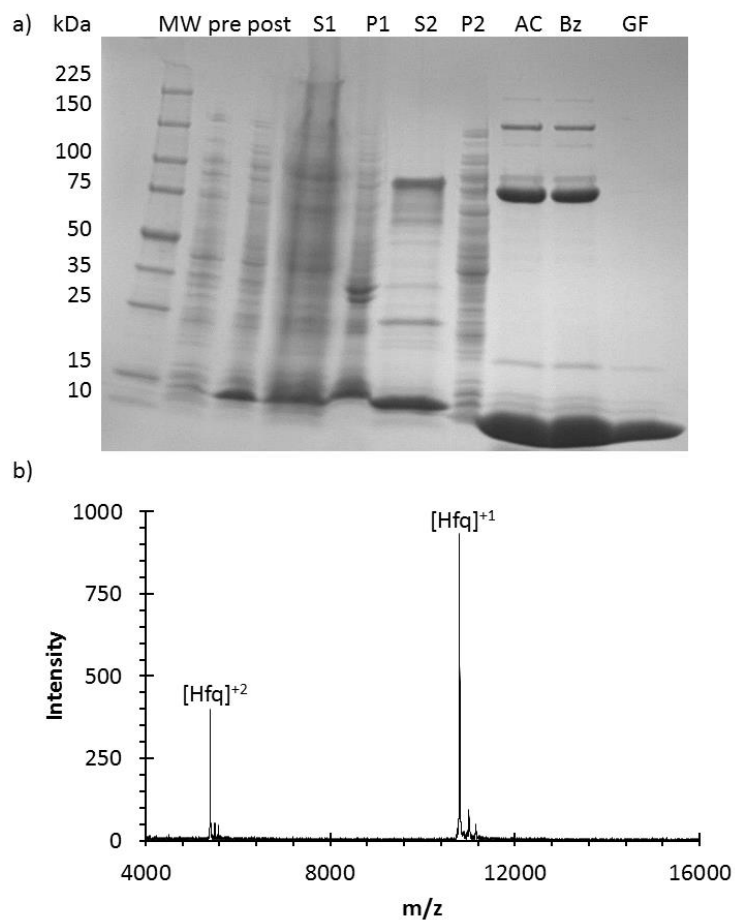
Figure 3.1 Representative SDS-PAGE and MALDI-TOF MS spectrum of recombinant *Tma* Hfq. a)

SDS-PAGE was performed on samples from the major steps in the recombinant *Tma* Hfq

purification protocol to monitor the purity of the sample. The samples run on 4-20% w/v TGX gel

were as follows: Promega Broad Range Protein Molecular Weight Marker (MW), pre-induction

(pre), post-induction (post), the soluble (S1) and pellet (P1) fractions from lysis, the soluble (S2)

and pellet (P2) fractions from the heat cut step, thrombin-treated sample (AC), benzamine

column flow-through (Bz), and gel filtration eluent (GF). b) Two peaks are detected in the

MALDI-TOF MS spectrum that can be contributed to the +1 and +2 charge states of the *Tma* Hfq

monomer, which has a molecular weight of 10,797.2 Da based on the amino acid sequence. No

peaks were observed in the higher molecular weight range (20-100 kDa).
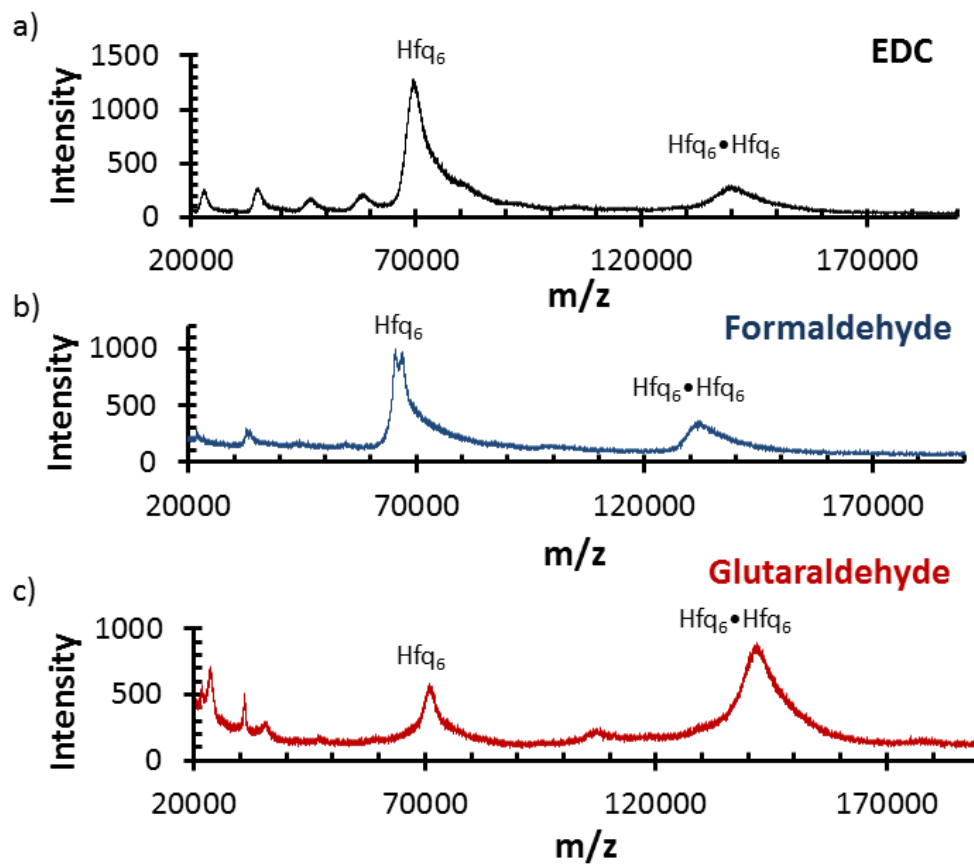
Figure 3.2 MALDI-TOF MS spectra of crosslinked *Tma* Hfq. Spectra were collected via MALDI-TOF MS. Two distinct oligomeric states, hexamer and dodecamer, were observed for all three cross-linkers: (a) 1-ethyl-3-[3-dimethylamino-propyl] carbodiimide hydrochloride (EDC), (b) formaldehyde, and (c) glutaraldehyde.
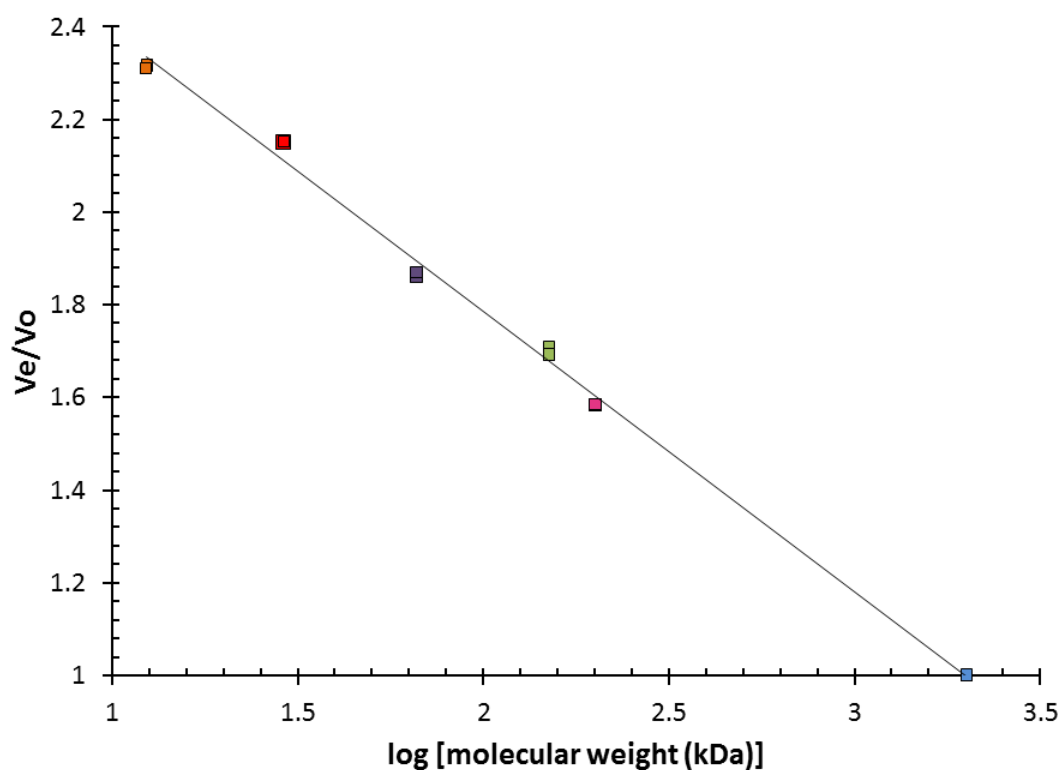
Figure 3.3 Standard curve for analytical size exclusion chromatography (AnSEC). The chromatography system consisted of a Superdex 200 10/300GL column and a Biologic DuoFlow™ system at 4°C. The molecular weight standards used to generate the standard curve were **blue dextran**, **β-amylase**, **alcohol dehydrogenase**, **albumin**, **carbonic anhydrase** and **cytochrome C** at a concentration of 25 mg/mL. Standards were run in duplicate and their average was fit using a linear regression with an R^2 of 0.997.
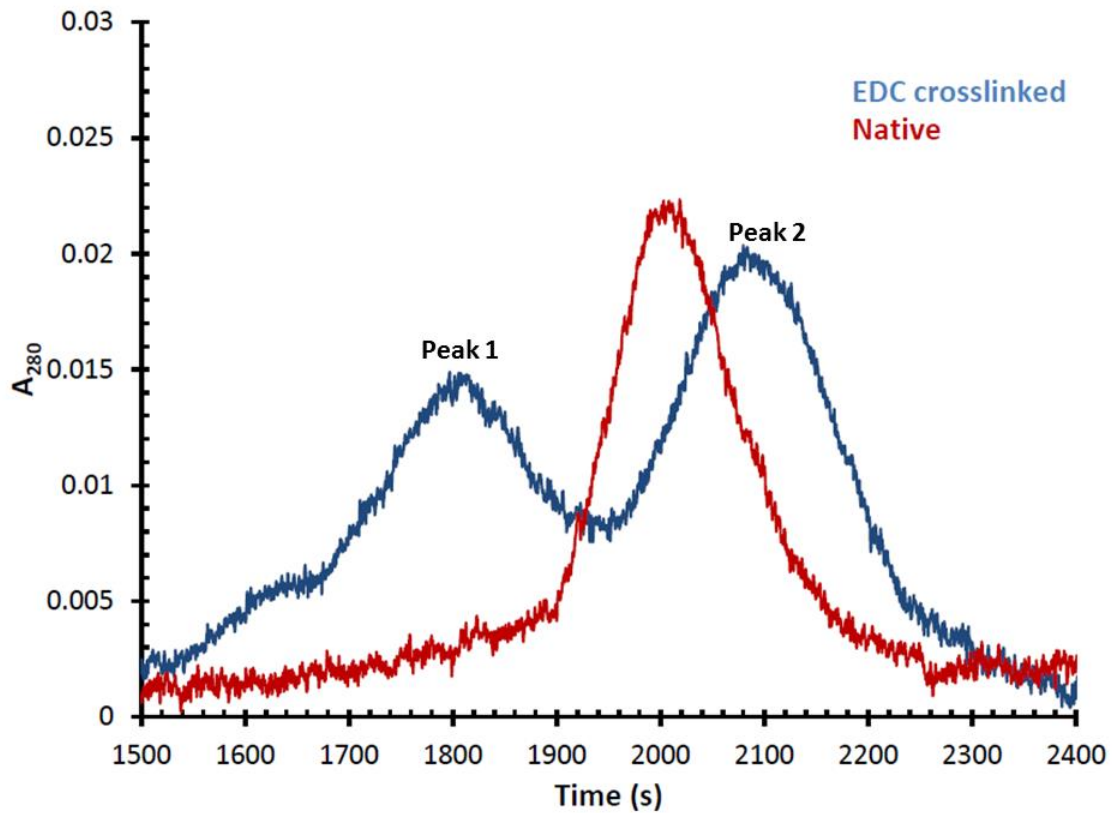
Figure 3.4 Representative chromatogram of AnSEC separation of native and crosslinked *Tma*

Hfq. A chromatogram of the native sample (red trace) contained a single peak corresponding to

a 9-mer (9.11 ± 0.07 subunits) of *Tma* Hfq, whereas the crosslinked sample(blue trace)

contained two peaks corresponding an Hfq hexamer (6.75 ± 0.09 subunits) and ~17-mer (17.4

±0.6 subunits). The presence of a single peak with an intermediate retention time in the native

sample suggests that an equilibrium exists between the two oligomeric states.
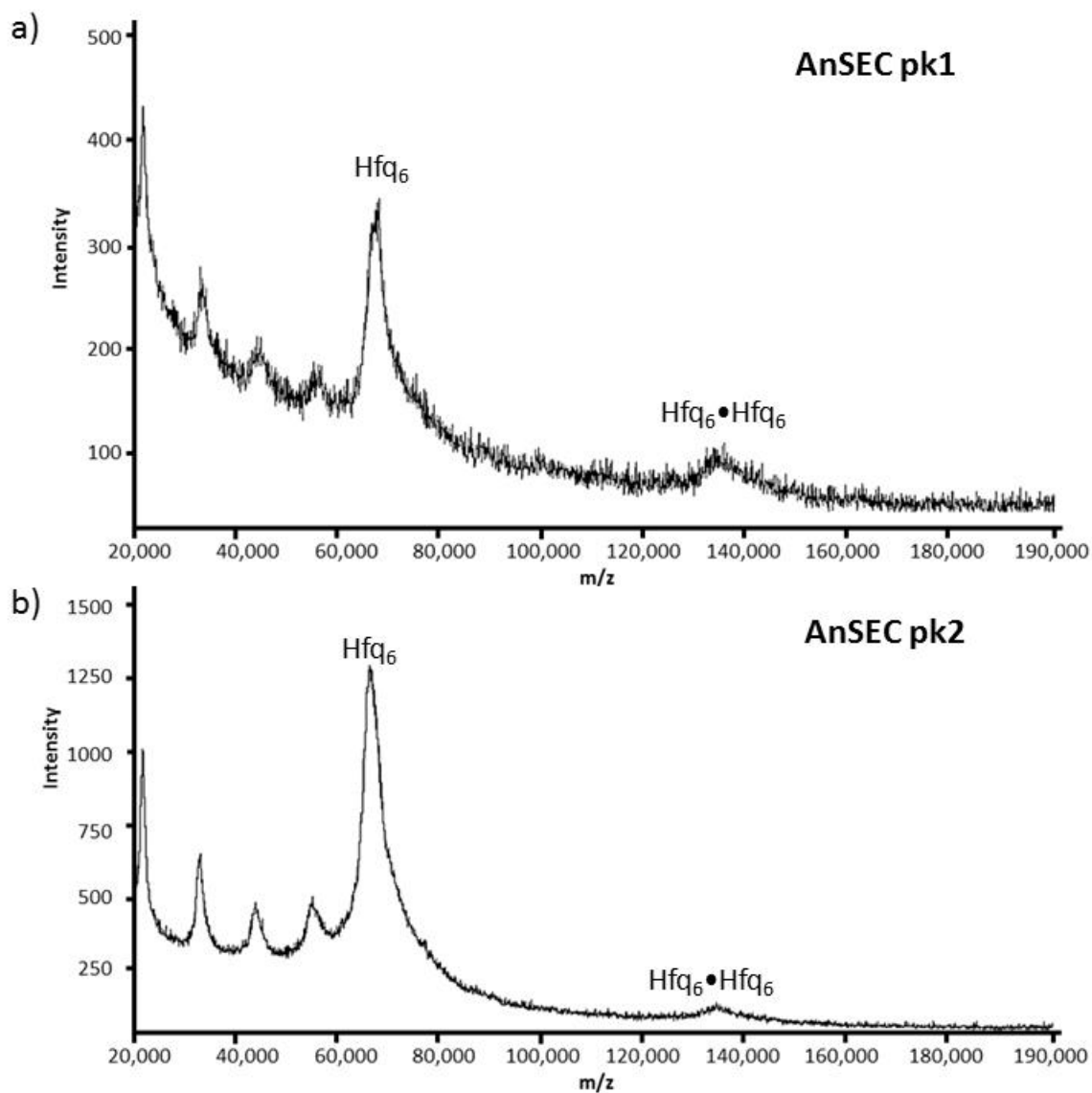
Figure 3.5 MALDI-TOF spectra of peak 1 and peak 2 from the AnSEC separation of EDC-crosslinked Recombinant *Tma* Hfq. The predominant peaks observed in the MALDI-TOF spectrum of recombinant *Tma* Hfq from peak 1 (a) and peak 2 (b) of the AnSEC separation can be contributed to the hexameric form of *Tma* Hfq. A significantly smaller peak corresponding to the *Tma* Hfq dodecamer is observed in both spectra. Additional peaks are observe below 60 kDa that correspond to lower level oligomeric states of Hfq that are presumably due to incomplete cross-linking. No peak was observed around 183.6 kDa, which would correspond to a *Tma* Hfq 17-mer.
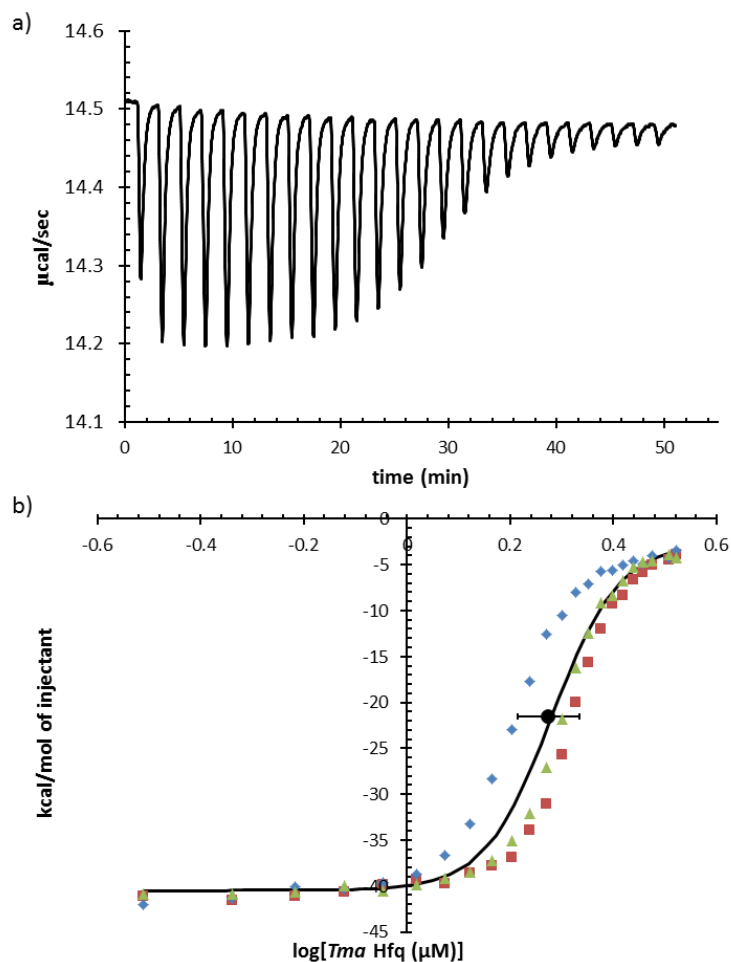
Figure 3.6 Representative data from isothermal titration calorimetry (ITC) experiments of *Tma* Hfq dissociation (dodecamer to hexamer) at 25°C. A representative thermograph of thermal power (μcal/sec) as a function of time (minutes); peaks are at constant intervals correspond to the injection of *Tma* Hfq (titrant) into dialysis buffer (sample cell) (a). The magnitude of these peaks corresponds to the energy of dissociation. After saturation, residual heat is still observed due to mechanical and dilution phenomena [40]. The area under each peak is integrated then normalized to the moles of injected *Tma* Hfq (kcal/mol of injectant), which is plotted against the concentration of *Tma* Hfq (b). Data were fit to a modified Hill's equation (Equation 1) and the K_a and ΔH were determined to be 0.5 ± 0.1μM and 36.7 ± 0.6 kcal/mol (N=4), respectively.
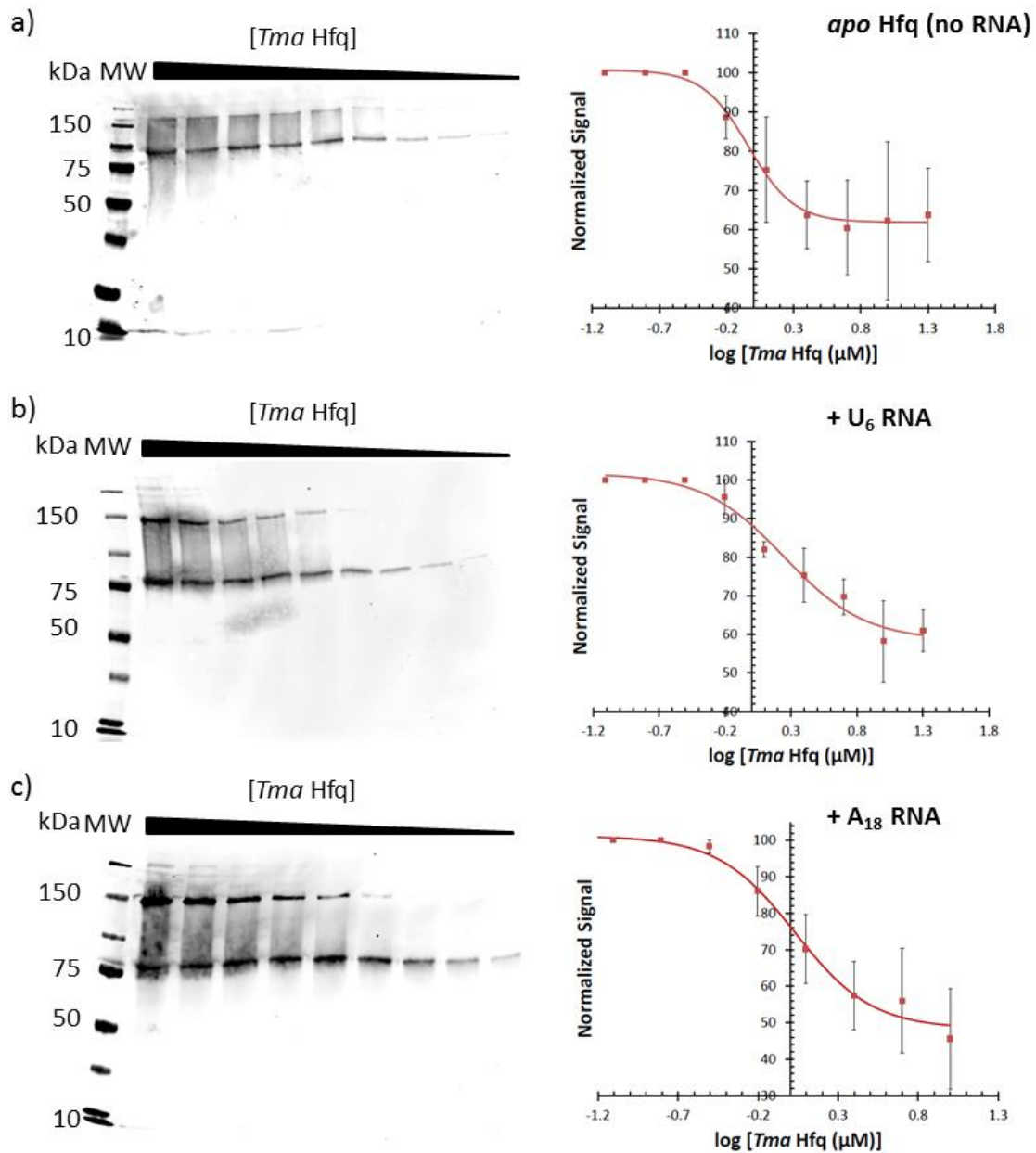
Figure 3.7 Semi-native Western blot of recombinant *Tma* Hfq. *Tma Hfq* was serially diluted

from 20 μM to 78 nM in 25 mM Tris pH 8.00, 350 mM NaCl and separated on a 7.5% (w/v) TGX

gel along with the Odyssey one-color protein molecular weight marker (MW). *Tma* Hfq was

detected using an indirect Western blot, which used rabbit anti-*Tma* Hfq polyclonal antibodies

pAb (Covance) and goat anti-rabbit IgG IRDye 800CW secondary antibodies (Li-Cor) for

visualization. Panel (a) shows a representative semi-native Western blot of apo Hfq and panels

(b) and (c) show representative semi-native Western blots of Hfq with U_6 and A_{18}, respectively.
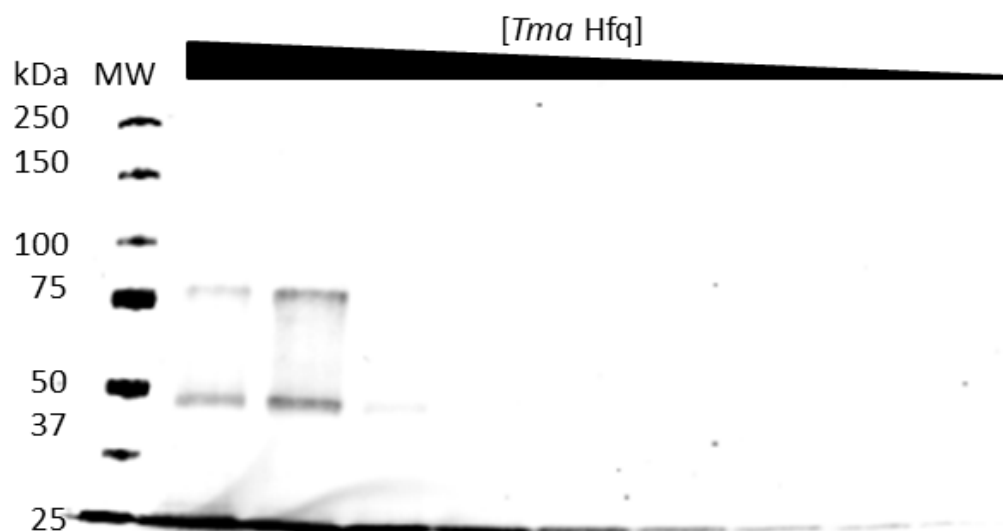
Figure 3.8 Western blot of denatured *Tma* Hfq. *Tma Hfq* was serially diluted from 2.5 µM to 78

nM under denaturing conditions (SDS loading buffer and heat incubation of samples at 95°C)

and separated on a 7.5% (w/v) TGX gel along with the Odyssey one-color protein molecular

weight marker (MW). Western blot analysis indicates that Rabbit anti-*Tma* Hfq pAb bind to the

denatured Hfq monomer with high affinity. It should be noted that separation of the Odyssey

one-color protein molecular weight marker (MW) below 25 kDa was not observed this is

because these molecular weights are below the separation range of the gel.
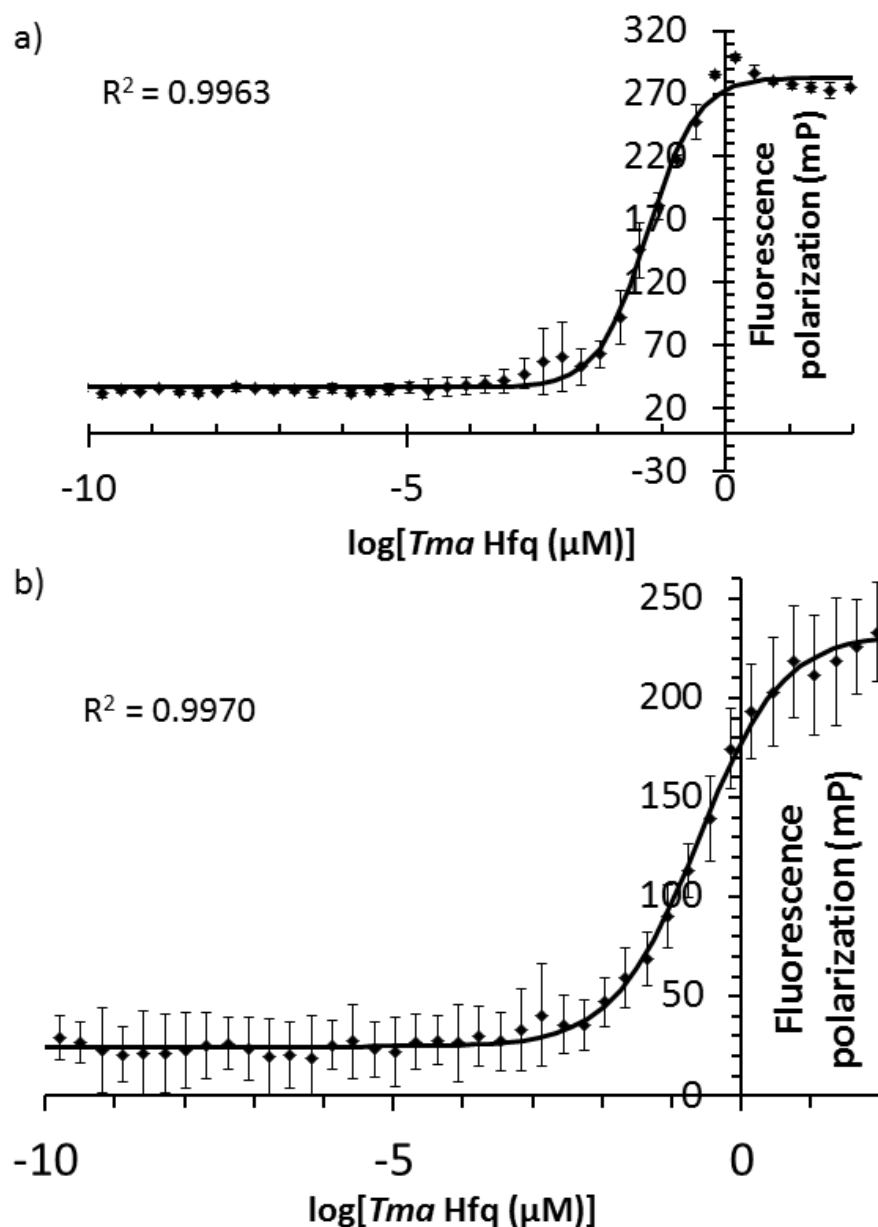
Figure 3.9 Binding curve for native _Tma_ Hfq with FAM-U$_6$ (a) and FAM-A$_{18}$ (b). Binding was measured by fluorescence polarization in 25 mM Tris pH 8.00 and 350 mM NaCl at RT, using 5 nM FAM-U$_6$ or 5nM FAM-A$_{18}$ and 180 μM − 2.4 pM native _Tma_ Hfq monomer. Fluorescence polarization from four independent experiments were averaged and plotted against the Hfq monomer concentration. The data was fit to a modified hill equation (Equation 1) yielding a K_d of 61 ± 10 nM and 236 ± 38 nM for FAM-U$_6$ and FAM-A$_{18}$, respectively.
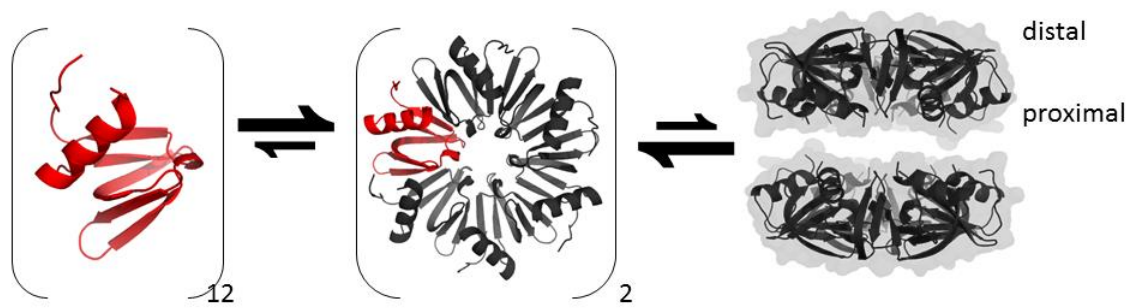
Figure 3.10 A working model for the equilibria that exist between the oligomeric states of Hfq in solution. The dodecamer forms with the proximal faces facing the interface between the two Hfq hexamers. The crystal structure of *T. maritima* Hfq (Chapter 4) was used in creating this schematic illustration.

Table 3.1 Summary of molecular weights and oligomeric states observed by MALDI-TOF for crosslinked *Tma* Hfq				
Crosslinking reagent	**Molecular weight (kDa)**		**Oligomeric states**	
	Pk1	**Pk 2**	**Pk1**	**Pk 2**
Formaldehyde	66 ± 1	133 ± 2	6.15 ± 0.09	12.3 ± 0.2
Glutaraldehyde	71.5 ± 0.3	141 ± 7	6.60 ± 0.3	13.1 ± 0.6
EDC	68.0 ± 0.6	136 ± 2	6.29 ± 0.06	12.6 ± 0.2

Table 3.2 Summary of the apparent thermodynamic parameters for the dodecamer-hexamer equilibria for *Tma* Hfq calculated from ITC dilution experiments and semi-native western blot analysis.					
Method	**Equilibrium**	K_d^{*} **(μM)**	**ΔG$_d$ (kcal/mol)**	**ΔH$_d$ (kcal/mol)**	**ΔS$_d$ (cal/mol·K)**
ITC	**Hfq$_6$ ⇌ Hfq$_6$::Hfq$_6$**	1.9 ± 0.3	7.81 ± 0.08	−36.9 ± 0.1	−150.1 ± 0.7
Semi-native western blot	**Hfq$_6$::Hfq$_6$ ⇌ Hfq$_6$**	1.0 ± 0.2	–	–	–
	Hfq$_6$::Hfq$_6$ + U$_6$ ⇌ Hfq$_6$ + U$_6$	1.8 ± 0.4	–	–	–
	Hfq$_6$::Hfq$_6$ + A$_{18}$ ⇌ Hfq$_6$ + A$_{18}$	1.0 ± 0.2	–	–	–

* K_d values were calculated using Hfq monomer concentrations.

