11-12-2012 12:00 AM

The Regulation of Prokaryotic Transposable Elements by the RNA Binding Protein Hfq

Brian J. Munshaw
The University of Western Ontario

Supervisor
Dr. David Haniford
The University of Western Ontario

Graduate Program in Biochemistry
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

The Regulation of Prokaryotic Transposable Elements by the RNA Binding Protein Hfq

(Spine title: Hfq and Transposable Elements)

(Thesis format: Monograph)

by

Brian Munshaw

Graduate Program in Biochemistry

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES

CERTIFICATE OF EXAMINATION

Supervisor	Examiners
_____	_____
Dr. David Haniford	Dr. Christopher Brandl
	_____
Supervisory Committee	Dr. Ze-Chun Yuan
_____	_____
Dr. Christopher Brandl	Dr. Stan Dunn
_____	
Dr. David Edgel	

The thesis by

Brian James Munshaw

entitled:

THE REGULATION OF PROKARYOTIC TRANSPOSABLE ELEMENTS BY THE RNA BINDING PROTEIN HFQ

is accepted in partial fulfilment of the
requirements for the degree of
Master of Science

Date_____ _____
Chair of the Thesis Examination Board

Abstract

This thesis investigates the role that the RNA binding protein Hfq has in regulating the mobility of IS Elements and the expression of their transposase genes. Hfq is a known regulator of the IS Element IS*10* and is a major component of sRNA pathways in *E.coli*. In this work, I demonstrate that through the use of a bioinformatics approach to identify a list of additional IS Elements from the IS Finder database, one can successfully identify IS Elements containing Hfq binding sites which may implicate Hfq regulation. One of the identified elements, IS*1413*, was characterized by transposition assays and transposase expression assays. These experiments demonstrated that the activity level of IS*1413* decreases in the presence of Hfq. Another identified IS Element, IS*10*, had its 3' end of its transposase mRNA transcript probed for the relevance of a putative Hfq binding site through mutational analysis, and through identifying that there are no post-transcriptional modifications at the 3' end of the transcript. Although Hfq bound IS*10* mRNA *in vitro*, genetic mutations to the putative Hfq binding site resulted in no difference in transposition activity. A growth condition was identified that increased IS*10* transposition frequencies in wild type cells, but not in cells containing a *rpoS* gene disruption. Overall, these studies have identified another IS Element that changes activity in the presence of Hfq and have linked slow unaerated growth conditions to an increase in IS*10* transposition frequencies in *E.coli*.

Key words: IS*10*/Tn*10*, IS*1413,* Hfq, Stress Conditions, Transposition, *E.coli*, *Burkholderia phenoltripix,* IS Finder

Acknowledgements

I would like to thank my supervisor David Haniford. Thank you for your guidance, feedback, and great conversations about science (and fishing) at lab coffee meetings. I would also like to extend my thanks to my colleagues (past and present) Joe Ross, Crystal McLellan, Ryan Trussler, Shahan Hossain, Mike Ellis and Morgan Black for the support and great conversations. I like to thank my graduate committee Dr. Edgell, and Dr. Brandl for their support in terms of motivation and research considerations. I would also like to thank Barb Green and her administrative team for their help in answering questions regarding department policies and procedures. A special thank you goes out to my wife Brie Munshaw who has supported me throughout this thesis and has offered me nothing but encouragement and motivation to continue this journey. I would also like to thank my family and friends for all their love and support.

Thank you to the University of Western Ontario and the Province of Ontario for providing me with this valuable opportunity as well as funding to pursue it.

Table of Contents

List of Tables

List of Figures

Abbreviation	Meaning
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DAM	DNA Adenine Methyl Transferase
DksA^{-}	NK5830F' (*dska*::Ωtet)
HBS	Hfq binding site
Hfq^{-}	NK5830F' (*hfq*-1::Ω*cat*)
I.E	Inside End
IHF	Integrative Host Factor
IR	inverted repeat
IS	Insertion Sequences
IS10L	Left end of the IS10 sequence (codes a non-functional transposase)
IS10R	Right end of the IS10 sequence
Mci	multi-copy inhibition
miRNAs	micro RNA
M-MuLV-RT	Moloney Murine Leukemia Virus Reverse Transcriptase
Normal	Mating out conditions grown at 37°C with aeration
O.E	Outside End
PAP	Polyadenylate Polymerase I
PNP	Polynucleotide Phosphorylase
PolyA	Polyadenylation
R	Purine
R Gene	resistance gene
RBS	ribosome binding site
RNA-IN	IS*10* Transposase mRNA
RNA-OUT	IS*10* antisense-RNA
RpoS^{-}	NK5830F' nlpD-rpoS::CmR
siRNAs	small interfering RNA
sRNA	small RNA
Starved	Mating out conditions grown at room temperature without aeration
UTR	untranslated region

Chapter 1 - Introduction

1.1 Lateral gene transfer in bacterial populations:

One of the driving forces of bacterial evolution is thought to be the ability to exchange DNA between populations via lateral gene transfer (reviewed in [1]). Promiscuous DNA transfer can result in foreign genes being transferred from species to species, potentially conferring new survival traits in the recipient (including pathogenicity or antibiotic resistance) [1]. By analyzing the genome of *E.coli* MG1655 for DNA with different base compositions and different codon bias, *Lawrence et al.* inferred that 17.6% of open reading frames of the genome were acquired from at least 234 lateral gene transfer events [2]. Mechanisms of lateral gene transfer include transduction, conjugation and transformation (See **Figure 1.1**). Transduction involves a bacteriophage incorporating some host DNA into its capsid and upon infection, incorporating it into the recipient's genome. Transformation entails the uptake of DNA from the environment. Conjugation refers to bacteria transferring self-mobilizing plasmids through sex pili from one cell to another. Conjugation can occur between bacteria, bacteria and plants [3], or bacteria and yeast [4], increasing the diversity of transmissible genes. Transposons are often carriers of antibiotic resistance genes; their ability to transpose onto phage or self-transmissible plasmids make them a major component in lateral gene transfer [5].

1.2 Transposons:

Many bacteria carry DNA transposons which are mobile genetic elements that can move to various loci throughout the genome. Transposition events can have consequences on the host such as gene deletions, inversions [6] and gene activation [7].

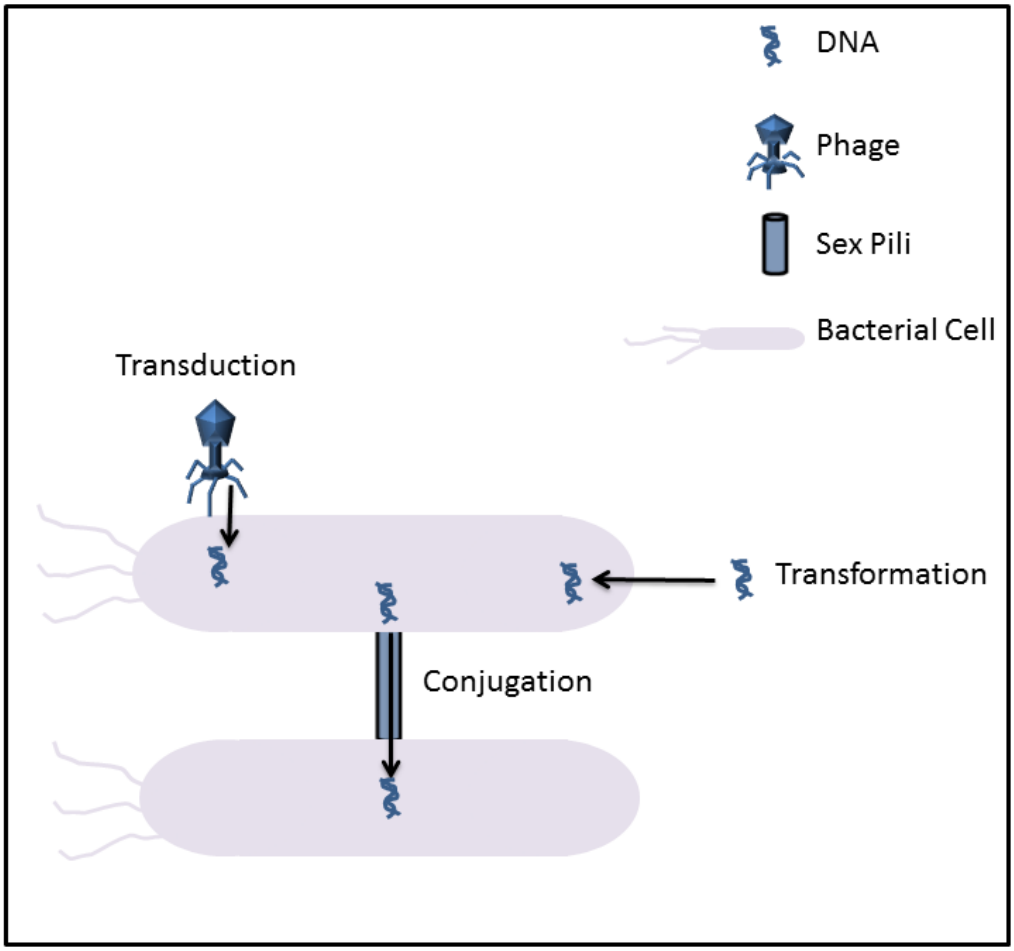


Figure 1.1 Mechanisms of lateral gene transfer. Transduction involves a bacteriophage injecting its genetic material into a recipient bacterial cell. Transformation involves the cellular uptake of DNA from the environment. Conjugation involves the active transfer of DNA from one cell to another through a sex pilus.

There are three major classes of mobile genetic elements found in bacteria: Insertion sequences (IS Elements), composite transposons, and conjugative transposons (see **Figure 1.2**). IS Elements (top of **Figure 1.2**) are comprised of genes necessary to facilitate their own mobility, such as a transposase. Many IS Elements are flanked by inverted repeats which often act as the boundaries of the IS Element and the transposase binding region. When two IS Elements flank a common genomic region they can form a composite transposon (bottom of **Figure 1.2**). The two IS Elements of the composite transposon may or may not be completely identical; however, as long as one of the IS Elements encodes a functional transposase, the entire composite transposon can move as a single unit. Conjugative transposons are larger mobile elements than IS Elements and composite transposons (middle of **Figure 1.2**). Conjugative transposons, such as Tn914, encode genes required for cell-to-cell transfer, an integrase for integration into the chromosome, an excisionase which excises the element from the host's chromosome and often antibiotic resistance genes such as tetracycline (reviewed in [8]). The abundance of IS Elements in a genome is exemplified with the 37 IS Elements present in *E.coli* MG1655 [2]. Mobile genetic elements are also very prevalent in *Burkholderia cephacia,* where the study of catabolite gene activation by IS Elements has been extensively studied [9]. Overall, the abundance of mobile genetic elements in bacterial genomes may play a role in changing the overall architecture of the host over time and may confer survival advantage to the host in times of environmental stresses. The tetracycline carrying conjugative transposon CTnDot is an example of a mobile genetic element playing a role in stress

adaptation. Upon exposure to tetracycline, the rates of CTnDOT excision from the host chromosome increases [10].

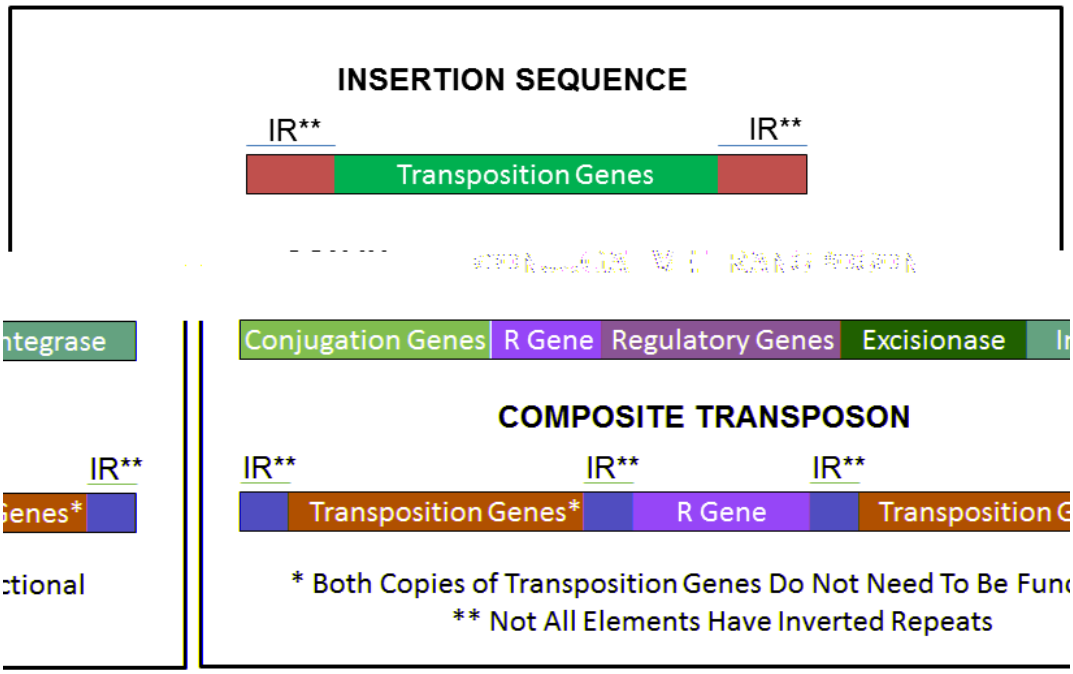


Figure 1.2 The differences between insertion sequences, transposons and composite transposons. Top: Insertion Sequences (IS Elements) are DNA sequences that only encode genes involved in their mobilization (such as a transposase). Many (but not all) IS Elements are flanked by inverted repeats which act as the IS Elements' boundary and often serve as the binding regions for the transposase. **Middle:** Conjugative transposons are much larger than IS Elements and contain gene clusters of similar function. Conjugation genes (light blue) are involved in the transfer of the excised circular form of the transposon to a recipient cell. Antibiotic resistance genes (R Gene) can also be carried within the transposon (orange). Regulatory genes (brown) are often general transcription factors involved in regulating the expression of the element's own genes. Excisionase (dark blue) is involved in excising the element from the host's chromosome to allow the element to circularize into its transmissible form. The integrase genes are involved in integrating the element into the bacterial chromosome. **Bottom**: A composite transposon is made when two IS Elements mobilize to a genomic region within close proximity to one another. The region that is trapped between the two IS Elements becomes part of the mobile element when the transposase binds the outer boundaries of the composite transposon. Often the genomic region between the two IS Elements is often an R Gene, or another survival enhancing trait.

1.3 Regulation of mobile elements:

Mobile genetic elements depend on the survival of the host to maintain their existence. With host gene rearrangements and deletions being possible consequences of transposition, mobile genetic elements are under tight regulation by both the host and the mobile genetic elements themselves to maintain the viability of both. Ultimately, the frequencies of transposition are dependent on the amount of transposase (the enzyme responsible for element mobilization) present in the cell [11]. Down regulation of the transposase at the expression level and transposon-binding level makes transposition events a rare occurrence. For example, IS*10* transposes at a rate of 10^{-4} events per cell generation [12], and Tn10 has an even lower transposition rate of 10^{-7} events per cell generation [13].

1.3.1 Regulation of transposase expression:

Transposons and IS Elements are often poorly expressed from moderate endogenous promoters coupled with suboptimal ribosome binding sites [14], resulting in small amounts of transposase available in the cell to perform the transposition reaction. Transcription levels of transposases can also be affected by host factors. For example, DNA Adenine Methyltransferase (DAM) in *E.coli* down regulates IS*10* transposase expression at the transcriptional level through the methylation of an adenine within the -10 region of the promoter, resulting in reduced promoter activity [15]. Anti-sense RNAs can also regulate transposon

expression, such as in IS*10*/Tn10, where an anti-sense RNA (RNA-OUT) base pairs with the transposase mRNA (RNA-IN) resulting in the occlusion of the RBS.

1.3.2 Post translational regulation of transposons:

Host factors can also play a role on mobilization of transposons at a post translational level. Host factors play an important role in the regulation of transposition (reviewed in [16]). Tn10 is both positively and negatively regulated by host factors. DAM, in addition to regulating IS*10* transcription, also methylates an adenine found within the inside end of IS*10* which reduces transposase binding [15]. Another host factor, Integration Host Factor, positively regulates Tn10 transposition through binding the transposon ends promoting the formation of the paired ends complex (a step in IS*10*/Tn10 transposition) [17]. The host factor Hfq has also been implicated in IS*10* regulation; *E. coli* strains with a *hfq* gene disruptions (Hfq^-) show a 50-86 fold increase in IS*10* transposition when IS*10* is on a multi-copy plasmid, compared to cells with functional Hfq [18].

1.4 Hfq and sRNAs:

Hfq is an RNA binding protein that forms a homo-hexameric ring from 11.2 kDa monomers [19]**.** It is a member of the SM/LSM family of proteins, containing the SM1 motif found in Eukaryotic SM and LSM proteins [20] . In Eukaryotes, SM and LSM proteins are involved in DNA and RNA processing events such as pre-mRNA splicing and telomere synthesis [21]. In bacteria, Hfq is a key component in small RNA (sRNA) regulation in bacteria. One of the more defined roles of Hfq is in promoting pairing of trans-encoded sRNAs with their targets. Recently, Hfq

has been shown to be involved in accelerating the rate of pairing a cis-encoded sRNA (IS*10*'s RNA-OUT) to RNA-IN *in vitro*[18]. Both trans-encoded and cis-encoded sRNAs are analogous to Eukaryotic micro RNAs (miRNAs) and small interfering RNAs (siRNAs) in that they can alter the expression of their mRNA targets through RNA destabilization or translation inhibition [22].

Hfq can affect RNA molecules in at least 6 different ways (see **Figure 1.3**). Firstly, Hfq can act in conjunction with an sRNA to block the ribosome binding site (RBS) on its target mRNA which results in translation inhibition**.** In this case, it is thought that Hfq acts to either promote pairing of the two RNA species through binding and secondary structure alteration, or by physically increasing the local concentrations of the two RNA species [23] (**Figure 1.3A**)**.** RBS blocking by an sRNA and Hfq is found in the regulation of *fhl* translation (an mRNA coding for a transcriptional activator involved in the formate pathway in Enterobacteria) by the sRNA OxyS; in this system, the OxyS binding site overlaps the RBS. Upon binding to the *fhl* mRNA, OxyS directly competes with the 30S ribosomal subunit for binding in an Hfq dependent manner [24]. Secondly, Hfq and sRNAs can actually increase translation initiation as well. An example of this is found in the study of the DsrA:RpoS interaction. RpoS mRNA contains a secondary structure in its 5' UTR that results in the occlusion of the RBS. DsrA relieves this secondary structure by pairing RpoS mRNA in an Hfq dependent manner making the ribosome binding site accessible and allowing for ribosome binding [25] **(Figure 1.3C).** Thirdly, Hfq can also directly target the RBS of mRNA independent of an

sRNA, which in turn has a destabilizing effect on the half-life of the mRNA and reduced translation (**Figure 1.3B**).

RNA stability which is affected by the frequency of translation initiation on an mRNA by the 16S ribosomal subunit, and the initiator tRNA [26]**.** A reduced frequency of translation initiation results in reduced mRNA levels in the cell [26]**.** This is exemplified in the mRNA *ompA*, a message for an outer membrane protein, where *Vytvytska et al* show Hfq can bind to *ompA* mRNA, preventing the initiator complex from forming and ultimately making the mRNA vulnerable to RNase E mediated degradation [27]. Hfq is not just a promoter of RNA degradation; in a fourth mechanism Hfq can also protect sRNAs from RNases by blocking cleavage sites (**Figure 1.3 F**). For example, the sRNAs DsrA and RhyB are protected from RNase E cleavage when Hfq binds and protects the RNase E cleavage site on the sRNAs [28]. In the fifth mechanism of how Hfq can affect RNA, Hfq facilitated pairings of sRNAs can result in the degradation of both the sRNA and target mRNA (**Figure 1.3D**). A study by *Massé et al* demonstrates that in the Hfq facilitated binding of the sRNA RhyB to the mRNA *sodB*, both RNA species are degraded in an RNase E dependent manner [29]. A sixth proposed mechanism of Hfq is its involvement in the polyadenylation of mRNA transcripts (**Figure 1.3E)**. In a study by *Mohanty et al,* total cellular RNA from log phase *E. coli* was analyzed for polyAdenylated (polyA) tail length; Hfq disruptions resulted in a decrease in polyA tail length of greater than 20 nucleotides [30]. A consequence of polyadenylation of mRNA in *E.coli* is the decrease of mRNA stability [31].

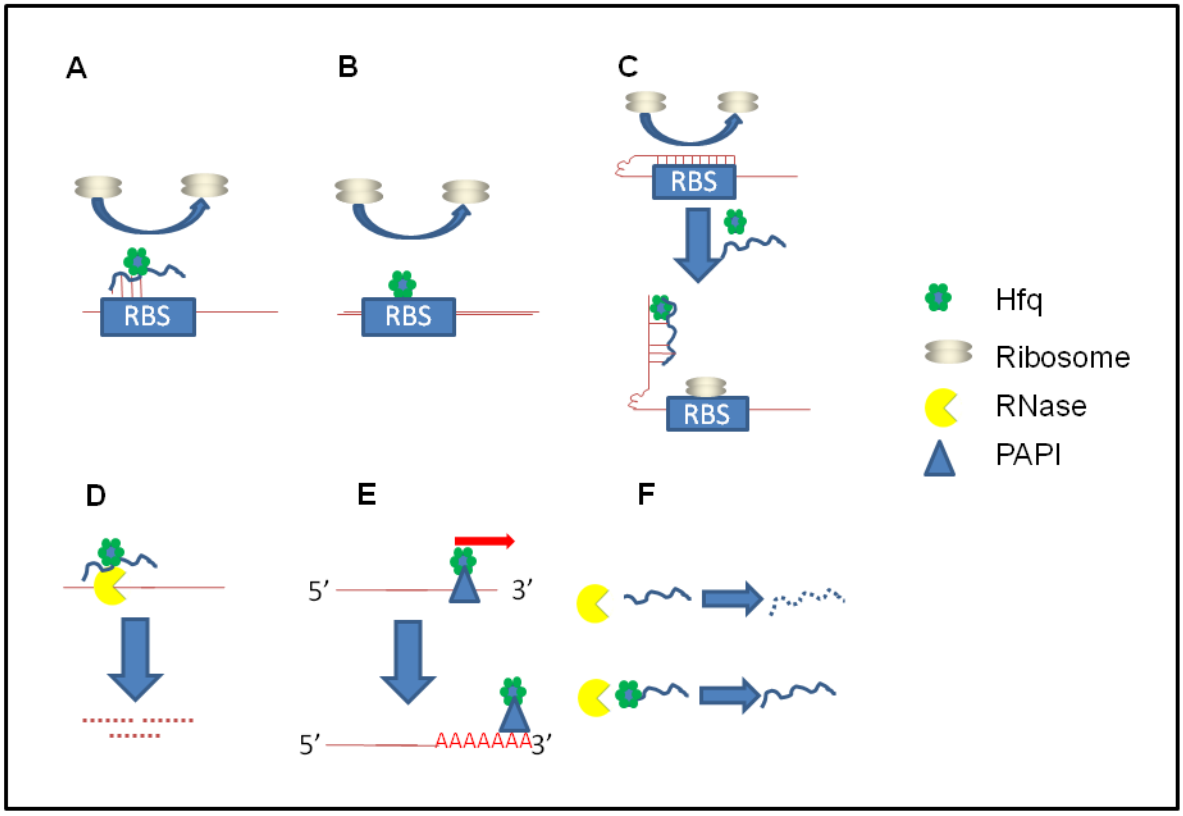


Figure 1.3 Hfq modes of action. A) Hfq and an sRNA can bind the ribosome binding site (RBS) of a target mRNA and inhibit translation. B) Hfq can directly bind the ribosome binding site of an mRNA preventing the 30S ribosomal subunit from binding. C) Hfq and an sRNA can positively regulate translation by altering the secondary structure of an mRNA that is self-occluding the RBS. D) RNase E can associate with Hfq's C-terminus and degrade mRNA that the Hfq:sRNA binary complex is base paired with. E) Hfq can recruit poly(A) polymerase to the 3' end of an mRNA, resulting in polyadenylation and subsequent destabilization of the transcript. F) Hfq can directly bind sRNAs and protect them from degradation

The impact of Hfq and sRNAs on gene regulation in bacteria is becoming apparent. To date, there are numerous classes of regulatory RNAs, including: cis-encoded RNAs encoded on the opposite stand of the mRNA, trans-encoded sRNAs which are often found in intergenic regions of the genome , and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) RNAs (reviewed in [22]). These 3 subclasses of small RNAs are transcribed from different parts of a bacterial genome and have different mechanisms of action. Cis-encoded RNAs

are encoded from the opposite DNA strand that codes for the mRNA, so they have perfect complementarity to a region of their target mRNA (reviewed in [32]). Trans-encoded sRNAs are often encoded in intergenic regions of the genome and can have multiple targets with limited complementarity; many trans-encoded sRNAs are dependent on Hfq to effectively basepair with their target mRNA (reviewed in [22]). CRISPR RNAs are found clustered in 1 or 2 loci of a bacterial genome and are thought to protect the cell from phage infection; research into the mechanism of action these CRISPR RNAs is still in preliminary stages (reviewed in [33]).

The first cis-encoded sRNA reported to regulate a transposon was the RNA-OUT transcript of Tn10 [34]**.** RNA-OUT is an anti-sense RNA that base pairs with RNA-IN (the transposase mRNA) ultimately resulting in the occlusion of the ribosome binding site and start codon, leading to reduced transposase expression [34]. Hfq may be involved in pairing of IS*10*'s RNA-IN and RNA-OUT [18] by increasing the half-life of RNA-OUT, by altering the secondary structure of the two RNA species making pairing more favourable or by increasing the local concentrations of the two RNA species making binding more probable. Other examples of cis-encoded RNA control have since been identified and characterized. An interesting study that found a connection between cis and trans-encoded sRNAs was on the *luxS* mRNA and its interaction with the sRNA MicA. MicA interacts with *OmpA* as a trans-sRNA; MicA is also a cis-encoded sRNA that, upon pairing with its target mRNA *luxS*, results in a processed transcript [35]. This interaction suggests that cis-encoded sRNAs may act on more

targets than their perfectly complementary partner mRNA and thus could be regulating a multitude of mRNAs by acting as a trans-sRNA as well. Generally, Hfq is not thought to be required for cis-encoded RNAs to pair with their target mRNA. However, Hfq's role in sRNA stability suggests that Hfq may be required for these sRNAs to have sufficient stabilities *in vivo* to act on their targets.

Trans-encoded sRNAs in *E.coli* are encoded from regions distant from their numerous targets. Trans-encoded sRNAs are frequently expressed in response to stress stimuli to the cell which can include oxidative stress (which induces the sRNA OxyS) [36] or low temperature growth (which induces the sRNA MicC) [37]. Trans-encoded sRNAs often require the RNA binding protein Hfq to promote pairing to their mRNA targets [38]. A well characterized sRNA:mRNA pairing reaction is between the sRNA GlmZ and the *glmS* mRNA (which encodes an enzyme involved in glucosamine-6-phosphate synthesis) [39]. Without the activating sRNA glmZ, the 5' untranslated region of the *glmS* RNA forms a hairpin occluding the ribosome binding site; Hfq binds to both GlmZ and *glmS* RNA species. GlmZ binds to *glmS* mRNA upstream of the Shine-Dalgarno in an Hfq dependent manner, resulting in an secondary structure change of the mRNA and an available ribosome binding site [39]. Overall, sRNAs allow the cell to quickly respond to cellular stresses through a network of sRNAs without the need to invest energy in regulatory proteins. Given that there is still a 7-fold increase in transposition in the absence of Hfq and the anti-sense RNA-OUT, there may be yet another undiscovered RNA-Hfq interaction involved in down regulating IS*10* transposition such as an endogenous sRNA interacting with RNA-IN.

The Hfq hexamer has two distinct RNA binding surfaces, known as the proximal and distal faces (reviewed in [40]). The two different faces of the Hfq hexamer are thought to bind different RNA sequences with the proximal face binding polyA/U sequences and the distal face binding sequences rich in the sequence ARNx (where R is a purine and N represents any nucleotide)[41]. Hfq is also able to bind DNA in a sequence independent manner [42]. Physiological concentrations of Hfq favour hexameric and multimeric forms of Hfq; the hexameric form is the active RNA and DNA binding form of the protein[43]. A study utilizing Genomic systematic evolution of ligands by exponential enrichment (SELEX) identified additional sequences that Hfq bound with high affinity: AAYAAYAA and AYAAUAA [44].

1.5 Regulation of Hfq expression:

Transcription of *hfq* is driven from 3 promoters (two σ^{70}-dependent and one σ^{35}-dependent) from the *hfq* operon [45]. Hfq self-regulates the amount of *hfq* transcripts in the cell though the binding and destabilization of its mRNA in exponentially growing cells [46]. Hfq also self regulates its own translation by binding to its own mRNA [47]. Regardless of such regulation, Hfq remains abundant in the exponentially growing cell with estimates ranging from 2500 molecules of monomer per cell [48] to 55,000 molecules of monomer per cell [49]. Deletions and disruptions of *hfq* result in shorter half-lives of small RNAs (sRNA) [28]**,** pleiotropic responses to cellular stress [50], and increased transposition of Tn*10*/IS*10*. Tn*5* transposition frequencies also increase in cells lacking functional Hfq [51].

1.6 Cellular factors that affect Hfq levels:

Recent studies by Hussein *et al.* and Moon *et al.* have shown that overexpression of sRNAs can result in a titration effect on cellular Hfq concentration. Most cellular Hfq is already bound to nucleic acids with 80-90% of cellular Hfq associated with ribosomes and 10-20% being found in the nucleoid [52]. Further depleting the pool of free cellular Hfq may result in the expression of genes that are normally repressed by Hfq thus allowing the cell to mount a quick response to cellular stress [53,54]. DksA is a transcription factor that responds to amino acid starvation conditions in conjunction with ppGpp, which modulates RNA polymerase activity; *dksA* deletion strains in *E.coli* and *Shigella flexneri* reduce Hfq protein levels by 50% [54,55]. Another cellular stress that results in reduced Hfq levels, is the transition between exponential phase to stationary phase which results in a 33% reduction of Hfq protein levels [49].

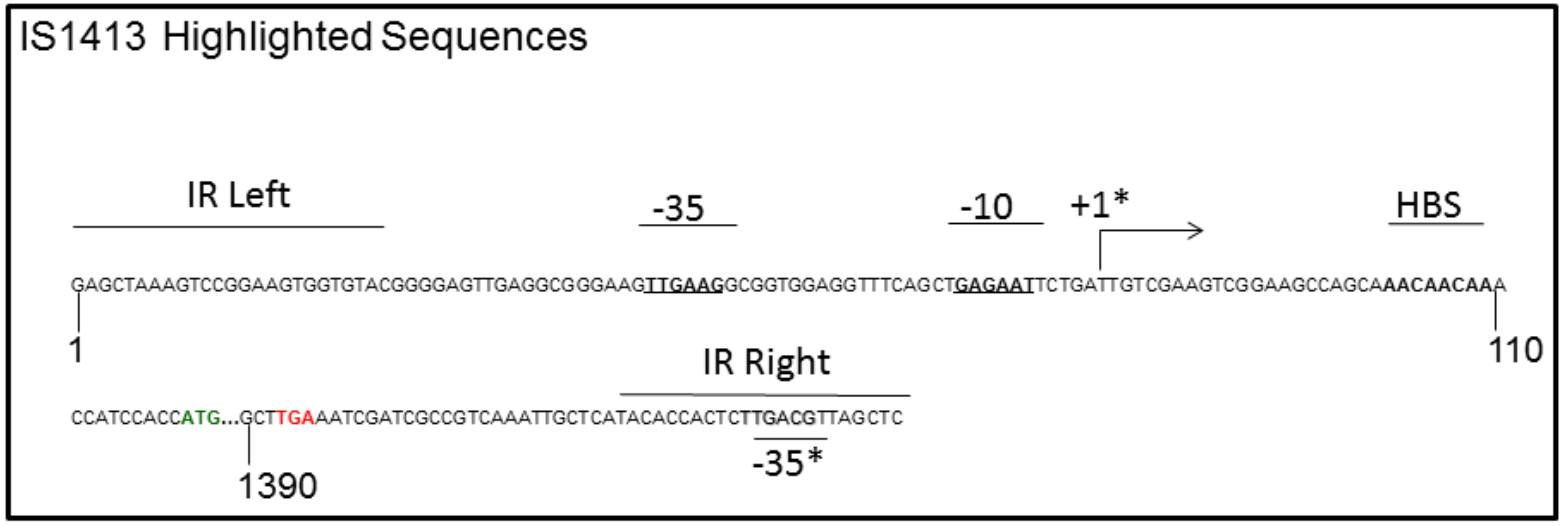


Figure 1.4 Features of IS*1413*: IS*1413* contains two flanking inverted repeats (denoted IR Left and IR Right) denoting the boundaries of the element. The +1* represents the transcriptional start based on 5' RACE data (lab data). The predicted start codon is found at bp 120. The … represents a break in the sequence. The stop codon is denoted in red at bp 1391. A predicted Hfq binding site is bolded and denoted HBS using the predicted Hfq binding aptamer (AAYAAYAA) [44]. The IR Right portion of IS*1413* contains a sequence TTGACG denoted -35* which can act as the -35 portion of an *E.coli* promoter when inserting in front of another gene.

1.7 IS*1413:*

IS*1413* is an IS Element (**Figure 1.**) originally identified in *Burkholderia cephacia AC1100* (now known as *Burkholderia phenoliruptrix)* which was found to occasionally create a promoter fusion with a promoterless streptomycin gene [56]. The promoter fusion occurs by the -35 sequence found in the right IR sequence (**Figure 1.**, shadowed text) inserting in front of a gene with a sequence resembling the -10 region of a promoter. There are 10 to 12 copies of IS*1413* in the genome [56] and very little is known about the mechanism of transposition or its regulation. IS*1413* was discovered by its 83% similarity to IS1*490* (the IS element thought to be responsible for *Burkholderia phenoltripix's* ability to break down DDT) [56]. IS*1413* contains a perfect match for a potential Hfq binding site (HBS) within its 5' untranslated region (UTR) (**Figure 1.**). This Hfq binding site may be indicative of a regulatory mechanism of IS*1413* that involves Hfq.

1.8 Tn*10:*

Tn*10* is a well-studied transposon that confers tetracycline resistance to its host. Tn*10* is a composite transposon comprised of two IS*10* copies (IS*10*L and IS*10*R) flanking tetracycline resistance genes (**Figure 1.**). IS*10*R is a functional IS element that can mobilize either as an IS element when the transposase acts at its inside and outside ends (**Figure 1.**), or mobilize the entire Tn*10* element when the transposase acts at the two outside ends. IS*10*L is a degenerate copy of IS*10* and does not encode a functional transposase [57]. The expression of IS*10*'s transposase is under tight regulation from host factors and an antisense RNA system. The antisense RNA system is self-encoded in the IS*10*R

transposase gene which encodes two RNA species (**Figure 1.**); RNA-IN (the mRNA of the IS*10* transposase) and RNA-OUT (an anti-sense RNA that contains 35 bases that are complementary to the 5' region of RNA-IN)[34]. Pairing of RNA-IN and RNA-OUT results in the occlusion of the ribosome binding site of RNA-IN [58] and an RNase II dependent decrease in the half-life of RNA-IN [59]. The anti-sense control is of particular importance when multiple copies of IS*10* are expressed. When expressing IS*10* on a multi-copy plasmid, the transposition frequency of a single chromosomal copy of Tn*10* decreases by about 20 fold [34]. This effect is known as multi-copy inhibition (mci). The mci effect can be explained by the transposase encoded by RNA-IN being most effective when acting on sequences close to where it was synthesized [13], in contrast to RNA-OUT which can freely distribute through the cell. Furthermore, RNA-IN levels are less than 1% that of RNA-OUT [11]. The differences in RNA-IN and RNA-OUT levels are due to RNA-OUT having a stronger promoter than RNA-IN and RNA-OUT having a longer half-life than RNA-IN (70 minutes versus 40 seconds, respectively) [11].

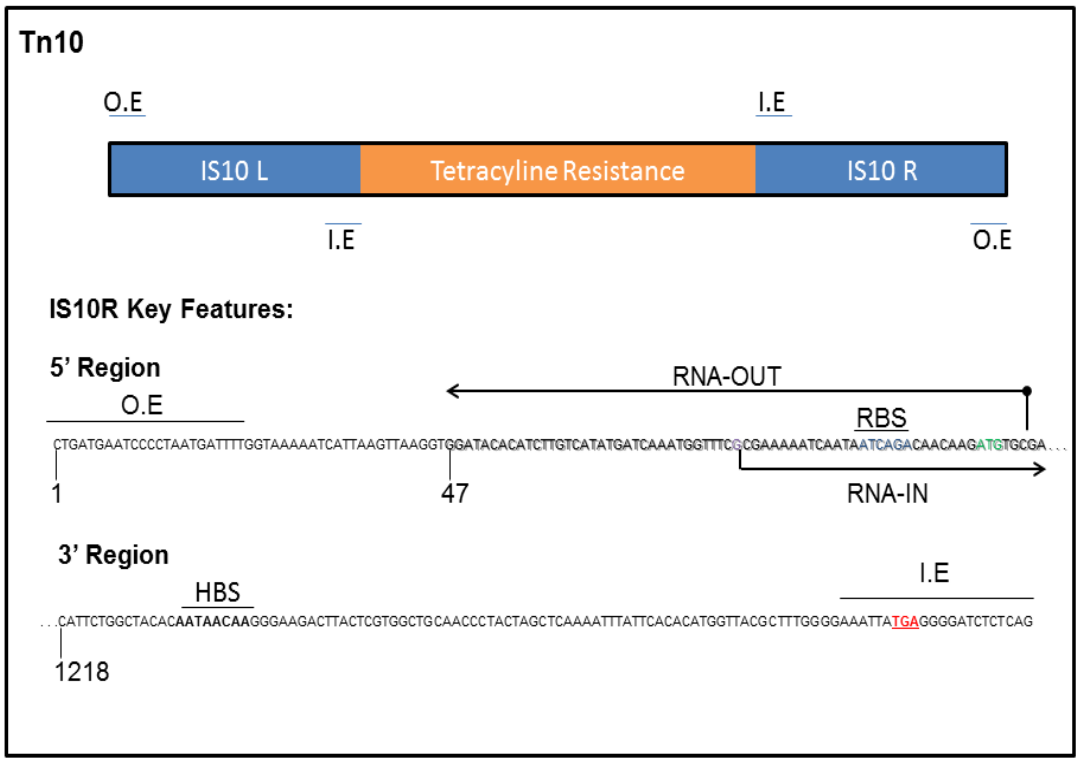


Figure 1.5 Structure of Tn*10* and key sequences of IS*10*R. O.E and I.E are the outside ends and inside ends respectively. The IS*10*R encoded transposase binds to the I.E and O.E. RNA-IN (the transposase mRNA) starts at the G denoted in purple. The ribosome binding site (RBS) of RNA-IN is denoted in blue and the translational start denoted in green. Downstream towards the 3' region of RNA-IN is a predicted Hfq binding site (HBS). The translational stop codon is denoted in red within the I.E. RNA-OUT is the antisense RNA and is denoted in the shadowed text. RNA-OUT is transcribed from the opposite DNA strand of RNA-IN.

1.9 Thesis topics addressed:

Transposons are heavily regulated by host factors. Hfq is a host factor that negatively regulates IS*10*/Tn*10* and IS*50*/Tn*5*. Hfq may be a global regulator of many IS Elements. An RNA interference mechanism has already been characterized in Eukaryotes that repress mobile genetic elements such as viruses and transposable elements. Small interfering RNAs (siRNAs) and microRNAs (miRNAs) are two defense mechanisms eukaryotes use to silence genes (including transposons and viral genes); siRNAs result in the cleavage of a target mRNA and miRNAs result in repressed translation or sometimes cleavage of the target mRNA (reviewed in [60,61]). Perhaps sRNAs could be functioning as a Prokaryotic method of silencing transposable elements. Tn*10*/IS*10* and Tn*5*/IS*50* have already been shown to be repressed by Hfq which is closely connected to the sRNA regulation in Bacteria. In this thesis I follow up on three main questions:

1) Can transposons that are subject to Hfq dependent regulation be identified by searching for putative Hfq binding sites in the transposase transcript?
2) Is an endogenous *E.coli* sRNA involved in Hfq-mediated regulation of Tn*10*/IS*10* transposition?
3) Does limiting the amount of synthesis or availability of Hfq influence Tn*10*/IS*10* transposition?

In this work I explored the effects Hfq has on the regulation of the two IS Elements IS*10* and IS*1413* which were identified by searching for putative Hfq

binding sites in the IS Finder database. IS*10* contains an Hfq binding site in its 3' region, however disrupting this Hfq binding site through mutation does not result in a de-repression of Hfq in a transposition assay. IS*1413* has an Hfq binding site found within its 5' untranslated region and for the first time I show it can undergo transposition in *E.coli* at an average frequency of $3.7 \times 10^{-5} \pm 7.8 \times 10^{-5}$ in the presence of Hfq and increases its average transposition frequency to $5.0 \times 10^{-4} \pm 3.8 \times 10^{-4}$ in the absence of Hfq. I also identified that IS*10* transposition increases in wildtype cells by 9.81 ±1.64 fold when cells are starved and put under non-aerating conditions. Finally I analyzed the 3' region of IS*10* to determine where the transcript terminates and looked for a potential mechanism in which Hfq could be acting on the RNA (such as facilitating polyadenylation). A defined stop point which appears to either be due to a processing event or run off transcription was found using the 3' RACE technique.

Chapter 2 - Materials and Methods

2.1 Bacterial strains used in this thesis:

All bacterial strains (*E.coli* and *Burkholderia*) used in this thesis are compiled in **Table 2.1**. All plasmids used in this study are compiled in **Table 2.2**. Bacterial strains with gene disruptions were made via generalized transduction. Phage P1 was used for generalized transduction [62]. P1 Phage was used to infect *E. coli* cells containing a "marked" gene of interest (containing an antibiotic resistance gene). A culture of *E.coli* containing the "marked" gene was grown overnight in LB broth with selection and subcultured in the absence of antibiotic but supplemented with 25 mM $MgCl_2$, 5 mM $CaCl_2$, and 0.2% glucose to grow to early log phase. Forty µL of P1 phage were added to the subculture of cells, and further incubated for 3 hours at 37°C with aeration. Fifty µL of chloroform was added to the infected cells, vortexed, and centrifuged at 6000 rcm for 5 minutes. The supernatant was removed and stored at 4°C. Phage lysate was titred to a multiplicity of infection of 1-5 plaque forming units/cell. NK5830 F' cells transformed with a temperature sensitive plasmid encoding recA (pET001) were grown in LB broth overnight at 30°C with aeration. Cells were subcultured (1:20) in LB broth containing 25 mM $MgCl_2$, 5 mM $CaCl_2$, and 0.2% glucose at 30°C with aeration and grown to exponential phase. Cells and phage lysate were mixed at a 1:1 ratio (100 µL of exponentially growing cells (undiluted, diluted by a factor of 10, or diluted by a factor of 100) and 100 µL of phage lysate) and incubated at 30°C for 60 minutes. Two-hundred microlitres of 1 M sodium citrate (pH 5.5) was added to the transduction mix, followed by 1 mL of LB broth. Cells were

incubated at 37°C with aeration for 60 minutes, spun down at 6000 rcm for 5 minutes and resuspended in 0.85% NaCl in H_2O. Cells were plated on LB agar containing 20 µg/mL of chloramphenicol and incubated at 37°C. Individual cells picked from overnight plates and were serially passaged on LB agar containing either 20 µg/mL of chloramphenicol as well as LB agar plates containing 100 µg/mL of ampicillin (to ensure the recA plasmid was lost) until colonies were cured of the recA containing plasmid pET001.

Table 2.1 Bacterial strains used in this thesis

Strain	Description	Source/ Reference
NK5830 F'	*recA56* su° *lacproXII1*, Arg^-, Ara^-, Nal^R, Rif^R/F' *laci*^{q} L8 *pro*	[63]
DBH16	NK5830F' transduced to hfq-1::ΩCm	[18]
DBH48	NK5830F' Transduced to rpoS::cat (phage from NM22542 ΔnlpD-rpoS Cm^R)	(Laboratory stock)
DBH144	Burkholderia phenoliruptix AC1100 (Gift from Dr. Zlosnik)	A Gift from Dr. Zlosnik (UBC)
DBH174	NK5830 ΔdksA::Tet (using phage lysate from MG1655 ΔdksA::Tet from M.Cashel via S Gottesman (CF9239)	This Study
DBH 181	NK5830Δ SroG::IS50-Kan-LacZ	Laboratory stock
HB101	F^-, leu^- ; Str^R	[64]

2.2 Plasmids used in this thesis:

All plasmids used in this thesis are compiled in **Table 2.2.** Plasmids pDH694-696 are pACYC184 derivatives; they contain pLlacO1 inducible sRNAs (DsrA. MicC, and RhyB respectively). These sRNA inducible plasmids were constructed by PCR amplifying DsrA, MicC or RhyB from a genomic preparation of NK5830 DNA using primer pairs BM43A and BM44A for DsrA, BM41A and

BM42A for MicC and BM40A and BM39A for RhyB. PCR products were digested using *Nco* I and *Xmn* I and cloned into *Nco* I/*Xmn* I digested pACYC184 plasmid. For DsrA, MicC and DsrA PCR reactions the reaction conditions were as follows: Genomic DNA preparations of NK5830 were done using the GenElute™ Bacterial Genomic DNA Kit (Sigma) following the manufacterer's protocol. PCR reactions were the same for all three pLlacO1 inducible sRNAs. Reactions consisted of 10 ng of NK5830F' gDNA,, the sRNA respective forward and reverse primers [1 µM each], combined with GoTaq Mastermix (Promega) [1x]. PCR conditions were: 98°C (2 min) for the initial denaturation, followed by 98°C (30s), 65°C (40s), and 72°C (1 min) repeated 30 times. A final extension step was done at 72°C for 5 min. The resulting amplicon was gel purified using the QIAEX II Gel Extraction Kit (Qiagen).

The new IS*10*-Kan construct pDH725 and the BM1 mutant pDH734 were generated by PCR amplifying IS*10* on pDH669 (for WT) or pDH703 (for BM1 Mutant) from bp1 to the end of the stop codon at bp1316 using primers BM27A and BM26A. Reactions consisted of 10 ng of template (pDH669 or pDH703), forward and reverse primers [1 µM each], combined with GoTaq Mastermix (Promega) [1x]. PCR conditions were: 98°C (2 min) for the initial denaturation, followed by 98°C (30s), 62.2°C (45s), and 72°C (2 min) repeated 35 times. A final extension step was done at 72°C for 5 min. The resulting amplicon was gel purified using the QIAEX II Gel Extraction Kit (Qiagen). This IS*10* amplicon was cloned into pGEM-T-Easy (Promega). The kanamycin gene containing IS*10*'s inside end was PCR amplified from pDH502 using the primers BM23A and BM22A. The BM22A primer contains 27 bases of IS*10*'s inside end (IS*10* I.E), 5'

to the complementary sequence of the kanamycin gene. PCR reactions consisted of 10 ng of template (pDH502), forward and reverse primers [1 μM each], combined with GoTaq Mastermix (Promega) [1x]. PCR conditions were: 98°C (2 min) for the initial denaturation, followed by 98°C (30s), 55°C (45s), and 72°C (2 min) repeated 5 times. Another set of PCR parameters of 98°C (30s), 68°C (45s), and 72°C (2 min) was then followed for 30 cycles: A final extension step was done at 72°C for 5 min. The resulting amplicon was gel purified using the QIAEX II Gel Extraction Kit (Qiagen).This kanamycin-IS*10* I.E amplicon was digested with *Spe* I and *Xba* I and cloned into the IS*10*-containing pGEM -T-easy construct digested with *Spe* I.

The LacZ translational fusion plasmids pDH703 and 704 were made through site directed mutagenesis of pDH669. For pDH703 primers BM10A and BM11A were used with pDH669 and Pfu polymerase (Stratagene) according to manufacturer's guidelines for site-directed mutagenesis. For pDH704 primers BM12A and BM13A were used with pDH669 and Pfu polymerase (Stratagene) according to manufacturer's guidelines for site-directed mutagenesis

The pDH734 (IS*10*-Kan BM1 mutation) was constructed by PCR amplifying the IS*10* fragment on pDH703 with primers BM26A and BM18A. PCR reactions consisted of 10 ng of template (pDH703), forward and reverse primers [1 μM each], combined with GoTaq Mastermix (Promega) [1x]. PCR conditions were: 98°C (2 min) for the initial denaturation, followed by 98°C (30s), 62.2°C (45s), and 72°C (2 min) repeated 25 times. A final extension step was done at 72°C for 5 min. The resulting amplicon was cloned into pGEM T-easy vector

(Promega). The construct was digested with SpeI and the gel purified using the QIAEX II Gel Extraction Kit (Qiagen).

The IS*1413* containing plasmid pDH735 was constructed by PCR amplifying IS*1413* from a genomic DNA preparation from *Burkholderia phenoltripix* AC1100 using primers BM1A and BM2A. The Taq polymerase amplified amplicon was cloned into pGEM T-easy (Promega). The mutated IS*1413* LacZ translational fusion plasmids pDH760 and pDH794 were both made by amplifying the IS*1413* portion of pDH757. The forward primers BM4A and BM3A contain a *Nhe* I cut site as well as the IS*1413* SELEX mutations C and B respectively **(see Figure 3.1B)**. The reverse primer used to amplify IS*1413* for LacZ translational fusions was BM5A which contains a *Sph* I cut site. Amplicons of IS*1413* mutants B and C were digested with *Nhe* I and *Sph* I and cloned into a *Nhe* I/*Sph* I digested pBad 24 vector. The kanamycin 'marked' IS*1413* construct pDH797, was made by first doing site-directed mutagenesis on IS*1413* using primers BM1A and BM7A with pDH735 as a template to introduce a *Xho* I cut site after the stop codon in the IS*1413* transposase gene. The resulting amplicon contained a truncated IS*1413* (missing 27 nucleotides), to rectify this, the *Xho* I containing amplicon was used as a template for another PCR reaction using BM1A and BM6A primers. The resulting amplicon was a full length IS*1413* sequence containing a *Xho* I cut site and was cloned into pGEM-T Easy (Promega) to make pDH796.

Table 2.2 Plasmids used in this thesis.

Plasmid	Description	Source or Reference
pDH502	pBR322-*h*IS*G1::*IS*10KAn* (pNK1219)	[65]
pDH602	pACYC184- HisG1::IS*10*Kan	[18]
pDH662	1-121 IS*10* translational fusion to the 9th codon of LacZ	Unpublished, a gift from C. Jain
pDH663	1-180 IS*10* translational fusion to the 9th codon of LacZ	Unpublished, a gift from C. Jain
pDH664	1-290 IS*10* translational fusion to the 9th codon of LacZ	Unpublished, a gift from C. Jain
pDH668	1-1020 IS*10* translational fusion to the 9th codon of LacZ	Unpublished, a gift from C. Jain
pDH669	1-1320 IS*10* translational fusion to the 9th codon of LacZ	Unpublished, a gift from C. Jain
pDH676	Tn10 in a broad host range plasmid (pBE10)	[66]
pDH694	pACYC184 encoding pLlacO-1 promoter driving expression of DsrA (as a pcr fragment cloned into *Nco* I/*Xmn* I-cut pACYC184)	This Study
pDH695	Same as 694 but with MicC	This Study
pDH696	Same as 695 but with RyhB	This Study
pDH703	pDH669 containing the BM1 mutation	This Study
pDH704	pDH669 containing the BM2 mutation	This Study
pDH725	Kanamycin marked IS*10* with no repeated sequence	This Study
pDH734	pDH725 with a mutated putative Hfq binding site Tyr 374 TAC→TAT (found just outside of the SELEX site) and an Arg 378 AGG →CGG	This Study
pDH735	IS*1413* in pGEM T-easy	This Study
pDH757	IS*1413* (up to bp 445) lacking -10/-35 ligated into MCS of pBAD24 ampR	Morgan Black's Fourth Year Thesis

pDH760	IS*1413*-LacZ (from 757) lacking native promoter sequence with a mutated Hfq binding site (GAACACCAA) cloned into MCS (NheI/SphI) of pBAD24	This Study
pDH794	Same as pDH760 only SELEX mutation = TAAACTACAA	This Study
pDH796	*1413* in pGEM mutations at A1397C,T1402G to make a XhoI cut	This Study
pDH797	*Sph* I-Kan-*Sph* I PCR product (from 602) cloned into pDH796 *Xho* I site	This Study
pDH763	Empty pLlacO1 vector (pBR322 derivative	[67]
pDH783	plLac-GlmZ (pBR322 derivative)	[67]
pET001	pRR10-ts97-derived recA; temperature-sensitive plasmid; Amp^R	[68]

A kanamycin resistance gene was amplified from pDH502 using primers BM24A and BM25A. The resulting amplicon was digested with *Sph* I and cloned into the *Xba* I site of pDH796 to make pDH797.

2.3 Mating out assay:

For the mating out assays, NK5830 F' and its isogenic mutant, Hfq^-, $RpoS^-$, H-NS^-, or $DksA^-$, derivatives were used as donors and HB101 as the recipient. IS*10*/IS*1413*-Kan encoding plasmids (pDH502, pDH602, pDH725, pDH734, or pDH797) were transformed into donor strains through electroporation and transformants were selected for on LB plates containing kanamycin [25 µg/mL]. Under normal circumstances (Normal Mating Out), overnight cultures of donor cells were grown overnight at 37°C with aeration to saturation in LB broth containing kanamycin [25 µg/mL]. Overnight cultures were subcultured in LB broth in the absence of antibiotics. For the mating out assay under starvation conditions (Starved), donor cells were grown at room temperature (~25°C) on benchtop (no aeration) in LB containing kanamycin [25 µg/mL] for 110 hours then subcultured in LB broth in the absence of antibiotics. The first step of subculturing consisted of diluting overnight cultures (1:20) in LB broth and growing for 2 hours with aeration (250 rpm) for Hfq^+ and $RpoS^-$). For Hfq^- and $DksA^-$(cells), the first step consisted of diluting overnight cultures (1:20) in LB broth and growing for 4 hours with aeration (250 rpm). After the first step of subculturing, Hfq^+ and $RpoS^-$ cell cultures were moved to a gentle shaker (50 rpm) at 37°C for an additional 2 hours. Cell cultures containing Hfq^- and $DksA^-$ donor cells were grown with gentle shaking for 50 rpm for 4 additional hours. The

extended growth time for the mutant strains allowed the slow growing Hfq$^-$ and DksA$^-$ donor cells to reach a similar cell density as the NK5830F' cells. Recipient HB101 cells were grown in LB broth at 250 rpm at 37°C to reach an OD_{600} of 0.5. Recipient and donor cells were gently mixed at a ratio of (2.5 mL : 1 mL) respectively and grown at 50 rpm at 37°C for an additional hour. Cell mixtures were vortexed for 30 seconds. One mL of cell mixture was removed, centrifuged (4000 rcm for 4 minutes) and resuspended in 1 mL of 0.85% saline. To determine the number of exconjugates (recipient cells that received the F-plasmid), cell mixtures were plated on M9 plates containing glucose [0.2%], leucine [40 µg/mL], thiamine [1 µg/mL] and streptomycin [150 µg/mL]. To determine the number of transposition events (cells that received an F-plasmid that also contained a marked IS element), cell mixtures were plated on M9 plates containing glucose [0.2%], leucine [40 µg/mL], thiamine [1µg/mL] streptomycin [150 µg/mL], and kanamycin [25 µg/mL]. The transposition frequency was determined by the following equation:

$$Transposition\ Frequency = \frac{\#\ of\ transposition\ events}{\#\ of\ exconjugates}$$

For the sRNA overexpression mating out assays, the same procedure outlined above was followed with the following exceptions. NK5830 F' and its Hfq$^-$ isogenic mutant DBH16 were co-transformed with pDH502 as an IS*10*-Kan source and one of the sRNA expression plasmids (pDH694-696) for DsrA, RhyB or MicC. Transformants were plated on LB-selective media containing tetracycline [10 µg/mL] and kanamycin [25 µg/mL] either in the presence or absence of 1mM IPTG for induction. Selection and induction conditions

remained the same for overnight cultures. The rest of the procedure remained the same as outlined above. For GlmZ overexpression, the same procedure was followed with the exceptions outlined below. The source of IS*10*Kan was pDH602 (to avoid plasmid incompatibilities). IPTG was used at a concentration of 100 μM and the antibiotics used were ampicillin [50 μg/mL] and kanamycin [25 μg/mL]. Cells were checked after the overnight incubation (by using the Sigma-Aldrich Mini-prep kit) to ensure cells maintained both plasmids.

Table 2.3 Oligos used in this thesis.