Table 3.3 Summary of the dissociation constants (K_d) determined for FAM-U$_6$ and FAM-A$_{18}$ with the different oligomeric states of *Tma* Hfq.

	Description of Sample	FAM –U$_6$ $K_d{}^*$ (nM)	FAM-A$_{18}$ $K_d{}^*$ (nM)
Native	Equilibrium	61 ± 10	236 ± 38
EDC	'static'	116 ± 29	231 ± 30
AnSEC pk1	Dodecamer	79 ± 10	163 ± 24
AnSEC pk 2	Hexamer	57 ± 23	333 ± 31

* K_d values were calculated using Hfq monomer concentrations.

Chapter 4

Crystallographic Studies of *Thermotoga maritima* Hfq[*]

Introduction

The three-dimensional (3D) structure of Hfq homologs from mesophilic bacteria has provided significant information concerning the functionally relevant RNA-binding surfaces and oligomer interfaces of Hfq. Structural studies of *Tma* Hfq can reveal the structural origin of the known differences in RNA-binding specificities and affinities; this is motivated by the fact that biochemical *function* arises from biomolecular 3D *structure* [1,2,3]. Structural studies can also provide insight into the role of nanoRNAs in Hfq-mediated sRNA-mRNA annealing as well as elucidate the RNA sequence motifs and end-chemistries recognized by *Tma* Hfq.

Initial insights into the quaternary structure of Hfq were obtained via electron microscopy, which revealed that the Hfq protein forms a hexameric toroid structure [4]. Crystal structures of Hfq further revealed that the toroid structure is approximately 70-80 Å in diameter and 30 Å thick, with an 8-10 Å central hole (Figure 1.3) [5,6,7]. The Hfq monomer has an N-terminal α-helix followed by five antiparallel β-strands with β5-β1-β2-β3-β4 topology (Figure 1.3). In the hexamer, β4 and β5 form the interface between adjacent subunits (Figure 4.1); the α-helix of each monomer lies on one side of the toroidal structure referred to as the proximal face.

Proteins in the Sm/LSm Superfamily are characterized by two highly conserved sequence motifs, referred to as the Sm1 and Sm2 motifs, corresponding to β1-3 and β4-5, respectively [8,9]. The Sm1 sequence signature can be identified in Sm/LSm homologs from all three domains of life, whereas the Sm2 motif is divergent in the bacterial Hfq proteins [4,10,11,12,13,14]. Despite weak sequence similarity in β4 and β5 between bacterial Hfq and the archaeal and eukaryotic Sm proteins, these two strands are the most conserved regions among bacterial Hfq proteins [5,7].

Hfq is known to preferentially bind uridine-rich (U-rich) and adenine-rich (A-rich) RNAs. U-rich sequences of sRNAs interact with the proximal face of Hfq. In contrast, A-rich sequences,

which are often located either internally in mRNAs [15,16,17] or at the 3' end of RNA degradation intermediates [18,19], interact with the opposite side of the Hfq toroid, referred to as the distal face. The precise RNA sequence motif recognized by the distal face of Hfq differs between Gram-negative and Gram-positive bacteria: Gram-negative Hfq homologs recognize an $(ARN)_n$ repeat (A, adenosine; R, purine; N, any nucleotide) [20], while Gram-positive Hfq homologs preferentially bind $(AN)_n$ repeats [21]. Furthermore, transcriptome studies of the RNA-association profiles of Hfq homologs from various Gram-negative bacterial species indicate that the RNA-binding behavior profiles of Hfq can vary in a subtle, species-specific manner [22]. Since the overall 3D structure of the Hfq protein backbone is highly conserved (root mean squared deviations (RMSD) ranging from 0.37 Å to 0.52 Å) these differences occur primarily near the RNA binding site [10].

There are currently 34 Hfq structures in the protein data bank (PDB) from eight distinct bacterial species: the Gram-negative *Escherichia coli* [7], *Pseudomonas aeruginosa* [5], *Salmonella typhimurium* [23], *Herbaspirillum seropedicae* [24], *Anabaena sp.* [25], and *Synechocystis sp.* [25], and the Gram-positive *Bacillus subtilis* [26] and *Staphylococcus aureus* [6]. Taken together, the past decade of atomic-resolution crystal structures [6,7,20,23,25,27,28,29] and biophysical characterization of RNA-binding to Hfq [15,20,21,23,30,31,32] have enabled the development of various mechanistic models for the RNA chaperone activity of Hfq [33]. Notably, not all of these models are mutually exclusive. A limitation of the current models is that they are derived from studies of Hfq from a very limited number of species (e.g. *E. coli*). Therefore, additional studies of phylogenetically distant homologs would significantly broaden our understanding of the functional mechanism of Hfq.

To date there is no structure of an Hfq protein from a thermophilic or a phylogenetically deep-branching bacterium. *Thermotoga maritima* represents one of the deepest lineages in the

bacterial domain of life [34,35] and was the first hyper-thermophilic bacteria discovered [34], making it an interesting species for evolutionary studies of the structure/function relationships of Hfq proteins as well as the Sm superfamily more broadly [36,37]. In this chapter, the crystal structure of *Tma* Hfq is reported to 2.1 Å resolution, the protein having been successfully carried through our cloning, expression, purification, and crystallization workflow.

Methods

Cloning, expression, and purification. Our protocols for the cloning, over-expression, and purification of recombinant *Tma* Hfq are described in Chapter 2. For crystallographic studies, our initial purification efforts used the protocol as described for oligonucleotide (nanoRNA) purification in Chapter 2, with the exception that fractions corresponding to peak 1 in the anion exchange column (with a A_{260}/A_{280} ratio ≤ 0.85) were combined and used for crystallization screens. Sample purity was assessed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), using the protocol described in Chapter 2.

In preparation for crystallization screens, purified and pooled *Tma* Hfq was dialyzed into 25 mM Tris pH 8.5, 100 mM NaCl. Protein concentration was determined by measuring the absorbance at 280 nm using a theoretical extinction coefficient of 5960 $M^{-1}cm^{-1}$ based on amino acid composition per monomer. Protein samples were concentrated to 15.3 mg/mL using Millipore centrifugal filter units with a molecular weight cut-off of 3350 Da following the manufacturer's protocol.

Chemical crosslinking assays. Purified *Tma* Hfq was dialyzed into 25 mM HEPES pH 8.0 and 200 mM NaCl and crosslinked with 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC), using the direct cross-linking protocol described in Chapter 3. The extent of crosslinking (monomers,

dimers, hexamers, etc.) was determined via MALDI-TOF MS using the protocol described in Chapter 2.

Crystallization. Crystallization trials were pursued using the hanging-drop vapor diffusion method [38]. Commercial sparse matrix screens (JCSG Core Suite 1-4 (Qiagen) and PEG/Ion 1-2 (Hampton Research)) were utilized to identify potential conditions for crystallization. Screens were set-up using a small volume liquid-handling robot (Mosquito, TTC Inc) in 96-well plates; the experiments utilized 100 μl of mother liquor and a 200 nL drop of 1:1 *Tma* Hfq:mother liquor. The conditions for initial leads were further refined using 24-well grid screens (parameters scanned: pH, salt concentration, buffer concentration, and PEG concentration) and a 96-well additive screen (HR2-428, Hampton Research). Two optimal conditions were identified: 150 mM tri-potassium citrate (pH was not adjusted, but was measured to be pH 8.28) and 30% w/v PEG-3350, with either glycine or sarcosine as an additive at a final concentration of 100 mM or 10 mM, respectively. Grid screens were also set-up using 28 nM U_5 RNA with a 5' monophosphate and 3' hydroxyl and 1.42 mM *Tma* Hfq (15 mg/mL), centered around these two optimal crystallization conditions. Diffraction quality crystals were obtained after two weeks of incubation at 17 ˚C.

For diffraction data collection at cryogenic temperatures, cryoprotectant conditions were screened using a Rigaku MicroMax 007 generator with a Saturn 944 CCD detector. The optimal cryoprotectant for both of the above crystallization conditions was determined by incubating crystals for approximately 10-15 seconds in 8 μL of mother liquor spiked with 0.6 μL 100% v/v PEG-400.

X-ray diffraction studies. Crystals were rapidly transferred from the hanging drop in which they formed to the artificial cryoprotectant mother liquor using nylon loops. Crystals were incubated in the cryoprotectant for 10-15 seconds and then rapidly flash-cooled in a bath of liquid nitrogen

(T≈77K). Protein crystals were shipped to the Advanced Photon Source at Argonne National Laboratories for synchrotron X-ray diffraction data collection. Data sets were collected on the Southeast Regional Collaborative Access Team (SER-CAT) 22ID and 22BM beamlines, equipped with MAR 300 and MAR 225 charge-coupled device (CCD) detectors, respectively. Diffraction data sets were collected for a full 360° crystal rotation at a detector distance of 300 mm with 1° oscillations; the X-ray wavelength was maintained at 0.9879 Å for these experiments.

Tma Hfq crystals grown in the presence of U_5 RNA were shipped to the Advanced Photon Source and data sets were collected on the Northeastern Collaborative Access Team (NE-CAT) 24ID-C beamline, using an ADSC Q315 3x3 CCD detector. These data-sets were collected with a 180° crystal rotation at a detector distance of 300 mm and with 0.5° oscillations.

Data processing. X-ray diffraction data were indexed, integrated and scaled in HKL2000 [39]. CTRUNCATE was used to convert the Scalepack output to an MTZ file [40,41]. The Matthews coefficient   (V_M=2.08 $Å^3$/Da) and non-crystallographic symmetry were examined using the Phaser module of the PHENIX software package [42]. In preparation for molecular replacement, the structure of *Bacillus subtilis* Hfq (3HSB) was trimmed using Sculptor [42], preserving only side-chains of conserved residues based on se   quence alignment with *Tma* Hfq. This trimmed model was subsequently used in Phaser as a probe model for molecular replacement phasing. The structural model was refined using REFMAC5 [43] and the phenix.refine module. Validation of the refined structural model included inspection of the Ramachandran plot for outliers. Model building and structure manipulation steps were performed in the Coot software environment [44,45].

Molecular modeling & analysis. Three-dimensional structures were visualized using Coot and PyMOL [46]. For comparison of surface electrostatic potentials, electrostatic potentials calculations were performed using the Adaptive Poisson-Boltzmann Solve (APBS) plugin to

PyMol [12,47]. For comparison of hydrophobicity, residues were colored according to the normalized consensus hydrophobicity scale using the color_h script (http://www.pymolwiki.org/index.php/Color_h) [48,49]. The 3D crystal structures of *Tma* Hfq, *E. coli* Hfq (PDB = 1HK9), and *S. aureus* Hfq (PDB = 1KQ2) were used for molecular modeling comparisons.

Results

Purification and oligomerization of *Tma* Hfq

Sequencing of the expression vector by Genewiz confirmed that *Tma* Hfq was properly cloned with no mutations. The recombinant protein construct is shown in Figure 4.2. This construct contains a N-terminal (His)$_6$ tag, which was employed for affinity chromatography and subsequently cleaved using thrombin. MALDI-TOF MS analysis of the purified protein (Figure 4.3a) verified that cleavage of the His-tag was complete, with no spurious cleavage of the *Tma* Hfq polypeptide.

The oligomeric state of *Tma* Hfq was assessed by chemical crosslinking with EDC for four hours. MALDI-TOF MS analysis of crosslinked *Tma* Hfq samples (Figure 4.3b) indicated the presence of two distinct species with molecular masses of 68.0 ± 0.6 kDa and 136 ± 2kDa, which correspond to the masses of a hexamer and dodecamer, respectively.

Structure determination and description

Preliminary crystallization conditions were determined by sparse-matrix screening. Initial crystals were found to be sensitive to mechanical manipulation and environmental changes (e.g. changes in temperature), making them difficult to cryo-loop or transport. Grid screens and additive screens were utilized to further optimize the initial conditions to reproducibly yield robust, well-diffracting crystals. The addition of glycine or sarcosine (to a final concentration of

100 mM or 10 mM, respectively) to the hanging drop was shown to substantially increase the overall visual quality of the crystals formed in that drop (Figure 4.4a).

Diffraction data (Figure 4.4b) were collected to a resolution of 2.6 Å for a *Tma* Hfq crystal (Figure 4.4a) grown under the following conditions: 150 mM tri-potassium citrate and 30% w/v PEG-3350 with glycine as an additive in the hanging drop at a final concentration of 100 mM. The data could be successfully indexed and scaled in space group $P2_12_12_1$ with Hfq packed as two homohexamer rings per asymmetric unit (Figure 4.5a). The rings arrange to give an overall crystalline lattice that contains parallel fibers of the Hfq hexamer (Figure 4.5b). Within each fibrillar tube, the Hfq hexamers stack head to tail, and neighboring fibers run in opposite directions within a single crystal lattice layer (Figure 4.5b). The X-ray diffraction statistics for the apoprotein crystal are summarized in Table 4.1.

The structure of *Tma* Hfq (Figure 4.4c,d) was determined by molecular replacement using *B. subtilis* Hfq (3HSB) as the probe to a resolution of 2.6 Å. The final structural model converged to R_{work} and R_{free} values of 0.21 and 0.23, respectively. In the final structure, chains A, E, and F contained electron density for residues R11 to M77, chain B contained electron density for residues F10 to P79, and chain D contained electron density for residues R11 to L76 of the *Tma* Hfq construct (Figure 4.2). The Ramachandran plot shows that in the final refined model 95.87% of residues are in preferred regions and 4.13% of residues are in allowed regions of Φ, ψ space (Figure 4.6a).

The hexameric ring exhibits a diameter of 68 Å and a 12 Å pore, consistent with other Hfq structures [4,7,12,27]. The *Tma* Hfq monomer contains an N-terminal α-helix followed by a β-sheet with the standard β5-β1-β2-β3-β4 topology of the Sm fold (Figure 4.4c). The subunit interface is formed by strands β4 and β5 of adjacent monomers (Figure 4.1). Residues L12 to N23 form the N-terminal α-helix, V27 to L31 form β1, G34 to F44 form β2, T48 to S53 form β3,

Q56 to Y61 form β4, and S66 to P70 form β5. The β-sheet is twisted due to the curvature of strand β2 (Figure 4.1).

Crystallization of *Tma* Hfq in the presence of U$_5$ RNA

Diffraction data (Figure 4.7b) were collected to a resolution of 2.1 Å for a crystal (Figure 4.7a) grown with 1.42 mM Hfq monomer (15.3 mg/mL) in the presence of 28 nM U$_5$ RNA under the following conditions: 130 mM Tri-potassium citrate and 32% w/v PEG 3350, with glycine as an additive in the hanging drop at a final concentration of 100 mM. The inhibitory constant (K_i) of U$_5$ RNA for *Tma* Hfq was determined in Chapter 2 to be 6 ± 1 nM (in terms of Hfq monomer) via fluorescence polarization using 5-Carboxyfluorescein labeled U$_6$ as a probe (Table 2.1). Therefore, under the conditions described earlier in this paragraph the RNA was used as more of an additive to improve the quality of the crystal, rather than in a deliberate attempt to determine the RNA-bound *Tma* Hfq structure.

The X-ray diffraction statistics for the U$_5$ RNA containing crystal are summarized in Table 4.2. The structure of *Tma* Hfq crystallized in the presence of U$_5$ RNA was determined by molecular replacement, as described above, to a resolution of 2.1 Å. The final structure of *Tma* Hfq (Figure 4.7c,d) in the presence of U$_5$ RNA converged to R_{work} and R_{free} values of 0.20 and 0.22, respectively. Although the crystallization condition contained U$_5$ RNA, the RNA was not found in difference electron density (Figure 4.7c,d). The Ramachandran plot shows that in the refined model 97.28% of residues (in the dodecamer found in the asymmetric unit) are in preferred regions, 2.59% of residues are in allowed regions, and 0.13% of residues are in a sterically disallowed regions of Φ, ψ space (Figure 4.6b). Notably, the single Ramachandran outlier (M75) is the final C-terminal residue that was built into electron density for its respective chain; unsurprisingly, electron density maps in this C-terminal region were less intense/unambiguous than elsewhere in the polypeptide chain.

Discussion

Tma Hfq was purified using a multistep purification that utilized a heat treatment, immobilized metal affinity chromatography (IMAC) under denaturing conditions (6M guanidinium-HCl), and gel filtration to remove both protein and nucleic acid contaminants. The addition of gnd-HCl to the protein purification protocol, specifically during the IMAC step, drastically decreased the extent of nucleic acid contamination. Notably, an earlier purification protocol that utilized IMAC under more native conditions (i.e. no Gnd-HCl) yielded Tma Hfq protein samples that crystallized under similar conditions, and with similar unit cell parameters (suggesting the two crystal specimens are isomorphous). The extent of nucleic acid contamination was assessed by the 260nm/280nm absorbance ratio, which was around 1.2 or higher for samples purified under native conditions, versus 0.85 or lower when purified in the presence of gnd-HCl. Samples with large amounts of nucleic acid contaminants resulted in crystals with poor diffraction quality (high mosaicity). Interestingly, the rate of crystal formation was significantly affected by the extent of RNA contamination: Crystals formed almost immediately when contaminated with RNA versus taking two weeks or more when RNA contamination was diminished.

Tma Hfq crystals grown in the presence of U_5 RNA diffracted to a resolution of 2.1 Å. Despite the addition of U_5 to the crystallization condition, no electron density corresponding to RNA was observed and the overall structure was not perturbed. Co-crystallization efforts are ongoing because a co-crystal structure could provide substantial new information regarding the interatomic interactions between *Tma* Hfq and nanoRNAs.

The crystal structure of *Tma* Hfq (Figure 4.4c and Figure 4.7c) exhibits the Sm fold that is characteristic of Hfq proteins. This is an expected result, as molecular replacement was used for crystallographic phasing. Structural alignment of *Tma* Hfq with Hfq from *E. coli* (1HK9) and *S. aureus* (1KQ2) indicated an RMSD of 0.286 Å and 0.373 Å, respectively. In contrast, sequence

121

alignment of *Tma* Hfq with *Eco* Hfq and *Sau* Hfq indicated that *Sau* Hfq and *Tma* Hfq share slightly higher sequence identity (53% vs 48%).

Comparison of the surface hydrophobicity of *Tma* Hfq, *Sau* Hfq, and *Eco* Hfq indicates that, for the most part, the localization of hydrophobic residues is conserved in all three structures, with the extent of hydrophobicity varying most on the lateral face (Figure 4.8). Distinctly, *Eco* Hfq contains hydrophilic patches on its lateral face that spans between the distal and proximal face; these patches are not observed in either *Tma* Hfq or *Eco* Hfq, which are predominantly hydrophobic on their lateral face.

Comparison of the molecular electrostatic potentials of Hfq from *T. maritima, S. aureus,* and *E. coli* (Figure 4.9) reveal differences between the three structures that are less apparent from sequence and structure alignments. The distal face (Figure 4.9, Top panel) of *Tma* Hfq has six isolated regions with strongly positive electrostatic potentials oriented around the central pore. In contrast, the negative electrostatic potential is evenly distributed on the distal face of *Eco* Hfq, and *Sau* Hfq has a more negative electrostatic potential oriented around the periphery of the ring. Binding data indicate that *Tma* Hfq has a weaker affinity than both *Sau* Hfq and *Eco* Hfq for poly-A RNA, which has been shown to bind to the distal face of Hfq in other organisms [31]. The dissociation constant (K_d) of *Tma* Hfq for A_{18} is 236 ± 38 nM (in terms of Hfq monomers) as reported in Chapter 3, whereas *Eco* Hfq and *Sau* Hfq have a K_d for A_{16} of 0.4 ± 0.5 nM and 4.2 ± 0.5 nM, binding respectively [21]. In order to further elucidate the weaker of *Tma* Hfq to A_{18}, co-crystallization of *Tma* Hfq with A_{18} RNA could be pursued to determine the detailed conformation of RNA in the distal RNA-binding site of *Tma* Hfq. The orientation of A_{18} RNA in the binding site could reveal the RNA sequence motifs for poly-A RNA recognition by *Tma* Hfq. The identified motif can then by compared to the poly-A RNA recognition sequences of Gram-negative $((ARN)_n)$ [20] and Gram-positive $((AN)_n)$ [21] Hfq homologs to determine which

class of Hfq homologs *Tma* Hfq is more similar to in regards to A-rich RNA binding, or if *Tma* Hfq has a distinct recognition sequence from those previously identified.

The lateral rim of each of these hexamers (Figure 4.9, middle panel) is fairly distinct; *Sau* Hfq is highly negatively charged in this region, whereas *E. coli* has positively charged patches that span between the distal and proximal face. The lateral face of *Tma* Hfq has less defined positive patches that do not make a continuous path between the distal to proximal face. It has been proposed that the lateral surface of *Eco* Hfq may be involved in orienting sRNA and mRNA [32]. This proposal was presaged in the earlier work of Sun & Wartell, which identified a rim mutation (R16A) important in DsrA binding [30]. It has since been shown that R16, R17, E18, and R19 are all involved in RNA interactions at the lateral surface of *E. coli* Hfq [50]. An alignment of 384 Hfq sequences revealed that these residues are highly conserved, with position 16 being nearly invariant, position 17 typically containing an R or K, and an R, K, or N frequently in position 19 [50]. The equivalent surface in *Tma* Hfq consists of R18, V19, N20, and K21, suggesting that the lateral surface of *Tma* Hfq may have interesting binding properties. Future studies could concentrate on characterizing the lateral binding site through the use of *Tma* Hfq mutants and strategically designed RNA-probes to determine the implication of these sequence variations on RNA-binding at the lateral surface. Additionally, longer RNA ligands that extend into the lateral surface can be utilized in co-crystallization efforts to develop an atomic-level structural model for RNA-binding at the lateral surface of *Tma* Hfq.

Both *Eco* Hfq and *Tma* Hfq have a predominantly positive electrostatic potential on their proximal faces (Figure 4.9, bottom panel), whereas *Sau* Hfq has negatively charged regions along the periphery of the proximal face and positively charged patches oriented around the central pore. The considerably higher positive electrostatic potential of the proximal face of *Tma* Hfq is consistent with our determination of a higher affinity for poly-U RNA (Chapter 2), which has

been shown to bind to the proximal face [31]. The K_d of Tma Hfq for FAM-U$_6$ is 10 ± 2 nM as

reported in Chapter 3, whereas Eco Hfq and Sau Hfq have a K_d for U$_6$ of 766.0 ± 73.7 nM [51]

and 69.8 ± 7.0 nM [21,51], respectively. Interestingly, Tma Hfq was also shown to have a

nanomolar affinity for U/C-rich nanoRNAs (5-6 nucleotides long) that were found to co-purify

with recombinant Tma Hfq over-expressed in $E.\ coli$ (Chapter 2). Future studies could focus on

co-crystallizing Tma Hfq with these previously identified nanoRNAs to determine their

orientation in the proximal RNA binding site. These studies could identify the RNA sequence

motif and end chemistries that preferentially associate with the proximal RNA binding surface of

Tma Hfq.

Crystal structures of Sau Hfq with AU$_5$G RNA (PDB 1KQ2), $Salmonella\ typhimurium$ Hfq

(Sty Hfq) with U$_6$ RNA (PDB 2YLC), and $Eco\ Hfq$ with A$_7$ and AU$_6$A RNA (PDB 4HT9) have revealed

that the amino acid residues Q8, K/Q41, F/Y42, K56, and H57 (number based on the Eco Hfq

sequence) are directly involved in RNA-binding and selection at the proximal RNA-binding

surface [6,23,29]. Furthermore, the crystal structure of Sty Hfq bound to U$_6$ RNA revealed that a

hydrogen bond could form between H57 and the 3' hydroxyl of this RNA, thereby stabilizing the

Hfq-RNA interaction [23]. This finding suggested that Hfq preferentially binds to the 3' hydroxyl

end of RNA. The importance of a 3' hydroxyl interaction was further supposed by $in\ vitro$

binding studies that indicated the presence of a propyl-phosphate group or a 2'-3' cyclic

phosphate significantly hindered RNA-binding at the proximal RNA-binding site of Sty Hfq [23].

High-resolution crystallographic studies could be pursued with Tma Hfq and the U/C-rich

nanoRNAs described in Chapter 2, in order to elucidate the amino acid residues involved in

nanoRNA recognition and to further evaluate the mode of RNA 3' end recognition by Hfq

proteins. Such studies could have implications for the function of nanoRNAs in Hfq-mediated

mRNA-sRNA annealing as well as the possible roles of Hfq in nanoRNA metabolism.

References

1. Hvidsten TR, Laegreid A, Kryshtafovych A, Andersson G, Fidelis K, et al. (2009) A Comprehensive Analysis of the Structure-Function Relationship in Proteins Based on Local Structure Similarity. Plos One 4.

2. Kinoshita K, Nakamura H (2003) Protein informatics towards function identification. Current Opinion in Structural Biology 13: 396-400.

3. Voet D, Voet JG (2004) Biochemistry. Hoboken: John Wiley & Sons, Inc.

4. Zhang A, Wassarman KM, Ortega J, Steven AC, Storz G (2002) The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. Mol Cell 9: 11-22.

5. Nikulin A, Stolboushkina E, Pederderina A, Vassilieva I, Blaesi U, et al. (2005) Structure of *Pseudomonas aeruginosa* Hfq protein. Acta Crystallogr D Biol Crystallogr 61: 141-146.

6. Schumacher MA, Pearson RF, Moller T, Valentin-Hansen P, Brennan RG (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. EMBO J 21: 3546-3556.

7. Sauter C, Basquin J, Suck D (2003) Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*. Nucleic Acids Research 31: 4091-4098.

8. Seraphin B (1995) Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. EMBO J 14: 2089-2098.

9. Branlant C, Krol A, Ebel JP, Lazar E, Haendler B, et al. (1982) U2 RNA shares a structural domain with U1, U4, and U5 RNAs. EMBO J 1: 1259-1265.

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

11. Sauer E (2013) Structure and RNA-binding properties of the bacterial LSm protein Hfq. RNA Biol 10: 610-618.

12. Moller T, Franch T, Hojrup P, Keene DR, Bachinger HP, et al. (2002) Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. Mol Cell 9: 23-30.

13. Sun X, Zhulin I, Wartell RM (2002) Predicted structure and phyletic distribution of the RNA-binding protein Hfq. Nucleic Acids Res 30: 3662-3671.

14. Arluison V, Derreumaux P, Allemand F, Folichon M, Hajnsdorf E, et al. (2002) Structural Modelling of the Sm-like Protein Hfq from Escherichia coli. J Mol Biol 320: 705-712.

15. Salim NN, Faner MA, Philip JA, Feig AL (2012) Requirement of upstream Hfq-binding (ARN)x elements in glmS and the Hfq C-terminal region for GlmS upregulation by sRNAs GlmZ and GlmY. Nucleic Acids Res 40: 8021-8032.

16. Soper TJ, Woodson SA (2008) The rpoS mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. Rna-a Publication of the Rna Society 14: 1907-1917.

17. Salim NN, Feig AL (2010) An upstream Hfq binding site in the fhlA mRNA leader region facilitates the OxyS-fhlA interaction. PLoS One 5.

18. Hajnsdorf E, 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.

19. Hajnsdorf E, Regnier P (1999) E. coli RpsO mRNA decay: RNase E processing at the beginning of the coding sequence stimulates poly(A)-dependent degradation of the mRNA. J Mol Biol 286: 1033-1043.

20. Link TM, Valentin-Hansen P, Brennan RG (2009) Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. Proceedings of the National Academy of Sciences of the United States of America 106: 19286-19291.

21. Horstmann N, Orans J, Valentin-Hansen P, Shelburne SA, 3rd, Brennan RG (2012) Structural mechanism of *Staphylococcus aureus* Hfq binding to an RNA A-tract. Nucleic Acids Res 40: 11023-11035.

22. Sittka A, Sharma CM, Rolle K, 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.

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

24. Kadowaki MA, Iulek J, Barbosa JA, Pedrosa Fde O, de Souza EM, et al. (2012) Structural characterization of the RNA chaperone Hfq from the nitrogen-fixing bacterium *Herbaspirillum seropedicae* SmR1. Biochim Biophys Acta 1824: 359-365.

25. Boggild A, Overgaard M, Valentin-Hansen P, Brodersen DE (2009) Cyanobacteria contain a structural homologue of the Hfq protein with altered RNA-binding properties. FEBS J 276: 3904-3915.

26. Baba S, Someya T, Kawai G, Nakamura K, Kumasaka T (2010) Expression, crystallization and preliminary crystallographic analysis of RNA-binding protein Hfq (YmaH) from *Bacillus subtilis* in complex with an RNA aptamer. Acta Crystallogr Sect F Struct Biol Cryst Commun 66: 563-566.

27. Nikulin A, Stolboushkina E, Perederina A, Vassilieva I, Blaesi U, et al. (2005) Structure of Pseudomonas aeruginosa Hfq protein. Acta Crystallographica Section D-Biological Crystallography 61: 141-146.

28. Nielsen JS, Boggild A, Andersen CB, Nielsen G, Boysen A, et al. (2007) An Hfq-like protein in archaea: crystal structure and functional characterization of the Sm protein from Methanococcus jannaschii. RNA 13: 2213-2223.

29. Wang W, Wang L, Wu J, Gong Q, Shi Y (2013) Hfq-bridged ternary complex is important for translation activation of rpoS by DsrA. Nucleic Acids Res 41: 5938-5948.

30. Sun XG, Wartell RM (2006) Escherichia coli Hfq binds A(18) and DsrA domain II with similar 2 : 1 Hfq(6)/RNA stoichiometry using different surface sites. Biochemistry 45: 4875-4887.

31. Mikulecky PJ, Kaw MK, Brescia CC, Takach JC, Sledjeski DD, et al. (2004) Escherichia coli Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. Nat Struct Mol Biol 11: 1206-1214.

32. Sauer E, Schmidt S, Weichenrieder O (2012) Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. Proceedings of the National Academy of Sciences of the United States of America 109: 9396-9401.

33. Wagner EG (2013) Cycling of RNAs on Hfq. RNA Biol 10: 619-626.

34. Huber R, Langworthy TA, Konig H, Thomm M, Woese CR, et al. (1986) *Thermotoga-Maritima* Sp-Nov Represents a New Genus of Unique Extremely Thermophilic Eubacteria Growing up to 90-Degrees-C. Archives of Microbiology 144: 324-333.

35. Woese CR (1987) Bacterial Evolution. Microbiological Reviews 51: 221-271.

36. Sun XG, Wartell RM (2003) Predicted structures and phyletic distribution of Hfq, a RNA-binding protein. Biophysical Journal 84: 459a-459a.

37. Mura C, Randolph PS, Patterson J, Cozen AE (2013) Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function. RNA Biol 10: 636-651.

38. McPherson A (1999) Crystallization of Biological Macromolecules: Cold Spring Harbor Laboratory Press. 586 p.

39. Otwinowski Z, Minor W (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode; Carter CWJ, Sweet RM, editors. New York: Academic Press.

40. French S, Wilson K (1978) On the treatment of negative intensity observations. Acta Crystal A 34: 517-525.

41. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, et al. (2011) Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 67: 235-242.

42. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66: 213-221.

43. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, et al. (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 67: 355-367.

44. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallographica Section D-Biological Crystallography 60: 2126-2132.

45. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallographica Section D-Biological Crystallography 66: 486-501.

46. DeLano WL (2002) The PyMOL molecular graphics system.

47. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A 98: 10037-10041.

48. Eisenberg D, Schwarz E, Komaromy M, Wall R (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J Mol Biol 179: 125-142.

49. Hermann H, Fabrizio P, Raker VA, Foulaki K, Hornig H, et al. (1995) snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. EMBO J 14: 2076-2088.

50. Panja S, Schu DJ, Woodson SA (2013) Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. Nucleic Acids Res 41: 7536-7546.