Oligo	Use	Sequence
BM1A	Forward primer for IS*1413*	GAGCTAAAGTCGGAAGTGGTGTACGGGGAGTTGAGGC
BM2A	Reverse primer for IS*1413* (Full Length)	GAGCTAACGTCAAGAGTGGTGTATGAGCAATTTGA
BM3A	Forward primer for IS*1413* SELEX site mutation B in the LacZ translational fusion)	CGTAGCTAGCAGCTGATTGTCGAAGTCGGAAGCCAGCGAACACCAA
BM4A	Forward primer for IS*1413* for mutation C in the SELEX Site for LacZ-Translational fusions	CGTAGCTAGCAGCTGATTGTCGAAGTCGGAAGCCAGTAAACTACAA
BM5A	Reverse primer for LacZ gene for introduction into pBad24 vector	CGATGCATGCACTGTTATTTTTGACACCAGACC
BM6A	Reverse primer to complete the IS*1413* sequence in the Xho I mutation	TCAAATTGCTCATACACCACTCTTGACGTTAGCTC
BM7A	Reverse primer for introduction of a *Xho* I site into IS*1413*	AGCAATTTGACGGCGCTCGAGTTCAAGC
BM8A	Forward primer #1 for IS*10* 3' RLM-RACE nested PCR	GAGCCATGGGTTCTAGCAACTAACTTACC
BM9A	Forward primer #2 for IS*10* 3' RLM-RACE nested PCR	TACGCCATAGCCGAACGAGCAGC
BM10A	Forward primer for introduction of the BM1 mutation in IS*10*	TTGCGGCATTCTGGCTATACAATAACACGGGAAGACTTACTCG
BM11A	Reverse primer for introduction of the BM1 mutation in IS*10*	CGAGTAAGTCTTCCCGTGTTATTGTATAGCCAGAATGCCGCAA

BM12A	Forward primer for introduction of the BM2 mutation in IS*10*	TTGCGGCATTCTGGCTAC ACGATAACGAGGGAAGA CTTACT
BM13A	Reverse primer for introduction of the BM2 mutation in IS*10*	AGTAAGTCTTCCCTCGTT ATCGTGTAGCCAGAATGC CGCAA
BM14A	Forward primer for 3' RNA-IN (includes a T7 RNA Polymerase promoter)	TAATACGACTCACTATAG GGCTGGCTACACAATAAC AAGGGAAG
BM15A	Reverse primer for 3' RNA-IN (includes a T7 RNA Polymerase promoter)	CTGAGAGATCCCCTCATA ATTTCCCCA
BM16A	Forward primer for 3' RNA-IN BM1 (includes a T7 RNA Polymerase promoter)	TAATACGACTCACTATAG GGCTGGCTATACAATAAC ACGGGAAG
BM17A	Reverse primer for 3' RNA-IN BM1 (includes a T7 RNA Polymerase promoter)	TAATACGACTCACTATAG GGCTGGCTACACGATAAC GAGGGAAG
BM18A	Reverse primer to introduce a stop codon in the IS*10*-LacZ translational fusions, for subsequent cloning to make the new IS*10*-Kan construct	TCATAATTTCCCCAAAGC GTAACC
BM19A	Reverse primer (1) for the RNA-Adapter in 3' RLM RACE nested pcr (DE-193)	TTTCATCAAAGCCAGCAA ACGC
BM20A	Reverse primer (2) for the RNA Adapter in 3' RLM RACE nested PCR RNA Adapter (DE-194)	CAAACGCAGTGTTCATTC ATCGCC
BM21A	RNA-Adaptor for 3' RACE	GCUGAUGGCGAUGAAUG AACACUGCGUUUGCUGG CUUUGAUGAAA
BM22A	Forward primer for kanamycin resistance gene	GCTCTAGACTGAGAGATC CCTCATAATTTCCCCTCA GCAAAAGTACGATTTATT CAAC
BM23A	Reverse primer for kanamycin resistance gene	GGACTAGTGCCAGTGTTA CAACCAATTAACCAATTC TGA
BM24A	Forward primer for kanamycin gene (with a *Sal* I cut site introduced)	ACGCGTCGACTCAGCAAA AGTACGATTTATTCAAC
BM25A	Reverse primer for kanamycin gene (with a *Sal* I cut site introduced)	ACGCGTCGACGCCAGTG TTACAACCAATTAACCAAT TCTGA
BM26A	Forward primer for IS*10*	NNCAGCTGATGAATCCCC TAATGATTTTGG
BM27A	Reverse primer for IS*10*	GGACTAGTTCATAATTTC

		CCCAAAGCGTAAC
BM28A	Forward primer for *rrsb* (16S rRNA) RT-PCR	CGTGTTGTGAATGTTGGGTTAAGT
BM29A	Reverse primer for *rrsb* (16S rRNA) RT-PCR	AACCCACTCCCATGGTGTGACGGGC
BM30A	Forward primer for *glmZ* RT-PCR	GTAGATGCTCATTCCATCTCTTATGTTC
BM31A	Forward primer for *glmZ* (includes a T7 RNA Polymerase promoter)	TAATACGACTCACTATAGGGTAGATGCTCATTCC
BM32A	Reverse primer for *glmZ*	AAAACAGGTCTGTATGACAACAAGTGG
BM33A	Forward primer for *rdlC* (includes a T7 RNA Polymerase promoter)	TAATACGACTCACTATAGGGTCTGGTTTCAAGATTAGCC
BM34A	Reverse primer for *rdlC*	AGAAAACCCCCGCACGTTG
BM35A	Forward primer for *sroG* (includes a T7 RNA Polymerase promoter)	TAATACGACTCACTATAGGGCTTATTCTCAGGGCG
BM36A	Reverse primer for *sroG*	CGTTACTCTCTCCCATCCGGAC
BM37A	Forward primer for *micC* RT-PCR	GTTATATGCCTTTATTGTCACAG
BM38A	Reverse primer for *micC*-RT-PCR	AAAGCCCGGACGACTGTTCG
BM39A	Forward primer for pLlacO1 *rhyB*	NNCCATGGNNATAAATGTGAGCGGATAACATTGACATTGTGAGCGGATAACAAGATACTGAGCACAAACACATCAGATTTCCTG
BM40A	Reverse primer for pLlacO1 *ryhB*	NNGAACCGGTTCNNAAAAGCCAGCACCCGGCT
BM41A	Forward primer for pLlacO1 *micC*	NNCCATGGNNATAAATGTGAGCGGATAACATTGACATTGTGAGCGGATAACAAGATACTGAGCACAGTTATATGCCTTTATTGTC
BM42A	Reverse primer for pLlacO1 *micC*	NNGAACCGGTTCNNAAAAAAAGCCCGGACGAC
BM43A	Forward primer for pLlacO1 *dsrA*	NNCCATGGNNATAAATGTGAGCGGATAACATTGACATTGTGAGCGGATAACAAGATACTGAGCACAGCGATCAGGAAGACCCTC
BM44A	Reverse primer for pLlacO1 *dsrA*	NNGAACCGGTTCNNAAATCCCGACCCTGAGGG

2.4 Finding putative Hfq binding sites in IS Elements:

The following families were searched for on the IS Finder database (http://www-is.biotoul.fr/): IS*1*, IS*3*, IS*4*, IS*5*, IS*6*, IS*21*, IS*30*, IS*66*, IS*91*, IS*110*, IS*200*, IS*605*, IS*256*, IS*481*, IS*630*, IS*982*, IS*1380*, IS*As1*, IS*L3*, Tn*3*, IS*1595*, IS*1182*, IS*1634*, IS*H3*, and IS*701*. The search results were given as a table and copied and pasted into Microsoft Excel. The URLs of each specific IS Element were extracted to create lists of the URLs of all elements within a family. WebSphinx Webcrawler (http://www.cs.cmu.edu/~rcm/websphinx/) was then used to compile all of the URLs within each family into one concatenated webpage saved as a text file. A Microsoft Word 2010 Macro was then used to extract the following fields from the compiled text files: Family, Element, Hosts, Transposition (Y or ND) IS_SEQ Comments and References, and put into a tab delimitated excel file (see **Script A1.1** in Appendix I) and more details. This Microsoft Excel file was then searched using the SELEX site AAYAAYAA, yielding the above results. In the database there were 3478 active links; 226 of the IS-Elements were tested for transposition (Y) = 6.50% of submitted sequences being actually transposing elements.

2.5 β-Galactosidase assays:

E. coli cells (NK5830F' or its isogenic *hfq*$^{-}$ derivative DBH16) were electroporated with transposase-LacZ translational fusions containing either an arabinose inducible IS*1413* transposase gene portion (pDH757, pDH760 or pDH794) or an IS*10* transposase gene portion (pDH662,664,668,669,703 or

704). Transformants were grown overnight at 37°C on LB agar plates containing ampicillin (50 µg/mL). Individual colonies were used to inoculate LB broth with ampicillin [50 µg/mL] and grown overnight at 37°C with aeration (250 rpm). Arabinose inducible constructs were grown either in the presence of arabinose [0.2%] or glucose [0.2%]. Overnight cultures were subcultured by diluting cultures 1:50 in LB broth in the absence of antibiotic and grown for 2 hours (for hfq^+) or 4 hours (for hfq^-) at 37°C. Cultures were then placed on ice for 20 minutes. One mL of subculture was centrifuged (4000 rcm for 4 minutes), aspirated, and resuspended in MinA broth. The turbidity of the cultures (OD_{600}) was recorded on a spectrophotometer (Montreal Biotech NanoPhotometer). Two hundred and fifty µL of resuspended cells were resuspended in 250 µL of LacZ Buffer permeablization buffer (60 mM Na_2HPO_4, 34 mM $NH_2PO_4 \bullet H_2O$, 2.5 mM KCl and 50 µM $MgSO_4 \bullet 7H_2O$, 1% SDS, 0.3% β-mercaptoethanol). Fifty µL of chloroform was added to each tube of permeabilized cells. Cells were lysed at 30°C for 5 minutes, followed by the addition of 100 µL of ortho-Nitrophenyl-β-galactosidase [4 mg/mL in LacZ Buffer] (this is considered the start time of the reaction). Samples were incubated at 30°C until they took on a yellow hue after which 250 µL of sodium carbonate [100 µg/mL] was added (this is considered the stop point of the reaction). Samples were further incubated for 5 minutes at 30°C, centrifuged (max speed, 5 minutes) and decanted into a cuvette. The optical density was measured for each sample at wavelength 420 nm and 550 nm. Miller units were calculated using the following equation:

$$Miller\ Units = 1000 \times \frac{(OD_{420} - (1.75 \times OD_{550})}{(t\ x\ v\ x\ OD_{600})}$$

Where

t = Elapsed time from start to finish

v= volume of cells used

2.6 Hfq protein purification:

Hfq protein was purified as described in [18].

2.7 *In vitro* transcription and RNA labeling:

DNA templates for *in vitro* transcription were made via PCR. Wild type, BM1 mutant and BM2 mutant RNA-IN run off transcription templates (bp 1222-1329) were prepared using a T7 RNA polymerase (RNAP) promoter containing forward primer BM14A, BM16A or BM17A respectively. A common reverse primer BM15A, was used for each transcript. The sRNAs GlmZ, RdlC, and SroG run off transcription templates were prepared using the primer pairs BM31A/BM32A, BM33A/BM34A, and BM35A/BM36A. The forward primers BM31A, BM33A and BM35A all contain T7RNAP promoters. The resulting amplicons were gel purified using the QIAEX II Gel Extraction kit and protocol (Qiagen).

Run off i*n vitro* transcription reactions consisted of 100 ng of linear DNA template, rNTP mix {NEB) (rATP [0.5 mM], rCTP [0.5mM], rGTP [0.5 mM]), 2.5µL of α^{32}P- UTP [3000Ci/mMol] (PerkinElmer) 20 units of RNasin (Promega), 50 units of T7 RNA polymerase (NEB), 1X BSA (NEB), 16 µM $MgCl_2$, 10 mM

DTT, inorganic pyrophosphatase [0.25U] (NEB) and T7 RNA polymerase buffer [1x] (NEB). For *in vitro* transcription of unlabeled RNA, α^{32}P- UTP was replaced with 0.5 mM rUTP. The reaction was incubated at 37°C for 30 minutes. Two µL of Turbo DNase (Invitrogen) was added to the reaction mixture and incubated for another 30 minutes. The reaction mixtures were run through a BioSpin 30 column (BioRad). An equal volume of stop dye (10 mM EDTA [pH 7.5]; 97.5% formamide (v/v); 3 mg (w/v) xylene cyanol and bromophenol blue) was added to reaction mixtures. The reaction mixtures were then loaded on a 10% denaturing (3M urea) polyacrylamide gel and ran at 21V/cm until the bromophenol blue dye reached the bottom of the gel. RNA bands were cut out of the gel and eluted overnight at 42°C in elution buffer (0.25M ammonium acetate [pH 5.8], 1 mM EDTA [pH 8], 0.1% SDS). The RNA was ethanol precipitated in 200 µL aliquots by adding 8 µL of DEPC treated NaCl [5M], 1 µL of glycogen (20 mg/mL), 4 µL EDTA (100 µM, pH 8.0), and 700 µL of 100% ethanol. Samples were incubated at -72°C for 30 minutes, centrifuged at maximum speed for 30 minutes at 4°C, washed twice with 70% ethanol and dried until no liquid remained (~ 5-10 minutes). RNA was resuspended in Hfq buffer (50 mM Tris [ph 7.5]; 1 mM EDTA; 250 mM ammonium chloride; 10% glycerol)) and quantified using a spectrophotometer (Montreal Biotech NanoPhotometer).

2.8 Electrophoretic mobility shift assays:

RNA (^{32}P labeled or unlabelled) was thawed on ice, heated at 95°C for 2 minutes, and then immediately put back on ice. Hfq protein was diluted in Hfq buffer to appropriate concentrations. RNA concentrations are indicated in the

figure legends of the appropriate experiments. When two species of RNA were being looked at simultaneously (pairing experiments), the two RNA species were mixed together and incubated at 37°C for 15 minutes. For competition experiments, Hfq was preincubated with the competitor RNA (*dsrA* or *polyA*$_{16}$) for 10 minutes at room temperature followed by the addition of RNA-IN for an additional 10 minute incubation at room temperature. For EMSAs looking at only one RNA species, Hfq the RNA were incubated for 10 minutes at room temperature together in 10 µL. Binding reaction mixes had load mix added (20 mM Tris-HCl pH 7.5, 10 mM DTT, 100 mM KCl, 30% glycerol; 0.05% w/v bromophenol blue) and were loaded onto a 6% native polyacrylamide gel. Gels were run at 21V/cm until the bromophenol blue ran to the bottom. The gel was then dried, and analyzed using a Phosphoimager (Storm 860 Scanner) and ImageQuant software (GE Life Sciences).

2.9 RNA-sRNA basepairing predictions:

The sRNA map database [69] was downloaded from (http://srnamap.mbc.nctu.edu.tw/download.php). The sRNA sequences were then merged into one file using MS-DOS's copy function. The resulting text file was used as the source of RNA sequences and was pasted into the ncRNA field of IntaRNA; IS*10* transposase DNA sequence was then pasted into the mRNA field of IntaRNA. The results table from INTARNA was copied into MS-Excel 2010, and sorted according to the stability of the resulting paired species

between the mRNA and sRNA. The parameters set for IntaRNA were: 37°C and a minimum seed region of 7 basepairs.

2.10 3' RNA ligase-mediated rapid amplification of cDNA ends (3' RLM-RACE):

Three prime RLM-RACE was conducted as described in [70]. NK5830 F' cells containing the Tn10 containing plasmid pBE10 was grown to exponential phase (OD_{600} = 0.6) in LB broth containing tetracycline [10 µg/mL]. Cells were grown at 37°C with aeration. Total cellular RNA of the Tn10 expressing NK5830 F' cells was extracted using the RNEasy kit (Qiagen) according to manufacturer's protocol. The RNA ligation reaction mixture contained: 2 µg of total cellular RNA,T4 RNA Ligase Buffer (1X, NEB), DMSO (10% v/v), 20 U of RNAsin (Promega), RNA adapter BM21A [1.5 ng/µL], 20 units of T4 RNA ligase (NEB), and DEPC H_2O to bring the reaction mixture volume to 50 µL. The RNA ligation mixture was incubated for 1 hour at 37°C and then cleaned up using the RNEasy kit (Qiagen) eluted in 30 µL of RNase free H_2O. The RNA was then subjected to reverse transcription PCR. Five µL of each sample containing the ligated RNA adapter was mixed with2 µL of BM19A [20 µM] and 4.5 µL of nuclease free H_2O. The reactions were heated to 75°C for 5 minutes and then immediately cooled on ice. To the reaction mix, 1 µL of Moloney Murine Leukemia Virus Reverse Transcriptase (M-MuLV RT) [200000 units/mL] (NEB), 5 µL of dNTP mix [10 mM], 2 µL of M-MuLV RT Buffer [10x] (NEB) and 0.5 µL of RNasin inhibitor was added. The reverse transcription reaction was incubated at 25°C for 5 minutes, 37°C for 1 hour, and 72°C for 10 minutes. Reaction mixtures were put through a PCR

clean up column (Qiagen) and eluted in 30 µL of nuclease free H_2O. The resulting cDNA mixture underwent a nested PCR reaction. For the first reaction 5 µL of cDNA was mixed with 1.25 µL of primers BM8A/BM19A [20 µM each], 5 µL H_2O and 12.5 µL of GoTaq Master Mix (Promega) [2X]. PCR cycling parameters were as follows: 98°C for 2 minutes for an initial denaturing, followed by 10 cycles of 98°C for 30 seconds, 64.5°C for 40 seconds, and 72°C for 2 minutes. The PCR mixture (mix 1) was diluted with 180 µL of ddH_2O. This diluted mix 1 was then used for a second round of amplification consisting of 5 µL of diluted mix, 1.25 µL of primers BM9A/BM20A [20 µM each], 5 µL of H_2O and 12.5 µL of GoTaq Master Mix (Promega). PCR cycling parameters were as follows: 98°C for 2 minutes for an initial denaturing, followed by 30 cycles of 98°C for 30 seconds, 68°C for 40 seconds, and 72°C for 2 minutes. A final extension step was performed at 72°C for 5 minutes. The PCR amplicons were gel purified using the QIAEX II Gel Extraction Kit (Qiagen) according to the manufacturer's protocol. Gel extracted amplicons were cloned into the T-vector pGEM-T-EASY (Promega) and used for electroporation of DH5α electrocompetent *E.coli* cells. Plasmids were extracted from individual clones using the GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich) and sent for sequencing using primer BM9A.

2.11 Western blots:

For Hfq semi-quantification in NK5830 F' cells and its *rpoS*$^-$ isogenic derivative in the mating out assays, individual colonies were grown in LB broth containing kanamycin [25 µg/mL] in conditions identical to the mating out assay. Three mL of cell cultures were centrifuged (1 min, maximum speed) and

resuspended in 20 µL of 1X Laemmli load dye and boiled for 5 minutes. Cell lysate was briefly centrifuged to collect lipids to the bottom of the tube. Ten µL of supernatant was loaded on a 12% Tricine SDS Polyacrylamide gel. Samples were run at 100V on the Mini-Protean II system (BIO-RAD) with 1X tricine buffer (0.1 M Tris, 0.1 M tricine and 0.1% SDS) until the samples passed through the stacking gel. The voltage was then increased to 150V and run until the bromophenol blue dye front reached the bottom of the gel. Proteins were transferred to a PVDF membrane (45 mA for 40 minutes). Hfq was detected using a rabbit polyclonal Hfq antibody (diluted 1:100,000 in 2.5% powdered non-fat milk (Nelson) in TBST (2 mM Tris pH7.5, 0.5% Tween-20, 15 mM NaCl)) and a Goat Anti-Rabbit IgG (H+L), conjugated to horseradish peroxidase (1:4000 dilution in 2.5% powdered non-fat milk (Nelson) in TBST) (Pierce). PVDF membranes were imaged using the Storm 860 Scanner. PVDF membranes were then stained by Coomassie stain, and a picture was obtained for sample normalization on an Alpha Imager. Western blot and coomassie images were analyzed for band intensity using ImageJ software (NIH).

2.12 Verification of sRNA induction:

Plasmids containing IPTG inducible sRNAs (pDH694-696, pDH673), were verified to induce via IPTG by extracting RNA from NK5830F' *E. coli* in the presence or absence of IPTG RNA was extracted from exponentially growing NK5830F' cells (OD_{600}=0.6) and isolated using the hot phenol extraction method as described in [71]. RNA was quantified using a spectrophotometer and RNA purity was ensured by ensuring an OD260/280 ratio of ≥1.8 and an

OD260/OD230 ratio ≥1.5. RNA samples were also checked for integrity by resolving 5S, 16S and 23S ribosomal RNA molecules on a 1% agarose gel. Reverse transcription PCR was performed using M-MuLV reverse transcriptase (NEB), Taq DNA polymerase (Promega), 1 µg of cellular RNA and primers BM37A/BM38A [2 µM each] for MicC and BM30A/BM32A [2 µM each] for GlmZ. The primer pairs for 16S rRNA were BM28A/BM29A. PCR was performed for MicC using the following parameters: 98°C, 30 s; 55°C, 45 s; 72°C, 60 s and cycled 19 times. The 16S rRNA PCR parameters are the same as above except that the annealing temperature was 62°C and cycled 7 times. The GlmZ PCR parameters were the same as above except the annealing temperature 60°C and cycled 11 times. To ensure PCR was within linear range, 2 more subsequent PCR cycles were done on each transcript (1 at a time) and the resulting amplicon approximately doubled in intensity.

Chapter 3 - Results

3.1 How common is the Hfq mediated repression of transposase expression in IS Elements?

Hfq is negative regulator of transposase gene expression in both IS*10* and IS*50* elements [18,51]. In this work I asked the question of how common is it for a transposase to be regulated by Hfq? To start addressing this question, I had to devise a method of how to systematically identify known IS Elements for potential Hfq regulation. One method of doing this would be to individually clone random IS Elements into an expression vector and look at either transposase expression levels or transposition frequencies in the presence or absence of Hfq. One issue with this method is the possibility that it may be difficult to express IS Elements from certain organisms in *E.coli*. To overcome this expression issue, a collection of different bacterial species with the appropriate *hfq* gene disruptions would be required. The other issue is that some of these species may require growth media and conditions that cannot be achieved with the equipment available in the laboratory. To prevent these possible technical difficulties, I chose to search the IS Finder database [72] for transposase genes that contain putative Hfq binding sites to give a list of candidates that may be Hfq regulated. This list is a starting point to determine which elements to then test for transposition activity in the presence and absence of Hfq. A possible caveat of this approach is that it is using Hfq binding motifs determined for *E. coli* Hfq. The possibility remains that other organisms' Hfq may have different binding preferences. Studies have been done on how effective interspecies Hfq can complement Hfq deletions in *E.coli*.

One can presume that if a xenotropic *hfq* gene can complement an *hfq* deletion mutant, the two proteins are acting in similar manner. A study by *Schilling et al*, shows that one can complement an *E.coli* strain MC4100 AM111 *hfq* deletion with the *hfq* gene from *Acinetobacter baylyi* to relieve the slow growth phenotype in LB media. In the *Acinetobacter* case, the Hfq protein is larger than the *E.coli* homologue (174 amino acids verses 102 amino acids respectively) with an elongated glycine rich C-terminus[73], yet still can complement the growth defects of an *E. coli* Hfq deletion. Another study by *Sonnleitner et al.* involving *Pseudomonas aeruginosa rpoS* gene activation (which is an Hfq dependent process), showed that *Pseudomas'* Hfq can functionally complement *rpoS* gene activation defects in *hfq* disruption strains of *E.coli. RpoS* gene activation and *rpoS* independent pathways that are affected by Hfq deletions were all restored when expressing the *Pseudomonas* Hfq homolog in *E.coli* [74]. This database searching strategy is a little more simplified than the successful approaches used for sRNA discovery that consist often of searching intergenic regions of bacterial genomes for orphan promoters. For example, a study by *Chen et al*. searched *E.coli* intergenic regions for genes between a σ^{70} promoter and a Rho-independent terminator. These searches resulted in 227 candidates to be sRNA genes [75].

The goal of my bioinformatics approach was to identify potential Hfq binding sites in IS Elements which may imply binding to and regulation of the element. To do this, all IS Elements submitted to the IS Finder database [72] were compiled into a series of files which were then searched for the two Hfq binding

sequences AAYAAYAA and AYAATAA. These consensus sequences were identified by *Lorenz et al* in a genomic SELEX screen for Hfq binding RNAs [44], and are consistent with the ARN_x motif of Hfq previously identified by *Link et al* to bind the distal face [41]. It is important to note that Hfq binding is sequence and structurally dependent. Hfq often binds motifs consisting of U rich regions upstream of a single or double hairpin [76]. The IS Finder database contained 3478 active IS entries (as of February 2011). Of these 3478 entries only 226 of the entries were actually tested to be active IS Elements. Both AAYAAYAA and AYAATAA motifs were found at a variety of positions within the IS Elements' transposase transcripts. For the AAYAAYAA motif, 3 of the motifs are in the 5' region of the transcripts, 7 of the motifs are in the middle region of the transcripts, and 5 of the motifs are in the 3' region of the transcripts **(Table 3.1**) for a total of 14 identified sequences. The family distribution of AAYAAYAA containing IS Element transposases' include IS*3*, IS*4*, IS*21*, IS*30*, IS*256* and lS*l3*. For the AYAATAA motif, 6 of the motifs are in the 5' region of the transcripts, 6 of the motifs are in the middle region of the transcripts, and 6 of the motifs are in the 3' region of the transcripts **(Table 3.2**) for a total of 18 identified sequences. The family distribution of AYAATAA containing IS element transposases include IS*3*, IS*4*, IS*5*, IS*21*, IS*30*, IS*110*, IS*630*, and IS*1380*. Since, IS Elements have different DNA base compositions, I also calculated what the probability of finding the number of SELEX sites in each element based on their length and DNA base composition. The binomial cumulative distribution function of Microsoft Excel 2010 was used to calculate the probability of finding exactly the number of

putative Hfq binding sites that were reported in each element (See Appendix I for sample calculations.)

Using this bioinformatics approach, I have identified 32 IS Element transposase genes that may be under the regulation of Hfq using putative Hfq binding sites. This provides a list of IS Element transposase genes that are candidates for further characterization of their regulatory sequences and interactions with Hfq.

3.2 Hfq regulation of IS*1413:*

To verify the effectiveness of my bioinformatics search to predict Hfq regulation in IS Elements the effect of Hfq on IS*1413* transposase expression was tested. Our fourth year honours student Morgan Black used my bioinformatics analysis as the basis to perform electrophoretic mobility assays (EMSA) to demonstrate that both *E.coli* and *Burkholderia* Hfq bind IS*1413*'s transposase transcript *in vitro*. She also showed using an IS*1413* transposase-LacZ translational fusion, that the expression of the IS*1413* transposase is 17.8 fold higher in *hfq*$^{--}$ *E.coli* cells than in *E.coli* cells encoding a functional *hfq* gene. In this section of the thesis I address 2 questions arising from Morgan's initial observations:

1. Is the Hfq repression of IS*1413* transposase expression dependent on the AAYAAYAA sequence found in the 5' UTR?
2. Does Hfq repress IS*1413* transposition in *E.coli*

Table 3.1 Elements in the IS Finder database containing the predicted Hfq binding sequence AAYAAYAA. IS Elements were merged into one file from the IS Finder Database (http://www-is.biotoul.fr/) utilizing the WebCrawler Web Sphinx [77]. Relevant fields were extracted utilizing the find and replace functions in Microsoft Word 2010, and finally the location of the SELEX defined Hfq binding site was determined. This information is current as of February 2011. The P-Value represents the probability that one would find exactly the number of binding sites found for the given element (the probability mass function) and was calculated using Microsoft Excel 2010. A probability of 1 would represent a SELEX site that is probably unimportant; as it appears the number of times one would expect it to. A probability of 0 would represent a SELEX site that you would expect to see either much more or less than one observed. See Appendix I for P-value calculations. SELEX Locations are denoted in the far right column preceded by the number of SELEX sites found in the IS Element's transposase mRNA sequence.