51. Robinson KE, Orans J, Kovach AR, Link TM, Brennan RG (2014) Mapping Hfq-RNA interaction

surfaces using tryptophan fluorescence quenching. Nucleic Acids Res: 2736-2749.
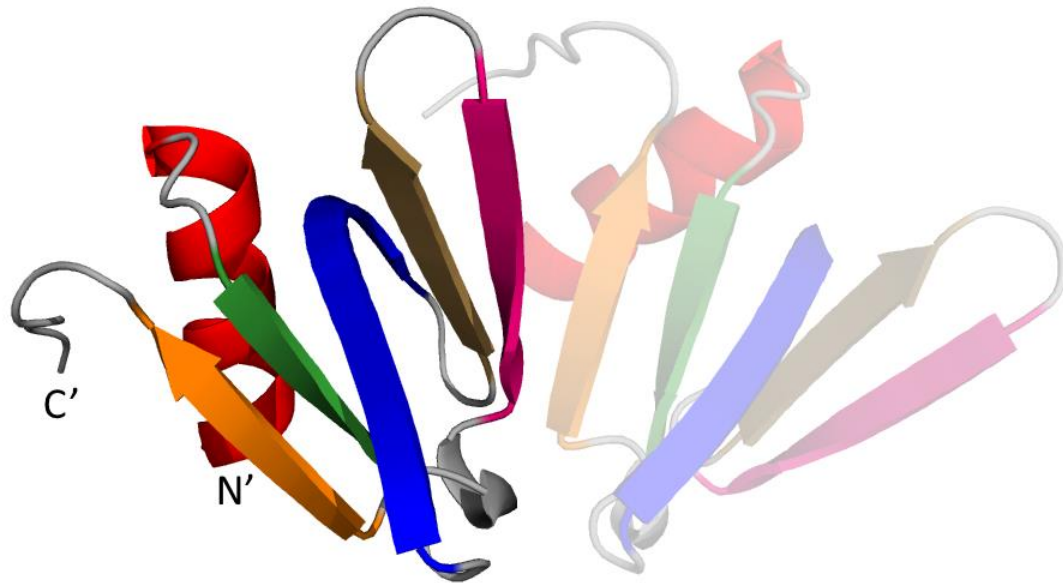
Figure 4.1 Tertiary structure of *Tma* Hfq with an adjacent subunit in the Hfq. The *Tma* Hfq monomer consists of an N-terminal α-helix follow by a bent β-sheet with β5- β1-β2-β3-β4 topology. The α-helix spans from L12 to N23, β1 consists of residues V27 to L31, β2 consisting G34 to F44, β3 consists of T48 to S53, β4 consists of residues Q56 to Y61, and β5 consists of S66 to P70. The interface between adjacent monomers is between β4 (subunit *n*) and β5 (subunit *n*+1).

```
        10              20              30              40
mgsshhhhhh ssslvprgsh MALAEKFNLQ DRFLNHLRVN

        50              60              70              80
KIEVKVYLVN GFQTKGFIRS FDSYTVLLES GNQQSLIYKH

        90             100             110
AISTIIPSSY VMLMPKKQET AQEAETSENE GS
```

Figure 4.2 The recombinant *Tma* Hfq construct used for crystallization. A His$_6$-tag (red) was genetically appended to the protein for affinity purification, and then removed by thrombin digestion (**recognition sequence**).   The cleaved construct (underlined) was used for crystallization, which contained three additional amino acid residues on the N-terminus of the protein that are not present in the native protein (the wild-type protein sequence is CAPITALIZED).
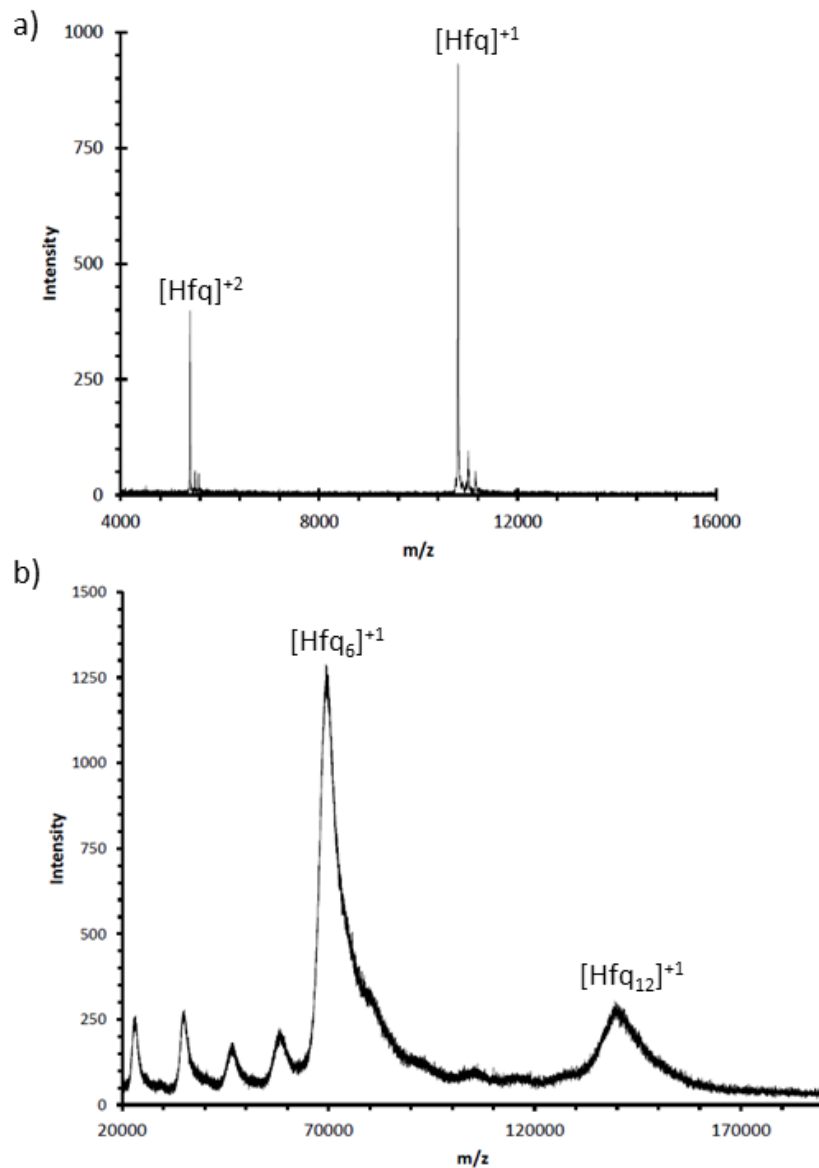
Figure 4.3 MALDI-TOF MS spectrum of purified (a) and crosslinked (b) *Tma* Hfq. In part (a), the two peaks correspond to the +1 and +2 charge states of the *Tma* Hfq monomer. The z=+1 charge state is within 0.01% of the molecular weight predicted by the amino acid sequence. In part (b), *Tma* Hfq was crosslinked with EDC for four hours at room temperature. The predominant peaks in the spectra correspond to a hexamer and dodecamer. Lower molecular weight peaks in the crosslinked Hfq spectra correspond to lower level oligomeric states, suggesting that cross-linking did not go to completion.

133
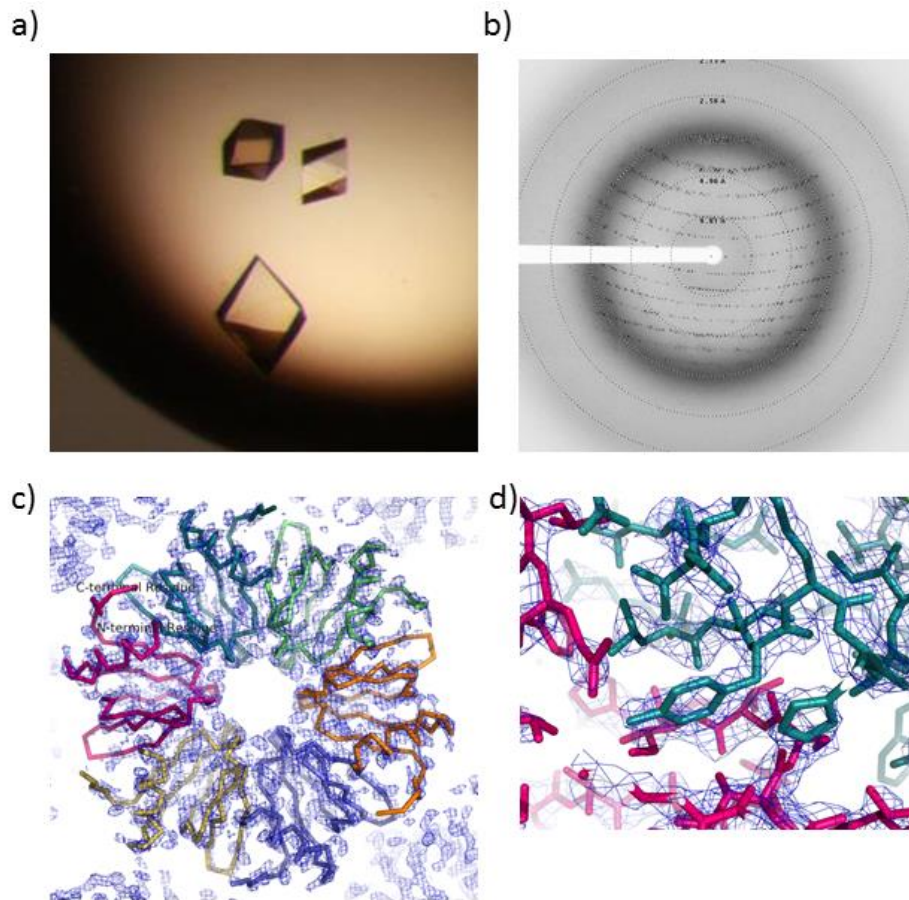
Figure 4.4 Crystals (a), diffraction pattern (b), and crystal structure (c and d) of *Tma* Hfq. (a)

Crystals were grown at 17 °C for two weeks using the hanging-drop vapor-diffusion method

under the following conditions: 15.3 mg/mL *Tma* Hfq with 25 mM Tris pH 8.5, 100mM NaCl and

150 mM tri-potassium citrate and 30% w/v PEG-3350 as the mother liquor and glycine as an

additive in the hanging drop at a final concentration of 100 mM. The crystals were harvested

using nylon loops and cryoprotected using mother liquor supplemented with 4.4% v/v PEG-400.

(b) Diffraction data was collected on the SER-CAT 22-ID beamline at the Advanced Photon

Source to a resolution of 2.6 Å. (c) Crystal structure of *Tma* Hfq shown as a ribbon model with

each chain in the hexamer shown as a different color and the electron density map ($2F_o$-F_c)

shown as a blue mesh contoured at 2.0 [sigma]. A zoomed in view of the electron density map

is shown in part (d).
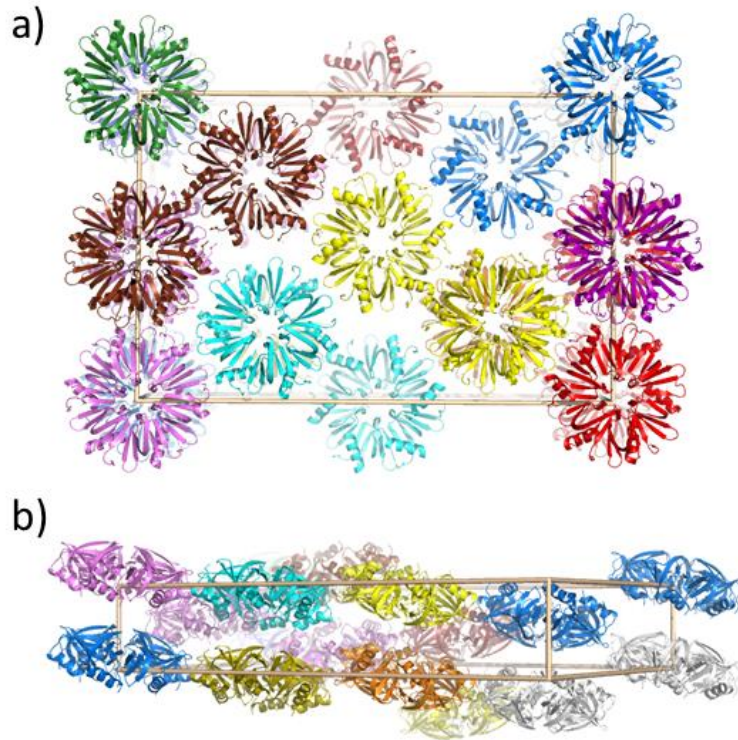
Figure 4.5 The *Tma* Hfq crystal lattice. Axial (a) and lateral (b) view of the *Tma* Hfq crystal lattice. The contents of one unit cell are shown. Hfq's from the same asymmetric unit are colored the same, with the unit cell shown as a box. Hfq hexamers stack head to tail within the crystal lattice and pack to form anti-parallel fibers.
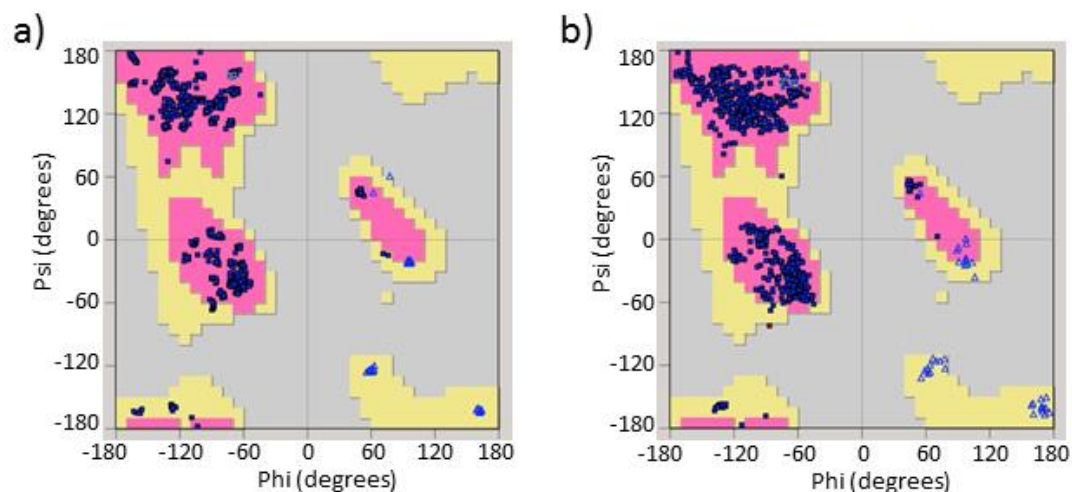
Figure 4.6 Ramachandran plot for the refined models solved from *Tma* Hfq crystals grown in the absence (a) and presence (b) of U_5 RNA. In the Ramachandran plot above, the grey areas correspond to regions of the diagram that are sterically disallowed for all amino acids except glycine (triangles), the pink areas correspond to preferred conformations, and the yellow areas correspond to regions that are allowed. For *Tma* Hfq crystals grown in the absence of RNA (a), the refined model has 95.87% of residues in preferred regions and 4.13% of residues in allowed regions of Φ, ψ space. For *Tma* Crystals grown in the presence of U_5 RNA (b), the refined model has 97.28% of residues in preferred regions, 2.59% of residues in allowed regions, and 0.13% of residues in a sterically disallowed regions of Φ, ψ space.
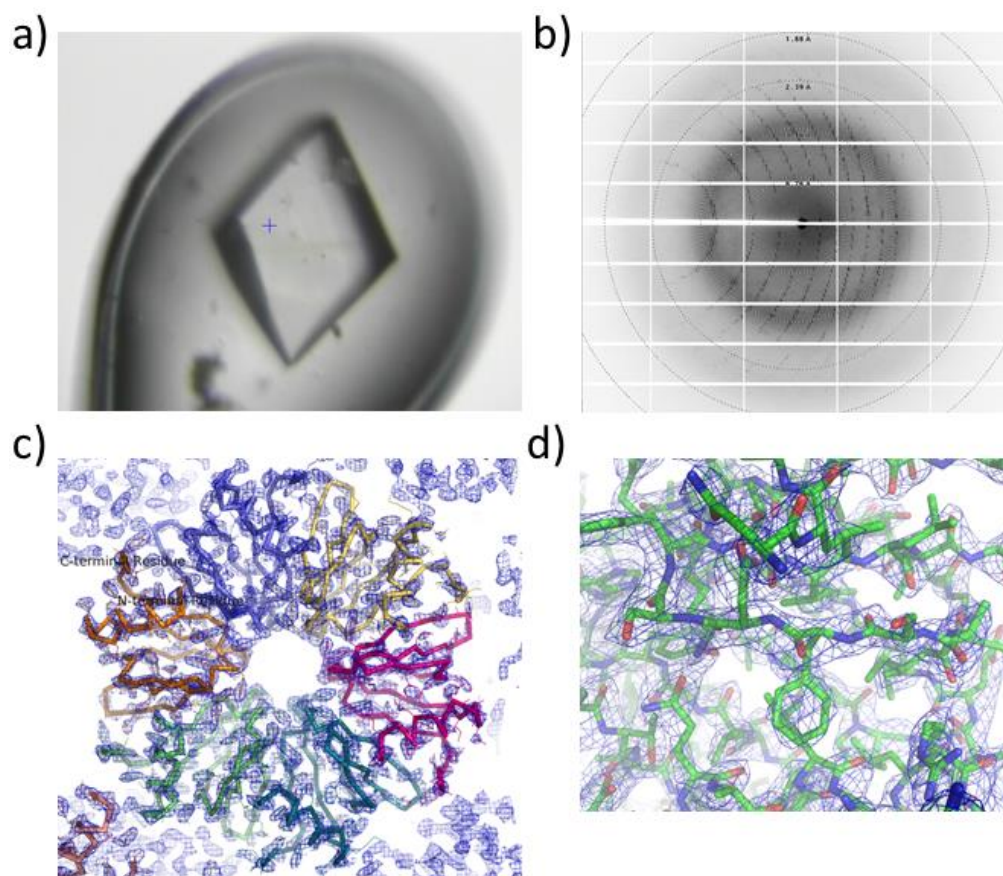
Figure 4.7 Crystal (a), diffraction pattern (b), and crystal structure (c) of *Tma* Hfq in presence of

U$_5$ RNA. (a) Image of crystal in nylon loop prior to collection of diffraction data. The crystal was

grown at 17 °C for 2 weeks using the hanging-drop vapor-diffusion method in the following

conditions: 15.3 mg/mL of *Tma* Hfq and 28 nM U$_5$ RNA with 25 mM Tris pH 8.5, 100mM NaCl

and 130 mM tri-potassium citrate and 32% w/v PEG-3350 as the mother liquor and glycine as an

additive in the hanging drop at a final concentration of 100 mM. The crystals were harvested

using nylon loops and cryoprotected using mother liquor supplemented with 4.4% v/v PEG-400.

(b) Diffraction data was collected on the NE-CAT 24ID-C beamline at the Advanced Photon

Source to a resolution 2.1 Å. (c) Crystal structure of *Tma* Hfq shown as a ribbon model with each

chain in the hexamer shown as a different color and the electron density map (2F$_o$-F$_c$) shown as

a blue mesh contoured at 2.0 [sigma]. A zoomed in view of the electron density map is shown in

part (d).
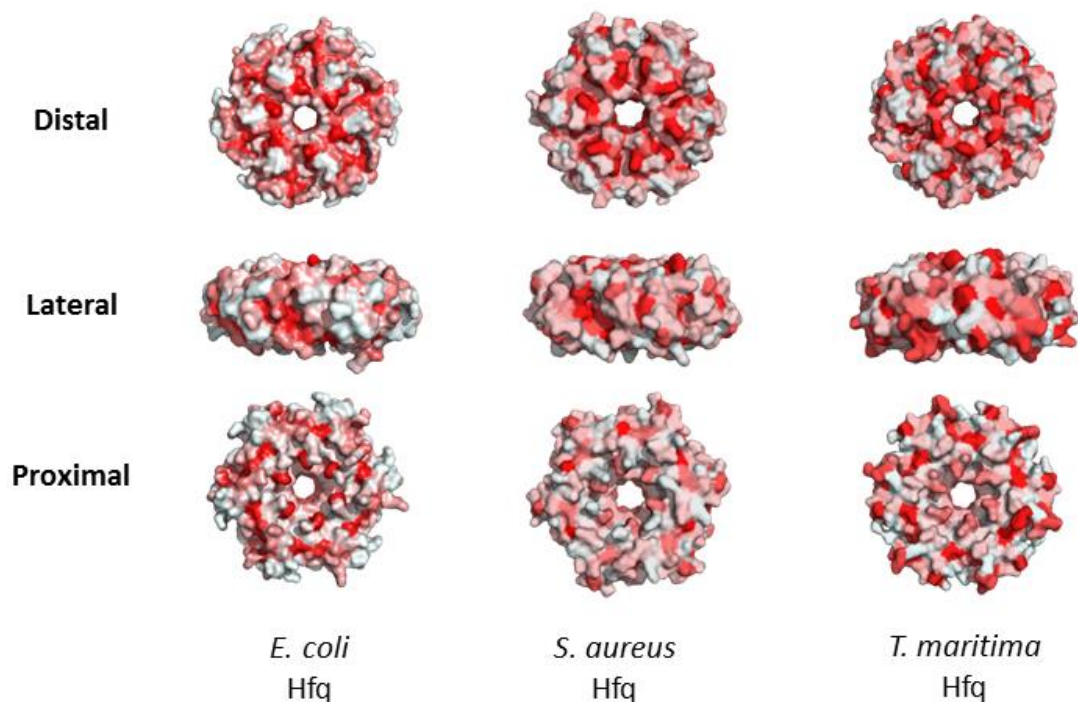
Figure 4.8 Surface hydrophobicity of Hfq from three distinct species. The surface hydrophobicity is shown for *Eco* Hfq, *Sau* Hfq, and *Tma* Hfq as seen from distal (top), lateral (middle), and proximal (bottom) views. Comparison of surface hydrophobicity of *Tma* Hfq, *Sau* Hfq, and *Eco* Hfq indicates that, for the most part, the localization of hydrophobic residues in the 3D structure of Hfq is well conserved with the extent of hydrophobicity on the lateral surface varying most between the three homologs. Hydrophobicity is visualized using a color scale from red to white in which the most hydrophilic residues are colored white and the most hydrophobic residues are in red. The color scale was generated with the color_h script [48] and visualized using PyMOL [46].
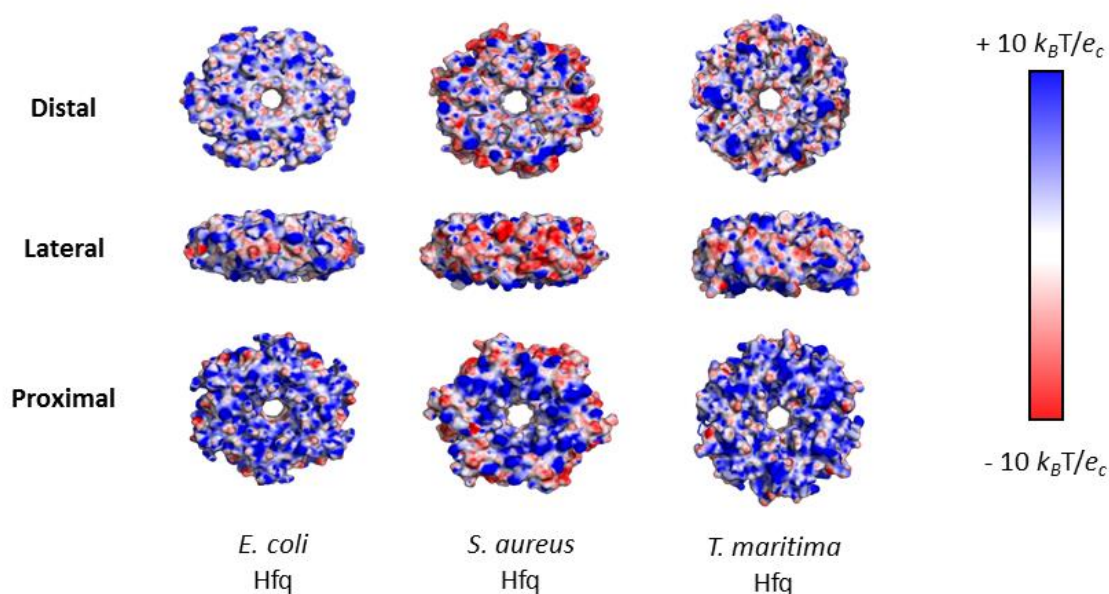
Figure 4.9 Surfaces of Hfq homologs from three distinct species are colored by electrostatic potential. The electrostatic potential is mapped onto the surface of *Eco* Hfq, *Sau* Hfq, and *Tma* Hfq as seen from distal (top), lateral (middle), and proximal (bottom) views. The distribution of electrostatics potentials on the *Tma* Hfq and *Eco* Hfq is well conserved across all three surfaces, with the highest degree of variability observed on the distal surface. In contrast, the distribution of electrostatic potentials on the *Sau* Hfq is distinct on all three surfaces with significantly more negatively charged electrostatic potentials localized toward the outer rim of the hexamer (on both the distal and proximal surfaces, as well as on the lateral surface). The electrostatic potentials are visualized using the color scale shown. Electrostatic potentials were calculated for each Hfq homolog with APBS [47] then visualized in PyMOL [46].

Table 4.1 Data collection statistics for the apoprotein crystal

Synchrotron	APS, SER-CAT 22-ID beamline
Wavelength (Å)	0.9879
Resolution range (Å)	50 – 2.60
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	
a	39.1
b	133.6
c	206.3
Total No. of reflections	444,056
No. of unique reflections	33,526
Completeness (%)	98.0 (88.4)
Average I/σ (I)	27.4 (1.9)
Redundancy	13.2 (10.7)
R_{merge} † (%)	13.7 (92.5)
Matthews coefficient (Å^3/Da)	2.08
Molecules per asymmetric unit	12

Value in parentheses are for the last resolution shell

† Remerge = $\Sigma_{hkl}\Sigma_i|I_i(hkl)-\langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I(hkl)$, where $I_i(hkl)$ is the intensity of the reflection *hkl* and $\langle I(hkl)\rangle$ is the average intensity of reflection *hkl*.

Table 4.2 Data collection statistics for the U_5 RNA containing crystal

Synchrotron	APS, NE-CAT 24ID-C beamline
Wavelength (Å)	0.9879
Resolution range (Å)	50 – 2.15
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	
a	38.7
b	133.2
c	205.8
Total No. of reflections	194,200
No. of unique reflections	58,420
Completeness (%)	98.4 (95.3)
Average $I/\sigma(I)$	16.1 (2.5)
Redundancy	3.32 (3.00)
R_{merge} † (%)	3.7 (46.4)
Matthews coefficient ($Å^3$/Da)	2.08
Molecules per asymmetric unit	12

Value in parentheses are for the last resolution shell

† Remerge = $\Sigma_{hkl}\Sigma_i|I_i(hkl)-\langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I(hkl)$, where $I_i(hkl)$ is the intensity of the reflection *hkl* and $\langle I(hkl)\rangle$ is the average intensity of reflection *hkl*.

Chapter 5

Initial *in vivo* studies of the expression levels, subcellular localization, and native binding partners of Hfq in *Thermotoga maritima*

Introduction

The central role of Hfq in numerous cellular regulatory pathways is most evident from the pleiotropic phenotype observed for many Hfq knock-out mutants [1,2]. As an alternative to deletion mutants, genome-wide approaches, such as deep sequencing [3,4] and microarray analysis [5], have also been used to identify Hfq-related pathways. Genome-wide approaches have the advantage of not requiring extensive screening of phenotypic effects that knock-out mutants inherently require. Both types of approaches suggest that Hfq is involved in many cellular pathways, including biofilm formation [1,6], response to temperature stress [7,8], sugar and nutrient uptake [9,10,11,12], and general cellular metabolism [7,13,14] (Figure 1.2). Despite the role of Hfq as a global RNA regulator in many bacterial species, there are also many species in which the role of Hfq appears to be quite limited or in which no Hfq homolog has been detected [2].

Though largely unexplored, the level of Hfq protein expression can vary significantly between different bacterial species and also between strains of the same species. For instance, Western blot analysis demonstrated the presence of Hfq protein in eight strains of *Staphylococcus aureus*, but not in two other strains [15]. This variability in Hfq protein expression levels explains contradictory findings regarding the role of Hfq in *S. aureus* [15,16], suggesting that the phenotypic role of Hfq in a given species should not be extrapolated to include all strains of that species. The level of Hfq protein expression has also been reported to be growth phase dependent: the level of Hfq expression in *Escherichia coli* is approximately 55,000 monomers per cell during exponential phase, and is approximately 30,000 monomers per cell during stationary phase [17,18,19,20].

In addition to expression levels, the cellular effect of a protein will depend on its subcellular localization. And, conversely, knowledge of protein localization can provide significant insight into the cellular function of a protein [21,22,23]. Nevertheless, very few studies have examined the localization of Hfq *in vivo.* Our current understanding of Hfq subcellular localization is limited to one proteomic [20] and one imaging [24] study of *E. coli* Hfq. Proteomic analysis of nucleoid-associated proteins in *E. coli* identified Hfq as one of the most abundant proteins in the nucleoid; in fact, approximately 10-20% of the cellular Hfq was estimated to be localized to the nucleoid [20]. Electron microscopy (EM) combined with immunogold labeling of *E. coli* cells expressing recombinant *Eco* Hfq with a C-terminal metallothionein tag indicated that Hfq localizes primarily to the cytoplasm and nucleoid [24]; in addition, less than 50% of the cellular Hfq was found proximal to the inner membrane

As suggested by the citations throughout this thesis, most of the literature on Hfq systems concerns one species- *E. coli*; therefore, data from additional bacterial species are required to broaden our understanding of Hfq systems. *Thermotoga maritima* was the first extremophilic bacterial species identified [25]. *T. maritimia* can survive over a broad temperature range, spanning 55 °C to 90 °C, with optimal growth at 80 °C [25]. *T. maritima* enzymes are attractive targets for various biotechnology applications due to their thermostability [26]. Furthermore, *T. maritima* is a heterotroph, capable of metabolizing both simple and complex carbohydrates; the main byproducts of *T. maritima* fermentation pathways are acetic acid, carbon dioxide, and hydrogen gas [25,27,28]. Hydrogen gas liberated by *T. maritima* during the fermentation of complex carbohydrates, such as cellulose and xylan, may also have potential as a source of renewable energy [26,29,30].

In addition to biotechnological applications, *T. maritima* biology can potentially illuminate the evolution of early life [26]. The extreme ecosystems found on modern Earth, such

as high-temperature and pressure hydrothermal vents, are more akin to those which might have been found in the more primitive biosphere of the early Earth [31,32]. An examination of the phylogenetic tree of life also shows the appeal of extremophiles as model organisms: As one traces the branches from extant species towards the last universal common ancestor (LUCA), virtually all early-branching species are found to be thermophiles [31]. The exact position of the phylum Thermotogae within the universal phylogenetic tree is uncertain due to potentially extensive lateral gene transfer with archaeal organisms known to inhabit the same ecosystem as *T. maritima* [33]. In fact, 24% of predicted protein-coding sequences in the *T. maritima* genome are most similar to archaeal proteins rather than eubacterial counterparts [33,34].

Another interesting feature of *T. maritima* is that it features one of the most compact bacterial genomes sequenced, having a median intergenic length of only 5 bases [34]. To date, 13 putative non-coding RNAs (ncRNAs), all of unknown function, have been identified by transcriptomic studies of *T. maritima* [34]. In numerous bacterial organisms, small non-coding RNAs, aided by the RNA chaperone Hfq, have been shown to mediate important cellular pathways via post-transcriptional regulation of mRNA transcripts [35,36].

In order to develop an integrated, systems-level understanding of any organism, the interactions and linkages between its cellular components must be identified [34,37,38]. In this study, we have begun to examine the growth phase dependent expression and subcellular localization of Hfq, and to map out the Hfq-dependent cellular pathways in *T. maritima*. In particular, quantitative Western blot analysis will be used to determine the amount of Hfq protein in well-defined samples of *T. maritima*; these samples will then be analyzed by flow cytometry to determine the number of *T. maritima* cells in the sample, so that the *in vivo* concentration of Hfq can then be calculated. The localization of Hfq in *T. maritima* will be determined by immunogold labeling of Hfq, followed by EM. Expression and localization studies

of Hfq will be performed at three distinct points in the growth curve of *T. maritima* (log phase, early stationary phase, and late stationary phase) to determine if either of these properties are growth-phase dependent. Finally, Hfq-linked cellular pathways in *T. maritima* will be identified by co-immunoprecipitation followed by either liquid chromatography-tandem mass spectrometry (LC-MS/MS) or Next-Generation Sequencing (NGS) in order to identify the protein and RNA binding partners, respectively.

Initial progress toward the above goals is reported in this chapter. Growth studies were performed in order to determine the growth curve of *T. maritima* under laboratory conditions. Cells were initially visualized by fluorescence confocal microscopy to observe any changes in cellular morphology (e.g. rod-shaped, cocci-shaped) and other features of the culture (e.g. aggregates and chains) that may occur during the different growth phases of *T. maritima.* The localization of Hfq was initially examined by subcellular fractionation and Western blot analysis. However, the lysozyme used during subcellular fractionation was found to cross-react with rabbit anti-*Tma* Hfq polyclonal antibodies (pAb; see Appendix 2), making Western blot analyses unreliable. Therefore, the localization of Hfq was further examined by immunogold labeling of *in vivo* Hfq, followed by negative staining and transmission EM imaging of *T. maritima* cells. The initial EM micrographs from this work indicated that cells were partially lysed during harvesting and fixation, so careful optimization and follow-up studies will be necessary. For expression studies, a quantitative Western blot was optimized so as to determine the concentration of Hfq in a complex sample, such as cell lysate. Finally, Hfq-binding RNAs were co-immunoprecipitated and separated from protein components in preparation for RNA-sequencing experiments.