IS Family	IS Element	Length bp	GC Content	P-Value	Organism	SELEX Location
IS3	IS1223	1494	32%	3.23×10^{-1}	*Lactobacillus johnsonii*	(1) Middle
IS3	IS1550	1408	24%	6.81×10^{-2}	*Mycoplasma fermentans*	(3) Middle
IS3	IS199	1220	43%	1.38×10^{-1}	*Streptococcus mutans*	(1) 5' end UTR
IS3	ISMmy2	1374	28%	3.54×10^{-1}	*Mycoplasma mycoides*	(1) 3' end
IS4	IS*10*R	1329	44%	1.36×10^{-1}	*Salmonella typhimurium*	(1) 3' end
IS4	IS186A	1341	51%	6.71×10^{-2}	*Escherichia coli*	(1) 3' end
IS4	IS186B	1338	51%	6.70×10^{-2}	*Escherichia coli*	(1) 3' end
IS21	IS21	2131	52%	9.17×10^{-2}	*Pseudomonas aeruginosa*	(1) Middle
IS21	IS232A	2184	34%	2.86×10^{-2}	*Bacillus thuringiensis*	(1) Middle (2) 5' end of ORF2
IS21	ISEc12	2484	48%	1.58×10^{-1}	*Escherichia coli*	(1) Middle
IS30	ISCg2	1636	53%	6.40×10^{-2}	*Corynebacterium glutamicum*	(1) Middle
IS256	IS*1413*	1428	60%	2.22×10^{-2}	*Burkholderia phenoltripix*	(1) 5' UTR
IS256	IS905A	1313	38%	2.17×10^{-1}	*Lactococcus lactis*	(1) Middle
IsL3	IS1181	1513	32%	9.45×10^{-2}	*Staphlococcus aureus*	(2) 3' end

Table 3.2 Elements in the IS Finder database containing the predicted Hfq binding sequence AYAATAA. IS Elements were merged into one file from the IS Finder Database (http://www-is.biotoul.fr/) utilizing the webcrawler Web Sphinx [77]. Relevant fields were extracted utilizing the find and replace functions in Microsoft Word 2010, and finally the location of the SELEX defined Hfq binding site was determined. This information is current as of February 2011. The P-Value represents the probability that one would find exactly the number of binding sites found for the given element (the probability mass function) and was calculated using Microsoft Excel 2010. A probability of 1 would represent a SELEX site that is probably unimportant; as it appears the number of times one would expect it to. A probability of 0 would represent a SELEX site that you would expect to see either much more or less than one observed. See Appendix I for P-value calculations. SELEX Locations are denoted in the far right column preceded by the number of SELEX sites found in the IS Element's transposase mRNA sequence.

IS Family	IS Element	Length (bp)	GC Content	P-Value	Organism	SELEX Location
IS3	IS1138	1288	23%	2.59E-01	*Mycoplasma pulmon*is	(1) 3' end
IS3	IS1520	1302	38%	3.24E-01	*Lactobacillus sakei*	(1) 5' UTR
IS3	IS1550	1408	24%	1.00E-01	*Mycoplasma fermentans*	(3) Middle (1) 3' end
IS3	ISMmy2	1374	28%	3.36E-01	*Mycoplasma mycoides*	(1) 3' end
IS4	IS*10*R	1329	44%	2.32E-01	*Salmonella typhimurium*	(1) 3' end
IS4	IS4Bsu1	1406	40%	3.06E-01	*Bacillus subtilis*	(1) 5' UTR
IS4	ISBth4	2024	35%	3.62E-01	*Bacillus thuringiens*is	(1) 5' UTR
IS5	IS903B	1057	53%	8.09E-02	*Escherichia coli*	(1) 5' end
IS5	ISC1058	1058	43%	2.13E-01	*Sulfolobus solfataricus*	(1) Middle
IS21	IS232A	2184	34%	3.45E-01	*Bacillus thuringiens*is	(1) Middle of ORF 1
IS30	ISAs2	1084	59%	3.84E-02	*Aeromonas salmonicida*	(1) 5' UTR
IS30	ISCg2	1636	53%	1.20E-01	*Corynebacterium glutamicum*	(1) Middle
IS110	IS492	1202	42%	2.49E-01	*Pseudomonas atlantica*	(1) 5' UTR
IS630	IS630	1153	52%	9.82E-02	*Shigella sonnei*	(1) 3' end
IS1380	IS942	1598	40%	3.25E-01	*Bacteroides fragilis TAL3636*	(1) 3' UTR
IS1380	ISEc9	1656	33%	3.63E-1	*Escherichia coli*	(1) Middle
IsH3	ISC1439A	1439	39.00%	3.24E-01	*Sulfolobus solfataricus PH1*	(1) 3' end
IsL3	IS1181	1513	32.00%	2.11E-01	*Staphylococcus aureus BM3121*	(1) Middle (1) 3' End

To address the first question, two mutations were made in the SELEX site of IS*1413*'s transposase gene. Mutations were made to disrupt the SELEX site or secondary structure around the SELEX site. Amino acid changes or codon bias was not considered in these mutations, as the SELEX site is in the 5' UTR of IS*1413* transposase. Mutant B contained an adenine to a guanine mutation at position 108 and an adenine to a cytosine mutation at position 113 **(Figure 3.1 B, sequence B).** Mutant C contained a cytosine to thymine mutation at position 107 and an adenine to a thymine mutation at position 112 **(Figure 3.1 B, sequence C).** To determine the effect of IS*1413* transposase's SELEX site mutations on Hfq regulation, transposase expression was determined in cells containing an intact *hfq* gene as well as isogenic cells containing an *hfq* disruption. Transposase expression was measured through the use of transposase-LacZ translational fusions. The first 445 base pairs of IS*1413* and the two mutants were fused to the sequence encoding the ninth codon of LacZ. These mutants, as well as the wild type sequence were then put under control of an arabinose inducible promoter (p-Ara) to increase transcription and transposase expression was tested in both hfq^+ and hfq^- backgrounds. For the wild type construct, there is a 3 fold increase in expression (i.e. β-galactosidase activity) in hfq^- cells when comparing isogenic hfq^+ and hfq^- strains **(Figure 3.1A).** In comparison, both mutant constructs gave an increase in transposase expression of around 10-fold in both genetic backgrounds **(Figure 3.1A).** An increase of transposase expression in the mutant constructs was surprising, as the hypothesis predicted that there only should have been an increase in activity in hfq^+ cells. The overall increase in transposase expression may be explained by the mutations providing

a better ribosome binding site than the wild type sequence. To date, the actual ribosome binding site of IS*1413* has not been characterized, but it would likely be around the proximity of the SELEX site. Another possibility is the mutations may increase either steady state levels of the transcript or increase the half-life of the transcript. Mutant B shows a 3 fold increase in activity in the *hfq*$^{-}$ strain relative to the *hfq*$^{+}$ strain **(Figure 3.1A).** Mutant C shows a 1.5 fold increase in activity in the *hfq*$^{-}$ strain relative to the *hfq*$^{+}$ strain. Overall, Hfq has an effect on expression of IS*1413*, however, more mutant forms of IS*1413* will have to be tested to hone into the sequence that Hfq is potentially binding to. There is a possibility that other Hfq binding sites may exist in other regions of the IS*1413* transposase transcript which may be the reason why mutation analysis did not abolish Hfq repression. Another possibility is that the Hfq effect is actually indirect. For example, Hfq is a known regulator of many genes within bacteria such as *rpoS* (a gene encoding a sigma factor often associated with stress responses) [78], and it could be a gene under RpoS' control that is the cause of the increased transposase expression in *hfq*$^{-}$ cells. These remain possible avenues for future investigation.

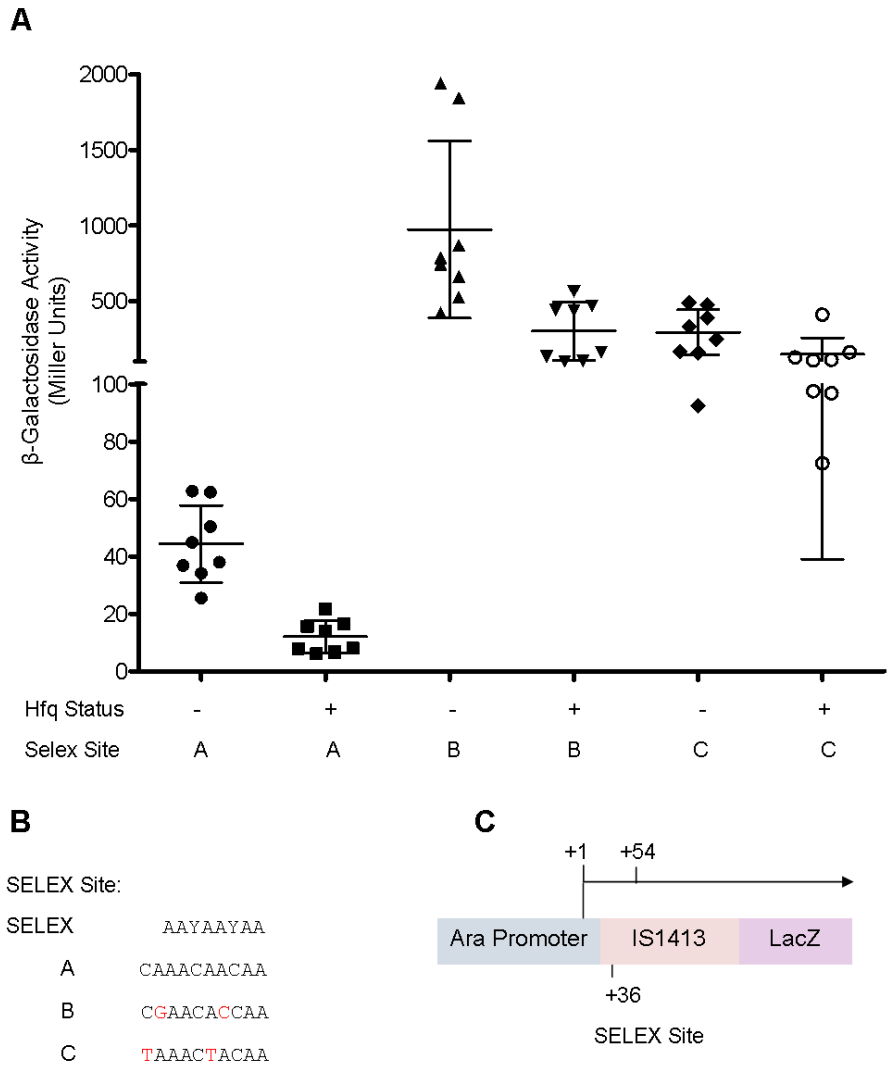


Figure 3.1 β-galactosidase activity for IS*1413*-LacZ translational fusions. Translational fusions were transformed into NK5830 F' (Hfq^+) or its isogenic Hfq^- derivative and Miller assays were performed. This data shows the means and standard deviations of at least 2 independent experiments. **B**. Mutations were made in the putative Hfq binding region (SELEX) of the IS*1413* LacZ fusion under the control of an arabinose inducible promoter. Construct A shows the wild type sequence of *1413*. Sequence B and C show the mutations made, with the affected nucleotide highlighted in red. **C**. A schematic of the *1413*-LacZ portion of the β-galactosidase construct**.**

To test the transposase activity of IS*1413* in *E. coli* and to determine if the Hfq status affects the activity, a 'marked' IS*1413* element (IS*1413*-Kan) was made by inserting a kanamycin resistance gene between the stop codon of the transposase and the adjacent transposase end of IS*1413* (see Materials and Methods). NK5830 F' cells were transformed with a plasmid containing IS*1413*-Kan and a mating out assay was performed (see Materials and Methods). Briefly, the mating out assay involves transforming Hfq^+ or Hfq^- NK5830 F' *E.coli* cells (Arg^-/F' *laciq* L8 *pro)* with a plasmid containing a kanamycin 'marked' IS*1413* element. Transformants (donors) and HB101 (leu^-, str^R, F^-) cells (recipients) were grown overnight on LB agar with kanamycin or LB agar with streptomycin respectively at 37°C. Individual colonies of donors were used to inoculate LB broth containing kanamycin and grown overnight at 37°C with aeration (250 rpm). Individual recipient colonies were also picked and grown overnight at 37°C with aeration (250 rpm) in LB broth containing streptomycin. Overnight cultures of donors and recipients were individually subcultured into selection free LB media. Donors were grown for 2 hours (4 hours for Hfq^-) at 250 rpm at 37°C, then culture tubes were transferred to a slower aerator at about 50 rpm to facilitate sex piIli formation for 2 hours (4 hours for Hfq^-) . Recipient cells were grown for 4 hours at 37°C at 250 rpm. Recipient and donor cells were mixed together at a 5:2 ratio and allowed to undergo conjugation for 1 hour at 37°C at 50 rpm. Cell mixtures were then vortexed, centrifuged (4000 rcm, 4 minutes), and resuspended in 0.85% saline. To determine the number of successful transfers through conjugation of the F-factor (which can be tracked by the proline gene), cells were plated on M9 agar plates containing glucose, streptomycin, leucine and thiamine.

The number of transposition events is measured by selecting for cells that have had the marked IS*1413* element transpose onto the F-factor. Recipients containing the F-factor with an IS*1413* element are selected for by plating on M9 agar plates containing glucose, streptomycin, kanamycin, leucine and thiamine. The transposition frequency of IS*1413* is calculated by dividing the number of IS*1413* transposition events by the total number of successful F-factor transfers. IS*1413* has an average transposition frequency in NK5830 F' *E.coli* cells of 3.7 x 10^{-5} ± 7.8 x 10^{-5} transposition events per mL of mating mix (in the presence of Hfq) and 5.0 X 10^{-4} ± 3.8 X 10^{-4} transposition events per mL of mating mix (in the absence of Hfq) (**Figure 3.2**). Thus, Hfq has a repressive effect on IS*1413* transposition, down regulating transposition events by approximately 13 fold. When comparing the medians of the two conditions, the trend of Hfq repression continues. The median of IS*1413* transposition frequency in an Hfq^+ background being 7.85 x10^{-6} and the median of IS*1413* transposition frequency in an Hfq^- background being 4.85 x 10^{-4} resulting in a 61.7 fold increase (by median) of IS*1413* transposition frequency in *hfq*$^-$ cells relative to *hfq*$^+$.

3.3 Alternative pathways of IS*10* regulation by Hfq at the 3' end of the mRNA transcript.

The 3' region of IS*10* transposase also contains the AAYAAYAA motif as identified in section 3.1. Most regulatory studies of IS*10* have been focused on the 5' region of the transposase gene. Hfq has already been established as a negative regulator of IS*10* transposition [18]. This section of the thesis explores the

potential contribution the IS*10* transposase's 3' putative Hfq binding site has in Hfq repression.

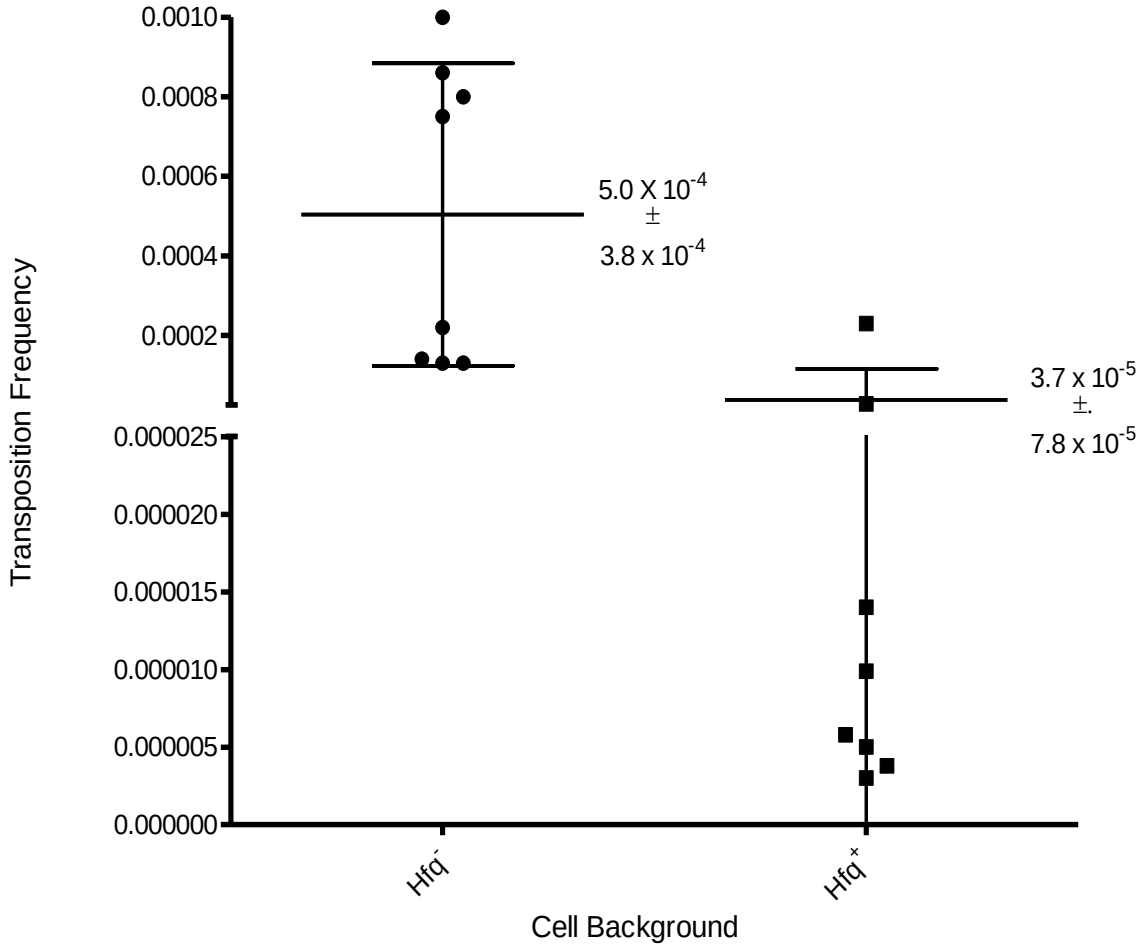


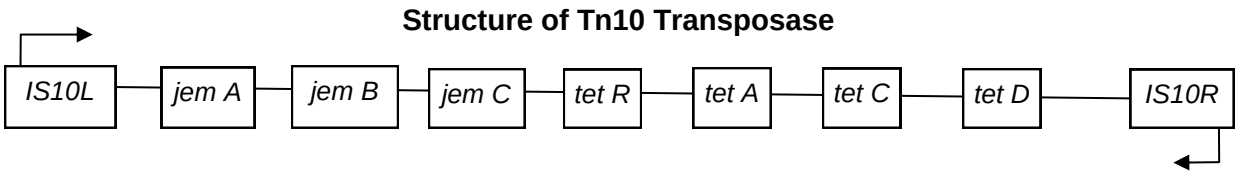
Figure 3.2 Hfq represses IS*1413* transposition. NK5830 F' and its Hfq$^-$ isogenic strain (DBH16) was transformed with a plasmid containing a kanamycin marked IS*1413* element and grown overnight on LB medium with kanamycin [25 µg/mL]. Cells were subcultured in LB to an OD_{600} of ~ 0.6 and the mating out procedure was followed. Total exconjugates and total hops were calculated and transposition frequencies were calculated by dividing the total hops by the total exconjugates. This data shows the means and standard deviations from 2 independent experiments both with n=4.

One proposed mechanism of Hfq regulation on IS*10*, is that Hfq accelerates the pairing of RNA-IN and RNA-OUT (see introduction) [18]. One issue still remains with this proposed mechanism; when you disrupt *hfq* in a cell harbouring a single copy of IS*10* (where the effect of RNA-OUT is thought to be negligible), Hfq still exhibits a repressive effect on IS*10* transposition [18]. An observation from our laboratory technician, Ryan Trussler, is that transposase mRNA stability increases in the absence of Hfq. The half-life of RNA-IN in the presence of Hfq is 0.79 minutes versus 1.1 minutes in the absence of Hfq. These observations suggest that Hfq plays a role in the stability of RNA-IN. There are at least two known mechanisms for how Hfq can alter the stability of RNA. Polyadenylation of mRNA by poly(A) polymerase I (PAP) in *E. coli* is stimulated by Hfq. *In vivo, hfq* gene deletions result in shorter polyA tails on the *rpoS* mRNA transcript [79]. Polynucleotide phosphorylase (PNP), is an 3 ' → 5' exonuclease that preferentially degrades RNA with 3' polyadenylated or polyurindinylated tails *in vitro* [80]. Taken together, these observations suggest that Hfq can facilitate the polyadenylation of RNA transcripts resulting in polynucleotide phosphorylase mediated degradation [81]. Another Hfq mediated RNA degradation mechanism involves Hfq, sRNAs, and RNase E. *Morita et al* demonstrated that RNase E co-purifies with Hfq. They also showed that Hfq in conjunction with an sRNA (in their case RhyB) is able to target the RNase E-Hfq complex to a target mRNA (*sodB*) resulting in mRNA degradation [81]. I sought to determine if Hfq is working through either of these two mechanisms to promote IS*10* mRNA turnover.

Hfq binding sites on the 3' end of mRNA transcripts have been associated with promoting the turnover of mRNA transcripts through increased

polyadenylation which facilitates PNP-dependent degradation. Since this mechanism depends on the transcript of interest to contain a polyadenylated tail, I sought to define the 3' end of the IS*10* transcript and establish the presence or absence of a polyA tail. The transcriptional stop points of IS*10* transposase have not previously been defined, nor has it been established if a polyA tail is present on the transcript. I employed the 3' RLM-RACE technique to analyze the 3' end of the IS*10* transposase transcript. The source of transposase RNA for this experiment was a plasmid (pBE10) encoding the Tn10 transposase. Cells containing pBE10 were grown to exponential phase (OD600 ~ 0.6). Total cellular RNA was isolated and a RNA-adaptor (BM21A) (GTGATGGGCGATGAATGAACACTGCGTT) was ligated to the total cellular RNA molecules using T4 RNA Ligase (NEB). RNA underwent reverse transcription, followed by a nested PCR reaction. Briefly, the nested PCR reaction consisted of the first PCR reaction using primers BM8A and BM19A for 10 cycles of amplification. This first PCR reaction was diluted ten-fold in H_20 and then 5 µL of the first round of amplification was used as a template for the second round of amplification. The second round of PCR amplification used primers BM9A and BM20A which anneal closer together than primers BM8A and BM19A to reduce the chance of non-specific amplification. The second round of PCR amplification was done for 30 cycles. Specific details of the PCR reactions are described in the Materials and Methods section. PCR products were run out on an agarose gel, gel purified, cloned into the pGEM T-easy vector (Promega) and transformed into DH5α. A total of 51 clones were sent for sequencing. Twenty-seven of the clones had IS*10*R sequence, 13 of the clones had IS*10*L sequence

and 11 of the clones did not contain IS*10* sequence. Analysis of the IS*10*R sequences reveals a favoured transcription stop point in the *tetD* gene at bp 7718 (relative to bp 1 of Tn*10*) with no signs of polyadenylation **(Figure 3.3**). Analysis of the IS*10*L sequences shows that there appears to be a favoured transcription stop site upstream of the *jemA* gene at bp 2750 relative to bp 1 of Tn*10* and no polyadenylation.



3’ RLM-RACE Alignment of IS*10*R Sequences

3’ RLM-RACE Alignment of IS*10*L Sequences

Figure 3.3 Three prime RLM-RACE studies show that IS*10* transcripts from Tn*10* have defined stop points, with no indication of polyadenylation or Rho-independent termination sequences. Total RNA was isolated from NK5830F’ cells expressing Tn*10*. Complementary DNA was synthesized using M-MLV Reverse Transcriptase (NEB) and IS*10* mRNA was PCR amplfied. PCR amplicons were cloned into pGEM T-Easy vector (Promega) and sequenced. Sequences were aligned and compared in MS Word 2010.

Overall IS*10* has a defined transcriptional stop with no indications of a Rho independent terminator (verified by ARNold terminator predictor [82]). There is also no polyadenylation of the IS*10* transcript. Interestingly, there is a predicted Rho dependent terminator early on in the IS*10* (L/R) transcript at bps 357-396 relative to the Tn*10* DNA sequence with a predicted stability of -9 Kcal/mol (ARNold terminator predictor) [82], which would not be picked up with the primers used in this method. With the absence of any polyadenylation on the IS*10* transposase mRNA, it is unlikely that the reduced half-life in the presence of Hfq is due to neither the recruitment of PAP nor the PNPase degradation pathway. The Hfq-RNase E mediated degradation of IS*10* transposase mRNA still remains a possibility and is explored in the next section.

3.4 Does Hfq bind the 3' end of IS*10*?

Absence of polyadenylation on the IS*10* transposase mRNA rules out the PAP-PNP model of RNA degradation. I considered the alternative model that an Hfq-RNase E complex might bind the 3' end of the IS*10* transposase transcript resulting in the degradation of the mRNA. I sought to determine if Hfq is actually able to bind the 3' end of the IS*10* transcript. Furthermore, since this mechanism of RNA degradation would be disrupted if Hfq was unable to bind the transcript, I looked at the effects of mutating the putative Hfq binding site. The 3' region of RNA-IN (bp 1222-1329) was PCR amplified with a T7 promoter included in the forward primer (see **Figure 3.4**). This amplicon was used as a template for 'run off' *in vitro* transcription using T7 RNA polymerase and rNTPs including ^{32}P α-CTP.

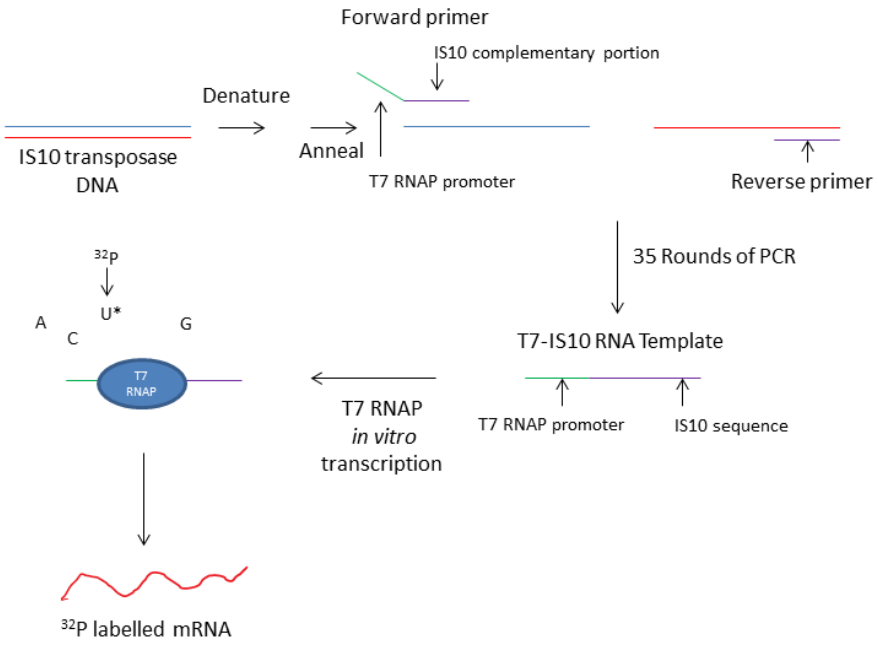


Figure 3.4 *In vitro* transcription schematic. A T7 RNA Polymerase (RNAP) promoter is incorporated in a forward primer for PCR amplification of the gene of interest. After purification of the PCR amplicon, T7 RNA Polymerase is incubated with rNTPs (including ^{32}P labelled rUTP) to make the resulting radio-labelled RNA molecule. Unlabelled RNA molecules can be made by substituting the ^{32}P labelled rUTP with an unlabelled version.

Labelled RNA was isolated by loading on a 10% denaturing polyacrylamide and the prominent RNA species was cut out and eluted overnight in elution buffer. Eluted RNA was tested for purity by spectrophotometer and integrity on another 10% denaturing polyacrylamide gel. RNA preparations were quality controlled to ensure they gave a clean band on the denaturing gel. Hfq binding was determined by EMSA. Competition assays were used to follow up initial Hfq binding assays to define which face of Hfq the 3' end of IS*10* transposase mRNA and its mutants bind. DsrA is a sRNA known to bind the proximal face of Hfq and polyA RNAs bind Hfq at the distal face [83]. By preincubating Hfq with either DsrA or polyA$_{16}$ RNA and then adding the 3' end of the IS*10* transposase mRNA, we

can determine if there is any competition for Hfq binding sites. Competition for Hfq binding provides evidence of which face the non-competing RNA is binding. For example, if one is unable to see a displacement of DsrA from Hfq, even in very high concentrations of RNA-IN, it is very likely that RNA-IN binds the distal face instead of the proximal face (where DsrA is) of Hfq. Knowing which face of Hfq the 3' end of RNA-IN binds, may help us elucidate possible roles the RNA-Hfq complex may have in the cell. For example, if the 3' end of RNA-IN binds the proximal site of Hfq (similar to how the 5' end binds), a potential explanation for its role could be that it chelates out Hfq from acting on the 5' end of the transcript (see Discussion).

The wild type sequence is predicted by Mfold [84] to form a stem loop right before and after the SELEX site (**Figure 3.5A)**. Hfq binds RNA-IN 3' with an approximate K_D between 6.08 and 12 nM, as suggested by over 50% of free RNA being bound by Hfq between Hfq concentrations of 6.08 nM and 12 nM **(Figure 3.6A +B).** The wild type sequence binds Hfq presumably on the proximal site as shown by it outcompeting DsrA (an sRNA known to bind the proximal sites of Hfq) in competition studies until DsrA is in 25 fold excess to RNA-IN **(Figure 3.7A).**

To test the specificity of the Hfq binding site on the 3' end of RNA-IN mutant, I looked at two different mutations. Since I eventually wanted to look at the effect of Hfq binding disruption *in vivo,* mutations were constrained to be silent and have little effect on codon bias in *E. coli.* I also considered potential positions around the SELEX site that may affect the secondary structure context of the site.