Methods

Culture growth conditions. *Thermotoga maritima* strain MSB8 was grown at 80 °C, under anaerobic conditions, in artificial salt water (ASW) media supplemented with 0.24% w/v maltose. The media composition is as follows (per liter): 1 g yeast extract; 5 g tryptone; 15 g NaCl; 2 g Na_2SO_4; 2 g $MgCl_2 \cdot 6H_2O$; 0.5 g $CaCl_2 \cdot H_2O$; 0.25 g $NaHCO_3$; 50 mg KBr; 20 mg KI; 20 mg H_3BO_3 ; 3 mg Na_2WO_4; 2 mg $NiCl_2 \cdot 6H_2O$ [39]. Media is prepared as a 2x stock and sterilized by autoclaving. Maltose, KH_2PO_4, and resazurin are added after sterilization, to a final concentration of 0.24% w/v, 0.1% w/v, and 0.1 µg/mL, respectively. (Maltose, KH_2PO_4, and resazurin were filter sterilized using 0.2 µm syringe filters (Millipore) prior to addition.) Media was prepared in culture bottles, sealed with a robber stopper, then sparged with N_2 and reduced with 0.05% w/v $Na_2S \cdot 9H_2O$ pH 8.0 then incubated at 80 °C for approximately 30 minutes. The media changed from dark blue to light yellow after being reduced with Na_2S. The color change is due to the irreversible reduction of the blue dye, resazurin, to resorufin, which is pink at pH values near neutral [40]. Resorufin undergoes a reversible reduction to the colorless product, hydroresorufin [40]. Once the media was completely reduced, the culture was inoculating with 1% v/v inoculum. Note that this protocol is based most closely on that of Chhabra *et al.* [41].

The *T. maritima* inoculum came from starter cultures that were grown in ASW media for 16 hours at 80 °C then moved to room temperature, where they can be stored for up to three months without compromising their viability. Starter cultures were prepared using a culture volume of 50 mL, whereas additional cell cultures were typically grown using a culture volume of 100 mL, 300 mL, or 400mL. Depending on the volume of the sample either 125 mL serum bottles (Wheaton Science Products) or 500 mL culture bottles (Pyrex) were used.

Harvesting Cells. The culture was transferred to a centrifuge tube and quickly cooled in a dry ice-ethanol bath until it reached approximately room temperature; this cooling step typically lasted 5-10 minutes. The cooled culture was then centrifuged at 10,000g for five minutes at 4 °C, the supernatant discarded, and the cell pellet stored at −20 °C.

Confocal microscopy. A 250 μL aliquot of *T. maritima* cells from a live culture were immediately fixed with 750 μL of 4% w/v paraformaldehyde in 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 2.7 mM KCl, and 137 mM NaCl pH 7.40 (1x PBS) and incubated overnight at 4 °C. Fixative-treated samples were then centrifuged at 10,000g for 5 minutes at room temperature (RT) and the supernatant was discarded. The cell pellet was washed with 1x PBS and then resuspended in 50 μL 1x PBS supplemented with 0.02 μg/μL N-(3-7riethylammoniumpropyl)-4-(6-(4-diethylamino) phenyl) hexatrienyl) pyridinium dibromide (FM4-64; Molecular Probes) and incubated for 30 minutes at RT. (FM4-64 binds the outer leaf of the plasma membrane [42].) The sample was centrifuged at 10,000g for 5 minutes at RT and the supernatant was discarded. The cell pellet was washed with 1x PBS then resuspended in 300 μL 285 nM 1H-Indole-6-carboximidamide, 2-[4-(aminoiminomethyl)phenyl]-, dihydrochloride (DAPI; Molecular Probes) and incubated for 5 minutes at RT. (DAPI forms highly fluorescent complexes with AT-rich regions of DNA [43].) Finally , the sample was then centrifuged at 12,000g for 5 minutes at RT, the supernatant discarded and cell pellet resuspended in 500 μL 1x PBS.

Glass slides (Fisherfinest) were first coated with 1% w/v agarose prepared in distilled and deionized water (ddH_2O); cells were then aliquoted onto the slide and covered with an uncoated glass coverslip (Fisherbrand). Fluorescence microscopy images were collected using a Leica SP5 WLL confocal microscope with the Leica application suite (LAS) software. FM 4-64 stain was excited at 510 nm and emission was monitored at 700 nm; DAPI stain was excited at

350 nm and emission was monitored at 470 nm. Cellular images of both stains, as well as a phase contrast image, were overlayed.

Production and purification of rabbit anti-*Tma* Hfq pAb. The antibody production and purification protocols described in Appendix 2 were used here.

Immunogold labeling and electron microscopy (EM). A 750 µL aliquot of *T. maritima* cells from a live culture were immediately fixed with 750 µL of 2x fixation solution containing Tris buffer saline supplemented with 4% paraformaldehyde, 2% glutaraldehyde, and 1% uranyl acetate, as per the method of Liebl *et al* [44]. (Note fixation solution was prepared fresh and insoluble uranyl acetate was removed by centrifugation at 10,000g for 5 minutes at RT prior to adding to the sample.) The cells were centrifuged at 3,350g at RT to remove fixative solution, resuspended in 30% v/v ethanol and incubated for 10 minutes on ice. Centrifugation and 10 minute incubation on ice steps were repeated in a series of steps for 50% v/v, 70% v/v, and 95% v/v ethanol. The cells were then centrifuged, resuspended in 100% ethanol, and incubated at room temperature with gentle rocking for 20 minutes. The cells were centrifuged and resuspended in 1:1 ethanol:LR White (London Resin Company) overnight followed by 1:2 ethanol:LRWhite for 2 hours and 1:4 ethanol:LRWhite for four hours. (LR White is a resin routinely used for electron microscopy [45].) The sample was then centrifuged, resuspended in 100% LR white, and incubated overnight. After transferring to fresh 100% LR White, the cells were embedded in gelatin capsules and cured at 65 °C for 24 hours. Sections were then cut at 75-80 nm using a Leica EM UC7 Ultramicrotome.

Thin-sectioned *T. maritima* cells were transferred to nickel grids and then incubated in 0.1M PBS for 10 minutes, followed by incubation in 0.1M glycine in PBS for 10 minutes. Grids were incubated in Aurion goat blocking solution A for 15 minutes at room temperature. Grids were incubated overnight in rabbit anti-*Tma* Hfq pAb at 4 °C followed by three washes with

0.1M PBS supplemented with 0.2% Aurion-BSA-C (incubation buffer) for a total of 30 minutes. Grids were then incubated with 20 µL 2° Ab, goat anti-rabbit Immunoglobulin G (IgG) gold (10nm) conjugate diluted 1:20 with incubation buffer. Grids then washed twice with incubation buffer and filtered water followed by 10 rinses in three separate containers of filtered water (1 mL) and then air dried. The sections were then contrast stained with 2% uranyl acetate and 0.2% lead citrate.

EM micrographs were generated using a JEOL JEM-1230 transmission electron microscope operated at 80kV and captured on an SIA (Scientific Instruments and Applications, Inc.) 4K X 4K digital camera.

Cy5-labeling rabbit anti-*Tma* Hfq pAb. Samples of pAbs purified as described in Appendix 2 were run applied to a pre-packed PD-10 column containing 8.3 mL of Sephadex G-25 medium (GE Healthcare) for buffer exchange into 1x PBS buffer. The eluted pAb sample were incubated at room temperature for 30 minutes before adding 100 molar excess of tris(2-carboxyethyl)phosphine (TCEP). The sample was flushed with nitrogen, capped in a glass vial, and incubated at room temperature for 10 minutes.   One pack of Cy5 maleimide (GE Healthcare) was solubilized in 50 µL dimethylformamide before adding to 1 mg of pAb sample. The sample was flushed with nitrogen then incubated at room temperature for 2 hours with gentle shaking, followed by an overnight incubation at 4 °C. Cy5-labeled pAbs were then applied to a PD-10 column to remove free Cy5; running buffer was 10 mM HEPES pH 7.80, 150 mM NaCl, 100 µg/mL bovine serum albumin (BSA), and 25% v/v glycerol. To mitigate photobleaching, Cy5-labeled pAbs were stored in the dark at 4 °C until being used.

Sandwich ELISA. 0.01 % v/v rabbit anti-*Tma* Hfq pAb in 1x PBS (Final concentration at ~382 nM) was aliquoted into an Immulon™ 2 HB 96-well plate (Thermo Lab Systems) and incubated overnight at 4 °C. Wells were washed three times with 100 µL 1x PBS followed by blocking with

100 μL 5% w/v dry milk in 1x PBS for 1 hour at 37 °C. Wells were washed twice with 100 μL 1x

PBS and then incubated with 100 μL of sample for 2 hours at 37 °C. Wells were washed three

times with 100 μL 1x PBS supplemented with 0.1% v/v tween-20 (1x PBS-T) followed by a two

hour incubation with 100 μL 0.05% v/v Cy5-labeled rabbit anti-*Tma* Hfq pAb in 1x PBS-T

containing 5% w/v dry milk at 37 °C. Wells were washed four times with 100 μL 1x PBS-T and

two times with 100 μL 1x PBS. The plate was imaged using an Odyssey Li-Cor imaging system

with the 700 channel, because emission of Cy5 is at 670 nm [46].

Western blots. The Western blot protocol used in this study is described in Appendix 2. For

quantitative analysis, a quantitative Western blot using recombinantly expressed *Tma* Hfq to

generate a standard curve was used in this study. Recombinantly expressed *Tma* Hfq was

purified according to the protein purification protocol described in Chapter 2 and then dialyzed

into 25 mM Tris pH 8.00 and 350 mM NaCl. *Tma* Hfq was serially diluted from 2.5 μM to 78 nM

and run on a 7.5% w/v TGX gel (Bio-Rad). The samples were transferred and immunoblotted

following the Western blot protocol described in Appendix 2. The signal of each band was

quantified using the Image Studio software.

Subcellular fractionation. The protocol used here is based on the method described in Thein *et*

al [47]. *T. maritima* cells from a 300 mL culture were harvested as described above,

resuspended in 10 mL of 0.2 M Tris pH 8.0, 1 M sucrose, 1 mM ethylenediaminetetraacetic acid

(EDTA), and 1mg/mL Hen egg white lysozyme (Fisher), and then incubated at room temperature

for 5 minutes. 40 mL of ddH$_2$O was added and the sample container was placed on ice for

approximately 5 minutes. In order to isolate the periplasmic fraction, the sample was

centrifuged in an ultracentrifuge (Beckmann Optima XPN 80) at 200,000*g* for 45 minutes at 4 °C.

The supernatant (presumably containing the periplasmic fraction) was stored at 4 °C and the

pellet was resuspended in 20 mL of ice-cold 10 mM Tris pH 7.5, 5 mM EDTA, and 0.2 mM DTT.

In order to isolate the cytoplasmic fraction, the resuspended pellet was passed through a French press twice and then centrifuged at 300,000g for 2 hours at 4 °C. The supernatant (presumably containing the cytoplasmic fraction) was stored at 4 °C and the pellet was resuspended with inner membrane solubilization buffer (20 mL 50 mM Tris pH 8.0, 2% w/v triton X-100, and 10 mM MgCl$_2$). In order to isolate the inner membrane fraction, this resuspension was centrifuged at 85,000g for 30 minutes at 4 °C. The supernatant (presumably containing the inner membrane fraction) was stored at 4 °C. The remaining pellet was washed with inner membrane solubilization buffer, centrifuged at 85,000g for 20 minutes at 4 °C, and the supernatant was discarded. The final pellet (presumably containing the outer membrane fraction) was resuspended in 50 mM Tris pH 7.5 and 2% w/v sodium dodecyl sulfate (SDS).

Subcellular fractionation samples from each step above were concentrated by acetone precipitation. A 5 mL aliquot from each fraction isolate above was added to 25 mL of ice-cold acetone in a 50 mL polypropylene centrifuge tube. Samples were vortexed and then incubated at -20 °C for 1 hour. The samples were then centrifuged at 15,000g for 10 minutes at 4 °C, and the supernatants discarded. Pellets were dried at room temperature for 30 minutes and then resuspended into the buffer associated with the subcellular fractionation step from which the sample was obtained. The final samples were then analyzed via Western blot analysis for the presence of *Tma* Hfq.

Immunoprecipitation of Hfq-binding RNAs from *T. maritima* and enrichment for nanoRNAs. *T. maritima* cultures were grown at 80 °C for 14-16 hours the harvested as described above. The cell pellet from 300 mL or 2.4 L of cell culture (Table 5.1) was resuspended in 50 mM Tris pH 8.5, 150 mM NaCl, 10 mM imidazole (lysis buffer); to mitigate proteolysis, phenylmethylsulfonyl fluoride (PMSF) was added to a final concentration of 0.4 mM. Cells were mechanically lysed using a microfluidizer and centrifuged at 35,000g for 20 minutes at 4 °C. The supernatant was

diluted with an equal volume of 20 mM sodium phosphate pH 7.00 then loaded onto a protein G

column that was pre-charged with rabbit anti-*Tma* Hfq pAb. Using an NGC chromatography

system (BioRad), this sample loading step was followed by 10 column volumes of wash buffer

(20 mM sodium phosphate pH 7.00) and then elution buffer (0.1 M glycine pH 2.7). The eluted

fractions were immediately neutralized with 0.2 volumes of 1M Tris-HCl pH 9.0. For the

negative control, the exact same procedure was followed except that the protein G column was

not charged with pAb. Fractions containing *Tma* Hfq and binding partners as indicated by a

strong absorbance at 280 nm and 260 nm were pooled for subsequent steps. The pooled

sample was either enriched for nanoRNAs (below) or further purified for other RNAs using a

Direct-zol RNA MiniPrep kit (Zymo Research) and stored at -80 °C.

For enrichment of nanoRNAs (see Chapter2), the pooled fractions from the protein G

column were dialyzed into 50 mM Tris pH 8.5, 150 mM NaCl, and 12.5 mM EDTA. The sample

was then diluted with 3 volumes (per volume sample) of buffer A (25 mM Tris pH 8.50, 50 mM

NaCl) and loaded onto a quaternary amine anion exchange column (GE Healthcare) using a

NGC^{TM} chromatography system (BioRad), followed by 10 column volumes of buffer A wash and

elution with buffer B (25 mM Tris pH 8.50, 2 M NaCl) using a stepwise gradient. Based on the

results described in Chapter 2, fractions that eluted at 20% buffer B (440 mM NaCl) were

combined and concentrated. The samples were then purified using a Direct-zol RNA MiniPrep

kit (Zymo Research) and stored at -80 °C.

Preliminary Results & Discussion

Growth curve for *T. maritima*

T. maritima cultures were successfully grown in ASW media using resazurin as a redox indicator

(Figure 5.1). A growth curve for the *T. maritima* culture was generated by plotting the optical

density at 600 nm (OD600) against incubation time (Figure 5.2). OD600 measurements were taken as quickly as possible by removing 750 μL of culture using a sterile needle and syringe, since the media turns pink as the redox potential increases upon exposure to air. The OD600 of the culture was found to reach a maximum at 12-14 hours, and then decrease gradually over time. The sigmoidal region of the observed growth curve could be fit to a logistic model with an R^2 equal to 0.99 [48,49]. From this fit, mid-log phase was calculated to be near 6.7 hours.

The dip in OD600 during late stationary phase, evident in the growth curve of *T. maritima* (Figure 5.2), suggests that the culture has entered a death phase at these later time points. The culture could enter a death phase for multiple reasons including depletion of necessary nutrients or the production of a toxic byproduct during metabolism. For instance, *T. maritima* is known to produce H_2 as an autoinhibitory byproduct of its metabolism [25]. Additionally, a change in cellular morphology, from rod-shaped to more compact cocci, during late stationary phase could alter the correlation between OD600 measurements and cell density [50].

T. maritima cells have two distinct morphologies

Confocal microscopy was used to examine the morphology of *T. maritima* cells at various time points (Figure 5.3). Initial images were collected for samples in early stationary phase to determine suitable fixative condition (Figure 5.3a). Samples fixed with glutaraldehyde displayed numerous features indicative of partial lysis, including free-floating nucleic acid material and visible breaks in both the outer and inner cell membranes (Figure 5.3a, left panel). In contrast, samples fixed with paraformaldehyde contained rod-shaped cells with some aggregates and chains present in the sample (Figure 5.3a, right panel). To maintain cell morphology and structural integrity, paraformaldehyde fixation was used for all subsequent samples.

T. maritima cells from a starter culture and cells from late stationary phase contained both rod-shaped and cocci-shaped cells (Figure 5.3b). There was little evidence of aggregates or chains in the starter culture, but some aggregates and chains of rod-shaped cells were observed in the late stationary phase sample. *T. maritima* cells from early stationary phase contained only rod-shaped cells with few aggregates and numerous short chains of rod-shaped cells throughout the sample.

Toward determining the localization of *Tma* Hfq *in vivo*

In order to elucidate the subcellular localization of Hfq in *T. maritima* cells, initial efforts were made to isolate various cellular components by ultracentrifugation and selective detergent treatments of *T. maritima* cells using a protocol previously established for gram-negative bacteria (Figure 5.4a)   [47].   Specifically, *T. maritima* was separated into periplasmic, cytoplasmic, inner membrane, and outer membrane fractions. Isolated fractions were then assayed for the presence of *Tma* Hfq via Western blot analysis with rabbit anti-*Tma* Hfq pAb (Figure 5.4b). The recurrent, high-intensity band at approximately 12 kDa in all fractions does not match what is known about the cellular function and localization of Hfq in other organisms [24]. Intriguingly, the intensity of this band in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 5.4c) showed it to be the most predominant protein in the whole cell fraction, which is inconsistent with initial assays of Hfq expression in *T. maritima* (Figure 5.5) as well as what is known about expression levels of Hfq in *E. coli* [17,19].

Because lysozyme was added to the cells in the first step of lysis, a Western blot was performed using purified lysozyme (commercially available) to see if rabbit anti-*Tma* Hfq pAbs cross-react. A serial dilution of lysozyme demonstrates cross-reactivity with the pAbs (Figure 5.6). The similar molecular weights of *Tma* Hfq (10.5 kDa) and lysozyme (14.3 kDa) make this cross-reactivity especially problematic, as the lysozyme signal masks the signal from Hfq in both

SDS-PAGE and Western blot analysis. For this reason, in future studies *T. maritima* cells will be lysed using mechanical means, such as gentle sonication or microfluidization, in order to avoid this problem of lysozyme cross-reactivity.

As an alternative means to avoid issues related to the antibody cross-reactivity with lysozyme, the localization of Hfq *in vivo* was visualized using EM with immunogold labeling of Hfq. EM micrographs were collected at different time points in the growth phase of *T. maritima* (Figure 5.7). All observed *T. maritima* cells had breaks in their membranes indicating partial lysis. For this reason, conditions need to be optimized so that localization of Hfq in intact cells can be observed. Despite these caveats, distinct differences in the cellular morphology during log phase (Figure 5.7a) and early stationary phase (Figure 5.7b) can already be seen. For example, more examples of cellular division (Figure 5.7b, right panel) were found in early stationary phase then log phase. Additionally, the toga (i.e. outer membrane) was more extended during early stationary phase (Figure 5.7b). It should be noted that, contrary to what was observed by confocal microscopy, the cells are more cocci-shaped during both phases when visualized by transmission EM. However, the overall shape observed by electron microscopy may not be accurate due to the large breaks in the membrane (Figure 5.7, indicated by black arrows).

Quantification of *Tma* Hfq in a complex sample

In order to quantify the amount of Hfq protein in a sample of *T. maritima* cell lysate*,* a sandwich ELISA (Figure A2.5) was developed using unlabeled rabbit anti-*Tma* Hfq pAbs as the capture Ab and Cy5-labeled rabbit anti-*Tma* Hfq pAbs as the detection Ab. The assay was able to detect Hfq at concentrations above 81 pM, making it ideal for the detection and quantification of low levels of *Tma* Hfq in complex samples, such as cell lysate (Figure 5. 8). Unfortunately, the linear range

of the assay was relatively limited (indicated by dashed lines in Figure 5.8) and the standard deviation of the signal was too large for reliable quantitative analysis.

Given the limited dynamic range of the sandwich ELISA, a quantitative Western blot, using unlabeled rabbit anti-*Tma* Hfq pAbs as the primary antibody and goat anti-rabbit IgG monoclonal antibodies (mAbs) as the secondary, was optimized for the quantification of *Tma* Hfq in complex samples, such as cell lysate. Recombinantly expressed and purified *Tma* Hfq was serially diluted and run on a 7.5% w/v TGX gel for quantitative Western blot analysis (Figure 5.9a). The signal for each lane was plotted against the amount of *Tma* Hfq to generate a standard curve (Figure 5.9b). Data were collected for three independent replicates and fit to a linear regression line with an R^2 of 0.9995. In future work, this assay may be useful to quantify *Tma* Hfq in samples of *T. maritima* culture.

Co-immunoprecipitation of Hfq-binding RNAs from *T. maritima*

The RNAs that physically associate with Hfq *in vivo* can be used to infer Hfq function and involvement in particular cellular pathways. RNA binding partners that physically interact with Hfq can be co-immunoprecipitated with Hfq specific antibodies [51,52] (such as the rabbit anti-*Tma* Hfq pAb described in Appendix 2) or with antibodies specific to a protein-tag genetically added to a recombinantly expressed Hfq [3,53,54,55]. These co-immunoprecipitated Hfq-binding RNAs can then be identified using well-established protocols, such as deep sequencing (also known as NGS) [3,54,55] or microarrays [51,52].

In this study, Hfq-binding RNAs were co-immunoprecipitated using a protein G column pre-charged with rabbit anti-*Tma* Hfq pAb (Figure 5.10); an uncharged protein G column was used as a negative control. Cell lysate from a *T. maritima* culture (grown for 14-16 hours at 80 °C) was added to the pre-charged column, which specifically interacts with Hfq in the cell lysate. Non-associated cellular components were washed away with a mild wash buffer, leaving Hfq

157

along with its associated binding partners bound to the column. Hfq and its binding partners were then eluted along with the rabbit anti-*Tma* Hfq pAb with a low pH elution buffer, which disrupts the association between protein G and the pAbs [56]. The eluted fractions were immediately neutralized with 1M Tris-HCl pH 9.0.

To enrich for nanoRNAs, the eluent from the charged protein G column was dialyzed into a low salt (150 mM NaCl) buffer at pH 8.5 and applied to an anion exchange column as described in the method section. The anion exchange column separated the Hfq-associated nanoRNAs from more negatively charged components of the sample (Figure 2.1), such as longer RNAs (Figure 5.11). All samples (i.e. negative control, Co-IP, and nanoRNA enriched Co-IP samples) were subsequently purified from protein components using a Direct-zol MiniPrep Kit.

The quality of each sample was assessed via nanodrop absorbance measurements (absorbance ratios: 260 nm/280nm and 260nm/230nm) [57,58]. A 260 nm/280 nm absorbance ratio >1.8 is an accepted indicator of high RNA purity [59,60,61,62], the RNA sample met this quality control whereas the 260 nm/280 nm ratio for the nanoRNA-enriched sample suggested possible protein contaminants. The 260 nm/230 nm absorbance ratio is used as an indicator of organic contaminants in the sample; a 260 nm/230 nm ratio <1.8 indicates that the sample contains residual organic contaminants [61,62]. Both the RNA and nanoRNA-enriched samples had low 260 nm/280 nm absorbance ratios indicating residual organic contamination. The approximate concentration, 260 nm/280 nm absorbance ratio, and 260 nm/230 nm absorbance ratio for each sample is summarized in Table 5.1.

Future Directions

Determining the concentrations of *Tma* Hfq *in vivo*

To determine the concentration of Hfq in *T. maritima* cells, the amount of Hfq in a sample of know cellular density needs to be quantified. In order to determine the cell density of a sample, cell counting via confocal microscopy was initially pursued as one method for determining the cellular density of *T. maritima* in a given sample. Unfortunately, the small size of *T. maritima* cells (~ 1 μm) precluded the use of a cell counting grid or a hemocytometer for technical reasons, as those instruments were incompatible with the 63x optic that was required for visualization of the cells. Future work will instead utilize imaging flow cytometry to count the number of cells in a sample [63,64] and thereby compute the cellular density of the culture at different time points in the growth phase of *T. maritima.*

In tandem, the concentration of Hfq in the sample will then be determined using the quantitative Western blot described above (Figure 5.9). Dividing the Hfq concentration in the sample (from quantitative Western blot) by the cellular density of the sample (from cell counting) will provide the concentration of Hfq *in vivo*, given that the volumes are known quantities in each case*.* The concentration of Hfq in *T. maritima* will be quantified at different time points to determine if Hfq expression levels are growth phase dependent.

Determining the localization of *Tma* Hfq *in vivo*

To determine the localization of *Tma* Hfq *in vivo*, samples will be immunogold labeled, then visualized using EN. Initial EM images indicated that the cells were partially lysed (Figure 5.7), therefore fixative conditions need to be optimized to maintain intact cells. Once optimized, the fixed cells will be immunoblotted using rabbit anti-*Tma* Hfq pAbs (1° Ab) and donkey anti-rabbit IgG conjugated to 10 nm gold particles (2° Ab). Samples that are not exposed to the 1° Ab can be used as a negative control. EM images collected for cells at different time points in the growth

phase of *T. maritima* can be used to determine if cellular localization is growth phase dependent.

Identifying binding partner of Hfq in *T. maritima*

Identification of the proteins and RNAs associated with Hfq *in vivo* can be used to elucidate the involvement of Hfq in particular cellular pathways in *T. maritima*. Proteins and RNAs physically associated with Hfq can be, in principle, isolated by co-immunoprecipitation using anti-Hfq pAbs [51,52]. The protein interacting partners co-immunoprecipitate with Hfq can then be identified by well-established proteomics experiments [65,66,67]. Immunoprecipitated RNA components can be further purified from the protein using phenol-chloroform extraction based technologies and identified using next-generation sequencing [3,53,54,55].

The RNA samples described in Table 5.1 will be used to reverse transcribe a cDNA library containing adapters for high-throughput sequencing (Figure 5.12). The adapters will include a primer sequence and a sequence complementary to the ligation fragments on the MiSeq chip that will be used for sequencing. One of the adapters will also contain a "barcode" sequence, which is a short sequence that is distinct for each sample (enabling determination of which samples each sequence originated from). An Illumina MiSeq desktop sequencer will be used to determine the sequences of the RNAs in the library. These RNA sequences will then be aligned with the genome of *T. maritima* strain MSB8 to determine if they are located in intergenic regions, or associated with specific genes. Bioinformatic analyses of these data will be used to identify any trends with regards to sequence or function of the Hfq-interacting RNAs identified by this workflow.

Additional studies will focus on the identification of proteins that physically associate with Hfq in *T. maritima*. Similar to the RNA binding partner studies, proteins-associated with Hfq will be co-immunoprecipitated using a protein G column pre-charged with rabbit anti-*Tma*

Hfq pAb as illustrated in Figure 5.10 with the exception that the pAbs will be covalently crosslinked to the protein G column so that they do not elute from the column with Hfq and the Hfq-associated molecules. Protein binding partners will then be identified using LC-MS/MS, which would be impeded by a high concentration of antibodies in the eluent. The proteins identified in these studies will be used to map out cellular interaction and to infer the function of Hfq *in vivo.*

References

1. Kulesus RR, Diaz-Perez K, Slechta ES, Eto DS, Mulvey MA (2008) Impact of the RNA chaperone Hfq on the fitness and virulence potential of uropathogenic *Escherichia coli*. Infect Immun 76: 3019-3026.

2. Jousselin A, Metzinger L, Felden B (2009) On the facultative requirement of the bacterial RNA chaperone, Hfq. Trends Microbiol 17: 399-405.

3. Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, et al. (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. PLoS Genet 4: e1000163.

4. Sharma CM, Vogel J (2009) Experimental approaches for the discovery and characterization of regulatory small RNA. Curr Opin Microbiol 12: 536-546.

5. Zhang A, Wassarman KM, Rosenow C, Tjaden BC, Storz G, et al. (2003) Global analysis of small RNA and mRNA targets of Hfq. Mol Microbiol 50: 1111-1124.

6. Wang MC, Chien HF, Tsai YL, Liu MC, Liaw SJ (2014) The RNA Chaperone Hfq Is Involved in Stress Tolerance and Virulence in Uropathogenic *Proteus mirabilis*. PLoS One 9: e85626.

7. Tsui HC, Leung HC, Winkler ME (1994) Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in *Escherichia coli* K-12. Mol Microbiol 13: 35-49.

8. Cui M, Wang T, Xu J, Ke Y, Du X, et al. (2013) Impact of Hfq on global gene expression and intracellular survival in *Brucella melitensis*. PLoS One 8: e71933.

9. Wilms I, Moller P, Stock AM, Gurski R, Lai EM, et al. (2012) Hfq influences multiple transport systems and virulence in the plant pathogen Agrobacterium tumefaciens. J Bacteriol 194: 5209-5217.

10. Guisbert E, Rhodius VA, Ahuja N, Witkin E, Gross CA (2007) Hfq modulates the sigmaE-mediated envelope stress response and the sigma32-mediated cytoplasmic stress response in Escherichia coli. J Bacteriol 189: 1963-1973.

11. Geng J, Song Y, Yang L, Feng Y, Qiu Y, et al. (2009) Involvement of the post-transcriptional regulator Hfq in Yersinia pestis virulence. PLoS One 4: e6213.

12. Sharma CM, Papenfort K, Pernitzsch SR, Mollenkopf HJ, Hinton JC, et al. (2011) Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. Mol Microbiol 81: 1144-1165.

13. Kakoschke T, Kakoschke S, Magistro G, Schubert S, Borath M, et al. (2014) The RNA Chaperone Hfq Impacts Growth, Metabolism and Production of Virulence Factors in *Yersinia enterocolitica*. PLoS One 9: e86113.

14. Torres-Quesada O, Oruezabal RI, Peregrina A, Jofre E, Lloret J, et al. (2010) The *Sinorhizobium meliloti* RNA chaperone Hfq influences central carbon metabolism and the symbiotic interaction with alfalfa. BMC Microbiol 10: 71.

15. Liu Y, Wu N, Dong J, Gao Y, Zhang X, et al. (2010) Hfq is a global regulator that controls the pathogenicity of *Staphylococcus aureus*. PLoS One 5.

16. Bohn C, Rigoulay C, Bouloc P (2007) No detectable effect of RNA-binding protein Hfq absence in *Staphylococcus aureus*. BMC Microbiol 7: 10.

17. Kajitani M, Kato A, Wada A, Inokuchi Y, Ishihama A (1994) Regulation of the *Escherichia coli* hfq gene encoding the host factor for phage Q beta. J Bacteriol 176: 531-534.

18. McNealy TL, Forsbach-Birk V, Shi C, Marre R (2005) The Hfq homolog in *Legionella pneumophila* demonstrates regulation by LetA and RpoS and interacts with the global regulator CsrA. J Bacteriol 187: 1527-1532.

19. Ali Azam T, Iwata A, Nishimura A, Ueda S, Ishihama A (1999) Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. J Bacteriol 181: 6361-6370.

20. Ali Azam T, Ishihama A (1999) Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. J Biol Chem 274: 33105-33113.

21. Yu CS, Chen YC, Lu CH, Hwang JK (2006) Prediction of protein subcellular localization. Proteins 64: 643-651.

22. Lu Z, Hunter L (2005) Go molecular function terms are predictive of subcellular localization. Pac Symp Biocomput: 151-161.

23. Parras-Molto M, Campos-Laborie FJ, Garcia-Dieguez J, Rodriguez-Grinolo MR, Perez-Pulido AJ (2013) Classification of protein motifs based on subcellular localization uncovers evolutionary relationships at both sequence and functional levels. BMC Bioinformatics 14: 229.

24. Diestra E, Cayrol B, Arluison V, Risco C (2009) Cellular electron microscopy imaging reveals the localization of the Hfq protein close to the bacterial membrane. PLoS One 4: e8301.

25. Huber R, Langworthy TA, Konig H, Thomm M, Woese CR, et al. (1986) *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90^{0}C. Arch Microbiol: 324-333.

26. Conners SB, Mongodin EF, Johnson MR, Montero CI, Nelson KE, et al. (2006) Microbial biochemistry, physiology, and biotechnology of hyperthermophilic *Thermotoga* species. FEMS Microbiol Rev 30: 872-905.

27. Chhabra SR, Shockley KR, Conners SB, Scott KL, Wolfinger RD, et al. (2003) Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*. J Biol Chem 278: 7540-7552.

28. Conners SB, Montero CI, Comfort DA, Shockley KR, Johnson MR, et al. (2005) An expression-driven approach to the prediction of carbohydrate transport and utilization regulons in the hyperthermophilic bacterium *Thermotoga maritima*. J Bacteriol 187: 7267-7282.

29. Van Ooteghem SA, Beer SK, Yue PC (2002) Hydrogen production by the thermophilic bacterium *Thermotoga neapolitana*. Appl Biochem Biotechnol: 177-189.

30. Sybirna K, Bottin H (2013) New insights into the systems for heterologous synthesis and maturation of hydrogenases, the most efficient biohydrogen producers. Int J Hydrogen Energ: 13164-13171.

31. Javaux EJ (2006) Extreme life on Earth--past, present and possibly beyond. Res Microbiol 157: 37-48.

32. Lal AK (2008) Origin of Life. Astrophys Space Sci: 267-278.

33. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, et al. (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature 399: 323-329.

34. Latif H, Lerman JA, Portnoy VA, Tarasova Y, Nagarajan H, et al. (2013) The genome organization of *Thermotoga maritima* reflects its lifestyle. PLoS Genet 9: e1003485.

35. Gottesman S, Storz G (2011) Bacterial small RNA regulators: versatile roles and rapidly evolving variations. Cold Spring Harb Perspect Biol 3.

36. Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. Nat Rev Microbiol 9: 578-589.

37. Kitano H (2002) Systems biology: a brief overview. Science 295: 1662-1664.

38. Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO (2009) Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol 7: 129-143.

39. Chhabra S, Parker KN, Lam D, Callen W, Snead MA, et al. (2001) Beta-mannanases from *Thermotoga* species. Methods Enzymol 330: 224-238.

40. Guerin TF, Mondido M, McClenn B, Peasley B (2001) Application of resazurin for estimating abundance of contaminant-degrading micro-organisms. Lett Appl Microbiol 32: 340-345.

41. Chhabra SR, Shockley KR, Ward DE, Kelly RM (2002) Regulation of endo-acting glycosyl hydrolases in the hyperthermophilic bacterium Thermotoga maritima grown on glucan- and mannan-based polysaccharides. Appl Environ Microbiol 68: 545-554.

42. Ueda T, Yamaguchi M, Uchimiya H, Nakano A (2001) Ara6, a plant-unique novel type Rab GTPase, functions in the endocytic pathway of Arabidopsis thaliana. EMBO J 20: 4730-4741.

43. Eriksson S, Kim SK, Kubista M, Norden B (1993) Binding of 4',6-diamidino-2-phenylindole (DAPI) to AT regions of DNA: evidence for an allosteric conformational change. Biochemistry 32: 2987-2998.

44. Liebl WG, Winterhalter C, Baumeister W, Armbrech M, Valdez M (2008) Xylanase attachment to the cell wall of the hyperthermophilic bacterium *Thermotoga maritima*. Journal of Bacteriology 190: 1350-1358.

45. Timms BG (1986) Postembedding immunogold labeling for electron microscopy using "LR White" resin. Am J Anat 175: 267-275.

46. Evangelista RA, Chen FT (1994) Analysis of structural specificity in antibody-antigen reactions by capillary electrophoresis with laser-induced fluorescence detection. J Chromatogr A 680: 587-591.

47. Thein M, Sauer G, Paramasivam N, Grin I, Linke D (2010) Efficient subfractionation of Gram-negative bacteria for proteomics studies. J Proteome Res 9: 6135-6147.

48. Winsor CP (1932) The Gompertz Curve as a Growth Curve. Proc Natl Acad Sci U S A 18: 1-8.

49. Zwietering MH, Jongenburger I, Rombouts FM, van 't Riet K (1990) Modeling of the bacterial growth curve. Appl Environ Microbiol 56: 1875-1881.

50. Johnson MR, Conners SB, Montero CI, Chou CJ, Shockley KR, et al. (2006) The *Thermotoga maritima* phenotype is impacted by syntrophic interaction with *Methanococcus jannaschii* in hyperthermophilic coculture. Appl Environ Microbiol 72: 811-818.

51. Zhang AX, Wassarman KM, Rosenow C, Tjaden BC, Storz G, et al. (2003) Global analysis of small RNA and mRNA targets of Hfq. Molecular Microbiology 50: 1111-1124.

52. Zhang AX, Schu DJ, Tjaden BC, Storz G, Gottesman S (2013) Mutations in Interaction Surfaces Differentially Impact E. coli Hfq Association with Small RNAs and Their mRNA Targets. Journal of Molecular Biology 425: 3678-3697.

53. Dambach M, Irnov I, Winkler WC (2013) Association of RNAs with Bacillus subtilis Hfq. PLoS One 8.

54. Li SK, Ng PKS, Qin H, Lau JKY, Lau JPY, et al. (2013) Identification of small RNAs in Mycobacterium smegmatis using heterologous Hfq. Rna-a Publication of the Rna Society 19: 74-84.

55. Sittka A, Sharma CM, Rolle K, 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 Biology 6: 266-275.

56. Akerstrom B, Bjorck L (1986) A physicochemical study of protein G, a molecule with unique immunoglobulin G-binding properties. J Biol Chem 261: 10240-10247.

57. Gayral P, Weinert L, Chiari Y, Tsagkogeorga G, Ballenghien M, et al. (2011) Next-generation sequencing of transcriptomes: a guide to RNA isolation in nonmodel animals. Molecular Ecology Resources 11: 650-661.

58. Tang CR, Qi JY, Li HP, Zhang CL, Wang YK (2007) A convenient and efficient protocol for isolating high-quality RNA from latex of Hevea brasiliensis (para rubber tree). Journal of Biochemical and Biophysical Methods 70: 749-754.

59. Manchester KL (1996) Use of UV methods for measurement of protein and nucleic acid concentrations. Biotechniques 20: 968-&.

60. Imbeaud S, Graudens E, Boulanger V, Barlet X, Zaborski P, et al. (2005) Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. Nucleic Acids Research 33.

61. Manchester KL (1995) Value of A260/A280 ratios for measurement of purity of nucleic acids. Biotechniques 19: 208-210.

62. Glasel JA (1995) Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. Biotechniques 18: 62-63.

63. Volkmer B, Heinemann M (2011) Condition-dependent cell volume and concentration of Escherichia coli to facilitate data conversion for systems biology modeling. PLoS One 6: e23126.

64. Lacroix R, Robert S, Poncelet P, Dignat-George F (2010) Overcoming Limitations of Microparticle Measurement by Flow Cytometry. Seminars in Thrombosis and Hemostasis 36: 807-818.

65. Ngounou Wetie AG, Sokolowska I, Woods AG, Roy U, Deinhardt K, et al. (2014) Protein-protein interactions: switch from classical methods to proteomics and bioinformatics-based approaches. Cell Mol Life Sci 71: 205-228.

66. Beaufour M, Godin F, Vallee B, Cadene M, Benedetti H (2012) Interaction proteomics suggests a new role for the tfs1 protein in yeast. J Proteome Res 11: 3211-3218.

67. Free RB, Hazelwood LA, Sibley DR (2009) Identifying novel protein-protein interactions using co-immunoprecipitation and mass spectroscopy. Curr Protoc Neurosci Chapter 5: Unit 5 28.
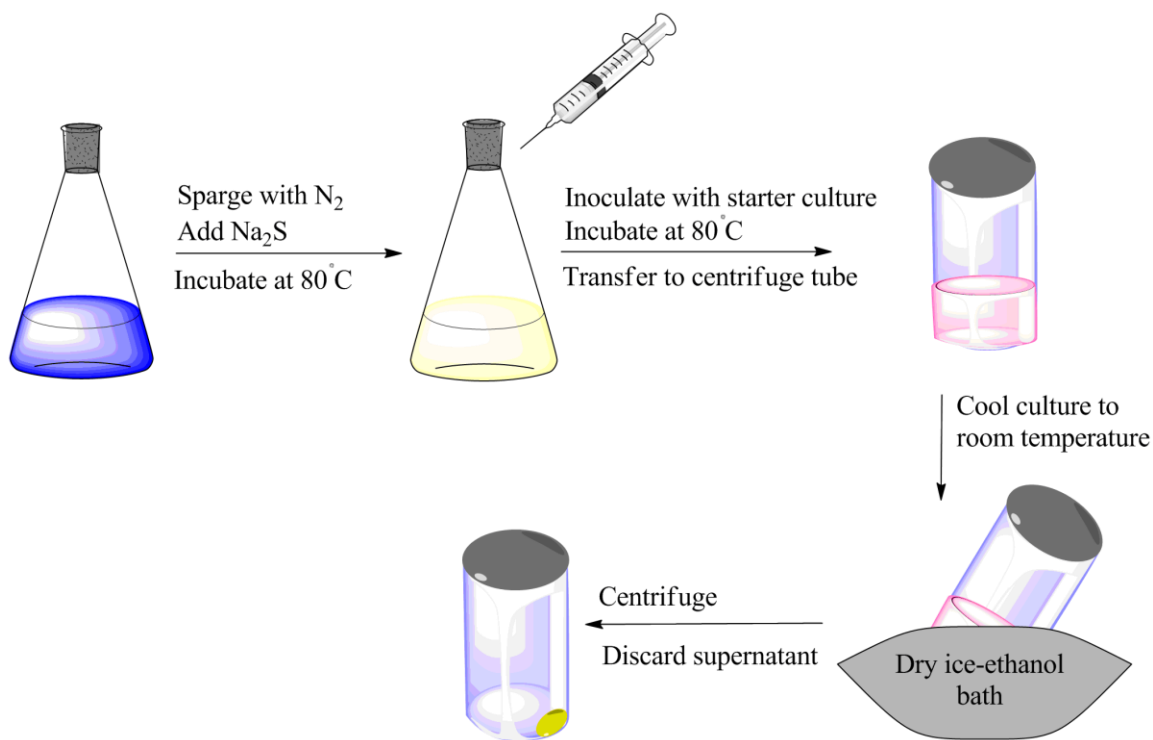
Figure 5.1 Culturing *T. maritima*. ASW media is prepared and resazurin is added as an indicator of the redox potential. The culture tube is sealed and sparged with nitrogen to remove oxygen. Na₂S is used as a reducing agent and the media is incubated at 80 °C to remove dissolved oxygen. The media can then be inoculated using a *T. maritima* starter culture once the ASW media is completely reduced (media is slightly yellow when fully reduced). Growth cultures are incubated at 80 °C without mechanical agitation for 14-16 hours. To harvest cells, the culture is transferred to a centrifuge tube and quickly cooled to room temperature in a dry ice-ethanol bath. The culture is centrifuged, supernatant is discarded, and the pellet is stored at -20 °C.
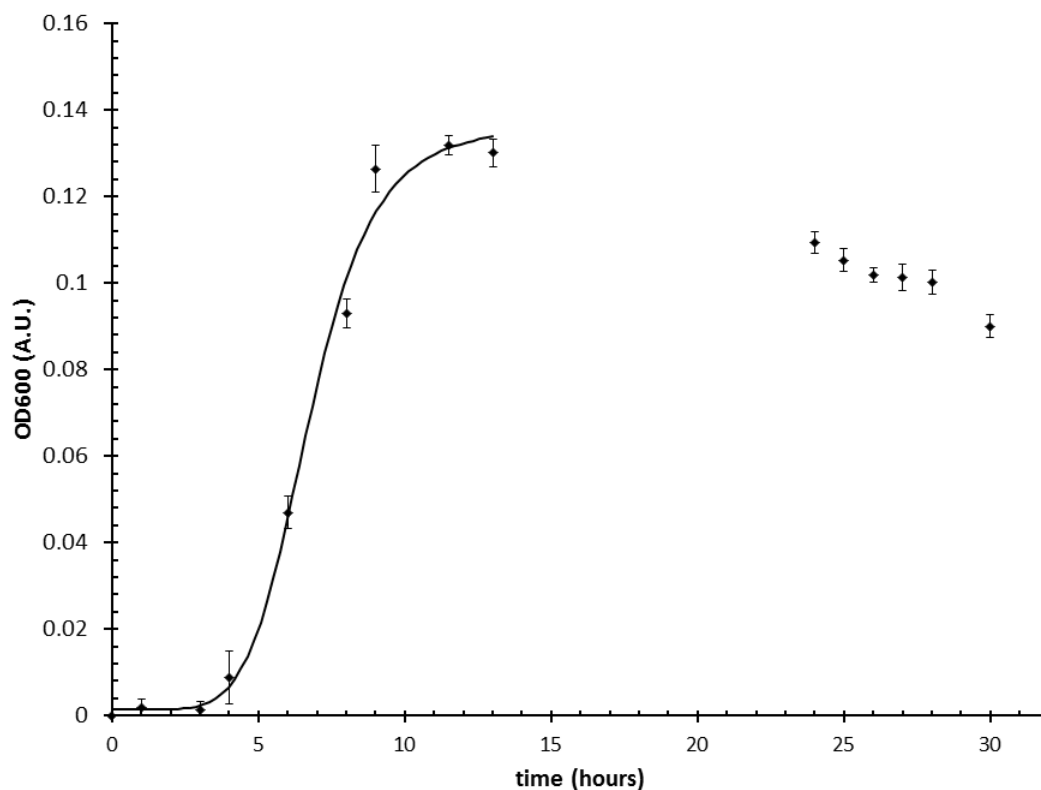
Figure 5.2 Growth curve for *T. maritima*. *T. maritima* was grown in ASW media at 80 °C under anaerobic conditions and absorbance was measured at 600 nm (*N*=3). The culture reached saturation at 12-14 hours, with subsequent readings decreasing over time. Note that the decrease in absorbance during late stationary phase is partially due to a change in cell morphology of *T. maritima*, from rod-shaped to cocci [50]. The first 13 hours of growth were fit using the logistic function with an R^2 of 0.99 [48,49].
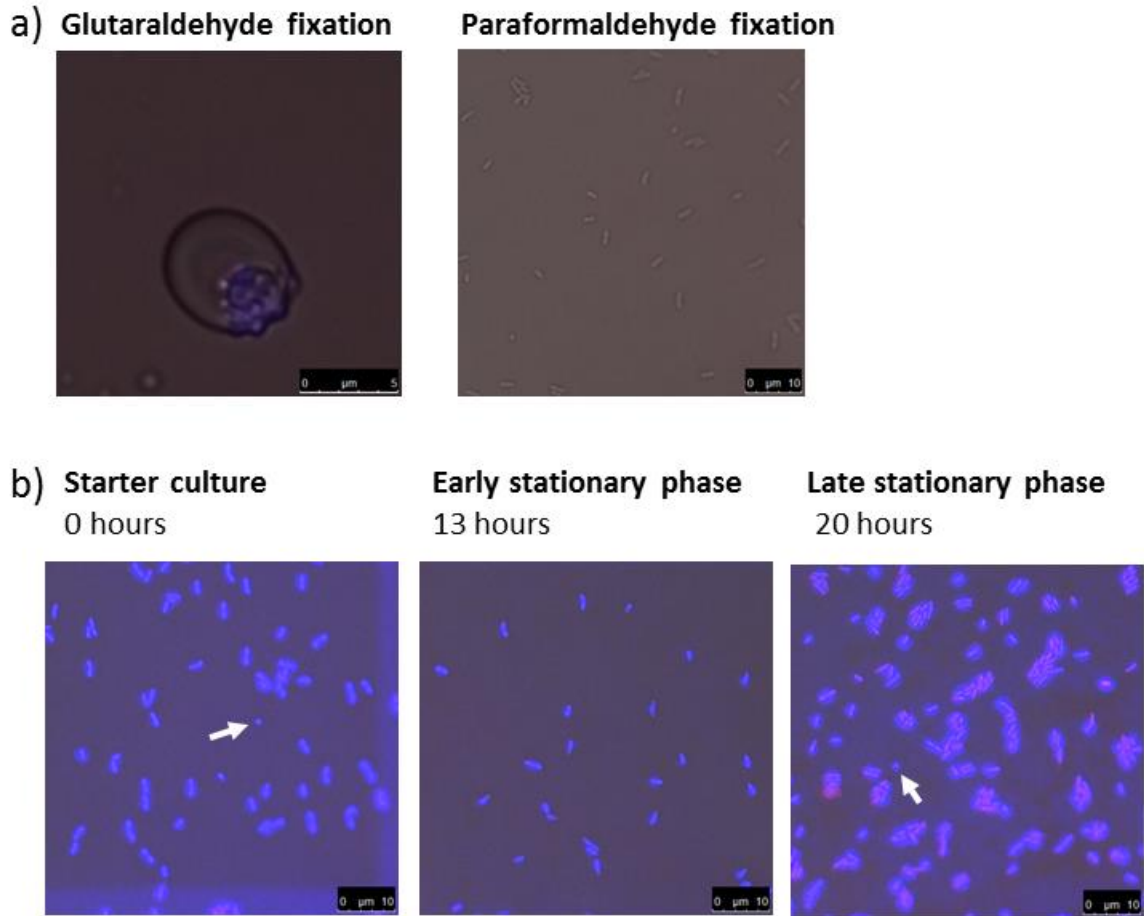
Figure 5.3 Confocal microscopy of _T. maritima_. (a) Confocal images of _T. maritima_ after fixing with either 1.9% w/v glutaraldehyde (left panel, 5 μm scale bar) or 3% w/v paraformaldehyde (right panel, 10 μm scale bar). Glutaraldehyde fixed cells were stained with DAPI (blue) to visualize the localization of cellular nucleic acid. Both samples in (a) were taken at early stationary phase (13 hours). (b) Confocal images of _T. maritima_ at different time points in its growth phase after fixation with 3% w/v paraformaldehyde. Cells were stained with DAPI (blue) to visualize the nucleic acid material. The white arrow in the starter culture and late stationary phase panel are pointing at an example of a cocci-shaped cell in each sample. Scale bars in each image of (b) are for 10 μm. All images were collected using a 63x objective.
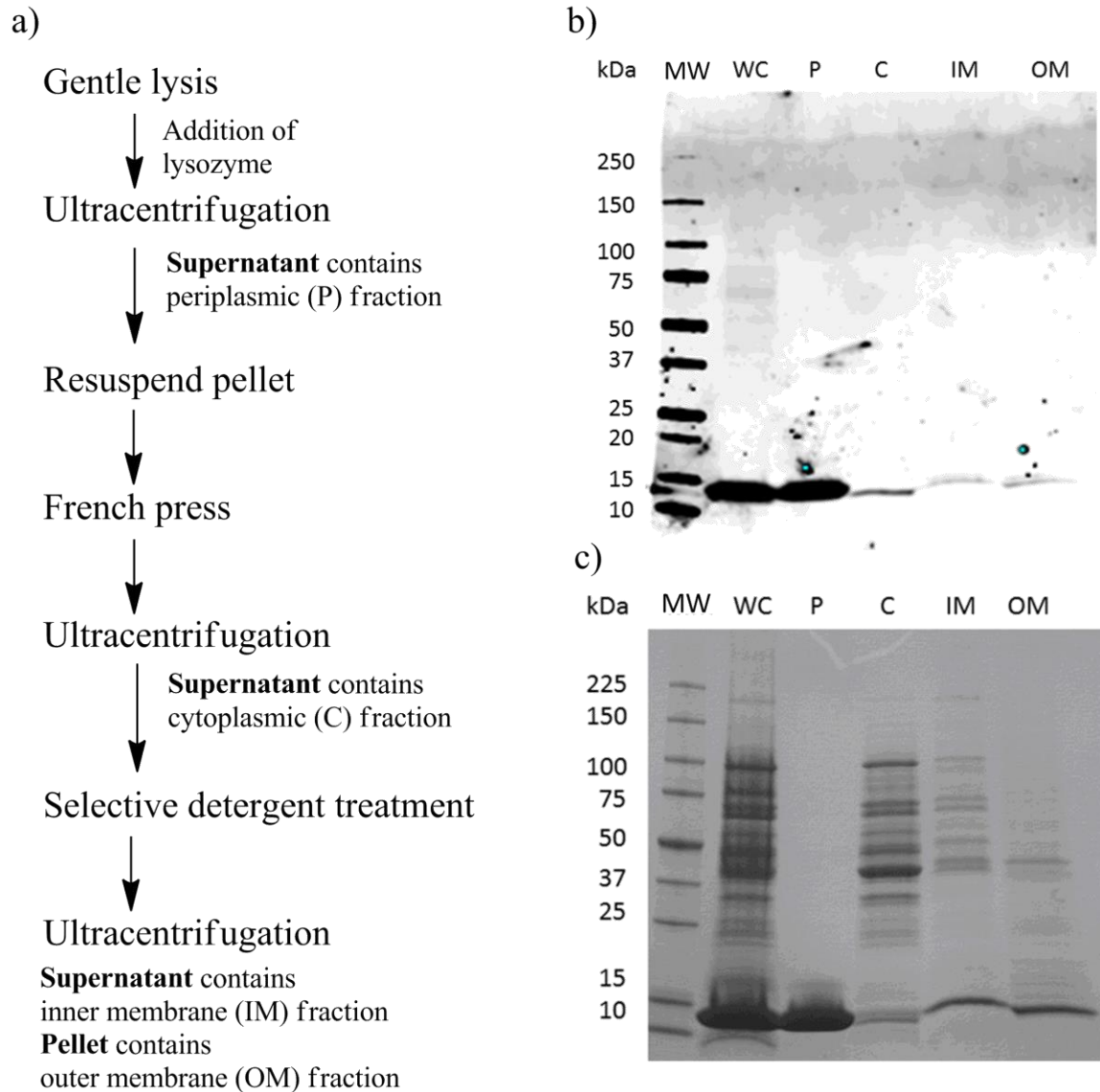
a)

Gentle lysis

→ Addition of lysozyme

Ultracentrifugation

→ **Supernatant** contains periplasmic (P) fraction

Resuspend pellet

↓

French press

↓

Ultracentrifugation

→ **Supernatant** contains cytoplasmic (C) fraction

Selective detergent treatment

↓

Ultracentrifugation

Supernatant contains inner membrane (IM) fraction
Pellet contains outer membrane (OM) fraction

b)

c)

Figure 5.4 Subcellular fractionation of *T. maritima* and Western blot analysis. (a) Workflow for the subfractionation of *T. maritima* cells into periplasmic (P), cytoplasmic (C), inner membrane (IM) and outer membrane (OM) components. (b) A Western blot was performed on the isolated fractions and a whole cell (WC) sample, using rabbit anti-*Tma* Hfq pAbs. A band near 12 kDa was observed for all fractions. (c) SDS-PAGE for the same isolated fractions and whole cell (WC) sample; a band around 12 kDa was also observed for all fractions.
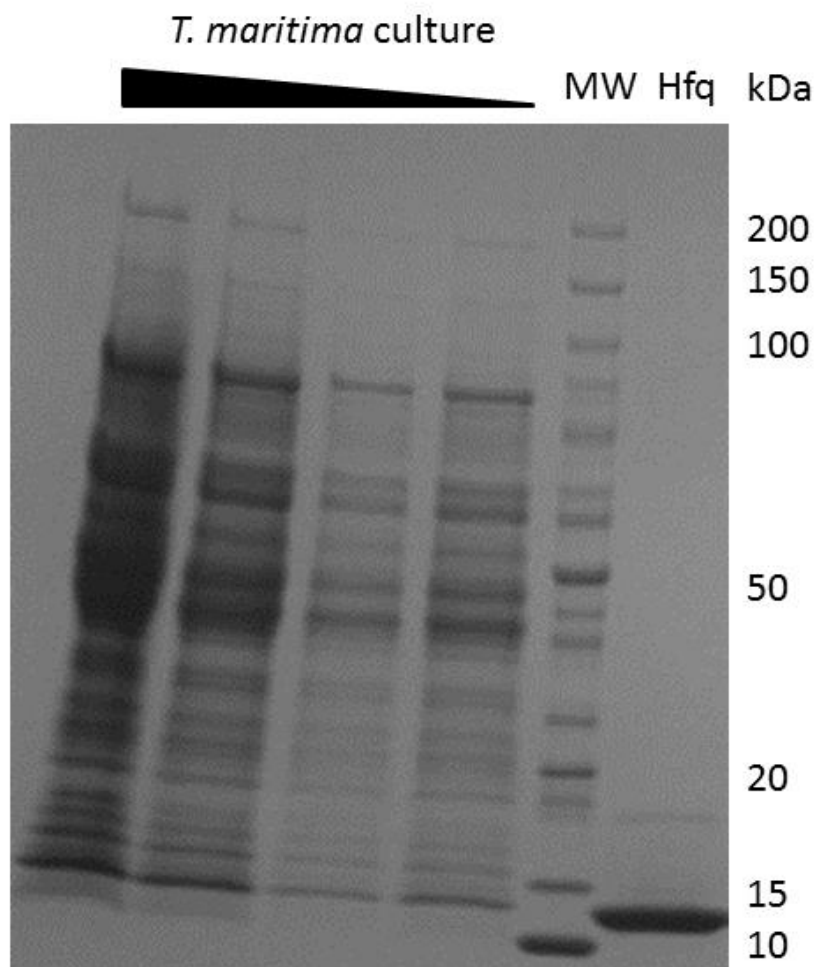
Figure 5.5 Western blot of *T. maritima* culture whole cell sample using rabbit anti-*Tma* Hfq pAbs. A whole cell sample of *T. maritima* from a culture grown at 80 °C for approximately 14-16 hours was serial diluted then separated by SDS-PAGE. Odyssey one-color protein molecular weight marker (MW) and purified recombinant *Tma* Hfq were run in parallel as standards for comparison. The protein band around 12 kDa is clearly not the most predominant component of the whole cell sample.
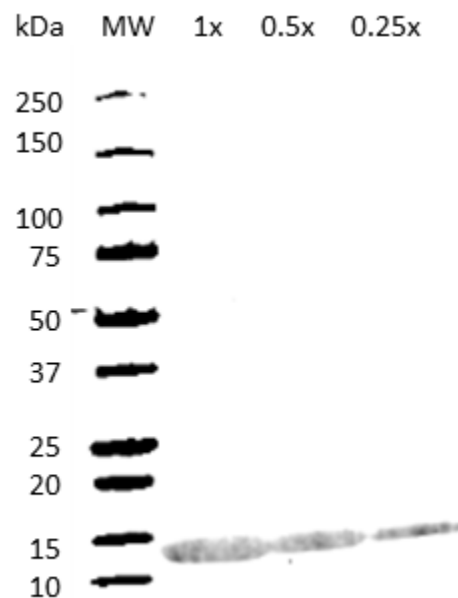
Figure 5.6 Western blot of hen egg white lysozyme using rabbit anti-*Tma* Hfq pAbs. Pure lysozyme was run on a 4-20% w/v TGX gel and transferred to a nitrocellulose membrane as describe in the Methods. A Western blot of a serial dilution of lysozyme, from 5.4 mg/mL (1x) to 1.4 mg/mL (0.25x) reveals a band near 12 kDa at each concentration. Odyssey one-color protein molecular weight marker (MW) was run in parallel as standards for comparison.
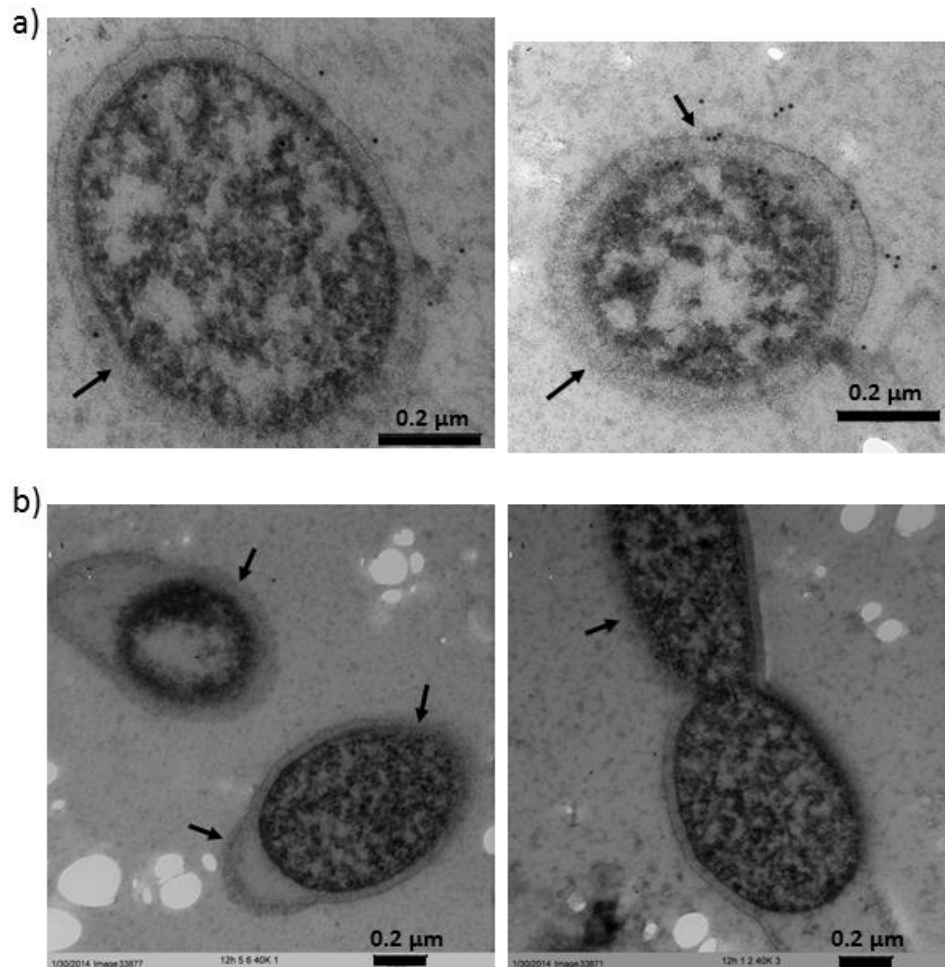
Figure 5.7 EM images of *T. maritima*. Representative images of *T. maritima* cells: (a) log phase (6 hours) and (b) early stationary phase (12 hours). The cells were fixed with a solution containing 2% w/v paraformaldehyde, 1% w/v glutaraldehyde, and 0.5% w/v uranyl acetate. Cells were then immunogold labeled using rabbit anti-*Tma* Hfq pAb as the primary Ab and donkey anti-rabbit IgG mAb conjugated to 10 nm gold beads as the secondary Ab. Black sphere in the images above are the gold beads conjugated to the secondary Ab and indicate the presence of Hfq. The presence of Hfq outside the cell is likely due to breaks in the cell membrane, highlighted by black arrows in the images above. These breaks in the cell membrane suggest partial lysis of the cells and loss of structural integrity.
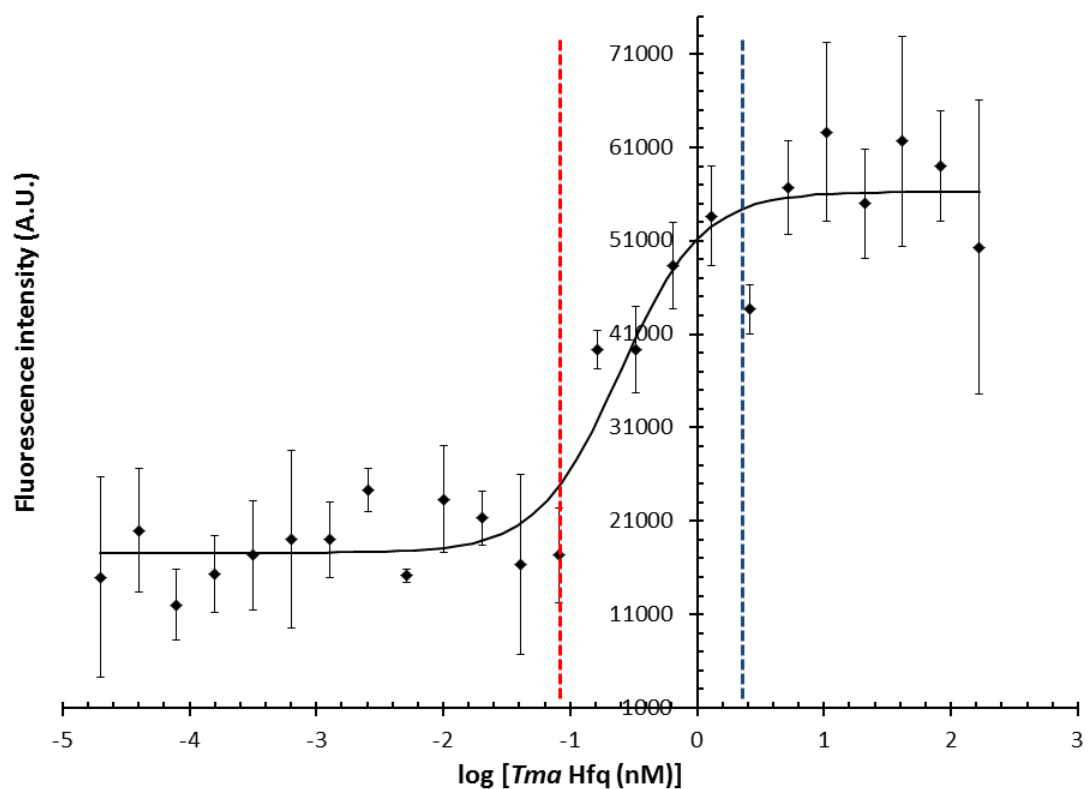
Figure 5.8 Sandwich ELISA for the quantification of *Tma* Hfq in a complex sample. A sandwich

ELISA was developed using unlabeled rabbit anti-*Tma* Hfq pAb as the capture Ab, which coat the

microplate well and Cy5-labeled rabbit anti-*Tma* Hfq pAb for the detection of bound antigen. A

standard curve was generated using recombinantly expressed and purified recombinant *Tma*

Hfq as a known antigen. The limit of detection was found to be roughly 81 pM (dashed red line)

and detection saturated near 5 nM (dashed blue line). Data were collected in three independent
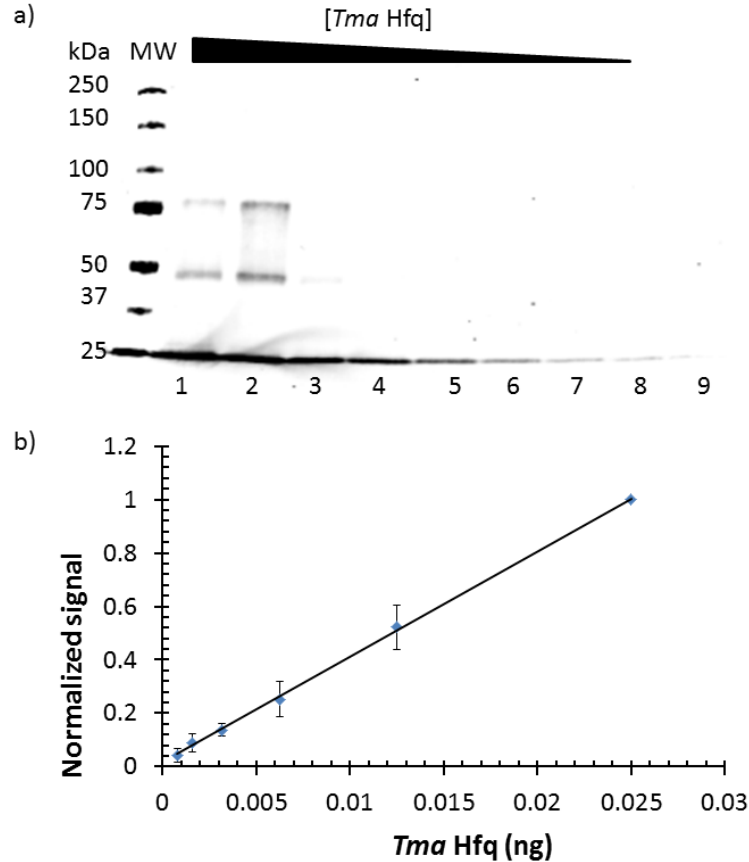
replicates.