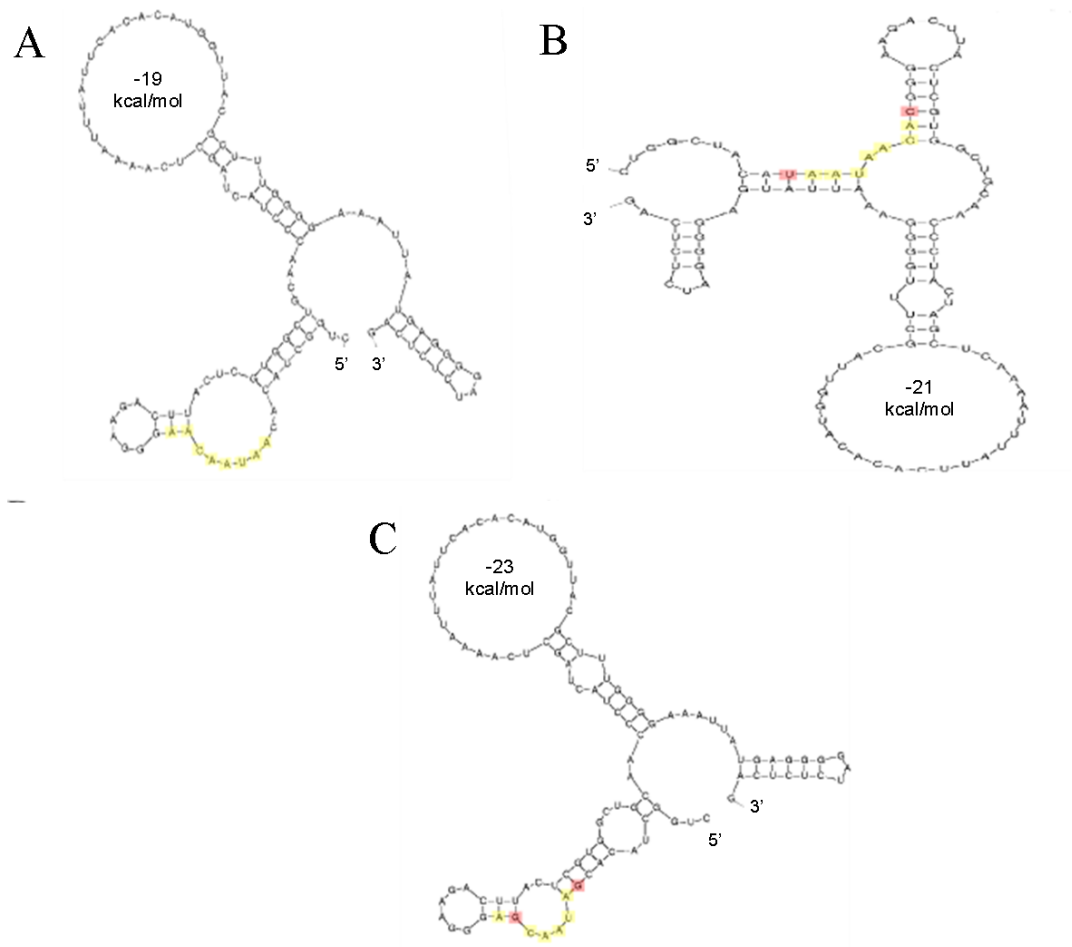


Figure 3.5 The proposed structure for each of the RNA-IN 3' species. RNA structures were predicted using Mfold [84] and their predicted stabilities are noted within the structures. The proposed Hfq binding site is highlighted in yellow and any mutations are highlighted in red **A.** RNA-IN 3' (WT). **B.** RNA-IN 3' (BM1) **C.** RNA-IN 3' (BM2)

The first mutant BM1 has a C1229T mutation (found just outside of the SELEX site) and an A1239C mutation (which disrupts the last A of the AAYAAYAA motif to a C). The BM1 mutant has its modified SELEX site on a predicted single stranded region close to an RNA-hairpin (**Figure 3.5B** shown with changed bases highlighted in red). The BM1 mutation resulted in an increase in apparent K_D relative to the WT sequence, with an approximate K_D

between 314-614 nM **(Figure 3.6C).** The BM1 mutated RNA-IN likely binds the proximal site of Hfq due to the inability of polyA$_{16}$ RNA (an RNA known to bind the distal site of Hfq) to outcompete the RNA-Hfq complex 2 formation at a 4.3 fold excess. DsrA starts to compete with RNA-IN 3' BM1 at a 4.3 fold excess, as shown by the loss of RNA-IN 3' BM1 Complex II **(Figure 3.7B).**

The RNA-IN 3' mutant BM2 has an A1232G mutation; (mutates the first A of the AAYAAYAA motif to a G) and an A1238G (mutates the second last A of the AAYAAYAA Motif to a G). The BM2 mutant has its modified SELEX site on a predicted single stranded region between 2 hairpins (shown in **Figure 3.5C** with changed bases highlighted in red), has its modified SELEX site on a single stranded region flanked directly by 2 hairpins and on an extend arm compared to the other two RNA species. The BM2 mutation resulted in an apparent K_D between 80-160 nM **(Figure 3.6D).** RNA-IN 3' BM2 appears to bind the distal side of Hfq, suggested by the competition with only a 0.5 excess of polyA$_{16}$ and no competition with DsrA at a 10 fold excess **(Figure 3.7C).** Hfq does specifically bind the 3' region of RNA-IN, and this specificity is sequence and possibly structurally dependent (based on mfold predictions) as shown by the modulation of binding affinity through the BM1 and BM2 mutations.

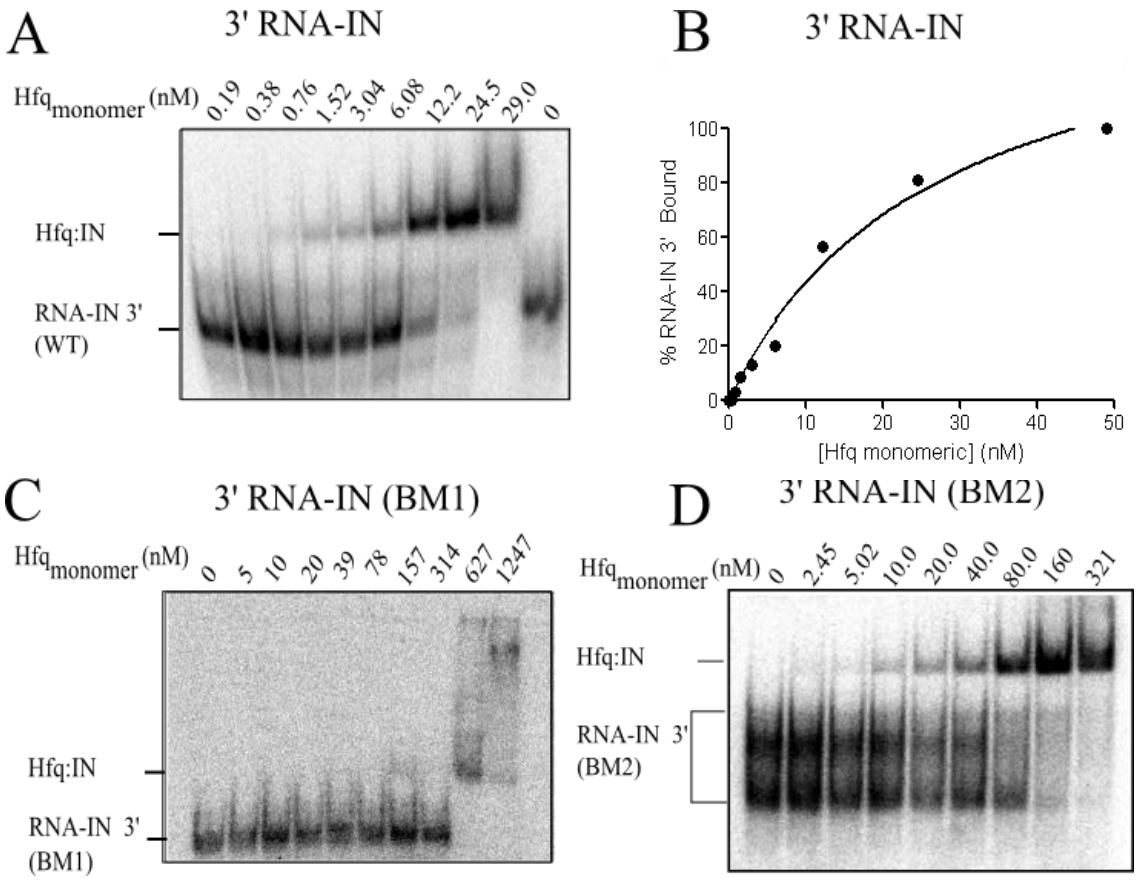


Figure 3.6 Hfq binds the last 107 nucleotides of IS*10* and mutating the SELEX site reduces binding. ^{32}P labelled RNA-IN 3' and its mutants were mixed with varying Hfq concentrations and incubated at room temperature for 10 minutes. Reactions were resolved on a 6% native polyacrylamide gel. [RNA-IN3'] = 1.52 nM/Reaction, [BM1] = 3.45 nM/Reaction, [BM2] = 2.15 nM/Reaction. B. The binding curve of Hfq-3' RNA-IN was plotted using Prism 5.0 software and fit to the equation Y=Bmax*X/ (Kd + X). Where Y = % of 3' RNA-IN bound to Hfq, Bmax = the minimum concentration of Hfq in which 100% of 3' RNA-IN is bound, X = the concentration of Hfq, and K_D = the concentration of Hfq_1 (in nanomoles/L) where 50% of 3' RNA-IN is bound.

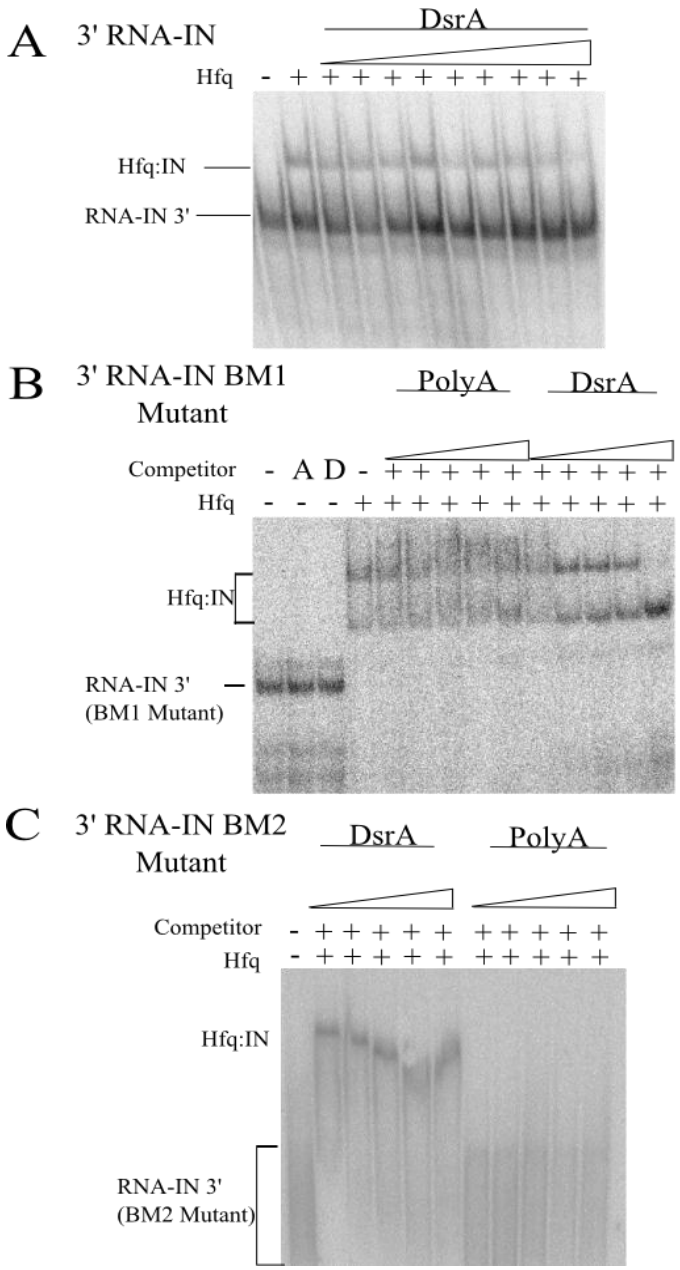


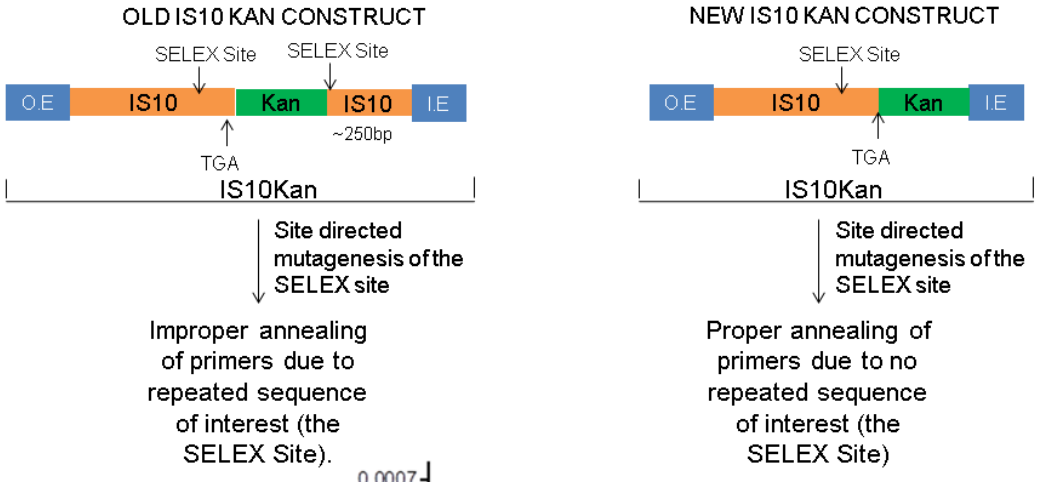
Figure 3.7 Hfq binds the last 107 nucleotides of IS*10* and binding can be competed using the proximal binding sRNA DsrA or the distal binding polyA$_{16}$ RNA. A. ^{32}P labelled RNA-IN 3' [1.52 nM/reaction] and was mixed with varying concentrations (0.154 nM, 0.308 nM, 0.616 nM, 1.32 nM, 2.46 nM, 4.93 nM, 9.86 nM, 19.7 nM, 39.4 nM) and with an Hfq concentration of 6.08 nM/reaction and incubated at room temperature for 10 minutes. EMSAs were performed on a 6% native polyacrylamide gel. **B**. ^{32}P labelled RNA-IN 3' (BM1) [3.45 nM/reaction] was mixed with varying DsrA or polyA$_{16}$ concentrations (3 nM, 6 nM, 12 nM, 15 nM) and with an Hfq concentration of 1.05 µM/reaction and EMSAs were performed on a 6% gel. **C.** ^{32}P labelled RNA-IN 3' (BM2) [2.15 nM/reaction] was mixed with varying DsrA or A$_{16}$ concentrations (1.3 nM, 2.5 nM, 5 nM, 10 nM, 20 nM) and EMSAs were performed.

3.5 Does the BM1 mutation affect IS*10* transposition frequency?

Since the 3' mutations reduce Hfq binding and Hfq binds the 3' end of RNA-IN tightly, the *in-vivo* effects of the 3' mutations in a mating out assay were tested. A new marked IS*10* construct had to be made to ensure a unique 3' region of IS*10* was present for site directed mutagenesis. The construct that was available in the lab (pDH502/pDH602) had a large portion of the 3' region of IS*10* (including the SELEX site) repeated following the kanamycin gene **(Figure 3.8A).** This repeated region resulted in amplification issues when trying to do site directed mutagenesis to the SELEX site. The new construct included the IS*10* sequence up to the stop codon, followed by a kanamycin gene transcribed off the opposite DNA strand followed by the remainder of the inside end of IS*10* to allow for transposition **(Figure 3.8A).** The IS*10* portion was mutated to make the BM1 mutation which was chosen due to its similar Hfq binding profile to the wild type transcript (likely binding the proximal site of Hfq, unlike the BM2 mutant which showed a different binding preference).

A mating out assay was performed with NK5830 cells transformed with either an IS*10*-Kan element containing the native IS*10* sequence or containing the BM1 mutation. There was no difference in transposition frequencies between either of the constructs **(Figure 3.8B).** These results suggest that the 3' SELEX site of IS*10* transposase RNA is not a region of regulation for RNA-IN 3' in the environmental conditions of the mating out assay.

A



B

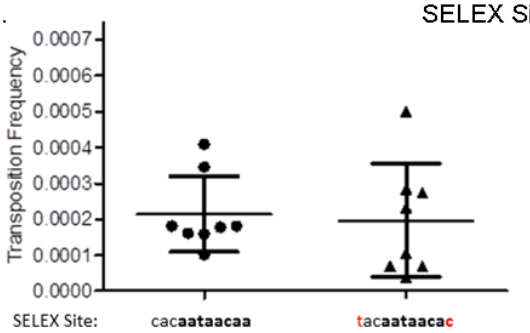


Figure 3.8 The BM1 mutation does not affect transposition frequencies of a high copy IS*10* element as measured by a mating out assay. A. A comparison of the old IS*10*-kan construct (pDH502) and the new one made for this study (pDH725). The new construct is more modular and has less repeated sequence. This makes gene manipulation more efficient. **B.** NK5830 F' cells were transformed with either a wildtype IS*10*-kan construct or an IS*10*-kan with a mutated SELEX site construct and grown overnight on LB medium with kanamycin selection [25 µg/mL]. Cells were subcultured in LB to an OD_{600} of ~ 0.6 and the mating out procedure was followed. Total exconjugates and total hops were calculated and transposition frequencies were calculated by dividing the total hops by the total exconjugates. This data is compiled from 2 independent experiments both with n=4.

As an additional method to determine which region of the transcript Hfq may be acting upon, a series of IS*10* truncations and mutations were explored in context of IS*10* transposase expression using IS*10*-LacZ translation fusions. The IS*10*-LacZ translational fusions are comprised of varying lengths of the IS*10* gene sequence, fused to DNA sequence starting at the 9th codon of LacZ. The two 3' mutants BM1 and BM2 were also tested in this context. The IS*10*-LacZ translational fusions include the RNA-IN promoter as well as the entire RNA-OUT promoter and sequence **(see Figure 3.9).** These data show a general trend of overall transposase expression increases as the IS*10* transcript gets longer (see **Figure 3.9).** This trend is true for all translational fusions in both the Hfq^- , Hfq^+ cells. Structural predictions of the RNA using mFold suggest the increase in activity between the different transcripts is not due to occlusion of the Shine Dalgarno sequence nor a difference in the position of the translational start (Data not shown). The longest IS*10* translational fusion contains the predicted Hfq binding site AAYAAYAA found in the IS Finder search, and there is about a 3 fold increase in activity in Hfq^- cells relative to wild type cells. Both BM1 and BM2 mutants resulted in about a 3 fold decrease in overall expression in both Hfq^- and Hfq^+ strains **(Figure 3.9).** The RNA structure of the promoter region and translational start region in the two mutants are no different than in the wild type transcript according to the most stable structures using mFOLD structural predictions, thus is likely not the reason for the reduced activity. Overall, it appears that mutating the 3' end of IS*10* in the context of a transposase expression assay and a mating out assay does not affect transposase activity.

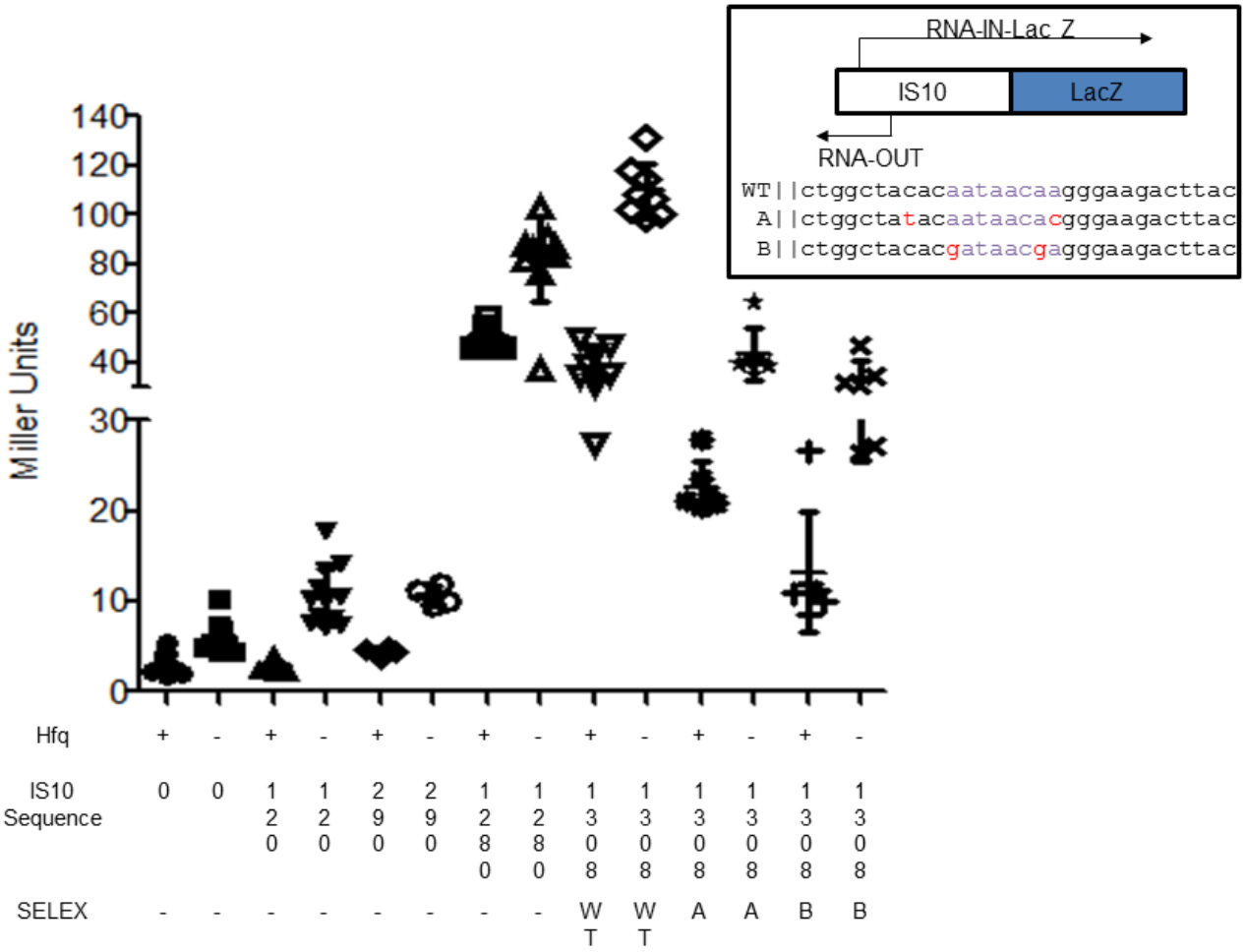


Figure 3.9 β-Galactosidase activity for IS*10*-LacZ translational fusions of varying lengths. Translational fusions were transformed into NK5830 F' (Hfq+) or its isogenic derivative DBH16 (Hfq$^-$) and transposase expression assays were performed. These data are a compilation of at 2 independent experiments with each geometric shape representing one colony in a β-galactosidase assay. The SELEX site refers to the identified Hfq binding sequence AAYAAYAA from searching the ISFINDER database. Increasing amounts of IS*10* sequence were fused to the 9th codon of a LacZ gene, as well as a mutated version of IS*10* that contained 2 base changes in the 3' region of the transcript.

3.6 Is the expression of RNA-IN regulated by an sRNA?

Manipulation of the 3' SELEX site of RNA-IN has a minimal effect on transposase expression and activity regardless of the observation that Hfq binds RNA-IN 3' *in vitro.* The problem being looked at is that IS*10* transposition is down regulated by Hfq up to 86 fold on a multi-copy plasmid but only up to 7 fold when on a single chromosomal copy [18]. RNA-OUT pairing with RNA-IN is also considered negligible at a single copy situation with hfq disruptions resulting in a 2 fold decrease in transposition [11]**.** This led to the question: What else could Hfq be working in conjunction with to repress IS*10* expression? Accordingly, I tested an alternative mechanism for how Hfq might be regulating IS10 transposase, independent of the anti-sense system. One of the major roles for Hfq in the cell is to bind RNA species and promote their pairing with mRNAs. I therefore decided to search for *E. coli* sRNAs that regulate RNA-IN expression. To address this possibility, a sRNA database sRNAMAP [69] was downloaded and the sequences searched against RNA-IN using INTARNA [85] (http://rna.informatik.unifreiburg.de:8080/IntaRNA.jsp). INTARNA is an online bioinformatics tool that allows one to predict the interactions between RNA molecules. INTARNA analyzes the hybridization energy between the two prospective pairing RNA species as well as unfolding energies required for the regions of interaction of each molecule to become accessible. This program is useful to find candidate RNAs that may interact with RNA of interest. An interaction stability of -8 Kcal/mol cut off was used to reduce the number of matches, and non-*E. coli* sRNAs were removed from the list. The top 20 most

stable sRNA-mRNA predicted interactions are reported here **(See Table 3.3).** Of these matches an sRNA from the most stable (GlmZ), mid-stable (SroG) and least stable range (RdlC) were chosen to test experimentally for pairing with 5' region of IS*10*. The predicted regions of interaction of GlmZ, SroG and RdlC are shown in **Figure 3.10.**

Table 3.3 Compiled sRNAs predicted to interact with IS*10* according to INTARNA. Start and stop refer to the base region on the sRNA where pairing may occur between the sRNA and IS*10*. Predictions are based on a minimum seed region of 7 bases for pairing.

sRNA	**Kcal/mol**	**sRNA Start**	**sRNA Stop**	**IS*10* Start**	**IS*10* End**
sroG	-20.232	102	130	53	85
rseX	-18.252	11	66	709	767
omrB	-15.919	6	46	751	790
C0343	-15.661	4	60	541	621
sibD	-14.587	1	24	739	761
glmZ	-14.308	132	157	20	43
oxyS	-14.230	48	82	107	145
sgrS	-14.124	168	178	364	374
ssrA	-13.967	275	326	67	123
tmRNA	-13.708	153	164	601	612
crsB	-13.414	133	181	452	505
C0719	-12.984	62	98	889	922
tpke70	-12.528	118	160	570	620
ryfB	-12.510	123	140	58	79
C0299	-12.039	1	51	252	292
rdlD	-11.877	1	10	664	673
dsrB	-11.748	55	73	769	786
C0465	-11.469	12	31	608	633
omrB	-11.428	6	46	751	790
ssrS	-11.420	25	61	750	785
*rdlC	-8.737	1	8	15	22

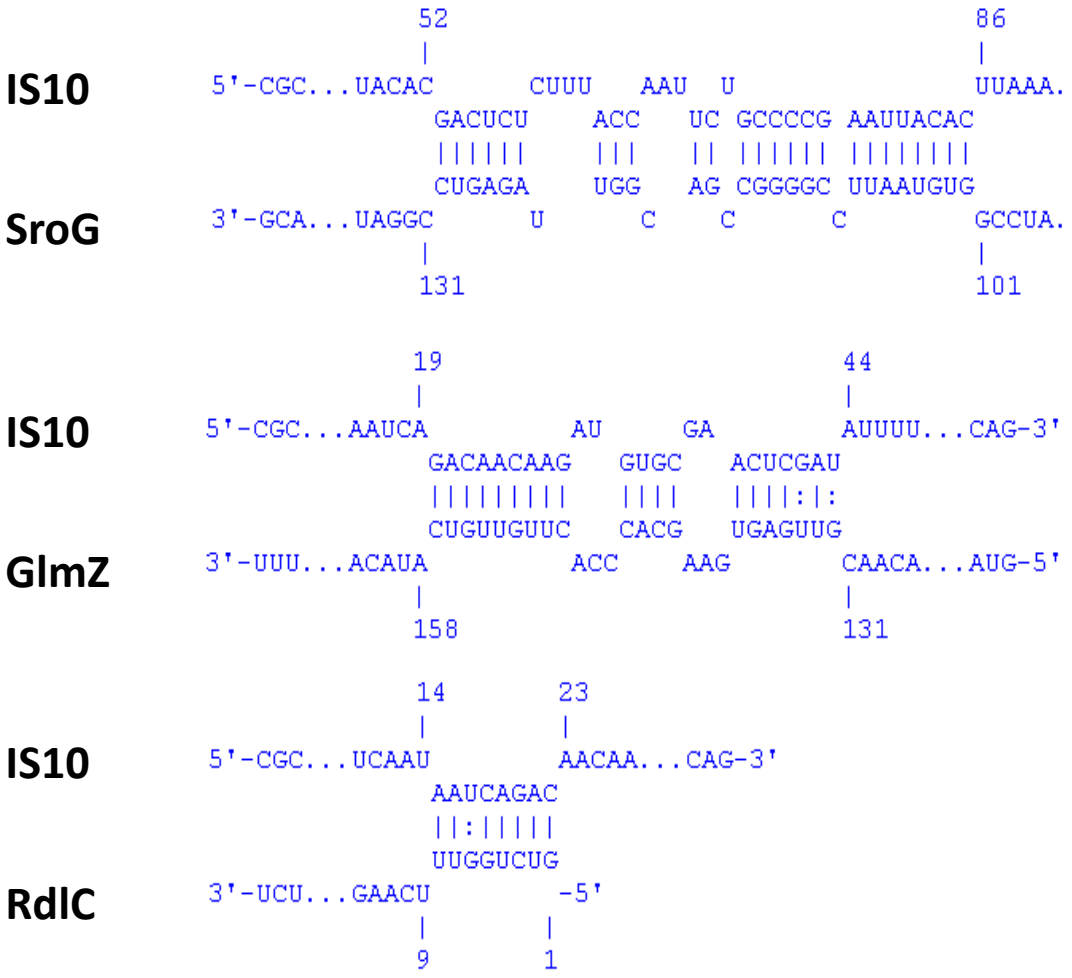


Figure 3.10 Predicted regions of pairing between sRNAs and RNA-IN according to INTARNA. Predictions are based on pairing at 37°C with a minimum seed region of 7.

The strategy I took to test potential sRNA-IS*10* transposase mRNA interactions, was to make an *in vitro* transcribed transcript of both the sRNA candidate and IS*10* transposase mRNA. These transcripts were then incubated together at 37°C (the same temperature employed in the INTARNA program), and ran on a polyacrylamide gel to see if base pairing occurred. A positive result in this 'pairing' assay would provide impetus for genetic studies. An *in vitro* transcription template was made for these sRNAs by PCR amplifying a genomic preparation of NK5830F' with the primers specific for the corresponding RNA. The primer pairs BM31A/BM32A, BM33A/BM34A and BM35A/BM36A were used to amplify GlmZ, RdlC and SroG respectively. The forward primer included the T7 RNA polymerase promoter (**see Figure 3.4** for the general strategy). RNA transcripts were made by T7 RNA Polymerase (NEB) using the T7 RNA polymerase promoter containing template of each sRNA. The sRNAs were heat treated at 90°C for 2 minutes, placed on ice for 5 minutes then the pipetted on the side of a microfuge tube along with the first 160 nucleotides of ^{32}P labeled RNA-IN (IN-160) and centrifuged into Hfq buffer to ensure all binding reactions were started at the same time. The reactions were incubated for 15 minutes at *37*°C and loaded on to a 5.5% native polyacrylamide gel at 21V/cm to test for base pairing of the sRNA and RNA-IN (first 160 bp) transposase RNA. Pairing was not detected between IS*10* and SroG, GlmZ, or RdlC at any of the concentrations of sRNA:IN-160 tested (where the sRNA was in as much as 100 fold excess over RNA-IN) **(Figure 3.11**). In contrast, when RNA-IN was in excess to GlmZ, some pairing was detected (**Figure 3.11)**. It is important to note that experiments

shown here are experiments done in the absence of Hfq. I was able to show that Hfq binds GlmZ (**Figure 3.12**), but I did not see stimulation in pairing efficiency when Hfq was titrated into a GlmZ-RNA-IN Pairing reaction (Data not shown). The *in silico* predictions I used to devise a list of potential sRNAs that could base pair with RNA-IN were based on conditions with the sRNA and mRNA at 37°C, and did not include Hfq. If GlmZ is involved in IS*10* regulation, we expect it to be in an Hfq dependent manner which would explain the increase in transposition in the absence of RNA-OUT and Hfq. To address this we tested if Hfq binds GlmZ transcript. ^{32}P labelled GlmZ was incubated for 15 minutes at 37°C with increasing concentrations of Hfq. GlmZ binds Hfq and depending on the Hfq concentration can form up to 3 complexes (**Figure 3.12)**.

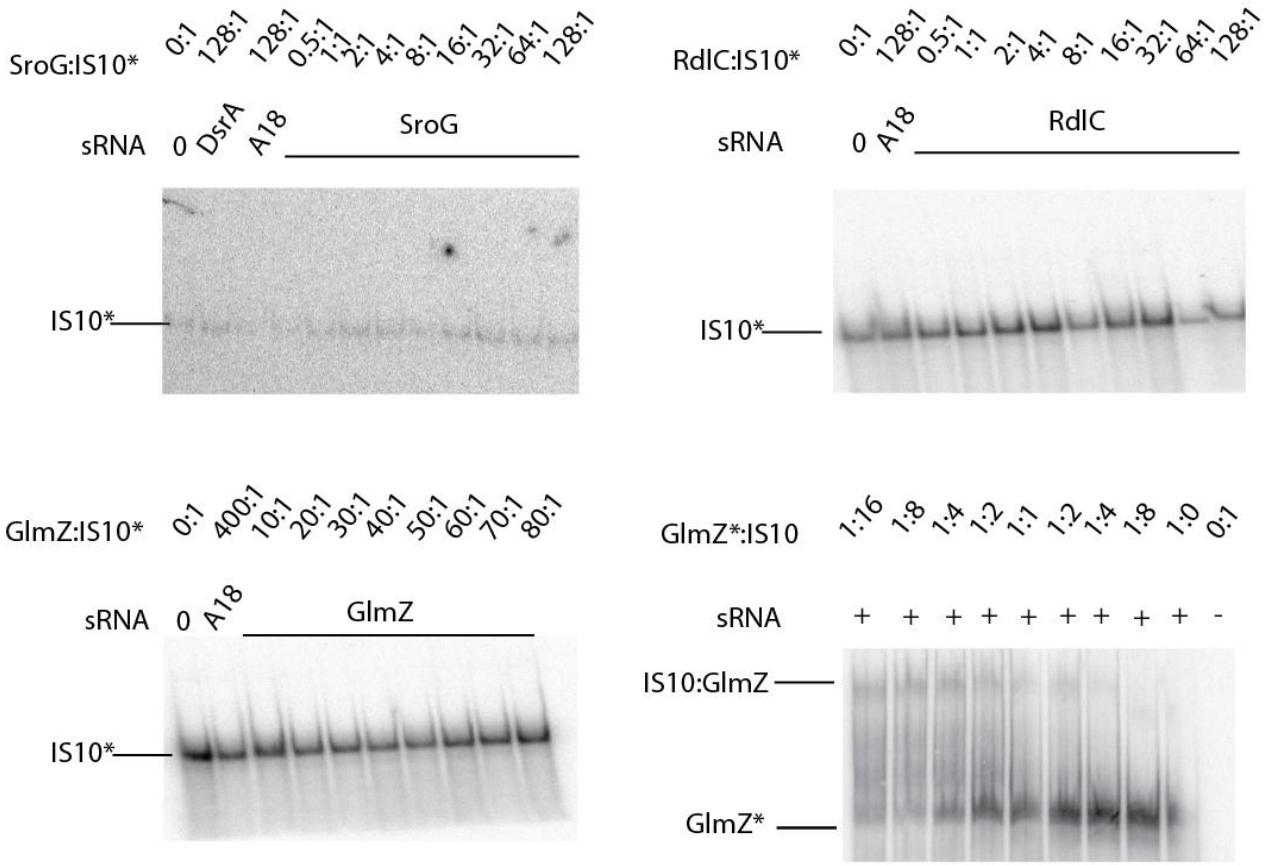


Figure 3.11 Pairing of sRNAs to IS*10* (bp1-160) only occurs when IS*10* is in excess. sRNAs were incubated at 37°C in Hfq buffer and ^{32}P labelled IS*10* at the given ratios for 15 minutes. An equal volume of 2x load mix was added to each reaction and samples were run on a 6% native polyacrylamide gel. **Bottom Right:** ^{32}P labelled GlmZ was incubated at 37°C in Hfq buffer and IS*10* RNA at constant ratios for 15 minutes. An equal volume of 2x load mix was added to each reaction and samples were run on a 6% native polyacrylamide gel.

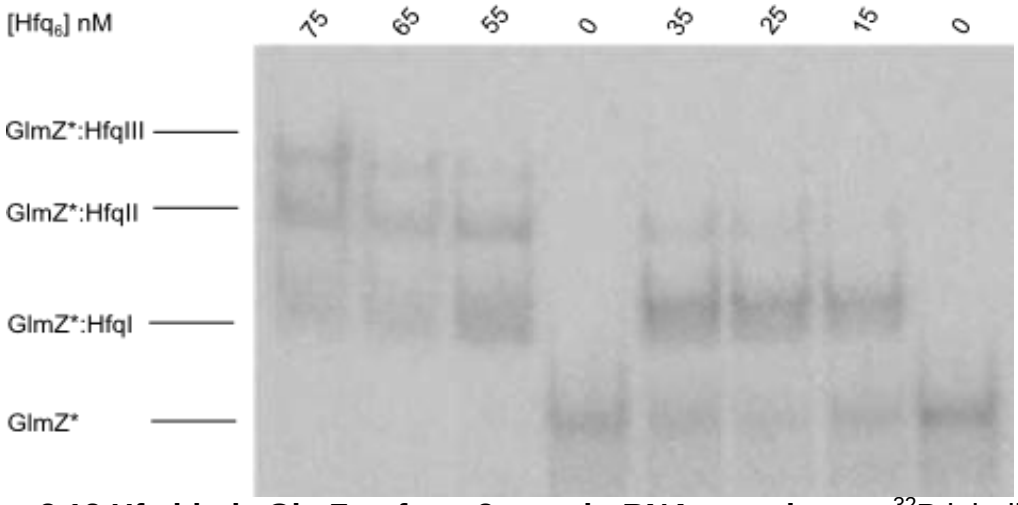


Figure 3.12 Hfq binds GlmZ to form 3 protein-RNA complexes. ^{32}P labelled GlmZ (GlmZ*) was incubated at 37°C in Hfq buffer with varying Hfq concentrations for 15 minutes. An equal volume of 2x load mix was added to each reaction and samples were run on a 6% native polyacrylamide gel.

3.7 Does limiting the availability of Hfq proteins in *E.coli* impact on IS10 transposition frequency?

In this section of the thesis I asked if reducing Hfq levels in the cell increases IS*10* transposition frequency. A proposed IS*10* Hfq regulation model from *Ross et a/.* suggested that cellular stress may result in the titration of Hfq through over expression of sRNAs [18]. In support of this model, data suggest that the overexpression of an Hfq binding sRNA or mRNA can decrease sRNA directed regulation of a target gene [54],[53]. The sRNAs DsrA, RhyB and MicC were chosen to titrate Hfq as they were effective Hfq titrators in the previously mentioned studies. We also explored the effect of the sRNAs GlmZ and SroG on IS*10* transposition frequencies. GlmZ was chosen, due to the *in vitro* data suggesting IS*10* can pair with it **(Figure 3.11),** and SroG due to its predicted binding to IS*10* from IntaRNA **(Figure 3.10),** as well as the availability of a SroG

disruption strain in the laboratory. To test if reduced Hfq levels result in differing transposition frequencies, 3 sRNAs, MicC, RhyB and DsrA, were cloned into pACYC184 under the control of the pLlac01 promoter. Additionally, an IPTG inducible pLlac01 GlmZ construct was generously given to us by the Gottesman lab (NIH, Bethesda, MD). The inducible sRNA containing plasmid were co-transformed into NK5830 with a plasmid containing a kanamycin marked IS*10* element and plated overnight on LB containing 25 µg/mL of kanamycin and 10 µg/mL of tetracycline at *37*°C either in the presence of absence of IPTG. MicC, RhyB and DsrA were induced with 1 mM IPTG, and GlmZ was induced with 100 µM IPTG. Resulting transformants were used to inoculate overnight cultures at 37°C in LB broth with identical additives as the plates the colonies were picked from. Overnight cultures were subcultured in 1 ml of LB broth and a standard mating out was performed (see Materials and Methods). One mL of the overnight culture was used for total RNA isolation. Total RNA was isolated using acid-phenol extraction, quantified and equal amounts of total RNA underwent reverse transcription using primers specific to the sRNA. A NK5830 F' containing a *sroG* gene disruption was also tested for a change in transposition frequency. Induction of MicC resulted in a 1.7 fold increase in transposition compared to non-induced. GlmZ, RhyB and DsrA induction had no significant effect on transposition compared to their non-induced counter parts **(Figure 3.14).** Semi-quantitative RT PCR shows approximately a 110 fold increase in MicC RNA in the induced cells versus the uninduced cells and an approximately 20 fold increase in GlmZ RNA in the included cells verses the uninduced cells **(Figure 3.13**). Overall, MicC overexpression caused a small increase in IS*10*

transposition and I suggest this may be due to limiting the availability of Hfq in the cell. In contrast, overexpression of a 2nd sRNA, GlmZ, did not influence IS*10* transposition frequencies.

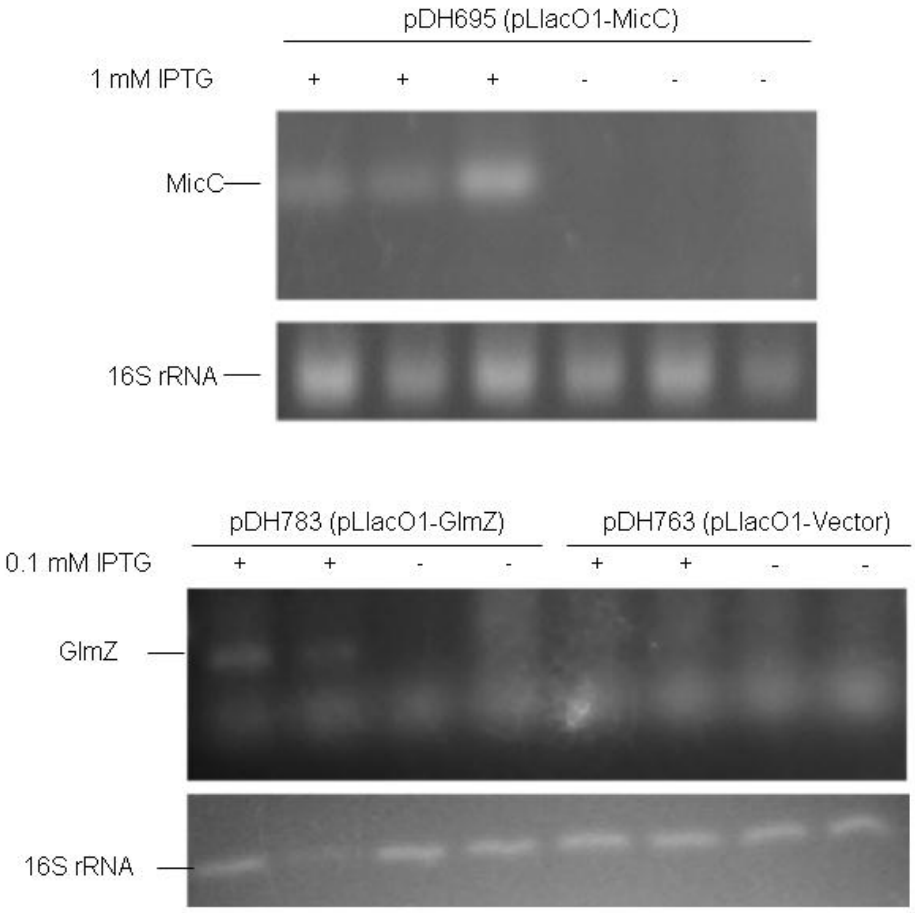


Figure 3.13 IPTG induced MicC expression (Top) and IPTG induced GlmZ expression (Bottom). Cells NK5830 F' cells transformed with a plasmid containing plLacO1 MicC or pLlacO1 GlmZ and were grown to exponential phase in the presence of 1 mM IPTG for MicC or 100 µM for GlmZ. Total RNA was isolated and RT PCR was performed using reverse primers specific to MicC, 16S rRNA and GlmZ.

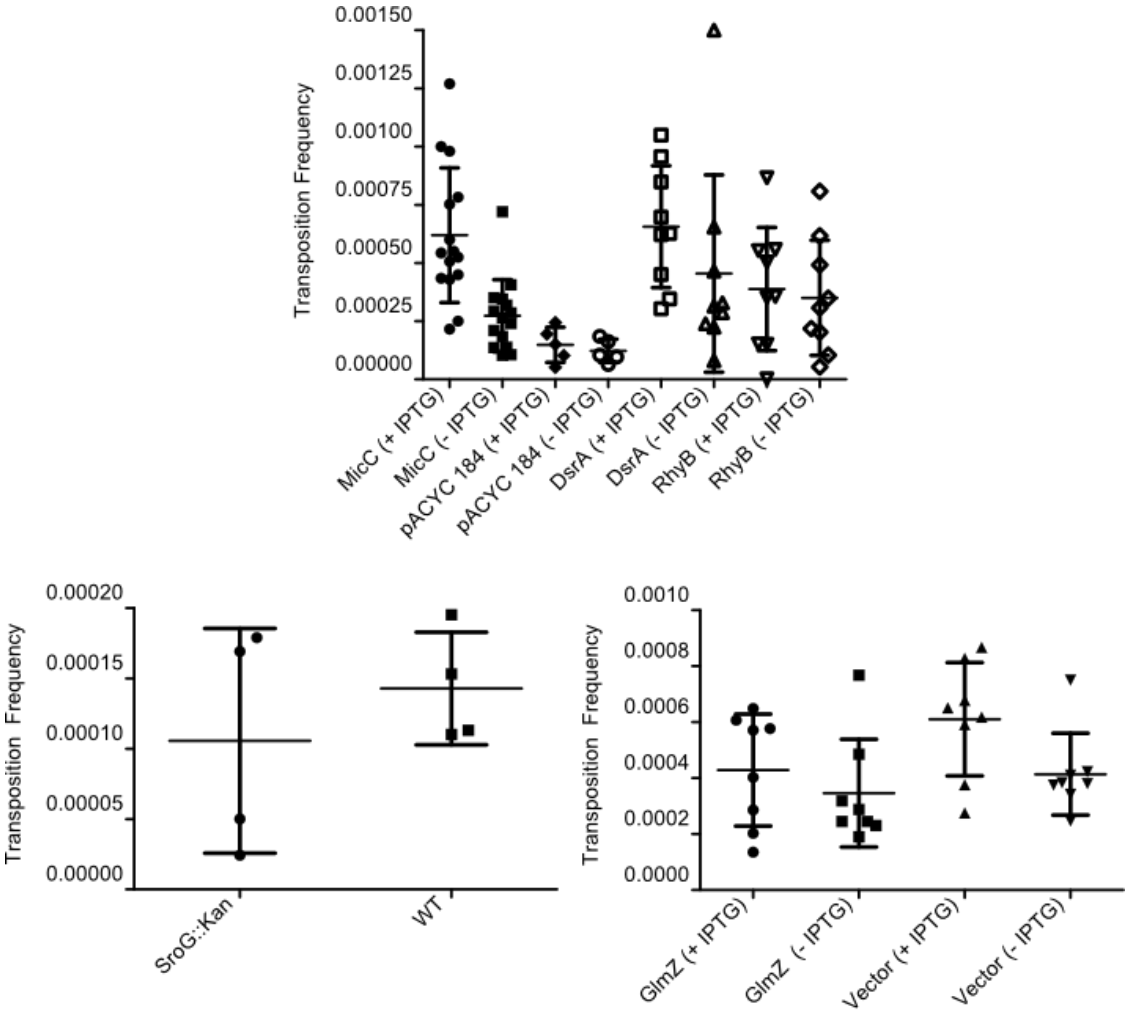


Figure 3.14 sRNA overexpression mating out assay (top and bottom-right). NK5830 F' cells were co-transformed with an inducible sRNA plasmid and a marked IS*10* element IS*10* Kan (pDH502). Cells were grown overnight in kanamycin and tetracycline (25 µg/mL and 10 µg/mL respectively) LB in either the presence or absence of IPTG (1 mM for top graph, 100 µM for bottom right) and then subcultured in LB in the presence or absence of IPTG (1 mM for top graph, 100 µM for bottom right). The mating out assay was then performed. **SroG disruption strain mating out assay (Bottom-Left):** NK5830 cells as well as NK5830F' with SroG replaced with a kanamycin gene were transformed with pDH502. Cells were grown overnight in LB containing Ampicillin [50 µg/mL]. The mating out assay was then performed.

Overexpression of sRNAs can titrate out available cellular Hfq however; the physical amount of Hfq is unchanged. To test if reducing Hfq expression levels (a percentage of the amount of Hfq in the cell) impacts on IS*10* transposition frequency I utilized a *dksA* disruption strain. Loss of DksA has been shown to reduce Hfq levels by 50% in *Shigella flexinari* [55] and *E. coli* [54]. To introduce a *dksA* disruption into a strain suitable for the mating out assay, a P1 phage lysate was produced by infecting *MG1655 Δdks::tet* (a gift from M. Cashel via S. Gottesman) with P1 phages and transducing NK5830 F' with the resulting lysate. NK5830F' transductants were verified by PCR to contain the tetracycline gene insertion in the *dksA* gene (DksA$^-$).

The dksA::tet insertion strain of our mating out strain was transformed with a 'marked' IS*10* element (pDH502), grown overnight on LB plates containing kanamycin and then overnight cultures were set up with individual colonies in LB broth containing kanamycin. The overnight cultures were subcultured in LB broth or LB broth containing tetracycline (to ensure the disruption was present). The standard mating out protocol described in Materials and Methods was followed. There was a 3-fold increase in IS*10* transposition seen between NK5830F' and DksA$^-$ NK5830F' cells during our standard mating out conditions (37°C, grown to exponential phase on LB broth) **(Table 3.4).** This result indicates that down-regulating Hfq expression results in a small increase in IS10Kan transposition.

Table 3.4 Ratios of transposition frequencies of NK5830F' and its isogenic mutants. IS*10* transposition frequencies in NK5830 F' wildtype and mutant cells were determined from a mating out assay at 37°C (Normal conditions) based on two independent experiments with n=4, except the H-NS mutant which had only 1 experiment with n=4.

Mutant/WT'	Ratio	Standard Deviation
***hfq*⁻/WT**	33.14	0.84
***dksA*⁻/WT**	3.03	1.41
***rpoS*⁻/WT**	2.12	0.60
***hns*⁻/WT**	1.03	0.57

The transition of *E. coli* from exponential growth phase to stationary phase changes the expression profile and levels of many cellular proteins. For example *Azam et a/* suggest that Hfq, H-NS and IHF (all proteins shown to be involved in Tn10 transposition) levels all decreased after prolonged stationary phase [86]. Additionally the sRNA expression profile of a cell will change under different growth conditions. *RpoS* encodes a sigma factor that plays an important role in orchestrating changes in gene expression that facilitate a cell's transition into stationary phase. In this section, I asked if stationary phase stress impacts IS*10* transposition frequency. A modified mating out protocol was devised in order to determine the effect of stationary phase stress on IS10 transposition frequencies. Our mating out strain, NK5830 F', was transformed with pDH502 and grown overnight on LB kanamycin plates at 37°C then individual colonies were used to inoculate LB kanamycin broth [25 µg/ml] at 37°C. The inoculum was then left for 5 days without aeration on a bench top for growth, and then processed in a mating out assay. These stationary phase growth conditions resulted in a 9.8 fold

increase in IS10 transposition compared to our 'standard' mating out conditions where cells are grown to exponential phase prior to the mating procedure **(Table 3.5)**. This result suggests that growth conditions can alter IS*10* transposition frequencies.

To follow up on this observation that growth conditions can alter IS*10* transposition frequencies, I asked the question: does this 9.8 fold increase transposition events still occur in different genetic backgrounds? The goal of this study was to determine the genetic dependencies on IS*10* transposition frequencies under starvation conditions. Three different genetic backgrounds (*rpoS*$^-$*, dksA*$^-$ and *hfq*$^-$*)* of our mating out strain (NK5830) were tested for an increase in IS*10* transposition frequencies under 'starvation' conditions. When analyzing the growth curves of these mutants, it became apparent that only the *rpoS*$^-$ mutant provided a good comparison to the wildtype mating out strain as *dksA*$^-$ and *hfq*$^-$ *cells* did not grow well under these conditions (see **Figure 3.15)**. Interestingly, there was no difference in transposition frequencies in the *rpos* mutant when comparing to room temperature growth to 37° C, but there is an increase in transposition in the wild type background. A possible explanation for this could be the Hfq levels present in the two backgrounds, whereas higher Hfq levels may reduce transposition frequencies in the *rpoS*$^-$ mutant. At 50 hours, both *rpoS*$^-$ and NK5830 cells appear to be towards the end of exponential phase, and at 110 hours, they both appear to be approaching death phase. One ml aliquots of cells were obtained from RpoS or NK5830F' backgrounds and used to quantify Hfq levels by a semi-quantitative Western Blot **(Figure 3.16).** There was no significant difference in levels in Hfq levels at either time point. Therefore, the

inability of *rpoS*$^{-}$ cells to support an increase in IS*10* transposition upon 'stationary phase' stress does not appear to be linked to Hfq levels. Nevertheless, this experiment has provided the 1st evidence that IS*10* transposition is induced by stationary phase stress and that this induction is dependent on the transcription factor RpoS.

Table 3.5: Transposition frequency ratios in different growth conditions. Ratio of transposition frequencies determined from a mating out assay at room temperature with no aeration divided (Starved) by the transposition frequency determined at 37°C with aeration (Normal) based on two independent experiments with n=4.

Starved/Normal	Ratio	Standard Deviation	N= 'Starved'	N= 'Normal'
5830/5830	9.81	1.64	14	8
***dksA*$^{-}$/*dksA*$^{-}$**	2.65	1.59	9	8
***rpoS*$^{-}$/*rpoS*$^{-}$**	1.02	0.57	8	8
***hfq*$^{-}$/*hfq*$^{-}$**	0.15	1.23	10	10

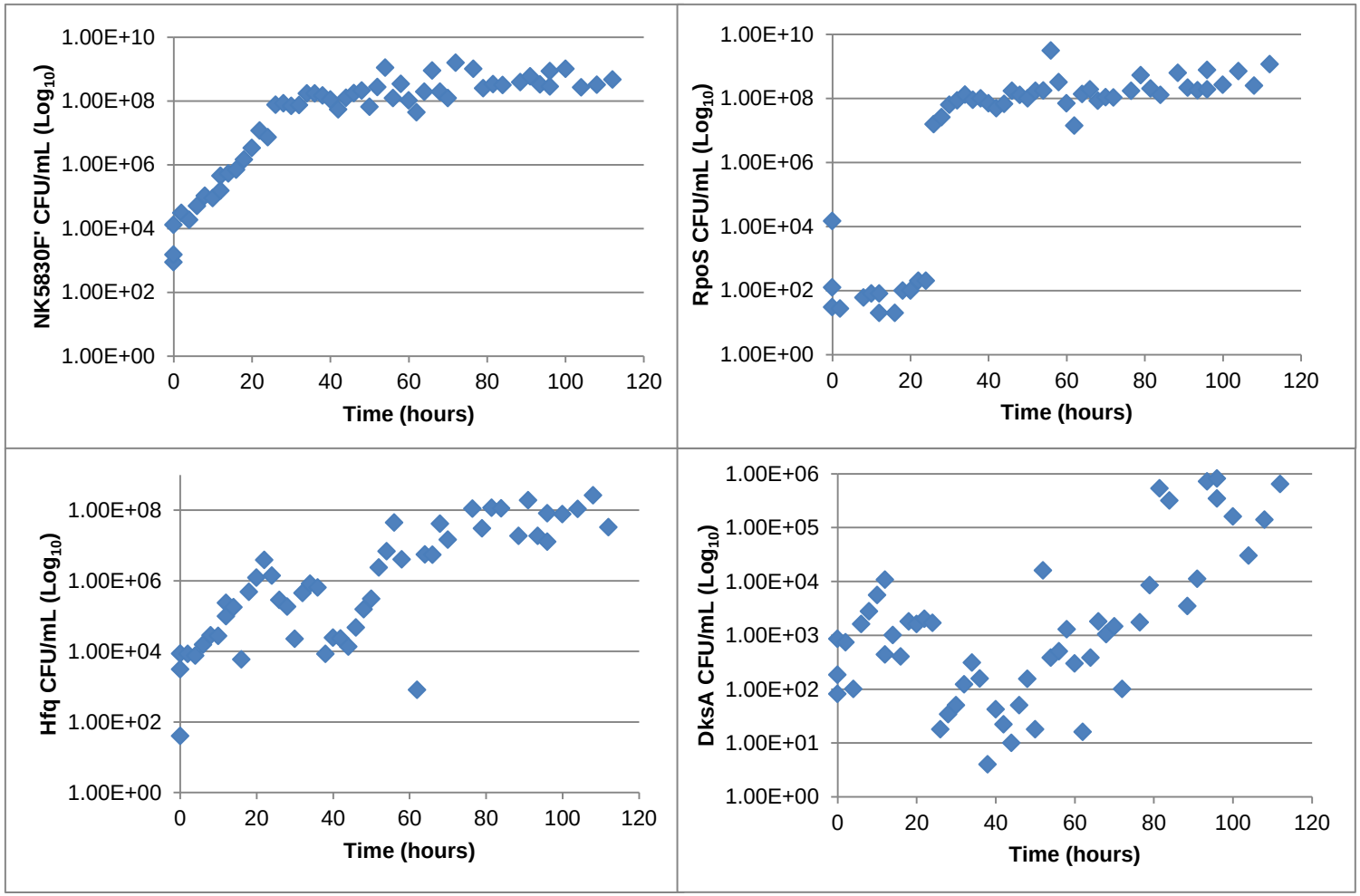


Figure 3.15 Growth curves of NK5830 F' and its isogenic mutants with disruptions in r*poS*, *dksA, a*nd *hfq t*ransformed with pDH502. Cultures were inoculated with a 12 hour offset in order to obtain a time point every two hours (3 independent colonies at varying times). Cells were taken out in 1 mL aliquots at each time point, spun down, suspended in 0.85% NaCl and diluted for plating. LB plates containing kanamycin [25 µg/mL] were incubated overnight at 37˚C and colonies were counted. The RpoS growth curve contains a large increase in cells per mL between 22 and 26 hours. This is likely due to having the 12 hour offset and missing the time point at 24 hours. One can see at time 0, there is one culture that started at 1 x 10^4 CFU/mL.This culture contains 100 times more starting cells than the other 3 cultures being analyzed, so after 26 hours of growth it will contain more cells than the other culture being measured.