Figure 5.9 Quantitative Western blot for *Tma* Hfq. (a) Representative Western blot for a serial

dilution (20 µM to 78 nM) of purified recombinant *Tma* Hfq. (b) Standard curve generated using

the normalized signal measured for lanes 4-9 of the Western blot, versus the amount of *Tma*

Hfq loaded in the corresponding lane. Signal was normalized using the signal measured for

0.025 ng (lane 4) of *Tma* Hfq. Signal for lanes 1-3 were not used to generate the standard curve

due to the presence of higher order oligomeric states that were not fully denatured at these

concentrations, which complicated the quantification and reproducibility of the signal

associated with these samples. Data were collected in three independent replicates, and were

fit to a simple linear regression with an R^2 of 0.9995. It should be noted that separation of the

Odyssey one-color protein molecular weight marker (MW) below 25 kDa was not observed this

is because these molecular weights are below the separation range of the gel.
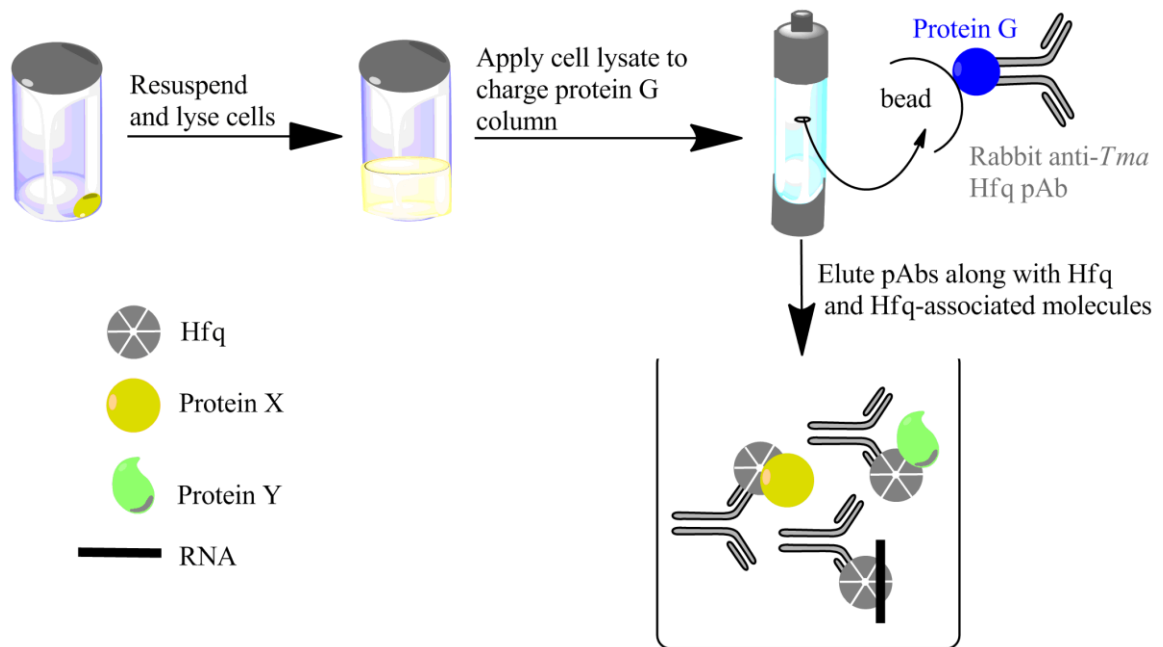
Figure 5.10 Co-immunoprecipitation of *Tma* Hfq and its binding partners. *T. maritima* cells are resuspended and lysed using a microfluidizer. The lysate is applied to a protein G column charged with rabbit anti-*Tma* Hfq pAbs. The pAbs bind to epitopes of *Tma* Hfq, retaining *Tma* Hfq and its binding partners on the column; all non-associated cellular components are washed away. The antibodies and associated *Tma* Hfq and binding-partners are then eluted.

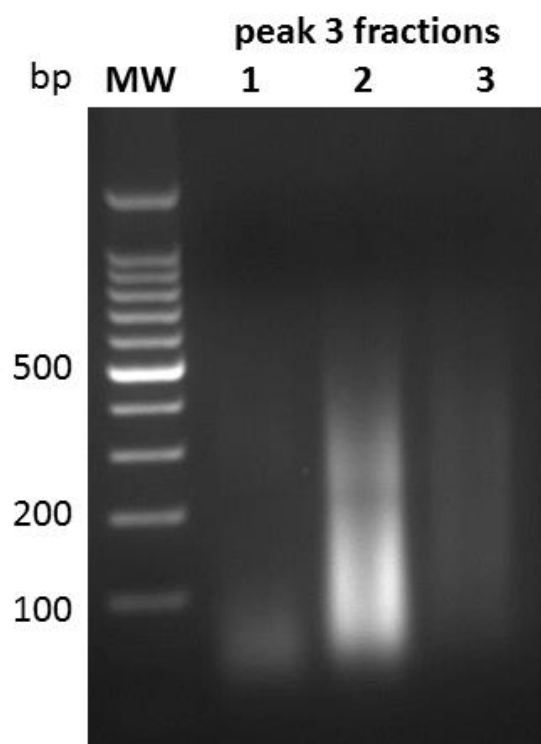**Figure 5.11 Agarose gel of Hfq-binding RNA components removed during nanoRNA enrichment.** RNA components were purified by phenol chloroform extraction of the three fractions that spanned peak 3 in the quaternary amine ion exchange chromatogram (Figure 2.1). RNA was separated on a 2% w/v agarose gel containing 0.1% v/v ethidium bromide and the Promega 100 base pairs (bp) DNA ladder (MW) was run in parallel as a standard for comparison. A discrete band is not observed in any of the fractions instead a smeared band that spans from approximately 50 bp to 500 bp is observed in all three fractions with fraction 2 having the highest concentration of RNA.
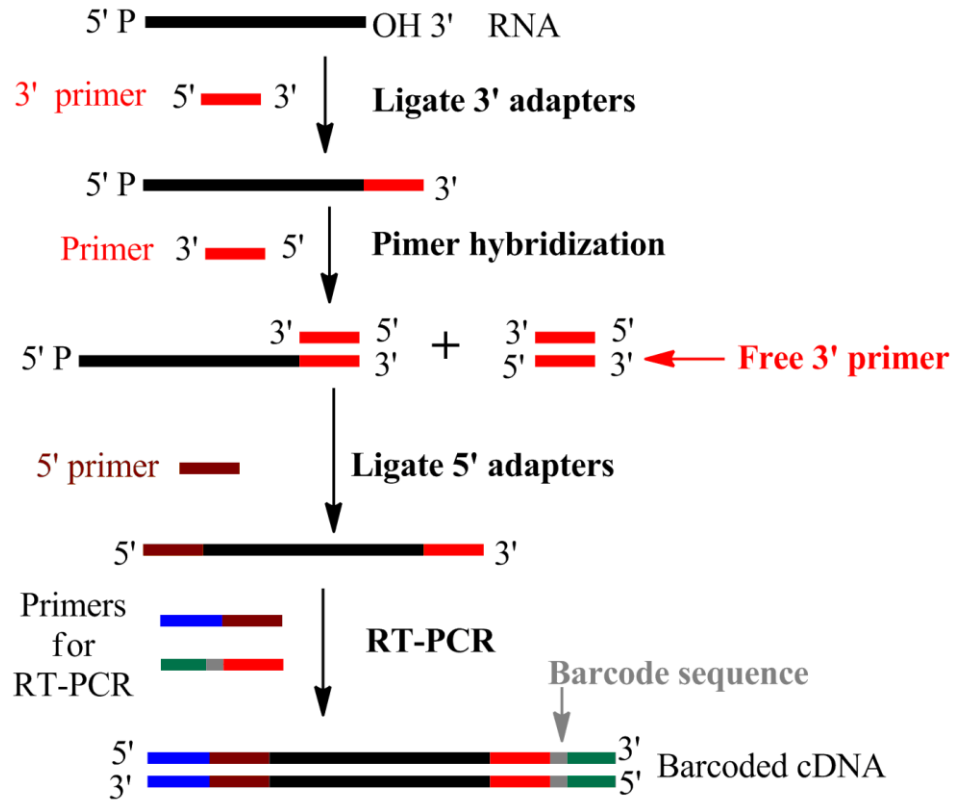
Figure 5.12 cDNA library generation in preparation for sequencing. The 3' adapter was ligated to the 3' end (red line) of RNA transcripts (black line) that were co-immunoprecipitated with Hfq. To prevent adapter dimer formation, a complementary primer (red line) to the 3' adapter was hybridized to the RNA product. The 5' adapter (maroon line) was then ligated with the 5' end of single stranded RNAs. Reverse transcription polymerase chain reaction (RT-PCR) primers containing a region complementary to the 5' and 3' adapter sequences (red and maroon segment of RT-PCR primers) were used to generate the complementary DNA (cDNA) library. The RT-PCR primers also contain an adapter sequence complementary to fragments of RNA on the surface of the MiSeq lane (blue and green segments of RT-PCR primers). For samples that will be mixed and run simultaneously with other samples in a single sequencing lane, a "barcode" sequence (grey segment of line) can be added to the RT-PCR primers in order to identify the sample that the sequence originated from.

Sample type	Amount of *Tma* cells (mL)	Amount of purified RNA (µg)†	A260/A280[*]	A260/A230[*]
Table 5.1 Yields and absorbance properties of the co-immunoprecipitated RNA samples that will be used for NGS.				
Negative control	300	0	2.14	2.01
RNA	300	34.27	1.98	0.79
nanoRNA	2400	3.14	1.48	0.87

.[*] Absorbance ratios (260 nm/280 nm and 260 nm/280 nm) were measured using a nanodrop instrument.

† The concentration (in ng/µL) of purified RNA was determined using the default settings for RNA on a nanodrop instrument, which use an extinction co-efficient of 40 ng cm μL^{-1} to estimate the concentration in of RNA in a sample. The volume of the sample was then used to determine the amount of purified RNA (in µg).

Appendix 1

Rapid Colorimetric Assays to Qualitatively Distinguish

RNA and DNA in Biomolecular Samples[*]

Introduction

Much of cell biology occurs via molecular interactions involving DNA and RNA [1]. These naturally occurring nucleic acids interact with one another [2], with proteins [3], and with a host of small-molecule compounds *in vivo.* The lifetime, affinity, and specificity of these interactions can vary significantly. Non-specific interactions with nucleic acids can have practical consequences for *in vitro* experiments involving mixtures of biomolecules, as it is possible, and even likely, that some nucleic acids will associate with the biomolecules of interest under a subset of the experimental conditions being used (ionic strength, solution pH, *etc.*). The presence of nucleic acid in the sample can affect downstream analysis; therefore, it is often desirable to remove nucleic acids from samples or at least identify their presence.

Protein production via heterologous over-expression of the recombinant protein in a host expression system is routinely performed in structural biology, biochemistry, and biophysics laboratories [4]. In preparing for further experiments, initial efforts generally focus on obtaining a sufficient quantity of the protein of interest (POI) in as pure a form as possible, ideally as a chemically homogeneous and biophysically monodisperse specimen. After disruption of the host cells, the early stages of a typical purification aim to isolate the POI from proteins, nucleic acids, and other components of the host system. However, some of these components may co-purify with the POI either due to association with the target protein or a chromatography matrix used in the purification scheme. Indiscriminate, high-affinity binding of host nucleic acids to a POI can cause issues in downstream experiments, such as fluorescence anisotropy assays of POI and RNA binding [5] and electrophoretic mobility shift assays [6]. Alternatively, unanticipated POI···nucleic acid associations can also be viewed fortuitously, as they illuminate the POI's nucleic acid-binding capacity. Either way, whether nucleic acids are

desired components or unwanted contaminants, one must quantify and identify co-purifying nucleic acids in preparation for downstream experiments.

Several analytical methods exist for detecting and quantifying nucleic acids in a sample, but most require a relatively pure sample or a separation step via electrophoresis or chromatography. Most of these methods utilize spectrophotometric or fluorometric measurements to quantitate the amount of nucleic acid in solution as recently described by De Mey *et al* [7]. However, the crucial step of distinguishing the polynucleotide as either RNA or DNA is beyond the scope of many of these quantitation approaches.

Here we provide a set of colorimetric assays to rapidly identify the types of nucleic acid components in a proteinaceous sample. The protocols described here (Figure A1.1) can be efficiently executed without isolation of the potential nucleic acid impurities, The methods utilize the Benedict's assay [8], the Bial's orcinol assay [9,10], and the Dische's diphenylamine reaction [11,12] to test for reducing sugars, pentose rings, and 2'-deoxypentoses, respectively. The Benedict's test (Figure A1.2a) utilizes the ability of the linear, open-chain (aldehyde) form of an aldose sugar to reduce Cu^{2+}, with concomitant oxidation of the sugar's carbonyl to a carboxylate moiety and production of Cu_2O as a precipitate. This reaction will test positive with free reducing sugars, but not with pentose sugars that are locked into cyclic form as part of the covalent backbone of a DNA or RNA polynucleotide. Both the Bial's orcinol and Dische's diphenylamine reactions (Figure A1.2b-c), are based on initial destruction of the polynucleotide backbone, via depurination of the nucleoside and acid- or base-catalyzed hydrolysis of the parent nucleotides, to yield furan-2-carbaldehyde (*furfural*) derivatives; these derivatives then react with either orcinol (Bial's) or diphenylamine (Dische's) to form colored condensation products of unknown chemical structure. The DNA versus RNA specificity of Dische's diphenylamine reaction stems from the fact that the pentose sugar must be 2'-deoxygenated in

order to be susceptible to oxidation to ω-hydroxylevulinyl aldehyde, which further reacts with diphenylamine under acidic conditions to yield a bright blue condensate (Figure A1.2c). Using the streamlined protocols described here, we have found that these sugar-specific colorimetric reactions can differentiate between RNA and DNA, and will also indicate the presence of free reducing sugars such as glucose, fructose, or ribose in a biomolecular sample of unknown composition.

Methods

Benedict's Assay. This assay begins by preparing 6x Benedict's reagent — 940 mM anhydrous sodium carbonate, 588 mM sodium citrate dehydrate, 68 mM copper (II) sulfate pentahydrate. This reagent can be stored at room temperature (RT) for at least six months with no noticeable change in reactivity. The reactions consisted of one volume Benedict's reagent and five volume sample. Reactions were incubated for 20 min in a boiling water bath. The heated sample was then removed from the bath and allowed to cool at RT for 10 min. For qualitative measurements, samples were visually compared to negative (DNA, RNA, water, bovine serum albumin (BSA)) and positive (ribose) controls. For quantitative measurements, samples were centrifuged at > 9,300g for 5 min in order to sediment any particulate material. The absorbance of the supernatant was then measured at 475 nm.

Bial's orcinol assay. This assay begins by preparing Bial's reagent A and B—*Reagent A* [0.05% w/v FeCl$_3$•6H$_2$O in concd HCl] and (ii) Reagent B [422 mM orcinol monohydrate prepared in 95% ethanol]. *Reagent A* can be stored at RT for six months; *Reagent B* can be stored at 4 °C for one month, covered with foil to limit light exposure. To prepare a 2x Bial's reagent the stock solutions were mixed in a 15:1 volumetric ratio of A:B prior to use. The reaction consisted of 1 volume Bial's reagent and 1 volume sample. Reactions were incubated for 20 min in a boiling

water bath. The heated sample was then removed from the bath and allowed to cool at RT for 10 min. For qualitative measurements, samples were visually compared to negative (water, BSA) and positive (ribose, RNA, DNA) controls. For quantitative measurements, samples were centrifuged at > 9,300g for 5 min in order to sediment any particulate material. The absorbance of the supernatant was then measured at 600 nm.

Dische's diphenylamine assay. Prepared 2x *Dische's diphenylamine reagent* — 60 mM diphenylamine, 11 M glacial acetic acid, 179 mM sulfuric acid, 0.62% v/v ethanol. This reagent can be prepared in advance and stored at RT in a dark container or covered with foil, to limit light exposure for up to three months with no apparent change in reactivity. The reaction consisted of 1 volume Dische diphenylamine reagent and 1 volume sample. Reactions were incubated for 20 min in a boiling water bath. The heated sample was then removed from the bath and allowed to cool at RT for 10 min. For qualitative measurements, samples were visually compared to negative (water, BSA) and positive (ribose, RNA, DNA) controls. For quantitative measurements, samples were centrifuged at > 9,300g for 5 min in order to sediment any particulate material. The absorbance of the supernatant was then measured at 600 nm.

Representative Results

Representative qualitative data are shown in Figures A1.2 and A1.3 for the Benedict's (a), Bial's orcinol (b), and Dische's diphenylamine (c) assays, and standard curves for these three assays are shown in Figure A1.4. In Figure A1.2, the right panels show positive and negative control experiments using suitably reactive and unreactive analytes. In the Benedict's assay (a), the positive control is ribose (0.42 mg/ml), while negative controls are a generic protein (0.75 mg/ml BSA) and two non-reducing sugars (DNA at 0.75 mg/ml and RNA at 12.5 mg/ml). In the orcinol assay (b), the positive controls are ribose (0.15 mg/ml), RNA (7.5 mg/ml), and DNA (0.45 mg/ml),

while BSA protein (0.45 mg/ml) is a negative control. In the diphenylamine assay (c), DNA (0.45 mg/ml) is shown as a positive control, and RNA (7.5 mg/ml), ribose (0.15 mg/ml), and BSA (0.45 mg/ml) serve as negative controls.

Figure A1.3 shows titration series for each colorimetric assay using a corresponding positive control: ribose, RNA, and DNA for Benedict's, Bial's, and Dische's reactions, respectively. The dilution ranges in Figure A1.3 vary because each reaction has a distinct visual detection limit, depending on the type of sugar being assayed. Samples were clarified by centrifugation prior to visual analysis. It should be noted that two distinct color changes are observed for Bial's reaction in the concentration range shown; at RNA concentrations between 0.50 mg/mL and 2.5 mg/mL a blue product is observed, whereas at higher concentration an aqua product is observed.

Spectrophotometric, rather than visual, detection can be used to improve the limit of detection as shown in Figure A1.4 for the (a) Benedict's, (b) Bial's, and (c) Dische's assays. Though the protocols described here are intended as primarily qualitative assays, the Beer-Lambert relation between absorbance and concentration enables a semi-quantitative estimation of the sugar or nucleic acid content. A practical consideration when using these assays for semi-quantitative analysis is that precipitated product must be removed prior to absorbance measurements. For all three assays, centrifugation at maximal speeds (e.g. 9,300g) for approximately 5 minutes was found to sufficiently clarify samples. After centrifugation, clarified samples can be transferred to a cuvette or microplate for absorbance measurements.

The linear ranges of each of the assays should be considered when using these assays for semi-quantitative analyses because the samples may need to be diluted prior to analysis. For Benedict's assay (Figure A1.4a) the reaction readout (A_{475nm}) for ribose was found to be in the range of 0.04 to 0.5 mg/ml. The absorbance data (A_{660nm}) for Bial's assay using baker's yeast RNA

(Figure A1.4b) were linear from 0 to 0.5 mg/ml analyte. Though Bial's reaction samples were centrifuged for clarification before absorbance measurements, no precipitant was found at these low concentrations of analyte and therefore the centrifugation step was not strictly necessary. For the Dische's assay, the absorbance measurements (A_{600nm}) for calf thymus DNA were determined to be linear over the range 0.15 to 0.75 mg/ml of analyte (Figure A1.4c).

Finally, Figure A1.5 illustrates the robustness of the assays by showing the Dische's reaction with samples of varying heterogeneity: two concentrations of DNA are shown as a positive control in samples 1 and 2, samples containing various ratios of DNA and RNA are shown in sample 3 through 5, and samples 6 and 7 show DNA in the absence or presence of the nucleic acid-binding protein Hfq [13], respectively. Note that the positive result of the Dische's assay is preserved for DNA-containing samples even in the presence of 'contaminating' RNA or protein.

Discussion

The colorimetric assays presented here offer a simple approach to rapidly assess the chemical nature of biomolecular mixtures, such as those encountered when purifying proteins, RNAs or complexes from whole-cell lysate. As structural biology pursues more native-like assemblies, progressively greater challenges - such as sample heterogeneity - will be posed by the intricate and multi-component complexes. Supramolecular assemblies are often only marginally stable, and their successful isolation may demand less stringent purification conditions. Under such milder conditions, both authentic and spurious POI⋯nucleic acid interactions are more likely to persist and thereby interfere in downstream assays focused on a target protein or nucleic acid. A necessary first step is identification of the types of chemical components in these samples.

Methods to detect RNA, DNA, and protein vary depending on quantity and concentration of analytes, available resources, time constraints and, perhaps most importantly, one's prior knowledge about the likely chemical composition of the analytes. Proteinaceous material can be detected by many well-established methods, including gel electrophoresis [14], colorimetric protein assays [15], and mass spectrometry [16,17]. Similarly, well-established methods exist for the detection of nucleic acids, including electrophoresis with fluorescent dyes [18,19] and Northern or Southern blots [20].

Each type of method features characteristic strengths and weaknesses. For instance, PicoGreen, SYBR-Gold, and other cyanine-based fluorescent dyes are sensitive to nucleic acid at many orders of magnitude beyond the colorimetric assays described here and feature broad dynamic ranges for quantitation purposes [21,22]. Additionally, these dyes exhibit various degrees of selectivity allowing for the differentiation of nucleic acids not only by type (DNA, RNA), but also by secondary structure (single stranded, ss; double-stranded, ds). For instance, SYBR-Gold binds to most nucleic acids, whereas PicoGreen and OliGreen bind specifically to dsDNA and ssDNA, respectively. However, cyanine-based fluorescent dyes are not without limits: more advanced equipment is required for detection versus the simple visual readout of colorimetric assays and there are restrictions on assay conditions (*e.g.*, a recommended pH range of 7–8.5 for SYBR-Gold; a 30% reduction of PicoGreen signal intensity at >200 mM NaCl).

The strengths and weaknesses of colorimetric and fluorescence-based assays make them complementary techniques. The sensitivity of cyanine-based fluorescent dyes makes them particularly useful in case of negative results with the rapid colorimetric assays or as a way to more carefully (quantitatively) expand on initial results from colorimetric assays. A fundamental benefit of the colorimetric nucleic acid protocols, in addition to simplicity of use, is that they rely

solely on the intrinsic covalent structure of nucleic acids rather than secondary structure, reactivity, or any other properties of the biopolymer that might vary with sequence.

The protocols described here are robust against most difficulties, particularly when performed with simple visual inspection of the reaction products; special care must be taken for semi-quantitative analyses. For instance, in the Benedict's assay the red precipitant that forms in the presence of high concentrations of ribose must be removed prior to spectrophotometric analysis; this is readily achieved via centrifugation. For the Dische's assay, the addition of diphenylamine reagent is accompanied by formation of a white precipitant that can be solubilized by heating; a green precipitant that forms in the presence of DNA may interfere with spectrophotometric measurements and therefore must be removed prior to semi-quantitative analysis. In addition, each assay is based on chemical reactions that are susceptible to potential interferents, as mentioned in the Introduction and further detailed below. Theoretically, a high relative concentration of RNA in a DNA/RNA mixture could mask the expected positive result of Dische's assay; this false negative for DNA stems from the fact that a high mole fraction of RNA also reacts, via furfural intermediates, with the Dische's reagents, but yields a colorless product rather than the blue adduct produced by reaction with the 2'-deoxypentose sugar of DNA.

Beyond the obvious case of sugars, some classes of lipids and proteins could hypothetically results in false positives with these colorimetric assays due to cross-reactivity. In principle, classes of cellular lipids that could conceivably interfere include glycosyl glycerols and other glycerolipids conjugated to sugars, saccharolipids, and lipopolysaccharides. In practice, these classes of lipids are unlikely to pose a problem in working with lipophilic proteins because they are relatively rare, compared to the much more abundant phospholipids, and therefore below the detection limits of these assays. Because most of the aforementioned cellular lipids contain hexoses (galactose, glucose) rather than pentoses, they are not expected to interfere as

false-positives in the Bial's or Dische's assays. Free monosaccharides that may give false-positives include fructose, galactofuranose, or other furanoses.

On a related note, interference from unwanted sugars could become an issue for recombinant proteins expressed with the maltose-binding protein (MBP) tag since maltose is used as an eluent during affinity chromatography. Thus, one would have to exercise caution in the post-chromatographic steps to ensure that residual maltose did not yield spurious results since residual maltose in downstream preparations could interfere with the Benedict's assay.

Applications of the protocols presented here include the identification of co-purifying compounds, assessment of RNA or DNA purity, and the detection of residual nucleic acid or sugar contamination in protein samples. Although there are alternative approaches to such applications, the methods described here are highly efficient in terms of both time and cost, and can therefore be easily integrated into an experimental workflow. The early identification of co-purifying compounds as free reducing sugar, RNA, DNA, or protein can guide the design of downstream purification steps. In this and other ways, the colorimetric assays described here can be combined with well-established methods for protein and nucleic acid determination to achieve a rapid and robust system for elucidating the RNA, DNA, and protein components in heterogeneous biomolecular samples.

References

1. Voet D, Voet JG (2011) Biochemistry. Hoboken, NJ: John Wiley & Sons. xxv, 1428, 1453 p. p.

2. Adams RLP, Knowler JT, Leader DP (1986) The Biochemistry of the Nucleic Acids. London ; New York: Chapman and Hall. xviii, 526 p. p.

3. Rice PA, Correll CC (2008) Protein-nucleic acid interactions: Structural biology. Cambridge: RSC Pub. xvii, 397 p. p.

4. Graslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, et al. (2008) Protein production and purification. Nat Methods 5: 135-146.

5. Pagano JM, Clingman CC, Ryder SP (2011) Quantitative approaches to monitor protein-nucleic acid interactions using fluorescent probes. RNA 17: 14-20.

6. Hellman LM, Fried MG (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nat Protoc 2: 1849-1861.

7. De Mey M, Lequeux G, Maertens J, De Maeseneire S, Soetaert W, et al. (2006) Comparison of DNA and RNA quantification methods suitable for parameter estimation in metabolic modeling of microorganisms. Anal Biochem 353: 198-203.

8. Benedict SR (2002) A reagent for the detection of reducing sugars. 1908. J Biol Chem 277: e5.

9. Endo Y (1970) A simultaneous estimation method of DNA and RNA by the orcinol reaction and a study on the reaction mechanism. J Biochem 67: 629-633.

10. Almog R, Shirey TL (1978) A modified orcinol test for the specific determination of RNA. Anal Biochem 91: 130-137.

11. Dische Z (1955) New color reactions for determination of sugars in polysaccharides. Method Biochem Anal 2: 313-358.

12. Burton K (1956) A study of the conditions and mechanism of the diphenylamine reaction for the colorimetric estimation of deoxyribonucleic acid. Biochem J 62: 315-323.

13. Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. Nat Rev Microbiol 9: 578-589.

14. Hakomori SI, Jeanloz RW (1961) Isolation and characterization of glycolipids from erythrocytes of human blood A (plus) and B (plus). J Biol Chem 236: 2827-2834.

15. Sapan CV, Lundblad RL, Price NC (1999) Colorimetric protein assay techniques. Biotechnol Appl Bioc 29 (Pt 2): 99-108.

16. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422: 198-207.

17. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20: 3551-3567.

18. Suenaga E, Nakamura H (2005) Prestaining method as a useful tool for the agarose gel electrophoretic detection of polymerase chain reaction products with a fluorescent dye SYBR gold nucleic acid gel stain. Anal Sci 21: 619-623.

19. Huang Q, Baum L, Fu WL (2010) Simple and practical staining of DNA with GelRed in agarose gel electrophoresis. Clin Lab 56: 149-152.

20. Wahl GM, Meinkoth JL, Kimmel AR (1987) Northern and Southern blots. Methods Enzymol 152: 572-581.

21. Gagna CE, Kuo HR, Chan NJ, Mitacek EJ, Spivak A, et al. (2007) Novel DNA staining method and processing technique for the quantification of undamaged double-stranded DNA in epidermal tissue sections by PicoGreen probe staining and microspectrophotometry. J Histochem Cytochem 55: 999-1014.