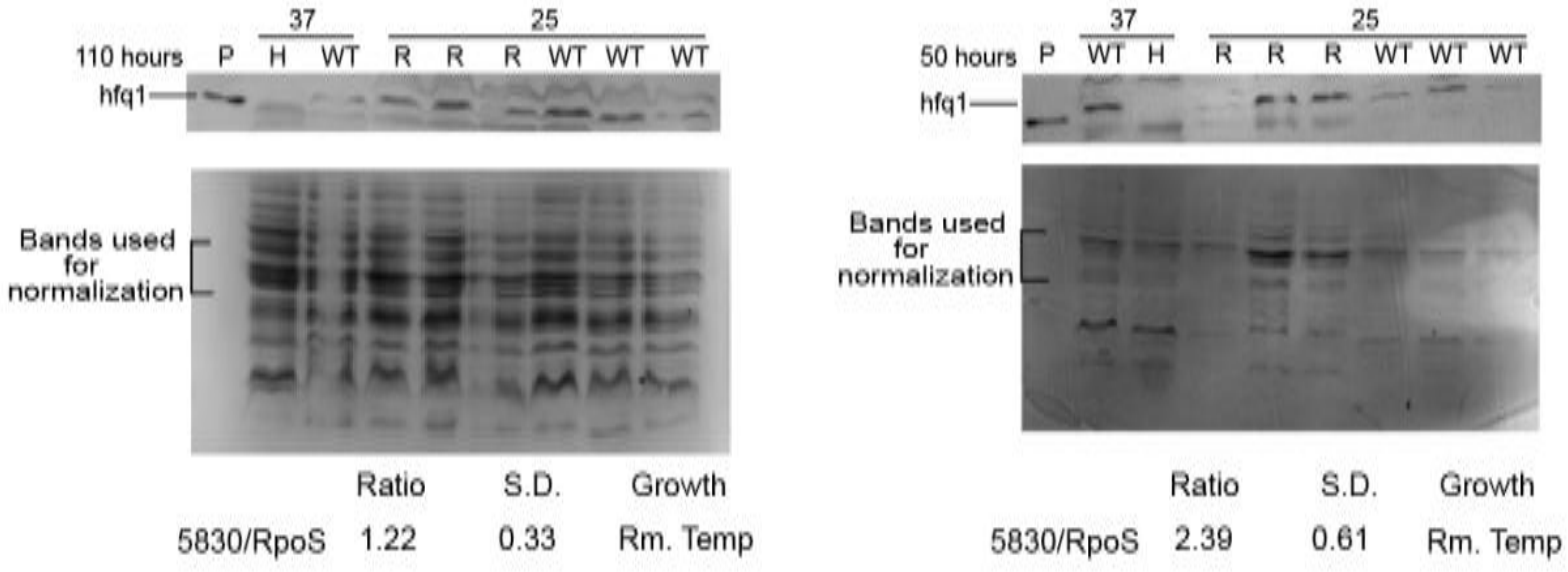


Figure 3.16 Western blots of NK5830F' and *rpoS*$^-$ cell lysates. Cells were grown to either 50 hours or 110 hours at room temperature without aeration. Cell volumes were normalized to 1 mL at an OD_{600} =0.6, and that volume was spun down and resuspended in 20µL of SDS load buffer (2% SDS, 10% glycerol, .06M Tris HCl pH6.8, .025% Bromophenol Blue), boiled for 300 seconds then loaded on a 14% SDS-PAGE gel (10µL load) and ran at 17V/cm until the Bromophenol Blue dye front reached the bottom of the gel. Gel was transferred on PVDF membrane at 0.6mA/cm^2 of PVDF membrane. PVDF membrane was blocked overnight in 5% fatfree milk (Nelson) dissolved in TBST buffer. Western blotting was done with a 1/100,000 dilution of α-hfq antibody (rabbit polyclonal) and a 1/4000 dilution of a mouse α-rabbit IgG (Promega). P-lane represents 100 ng of purified Hfq. H-lane contains cell lysate from *hfq*$^-$ cells (DBH16). WT-lane contains cell lysate from wildtype cells (NK5830). R-lane contains cell lysate from rpos$^-$ cells (DBH 48). P-lane contains purified Hfq protein only.

Chapter 4 - Discussion

4.1 Using bioinformatics to identify IS Elements that are under Hfq regulation.

In this work we successfully identified another Hfq regulated element, IS*1413*, by searching for Hfq binding sites of in IS Element transposase mRNA sequences, instead of physically working with the element first. Further characterization of other elements found in **Tables 3.1** and **3.2** may expand the list of Hfq regulated IS Elements from the current 3 (IS*10* [18], IS50 (Crystal McClellan thesis), and now IS*1413* (this work)).

This bioinformatics approach to finding IS Elements that are candidates for Hfq regulation demonstrated several advantages over individually testing each IS Element for a change in transposition frequencies in the presence or absence of Hfq. First of all, there is a large data base of IS Element sequences available in one location in the IS finder database. In its original form, the IS finder data base makes a meta-analysis of all entries difficult; however this was overcome using a webcrawler and a parsing macro to put all of the Element information in one easy to search document. Once this document of all the IS-Elements was compiled, all of the IS Element information are found in a sortable and searchable spreadsheet. Second of all, the Hfq binding site has been rigorously defined in sequence and structure composition. Hfq binding sites on mRNA species have been defined in terms of sequence such as the SELEX binding site [44], the polyA binding site [41], and the ARN_x binding sites [41] . Regions of RNA that bind Hfq are often followed by a hairpin (reviewed in [38]). Finally, the high degree of

conservation between Hfq proteins in different bacteria, lead me to the expectation that Hfq binding sites will also be highly conserved in different bacteria. As mentioned previously, Pseudomonas Hfq relieves RpoS gene inactivation in *E.coli* containing Hfq disruptions[74] and *Burkholderia cephacia* IST408-SS7 Hfq relieves the slow growth phenotype that *E.coli* MC4100 ΔHfq exhibits [87]. Overall, this method of identifying IS Elements that may be under Hfq control, remains a valuable tool in the study of how prevalent Hfq regulation is throughout the different families of IS Elements.

There still remains some issues to address regarding using the ISFINDER searching strategy for Hfq regulated IS Elements. The first concern of this approach is it does not take into account the secondary structure of the IS Elements' transposase mRNA. Hfq binding is dependent on both sequence and structure of the RNA species (reviewed in [38]). I would predict that by adding the secondary structure parameter into this approach would result in fewer candidates for Hfq regulation. Future modifications of this approach to overcome this concern could include, a scoring system that first scores candidate IS Elements for containing the Hfq binding sequence, and then analyzing that region for secondary structures. Conversely, although we may have more candidates by not including the secondary structure parameter in our search, we are biasing the results for a high percentage of type II statistical errors, due to the stringency in the Hfq binding site searched. Not all Hfq binding sites will contain the perfect consensus sequence AAYAAYAA found by *Lorenz et al.* [44], We are thus increasing the possibility of false negatives in this search. Finally, the capacity of

Hfq to bind an RNA site tightly does not always translate into the interaction being biologically relevant. This phenomenon was observed in the Hfq binding site on the 3' end of IS10's transposase mRNA, which upon mutation did not result in any change in Hfq sensitivity. Despite these potential caveats of this bioinformatics approach to identifying Hfq regulation of IS Elements, I was successful in identifying IS*1413*, which did exhibit a change in transposition frequencies in the presence of Hfq. This approach does remain a good starting point to identify Hfq sensitive IS Elements to be followed up on using traditional genetic and *in vitro* studies.

4.2 Hfq and IS*1413* transposition:

These studies show that the frequency of IS*1413* transposition and IS*1413* transposase expression both increase in the absence of Hfq. Following up on this observation, I tried to correlate this 'Hfq effect' with a validated Hfq binding site in the 5' UTR of the IS*1413* transposase mRNA. I did this by looking at mutations in the putative 5' UTR of the IS*1413* transposase mRNA. These mutations did not result in an increase of transposase expression in the presence of Hfq (unlike the hfq disruption). There are several possibilities why mutating the putative Hfq binding site did not result in the de-repression of IS*1413* transposase expression in the presence of Hfq. The most obvious possibility is our identified binding site may not be a bona fide Hfq binding site *in vivo.* Our bioinformatics data and *in vitro* data support that Hfq does bind to this region, however *in vivo* several other factors would be involved in a successful binding interaction such as the accessibility of the binding site, or the availability of cellular Hfq. An interesting

result of the IS*1413* transposase mutations we looked at was that it caused an overall increase in IS*1413* transposase expression in both the presence and absence of Hfq **(see Figure 3.1).** The expectation of this study was that the mutants would have lower transposase expression in the presence of Hfq relative to the wild type translational fusions. These mutations we introduced may have inadvertently improved either the translation initiation or the transcript stability of IS*1413*. This overall increase in IS*1413* transposase expression may overshadow any positive effects linked to the loss of Hfq binding to the transposase mRNA, and thus we would be unable to see any subsequent repression. More mutations will need to be made and analyzed for IS*1413* to see if we are able to alleviate Hfq repression. Our identified putative Hfq binding site in IS*1413* was based on the stringent sequence AAYAAYAA, which did not allow for any substitutions in any position. The possibility remains that there may be other Hfq binding sites, that are not a perfect match, are downstream of the 5'UTR that Hfq is binding to down regulate IS1*413* transposase expression. Finally, it is possible the Hfq effect on IS*1413* transposition frequencies is indirect, and is actually due to the different genetic expression profiles between the wildtype NK5830 F' cells, and its isogenic *hfq* disruption strain DBH16.

In this work on IS*1413* we were able to show that IS*1413* is active in *E.coli* and we were able to measure the expression of the IS*1413* transposase gene in *E.coli* using a translational fusion. The Hfq status of *E.coli* influenced both IS*1413* transposition frequencies and transposase expression; both of these increased in an *hfq*$^{-}$ *E.coli* strain. We have has shown that IS*1413* is an active element that

can transpose in *E.coli* with a transposition frequency of 3.7 x 10^{-5} transposition events per mL of mating mix in the presence of Hfq and 5.0 x 10^{-4} transposition events per mL of mating mix in the absence of Hfq. These results suggest that Hfq has a repressing role on IS*1413* expression. This is more evidence for the notion that Hfq is a general regulator of IS Elements. Subsequently, with Hfq being linked to cellular stress responses [88], these results provide a possible pathway for linking IS1413 transposition to cellular stress pathways. In other words, under certain conditions the host may be more permissive to transposition events which may result in the transposon induced gene expression changes, such as the expression of the *tft* operon to allow for the utilization of phenol compounds as a carbon source.

The observation that IS*1413* is active in *E. coli* is an example of how IS Elements can be promiscuous carriers of genes from one bacterial population to another. Since IS*1413* is also active in *Burkholderia* [56] *,* this opens up the possibility that *Burkholderia* and *E. coli* could potentially exchange genes through IS*1413* mediated transposition. This gene exchange would entail IS*1413* carrying additional genes with it during a transposition event onto a conjugative plasmid and transferring it to another species. Species of the genus *Burkholderia* are commonly found in water in soil and sometimes as plant or human pathogens [89]. *E. coli* is primarily found in the lower intestine of warm blooded animals as well as in water, sediment and soil [90]. It is very possible these two organisms could be found in a common habitat such as soil, and share genes through conjugation and transposition events. The identification of IS*1413* being active in

E. coli opens up the possibility that it could be an agent of pathogenicity transfer between the two organisms. For example, if IS*1413* flanks a pathogenicity gene in *E.coli* or an antibiotic resistance factor, it may move onto a transmissible plasmid (such as the F-factor). If this IS*1413* harbouring pathogenic *E.coli* is passed into soil by an infected animal and *Burkholderia phenoltripix* is also present, through conjugation *Burkholderia phenoltripix* may acquire the ability to infect an animal that consumes that soil (See **Figure 1.1** for mechanisms that could account for this).

Our results show Hfq influences IS*1413* expression and transposition frequencies; this provides us with further understanding of how IS*1413* transposition is regulated. *IS1413* is not a well-studied transposon. Prior to this study, IS*1413* was identified to be active in *Burkholderia phenoltripix* [56]. Subsequently, it was demonstrated that IS*1413* can up regulate the *tft* operon in *Burkholderia phenoltripix*. The first gene in the *tft* operon is involved in the breakdown of 2,4,5 tricholorophenol (i.e. Agent Orange) [91]. By further understanding the regulation of IS*1413* and its activation of the *tft* operon, one could potentially utilize *Burkholderia phenoltripix* for bioremediation of areas polluted with 2, 4, 5 tricholorophenol compounds such as Agent Orange. Ultimately, one would hope on identifying conditions that would facilitate the up regulation of the *tft* operon and thus facilitate the degradation of these toxic compounds.

4.3 Hfq regulation of IS*10*

An objective of my thesis was to fully identify how Hfq represses IS*10* transposase. There were two main reasons for why I looked at Hfq regulation of IS*10* transposase. First of all, my bioinformatics screen predicted that there is an Hfq binding site in the 3' end of the transposase transcript. Secondly, previous studies by *Ross et al.* suggest there is Hfq regulation independent of the IS*10* transposase transcript. I looked at the possibility that Hfq regulates IS*10* transposase, at least in part, by binding the 3' end of the transposase transcript where my bioinformatics analysis predicted the existence of an Hfq binding site.

An example of Hfq binding the 3' end of a transcript and the effects it can have are seen in the interactions of *cspA* mRNA and Hfq. Hfq protects *cspA* mRNA from RNase E mediated degradation. It also increases the length of a polyA tail due to polyadenylate polymerase [92]. With a putative 3' binding site in the IS*10 transposase*, I decided to characterize the 3' end of the IS*10* transposase transcript to look for evidence of polyadenylation. The presence of polyadenylation on the transposase transcript would have been consistent with the 3' SELEX site and Hfq playing a role in processing of the transcript. Results the 3' RLM-RACE did not show any evidence of polyadenylation of either IS*10*L or IS*10*R. The identified end may be a processed transcript that occurs following transcription. Nonetheless, I did successfully define a 3' end for both IS*10*R and IS*10*L. Previous studies have never determined this prior to this work. Future directions to confirm the true 3' end of IS*10* could include an RNA protection assay to verify the predicted 3' end of IS*10* predicted by the 3' RLM-RACE.

Understanding precisely how the IS*10* transcript terminates may provide some insight into the effect Hfq has on reducing its half-life.

To characterize the putative Hfq binding site in the 3' end of IS*10* transposase transcript, I utilized RNA-Hfq binding studies and transposase expression studies. Through binding studies, I found that Hfq binds to the 3' site when a portion of the transposase transcript was studied. When trying to initially determine the regions that are important for Hfq regulation of IS*10*, different lengths of IS*10* were used to make translational fusions with LacZ. A general trend observed in this set up was the longer the transcript, the more active the fusion was. Generally, Hfq repressed the overall activity of the transcript in all of the translational fusions about 3 fold. However, the fusions containing very small amounts of IS*10* (180 bp or 290 bp) have transposase expression barely above the 'cells only' control. One possibility as to why the longer fusions were more active is due to the shorter transcripts not being able to fold properly. Improperly folded transcripts would have a greater chance of being degraded by single stranded specific RNases. When the 3' SELEX site in the IS*10* transposase LacZ translational fusions was disrupted through site directed mutagenesis, Hfq repression was not relieved. There are a few possible reasons for why we were unable to de-repress Hfq regulation of IS*10* transposase expression. First, Hfq may be unable to bind the 3' SELEX site in the IS*10* transposase transcript *in vivo*. Second, Hfq may be able to bind the 3' SELEX site in the IS*10* transposase transcript *in vivo*, however under the conditions we tested it may be missing another factor required to elicit a biological response. This missing biological factor could be an sRNA that is only expressed under certain stress conditions.

Overall, the bioinformatics approach I took to identify Hfq regulation sites in transposase transcripts was successful in the case of IS*1413*, but not for IS*10*. This project successfully identified a new Hfq-regulated transposase, IS*1413*, however this may be fortuitous. Further characterization of the sensitivity of transposase expression to Hfq of my bioinformatics identified elements and a random list of IS Elements would help put my approach to the ultimate test. If my bioinformatics approach results in a higher proportion of Hfq sensitive transposases compared to the random list, this would suggest that the initial *in silico* screen provides a list of candidates that one should follow up on.

4.4 RNA-IN has an Hfq binding site in the 3' end and it may be relevant under certain physiological conditions.

The data from the 3' RLM-RACE data suggests that there may not be any post transcriptional modifications of IS*10* (**Figure 3.3)**. *In vitro* binding studies show that Hfq is able to tightly bind the 3' end of RNA-IN and mutating the putative Hfq binding sequence reduces the apparent binding affinity Hfq. So if Hfq binds the transcript, why do we see no effect in the mating out assay when we mutate the 3' end? One possibility is that under the conditions used for the *in vitro* data favoured the binding of Hfq whereas as the mating out conditions did not permit Hfq binding. Hfq binds not only in a sequence specific manner but also in a secondary structure dependent manner. For example, in the case of Hfq dependent sRNAs the basic requirements for productive Hfq binding include a hairpin flanked by U rich regions [76]. One could envision conditions that may alter secondary structures of RNA molecules such as cold shock or extreme heat.

An example of this is in RNA thermometers where mRNA transcripts form self-inhibiting structures blocking their RBS, but when presented with specific environmental conditions they undergo a conformational change which permits translation (reviewed in [93]). It is possible that the conditions we tested for a phenotype in both the β-galactosidase assay and the mating out assay are not permissive for the appropriate secondary structure to form and thus no productive Hfq binding occurs (**See Figure 4.1)**. Another possibility remains that the function of the 3' end of the RNA molecule may act as an Hfq chelator under certain conditions, as a method of the element to modulate cellular Hfq availability and thus increasing the amount of RNA-IN transcript available in the cell and thus more transposition under times of stress (**See Figure 4.1).** Future directions should be to identify conditions that alter the IS*10* transposition frequency of Hfq^+ cells but not Hfq^- cells. Such a condition may have inadvertently been identified already with the room temperature data (discussed further below).

4.5 Is IS*10* transposition regulated by an sRNA

In this study I utilized another bioinformatics approach (INTARNA) to define sRNAs that may regulate IS*10* transposase expression. From the list of potential sRNA-IS*10* mRNA interactions, I studied 3 sRNAs: GlmZ, SroG and RdlC. *In vitro* binding studies with HFQ revealed that GlmZ was the only of the 3 sRNAs that pairs with the 5' end of IS*10* transposase mRNA. However, overexpression of GlmZ did not have an impact on IS*10* transposition frequencies

in a mating out assay. These results suggest that the weak *in vitro* interaction between IS*10* and GlmZ is not biologically relevant.

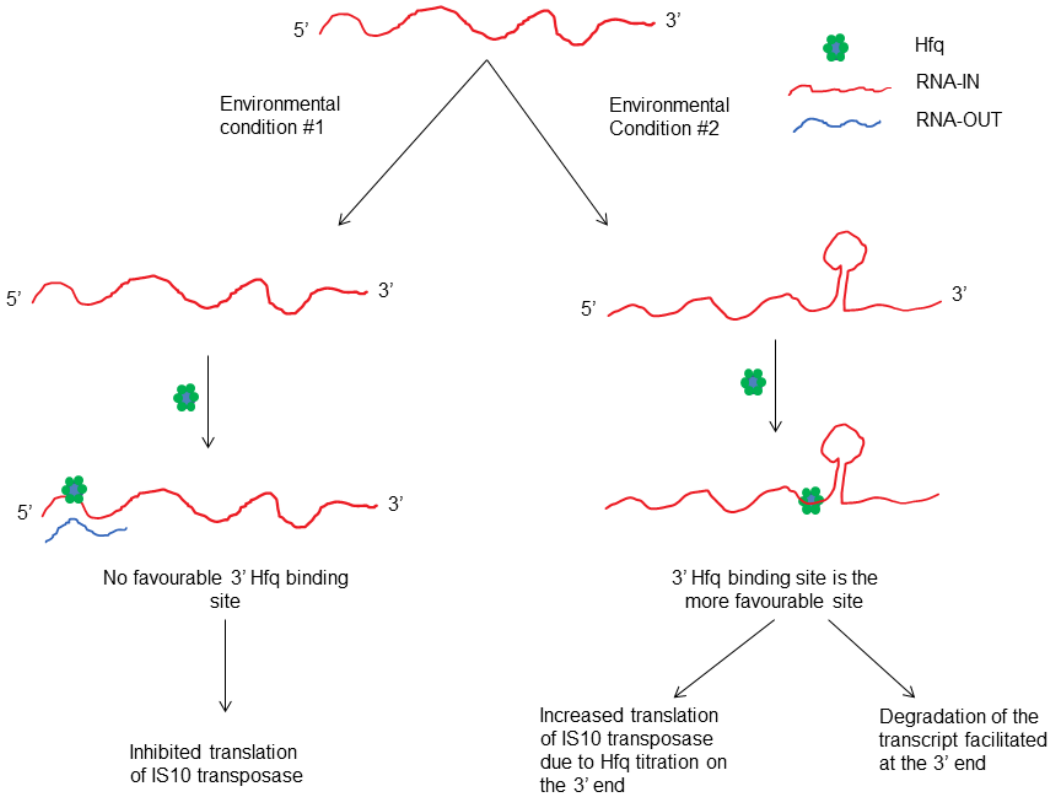


Figure 4.1 Model for potential conditions that may change the ability for Hfq to bind a transcript. (Left) Growth of *E. coli* at our standard lab conditions (37°C, with aeration) is not permissive to 3' RNA-IN Hfq regulation, resulting in Hfq being available to promote base pairing of RNA-IN and RNA-OUT, resulting in inhibited IS*10* transposase translation and subsequence degradation of the mRNA transcript. **(Right)** In growth conditions that alter the folding of the RNA-IN transcript, Hfq may be able to bind the 3' region resulting in 2 possible mechanisms. The first possibility is that Hfq may get titrated away from the 5' binding site, resulting in reduced RNA-IN/RNA-OUT pairing and ultimately increased translation. The second possibility is that Hfq may facilitate degradation under certain growth conditions either through an expressed sRNA specific to those conditions, or a general nuclease mediated degradation (such as PNPase).

4.6 Does modulation of Hfq levels have an impact on IS*10* transposition?

Ross *et al* suggest a mechanism of Hfq regulation of IS*10* where stress environments promote the overexpression of sRNAs which in turn occupy Hfq RNA binding sites. This 'chelating' effect on Hfq may result in less Hfq available in the cell to promote the pairing of RNA-OUT and RNA-IN, resulting in a reduction of IS*10* transposase expression [18]. The initial inducer of sRNA expression is often a change in environmental conditions, and the bacterial cell can modulate gene expression through the expression of sRNAs. Therefore, under certain stresses, Hfq binding sRNAs can be expressed which bind Hfq, de-repressing the Hfq down regulation of IS*10* transposase expression and consequently increasing IS*10* transposition frequencies.

The overexpression of MicC in our wildtype (Hfq^+) mating out *E.coli* strain resulted in an increase of in IS*10* transposition of 1.7 fold compared to the uninduced and vector only controls. MicC has already been shown in prior studies ([53],[54]), to alter the expression of genes that are subject to Hfq repression or activation. Furthermore, I have identified a growth condition where forcing cells to incubate for prolonged periods in stationary phase that results in a 10 fold increase in IS*10* transposition frequencies in wildtype cells, compared to normal growth conditions but not in an *rpoS* gene disruption strain. This is the first study done that links RpoS to a decrease in IS*10* transposition frequencies when cells are in stationary phase.

The observation that MicC overexpression results in an increase of 10 fold in IS*10* transposition frequencies suggests a possible link between cellular stress

and IS*10* transposition. For example, MicC sRNA expression is induced when *E.coli* cells are exposed to cellular stresses such as low temperatures (24°C), osmotic shock, the herbicide Paraquat, and when grown in conditions where glycerol is the only carbohydrate source [37]. If these cellular stresses induce MicC expression, and MicC expression results in an increase in IS*10* transposition, is it possible that these cellular stresses result in increased IS*10* transposition? Follow up experiments on the effect of sRNA overexpression on IS*10* transposition studies should focus on these stresses that may increase IS*10* transposition frequencies. These studies will help us understand the role transposition of transposable elements as a possible mechanism for cells to adapt to cellular stress.

Previous studies have already been done on the role of IS*10* and stress adaptation of the cell. For example, studies have been to on the effects of UV light induction of IS*10* transposition [94]. Another study was done on the osmotic stress adaption of *rpoS* deletion strains of *E.coli*, where the authors found that RpoS deletion strains consistently adapted to osmotic stress through IS*10* transposition in front of the *otsBA* operon (an RpoS dependent operon)[95]. This study is the first one to directly link RpoS as a regulator of IS*10* transposition in terms of transposition frequencies in stationary phase. It is also the first study to link the sRNA MicC to IS*10* transposition.

Follow up experiments to test the how Hfq levels impact IS*10* transposition rates should include using microarray data to determine how gene expression profiles change in wild type and mutants with the *rpos* gene disruption in the two growth conditions we analyzed in this work. Using these expression profiles and

through gene deletion strains of *E.coli*, we could focus on specifically how the *rpoS* mutant is down regulating IS*10* transposition. It would be interesting to see the effects of our stationary phase growth conditions on known regulators of IS*10* transposition such as H-NS (a positive regulator of IS*10* transposition) and Integrative Host Factor (IHF) (a host factor that promotes end pairing of the IS*10* transpososome.). The effect of stationary phase and IHF and H-NS have already been studied by Azam *et al.* where the abundance of IHF actually increase as well as the abundance of H-NS actually decreased by 40% at stationary phase. Follow up experiments to how growth conditions can affect IS*10* transposition frequencies should include the characterization of the minimal amount of time of cell growth required to observe the increase of IS*10* transposition frequencies in wildtype cells. Furthermore, an experiment testing different conditions should be conducted so we can really understand the specific factors that contribute to the increase in IS*10* transposition, for example was it the lower temperature that resulted in the increase of transposition frequencies or was it the stationary phase? The most interesting follow up experiments should really characterize the genetic landscape of the *rpoS* disruption strain and the wildtype strain under the stationary phase growth conditions. It is intriguing as to why the wildtype strain has an increase in IS*10* transposition yet the *rpoS* deletion strain maintains the same IS*10* transposition frequency when grown into exponential phase. Further experiments are required to elucidate just how RpoS is down regulating IS*10* under stationary phase growth conditions and whether or not this is a direct effect or due to the induction of an sRNA or the up regulation of a stationary phase response gene.

This work has further explored the how host factors in bacteria regulate the expression of mobile genetic elements. In this work I have provided another example of an Hfq repressed IS Element (IS*1413*) as well as a growth condition that increases IS*10* transposition elements in the presence of RpoS but not in the absence. These findings provide a greater glimpse into factors that affect the mobility and expression of IS Elements in bacterial populations. By understanding what permits IS Element mobility and conditions that promote IS Element mobility, we are one step closer to understanding how we can control the movement of mobile genes in populations. If we eventually can control the movement of mobile genetic elements in populations, we can use these mobile genes for the benefit of the environment through bioremediation (such as in the case of IS*1413*) or to prevent the rampant spread of antibiotic resistance genes through bacterial populations in our health care facilities.

Appendix 1- Obtaining a compiled list of IS Elements from the IS Finder database.

In order to search the IS Finder database for sequences that contained putative Hfq binding sites, the sequences had to be compiled into a series of files that could be sorted and searched. In their native state, the sequences had to be accessed one by one on the IS Finder database; searching the 3478 active links in the database one by one was not a viable option. Firstly, a list of urls of each element entry had to be compiled. This was achieved by searching for each IS Family number one by one (IS*1*, IS*3*, IS*4*, IS*5*, IS*6*, IS*21*, IS*30*, IS*66*, IS*91*, IS*110*, IS*200*, IS*605*, IS*256*, IS*481*, IS*630*, IS*982*, IS*1380*, IS*As1*, IS*L3*, Tn*3*, IS*1595*, IS*1182*, IS*1634*, IS*H3*, IS*701*) and copying the IS Element names into a Microsoft Excel 2010 worksheet in column A. All IS Finder urls have a common address of "http://www-is.biotoul.fr/index.html?is_special_name=" so by using the concatenate function of MS Excel 2010, one can paste the following function in column B to obtain a list of URLs to be searched:

"CONCATENATE("http://wwwis.biotoul.fr/index.html?is_special_name=",A 1)"

This list of URLs was then pasted into WEBSPHINX[77]. Using the concatenate function within WEBSPHINX[77], all the information about the elements within each family was merged into one text file (.txt), which was then parsed as described below.

Script A1.1 contains the find and replace script I used to parse the IS Finder Data base into a tab deliminated document. The scripting language is Microsoft Visual Basic for Applications. Extra text was removed from the compiled IS Finder database to conserve computer resources and the following fields remained: Element, Host, Transposition Determined, IS_Sequence, Comments, and References. The resulting document allowed for an Excel 2010 workbook to be made, so all elements within each IS-Family could be searched for the putative Hfq binding sites AAYAAYAA and AYAATAA, using Microsoft Excel's built in find function. Due to the variability of IS Elements in GC content, a probability (depicted as a P-Value in Tables 3.1 and 3.2) was also calculated of how likely it was to find the number of SELEX sites I did, taking into account the length and GC content of each element. For example, if an element is very long and has an abundance of adenine and thymine bases, one would expect to find more SELEX sites compared to an IS Element with a very high GC content. This was accomplished using a cumulative binomial distribution which gives the probability of finding exactly 1 SELEX site in IS*10* for example. Outlined below are two calculations for the IS*10* Element for both SELEX sites AAYAAYAA (**Calculation A1.1**) and AYAATAA (**Calculation A1.2**).

Calculation A1.1: The calculation of the probability of finding exactly 1 AAYAAYAA SELEX site in the IS*10* Insertion Sequence transposase mRNA.

Let p represent probability
Sequence = AAYAAYAA
Length = 1329 bp
Number of potential sequences with a stretch of 8 nucleotides = 1329-7
Number of trials = 1322
GC Content = 44%
p of G = 0.22 (or 22%)
p of C = 0.22 (or 22%)
p of A = 0.28 (or 28%)
p of T = 0.28 (or 28%)

$$pAAYAAYAA = pA * pA * p(C\ or\ T) * pA * pA * p(C\ or\ T) * pA * pA$$

pAAYAAYAA= 0.0001205

In this case of using the binomial distribution a success would be the SELEX site is present.

The binomial distribution is represented by:

$$p(x) = \frac{N!}{x!\,(N-x)!} p^x (1-p)^{N-x}$$

For IS*10* we found 1 SELEX site over 1329bp so:

x = 1

N = 1322

$$p(1) = \frac{1322!}{1!\,(1322-1)!} 0.0001205^1 (1-0.0001205)^{1322-1}$$

$$p(1) = 1322(0.0001205)(0.9998795)^{1321}$$

$$p(1) = 0.14$$

Calculation A1.2: The calculation of the probability of finding exactly 1

AYAATAA SELEX site in the IS*10* Insertion Sequence transposase mRNA.

Let P represent probability
Sequence = AYAATAA
Length = 1329 bp
Number of potential sequences with a stretch of 8 nucleotides = 1329-7

Number of trials = 1322
GC Content = 44%
P of G = 0.22 (or 22%)
P of C = 0.22 (or 22%)
P of A = 0.28 (or 28%)
P of T = 0.28 (or 28%)

$$pAYAATAA = pA * \ p(C\ or\ T) * pA * pA * pT * pA * pA$$

pAYAATAA= 0.00024

In this case of using the binomial distribution a success would be the SELEX site is present.

The binomial distribution is represented by:

$$p(x) = \frac{N!}{x!\,(N-x)!} p^x (1-p)^{N-x}$$

For IS*10* we found 1 SELEX site over 1329bp (1322 trials) so:

x = 1

N = 1322

$$p(1) = \frac{1322!}{1!\,(1322-1)!} 0.00024^1 (1-0.00024)^{1322-1}$$

$$p(1) = 1322(0.00024)(0.99976)^{1321}$$

$$pr(1) = 0.23$$

Script A1.1: Visual Basic for Applications macro code used to parse IS Finder files for export into searchable files.

```
Sub IS_PARSER()

'Start the script at the very top of the document
Selection.HomeKey Unit:=wdStory

' allows Wild card to recognize ^p as $
'Find within the selected text.  In this case because no text is selected, the
whole document is searched.
 With Selection.Find
 ' What string are you looking for
  .Text = "^p"
 ' What would you like to replace it with?
  .Replacement.Text = " $^p"
 ' Search forward in the document
  .Forward = True
 ' Ensures the whole document is searched
  .Wrap = wdFindContinue
 ' Wildcards are disabled
  .MatchWildcards = False
' Every instance of the word is replaced
  .Execute Replace:=wdReplaceAll
 End With
 ' allows Wild card to recognize ^v as $
 With Selection.Find
  .Text = "^v"
  .Replacement.Text = " $^v"
  .Forward = True
  .Wrap = wdFindContinue
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

 'Remove the First line of Each Entry
 With Selection.Find
  .Text = "Result of your query $^p"
  .Replacement.Text = ""
  .Forward = True
  .Wrap = wdFindContinue
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

 'remove tabs to deliminate wanted fields
  With Selection.Find
  .Text = "^t"
  .Replacement.Text = ""
  .Forward = True
  .Wrap = wdFindContinue
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With
```

```
 'Remove the First line of Each Entry
 With Selection.Find
  .Text = "Result of your query $^p"
  .Replacement.Text = ""
  .Forward = True
  .Wrap = wdFindContinue
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

 'Removes all text between "Synonyms" and "Hosts" and replaces it with
"Hosts:"
  With Selection.Find
  .Text = "SynonymsIsoFamilyGroup*OriginHosts"
  .Replacement.Text = "Hosts:"
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters enabled
  .MatchWildcards = True
  .Execute Replace:=wdReplaceAll
 End With

 'Put a tab character in front of "Hosts:"
  With Selection.Find
  .Text = "Hosts:"
  .Replacement.Text = "^tHosts:"
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Disabled again
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

'Enable Wildcard, removes text between Accension and Transposition
  With Selection.Find
  .Text = "Accession NumberLengthIRDR*Transposition:"
  .Replacement.Text = "Transposition:"
  .Forward = True
  .Wrap = wdFindContinue
  .MatchWildcards = True
  .Execute Replace:=wdReplaceAll
 End With

'Disable Wildcard, Tab deliminate Transposition field
  With Selection.Find
  .Text = "Transposition:"
  .Replacement.Text = "^t Transposition"
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Disabled again
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With
```

```
'Delete "*" make sure wild cards are disabled, or you will lose all your text
  With Selection.Find
  .Text = "*"
  .Replacement.Text = ""
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Disabled again
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

'Delete Sequence BTW ORF and IS_SEQ; Enable Wild Card,
  With Selection.Find
  .Text = "ORF:*IS_SEQ:"
  .Replacement.Text = "IS_SEQ:"
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Enabled again
  .MatchWildcards = True
  .Execute Replace:=wdReplaceAll
 End With

'Tab Deliminate IS_SEQ field
  With Selection.Find
  .Text = "IS_SEQ:"
  .Replacement.Text = " ^t IS_SEQ:"
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Disabled again
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

'Delete the text between IS_PEP: and Comments
  With Selection.Find
  .Text = "IS_PEP:*Comments:"
  .Replacement.Text = "Comments"
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Enabled again
  .MatchWildcards = True
  .Execute Replace:=wdReplaceAll
 End With

'Tab deliminates Comments:
  With Selection.Find
  .Text = "Comments"
  .Replacement.Text = "^t Comments:"
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Enabled again
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With
```

```
' Remove all paragraph breaks #This will take awhile
  With Selection.Find
  .Text = "^p"
  .Replacement.Text = ""
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Disabled again
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

 ' Separates each entry as a new paragraph
  With Selection.Find
  .Text = "Page ^#"
  .Replacement.Text = "^p Page 1"
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Disabled again
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

' Tab Deliminates References Section
  With Selection.Find
  .Text = "References:"
  .Replacement.Text = "^t References:"
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Disabled again
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

' Delete $ place holders
  With Selection.Find
  .Text = "$"
  .Replacement.Text = ""
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Disabled again
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

' Delete webpage breaks
  With Selection.Find
  .Text = "-"
  .Replacement.Text = ""
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Disabled again
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With
```

```
 'Replace Page # with IS_ELEMENT Tab delimited

 With Selection.Find
  .Text = "Page ^#^#^#"
  .Replacement.Text = "^t IS_ELEMENT: "
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Disabled again
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

 With Selection.Find
  .Text = "Page ^#^#"
  .Replacement.Text = "^t IS_ELEMENT: "
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Disabled again
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

 With Selection.Find
  .Text = "Page ^#"
  .Replacement.Text = "^t IS_ELEMENT: "
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Disabled again
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

'Replace IS Finder URL with ISELEMENT name
With Selection.Find
  .Text = "<[http://wwwis.biotoul.fr/is/scripts/is_spec.php?name="
  .Replacement.Text = "^p["
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Disabled again
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

 'Replace Delete double IS_Element
With Selection.Find
  .Text = "^p"
  .Replacement.Text = "$"
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Disabled again
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

With Selection.Find
```

```
  .Text = "$*$"
  .Replacement.Text = "^p"
  .Forward = True
  .Wrap = wdFindContinue
  'Wildcard characters Enabled again
  .MatchWildcards = True
  .Execute Replace:=wdReplaceAll
 End With

 'Remove Field Names for proper Export as a Table
 With Selection.Find
  .Text = " IS_ELEMENT:  "
  .Replacement.Text = ""
  .Forward = True
  .Wrap = wdFindContinue
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

With Selection.Find
  .Text = "Hosts: "
  .Replacement.Text = ""
  .Forward = True
  .Wrap = wdFindContinue
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

 With Selection.Find
  .Text = " IS_SEQ: "
  .Replacement.Text = ""
  .Forward = True
  .Wrap = wdFindContinue
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

With Selection.Find
  .Text = " Comments: "
  .Replacement.Text = ""
  .Forward = True
  .Wrap = wdFindContinue
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

 With Selection.Find
  .Text = " References: "
  .Replacement.Text = ""
  .Forward = True
  .Wrap = wdFindContinue
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With
```

```
 With Selection.Find
  .Text = " Transposition "
  .Replacement.Text = ""
  .Forward = True
  .Wrap = wdFindContinue
  .MatchWildcards = False
  .Execute Replace:=wdReplaceAll
 End With

'Move to the top of the document and Title Fields for Excel export

 Selection.HomeKey Unit:=wdStory
  Selection.TypeText (vbTab & "IS_Element" & vbTab & "Host" & vbTab &
"Transposase Tested" & vbTab & "IS_SEQ" & vbTab & "Comments" & vbTab &
"References" & vbCrLf)


 End Sub
```

References

1. Ochman, H., Lawrence, J. G. & Groisman, E. a Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).

2. Lawrence, J. G. & Ochman, H. Molecular archaeology of the Escherichia coli genome. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 9413–7 (1998).

3. Buchanan-Wollaston, V., Passiatore, J. E. & Cannon, F. The mob and oriT mobilization functions of a bacterial plasmid promote its transfer to plants. *Nature* **327**, 617–618 (1987).

4. Heinemann, J. A. & Sprague Jr, G. F. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature* **340**, 205–209 (1989).

5. Davies, J. Inactivation of antibiotics and the dissemination of resistance genes. *Science (New York, N.Y.)* **264**, 375–82 (1994).

6. Ross, D. G., Swan, J. & Kleckner, N. Physical structures of Tn10-promoted deletions and inversions: role of 1400 bp inverted repetitions. *Cell* **16**, 721–31 (1979).

7. Wang, A. & Roth, J. R. Activation of silent genes by transposons Tn5 and Tn10. *Genetics* **120**, 875–885 (1988).

8. Burrus, V., Pavlovic, G. & Decaris, B. Conjugative transposons: the tip of the iceberg. *Molecular* **46**, 601–610 (2002).

9. Lessie, T. G., Hendrickson, W., Manning, B. D. & Devereux, R. Genomic complexity and plasticity of Burkholderia cepacia. *FEMS microbiology letters* **144**, 117–28 (1996).

10. Shoemaker, N. Tetracycline-dependent appearance of plasmidlike forms in Bacteroides uniformis 0061 mediated by conjugal Bacteroides tetracycline resistance elements. *Journal of bacteriology* **170**, 1651–7 (1988).

11. Kleckner, N. Mobile DNA. *Cell* **60**, 227–268 (1989).

12. Shen, M. & Raleigh, E. Physical analysis of Tn10-and IS10-promoted transpositions and rearrangements. *Genetics* **369**, 359–369 (1987).

13. Morisato, D., Way, J. C., Kim, H. J. & Kleckner, N. Tn10 transposase acts preferentially on nearby transposon ends in vivo. *Cell* **32**, 799–807 (1983).

14. Kleckner, N. Regulation of transposition in bacteria. *Annu. Rev. Cell Bioi. 1990.* **6**, 297–327 (1990).

15. Roberts, D., Hoopes, B. C., McClure, W. R. & Kleckner, N. IS10 transposition is regulated by DNA adenine methylation. *Cell* **43**, 117–30 (1985).

16. Haniford, D. B. Transpososome dynamics and regulation in Tn10 transposition. *Critical reviews in biochemistry and molecular biology* **41**, 407–24 (2006).

17. Sakai, J., Chalmers, R. M. & Kleckner, N. Identification and characterization of a pre-cleavage synaptic complex that is an early intermediate in Tn 10 transposition end fragments standard (S) long (L). *EMBO Journal* **14**, 4374–4383 (1995).

18. Ross, J., Wardle, S. J. & Haniford, D. B. Tn10/IS10 transposition is downregulated at the level of transposase expression by the RNA-binding protein Hfq. *Molecular microbiology* **78**, 607–21 (2010).

19. Beich-Frandsen, M., Večerek, B., Sjöblom, B., Bläsi, U. & Djinović-Carugo, K. Structural analysis of full-length Hfq from Escherichia coli. *Acta crystallographica. Section F, Structural biology and crystallization communications* **67**, 536–40 (2011).

20. Møller, T. *et al.* Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Molecular cell* **9**, 23–30 (2002).

21. Séraphin, B. 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. *The EMBO journal* **14**, 2089–98 (1995).

22. Gottesman, S. & Storz, G. Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations. *Cold Spring Harbor perspectives in biology* (2010).doi:10.1101/cshperspect.a003798

23. Mikulecky, P. J. *et al.* Escherichia coli Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nature structural & molecular biology* **11**, 1206–14 (2004).

24. Zhang, A., Wassarman, K. M., Ortega, J., Steven, A. C. & Storz, G. The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Molecular cell* **9**, 11–22 (2002).

25. Soper, T. J. & Woodson, S. a The rpoS mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. *RNA (New York, N.Y.)* **14**, 1907–17 (2008).

26. Yarchuk, O., Jacques, N. & Guillerez, J. Interdependence of translation, transcription and mRNA degradation in the lacZ gene. *Journal of molecular* 581–596 (1992).at <http://www.sciencedirect.com/science/article/pii/002228369290617S>

27. Vytvytska, O., Moll, I. & Kaberdin, V. Hfq (HF1) stimulates ompA mRNA decay by interfering with ribosome binding. *Genes &* **14**, 1109–1118 (2000).

28. Isabella Moll, Taras Afonyushkin, Oresta Vytvytska, V. R. K. Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs. *Rna* **9**, 1308–1314 (2003).

29. Massé, E., Escorcia, F. E. & Gottesman, S. Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. *Genes & development* **17**, 2374–83 (2003).

30. Mohanty, B. K., Maples, V. F. & Kushner, S. R. The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in Escherichia coli. *Molecular microbiology* **54**, 905–20 (2004).

31. O'Hara, E. B. *et al.* Polyadenylylation helps regulate mRNA decay in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 1807–11 (1995).

32. Thomason, M. K. & Storz, G. Bacterial antisense RNAs: how many are there, and what are they doing? *Annual review of genetics* **44**, 167–88 (2010).

33. Sorek, R., Kunin, V. & Hugenholtz, P. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature reviews. Microbiology* **6**, 181–6 (2008).

34. Simons, R. W. & Kleckner, N. Translational control of IS10 transposition. *Cell* **34**, 683–91 (1983).

35. Udekwu, K. I. Transcriptional and post-transcriptional regulation of the Escherichia coli luxS mRNA; involvement of the sRNA MicA. *PloS one* **5**, e13449 (2010).

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

37. Chen, S., Zhang, A., Blyn, L. B. & Storz, G. MicC , a Second Small-RNA Regulator of Omp Protein Expression in Escherichia coli. *Society* **186**, 6689–6697 (2004).

38. Vogel, J. & Luisi, B. F. Hfq and its constellation of RNA. *Nature reviews. Microbiology* **9**, 578–89 (2011).

39. Salim, N. N., Faner, M. a, Philip, J. a & Feig, A. L. Requirement of upstream Hfq-binding (ARN)x elements in glmS and the Hfq C-terminal region for GlmS upregulation by sRNAs GlmZ and GlmY{dagger}. *Nucleic acids research* 1–12 (2012).doi:10.1093/nar/gks392

40. Brennan, R. G. & Link, T. M. Hfq structure, function and ligand binding. *Current opinion in microbiology* **10**, 125–33 (2007).

41. Link, T. M., Valentin-Hansen, P. & Brennan, R. G. Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 19292–7 (2009).

42. Takada, a, Wachi, M., Kaidow, a, Takamura, M. & Nagai, K. DNA binding properties of the hfq gene product of Escherichia coli. *Biochemical and biophysical research communications* **236**, 576–9 (1997).

43. Panja, S. & Woodson, S. a Hexamer to Monomer Equilibrium of E. coli Hfq in Solution and Its Impact on RNA Annealing. *Journal of molecular biology* **417**, 406–412 (2012).

44. Lorenz, C. *et al.* Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts. *Nucleic acids research* **38**, 3794–808 (2010).

45. Tsui, H. & Feng, G. 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. *Journal of bacteriology* (1996).at <http://jb.asm.org/content/178/19/5719.short>

46. Tsui, H. T., Feng, G. & Winkler, M. E. regulators of Escherichia coli K-12. gene expression by the Hfq and RpoS global Negative regulation of mutS and mutH repair. **179**, (1997).

47. Vecerek, B., Moll, I. & Bläsi, U. Translational autocontrol of the Escherichia coli hfq RNA chaperone gene. *RNA (New York, N.Y.)* **11**, 976–84 (2005).

48. Carmichael, G., Weber, K. & Niveleau, A. The host factor required for RNA phage Qbeta RNA replication in vitro. Intracellular location, quantitation, and purification by polyadenylate-cellulose chromatography. *Journal of Biological* **250**, 3607–3612 (1975).

49. Ali Azam, T., Iwata, a, Nishimura, a, Ueda, S. & Ishihama, a Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *Journal of bacteriology* **181**, 6361–70 (1999).

50. Tsui, H.-C. T., Leung, H.-C. E. & Winkler, M. E. Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in Escherichia coli K-12. *Molecular Microbiology* **13**, 35–49 (1994).

51. McLellan, C. R. Regulation of the bacterial transposon Tn5. (2012).

52. Kajitani, M., Kato, a, Wada, a, Inokuchi, Y. & Ishihama, a Regulation of the Escherichia coli hfq gene encoding the host factor for phage Q beta. *Journal of bacteriology* **176**, 531–4 (1994).

53. Hussein, R. & Lim, H. N. Disruption of small RNA signaling caused by competition for Hfq. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 1110–5 (2011).

54. Moon, K. & Gottesman, S. Competition among Hfq-binding small RNAs in Escherichia coli. *Molecular Microbiology* **82**, 1545–1562 (2011).

55. Sharma, A. K. & Payne, S. M. Induction of expression of hfq by DksA is essential for Shigella flexneri virulence. *Molecular microbiology* **62**, 469–79 (2006).

56. Hübner, a & Hendrickson, W. A fusion promoter created by a new insertion sequence, IS1490, activates transcription of 2,4,5-trichlorophenoxyacetic acid catabolic genes in Burkholderia cepacia AC1100. *Journal of bacteriology* **179**, 2717–23 (1997).

57. Simons, R. W. DNA sequence organization of IS10-right of Tn10 and comparison with IS10-left. *Proc. Natl. Acad. Sci* **79**, 2608–2612 (1982).

58. Ma, Charles and Simons, R. W. The IS10 antisense RNA blocks ribosome binding at the transposase translation initiation site. *The EMBO journal* **9**, 1267–1274 (1990).

59. Case, C. & Simons, E. The IS10 transposase mRNA is destabilized during antisense RNA control. *The EMBO journal* **9**, 1259–1266 (1990).

60. Buchon, N. & Vaury, C. RNAi: a defensive RNA-silencing against viruses and transposable elements. *Heredity* **96**, 195–202 (2006).

61. Obbard, D. J., Gordon, K. H. J., Buck, A. H. & Jiggins, F. M. The evolution of RNAi as a defence against viruses and transposable elements. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **364**, 99–115 (2009).

62. Rosner, J. L. Formation, induction, and curing of bacteriophage P1 lysogens. *Virology* **48**, 679–689 (1972).

63. Foster, T. J., Davis, M. a, Roberts, D. E., Takeshita, K. & Kleckner, N. Genetic organization of transposon Tn10. *Cell* **23**, 201–13 (1981).

64. Bolivar, F. & Backman, K. Complementation analysis of the restriction and modification of DNA in Escherichia coli. *Journal of molecular biology* **68**, 245–267 (1979).

65. Bender, J. & Kuo, J. Genetic evidence against intramolecular rejoining of the donor DNA molecule following IS10 transposition. *Genetics* (1991).at <http://www.genetics.org/cgi/content/abstract/128/4/687>

66. Ely, B. Vectors for transposon mutagenesis of non-enteric bacteria. *Molecular & general genetics : MGG* **200**, 302–4 (1985).

67. Guillier, M. & Gottesman, S. Remodelling of the Escherichia coli outer membrane by two small regulatory RNAs. *Molecular microbiology* **59**, 231–47 (2006).

68. Coros, A. & Twiss, E. Genetic evidence that GTP is required for transposition of IS903 and Tn552 in Escherichia coli. *Journal of …* **187**, 4598–4606 (2005).

69. Huang, H.-Y. *et al.* sRNAMap: genomic maps for small non-coding RNAs, their regulators and their targets in microbial genomes. *Nucleic acids research* **37**, D150–4 (2009).

70. Gibb, E. a & Edgell, D. R. Multiple controls regulate the expression of mobE, an HNH homing endonuclease gene embedded within a ribonucleotide reductase gene of phage Aeh1. *Journal of bacteriology* **189**, 4648–61 (2007).

71. Celesnik, H., Deana, A., And & Belasco, Joel, G. Extraction of bacterial RNA by the hot phenol method. *RNA Turnover in Bacteria, Archaea and Organelles* 92 (2008).

72. Siguier, P., Perochon, J., Lestrade, L., Mahillon, J. & Chandler, M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic acids research* **34**, D32–6 (2006).

73. Schilling, D. & Gerischer, U. The Acinetobacter baylyi Hfq gene encodes a large protein with an unusual C terminus. *Journal of bacteriology* **191**, 5553–62 (2009).

74. Sonnleitner, E., Moll, I. & Bläsi, U. Functional replacement of the Escherichia coli hfq gene by the homologue of Pseudomonas aeruginosa. *Microbiology (Reading, England)* **148**, 883–91 (2002).

75. Chen, S. *et al.* A bioinformatics based approach to discover small RNA genes in the Escherichia coli genome. *Bio Systems* **65**, 157–77 (2002).

76. Ishikawa, H., Otaka, H., Maki, K., Morita, T. & Aiba, H. 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 (New York, N.Y.)* **18**, 1062–74 (2012).

77. Miller, R. & Bharat, K. SPHINX: a framework for creating personal, site-specific Web crawlers. *Computer Networks and ISDN Systems* **30**, 119–130 (1998).

78. Brown, L. & Elliott, T. Efficient translation of the RpoS sigma factor in Salmonella typhimurium requires host factor I, an RNA-binding protein encoded by the hfq gene. *Journal of bacteriology* **178**, 3763–3770 (1996).

79. Hajnsdorf, E. & Régnier, P. Host factor Hfq of Escherichia coli stimulates elongation of poly(A) tails by poly(A) polymerase I. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 1501–5 (2000).

80. Lisitsky, I. & Schuster, G. Preferential degradation of polyadenylated and polyuridinylated RNAs by the bacterial exoribonuclease polynucleotide phosphorylase. *European Journal of Biochemistry* **261**, 468–74 (1999).

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

82. Naville, M., Ghuillot-Gaudeffroy, A., Marchais, A. & Gautheret, D. ARNold: A web tool for the prediction of Rho-independent transcription terminators. *RNA Biology* **8**, 11–13 (2011).

83. Brescia, C. C. Identification of the Hfq-binding site on DsrA RNA: Hfq binds without altering DsrA secondary structure. *Rna* **9**, 33–43 (2003).

84. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**, 3406–3415 (2003).

85. Smith, C., Heyne, S., Richter, A. S., Will, S. & Backofen, R. Freiburg RNA Tools: a web server integrating INTARNA, EXPARNA and LOCARNA. *Nucleic acids research* **38**, W373–7 (2010).

86. Azam, T. A., Iwata, A., Nishimura, A., Azam, T. A. L. I. & Ueda, S. Growth Phase-Dependent Variation in Protein Composition of the Escherichia coli Nucleoid Growth Phase-Dependent Variation in Protein Composition of the Escherichia coli Nucleoid. *Society* (1999).

87. Sousa, S. a, Ramos, C. G., Moreira, L. M. & Leitão, J. H. The hfq gene is required for stress resistance and full virulence of Burkholderia cepacia to the nematode Caenorhabditis elegans. *Microbiology (Reading, England)* **156**, 896–908 (2010).

88. Waters, L. S. & Storz, G. Regulatory RNAs in bacteria. *Cell* **136**, 615–28 (2009).

89. Coenye, T. & Vandamme, P. Diversity and significance of Burkholderia species occupying diverse ecological niches. *Environmental Microbiology* **5**, 719–729 (2003).

90. Savageau, M. Escherichia coli habitats, cell types, and molecular mechanisms of gene control. *American Naturalist* **122**, 732–744 (1983).

91. Hübner, A., Danganan, C., Xun, L., Chakrabarty, A. M. & Hendrickson, W. 2, 4, 5-trichlorophenoxyacetic acid metabolism in Burkholderia cepacia AC1100: characterization of the tftC and tftD genes and locations of the tft operons on multiple. *Applied and …* **64**, 2086–93 (1998).

92. Hankins, J. S., Denroche, H. & Mackie, G. a Interactions of the RNA-binding protein Hfq with cspA mRNA, encoding the major cold shock protein. *Journal of bacteriology* **192**, 2482–90 (2010).

93. Narberhaus, F., Waldminghaus, T. & Chowdhury, S. RNA thermometers. *FEMS microbiology reviews* **30**, 3–16 (2006).

94. Eichenbaum, Z. & Livneh, Z. UV light induces IS10 transposition in Escherichia coli. *Genetics* **149**, 1173–81 (1998).

95. Stoebel, D. M., Hokamp, K., Last, M. S. & Dorman, C. J. Compensatory evolution of gene regulation in response to stress by Escherichia coli lacking RpoS. *PLoS genetics* **5**, e1000671 (2009).

Curriculum Vitae – Brian Munshaw

Education

M.Sc 2010 – 2012 – University of Western Ontario – Biochemistry

B.Ed: 2008-2009 University of Western Ontario – Intermediate/Senior Chemistry and Biology

B.MSc: 2003-2008 University of Western Ontario – Honours Microbiology and Immunology

Awards

Ontario Graduate Scholarship

Western Graduate Scholarship