22. Yan X, Grace WK, Yoshida TM, Habbersett RC, Velappan N, et al. (1999) Characteristics of different nucleic acid staining dyes for DNA fragment sizing by flow cytometry. Anal Chem 71: 5470-5480.
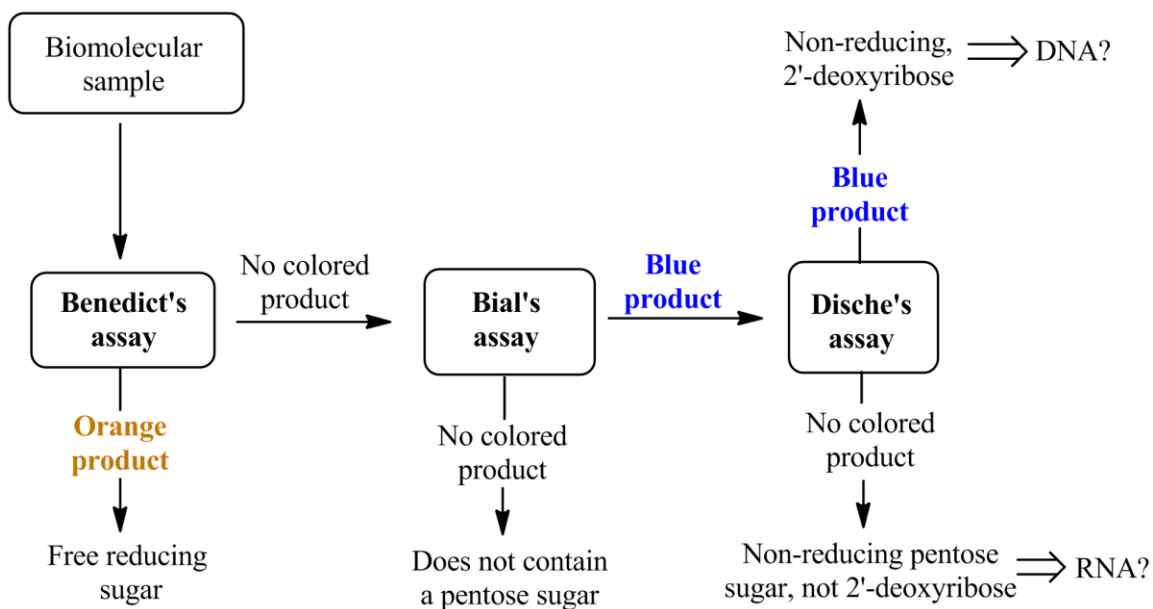
Figure A1.1 Decision tree for analyzing the colorimetric results obtained from the Benedict's, Bial's, and Dische's reactions. An orange product in the Benedict's assay indicates the presence of free-reducing sugars in the biomolecular sample, such as glucose or ribose. A blue product in the Bial's assay indicates the presence of a pentose sugar. Finally, a blue product in the Dische's assay indicates the presence of deoxyribose. The decision tree above is designed with the goal of distinguishing between RNA and DNA in a biomolecular sample and ruling out the presence of free reducing sugars which can also produce positive results in both the Bial's and Dische's assays. For instance, the presence of free ribose in the sample would yield an orange product and blue-green product for the Benedict's and Bial's assay, respectively.
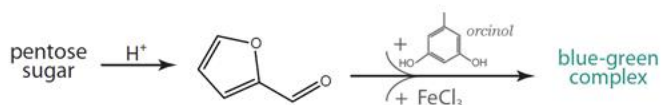
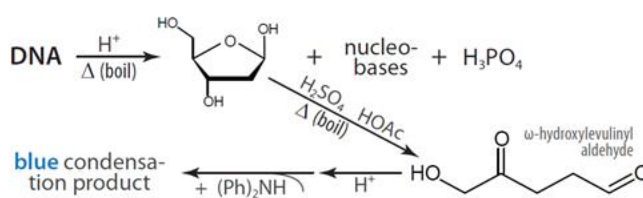**a) Benedict's assay** for free reducing sugars
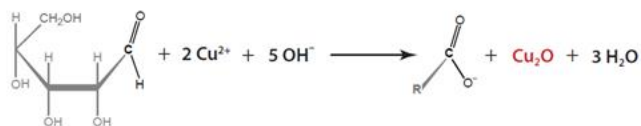
ribose BSA RNA DNA
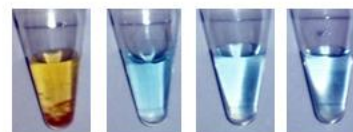
b) Bial's assay for pentose rings

ribose BSA RNA DNA

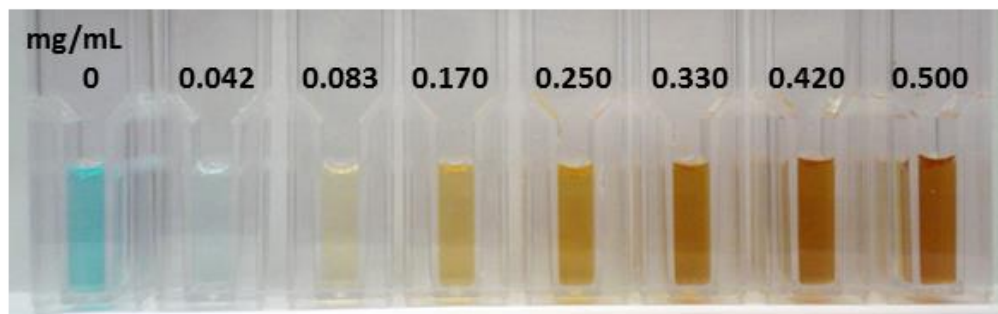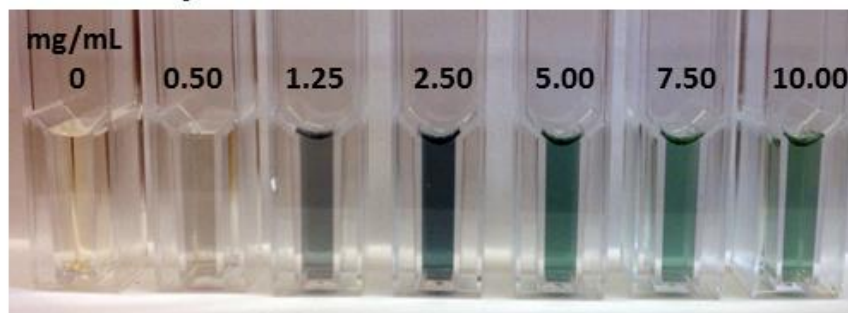**c) Dische's Assay** for 2'-deoxyribose

ribose BSA RNA DNA

Figure A1.2 Underlying chemical reactions and controls are shown for the colorimetric assays, with the detectable colored product indicated. (a) An insoluble, red precipitate of Cu_2O is the positive result of Benedict's assay for reducing sugars. Ribose is used a positive control for the Benedict's assay, whereas RNA, DNA, and BSA serve as negative controls. (b) Upon heating and acidification, sugars containing pentose rings will decompose to furfural, which then reacts with orcinol, a component of the Bial's reagent, to yield a soluble blue-green adduct. Ribose, RNA, and DNA are used as positive controls for the Bial's assay and BSA is used as a negative control. (c) In Dische's assay, hydrogen at the 2' position enables an acid-catalyzed ring-opening oxidation reaction, the aldehyde product of which further reacts with diphenylamine [$(Ph)_2NH$] to yield a bright blue product. DNA is used as a positive control for the Dische's assays and ribose, BSA, and RNA are used as negative controls. Chemical structures remain unknown for the large, multi-ring condensation products of the orcinol (b) and diphenylamine (c) reactions.

a) Benedict's assay: Ribose serial dilution

mg/mL							
0	0.042	0.083	0.170	0.250	0.330	0.420	0.500

b) Bial's assay: RNA serial dilution

mg/mL						
0	0.50	1.25	2.50	5.00	7.50	10.00

c) Dische's assay: DNA serial dilution

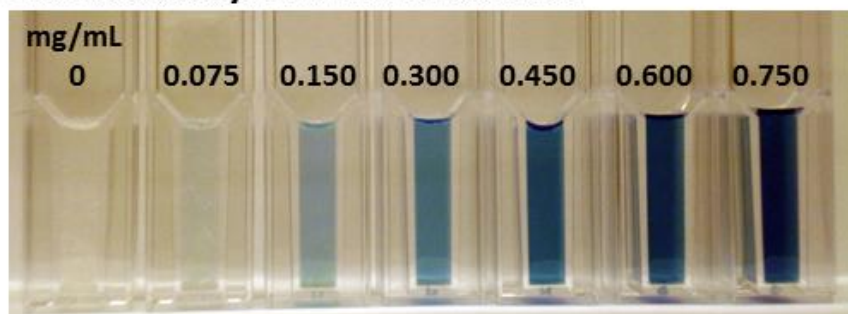| mg/mL						
0	0.075	0.150	0.300	0.450	0.600	0.750

Figure A1.3 Serial dilution for qualitative analysis. Sample results are shown for the Benedict's (a), Bial's (b), and Dische's (c) colorimetric assays. The serial dilutions illustrate the range of detection for qualitative analysis of the three assays. Standard baker's yeast RNA (b) and calf thymus DNA (c) were the analytes in these titration series.

a) Benedict' Assay: Ribose standard curve

$y = 0.5186x - 0.1713$
$R^2 = 0.9986$
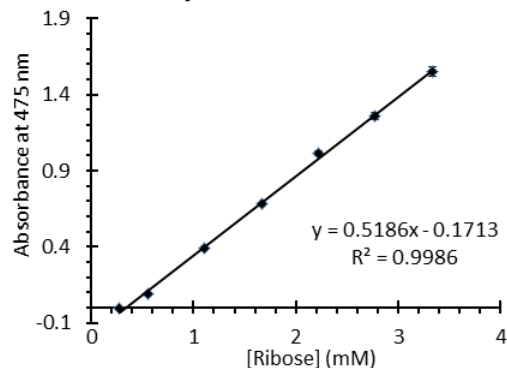
c) Dische's Assay: DNA standard curve

$y = 1.8836x - 0.0186$
$R^2 = 0.9979$
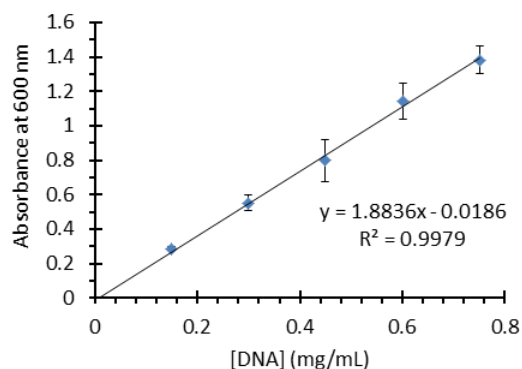
b) Bial's Assay: RNA standard curve

$y = 0.406x - 0.0554$
$R^2 = 0.9983$
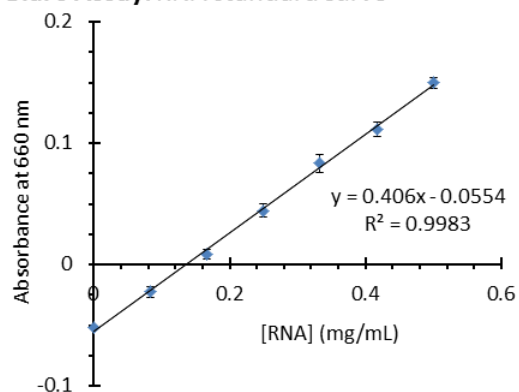
Figure A1.4 Standard curves of reference compounds for each colorimetric assay. Calibration curves are shown for the Benedict's (a), Bial's (b), and Dische's (c) assays, depicting the linear response region for each assay. The linear regression fits and corresponding correlation coefficients are indicated for each assay. Standard baker's yeast RNA (b) and calf thymus DNA (c) were used as the analyte in these titration series. Each assay was repeated in triplicate.
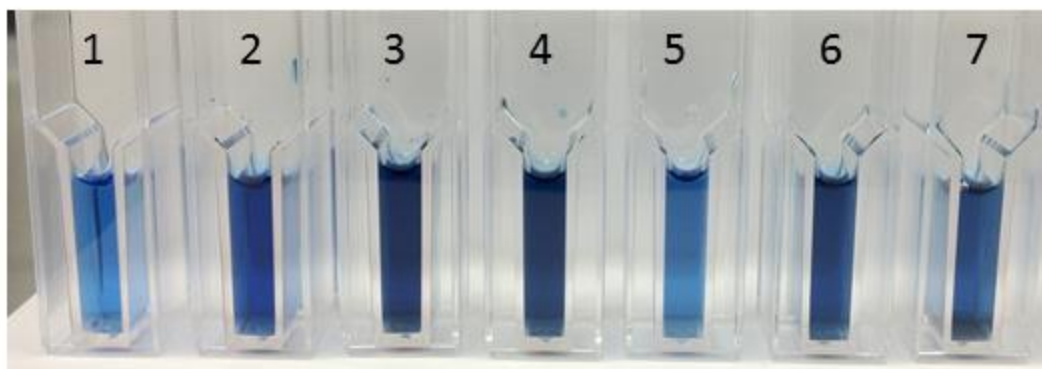
Figure A1.5 Dische's reaction for samples of varying heterogeneity. Samples (1) and (2) are homogeneous controls containing 0.75 mg/mL and 1.50 mg/mL DNA, respectively. Samples (3), (4), and (5) contain a mixture of RNA and DNA with (3) containing 1.5 mg/mL of both RNA and DNA, (4) containing 1.5 mg/mL DNA and 0.75 mg/mL RNA, and (5) containing 1.5 mg/mL RNA and 0.75 mg/mL DNA. Sample (6) contains 1.5 mg/mL DNA in 12.5 mM Tris pH 8.0 and 175 mM NaCl. Sample (7) contains 1.5 mg/mL DNA in 12.5 mM Tris pH 8.0 and 175 mM NaCl in the presence of 0.485 mg/mL Hfq from *Thermotoga maritima*.

Appendix 2

Purification and characterization of rabbit anti-

Thermotoga maritima Hfq polyclonal antibodies

Introduction

Immunoglobulins (Ig), also known as antibodies (Ab), are components of the immune system that bind foreign substances such as toxins, bacteria, or viruses. Antibodies are a component of the adaptive immune system found in vertebrates [1,2]. Antibodies are Y-shaped protein assemblies comprised of two identical heavy chains and two light chains that are linked by disulfide bonds (Figure A2.1) [3,4,5]. In mammals, the five classes of antibodies are distinguished by the type of heavy chain (γ, μ, α, ε, or δ) present in the molecule [6,7]. The heavy chain has two characteristic regions (Figure A2.1b): a "crystallizable fragment" region (Fc-region), conserved within a species, and a variable region, which is involved in specific antigen recognition [8,9]. Each heavy chain can associate with one of two types of light chains (κ and λ), which are also comprised of a constant and variable region (Figure A2.1b) [3,6]. The light and heavy chains consist of two and four immunoglobulin protein domains, respectively, with each domain containing two beta sheets (Figure A2.1a) [10].

The natural molecular specificity of antibodies allows for their use in biochemistry for the detection of an antigen of interest within a cell [11] or complex sample, such as cell lysate [12,13]. For biotechnological applications, antibodies to a target antigen are generally produced in small mammals such as rabbits, mice, or rats; generally by immunization with the purified antigen (Figure A2.2). For this reason the antigen may also be referred to as the "immunogen".

The first step in the *in vivo* Ab production pathway is engulfment of the antigen by an antigen-presenting cell (APC). After engulfment, the antigen is degraded and fragments of the antigen, along with a major histocompatibility complex (MHC) class II protein (antigen-MHC II complex), are displayed on the cell surface of the APC [14]. Helper T cells that specifically recognize the antigen-MHC II complex will associate with the APCs, initiating proliferation of helper T cells [15]. Additionally, association of the antigen with surface Abs on B cells leads to

engulfment, degradation, and presentation of fragments of the antigen along with MHC class II proteins [6,16]. Recognition of the antigen-MHC class II complex by helper T cells initiates proliferation and differentiation of the B cells into plasma and memory cells [17]. Plasma cells are terminally differentiated cells that live for three to four days and secrete large amounts of Abs. Memory cells are long-lived, differentiated cells that retain the surface Ab specific to a given antigen; upon subsequent exposure to the antigen, memory cells can activate a more rapid immune response [6,16].

In order to produce sufficient quantities of Ab for research purposes, the host animal is exposed to the immunogen multiple times after the primary immunization; these subsequent antigen exposures are referred to as "boosters". The first immunization produces a slow and weak immune response, with low to moderate levels of Ab secreted into the serum. Subsequent exposures produce a faster and more aggressive immune response, and higher levels of Ab in the serum. After multiple exposures, sera are obtained from the host animal and immunoglobulin G (IgG) is purified from the sera samples using established techniques, such as affinity chromatography with a protein G column [6]. Protein G affinity chromatography methods are based on the naturally occurring immunoglobulin-binding affinity of protein G [18]. IgG Abs are used in most biochemical assays because they are present at the highest concentration in serum [6] and therefore can be obtained at sufficiently high levels.

Abs purified from animal serum are polyclonal (pAb), which means a chemically heterogeneous mixture (in terms of the antigen-binding sites), with specificity to various epitopes of the antigen. Alternatively, monoclonal antibodies (mAb) are a homogeneous line of Abs produced by generating a clone of a single B cell [6,7]. There are advantages and disadvantages to both types of Abs. For example, pAbs are more robust to different sample conditions, as different individual Abs recognize different epitopes on the antigen, whereas

mAbs may only work under a subset of sample conditions (*e.g.* denaturing or native conditions). Polyclonal antibodies have an increased risk of cross-reacting with other antigens because they recognize multiple epitopes [7]. In addition to less cross-reactivity, mAbs have the added advantage of being unlimited and consistent between productions; however, they are more time consuming and expensive to produce than pAbs [7].

The unparalleled ability of Abs to recognize a specific antigen, such as a protein, nucleic acid, or small molecule ligand, has been employed in the development of numerous clinical and research techniques [7]. The goal of many of these techniques is the detection of the antigen (often at trace levels) in a prepared sample. Immunoblots and enzyme-linked immunosorbent assays (ELISA) are two common and well-established methods for detecting an antigen or Ab in a sample. In an immunoblot (Figure A2.3), a sample is transferred to a suitable membrane (e.g. nitrocellulose for protein antigens) and the membrane is blocked with an inert molecule (*e.g.* milk or bovine serum albumin (BSA)) to minimize non-specific interactions with the antigen-specific Ab. In order to detect an interaction between the Ab and antigen, the Ab is often labeled with a fluorophore or an enzyme capable (e.g. horse-radish peroxidase [19,20]) of being easily read out. Alternatively, a labeled secondary Ab, specific to the antigen-specific Ab (primary Ab), can be used for detection. Western blots and dot blots are two types of immunoblots used for the detection of a protein antigen. In a Western blot, the sample is separated by polyacrylamide gel electrophoresis prior to transfer to the membrane (Figure A2.3); in contrast, samples (often already purified) are directly applied to the membrane in a dot blot.

ELISAs are a high-throughput immunoassay that can be used for either qualitative or quantitative analysis of a soluble sample. There are four general types of ELISAs: direct, indirect, sandwich, and competitive [21]. In a direct ELISA, purified antigen is aliquoted into a 96-well

plate which is subsequently blocked and then probed with a primary Ab [22]. In an indirect ELISA (Figure A2.4), the protocol is the same as the direct ELISA except that a secondary antibody is used for detection [23]. Direct and indirect ELISAs are often used for the detection of Abs to a specific antigen in sera or for evaluating the quality (*e.g.* specificity and affinity) of an Ab product [22,24]. Indirect ELISAs have the advantage of not requiring conjugation of a label to the primary Ab, which can be both time consuming and costly, instead commercially available secondary antibodies, can be obtained at a substantially lower cost. A direct ELISA should be used if there is substantial cross-reactivity between the secondary Ab and the antigen containing sample. In a sandwich ELISA (Figure A2.5), also known as a capture ELISA, a 96-well plate is coated with a capture Ab (i.e. primary Ab specific to antigen) and blocked for non-specific interactions prior to incubation with an antigen-containing sample; a second primary Ab (generated in a different host organism then the capture Ab) is then incubated with the sample followed by a labeled secondary Ab specific to the second primary Ab [7]. Alternatively, the second primary Ab can be labeled for detection. A sandwich ELISA is commonly used for complex samples, such as cell lysate or sera, to determine if an antigen of interest is present [7]. Finally, in a competitive ELISA the plate is coated with purified antigen and blocked prior to the addition of primary Ab, which has previously been incubated with an antigen-containing sample. Ab not bound to antigen in the sample will bind to the antigen bound to the surface of the well, Ab not bound to the well is then washed away, and the bound Ab is detected by the addition of a labeled secondary Ab [23,24]. A competitive ELISA can also be used to detect antigen in a complex sample, but has the advantage over sandwich ELISAs of not requiring two primary Abs [23].

In the work of this dissertation, pAbs specific to recombinant *Thermotoga maritima* Hfq (*Tma* Hfq) were produced in rabbits and purified using affinity chromatography on a protein G

conjugated resin (this resin specifically binds rabbit IgG). Dot blot showed that the pAbs bind with high specificity to *Tma* Hfq. An indirect ELISA indicates that the Abs have a picomolar affinity for *Tma* Hfq. Conditions were successfully optimized for the detection of *Tma* Hfq in a Western blot.

Methods

Production of anti-*Tma* Hfq pAb. Recombinant *Tma* Hfq, purified according to the protein purification protocol described in Chapter 2, was dialyzed into 10 mM Na_2HPO_4, 1.8 mM NaH_2PO_4, 2.7 mM KCl, and 137 mM NaCl at pH 7.40 (PBS) and concentrated to 1 mg/mL using a 3350-Da molecular weight cut off (MWCO) concentrator. After verifying protein purity by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using a 4-20% w/v TGX gel and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), pAbs were then produced against this protein sample by Covance (Denver, PA) using their 77-day production protocol in four specific-pathogen-free New Zealand White rabbits. Sera were obtained from a pre-inoculation bleed, three production bleeds, and a terminal bleed from each rabbit, and were stored at −80 °C.

Purification of pAb. In preparation for purification, serum was thawed at 4 °C with gentle mixing and then clarified by centrifugation at 5,000*g* for 15 minutes. The serum was diluted threefold with binding buffer (20 mM Phosphate pH 7.00) and then loaded onto a HiTrap™ protein G column using an AKTAprime HPLC. This step was followed by washing with 10 column volumes of binding buffer, and then elution with 10 column volumes of 0.1 M Glycine pH 2.70 using a step gradient. Immediately upon elution, the samples were neutralized with 1 M Tris pH 9.00 at 0.2 volumes per volume of eluent. Fractions containing eluted pAbs were combined and dialyzed into 20 mM HEPES pH 8.0 with 300 mM NaCl. Purified pAbs (at ~38.2 µM) were stored

at −20 °C in storage buffer consisting of 10 mM HEPES pH 7.80, 150 mM NaCl, 100 µg/mL BSA, and 25% v/v glycerol.

Dot blots. A nitrocellulose membrane and two pieces of filter paper were soaked in 50 mM Tris pH 7.50 and 150 mM NaCl (1x TBS) for approximately 2 minutes. Next, the nitrocellulose membrane was placed on the two pieces of filter paper and excess liquid was allowed to evaporate. Protein samples intended for analysis were then applied to the nitrocellulose membrane. Once the sample dried, the membrane was rinsed with 1x TBS and blocked with 5% w/v dry milk in 1x TBS for one hour at 4 °C with gentle shaking. The nitrocellulose membrane was then incubated with rabbit anti-*Tma* Hfq pAb in 1x TBS supplemented with 0.1% v/v tween-20 (1x TBS-T) and 5% w/v dry milk at 4 °C overnight, with gentle shaking. The nitrocellulose membrane was thoroughly washed with 1x TBS-T, followed by incubation with 0.01% v/v goat anti-rabbit IgG that was pre-labeled with IRdye 800CW (LI-COR) in 1x TBS-T with 5% w/v dry milk for 1 hour at 4 °C. The nitrocellulose membrane was then washed three times with 1x TBS-T and dried. Dot blots were visualized using an Odyssey LI-COR imaging system. Image Studio Ver2.1 was used for data acquisition and processing; default Western settings were used with adjustments to the channel intensities based on the signal intensity to minimize saturation of the detector.

ELISA. For indirect ELISAs (Figure A2.4), protein samples in 1x PBS were aliquoted into an immunolon 2 HB 96-well plate (Thermo Lab Systems) and incubated overnight at 4 °C. Sample wells were then washed three times with 1x PBS. Wells were then blocked with 5% w/v dry milk in 1x PBS for 1 hour at 37 °C. Wells were washed with 1x PBS with 0.1% v/v tween-20 (1x PBS-T) followed by a one hour incubation with 0.005% v/v rabbit anti-*Tma* Hfq pAb in 1x PBS-T containing 5% w/v dry milk at 37 °C. Wells were washed again with 1x PBS-T, followed by 30 minute incubation with 0.005% v/v goat anti-rabbit IgG conjugated to horseradish peroxidase

(Millipore)(in 1x PBS-T containing 5% w/v dry milk) at 37 °C. Wells were washed four times with 1x PBS-T and then incubated with 100 µL o-phenylenediamine (OPD) dihydrochloride (Thermo Scientific) in 50 mM citric acid and 50 mM sodium phosphate at pH 5.00 containing 0.03% v/v hydrogen peroxide for 10 minutes at room temperature. To stop the reaction, 50 µL of 2.5 M sulfuric was added to each well and the absorbance of the product was measured at 490 nm using an iMark microplate absorbance reader (Bio-Rad).

Western blots. 4x loading buffer (final 1x working concentration: 50 mM Tris pH 6.80, 10% v/v glycerol, 12.5mM ethylenediaminetetraacetic acid (EDTA), 2% w/v SDS, and 0.02% bromophenol blue) was added to samples. Samples were incubated for 15 minutes at 90 °C prior to electrophoretic separation on a 4-20% w/v TGX gel (Bio-Rad) using 1x Tris-glycine SDS running buffer. The Odyssey protein molecular weight marker (Li-Cor) was run in a parallel lane to the protein samples. The separated samples were transferred to a nitrocellulose membrane using the Trans-Blot Turbo transfer system (Bio-Rad) with a Trans-Blot Turbo Transfer pack (Bio-Rad) at 2.5 A (up to 25 V) for 3 minutes. After transfer, the membrane was blocked with 5% w/v dry milk and then probed with rabbit anti-*Tma* Hfq pAb (in 1x TBS-T with 5% w/v dry milk). The nitrocellulose membrane was then thoroughly washed with 1x TBS-T and then incubated with secondary Ab for 1 hour at 4 °C (0.01% v/v goat anti-rabbit IgG labeled with IRdye 800CW (LI-COR) in 1x TBS-T with 5% w/v dry milk). The nitrocellulose membrane was washed three times with 1x TBS-T and two times with 1x TBS at 4 °C with gentle shaking, and then dried at room temperature. Western blots were visualized using an Odyssey LI-COR imaging system; default Western settings were used with adjustments to the channel intensities based on the signal intensity to minimize saturation of the detector.

Results & Discussion

Polyclonal antibodies were successfully purified from rabbit serum using a protein G affinity column (Figure A2.6a). Two bands corresponding to molecular weights of approximately 50 kDa and 25 kDa were found by SDS-PAGE (Figure A2.6b), consistent with the expected molecular weight of the heavy and light chains of an IgG molecule (approximately 55 kDa and 25 kDa, respectively) [7].

A dot-blot was used to initially screen the interaction specificity of the purified rabbit anti-*Tma* Hfq pAb. As negative controls, BSA, *Tma* Hfq, *Aquifex aeolicus* Hfq, *Escherichia coli* Hfq, *Methanobacterium thermautotrophicum* SmAP1, and *Pyrobaculum aerophilum* SmAP2 were all assayed for interactions with the pAbs (SmAPs are archaeal homologs of Hfq [25]). Only *Tma* Hfq was found to interact with the pAbs under the conditions tested (Figure A2.7). As a negative control, a dot plot was performed using Abs purified from pre-inoculation serum; pre-inoculation serum did not interact with any of the proteins tested (Figure A2.7b).

An indirect ELISA (Figure A2.4) was used to quantitatively determine the affinity of the pAbs for recombinant *Tma* Hfq. The interaction between *Tma* Hfq and the pAbs was spectrophotometrically monitored with goat anti-rabbit IgG conjugated to horseradish peroxidase (HRP). In the presence of the substrate OPD dihydrochloride and hydrogen peroxide, HRP produces 2,3-diaminophenazine, a yellow product of OPD oxidation with a maximum absorbance at 492 nm [26]. The apparent dissociation constant (K_d) for Hfq-pAb complex was determined to be 3 pM (N=3) by nonlinear fitting of the data to the modified Hill's equation (Figure A2.8).

Western blot conditions were optimized for use in semi-native Western blot analysis of the Hfq hexamer to Hfq dodecamer equilibrium (Chapter 3) as well as for the development of a quantitative Western blot for quantifying the amount of Hfq protein in *Thermotoga maritima*

cells (Chapter 5). A dilution series of pAbs was tested to determine the concentration of rabbit anti-*Tma* Hfq pAb that will be optimal in the sense of yielding a strong analyte signal while minimizing background signal (Figure A2.9). Diluting the pAb stock by 10,000-fold resulted in the lowest background signal, while still exhibiting binding to the target *Tma* Hfq protein. In addition to the applications mentioned above, the Western blot optimized here can be used to verify the presence of *Tma* Hfq in co-immunoprecipitation samples (Chapter 5).

References

1. Rimer J, Cohen IR, Friedman N (2014) Do all creatures possess an acquired immune system of some sort? Bioessays.

2. Deng L, Luo M, Velikovsky A, Mariuzza RA (2013) Structural insights into the evolution of the adaptive immune system. Annu Rev Biophys 42: 191-215.

3. Tonegawa S (1983) Somatic Generation of Antibody Diversity. Nature 302: 575-581.

4. Amzel LM, Poljak RJ (1979) Three-dimensional structure of immunoglobulins. Annu Rev Biochem 48: 961-997.

5. Saphire EO, Parren PW, Pantophlet R, Zwick MB, Morris GM, et al. (2001) Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design. Science 293: 1155-1159.

6. Harlow E, Lane D (1988) Antibodies : a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory. xiii, 726 p. p.

7. Lipman NS, Jackson LR, Trudel LJ, Weis-Garcia F (2005) Monoclonal versus polyclonal antibodies: Distinguishing characteristics, applications, and information resources. Ilar Journal 46: 258-268.

8. Hilschmann N, Craig LC (1965) Amino acid sequence studies with Bence-Jones proteins. Proc Natl Acad Sci U S A 53: 1403-1409.

9. Putnam FW, Easley CW (1965) Structural Studies of the Immunoglobulins. I. The Tryptic Peptides of Bence-Jones Proteins. J Biol Chem 240: 1626-1638.

10. Janeway CA, Travers P, Walport M, Schlomchik MJ (2001) The structure of a typical antibody molecule. Immunobiology. New York: Garland Science.

11. Glynn MW, McAllister AK (2006) Immunocytochemistry and quantification of protein colocalization in cultured neurons. Nat Protoc 1: 1287-1296.

12. Voller A, Bartlett A, Bidwell DE (1978) Enzyme immunoassays with special reference to ELISA techniques. J Clin Pathol 31: 507-520.

13. Angenendt P (2005) Progress in protein and antibody microarray technology. Drug Discov Today 10: 503-511.

14. Unanue ER, Allen PM (1987) The basis for the immunoregulatory role of macrophages and other accessory cells. Science 236: 551-557.

15. Marrack P, Kappler J (1987) The T cell receptor. Science 238: 1073-1079.

16. Silverthorn DU (2001) Human physiology : an integrated approach. Upper Saddle River, NJ: Prentice Hall. xxxi, 816 p. p.

17. Howard M, Nakanishi K, Paul WE (1984) B cell growth and differentiation factors. Immunol Rev 78: 185-210.

18. Ayyar BV, Arora S, Murphy C, O'Kennedy R (2012) Affinity chromatography as a tool for antibody purification. Methods 56: 116-129.

19. Rook GAW (1981) Chromogens for the Enzyme-Linked Immunosorbent-Assay (Elisa) Using Horseradish-Peroxidase. Leprosy Review 52: 281-283.

20. Khatkhatay MI, Desai M (1999) A comparison of performances of four enzymes used in elisa with special reference to beta-lactamase. Journal of Immunoassay 20: 151-183.

21. Grange RD, Thompson JP, Lambert DG (2014) Radioimmunoassay, enzyme and non-enzyme-based immunoassays. Br J Anaesth 112: 213-216.

22. Schots A, Van der Leede BJ, De Jongh E, Egberts E (1988) A method for the determination of antibody affinity using a direct ELISA. J Immunol Methods 109: 225-233.

23. Gan SD, Patel KR (2013) Enzyme immunoassay and enzyme-linked immunosorbent assay. J Invest Dermatol 133: e12.

24. Goldberg ME, Djavadi-Ohaniance L (1993) Methods for measurement of antibody/antigen affinity based on ELISA and RIA. Curr Opin Immunol 5: 278-281.

25. Mura C, Randolph PS, Patterson J, Cozen AE (2013) Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function. Rna Biology 10: 636-651.

26. Hempen C, van Leeuwen SM, Luftmann H, Karst U (2005) Liquid chromatographic/mass spectrometric investigation on the reaction products in the peroxidase-catalyzed oxidation of o-phenylenediamine by hydrogen peroxide. Anal Bioanal Chem 382: 234-238.
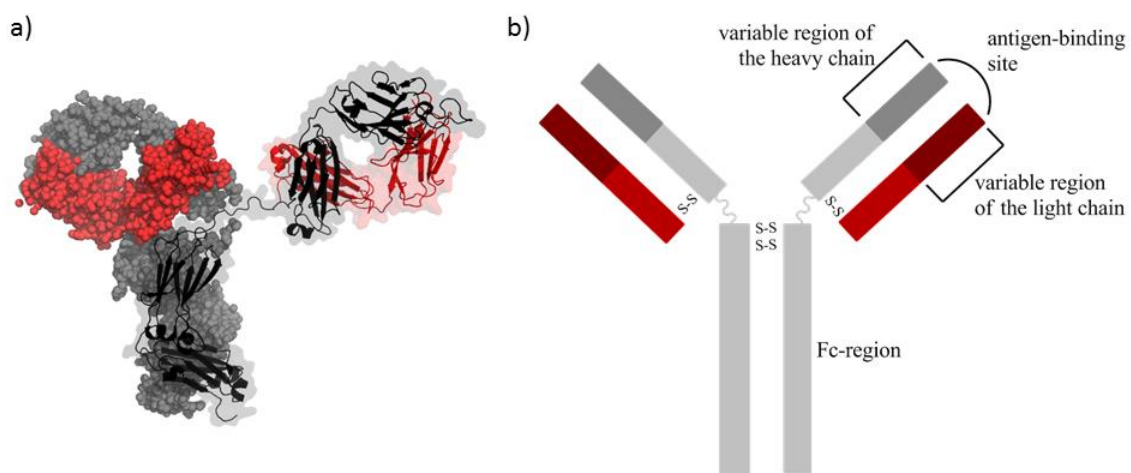
Figure A2.1 Antibodies are Y-shaped proteins composed of two identical heavy chains and two light chains. (a) The crystal structure of human IgG (1HZH) against the human immunodeficiency virus (HIV) is shown, modeled in PyMOl [5]. The light chains and heavy chains are highlighted in **red** and **grey**, respectively. The image was generated using a cartoon and sphere representation of each chain. (b) A schematic representation of the basic structure of an Ab with the light chains and heavy chains in **red** and **grey**, respectively. The variable regions of each chain are indicated as well as the fragment crystallizable region (Fc-region) of the heavy chain.
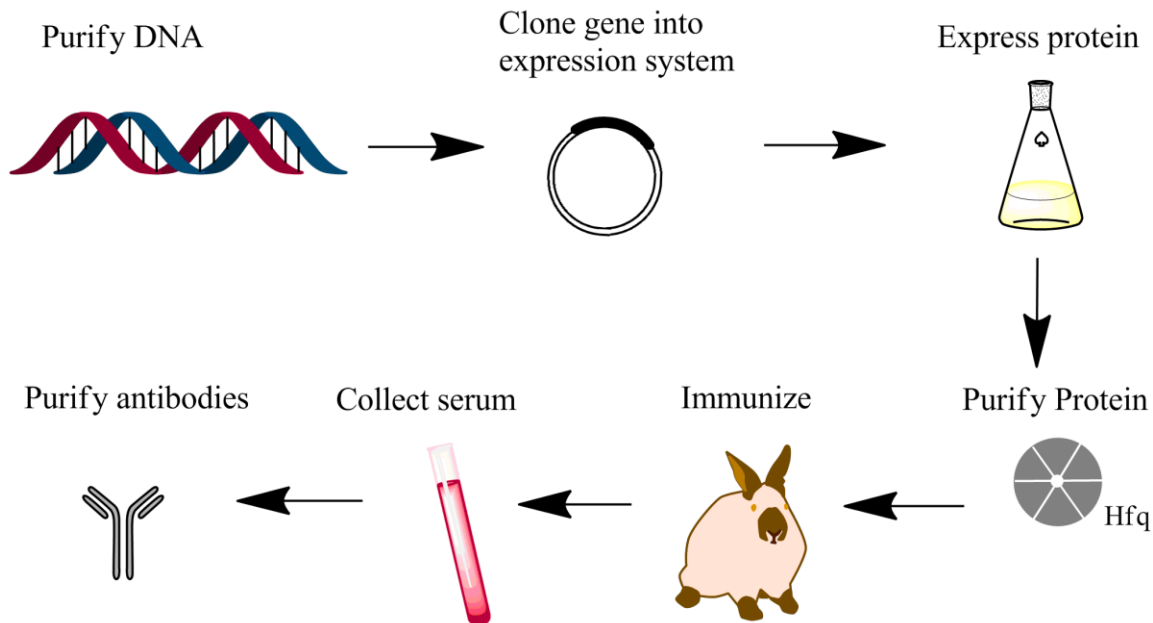
Figure A2.2 Basic workflow for generating polyclonal antibodies to a recombinant protein. The first step is to obtain and purify genomic DNA encoding the protein of interest. The gene is then cloned into an expression system (*e.g.* a plasmid) and transformed into a host system (*e.g. E. coli*). The recombinantly expressed protein is purified and dialyzed into a suitable buffer for immunization. The animal (often rabbits or mice) is given a primary immunization, followed by multiple boosters over the course of several weeks in order to generate a sufficient immune response. Finally, serum is collected and Abs are purified by a method such as affinity chromatography.
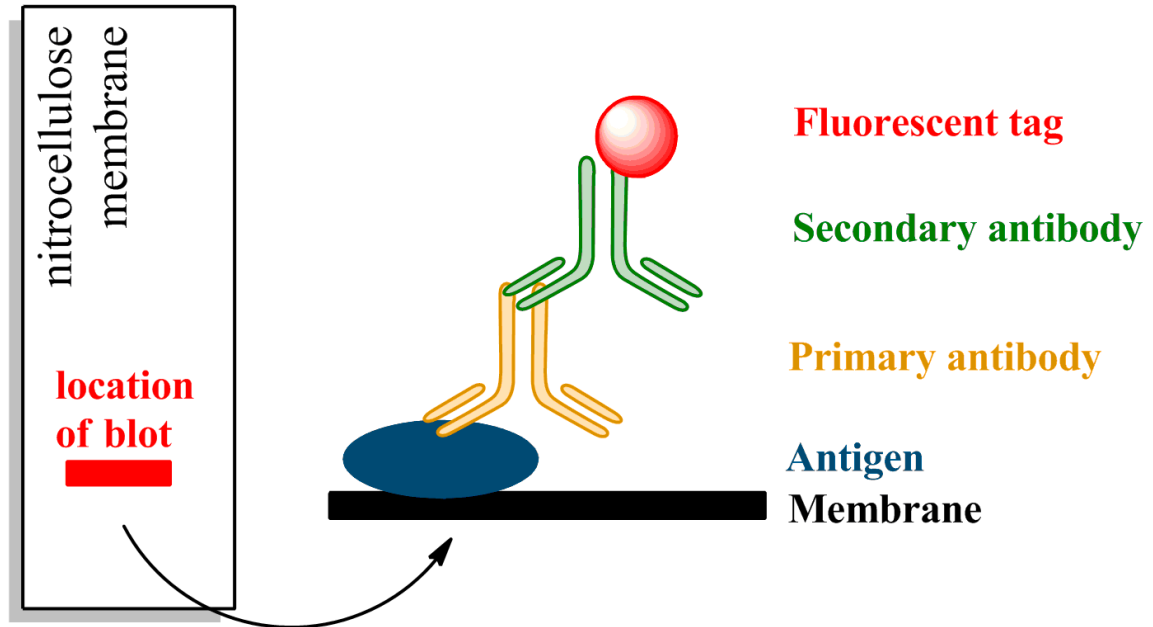
Figure A2.3 Schematic of a basic immunoblot assay. The antigen is transferred to a suitable membrane (e.g. nitrocellulose for protein antigens); the membrane is then blocked to inhibit nonspecific interactions with the Abs used in the immunoblot. A primary Ab that binds to an epitope of the antigen is then added. This is followed by a labeled (*e.g.* fluorescently tagged) secondary Ab that binds to a known epitope of the primary Ab. Alternatively, the primary Ab may be directly labeled for read-out.
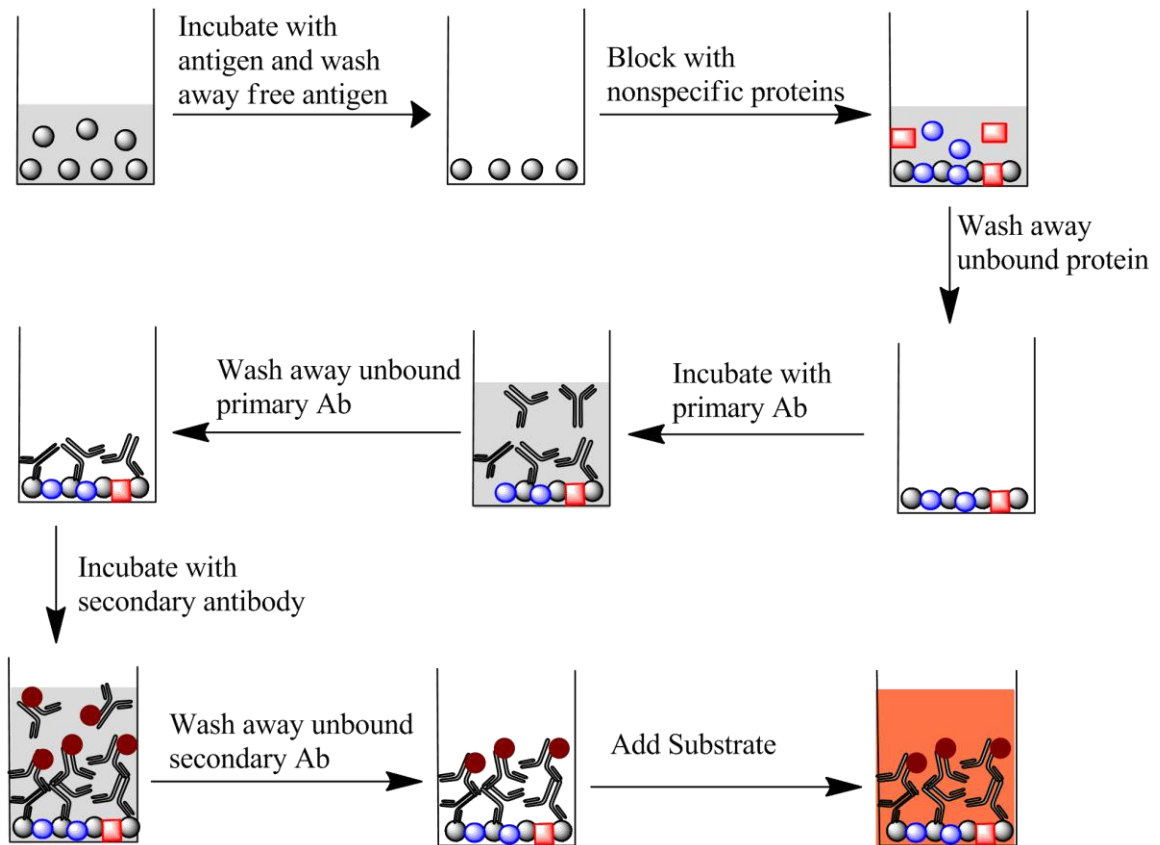
Figure A2.4 Workflow for an indirect ELISA using an enzyme-conjugated secondary antibody.
Wells are incubated with antigen (grey spheres) which associates with the surface of the well.
The well is then blocked with nonspecific proteins (red squares and blue spheres), such as dry
milk, to prevent nonspecific interactions between the surface of the well and the Abs used in
the assay. After washing, the primary Ab is added and binds to an epitope in the antigen.
Unbound primary Ab is washed off prior to incubation with the enzyme-conjugated secondary
Ab (red sphere attached to Ab); the secondary Ab recognizes an epitope in the primary Ab.
Unbound secondary Ab is then washed prior to the addition of substrate. Substrate is added,
the enzymatic reaction is then stopped at a predetermined time, and the colored product is
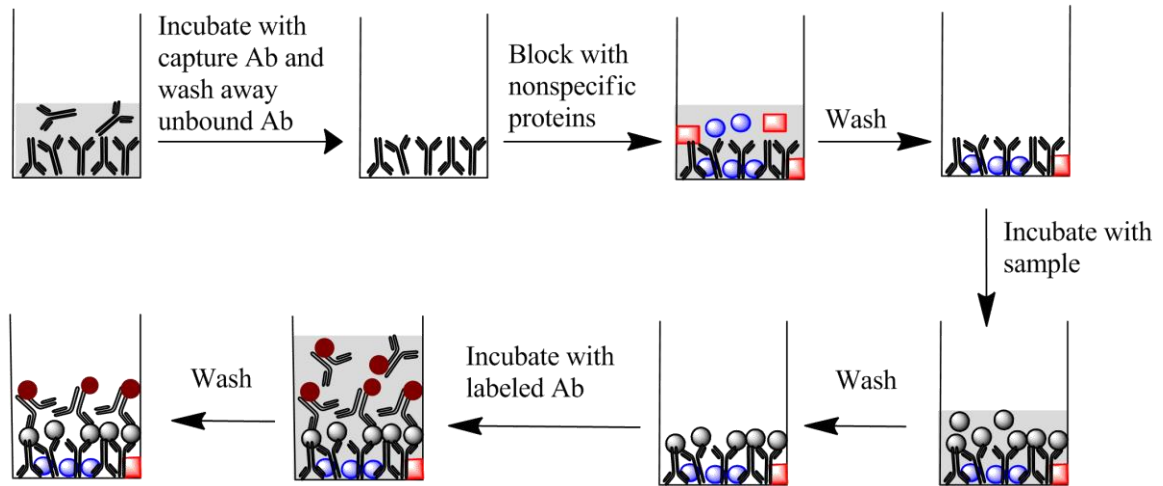measured using a spectrophotometer.

Figure A2.5 Workflow for a sandwich ELISA. Wells are incubated with a capture Ab that associates with the well's surface. The well is then blocked with nonspecific proteins (red squares and blue spheres), such as dry milk, to prevent nonspecific interactions with the surface of the well. The sample is then added and antigen (grey spheres) associates with the capture Ab; unbound components of the sample are washed away. A second labeled (*e.g.* fluorescent tag) Ab that recognizes an epitope in the antigen is added for detection. Note that both the capture Ab and the labeled detection Ab are primary Abs in the sense that they directly interact with the antigen.
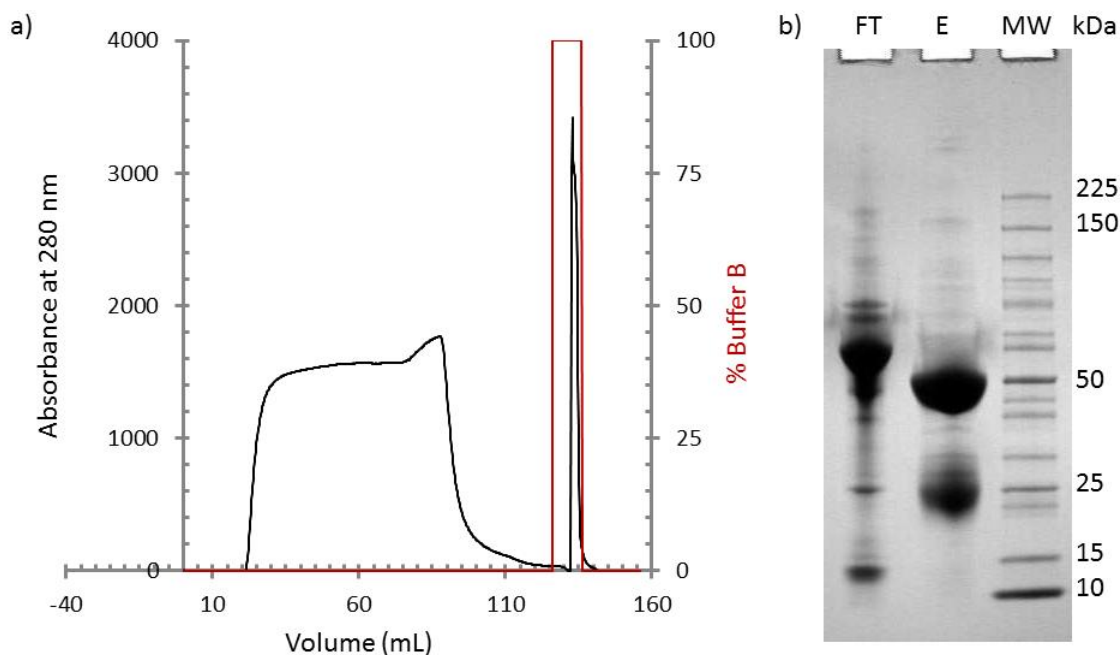
Figure A2.6 Purification of anti-*Tma* Hfq polyclonal antibodies. (a) Representative chromatogram for the purification of IgG from rabbit serum using HiTrap™ protein G columns. The absorbance trace (black) shows a single peak eluting after the addition of buffer B (0.1 M glycine pH 2.7). (b) The flow-through (FT) and elution peak (E) from the protein G column were run on a 4-20% w/v TGX gel along with a molecular weight marker (MW). As expected, two predominant bands occur in the elution peak, corresponding to the expected molecular weight of a light chain (~25 kDa) and heavy chain (~55 kDa) of an IgG molecule.
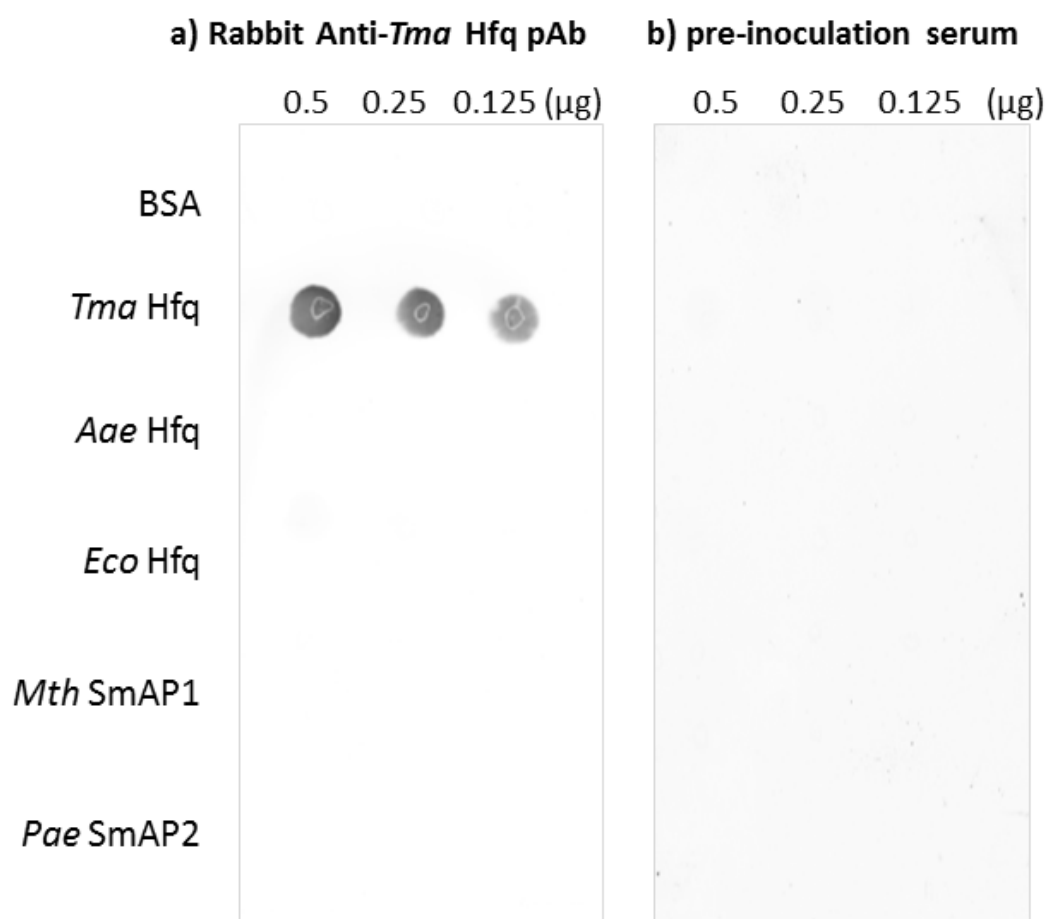
Figure A2.7 Representative dot blots evaluating the specificity of pAbs generated against *Tma* Hfq. a) Recombinantly expressed and purified *T. maritima* Hfq (*Tma* Hfq), *A. aeolicus* Hfq (*Aae* Hfq), *E. coli* Hfq (*Eco* Hfq), *M. thermautotrophicum* SmAP1 (*Mth* Hfq), and *P. aerophilum* SmAP2 (*Pae* Hfq) along with the generic protein BSA were screened for interactions with rabbit anti-*Tma* Hfq pAb using goat anti-rabbit IgG labeled with IRdye 800CW for detection. The pAbs were found to interact specifically with *Tma* Hfq. b) Dot plot using Abs purified from pre-inoculation serum. Abs from pre-inoculation serum did not interact with any of the proteins tested.
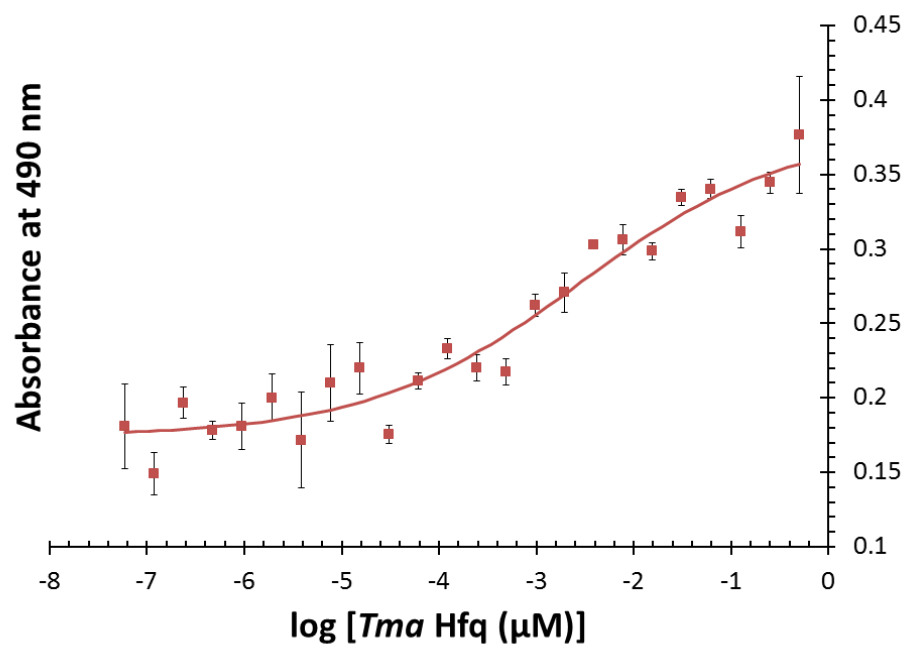
Figure A2.8 Binding curve for *Tma* Hfq with rabbit anti-*Tma* Hfq pAb. An indirect ELISA was used to evaluate the binding affinity of the pAbs for *Tma* Hfq. Data from three independent assays were averaged and then fit using the modified Hill's equation with an R^2 equal to 0.94. The computed K_d is 3 pM.
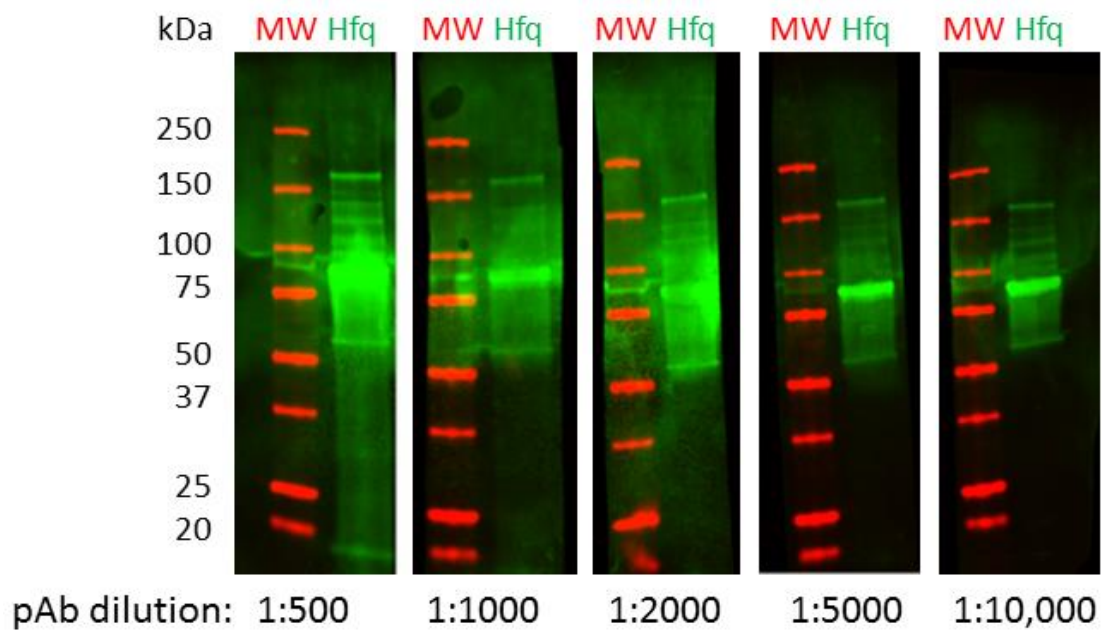
Figure A2.9 Optimization of the Western blot protocol. Molecular marker (MW; red bands) and *Tma* Hfq (green bands) were run on a 4-20% w/v TGX gel and transferred to a nitrocellulose membrane. Multiple dilutions of rabbit anti-*Tma* Hfq pAb (specified below the Western blot images) were evaluated to determine the optimal concentration for detection of *Tma* Hfq. A 10,000-fold dilution of the Ab stock (at ~38.2 μM) had the lowest background of the concentrations tested.

Curriculum vitae

Jennifer Patterson West

Department of Chemistry

University of Virginia

PO Box 400319

Charlottesville, VA 22904

15 Fairwood Place

Palmyra, VA 22963

(570) 972-3612

jlp7fp@virginia.edu

Education

Ph.D. Chemistry

Department of Chemistry, University of Virginia, Charlottesville, VA, U.S., 2009-2014

B.S. Chemistry, Biochemistry concentration

Department of Chemistry, Shippensburg University, Shippensburg, PA, U.S., 2005-2009

B.S. Biology, Pre-health concentration

Department of Biology, Shippensburg University, Shippensburg, PA, U.S., 2005-2009

Research Experience

University of Virginia, Department of Chemistry, Charlottesville, VA (2010-present)

Graduate Researcher, Advisor: Cameron Mura, Ph.D., **Biochemical & Biophysical research on RNA-associated Proteins.** Characterized the oligomerization and ligand-binding properties of the RNA-associated 'Hfq' protein from extremophilic bacteria; this protein family is involved in bacterial quorum sensing and numerous other metabolic pathways. I have gained experience with PCR, gene cloning, protein and RNA purification, assay development, electrophoretic mobility gel-shift assays, mass spectrometry (MALDI-TOF), fluorescence polarization binding assays, size-exclusion chromatography, static light scattering, spot blots, Western blot analysis, enzyme-linked immunosorbent assay, and protein crystallization and crystallography.

University of Virginia, Department of Chemistry, Charlottesville, VA (Summer 2009)

Graduate Researcher, Advisor: Timothy Macdonald, Ph.D., **Organic Synthesis.** Experience with organic synthesis, column purification, and NMR. Collected and analyzed data, presented at group meeting.

Shippensburg University, Departments of Chemistry and Biology, Shippensburg, VA (2008-2009)

Undergraduate Research, Advisors: Robin McCann, Ph.D. and Pablo Delis, Ph.D., **ELISA development.** Worked on the development of an enzyme-linked immunosorbent assay (ELISA) for the detection of tetrodotoxin that was extracted from *Notophthalmus viridescens.*

United States Army Medical Research Institute of Infectious Diseases (USAMRIID), Frederick, MD (Summer 2008)

Summer Intern – Junior Researcher, Mentor: David Rozak, Ph.D., **Quorum Sensing Research.** Analyzed the ability of various algal species to inhibit quorum sensing using CV026 as a model bacterium and extracted the inhibitory factors from applicable species of algae. Helped develop a high-throughput colorimetric screen for detecting potential quorum sensing inhibitors.

Collected and analyzed data, participated in safety training and animal training courses, presented findings at group meetings.

<u>**Awards, Honors, and Funding**</u>

- Received a Robert J. Huskey Travel Fellowship to attend the South East Regional Meeting of the American Chemical Society in Atlanta, Georgia (November 2013)

- Received RNA Society Travel Award to attend the North Carolina RNA Society's symposium at the University of North Carolina, Chapel Hill, NC (October 2013)

- Selected to attend and present research poster at the Eighth Annual NIH National Graduate Student Research Conference (October 2013); canceled due to government shutdown.

- Selected from a competitive pool of candidates to attend the 2013 post-doctoral training program at the University of Alabama at Birmingham (May 2013).

- Robert J. Huskey Graduate School Research Exhibition, University of Virginia (March 2012) Oral presentation in Physical Science and Mathematics, *Second place award*

- Virginia Academy of Sciences Annual Meeting, University of Richmond, VA (May 2011) *Best Graduate Poster Award* in the *Structural Biology, Biochemistry & Biophysics* division

- Research grant from Shippensburg University Undergraduate Research Program (2008) *Along with Dr. McCann (advisor) and Dr. Delis (advisor); Assessment of skin toxicity, tetrodotoxin levels, in a population of spotted newt, Notophthalmus viridescens, in South Central Pennsylvania.*

- Janice Tarner Crawford Memorial Scholarship (2007)
- Academic scholarship from Coca-Cola Corporation (2005)

<u>**Teaching Experience**</u>

Teaching Assistant, Chem4411/4421, Upper-division Undergraduate Biochemistry Laboratory University of Virginia, Department of Chemistry, Charlottesville, VA (5 semesters; 2009-2011)

<u>**Publications**</u>

J. Patterson, C. Mura (2013) "Rapid Colorimetric Assays to Qualitatively Distinguish RNA and DNA in Biomolecular Samples" *J. Vis. Exp.* 72: 1-10 (doi: 10.3791/50225**)**

C. Mura, P.S. Randolph, **J. Patterson**, A.E. Cozen (2013) "Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function" *RNA Biology.* **10**: 636 - 651 (PMID: 23579284)

J. Patterson, S. Cao, D. Fox, L. Columbus, J. Clardy, C. Mura (201X) "Discovery and Identification of a Novel Class of Hfq-binding nanoRNAs in *Escherichia coli*" [In preparation]

P. Randolph, **J. Patterson**, C. Mura (201X) "Preliminary crystallographic analysis of the RNA-associated Hfq protein from *Thermotoga maritima*" *Acta F* [In preparation]

J. Patterson, C. Mura (201X) "Characterization of *Thermotoga maritima* Hfq self-assembly: A regulatory mechanism involving two distinct oligomeric states?" [In preparation]

Conference, Meeting, & Invited Presentations

K. Stanek, **J. Patterson**, P. Randolph and C. Mura**.** Crystal Structure and Functional Characterization of an Hfq Homolog from *Aquifex aeolicus*. Huskey Research Exhibition, University of Virginia, Charlottesville, VA, March 2014 (poster)

J. Patterson, S. Cao, P. Randolph, L. Columbus, J. Clardy, and C. Mura. Identification and Characterization of Hfq-binding nanoRNAs. Southeastern Regional Meeting of the American Chemical Society, Atlanta, GA, November 2013 (poster)

J. Patterson, S. Cao, P. Randolph, L. Columbus, J. Clardy, and C. Mura. Identification and Characterization of Hfq-binding nanoRNAs. Symposium on RNA biology X: RNA Tool & Target, Chapel Hill, NC, October 2013 (poster)

J. Patterson, S. Cao, D. Fox, L. Columbus, J. Clardy, C. Mura. Discovery and identification of Hfq-binding nanoRNAs in *Escherichia coli*. Annual Meeting of the RNA Society, Davos, Switzerland, June 2013 (poster)

P.S. Randolph, **J. Patterson**, and C. Mura. Structural analysis of an Hfq from *Thermatoga maritima*. Mid-Atlantic Macromolecular Crystallography Meeting, Duke University, Durham, NC, June 2013 (poster)

K. Stanek, **J. Patterson**, and C. Mura**.** Towards the structural and functional characterization of the Hfq homolog from *Aquifex aeolicus*. Mid-Atlantic Macromolecular Crystallography Meeting, Duke University, Durham, NC, June 2013 (poster)

K. Stanek, **J. Patterson**, and C. Mura**.** Towards the structural and functional characterization of the Hfq homolog from *Aquifex aeolicus*. Annual Meeting of the Virginia Academy of Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA June 2013 (poster)

J. Patterson, S. Cao, P. Randolph, J. Clardy, and C. Mura. Biophysical & Biochemical Characterization of an Hfq Ortholog from *Thermotoga maritima*. University of Alabama at Birmingham's Postdoc Prep Conference, Birmingham, AL, 2013 (poster)

J. Patterson, C. Mura. Biochemical & Biophysical Studies of Hfq Homologs from Deep-branching Bacterial Species. Shippensburg University, Shippensburg, VA, 2013 (invited talk)

K. Stanek, **J. Patterson**, and C. Mura**.** Towards the Structural and Functional Characterization of the *Aquifex aeolicus* Hfq Homolog. Southeastern Regional Meeting of the American Chemical Society, Raleigh, NC, November 2012 (poster)

J. Patterson, C. Mura. Oligomerization of the RNA-associated *Thermotoga maritima* Hfq Protein: A complex regulatory mechanism involving two distinct assemblies. Southeastern Regional Meeting of the American Chemical Society, Raleigh, NC, November 2012 (talk)

J. Patterson, C. Mura. Characterization of the oligomeric states of Hfq from *Thermotoga maritima* alludes to a complex regulatory mechanism involving two distinct oligomeric states. Mid-Atlantic Macromolecular Crystallography Meeting, University of Virginia, Charlottesville, VA, June 2012 (poster)

J. Patterson, C. Mura. Characterization of the oligomeric states of Hfq from *Thermotoga maritima* alludes to a complex regulatory mechanism involving two distinct oligomeric states. Annual Meeting of the Virginia Academy of Sciences, Norfolk University, Norfolk, VA, May 2012 (talk)

J. Patterson, C. Mura. Hfq Oligomerization: Does the RNA-based regulatory mechanism involve two distinct oligomeric states? Huskey Research Exhibition, University of Virginia, Charlottesville, VA, March 2012 (talk)

J. Patterson, C. Mura. Identification of binding partners for Hfq from *Thermotoga maritima* using traditional colorimetric assays and enzymatic techniques. Southeastern Regional Meeting of the American Chemical Society, Richmond, VA, October 2011 (talk)

J.L. Patterson, T.K Kabzinski, C. Mura. Evolution of Nucleic acid–Binding by the Sm Superfamily: Hfq Homologs from Ancient Bacterial Species. Mid-Atlantic Macromolecular Crystallography Meeting, National Cancer Institute, Frederick, MD, June 2011 (poster)

J.L. Patterson, T.K Kabzinski, C. Mura. Evolution of Nucleic acid–Binding by the Sm Superfamily: Hfq Homologs from Ancient Bacterial Species. Annual Meeting of the Virginia Academy of Sciences, University of Richmond, Richmond, VA, May 2011 (poster)

<u>Advance Training</u>

Bio-Trac1: Biotechnology Techniques, Foundation for Advanced Education in the Sciences (FAES), National Institutes of Health (NIH), Bethesda, MD, September 2012

Robert Kelly's Lab: Spent 2 days in the Kelly Lab learning how to culture *Thermotoga maritima* and harvest cells for downstream experiments, North Carolina State University, Raleigh, NC, December 2011.

Todd Lowe's Lab: Spent 2 weeks in the Lowe Lab learning how to perform RNA extractions (phenol-chloroform), endonuclease titrations, western blots, and crosslinking immunoprecipitation (CLIP), University of California, Santa Cruz, CA, August 2011.

Service & Community Outreach

GWIS Fellowship Reviewer: Evaluated Fellowship application for the National Fellowships committee for Sigma Delta Epsilon Chapter of Graduate Women in Science, March 2014

Chemistry Camp: Prepared lessons and instructed groups of children (K-2) on spy themed chemistry experiments, University of Virginia, Charlottesville, VA March 2013

Kid-Vention: Assisted Children with Strawberry DNA extraction, Virginia Discovery Museum's free family science festival, Charlottesville, VA February 2013

Professional Memberships

Biochemical Society (U.K.)
RNA Society
Graduate Women in Science (omega chapter)